

Article

Minimum Residual Sum of Squares Estimation Method for High-Dimensional Partial Correlation Coefficient

Jingying Yang ¹, Guishu Bai ¹ and Mei Yan ^{2,*}¹ School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu 611731, China² School of Mathematics, Yunnan Normal University, Kunming 650500, China

* Correspondence: myanlx920@hotmail.com

Abstract: The partial correlation coefficient (Pcor) is a vital statistical tool employed across various scientific domains to decipher intricate relationships and reveal inherent mechanisms. However, existing methods for estimating Pcor often overlook its accurate calculation. In response, this paper introduces a minimum residual sum of squares Pcor estimation method (MRSS), a high-precision approach tailored for high-dimensional scenarios. Notably, the MRSS algorithm reduces the estimation bias encountered with positive Pcor. Through simulations on high-dimensional data, encompassing both sparse and non-sparse conditions, MRSS consistently mitigates the arithmetic bias for positive Pcors, surpassing other algorithms discussed. For instance, for large sample sizes ($n \geq 100$) with $Pcor > 0$, the MRSS algorithm reduces the MSE and RMSE by about 30–70% compared to other algorithms. The robustness and stability of the MRSS algorithm is demonstrated by the sensitivity analysis with variance and sparsity parameters. Stocks data in China's A-share market are employed to showcase the MRSS methodology's practicality.

Keywords: partial correlation; high-dimensional data; variable selection; MCP**MSC:** 62H20; 62J07

Citation: Yang, J.; Bai, G.; Yan, M. Minimum Residual Sum of Squares Estimation Method for High-Dimensional Partial Correlation Coefficient. *Mathematics* **2023**, *11*, 4311. <https://doi.org/10.3390/math11204311>

Academic Editor: Jin-Ting Zhang

Received: 20 September 2023

Revised: 10 October 2023

Accepted: 13 October 2023

Published: 16 October 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The partial correlation coefficient (Pcor) measures the correlation between two random variables, X and Y , after accounting for the effects of controlling variables Z , denoted by $\rho_{XY|Z}$. The Pcor essentially quantifies the unique relationship between X and Y , after removing the correlations between X and Z , and between Y and Z [1]. This correlation coefficient provides a more thorough comprehension of the connection between variables, untainted by the influence of confounding factors. Unlike the Pearson correlation coefficient, which only captures the direct correlation between random variables, the Pcor enables the identification of whether correlations stem from intermediary variables. This distinction enhances the precision and validity of statistical analyses.

The Pcor is a fundamental statistical tool for investigating intricate relationships and gaining a more profound comprehension of the underlying mechanisms in a variety of scientific fields, such as psychology, biology, economics, and social sciences. When examining genetic markers and illness outcomes, biologists used the Pcor to identify correlations while accounting for potential confounding factors [2–4]. Marrelec et al. utilised the partial correlation matrix to explore large-scale functional brain networks through functional MRI [5]. In the field of economics, Pcor assists in comprehending complex connections, including the interplay between interest rates and inflation, while considering other variables' influence [6]. The financial industry also employs Pcor to interpret connections and relationships between stocks in the financial markets [7,8]. For example, Michis proposed a wavelet procedure for estimating Pcor between stock market returns over different time scales and implemented it for portfolio diversification [9]. Using

partial correlations within a complex network framework, Singh et al. examined the degree of globalisation and regionalisation of stock market linkages and how these linkages vary across different economic or market cycles [10]. Meanwhile, the employment of the Gaussian graphical model (GGM) technique in psychology has recently gained popularity for defining the relationships between observed variables. This technique employs Pcors to represent pairwise interdependencies, controlling the influence of all other variables [11–13]. In the field of geography, a correlation analysis based on the Pcor of the fractal dimension of the variations of HZD components is implemented to study the geomagnetic field component variations in Russian [14].

Several methodologies have been proposed over the years to estimate the Pcor in statistical analyses. For instance, Peng et al. introduced a Pcor estimation technique that relies on the sparsity property of the partial correlation matrix and utilises sparse regression methods [3]. Khare et al. suggested a high-dimensional graphical model selection approach based on the use of pseudolikelihood [15]. Kim provided an R package “ppcor” for a fast calculation to semi-Pcor [16]. Huang et al. introduced the kernel partial correlation coefficient as a measure of the conditional dependence between two random variables in various topological spaces [17]. Van Aert and Goos focused on calculating the sampling variance of Pcor [18]. Hu and Qiu proposed a statistical inference procedure for Pcor under the high-dimensional nonparanormal model [19]. However, these methods mainly centre around determining whether or not the partial correlation coefficient is zero, without adequate regard for the precision of the Pcor calculation and the algorithm’s efficacy. We analysed multiple high-dimensional algorithms and discovered notable Pcor estimation biases, particularly for positive Pcor. Even with larger sample sizes, these biases persisted. Motivated by these findings, our primary goal is to put forward a Pcor estimation algorithm to increase the precision of the Pcor estimation algorithm and diminish the estimation bias for positive Pcor values.

This paper reviews current methods for estimating Pcor in high-dimensional data. We introduce a novel minimum residual sum of squares (MRSS) Pcor estimation method under high-dimensional conditions, aiming to mitigate the estimation bias for positive Pcor. The algorithm’s effectiveness is validated through simulation studies under sparse and non-sparse conditions and real data analysis on stock markets.

The sections are structured as follows: Section 2 outlines definitions and corresponding formulae for calculating Pcor, and examines common algorithms for estimating Pcor. Section 3 presents our Minimum Residual Sum of Squares Pcor estimation, designed to mitigate estimation bias for positive Pcor. In Section 4, we demonstrate the effectiveness of our proposed algorithm through simulation studies on high-dimensional data under both sparse and non-sparse conditions. Section 5 provides an analysis of real data related to stock markets, while Section 6 contains the conclusion.

2. Estimation for Partial Correlation Coefficient

2.1. Definition of Pcor

The classical definition of the partial correlation coefficient is defined by the correlation coefficient between the regression residuals from the linear models of two variables with the controlling variable, respectively. Let X and Y be two random variables, and $Z = [Z_1, Z_2, \dots, Z_p]$ be p -dimensional controlling variables. Consider the linear regression models of X and Y , respectively, with the controlling variable Z ,

$$\begin{aligned} X &= \alpha_0 + \sum_{i=1}^p \alpha_i Z_i + \varepsilon, \\ Y &= \beta_0 + \sum_{i=1}^p \beta_i Z_i + \zeta, \end{aligned}$$

where ε and ζ are error terms. The partial correlation coefficient between X and Y is conditional on Z , denoted by $\rho_{XY|Z}$, and defined by the correlation coefficient between the regression residuals ε and ζ , as follows

$$\rho_{XY|Z} = \text{cor}(\varepsilon, \zeta) = \frac{\text{cov}(\varepsilon, \zeta)}{\sqrt{\text{var}(\varepsilon)}\sqrt{\text{var}(\zeta)}}. \quad (1)$$

where $\text{cor}(\cdot, \cdot)$ is the correlation coefficient of two random variables; $\text{cov}(\cdot, \cdot)$ is the covariance of two random variables; $\text{var}(\cdot)$ is the variance of a random variable. Let the sample size be n . In conventional low-dimensional cases ($p < n$), the ordinary least squares (OLS) is used to compute the residuals ε and ζ . Subsequently, the Pcor is computed from the correlation coefficient of residuals. However, the OLS method is not practical for high-dimensional cases ($p > n$). Regularisation methods are introduced to deal with such cases later.

2.2. Calculation Formulae of Pcor

2.2.1. Based on Concentration Matrix

The concentration matrix can also be used to calculate Pcor. Let $U_1 = [X, Y]$, $U_2 = Z = [Z_1, \dots, Z_p]$, $U = [U_1, U_2] = [X, Y, Z]$ and $\Sigma = \text{cov}(U)$ be the covariance matrix. When assuming that Σ is a non-singular matrix, the concentration matrix is denoted as $\Omega = (\omega^{ij})_{i,j=1}^{p+2} = \Sigma^{-1}$. Consider the following linear regression,

$$U_1 = U_2 b + e,$$

where $b = (b_1, b_2)$ is the regression coefficient and $e = (e_1, e_2)$ is the regression error. We have

$$\hat{e} = (\hat{e}_1, \hat{e}_2) = U_1 - U_2 \hat{b} = U_1 - \hat{U}_1,$$

where $\hat{b}$ is the estimator of b and $\hat{U}_1$ is the estimator of U_1 . The regression residual $\hat{e} \sim N(0, V)$ is independent of $\hat{U}_1$. The covariance matrix of $\hat{e}$ can be computed by

$$\begin{aligned} \text{cov}(\hat{e}) &= \text{cov}(U_1) + \text{cov}(\hat{U}_1) - 2\text{cov}(U_1, \hat{U}_1) \\ &= \Sigma_{11} + \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{22}\Sigma_{22}^{-1}\Sigma_{21} - 2\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \\ &= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} = \Omega_{11}^{-1}. \end{aligned}$$

According to the definition in Equation (1) and $\Omega_{11}^{-1} = \frac{1}{\omega^{11}\omega^{22} - \omega^{12}\omega^{21}} \begin{bmatrix} \omega^{22} & -\omega^{12} \\ -\omega^{21} & \omega^{11} \end{bmatrix}$, the partial correlation coefficient can be computed by

$$\rho_{XY|Z} = \text{cor}(\hat{e}_1, \hat{e}_2) = \frac{\text{cov}(\hat{e}_1, \hat{e}_2)}{\sqrt{\text{var}(\hat{e}_1)}\sqrt{\text{var}(\hat{e}_2)}} = -\frac{\omega^{12}}{\sqrt{\omega^{11}}\sqrt{\omega^{22}}}. \quad (2)$$

2.2.2. Based on Additional Regression Models

Additional linear regression models are introduced to calculate the Pcor. Consider new linear regression models of X with $[Y, Z]$ and Y with $[X, Z]$, respectively,

$$X = \lambda_0 Y + \sum_{i=1}^p \lambda_i Z_i + \eta, \quad (3)$$

$$Y = \gamma_0 X + \sum_{i=1}^p \gamma_i Z_i + \tau, \quad (4)$$

where η and τ are regression error terms. Peng et al. [3] established the correlation between the aforementioned regression coefficients and Pcor, while verifying that formulas $\lambda_0 = \rho_{XY|Z}\sqrt{\frac{\omega^{22}}{\omega^{11}}}$, $\gamma_0 = \rho_{XY|Z}\sqrt{\frac{\omega^{11}}{\omega^{22}}}$, $\text{var}(\eta) = \frac{1}{\omega^{11}}$ and $\text{var}(\tau) = \frac{1}{\omega^{22}}$ hold. Then, we derive $\lambda_0\gamma_0 = \rho_{XY|Z}^2$. Thus, the partial correlation coefficient between X and Y can be calculated by the formula below,

$$\rho_{XY|Z} = \lambda_0 \sqrt{\text{var}(\tau) / \text{var}(\eta)}, \quad (5)$$

$$= \text{sign}(\lambda_0) \sqrt{\lambda_0 \gamma_0}, \quad (6)$$

where $\text{sign}(\cdot)$ is the sign function.

Consider linear regression models of Y with $[1, Z]$ and Y with $[1, X, Z]$, respectively,

$$Y = \beta_0 + \sum_{i=1}^p \beta_i Z_i + \zeta,$$

$$Y = \gamma_{-1} + \gamma_0 X + \sum_{i=1}^p \gamma_i Z_i + \tau,$$

where ζ and τ are error terms. The partial correlation coefficient can also be calculated as follows [20]

$$\rho_{XY|Z} = \text{sign}(\gamma_0) \sqrt{\frac{\text{var}(\zeta) - \text{var}(\tau)}{\text{var}(\zeta)}}. \quad (7)$$

Here, we present five distinct formulae, (1), (2), (5), (6), and (7), for calculating Pcor based on diverse regression models. Specific algorithms applicable to high-dimensional scenarios will be presented in the following section.

2.3. Regularisation Regression for High-Dimensional Cases

Suppose we have centralised samples $\{x_j, y_j, z_{j1}, \dots, z_{jp}\}_{j=1}^n$ i.i.d. observed from $[X, Y, Z]$ with $Z = [Z_1, \dots, Z_p]$. Let $\mathbf{X} = [x_1, \dots, x_n]^T$, $\mathbf{Y} = [y_1, \dots, y_n]^T$ and $\mathbf{Z} = [Z_1, \dots, Z_p] = (z_{ji})_{n \times p}$. We consider matrix-type linear regression models as follows,

$$\mathbf{X} = \sum_{i=1}^p \alpha_i \mathbf{Z}_i + \varepsilon, \quad (8)$$

$$\mathbf{Y} = \sum_{i=1}^p \beta_i \mathbf{Z}_i + \zeta, \quad (9)$$

where $\varepsilon = [\varepsilon_1, \dots, \varepsilon_n]^T$ and $\zeta = [\zeta_1, \dots, \zeta_n]^T$ are error terms. If we estimate regression coefficients $\hat{\alpha} = [\hat{\alpha}_1, \dots, \hat{\alpha}_p]^T$ and $\hat{\beta} = [\hat{\beta}_1, \dots, \hat{\beta}_p]^T$, then we can calculate the estimated residuals $\hat{\varepsilon} = \mathbf{X} - \hat{\mathbf{X}}$ and $\hat{\zeta} = \mathbf{Y} - \hat{\mathbf{Y}}$, with $\hat{\mathbf{X}} = \mathbf{Z}\hat{\alpha} = \sum_{i=1}^p \hat{\alpha}_i \mathbf{Z}_i$ and $\hat{\mathbf{Y}} = \mathbf{Z}\hat{\beta} = \sum_{i=1}^p \hat{\beta}_i \mathbf{Z}_i$. According to the definition of Pcor, we can estimate the Pcor as follows

$$\hat{\rho} = \frac{\text{cov}(\hat{\varepsilon}, \hat{\zeta})}{\sqrt{\text{var}(\hat{\varepsilon})\text{var}(\hat{\zeta})}}, \quad (10)$$

where $\text{cov}(\hat{\varepsilon}, \hat{\zeta}) = \sum_{j=1}^n (\hat{\varepsilon}_j - \bar{\varepsilon})(\hat{\zeta}_j - \bar{\zeta})$, $\text{var}(\hat{\varepsilon}) = \sum_{j=1}^n (\hat{\varepsilon}_j - \bar{\varepsilon})^2$ and $\text{var}(\hat{\zeta}) = \sum_{j=1}^n (\hat{\zeta}_j - \bar{\zeta})^2$ with $\bar{\varepsilon} = \frac{1}{n} \sum_{j=1}^n \hat{\varepsilon}_j$, $\bar{\zeta} = \frac{1}{n} \sum_{j=1}^n \hat{\zeta}_j$.

In high-dimensional ($p > n$) situations, the penalty function and regularisation regression methods can be introduced to estimate the regression coefficients for regression models. Regularisation regression methods address overfitting in statistical modelling by adding a penalty to the loss function, constraining the coefficient magnitudes. Let $p_\lambda(\beta)$ be the penalty function with a tuning parameter λ , for example, the regularisation estimate of model (8) is given by

$$\hat{\alpha} = \arg \min_{\alpha} \frac{1}{n} \|\mathbf{X} - \mathbf{Z}\alpha\|^2 + p_\lambda(\alpha),$$

where the penalty $p_\lambda(\alpha)$ could widely choose the Lasso penalty [21], the Ridge penalty [22], the SCAD penalty [23], the Elastic net [24], the Fused lasso [25], the MCP penalty [26], and other penalty functions. In this paper, the Lasso regularisation with penalty as $p_\lambda(\alpha) = \lambda \|\alpha\|_1$ is implemented by the R-package “glmnet” [27], and the MCP with penalty as $p_\lambda(\alpha) = \frac{1}{t} (t\lambda - \alpha)_+$, ($t > 1$) is implemented by the R-package “ncvreg”.

2.4. Existing Pcor Estimation Algorithms

To investigate high-dimensional Pcor estimation methods, we present some existing methods that are suitable for both sparse and non-sparse conditions. Combining the advantages and disadvantages of these methods, we propose a new high-dimensional

Pcor estimation method: MRSS—minimum residual sum of squares partial correlation coefficient estimation algorithm.

2.4.1. Res Algorithm

The Res algorithm is primarily defined by the Pcor definition. This algorithm is implemented as follows. First, we use the regularisation regression (Lasso and MCP) on linear models (8) and (9) to obtain the estimated regression coefficients $\hat{\alpha}$ and $\hat{\beta}$; then calculate estimated residuals $\hat{\varepsilon} = \mathbf{X} - \hat{\mathbf{X}}$ and $\hat{\zeta} = \mathbf{Y} - \hat{\mathbf{Y}}$, with $\hat{\mathbf{X}} = \mathbf{Z}\hat{\alpha}$ and $\hat{\mathbf{Y}} = \mathbf{Z}\hat{\beta}$; at last, estimate Pcor $\hat{\rho}_{res}$ by formula (10).

2.4.2. Reg2 Algorithm

The Reg2 algorithm can more effectively remove the influence of $\mathbf{Z}$ in $\mathbf{X}$ and $\mathbf{Y}$ using the new regressions below. Consider new linear regression models as follows

$$\mathbf{X} = a_1\hat{\mathbf{X}} + a_2\hat{\mathbf{Y}} + \eta_1, \quad (11)$$

$$\mathbf{Y} = b_1\hat{\mathbf{X}} + b_2\hat{\mathbf{Y}} + \tau_1, \quad (12)$$

where η_1 and τ_1 are error terms, the estimators $\hat{\mathbf{X}} = \sum_{i=1}^p \hat{\alpha}_i \mathbf{Z}_i$ and $\hat{\mathbf{Y}} = \sum_{i=1}^p \hat{\beta}_i \mathbf{Z}_i$ are estimated by the Lasso or MCP regularisation regressions of models (8) and (9). Then, we implement the ordinary least squares (OLS) on models (11) and (12), and denote new estimators of $\mathbf{X}$ and $\mathbf{Y}$ by $\hat{\mathbf{X}}_{Reg2}$ and $\hat{\mathbf{Y}}_{Reg2}$. Computing new residuals $\hat{\eta}_1 = \mathbf{X} - \hat{\mathbf{X}}_{Reg2}$ and $\hat{\tau}_1 = \mathbf{Y} - \hat{\mathbf{Y}}_{Reg2}$, we finally estimate Pcor by the Reg2 algorithm as $\hat{\rho}_{reg2} = cor(\hat{\eta}_1, \hat{\tau}_1)$.

2.4.3. Coef and Var Algorithm

The Coef and Var algorithm is generated through the introduction of novel regression coefficients based on the Pcor definition formula (5) and (6). Consider linear regression models as follows

$$\mathbf{X} = \lambda_0 \mathbf{Y} + \sum_{i=1}^p \lambda_i \mathbf{Z}_i + \eta_2, \quad (13)$$

$$\mathbf{Y} = \gamma_0 \mathbf{X} + \sum_{i=1}^p \gamma_i \mathbf{Z}_i + \tau_2, \quad (14)$$

where η_2 and τ_2 are error terms. Then, we implement MCP regularisation on these models (13) and (14) and obtain estimated first-term regression coefficients $\hat{\lambda}_0$, $\hat{\gamma}_0$ and the estimated variance $var(\hat{\eta}_2)$, $var(\hat{\tau}_2)$. Finally, we can obtain the Pcor estimate by Coef algorithm as $\hat{\rho}_{coef} = sign(\hat{\lambda}_0) \sqrt{\hat{\lambda}_0 \hat{\gamma}_0}$ and the Pcor estimate by Var algorithm as $\hat{\rho}_{var} = \hat{\lambda}_0 \sqrt{var(\hat{\tau}_2)/var(\hat{\eta}_2)}$.

2.4.4. RSS2 Algorithm

The RSS2 algorithm is given by the residual sum of squares in formula (7). First, we implement the MCP regularisation on model (9): $\mathbf{Y} = \mathbf{Z}\beta + \zeta$ and estimate the residual $\hat{\zeta}$ and the residual sum of squares (RSS) $R_1 = \|\hat{\zeta}\|_2^2$. Similarly, we implement the MCP regularisation on model (14): $\mathbf{Y} = \gamma_0 \mathbf{X} + \sum_{i=1}^p \gamma_i \mathbf{Z}_i + \tau$ and estimate the first-term regression coefficient $\hat{\gamma}_0$, the residual $\hat{\tau}$, and the RSS $R_2 = \|\hat{\tau}\|_2^2$. Then, we obtain the Pcor estimate $\hat{\rho}_Y = sign(\hat{\gamma}_0) \sqrt{\max(0, R_1 - R_2)/R_1}$. Switch the position of $\mathbf{X}$ and $\mathbf{Y}$ similarly as the above steps. Then, we implement the MCP regularisation on model (8): $\mathbf{X} = \mathbf{Z}\alpha + \varepsilon$ and model (13): $\mathbf{X} = \lambda_0 \mathbf{Y} + \sum_{i=1}^p \lambda_i \mathbf{Z}_i + \eta$ and obtain the RSS $R_3 = \|\hat{\varepsilon}\|_2^2$, $R_4 = \|\hat{\eta}\|_2^2$ and the estimated first-term regression coefficient $\hat{\lambda}_0$. We obtain another Pcor estimate $\hat{\rho}_X = sign(\hat{\lambda}_0) \sqrt{\max(0, R_3 - R_4)/R_3}$. Finally, we have the estimate Pcor by RSS2 algorithm as $\hat{\rho}_{rss2} = (\hat{\rho}_X + \hat{\rho}_Y)/2$.

3. Minimum Residual Sum of Squares Pcor Estimation Algorithm

3.1. Motivation

From the comprehensive simulations in this paper, it is evident that the Pcor estimation methods discussed exhibit significant bias. This bias becomes more pronounced as the true Pcor increases, especially when the Pcor is positive. Therefore, further research is necessary to address this estimation bias in positive Pcor scenarios. While each algorithm has its merits, the Reg2 algorithm performs notably well when Pcor is below approximately 0.5. In contrast, the Coef and Var algorithm stands out with minimal bias when Pcor exceeds roughly 0.5. Our goal is to develop a method that synergises the strengths of both the Reg2 and Var algorithms.

The models introduced in the Reg2 algorithm, (11) and (12), can be represented as,

$$\mathbf{X} = a_1 \hat{\mathbf{X}} + a_2 \sum_{i=1}^p \hat{\beta}_i \mathbf{Z}_i + \eta_1, \quad (15)$$

$$\mathbf{Y} = b_2 \hat{\mathbf{Y}} + b_1 \sum_{i=1}^p \hat{\alpha}_i \mathbf{Z}_i + \tau_1, \quad (16)$$

When compared to models (13) and (14) from the Coef and Var algorithm, it is evident that the residuals η_1 and η_2 share commonalities. Both provide insights into the information in X after the exclusion of Y and Z effects in some sense. Similarly, τ_1 and τ_2 capture the essence of Y after removing for X and Z influences. If we choose a η_i and τ_i with a smaller residual sum of squares, then this will lead to a better estimation for the corresponding regression models. A reduced residual sum of squares in the corresponding regression models signifies enhanced precision in eliminating controlling variables effects, leading to a more accurate Pcor estimator. Guided by the objective of minimising the residual sum of squares, we introduce a novel algorithm for high-dimensional Pcor estimation in the subsequent subsection.

3.2. MRSS Algorithm and Its Implementation

We propose a novel Minimum Residual Sum of Squares partial correlation coefficient estimation algorithm, denoted by MRSS. This algorithm aims to diminish the estimation bias for positive Pcor values under high-dimensional situations. Our MRSS algorithm amalgamates the strengths of the Reg2, Coef, and Var algorithms, effectively curtailing bias in Pcor estimation.

Define $RSSX = \|\eta_k\|_2^2$ and $RSSY = \|\tau_k\|_2^2$ as the residual sum of squares of X after removing the effects of X and Z , and the residual sum of squares of Y after removing the effects of X and Z , respectively. The tuning parameter k is chosen by minimising the sum of squares of the residuals, so as to remove more associated effects and ensure a more efficient Pcor estimator. For $k = 1$, the pair (η_1, τ_1) represents the residuals from the Reg2 algorithm's models (11) and (12). For $k = 2$, (η_2, τ_2) corresponds to the residuals from the Coef and Var algorithms' models (13) and (14). Then, the residuals estimated by the MRSS algorithm satisfy the minimum residual sum of squares of both X and Y for a more efficient Pcor estimator as follows

$$\begin{aligned} \eta_{mrss} &= \arg \min_{k=1,2} RSSX = \arg \min_{k=1,2} \|\eta_k\|_2^2, \\ \tau_{mrss} &= \arg \min_{k=1,2} RSSY = \arg \min_{k=1,2} \|\tau_k\|_2^2. \end{aligned} \quad (17)$$

The Pcor estimated by MRSS is then given by

$$\rho_{mrss} = cor(\eta_{mrss}, \tau_{mrss}) I_{k=1} + \lambda_0 \sqrt{var(\tau_{mrss}) / var(\eta_{mrss})} I_{k=2} \quad (18)$$

where I is the indicator function and λ_0 is the primary regression coefficient in model (13). If $k = 1$, then ρ_{mrss} is estimated following the idea of Reg2 algorithm; if $k = 2$, then ρ_{mrss} is estimated following the idea of the Coef and Var algorithm. If the two k estimates in (17) differ, the more stable Reg2 algorithm is preferred, setting $k = 1$ in (18). Given that MRSS integrates two existing algorithms, its convergence should align with their rates.

During the implementation of the MRSS algorithm (Algorithm 1), the Coef and Var algorithm often misestimates Pcor as 0 or ± 1 when the true Pcor is close to 0 or ± 1 , affecting the algorithms' precision. To address this, we incorporate a discriminative condition in the MRSS pseudo-code. If the estimated Pcor $\hat{\rho}_{coef}$ or $\hat{\rho}_{var}$ is zero or ± 1 , the Coef and Var algorithm is deemed unreliable, and the Reg2 algorithm's estimate is adopted.

Algorithm 1: MRSS algorithm

Data: $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ with the dimension (n, p)
Result: Pcor estimate $\hat{\rho}_{mrss}$

- 1 Implement MCP regularisation on models (8) and (9), and obtain $\hat{\mathbf{X}}$ and $\hat{\mathbf{Y}}$;
- 2 Implement ordinary least squares (OLS) on models (11) and (12), and obtain $\hat{\mathbf{X}}_{Reg2}$ and $\hat{\mathbf{Y}}_{Reg2}$ with residuals $\hat{\eta}_1 = \mathbf{X} - \hat{\mathbf{X}}_{Reg2}$ and $\hat{\tau}_1 = \mathbf{Y} - \hat{\mathbf{Y}}_{Reg2}$. Calculate the RSS by $RSSX_1 = \|\hat{\eta}_1\|_2^2$ and $RSSY_1 = \|\hat{\tau}_1\|_2^2$;
- 3 Implement MCP regularisation on models (13) and (14), obtain estimated coefficients $\hat{\lambda}_0, \hat{\gamma}_0$ and residuals $\hat{\eta}_2, \hat{\tau}_2$. Calculate RSS by $RSSX_2 = \|\hat{\eta}_2\|_2^2$, $RSSY_2 = \|\hat{\tau}_2\|_2^2$;
- 4 Estimate the Coef Pcor by $\hat{\rho}_{coef} = \text{sign}(\hat{\lambda}_0) \sqrt{\hat{\lambda}_0 \hat{\gamma}_0}$ and the Var Pcor by $\hat{\rho}_{var} = \hat{\lambda}_0 \sqrt{\text{var}(\hat{\tau}_2) / \text{var}(\hat{\eta}_2)}$;
- 5 **if** $\hat{\rho}_{coef} = 0$ or $\hat{\rho}_{var} = 0$ or ± 1 **then**
- 6 | Estimate Pcor by $\hat{\rho}_{mrss} = \text{cov}(\hat{\eta}_1, \hat{\tau}_1)$;
- 7 **else if** $RSSX_2 \leq RSSX_1$ and $RSSY_2 \leq RSSY_1$ **then**
- 8 | Estimate Pcor by $\hat{\rho}_{mrss} = \hat{\lambda}_0 \sqrt{\text{var}(\hat{\tau}_2) / \text{var}(\hat{\eta}_2)}$;
- 9 **else**
- 10 | Estimate Pcor by $\hat{\rho}_{mrss} = \text{cov}(\hat{\eta}_1, \hat{\tau}_1)$;
- 11 **end**

The proposed MRSS algorithm selects the most suitable residuals by minimising RSS and removing the impact of control variables to optimise the estimation of residuals in the regression model. As such, the estimated Pcor generated by the MRSS algorithm combines the advantages of both algorithms, resulting in a more accurate estimate. Notably, our MRSS algorithm effectively addresses the Pcor estimation bias in cases where $\text{Pcor} \geq 0$. For instance, when the Coef and Var algorithms estimate Pcor as 0 for true Pcor near 0, the MRSS algorithm utilises the minimum RSS principle to select the Reg2 algorithm, which performs better in the vicinity of $\text{Pcor} = 0$, and thereby efficiently avoids such misestimations. Around $\text{Pcor} = 0.5$, the MRSS algorithm employs the minimum RSS principle to determine the more accurate method between Reg2 and Var for exact selection. This selection conforms to the minimum RSS principle, where the regression model and accompanying residuals are selected to provide optimal estimation accuracy, leading to a more precise Pcor estimate. When Pcor lies close to 1, the Reg2 algorithm's estimates are typically lower with a high RSSs. Thereafter, the MRSS method selects the Var algorithm with small RSSs, which performs better based on the minimum RSS principle. In essence, the MRSS method amalgamates the merits of the Reg2 and Var algorithms. By reducing the sum of squares of the residuals, MRSS can choose the algorithm with a smaller estimation error for $\text{Pcor} \geq 0$, which allows for the proficient regulation of the estimation bias of Pcor.

4. Simulation

4.1. Data Generation

To study the estimation efficiency of Pcor estimation algorithms under high-dimensional conditions, we generate n centralised samples $\{x_j, y_j, z_{j1}, \dots, z_{jp}\}_{j=1}^n$ i.i.d from $[X, Y, Z]$ with $Z = [Z_1, \dots, Z_p]$. Let $\mathbf{X} = [x_1, \dots, x_n]^T$, $\mathbf{Y} = [y_1, \dots, y_n]^T$ and $\mathbf{Z} = [\mathbf{Z}_1, \dots, \mathbf{Z}_p] = (z_{ji})_{n \times p}$. Initially, we produce n controlling samples $\{\mathbf{Z}_i\}_{i=1}^p$ independently and identically by

$$\mathbf{Z}_i = 0.5\mathbf{u} + \mathbf{e}_i$$

where $\mathbf{u} = [u_1, \dots, u_n]^T$ and $\mathbf{e}_i = [e_{1i}, \dots, e_{ni}]^T$ with u_j and e_{ji} generated independently from the normal distribution $N(0, \sigma^2)$ with variance σ^2 for $i = 1, \dots, p$. The samples $\mathbf{X}$ and $\mathbf{Y}$ are then generated by

$$\mathbf{X} = \sum_{i=1}^p \alpha_i \mathbf{Z}_i + \varepsilon, \quad \text{and} \quad \mathbf{Y} = \sum_{i=1}^p \beta_i \mathbf{Z}_i + \zeta,$$

where $\varepsilon = [\varepsilon_1, \dots, \varepsilon_n]^T$ and $\zeta = [\zeta_1, \dots, \zeta_n]^T$ with $\zeta_j = \frac{\omega \varepsilon_j + \eta_j}{\sqrt{1 + \omega^2}}$ and ε_j, η_j drawn i.i.d. from $N(0, \sigma^2)$. The Pearson correlation of ε and ζ gives the partial correlation coefficient $\text{Pcor} \rho_{XY|Z} = \frac{\omega}{\sqrt{1 + \omega^2}}$. Notably, there is a one-to-one mapping between the true Pcor and the ω parameter.

Since our MRSS algorithm and the Reg2 algorithm perform essentially the same for $\text{Pcor} < 0$, our simulation focuses on real Pcor values in the range $[0, 1]$, an interval prone to significant biases with existing methods. Let the true partial correlation coefficient vary as $\text{Pcor} = 0, 0.05, 0.1, \dots, 0.95$ with the sample size $n = 50, 100, \dots, 400$, the controlling variable size $p = 200, 500, 1000, 2000, 4000$ and the normal distribution variance $\sigma^2 = 1, 10, 40$. For each n, p combination, we estimate the partial correlation coefficient for 200 replications using the aforementioned estimation algorithms. We use the software R (4.3.1) for our simulation.

Recognising that both sparse and non-sparse conditions are prevalent in real-world applications [3,28], we present examples under both conditions. To ensure comparability between the examples, the initial l coefficients of α and β are fixed under both conditions, where we select the high-correlated numbers of controlling variables as $l = 6, 10, 14$. For non-sparse examples, the coefficients of α and β asymptotically converge to 0 at varying rates, with coefficients beyond the $(l + 1)$ -th starting at 0.05, which is significantly smaller than the initial l coefficients.

- Example 1: under sparse conditions
Let the coefficients α and β be non-zero for the initial l elements and zero for the rest as follows

$$\alpha = -\beta = (-0.1, -0.2, \dots, -\frac{l}{20}, 0.1, 0.2, \dots, \frac{l}{20}, 0, \dots, 0).$$
- Example 2: under non-sparse conditions
Let the coefficients α and β be the same as Example 1 for the initial l elements with a convergence rate of $O(1/2^p)$ for the remaining elements as follows

$$\alpha = -\beta = (-0.1, -0.2, \dots, \frac{l}{20}, \frac{r}{2^{l/2+1}}, \frac{r}{2^{l/2+2}}, \dots, \frac{r}{2^{p/2}}, -\frac{r}{2^{l/2+1}}, \dots, -\frac{r}{2^{p/2}}),$$
where r is a tuning parameter to make the $(l + 1)$ -th element close to 0.05.
- Example 3: under non-sparse conditions
Let the coefficients α and β be the same as Example 1 for the initial l elements with a convergence rate of $O(1/p)$ for the remaining elements as follows,

$$\alpha = -\beta = (-0.1, -0.2, \dots, \frac{l}{20}, \frac{r}{l/2+1}, \frac{r}{l/2+2}, \dots, \frac{r}{p/2}, -\frac{r}{l/2+1}, \dots, -\frac{r}{p/2}),$$
where r is a tuning parameter to make the $(l + 1)$ -th element close to 0.05.
- Example 4: under non-sparse conditions
Let the coefficients α and β be the same as Example 1 for the initial l elements with a convergence rate of $O(1/\sqrt{p})$ for the remaining elements as follows,

$$\alpha = -\beta = (-0.1, -0.2, \dots, \frac{l}{20}, \frac{r}{\sqrt{l/2+1}}, \frac{r}{\sqrt{l/2+2}}, \dots, \frac{r}{\sqrt{p/2}}, -\frac{r}{\sqrt{l/2+1}}, \dots, -\frac{r}{\sqrt{p/2}}),$$
where r is a tuning parameter to make the $(l + 1)$ -th element close to 0.05.

4.2. Simulation Results

4.2.1. By MSE and RMSE

We will assess the efficacy of the Pcor estimation algorithms using the mean square error (MSE) and root mean square error (RMSE) indices as follows. These evaluation indicators may indicate the performance of Pcor estimation algorithms from various perspectives.

$$MSE(\rho_0) = \frac{1}{R} \sum_{i=1}^R (\hat{\rho}_{(i)} - \rho_0)^2, \quad \text{and} \quad RMSE(\rho_0) = \sqrt{\frac{1}{R} \sum_{i=1}^R (\hat{\rho}_{(i)} - \rho_0)^2},$$

where ρ_0 is the true Pcor, and $\hat{\rho}_{(i)}$ is the estimated Pcor in the (i) -th replication of $R = 200$ replications.

Table 1 displays the mean of MSE and RMSE ($\times 10^2$) for the estimated Pcors of the true Pcor = 0, 0.05, ..., 0.95 with $l = 10$, $\sigma^2 = 1$, $n = 50, 100, 200, 400$ and $p = 200, 500, 1000, 2000, 4000$ across Examples 1–4 using various methods. Tables A1 and A2, which consider the means of MSE and RMSE ($\times 10^2$) for the estimated Pcors for high correlation controlling variables number $l = 6, 14$, can be found in the Appendix.

Table 1. The mean of MSE ($\times 10^2$) and RMSE ($\times 10^2$) for estimated Pcors of real Pcor = 0, 0.05, ..., 0.95 with $l = 10$, $\sigma^2 = 1$, $n = 50, 100, 200, 400$ and $p = 200, 500, 1000, 2000, 4000$ in Examples 1–4.

		MSE ($\times 10^2$)								RMSE ($\times 10^2$)							
Example 1		Lasso		MCP				Lasso				MCP					
n	p	Res	Reg2	Res	Reg2	Coef	Var	RSS2	MRSS	Res	Reg2	Res	Reg2	Coef	Var	RSS2	MRSS
50	200	36.7	21.7	37.6	27.4	34.6	35.4	34.6	27.1	59.2	44.4	59.9	50.8	55.2	56.3	55.6	50.5
	500	46.8	29.6	47.2	36.0	34.4	35.3	36.4	35.5	66.9	51.9	67.2	58.2	54.9	56.0	57.3	57.5
	1000	51.4	33.0	50.7	39.3	34.1	34.8	36.6	38.8	70.2	54.8	69.7	60.8	54.3	55.3	57.5	60.3
100	500	22.1	11.7	15.2	9.5	24.2	21.5	14.0	7.7	45.9	32.4	38.1	29.9	47.9	46.0	35.5	27.3
	1000	29.6	16.1	22.4	13.9	32.5	30.0	22.6	13.4	53.1	37.8	46.1	36.1	54.5	53.5	45.2	35.4
	2000	36.1	19.7	29.6	18.4	35.0	34.3	29.4	18.2	58.8	41.7	53.2	41.5	56.1	56.3	51.1	41.3
200	500	6.6	3.1	1.7	1.5	2.4	2.6	1.8	0.7	25.1	17.3	12.8	12.1	12.7	13.3	11.5	7.6
	1000	9.6	5.2	2.5	2.2	3.8	4.3	2.4	0.9	30.2	22.1	15.6	14.5	16.4	17.1	13.1	8.9
	2000	13.0	7.9	3.6	3.1	6.0	6.6	3.3	1.4	35.3	27.0	18.6	17.3	21.0	21.5	15.3	10.5
400	1000	2.7	1.3	0.5	0.5	0.7	0.6	0.7	0.2	16.1	11.3	6.7	6.8	6.3	6.0	6.7	4.4
	2000	3.8	2.2	0.6	0.6	0.8	0.7	0.9	0.3	19.2	14.6	7.8	8.0	6.8	6.4	7.5	4.9
	4000	5.2	3.5	0.8	0.9	1.0	0.9	1.1	0.4	22.3	18.4	9.0	9.3	7.7	7.5	8.7	5.6
Example 2																	
50	200	37.0	21.8	38.3	28.1	34.7	35.8	34.9	27.9	59.4	44.4	60.4	51.3	55.3	56.7	56.0	51.2
	500	47.1	29.5	47.6	36.5	34.5	35.4	36.4	36.0	67.1	51.7	67.4	58.5	55.0	56.1	57.2	58.0
	1000	51.7	33.6	51.3	39.8	34.1	34.8	36.9	39.2	70.4	55.4	70.1	61.2	54.4	55.2	57.8	60.5
100	500	22.6	12.0	15.7	9.9	25.1	22.6	14.9	8.1	46.5	32.8	38.7	30.5	48.8	47.1	36.7	28.0
	1000	30.0	16.5	23.0	14.4	33.3	30.8	23.4	13.9	53.5	38.1	46.8	36.6	55.1	54.1	45.8	36.2
	2000	36.6	20.1	30.3	19.2	35.3	34.7	30.0	19.0	59.2	42.1	53.8	42.4	56.4	56.6	51.6	42.2
200	500	6.9	3.4	1.9	1.7	2.6	2.8	1.9	0.8	25.8	17.9	13.7	12.9	13.5	14.0	11.8	8.2
	1000	9.9	5.6	2.8	2.4	4.1	4.6	2.6	1.1	30.8	22.8	16.4	15.4	17.0	17.9	13.4	9.5
	2000	13.5	8.3	3.9	3.4	6.4	6.9	3.5	1.5	35.9	27.6	19.5	18.1	21.6	22.3	15.8	11.2
400	1000	2.9	1.5	0.6	0.6	0.7	0.6	0.7	0.3	16.8	12.1	7.5	7.5	6.7	6.4	6.9	4.9
	2000	4.1	2.5	0.8	0.8	0.9	0.8	0.9	0.4	19.9	15.4	8.6	8.8	7.3	7.0	7.8	5.5
	4000	5.5	3.9	1.0	1.1	1.1	1.0	1.2	0.5	23.0	19.3	9.8	10.3	8.3	8.1	9.1	6.3

Table 1. Cont.

		MSE ($\times 10^2$)								RMSE ($\times 10^2$)							
		Lasso		MCP				Lasso				MCP					
n	p	Res	Reg2	Res	Reg2	Coef	Var	RSS2	MRSS	Res	Reg2	Res	Reg2	Coef	Var	RSS2	MRSS
Example 3																	
50	200	38.5	23.0	39.9	29.5	35.1	36.2	35.5	29.4	60.6	45.7	61.6	52.7	55.7	57.2	56.5	52.5
	500	48.7	30.6	49.2	37.9	34.8	35.7	37.1	37.3	68.3	52.7	68.5	59.6	55.4	56.5	58.0	59.0
	1000	53.4	34.5	53.0	41.4	34.5	35.4	37.4	40.9	71.5	56.0	71.2	62.4	54.9	56.1	58.4	61.9
100	500	24.4	13.3	17.8	11.5	27.9	25.5	17.3	10.1	48.3	34.6	41.2	32.8	51.3	50.0	39.6	31.1
	1000	31.9	17.5	25.2	16.1	34.6	33.1	26.5	15.8	55.2	39.3	49.0	38.8	56.2	55.9	48.6	38.5
	2000	38.8	21.5	32.4	20.7	36.0	35.8	31.7	20.6	61.0	43.6	55.6	44.1	57.1	57.6	53.0	44.0
200	500	8.3	4.4	2.8	2.6	3.6	3.9	2.5	1.3	28.1	20.3	16.5	15.8	17.1	17.7	14.0	11.1
	1000	11.6	6.9	4.0	3.5	5.5	6.2	3.3	1.7	33.3	25.2	19.5	18.3	21.1	22.1	16.0	12.8
	2000	15.4	9.7	5.4	4.7	8.3	8.9	4.6	2.3	38.3	29.8	22.6	21.1	26.2	26.8	18.8	14.6
400	1000	4.0	2.3	1.1	1.2	1.2	1.0	1.0	0.6	19.6	15.1	10.6	10.6	9.5	9.2	8.9	7.7
	2000	5.4	3.6	1.4	1.5	1.4	1.3	1.2	0.8	22.8	18.6	11.9	12.0	10.4	10.1	9.9	8.5
	4000	7.0	5.2	1.8	1.9	1.7	1.6	1.6	0.9	25.9	22.3	13.1	13.5	11.6	11.5	11.2	9.4
Example 4																	
50	200	41.7	25.4	43.6	32.8	36.3	37.6	37.2	32.6	63.1	48.1	64.4	55.6	57.2	58.7	58.3	55.4
	500	53.8	34.3	54.3	42.6	36.1	37.2	38.9	41.9	71.7	56.0	72.0	63.3	57.0	58.2	60.0	62.6
	1000	58.9	38.3	58.9	46.8	35.9	36.9	39.9	45.7	75.2	59.3	75.1	66.4	56.7	57.8	61.1	65.3
100	500	30.4	17.3	24.8	16.9	35.5	34.6	25.9	16.6	53.9	39.4	48.6	39.9	57.6	57.8	48.1	39.7
	1000	39.7	22.5	34.6	23.5	38.4	39.5	33.6	23.4	61.6	44.8	57.4	46.8	59.8	61.1	54.5	46.8
	2000	48.0	27.1	43.2	29.7	39.5	40.8	37.0	29.5	67.8	49.3	64.3	52.9	60.9	62.1	57.9	52.7
200	500	12.9	8.1	6.8	6.3	8.7	9.3	5.5	4.5	35.1	27.4	25.5	24.4	28.7	29.8	22.1	21.1
	1000	18.5	12.4	9.8	8.8	13.9	15.0	8.2	6.7	42.1	33.7	30.6	28.9	36.5	38.0	26.9	25.6
	2000	24.6	16.6	13.5	11.8	21.6	22.1	12.3	9.7	48.5	38.9	36.0	33.5	45.8	46.5	32.7	30.7
400	1000	9.1	7.0	5.2	5.2	5.2	5.0	4.0	4.2	29.6	25.9	22.3	22.4	22.2	21.9	19.7	20.4
	2000	12.7	10.4	7.0	7.1	7.2	7.1	5.8	5.8	34.9	31.4	25.9	26.0	26.2	26.2	23.7	24.0
	4000	16.5	14.1	8.9	9.2	9.7	9.9	7.8	7.8	39.8	36.3	29.3	29.6	30.4	31.1	27.2	27.6

For small sample sizes ($n < 100$), all algorithms tend to underperform due to the limited data information, with the mean MSE and RMSE being approximately ten times higher than that of large sample size $n > 100$. And, our MRSS algorithm remains competitive, with both MSE and RMSE in the same order of magnitude as the best performance Lasso.Reg2. However, for large sample size ($n \geq 100$), the MRSS algorithm's performance becomes notably superior. Specifically, the MRSS reduces the MSE by around 40% compared to the suboptimal MCP.Reg2, and this percentage grows with increasing n . The MRSS represents a significant improvement in algorithmic performance. Additionally, the MSE of the MRSS algorithm exhibits a slower increase with increasing controlling size p , implying improved stability to some extent.

To compare the performance of different algorithms more intuitively, we calculated the percentage difference of MSE by $\frac{MSE_{MRSS} - MSE_{ALG}}{MSE_{ALG}} \times 100\%$ with ALG be algorithms listed above. Similarly, the percentage difference of RMSE can be calculated. And, Table 2 shows the average percentage difference of MSE and RMSE compared to the MRSS algorithm for a small sample size ($n = 50$) and large sample size ($n = 100, 200, 400$) with the same settings in Table 1. For a small sample size ($n = 50$), we observe a 10–20% decrease in MSE and RMSE for an MRSS algorithm relative to the Res algorithm, a 10–20% increase relative to Lasso.Reg2, and a slight change relative to other algorithms. For large sample size ($n = 100, 200, 400$), the MRSS algorithm reduces MSE by about 30–70% and RMSE by 20–60% relative to other algorithms, achieving effective control of the Pcor estimation error. These results further illustrate the superiority of the MRSS algorithm. For optimal Pcor

estimation performance, we suggest using the MRSS algorithm with a minimum sample size of $n = 100$.

Table 2. The average percentage difference of the MSE and RMSE compared to the MRSS algorithm for a small sample size ($n = 50$) and a large sample size ($n = 100, 200, 400$) with the same settings in Table 1.

Example	(%)		For MSE						For RMSE					
			Lasso			MCP			Lasso			MCP		
	Res	Reg2	Res	Reg2	Coef	Var	RSS2	Res	Reg2	Res	Reg2	Coef	Var	RSS2
Small sample size ($n = 50$)														
Example 1	−25	21	−25	−1	−2	−4	−6	−14	12	−15	−1	2	0	−1
Example 2	−24	22	−25	−1	0	−3	−5	−14	12	−14	−1	3	1	−1
Example 3	−23	23	−24	−1	3	0	−2	−13	12	−14	−1	4	2	0
Example 4	−22	23	−23	−1	11	8	3	−13	12	−13	−1	7	5	2
Large sample size ($n = 100, 200, 400$)														
Example 1	−79	−62	−52	−39	−65	−63	−56	−60	−44	−34	−27	−36	−35	−30
Example 2	−78	−61	−51	−39	−63	−61	−53	−58	−43	−34	−26	−34	−34	−27
Example 3	−74	−55	−47	−34	−56	−55	−42	−52	−35	−28	−20	−30	−30	−18
Example 4	−53	−28	−27	−14	−37	−38	−15	−31	−14	−14	−6	−20	−21	−6

For Examples 1–4, shifting from sparse to non-sparse conditions with increasing non-sparsity, we observe that all algorithms exhibit a higher MSE and RMSE under non-sparse conditions compared to sparse conditions, and the MSE and RMSE increase with increasing non-sparsity. This could be attributed to the greater impact and more complicated correlations of the controlling variables, resulting in a less accurate estimate of the partial correlation. However, even in Example 4 with the strongest non-sparsity, the MRSS algorithm still performs well, possessing the smallest MSE and RMSE and outperforming conventional algorithms. Especially under non-sparse conditions, the MRSS algorithm provides a dependable and accurate estimation of Pcor despite the influence of complex controlling variables.

4.2.2. For Pcor Values on $[0, 1]$

To investigate the effectiveness of Pcor estimation algorithms for various Pcor values, we set a constant ratio of the dimension of controlling variables to the sample size (i.e., a fixed $p/n = 2, 10$). Figure 1 displays the average estimated Pcor of 200 repetitions compared to the true Pcor for $n = 100, 200, 400$ and $l = 6$ in Example 1. The MRSS, MCP.Reg2, and MCP.Var are denoted in red, green and blue, respectively. When Pcor is small around $Pcor < 0.5$, the MRSS accurately simulates the true Pcor, performing similarly to the MCP.Reg2. When Pcor is large, like about $Pcor > 0.5$, the MRSS performs sub-optimally and comparable to the MCP.Var, falling slightly behind the RSS2. Essentially, the MRSS effectively amalgamates the strengths of both MCP.Reg2 and MCP.Var algorithms, reducing potential weaknesses for Pcor estimation. For a small sample size $n = 100$, the MRSS leads to a significant improvement in the estimation for a large Pcor in $[0, 1]$, but still a considerable estimation bias for small Pcor in $[0, 1]$ owing to the limited sample size and information. For a large sample size $n \geq 200$, the MRSS effectively reduces the Pcor estimation bias for $Pcor > 0$. Consequently, greatly enhancing the sample size substantially boosts the MRSS estimation accuracy, even if the ratio of the controlling variables dimension to the sample size p/n increases from 2 to 10.

4.3. Parameter Sensitivity

We investigate the sensitivity of the performance of the MRSS algorithm to different parameter settings, such as variance and sparsity. This allows us to explore the robustness of algorithms under different parameter configurations.

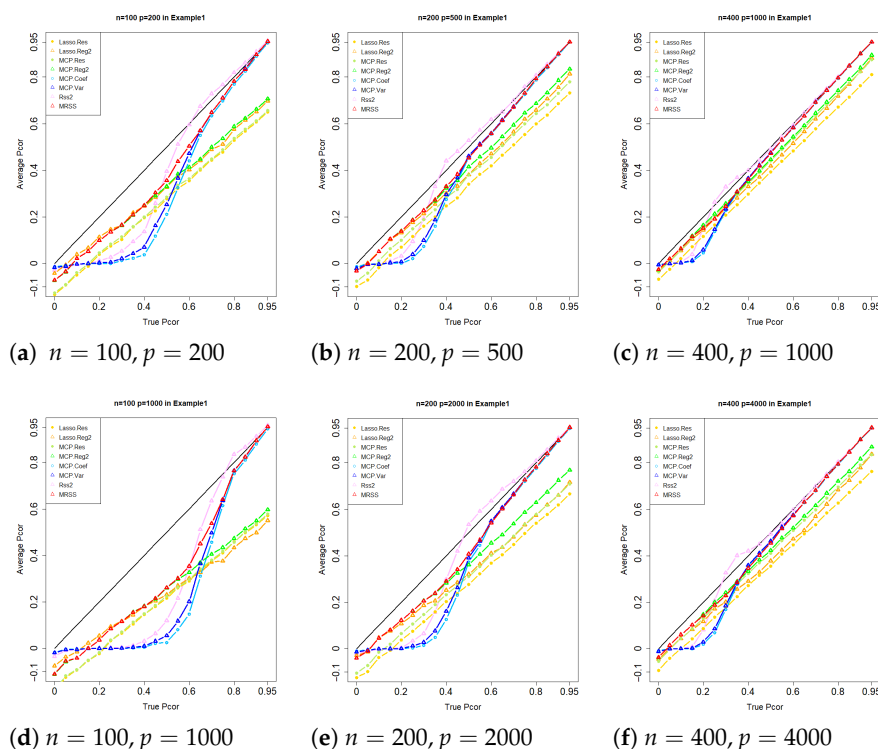


Figure 1. Average Pcor against true Pcor of each true Pcor = 0, 0.1, ..., 0.95 for $p = 2n$ in first row and $p = 10n$ in second row with $n = 100, 200, 400$ and $l = 6$ in Example 1.

4.3.1. For Variance

We set a variance parameter σ^2 in data generation to test the stability of our algorithm under varying variance. Table 3 shows the mean of MSE ($\times 10^2$) and RMSE ($\times 10^2$) for the estimated Pcors of real Pcor = 0, 0.05, ..., 0.95 with different variances $\sigma^2 = 1, 10, 40$ and $l = 10$ for a large sample size ($n = 50, 100$) and small sample size ($n = 200, 400$) in Examples 1–4. We discover that, as the variance increases σ^2 from 1 to 40, the MSE and RMSE remain consistent for various examples and sample sizes. This indicates that our MRSS algorithm is highly robust to variance and retains good stability.

Table 3. The means of MSE ($\times 10^2$) and RMSE ($\times 10^2$) for the estimated Pcors of real Pcor = 0, 0.05, ..., 0.95 with different variances $\sigma^2 = 1, 10, 40$ and $l = 10$ for large sample size ($n = 50, 100$) and small sample size ($n = 200, 400$) in Examples 1–4.

Small Sample Size		MSE ($\times 10^2$)								RMSE ($\times 10^2$)									
Example	σ^2	Lasso		MCP				Lasso		MCP				Lasso		MCP			
		Res	Reg2	Res	Reg2	Coef	Var	RSS2	MRSS	Res	Reg2	Res	Reg2	Coef	Var	RSS2	MRSS		
Example 1	1	37.10	21.99	33.79	24.08	32.47	31.89	28.93	23.44	59.01	43.83	55.70	46.21	53.82	53.89	50.38	45.38		
	10	37.04	21.74	33.66	24.02	32.33	31.84	28.73	23.47	58.96	43.64	55.58	46.13	53.70	53.88	50.16	45.39		
	40	37.02	21.84	33.80	24.17	32.40	31.88	28.83	23.57	58.97	43.68	55.68	46.20	53.68	53.82	50.23	45.43		
Example 2	1	37.51	22.24	34.36	24.62	32.86	32.34	29.39	24.03	59.36	44.11	56.21	46.76	54.17	54.31	50.85	46.00		
	10	37.54	22.18	34.25	24.49	32.74	32.16	29.17	23.84	59.39	44.06	56.11	46.63	54.12	54.19	50.63	45.82		
	40	37.40	22.16	34.29	24.64	32.80	32.27	29.15	24.08	59.27	44.00	56.15	46.75	54.08	54.22	50.50	46.05		
Example 3	1	39.28	23.39	36.25	26.19	33.82	33.62	30.91	25.70	60.80	45.34	57.87	48.40	55.11	55.55	52.35	47.83		
	10	39.26	23.31	36.14	26.10	33.73	33.63	30.70	25.59	60.81	45.26	57.83	48.36	55.06	55.61	52.15	47.77		
	40	39.16	23.26	36.35	26.31	33.88	33.71	30.73	25.82	60.71	45.20	57.94	48.47	55.13	55.58	52.16	47.91		
Example 4	1	45.40	27.50	43.23	32.04	36.97	37.76	35.42	31.63	65.54	49.50	63.65	54.16	58.21	59.28	56.64	53.76		
	10	45.43	27.40	43.20	32.00	36.98	37.76	35.50	31.65	65.56	49.40	63.65	54.10	58.21	59.34	56.77	53.76		
	40	45.38	27.42	43.15	31.89	37.06	37.94	35.30	31.59	65.49	49.36	63.56	53.98	58.23	59.43	56.50	53.68		

Table 3. Cont.

Large sample size		MSE ($\times 10^2$)								RMSE ($\times 10^2$)									
Example	σ^2	Lasso		MCP				Lasso		MCP				Lasso		MCP			
		Res	Reg2	Res	Reg2	Coef	Var	RSS2	MRSS	Res	Reg2	Res	Reg2	Coef	Var	RSS2	MRSS		
Example 1	1	6.81	3.87	1.62	1.47	2.46	2.60	1.71	0.65	24.70	18.44	11.77	11.34	11.80	11.97	10.47	6.98		
	10	6.80	3.84	1.65	1.49	2.51	2.71	1.76	0.68	24.68	18.39	11.85	11.41	11.87	12.18	10.55	7.08		
	40	6.77	3.88	1.64	1.46	2.43	2.55	1.70	0.66	24.61	18.43	11.78	11.29	11.65	11.86	10.34	6.98		
Example 2	1	7.16	4.17	1.83	1.67	2.63	2.78	1.80	0.75	25.38	19.18	12.58	12.16	12.42	12.60	10.80	7.60		
	10	7.15	4.13	1.87	1.69	2.72	2.89	1.85	0.77	25.36	19.11	12.68	12.20	12.62	12.81	10.88	7.66		
	40	7.14	4.15	1.84	1.67	2.63	2.76	1.80	0.75	25.32	19.10	12.59	12.11	12.32	12.55	10.77	7.56		
Example 3	1	8.60	5.34	2.75	2.55	3.62	3.82	2.37	1.28	28.00	21.88	15.69	15.22	15.99	16.23	13.12	10.69		
	10	8.58	5.30	2.78	2.57	3.61	3.82	2.37	1.31	27.97	21.79	15.76	15.26	15.93	16.18	13.01	10.74		
	40	8.57	5.33	2.76	2.53	3.59	3.76	2.33	1.28	27.94	21.84	15.68	15.14	15.93	16.13	12.84	10.63		
Example 4	1	15.73	11.44	8.53	8.07	11.05	11.40	7.27	6.45	38.31	32.27	28.25	27.46	31.62	32.26	25.37	24.88		
	10	15.71	11.43	8.56	8.07	11.04	11.40	7.22	6.44	38.29	32.25	28.29	27.45	31.61	32.24	25.28	24.86		
	40	15.70	11.42	8.56	8.05	11.09	11.36	7.29	6.43	38.28	32.25	28.28	27.42	31.68	32.20	25.39	24.83		

4.3.2. For Sparsity

To evaluate the effectiveness of algorithms under different sparsity conditions, we set the data generation conditions to develop from sparse to non-sparse, with an increasingly non-sparse convergence rate from Example 1 to Example 4. This suggests a greater inclusion of controlling variables as we progress through the examples. From the above Tables 1–3, our observations show that the MRSS algorithm performs well for all examples. For moderate non-sparse convergence rates, as witnessed in Examples 2–3, MRSS demonstrates both low MSE and RMSE, comparable to the sparse conditions of Example 1. As the rate of non-sparsity convergence and the impact of controlling variables increase in Example 4, the best-performing MRSS also encounters difficulties in reducing the estimation bias. Therefore, the best-performing MRSS algorithm remains the most favoured choice for estimating Pcor under both sparse and non-sparse conditions. If it is possible to analyse the degree of non-sparsity the initial data, then we can obtain a better understanding of the algorithm's error margin.

Another indication of the sparsity strength is the number of high correlation controlling variables l . Figure 2 illustrates the performance of the featured algorithms for varying numbers $l = 6, 10, 14$. The figure contrasts the average Pcor with the true Pcor for $l = 6, 10, 14$ in Example 2 with the first row $n = 100, p = 200$ and the second $n = 200, p = 2000$. As l increases, the interference from controlling variables in the estimation process becomes more pronounced, leading to a heightened estimation bias. However, the MRSS algorithm consistently showcases an optimal performance throughout the entire $[0, 1]$ interval. Remarkably, despite encountering a high interference level at $l = 14$, MRSS keeps the bias in close alignment with the diagonal, in contrast to its counterparts. Table 4 shows the mean of the MSE and RMSE for $l = 6, 10, 14$. As l increases, both the MSE and RMSE of the MRSS algorithm increase, but always remain slightly weaker than optimal in small samples and significantly more optimal than the other algorithms in large samples. These results demonstrate the robustness, stability, and precision advantages of the MRSS algorithm.

4.4. Summaries

Based on numerous simulations, our study examines the practicality and effectiveness of the MRSS algorithm in a variety of scenarios. Through extensive simulations, we provide valuable insights into the accuracy and effectiveness of the MRSS algorithm. We provide empirical evidence that MRSS effectively incorporates the strengths of the MCP.Reg2 and MCP.Var algorithms and reduces the potential weaknesses of Pcor estimation, especially in challenging environments with high-dimensional sparse and non-sparse conditions. For larger sample sizes ($n \geq 100$), the MRSS algorithm reduces the MSE and RMSE

by approximately 30–70% compared to other algorithms and effectively controls Pcor estimation errors. For small sample sizes ($n < 100$), a reduction of 10–20% is observed in MSE and RMSE for the MRSS algorithm compared to the Res algorithm, an increase of 10–20% compared to Lasso.Reg2, and a slight change compared to other algorithms.

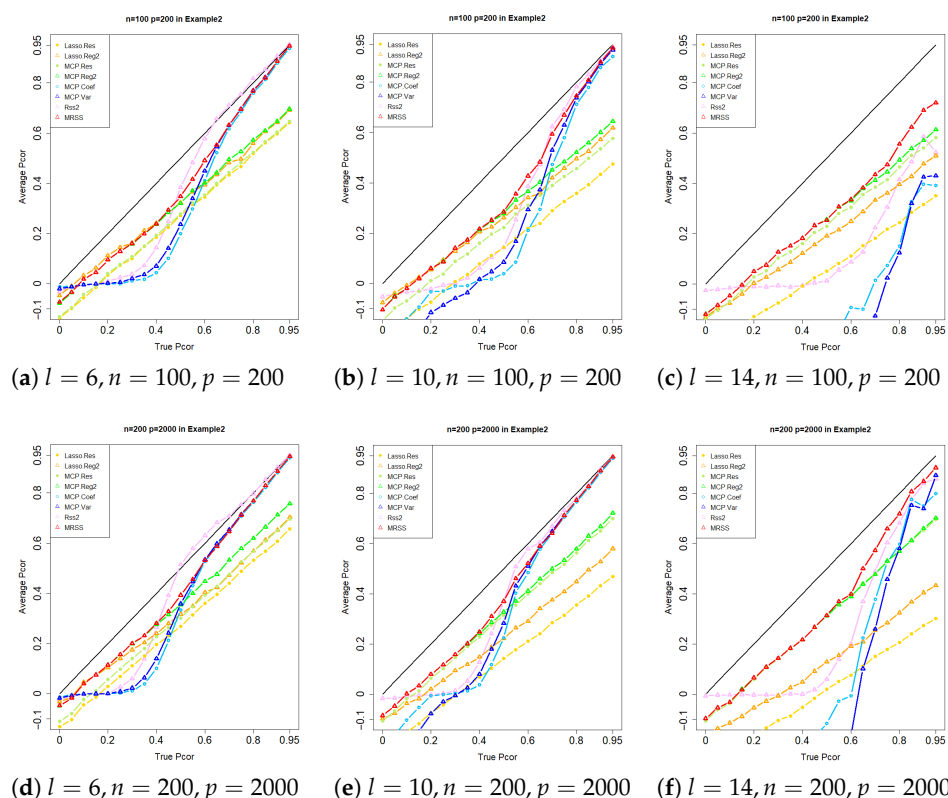


Figure 2. Average Pcor against true Pcor for $n = 100, p = 200$ in the first row and $n = 200, p = 2000$ in the second row with $l = 6, 10, 14$ in Example 2.

Table 4. The mean of MSE ($\times 10^2$) and RMSE ($\times 10^2$) for estimated Pcors of real Pcor = 0, 0.05, . . . , 0.95 with $l = 6, 10, 14$ and $\sigma^2 = 1$ for a large sample size ($n = 50, 100$) and small sample size ($n = 200, 400$) in Examples 1–4.

Small Sample Size		MSE ($\times 10^2$)								RMSE ($\times 10^2$)							
Example	l	Lasso		MCP						Lasso		MCP					
		Res	Reg2	Res	Reg2	Coef	Var	RSS2	MRSS	Res	Reg2	Res	Reg2	Coef	Var	RSS2	MRSS
Example 1	6	9.8	11.2	9.9	9.5	14.4	13.2	11.3	7.4	30.7	31.4	30.8	29.6	32.2	30.9	28.9	25.2
	10	37.1	22.0	33.8	24.1	32.5	31.9	28.9	23.4	59.0	43.8	55.7	46.2	53.8	53.9	50.4	45.4
	14	70.6	37.8	59.6	46.1	91.2	98.5	55.0	47.1	81.2	58.2	72.6	63.6	94.6	98.2	69.6	64.5
Example 2	6	10.2	11.6	10.3	9.9	15.1	13.8	11.6	7.8	31.3	31.9	31.4	30.2	33.1	31.8	29.5	25.9
	10	37.5	22.2	34.4	24.6	32.9	32.3	29.4	24.0	59.4	44.1	56.2	46.8	54.2	54.3	50.8	46.0
	14	71.0	38.2	60.1	46.6	91.5	98.8	55.3	47.5	81.5	58.4	73.0	64.0	94.7	98.4	69.8	64.9
Example 3	6	11.1	12.3	11.3	10.8	16.4	15.1	12.9	9.0	32.6	33.0	32.9	31.6	35.0	33.6	31.4	28.1
	10	39.3	23.4	36.2	26.2	33.8	33.6	30.9	25.7	60.8	45.3	57.9	48.4	55.1	55.5	52.3	47.8
	14	73.0	39.5	62.6	48.8	93.2	100.3	56.7	49.9	82.7	59.6	75.0	65.9	95.6	99.1	70.8	66.9
Example 4	6	15.3	15.6	15.6	14.7	22.1	20.8	18.3	13.8	38.3	37.3	38.6	37.0	41.5	40.4	38.6	35.7
	10	45.4	27.5	43.2	32.0	37.0	37.8	35.4	31.6	65.5	49.5	63.6	54.2	58.2	59.3	56.6	53.8
	14	78.9	43.5	70.9	56.0	98.9	105.3	61.0	57.5	86.3	62.8	80.7	71.5	98.3	101.4	73.9	72.8

Table 4. Cont.

Large Sample Size		MSE ($\times 10^2$)								RMSE ($\times 10^2$)							
Example	l	Lasso		MCP						Lasso		MCP					
		Res	Reg2	Res	Reg2	Coef	Var	RSS2	MRSS	Res	Reg2	Res	Reg2	Coef	Var	RSS2	MRSS
Example 1	6	2.9	1.9	1.8	1.1	1.6	1.4	1.4	0.5	16.3	12.8	12.4	9.9	9.5	9.0	9.7	6.5
	10	6.8	3.9	1.6	1.5	2.5	2.6	1.7	0.6	24.7	18.4	11.8	11.3	11.8	12.0	10.5	7.0
	14	11.7	6.8	1.6	1.6	12.6	17.9	3.2	0.9	32.2	24.3	11.8	11.9	25.3	29.9	14.1	8.1
Example 2	6	3.1	2.1	2.0	1.3	1.8	1.5	1.5	0.6	17.0	13.5	13.2	10.6	10.2	9.6	9.9	7.1
	10	7.2	4.2	1.8	1.7	2.6	2.8	1.8	0.7	25.4	19.2	12.6	12.2	12.4	12.6	10.8	7.6
	14	12.1	7.2	1.8	1.8	13.1	18.3	3.4	1.0	32.8	25.0	12.7	12.7	26.2	30.7	14.6	8.8
Example 3	6	3.8	2.6	2.6	1.7	2.2	1.9	1.7	0.9	18.8	15.2	15.1	12.4	12.2	11.5	11.0	8.9
	10	8.6	5.3	2.7	2.5	3.6	3.8	2.4	1.3	28.0	21.9	15.7	15.2	16.0	16.2	13.1	10.7
	14	14.4	9.1	3.2	3.1	15.6	21.4	4.5	1.9	36.0	28.5	16.9	16.9	31.0	35.7	17.7	13.0
Example 4	6	8.1	6.1	6.5	5.0	6.0	5.6	4.2	3.9	27.6	23.5	24.7	21.7	22.8	22.2	19.2	19.3
	10	15.7	11.4	8.5	8.1	11.0	11.4	7.3	6.4	38.3	32.3	28.2	27.5	31.6	32.3	25.4	24.9
	14	23.8	17.6	10.7	10.6	29.9	37.4	12.6	9.4	47.0	40.0	31.7	31.5	50.0	55.2	32.6	29.9

Conducting a sensitivity analysis with various variance and sparsity parameters, the outcomes demonstrate the benefits of the MRSS algorithm in terms of robustness, stability, and accuracy. As the variance increases from 1 to 40, the MSE and RMSE remain consistent for distinct examples and sample sizes. This demonstrates that our MRSS algorithm is remarkably resilient to variability and maintains excellent stability. As the level of sparsity decreases (from Examples 1–4, or from $l = 6$ to 14), it is noticeable that the MSE and RMSE of the MRSS algorithm increase, but remain within the same order of magnitude. Even the optimal MRSS algorithms undergo a significant rise in MSE and RMSE for Example 4 and $l = 14$, as an escalation of non-sparse and intricate controlling variables brings forth certain systematic errors.

5. Real Data Analysis

A distinguishing feature of financial markets is the observed correlation among the price movements of various financial assets. A prevalent feature entails the existence of a substantial cross-correlation between stock returns' simultaneous time evolution [29]. In numerous instances, a strong correlation does not necessarily imply a significant direct relationship. For instance, two stocks in the same market may be subject to shared macroeconomic or investor psychology influences. Therefore, to examine the direct correlation between these stocks, it is necessary to eliminate the common drivers represented by the market index. The Pcor meets this requirement by assessing the direct relationship between the two stocks after removing the market impacts of controlling variables. When accurately estimating the Pcor, it is possible to evaluate the impact of diverse factors (e.g., economic sectors, other markets, or macroeconomic factors) on a specific stock. The resulting partial correlation data may be utilised in fields, such as stock market risk management, stock portfolio optimisation, and financial control [7,8]. Moreover, the Pcor can also indicate the interdependence and influence of industries in the context of global integration. These techniques for analysing Pcor can provide valuable information on the correlations between different assets and different sectors of the economy, as they are generalisable and can be applied to other asset types and cross-asset relationships in financial markets. This information is beneficial for practitioners and policymakers.

We chose 100 stocks with substantial market capitalisation and robust liquidity from the Shanghai Stock Exchange (SSE) market. These stocks can comprehensively represent the overall performance of listed stock prices in China's A-share market. We then downloaded their daily adjusted closing prices from Yahoo Finance from January 2018 to August 2023 and removed the missing data. Here, a sufficient sample size of $n = 1075$ was chosen to ensure the effectiveness of algorithms and limit the bias in Pcor estimation. For each pair of the 100 stocks, we estimate their Pcor by setting the remaining stocks as the corresponding

controlling variables and construct the estimated Pcor matrix. The Pcor matrix shows the better internal correlation between two stocks after removing the influence of the stock market.

Figure 3 presents the estimated Pcor matrices for 100 stocks from SSE markets using MCP.Reg2, MCP.Var and MRSS algorithms. Blue signifies $Pcor = 1$, while red represents $Pcor = -1$. Whilst the MCP.Coef, MCP.Var, and RSS2 algorithms all estimate Pcor as 0 when true Pcor approaches 0, our proposed MRSS algorithm resembles the MCP.Reg2, which estimates an accurate Pcor for weak partial correlation. Thus, the MRSS is capable of effectively estimating weak partial correlations. When dealing with high Pcor values and strong partial correlation, we find that the MCP.Var algorithm overestimates Pcor as a result of the divergence in stock prices. For two stocks with a higher stock price, the Pcor estimated by the Var algorithm to be overestimated or even most at 1. MRSS effectively solves this problem. Notably, as a result of incorporating the MCP.Var algorithm, the MRSS algorithm amplifies certain partial correlations that are not significant by MCP.Reg2. These results can also be seen in Table 5. The MRSS estimates these correlations to be stronger partial correlations resulting in improved clarity in the partial correlations.

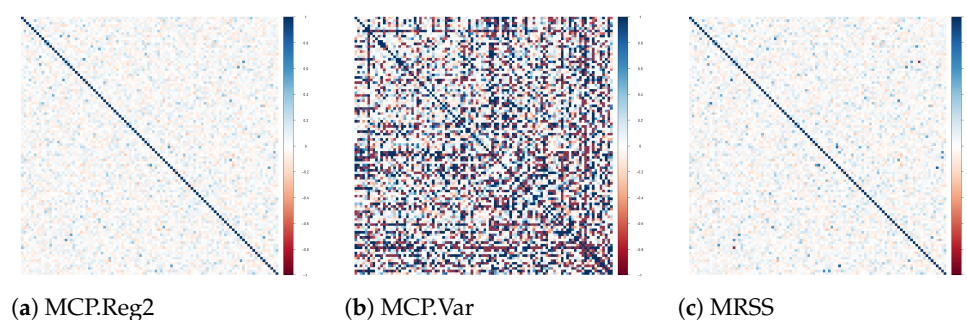


Figure 3. Estimated Pcor matrix of 100 HKSE stocks, with blue representing $Pcor = 1$ and red representing $Pcor = -1$.

Figure 4 shows the stocks' Pcor network for the top-100 and top-50 pairs of Pcor estimates by the MRSS algorithm from 100 SSE stocks. The node represents the stock, coloured with its sector. The edge thickness represents the Pcor estimate between two nodes, with the thick-edge $Pcor > 0.4$ and the thin-edge $Pcor < 0.4$. Table 5 shows the stock pairs with their sector and Pcor estimates for all the MRSS estimated $Pcor > 0.4$ from 100 SSE stocks, and Table 6 shows the corresponding stock pairs with their company name, business, and sector. Here, we use industry classifications from the Global Industry Classification Standard (GICS) with Communication Services, Consumer Discretionary (C.D.), Consumer Staples, Energy, Financials, Health Care, Industrials, Information Technology (I.T.), Materials, Real Estate and Utilities. We find that two stocks connected in the partial correlation network with a high Pcor are almost in the same sector and operate in the same business. In addition, high Pcor values may indicate shareholding relationships between companies. For instance, the highly correlated 601398–601939–601288–601988–601328 (financials) are all state-controlled banks that do not have a direct high Pcor link with the city banks 601009–601166 (financials). And, stocks that do not belong to the same industry under a high Pcor may have certain other links behind them, such as 601519 (I.T.)–601700 (industrials) having a common major shareholder. After stripping out the other factors influencing the market, Pcor represents the inherent and intrinsic correlation between two stocks because they are in the same sector.

As societies become increasingly integrated, the productive activities of different industries become interdependent and interact with each other. Categorising a company into only one industry does not reflect its overall performance and associated risks. Many listed companies in the stock market belong to conglomerates and operate in different industry sectors, so it is natural for the performance of these companies to be affected

by multiple industries. Therefore, we will also find that Pcor, apart from showing the correlation between industries, will also reveal the correlation between two industries that are linked together by two stocks in different industries. For example, the partial correlation between the Bank of Communications (601328) and PetroChina (601857) with $Pcor = 0.258$ links the Energy (600028–601857 in orange) and Financial (601398–601939–601288–601988–601328 in dark blue) sectors of state-owned assets.

Overall, the MRSS algorithm amalgamates the characteristics of MCP.Reg2 and MCP.Var, enhancing the estimation of strong partial correlations, while effectively estimating those weak partial correlations, ultimately revealing the stock correlations.

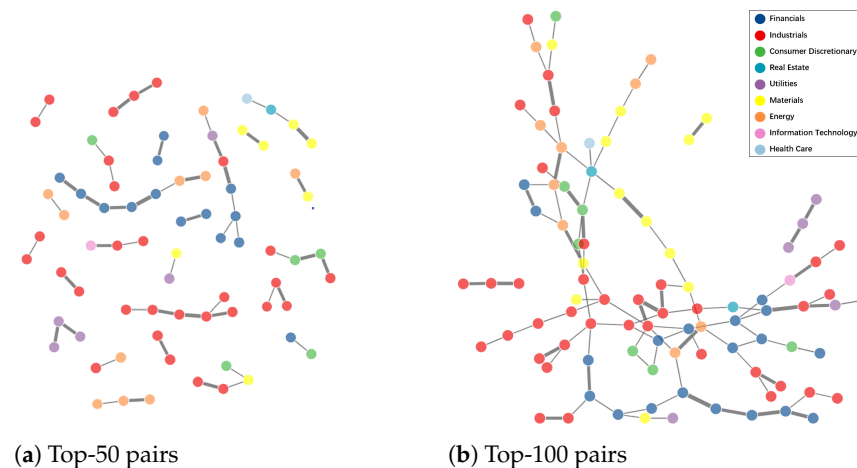


Figure 4. Stocks' Pcor network for the top-100 and top-50 pairs of Pcor estimates by MRSS algorithm from 100 SSE stocks. The node represents the stock, coloured with its sector. The edge thickness represents the Pcor estimate between two nodes, with the thick edge $Pcor > 0.4$ and the thin edge $Pcor < 0.4$.

Table 5. Stock pairs with their sector and Pcor estimates for all the MRSS estimated $Pcor > 0.4$ by different algorithms from 100 SSE stocks.

Stock1		Stock2		Lasso				MCP			
Symbol	Sector	Symbol	Sector	Res	Reg2	Res	Reg2	Coef	Var	RSS2	MRSS
601398	Financials	601939	Financials	0.526	0.526	0.535	0.535	0.533	0.840	0.527	0.840
600022	Materials	601005	Materials	0.569	0.569	0.581	0.581	0.580	0.769	0.590	0.769
601186	Industrials	601390	Industrials	0.589	0.589	0.566	0.566	0.587	0.748	0.584	0.748
600012	Industrials	601099	Financials	0.405	0.405	0.399	0.399	0.404	0.697	0.414	0.697
601288	Financials	601988	Financials	0.473	0.473	0.476	0.476	0.490	0.646	0.473	0.646
600028	Energy	601857	Energy	0.550	0.550	0.545	0.545	0.569	0.607	0.534	0.607
601098	C.D.	601801	C.D.	0.468	0.468	0.474	0.474	0.475	0.606	0.476	0.606
601328	Financials	601988	Financials	0.357	0.357	0.316	0.316	0.369	0.600	0.322	0.600
600017	Industrials	601880	Industrials	0.372	0.372	0.382	0.382	0.384	0.574	0.394	0.574
600026	Industrials	601872	Industrials	0.590	0.590	0.572	0.573	0.590	1	0.593	0.573
601866	Industrials	601919	Industrials	0.552	0.552	0.545	0.545	0.562	1	0.554	0.545
601179	Industrials	601390	Industrials	0.291	0.291	0.275	0.275	0.285	0.543	0.284	0.543
600011	Utilities	600021	Utilities	0.535	0.535	0.522	0.522	0.535	0.543	0.529	0.543
601333	Industrials	601801	C.D.	0.526	0.526	0.526	0.526	0.525	1	0.528	0.526
600011	Utilities	600027	Utilities	0.514	0.514	0.517	0.517	0.514	1	0.501	0.517
601088	Energy	601666	Energy	0.289	0.289	0.326	0.326	0.359	0.492	0.349	0.492
601288	Financials	601398	Financials	0.353	0.353	0.338	0.338	0.349	0.491	0.339	0.491
601168	Materials	601899	Materials	0.515	0.515	0.490	0.490	0.535	1	0.497	0.490
601186	Industrials	601618	Industrials	0.260	0.260	0.245	0.245	0.250	0.488	0.249	0.488
600018	Industrials	601018	Industrials	0.480	0.480	0.483	0.483	0.486	1	0.485	0.483
600008	Utilities	600012	Industrials	0.319	0.319	0.309	0.309	0.312	0.436	0.304	0.436

Table 5. Cont.

Stock1		Stock2		Lasso				MCP			
Symbol	Sector	Symbol	Sector	Res	Reg2	Res	Reg2	Coef	Var	RSS2	MRSS
601009	Financials	601166	Financials	0.300	0.300	0.298	0.298	0.303	0.427	0.307	0.427
600020	Industrials	601177	Industrials	0.431	0.431	0.430	0.430	0.421	0.422	0.429	0.422
601001	Energy	601137	Materials	0.268	0.268	0.261	0.261	0.283	0.421	0.260	0.421
601519	I.T.	601700	Industrials	0.224	0.224	0.219	0.219	0.217	0.410	0.203	0.410
600017	Industrials	601008	Industrials	0.404	0.404	0.405	0.405	0.403	0.979	0.407	0.405
601318	Financials	601601	Financials	0.414	0.414	0.403	0.403	0.414	1	0.414	0.403

Table 6. Stock pairs with their company name, business, and sector for all the MRSS estimated Pcor > 0.4 from 100 SSE stocks.

Symbol	Company	Business	Sector	Symbol	Company	Business	Sector
601098	South central Media	Media	C.D.	601801	Anhui Xinhua Media	publishing	C.D.
600028	Sinopec	Refining and Trading	Energy	601857	PetroChina	Refining and Trading	Energy
601088	China Shenhua Energy	Coal Mining	Energy	601666	Pingdingshan Tianan Coal Mining	Coal Mining	Energy
601001	Datong Coal Industry	Coal Mining	Energy	601137	Ningbo Boway Alloy Material	Industrial Metals	Materials
601398	Industrial and Commercial Bank of China	Banks	Financials	601939	China Construction Bank	Banks	Financials
601288	Agricultural Bank of China	Banks	Financials	601988	Bank of China	Banks	Financials
601328	Bank of Communications	Banks	Financials	601988	Bank of China	Banks	Financials
601288	Agricultural Bank of China	Banks	Financials	601398	Industrial and Commercial Bank of China	Banks	Financials
601009	Bank of Nanjing	Banks	Financials	601166	Industrial Bank of China	Banks	Financials
601318	Ping An Insurance of China	Insurance	Financials	601601	China Pacific Insurance	Insurance	Financials
601186	China Railway Construction	Infrastructure	Industrials	601390	China Railway Engineering Group	Infrastructure	Industrials
600012	Anhui Expressway	Railway and Highway	Industrials	601099	China Pacific Insurance	certificate	Financials
600017	Rizhao Port	Shipping Ports	Industrials	601880	Dalian Port	Shipping Ports	Industrials
600026	China Shipping Development	Shipping Ports	Industrials	601872	China Merchants Energy Shipping	Shipping Ports	Industrials
601866	China Shipping Container Lines	Shipping Ports	Industrials	601919	China Ocean Shipping	Shipping Ports	Industrials
601179	China Xidian Electric	Grid Equipment	Industrials	601390	China Railway Engineering	Infrastructure	Industrials
601333	Guangshen Railway	Railway and Highway	Industrials	601801	Anhui Xinhua Media	publishing	C.D.
601186	China Railway Construction	Infrastructure	Industrials	601618	Metallurgical Corporation of China	Professional Engineering	Industrials
600018	Shanghai International Port Group	Shipping Ports	Industrials	601018	Ningbo Port	Shipping Ports	Industrials
600020	Zhongyuan Expressway	Railway and Highway	Industrials	601177	Hangzhou Advance Gearbox	machine	Industrials
600017	Rizhao Port	Shipping Ports	Industrials	601008	Lianyungang Port	Shipping Ports	Industrials
601519	Shanghai DZH	Software Development	I.T.	601700	Changshu Fengfan Power Equipment	Grid Equipment	Industrials
600022	Jinan Iron and Steel	Plain Steel	Materials	601005	Chongqing Iron and Steel	Plain Steel	Materials
601168	Western Mining	Industrial Metals	Materials	601899	Zijin Mining	Industrial Metals	Materials
600011	Huaneng Power International	Electricity	Utilities	600021	Shanghai Electric Power	Electricity	Utilities
600011	Huaneng Power International	Electricity	Utilities	600027	Huadian Power International	Electricity	Utilities
600008	Beijing Capital	Water	Utilities	600012	Anhui Expressway	Railway and Highway	Industrials

6. Conclusions

This paper presents a novel minimum residual sum of squares (MRSS) algorithm for estimating partial correlation coefficients. Its purpose is to reduce the estimation bias of positive partial correlation coefficients in high-dimensional settings under both sparse and non-sparse conditions. The MRSS algorithm is effective in mitigating a Pcor estimation bias by synergistically harnessing the strengths of the Coef, Reg2, and Var algorithms. We also discuss the MRSS algorithm mathematical foundation and its performance in various scenarios compared to some existing algorithms. Through rigorous simulations and real data analysis, it becomes evident that the MRSS algorithm consistently outperforms its constituent and listed algorithms, particularly in challenging environments characterised by non-sparse conditions and high dimensionality. The sensitivity analysis with variance and sparsity parameters demonstrate the robustness, stability, and precision advantages of the MRSS algorithm. Further evidence of the effectiveness of the MRSS algorithm in the correlation analysis of stock data is provided by real data analyses.

7. Future Work

Our proposed MRSS algorithm combines the benefits of two existing algorithms by reducing the total squared residuals and enhancing the accuracy of Pcor estimation. In upcoming studies, we may explore the integration of additional algorithms by minimising the RSS to achieve a greater amalgamation of benefits from various algorithms and improve the

estimation accuracy of the integrated algorithm. Reducing the computational complexity of our minimised RSS integration algorithm to decrease computing time represents a core issue in future research. Additionally, conducting in-depth theoretical research on MRSS algorithms, including a proof analysis of consistency and convergence, will be an essential direction for our next steps. Further refinement of theoretical proofs and an in-depth investigation of error convergence speed may uncover reasons for the systematic estimation bias that cannot be ignored when P_{cor} is positive in all current algorithms. Meanwhile, expanding the use of the MRSS algorithm to a wider range of fields is a focal point of our future research. Concerning financial data, we intend to thoroughly examine the biased correlations between financial data besides stocks and advise on relevant policies.

Author Contributions: Conceptualisation and methodology, J.Y. and M.Y.; software, G.B. and J.Y.; validation and formal analysis, G.B.; data curation, writing—original draft preparation, review and editing, and visualisation, J.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Doctoral Foundation of Yunnan Normal University (Project No.2020ZB014) and the Youth Project of Yunnan Basic Research Program (Project No.202201AU070051).

Data Availability Statement: The authors confirm that the data supporting the findings of this study are available within the article.

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

Appendix A

Tables for the mean of $MSE (\times 10^2)$ and $RMSE (\times 10^2)$ of estimated P_{cors} of real $P_{cor} = 0, 0.05, \dots, 0.95$ with $\sigma^2 = 1$, $n = 50, 100, 200, 400$, $p = 200, 500, 1000, 2000, 4000$ and the numbers of high correlation controlling variables $l = 6, 14$ in Examples 1–4.

Table A1. The mean of $MSE (\times 10^2)$ and $RMSE (\times 10^2)$ for the estimated P_{cors} of real $P_{cor} = 0, 0.05, \dots, 0.95$ with $l = 6$, $\sigma^2 = 1$, $n = 50, 100, 200, 400$ and $p = 200, 500, 1000, 2000, 4000$ in Examples 1–4.

n	p	$MSE (\times 10^2)$								$RMSE (\times 10^2)$							
		Lasso				MCP				Lasso				MCP			
Example 1	Method	Res	Reg2	Res	Reg2	Coef	Var	RSS2	MRSS	Res	Reg2	Res	Reg2	Coef	Var	RSS2	MRSS
50	200	10.0	11.1	10.6	10.2	15.8	14.3	12.1	8.1	31.2	31.9	32.0	31.0	35.3	33.8	31.5	27.6
	500	11.4	14.3	11.7	12.3	20.5	19.1	16.2	11.3	33.2	36.0	33.5	33.9	40.2	39.1	36.4	32.6
	1000	11.8	15.8	12.0	13.1	22.5	21.5	18.6	12.8	33.8	37.6	34.1	35.0	42.0	41.2	39.0	34.7
100	500	7.4	6.6	7.2	5.8	7.3	6.2	5.3	3.1	26.7	24.5	26.4	23.4	22.3	20.9	19.6	16.3
	1000	8.6	8.6	8.5	7.2	9.2	8.0	6.8	4.0	28.9	27.7	28.6	26.0	25.2	23.8	22.2	18.4
	2000	9.6	10.7	9.4	8.7	11.2	10.0	8.5	5.4	30.4	30.9	30.1	28.4	28.1	26.8	25.0	21.4
200	500	3.1	1.7	2.0	1.1	1.8	1.5	1.6	0.6	17.3	12.6	14.1	10.6	10.7	10.0	10.7	7.2
	1000	4.1	2.6	2.9	1.6	2.4	2.0	2.1	0.8	19.8	15.6	16.6	12.5	12.3	11.6	12.2	8.1
	2000	5.0	3.7	3.7	2.2	3.1	2.6	2.5	1.0	21.9	18.3	19.0	14.4	13.9	13.0	13.3	9.0
400	1000	1.3	0.7	0.5	0.4	0.6	0.5	0.6	0.3	11.3	8.3	7.0	6.3	6.0	5.8	6.5	4.6
	2000	1.7	1.1	0.7	0.5	0.8	0.7	0.8	0.3	12.9	10.1	8.1	7.2	6.7	6.4	7.2	5.0
	4000	2.2	1.5	0.9	0.7	0.9	0.8	0.9	0.4	14.5	12.1	9.3	8.1	7.4	7.0	8.1	5.4

Table A1. Cont.

<i>n</i>	<i>p</i>	MSE ($\times 10^2$)								RMSE ($\times 10^2$)							
		Lasso		MCP				Lasso				MCP					
		Res	Reg2	Res	Reg2	Coef	Var	RSS2	MRSS	Res	Reg2	Res	Reg2	Coef	Var	RSS2	MRSS
Example 2																	
50	200	10.5	11.5	11.0	10.6	16.5	15.0	12.4	8.6	31.8	32.5	32.5	31.6	36.4	35.0	31.9	28.7
	500	11.8	14.8	12.1	12.7	21.6	19.9	16.7	11.8	33.7	36.5	34.1	34.4	41.3	39.9	37.1	33.3
	1000	12.2	16.0	12.4	13.5	23.1	22.0	19.2	13.1	34.3	37.9	34.6	35.5	42.5	41.7	39.7	35.1
100	500	7.8	6.9	7.6	6.1	7.8	6.7	5.5	3.3	27.4	25.0	27.1	23.9	23.4	21.9	20.1	16.8
	1000	9.0	9.0	8.9	7.6	9.7	8.5	7.1	4.4	29.5	28.4	29.2	26.6	26.3	24.8	22.8	19.5
	2000	10.0	11.0	9.8	9.0	11.6	10.5	8.8	5.7	31.0	31.4	30.7	29.0	28.9	27.6	25.5	22.3
200	500	3.3	1.8	2.3	1.3	2.0	1.7	1.7	0.7	18.0	13.2	14.9	11.4	11.3	10.6	11.1	7.7
	1000	4.4	2.9	3.2	1.8	2.7	2.3	2.1	0.9	20.5	16.2	17.5	13.3	13.3	12.4	12.4	8.6
	2000	5.3	3.9	4.1	2.4	3.4	2.9	2.6	1.1	22.7	18.9	19.8	15.1	14.8	13.8	13.6	9.6
400	1000	1.5	0.8	0.6	0.5	0.7	0.6	0.6	0.3	12.1	9.1	7.7	7.0	6.5	6.3	6.7	5.1
	2000	1.9	1.2	0.8	0.6	0.8	0.7	0.8	0.4	13.7	10.9	8.9	8.0	7.2	6.9	7.4	5.5
	4000	2.4	1.7	1.1	0.8	1.0	0.9	1.0	0.4	15.3	12.8	10.2	8.9	8.0	7.6	8.3	6.0
Example 3																	
50	200	11.4	12.3	12.0	11.5	18.1	16.6	13.9	10.0	33.1	33.5	34.0	32.9	38.1	36.8	34.2	30.8
	500	12.7	15.5	13.0	13.7	22.9	21.4	18.2	13.2	34.9	37.3	35.4	35.8	42.4	41.4	38.9	35.2
	1000	13.2	17.0	13.5	14.5	24.6	23.4	20.6	14.5	35.7	39.1	36.1	36.8	43.8	42.9	40.9	36.8
100	500	8.6	7.7	8.6	7.0	8.7	7.6	6.4	4.0	28.8	26.3	28.7	25.5	25.5	24.0	21.9	18.9
	1000	9.9	9.5	9.8	8.5	10.9	9.7	7.9	5.2	30.9	29.2	30.7	28.1	28.4	26.9	24.4	21.7
	2000	10.9	11.8	10.8	9.9	13.1	11.9	10.1	6.8	32.5	32.5	32.2	30.5	31.5	30.0	27.8	24.9
200	500	4.0	2.3	2.9	1.7	2.4	2.1	1.9	0.9	19.7	14.8	16.7	13.0	13.3	12.5	11.8	9.3
	1000	5.1	3.4	3.9	2.4	3.2	2.8	2.5	1.2	22.3	17.7	19.4	15.0	15.3	14.3	13.6	10.5
	2000	6.2	4.6	4.9	3.1	4.1	3.5	3.1	1.5	24.4	20.4	21.7	17.0	17.0	15.9	15.0	11.7
400	1000	2.0	1.2	1.0	0.8	0.9	0.9	0.8	0.5	13.9	10.9	9.7	8.9	8.3	8.0	7.7	6.7
	2000	2.5	1.7	1.2	1.0	1.1	1.0	0.9	0.6	15.6	12.8	10.9	9.9	9.1	8.8	8.3	7.3
	4000	3.1	2.2	1.5	1.2	1.4	1.2	1.2	0.7	17.2	14.6	12.2	10.8	10.0	9.6	9.3	7.9
Example 4																	
50	200	13.9	14.2	14.7	14.0	22.4	20.7	17.7	13.3	36.6	36.0	37.6	36.3	42.3	41.0	39.0	35.5
	500	16.5	18.4	16.9	17.2	27.5	26.4	23.8	17.5	39.8	40.7	40.3	40.1	45.9	45.2	44.1	40.3
	1000	18.0	20.5	18.2	18.8	28.6	28.0	26.1	19.2	41.6	42.9	41.9	42.0	46.5	46.3	45.7	42.3
100	500	12.4	10.8	12.4	10.3	13.7	12.4	10.2	7.8	34.5	31.1	34.5	31.0	33.4	31.9	29.2	27.2
	1000	14.6	13.4	14.6	12.7	18.3	16.8	14.0	10.8	37.5	34.8	37.5	34.5	38.7	37.3	34.6	32.0
	2000	16.6	16.5	16.6	15.4	22.1	20.6	17.9	14.2	40.0	38.5	40.0	38.0	42.1	41.0	39.1	36.7
200	500	7.0	4.7	5.9	4.1	5.1	4.7	3.6	2.9	26.0	20.8	23.8	19.9	21.3	20.5	17.7	16.9
	1000	9.4	6.8	8.1	5.7	7.3	6.6	5.1	4.1	30.1	24.8	27.9	23.2	25.1	24.2	21.1	20.0
	2000	11.6	8.9	10.3	7.4	9.4	8.7	6.8	5.5	33.4	28.2	31.4	26.4	28.4	27.5	24.3	23.1
400	1000	5.2	3.9	3.7	3.3	3.5	3.3	2.4	2.7	22.3	19.2	18.8	17.7	18.1	17.8	15.2	16.2
	2000	6.8	5.5	4.9	4.3	4.6	4.4	3.2	3.6	25.5	22.7	21.6	20.3	20.7	20.3	17.4	18.7
	4000	8.5	7.0	6.2	5.4	5.9	5.6	4.2	4.5	28.5	25.4	24.4	22.6	23.4	23.0	19.6	20.9

Table A2. The mean of MSE ($\times 10^2$) and RMSE ($\times 10^2$) for estimated Pcors of real Pcor = 0, 0.05, ..., 0.95 with $l = 14$, $\sigma^2 = 1$, $n = 50, 100, 200, 400$ and $p = 200, 500, 1000, 2000, 4000$ in Examples 1–4.

n	p	MSE ($\times 10^2$)								RMSE ($\times 10^2$)							
		Lasso				MCP				Lasso				MCP			
Example 1	Method	Res	Reg2	Res	Reg2	Coef	Var	RSS2	MRSS	Res	Reg2	Res	Reg2	Coef	Var	RSS2	MRSS
50	200	68.5	36.8	70.4	55.2	87.9	96.0	61.5	57.4	80.9	58.2	81.9	72.4	92.9	97.1	76.3	74.3
	500	91.0	49.5	91.2	71.9	86.8	93.2	77.3	73.7	93.4	67.7	93.4	82.7	92.3	95.7	86.2	84.4
	1000	100.9	56.4	98.8	78.3	83.5	89.3	83.0	79.0	98.3	72.5	97.3	86.3	90.5	93.6	89.6	87.2
100	500	39.4	21.4	18.2	14.2	91.1	103.2	30.5	14.4	61.3	44.1	41.5	36.6	94.9	100.6	49.9	36.8
	1000	54.4	28.3	30.9	22.9	97.9	104.7	35.7	23.3	72.1	50.6	54.1	46.4	98.1	101.3	54.6	46.9
	2000	69.6	34.6	47.9	34.4	99.8	104.6	41.9	34.8	81.5	55.9	67.4	56.9	98.9	101.1	60.8	57.5
200	500	10.6	5.3	1.7	1.7	7.1	13.2	2.8	0.8	31.8	22.5	13.0	12.8	21.9	28.4	14.1	8.2
	1000	16.4	9.4	2.5	2.4	20.9	32.3	4.5	1.3	39.5	29.6	15.6	15.4	40.8	49.8	18.2	10.5
	2000	23.4	14.3	3.7	3.7	43.6	56.7	8.4	2.4	47.3	36.5	18.8	18.7	63.4	72.4	25.9	14.9
400	1000	4.4	2.1	0.5	0.5	0.9	0.8	0.9	0.2	20.5	14.2	6.9	6.9	7.3	6.9	7.8	4.5
	2000	6.4	3.7	0.6	0.7	1.2	1.2	1.2	0.3	24.8	19.0	7.8	8.0	8.3	8.6	8.8	4.9
	4000	9.0	6.0	0.8	0.9	1.8	3.2	1.5	0.4	29.3	24.2	9.0	9.4	10.3	13.4	9.8	5.6
Example 2																	
50	200	68.6	37.0	70.5	56.0	88.1	95.9	61.4	58.1	80.9	58.4	82.0	73.0	93.0	97.1	76.3	74.9
	500	91.5	49.9	91.3	71.9	87.3	94.4	77.6	73.9	93.6	68.2	93.4	82.7	92.5	96.3	86.3	84.5
	1000	101.0	56.9	99.2	78.2	83.5	89.4	83.5	78.9	98.4	72.7	97.5	86.3	90.4	93.7	89.9	87.2
100	500	39.9	21.7	18.8	14.7	92.2	103.2	30.7	14.8	61.7	44.4	42.2	37.2	95.4	100.6	50.1	37.4
	1000	55.0	28.7	31.9	23.6	98.0	104.9	36.0	23.8	72.4	51.0	55.0	47.1	98.1	101.3	54.9	47.5
	2000	70.0	34.8	48.8	35.2	99.8	104.8	42.4	35.8	81.8	56.0	68.1	57.6	98.9	101.2	61.2	58.3
200	500	11.0	5.6	1.9	1.9	7.6	13.8	3.0	0.9	32.4	23.1	13.8	13.5	22.8	29.3	14.5	8.7
	1000	16.8	9.8	2.8	2.7	22.1	32.7	4.9	1.4	40.0	30.2	16.4	16.2	42.5	50.6	19.3	11.3
	2000	23.9	14.9	4.0	4.0	44.7	57.8	8.8	2.7	47.8	37.2	19.6	19.5	64.4	73.0	26.6	16.1
400	1000	4.7	2.3	0.6	0.6	1.1	0.9	1.0	0.3	21.2	14.9	7.6	7.7	7.9	7.6	8.1	5.0
	2000	6.8	4.0	0.8	0.8	1.3	1.4	1.3	0.4	25.4	19.8	8.7	8.8	8.8	9.2	9.1	5.5
	4000	9.4	6.5	1.0	1.1	1.9	3.5	1.6	0.5	30.0	25.0	9.8	10.2	10.9	14.2	10.2	6.2
Example 3																	
50	200	70.3	38.1	72.5	57.6	89.4	97.2	63.2	60.2	81.9	59.3	83.1	74.0	93.7	97.7	77.5	76.2
	500	93.1	51.3	92.8	73.4	88.4	95.0	79.0	75.6	94.4	69.1	94.2	83.6	93.1	96.6	87.1	85.5
	1000	102.5	58.2	100.7	80.0	85.1	91.0	85.3	80.3	99.1	73.7	98.2	87.3	91.3	94.5	90.7	87.8
100	500	42.4	23.2	21.9	17.3	94.4	105.3	31.9	17.5	63.6	46.0	45.5	40.4	96.5	101.6	51.1	40.7
	1000	57.6	30.2	35.7	26.8	100.1	106.7	37.0	27.2	74.1	52.2	58.2	50.3	99.1	102.1	55.8	50.8
	2000	72.2	36.0	52.3	37.9	101.8	106.5	43.7	38.5	83.1	57.1	70.6	59.8	99.8	102.0	62.3	60.6
200	500	13.0	7.3	3.3	3.2	10.4	17.5	4.0	1.7	35.3	26.2	17.7	17.4	28.7	35.5	17.6	12.8
	1000	19.3	11.9	4.4	4.3	26.0	37.6	6.3	2.6	42.9	33.3	20.5	20.3	47.3	55.9	22.2	15.6
	2000	26.9	17.1	5.9	5.9	49.8	64.1	10.8	4.3	50.7	39.8	23.9	23.7	68.6	77.7	29.5	20.3
400	1000	6.3	3.7	1.4	1.5	1.8	1.7	1.6	0.8	24.6	19.0	11.9	11.9	12.1	11.7	11.1	9.0
	2000	8.8	6.0	1.8	1.8	2.2	2.3	1.9	1.0	29.1	24.0	13.0	13.2	13.3	14.0	12.4	9.8
	4000	11.8	8.8	2.1	2.2	3.3	5.2	2.4	1.2	33.6	29.1	14.3	14.7	16.1	19.6	13.6	10.7
Example 4																	
50	200	73.4	40.2	76.7	61.4	92.4	99.9	66.4	64.5	83.8	61.0	85.5	76.4	95.2	99.0	79.5	78.9
	500	96.6	53.6	96.6	77.3	92.5	99.2	82.7	80.1	96.3	70.8	96.1	85.8	95.2	98.6	89.1	88.0
	1000	106.6	61.8	105.8	85.4	90.4	96.4	91.1	86.4	101.1	76.0	100.7	90.2	94.1	97.2	93.8	91.2
100	500	49.2	27.9	31.0	25.1	102.1	111.1	34.7	25.4	68.5	50.5	54.2	48.7	100.2	104.2	53.7	49.1
	1000	66.0	35.5	48.4	37.1	106.9	112.1	41.0	37.8	79.4	56.8	67.8	59.3	102.3	104.6	59.8	60.0
	2000	81.7	41.8	66.8	49.8	109.2	112.9	50.4	50.8	88.5	61.9	79.9	68.7	103.2	104.9	67.8	69.7
200	500	18.9	12.2	8.3	8.1	20.8	29.7	8.9	6.3	42.4	33.8	28.2	27.8	44.0	51.6	27.8	24.8
	1000	28.3	19.4	11.7	11.5	45.0	59.2	14.7	9.9	51.9	42.4	33.5	33.0	66.1	75.3	35.1	31.0
	2000	38.8	26.4	16.2	15.8	76.1	92.8	23.6	15.2	60.9	49.3	39.3	38.7	87.0	95.7	43.2	38.1
400	1000	13.2	10.3	6.9	6.9	8.0	7.8	6.4	6.0	35.5	31.2	25.7	25.8	27.5	27.5	24.9	24.3
	2000	18.8	15.7	9.2	9.4	11.6	12.5	9.7	8.2	42.4	38.5	29.7	29.9	33.4	34.9	30.6	28.4
	4000	24.9	21.4	11.7	12.0	18.3	22.4	12.3	10.8	48.8	44.9	33.6	34.0	42.0	46.4	34.0	32.5

References

1. Tabachnick; Barbara, G.; Linda, S.F.; Jodie, B.U. *Using Multivariate Statistics*, 6th ed.; Pearson: Boston, MA, USA, 2013.
2. Huang, Y.; Chang, X.; Zhang, Y.; Chen, L.; Liu, X. Disease characterization using a partial correlation-based sample-specific network. *Brief. Bioinform.* **2021**, *22*, bbaa062. [[CrossRef](#)]
3. Peng, J.; Wang, P.; Zhou, N.; Zhu, J. Partial correlation estimation by joint sparse regression models. *J. Am. Stat. Assoc.* **2009**, *104*, 735–746. [[CrossRef](#)] [[PubMed](#)]
4. De La Fuente, A.; Bing, N.; Hoeschele, I.; Mendes, P. Discovery of meaningful associations in genomic data using partial correlation coefficients. *Bioinformatics* **2004**, *20*, 3565–3574. [[CrossRef](#)] [[PubMed](#)]
5. Marrelec, G.; Kim, J.; Doyon, J.; Horwitz, B. Large-scale neural model validation of partial correlation analysis for effective connectivity investigation in functional MRI. *Hum. Brain Mapp.* **2009**, *30*, 941–950. [[CrossRef](#)] [[PubMed](#)]
6. Wang, G.J.; Xie, C.; Stanley, H.E. Correlation structure and evolution of world stock markets: Evidence from Pearson and partial correlation-based networks. *Comput. Econ.* **2018**, *51*, 607–635. [[CrossRef](#)]
7. Kenett, D.Y.; Tumminello, M.; Madi, A.; Gur-Gershgoren, G.; Mantegna, R.N.; Ben-Jacob, E. Dominating clasp of the financial sector revealed by partial correlation analysis of the stock market. *PLoS ONE* **2010**, *5*, e15032. [[CrossRef](#)]
8. Kenett, D.Y.; Huang, X.; Vodenska, I.; Havlin, S.; Stanley, H.E. Partial correlation analysis: Applications for financial markets. *Quant. Finance* **2015**, *15*, 569–578. [[CrossRef](#)]
9. Michis, A.A. Multiscale partial correlation clustering of stock market returns. *J. Risk Financ. Manag.* **2022**, *15*, 24. [[CrossRef](#)]
10. Singh, V.; Li, B.; Roca, E. Global and regional linkages across market cycles: evidence from partial correlations in a network framework. *Appl. Econ.* **2019**, *51*, 3551–3582. [[CrossRef](#)]
11. Epskamp, S.; Fried, E.I. A tutorial on regularized partial correlation networks. *Psychol. Methods* **2018**, *23*, 617–634. [[CrossRef](#)]
12. Williams, D.R.; Rast, P. Back to the basics: Rethinking partial correlation network methodology. *Brit. J. Math. Stat. Psy.* **2020**, *73*, 187–212. [[CrossRef](#)] [[PubMed](#)]
13. Waldorp, L.; Marsman, M. Relations between networks, regression, partial correlation, and the latent variable model. *Multivariate Behav. Res.* **2022**, *57*, 994–1006. [[CrossRef](#)]
14. Gvozdev, A.; Parovik, R. On the relationship between the fractal dimension of geomagnetic variations at Altay and the space weather characteristics. *Mathematics* **2023**, *11*, 3449. [[CrossRef](#)]
15. Khare, K.; Oh, S.Y.; Rajaratnam, B. A convex pseudolikelihood framework for high dimensional partial correlation estimation with convergence guarantees. *J. R. Stat. Soc. B* **2015**, *77*, 803–825. [[CrossRef](#)]
16. Kim, S. ppcor: An R package for a fast calculation to semi-partial correlation coefficients. *Commun. Stat. Appl. Methods* **2015**, *22*, 665. [[CrossRef](#)] [[PubMed](#)]
17. Huang, Z.; Deb, N.; Sen, B. Kernel partial correlation coefficient—A measure of conditional dependence. *J. Mach. Learn. Res.* **2022**, *23*, 9699–9756.
18. Van Aert, R.C.; Goos, C. A critical reflection on computing the sampling variance of the partial correlation coefficient. *Res. Synth. Methods* **2023**, *14*, 520–525. [[CrossRef](#)]
19. Hu, H.; Qiu, Y. Inference for nonparanormal partial correlation via regularized rank based nodewise regression. *Biometrics* **2023**, *79*, 1173–1186. [[CrossRef](#)]
20. Cox, D.R.; Wermuth, N. *Multivariate Dependencies—Models, Analysis and Interpretation*; Chapman and Hall: London, UK, 1996.
21. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. B* **1996**, *58*, 267–288. [[CrossRef](#)]
22. Owen, A.B. A robust hybrid of lasso and ridge regression. *Contemp. Math.* **2007**, *443*, 59–72.
23. Fan, J.; Li, R. Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Am. Stat. Assoc.* **2001**, *96*, 1348–1360. [[CrossRef](#)]
24. Zou, H.; Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. B.* **2005**, *67*, 301–320. [[CrossRef](#)]
25. Tibshirani, R.; Saunders, M.; Rosset, S.; Zhu, J.; Knight, K. Sparsity and smoothness via the fused lasso. *J. R. Stat. Soc. B.* **2005**, *67*, 91–108. [[CrossRef](#)]
26. Zhang, C.H. Nearly unbiased variable selection under minimax concave penalty. *Ann. Stat.* **2010**, *38*, 894–942. [[CrossRef](#)] [[PubMed](#)]
27. Wang, H. Coordinate descent algorithm for covariance graphical lasso. *Stat. Comput.* **2014**, *24*, 521–529. [[CrossRef](#)]
28. Fan, J.; Fan, Y.; Lv, J. High dimensional covariance matrix estimation using a factor model. *JoE.* **2008**, *147*, 186–197. [[CrossRef](#)]
29. Elton, E.J.; Gruber, M.J.; Brown, S.J.; Goetzmann, W.N. *Modern Portfolio Theory and Investment Analysis*; John Wiley and Sons: Hoboken, NJ, USA, 2009.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.