Taylor & Francis
Taylor & Francis Group

Check for updates

Robust Covariance Matrix Estimation for High-Dimensional Compositional Data with Application to Sales Data Analysis

Danning Li[a], Arun Srinivasan[b], Qian Chen[c], and Lingzhou Xue[b]

[a]School of Mathematics and Statistics and KLAS, Northeast Normal University, Changchun, China; [b]Department of Statistics, Penn State University, University Park, PA; [c]Department of Supply Chain and Information Systems, Penn State University, University Park, PA

ABSTRACT

Compositional data arises in a wide variety of research areas when some form of standardization and composition is necessary. Estimating covariance matrices is of fundamental importance for high-dimensional compositional data analysis. However, existing methods require the restrictive Gaussian or sub-Gaussian assumption, which may not hold in practice. We propose a robust composition adjusted thresholding covariance procedure based on Huber-type M-estimation to estimate the sparse covariance structure of high-dimensional compositional data. We introduce a cross-validation procedure to choose the tuning parameters of the proposed method. Theoretically, by assuming a bounded fourth moment condition, we obtain the rates of convergence and signal recovery property for the proposed method and provide the theoretical guarantees for the cross-validation procedure under the high-dimensional setting. Numerically, we demonstrate the effectiveness of the proposed method in simulation studies and also a real application to sales data analysis.

1. Introduction

Compositional data analysis gives a powerful multivariate statistical approach for the analysis of high-dimensional data, especially when some form of standardization and composition is necessary. In this compositional framework, raw data are transformed into compositional data after calculating their relative proportions or percentages, and the sum of compositional data is constrained to the be some constant. This data structure is widely used in many research fields such as biological sciences (Tsilimigras and Fodor 2016; Gloor et al. 2017), business and economics (Fry, Fry, and McLaren 2000; Arata and Onozaki 2017), and geological sciences (Thomas and Aitchison 2005). As the rate of data collection increases exponentially, the need to analyze data under a compositional framework increases as well. The comparison of raw data across experimental units can be misleading when raw totals may vary greatly from experimental unit to experimental unit or be accidentally large. To facilitate the comparison between experimental units, researchers usually normalize the data matrix into a matrix of proportions or percentages. For example, a sales data analysis can provide helpful insights to design data-driven marketing strategies, and the comparison of sales data of various products/categories can be biased by different scales of sales across channels or store locations. In practice, compositions are commonly used in sales data to effectively compare the sales of products through different sales channels or at different store locations.

The constant-sum structure in compositional data leads to the so-called negative bias problem (also known as the constant-sum problem) in the covariance or correlation matrix (Aitchison 1982). Most of existing multivariate statistical methods do not take into account the special nature of compositional spaces (i.e., the sum is constrained to be a constant), and they can lead to inappropriate inferences and spurious findings. This problem has been demonstrated in the statistical literature over the past four decades (Aitchison 1982), and we will also illustrate this issue in the empirical analysis of sales data in Section 5. As a result, alternative multivariate statistical methods have been developed to incorporate the special constant-sum structure of compositional data, including compositional regression analysis, logcontrast principal component analysis, and many others (Aitchison 1982, 1983; Aitchison and Bacon-Shone 1984). In the last decade, new multivariate statistical methods such as high-dimensional regression analysis (Lin et al. 2014; Srinivasan, Xue, and Zhan 2021) and large covariance matrix estimation (Cao, Lin, and Li 2019; Li et al. in press) have been proposed for the analysis of compositional data in the high-dimensional setting, where the number of samples can be smaller or much smaller than the number of variables.

Estimating covariance matrices is of fundamental importance for multivariate statistical methods since covariance matrices characterize the important statistical relationships between variables. The accurate estimation of the covariance matrix is essential for a variety of downstream statistical methods such as clustering, discriminant analysis, and principal

components analysis, especially for high-dimensional data. Recently, Cao, Lin, and Li (2019) proposed the composition adjusted thresholding (COAT) procedure to estimate the sparse covariance matrix of compositional data in the high-dimensional setting. However, existing methods including Lin et al. (2014), Cao, Lin, and Li (2019), Srinivasan, Xue, and Zhan (2021), and Li et al. (in press) require the restrictive Gaussian or sub-Gaussian assumption on the data-generating distribution for compositional data. In real-world applications, we often need to deal with nonnormal and heavy-tailed data. For example, the issue of nonnormal and heavy-tailed data is demonstrated in the application to sales data analysis in Section 5. The violation of Gaussian or sub-Gaussian assumption can greatly diminish the effectiveness of existing methods (Huber 1964, 1981; Fan, Li, and Wang 2017; Avella-Medina et al. 2018).

In this article, we study the robust estimation of the large covariance matrix for high-dimensional compositional data to handle potential deviations from the normality. Synthesizing robust estimation and compositional analysis provides an promising approach without requiring the restrictive Gaussian or sub-Gaussian assumption. We propose a robust composition adjusted thresholding based on Huber-type M-estimation, denoted by M-COAT, to estimate the sparse covariance matrix for high-dimensional compositional data. M-COAT is a two-step regularized estimation procedure. In the first step, we combine both strengths of the centered log-ratio (clr) transformation and Huber-type M-estimation to derive a pilot estimator of the covariance matrix for compositional data. In the second step, given this robust pilot estimate, we employ a positive-definite thresholding scheme to induce sparsity and ensure the positive-definiteness of our thresholding covariance matrix estimator, which is vital for many downstream analysis. The numerical properties of the proposed M-COAT method are demonstrated in simulation studies and a real application to the store-scanner Dominick's Finer Foods dataset.

Theoretically, we study the asymptotic properties for the proposed M-COAT method under the high-dimensional setting where the dimension is nearly on the exponential order of the sample size. We assume a bounded fourth moment condition, allowing for a richer class of distributions beyond the Gaussian or sub-Gaussian assumption. The proposed estimator achieves the same convergence rates under both the spectral norm and Frobenius norm as the COAT does under the sub-Gaussian assumption. Also, we prove the correct sparsity and sign consistency properties for the proposed M-COAT method.

The proposed M-COAT method includes two tuning parameters: H in the piece-wise Huber loss function $\psi_H(z) = \min\{\max(-H, z) H\}$ and λ in the thresholding estimation. We introduce a cross-validation procedure using the squared Frobenius loss to choose both H and λ. The proposed cross-validation procedure enjoys the theoretical guarantees under the high-dimensional setting. We prove that the tuning parameters selected by this cross-validation procedure perform comparably to the oracle tuning parameters that are based on the true covariance matrix under an elliptical distribution assumption. To the best of our knowledge, this result makes a separate contribution to the literature, as it fills the gap of choosing the tuning parameters with theoretical guarantees for robust large covariance matrix estimation via Huber-type M-estimation (Fan, Li, and Wang 2017; Avella-Medina et al. 2018).

The rest of this article is organized as follows. In Section 2 we first introduce the compositional framework and then present the methodological details for the proposed M-COAT method and its cross-validation procedure. In Section 3 we outline the theoretical results including convergence rates, correct sparsity, sign consistency, and a guarantee for cross-validation under the compositional and nonnormal framework in high dimensions. Sections 4 and 5 evaluate the numerical performance of the proposed M-COAT through simulation studies and an application to sales data analysis. Finally we conclude with a summary of our results in Section 6. The proofs and additional numerical results are presented in a supplementary file.

2. Methodology

We first describe the compositional framework (Aitchison 1982; Aitchison and Bacon-Shone 1984) in Section 2.1. After these preliminaries, we present the proposed two-step M-COAT method in Section 2.2 and a new cross-validation procedure to choose the tuning parameters of the M-COAT in Section 2.3.

2.1. The Compositional Framework

Following the compositional framework (Aitchison 1982; Aitchison and Bacon-Shone 1984; Cao, Lin, and Li 2019), we will use latent variables and the centered log-ratio (clr) transformation to account for the special constant-sum structure of compositional data.

Let $\mathbf{W} = (W_1, \ldots, W_p)$ with $W_k > 0, k = 1, \ldots, p$, be the vector of latent variables known as the basis. We consider the following normalization of $\mathbf{W}$ to obtain the vector of compositional variables $\mathbf{X} = (X_1, \ldots, X_p)$, where

$$X_j = \frac{W_j}{\sum_{k=1}^p W_k} \tag{1}$$

for $j = 1, \ldots, p$. With this normalization procedure, $\mathbf{X}$ lies on the simplex $\mathcal{A}_p$ in $\mathbb{R}^p$, which is known as the Aitchison Simplex (Aitchison 1982) and is defined as follows:

$$\mathcal{A}_p = \{\mathbf{X} = (X_1, \ldots, X_p)' \in \mathbb{R}^p : X_k > 0, \, k = 1, \ldots, p, \sum_{k=1}^p X_k = 1\} \tag{2}$$

One critical issue arising from this simplex is that the compositional vector $\mathbf{X}$ is non-identifiable. To combat this issue, we define $\mathbf{Y} = \log(\mathbf{W}) = (\log(W_1), \ldots, \log(W_p))^T$ as the log-basis and assume that $\mathbf{Y}$ follows a distribution F_Y with mean vector μ and covariance matrix $\Sigma = (\sigma_{ij})_{p \times p}$. The basis covariance matrix Σ is the parameter of interest that characterizes the statistical relationship in the compositional data.

We will study the robust estimation of the sparse basis covariance matrix Σ with the iid high-dimensional compositional data $\mathbf{X}_1, \ldots, \mathbf{X}_n$, where p can be much larger than n. However, it is infeasible to use $\mathbf{W}$ or $\mathbf{X}$ for estimating the basis covariance matrix (Cao, Lin, and Li 2019). The clr transformation provides a promising approach by transforming the compositional data away from $\mathcal{A}_p$, which is defined as $\mathbf{Z} = \text{clr}(\mathbf{X}) = (Z_1, \ldots, Z_p)'$ with

$$Z_j = \log(X_j) - \frac{1}{p}\sum_{k=1}^{p}\log(X_k) = \log(X_j/g(\mathbf{X})) \qquad (3)$$

for $j = 1, \ldots, p$, where $g(\mathbf{X}) = (\prod_{j=1}^{p} X_j)^{1/p}$ is the geometric mean of $\mathbf{X}$. The magnitude of each Z_j is not restricted to $(0, 1]$, but $\mathbf{Z}$ is still rank-deficient due to the simple fact that $\sum_{j=1}^{p} Z_j = 0$. The covariance matrix of $\mathbf{Z}$, denoted by Γ, is also known as the centered log-ratio covariance matrix. By definition, we know that $\mathbf{Z} = A\mathbf{Y}$ with $A = \mathbf{I}_{p \times p} - (1/p)\mathbf{1}_p\mathbf{1}_p^T$. Thus, we have that $\Gamma = A\Sigma A^T$.

As shown in Proposition 1 of Cao, Lin, and Li (2019), two covariance parameters Σ and Γ are asymptotically indistinguishable if they belong to a class of sparse covariance matrices. More specifically, the sparse basis covariance parameter Σ is approximately identifiable by estimating the sparse covariance parameter Γ. The sparse estimation of Γ with the clr transformed data $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ serves as an effective proxy for the sparse estimation of Σ.

2.2. The Proposed Method

The proposed M-COAT method consists of two steps. In the first step, we use the Huber-type M-estimation to construct a pilot estimator denoted by $\hat{\Gamma}_H$ based on the clr transformed data $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$. In the second step, we use a positive-definite thresholding scheme to obtain a positive-definite sparse covariance matrix estimator denoted by $T_\lambda(\hat{\Gamma}_H)$. The resulting sparse covariance estimator $T_\lambda(\hat{\Gamma}_H)$ can approximately estimate the desired basis covariance parameter Σ.

The pilot estimator $\hat{\Gamma}_H$ plays an important role in accounting for both robustness and the compositional constraint before progressing into the thresholding estimation. As will be shown through numerical analysis, an accurate pilot estimator is necessary to ensure that large errors from the pilot step do not unduly influence thresholding. Cao, Lin, and Li (2019) used the sample covariance matrix of the clr transformed data $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ to construct the pilot estimator in the COAT method. If the log-basis $\mathbf{Y}$ satisfies the Gaussian or sub-Gaussian assumption, the sample covariance matrix of $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ performs very well to construct the pilot estimator in the high-dimensional regime. However, the nonnormal, heavy-tailed, and leptokurtic distribution of the log-basis $\mathbf{Y}$ will negatively impact the accuracy of the pilot estimator for the COAT, which we will demonstrate in the simulation studies in Section 4.

In the presence of nonnormal and heavy-tailed data, we need to construct a robust pilot estimator from $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$. In the literature, Huber's M-estimation (Huber 1981) provides a broad class of robust estimators in many statistical problems. Recently, Avella-Medina et al. (2018) explored the Huber-type M-estimation of the large covariance matrix. We extend the Huber-type M-estimation to construct a robust pilot estimator for high-dimensional compositional data in the sequel.

Recall that Γ is the centered log-ratio covariance matrix of $\mathbf{Z} = \text{clr}(\mathbf{X}) = (Z_1, \ldots, Z_p)'$. Since sparse Γ and sparse Σ are asymptotically indistinguishable, we use the Huber-type M-estimation to estimate each entry of $\Gamma = (\gamma_{uv})_{p \times p}$, where

$$\gamma_{uv} = \text{cov}(Z_u, Z_v) = E(Z_u Z_v) - E(Z_u)E(Z_v) := \mu_{uv}^z - \mu_u^z \mu_v^z \qquad (4)$$

for any $u, v = 1, \ldots, p$. In view of (4), it is sufficient to provide the robust estimation for μ_{uv}^z and μ_u^z with $\mathbf{Z}_i = (Z_{i1}, \ldots, Z_{ip})'$ ($i = 1, \ldots, n$), for any $u, v = 1, \ldots, p$, respectively. Given the Huber function $\psi_H(z) = \min\{\max(-H, z), H\}$ at level H, we construct the Huber-type M-estimators of μ_{uv}^z and μ_u^z by solving the solutions $\hat{\mu}_u^z$ and $\hat{\mu}_{uv}^z$ to the following estimating equation, respectively:

$$\sum_{i=1}^{n} \psi_H(Z_{iu} - \mu_u^z) = 0 \qquad (5)$$

and

$$\sum_{i=1}^{n} \psi_H(Z_{iu} Z_{iv} - \mu_{uv}^z) = 0 \qquad (6)$$

for $u, v = 1, \ldots, p$. Here, H is a tuning parameter that needs to be chosen properly. Therefore, the robust pilot estimator can be constructed as follows:

$$\hat{\Gamma}_H = (\hat{\gamma}_{uv})_{p \times p} := (\hat{\mu}_{uv}^z - \hat{\mu}_u^z \hat{\mu}_v^z)_{p \times p}. \qquad (7)$$

The robust pilot estimator $\hat{\Gamma}_H$ enjoys theoretical guarantees which are explored in Section 3. In addition, $\hat{\Gamma}_H$ outperforms other rank-based methods (Xue and Zou 2012, 2014) through the simulation studies in Section 4.

Next, given the robust pilot estimator $\hat{\Gamma}_H$ in the first step, we use a positive-definite thresholding procedure to construct a sparse estimator $\hat{\Sigma}$ for the basis covariance parameter in the second step. Due to the rank-deficiency of $\mathbf{Z}$, the rank of $\hat{\Gamma}_H$ is at most $p - 1$. Nonetheless, both the covariance parameter of interest Σ and the proxy covariance parameter Γ are positive definite. The positive-definiteness of $\hat{\Sigma}$ is essential to avoid the potential identifiability issue or degeneracy in downstream analysis. To this end, we use the positive-definite ℓ_1-penalization (Xue, Ma, and Zou 2012) to compute the sparse estimator $\hat{\Sigma}$ as

$$T_\lambda(\hat{\Gamma}_H) := \arg\min_{\Sigma \geq \varepsilon I} \frac{1}{2}||\Sigma - \hat{\Gamma}_H||_F^2 + \lambda|\Sigma|_{1,\text{off}} \qquad (8)$$

where $\varepsilon > 0$ is an arbitrarily small constant, λ is a tuning parameter, and $|\cdot|_{1,\text{off}}$ denotes the ℓ_1-norm of the off-diagonal entries. The constraint that $\varepsilon > 0$ ensures that the resulting covariance estimator is positive-definite. Note that ε can be set as small as possible, and it is not a tuning parameter. Through this constraint, we recover the optimal positive-definite estimate of Γ, even though $\hat{\Gamma}_H$ itself is rank-deficient. We also note that (8) is a convex but nonsmooth function that can be efficiently solved by an alternating direction method of multipliers (ADMM) as outlined in Algorithm 1 of Xue and Zou (2012). We summarize the details about the proposed M-COAT method in Algorithm 1.

Algorithm 1: The proposed M-COAT method.

Input. the compositional data matrix $\mathbf{X}$ and tuning parameters H & λ.

Preprocessing. use the clr transformation (3) to obtain $\mathbf{Z} = \text{clr}(\mathbf{X})$.

Step 1. use Huber-type M-estimation to construct a robust pilot estimator

$$\hat{\Gamma}_H = (\hat{\mu}^z_{uv} - \hat{\mu}^z_u \hat{\mu}^z_v)_{p \times p}$$

where $\hat{\mu}^z_{uv}$ and $\hat{\mu}^z_u$ are solved from estimating equations $\sum_{i=1}^n \psi_H(Z_{iu} - \mu^z_u) = 0$ and $\sum_{i=1}^n \psi_H(Z_{iu}Z_{iv} - \mu^z_{uv}) = 0$ as in (5)–(6) for $u, v = 1, \ldots, p$.

Step 2. use an alternating direction method of multipliers (ADMM) to solve the positive-definite thresholding estimator $\hat{\Sigma}$ from

$$\min_{\Sigma \geq \varepsilon I} \frac{1}{2} ||\Sigma - \hat{\Gamma}_H||^2_F + \lambda |\Sigma|_{1, \text{off}}.$$

Output: a positive-definite sparse basis covariance estimator $\hat{\Sigma}$.

2.3. A Cross-Validation Procedure

The effectiveness of our proposed M-COAT method lies in the proper selection of H in the Huber loss function and λ in the positive-definite thresholding. The tuning parameter λ is usually chosen by cross-validation that enjoys theoretical guarantees (Bickel and Levina 2008; Cai and Liu 2011; Cao, Lin, and Li 2019). The tuning parameter H affects the accuracy of the pilot estimator $\hat{\Gamma}_H$ and thus the sparse basis covariance estimator $\hat{\Sigma}$. The proper choice of H involves a tradeoff between bias and robustness. Unlike the classical robust statistical methods (Huber 1981), in order to alleviate the bias, we need to allow H to grow as the sample size increases in the high-dimensional setting (Avella-Medina et al. 2018).

However, Avella-Medina et al. (2018) do not provide a procedure with theoretical guarantees to choose this critical tuning parameter for Huber-type M-estimation. Specifically, Avella-Medina et al. (2018) used a conservative choice by assuming the true distribution to be a Student t-distribution with five degrees of freedom. In what follows, we fill this gap with a cross-validation procedure with theoretical guarantees. Thus, this is a separate contribution to the literature of Huber-type M-estimation with high-dimensional data.

We propose a cross-validation procedure using the squared Frobenius norm to select both H and λ for the proposed M-COAT method. We randomly partition the sample into two subsamples of size n_1 and $n_2 = n - n_1$ and repeat this random partition K times, where $n_1 \asymp n_2 \asymp n$ and K is a fixed integer. Let $\hat{\Gamma}^k_{1,H}$ be the robust pilot estimator defined as (7) using the training dataset with n_1 observations in the kth split, and $T_\lambda(\hat{\Gamma}^k_{1,H})$ is the corresponding M-COAT estimator. Let $\hat{\Gamma}^k_2$ be the sample covariance matrix computed from the hold-out dataset with n_2 observations in the kth split.

For any given tuning parameters H and λ, we define the squared Frobenius loss to assess the discrepancy between

$T_\lambda(\hat{\Gamma}^k_{1,H})$ and $\hat{\Gamma}^k_2$ for $k = 1, \ldots, K$ as follows:

$$\hat{R}(\lambda, H) = \frac{1}{K} \sum_{k=1}^K ||T_\lambda(\hat{\Gamma}^k_{1,H}) - \hat{\Gamma}^k_2||^2_F. \tag{9}$$

To find a data-driven choice of (λ, H), we optimize (9) by searching over a set of candidate parameters $(\lambda_1, \ldots, \lambda_{J_1}) \times (H_1, \ldots, H_{J_2})$, where J_1 denotes the size of the grid for λ and J_2 denotes the size of the grid for H. More specifically, we choose the pair $(\hat{\lambda}, \hat{H})$ that minimizes (9) as follows:

$$(\hat{\lambda}, \hat{H}) = \text{argmin}_{\lambda \in (\lambda_1, \ldots, \lambda_{J_1}), H \in (H_1, \ldots, H_{J_2})} \hat{R}(\lambda, H). \tag{10}$$

The final estimator of the sparse basis covariance parameter Σ is given by $\hat{\Sigma} = T_{\hat{\lambda}}(\hat{\Gamma}_{\hat{H}})$. In Section 3, we will prove that the data-driven choice $(\hat{\lambda}, \hat{H})$ as in (10) performs comparably to the oracle tuning parameters that are based on the true covariance matrix under an elliptical distribution assumption.

In practice, we may define the grid points for λ to be evenly spaced between a small value near 0 and a large value such that the estimated covariance matrix would be close to a diagonal matrix. As for the choice of H, we define the grid according to the theoretical order (see Theorem 4 in the next section). Although we need to choose two parameters λ and H, it is very efficient to solve the penalized covariance estimation for each given (λ, H) pair, so the overall computing time is still feasible for large-scale real applications. We may also use a more efficient C/C++ implementation to reduce the computing time of the proposed method, and parallel computing can further reduce the computing time.

3. Theoretical Properties

In this section, we study the theoretical properties of our proposed M-COAT method and cross-validation procedure. We first derive the desired convergence rate and signal recovery properties in Section 3.1. Further, we provide a new theoretical guarantee for using the cross-validation to simultaneously choose H and λ in Section 3.2.

We consider the high-dimensional setting that $\log p/n \to 0$, where the dimension p is on a nearly exponential order of the sample size n. Also, we consider the following class of covariance matrices defined by the ℓ_q norm with $q \in [0, 1)$:

$$U(q, s_0(p), M) \equiv \left\{ \Sigma \; : \; \Sigma \succ 0, \; \max_j \sigma_{jj} \leq M, \; \max_i \sum_{j=1}^p |\sigma_{ij}|^q \leq s_0(p) \right\}.$$

$U(0, s_0(p), M)$ denotes the class of sparse covariance matrices and $U(q, s_0(p), M)$ with $q \in (0, 1)$ denotes the class of approximately sparse covariance matrices. $U(q, s_0(p), M)$ was first introduced by Bickel and Levina (2008) and then widely explored in the large covariance matrix estimation (Rothman, Levina, and Zhu 2009; Cai and Liu 2011; Xue, Ma, and Zou 2012; Cao, Lin, and Li 2019).

3.1. Convergence Rates and Signal Recovery

We study the convergence rates and signal recovery of the M-COAT without requiring the Gaussian or sub-Gaussian assumption. Instead, we assume that the first four moments of $\mathbf{Z}$ are

finite, that is, $\max_{1 \le u \le p} E[Z_u]^4 = \kappa^2 < \infty$. Our theoretical analysis allows for a richer class of heavy-tailed distributions such as the Laplace distribution and t-distributions with five or more degrees of freedom.

Theorem 1 and Corollary 1 present the rate of convergence of our proposed M-COAT estimator under the spectral norm and under the Frobenius norm, respectively.

Theorem 1. Suppose that $\log p / n \to 0$ and $\max_{1 \le u \le p} E[Z_u]^4 = \kappa^2 < \infty$. Then, uniformly on $U(q, s_0(p), M)$ with $q \in [0, 1)$, if $H = K_0 \sqrt{\frac{n}{\log p}}$ and $\lambda = C_1 \sqrt{\frac{\log p}{n}} + C_2 \frac{s_0(p)}{p}$ for some constants $K_0, C_1, C_2 > 0$, we have

$$||\hat{\Sigma} - \Sigma||_2 = O_p \left(s_0(p)(\sqrt{\frac{\log p}{n}} + \frac{s_0(p)}{p})^{1-q} \right)$$

Corollary 1. Suppose that $\log p / n \to 0$ and $\max_{1 \le u \le p} E[Z_u]^4 = \kappa^2 < \infty$. Then, uniformly on $U(q, s_0(p), M)$ with $q \in [0, 1)$, if $H = K_0 \sqrt{\frac{n}{\log p}}$ and $\lambda = C_1 \sqrt{\frac{\log p}{n}} + C_2 \frac{s_0(p)}{p}$ for some constants $K_0, C_1, C_2 > 0$, we have

$$\frac{1}{p}||\hat{\Sigma} - \Sigma||_F^2 = O_p \left(s_0(p)(\sqrt{\frac{\log p}{n}} + \frac{s_0(p)}{p})^{2-q} \right).$$

The above rates of convergence can be decomposed into two parts, the first matching the optimal rate proposed by Cai and Liu (2011) and the additional $\frac{s_0(p)}{p}$ denoting the necessary additional error from using $\mathbf{Z}$ instead of the unattainable $\mathbf{Y}$. An interesting feature of this rate is that in high dimensions the identification error $\frac{s_0(p)}{p}$ vanishes as the dimensionality increases. Instead, for large p, the above rate is dominated by $\sqrt{\frac{\log p}{n}}$ which matches the rates on unconstrained data studied by Cai and Liu (2011) and Bickel and Levina (2008).

As shown in Theorem 1 and Corollary 1, the proposed M-COAT achieves the same convergence rates under the bounded fourth moment condition as the COAT (Cao, Lin, and Li 2019) does under the sub-Gaussian assumption.

Next, we present the signal recovery and sign consistency of the M-COAT in Theorems 2 and 3, respectively. We show that the proposed M-COAT estimator attains desirable guarantees for the recovery of the support of the true signal and also the sign of the true signal with high probability.

Theorem 2. Suppose that $\log p / n \to 0$ and $\max_{1 \le u \le p} E[Z_u]^4 = \kappa^2 < \infty$. Then, uniformly on $U(q, s_0(p), M)$ with $q \in [0, 1)$, if $H = K_0 \sqrt{\frac{n}{\log p}}$ and $\lambda = C_1 \sqrt{\frac{\log p}{n}} + C_2 \frac{s_0(p)}{p}$ for some constants $K_0, C_1, C_2 > 0$, we have

$$P(\hat{\sigma}_{ij} = 0 \text{ for all } \sigma_{ij} = 0) \to 1 \qquad (11)$$

Theorem 3. Suppose that $\log p / n \to 0$ and $\max_{1 \le u \le p} E[Z_u]^4 = \kappa^2 < \infty$. Then, uniformly on $U(q, s_0(p), M)$ with $q \in [0, 1)$, if $H = K_0 \sqrt{\frac{n}{\log p}}$ and $\lambda = C_1 \sqrt{\frac{\log p}{n}} + C_2 \frac{s_0(p)}{p}$ satisfying that $\lambda \le \frac{3}{2} \min_{(i,j): \sigma_{ij} \ne 0} |\sigma_{ij}|$ for some constants $K_0, C_1, C_2 > 0$, we have

$$P(\text{sign}(\hat{\sigma}_{ij}) = \text{sign}(\sigma_{ij}) \text{ for all } (i, j)) \to 1 \qquad (12)$$

3.2. Cross-validation Guarantees

To derive the theoretical guarantees for the cross-validation procedure, we need an additional assumption on the distribution by assuming that $\mathbf{Y}_i$ is drawn from the broad class of elliptical distributions. Specifically, we assume that each $\mathbf{Y}_i \in \mathbb{R}^p$ follows an elliptical distribution such that $\mathbf{Y}_i \sim \mathcal{E}_p(\mu, \Sigma, \phi)$, where $\mathcal{E}_p(\mu, \Sigma, \phi)$ is a p-dimensional elliptical distribution with the location parameter $\mu \in \mathbb{R}^p$, positive-definite shape matrix Σ, and density generator ϕ (Cambanis, Huang, and Simons 1981; Tyler 1987). By Cambanis, Huang, and Simons (1981), $\mathbf{Y}_i$ is equivalently represented as

$$\mathbf{Y}_i = \mu + u_i(\Sigma)^{1/2} \xi_i,$$

where u_i be a scalar random variable, and $\xi_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{p_0 \times p_0})$. This is not an overly restrictive assumption as commonly used heavy-tailed distributions such as the t-distribution and Laplace distribution are special examples of this class of elliptical distributions (Goes, Lerman, and Nadler 2020).

To demonstrate the effectiveness of our proposed cross-validation procedure, we define the oracle tuning parameters $\hat{\lambda}_o$ and $\hat{H}_o$ as the optimal choices that has the oracle information about Σ as follows:

$$(\hat{\lambda}_o, \hat{H}_o) = \underset{(\lambda, H)}{\text{argmin}} ||T_\lambda(\hat{\Gamma}_H) - \Sigma||_F^2.$$

Theorem 4 shows that the error of the M-COAT estimator computed using the selected tuning parameters $\hat{\lambda}$ and $\hat{H}$ is close to the optimal error with the oracle tuning parameters $\hat{\lambda}_o$ and $\hat{H}_o$. Following Bickel and Levina (2008) and Cai and Liu (2011), we just need to study the $K = 1$ setting in Theorem 4, and this result holds for the general K.

Theorem 4. Let $J = J_1 \times J_2$. Suppose that $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ follows the iid elliptical distribution $\mathcal{E}_p(\mu, \Sigma, \phi)$ with $\Sigma \in U(q, s_0(p), M)$, $Eu_i^4 < \infty$, $\frac{\log p}{n} \to 0$, and $p(\frac{\log p}{n})^{1-q/2} \to \infty$. When $n_1 \asymp n_2 \asymp n$, $s_0(p) = o(p(\frac{\log p}{n})^{1-q/2})$ and $(\log J)^3 = o(n(\frac{\log p}{n})^{1-q/2})$, we have

$$\frac{1}{p}||T_{\hat{\lambda}}(\hat{\Gamma}_{1,\hat{H}}) - \hat{\Gamma}_2||_F^2 = \frac{1}{p}||T_{\hat{\lambda}_o}(\hat{\Gamma}_{\hat{H}_o}) - \Sigma||_F^2(1 + o_p(1))$$

and

$$\frac{1}{p}||\hat{\Sigma} - \Sigma||_F^2 = O_p \left(s_0(p)(\sqrt{\frac{\log p}{n}} + \frac{s_0(p)}{p})^{2-q} \right).$$

Note that Theorem 4 includes both settings: fixed J and diverging J. When J_1 and J_2 are fixed, in view of $p \to \infty$, the condition that $(\log J)^3 = o(n(\log p/n)^{1-q/2})$ naturally holds. When J_1 or J_2 diverges, the upper bound on J is required such that the size of tuning set is proper with respect to the dimension and sample size.

We would like to point out that Avella-Medina et al. (2018) did not provide any theoretical guarantee to choose the tuning parameter for Huber-type M-estimation. In the current literature, Bickel and Levina (2008) and Cai and Liu (2011) provided the theoretical guarantees for their cross-validation procedure to choose the thresholding parameter under the Gaussian assumption. More specifically, Bickel and Levina (2008)
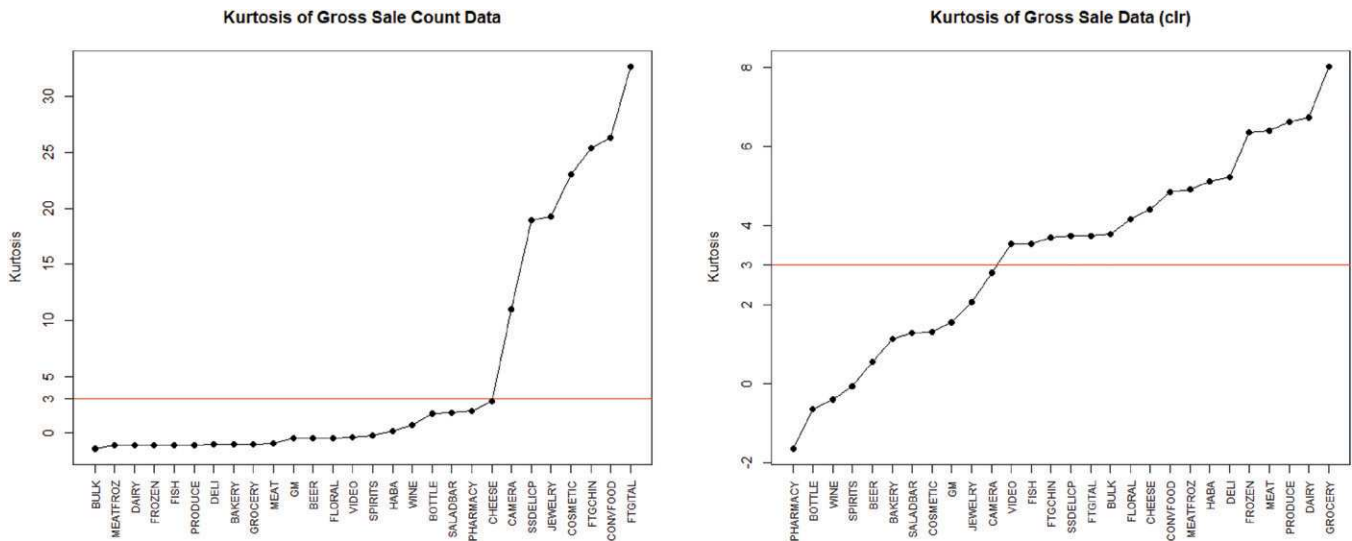
Figure 1. Plots of the kurtosis for department-level gross sale count data and clr transformed data. When kurtosis is greater than 3, the distribution is leptokurtic (heavy-tailed).

Table 1. Comparison of estimation performance for the AR covariance structure over 100 independent repetitions under the Laplace and t_5 distributions.

	Laplace(0, 1)			t_5		
Method	$p = 150$	$p = 200$	$p = 250$	$p = 150$	$p = 200$	$p = 250$
			Forbenius Norm			
M-COAT	13.316	15.255	16.994	13.209	15.207	17.051
S-COAT	15.754	18.148	20.225	15.513	17.906	20.039
K-COAT	16.682	19.240	21.442	16.416	18.953	21.192
COAT	17.982	21.423	24.404	19.056	21.833	22.799
			Max Norm			
M-COAT	6.070	6.001	6.413	5.746	5.743	6.045
S-COAT	6.452	6.753	7.394	6.374	6.805	7.386
K-COAT	6.379	6.712	7.317	6.354	6.817	7.382
COAT	8.523	10.660	13.726	12.006	12.886	10.553
			Spectral Norm			
M-COAT	4.354	4.308	4.296	4.322	4.289	4.297
S-COAT	4.892	4.874	4.869	4.868	4.858	4.864
K-COAT	5.061	5.052	5.049	5.037	5.034	5.039
COAT	5.605	6.352	7.082	7.259	7.479	6.035

Table 2. Comparison of estimation and selection performance for the sparse covariance structure over 100 independent repetitions under the Laplace and t_5 distributions.

	Laplace(0, 1)			t_5		
Method	$p = 150$	$p = 200$	$p = 250$	$p = 150$	$p = 200$	$p = 250$
			Forbenius Norm			
M-COAT	38.234	44.527	51.135	38.717	44.048	50.499
S-COAT	42.355	48.805	55.899	42.200	47.983	54.867
K-COAT	43.728	50.219	57.481	43.355	49.281	56.309
COAT	77.447	95.660	114.620	64.823	79.295	101.312
			Max Norm			
M-COAT	16.311	16.854	18.390	16.353	16.658	18.245
S-COAT	16.763	17.248	18.774	16.646	16.951	18.519
K-COAT	17.030	17.361	18.826	16.755	17.030	18.604
COAT	90.172	117.594	151.836	72.408	94.700	140.004
			Spectral Norm			
M-COAT	10.358	10.010	11.360	10.419	9.974	11.315
S-COAT	10.655	10.290	11.619	10.674	10.236	11.556
K-COAT	10.745	10.375	11.702	10.748	10.312	11.623
COAT	37.055	46.879	55.887	31.033	37.666	51.301
			TPR			
M-COAT	0.614	0.615	0.633	0.609	0.608	0.627
S-COAT	0.655	0.642	0.657	0.636	0.625	0.638
K-COAT	0.683	0.666	0.673	0.662	0.648	0.658
COAT	0.737	0.724	0.730	0.714	0.706	0.716
			FPR			
M-COAT	0.039	0.035	0.026	0.024	0.011	0.000
S-COAT	0.156	0.116	0.117	0.107	0.064	0.038
K-COAT	0.234	0.191	0.178	0.178	0.140	0.108
COAT	0.419	0.388	0.379	0.372	0.347	0.325

assumed a diverging size of the grid under the Gaussian assumption, and Cai and Liu (2011) assumed a fixed size of the grid under the Gaussian assumption. See Theorem 4 of Bickel and Levina (2008) and Theorem 5 of Cai and Liu (2011). To the best of our knowledge, Theorem 4 fills a gap to choose the tuning parameters with theoretical guarantees for the robust estimation of large covariance matrices, which makes a separate contribution to the literature.

4. Simulation Studies

We compare the numerical performance of the proposed M-COAT against the COAT and also its rank-based alternatives in simulation studies with different covariance structures and sparsity patterns. The rank-based alternatives use the adjusted Spearman's or Kendall's rank correlation as the pilot estimator (Xue and Zou 2012, 2014; Avella-Medina et al. 2018), and we denote them by S-COAT and K-COAT, respectively. The tuning parameters for K-COAT, S-COAT and COAT were selected in

a similar manner as M-COAT; however, we replace the Huber-based pilot estimator $\hat{\Gamma}_{1,H}^k$ used in (9) with the adjusted Kendall rank-based, the adjusted Spearman rank-based, and the standard covariance estimators accordingly.

We consider the following data generating process to compare the numerical performance. We generate $\mathbf{Y}$ through a Gaussian scale mixture framework. Gaussian scale mixtures fall into the class of elliptical distributions and is useful for efficiently generating elliptically distributed data (Andrews and Mallows 1974; Müller and Richter 2019; Goes, Lerman, and Nadler 2020; Li et al. in press). Thus, $\mathbf{Y}_i = \mu + u_i \Sigma^{1/2} \xi_i$ where the heavy-
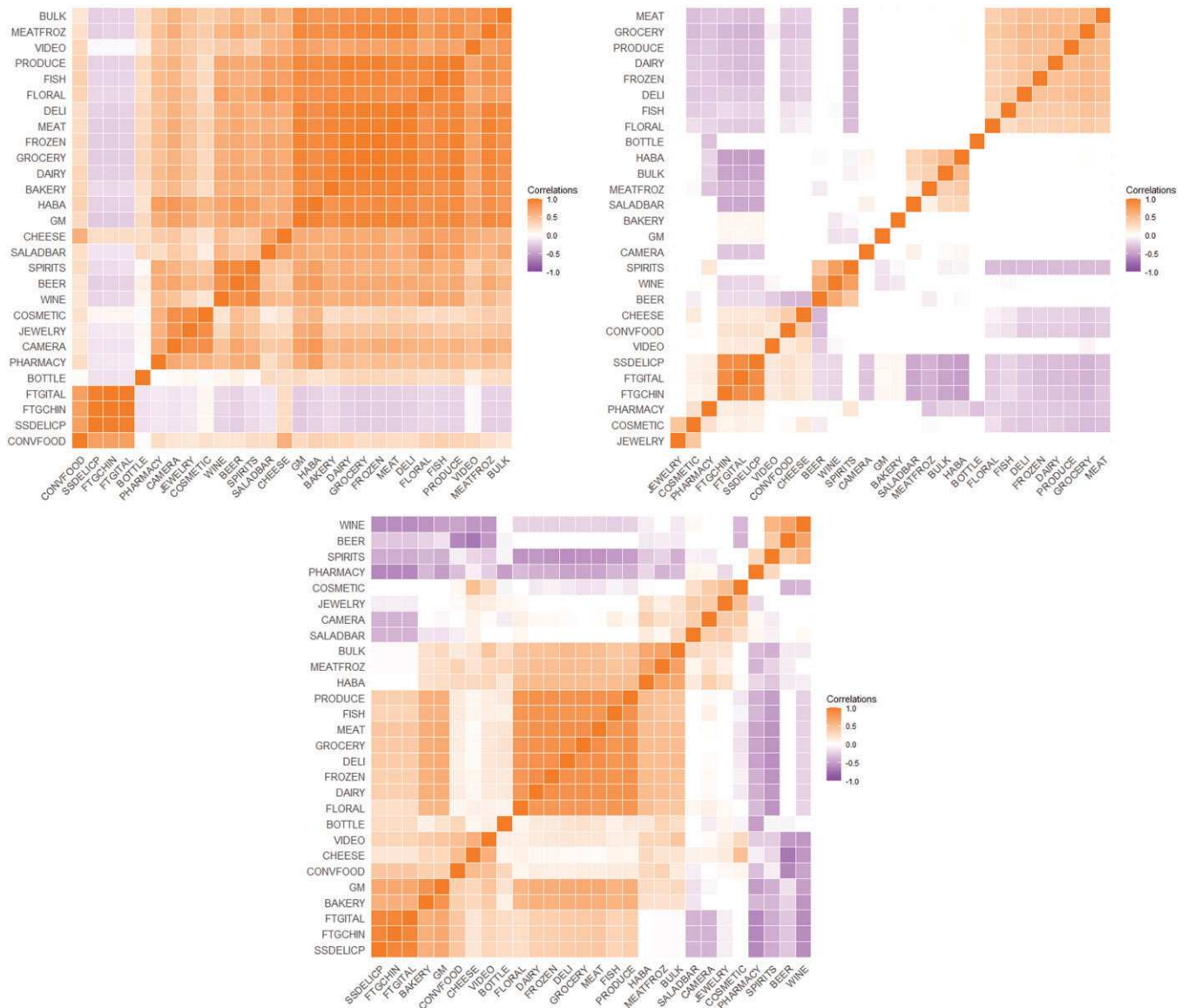
Figure 2. Heatmaps for the thresholding estimator on the raw count data (on the top left panel), the COAT procedure (on the top right panel), and the M-COAT procedure (on the bottom panel). The colors of the heatmap represent the correlation of the total sales between departments. Positive correlations are visualized by the orange color, and negative correlations are visualized by the purple color. The departments are ordered by hierarchical clustering using the estimated correlations.

tailed behavior is captured by the choice of u_i. We note that $\xi_{i,k} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{p \times p})$ and set $\mu = 0$ and Σ denotes the true covariance matrix of $\mathbf{Y}$. When $u_i = 1$ for all $i = 1, \ldots, n$, we have that $\mathbf{Y}$ follows a Gaussian distribution. We explore two heavy-tailed settings where $u_i \sim \text{Laplace}(0, 1)$ and $u_i \sim t_5$ for all $i = 1, \ldots, n$.

We consider two different covariance structures for Σ: the autoregressive (AR) covariance structure with $\Sigma = (0.7^{|i-j|})_{p \times p}$, and a sparse covariance structure with a block-diagonal matrix $\Sigma = \text{bdiag}(A_1, A_2)$, where $p = p_1 + p_2$, $A_1 = B + \varepsilon I_{p_1}$ and $A_2 = 4 I_{p_2}$. Here, B is a symmetric matrix whose lower triangle entries are independent and drawn uniformly from $[-3, -1.5] \cup [1.5, 3]$ with probability 0.15 and are zero otherwise. Let $\varepsilon = \max(-\lambda_{\min}, 0) + .01$ where $\lambda_{\min}$ is the minimum eigenvalue of B to ensure positive-definiteness. Both covariance structures have served as the benchmarks in the literature such as Bickel and Levina (2008), Cai and Liu (2011), Xue, Ma, and Zou (2012), and Cao, Lin, and Li (2019).

We set $n = 100$, $p = 100, 150, 200$, and $p_1 = \lfloor 2\sqrt{p} \rfloor$. To assess the effectiveness of different methods, we calculate the Forbeinus norm $||\hat{\Sigma} - \Sigma||_F$, the entry-wise maximum norm $||\hat{\Sigma} - \Sigma||_{\max}$, and the spectral norm $||\Sigma - \hat{\Sigma}||_2$. In the sparse covariance setting, we also assess the true positive rate (TPR) and false positive rate (FPR) that are defined as:

$$\text{TPR} = \frac{\#\{(i,j) : \hat{\sigma}_{ij} \neq 0 \text{ and } \sigma_{ij} \neq 0\}}{\#\{(i,j) : \sigma_{ij} \neq 0\}}$$

$$\text{FPR} = \frac{\#\{(i,j) : \hat{\sigma}_{ij} \neq 0 \text{ and } \sigma_{ij} = 0\}}{\#\{(i,j) : \hat{\sigma}_{ij} = 0\}}$$

These measures are averaged over 100 independent repetitions. The numerical results under the Laplace and t_5 distributions are found in Tables 1 and 2, respectively. The results for the Gaussian error setting across both covariance models is found in the supplementary materials.

As shown in both tables, regardless of the covariance structure, the M-COAT attains the smallest estimation errors and the
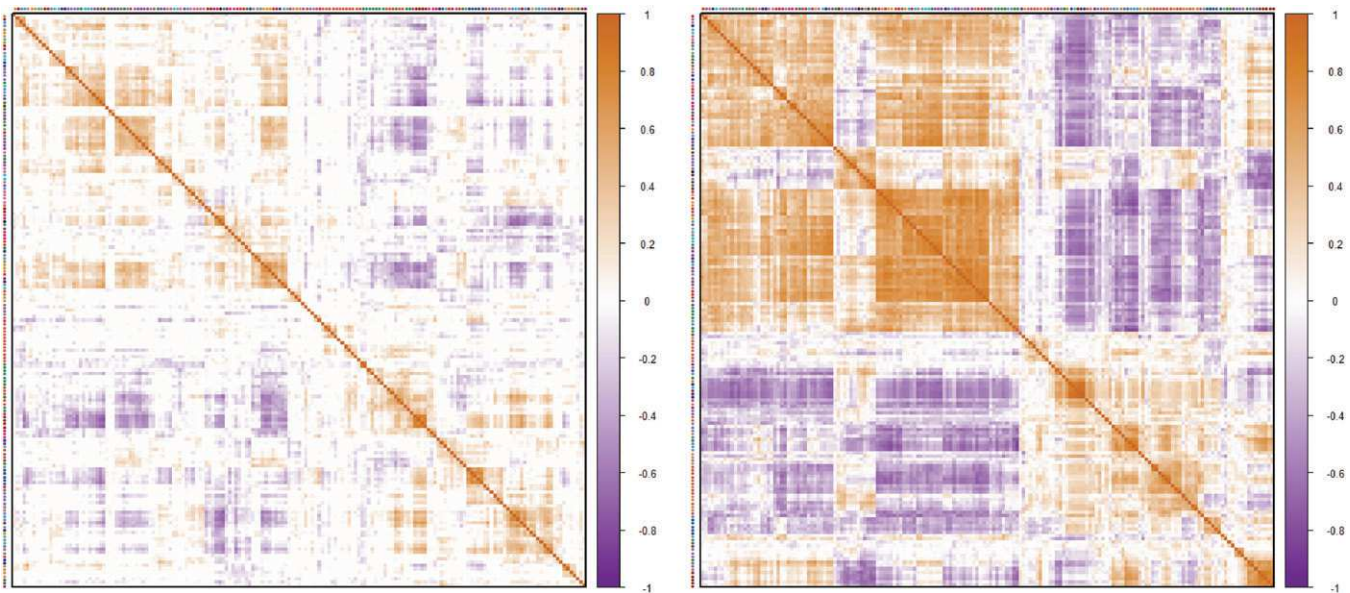
Figure 3. Heatmaps for the sparse covariance matrices of the product-level sales data in 1989 (left) and 1995 (right), estimated by the M-COAT procedure. The color of the circular dots indicates which of the 27 departments the product is from. Products from the same department are given the same color dot label on the rows and columns.

COAT has the largest estimation errors under both the Laplace and t_5 distributions. The COAT uses the sample covariance matrix as the pilot estimator, which is inappropriate for non-normal and heavy-tailed distributions. As a result, the COAT has much higher estimation errors than other methods across all settings. The rank-based estimators S-COAT and K-COAT are comparable to one another. The M-COAT provides further improvement over S-COAT and K-COAT, especially with the sparse covariance structure in Table 2. As for the selection performance, the M-COAT attains similar TPRs as S-COAT and K-COAT but achieves the significantly lower FPRs. The COAT has the slightly better TPRs but much worse FPRs than other methods across all settings. In summary, the M-COAT outperforms the COAT and also other robust estimators based on rank correlation in terms of both estimation and selection performances.

5. Application to Sales Data Analysis

In this section, we examine the numerical performance of the proposed M-COAT using the Dominick's Finer Foods scanner dataset, which is downloaded from Chicago Booth's Kilts Center for Marketing. This dataset contains a large amount of scanner data from 1989 to 1994 across multiple locations of the Dominick's Finer Foods store chain in the Chicago area. In this sales dataset, the total sales per store vary greatly due to store location. Besides, sales data is high-dimensional because of the large product variety, and can often be nonnormal. Thus, to effectively study the statistical associations of sales data, normalization is a required preprocessing step, and the compositional data analysis is then necessary.

This dataset consists the sales data of individual products as well as their associated departments (e.g., FISH, MEAT, GROCERY, DAIRY, and so on). Thus, researchers can perform the analysis of sales data at the department level or at the product level. In what follows, we first conduct the analysis of department-level sales data to motivate the importance of compositional analysis and robust estimation in Section 5.1, and we explore the high-dimensional compositional analysis of product-level sales data in Section 5.2. We demonstrate the effectiveness of our proposed approach via this real-world application. We also show that our method is a powerful tool for cross-selling analytics to measure the linkage between the sales of various products and provides a more accurate way of visualizing shoppers' behavior.

5.1. Department-Level Sales Analysis

Department-level sales were retrieved from the customer count files in the scanner dataset. We preprocessed the sales data by aggregating across years by store, and we formed a dataset consisting of the total sales of $p = 28$ departments across $n = 70$ different stores.

Through the exploratory analysis, we computed the kurtosis measures of the sales data at the department level. We plotted the department-level kurtosis values in Figure 1 for both the raw sales data, $\mathbf{W}$, and the clr-transformed sales data, $\mathbf{Z}$. As shown in Figure 1, it is evident that about 25% of the department-level observations are heavy-tailed or leptokurtic (i.e., kurtosis is greater than 3) in terms of raw counts and about half of them are heavy-tailed after the clr transformation. This suggests the importance of robust estimation.

Next, we demonstrate the need for the compositional analysis. We computed the naive thresholding estimator on the raw count data, the COAT on the compositional data, and the M-COAT on the compositional data. All these covariance estimators were normalized to correlation matrices for a better visualization. In the heatmaps, the departments are ordered by hierarchical clustering using the estimated correlations. The thresholding estimator, the COAT, and the M-COAT are all visualized in Figure 2.
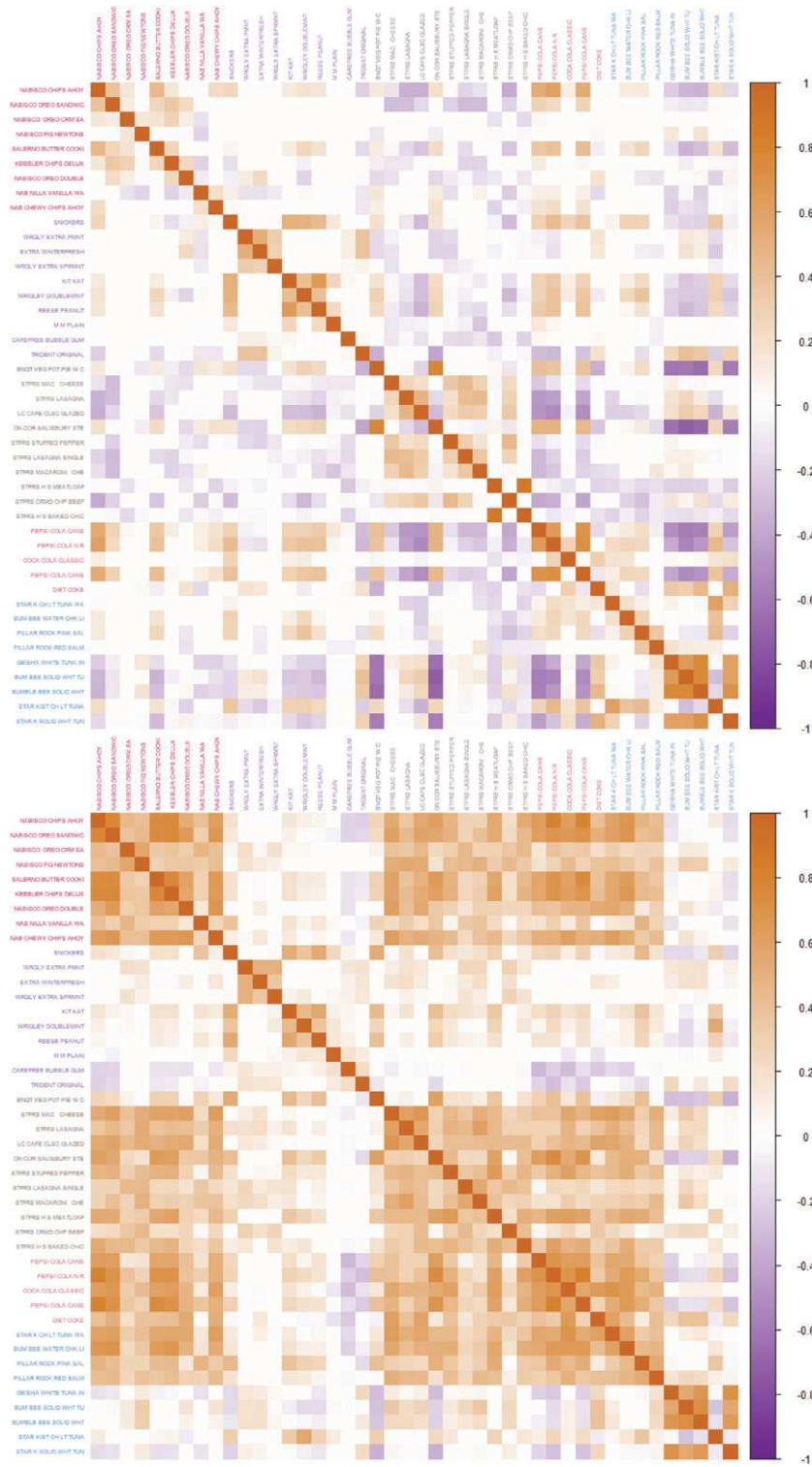
Figure 4. Heatmap for a subset of categories of products labeled by brand. The plots shows the correlation matrix in the year 1989 (left) during the recession and the year 1995 (right) after the recession. Products from the same category are given the same color on their labels.

When the compositional structure is ignored, as shown in Figure 2, the naive thresholding estimator leads to a very dense correlation structure, and it fails to identify any meaningful dependence structure between departments. The COAT improves the naive thresholding and captures the block structure after performing the hierarchical clustering using the estimated correlations. M-COAT provides the most

meaningful result that reveals rich associations of sales data between departments. The COAT and M-COAT estimates share some similarities (such as the dependencies among the core departments of FISH, MEAT, GROCERY, and DAIRY) in their heatmaps. Compared with COAT, M-COAT demonstrates a better capability to capture an interconnected market by identifying the significant associations between the core
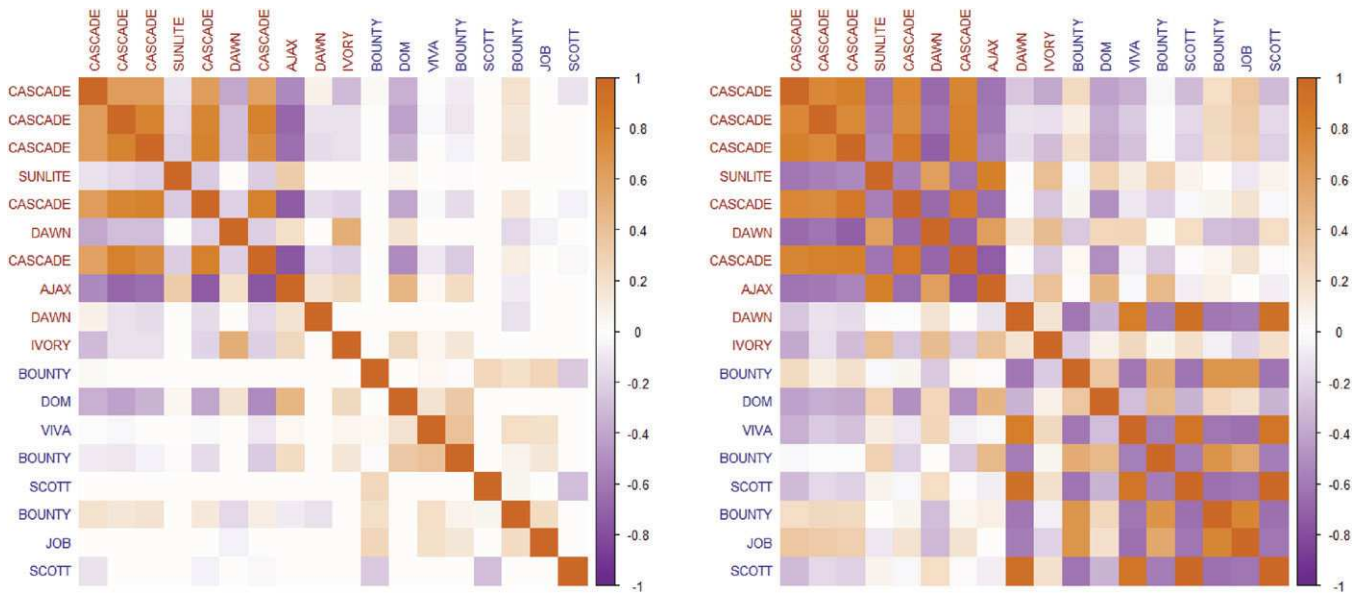
Figure 5. Heatmap for products in the categories of dish detergents and paper towels labeled by brand. The plots shows the correlation matrix in the year 1989 (left) during the recession and the year 1995 (right) after the recession. Products from the same category are given the same color on their labels.

departments and other departments such as the food-to-go (FTG) sections and BAKERY.

5.2. Product-Level Sales Analysis

This dataset also provides in-depth weekly sales data for individual products within each store. First, we aggregate the weekly sales data by year in order to capture larger scale economic factors that influence sales per year, and aid in our interpretation. Next, we select the top 10 products for each of the 27 departments and only keep those products that were sold across all 8 years for further investigation. Now, we have $p = 173$ unique products across $n = 70$ different stores, which falls into the high-dimensional setting.

We used the proposed M-COAT method to robustly estimate the high-dimensional covariance matrices for the yearly sales of each product and provide insights on cross-selling patterns. Figure 3 plots the heatmaps of the estimated covariance matrix for year 1989 and 1995. Figure 4 shows the heatmaps of the submatrices about five specific categories from the estimated covariance matrix for year 1989 and 1995, including cookies (red), front-end candies (purple), frozen entrees (black), frozen dinners (green), soda (yellow), and tuna (blue). Figure 5 shows the sales of paper towels (dark red) and dish detergent (dark blue) from the estimated covariance matrix for year 1989 and 1995. It is worth pointing out that an economic recession affected the United States the early 1990s (1989–1992) and the mid-nineties (1994–1996) was an era of strong economic growth. By comparing these heatmaps, we can have the following insights about product sales and consumer behavior during or after the economic recession.

- First, as shown in Figure 3, the estimated covariance structure in 1995 shows many more dependencies than in 1989. It is likely that more products were sold together in the mid-nineties thanks to the improved economic prospects in the

post-recession years. Given the significant amount of sparsity in the estimated covariance structure in 1989, products were likely purchased individually or in smaller groups due to the economic downturn and a jobless recovery during the recession years.

- Second, Figure 4 shows the different co-purchasing behavior in 1989 versus 1995. Products such as tuna and frozen entrees are often considered the essential grocery products, while soda, cookies and candies are generally sweets or treats. In 1989, we see that cookies and soda products are positively correlated, suggesting they are often bought together, but negatively correlated to the frozen meals. In 1995, after the recession, the sales of sweets are more positively correlated with meals.

- Third, another interesting finding discovered by the M-COAT procedure is the effect of brand loyalty. As shown in Figure 5, in both 1989 and 1995, *Cascade* brand product sales are positively correlated with one another, but negatively correlated with other dish detergent product sales. This suggests some form of brand loyalty where an individual who purchases *Cascade* products is more likely to purchase other *Cascade* products. Notably, brand loyalty may develop over time. In 1989, there does not appear to be a strong dependence structure among paper towel varieties. However, in 1995, we see that the *Bounty* brand product sales are positively correlated with each other, and a similar pattern occurs for the sales of *Scott* brand towels.

6. Conclusion

In this article we propose the M-COAT procedure to robustly estimate the covariance structure of high-dimensional compositional data. The proposed method attains improved accuracy in the presence of nonnormal and heavy-tailed distributions by employing Huber-type M-estimation. We also introduce a cross-validation procedure to choose the tuning parameters of

the proposed M-COAT. Without requiring the Gaussian or sub-Gaussian assumption, we obtain the rates of convergence and signal recovery property under the high-dimensional setting. We provide the theoretical guarantees for the cross-validation procedure, which makes a separate contribution to the literature of Huber-type M-estimation with high-dimensional data. The finite-sample properties of the proposed M-COAT are demonstrated in simulation studies. Finally, we apply the proposed M-COAT in a real application to the Dominick's Finer Foods scanner dataset and show how the M-COAT can give insights about product sales and consumer behavior.

Supplementary Materials

The complete proofs and additional numerical results are presented in a supplementary file.

Acknowledgments

We would like to thank the editor, associate editor and referees for their helpful comments and suggestions.

ORCID

Lingzhou Xue https://orcid.org/0000-0002-8252-0637

References

Aitchison, J. (1982), "The Statistical Analysis of Compositional Data," *Journal of the Royal Statistical Society*, Series B, 44, 139–160. [1090,1091]
——— (1983), "Principal Component Analysis of Compositional Data," *Biometrika*, 70, 57–65. [1090]
Aitchison, J., and Bacon-Shone, J. (1984), "Log Contrast Models for Experiments with Mixtures," *Biometrika*, 71, 323–330. [1090,1091]
Andrews, D. F., and Mallows, C. L. (1974), "Scale Mixtures of Normal Distributions," *Journal of the Royal Statistical Society*, Series B, 36, 99–102. Available at: *https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1974.tb00989.x*. [1095]
Arata, Y., and Onozaki, T. (2017), "A Compositional Data Analysis of Market Share Dynamics," Discussion papers, Research Institute of Economy, Trade and Industry (RIETI). Available at: *https://EconPapers.repec.org/RePEc:eti:dpaper:17076*. [1090]
Avella-Medina, M., Battey, H. S., Fan, J. and Li, Q. (2018), "Robust Estimation of High-Dimensional Covariance and Precision Matrices," *Biometrika*, 105, 271–284. *https://doi.org/10.1093/biomet/asy011*. [1091,1092,1093,1094,1095]

Bickel, P. J., and Levina, E. (2008), "Covariance Regularization by Thresholding," *The Annals of Statistics*, 36, 2577 – 2604. *https://doi.org/10.1214/08-AOS600*. [1093,1094,1095,1096]
Cai, T., and Liu, W. (2011), "Adaptive Thresholding for Sparse Covariance Matrix Estimation," *Journal of the American Statistical Association*, 106, 672–684. *https://doi.org/10.1198/jasa.2011.tm10560*. [1093,1094,1095,1096]
Cambanis, S., Huang, S., and Simons, G. (1981), "On the Theory of Elliptically Contoured Distributions," *Journal of Multivariate Analysis*, 11, 368–385. Available at: *https://www.sciencedirect.com/science/article/pii/0047259X81900828*. [1094]
Cao, Y., Lin, W., and Li, H. (2019), "Large Covariance Estimation for Compositional Data via Composition-Adjusted Thresholding," *Journal of the American Statistical Association*, 114, 759–772. *https://doi.org/10.1080/01621459.2018.1442340*. [1090,1091,1092,1093,1094,1096]
Fan, J., Li, Q., and Wang, Y. (2017), "Estimation of High Dimensional Mean Regression in the Absence of Symmetry and Light Tail Assumptions," *Journal of the Royal Statistical Society*, Series B, 79, 247–265. [1091]
Fry, J. M., Fry, T. R. L., and McLaren, K. R. (2000), "Compositional Data Analysis and Zeros in Micro Data," *Applied Economics*, 32, 953–959. *https://doi.org/10.1080/000368400322002*. [1090]
Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V., and Egozcue, J. J. (2017), "Microbiome Datasets are Compositional: And this is not Optional," *Frontiers in Microbiology*, 8, 2224. Available at *https://www.frontiersin.org/article/10.3389/fmicb.2017.02224* [1090]
Goes, J., Lerman, G., and Nadler, B. (2020), "Robust Sparse Covariance Estimation by Thresholding Tyler's M-estimator," *The Annals of Statistics* 48, 86–110. *https://doi.org/10.1214/18-AOS1793*. [1094,1095]
Huber, P. J. (1964), "Robust Estimation of a Location Parameter," *The Annals of Mathematical Statistics*, 35, 73–101. [1091]
——— (1981), *Robust Statistics*, New York: Wiley. [1091,1092,1093]
Li, D., Srinivasan, A., Xue, L. and Zhan, X. (in press), "Robust Shape Matrix Estimation for High-Dimensional Compositional Data with Application to Microbial Inter-Taxa Analysis," *Statistica Sinica*. [1090,1091,1095]
Lin, W., Shi, P., Feng, R., and Li, H. (2014), "Variable Selection in Regression with Compositional Covariates," *Biometrika*, 101, 785–797. [1090,1091]
Müller, K., and Richter, W.-D. (2019), "On p-generalized Elliptical Random Processes," *Journal of Statistical Distributions and Applications*, 6, 2195–5832. *https://doi.org/10.1186/s40488-019-0090-6*. [1095]
Rothman, A. J., Levina, E., and Zhu, J. (2009), "Generalized Thresholding of Large Covariance Matrices," *Journal of the American Statistical Association*, 104, 177–186. Available at: *http://www.jstor.org/stable/40591909*. [1093]
Srinivasan, A., Xue, L., and Zhan, X. (2021), "Compositional Knockoff Filter for High-Dimensional Regression Analysis of Microbiome Data," *Biometrics*, 77, 984–995. [1090,1091]
Thomas, C., and Aitchison, J. (2005), "Compositional Data Analysis of Geological Variability and Process: A Case Study," *Mathematical Geology*, 37, 753–772. [1090]
Tsilimigras, M. C., and Fodor, A. A. (2016), "Compositional Data Analysis of the Microbiome: Fundamentals, Tools, and Challenges," *Annals of Epidemiology*, 26, 330–335. [1090]
Tyler, D. E. (1987), "A Distribution-Free M-Estimator of Multivariate Scatter," *The Annals of Statistics*, 15, 234–251. *https://doi.org/10.1214/aos/1176350263*. [1094]
Xue, L., Ma, S., and Zou, H. (2012), "Positive-Definite ℓ_1-penalized Estimation of Large Covariance Matrices," *Journal of the American Statistical Association*, 107, 1480–1491. [1092,1093,1096]
Xue, L., and Zou, H. (2012), "Regularized Rank-based Estimation of High-Dimensional Nonparanormal Graphical Models," *The Annals of Statistics*, 40, 2541–2571. *https://doi.org/10.1214/12-AOS1041*. [1092,1095]
——— (2014), "Rank-based Tapering Estimation of Bandable Correlation Matrices," *Statistica Sinica*, 24, 83–100. [1092,1095]