Robust Dependence Modeling for High-Dimensional Covariance Matrices with Financial Applications

Zhe Zhu
Operations Research Center, MIT Sloan School of Management, Cambridge, MA, 02142

Roy E. Welsch
MIT Sloan School of Management, Cambridge, MA, 02142

Abstract

A very important problem in finance is the construction of portfolios of assets that balance risk and reward in an optimal way. A critical issue in portfolio development is how to address data outliers that reflect very unusual, generally non-recurring, market conditions. Should we allow these to have a significant impact on our estimation and portfolio construction process or should they be considered separately as evidence of a regime shift and/or be used to adjust baseline results? In financial asset allocation, a fundamental step is often a mean-variance optimization problem that makes use of the location vector and dispersion matrix of the financial assets. In this paper, we introduce a new high dimensional covariance estimator that is much less sensitive to outliers compared to its classical counterparts. We then apply this estimator to the active asset allocation application, and show that our proposed new estimator delivers better results compared to many existing asset allocation methods. An important bonus is that on our examples, the method has a smaller proportion of stock weights greater than 10% and, in many cases, a higher alpha. Covariance estimation is more challenging than mean estimation and only locally and not globally optimal solutions are available. Our proposed new robust covariance estimator uses a regular vine dependence structure and only pairwise robust partial correlation estimators. The resulting robust covariance estimator delivers high performance for identifying outliers for large high dimensional datasets, has a high breakdown point, and is positive definite. When the full vine structure is not available, we propose using a minimal spanning tree algorithm to replace missing vine structure.

Key words: active asset allocation; portfolio selection; robust estimation; high-dimensional dependence modeling; covariance/correlation estimation; regular vine;

1. Introduction

In this paper, we propose a new model for active asset allocation and credit risk dependence modelling where the covariance matrix is a key quantity under the Capital Asset Pricing Model (CAPM) (Markowitz, 1952), which is a mean-variance optimization framework. Since the covariance matrix has to be estimated from the data, the model is subject to estimation errors. It is well-known that the naive implementation of the CAPM model, with the sample mean and sample covariance matrix as inputs, performs poorly out of sample (Garcia-Alvarez and Luger, 2011). This paper will demonstrate that the mean-variance portfolio selection using our proposed robust covariance estimator outperforms many popular allocation methods in the literature.

Classical sample covariance estimates are very sensitive to outliers, and therefore their robust counterparts are considered to address this problem. However, there are many challenges in the literature of robust covariance estimation for high dimensional datasets, which include low breakdown point (i.e., the maximum percentage of contamination that can be tolerated), computational inefficiency, and no guarantee of positive definiteness. Also, in many applications, there may be prior knowledge about the data dependence structure that can be used in the modelling. For example, in the asset allocation problem, we know which industry group each stock falls under, and we should be able to use such knowledge when estimating the covariance matrix.

Our proposed new robust covariance estimator uses a regular vine dependence structure and only pairwise robust partial correlation estimators. The resulting robust covariance estimator delivers high performance for identifying outliers for large high dimensional datasets, has a high breakdown point, and is positive definite. An important bonus is that on our examples, the method has a smaller proportion of stock weights greater than 10% and, in many cases, a higher alpha. When the full vine structure is not available, we propose using a minimal spanning tree algorithm to replace missing vine structure.

Our application includes 50 stocks picked from different sectors of the S&P 500, and we use our robust covariance estimator for the covariance to be used in the mean-variance optimization step. The result shows that our proposed method outperforms many other methods when measured using the single-index model based on measures of specific risk and portfolio risk as well as the information ratio.

2. Example of a Financial Vine Dependence Structure

Here is a financial example using a basic regular vine structure. Vines are discussed more fully in Section 4. Let s be the stock return for a company selling consumer products (seller), such as Apple. Let s_1 and s_2 be the stock returns of two competing suppliers for s, in this case TSMC and Foxconn.

Figure 2.1 gives a possible regular vine dependence structure for s_1, s_2 and s, and the pairwise correlations and partial correlation are displayed on the edges. For example, the correlation between s_1 and s is 0.8, which shows that supplier 1's company stock return is positively correlated to the seller's company stock return, and this assumption is reasonable. Similarly, supplier 2's company stock return is positively correlated to the seller's company stock return as well. Finally, the partial correlation between s_1 and s_2 given s is -0.8, which shows that supplier 1's company stock return is negatively correlated to supplier 2's company stock return while controlling for the seller's company stock return. This reflects that the two suppliers are competitors for supplying the same company.

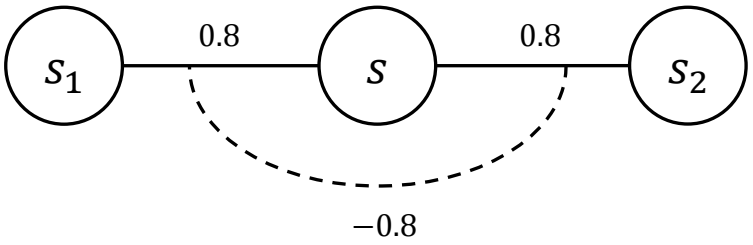

Figure 2.1 Financial example of a regular vine on three variables

3. Robust Covariance Literature Review

There are many existing robust procedures for estimating covariance matrices for multivariate datasets, but they suffer from one or more of the shortcomings mentioned in the previous section.

Multivariate M-estimators use generalized maximum likelihood estimators (MLEs) when estimating covariance matrices. M-estimators are computationally efficient in general. However, for monotone M-estimators, the breakdown point is $1/(p+1)$, where p is the number of variables. Tyler (1987) showed an M-estimator with breakdown being $1/p$, which is a bit higher, but when p is large, such an estimator does not make a difference. Also, S-estimators were introduced by Rousseeuw and Yohai (1984), and Lopuhaa (1989) showed the relation between S-estimators and M-estimators.

Later, multivariate covariance estimators with high breakdown point were introduced. For example, minimum covariance determinant (MCD) and minimum volume ellipsoid (MVE) were presented in Rousseeuw and Leroy (1987). These estimators have many attractive properties, such as high breakdown that can be specified, guaranteed positive definiteness of the covariance estimate, and affine-equivariance. Affine-equivariance is a desirable property of an estimator that any linear transformation of the data is paralleled by the same linear transformation of the estimator. Lopuhaa and Rousseeuw (1991) also

presented the breakdown properties of such affine-equivariant estimators. However, these methods require heavy computation. For example, MCD was not popular until the introduction of Fast-MCD (Rousseeuw and Driessen, 1999), but it is still computationally heavy compared to estimators like M-estimators, and it may get stuck in a local minimum. With a big data set with many variables, MCD may not be computationally feasible.

Finally, there have been discussions of using robust pair-wise covariances as entries for the covariance matrix. Since there are many computationally-efficient robust procedures for estimating pair-wise covariances, the estimation of the covariance matrix is also computationally efficient. However, the resulting covariance matrix may not be positive definite, and this violates the fundamental property of a covariance matrix. Many papers, such as Maronna and Zamar (2002), have investigated techniques to ensure positive-definiteness. The estimation of the covariance matrix is of great interest in many areas of application, including the field of finance. Fan, Fan, and Lv (2008) showed convariance matrix estimation using a factor model.

This paper will introduce a new robust covariance estimator using a dependence structure called a vine. Together with high-breakdown robust regression estimators to compute partial correlations, our proposed method achieves a high breakdown point, guaranteed positive-definiteness of the covariance matrix, as well as a faster run time compared to MCD for datasets with a large number of data points.

4. Formal Definition of a Vine

A vine is a graphical tool for modeling dependence structure in high dimensions, and it was introduced in Bedford and Cooke (2002) and Kurowicka and Cooke (2006). A regular vine is a special case where dependence structures are specified for two variables conditional on some other variables. Regular vines generalize tree structures, and combined with robust statistical techniques, they have proven to be a valuable tool in high-dimensional robust modeling.

Now, we are going to formally define the regular vine structure and its representations.

Definition 4.1 (Vine). $\mathcal{V}$ is a vine on p elements with $\mathcal{E}(\mathcal{V}) = \mathcal{E}_1 \cup \ldots \cup \mathcal{E}_{p-1}$ denoting the set of edges of $\mathcal{V}$ if

1. $\mathcal{V} = \{T_1, \ldots, T_{p-1}\}$;
2. T_1 is a connected tree (Diestel R, 2005) with nodes $N_1 = \{1, \ldots, p\}$, and edges $\mathcal{E}_1$;
3. For $i = 2, \ldots, p-1$, T_i is a tree with nodes $N_i = \mathcal{E}_{i-1}$.

There are usually two graphical ways to represent a vine structure. Figure 4.1 shows the two graphical representations of the same vine. The graph on the left follows directly from Definition 4.1, and it shows all the trees T_i. In particular, T_1 has nodes $N_1 = \{1,2,3,4\}$ and edges $\mathcal{E}_1 = \{(1,2),(2,3),(3,4)\}$; T_2 has nodes $N_2 = \mathcal{E}_1 = \{(1,2),(2,3),(3,4)\}$ and edges $\mathcal{E}_2 = \{((1,2),(3,4)),((2,3),(3,4))\}$ or in shorthand $\{(1,2,3,4),(2,3,4)\}$ by removing the inner brackets; T_3 has nodes $N_3 = \mathcal{E}_2 = \{(1,2,3,4),(2,3,4)\}$ and edges $\mathcal{E}_3 = \{((1,2,3,4),(2,3,4))\}$ or in short hand $\{(1,2,3,4)\}$ The graph on the right is a more compact way of representing the vine structure. It combines all the trees T_i by connecting the edges from the previous level.

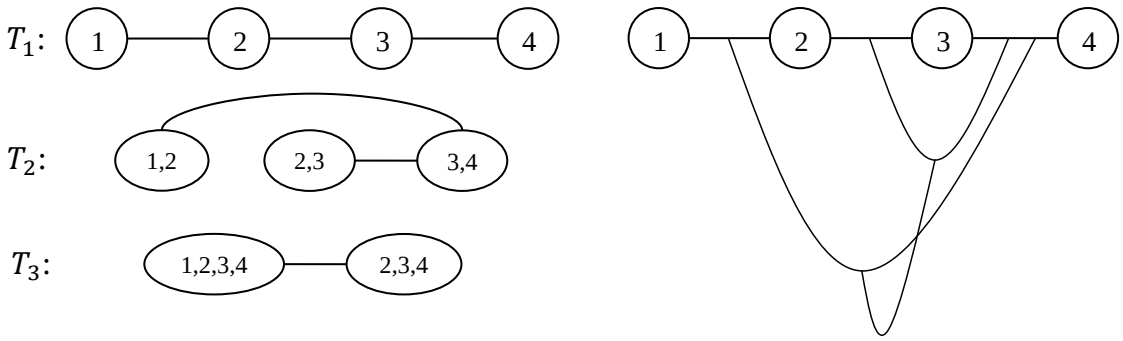

Figure 4.1 Two graphical representations of a vine on four variables

Definition 4.2 (Regular Vine). $\mathcal{V}$ is a regular vine on p elements with $\mathcal{E}(\mathcal{V}) = \mathcal{E}_1 \cup \ldots \cup \mathcal{E}_{p-1}$ denoting the set of edges of $\mathcal{V}$ if

1. $\mathcal{V}$ is a vine;
2. (**proximity**) For $i = 2, \ldots, p-1$, $\{a,b\} \in \mathcal{E}_i$, $\#(a\Delta b) = 2$ where Δ denotes the symmetric difference operator, and # denotes the cardinality of a set
   - $a\Delta b = (a \cup b) \backslash (a \cap b)$

On a regular vine, each edge connecting the nodes a and b corresponds to the dependence structure of the pair $a\Delta b$ conditional on $a \cap b$. The proximity property guarantees that $a\Delta b$ always has exactly two variables.

Following Definition 4.2, the vine in Figure 4.1 is not a regular vine because it violates the proximity condition. Specifically, at T_2, $\{1,2\}\Delta\{3,4\} = \{1,2,3,4\}$, so $\#(\{1,2\}\Delta\{3,4\}) = 4 \neq 2$.

Figure 4.2 shows a sample regular vine with the two graphical representations.

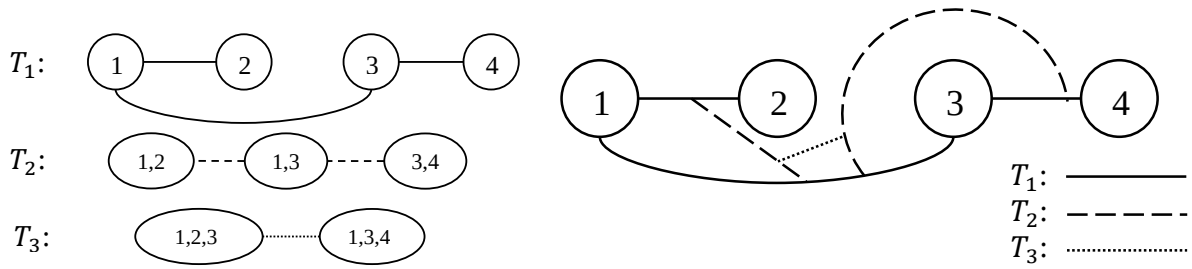

Figure 4.2 Two graphical representations of a regular vine on four variables

There are two special types of regular vine, namely C-vine and D-vine. They have very specific structures, and they are very useful as initial dependence structure models when there is not much prior knowledge about the dependence structure.

Definition 4.3 (C-vine). A regular vine is called a **Canonical** or **C-vine** if each tree T_i has a unique node of degree (Diestel R, 2005) $p - i$ for $i = 1, \dots, p - 2$. The node with the highest degree in T_i is the **root** for $i = 1, \dots, p - 2$.

Figure 4.3 is a sample C-vine on 4 variables. The roots are $\{1\}$ and $\{1,2\}$ for T_1 and T_2 respectively.

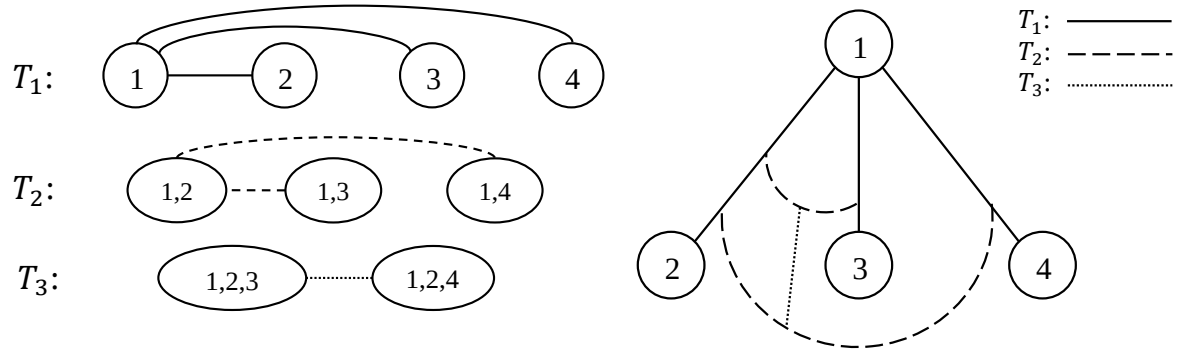

Figure 4.3 Example of a C-vine on four variables and its two graphical representations

Definition 4.4 (D-vine). A regular vine is called a **Drawable** or **D-vine** if each node in T_1 has a degree of at most 2.

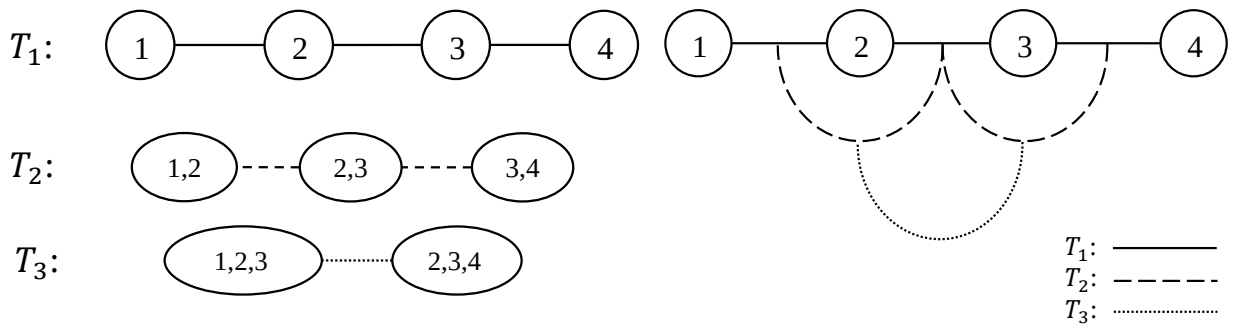

Figure 4.4 Example of a D-vine on four variables and its two graphical representations

Figure 4.4 is a sample D-vine on 4 variables. Each tree is a path, so it is the most drawable vine.

There are regular vines, for example the vine shown in Figure 4.2, which is neither a C-vine nor a D-vine.

Regular vines are very useful in modelling the correlation structure. The vine provides a graphical tree structure of the variables, and in section 6, we will show that each edge in this tree structure has a number associated with it, which is called the partial correlation, and these yield the overall structure.

5. Financial Application

Covariance/correlation estimation is of significant importance in the field of finance. Many areas, including but not limited to asset allocation and active portfolio management, hedging instrument selection, and credit derivative pricing/modeling, require covariance/correlation in the modeling phase, and they are subject to estimation errors because it is well known that most financial data series are not normally distributed and have skewed and non-symmetric tails. As we will demonstrate in an example of asset allocation, our proposed robust correlation estimator is superior to many existing methods.

The Capital Asset Pricing Model (CAPM) is a classical asset allocation method proposed by Markowitz (1952), and this model and its enhanced versions still remain a very prominent approach to asset allocation and active portfolio management (Garcia-Alvarez and Luger, 2011). The CAPM model minimizes the variance of the portfolio with a target return, and therefore the expected returns and the covariance matrix are the inputs for the model. In practice, both the expected return and the covariance matrix have to be estimated from the data, and therefore, the model is subject to estimation errors. It is well-known that the naive implementation of the CAPM model, with the sample mean and sample covariance matrix as inputs, performs poorly out of sample (Garcia-Alvarez and Luger, 2011). Furthermore, according to a study by DeMiguel, Garlappi and Uppal (2009), the CAPM model, together with many enhanced estimation methods for the mean and covariance matrix, does not perform better

than the simple equal-weight portfolio. The equal-weight portfolio simply invests equally in all N stocks under consideration, so it is not subject to estimation errors. This paper will demonstrate that the mean-variance portfolio selection using our proposed robust covariance estimator outperforms many popular allocation methods in the literature, including the equal-weight portfolio.

5.1 Modeling Set-up

Five stocks from each Global Industry Classification Standard (GICS) sector have been selected from the S&P 500 index based on their market capitalizations and availability of data. Therefore, a total of fifty stocks have been chosen for modeling. Historical stock price data were downloaded from Yahoo! Finance using the close price adjusted for dividends and splits.

Here is the list of the GICS sectors and the five stocks chosen from each sector:

1. Consumer Discretionary: MCD, DIS, CMCSA, HD, NKE
2. Consumer Staples: WMT, PG, KO, PEP, MO
3. Energy: XOM, CVX, SLB, COP, OXY
4. Financials: WFC, JPM, C, BAC, AXP
5. Health Care: JNJ, PFE, MRK, ABT, BMY
6. Industrials: GE, UTX, CAT, MMM, BA
7. Information Technology: AAPL, MSFT, IBM, INTC, ORCL
8. Materials: DD, DOW, NEM, MOS, ECL
9. Telecommunications Services: T, VZ, CTL, S, FTR
10. Utilities: SO, D, DUK, NEE, EXC

At beginning of each week, the following optimization problem is solved with different estimated location vectors and dispersion matrices.

$$\begin{aligned} &\text{Minimize} && w'\hat{\Sigma}w \\ &\text{Subject to:} && w'\hat{\mu} = \mu_{target} \\ & && w'e = 1 \\ & && w \geq 0 \end{aligned}$$

where w is the weight vector, $\hat{\mu}$ is the estimated location (mean) vector, $\hat{\Sigma}$ is the estimated dispersion (covariance) matrix, μ_{target} is the target return location for the portfolio, and e is the vector of all ones.

We call this process of recalculation of weights and adjustment of stock allocations weekly rebalance, and we perform such process weekly for the 10-year period (2002-2012). For each weekly rebalance, we use the past 200 weeks of stock return observations for the estimation process. We will consider and compare returns of the following six portfolio allocation rules: Selected vine, D-vine, equal-weight, OGK,

MCD, CAPM and S&P 500. The full derivation of the Selected Vine and D-vine are in Section 6. Obviously, both equal-weight and S&P 500 portfolio allocation rules do not require any model or estimation. The remaining models use the mean-variance optimization model with different estimators for the location (mean) vector and the dispersion (covariance) matrix.

Estimation methods for each allocation rule:

- Selected vine
  - The location vector is estimated using the sample median vector;
  - The dispersion matrix is estimated using the optimally selected vine (Algorithm 6.3) as the partial correlation vine structure and LTS with 50% breakdown as the robust partial correlation estimator.
- D-vine
  - The location vector is estimated using the sample median vector;
  - The dispersion matrix is estimated using D-vine as the partial correlation vine structure and LTS with 50% breakdown as the robust partial correlation estimator.
    - The order of the first level tree for the D-vine is taken to be the same as the order of the stocks specified above.
- OGK
  - The location vector and the dispersion matrix are estimated using OGK (Maronna and Zamar, 2002) with median and median absolute deviation (MAD) for the univariate robust location and scale estimates.
- MCD
  - The location vector and the dispersion matrix are estimated using the Fast MCD algorithm (Rousseeuw and Driessen, 1999) with 50% breakdown.
- CAPM
  - The location vector is estimated using the classical sample mean;
  - The dispersion matrix is estimated using the classical sample covariance.

The target return is the return of the equal-weight portfolio. However, due to the differences in estimation methods for the location vector of the return, the estimated target return under different methods may not be identical. In addition, short selling is not allowed.

Finally, no transaction cost is included. Transaction costs and limits on stock turn-overs will be considered in future research.

5.2 Model Results and Comparisons using index model

The four different covariance estimation techniques result in different weight vectors in the mean-variance optimization step. One observation is that the traditional CAPM method using the sample mean and sample covariance estimators have a higher proportion of individual stock weights that are greater than 10% compared to some other approaches. Also, the MCD method produces a high proportion similar

to the CAPM method, while our proposed selected vine method produces the lowest proportion. This analysis shows that our proposed Selected Vine method does not produce many extreme weights compared to the other approaches, so it is most stable. Table 5.1 shows the detailed comparison figures.

Table 5.1 Proportion of Weights that are Greater Than 10%

Selected Vine	D-Vine	OGK	MCD	CAPM
4.22%	5.69%	5.75%	6.18%	6.15%

We also use the single-index model for model comparisons:

$$r_P - r_f = \alpha + \beta(r_m - r_f) + \epsilon, \epsilon \sim N(0, \sigma^2)$$

where:

- r_P is the portfolio return
- r_f is the risk free return, and the weekly return of the one-month treasury bill is used
- α is the portfolio alpha, estimated by linear regression
- β is the portfolio beta, estimated by linear regression
- r_m is the market return, and S&P 500 is used

Table 5.2 shows the results for different asset allocation methods. Our proposed Selected Vine method has produced the highest alpha, which measures the excess return after taking market risk into account. A high positive alpha is an indication of good performance of an asset allocation method. Also, the beta of a portfolio shows its correlation to the S&P 500 market return, so it measures the systematic risk of the asset portfolio compared to the S&P 500. Our proposed Selected Vine method has one of the highest betas, which indicates that its return moves more closely with the S&P 500 market return and it has more systematic risk compared to other asset allocation methods.

Table 5.2 Statistics of the Single-Index Model for Each Asset Allocation Method

	Selected Vine	D-Vine	Equal Weight	OGK	MCD	CAPM	S&P 500
α	**0.1634%**	0.1091%	0.1477%	0.0931%	0.1073%	0.1019%	0%
β	0.4560	0.3464	0.5074	0.2871	0.2905	0.2846	1

Another measure of the risk-adjusted return of the portfolios is the information ratio, and it is defined as the expected active return divided by the tracking error. This is a good measure for risk-adjusted return because it incorporates standard deviation, which is used to measure risk, into the equation, and given the same amount of risk/standard deviation, the higher the excess return, the better the asset allocation method is. From Table 5.3, the allocation method using the Selected Vine has the highest IR for the realized returns, and therefore, it is considered the best active asset allocation method under the IR criterion.

Table 5.3 Statistics of the Realized Weekly Log-Returns for Each Asset Allocation Method

	Selected Vine	D-Vine	Equal Weight	OGK	MCD	CAPM	S&P 500
Mean	**0.1948%**	0.1402%	0.1793%	0.1241%	0.1382%	0.1328%	0.0329%
s.d.	2.4518%	1.8732%	2.4492%	1.8024%	1.8078%	**1.7820%**	2.6126%
IR	**6.2976%**	4.5309%	6.0290%	3.6750%	4.2554%	4.0413%	N/A

where the information ratio $IR = \frac{E[R_p - R_b]}{\sqrt{Var(R_p - R_b)}}$, and the benchmark return, R_b, is the S&P 500 return.

The asset allocation application has demonstrated that our proposed robust covariance estimation can deliver better results than many existing methods. Also, by estimating the location vector and the dispersion matrix robustly, one can easily identify outliers using the robust version of the Mahalanobis distance, and any data point with a Mahalanobis distance above a certain threshold value would be considered an extreme observation. This is very useful in risk management when modeling the different correlation/dependence structures under normal and crisis situations.

6. Robust Correlation Estimation

In this section, we introduce a new robust correlation estimation method using pairwise partial correlations. We first provide the definition of partial correlation, and then describe the partial correlation vine structure and its relationship with the correlation matrix. Finally, we incorporate a robust estimation technique for partial correlations and use it to construct the robust correlation matrix that guarantees positive-definiteness.

6.1 Partial Correlation Vine Structure

Definition 6.1 (Partial correlation). Consider random variables $X_1, \dots, X_p$.
Let $X^*_{1;3,\dots,p}$ and $X^*_{2;3,\dots,p}$ be the best linear approximations to X_1 and X_2 based on the variables $X_3, \dots, X_p$. Let $Y_1 = X_1 - X^*_{1;3,\dots,p}$, $Y_2 = X_2 - X^*_{2;3,\dots,p}$ be the residuals. Then, the partial correlation between X_1 and X_2 given all other variables, denoted by $\rho_{1,2;3,\dots,p}$ is defined as the ordinary correlation coefficient between Y_1 and Y_2.

Therefore, the partial correlation $\rho_{1,2;3,\dots,p}$ can be interpreted as the correlation between the orthogonal projections of X_1 and X_2 on the plane orthogonal to the space spanned by $X_3, \dots, X_p$. Partial correlations can be computed from correlations using the following recursive formula (Yule and Kendall, 1965):

$$\rho_{1,2;3,\dots p} = \frac{\rho_{1,2;3,\dots,p-1} - \rho_{1,p;3,\dots,p-1} \cdot \rho_{2,p;3,\dots,p-1}}{\sqrt{1-\rho^2_{1,p;3,\dots,p-1}} \cdot \sqrt{1-\rho^2_{2,p;3,\dots,p-1}}} \tag{6.1}$$

Also, the partial correlation can be computed directly from the correlation matrix, Σ. Define $P = \Sigma^{-1}$, then:

$$\rho_{i,j;\{1,\dots p\}\backslash\{i,j\}} = -\frac{p_{ij}}{\sqrt{p_{ii}p_{jj}}} \tag{6.2}$$

Where p_{ij} is the $(i,j)^{\text{th}}$ entry of P.

Theorem 6.1 below provides another method for computing partial correlations.

Theorem 6.1 (Yule and Kendall, 1965). Consider a variable X_i with zero mean, $i = 1, \dots, p$. Let the numbers $b_{i,j;\{1,\dots p\}\backslash\{i,j\}}$ minimize

$$E\left[\left(X_i - \sum_{j:j\neq i} b_{i,j;\{1,\dots,p\}\backslash\{i,j\}} X_j\right)^2\right], i = 1, \dots, p$$

Then, the partial correlation can be computed as

$$\rho_{i,j;\{1,\dots,p\}\backslash\{i,j\}} = sgn\left(b_{i,j;\{1,\dots,p\}\backslash\{i,j\}}\right)\left|b_{i,j;\{1,\dots,p\}\backslash\{i,j\}} b_{j,i;\{1,\dots,p\}\backslash\{i,j\}}\right|^{\frac{1}{2}}$$

Theorem 6.2 (Bedford and Cooke, 2002). For any regular vine on p elements, there is a one-to-one correspondence between the set of $p \times p$ positive definite correlation matrices and the set of partial correlation specifications for the vine.

Therefore, by Theorem 6.2, any assignment of the numbers strictly between -1 and 1 to the edges of a partial correlation regular vine is consistent with a positive definite correlation matrix, and all positive definite correlation matrices can be obtained this way by allowing for all possible vines. The off-diagonal entries of a positive definite correlation matrix must be strictly between -1 and 1 as will be seen in the proof of Theorem 6.3.

It can be verified that the correlation between the i^{th} and j^{th} variables can be computed from the sub-vine generated by the constraint set of the edge whose conditioned set is $\{i,j\}$. The detailed proof is in the Bedford and Cooke (2002) paper.

Definition 6.2 (Partial correlation vine). A partial correlation vine is obtained by assigning a partial correlation ρ_ε, with a value chosen arbitrarily in the interval $(-1,1)$, to each edge ε in $\mathcal{E}(\mathcal{V})$ of the vine defined in Definition 4.1.

Theorem 6.2 shows that there is a bijection between the set of all possible regular partial correlation vines and the set of all (positive definite) correlation matrices.

6.2 Preservation of the Breakdown Point

As mentioned previously, in robust covariance estimation, having a high breakdown point is a favorable, yet difficult, property for a robust estimator. We will first provide a formal definition of the breakdown point, and then prove that with high-breakdown robust partial correlation estimators, which are readily available, we can produce robust covariance estimators with the same high breakdown as those robust partial correlation estimators.

Definition 6.3 (Breakdown point or BP). Let θ be the parameter (or vector of parameters) of interest that ranges over a set Θ. Let $\hat{\theta} = \hat{\theta}(\boldsymbol{x})$ be an estimate defined for samples $\boldsymbol{x} = \{x_1, \dots, x_n\}$. The replacement finite-sample breakdown point of $\hat{\theta}$ at $\boldsymbol{x}$ is the largest proportion ϵ^* of data points that can be arbitrarily replaced by outliers (i.e. contamination of data) without $\hat{\theta}$ leaving a set which is bounded and remains away from the boundary of Θ (Donoho and Huber, 1983). In other words, there exists a closed and bounded set $K \subset \Theta$ such that $K \cap \partial\Theta = \emptyset$ (where $\partial\Theta$ denotes the boundary of Θ), where for any $\epsilon \leq \epsilon^*$ contamination of data, $\hat{\theta}$ remains in the set K.

Roughly speaking, the breakdown point (BP) of an estimate is the largest amount of contamination that the data may contain such that the estimate still gives some information about the underlying parameter. Now, we will show that the BP for the robust correlation estimator is preserved using a partial correlation vine with a robust partial correlation estimator on each edge of the vine.

Theorem 6.3. Given a partial correlation vine structure $\mathcal{V}$ and a robust partial correlation estimate on each edge ε in $\mathcal{E}(\mathcal{V})$ of the vine, the correlation matrix can be derived from the partial correlation estimates using Theorem 6.2. If each robust partial correlation estimator has BP at least ϵ^*, then the resulting correlation matrix also has BP at least ϵ^*.

Proof:

There are $\binom{p}{2}$ correlations to be estimated, and each ranges from -1 to 1 (both inclusive). Therefore, the parameter space for correlations is $\Theta = [-1,1]^{\binom{p}{2}}$. Also, for any regular vine, there are $\binom{p}{2}$ number of edges (Kurowicka, Cooke and Callies, 2006), and therefore, the parameter space for partial correlations on the vine is $\widetilde{\Theta} = [-1,1]^{\binom{p}{2}}$. For each edge ε of the partial correlation vine, let $\widetilde{\Theta}_\varepsilon = [-1,1]$ be the parameter space for the partial correlation on that edge.

First we will show that, given a partial correlation vine structure $\mathcal{V}$, there exists a continuous function mapping from the correlation matrix to the partial correlations on the edges $\mathcal{E}(\mathcal{V})$ of the vine.

According to Formula 6.2 and its generalization, if Σ is the correlation matrix with $P = \Sigma^{-1}$, then $\rho_{i,j;\{1,\dots n\}\{i,j\}} = -\frac{p_{ij}}{\sqrt{p_{ii}p_{jj}}}$. Because both the matrix inversion and the computation of partial correlation

from the inverse matrix are continuous mappings, there exists a continuous function mapping, denoted by f, from the correlation matrix to the partial correlations on the edges $\mathcal{E}(\mathcal{V})$.

Now, since each robust partial correlation estimator has BP at least ϵ^*, by Definition 6.3 (BP), for each robust partial correlation estimator $\tilde{\theta}_\varepsilon$ on the edge ε, there exists a closed and bounded set $\widetilde{K}_\varepsilon \subset \widetilde{\Theta}_\varepsilon$ such that $\widetilde{K}_\varepsilon \cap \partial\widetilde{\Theta}_\varepsilon = \emptyset$, and for any $\epsilon \leq \epsilon^*$ contamination of data, $\tilde{\theta}_\varepsilon$ remains in the set $\widetilde{K}_\varepsilon$. In this case, $\widetilde{\Theta}_\varepsilon = [-1,1]$, and $\partial\widetilde{\Theta}_\varepsilon = \{-1,1\}$, so $-1 \notin \widetilde{K}_\varepsilon$ and $1 \notin \widetilde{K}_\varepsilon$.

Let $\widetilde{K} = \prod_\varepsilon \widetilde{K}_\varepsilon$, the Cartesian product of the sets. Let $K = f^{-1}(\widetilde{K})$, and without loss of generality (WLOG), say $K \subseteq \widetilde{\Theta}$, and if not, we can take $K = f^{-1}(\widetilde{K}) \cap \widetilde{\Theta}$, and the argument will still follow. Now, since the inverse image of a closed set under continuous mapping is also closed (Rudin, 1976), K is closed. Note that $\hat{\theta} \in K$.

Finally, we have to prove that $K \cap \partial\Theta = \emptyset$, and we will prove this by showing that none of the off-diagonal elements of the correlation matrix can be -1 or 1. Suppose, to reach a contradiction, that there is an off-diagonal element of the correlation matrix that is -1 or 1. Since we can reorder the variables, WLOG, let's say such element is ρ_{12}. Therefore, the determinant of the upper left 2-by-2 corner is:

$$\det\begin{bmatrix} 1 & \rho_{12} \\ \rho_{12} & 1 \end{bmatrix} = 1 - \rho_{12}^2 = 0$$

By Sylvester's criterion, the correlation matrix is not positive definite. However, this violates Theorem 6.2. Therefore, we can conclude that none of the off-diagonal elements of the correlation matrix can be -1 or 1, and hence $K \cap \partial\Theta = \emptyset$. ■

Theorem 6.3 shows that BP of the final robust correlation matrix estimator is inherited from the robust partial correlation estimators. There are many robust partial correlation estimators with high BP, and the resulting robust correlation matrix estimator will have high BP as well.

6.3 Algorithms for Robust Estimation

Now, we will propose a robust estimation procedure for computing robust covariance/correlation matrices following the principles described above.

We apply robust regression estimators, such as the MM estimator with high breakdown and efficiency (Yohai VJ, 1987), to estimate the robust partial correlations on the vine. Then according to Theorem 6.3, we can estimate the correlation matrix robustly using these robust partial correlation estimates. These robust regression estimators are generally faster to compute compared to MCD, and we will demonstrate this in the Benchmark section. Finally, the robust covariance matrix is constructed using the robust correlation matrix and the robust covariance for each individual variable.

Theorem 6.2 guarantees the construction of a correlation matrix from partial correlation vines, but there is no algorithm specified how to do so in real-life applications. There are special regular vine

structures where algorithms can be specified for such a process, and we will provide algorithms for both the C-vine and D-vine.

Algorithm 6.1 (Correlation Computation for C-vine)

For C-vine, the partial correlations are of the form $\rho_{i,j;1,\dots,i-1}$, where $i < j$.

From Formula 6.1, we have $\forall k < i$

$$\rho_{i,j;1,\dots,k} = \frac{\rho_{i,j;1,\dots,k-1} - \rho_{k,i;1,\dots,k-1} \cdot \rho_{k,j;1,\dots,k-1}}{\sqrt{1-\rho^2_{k,i;1,\dots,k-1}} \cdot \sqrt{1-\rho^2_{k,j;1,\dots,k-1}}} \tag{6.3}$$

Rearrange terms to get a recursive formula for $\rho_{i,j;1,\dots,k-1}$:

$$\rho_{i,j;1,\dots,k-1} = \rho_{i,j;1,\dots,k} \cdot \sqrt{1-\rho^2_{k,i;1,\dots,k-1}} \cdot \sqrt{1-\rho^2_{k,j;1,\dots,k-1}} + \rho_{k,i;1,\dots,k-1} \cdot \rho_{k,j;1,\dots,k-1} \tag{6.4}$$

Therefore, for $k = i-1, i-2, \dots, 1$, we can recursively compute $\rho_{i,j;1,\dots,k-1}$, and when $k = 1$, $\rho_{i,j} = \rho_{i,j;1,\dots,k-1}$.

This algorithm has $O(p^3)$ runtime.

Here is an example how this is done for $p = 4$. For a C-vine with 4 variables, robust partial correlation estimators give estimates for the following partial correlations: $\rho_{1,2}$, $\rho_{1,3}$, $\rho_{1,4}$, $\rho_{2,3;1}$, $\rho_{2,4;1}$ and $\rho_{3,4;1,2}$. We can compute the remaining correlations with the following formulas:

$$\rho_{2,3} = \rho_{2,3;1}\sqrt{(1-\rho^2_{1,2})(1-\rho^2_{1,3})} + \rho_{1,2}\rho_{1,3}$$

$$\rho_{2,4} = \rho_{2,4;1}\sqrt{(1-\rho^2_{1,2})(1-\rho^2_{1,4})} + \rho_{1,2}\rho_{1,4}$$

$$\rho_{3,4;1} = \rho_{3,4;1,2}\sqrt{(1-\rho^2_{2,3;1})(1-\rho^2_{2,4;1})} + \rho_{2,3;1}\rho_{2,4;1}$$

$$\rho_{3,4} = \rho_{3,4;1}\sqrt{(1-\rho^2_{1,3})(1-\rho^2_{1,4})} + \rho_{1,3}\rho_{1,4}$$

Algorithm 6.2 (Correlation Computation for D-vine)

For D-vine, the partial correlations are of the form $\rho_{i,j;i+1,\dots,j-1}$, where $i < j$.

By Anderson (1958)

$$\rho_{i,j} = r_1'(i,j)R_2(i,j)^{-1}r_3(i,j) + \rho_{i,j;i+1,\dots,j-1}D_{i,j} \quad (6.5)$$

Where

- $r_1'(i,j) = \left(\rho_{i,i+1}, \dots, \rho_{i,j-1}\right)$
- $r_3'(i,j) = \left(\rho_{j,i+1}, \dots, \rho_{j,j-1}\right)$
- $R_2(i,j) = \begin{pmatrix} 1 & \rho_{i+1,i+2} & \cdots & \rho_{i+1,j-1} \\ \rho_{i+2,i+1} & 1 & & \rho_{i+2,j-1} \\ \vdots & & \ddots & \vdots \\ \rho_{j-1,i+1} & \rho_{j-1,i+2} & \cdots & 1 \end{pmatrix}$
- $D_{i,j}^2 = \left(1 - r_1'(i,j)R_2(i,j)^{-1}r_1(i,j)\right)\left(1 - r_3'(i,j)R_2(i,j)^{-1}r_3(i,j)\right)$

We compute $\rho_{i,j}$ in the following order: i steps up from 3 to p, and for each i, j steps down from $i-2$ to 1.

Here is an example how this is done for $p = 4$. For a D-vine with 4 variables, robust partial correlation estimators give estimates for the following partial correlations: $\rho_{1,2}$, $\rho_{2,3}$, $\rho_{3,4}$, $\rho_{1,3;2}$, $\rho_{2,4;3}$ and $\rho_{1,4;2,3}$. We can compute the remaining correlations in the order: $\rho_{1,3}$, $\rho_{2,4}$, $\rho_{1,4}$.

We can first calculate $\rho_{1,3}$ because

- $r_1'(1,3) = \left(\rho_{1,2}\right)$ is known;
- $r_3'(i,j) = \left(\rho_{2,3}\right)$ is known;
- $R_2(i,j) = (1)$ is known;
- $\rho_{1,3;2}$ is known.

Using Formula 6.5, we can compute $\rho_{1,3}$.

Similarly, we can then calculate $\rho_{2,4}$.

Finally, we can calculate $\rho_{1,4}$ because

- $r_1'(1,4) = \left(\rho_{1,2}, \rho_{1,3}\right)$ is known;
- $r_3'(1,4) = \left(\rho_{2,4}, \rho_{3,4}\right)$ is known;
- $R_2(1,4) = \begin{pmatrix} 1 & \rho_{2,3} \\ \rho_{2,3} & 1 \end{pmatrix}$ is known;
- $\rho_{1,4;2,3}$ is known.

Matrix inversion operations, each with $O(p^3)$ runtime, are required, so naively implementing an algorithm for D-vine requires $O(p^5)$ runtime. However, one can cleverly use block matrix inversion techniques to improve the runtime to $O(p^4)$.

Algorithm 6.1 can only be applied when the C-vine structure is pre-specified, so it is necessary to choose a correct C-vine structure initially. Also, if the variable labels in the vine are permuted, a different

robust covariance estimate may be found depending on the robust partial correlation estimator. A similar statement applies to Algorithm 6.2 and the D-vine.

Therefore, we need to discuss how to select a regular vine robustly. We will use an inductive method to select the optimal tree at each level. A similar method has been proposed in the literature (Dissmann J, Brechmann E C, Czado C, Kurowicka D, 2012), but our method does not require any specification of the copula structure, and it is more robust against outliers.

Algorithm 6.3 (Robust Vine Selection). Consider random variables $X_1, \dots, X_p$. For a given robust partial correlation estimator, let $\tilde{\rho}_{i,j;D}$ be the robust partial correlation between X_i and X_j given all variables in the set D. (Note: $i \notin D$ and $j \notin D$, and when $D = \emptyset$, we shorten the notation to be $\tilde{\rho}_{i,j}$)

1. Select the first level tree, T_1, by solving the following maximum spanning tree problem:
$$\max_{\{i,j\}\ in\ spanning\ tree} \sum |\tilde{\rho}_{i,j}|$$
2. For $k = 2, \dots, p-1$ do
3. Select tree T_k by solving the following maximum spanning tree problem over all pairs $\{i, j|D\}$ that can be part of tree T_k, i.e. all edges satisfying the proximity condition (see **Definition 6.2**):
$$\max_{\{i,j|D\}\ in\ spanning\ tree} \sum |\tilde{\rho}_{i,j;D}|$$

The vine structure determined by Algorithm 6.3 will be regular vine, but it is not necessarily a C-vine or D-vine.

If a professional has prior knowledge about the first several level trees, Algorithm 6.3 can be easily adapted by carrying out the inductive steps from the first unspecified level tree. In this case, the professional would specify the first several level trees, say T_1,…,T_l, and then apply Step 3 of Algorithm 6.3 for $k = l+1, \dots, p-1$.

7. Benchmark

There have been many proposals made to estimate covariance matrices robustly. Some authors in the field of robust estimation have emphasized the importance of high breakdown point, and Theorem 6.3 guarantees that our estimator can achieve a high breakdown point. Others have emphasized speed for large p (Maechler and Stahel, 2009), and our method has an $O(p^3)$ runtime using a C-vine. However, there is usually one important criterion that is overlooked when evaluating a robust covariance estimator, namely the effectiveness of identifying the outlier part. The Barrow Wheel Benchmark (Maechler and Stahel, 2009) has been proposed as a benchmark to evaluate such a criterion.

Example 7.1 (Barrow Wheel). The "Barrow Wheel" distribution is a mixture of a flat multivariate normal distribution contaminated with a portion $\epsilon = 1/p$ (p is the dimension) of gross errors concentrated near a one-dimensional subspace. Formally, Let

$$G_0 = (1-\epsilon)\cdot \mathcal{N}_p\left(\mathbf{0}, \text{diag}(\sigma_1^2, 1, \dots, 1)\right) + \epsilon \cdot H$$

where $\mathcal{N}_p\left(\mathbf{0}, \text{diag}(\sigma_1^2, 1, \dots, 1)\right)$ is a p-dimensional multivariate normal distribution with mean $\mathbf{0}$ and diagonal covariance matrix with entries $(\sigma_1^2, 1, \dots, 1)$. H is the distribution of $\boldsymbol{Y}$, where $Y^{(1)}$ has a symmetric distribution with $\left(Y^{(1)}\right)^2 \sim \chi^2_{p-1}$ (i.e. Chi-squared distribution with $p-1$ degrees of freedom) and is independent of $Y^{(2)}, \dots, Y^{(p)} \sim \mathcal{N}_p\left(\mathbf{0}, \sigma_2^2 I_{p-1}\right)$. Then, this distribution is rotated such that the $X^{(1)}$ axis points in the space diagonal direction $(1,1,\dots,1)$, and the components are rescaled to obtain G, which is the Barrow Wheel distribution.

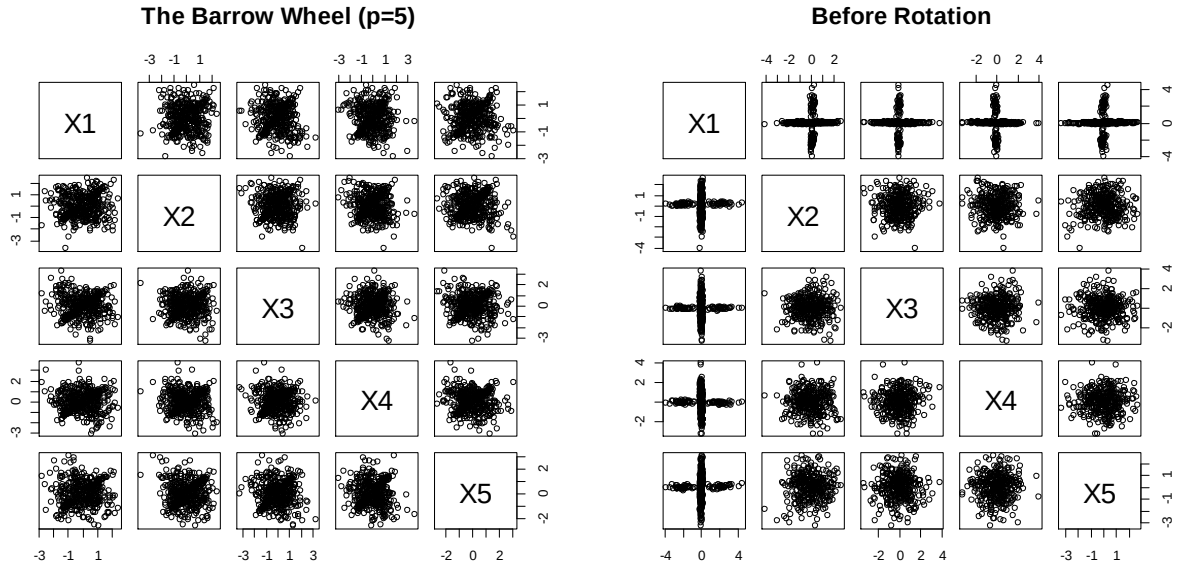

Figure 7.1 Pairwise scatter plot for the barrow wheel distribution (p=5), after rotation (left) and before rotation (right)

Figure 7.1 shows the data generated by the "Barrow Wheel" distribution with $p = 5$, $n = 400$, $\sigma_1 = 0.05$ and $\sigma_2 = 0.1$, where n is the number of data points, and p is the number of variables.

The "good" portion of data is $\mathcal{N}_p\left(\mathbf{0}, \text{diag}(\sigma_1^2, 1, \dots, 1)\right)$, i.e. the wheel, and the outlier portion is H, i.e. the axle. An effective robust estimator should be able capture the wheel and not the axle. We will use the concept of condition number or kappa to measure a robust estimator's ability to identify and isolate outliers.

Definition 7.1 (Condition number or kappa). For any positive definite matrix S, let $\lambda_{max}(S)$ and $\lambda_{min}(S)$ be the largest and smallest eigenvalues of S respectively. Then, the condition number or kappa of S is defined as:

$$\kappa(S) = \frac{\lambda_{max}(S)}{\lambda_{min}(S)}$$

With the convention of setting $\sigma_1 < 1$, the theoretical condition number of the covariance matrix for the wheel should be $1/\sigma_1^2$. An effective robust estimator should capture most of the data in the wheel, and hence the condition number of the estimated covariance matrix should be close to the theoretical value of $1/\sigma_1^2$.

As an experiment, we set $\sigma_1 = 0.05$ and $\sigma_2 = 0.1$ for the remainder of this section and, therefore, the theoretical condition number of the covariance matrix for the wheel is $1/0.05^2 = 400$. Maechler and Stahel (2009) provided box-plots (Figure 7.2, permission to use the plots granted by Stahel) of condition numbers for a list of robust covariance estimators, and they demonstrated that many cheap "robust" estimators, such as BACON (Billor, Hadi and Velleman, 2000), are as ineffective as the classical covariance estimator with condition number very close to 1. Also, methods like MCD and MVE perform really well, but they are computationally time-consuming. OGK (Maronna and Zamar, 2002) methods perform better than those cheap "robust" methods, but not by a significant margin.

We performed the Barrow Wheel benchmark using our proposed robust covariance estimator. We used C-vine, D-vine and the optimally selected vine as the regular vine structures, and we used both the MM-estimator and Least Trimmed Squares (LTS) as the robust regression estimator for the partial correlations. To make the result comparable with the ones in Figure 7.2, we used the same parameters: $n = 100$, $p = 5$, $n_{sim} = 50$, where n and p are the sample size and number of variables in each simulation respectively, and n_{sim} is the total number of simulations. We also included the "Oracle" estimator, which gives the estimation of the covariance matrix knowing which portion is the wheel part. The box-plots in Figure 7.3 show that our estimators perform much better than many robust covariance estimators in Figure 7.2. For example, the OGK estimator in Figure 7.2, which uses pair-wise estimation techniques for constructing the robust covariance, has a median condition number well below 100, and our proposed estimators, as shown in Figure 7.3, all have condition number above 200. The D-Vine (LTS) estimator even has similar performance to the Oracle one.

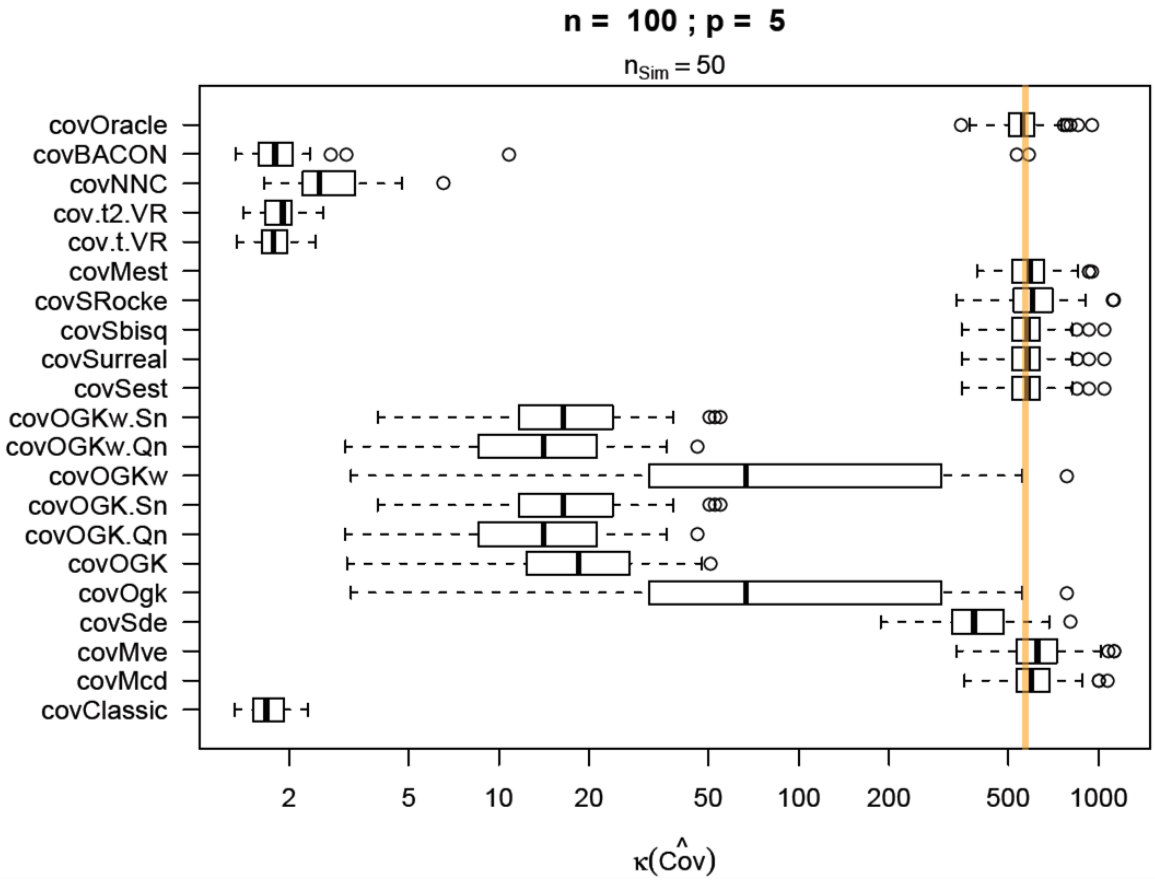

Figure 7.2 Boxplot of the benchmark results from Maechler and Stahel (2009)'s talk comparing twenty different robust estimators

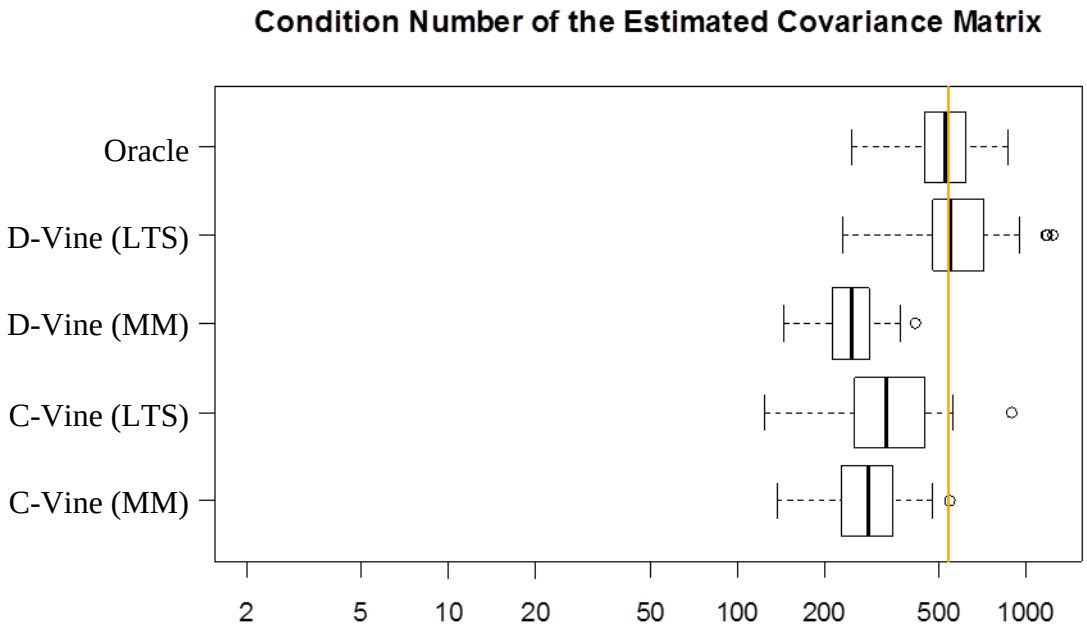

Figure 7.3 Boxplots of the benchmark results for our four robust estimators

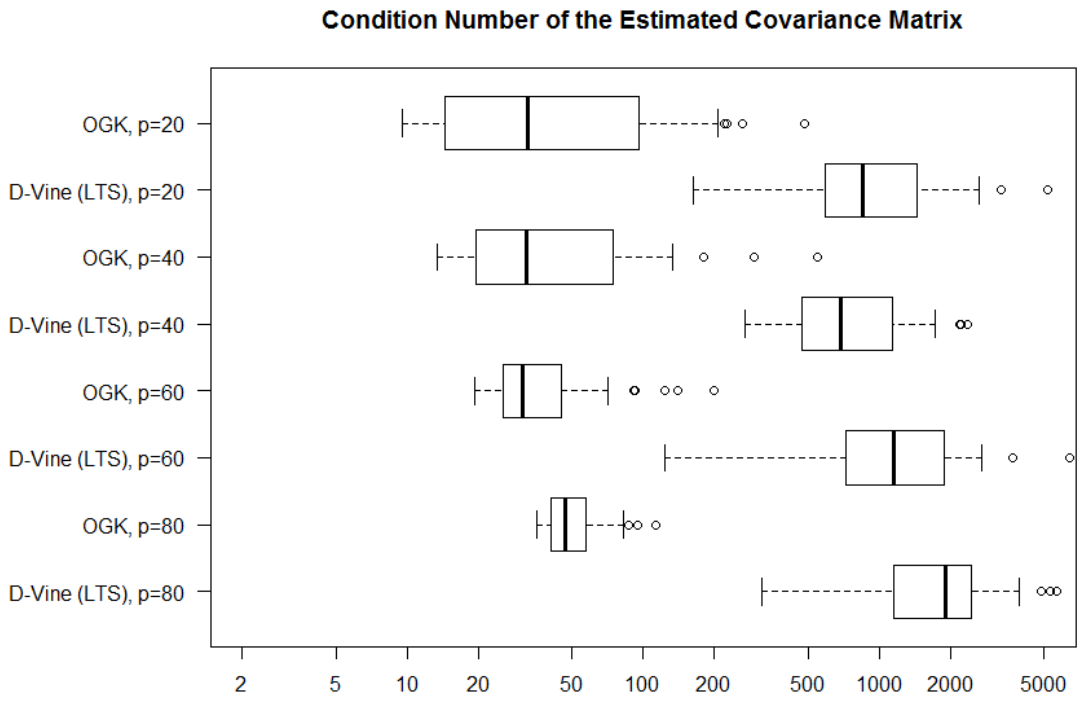

Figure 7.4 Boxplots of the benchmark results for different p under OGK and D-Vine (LTS) robust estimators

Figure 7.4 shows a comparison of the benchmark result for different dimensions in terms of condition number between the proposed D-Vine (LTS) estimator and the OGK estimator. We tested for an increasing p from 20 to 80. It shows in the figure that our proposed estimator is able to better capture the central portion of the data compared to the OGK estimator as measured by the condition number for all those tested dimensions.

Figure 7.5 shows a comparison of computation times for D-Vine (LTS) and MCD, where n_{sim} is the total number of simulations. It is clear that, with large n, D-Vine (LTS) runs much faster than MCD. Also, the difference grows much more significantly as n becomes larger. Finally, note that our proposed robust covariance estimator may not necessarily be affine equivariant.

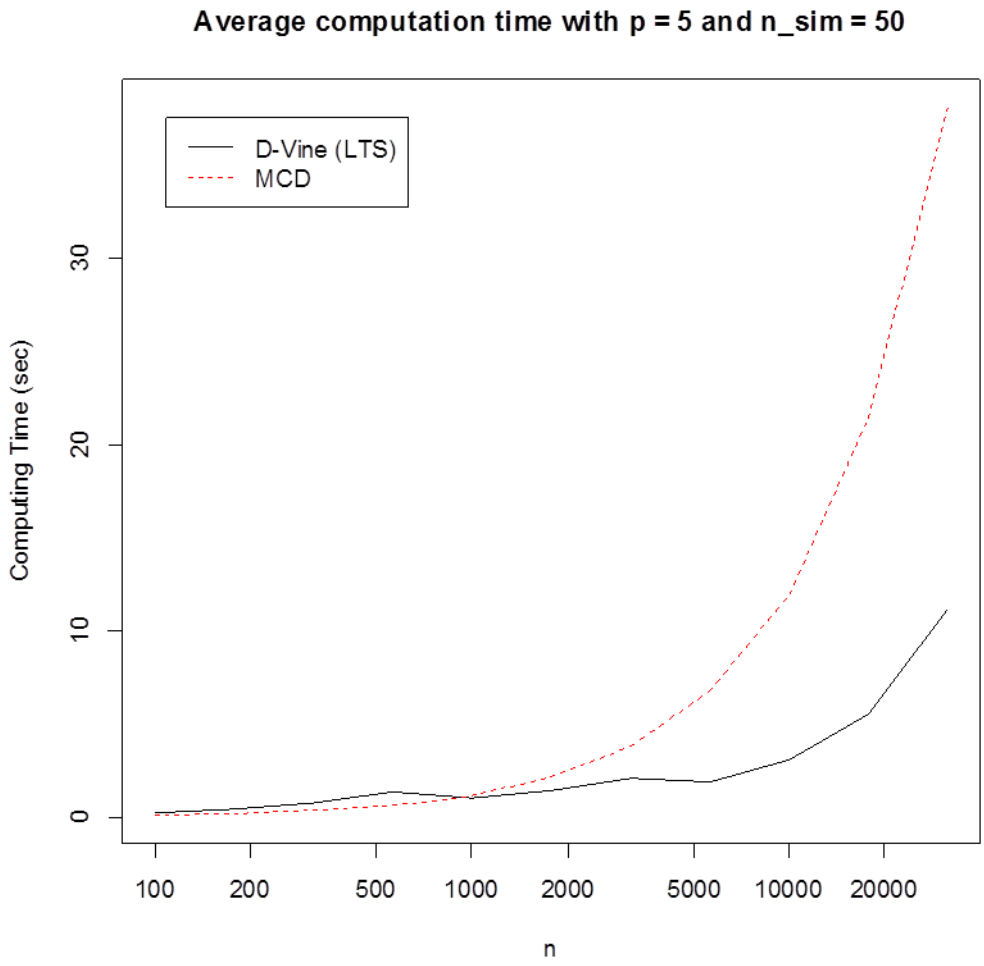

Figure 7.5 Comparison of computation time between D-vine (LTS) and MCD for different sample size n ranging from 100 to 30000

8. Conclusion

This paper introduced a new robust covariance estimation method for multivariate data with important applications in financial portfolio optimization, outlier detection, and other areas. We demonstrate the benefit of using our proposed estimation method in the application of active asset allocation with fifty stocks from the S&P 500 index and comparison of seven different asset allocation methods. Our proposed Selected Vine method for estimating the covariance matrix has more stable results by producing a smaller proportion of individual stock weights that are greater than 10%. Also, under the single-index model, the Selected Vine method has the highest alpha among all considered methods. Finally, using the information ratio criterion, our proposed Selected Vine method generates the highest information ratio for the realized returns while taking the risk of the asset portfolio, which is measured by the standard deviation, into consideration. This method combines the regular vine structure with robust partial correlation estimation

techniques and it allows the use of prior information such as knowing which industry group each stock falls under and the supplier relationships among the target companies. When the vine structure is not completely available, a minimal spanning tree algorithm is used to replace the missing portion. It guarantees the positive definiteness of the correlation/covariance matrix, which is an important property for robust covariance estimators using pair-wise covariance estimates. The breakdown point is also preserved in the estimation process, and this allows the building of multivariate estimators with high breakdown point. A Barrow Wheel benchmark shows that this new approach effectively captures the good portion of the data while improving the computing time for large datasets.

References

Anderson TW (1958) An Introduction to Multivariate Statistical Analysis. *Wiley*, NewYork.

Bedford T and Cooke R (2002) Vines--a new graphical model for dependent random variables. *Annuals of Statistics*, Volume 30, Number 4 (2002), 1031-1068.

Billor N, Hadi AS and Velleman PF (2000) BACON: blocked adaptive computationally efficient outlier nominators. *Computational Statistics & Data Analysis* 34, 279-298.

DeMiguel V, Garlappi L and Uppal R (2009) Optimal versus naïve diversification: how inefficient is the 1/N portfolio strategy?. *Review of Financial Studies* 22:1915-53.

Diestel R (2005) Graph Theory (3rd Ed.), Berlin, New York: Springer-Verlag.

Dissmann J, Brechmann E C, Czado C, Kurowicka D (2012) Selecting and estimating regular vine copulae and application to financial returns. *Computational Statistics & Data Analysis*, 59, 52-69, 2013.

Donoho DL and Huber PJ (1983) The notion of breakdown point. *A Festschrift for Erich L. Lehmann*, P.J. Bickel, K.A. Doksum and J.L. Hodges, (eds.), 157-184, Belmont, CA: Wadsworth.

Fan J, Fan Y and Lv J (2008), High dimensional covariance matrix estimation using a factor model. *Journal of Econometrics*, 147, 186-197.

Garcia-Alvarez L and Luger R (2011) Dynamic correlations, estimation risk, and portfolio management during the financial crisis. Working Paper. CEMFI. Madrid.

Hampel FR, Ronchetti EM, Rousseeuw PJ and Stahel WA (1986), Robust Statistics. The Approach Based on Influence Functions, *Wiley*, New York, 1986.

Kurowicka D, Cooke RM and Callies U (2006), Vines Inference. *Brazilian Journal of Probability and Statistics*, 20, pp. 103-120.

Kurowicka D and Cooke RM (2006). Uncertainty Analysis with High Dimensional Dependence Modelling. *Wiley*, Chichester.

Kurowicka D and Joe H (2011). Dependence Modeling: Vine Copula Handbook. World Scientific Publishing Co. Pte. Ltd.

Lopuhaa HP and Rousseeuw PJ (1991), Breakdown properties of affine equivariant estimators of multivariate location and covariance matrices, *Annuals of Statistics*, 19, 229-248.

Lopuhaa HP (1989), On the relation between S-estimators and M-estimators of multivariate location and covariance, *Annuals of Statistics*, 17, 1662-1683

Maechler M and Stahel W (2009) Robust Scatter Estimators - The Barrow Wheel Benchmark. *ICORS 2009*, Parma.

Markowitz H (1952), Portfolio Selection, *The Journal of Finance*, Vol. 7, No. 1. (Mar., 1952), pp. 77-91

Maronna RA, Martin DR and Yohai VJ (2006), Robust Statistics: Theory and Methods. *Wiley Series in Probability and Statistics*.

Maronna RA and Zamar RH (2002), Robust estimates of location and dispersion for high‐dimensional data sets. *Technometrics*, 44:307‐317.

Rousseeuw PJ and Driessen KV (1999), A fast algorithm for the minimum covariance determinant estimator. *Technometrics* 41, 212–223.

Rousseeuw PJ and Leroy AM (1987), Robust Regression and Outlier Detection. *Wiley*.

Rousseeuw PJ and Yohai VJ (1984), Robust regression by means of S-estimators. *Robust and Nonlinear Time Series Analysis. Lecture Notes in Statist.* 26, 256-272. Springer, New York.

Rudin W (1976), Principles of Mathematical Analysis. *McGraw-Hill*.

Tyler DE (1987), A distribution-free M-estimator of multivariate scatter. *The Annals of Statistics*, 15, 234–251.

Yohai VJ (1987), High breakdown-point and high efficiency robust estimates for regression. *The Annuals of Statistics*, Vol. 15, No. 20, 642-656.

Yule GU and Kendall MG (1965) An Introduction to the Theory of Statistics. 14th Ed, *Charles Griffin & Co.*, London