A Semiparametric Graphical Modelling Approach for Large-Scale Equity Selection

Han Liu* John Mulvey† Tianqi Zhao‡

September 23, 2015

Abstract

We propose a new stock selection strategy that exploits rebalancing returns and improves portfolio performance. To effectively harvest rebalancing gains, we apply ideas from elliptical-copula graphical modeling and stability inference to select stocks that are as independent as possible. The proposed elliptical-copula graphical model has a latent Gaussian representation; its structure can be effectively inferred using the regularized rank-based estimators. The resulting algorithm is computationally efficient and scales to large datasets. To show the efficacy of the proposed method, we apply it to conduct equity selection based on a 16-year healthcare stock dataset and a large 34-year stock dataset. Empirical tests show that the proposed method is superior to alternative strategies including a principal component analysis based approach and the classical Markowitz strategy based on the traditional buy-and-hold assumption.

Keyword: Equity selection; Machine learning; Graphical model; Rebalancing gains; Semiparametric methods; Elliptical copula; Stability selection; Markowitz strategy

1 Introduction

We propose a new stock selection strategy to improve the performance of investment selection tactics by harvesting "rebalancing gains". This concept has been well documented in both theory and practice. There has been a large literature (Dempster et al., 2008, 2010; Luenberger, 2013; Mulvey and Kim, 2009; Mulvey et al., 2001, 2004, 2007) that shows the benefits of rebalancing a portfolio as compared with the traditional buy-and-hold rule embedded in the classical Markowitz portfolio model. In particular, we can evaluate the level of rebalancing gains when stocks are approximated as correlated geometric Brownian motions.

*Department of Operations Research and Financial Engineering, Princeton University, Princeton, NJ 08544, USA; e-mail: hanliu@princeton.edu.

†Department of Operations Research and Financial Engineering, Princeton University, Princeton, NJ 08544, USA; e-mail: mulvey@princeton.edu.

‡Department of Operations Research and Financial Engineering, Princeton University, Princeton, NJ 08544, USA; e-mail: tianqi@princeton.edu.

The most desirable environment for achieving rebalancing gains requires several components. First, transaction costs (including market impact costs) must be relatively small, for example, via actively traded futures markets. Second and importantly, the returns of the instruments should be as independent as possible. Third, the instruments must display a positive drift. As an idealized example, suppose that we are able to discover ten investment tactics with the same growth level ρ , covariance matrix $\mathbf{\Sigma}$, and independent returns (as specified by white noise terms). Ignoring the transaction costs, rebalancing gains are strictly a function of the volatility of the individual instruments. Table 1 shows the level of rebalancing gains for several levels of volatility (Luenberger, 2013).

Table 1.1: Level of rebalancing gains for stylized example

Volatility	Rebalancing gains
20%	1.8%
40%	7.2%
60%	16.2%

Clearly, rebalancing gains are ideally suited to markets possessing low correlation, relatively independent returns, and high volatility. To some degree, these characteristics are part and parcel of commodity markets (Mulvey, 2012; Gorton et al., 2012; Gorton and Rouwenhorst, 2004; Higgs and Worthington, 2008).

There are several issues regarding practical implementation of a rebalancing strategy (Mulvey and Kim, 2009). The first issue is to find a set of relatively independent assets. This task is challenging and unlikely to be perfectly achieved in the real world since financial instruments are correlated. For example, all the stocks in the market are exposed to a degree of systemic risk, and exhibit co-movements with the overall market. As a second issue, correlations change over time. For example, financial instruments exhibit higher level of correlation during downside regimes than upside regimes.

To handle these challenges, we propose a new machine learning framework to construct equity portfolios whose members are as independent as possible. Our method exploits new ideas of elliptical-copula modeling, latent Markovian graphical models, and stability selection.

Copula models are popular in quantitative finance (Embrechts et al., 2003; Cherubini et al., 2004; Dias and Embrechts, 2004; Costa Dias, 2004). They are particularly useful for modeling dependency structures among financial instruments. For example, the Gaussian copula has attracted much interest because of its simplicity and mathematical elegance (Li, 2000). However, its flaw for not being able to accurately model tail dependence has been discovered during financial crises. In this paper, we exploit the more realistic elliptical-copula distribution family to model stock data. Unlike the Gaussian copula model, the elliptical-copula model is more general and can depict nontrivial tail dependence. In the statistical machine learning literature, the elliptical-copula model has been shown to be equivalent to the "transelliptical" model (Details are provided in Section 2.3).

We define a notion of an elliptical-copula graphical model which provides a graphical representation of multivariate elliptical distribution. To interpret the graph, we develop a hierarchical Gaussian representation for any elliptical distribution. Such a latent Gaussian representation allows

us to interpret the elliptical graphical model as a latent Markov random field. It also provides a formal mechanism to select a subset of stocks that are "as independent as possible".

We exploit a regularized rank-based estimator to depict the structure of the latent Gaussian Markov random field of a large number of stock returns. This graph encodes the conditional independence relationship of the latent Gaussian random variables. Next, we propose an efficient path-following algorithm that obtains a sequence of graph estimates corresponding to different sparsity levels. This algorithm is then embedded within a stability selection framework to obtain a subset of stocks. The selected stocks form an equally-weighted portfolio for investment.

We apply the proposed method on two stock datasets, showing that it significantly outperforms the classical buy-and-hold rule that underlies the Markowitz strategy. In addition, our graphical model outperforms a principal component analysis based strategy.

The rest of the paper is organized as follows: Section 2 describes the proposed stock selection method. In particular, §2.3 introduces the elliptical-copula graphical model and its hierarchical interpretation. §2.4 and §2.5 describe the estimator and graph estimation algorithm. §2.6 introduces the stability selection method, and §2.7 describes the complete algorithm. In Section 3 we provide some theoretical guarantees of our stock selection procedure. In Section 4 we apply the proposed method to stock data and evaluate the performance of rebalancing strategy compared with other strategies, showing the potential advantages of the proposed selection procedure.

Notation. Capital italic letters denote random variables (e.g., return of one stock X_1) and bold italic letters to denote random vectors (e.g., a random vector is denoted as $\boldsymbol{X}$). All matrices are denoted by normal bold letters (e.g., $\mathbf{A}$). For any square matrix $\mathbf{A}$ and $\mathbf{B}$, $\|\mathbf{A}\|_1 := \sum_{j,k} |\mathbf{A}_{jk}|$ is the elementwise ℓ_1-norm, $\|\mathbf{A}\|_{\mathsf{F}} = \sqrt{\sum_{j,k} \mathbf{A}_{jk}^2}$ is the Frobenius norm and $\langle \mathbf{A}, \mathbf{B} \rangle = \mathrm{Tr}(\mathbf{A}^T \mathbf{B})$ is the matrix inner product. Let $\mathbb{I}\{\cdot\}$ be the indicator function.

2 A Graphical Model Approach for Selection

We employ new tools from probabilistic graphical models and stability selection to construct a portfolio that can effectively exploit rebalancing returns. More specifically, we adopt the elliptical-copula graphical models for multivariate stock selection and develop efficient computational algorithms for parameter learning and graph estimation.

2.1 Preliminaries on Undirected Graphical Models

The undirected graphical model is a powerful tool for modeling the relationships among a large number of random variables. In this subsection, we introduce a simplest notion of undirected graphical model — Gaussian graphical model. Next, we extend the approach to the more sophisticated elliptical-copula graphical model.

Let $\boldsymbol{X} = (X_1, \ldots, X_d)^T \in \mathbb{R}^d$ be a d-dimensional Gaussian random vector $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Define an undirected graph $G = (V, E)$ with the vertex set $V := \{X_1, \ldots, X_d\}$ and edge encoded in the adjacency matrix $E \in \{0,1\}^{d \times d}$. We denote $E_{jk} = 1$ if there is an edge connecting X_j and X_k, and $E_{jk} = 0$ otherwise. $\boldsymbol{X}$ is Markov to G if $E_{jk} = 0$ if and only if the variables X_j and X_k are conditionally independent given the rest of variables, denoted by a subvector $\boldsymbol{X}_{\setminus\{j,k\}}$. Dempster (1972) shows that under the Gaussian distribution, the graph G is completely characterized by the

sparsity pattern of the inverse covariance matrix, also known as the precision matrix $\boldsymbol{\Theta} = \boldsymbol{\Sigma}^{-1}$. In particular, $E_{jk} = 0$ if and only if $\boldsymbol{\Theta}_{jk} = 0$.

The Gaussian graphical model, though simple and elegant, is unrealistic for our purposes. Specifically, stock returns are known to be heavy-tailed with nontrivial tail dependence, while the Gaussian distribution fails these properties. For example, in downside regimes, stock returns tend to drop drastically as many investors simultaneously seek refugee in safer assets such as cash or bonds, whereas in upside regimes, such collective behavior is less likely to be observed and stock returns are more stable. In the next subsection, we relax the Gaussian modeling assumptions.

2.2 Elliptical Copula Distribution family

A *copula* is a joint cumulative distribution function of random variables that have uniform marginal distributions. For any random vector $\boldsymbol{X} \in (X_1, \ldots, X_d)^T \in \mathbb{R}^d$ with marginal distribution functions $F_1, \ldots, F_d$, a key observation is that the transformed random variable $F_j(X_j)$ has a uniform distribution on $[0, 1]$ for any $j = 1, \ldots, d$. We define the copula for $\boldsymbol{X}$ as a function $C : [0, 1]^d \mapsto [0, 1]$ such that

$$C(u_1, u_2, \ldots, u_d) := \mathbb{P}\big(F_1(X_1) \leq u_1, \ldots, F_d(X_d) \leq u_d\big).$$

Sklar's theorem shows that the joint distribution function F of $\boldsymbol{X}$ can be expressed using its copula function and marginal distributions, i.e.,

$$F_{X_1,\ldots,X_d}(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d)). \tag{2.1}$$

Equation (2.1) shows that any joint multivariate distribution contains two ingredients: the marginal distribution and the copula function. In particular, the marginal distributions captures the marginal information of each individual component in $\boldsymbol{X}$, while the copula function captures the interrelationships among these components.

The concept of copula is not new. It has been extensively employed for market data calibration, stress tests, credit risk evaluation, derivatives pricing, etc. However, the way we apply copula to portfolio analysis is new: we apply the elliptical-copula distribution family to model the stock returns and define graphical models based on the elliptical copula.

To rigorously define elliptical-copula distribution family, we start with a formal definition of an elliptical distribution. To this end, we define $\boldsymbol{Y}$ to be elliptically distributed if its characteristic function $\Psi_{\boldsymbol{Y}}$ can be written as

$$\Psi_{\boldsymbol{Y}}(\boldsymbol{t}) := \mathbb{E}[e^{i\boldsymbol{t}^T\boldsymbol{X}}] = \exp(i\boldsymbol{t}^T\boldsymbol{\mu})\psi(\boldsymbol{t}^T\boldsymbol{\Sigma}\boldsymbol{t}),$$

where $\boldsymbol{\mu}$ is the location parameter (also the mean if it exists), $\boldsymbol{\Sigma}$ is a positive definite scatter matrix, and $\psi(\cdot)$ is specific to each elliptical random variable and is often called the "characteristic generator". Hence an elliptical distribution is uniquely determined by the above three ingredients, and we denote

$$\boldsymbol{Y} \sim E_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \psi),$$

for any random vector $\boldsymbol{Y}$ that follows an elliptical distribution. The density of $\boldsymbol{Y}$ has the following from: $f_{\boldsymbol{Y}}(\boldsymbol{y}) \propto h\big((\boldsymbol{y} - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{\mu})\big)$, where h is uniquely determined by ψ.

Significantly, the elliptical distribution family strictly extends the Gaussian family and contains many heavy-tailed distributions, for example, multivariate t-distribution, Cauchy, logistic, Kotz,

symmetric Pearson type-II and type-VII distributions. Intuitively, the elliptically distribution family includes all distributions whose density function has elliptical contours.

Next, we introduce the elliptical-copula distribution family:

Definition 2.1 (Elliptical Copula Distribution Family)**.** *The elliptical-copula distribution family contains all distributions with elliptical-copulas.*

Any distribution $\boldsymbol{X}$ in the elliptical-copula family is generated by combining an elliptical-copula and possibly non-elliptical marginal distributions. In fact, Liu et al. (2012b) proved that for any $\boldsymbol{X} = (X_1, \ldots, X_d)$ that belongs to the elliptical-copula family, there is a set of latent variables $\boldsymbol{Y} = (Y_1, \ldots, Y_d)$ corresponding to $\boldsymbol{X}$ with an elliptical distribution and zero mean. Concretely, there exists a set of strictly increasing functions $g_1, \ldots, g_d$ such that $(g_1(X_1), \ldots, g_d(X_d)) \sim E_d(\mathbf{0}, \boldsymbol{\Sigma}_0, \psi)$ where $\mathrm{diag}(\boldsymbol{\Sigma}_0) = \mathbf{I}_d$ so that the model is identifiable. Hence the elliptical-copula family is also called the transelliptical family, and can be represented as

$$\boldsymbol{X} \sim TE_d(\boldsymbol{\Sigma}_0, \psi; g_1, \ldots, g_d).$$

Here, $\boldsymbol{\Sigma}_0$ is the latent correlation matrix, as it is the correlation matrix of the latent random vector $\boldsymbol{Y}$. Modeling stock returns via elliptical copulae has many advantages: the elliptical-copula distribution can naturally depict heavy-tailed, asymmetric data, and tail dependencies – an important property of stock return data.

2.3 Elliptical Copula Graphical Model and its Latent Gaussian Representation

Let $\boldsymbol{X} := (X_1, \ldots, X_d)^T$ be an elliptical-copula distribution, denoted by $\boldsymbol{X} \sim TE_d(\boldsymbol{\Sigma}_0, \psi; g_1, \ldots, g_d)$. We define a graph G based on the sparsity pattern of the latent inverse correlation matrix $\boldsymbol{\Theta} := \boldsymbol{\Sigma}_0^{-1}$. There is an edge connecting nodes X_j and X_k if and only if $\boldsymbol{\Theta}_{jk} = 1$. Given any semiparametric transelliptical distribution, such a graph G is always well-defined. The key question is: how shall we interpret the obtained graph? Herein, we present a hierarchical latent Gaussian representation of the elliptical-copula distribution. Such a representation allows us to lever existing results developed under the Gaussian graphical model.

The elliptical-copula graphical model can be viewed as a three-layer generative model (Figure 2.1). The first layer encompasses all the observed elliptical-copula (or transelliptical) random variables. The second layer includes a set of latent elliptical variables, each of which can generate a corresponding observed transelliptical variable by marginal monotone transformation. The third layer includes a set of Gaussian random variables which could generate the second-layer elliptical random variables by global stochastic scaling. In particular, the three layers share the same scatter matrix $\boldsymbol{\Sigma}_0$, thus they have the same graph structure (since the graph is completely endowed by the sparsity pattern of $\boldsymbol{\Sigma}_0^{-1}$). In the following we provide more detailed interpretations of the graph at different layers.

- We start from the second layer – the latent elliptical variables. By definition, G is the sparsity pattern of the latent precision matrix $\boldsymbol{\Theta}$, which is the inverse of the correlation matrix $\boldsymbol{\Sigma}_0$ of the latent random vector $\boldsymbol{Y}$, where $Y_j = g_j^{-1}(X_j)$ for $j = 1, \ldots, d$. Hence G encodes the

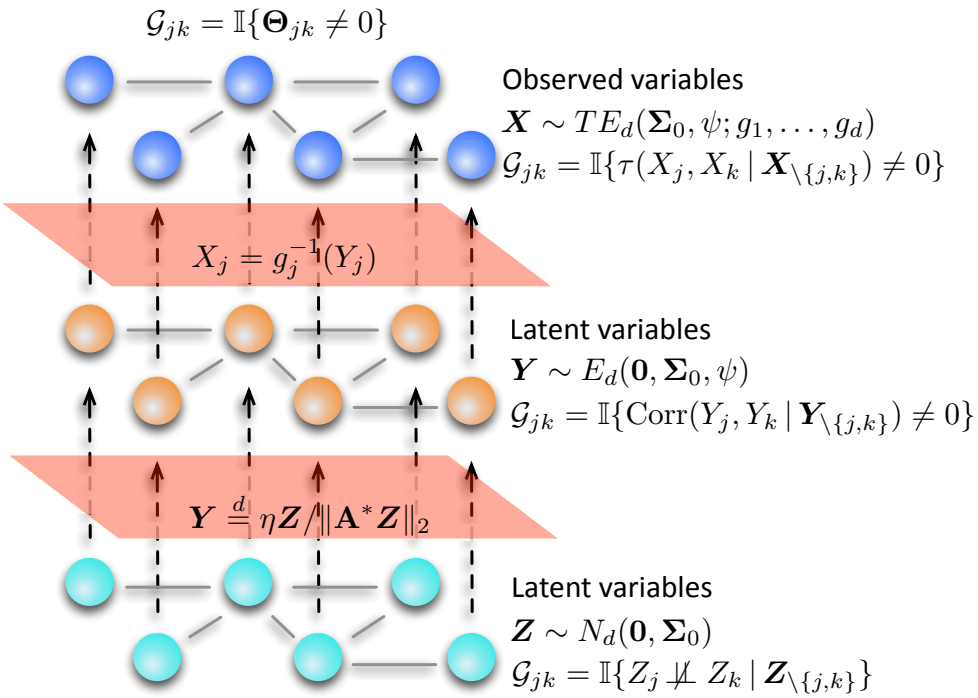

Figure 2.1: Latent hierarchical framework for elliptical-copula graphical model. All three layers share the same graph structure, which is defined by $\mathcal{G}_{jk} = \mathbb{I}\{\boldsymbol{\Theta}_{jk} \neq 0\}$, but have different interpretations.

correlation information of the latent variables in $\boldsymbol{Y}$. By definition, $\boldsymbol{Y} \sim E_d(\boldsymbol{0}, \boldsymbol{\Sigma}_0, \psi)$. In the proof of Lemma 3.7 in Liu et al. (2013), it is shown that

$$\boldsymbol{Y}_J \,|\, \boldsymbol{Y}_{J^c} \sim E_{|J|}(\widetilde{\boldsymbol{\mu}}, [\boldsymbol{\Theta}_{JJ}]^{-1}, \widetilde{\psi}). \tag{2.2}$$

for some $\widetilde{\boldsymbol{\mu}} \in \mathbb{R}^{|J|}$ and $\widetilde{\psi}$. As $\boldsymbol{\Theta}_{JJ}$ is diagonal if and only if $[\boldsymbol{\Theta}_{JJ}]^{-1}$ is diagonal, thus $\boldsymbol{\Theta}_{JJ}$ being diagonal is equivalent to $\boldsymbol{Y}_J$ being pairwise uncorrelated given $\boldsymbol{Y}_{J^c}$. When $J = \{j, k\}$, this implies that $\boldsymbol{\Theta}_{jk} = 0$ if and only if Y_j and Y_k are uncorrelated conditional on $\boldsymbol{Y}_{\setminus\{j,k\}}$. Hence we have $E_{jk} = \mathbb{I}\{\mathrm{Corr}(Y_j, Y_k \,|\, \boldsymbol{Y}_{\setminus\{j,k\}}) \neq 0\}$.

- We next consider the first layer – the observed transelliptical variables. For the observed random variables, $\boldsymbol{\Sigma}_0$ does not directly characterize the correlation between random variables in $\boldsymbol{X}$. This is because correlation matrix is not invariant under marginal transformation. Instead, we consider the following rank-based correlation statistic:

$$\tau_{jk} := \mathbb{P}\big((X_j - \widetilde{X}_j)(X_k - \widetilde{X}_k) > 0\big) - \mathbb{P}\big((X_j - \widetilde{X}_j)(X_k - \widetilde{X}_k) < 0\big) \tag{2.3}$$

where $\widetilde{\boldsymbol{X}} = (\widetilde{X}_1, \ldots, \widetilde{X}_d)$ is an independent copy of $\boldsymbol{X}$. τ_{jk} is the population version of Kendall's tau statistic. It is not difficult to see that τ_{jk} is invariant to monotone marginal transformation, i.e., $\tau_{jk} = \mathbb{P}\big((Y_j - \widetilde{Y}_j)(Y_k - \widetilde{Y}_k) > 0\big) - \mathbb{P}\big((Y_j - \widetilde{Y}_j)(Y_k - \widetilde{Y}_k) < 0\big)$, which shows that the rank correlation of the observed randoms variable is the same as that of the latent elliptical random variables. Lindskog et al. (2003) and Liu et al. (2013) showed that τ_{jk}

has the following simple relationship with $[\mathbf{\Sigma}_0]_{jk}$ for elliptically distributed random variables: $\tau_{jk} = (2/\pi)\arcsin\big([\mathbf{\Sigma}_0]_{jk}\big)$. Such a relationship also holds for the conditional version:

$$\tau(X_j, X_k \,|\, \boldsymbol{X}_{\backslash\{j,k\}}) = \tau(Y_j, Y_k \,|\, \boldsymbol{Y}_{\backslash\{j,k\}}) = \frac{2}{\pi}\arcsin\big(\big[[\mathbf{\Theta}_{\{j,k\}}]^{-1}\big]_{jk}\big),$$

where we employed the fact that $\boldsymbol{Y}_{\{j,k\}} \,|\, \boldsymbol{Y}_{\backslash\{j,k\}} \sim E_2(\widetilde{\boldsymbol{\mu}}, [\mathbf{\Theta}_{\{j,k\}}]^{-1}, \widetilde{\psi})$ given by (2.2). Thus, $\mathbf{\Theta}_{jk} = 0$ if and only if X_j and X_k are conditionally uncorrelated under rank association given $\boldsymbol{X}_{\backslash\{j,k\}}$, i.e., $E_{jk} = \mathbb{I}\{\tau(X_j, X_k \,|\, \boldsymbol{X}_{\backslash\{j,k\}}) \neq 0\}$.

- Lastly, we consider the third layer – the latent Gaussian variables. For any elliptically distributed random vector $\boldsymbol{Y} \sim E_d(\boldsymbol{\mu}, \mathbf{\Sigma}, \psi)$, it has a latent Gaussian representation (Liu et al., 2012b). Concretely, suppose $\mathbf{\Sigma} = \mathbf{A}\mathbf{A}^T$, then we have

$$\boldsymbol{Y} \stackrel{d}{=} \boldsymbol{\mu} + \eta\boldsymbol{Z}/\|\mathbf{A}^*\boldsymbol{Z}\|_2,$$

where $\boldsymbol{Z} \sim N_d(\mathbf{0}, \mathbf{\Sigma})$. $\eta \geq 0$ is a random variable that is independent of $\boldsymbol{Z}$, and $\mathbf{A}^*$ is the Moore-Penrose pseudo-inverse of $\mathbf{A}$. Hence the third level is composed of the latent random variables $\boldsymbol{Z} = (Z_1, \ldots, Z_d)$. By conditional Gaussian formula we have that $\boldsymbol{Z}_J \,|\, \boldsymbol{Z}_{J^c} \sim N(\mathbf{\Sigma}_{JJ^c}[\mathbf{\Sigma}_{J^cJ^c}]^{-1}\boldsymbol{Z}_{J^c}, [\mathbf{\Theta}_{JJ}]^{-1})$, therefore $\mathbf{\Theta}_{jk} = 0$ if and only if Z_j and Z_k are conditionally uncorrelated given $\boldsymbol{Z}_{\backslash\{j,k\}}$, and, by the Gaussian property, if and only if they are conditionally independent. Hence $E_{jk} = \mathbb{I}\{Z_j \not\perp\!\!\!\perp Z_k \,|\, \boldsymbol{Z}_{\backslash\{j,k\}}\}$.

This hierarchical framework characterizes the interdependency structure between stocks. Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ denote the observed historical returns of d stocks across n periods. To construct an "optimal" rebalancing portfolio, we seek to select stocks that are "as independent as possible". For this, we first estimate the undirected elliptical-copulal graph that underlies d random variables (stock returns), and then select a set of stocks each of which is "isolated" from the rest of the graph. The estimation and selection procedure is described in the next subsection. Notably, our method scales with very large dimension d. This feature is crucial, as it allows us to choose from a large pool of stocks, so the selected stocks could be very different from each other and potentially close to being independent.

2.4 Rank-based Estimator

As G is defined by the sparsity pattern of $\mathbf{\Theta}$, our graph estimating procedure is based on three steps: we first estimate the latent correlation matrix using an estimator $\widehat{\mathbf{\Sigma}}$, then solve for a sparse approximation $\widehat{\mathbf{\Theta}}$ of the inverse of $\widehat{\mathbf{\Sigma}}$, and lastly we estimate G via $\widehat{\mathbf{\Theta}}$ and a stability based method. In this and the following two sections, we describe the three steps, respectively.

From the previous subsection, it is not difficult to see that the rank-based Kendall's tau statistic is a good estimator for $\mathbf{\Sigma}_0$, because of its two desirable properties: (i) it is invariant under the monotone marginal transformation, so that the observed $\boldsymbol{X}$ and the latent variables in $\boldsymbol{Y}$ share the same Kendall's tau statistic; (ii) it has a simple relationship with the correlation matrix under the elliptical distribution, i.e., $[\mathbf{\Sigma}_0]_{jk} = \sin(\frac{\pi}{2}\tau_{jk})$. Liu et al. (2012a) and Xue and Zou (2012) exploited the same estimator for inferring the latent correlation matrix $\mathbf{\Sigma}$ in the Gaussian copula family (also known as nonparanormal family), which is an extension of the Gaussian distribution using marginal

monotone transformation. It was further extended to the elliptical-copula (transelliptical) family by Han and Liu (2012) and Liu et al. (2012b).

Formally, our estimator is defined as follows:

$$\widehat{\mathbf{\Sigma}}_{jk} = \sin\Big(\frac{\pi}{2}\widehat{\tau}_{jk}\Big), \tag{2.4}$$

where

$$\widehat{\tau}_{jk} := \frac{2}{n(n-1)} \sum_{1\le i<i'\le n} \Big[\mathbb{I}\{(X_{ij} - X_{i'j})(X_{ik} - X_{i'k}) > 0\} - \mathbb{I}\{(X_{ij} - X_{i'j})(X_{ik} - X_{i'k}) < 0\}\Big] \tag{2.5}$$

is the estimated Kendall's tau statistic in replacement of the population version (2.3).

Liu et al. (2012a) showed that $\widehat{\mathbf{\Sigma}}$ is a consistent estimator of the latent correlation matrix $\mathbf{\Sigma}$ (with respect to elementwise sup norm $\|\cdot\|_{\max}$), even when the order of d is nearly exponentially larger than n. Concretely, we have with probability at least $1 - 1/\max\{d, n\}$,

$$\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}_0\|_{\max} \le 2.45\pi\sqrt{\frac{\log d}{n}}.$$

We next introduce a statistical procedure that efficiently estimate the precision matrix $\mathbf{\Theta}$ by plugging in the Kendall's tau estimate $\mathbf{\Sigma}$. With a properly chosen tuning parameter λ, the estimated graph has a sparse representation, which facilitates the stock selection procedure. Moreover, when the true graph is indeed sparse, the estimated graph is also consistent. Thus, our selection procedure has a solid statistical foundation.

2.5 Smooth-Projected Neighborhood Pursuit Algorithm for Graph Estimation

We next introduce a statistical procedure that efficiently estimate sparsity of the precision matrix $\mathbf{\Theta}$ using the Kendall's tau estimate $\widehat{\mathbf{\Sigma}}$. Note that although $\mathbf{\Theta}$ is the inverse of $\mathbf{\Sigma}$, naively inverting $\widehat{\mathbf{\Sigma}}$ is not a good procedure, as $\widehat{\mathbf{\Sigma}}$ is not invertible when dimension d is larger than n, and even if it is, its inverse is not sparse.

There are many procedures for transelliptical graph estimation using the Kendall's statistics. We adopt the neighborhood pursuit method introduced by Meinshausen and Bühlmann (2006). The intuition originates from the Gaussian graphical model. Recall from §2.3 that for Gaussian random variables, $Z_j \,|\, \boldsymbol{Z}_{\setminus j} \sim N(\mathbf{\Sigma}_{j,\setminus j}[\mathbf{\Sigma}_{\setminus j,\setminus j}]^{-1}\boldsymbol{Z}_{\setminus j}, [\mathbf{\Theta}_{jj}]^{-1})$. Therefore we have

$$\mathbb{E}[Z_j \,|\, \boldsymbol{Z}_{\setminus j}] = \mathbf{\Sigma}_{j,\setminus j}[\mathbf{\Sigma}_{\setminus j,\setminus j}]^{-1}\boldsymbol{Z}_{\setminus j} = \mathbb{E}[\boldsymbol{\beta}^T\boldsymbol{Z}_{\setminus j}], \tag{2.6}$$

where $\boldsymbol{\beta} = [\mathbf{\Sigma}_{\setminus j,\setminus j}]^{-1}\mathbf{\Sigma}_{\setminus j,j}$. (2.6) is exactly a regression problem. Moreover, by matrix block inverse formula, we have $\boldsymbol{\beta} = \mathbf{\Theta}_{\setminus j,j}/\mathbf{\Theta}_{jj}$. This shows that $\boldsymbol{\beta}$ encodes the sparsity pattern of $\mathbf{\Theta}_{\setminus j,j}$. Hence the neighborhood pursuit algorithm is based on regressing Z_j onto other variables using an ℓ_1 penalized minimization method known as LASSO regression: let $\widehat{\mathbf{R}} \in \mathbb{R}^{d\times d}$, then we solve

$$\widehat{\mathbf{R}}_{\setminus j,j} = \underset{\boldsymbol{\beta}\in\mathbb{R}^{d-1}}{\operatorname{argmin}} \frac{1}{n}\|\mathbb{Z}_{\cdot,j} - \mathbb{Z}_{\cdot,\setminus j}\boldsymbol{\beta}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_1,$$

where $\mathbb{Z} = (\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n)^T \in \mathbb{R}^{n\times d}$ is the design matrix and λ is a tuning parameter. The ℓ_1-regularization is imposed so that the solution has a sparse representation, a common variable

selection method used in statistics. Solving the above problems for all j and letting $\mathrm{diag}(\widehat{\mathbf{R}}) = \mathbf{0}$, we derive an estimator $\widehat{\mathbf{R}}$ for $\boldsymbol{\Theta}$.

If we expand the objective function we find that it can be rewritten as follows:

$$\widehat{\mathbf{R}}_{\backslash j,j} = \underset{\boldsymbol{\beta}\in\mathbb{R}^{d-1}}{\operatorname{argmin}} \boldsymbol{\beta}^T \widehat{\mathbf{S}}_{\backslash j,\backslash j}\boldsymbol{\beta} - 2\widehat{\mathbf{S}}_{\backslash j,j}\boldsymbol{\beta} + \lambda\|\boldsymbol{\beta}\|_1.$$

where $\widehat{\mathbf{S}} = \mathbb{Z}^T\mathbb{Z}/n$ is the Pearson's sample covariance matrix. To solve for the elliptical-copula (transelliptical) graph, a natural idea is to replace $\widehat{\mathbf{S}}$ by the Kendall's tau covariance matrix $\widehat{\boldsymbol{\Sigma}}$. Although statistically this method enjoys good property, it is computationally very challenging: $\widehat{\boldsymbol{\Sigma}}$ is possible to be not positive definite, and when it is not, the above minimization with $\widehat{\boldsymbol{\Sigma}}$ plugged in is not a convex problem, which is very difficult to solve.

To handle this challenge, we apply the smooth-projected neighborhood pursuit algorithm proposed by Zhao et al. (2012b) for estimating Gaussian copula (nonparanormal) graphical models. In this algorithm, we first obtain a positive definite estimator $\widetilde{\boldsymbol{\Sigma}}$ which is the projection of the Kendall's tau estimate $\widehat{\boldsymbol{\Sigma}}$. Ideally, we want to project the matrix under the elementwise max norm $\|\cdot\|_{\max}$. However, the projection problem under this norm is difficult to compute due to the non-smooth nature of the max norm. Therefore, we first define a smooth norm that approximates the elementwise max norm: for any matrix $\mathbf{A}$, define

$$\|\mathbf{A}\|_\infty^\rho = \max_{\|\mathbf{B}\|_1\le 1} \big\{\langle \mathbf{B}, \mathbf{A}\rangle - \frac{\rho}{2}\|\mathbf{B}\|_{\mathsf{F}}^2\big\},$$

In particular, when $\rho = 0$, $\|\cdot\|_\infty^\rho$ simply becomes $\|\cdot\|_{\max}$ by norm duality. We then define the smooth projection $\widetilde{\boldsymbol{\Sigma}}$ of $\widehat{\boldsymbol{\Sigma}}$ as

$$\widetilde{\boldsymbol{\Sigma}} = \underset{\boldsymbol{\Sigma}\in \boldsymbol{S}\mathbb{R}_+^d}{\operatorname{argmin}} \|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_\infty^\rho, \tag{2.7}$$

where $\boldsymbol{S}\mathbb{R}_+^d$ denotes the set of symmetric positive definite $d\times d$ matrices. Now that $\widetilde{\boldsymbol{\Sigma}}$ is positive definite and is an approximation of $\widehat{\boldsymbol{\Sigma}}$, we can define the estimate $\widehat{\mathbf{R}}$ by solving

$$\widehat{\mathbf{R}}_{\backslash j,j} = \underset{\boldsymbol{\beta}\in\mathbb{R}^{d-1}}{\operatorname{argmin}} \boldsymbol{\beta}^T \widetilde{\boldsymbol{\Sigma}}_{\backslash j,\backslash j}\boldsymbol{\beta} - 2\widetilde{\boldsymbol{\Sigma}}_{\backslash j,j}\boldsymbol{\beta} + \lambda\|\boldsymbol{\beta}\|_1, \tag{2.8}$$

for all $j = 1,\ldots,d$ and setting $\mathrm{diag}(\widehat{\mathbf{R}}) = \mathbf{0}$.

The rest of this subsection is devoted to finding efficient algorithms for solving (2.7) and (2.8). The optimization problem (2.7) can be solved by the fast proximal gradient algorithm (Zhao et al., 2012b). We present the procedure in Algorithm 1.

In particular, Step 8 in Algorithm 1 requires to calculate the gradient of the smoothed elementwise ℓ_∞-norm. An explicit calculation gives that

$$\nabla\|\widehat{\boldsymbol{\Sigma}} - \mathbf{U}\|_\infty^\rho = -\widetilde{\mathbf{T}}, \tag{2.9}$$

where

$$\widetilde{\mathbf{T}}_{jk} = \mathrm{sign}(\widehat{\boldsymbol{\Sigma}}_{jk} - \mathbf{U}_{jk})\cdot \max\Big\{\Big|\frac{\widehat{\boldsymbol{\Sigma}}_{jk} - \mathbf{U}_{jk}}{\rho}\Big| - \theta, 0\Big\}$$

and θ is the minimum non-negative constant such that $\|\widetilde{\mathbf{U}}\|_1 \le 1$. $\mathbf{T}$ is essentially a "soft thresholding" function, which means that it thresholds the value of $\widehat{\boldsymbol{\Sigma}}_{jk} - \mathbf{U}_{jk}$ as 0 if its absolute value is less that $\rho\cdot\theta$. A pictorial explanation of the soft thresholding function is provided in Figure 2.2.

Algorithm 1 Fast proximal gradient descent algorithm for smooth projection

1: $\widetilde{\boldsymbol{\Sigma}} \leftarrow \text{Proximal-Gradient}\big(\widehat{\boldsymbol{\Sigma}}, \rho\big)$
2: **input:** A symmetric matrix $\widehat{\boldsymbol{\Sigma}}$, a smoothing parameter ρ
3: **output:** A positive definite matrix $\widetilde{\boldsymbol{\Sigma}}$, as the solution of (2.7).
4: **initialize:** $t \leftarrow 1$, $\mathbf{U}^{(0)} = \mathbf{V}^{(0)} = \boldsymbol{\Sigma}^{(0)}$
5: **repeat**
6: $\quad \gamma_t \leftarrow 2/(1+t)$
7: $\quad \mathbf{U}^{(t)} \leftarrow (1-\gamma_t)\boldsymbol{\Sigma}^{(t-1)} + \gamma_t \mathbf{V}^{(t-1)}$
8: $\quad \mathbf{G}^{(t)} \leftarrow \nabla \|\widehat{\boldsymbol{\Sigma}} - \mathbf{U}\|_\infty^\rho \,|_{\mathbf{U}=\mathbf{U}^{(t)}}$, where the gradient can be calculated by (2.9)
9: $\quad \mathbf{V}^{(t)} \leftarrow \Pi_+\Big(\mathbf{V}^{(t-1)} - \frac{\rho}{\gamma_t}\mathbf{G}^{(t)}\Big)$, where $\Pi_+(\cdot)$ can be calculated by (2.10)
10: $\quad \boldsymbol{\Sigma}^{(t)} \leftarrow (1-\gamma_t)\boldsymbol{\Sigma}^{(t-1)} + \gamma_t \mathbf{V}^{(t)}$
11: **until** Converge
12: **return** $\widetilde{\boldsymbol{\Sigma}} \leftarrow \boldsymbol{\Sigma}^{(t)}$

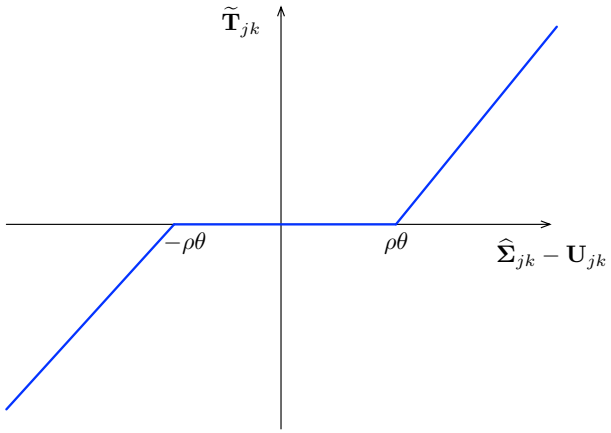

Figure 2.2: The soft thresholding function $\widetilde{\mathbf{T}}_{jk}$

Moreover, Step 9 in Algorithm 1 needs to solve the following positive definite projection problem under the Frobenius norm:

$$\Pi_+(\mathbf{A}) = \underset{\mathbf{B} \in \boldsymbol{S}\mathbb{R}_+^d}{\operatorname{argmin}} \|\mathbf{B} - \mathbf{A}\|_{\mathsf{F}}^2,$$

for any symmetric matrix $\mathbf{A} \in \mathbb{R}^{d\times d}$. Lemma 3.2 in Zhao et al. (2012b) showed that $\Pi_+(\mathbf{A})$ has a closed-form solution: As $\mathbf{A}$ is symmetric, it has the eigenvalue decomposition as $\mathbf{A} = \sum_{j=1}^d \sigma_j \boldsymbol{v}_j \boldsymbol{v}_j^T$, where σ_j's are the eigenvalues and $\boldsymbol{v}_j$ are the eigenvectors. The solution to the projection problem under Frobenius norm is

$$\Pi_+(\mathbf{A}) = \sum_{j=1}^d \max(\sigma_j, 0)\boldsymbol{v}_j \boldsymbol{v}_j^T. \tag{2.10}$$

This shows that the projection to the space of positive definite matrix under Frobenius norm is simply a thresholding of the eigenvalues.

The following theorem guarantees fast convergence of Algorithm 1:

Theorem 2.2. *(Zhao et al., 2012b) To achieve the accuracy ϵ such that $\|\widehat{\mathbf{\Sigma}}-\mathbf{\Sigma}^{(t)}\|_\infty^\rho - \|\widehat{\mathbf{\Sigma}}-\widetilde{\mathbf{\Sigma}}\|_\infty^\rho \leq \epsilon$, we need at most the following number of iterations:*

$$t = \sqrt{2\|\mathbf{\Sigma}^{(0)} - \widetilde{\mathbf{\Sigma}}\|_{\mathsf{F}}^2/(\rho\epsilon)} - 1 = O(\sqrt{1/(\rho\epsilon)}).$$

This convergence rate is actually optimal in the sense that it is unimprovable over the class of all gradient-based computational algorithms (Nesterov and Nesterov, 2004).

The minimization problem (2.8) can be efficiently solved using existing solvers. For example, it has the special structure of a ℓ_1-penalized regression (LASSO) and thus can be solved using many optimization algorithms, such as the coordinate descent (Friedman et al., 2007). Alternatively, we can use general convex programming algorithms such as the interior point method Vanderbei (1999).

Once $\widehat{\mathbf{R}}$ is obtained, we calculate the graph estimator $\widehat{G} = (V, \widehat{E})$, in which a pair (j, k) satisfies that $\widehat{E}_{jk} \neq 0$ if and only if $\widehat{\mathbf{R}}_{jk} \neq 0$. Algorithm 2 summarizes the approach. For each λ, this algorithm outputs the estimated graph that underlies the stock returns. Choosing a large λ will impose large penalization of the ℓ_1-norm of the $\widehat{\mathbf{\Theta}}$ and result in a sparse graph and vice versa.

Algorithm 2 Graph Estimation

1: $\widehat{G}^\lambda \leftarrow$ Graph-Estimation$(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n; \lambda)$
2: **input:** Historical return of d stocks over n periods $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \in \mathbb{R}^d$, tuning parameter λ
3: **output:** Estimated graph $\widehat{G}^\lambda$ corresponding to the tuning parameter λ
4: Calculate Kendall's tau correlation matrix by (2.4) and (2.5).
5: Let $\widetilde{\mathbf{\Sigma}} \leftarrow$ Proximal-Gradient$(\widehat{\mathbf{\Sigma}})$ as in Algorithm 1
6: Obtain $\widehat{\mathbf{R}}$ by solving

$$\widehat{\mathbf{R}} = \underset{\mathbf{R}\in\mathbb{R}^{d\times d}, \mathrm{diag}(\mathbf{R})=\mathbf{0}}{\mathrm{argmin}} \sum_{j=1}^{d} \left\{ \mathbf{R}_{\backslash j,j}^T \widetilde{\mathbf{\Sigma}}_{\backslash j,\backslash j} \mathbf{R}_{\backslash j,j} - 2\widetilde{\mathbf{\Sigma}}_{j,\backslash j} \mathbf{R}_{\backslash j,j} + \lambda \|\mathbf{R}_{\backslash j,j}\|_1 \right\}$$

7: Set $\widehat{G}^\lambda = (V, \widehat{E}^\lambda)$ where $\widehat{E}_{jk}^\lambda = \mathbb{I}\{\widehat{\mathbf{R}}_{jk} \neq 0\}$
8: **return** $\widehat{G}^\lambda$

2.6 Stability Selection

Using Algorithm 2, we estimate the graph by selecting edges that indicate strong conditional dependence between two nodes. In this section, we employ a stability-based method to enhance the estimation performance. Instead of based on only one data realization, we aim to select those edges that *always* indicate strong conditional dependence under multiple realizations.

The stability selection method was first proposed by Meinshausen and Bühlmann (2010), and has become very successful in variable selection and structure estimation (Meinshausen and Bühlmann, 2010; Liu et al., 2010; Shah and Samworth, 2013). Using this method, the estimated graphs are more stable, in the sense that another set of realized data under the same data generating process would produce a graph that is very similar to the current one. Empirical results show that the stability selection method improves the selection accuracy.

Specifically, we first sub-sample without replacement from the original data, and then apply Algorithm 2 for edge selection to each sub-sample. Next, we calculate the frequency that each edge is selected by the subsamples, and choose those with the highest frequencies to be selected. Suppose the size of each subsample is b. There are in total $\binom{n}{b}$ subsamples. Ideally, we should use all subsamples to get the best result. However, Politis et al. (1999) showed that it suffices to use a large number L of subsamples at random. For each subsample ℓ, we estimate the sparse precision matrix. We then count the frequency that each edge is added to the graph using all L subsamples. Define

$$\widehat{p}_{jk}^{(\lambda)} = \frac{\sum_{\ell=1}^{L} \mathbb{I}\{\widehat{\mathbf{R}}_{jk}^{(\ell)} \neq 0\}}{L}.$$

where $\widehat{\mathbf{R}}^{(\ell)}$ is obtained by the ℓ-th subsample using Algorithm (2). Taking different tuning parameter λ, we get a sequence of frequencies for each pair (j, k). We can plot these frequencies against λ; the resulted graph is called the stability path (Figure 2.3). Stability selection chooses those edges that have high probability to be selected:

$$\widehat{E}_{jk}^{(\lambda)} = \mathbb{I}\{\widehat{p}_{jk}^{(\lambda)} > \alpha\}.$$

In practice, the estimated graph is insensitive to the choice of α (Liu et al., 2010; Meinshausen and Bühlmann, 2010). The convention of $\alpha = 0.05$ is often chosen. For our purpose of independent variable selection, it suffices to choose $\alpha = 0$. The detailed procedure is presented in Algorithm 3, which is a simple version of the stability selection. More complex methods can be considered. For example, instead of subsampling across time horizon, we can take the variable dimension, i.e., for each subsample, we take $(\boldsymbol{X}_{1J}, \ldots, \boldsymbol{X}_{nJ})$ where $J \subset \{1, 2, \ldots, d\}$. Furthermore, we can employ a hybrid method, by taking subsamples from both the time and variable dimensions.

2.7 Independent Variable Selection

We are now ready to present the complete algorithm for stock selection. Using stability selection, we obtain a stability path corresponding to a sequence of tuning parameters λ. We start from the smallest λ, so that the estimated graph are the most dense. We next increase λ, so that the estimated graph gets sparser as λ increases. We scan this stability path until we obtain a graph in which there are 10 isolated nodes. Figure 2.3 shows the stability path and selection procedure of the example of health care stocks discussed in §4.1; Figure 2.4 shows the evolvement of the estimated graph when scanning the stability path with increasing λ.

Algorithm 4 lists the primary steps: 1) calculating Kendall's tau statistic, 2) smoothly projected neighborhood pursuit, 3) stability selection and 4) independent node selection. The input requires a sequence of λ that is dense enough. To improve efficiency, we employ the binary search method, as the number of isolated points is monotonically nondecreasing as λ increases.

3 Theoretical Guarantees

In this section, we provide theoretical guarantees for the stability selection procedure. Our main theorem establishes the statistical rate of convergence of the estimated index set $\widehat{J}$ to its population quantity $J := \mathbb{E}[\widehat{J}]$, where the expectation is taken over the randomness in subsampling. We define

Algorithm 3 Stability Selection

$\widehat{G}^\lambda \leftarrow$ Stability-Selection$(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n; \lambda)$

input: Historical return of d stocks over n periods $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \in \mathbb{R}^d$, tuning parameter λ, number of subsamples L, subsample size b, thresholding parameter α

output: Estimated graph $\widehat{G}^\lambda$ corresponding to the tuning parameter λ

for $\ell \in 1, \ldots, L$ **do**

Randomly draw b samples from the original data, denoted by $\boldsymbol{X}_{i_1}, \ldots, \boldsymbol{X}_{i_b}$, where $1 \le i_1 < i_2 < \ldots < i_b \le n$

Calculate kendall's tau correlation matrix $\widehat{\boldsymbol{\Sigma}} \in \mathbb{R}^{d \times d}$, where

$$\widehat{\boldsymbol{\Sigma}}_{jk} = \sin\Big(\frac{\pi}{2}\widehat{\tau}_{jk}(\boldsymbol{X}_{i_1}, \ldots, \boldsymbol{X}_{i_b})\Big)$$

and

$$\widehat{\tau}_{jk}(\boldsymbol{X}_{i_1}, \ldots, \boldsymbol{X}_{i_b}) := \frac{2}{b(b-1)} \sum_{1 \le \ell < \ell' \le b} \operatorname{sign}\left[(\boldsymbol{X}_{i_\ell j} - \boldsymbol{X}_{i_{\ell'} j}) \cdot (\boldsymbol{X}_{i_\ell k} - \boldsymbol{X}_{i_{\ell'} k})\right]$$

Let $\widetilde{\boldsymbol{\Sigma}} \leftarrow$ Proximal-Gradient$(\widehat{\boldsymbol{\Sigma}})$ as in Algorithm 1

Obtain $\widehat{\mathbf{R}}$ by solving

$$\widehat{\mathbf{R}} = \underset{\mathbf{R} \in \mathbb{R}^{d \times d}, \operatorname{diag}(\mathbf{R}) = \mathbf{0}}{\operatorname{argmin}} \sum_{j=1}^{d} \Big\{ \mathbf{R}_{\setminus j,j}^T \widetilde{\boldsymbol{\Sigma}}_{\setminus j, \setminus j} \mathbf{R}_{\setminus j,j} - 2\widetilde{\boldsymbol{\Sigma}}_{j, \setminus j} \mathbf{R}_{\setminus j,j} + \lambda \|\mathbf{R}_{\setminus j,j}\|_1 \Big\}$$

Set $s_{jk}^{(\lambda)} = s_{jk}^{(\lambda)} + \mathbb{I}\{\widehat{\mathbf{R}}_{jk} \neq 0\}$.

Let $\widehat{p}_{jk}^{(\lambda)} \leftarrow s_{jk}^{(\lambda)} / L$

Obtain $\widehat{G}^\lambda = (V, \widehat{E}^\lambda)$ where $\widehat{E}_{jk}^\lambda = \mathbb{I}\{\widehat{p}_{jk}^\lambda > 0\}$

return $\widehat{G}^\lambda$

$p_{jk}^{(\lambda)} := \mathbb{E}[\widehat{p}_{jk}^{(\lambda)}]$. Recall that n is the sample size, d is the total number of stocks, m is the number of regularization parameters λ used in obtaining the graph path, and b is the subsample size in the stability selection.

The following theorem shows that the estimated index set is consistent (converges to its population version):

Theorem 3.1. *Suppose that m is a polynomial of n, i.e. $m \le n^c$ for some constant c, and $|p_{jk}^{(\lambda)} - p_{j'k'}^{(\lambda)}| \ge 2\sqrt{\frac{c' b \log(n \vee d)}{n}}$ for any pairs (j,k) and (j',k') and λ, where c' is a universal constant. Then we have*

$$\mathbb{P}(\widehat{J} = J) \to 1$$

as $n \to 1$.

The second condition in Theorem 3.1 is satisfied if $b \log(n \vee d)/n \to 0$ and for large enough n. Therefore, our algorithm scales to high dimensional models, where d could potentially grow exponentially with n.

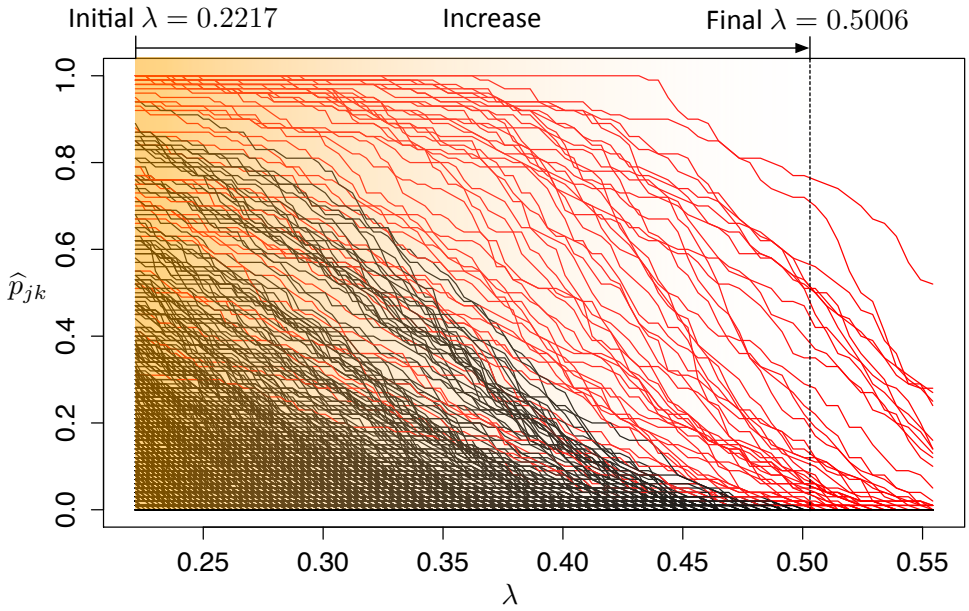

Figure 2.3: Stability path and the selection procedure of the healthcare stock example. In this figure, a curve is the stability path of an edge. There are in total $d(d-1)/2$ paths. As we scan the stability paths with increasing λ, fewer edges are selected, and more isolated nodes emerge. The final λ is chosen at a point where the corresponding graph has exactly 10 isolated nodes (when $\lambda = 0.5006$ in this example). The red curves are paths of edges selected in the final graph.

4 Application to Equity Selection

As mentioned, to harvest rebalancing gains, our investment goal is to select stocks to be as independent of each other as possible. We apply the above graphical modeling Algorithm 4 to select a portfolio of stocks.

Herein, we consider the problem of portfolio management. Standard portfolio theory advocates using a variety of stocks to diversify the risk. Instead of using all available stocks, we choose a few stocks that are as independent as possible. We then construct our portfolio only using the chosen stocks. This dynamic portfolio strategy is expected to perform better than the traditional Markowitz model implementing the buy-and-hold strategy in terms of return, Sharpe ratio, and other performance measures.

4.1 Healthcare Data

First, to illustrate the approach, we select data from the healthcare industry (50 stocks). The price and return data start from January 1992 and end at April 2014. As many of the stocks did not exist until recently, we truncate the data before 1998 and only look at stocks that existed on January 2nd, 1998. That gives us 43 stocks to depict the power of our method in terms of harvesting rebalancing gains. To address missing data, we use the k-nearest neighbor method to impute the data. This can be implemented by the function "impute.knn" in the package "impute" using the statistical analysis software R.

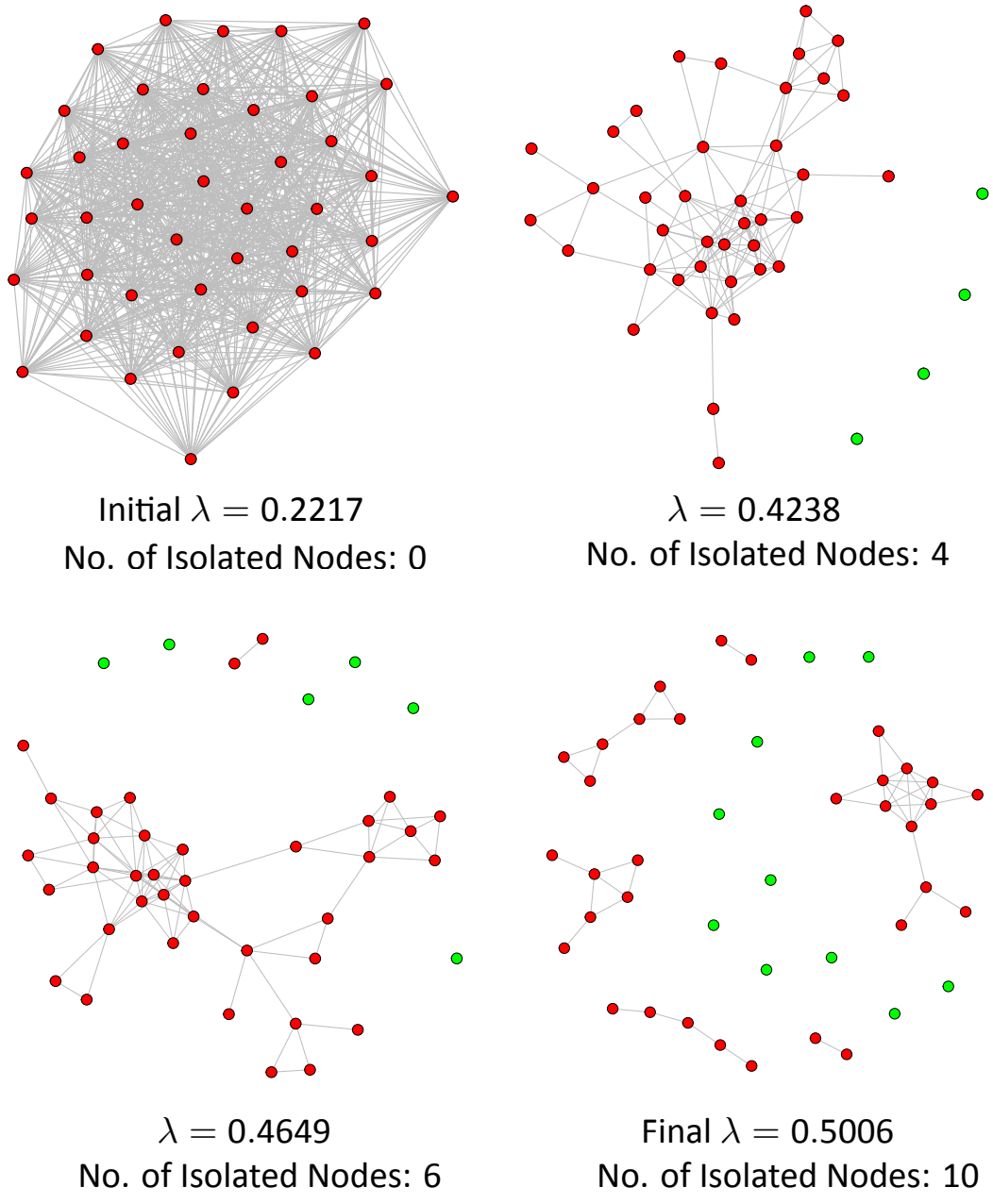

Figure 2.4: Evolvement of the graph generated by stability selection with increasing λ. The initial graph is very dense. As we increase λ, the graph becomes sparser and more isolated nodes emerge. The final graph is selected when there are exactly 10 isolated nodes, when $\lambda = 0.5006$ in this example.

Algorithm 4 Stock Selection

$\widehat{J} \leftarrow$ Stock-Selection$(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n; \lambda_1, \ldots, \lambda_m; k)$
input: Historical return of d stocks over n periods $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \in \mathbb{R}^d$; a sequence of tuning parameters $\lambda_1 < \lambda_2 < \ldots < \lambda_m$; desired number of stocks to be selected k
output: An index set $\widehat{J} \subset \{1, \ldots, d\}$ with $|\widehat{J}| = k$
for $r \in 1, \ldots, m$ **do**
  $\widehat{G}^{\lambda_r} \leftarrow$ Stability-Selection$(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_2; \lambda_r)$ as in Algorithm 3
  $N_r \leftarrow$ number of nodes that are not connected to other nodes in $\widehat{G}^{\lambda_r}$
  if $N_r \geq k$ **then** break
$\widehat{J} \leftarrow$ the indices of the nodes that are not connoted to other nodes in $\widehat{G}^{\lambda_r}$
return $\widehat{J}$

We compare three portfolio strategies. The first one is a rebalancing strategy, in which the stocks are selected using the graphical model and stability selection method described in Section 2. The second one is also a rebalancing strategy, but the stocks are selected using a factor model approach, which combines the principal component analysis and clustering technique. The comparison between these two approaches demonstrates the effectiveness of independent stock selection. The third one is the buy-and-hold Markowitz portfolio strategy. Comparison between the rebalancing strategies and the Markowitz approach demonstrates the benefits of rebalancing.

Graphical model strategy. The practical implementation of graphical model strategy is detailed as follows. At the beginning of the investment horizon, we select k stocks to put in the rebalancing portfolio using the graphical model and stability selection approach. The stocks are selected using previous years' returns as training data, and are expected to be as in dependent to each other as possible in terms of the graphical metric introduced in Section 2. We invest only in these stocks for the whole year, and the investment amount in each stock is rebalanced at some frequency (usually monthly, weekly or daily) to remain at the same weights as their initial values. For simplicity, we use equal weight for all stocks. The set of selected stocks is updated after one year, again based on a new training set of historical data, until the investment period ends.

Note that equal weight is a first benchmark for achieving rebalancing gains, with better performance than the traditional buy-and-hold under many time periods (Mulvey and Kim, 2009). Improved performance can be obtained by rendering robust estimates of the expected values of each stock, but this extension requires considerably more information.

The graphical model stock selection procedure is described in the following steps. First, obtain the historical returns of all candidate stocks. Then compute the graph path of all the stocks for a sequence of regularization parameters $\lambda_1 < \ldots < \lambda_m$ via the graphical model and stability selection method as described in Algorithms 2 and 3. From the graph path, select one graph whose number of isolated nodes is equal to k. Specifically, one can start from λ_1, when the graph is densest (e.g., see first graph in Figure 2.4). Usually, all the nodes are connected, so we proceed to count the number of isolated nodes of the next graphs, until this number is greater than or equal to k. The isolated nodes on the final graph are the stocks we select to put in the rebalancing portfolio.

In our real data experiment of the heathcare stocks, we consider two scenarios: one is based on monthly return data and the rebalancing frequency is monthly, the other is based on daily data the rebalancing frequency is daily. We use previous four year's stock returns as the training data, and

select 10 among the total 43 stocks to put in the rebalancing portfolio (i.e., $k = 10$). The historical returns are stored in a 48×43 data matrix for monthly returns, or 1008×43 for daily returns. We then compute the graph path using the data matrix and Algorithms 2 and 3. In practice, the procedures can be conveniently implemented by the functions in the High-dimensional Undirected Graph Estimation ("huge") package (Zhao et al., 2012a) using the statistical analysis software R. We briefly describe the usage of this package and the parameter settings in our example.

First, input the data matrix to the "huge" function, and select "mb" for the "method" option. This function computes a list of Markovian graphs corresponding to a sequence of regularization parameters, and the option "mb" stipulates that the smoothly projected neighbourhood selection method (Algorithm 2) is adopted for graph estimation. The sequence of regularization parameters $\lambda_1 < \ldots < \lambda_m$ is automatically generated by the function, where the two ends λ_1, λ_m corresponds to the densest and sparsest (empty) graphs of the 48 stock returns. The output of the function "huge" is an R object of the class "huge", which is then input into the function "huge.select" for implementing the stability selection. In this function, subsampling is performed to select edges and obtain a graph path as described in Algorithm 3, for the sequence of regularization parameters chosen before. An example of the graph path is shown in Figure 2.3. There are several arguments that controls the subsampling. The "stars.subsample.ratio" is the ratio between the sizes of the subsample (denoted by b in Algorithm 3) and the entire sample (n). The default is $10\sqrt{n}/n$ if $n > 144$ and 0.8 if $n < 144$. Another argument is "rep.num", which is the number of subsamples (denoted by L in Algorithm 3), which we take to be 50. Lastly, scan the graph path to obtain the graph with 10 isolated nodes. Sometimes this number might not be able to achieved if the sequence of regularization parameters is not refined enough, and we choose the closest number. In the function "huge", we can stipulate the refinement of this sequence. Specifically, "nlambda" controls the number of regularization parameters m in the sequence. In practice, $m = 30$ often results in $9 - 11$ stocks, and with $m = 50$, the number of selected stocks can be exactly 10.

PCA clustering strategy. The second strategy is also based on rebalancing, but the set of k stocks is selected using a factor model. Similar to the previous strategy, we use historical returns as our training data for stock selection. We project the stocks into the space spanned by the leading three principle components (factors), and cluster their low dimension projections into k groups. The intuition is that under the factor model, stock returns are mostly determined by the leading three factors. Hence those stocks belonging to different clusters should have low correlations as their factor loadings are rather different. We select the stocks closest to each of the k cluster centers and use the k stocks to construct our portfolio.

Specifically, we take the data matrix containing the stock returns, and perform the eigen-decomposition of its empirical covariance matrix. In particular, the eigenvectors are orthonormal. We take the three leading eigenvectors (that is, eigenvectors corresponding to the three largest eigenvalues). Each dimension represents one stock, so the j-th coordinates of the three vectors are the loadings (weights) of the three principal components corresponding to the j-th stock. We cluster the stocks into k groups by their factor loadings using the k-means algorithm, where the L_2-norm of the factor loadings is used as the distance metric for the stocks. The k-means algorithm is described as follows. Given an initial clustering of k groups, compute the mean of each group as the new cluster center. Then calculate the distance of each member to the k centers, and cluster the members into k new groups according to their closest centers. Next calculate the mean of the each

group to update the cluster centers. The process is repeated until convergence. Note that in our example, the members are the three dimensional factor loadings of each stock. In the R software, we use the build-in function "kmeans" to implement the algorithm. The inputs are the 48×3 factor loading matrix, and the parameter k to specify the number of clusters. We take $k = 10$. The output is an R object containing the information of the cluster centers. Finally, we select the ones closest to each cluster center as the 10 stocks to be invested in the rebalancing portfolio.

Markowitz strategy. The last strategy we consider is Markowitz portfolio selection. This traditional selection method is characterized by its "buy-and-hold" nature, which ignores the rebalancing gains. Stocks are selected once every year and are held for the entire year. The description of Markowitz strategy can be found in multiple literature, e.g., Section 6.6 of Luenberger (2013).

The comparison results appear in Tables 4.1 and 4.2. The results show that the selection method based on graphical model is better in return, Sharpe ratio and max drawdown. The "buy-and-hold" Markowitz strategy produces the least favorable performance, with low returns, high standard deviations and large maximum drawdown.

Table 4.1: Comparisons of stock strategies based on monthly data with monthly rebalancing.

	Ann. Return (%)	Ann. SD (%)	Sharpe	Max Drawdown (%)
Graphical Model	15.32	16.51	0.87	31.16
PCA Clustering	14.08	20.28	0.64	36.47
Markowitz	4.23	32.03	0.10	66.89

Table 4.2: Comparisons of stock strategies based on daily data with daily rebalancing.

	Ann. Return (%)	Ann. SD (%)	Sharpe	Max Drawdown (%)
Graphical Model	20.34	21.09	0.92	41.70
PCA Clustering	15.80	23.48	0.63	48.00
Markowitz	10.80	23.33	0.42	60.95

Transaction Cost. We discuss the impact of transaction cost on the portfolio returns. For the rebalancing strategy, there are two sources of transaction costs: one is from daily (monthly) rebalancing, and the other is from adjusting the set of stocks selected at the beginning of each year. For the Markowitz strategy, transaction cost only occurs at the beginning of each year for stock selection. In Tables 4.3 and 4.4 we calculate the annualized turnovers and the impact on the portfolio returns under different rates of transaction costs. In particular, the first column is the annualized turnovers of the three strategies, and the last three columns are the adjusted annualized returns after accounting for transaction costs, at the rates of 0.1%, 0.5% and 1%.

An investor can thereby see the impact of these costs on performance depending upon their circumstances. In the experiment result, we can see that the rebalancing strategy indeed requires significantly more trading amounts than the buy-and-hold Markowitz strategy. However, the impact of transaction cost on the portfolio return is relatively small for 1% transaction cost.

Table 4.3: Impact of transaction costs on portfolio returns based on monthly data.

	Ann. Turnover	0.1% Trans. Cost	0.5% Trans. Cost	1% Trans. Cost
Graphical Model	37.31%	15.27%	15.10%	14.88%
PCA Clustering	31.30%	14.04%	13.90%	13.72%
Markowitz	22.70%	4.20%	4.11%	3.99%

Table 4.4: Impact of transaction costs on portfolio returns based on daily data.

	Ann. Turnover	0.1% Trans. Cost	0.5% Trans. Cost	1% Trans. Cost
Graphical Model	112.6%	26.62%	26.31%	25.92%
PCA Clustering	101.1%	15.68%	15.20%	14.56%
Markowitz	3.08%	10.80%	10.78%	10.77%

4.2 Large Stock Data

In our second real data experiment, we apply our graphical model strategy to a large data set including 1500 stocks over a 34-year investment horizon from 1981 - 2014. In the first step, we implement a momentum-based investment procedure (Asness et al., 2013) to preselect 50 stocks. At selected time junctures, we rank stocks in the S&P 1500 index according to their performance over a predetermined back-test period (typically 3 to 12 months). Under momentum, stocks with the best historical performance have a higher probability of better performance over the subsequent holding period. For our purposes, we take the top 50 stocks out of 1500 as a target list based on a 6-month look-back. In the second step, we compare four strategies for choosing a portfolio among the 50 stocks: 1) rebalancing with graphical model selection; 2) rebalancing with PCA clustering selection; 3) traditional Markowitz model; and 4) rebalancing with all 50 stocks. Under alternatives 1, 2, and 4, we apply an equal weighted mix - either 10 stocks for 1 and 2, or 50 stocks for 4.

Table 4.5: Comparison of strategies using momentum-based pre-selected stocks.

	Ann. Return (%)	Ann. SD (%)	Sharpe	Max Drawdown (%)
Graphical Model	31.00	22.70	1.32	67.44
PCA Clustering	30.31	28.29	1.04	74.20
Markowitz	13.15	61.76	0.20	97.01
All 50 Stocks	27.66	22.61	1.18	74.86
S&P500	7.65	18.35	0.36	57.60

Herein, the ranking and independent stock selection are conducted annually, followed by daily rebalancing until the next ranking and selection date. We use two-year historical returns as the training data set to compute the independence set. For stability selection, we take $L = 50$ and $m = 30$. Our method exhibits significant positive results in the following aspects:

1. Our method substantially outperforms the classical Markowitz strategy. This shows that our approach of applying graphical model combined with rebalancing strategy is able to reap

Table 4.6: Impact of transaction costs on portfolio returns for large stock data example.

	Ann. Turnover	0.1% Trans. Cost	0.5% Trans. Cost	1% Trans. Cost
Graphical Model	70.57%	30.91%	30.51%	29.93%
PCA Clustering	50.59%	30.24%	29.96%	29.58%
Markowitz	1.95%	13.15%	13.14%	13.13%
All 50 Stocks	167.0%	27.44%	26.39%	24.39%

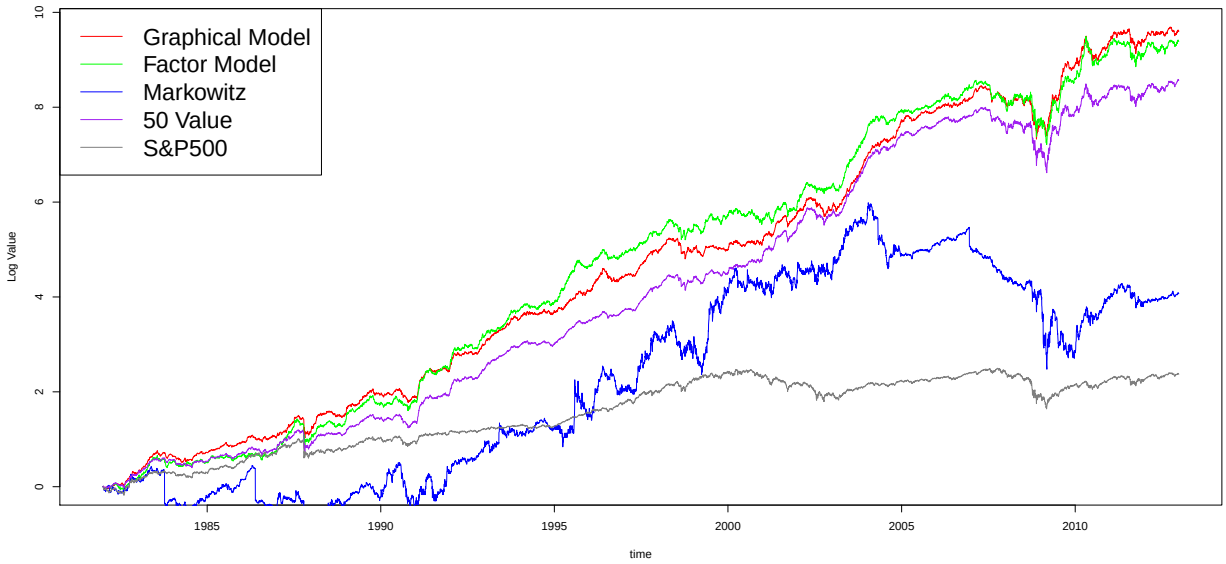

Figure 4.1: Comparison of strategies with stocks based on value with yearly selection and daily rebalancing.

rebalancing gains compared with the traditional "buy-and-hold" strategy. This empirically conforms with the theories of rebalancing gains discussed in Luenberger (2013) and Mulvey et al. (2001), etc.

2. Compared with rebalancing strategy using all 50 stocks, the rebalancing portfolio with 10 stocks selected by the graphical model has higher returns, better Sharpe ratios and smaller max-drawdowns. This not only confirms the advantage of using independent stocks in such rebalancing strategy, but also shows that the stocks selected by our graphical method indeed exhibit less dependence than the entire pool of stocks.

3. Compared with the factor based clustering method, the approach based on graphical model has better performance for the momentum-based investment procedure. (Figure 4.1, Table 4.5. Note that the y-axis of Figure 4.1 is log-scale). This suggests that our latent hierarchical framework is able to effectively capture the inter-dependency structure among stocks, while the factor model only partially captures such relationships.

Transaction cost. Similar to the previous example, we calculate the annualized turnovers and impact of transaction costs, which are summarized in Table 4.6.

Computation time. Our stability based stock selection algorithm potentially requires many iterations, due to subsampling in the stability path. However, as computation of each iteration is very fast, the overall procedure performs reasonable in time. In the healthcare example, for the monthly return data, we use a four-year sliding window as training set to select independent stocks. So $n = 48$ and $d = 43$. As mentioned above, the number of subsamples L is 50 and the number of tuning parameters m is 30. The computation time is 15.308 seconds.

With daily data, we have $n = 1008$ and $d = 42$. With the same L and m, the computation time is 15.680 seconds. Substantially, the computation time is relatively invariant to the sample size (n). This is because the data size is only involved in the computation of the Kendall's tau statistic, which costs little time. The rest of the algorithm is not related to the sample size.

Lastly, in the large-scale stock example, we take a two-year sliding window, so $n = 504$ and $d = 50$. Same as the above example, we take $L = 50$ and $m = 30$. The computation time is 16.616 seconds.

All the results are run on a Macintosh OS X with 2.7 GHZ Intel Core i5 CPU and 16GB 1600 MHZ DDR3 Memory.

5 Independence Measure

In previous sections, we presented our stock selection method based on graphical model approach to construct a portfolio that exploits rebalancing gains. The crucial idea is to select stocks that exhibit the highest level of independence. Empirical results show that our method for independent stock selection is able to harvest such rebalancing gains. In this section, we propose a new measure that characterizes the level of independence between stocks. To this end, we apply a novel distribution-free independence test. The level of independence is characterized by the p-value of this test. The stocks selected by graphical model method indeed exhibit higher level of independence compared with the methods of PCA selection, random selection or selecting all available stocks, which justifies the validity of our method.

Formally, this test is formulated as follows: let $\boldsymbol{X} = (X_1, \ldots, X_d)$ be a d-dimensional random vector. In our example, $\boldsymbol{X}$ is the return of d selected stocks. We aim to test the null hypothesis:

$$\mathbf{H}_0 : X_1, X_2, \ldots, X_d \text{ are mutually independent.} \tag{5.1}$$

This problem has been well studied in statistics literature under the Gaussian assumption, for example (Anderson, 1958; Roy and Roy, 1957; Nagao, 1973) in low dimensional settings and Bai et al. (2009); Cai et al. (2013); Jiang et al. (2013); Johnstone (2001); Ledoit and Wolf (2002); Schott (2005); Srivastava (2005) in high dimensions. Moreover, Jiang (2004) and Zhou (2007) relaxed the Gaussian assumption to distributions with high order moments.

We aim to test (5.1) based on the rank coefficient statistics. Our method is state-of-art, as it does not require any distributional assumption on $\boldsymbol{X}$, and works well under high dimensional settings where d is allowed to grow at the rate of $\log d = o(n^{1/3})$.

Given realizations $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$, the test statistic is defined as follows:

$$\widehat{T}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) := \max_{j<k} \big|\widehat{\tau}_{jk}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) - \mathbb{E}_0\big[\widehat{\tau}_{jk}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)\big]\big|,$$

where $\widehat{\tau}_{jk}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$ is the Kendall's tau statistic defined by (2.5) and $\mathbb{E}_0$ is the expectation taken under $\mathbf{H}_0$. Hence $\mathbb{E}_0[\widehat{\tau}_{jk}]$ is the population version of Kendall's tau rank correlation under the null hypothesis. It can be shown under $\mathbf{H}_0$ as defined in (5.1), $\mathbb{E}_0\big[\widehat{\tau}_{jk}\big] = 0$. The following theorem shows the asymptotic behavior of the test statistic $\widehat{T}$ under $\mathbf{H}_0$.

Theorem 5.1. *Under the scaling that* $\log d = o(n^{1/3})$ *and the null hypothesis* $\mathbf{H}_0$*, we have for any* $x \in \mathbb{R}$

$$\big|\mathbb{P}\big(\widehat{T}^2/\sigma_0^2 - 4\log d + \log\log d \le x\big) - \exp\big(-\exp(-x/2)/\sqrt{8\pi}\big)\big| = o(1),$$

where σ_0^2 *is the variance of* $\widehat{\tau}$ *under the* $\mathbf{H}_0$*, and when* $\mathbf{H}_0$ *is as in* (5.1)*,* $\sigma_0^2 = \frac{2(2n+5)}{9n(n-1)}$.

Theorem 5.1 shows that the distribution function of $\widehat{T}^2/\sigma_0^2 - 4\log d + \log\log d$ can be approximated by a Gumbel distribution when n is large. The proof can be found in the proof of Theorem 3.2 in Han and Liu (2014). Based on this theorem, we construct the test statistic as follows:

$$\phi(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) = \mathbb{I}\{\widehat{T}^2 > \sigma^2(4\log d - \log\log d - 2\log\sqrt{8\pi} - 2\log\log(1-\alpha)^{-1})\}.$$

The validity of this test ϕ follows immediately from Theorem 5.1, as the type I error is approximately α when n is large:

$$\mathbb{P}(\phi(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) = 1 \,|\, \mathbf{H}_0) = \alpha + o(1).$$

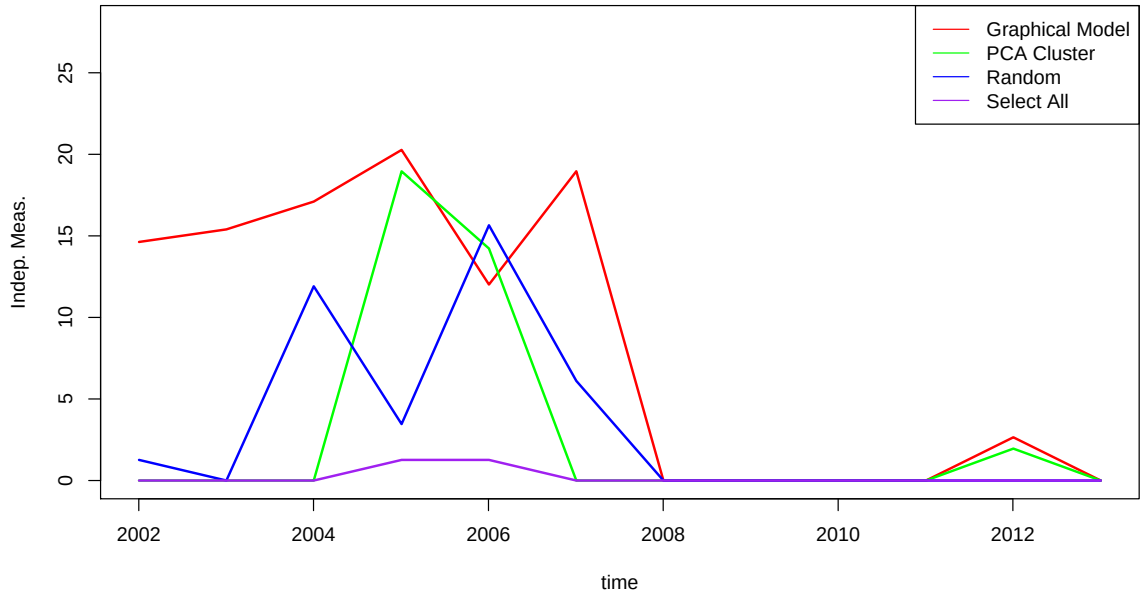

Figure 5.1: Comparisons of independence level of selected stocks based on yearly selection in the health care stock example.

We also provide power analysis that controls the type II error of the test ϕ. The following theorem shows that the power goes to 1 when n goes to infinity. We consider the following family

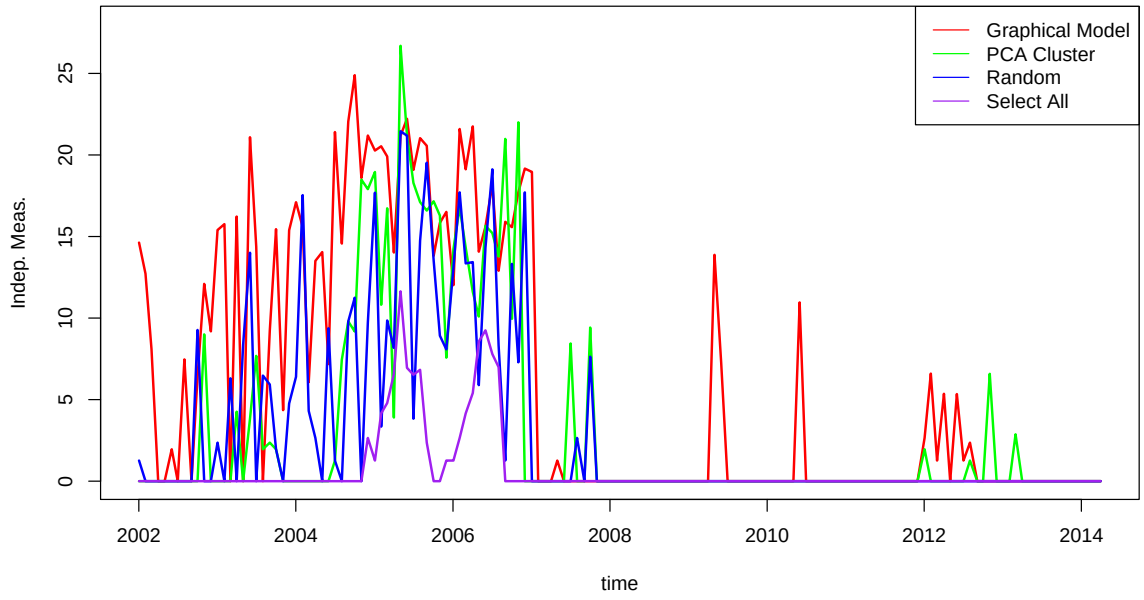

Figure 5.2: Comparisons of independence level of selected stocks based on monthly selection for the health care stock example.

of alternative hypotheses that are defined in terms of the population Kendal' tau matrix: $\mathcal{H}_1 = \{\mathcal{T} \in R(a)\}$, where $\mathcal{T} = \left[\mathbb{E}[\widehat{\tau}_{jk}]\right]_{\{1\le j,k\le d\}} \in \mathbb{R}^{d\times d}$, and

$$R(a) := \left\{\mathbf{M} \in \mathbb{R}^{d\times d} : diag(\mathbf{M}) = \mathbf{I}_d, \mathbf{M} = \mathbf{M}^T, \max_{1\le j<k\le d} |\mathbf{M}_{jk}| \ge a\sqrt{\log d/n}\right\}.$$

Theorem 5.2. *There exists some large enough constant $C > 0$ such that*

$$\inf_{\mathcal{T}\in R(C)} \mathbb{P}(\phi = 1) = 1 - o(1).$$

See Theorem 4.2 of Han and Liu (2014) for the proof.

This test is applied to the stock selection problem. We measure the independence level of stocks selected by different methods for the healthcare portfolio. Specifically, we compare the four methods: 1) graphical model approach; 2) PCA dimension reduction followed by clustering; 3) random selection; 4) selecting all stocks. The p-values demonstrate the independence level of the selected stocks. For clearance of presentation we rescale the p-values by first taking the logarithm and then adding a constant to make them positive. The rescaled p-values are shown in Figure 5.1 for yearly selection and in Figure 5.2 for monthly selection of the health care stock data example in §4.1. From the figures we can see that the p-values of stocks selected by graphical model approach are higher than other methods, indicating that our method indeed is able to select stocks that possess more independence. Applying this measure for large-scale stock example in §4.2, the result is shown in Figure 5.3.

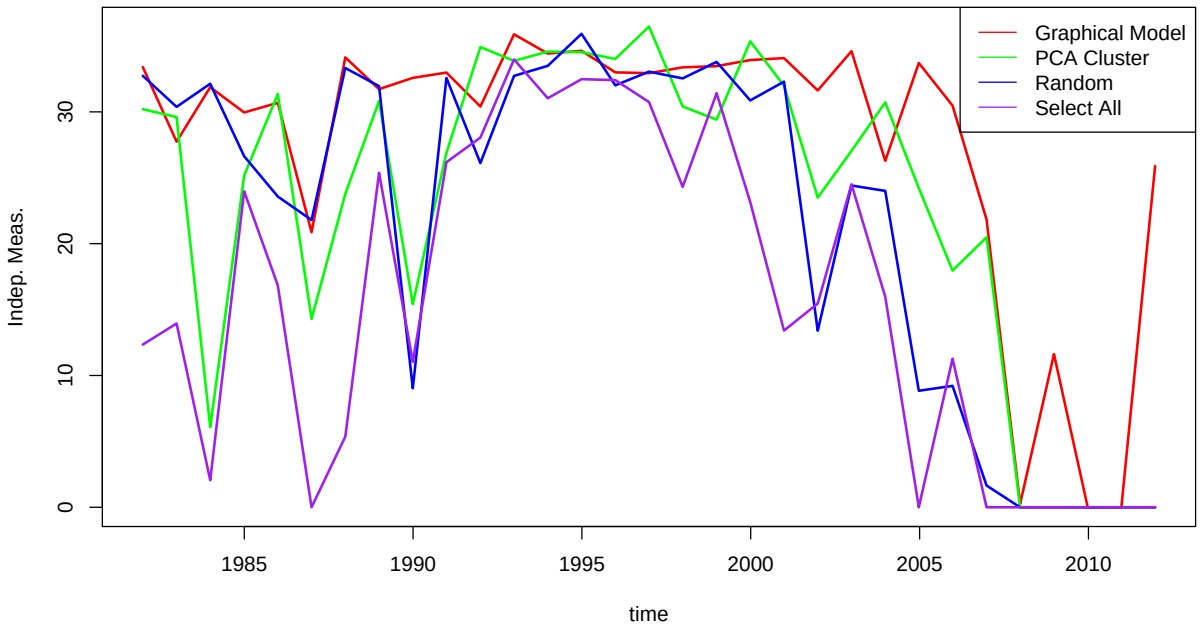

Figure 5.3: Comparisons of independence level of selected stocks in the large-scale stock data example.

Also, notice that in all of the examples, the independence level of stocks decreases drastically during 2008/2009 crash periods under the proposed independence measure, which agrees with theoretical and empirical evidence that assets are more correlated in crash periods than in normal periods. A more interesting fact revealed by all three figures appears to be that there is a permanent change in the correlation structure among financial assets before and after the crash. For example, during the period 2010 - 2014, the correlation level stayed high despite the rebouncing market, as opposed to the period 2002 - 2006 when the market was also growing. This may be explained by the fact that investors have become more cautious and behaved qualitatively different after the financial crisis.

6 Conclusion

Portfolio performance can be enhanced by means of rebalancing gains. Theoretically, when the assets are independent and transaction costs are small, we are able to harvest the rebalancing gains compared with a traditional buy-and-hold strategy. However, implementation of this strategy is difficult as stocks possess substantial co-movements. In this paper we considered the challenging problem of selecting stocks that are "as independent as possible", and applied our method to implement the rebalancing tactic. The method uses the ideas of elliptical-copula graphical models. Under such modeling strategy, the notion of independence is defined by the conditional independence relationships among the latent Gaussian random variables that underly the observed variables. We

propose an algorithm via ℓ_1-penalization and smoothed projection methods for estimating the latent graphical structure. This algorithm not only has fast convergence rate, but also scales to potentially very large number of stocks. We also apply stability selection to improve the estimation performance. The stocks to add to the portfolio are selected by a two-step thresholding method.

Empirically, the portfolio strategy implementing the graphical model selection significantly outperforms the traditional buy-and-hold strategy such as the Markowitz strategy, as well as alternative rebalancing strategies such as PCA dimension reduction combined with clustering. This not only empirically shows the advantages of investing independent stocks in the rebalancing strategy, but also depicts the power of latent Markovian graphical methods in modeling the dependence structure among financial assets.

A Appendix: Proof of Theorem 3.1

We first prove the convergence of $\widehat{p}_{jk}^{(\lambda)}$ for each j, k and λ. Note that $\widehat{p}_{jk}^{(\lambda)}$ is a U-statistic of order b. Then using Hoeffding's inequality, we have for any $t > 0$,

$$\mathbb{P}(|\widehat{p}_{jk}^{(\lambda)} - p_{jk}^{(\lambda)}| > t) \leq 2\exp\big(-2nt^2/b\big).$$

By union bound, we have

$$\begin{aligned}\mathbb{P}\Big(\exists r, j, k \text{ such that } |\widehat{p}_{jk}^{(\lambda_r)} - p_{jk}^{(\lambda_r)}| > t\Big) &\leq md(d-1)\exp\big(-2nt^2/b\big) \\ &\leq \exp(\log m + 2\log d - 2nt^2/b).\end{aligned}$$

For any $\delta \in (0,1)$, we let t satisfy $\delta = \exp(\log m + 2\log d - 2nt^2/b)$ in the above inequality. Hence, we get with probability $1-\delta$,

$$|\widehat{p}_{jk}^{(\lambda_r)} - p_{jk}^{(\lambda)}| \leq \sqrt{\frac{b(\log m + 2\log d - \log\delta)}{2n}}, \tag{A.1}$$

for all $r = 1, \ldots, m$, and j, k such that $1 \leq j < k \leq d$.

Define the event $\mathcal{E} = \Big\{|\widehat{p}_{jk}^{(\lambda_r)} - p_{jk}^{(\lambda_r)} \leq \sqrt{\frac{c' b\log(n\vee d)}{n}} \text{ for any } j, k, r\Big\}$, where c' is some universal constant. Letting $\delta = (n\vee d)^{-1}$ in (A.1) and by the fact that $m = n^c$, we have $\mathbb{P}(\mathcal{E}^c) \leq (n\vee d)^{-1}$. On the event $\mathcal{E}$, and by the condition that $|p_{jk}^{(\lambda)} - p_{j'k'}^{(\lambda)}| \geq 2\sqrt{\frac{c' b\log(n\vee d)}{n}}$, we have that for any λ_r, the order of $\widehat{p}_{jk}^{(\lambda_r)}$ is the same as $p_{jk}^{(\lambda_r)}$ for $1 \leq j < k \leq d$. Thus, on the event $\mathcal{E}$, we have $\widehat{\mathcal{G}}^{(\lambda_r)} = \mathcal{G}_0^{(\lambda_r)}$, where $\mathcal{G}_0^{(\lambda_r)} = \mathbb{E}[\widehat{\mathcal{G}}^{(\lambda_r)}]$, where the expectation is taken with respect to randomness in subsampling. Thus $\mathcal{E}$ implies the event that $\widehat{J} = J$. Therefore,

$$\mathbb{P}(\widehat{J} \neq J) \leq \mathbb{P}(\mathcal{E}^c) \to 0.$$

which completes the proof.

References

Anderson, T. W. (1958). An introduction to multivariate statistical analysis .

Asness, C. S., Moskowitz, T. J. and Pedersen, L. H. (2013). Value and momentum everywhere. *Journal of Finance* **68** 929–985.

Bai, Z., Jiang, D., Yao, J.-F. and Zheng, S. (2009). Corrections to lrt on large-dimensional covariance matrix by rmt. *The Annals of Statistics* 3822–3840.

Cai, T. T., Ma, Z. et al. (2013). Optimal hypothesis testing for high dimensional covariance matrices. *Bernoulli* **19** 2359–2388.

Cherubini, U., Luciano, E. and Vecchiato, W. (2004). *Copula methods in finance*. John Wiley & Sons.

Costa Dias, A. d. (2004). Copula inference for finance and insurance .

Dempster, A. P. (1972). Covariance selection. *Biometrics* **28** 157–175.

Dempster, M. A. H., Evstigneev, I. V. and Schenk-Hoppé, K. R. (2008). Financial markets. the joy of volatility. *Quantitative Finance* **8** 1–3.

Dempster, M. A. H., Evstigneev, I. V. and Schenk-Hoppé, K. R. (2010). Growing wealth with fixed-mix strategies. *The Kelly Capital Growth Investment Criterion: Theory and Practice* **3** 427–455.

Dias, A. and Embrechts, P. (2004). Dynamic copula models for multivariate high-frequency data in finance. *Manuscript, ETH Zurich* .

Embrechts, P., Lindskog, F. and McNeil, A. (2003). Modelling dependence with copulas and applications to risk management. *Handbook of heavy tailed distributions in finance* **8** 329–384.

Friedman, J., Hastie, T., Höfling, H., Tibshirani, R. et al. (2007). Pathwise coordinate optimization. *The Annals of Applied Statistics* **1** 302–332.

Gorton, G. and Rouwenhorst, K. G. (2004). Facts and fantasies about commodity futures. Tech. rep., National Bureau of Economic Research.

Gorton, G. B., Hayashi, F. and Rouwenhorst, K. G. (2012). The fundamentals of commodity futures returns. *Review of Finance* rfs019.

Han, F. and Liu, H. (2012). Transelliptical component analysis. In *Advances in Neural Information Processing Systems*.

Han, F. and Liu, H. (2014). Distribution-free tests of independence with applications to testing more complex structures. *Technical report* .

Higgs, H. and Worthington, A. (2008). Stochastic price modeling of high volatility, mean-reverting, spike-prone commodities: The australian wholesale spot electricity market. *Energy Economics* **30** 3172–3185.

Jiang, T. (2004). The asymptotic distributions of the largest entries of sample correlation matrices. *Annals of Applied Probability* 865–880.

Jiang, T., Yang, F. et al. (2013). Central limit theorems for classical likelihood ratio tests for high-dimensional normal distributions. *The Annals of Statistics* **41** 2029–2074.

Johnstone, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Annals of statistics* 295–327.

Ledoit, O. and Wolf, M. (2002). Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *Annals of Statistics* 1081–1102.

Li, D. X. (2000). On default correlation: A copula function approach. *The Journal of Fixed Income* **9** 43–54.

Lindskog, F., McNeil, A. and Schmock, U. (2003). *Kendall's tau for elliptical distributions*. Springer.

Liu, H., Han, F., Yuan, M., Lafferty, J. and Wasserman, L. (2012a). High dimensional semiparametric gaussian copula graphical models. *Annals of Statistics* **40** 2293–2326.

Liu, H., Han, F. and Zhang, C.-h. (2012b). Transelliptical graphical models. In *Advances in Neural Information Processing Systems*.

Liu, H., Han, F. and Zhang, C.-h. (2013). Transelliptical graphical modeling under a hierarchical latent variable framework. *Technical report* .

Liu, H., Roeder, K. and Wasserman, L. (2010). Stability approach to regularization selection (stars) for high dimensional graphical models. In *Advances in Neural Information Processing Systems*, vol. 23.

Luenberger, D. G. (2013). *Investment science (2nd edition)*. Cambridge University Press.

Meinshausen, N. and Bühlmann, P. (2006). High-dimensional graphs and variable selection with the lasso. *The Annals of Statistics* **34** 1436–1462.

Meinshausen, N. and Bühlmann, P. (2010). Stability selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **72** 417–473.

Mulvey, J. M. (2012). Long-short versus long-only commodity funds. *Quantitative Finance* 1779–1785.

Mulvey, J. M., Kaul, S. S. N. and Simsek, K. D. (2004). Evaluating a trend-following commodity index for multi-period asset allocation. *The Journal of Alternative Investments* **7** 54–69.

Mulvey, J. M. and Kim, W. C. (2009). Constantly rebalanced portfolio - is mean reversion necessary? *Encyclopedian of Quantitative Finance* .

Mulvey, J. M., Lu, N. and Sweemer, J. (2001). Rebalancing strategies for multi-period asset allocation. *The Journal of Wealth Management* **4** 51–58.

Mulvey, J. M., Ural, C. and Zhang, Z. (2007). Improving performance for long-term investors: wide diversification, leverage, and overlay strategies. *Quantitative Finance* **7** 175–187.

Nagao, H. (1973). On some test criteria for covariance matrix. *The Annals of Statistics* 700–709.

Nesterov, Y. and Nesterov, I. E. (2004). *Introductory lectures on convex optimization: A basic course*, vol. 87. Springer.

Politis, D., Romano, J. and Wolf, M. (1999). Subsampling. springer series in statistics.

Roy, S. N. and Roy, S. (1957). *Some aspects of multivariate analysis*. Wiley New York.

Schott, J. R. (2005). Testing for complete independence in high dimensions. *Biometrika* **92** 951–956.

Shah, R. D. and Samworth, R. J. (2013). Variable selection with error control: another look at stability selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **75** 55–80.

Srivastava, M. S. (2005). Some tests concerning the covariance matrix in high dimensional data. *Journal of the Japan Statistical Society* **35** 251–272.

Vanderbei, R. J. (1999). Loqo: An interior point code for quadratic programming. *Optimization methods and software* **11** 451–484.

Xue, L. and Zou, H. (2012). Regularized rank-based estimation of high-dimensional nonparanormal graphical models. *The Annals of Statistics* **40** 2541–2571.

Zhao, T., Liu, H., Roeder, K., Lafferty, J. and Wasserman, L. (2012a). The huge package for high-dimensional undirected graph estimation in r. *The Journal of Machine Learning Research* **13** 1059–1062.

Zhao, T., Roeder, K. and Liu, H. (2012b). Smooth-projected neighborhood pursuit for high-dimensional nonparanormal graph estimation. In *Advances in Neural Information Processing Systems*.

Zhou, W. (2007). Asymptotic distribution of the largest off-diagonal entry of correlation matrices. *Transactions of the American Mathematical Society* **359** 5345–5363.