The London School of Economics and Political Science

Three Essays in Financial Econometrics

Yu-Min Yen

A thesis submitted to the Department of Finance of the London School of Economics for the degree of Doctor of Philosophy, London, July 2012

Declaration

I certify that the thesis I have presented for examination for the MPhil/PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

I declare that my thesis consists of 35,000 words.

Statement of conjoint work (if applicable)

I confirm that Chapter 2 was jointly co-authored with Dr. Philippe Mueller and Dr. Andrea Vedolin, and I contributed 33.3% of this work.

Statement of inclusion of previous work (if applicable)

Statement of use of third party for editorial help (if applicable)

Abstract

Sparse Weighted Norm Minimum Variance Portfolio

In this paper, I propose a weighted L1 and squared L2 norm penalty in portfolio optimization to improve the portfolio performance as the number of available assets N goes large. I show that under certain conditions, the realized risk of the portfolio obtained from this strategy will asymptotically be less than that of some benchmark portfolios with high probability. An intuitive interpretation for why including a fewer number of assets may be beneficial in the high dimensional situation is built on a constraint between sparsity of the optimal weight vector and the realized risk. The theoretical results also imply that the penalty parameters for the weighted norm penalty can be specified as a function of N and sample size n. An efficient coordinate-wise descent type algorithm is then introduced to solve the penalized weighted norm portfolio optimization problem. I find performances of the weighted norm strategy dominate other benchmarks for the case of Fama-French 100 size and book to market ratio portfolios, but are mixed for the case of individual stocks. Several novel alternative penalties are also proposed, and their performances are shown to be comparable to the weighted norm strategy.

Bond Variance Risk Premia (Joint work with Philippe Mueller and Andrea Vedolin)

Using data from 1983 to 2010, we propose a new fear measure for Treasury markets, akin to the VIX for equities, labeled TIV. We show that TIV explains one third of the time variation in funding liquidity and that the spread between the VIX and TIV captures flight to quality. We then construct Treasury bond variance risk premia as the difference between the implied variance and an expected variance estimate using autoregressive models. Bond variance risk premia display pronounced spikes during crisis periods. We show that variance risk premia encompass a broad spectrum of macroeconomic uncertainty. Uncertainty about the nominal and the real side of the economy increase variance risk premia but uncertainty about monetary policy has a strongly negative effect. We document that bond variance risk premia predict excess returns on Treasuries, stocks, corporate bonds and mortgage-backed securities, both in-sample and out-of-sample. Furthermore, this predictability is not subsumed by other standard predictors.

Testing Jumps via False Discovery Rate Control

Many recently developed nonparametric jump tests can be viewed as multiple hypothesis testing problems. For such multiple hypothesis tests, it is well known that controlling type I error often unavoidably makes a large proportion of erroneous rejections, and such situation becomes even worse when the jump occurrence is a rare event. To obtain more reliable results, we aim to control the false discovery rate (FDR), an efficient compound error measure for erroneous rejections in multiple testing problems. We perform the test via a nonparametric statistic proposed by Barndorff-Nielsen and Shephard (2006), and control the FDR with a procedure proposed by Benjamini and Hochberg (1995). We provide asymptotical results for the FDR control. From simulations, we examine relevant theoretical results and demonstrate the advantages of controlling FDR. The hybrid approach is then applied to empirical analysis on two benchmark stock indices with high frequency data.

Contents

List of Tables

List of Figures

Chapter 1

Sparse Weighted Norm Minimum Variance Portfolio

Abstract: In this paper, I propose a weighted l_1 and squared l_2 norm penalty in portfolio optimization to improve the portfolio performance as the number of available assets N goes large. I show that under certain conditions, the realized risk of the portfolio obtained from this strategy will asymptotically be less than that of some benchmark portfolios with high probability. An intuitive interpretation for why including a fewer number of assets may be beneficial in the high dimensional situation is built on a constraint between sparsity of the optimal weight vector and the realized risk. The theoretical results also imply that the penalty parameters for the weighted norm penalty can be specified as a function of N and sample size n. An efficient coordinate-wise descent type algorithm is then introduced to solve the penalized weighted norm portfolio optimization problem. I find performances of the weighted norm strategy dominate other benchmarks for the case of Fama-French 100 size and book to market ratio portfolios, but are mixed for the case of individual stocks. Several novel alternative penalties are also proposed, and their performances are shown to be comparable to the weighted norm strategy.
KEYWORDS: Sparsity and diversity; Weighted norm portfolio; Coordinate-wise descent algorithm.
JEL Codes: C40, C61, G11.

1.1 Introduction

How to select assets to form an optimal portfolio is one of the central issues in financial studies. Since Markowitz (1952) formalized the mean-variance portfolio optimization in which an investor is allowed to consider a portfolio's return and risk, the mean-variance optimization has become one of the principles in portfolio management for the past 60 years. However, such a strategy is also well known to perform poorly when the number of available assets N is relatively large to the sample size n. In this paper, I propose to impose a weighted l_1 and squared l_2 norm penalty on the portfolio weights to improve the portfolio performance in such high dimensional situation. The l_1 norm penalty facilitates sparsity (zero components) of the portfolio weight vector, and in turn leads to automatically selecting and excluding certain assets. Nevertheless, such sparsity may cause problems of under diversification and extreme weights of the portfolio. On contrary, the squared l_2 norm does not produce any sparsity, but it can efficiently regularize size of the weight vector. Thus the squared l_2 can function as a solution to alleviate the problems of under diversification and extreme weights of the portfolio. I call the optimal portfolio obtained by solving such weighted l_1 and squared l_2 norm penalized portfolio optimization problem the weighted norm portfolio.

The motivation to use the l_1 norm penalty for sparse estimation of the weight vector is given as follows. In the situation when the number of available assets N becomes large relative to the sample size n, if one wants to reduce impacts of estimation errors, then she can consider to choose fewer number of assets, say $N' \leq N$, for the portfolio optimization. It is equivalent to imposing an l_0 norm penalty on the weight vector and then optimize the objective function with the l_0 norm penalty. However, to obtain the optimal N', it needs to solve a combinatoric optimization which in general is intractable for large N. An alternative way is to impose the l_1 norm penalty on the weight vector. One advantage of this replacement is that the l_1 norm is a convex function of the weights, and such convex relaxation makes the modified portfolio optimization more tractable. In fact, except for the l_1 norm, there does not exist a norm penalty which can simultaneously produce sparsity as well as being a convex function of the weight vector.

A main concern of the weighted norm penalty is that the resulting portfolio weight vector is sparse, and only a subset of the whole assets is used to construct the portfolio, which violates the principle of diversity. Besides reducing the parameter uncertainty, there are other reasons of why concentrating on fewer certain assets may not be a bad idea as one might initially think. For instance, an investor may want to limit the number of assets in her portfolio due to high management and monitor costs. Simultaneously making many different decisions on asset allocations may be harmful for quality of the overall decisions[1]. Cohen et al. (2009) find

[1] A related issue is that how the number of choices present affects an individual's welfare. Plenty behavioral economics and finance papers already have discussions on this issue, and conclude that a larger choice set may not necessarily make an individual better off (e.g., Abaluck and Gruber, 2011; Beshears et al., 2006; Iyengar and Kamenica, 2010).

that a few number of assets that a fund manager concentrates on, which they termed the fund manager's best ideas, can consistently have higher risk adjusted returns and outperform the rest of the positions in the manager's portfolios. Having information advantages is another reason of why investors are willing to concentrate on certain assets. For example, home investors can have superior access to information about domestic firms or economic conditions. Such information asymmetry is often used to explain home bias phenomenon in international portfolio management literature: the home investors often concentrate too much on domestic assets and are reluctant to hold foreign assets. Moreover, Nieuwerburgh and Veldkamp (2009) argue that even if the home investors can eliminate such home bias by learning, they will still choose to learn more about the home assets, since it deepens their information advantages and lets them enjoy more excess returns of home assets than the foreign investors.

Combining both the l_1 and squared l_2 norm penalties together in the portfolio optimization is rarely seen in previous literature[2], although using them separately is not a new idea. Brodie et al. (2009), DeMiguel et al. (2009a), Fan et al. (2009), and Welsch and Zhou (2007) show that superior portfolio performances can be obtained when the l_1 norm or squared l_2 norm penalty is used in the portfolio optimization. Imposing the norm penalties is often viewed as a way to reduce the size of the portfolio weight vector. It produces a similar effect as a shrinkage estimator does for reducing the estimation errors.

However, the issue that different norm penalties result in different patterns of sparsity in the portfolio weight vector seems not to be fully addressed in the relevant literatures. In this paper, I show that the number of assets included in the portfolio (or the number of assets with nonzero weights) is the key to explain why the weighted norm approach can work well in the situation when the number of available assets N is large. I compare the out-of-sample (oos) conditional variances of different portfolios. The oos conditional variance is a reasonable measure for the risk that an investor will immediately face in the next period when she adopts a certain portfolio strategy in the current period. I prove that under certain conditions, the oos conditional variance of the weighted norm portfolio will be less than that of some benchmark portfolios with high probability, as both of the number of available assets N and sample size n increase. One of the sufficient conditions explicitly puts a constraint between sparsity of the weight vector and the true conditional variances of these portfolios. While impacts from the estimation errors can be mitigated by limiting size of the portfolio weight vector, this condition characterizes a relationship between the other two fundamental factors affecting portfolio performances: namely how many assets and which assets should to be included in the portfolio, and in turn provides a heuristic justification on why including fewer assets may be beneficial in the high dimensional situation.

[2]Various norm penalties have been widely used in statistics for model selections and ill-posed problems, and there are plenty literatures devoted on studying their properties. Using the l_1 norm penalty can be dated to Tibshirani (1996). The squared l_2 norm penalty was originally proposed by Russian mathematician Andrey N. Tikhonov. Combining the l_1 and squared l_2 norm penalties for regression problems can be found in Zou and Hastie (2005). Summary for recent developments on the norm penalty approaches in statistics can be found in Hastie et al. (2009) and Buhlmann and van de Geer (2011).

Using the weighted norm strategy is different from using an ordinary shrinkage estimator that only shrinks particular elements of the estimated covariance matrix to their targets. The weighted norm approach is an estimation strategy that not only shrinks particular elements but also reduces dimension of the estimated covariance matrix.

In addition to the econometric side, I also give explanations on why using the norm penalty in the portfolio optimization is reasonable from other perspectives. Imposing the norm penalty is similar as an investor wants to limit transaction costs or exposure to risky assets, or faces liquidity constraint such as margin requirements. This approach also can be viewed as an investor making a decision based on marginal increment of the portfolio variance. The view is closely related to Gabaix (2011), who shows that many different bounded rationality and psychological phenomenons can be delineated by models that incorporate the l_1 norm penalty into individuals' optimization problems. I also discuss the relationships between the weighted norm approach, the maximum a posteriori probability (MAP) estimator and the minimum mean square deviation problem.

The theoretical results also imply that the penalty parameters for the weighted norms can be specified as a function of N and n. The specification helps us to tackle the problem of choosing the optimal penalty parameters. Previous research suggests that one can use nonparametric methods such as cross validation to choose the optimal penalty parameters (DeMiguel et al., 2009a). However, cross validation is computational intensive and may produce very unstable sequences of the penalty parameters. The instability of the penalty parameters may damage the performance of the weighted norm strategy. On contrary, my specification for the penalty parameters is easier for implementation and can produce stable sequences of the penalty parameters, leading to satisfactory results in the empirical analysis.

To solve the weighted norm minimum variance portfolio (mvp) optimization problem, I use coordinate-wise descent algorithm. The algorithm is fast, efficient and can be easily extended to other kinds of norm penalties. In addition, unlike the popular Least Angle Regression (LARS) type algorithm (Efron et al., 2004) adopted by previous studies (Brodie et al., 2009; Fan et al., 2009), the coordinate-wise descent algorithm can deliver an exact solution to the weighted norm mvp optimization problem.

I use two real data sets, Fama-French 100 size and book to market portfolios (FF100) and three hundred stocks randomly chosen from CRSP data bank (CRSP300), to demonstrate how the weighted norm approach performs in real world. The calibrated covariance matrix is estimated via the sample covariance estimator with an expanding window scheme. The covariance matrix estimated using the expanding window scheme may be less capable to capture dynamics of the true conditional covariance matrix than that using rolling window schemes. However, the expanding window scheme can make estimated covariances stable, leading to a decrease in the portfolio turnover rate. The other source for reducing the portfolio turnover rate is the l_1 norm penalty. Unlike explicitly imposing constraint on the portfolio turnover rate (e.g., DeMiguel et al., 2010; Kirby and Ostdiek, 2011), I find that the two sources are enough to prevent high transaction costs.

For the case of FF100, the weighted norm portfolio yields an annualized variance from 82.97% to 98.60% and an annualized Sharp ratio from 1.06 to 1.18. On the other hand, the annualized variances for other three benchmark strategies: $1/N$, no-shortsales mvp and global minimum variance portoflio (gmvp) are 311.16%, 176.21% and 77.66%, respectively. The corresponding annualized Sharpe ratios are 0.45, 0.74 and 0.94, respectively. For the case of CRSP300, the performances of the weighted norm mvp are mixed. Although it is able to produce an annualized variance lower than those of other benchmark strategies, it fails to achieve higher Sharpe ratio than the no-shortsales mvp. These empirical results are robust to a change in the frequency of balancing the portfolios.

I further examine whether the weighted norm mvp can deliver a higher expected utility than other benchmarks under the assumption that the investor's utility function is nondecreaseing and concave. I use the stochastic dominance tests with a subsampling scheme proposed by Linton et al. (2005) for the assessment. The result shows that the weighted norm mvp is less risky than other benchmark portfolios.

Previous literatures argue that adding the estimated return vector into the portfolio optimization often damages portfolio performances (Jagannathan and Ma, 2003; DeMiguel et al., 2009a)). I re-examine this argument by imposing an additional target return constraint into the weighted norm portfolio optimization. I find that the performances are not as good as the case without such target return constraint imposed, which are in line with what the previous literatures find.

Finally, I investigate whether imposing different forms of norm penalties on the portfolio optimization can obtain better portfolio performances than the weighted norm penalty does. Three novel alternative penalty functions are introduced, and the reasons why they can be used in the portfolio optimization are also discussed. I find they can deliver at least comparable performances as the weighted norm mvp and other benchmarks.

In addition to the norm penalty approach, there are many other methods proposed for improving portfolio performances as the number of asset N goes large. The most frequently used one may be to assign the weights via some simple rules, and avoid massive estimations. The value weighted and equally weighted ($1/N$) portfolios are such examples. DeMiguel et al. (2009b) show how such simple strategies can outperform other sophisticated strategies. The second way is to construct more robust statistical estimators for the mean vector and covariance matrix of the asset returns, such as bias-adjusted or Bayesian shrinkage estimators (El Karoui, 2010; Jorion, 1986; Kan and Zhou, 2007; Ledoit and Wolf, 2003; Lai, Xing, and Chen, 2011). Some of the improved estimations can be incorporated and transform the initial portfolio optimization to a modified one, and a new class of optimal weight vectors can be obtained. Frahm and Christoph (2011) and Tu and Zhou (2011) show that a suitable linear combination of weights of a benchmark portfolio and a more sophisticated strategy often provides better performances than either only one of them is considered. One also can treat the optimal portfolio weights as a function of parameters in certain structural models. These parameters can be directly estimated and then used to construct the optimal portfolio weights. Brandt et al. (2009) and Kirby and Ostdiek (2011) show such method can deliver better portfolio performances than some benchmark portfolio strategies.

The rest of the paper is organized as follows. In Section 1.2, I introduce the weighted norm approach and describe some basic properties of the weighted norm mvp. In Section 1.3 I give some explanations on the norm constrained mvp problem from various economic and statistical perspectives. Section 1.4 details the main theoretical results. I then describe the coordinate-wise descent algorithm for solving a benchmark case in Section 1.5. In Section 1.6, I present the empirical results. Section 1.7 is the conclusion.

1.2 Methodology

1.2.1 Weighted Norm MVP Optimization

The weighted norm mvp (minimum variance portfolio) optimization is defined as

$$\min_{\mathbf{w}} \mathbf{w}^{\mathbf{T}}\Sigma\mathbf{w} + \lambda_1 \|\mathbf{w}\|_1 + \lambda_2 \|\mathbf{w}\|_2^2 \,, \text{ subject to } A\mathbf{w} = \mathbf{u}, \tag{1.1}$$

where $\mathbf{w}$ is a $N \times 1$ portfolio weight vector, Σ is a $N \times N$ covariance matrix of asset returns (or asset excess returns). The i,jth off-diagonal term of Σ is denoted by σ_{ij}, $i,j = 1,\ldots,N$, and $i \neq j$. The ith diagonal term of Σ is denoted by $\sigma_{ii} = \sigma_i^2$, $i = 1,\ldots,N$. The optimization problem (1.1) has the portfolio variance plus a penalty function on the portfolio weights as its objective function, and subject to a set of linear constraints. The penalty function imposed here is called the weighted norm penalty. Without such penalty, the portfolio optimization is the global minimum variance portfolio (gmvp) optimization.

$\|\mathbf{w}\|_1 = \sum_{i=1}^N |w_i|$ and $\|\mathbf{w}\|_2^2 = \sum_{i=1}^N w_i^2$ are the l_1 norm and the squared l_2 norm of the portfolio weights. The parameters $\lambda_1, \lambda_2 \in \mathbb{R}^+$ are called the penalty parameters. $A\mathbf{w} = \mathbf{u}$ is a system of linear constraints on the weights, where $\dim(A) = k \times N$ and $\dim(\mathbf{u}) = k \times 1$. To guarantee that the objective function of (1.1) is a convex function of $\mathbf{w}$, Σ should be positive semidefinite (psd). In addition, if the set of solution for $A\mathbf{w} = \mathbf{u}$ is non-empty, then $\mathbf{w}$ is feasible. With positive semidefiniteness of Σ and feasibility of $\mathbf{w}$, (1.1) is a well defined convex optimization problem.

For the linear constraints, the most frequently used one is the full investment constraint, which requires sum of portfolio weights equals to 1: ${\mathbf{1}^{\mathbf{T}}}_N\mathbf{w} = \mathbf{w}^{\mathbf{T}}\mathbf{1}_N = 1$, where $\mathbf{1}_N$ is a $N \times 1$ vector which elements are all 1's. Another example is the target return constraint, in which the expected portfolio return should satisfy a certain desired level: $\mu^{\mathbf{T}}\mathbf{w} = \mathbf{w}^{\mathbf{T}}\mu = \overline{\mu}$, where μ is a $N \times 1$ vector of expected asset returns, and $\overline{\mu}$ is the desired portfolio return.

Except the above commonly used settings, the linear constraints can be specified for other purposes. Cochrane (2011) considers to constrain covariance of the portfolio return and a factor f,

$$cov\left(R_p, f\right) = \sum_{i=1}^{N} w_i cov\left(R_i, f\right) = \sigma^{\mathbf{T}}{}_{f,\mathbf{R}}\mathbf{w} = \xi_f, \tag{1.2}$$

where $\sigma_{f,\mathbf{R}}$ is a $N \times 1$ vector in which the jth element is the covariance of the jth asset return and the factor f. ξ_f is the desired level of comovement of the portfolio return and the factor f. The motivation of such linear constraint is that the investor may want to limit volatility of her wealth while the factor f fluctuates. For instance, a high-tech company employee may fear that her labor income declines simultaneously with values of her holding stocks. To prevent this happen, she would like to hold a portfolio which has a low correlation with her labour income. Such portfolio may be constructed by solving a mvp optimization with linear constraint (1.2) by treating the factor f as her labor income stream.

1.2.2 Basic Properties

The Lagrangian of the optimization problem (1.1) is given by

$$L\left(\mathbf{w}, \gamma; \Sigma, \lambda_1, \lambda_2\right) = \mathbf{w}^{\mathbf{T}}\Sigma\mathbf{w} + \lambda_1 \left\|\mathbf{w}\right\|_1 + \lambda_2 \left\|\mathbf{w}\right\|_2^2 + +\gamma^{\mathbf{T}}\left(\mathbf{u} - A\mathbf{w}\right),$$

where γ is a $k \times 1$ vector of the Lagrange multipliers. Let $\mathbf{w}^* = (w_1^*, \ldots, w_N^*)$ denote the optimal solution of (1.1), and $S = \{i : w_i^* \neq 0\}$ and $S^c = \{i : w_i^* = 0\}$ denote the sets of assets with nonzero and zero weights respectively. Also let $|S|$ and $|S^c|$ denote the cardinality of the set S and S^c. Without loss of generality, one can do some rearrangements on the optimal weight vector $\mathbf{w}^*$. I let the first $|S|$ elements of $\mathbf{w}^*$ be the nonzero weights, and the rest $|S^c| = N - |S|$ be the zero weights. Then the following partition for Σ can be obtained,

$$\Sigma = \begin{pmatrix} \Sigma_{ss} & \Sigma_{ss^c} \\ \Sigma_{s^cs} & \Sigma_{s^cs^c} \end{pmatrix}.$$

Here Σ_{ss} ($\Sigma_{s^cs^c}$) is a $|S| \times |S|$ ($|S^c| \times |S^c|$) sub-principal matrix of Σ, which can be constructed by deleting $N - |S|$ ($N - |S^c|$) columns and the corresponding rows from Σ. The matrix Σ_{ss} is the covariance matrix of the $|S|$ assets which optimal weights are nonzero, and $\Sigma_{s^cs^c}$ is the covariance matrix of the $|S^c|$ assets which optimal weights are zero. In the off-diagonal parts, $\Sigma_{ss^c} = \Sigma_{s^cs}^{\mathbf{T}}$, which have dimensions $|S| \times |S^c|$ and $|S^c| \times |S|$ respectively. They are matrices of covariances between the nonzero weighted and zero weighted assets. Finally one also can have the following similar partition for matrix A,

$$A = \left(A_s, A_{s^c}\right),$$

where $\dim(A_s) = k \times |S|$ and $\dim(A_{s^c}) = k \times |S^c|$.

Let $\mathbf{w}_s^*$ be the $1 \times |S|$ optimal nonzero weight vector from solving (1.1). At the stationary point, the following KKT conditions should hold,

$$2\left(\Sigma_{ss}+\lambda_2\mathbf{I}_{ss}\right)\mathbf{w}_s^* = A_s^{\mathbf{T}}\gamma^* - \lambda_1 \times sign\left(\mathbf{w}_s^*\right), \tag{1.3}$$

$$A_s\mathbf{w}_s^* = \mathbf{u}, \tag{1.4}$$

$$\left\|2\Sigma'_{s^c s}\mathbf{w}_s^* - A_{s^c}^{\mathbf{T}}\gamma^*\right\|_\infty \leq \lambda_1. \tag{1.5}$$

where $\mathbf{I}_{ss}$ is a $|S|\times|S|$ identity matrix, $sign\,(.)$ is the sign function, and $\|.\|_\infty$ is the sup norm. Let $\mathbf{0}_{|S^c|}$ be the $1\times|S^c|$ zero vector. The optimal weight vector is then given by

$$\mathbf{w}^* = \left(\mathbf{w}_s^*, \mathbf{0}_{|S^c|}\right).$$

Let $\Sigma'_{ss} = \Sigma_{ss} + \lambda_2\mathbf{I}_{ss}$, $M_s = A_s\Sigma_{ss}^{-1}A_s^{\mathbf{T}}$ and $M'_s = A_s\Sigma_{ss}'^{-1}A_s^{\mathbf{T}}$. The optimal nonzero weights can be solved by using the KKT conditions, which are shown in the following lemma.

Lemma 1 *Let $\mathbf{w}_s^*$ be the vector of nonzero portfolio weights and γ^* be the Lagrange multiplier of the constraint $A\mathbf{w}=\mathbf{u}$ from solving (1.1). Then*

$$\begin{aligned}\mathbf{w}_s^* &= \mathbf{w}_{2,s}^* + \frac{\lambda_1}{2}\left(\Sigma_{ss}'^{-1}A_s^{\mathbf{T}}\delta_{2,s} - \Sigma_{ss}'^{-1}sign\left(\mathbf{w}_s^*\right)\right),\\ \gamma^* &= \gamma_{2,s}^* + \lambda_1\delta_{2,s},\end{aligned}$$

where

$$\begin{aligned}\mathbf{w}_{2,s}^* &= \Sigma_{ss}'^{-1}A_s^{\mathbf{T}}M_s'^{-1}\mathbf{u},\\ \delta_{2,s} &= M_s'^{-1}A_s\Sigma_{ss}'^{-1}sign\left(\mathbf{w}_s^*\right),\\ \gamma_{2,s}^* &= 2M_s'^{-1}\mathbf{u}.\end{aligned}$$

and $sign\left(\mathbf{w}_s^*\right)$ is a $|S|\times 1$ column vector which elements are sign of the nonzero weights.

Proof of Lemma 1 can be found in Appendix 1.8.1. It is not difficult to see that $\mathbf{w}_{2,s}^*$ is the optimal solution for the following mvp optimization, in which only asset $i\in S$ are used,

$$\mathbf{w}_{2,s}^* = \arg\min_{\mathbf{w}} \mathbf{w}^{\mathbf{T}}\Sigma'_{ss}\mathbf{w}, \text{ subject to } A_s\mathbf{w}=\mathbf{u}, \tag{1.6}$$

and $\gamma_{2,s}^*$ is a vector of Lagrange multipliers for the linear constraints $A_s\mathbf{w}=\mathbf{u}$. Note that (1.6) is in fact an mvp optimization penalized by the squared l_2 norm penalty[3]. Without such squared l_2 norm penalty, (1.6) becomes a gmvp optimization. Define $\mathbf{w}_{un,s}^*$ as the optimal solution of such gmvp optimization,

$$\mathbf{w}_{un,s}^* = \arg\min_{\mathbf{w}} \mathbf{w}^{\mathbf{T}}\Sigma_{ss}\mathbf{w}, \text{ subject to } A_s\mathbf{w}=\mathbf{u}. \tag{1.7}$$

[3]The objective function of (1.6) is sum of a portfolio variance and squared l_2 norm penalty on the weight vector with penalty parameter λ_2.

Similar to $\mathbf{w}^*_{2,s}$, $\mathbf{w}^*_{un,s}$ and vector of the Lagrange multipliers $\gamma^*_{un,s}$ can be solved explicitly as

$$\begin{aligned} \mathbf{w}^*_{un,s} &= \Sigma_{ss}^{-1} A_s^{\mathbf{T}} M_s^{-1} \mathbf{u}, \\ \gamma^*_{un,s} &= 2 M_s^{-1} \mathbf{u}. \end{aligned}$$

Throughout the paper, I use the following notations to denote the portfolio variances via solving different mvp optimizations,

$$\begin{aligned} \sigma_s^2 : &= \mathbf{w}^{*\mathbf{T}} \Sigma \mathbf{w}^*, \\ \sigma_{un}^2 : &= \mathbf{w}_{un}^{*\mathbf{T}} \Sigma \mathbf{w}^*_{un}, \\ \sigma_{2,s}^2 : &= \mathbf{w}_{2,s}^{*\mathbf{T}} \Sigma_{ss} \mathbf{w}^*_{2,s}, \\ \sigma_{un,s}^2 : &= \mathbf{w}_{un,s}^{*\mathbf{T}} \Sigma_{ss} \mathbf{w}^*_{un,s}. \end{aligned}$$

σ_s^2 is the minimum portfolio variance can be achieved via solving (1.1) [4]. σ_{un}^2 is the gmvp portfolio variance and $\mathbf{w}^*_{un}$ is optimal gmvp weight vector. $\sigma^2_{2,s}$ and $\sigma^2_{un,s}$ are the portfolio variances via using the optimal weight vector from (1.6) and (1.7).

With the above notations, I introduce the following lemma, which provides a key inequality for the proof of the main results in Section 1.4.2.

Lemma 2 *Let* $\phi_j(\Sigma_{ss})$, $j = 1, \ldots, |S|$ *denote eigenvalues of* Σ_{ss}. *Let*

$$\begin{aligned} \Sigma'_{ss} &= \Sigma_{ss} + \lambda_2 \mathbf{I}_{ss}, \\ \phi^{\min}(\Sigma_{ss}) &= \arg \min_{j=1,\ldots,|S|} \phi_j(\Sigma_{ss}), \end{aligned}$$

where $\mathbf{I}_{ss}$ *denote a* $|S| \times |S|$ *identity matrix. Suppose* $\phi^{\min}(\Sigma_{ss}) > 0$. *Then*

$$0 \leq \sigma_s^2 - \sigma_{un,s}^2 \leq c_{s,1}(1 + c_{s,1}) \sigma_{un,s}^2 - \lambda_2 \|\mathbf{w}_s^*\|_2^2 + \frac{\lambda_1}{2} \left(\left\|\mathbf{w}^*_{2,s}\right\|_1 - \|\mathbf{w}_s^*\|_1 \right),$$

where $c_{s,1} = \lambda_2^{\frac{1}{2}} \left(\phi^{\min}(\Sigma_{ss})\right)^{-\frac{1}{2}}$.

Proof of Lemma 2 can be found in Appendix 1.8.2. The lemma indicates that σ_s^2 is bounded by a linear combination of the scaled $\sigma^2_{un,s}$, the penalty parameters and the norm penalties. Lemma 1 and Lemma 2 are stated in a deterministic way, but these results still hold if Σ is replaced by some positive semidefinite estimates of the covariance matrix. It is helpful, since it provides useful tools to construct further asymptotical results when the simple sample covariance estimator is calibrated for solving (1.1).

[4] Note that $\mathbf{w}^{*\mathbf{T}} \Sigma \mathbf{w}^* = \mathbf{w}_s^{*\mathbf{T}} \Sigma_{ss} \mathbf{w}_s^*$.

1.2.3 Relation to the Benchmark Portfolios

It can be easily seen that one can solve the weighted norm mvp optimization (1.1) as a l_1 penalized portfolio optimization with the covariance matrix $\Sigma' = \Sigma + \lambda_2 \mathbf{I}_{NN}$, where $\mathbf{I}_{NN}$ denotes a $N \times N$ identity matrix. In a special situation when $\lambda_1 = 0$ and $\lambda_2 \to \infty$, the optimal weights from solving (1.1) will converge to $1/N$ (DeMiguel et al., 2009a). When the linear constraint is the full investment constraint $\mathbf{w}^{\mathbf{T}}\mathbf{1}_N = 1$ and only the l_1 penalty is active ($\lambda_2 = 0$), it can be shown that as λ_1 is beyond some threshold $\overline{\lambda} > 0$, the optimal weight vectors of the weighted norm mvp and no-shortsales mvp will be identical (Brodie et al., 2009; DeMiguel et al., 2009a; Fan et al., 2009). On the other hand, in this situation, using any $\lambda_1 \geq \overline{\lambda}$ for solving (1.1) will only generate the optimal no-shortsales weight vector. Practically to pin down the upper bound $\overline{\lambda}$ is important if one wants to search the optimal λ_1 over a range of possible values. Yen and Yen (2011) show that the upper bound $\overline{\lambda}$ can be easily obtained by using the optimal no-shortsales solution.

Furthermore, from the KKT conditions, it can be shown that the optimal nonzero weights of the no-shortsales mvp can be obtained by solving (1.7) if one replaces S with $S_{ns} = \{i : w_{ns,i} > 0\}$, where $w_{ns,i}$ is the optimal no-shortsales weight for asset i and S_{ns} is the set of assets included in the optimal no-shortsales mvp. In other words, the no-shortsales mvp, as a mvp with a heavy l_1 norm penalty, can be equivalently obtained from the gmvp optimization via using a dimensional reduction covariance matrix Σ_{ss} with $S = S_{ns}$, and assigning zero weights to assets $i \notin S_{ns}$.

The way to construct the no-shortsales mvp is different from Jagannathan and Ma (2003), in which the authors argue that the no-shorsales mvp can be equivalently constructed from the gmvp optimization via using the following shrinkage type covariance matrix

$$\Sigma^{JM} = \Sigma - \left(\nu \mathbf{1}^{\mathbf{T}} + \nu^{\mathbf{T}} \mathbf{1}\right),$$

where ν is the vector of the Lagrange multipliers for the nonnegativity constraint $w_i \geq 0$, $i = 1, \ldots, N$. Nevertheless, here shows that the no-shorsales mvp is a typical sparse portfolio which can be constructed from casting the gmvp optimization on a suitable subset S_{ns} of the whole available assets, without shrinking any element in the covariance matrix used.

1.3 Explanations on the Weighted Norm MVP Optimization

1.3.1 Individual's Financing Constraint

If one sets $\lambda_1 = \lambda\alpha$ and $\lambda_2 = \lambda(1-\alpha)$, where $\lambda \in \mathbb{R}^+$ and $\alpha \in [0,1]$, the portfolio optimization of (1.1) is essentially the same as minimizing the portfolio variance subject to the linear constraints and the following norm constraint,

$$\alpha \|\mathbf{w}\|_1 + (1-\alpha)\|\mathbf{w}\|_1^2 \leq c. \tag{1.8}$$

When $\alpha = 1$, $\|\mathbf{w}\|_1 \leq c$ is called a gross exposure constraint in Fan, Zhang, and Yu (2009). It can be viewed as the investor wants to minimize the portfolio variance (or maximize the mean-variance utility when the mean return vector of assets is taken into account), but still trying to limit the investment positions exposed to the risky assets. The constant c is the maximum allowable amounts of investments on the risky assets, which reflects the investor's concern on parameter uncertainty due to statistical estimation errors.

Now consider a more general version of the l_1 norm constraint

$$\sum_{i=1}^{N} \nu_i |w_i| \leq c,$$

where ν_i is a nonnegative constant. In Brodie et al. (2009), ν_i is viewed as a measure of transaction cost, such as bid-ask spread of asset i. One can also interpret ν_i as the requirement of margin on asset i (Garleanu and Pedersen, 2011). As for the case of (1.1) with $\alpha = 1$, it is equivalent to treating all of such transaction costs being equal to one[5].

1.3.2 Decision Based on Marginal Increment of the Portfolio Variance

Now suppose that, due to fear of estimation errors, the investor believe that the corrected estimated covariance matrix is the regularized one $\hat{\Sigma}' = \hat{\Sigma} + \lambda_2 \mathbf{I}_{NN}$ rather than $\hat{\Sigma}$. Also to simplify the analysis, assume the investor only faces the full investment constraint $\mathbf{w}^{\mathbf{T}}\mathbf{1} = 1$. Then at the stationary point, the marginal change of the in sample portfolio variance due to including asset i is given by

$$\frac{\partial \mathbf{w}^{\mathbf{T}} \hat{\Sigma}' \mathbf{w}}{\partial w_i} = 2w_i \left(\hat{\sigma}_i^2 + \lambda_2\right) + 2\sum_{j \neq i}^{N} w_j \hat{\sigma}_{ij} = \gamma - \lambda_1 \tag{1.9}$$

if $w_i > 0$, and

$$\frac{\partial \mathbf{w}^{\mathbf{T}} \widehat{\Sigma}' \mathbf{w}}{\partial w_i} = 2w_i \left(\hat{\sigma}_i^2 + \lambda_2\right) + 2\sum_{j \neq i}^{N} w_j \hat{\sigma}_{ij} = \gamma + \lambda_1 \tag{1.10}$$

if $w_i < 0$. If $\lambda_1 = 0$, the marginal change is the Lagrangian multiplier γ, which is the shadow price to measure how the portfolio variance changes when the investor's wealth changes. If both $\lambda_1 > 0$, the marginal change due to buying an asset is $\gamma - \lambda_1$, and it is lower than the one due to shorting an asset, $\gamma + \lambda_1$. It can be shown that $\gamma > \lambda_1$ always holds at the stationary point. Therefore the marginal change is always positive. In general, absolute values of the optimal weights obtained from solving (1.1) is smaller than those obtained from solving the

[5]When $0 < \alpha < 1$, the optimization problem (1.1) can be viewed as a l_1 norm penalized mvp problem with covariance matrix $\Sigma + \lambda(1-\alpha)\mathbf{I}_{NN}$, therefore the explanations given above still apply

gmvp. In addition, from (1.5), if $w_i = 0$,

$$\gamma - \lambda_1 \leq \frac{\partial \mathbf{w}^T \widehat{\Sigma}' \mathbf{w}}{\partial w_i} \leq \gamma + \lambda_1. \tag{1.11}$$

If λ_1 becomes large, it is more possible that $\partial \mathbf{w}^T \widehat{\Sigma}' \mathbf{w} / \partial w_i$ will fall into the interval $[\gamma - \lambda_1,\ \gamma + \lambda_1]$. Therefore more assets will be excluded[6]. Meanwhile, it is less likely that (1.10) will hold, since $\gamma + \lambda_1$ will also increase. Furthermore, as mentioned, if λ_1 is beyond some upper bound $\bar{\lambda}$, only no-shortsales positions will be included. Note that in this extreme case, (1.9) still needs to hold. Thus the optimal no-shortsales solution is equivalent to the solution of a gmvp with a certain subset of the whole assets.

In summary, the weighted norm portfolio strategy can be viewed as a way that the investor, with the believe that $\hat{\Sigma}'$ is the corrected estimated covariance matrix, to choose the penalty parameter λ_1 to decide whether to include an asset or not. The decision is based on how the portfolio variance changes due to including an asset. If including an asset causes a large (small) enough increase in the portfolio variance, then the asset weight is negative (positive). If including an asset only causes a mild change in the portfolio variance, the asset will be excluded. One can interpret it as that the investor are concerned with both the sign and magnitude of the asset weight, and the decision on assigning the weight hinges on how the portfolio variance changes due to including the asset. If including the asset leads a change larger than $\gamma + \lambda_1$ or smaller than $\gamma - \lambda_1$ in the portfolio variance, then the investor will consider the information is sufficient to safely determine the sign of the asset weight. On contrary, if including an asset only makes the portfolio variance change mildly, say at some level between $\gamma - \lambda_1$ and $\gamma + \lambda_1$, then the investor will think that the information is too ambiguous to make any decisions, and she had better make the asset redundant. It can be seen as the investor's attitude to the parameter uncertainty, and the penalty parameter λ_1 controls the degree of such attitude.

When λ_1 becomes extremely large, only few assets will be included, and only assets which give a very large marginal increment to the portfolio variance can have negative weights. However, including such assets is risky, since a small change in their weights will cause an extremely large volatility in the portfolio return. Thus the investor will try to avoid such assets. Finally, only assets which give a small enough marginal increment to the portfolio variance will be included, and all their weights are positive.

1.3.3 Relation to Bounded Rationality and Psychological Phenomenons

Gabaix (2011) shows that adding the l_1 norm penalty to an individual's optimization problems can generate rich bounded rationality and psychological effects. In his model, the individual tries to simplify her optimal decision process by considering only a few important parameters in the utility function. To achieve this, at first the individual minimizes a cost function for

[6] Here one can safely ignore the case when $w_i = 0$ and $\partial \mathbf{w}^T \widehat{\Sigma}' \mathbf{w} / \partial w_i = \gamma - \lambda_1$ (or $\partial \mathbf{w}^T \widehat{\Sigma}' \mathbf{w} / \partial w_i = \gamma + \lambda_1$), since the probability that the two events occur simultaneously is negligible

choosing the important parameters in the utility function. The cost function has a quadratic form plus the l_1 norm penalty on the parameters. With the l_1 norm penalty, it is easy to achieve a sparse solution for the parameter vector. Such decision process can be explained as the individual prefer to frame her view on the real world not too complex. It is a reasonable setting in reality, since no one can freely consider a large amount of relevant information for her decisions. If some of the relevant information is not so important, the individual had better to damp it. Then the individual can consider her optimal actions based on the simplified utility function.

To get more insights of Gabaix (2011), one can further generalize the l_1 norm penalty used in 1.1 as

$$\|\mathbf{w} - \mathbf{w}_0\|_1 .$$

As in the portfolio optimization, $\mathbf{w}_0$ can be viewed as the investor's default decision on the asset allocation. In my case, I set such default weight vector equal to zero. With such generalized l_1 norm penalty, one can image that some elements in the optimal weight vector will naturally be the default weights, like the zero components resulted by $\|\mathbf{w}\|_1$. It means the investor's decision with respective to some of the assets will not be changed. In practice, this property is helpful on reducing portfolio turnover rate, as shown in DeMiguel et al. (2010), which sets $\mathbf{w}_0$ equal to optimal weight vector of the gmvp.

In economics, such sticking-to-default effect is called inattention, while in psychology, it is called an endowment effect. This effect often arises in real world when the investor faces too many assets to choose. The investor may be aware that simultaneously to make many different decisions perhaps will lower overall quality of these decisions. A better way to do it is to keep as many initial decisions as one can, and change some of them if it is really necessary.

1.3.4 The Maximum a Posteriori Probability (MAP) Estimator

Zou and Hastie (2005) show that regression coefficients regularized by the elastic net constraint can be viewed as having a compromised prior between the Gaussian and Laplace distributions. Based on this result, one can give the optimal weights in (1.1) a similar statistical explanation. Let $\mathbf{R}_t$ be the $N \times 1$ vector of asset returns at time t. Suppose that given $\mathbf{w}$, estimated mean return $\overline{\mathbf{R}}$, and portfolio variance σ^2_{por}, the investor believes that the portfolio return $\mathbf{w}^{\mathbf{T}}\mathbf{R}_t \mid \mathbf{w}, \overline{\mathbf{R}}, \sigma^2_{por} \overset{iid}{\sim} \mathcal{N}\left(\mathbf{w}^T\overline{\mathbf{R}}, \sigma^2_{por}\right)$ for all t. Also suppose that the investor has a prior belief that the weight vector $\mathbf{w}$ follows a distribution with the density proportional to

$$\exp\left(-\psi \times \left(\alpha \|\mathbf{w}\|_1 + (1-\alpha)\|\mathbf{w}\|_2^2\right)\right)\mathbb{I}\{A\mathbf{w} = \mathbf{u}\},$$

where $\psi > 0$ and $\alpha \in [0,1]$. Then conditional on σ^2_{por}, $\{\mathbf{R}_t\}_{t=1}^n$ and $\overline{\mathbf{R}}$, the density of the posterior distribution of the portfolio weights $\mathbf{w}$ will be proportional to

$$\exp\left(-\frac{T-1}{2\sigma_{por}^2}\mathbf{w}^T\widehat{\Sigma}\mathbf{w}-\psi\times\left(\alpha\left\|\mathbf{w}\right\|_1+(1-\alpha)\left\|\mathbf{w}\right\|_2^2\right)\right)\mathbb{I}\left\{A\mathbf{w}=\mathbf{u}\right\}, \tag{1.12}$$

where $\widehat{\Sigma}$ is the sample covariance matrix of $\mathbf{R}_t$. Thus maximizing the log posterior portfolio weight density with respect to $\mathbf{w}$ is equivalent to solving problem (1.1) with $\Sigma = \widehat{\Sigma}$ and

$$\begin{aligned}\lambda_1 &= \frac{2\sigma_{por}^2\psi\alpha}{T-1},\\ \lambda_2 &= \frac{2\sigma_{por}^2\psi\left(1-\alpha\right)}{T-1},\end{aligned}$$

and the optimal $\mathbf{w}$ is the maximum a posteriori probability (MAP) estimator for $\mathbf{w}$.

The above result is related to proposition 8 and 9 in DeMiguel et al. (2009a), in which they stated the cases when $\alpha = 1$. The posterior distribution of (1.12) can be rewritten as the one proportional to

$$\exp\left(-\frac{T-1}{2\sigma_{por}^2}\mathbf{w}^T\left(\widehat{\Sigma}+\frac{2\sigma_{por}^2\psi\left(1-\alpha\right)}{T-1}\mathbf{I}_{N\times N}\right)\mathbf{w}-\psi\alpha\left\|\mathbf{w}\right\|_{l_1}\right)\mathbb{I}\left\{A\mathbf{w}=\mathbf{u}\right\},$$

which implies the investor has a prior belief that $\mathbf{w}$ follows a distribution with the density proportional to

$$\exp\left(-\psi\alpha\left\|\mathbf{w}\right\|_{l_1}\right)\mathbb{I}\left\{A\mathbf{w}=\mathbf{u}\right\}.$$

In addition, the regularized covariance matrix estimator

$$\widehat{\Sigma}+\frac{2\sigma_{por}^2\psi\left(1-\alpha\right)}{T-1}\mathbf{I}_{N\times N}$$

is the calibrated covariance matrix estimation.

1.3.5 MVP Optimization as a Minimum Mean Square Deviation Problem

The mvp optimization has some similar characteristic to the squared loss-based linear regression estimation. Consider the following minimum mean square deviation problem:

$$\min_{\mathbf{w}}\mathbb{E}(Y-\mathbf{s}^{\mathbf{T}}\mathbf{w})^2, \text{ subject to } A\mathbf{w}=\mathbf{u}, \tag{1.13}$$

where the expectation is taken with respect to Y. To solve (1.13), one seeks a vector $\mathbf{w}$ to minimize the mean square deviation between $\mathbf{s}^{\mathbf{T}}\mathbf{w}$ and Y given that $A\mathbf{w} = \mathbf{u}$. Suppose Y is the dependent variable and $\mathbf{s}$ is a $N \times 1$ vector of predictors. $\mathbb{E}(Y - \mathbf{s}^{\mathbf{T}}\mathbf{w})^2$ can be interpreted as the expected squared prediction error, and $A\mathbf{w} = \mathbf{u}$ is the constraint on the coefficient vector. Now let $\mathbf{R}$ be a $N \times 1$ random asset return vector. If one sets $Y = \mathbf{R}^{\mathbf{T}}\mathbf{w}$, $\mathbf{s} = \mathbb{E}(\mathbf{R})$, the objective function becomes the portfolio variance $\mathbb{E}((\mathbf{R}-\mathbb{E}(\mathbf{R}))^{\mathbf{T}}\mathbf{w})^2$. Consequently (1.13) becomes the mvp optimization. It is equivalent to seeking the minimum mean square deviation between $(\mathbf{R}-\mathbb{E}(\mathbf{R}))^{\mathbf{T}}\mathbf{w}$ and zero, subject to $A\mathbf{w}=\mathbf{u}$.

When the number of covariates becomes relatively large to the sample size, recent research on large dimensional variable selections in the linear regression shows that regularization methods can work well not only for model selections but also for improving out-of-sample predictions. Bai and Ng (2008) and De Mol, Giannone, and Reichlin (2008) show regression penalized by the l_1 norm penalty can perform at least equally well or better than traditional methods on predicting important macroeconomic indicators when a large number of predictors are jointly considered. As for the mvp, when the number of assets becomes relatively large to the sample size, it is reasonable to see that the regularization methods can do the same improvements for reducing the out-of-sample portfolio variance as it does for the mean squared prediction error of the OLS regression, since the mvp optimization and the OLS estimation share a similar property on searching the optimal solution: namely, finding the optimal coefficient vector $\mathbf{w}$ to minimize the mean square deviation between two points.

1.4 An Econometric Analysis on the Weighted Norm MVP

1.4.1 Basic Settings

In this section I provide an analysis for the weighted norm mvp from econometric perspective. I explicitly introduce randomness into the portfolio optimization problem: the problem is solved via calibrating the estimated covariance matrix $\hat{\Sigma}$. Let $\mathbf{R}_t\in\mathbb{R}^N$, $t=1,\ldots,n$ denote the independent observed N asset return data points with mean $\mu=(\mu_1,\ldots,\mu_N)$ and covariance matrix Σ. Let $\widehat{\Sigma}$ denote some estimate of Σ from using these data points. The i,jth off-diagonal term of $\widehat{\Sigma}$ is denoted by $\widehat{\sigma}_{ij}$, $i,j=1,\ldots,N$, and $i\neq j$. The ith diagonal term of $\widehat{\Sigma}$ is denoted by $\widehat{\sigma}_{ii}=\widehat{\sigma}_i^2$, $i=1,\ldots,N$. In order to distinguish the deterministic case, I let all of the notations used in previous sections with a hat above to denote their analogues from solving the same mvp when Σ is replaced by $\widehat{\Sigma}$. For example , $\widehat{\mathbf{w}}^*=(\widehat{w}_1^*,\ldots,\widehat{w}_N^*)$ will be used to denote the optimal weight vector by solving (1.1) when $\widehat{\Sigma}$ is replaced by Σ, and $\widehat{S}=\{i:\widehat{w}_i^*\neq 0\}$ and $\widehat{S}^c=\{i:\widehat{w}_i^*=0\}$ will be used to denote the sets of nonzero and zero components in $\widehat{\mathbf{w}}^*$. Given $\widehat{S}$ and $\widehat{S}^c$, like the deterministic case, one can have the following partitions for $\widehat{\mathbf{w}}^*$, $\widehat{\Sigma}$ and A,

$$\begin{aligned}
\widehat{\mathbf{w}}^* &= \left(\widehat{\mathbf{w}}^*_{\widehat{S}},\mathbf{0}_{|\widehat{S}^c|}\right), \\
\widehat{\Sigma} &= \begin{pmatrix} \widehat{\Sigma}_{\widehat{S}\widehat{S}} & \widehat{\Sigma}_{\widehat{S}\widehat{S}^c} \\ \widehat{\Sigma}_{\widehat{S}^c\widehat{S}} & \widehat{\Sigma}_{\widehat{S}^c\widehat{S}^c} \end{pmatrix}, \\
A &= \left(A_{\widehat{S}},A_{\widehat{S}^c}\right).
\end{aligned}$$

Let the out-of-sample (oos) portfolio return of the weighted norm mvp at $t=n+1$ be $R_{wp,n+1}=\hat{\mathbf{w}}^{*\mathbf{T}}\mathbf{R}_{n+1}$. Given $\{\mathbf{R}_t\}_{t=1}^n$, the out-of-sample (oos) conditional portfolio variance of

(1.1) is given by

$$var\left(R_{wp,n+1}|\left\{\mathbf{R}_t\right\}_{t=1}^n\right) = \widehat{\mathbf{w}}_{\widehat{s}}^{*\mathbf{T}}\Sigma_{\widehat{s}\widehat{s}}\widehat{\mathbf{w}}_{\widehat{s}}^{*}. \tag{1.14}$$

(1.14) can be viewed as a measure of risk that an investor will immediately face at period $n+1$ if she allocates wealth according to $\widehat{\mathbf{w}}^*$. Thus the oos conditional portfolio variance is also called realized risk or out-of-sample risk (El Karoui, 2009, 2010). By similar fashion, let $R_{unp,n+1} = \widehat{\mathbf{w}}_{un}^{\mathbf{T}}\mathbf{R}_{n+1}$ be the oos portfolio return of the gmvp. The oos conditional portfolio variance of $R_{unp,n+1}$ is then given by

$$var\left(R_{unp,n+1}|\left\{\mathbf{R}_t\right\}_{t=1}^n\right) = \widehat{\mathbf{w}}_{un}^{\mathbf{T}}\Sigma\widehat{\mathbf{w}}_{un}, \tag{1.15}$$

and the oos conditional portfolio variance of the $1/N$ portfolio,

$$var\left(\frac{1}{N}\mathbf{1}_N^{\mathbf{T}}\mathbf{R}_{n+1}|\left\{\mathbf{R}_t\right\}_{t=1}^n\right) = \frac{1}{N^2}\mathbf{1}_N^{\mathbf{T}}\Sigma\mathbf{1}_N. \tag{1.16}$$

Let $\sigma_{1/N}^2 := N^{-2}\mathbf{1}_N^{\mathbf{T}}\Sigma\mathbf{1}_N$. Unlike (1.14) and (1.15), the oos conditional variance of $1/N$ portfolio is non-random, since its weights are always deterministic $1/N$.

1.4.2 Main Results

When the investor faces a large number of assets being available and has the estimated covariance matrix $\widehat{\Sigma}$, given two easily-implementable benchmark strategies: the gmvp and $1/N$, is it still worth to undertake the weighted norm mvp strategy? To answer this question, one can directly compare their realized risks, and see how large probability that (1.14) will be smaller than (1.15) (or (1.16)). Such comparisons are discussed in this subsection. Before the main results are stated, I shall introduce a definition and some conditions that are used in the analysis.

Definition 1 *(Asymptotical feasible active set) We call S the asymptotical feasible active set of assets for the weighted norm portfolio optimization, if S is a subset of $\{1,\dots,N\}$ such that as n, $N\to\infty$, at the stationary point, the following KKT conditions hold,*

$$\begin{aligned}
2\widehat{\Sigma}'_{ss}\widehat{\mathbf{w}}_s^* &= A_{ks}^{\mathbf{T}}\widehat{\gamma}_s^* - \lambda_1 \times sign\left(\widehat{\mathbf{w}}_s^*\right), \\
A_s\widehat{\mathbf{w}}_s^* &= \mathbf{u}, \\
\left\|2\widehat{\Sigma}'_{s^cs}\widehat{\mathbf{w}}_s^* - A_{s^c}^{\mathbf{T}}\widehat{\gamma}_s^*\right\|_\infty &\le \lambda_1.
\end{aligned}$$

In short, $S\subseteq\{1,\dots,N\}$ is a possible realization of $\widehat{S}$ as n, $N\to\infty$.

The following five conditions are needed for proof of the main results.

Condition 1 *All elements in Σ have finite values, and all eigenvalues of Σ are bounded away from 0 and ∞, i.e. $0 \ll \phi^{\min}(\Sigma) \le \phi^{\max}(\Sigma) \ll \infty$.*

Condition 2 *The number of linear constraints k is fixed, and the set of solution for $A\mathbf{w} = \mathbf{u}$ is non-empty. The elements in A and $\mathbf{u}$ are nonrandom.*

Condition 3 *When Σ and $\hat{\Sigma}$ are calibrated, for $0 \le \lambda_1, \lambda_2 < \infty$, the weighted norm optimization is feasible as n, $N \to \infty$. The definition of a feasible optimization problem is that there exists at least one feasible point such that the optimal value of the objective function is finite.*

Condition 4 *There exists a constant B_0 such that $\sup_{S \subseteq \{1,\ldots,N\}} P\left(\|\widehat{\mathbf{w}}_s^*\|_1 > B_0\right) = o(1)$.*

Condition 5 *$N < n$, and $\lim_{n\to\infty} N/n = \rho_N \in (0,1)$.*

The main results are stated in the following.

Theorem 1 *Suppose that $\mathbf{R}_t \overset{iid}{\sim} \mathcal{N}(\mu, \Sigma)$, $t = 1, \ldots, n$, $\widehat{\Sigma}$ is the sample covariance matrix, and conditions 1 to 5 hold. Let S be the asymptotical feasible active set of assets such that $|S| >> k$. Let $a^{\min} = \min_{i,j=1,\ldots,N} a_{ij}$, and a_{ij} has the same definition as in Lemma 4. Set*

$$\lambda_1 = \lambda_2 = \lambda_{n,N} = B_0 \sqrt{\frac{2\log N}{a^{\min} n}},$$

and assume the following maximum ratio of portfolio variances condition (MRPV) holds,

$$\sup_{S \subseteq \{1,\ldots,N\}} \left(\frac{\sigma_{un,s}^2 - \sigma_{un}^2}{\sigma_{un,s}^2} - \rho_s \right) \le 0, \tag{1.17}$$

as $n \to \infty$, where $\rho_s = \lim_{n\to\infty} |S|/n$. Then as n, $N \to \infty$,

$$P\left(\widehat{\mathbf{w}}^{*\mathbf{T}} \Sigma \widehat{\mathbf{w}}^* \le \widehat{\mathbf{w}}_{un}^{\mathbf{T}} \Sigma \widehat{\mathbf{w}}_{un}\right) \to 1.$$

If $0 < \underline{\sigma}^2 \left(N^{-1}\log N\right)^{\frac{1}{2}-\varepsilon} < \sigma_{\frac{1}{N}}^2$ holds and (1.17) is changed to

$$\sup_{S \subseteq \{1,\ldots,N\}} \left(\frac{\sigma_{un,s}^2 - \sigma_{\frac{1}{N}}^2}{\sigma_{un,s}^2} - \rho_s \right) \le 0, \tag{1.18}$$

where $0 < \underline{\sigma}^2$ and $0 < \varepsilon < 1/2$ are two constants. Then as $n \to \infty$,

$$P\left(\widehat{\mathbf{w}}^{*\mathbf{T}} \Sigma \widehat{\mathbf{w}}^* \le \sigma_{\frac{1}{N}}^2\right) \to 1.$$

Proof of Theorem 1 can be found in Appendix 1.8.6. In addition to Lemma 1 and Lemma 2, the proof also relies on Lemma 3, which shows bounded eignevalues of the sample estimate $\hat{\Sigma}$, and Lemma 4, which provides an upper bound for the tail probability of the estimation errors of elements in $\hat{\Sigma}$ [7].

Theorem 1 says that under certain conditions, the conditional variance of the weighted norm mvp will be less than that of the gmvp or $1/N$ with high probability as the number of available assets N and sample size n both go to large. The weighted norm penalty tends to result in sparsity of the weight vector, consequently only a smaller number of assets (say $|S|$) will be included in the portfolio. When only a smaller number of assets is included, the true risk of this portfolio $\sigma^2_{un,s}$, may be well approximated by the realized risk of the weighted norm mvp $\widehat{\mathbf{w}}^{*\mathbf{T}}\Sigma\widehat{\mathbf{w}}^{*}$. On contrary, the conditional variance of the gmvp $\widehat{\mathbf{w}}^{\mathbf{T}}_{un}\Sigma\widehat{\mathbf{w}}_{un}$ may be a bad approximation for its true risk σ^2_{un} due to large N. Note that the true risk of the gmvp σ^2_{un}, is the minimum value that any realized risk can achieve. So if the true risk of the portfolio with the smaller number of assets $\sigma^2_{un,s}$ is not too far from σ^2_{un}, then the realized risk of the weighted norm portfolio can have a large chance to be smaller than that of the gmvp, since the later may be more away from σ^2_{un} due to its bad empirical property.

The above result implies that, to ensure the weighted norm strategy work well, $\sigma^2_{un,s}$ and σ^2_{un} should not be too different. How closeness should the two have? It is stated by the maximum ratio of portfolio variances condition (MRPV) of (1.17): for every possible feasible active set of assets S, the maximum difference between the two should not exceed

$$\sigma^2_{un,s}\rho_s, \tag{1.19}$$

where $\rho_s = \lim_{n\to\infty} |S|/n$.

While impacts from the estimation errors can be mitigated by shrinking size of the weight vector, the MRPV condition characterizes a link between how many assets and which assets should to be included in the portfolio. The MRPV condition implies that the number of selected assets $|S|$ should increase with the sample size n. Consider the case when $|S|$ is fixed, the upper bound (1.19) goes to zero as $n \to \infty$, but $\sigma^2_{un,s} - \sigma^2_{un} > 0$ (or $\sigma^2_{un,s} - \sigma^2_{1/N} > 0$) holds, and this violates the MRPV condition.

However, this does not always mean including more assets is beneficial. In addition to inducing more estimation errors, including more assets might not ease the MRPV condition. Consider the case when $|S'|$ assets is selected, and $|S'| > |S|$ but $S \nsubseteq S'$, so $\rho_{s'} > \rho_s$. Nevertheless it is possible that $\sigma^2_{un,s} < \sigma^2_{un,s'}$, since S is not a subset of S'. In turn, the MRPV condition might hold for the portfolio with S assets, but not hold for that with S' assets. Therefore requiring a suitable set of assets included in the portfolio is also important for ensuring the weighted norm strategy work well [8].

[7] Some discussions on properties of the normal distributed returns, sample covariance matrix estimator, estimation errors are also given in Appendix 1.8.3 to 1.8.5.

[8] Note that the situation $|S'| > |S|$ but $S \nsubseteq S'$ is possible when the weighted norm strategy is used. Suppose that $\lambda_1 > \lambda'_1$. The l_1 norm penalty does not always guarantee that the set of assets being selected under λ_1 is a subset of those being selected under λ'_1, since the optimal weights are not always monotonically decreasing with the penalty parameter of the l_1 norm penalty.

For the case of comparing with $\sigma^2_{1/N}$, the MRPV condition (1.18) is similar as (1.17). Given every realized S, (1.18) can be explained as $\sigma^2_{un,s}$ should be closer to σ^2_{un} than $\sigma^2_{1/N}$ does, and the difference between the two distances $\sigma^2_{un,s} - \sigma^2_{un}$ and $\sigma^2_{un,s} - \sigma^2_{1/N}$ should not be over $\rho_s \sigma^2_{un,s}$. The MRPV condition (1.18) is even easier to hold than (1.17), since $\sigma^2_{un,s} - \sigma^2_{1/N} \geq 0$ may not hold but $\sigma^2_{un,s} - \sigma^2_{un} \geq 0$ always does. However, here $\sigma^2_{1/N}$ is required to not decrease too fast with N. Note that as n, N go to large, approximation error of $\widehat{\mathbf{w}}^{*\mathbf{T}} \Sigma \widehat{\mathbf{w}}^*$ to $\sigma^2_{un,s}$ vanishes with the rate $O_p\left(\sqrt{\log n/n}\right)$ (see section 1.8.6). If $\sigma^2_{1/N}$ were to decrease too fast with N, say $O_p\left((\log N/N)^{\frac{1}{2}+\varepsilon}\right)$ where $\varepsilon > 0$ is a constant[9], approximation error of $\widehat{\mathbf{w}}^{*\mathbf{T}} \Sigma \widehat{\mathbf{w}}^*$ to $\sigma^2_{un,s}$ would be larger than $\sigma^2_{1/N}$ itself. Therefore in this situation, the weighted norm mvp is unlikely to beat $1/N$. An immediate example for such situation can be given is when $\Sigma = \mathbf{I}_{NN}$. Suppose that the linear constraint is $\mathbf{w}^{\mathbf{T}}\mathbf{1} = 1$, then $\sigma^2_{1/N} = N^{-1} = \sigma^2_{un}$. It follows that

$$P\left(\widehat{\mathbf{w}}^{*\mathbf{T}} \Sigma \widehat{\mathbf{w}}^* \leq \sigma^2_{1/N}\right) = 0.$$

Another example is when Σ is a Toeplitz matrix. In this case, if $\sigma^2_i = 1$ and $\sigma^2_{ij} = c^{|i-j|}$, where $i \neq j$ and $0 < c < 1$, then

$$\sigma^2_{\frac{1}{N}} = \frac{1}{N^2}\left[2\left(\frac{(N-1)\,\rho - \frac{c^2 - c^{N+1}}{1-c}}{1-\rho}\right) + N\right] = O\left(N^{-1}\right),$$

which also violates the condition.

The number of assets selected is controlled by the penalty parameter of the l_1 norm penalty, which is specified as

$$\lambda_1 = \lambda_{n,N} = B_0 \sqrt{\frac{2\log N}{a^{\min} n}}.$$

The constant $a^{\min}$ is related to the elements in Σ (see proof of Lemma 3). Such specification satisfies the requirement that $|S|$ should increase with n, and more importantly, it provides a practical guideline to set the penalty parameters, as we will see in Section 1.6.2.

Finally, the property of causing sparsity makes the weighted norm strategy different from the ordinary shrinkage estimators. The ordinary shrinkage estimators shrink particular elements within the estimated covariance matrix to some targets, and in turn to reduce the estimation errors. The weighted norm strategy, however, not only shrinks particular elements within the estimated covariance matrix, but also reduces its dimension for the portfolio optimization. Consequently the resulting mvp is only constituted by a certain subset S of the whole assets. Thus it can be viewed that the ordinary shrinkage estimators aim to approximate σ^2_{un} directly, while the weighted norm strategy aims to approximate $\sigma^2_{un,s}$. As I show above, how such norm penalty strategy performs hinges on how far $\sigma^2_{un,s}$ deviates from σ^2_{un}.

[9]Note that here $\lim_{n\to\infty} N/n = \rho_N \in [0, 1]$ should hold.

1.4.3 Unconditional Portfolio Variance

If asset returns are iid, by using total variance formula, it can be shown that

$$\begin{aligned} var\left(R_{wp,n+1}\right) &= \mathbb{E}\left[\hat{\mathbf{w}}^{*\mathbf{T}}\Sigma\hat{\mathbf{w}}^{*}\right] + \mu^{\mathbf{T}} var\left(\hat{\mathbf{w}}^{*}\right)\mu, \\ var\left(R_{unp,n+1}\right) &= \mathbb{E}\left[\widehat{\mathbf{w}}_{un}^{\mathbf{T}}\Sigma\widehat{\mathbf{w}}_{un}\right] + \mu^{\mathbf{T}} var\left(\hat{\mathbf{w}}_{un}\right)\mu. \end{aligned}$$

By Theorem 1, under certain conditions, $P\left(\widehat{\mathbf{w}}^{*\mathbf{T}}\Sigma\widehat{\mathbf{w}}^{*} \leq \widehat{\mathbf{w}}_{un}^{\mathbf{T}}\Sigma\widehat{\mathbf{w}}_{un}\right) \to 1$ as N and n both go to large. It follows that

$$\mathbb{E}\left[\hat{\mathbf{w}}^{*\mathbf{T}}\Sigma\hat{\mathbf{w}}^{*}\right] \leq \mathbb{E}\left[\widehat{\mathbf{w}}_{un}^{\mathbf{T}}\Sigma\widehat{\mathbf{w}}_{un}\right].$$

as N and n both go to large. Thus if $\mu^{\mathbf{T}} var\left(\hat{\mathbf{w}}^{*}\right)\mu \leq \mu^{\mathbf{T}} var\left(\hat{\mathbf{w}}_{un}\right)\mu$, $var\left(R_{wp,n+1}\right) \leq var\left(R_{unp,n+1}\right)$ will also hold.

As for $1/N$, its unconditional variance also equals to $\sigma_{1/N}^2$, and therefore $var\left(R_{wp,n+1}\right) \leq \sigma_{1/N}^2$ if and only if

$$\mathbb{E}\left[\hat{\mathbf{w}}^{*\mathbf{T}}\Sigma\hat{\mathbf{w}}^{*}\right] + \mu^{\mathbf{T}} var\left(\hat{\mathbf{w}}^{*}\right)\mu \leq \sigma_{1/N}^2.$$

From Theorem 1, under certain conditions, $P\left(\hat{\mathbf{w}}^{*\mathbf{T}}\Sigma\hat{\mathbf{w}}^{*} \leq \sigma_{1/N}^2\right) \to 1$ as N and n both go to large. It implies

$$\mathbb{E}\left[\hat{\mathbf{w}}^{*\mathbf{T}}\Sigma\hat{\mathbf{w}}^{*}\right] \leq \sigma_{1/N}^2.$$

will hold as N and n both go to large. Thus to require $var\left(R_{wp,n+1}\right) \leq \sigma_{1/N}^2$, either $\mu^{\mathbf{T}} var\left(\hat{\mathbf{w}}^{*}\right)\mu$ should not be too large or $\mathbb{E}\left[\hat{\mathbf{w}}^{*\mathbf{T}}\Sigma\hat{\mathbf{w}}^{*}\right]$ should be small enough to $\sigma_{1/N}^2$.

$var\left(\hat{\mathbf{w}}^{*}\right)$ and $var\left(\hat{\mathbf{w}}_{un}^{*}\right)$ are covariance matrices of $\hat{\mathbf{w}}^{*}$ and $\hat{\mathbf{w}}_{un}^{*}$, respectively, and their analytical expressions are not derived here. To gain some meaningful insights, one can assume that covariances between elements in $\hat{\mathbf{w}}^{*}$ are small. Then $\mu^{\mathbf{T}} var\left(\hat{\mathbf{w}}^{*}\right)\mu$ will be mainly determined by

$$\sum_{i=1}^{N} \mu_i^2 var\left(\hat{w}_i^{*}\right).$$

By similar fashion, $\mu^{\mathbf{T}} var\left(\hat{\mathbf{w}}_{un}^{*}\right)\mu$ will be mainly determined by $\sum_{i=1}^{N} \mu_i^2 var\left(\hat{w}_{un,i}^{*}\right)$. $var\left(\hat{w}_i^{*}\right)$ and $var\left(\hat{w}_{un,i}^{*}\right)$ can be measured by their portfolio turnover rates, and we will see that empirically, the weighted norm mvp tends to have a lower turnover rate than the gmvp.

1.5 Coordinate Wise Descent Algorithm

In this section we introduce a coordinate-wise descent algorithm developed by Yen and Yen (2011), to solve the weighted norm mvp optimization problem. We focus on the benchmark case in which the objective function only subjects to the full investment constraint. How a coordinate-wise descent algorithm is implemented is as follows. Assume the objective function $f(\mathbf{w}) = f(w_1, \ldots, w_N)$ is convex. The algorithm starts by fixing w_i for $i = 2, \ldots, N$ and finding a value for w_1 to minimize $f(\mathbf{w})$. The iteration is then done over $i = 2, 3, \cdots, N$

before going back to start again for $i = 1$. The procedure is repeated until the value of $f(\mathbf{w})$ converges.

Friedman et al. (2007) demonstrates that coordinate-wise descent algorithms can be powerful tools in solving regression problems regularized by convex constraints. Since then, the approach has gradually become popular in statistics to solve various norm constrained regression problems. Theoretical properties of this type of algorithm can be found in Tseng (2001). Yen and Yen (2011) develops efficient coordinated-wise descent type algorithms for solving various norm constrained portfolio optimizations.

As for the case of when $\mathbf{w}^{\mathbf{T}}\mathbf{1}_N = 1$ is the only linear constraint, the proposed scheme to update w_i is

$$\frac{ST(\gamma - z_i, \lambda_1)}{2(\sigma_i^2 + \lambda_2)}, \tag{1.20}$$

where $ST(x, y) = sign(x)(|x| - y)_+$ is the soft thresholding function and $z_i = 2\sum_{j \neq i}^{N} w_j \sigma_{ij}$. Let $S_+ = \{i : w_i > 0\}$ and $S_- = \{i : w_i < 0\}$. The proposed scheme to update the Lagrangian multiplier γ is

$$\frac{1 + \sum_{i \in S_+ \cup S_-} \frac{z_i}{2(\sigma_i^2+\lambda_2)} - \lambda_1 \left(\sum_{i \in S_-} \frac{1}{2(\sigma_i^2+\lambda_2)} - \sum_{i \in S_+} \frac{1}{2(\sigma_i^2+\lambda_2)} \right)}{\left[\sum_{i \in S_+ \cup S_-} \frac{1}{2(\sigma_i^2+\lambda_2)} \right]}. \tag{1.21}$$

The derivations of (1.20) and (1.21) can be found in Appendix 1.8.7. The updating scheme is essentially the same as using the Gauss-Seidel method to solve a system of equations of the non-zero weights in the KKT conditions with γ fixed, and then using the solved weights and full investment constraint to update γ. Under the Gauss-Seidel method, if the iteration converges, the limit will be guaranteed to be the solution of the system (Saad, 2003). Furthermore, if the set of non-zero weights $S_+ \cup S_-$ is correctly identified, convergence of the algorithm will guarantee the minimum is attained (with γ fixed). Through the adjustment of γ, the full investment constraint can be further satisfied. Thus if $\mathbf{w}$ and γ both converge, we can obtain solution of (1.1).

The above approach of updating the weights and the Lagrangian multiplier can be generalized. Considering instead that $A\mathbf{w} = \mathbf{u}$ should be satisfied. The Lagrangian, as shown in Section 1.2.2 is the objective function plus $\gamma^{\mathbf{T}}(\mathbf{u} - A\mathbf{w})$, where $\gamma = (\gamma_1, \ldots, \gamma_k)$ is a $k \times 1$ vector of the Lagrange multipliers. One can follow procedures similar as above to derive the updated forms of w_i and γ, and then $\mathbf{w}$, γ_1, $\gamma_2, \ldots, \gamma_p$ can be updated sequentially. An example of such sequential update when an additional target return constraint is considered is shown in Yen and Yen (2011).

1.6 Empirical Results

1.6.1 Performance Measures

For estimating the covariance matrix, I adopt the expanding window scheme with initial window length $\tau_0 = 1.2N$. Suppose there are $\bar{T}$ period observations. Let $\widehat{\Sigma}_t$ denote the estimated covariance matrix for period $t = \tau_0, \ldots, \bar{T} - 1$. The length of testing period (from $\tau_0 + 1$ to $\bar{T}$) is $T = \bar{T} - 1 - \tau_0$. $\widehat{\Sigma}_t$ is calibrated into (1.1) at each period t, and the solved optimal weight is denoted by $\widehat{w}^*_{i,t}$, $i = 1, \ldots, N$. In the following, I use $\widehat{w}^*_{i,t}$ as an example to illustrate how to obtain the performance measures.

The out-of-sample portfolio return at period $t + 1$ is defined as

$$\widehat{R}_{oos,p,t+1} = \sum_{i=1}^{N} \widehat{w}^*_{i,t} R_{i,t+1}.$$

I then calculate turnover rate of the trading strategy. Suppose at the end of period $t - 1$, the investor has wealth Π_{t-1} that can be invested on the assets. Given the optimal weight $\widehat{w}_{i,t-1}$, at the end of period t, the holding value of asset i is $\Pi_{t-1}\widehat{w}^*_{i,t-1}(1 + R_{i,t})$, and the total wealth at period t is $\Pi_t = \Pi_{t-1}\left(1 + \widehat{R}_{oos,p,t}\right)$. Then given $\widehat{w}^*_{i,t}$, the amount of wealth to invest on asset i is $\Pi_t \widehat{w}_{i,t}$. I define the turnover rate of asset i between t to $t + 1$ as

$$TOR_{i,t+1} = \left| \widehat{w}^*_{i,t} - \widehat{w}^*_{i,t-1} \frac{(1 + R_{i,t})}{\left(1 + \widehat{R}_{oos,por,t}\right)} \right|, \tag{1.22}$$

which is the proportion of wealth at the end of period t needed to invest on asset i in order to satisfy the required amount $\Pi_t \widehat{w}_{i,t}$. I further define the portfolio turnover rate at period $t + 1$ as the sum of the turnover rate (1.22) over whole assets, i.e.

$$TOR_{p,t+1} = \sum_{i=1}^{N} TOR_{i,t+1}.$$

I then impose transaction fee ε, and the oos net portfolio return at period $t + 1$ is defined as

$$\widehat{R}^{net}_{oos,p,t+1} = (1 - \varepsilon TOR_{p,t+1}) \times \left(1 + \widehat{R}_{oos,p,t+1}\right) - 1.$$

The oos net portfolio return is then used to calculate its sample variance, the Sharpe ratio and certainty equivalent return [10].

To see how sparsity affects the portfolio performance, one can have a look of proportion of

[10] The certainty equivalent return used here is defined the same as in DeMiguel et al. (2009b) with the risk aversion parameter ψ.

active assets at period t,

$$PAC_t = \frac{\left|\widehat{S}_t^+ \cup \widehat{S}_t^-\right|}{N},$$

where $\widehat{S}_t^+ = \left\{i : \widehat{w}_{i,t}^* > 0\right\}$ and $\widehat{S}_t^- = \left\{i : \widehat{w}_{i,t}^* < 0\right\}$. One may also interest in whether the strategy concentrates too much on certain assets. To measure this, I calculate Herfindahl-Hirschman index for the portfolio weights at period t, which is defined as

$$HHI_t = \sum_{i=1}^{N} \frac{\left|\hat{w}_{i,t}^*\right|^2}{\left(\sum_{i=1}^{N}\left|\hat{w}_{i,t}^*\right|\right)^2} = \frac{\|\hat{\mathbf{w}}_t^*\|_2^2}{\|\hat{\mathbf{w}}_t^*\|_1^2}.$$

The norm penalty results in a sparse solution, and portfolio weights will concentrate on the active assets. To know whether some of weights of the active assets may be too extreme or not, measuring concentration among these active assets is more informative. For this purpose, I also calculate the adjusted normalized Herfindahl-Hirschman Index at period t,

$$ANHHI_t = \frac{HHI_t - \frac{1}{|\widehat{S}_t|}}{1 - \frac{1}{|\widehat{S}_t|}},$$

where $\widehat{S}_t = \left\{i : \widehat{w}_{i,t}^* \neq 0\right\}$[11]. Finally, define the shortsales-long ratio at period t as

$$SLR_t = \frac{\sum_{i \in \widehat{S}_t^-}\left|\hat{w}_{i,t}^*\right|}{\sum_{i \in \widehat{S}_t^+}\left|\hat{w}_{i,t}^*\right|}.$$

The ratio is helpful for clarifying how the norm penalty and the linear constraints have impacts on the overall short and long positions of the portfolio.

1.6.2 Setting the penalty parameters

Theorem 1 suggests that one can set the penalty parameters as a function of sample size and the number of assets. Let $\lambda_{1,t}$ and $\lambda_{2,t}$ denote penalty parameters for period t. Here, I propose the following settings

$$\begin{aligned}\lambda_{1,t} &= \alpha \hat{a}_t \hat{B}_t \sqrt{\frac{2\log N}{n_t}},\\ \lambda_{2,t} &= (1-\alpha)\,\hat{a}_t \widehat{B}_t \sqrt{\frac{2\log N}{n_t}},\end{aligned}$$

[11]Note that for $ANHHI_t$, I use $1/\left|\widehat{S}_t\right|$ rather than $1/N$ as the normalizing constant. The reason is that unlike HHI_t, which measures concentration of weights among all N assets, the $ANHHI_t$ measures concentration of weights among the active assets.

where $\hat{a}_t = \min_{i=1,\dots,N} \hat{\sigma}^2_{i,t}$, and $\hat{B}_t = \left\|\hat{\mathbf{w}}^*_{t-1}\right\|_1$. The parameter n_t is the sample size used at period t, which will increase with time under the expanding window scheme. The specification shown here keeps the form $\sqrt{2\log N/n_t}$, but also uses some scale factors, $\hat{a}_t$, $\hat{B}_t$ and α. $\hat{a}_t$ is the minimum diagonal term of the estimated covariance matrix (the minimum variance among all the N assets). The reason for such setting is that, as shown in the proof of Lemma 3, a_{ij} is inversely related to σ^2_{ij} or σ^4_i. Therefore $a^{\min}$ should also inversely relate to σ^2_{ij} or σ^4_i. I find approximate $a^{\min}$ by $\hat{a}_t$ works pretty well in practice. In addition, adding $\hat{a}_t$ also makes the penalty parameters are free to changes in units. The choice for $\hat{B}_t$ here is trying to satisfy condition 4 in Section 1.4.2, and I also find this choice works well. The parameter $\alpha \in [0,1]$ imposed here is used to adjust relative importance between the l_1 and squared l_2 norms. This can help us to see how changing the relative weight imposed on the two penalties affects the portfolio performances.

1.6.3 Main Results

I first compare performances of different portfolio strategies in which the full investment constraint $\mathbf{w}^{\mathbf{T}}\mathbf{1} = 1$ is imposed. The benchmark strategies I consider are the no-shortsales mvp, $1/N$ and the gmvp. The data used are Fama-French 100 size and book-to-market ratio portfolios (FF100) and 300 stocks randomly chosen from CRSP data bank (CRSP300)[12]. Figure 3.1 shows annualized mean returns and standard deviations of individual assets in FF100 and CRSP300. The mean returns and standard deviations are calculated with daily data over the whole sample period[13].

Table 3.6 and 3.6 show the results as daily FF100 and CRSP300 data are used, and the portfolios are balanced at daily basis. Testing period for FF100 is from Jan-02-1990 to Dec-31-2010 and for CRSP300 is from Jan-03-2000 to Dec-31-2010. The tables include annualized sample variance of the oos net returns (SV), the annualized Sharpe ratio (SR)[14], and average values of the other six measures: certainty equivalent return (CE), turnover rate (TOR), proportion of active assets (PAC), Herfindahl-Hirschman Index (HHI), adjusted normalized Herfindahl-Hirschman Index (ANHHI), and shortsales-long ratio (SLR). For each table, in the parentheses are the bootstrap standard errors of the corresponding quantities, obtained from using stationary bootstrap of Politis and Romano (1994). The net portfolio return is obtained as transaction fee $\varepsilon = 35$ basis points is deducted.

[12] The data of FF100 can be downloaded from Professor Kenneth French's website: `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/datalibrary.html`. I use average value weighted returns of the 100 portfolios. The return data of CRSP300 is available from the author.

[13] The sample period for FF100 is from July-12-1987 to Dec-31-2010 (5,415 observations), and for CRSP300 is from July-30-1998 to Dec-31-2010 (3,127 observations)

[14] Note that given SV and SR, and the annualized risk free rate, it is not difficult to calculate the annualized average (oos net) portfolio return. For example, as $\alpha = 0$, for FF100, the annualized average (oos net) portfolio return (%) is

$$1.0666 \times \sqrt{82.9718} + 3.63 \approx 13.3455,$$

and as $\alpha = 0.2$, the value is 14.2830%.

For daily FF100, it can be seen that the weighted norm strategy can deliver a lower SV than no-shortsales mvp (N.S.) and naive $1/N$ strategy. It is somehow surprising that$1/N$ has the highest volatility in this case. The result is different from DeMiguel et al. (2009b), in which the authors show how $1/N$ dominates other sophisticated portfolio strategies in many different performance measures. $1/N$ is often considered as a layman strategy: one just assigns $1/N$ to each asset without doing any adjustment according to the information one has. But its extremely high volatility implies that it in fact heavily depends on market timing, and one needs a professional skill to determine whether it is a good time or bad time to implement such strategy. As for the annualized Sharpe ratio, the weighted norm performs pretty well and completely dominates the three benchmarks strategies in this case. The annualized SR are all above 1 and varies from 1.07 to 1.19 as α varies from 0 to 1. The weighted norm mvp also dominates the benchmarks in the average certainty equivalent return (CE). Finally, one can see the proportion of active assets is inversely related to the annualized portfolio variance. It suggests that increasing sparsity seems not work for reducing portfolio volatility in this case.

As for the case of daily CRSP300, the weighted norm mvp can yield a lower SV than the other three benchmarks for a range of α. The lowest SV occurs when $\alpha = 0.2$, and on average only 76% assets are selected. It suggests that optimally increasing sparsity may work for reducing portfolio volatility in this case. However, now the weighted norm mvp does not enjoy the highest SR and CE. The highest SR and CE is achieved by the no-shortsales mvp, a special case of the mvp penalized by the l_1 norm penalty.

The average turnover rates of $1/N$ is obviously lower than that of other strategies, and the fact was widely documented in previous literatures. As for the weighted norm mvp, it can be seen that the average turnover rate declines with α, which suggests that imposing more l_1 penalty helps to stablize the weight vector. For the average proportion of active assets, the weighted norm mvp with $\alpha > 0$ and no-shortsales mvp constantly have less than N assets included. The PAC also decreases with α, since increasing α is equivalent to imposing more l_1 penalty, and it facilitates more sparsity in the portfolio weight vector. The no-shortsales mvp on average has the sparsest optimal weight vector. It is expected, since the no sortsales-mvp is essentially the same as the weighted norm mvp with a heavy l_1 penalty.

The $1/N$ strategy assigns equal weights on each asset, so its HHI and ANHHI have the lowest values ($1/N$ and 0) among all strategies. In general, the portfolios with sparse weight vectors will assign relatively more loadings on a few certain assets, and consequently problem of extreme weights will arise. This phenomenon can been seen from their HHI and ANHHI. For the weighted norm mvp, the HHI and ANHHI increase with α and the average proportion of active assets. It is not surprising that the no-shortsales mvp, which has the lowest average proportion of active assets, has the highest values of HHI and ANHHI. One thing worth to note here is that the weighted norm mvp with $\alpha = 0$, which only l_2 norm penalty is activated, has lower values of HHI and ANHHI than the gmvp. It suggests that putting l_2 norm indeed helps to alleviate the problem of extreme weights. Finally, the shortsales-long ratio SLR is positively related to α, which confirms that the l_1 norm penalty, together with the full investment constraint, facilitate long positions in the weighted norm portfolio.

1.6.4 Stochastic Dominance Test

The performance measures shown in Table 3.6 and 3.6, such as the Sharpe ratio or certainty equivalent return, does not take into a general framework of utility maximization into account. If an individual endowed with an arbitrary (nondecreasing) utility function, should she prefer the weighted norm strategy to other benchmarks? The concepts of first-order (FSD) and second-order (SSD) stochastic dominance can help us to answer the question. Given returns of two portfolio strategies 1 and 2, say R_1 and R_2, if R_1 first order stochastically dominates R_2, it is equivalent to saying that every expected utility maximizer will prefer strategy 1 to strategy 2. On the other hand, strategy 1 clearly deliver higher expected utility than strategy 2. If R_1 second order stochastically dominates R_2, it is equivalent to saying that every risk-averse expected utility maximizer will prefer strategy 1 to strategy 2. Or one can say strategy 1 is less risky than strategy 2.

In practice, to see whether R_1 first or second order stochastic dominates R_2, one can implement some formal statistical tests via comparing functionals of their cumulative distribution functions. Now let strategy 1 as the weighted norm mvp and strategy 2 as the other benchmarks. The null hypothesis used here for the test is

$$H_0 : \text{The weighted norm mvp FSD (SSD) the benchmarks.}$$

If we cannot reject the null, there is not enough evidence to say that the weighted norm mvp FSD (SSD) the benchmarks does not hold. To empirically construct critical values and p-values, I adopt subsampling method suggested by Linton et al. (2005) [15].

Table 3.6 and 3.6 show subsample p-values of the stochastic dominance tests for the cases of daily FF100 and CRSP300[16]. For the case of daily FF100, there is a strong evidence against the hypotheses that the weighted norm mvp FSD naive $1/N$ and the no-shortsales mvp, but in favour of the one that the weighted norm mvp FSD the gmvp. These results are inconsistent with those presented in Table 3.6, in which $1/N$ and no-shortsales mvp both have high volatility and low average annualized portfolio returns[17]. As for CRSP300, except the $1/N$, there is some evidence supporting that the weighted norm mvp FSD the other two benchmarks. Overall, the FSD test suggests that it is still hard to say whether the weighted norm mvp is preferred to the other three benchmarks by an individual endowed with an arbitrary risk preference.

Unlike the FSD test, the results for the SSD test are more consistent. For the two data sets,

[15] More detail discussions on the FSD, SSD and how the tests are implemented can be found in appendix 1.8.8.

[16] The data used here are daily net returns when transaction fee of 35 basis points is deducted.

[17] The annualized average (oos net) portfolio returns for $1/N$ and no-shortsales mvp are about 11.5416% and 13.4645%, respectively, while the lowest annualized average (oos net) portfolio returns for the weighted norm mvp are around 13.3457% ($\alpha = 0$) and 14.2830% ($\alpha = 0.2$).

the hypothesis that the weighted norm mvp second order stochastic dominates the benchmarks cannot be rejected at the significant level 0.05. It suggests that an individual with risk averse preference is more likely to choose the weighted norm strategy than the other benchmark ones as she is making decisions on allocating assets.

1.6.5 Lower Frequency Data

Table 1.5 to 1.7 show the results of weekly FF100 and CRSP300 and monthly FF100 data are used[18]. Since weekly and monthly data are used for estimating the covariance matrix, accordingly, the portfolio is balanced weekly or monthly here. The results are qualitatively similar as those when daily data are used. In the case of FF100, the weighted norm mvp consistently yields higher Sharp ratio than the other three benchmarks, but for the case of CRSP 300, it fails to achieve a higher Sharpe ratio than the no-shortsales mvp. Comparing with the daily results, the annualized portfolio variances and Sharpe ratio are worse in the lower frequency cases. One also can see that while the gmvp can yield the lowest annualized portfolio variance in the case of daily and weekly FF100, it loses such merit in the case of monthly FF100. This may be because a far fewer number of average sample size is used in the monthly case than the daily and weekly cases, and therefore bias the estimates.

1.6.6 With the Target Return Constraint

Previous literatures argue that empirically, adding the estimated return vector into the portfolio optimization often damages portfolio performances (Jagannathan and Ma, 2003; DeMiguel et al., 2009a)). Indeed, accurately estimating the means of returns is more difficult than accurately estimating the corresponding variances or covariances. This is the main reason why people often focus on the portfolio optimization without imposing the target return constraint $\mathbf{w}^{\mathbf{T}}\mu = \bar{\mu}$, in which the estimated mean return vector is needed. One may wonder whether imposing the weighted norm penalty can help to boost portfolio performances in this situation. Table 1.8 shows results when the additional target return constraint is imposed. The calibrated mean return vector μ is again estimated via the sample mean estimation with the expanding window scheme. The target return $\bar{\mu}$ (annualized) is set at high and low levels for each case. I keep the penalty parameter settings as I use in the full investment constraint case.

Overall, the annualized portfolio variance, Sharpe ratio and average certainty equivalent return shown here are still not as good as the case without such additional target return constraint. One exception is the case of CRSP300 with $\bar{\mu} = 5\%$, in which the SR and CE are higher than the case with the target return constraint. It also can be seen that the value of $\bar{\mu}$

[18] The testing periods are the same as the daily cases, but the sample periods are now different. For the weekly return data, I construct them by aggregating the daily returns over a week. The sample period for the weekly FF100 is from the 38th week of 1987 to the last week of 2010 (1,215 observations); for CRSP300 is from the sixth week of 1993 to the last week of 2010 (933 observations). Again, the weekly CRSP300 return data is available from the author. The sample period of monthly FF100 is from Jan-1980 to Dec-2010 (372 observations), and it can be downloaded from Professor Kenneth French's website.

can have significant impacts on the performance measures. Also, higher $\bar{\mu}$ does not necessarily result in higher SR and CE. For example, in CRSP300, the weighted norm mvp can yield the Sharpe ratio around 0.59 to 1 when the annualized required return is 5%. But as the required return is up to 10%, the resulting Sharpe ratio is only around 0.40 to 0.74. For the case of FF100, the situation becomes opposite. The weighted norm mvp can yield Sharpe ratio from 1.02 to 1.06 as the required return is 20%, while the required return is down to 10%, the resulting Sharpe ratio is only around 0.59 to 0.64. In sum, the results shown here are in line with what previous literatures find: adding the estimated mean returns into the portfolio optimization does little or no help on improving the portfolio performances. However, one should note that the penalty parameters used here do not incorporate any information about the estimated means, and this perhaps is another reason for the inferior performances of the weighted norm mvp with the target return constraint.

1.6.7 Alternative Norm Penalties

I then investigate whether imposing different forms of norm penalties on the portfolio optimization can obtain better portfolio performances than the weighted norm penalty does. As mentioned in Section 1.3.5, the penalized portfolio optimization can be viewed as a constrained regression problem. I therefore borrow some ideas from statistics. The first alternative penalty function I consider is the berhu penalty (Owen, 2007),

$$\lambda \sum_{i=1}^{N} \left(|w_i| \, \mathbb{I}\{|w_i| < \kappa\} + \frac{w_i^2 + \kappa^2}{2\kappa} \mathbb{I}\{|w_i| \geq \kappa\} \right), \tag{1.23}$$

where $\mathbb{I}\{.\}$ is the indicator function. The name berhu comes from the fact that (1.23) is the reverse of Huber's loss. The berhu penalty is convex and satisfies the additive separability condition. Yen and Yen (2011) propose an efficient coordinate-wise descent algorithm for solving the portfolio optimization with the berhu penalty. The penalty can be seem as a compromise between the l_1 and squared l_2 norm regularizations: if $|w_i|$ is less than some positive constant κ, then it will be regularized by the l_1 norm; if $|w_i|$ is larger than or equal to κ, then it will be regularized by the squared l_2 norm.

I also consider the following generalized l_1 norm penalty[19],

$$\lambda_1 \|\mathbf{w}\|_1 + \lambda_2 \|\mathbf{w} - \mathbf{w}_0\|_1 .$$

This is a more general version of the norm penalty considered in DeMiguel et al. (2010) in which only the part of $\|\mathbf{w} - \mathbf{w}_0\|_1$ is active ($\lambda_1 = 0$). The additional $\|\mathbf{w}\|_1$ imposed here is functioning as inducing more sparsity. We have seen the penalty $\|\mathbf{w} - \mathbf{w}_0\|_1$ in Section 1.3.3, where I relate it with individual's inattention or the endowment effect. The vector $\mathbf{w}_0$ in the

[19] The derivation of the coordinate-wise descent algorithm for solving the mvp penalized the generalized l_1 norm penalty can be found in Appendix 1.8.9

second l_1 norm penalty is called the target weight vector. DeMiguel et al. (2010) suggest to use the optimal gmvp weight vector as the target weight vector, and shows that such choice can help to reduce portfolio turnover rate. Unlike them, here I use $1/N$ or the optimal no-shortsales weight vector as the the target weight vector. Choosing $1/N$ is because such setting in general results in the lowest turnover rate among the benchmarks, while using the optimal no-shortsales weight is because its dominant performances in the CRSP300 data.

I also try to assign different penalty values to different assets. To do this, I consider a multiple-stage portfolio optimization with the following penalty

$$\lambda \sum_{i=1}^{N} \frac{\epsilon}{\exp\left(\epsilon \left|w_i^{*(l)}\right|\right)} |w_i| , \tag{1.24}$$

where $w_i^{*(l)}$, $l = 0, 1, \ldots,$ is the portfolio weight for asset i obtained from some portfolio optimizations, and $\epsilon > 0$ is a turning parameter. I call (1.24) the adaptive penalty. Note that all of the norm penalties I use so far treat each asset equally: the penalty parameter for each of them has the same value. The above modified l_1 norm penalty is trying to impose different degrees of penalty on the assets by adding the term $\epsilon \exp\left(-\epsilon \left|w_i^{*(l)}\right|\right)$. To obtain $w_i^{*(l)}$, I propose to use a multi-stage approach, which can be cast as follows.

- Step 1: Solve the l_1 norm penalized portfolio optimization. Denote the solution $w_i^{*(0)}$, $i = 1, \ldots, N$.
- Step 2: Plug $w_i^{*(0)}$ into (1.24). Then solve the portfolio optimization with penalty (1.24), and denote the solution $w_i^{*(1)}$, $i = 1, \ldots, N$.
- Step 3: Plug $w_i^{*(1)}$ into (1.24). Then solve the portfolio optimization with penalty (1.24), and the solution, denoted by $w_i^{*(2)}$, $i = 1, \ldots, N$ is used as the portfolio weight for asset i.

In fact, one can continue to step 4, 5,... and obtain $w_i^{*(3)}$, $w_i^{*(4)}, \ldots,$ and terminate until certain convergence condition is satisfied. However, I find in practice proceed to step 3 is enough to ensure good convergence. Imposing (1.24) can be viewed as using the majorization-minimization method to approximately solve an l_0 norm penalized portfolio optimization, and a more detail discussion on this can be found in Appendix 1.8.9.

For each penalty, I uniformly set the penalty parameters as $\hat{a}_t \hat{B}_t \sqrt{2 \log N / n_t}$. Table 1.9 and 1.10 show the results for the cases of daily FF100 and CRSP300, respectively. For FF100, using berhu penalty can yield the Sharpe ratio from 0.83 to 1.11, average certainty equivalent return from 0.04 to 0.05, as one vary the parameter κ from 0.02 to 0.1. For CRSP300, the SR is from 0.99 to 1.08, and the CE is from 0.033 to 0.037, as κ is varied within the same range. Both the SR and CE of the berhu penalty for the CRSP300 are better than those of the weighted norm mvp, but still slightly worse than those of the no-shortsales mvp. As for the generalized l_1 norm penalty, I use $TWN - l_1$ ($TWNS - l_1$) and TWN ($TWNS$) to denote the cases with and without the additional l_1 norm penalty when the target weight vector $\mathbf{w}_0$ is $1/N$ (or the optimal no-shortsales weight vector). It can be seen that using the optimal no-shortsales weight vector as $\mathbf{w}_0$ yields higher SR and CE than using the $1/N$ weight does, no matter whether the additional l_1 norm penalty is considered. Imposing the additional l_1 penalty works well for the CRSP300, but not for the FF100. As for the adaptive penalty, it is a little surprising that it can obtain higher Sharpe ratio than the weighted norm mvp for FF100 and CRSP300, as one carefully set the parameter ϵ. The number of iteration l seems not have effect on the performances. The SR for FF100 can achieve 1.21 as $\epsilon = 1$. For the CRSP300, setting $\epsilon = 1$ and 2.5 yields SR up to 0.96 and 0.99, and the number is similar to the case of the weighted norm mvp. Overall, the three alternative penalties used here have chances to obtain at least comparable performances as the weighted norm penalty and the benchmarks.

1.7 Conclusion

This paper provides ample evidence, both theoretically and empirically, why imposing a l_1 and squared l_2 norm penalties on the portfolio optimization can improve portfolio performance when the number of available assets N is large. The paper also links the weighted norm strategy to some interesting issues in finance, economics and statistics, and provides alternative explanations on why using the weighted norm strategy is reasonable. An efficient algorithm for solving the problem and optimal penalty parameter setting is introduced and tested and applied to financial data.

The paper remains agnostic on several issues but they provide exiting avenues for future research. For example, in the current version, I only use the sample covariance matrix and sample mean return vector with the expanding window scheme, to solve for the optimal portfolio. More sophisticated estimations, such as using high frequency data or factor models, might provide better results. Another important issue is on how the theoretical properties of the weighted norm portfolio changes when 1) one allows random components in the linear constraints, 2) the data generating process becomes more complex and 3) more sophisticated estimations are used. El Karoui (2009) and El Karoui (2010) provide some results on the first two issues for the global minimum variance portfolio. The above three modifications will make the statistical properties of the high dimensional covariance estimation far more complicated than in this paper, and to the best of my knowledge, until now there is very little literature focusing on these issues.

1.8 Appendix

1.8.1 Proof of Lemma 1

Proof. The optimal nonzero weights $\mathbf{w}_s^*$ and Lagrange multipliers can be solved via the following system of equations

$$\begin{pmatrix} 2\left(\Sigma_{ss}+\lambda_2\mathbf{I}_{ss}\right) & A_s^{\mathbf{T}} \\ A_s & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{w}_s^* \\ -\gamma^* \end{pmatrix} = \begin{pmatrix} -\lambda_1 \times sign\left(\mathbf{w}_s^*\right) \\ \mathbf{u} \end{pmatrix}.$$

By using matrix inverse formula, we can obtain expressions of $\mathbf{w}_s^*$ and γ^* as shown in Lemma 1. ■

1.8.2 Proof of Lemma 2

Proof. Multiplying both sides of (1.3) by $\mathbf{w}^{*\mathbf{T}}$ yields

$$2\sigma_s^2 + 2\lambda_1 \left\|\mathbf{w}_s^*\right\|_2^2 = \mathbf{w}_s^{*\mathbf{T}} A_s^{\mathbf{T}}\gamma^* - \lambda_1 \mathbf{w}_s^{*\mathbf{T}} sign\left(\mathbf{w}_s^*\right).$$

By (1.4) and $\mathbf{w}^{*\mathbf{T}} sign\left(\mathbf{w}^*\right) = \left\|\mathbf{w}^*\right\|_1$, It follows that

$$2\sigma_s^2 + 2\lambda_2 \left\|\mathbf{w}_s^*\right\|_2^2 = \mathbf{u}^T\gamma^* - \lambda_1 \left\|\mathbf{w}_s^*\right\|_1.$$

From Lemma 1, $\mathbf{u}^{\mathbf{T}}\gamma^* = \mathbf{u}^{\mathbf{T}}\gamma_{2,s}^* + \lambda_1\mathbf{u}^{\mathbf{T}}\delta_{2,s}$. It can be shown that

$$\mathbf{u}^{\mathbf{T}}\gamma_{2,s}^* = 2\left(\sigma_{2,s}^2 + \lambda_2 \left\|\mathbf{w}_{2,s}^*\right\|_2^2\right).$$

Therefore

$$\sigma_s^2 - \sigma_{un,s}^2 = \sigma_{2,s}^2 - \sigma_{un,s}^2 + \lambda_2\left(\left\|\mathbf{w}_{2,s}^*\right\|_2^2 - \left\|\mathbf{w}_s^*\right\|_2^2\right) + \frac{\lambda_1}{2}\left(\mathbf{u}^{\mathbf{T}}\delta_{2,s} - \left\|\mathbf{w}_s^*\right\|_1\right). \tag{1.25}$$

Also

$$\mathbf{u}^{\mathbf{T}}\delta_{2,s} = \mathbf{w}_{2,s}^{*\mathbf{T}} sign\left(\mathbf{w}_s^*\right) \le \sum_{i\in S}\left|w_{2,s,i}^* sign\left(w_{s,i}^*\right)\right| \le \left\|\mathbf{w}_{2,s}^*\right\|_1.$$

I then prove that

$$\sigma_{2,s}^2 - \sigma_{un,s}^2 + \lambda_2\left\|\mathbf{w}_{2,s}^*\right\|_2^2 \le c_{s,1}\left(c_{s,1}+1\right)\sigma_{un,s}^2.$$

To see this, since $\Sigma'_{ss} = \Sigma_{ss} + \lambda_2\mathbf{I}_{ss}$, it can be shown that

$$A_s\Sigma_{ss}^{-1}A_s^{\mathbf{T}} - A_s\Sigma_{ss}'^{-1}A_s^{\mathbf{T}} = \lambda_2 A_s\Sigma_{ss}'^{-1}\Sigma_{ss}^{-1}A_s^{\mathbf{T}}. \tag{1.26}$$

Furthermore,

$$\sigma_{2,s}^2 - \sigma_{un,s}^2 + \lambda_2\left\|\mathbf{w}_{2,s}^*\right\|_2^2 = \lambda_2\left[\mathbf{u}^{\mathbf{T}}M_s'^{-1}\left(A_s\Sigma_{ss}'^{-1}\Sigma_{ss}^{-1}A_s^{\mathbf{T}}\right)M_s^{-1}\mathbf{u}\right].$$

Note that if a $|S|\times|S|$ matrix M_s is positive semidefinite, one can have the following Cauchy-Schwartz type inequality

$$\left|\mathbf{x}^{\mathbf{T}}M_s\mathbf{y}\right| \le \sqrt{\mathbf{x}^{\mathbf{T}}M_s\mathbf{x}}\sqrt{\mathbf{y}^{\mathbf{T}}M_s\mathbf{y}},$$

where $\mathbf{x}$ and $\mathbf{y}$ are $|S| \times 1$ column vectors. Note that since $\Sigma_{ss}'^{-1}$ and Σ_{ss}^{-1} both positive semidefinite, given $|S| >> k$, the matrix $A_s\Sigma_{ss}'^{-1}\Sigma_{ss}^{-1}A_s^{\mathbf{T}}$ is also positive semidefinite. It follows

$$\mathbf{u}^{\mathbf{T}}M_s'^{-1}\left(A_s\Sigma_{ss}'^{-1}\Sigma_{ss}^{-1}A_s^{\mathbf{T}}\right)M_s^{-1}\mathbf{u} \leq \sqrt{\mathbf{u}^{\mathbf{T}}M_s'^{-1}\left(A_s\Sigma_{ss}'^{-1}\Sigma_{ss}^{-1}A_s^{\mathbf{T}}\right)M_s'^{-1}\mathbf{u}} \tag{1.27}$$

$$\times\sqrt{\mathbf{u}^{\mathbf{T}}M_s^{-1}\left(A_s\Sigma_{ss}'^{-1}\Sigma_{ss}^{-1}A_s^{\mathbf{T}}\right)M_s^{-1}\mathbf{u}} \tag{1.28}$$

In the following I derive the upper bound of (1.27) and (1.28). For (1.27), by $\phi^{\min}(\Sigma_{ss}) > 0$,

$$\mathbf{u}^{\mathbf{T}}M_s'^{-1}\left(A_s\Sigma_{ss}'^{-1}\Sigma_{ss}^{-1}A_s^{\mathbf{T}}\right)M_s'^{-1}\mathbf{u} \leq \frac{1}{\phi^{\min}(\Sigma_{ss})}\sum_{j=1}^{|S|}\frac{1}{(\phi_j(\Sigma_{ss})+\lambda_2)}\left\|\mathbf{x}_s\mathbf{q}_{j,s}\right\|_2^2$$

where $\mathbf{x}_s = \mathbf{u}^{\mathbf{T}}M_s'^{-1}A_s$. The vector $\mathbf{q}_{j,s}$ is the jth column of a square matirx $\mathbf{Q}_s$ such that $\Sigma_{ss}'^{-1}\Sigma_{ss}^{-1} = \mathbf{Q}_s\Lambda_s\mathbf{Q}_s^{\mathbf{T}}$, $\mathbf{Q}_s\mathbf{Q}_s^{\mathbf{T}} = \mathbf{I}_{ss}$ and Λ_s is a diagonal matrix which jth diagonal element is $\left(\phi_j(\Sigma_{ss})\left(\phi_j(\Sigma_{ss})+\lambda_2\right)\right)^{-1}$. Then

$$\sum_{i=1}^{s}\frac{1}{(\phi_j(\Sigma_{ss})+\lambda_2)}\left\|\mathbf{x}_s\mathbf{q}_{i,s}\right\|_2^2 = \sigma_{2,s}^2 + \lambda_2\left\|\mathbf{w}_{2,s}^*\right\|_2^2.$$

Combining the above results, (1.27) is bounded by $\left(\phi^{\min}(\Sigma_{ss})\right)^{-\frac{1}{2}}\left(\sigma_{2,s}^2+\lambda_2\left\|\mathbf{w}_{2,s}^*\right\|_2^2\right)^{\frac{1}{2}}$. For (1.28), since the matrix $A_s\Sigma_{ss}'^{-1}\Sigma_{ss}^{-1}A_s^{\mathbf{T}}$ is positive semidefinite, by (1.26),

$$\mathbf{x}^{\mathbf{T}}M_s\mathbf{x} - \mathbf{x}^{\mathbf{T}}M_s'\mathbf{x} = \lambda_2\mathbf{x}^{\mathbf{T}}A_s\Sigma_{ss}'^{-1}\Sigma_{ss}^{-1}A_s^{\mathbf{T}}\mathbf{x} \geq 0,$$

where $\mathbf{x}$ is a $k \times 1$ column vector. By $\mathbf{x}^{\mathbf{T}}M_s'\mathbf{x} \geq 0$,

$$\mathbf{x}^{\mathbf{T}}M_s\mathbf{x} \geq \lambda_2\mathbf{x}^{\mathbf{T}}A_s\Sigma_{ss}'^{-1}\Sigma_{ss}^{-1}A_s^{\mathbf{T}}\mathbf{x}$$

for any $k \times 1$ column vector $\mathbf{x}$. Therefore

$$\mathbf{u}^{\mathbf{T}}M_s^{-1}\left(A_s\Sigma_{ss}'^{-1}\Sigma_{ss}^{-1}A_s^{\mathbf{T}}\right)M_s^{-1}\mathbf{u} \leq \frac{\mathbf{u}^{\mathbf{T}}M_s^{-1}M_sM_s^{-1}\mathbf{u}}{\lambda_2} = \frac{\sigma_{un,s}^2}{\lambda_2}.$$

The second term (1.28) is bounded by $\lambda_2^{-\frac{1}{2}}\sigma_{un,s}$. Note that $\sigma_{2,s}^2 \geq \sigma_{un,s}^2$, since the former is obtained by restricting the portfolio weights with the l_2 norm, while the later is obtained without imposing any constraint on the portoflio weights. Therefore

$$\sigma_{2,s}^2 - \sigma_{un,s}^2 + \lambda_2\left\|\mathbf{w}_{2,s}^*\right\|_2^2 \leq c_{s,1}\sqrt{\left(\sigma_{2,s}^2+\lambda_2\left\|\mathbf{w}_{2,s}^*\right\|_2^2\right)\sigma_{un,s}^2},$$

where $c_{s,1} = \lambda_2^{\frac{1}{2}}\left(\phi^{\min}(\Sigma_{ss})\right)^{-\frac{1}{2}}$. Let $a = \sigma_{2,s}^2 + \lambda_2\left\|\mathbf{w}_{2,s}^*\right\|_2^2$, $b = \sigma_{un,s}^2$. The above result shows that $a - b \leq c_{s,1}\sqrt{ab}$. By a and b are both nonnegative, Then

$$\sqrt{a} - \sqrt{b} \leq \frac{c_{s,1}\sqrt{ab}}{\sqrt{a}+\sqrt{b}} \leq \frac{c_{s,1}\sqrt{ab}}{\sqrt{a}} = c_{s,1}\sqrt{b}.$$

Therefore $\sqrt{ab} \leq (c_{s,1}+1)\,b$, and

$$\sigma_{2,s}^2 - \sigma_{un,s}^2 + \lambda_2\left\|\mathbf{w}_{2,s}^*\right\|_2^2 \leq c_{s,1}(c_{s,1}+1)\,\sigma_{un,s}^2.$$

Combining the above results, the proof is completed. ■

1.8.3 Normality of Asset Returns

Suppose that $\mathbf{R}_t \overset{iid}{\sim} \mathcal{N}(\mu, \Sigma)$, $t = 1, \ldots, n$, where $\mu = (\mu_1, \ldots, \mu_N)$ is the mean return vector and Σ is the covariance matrix, and the calibrated $\widehat{\Sigma}$ is the sample covariance matrix. Let $\mathbf{1}_n$ denote an $n \times 1$ column vector in which all components are 1. It can be shown that

$$\widehat{\Sigma} \;=\; \frac{1}{n-1}\Sigma^{\frac{1}{2}}\mathbf{Z}^{\mathbf{T}}\mathbf{H}_n\mathbf{Z}\Sigma^{\frac{1}{2}},$$

where $\overline{\mathbf{R}} = \frac{1}{n}\sum_{t=1}^{n}\mathbf{R}_t$ is the sample mean return, $\mathbf{R} = (\mathbf{R}_1, \ldots, \mathbf{R}_n)^{\mathbf{T}}$ is an $n \times N$ return matrix, $\mathbf{H}_n = \mathbf{I}_{nn} - n^{-1}\mathbf{1}_n\mathbf{1}_n^T$ is an idempotent matrix, and $\mathbf{HR} = \mathbf{R} - \overline{\mathbf{R}}$. The vector $\mathbf{Z} = (\mathbf{Z}_1, \ldots, \mathbf{Z}_n)^{\mathbf{T}}$ is an $n \times N$ matrix for standard normal random vector,

$$\mathbf{Z}_t \overset{iid}{\sim} \mathcal{N}(\mathbf{0},\ \mathbf{I}_{NN}),$$

$i = 1, \ldots, n$. Obviously, $\mathbf{R}_t = \mu_t + \mathbf{Z}_t\Sigma^{\frac{1}{2}}$.

With iid normal distributed returns and the sample covariance estimator, it is well known that

$$\widehat{\Sigma} \sim \frac{\mathcal{W}(\Sigma, N, n-1)}{n-1}, \tag{1.29}$$

where $\mathcal{W}(\Sigma, N, n-1)$ is the Wishart distribution with parameter Σ, N, and $n-1$. If A is a $k \times N$ deterministic matrix, then

$$A\widehat{\Sigma}A^{\mathbf{T}} \sim \frac{\mathcal{W}\left(A\Sigma A^{\mathbf{T}}, k, n-1\right)}{n-1}. \tag{1.30}$$

If $\widehat{\Sigma}$ is positive definite with probability one, then

$$\left(A\widehat{\Sigma}^{-1}A^{\mathbf{T}}\right)^{-1} \sim \frac{\mathcal{W}\left(\left(A\Sigma^{-1}A^{\mathbf{T}}\right)^{-1}, k, n-1-N+k\right)}{n-1}. \tag{1.31}$$

Following similar way in previous section, one can also do partition on the return vector $\mathbf{R}_i$. Given some nonrandom sets S and S^c, let $\mathbf{R}_{s,t} \in \mathbb{R}^{|S|}$, and $\mathbf{R}_{s^c,t} \in \mathbb{R}^{|S^c|}$, $t = 1, \ldots, n$ be the first $|S|$ and the rest $|S^c| = N - |S|$ elements of $\mathbf{R}_t$, then

$$\begin{aligned}
&\mathbf{R}_{s,t} \overset{iid}{\sim} \mathcal{N}(\mu_s, \Sigma_{ss}), \\
&\mathbf{R}_{s^c,t} \overset{iid}{\sim} \mathcal{N}(\mu_{s^c}, \Sigma_{s^cs^c}),
\end{aligned}$$

where

$$\mu = (\mu_s, \mu_{s^c}), \text{ and } \Sigma = \begin{pmatrix} \Sigma_{ss} & \Sigma_{s^cs} \\ \Sigma_{s^cs} & \Sigma_{s^cs^c} \end{pmatrix}.$$

Furthermore, when $\widehat{\Sigma}$ is the sample covariance matrix, then

$$\begin{aligned}
\widehat{\Sigma}_{ss} &\sim \frac{\mathcal{W}(\Sigma_{ss}, |S|, n-1)}{n-1}, \\
\widehat{\Sigma}_{s^cs^c} &\sim \frac{\mathcal{W}(\Sigma_{s^cs^c}, |S^c|, n-1)}{n-1},
\end{aligned}$$

where $\widehat{\Sigma}_{ss}$ and $\widehat{\Sigma}_{s^cs^c}$ are sample covariance estimates of Σ_{ss} and $\Sigma_{s^cs^c}$ respectively. Furthermore, (1.30) and (1.31) also hold when $\widehat{\Sigma}$ and Σ are replaced by these submatrices $\widehat{\Sigma}_{ss}$ and $\widehat{\Sigma}_{s^cs^c}$ and A replaced by deterministic matrices A_s (with dimension $k \times |S|$) and A_{s^c} (with dimension $k \times |S^c|$). The above property holds for any $S \subseteq N$, and is useful on the following proof as we want to construct probability of a certain event.

1.8.4 Eigenvalues of the Sample Covariance Matrix

I then discuss some issues on the convergence of the extreme eigenvalues of a sample covariance matrix. By using Theorem 5.11 in Bai and Silverstein (2010), one can obtain an useful result on bounded eigenvalues of the sample covariance matrix with iid normal samples.

Lemma 3 *Suppose* $\mathbf{Z}_i \overset{iid}{\sim} \mathcal{N}\left(\mathbf{0}, \sigma^2 \otimes \mathbf{I}_{NN}\right)$, *and*

$$\hat{\Sigma} = \frac{\Sigma^{\frac{1}{2}}\mathbf{Z}^{\mathbf{T}}\mathbf{H}_n\mathbf{Z}\Sigma^{\frac{1}{2}}}{n-1},$$

where $\mathbf{Z}$ *is a* $n \times N$ *matrix which ith row is* $\mathbf{Z}_i$, Σ *is a* $N \times N$ *symmetric and positive semidefinite matrix, and* $\mathbf{H}_n = \mathbf{I}_{nn} - n^{-1}\mathbf{1}_n\mathbf{1}_n^{\mathbf{T}}$. *Let* $\phi_j(M)$, $j = 1, \ldots, N$, *denote eigenvalues of matrix* M, *and* $\phi^{\min}(M)$ *and* $\phi^{\max}(M)$ *denote the smallest and largest eigenvalues of* M, *respectively. As* $\lim_{n\to\infty} N/n = \rho_N \in (0,1)$, *almost surely*

$$\sigma^2\left(1-\sqrt{\rho_N}\right)^2\phi^{\min}(\Sigma) \leq \lim_{n\to\infty}\phi_j\left(\hat{\Sigma}\right) \leq \sigma^2\left(1+\sqrt{\rho_N}\right)^2\phi^{\max}(\Sigma),$$

for every j, $j = 1, \ldots, N$.

To prove Lemma 3, I will use Theorem 5.11 in Bai and Silverstein (2010). For completeness, I restate the theorem here.

Theorem 2 *(Bai and Silverstein, 2010) Let* C^N *denote the set of* N *dimensional complex numbers. Assume* $\mathbf{Z}_i \in C^N$, $i = 1, \ldots, n$ *be iid with mean* $\mathbf{0}_N$ *and covariance matrix*Σ *in which its diagonal term* $\Sigma_{jj} = \sigma^2$, *and off-diagonal terms* $\Sigma_{jl} = 0$, $j, l = 1, \ldots, N$, $j \neq l$. *Suppose* $\mathbf{Z}_i$ *also has finite fourth moment. As* $\lim_{n\to\infty} N/n = \rho_N \in (0,1)$, *then almost surely one can have*

$$\begin{aligned}\lim_{n\to\infty}\phi^{\min}\left(\hat{\Sigma}\right) &= \sigma^2\left(1-\sqrt{\rho_N}\right)^2, \\ \lim_{n\to\infty}\phi^{\max}\left(\hat{\Sigma}\right) &= \sigma^2\left(1+\sqrt{\rho_N}\right)^2,\end{aligned}$$

where$\hat{\Sigma} = (n-1)^{-1}\sum_{i=1}^n \mathbf{Z}\mathbf{Z}_i^{\mathbf{T}}$.

The proof of Lemma 3 is straightforward, and can be accomplished by applying the above result to the matrix $\mathbf{Z}^{\mathbf{T}}\mathbf{H}_n\mathbf{Z}$.

Proof. Since Σ and $\hat{\Sigma}$ are both positive semidefinite, for any arbitrary $1 \times N$ vector$\mathbf{x}$

$$0 \leq \phi^{\min}\left(\frac{\mathbf{Z}^{\mathbf{T}}\mathbf{H}_n\mathbf{Z}}{n-1}\right)\mathbf{x}^{\mathbf{T}}\Sigma\mathbf{x} \leq \frac{\mathbf{x}^{\mathbf{T}}\hat{\Sigma}\mathbf{x}}{n-1} \leq \phi^{\max}\left(\frac{\mathbf{Z}^{\mathbf{T}}\mathbf{H}_n\mathbf{Z}}{n-1}\right)\mathbf{x}^{\mathbf{T}}\Sigma\mathbf{x}.$$

Then it can be shown that

$$\begin{aligned}\phi^{\max}\left(\hat{\Sigma}\right) &\leq \phi^{\max}\left(\frac{\mathbf{Z^T H}_n \mathbf{Z}}{n-1}\right)\phi^{\max}(\Sigma),\\ \phi^{\min}\left(\hat{\Sigma}\right) &\geq \phi^{\min}\left(\frac{\mathbf{Z^T H}_n \mathbf{Z}}{n-1}\right)\phi^{\min}(\Sigma).\end{aligned}$$

Note that the above inequalities hold for every $n > 1$ and $N \geq 1$. Since $\mathbf{H}_n$ is idempotent, it can be shown that

$$\mathbf{Z^T H}_n \mathbf{Z} = \left(\mathbf{Z^T H}_n\right)\left(\mathbf{H}_n \mathbf{Z}\right) = \sum_{i=1}^{n}\left(\mathbf{Z}_i - \overline{\mathbf{Z}}\right)\left(\mathbf{Z}_i - \overline{\mathbf{Z}}\right)^T,$$

where $\overline{\mathbf{Z}} = n^{-1}\sum_{i=1}^{n}\mathbf{Z}_i$. Since $\mathbf{Z}_i \overset{iid}{\sim} \mathcal{N}\left(\mathbf{0}, \sigma^2 \otimes \mathbf{I}_{NN}\right)$, $\mathbf{Z}_i - \overline{\mathbf{Z}}$ has mean $\mathbf{0}_N$ and diagonal covariance matrix which diagonal elements equal to $n^{-1}(n-1)\sigma^2$. By Theorem 5.11 in Bai and Silverstein (2010), it can be shown that as $\lim_{n\to\infty} N/n = \rho_N \in (0,1)$, almost surely

$$\begin{aligned}\lim_{n\to\infty}\phi^{\min}\left(\frac{\mathbf{Z^T H}_n \mathbf{Z}}{n-1}\right) &= \sigma^2\left(1-\sqrt{\rho_N}\right)^2,\\ \lim_{n\to\infty}\phi^{\max}\left(\frac{\mathbf{Z^T H}_n \mathbf{Z}}{n-1}\right) &= \sigma^2\left(1+\sqrt{\rho_N}\right)^2,\end{aligned}$$

by $n^{-1}(n-1)\sigma^2 \to \sigma^2$. Therefore as $\lim_{n\to\infty} N/n = \rho_N \in (0,1)$, almost surely

$$\begin{aligned}\lim_{n\to\infty}\phi^{\max}\left(\hat{\Sigma}\right) &\leq \sigma^2\left(1+\sqrt{\rho_N}\right)^2\phi^{\max}(\Sigma),\\ \lim_{n\to\infty}\phi^{\min}\left(\hat{\Sigma}\right) &\geq \sigma^2\left(1-\sqrt{\rho_N}\right)^2\phi^{\min}(\Sigma).\end{aligned}$$

■

1.8.5 Estimation Errors

Suppose that $\widehat{\sigma}_{ij}$ and $\widehat{\sigma}_i^2$ and are consistent estimators for σ_{ij} and σ_i^2, respectively. Without loss of generality, one may assume $\widehat{\sigma}_{ij}$ and σ_{ij} have the following linear relationship,

$$\begin{aligned}\sigma_{ij} &= \widehat{\sigma}_{ij} + \omega_{ij}(n),\\ \sigma_i^2 &= \widehat{\sigma}_i^2 + \omega_{ii}(n),\end{aligned}$$

where $\omega_{ij}(n)$ ($\omega_{ii}(n)$) is the estimation error for σ_{ij} (σ_i^2) by using $\widehat{\sigma}_{ij}$ ($\widehat{\sigma}_{ij}$), and is a function of sample size n. If σ_{ij} and σ_i^2 are consistent, as $n \to \infty$, $\omega_{ij}(n)$ and $\omega_{ii}(n)$ should vanish to zero with a high probability. Σ may be expressed as

$$\Sigma = \widehat{\Sigma} + \Omega(n) \tag{1.32}$$

where $\Omega(n)$ is an $N \times N$ estimation error matrix with the (i,j)th elements $\omega_{ij}(n)$, $i,j = 1,\ldots,N$. It is also symmetric but may not be positive semidefinite.

Furthermore, by similar partition used in $\widehat{\Sigma}$, it follows that

$$\Omega(n) = \begin{pmatrix}\Omega_{\widehat{S}\widehat{S}}(n) & \Omega_{\widehat{S}\widehat{S}^c}(n)\\ \Omega_{\widehat{S}^c\widehat{S}}(n) & \Omega_{\widehat{S}^c\widehat{S}^c}(n)\end{pmatrix}.$$

Then the difference between the oos variances of weighted norm mvp and gmvp can be expressed as

$$\widehat{\mathbf{w}}_{\widehat{s}}^{*\mathbf{T}}\Sigma_{\widehat{s}\widehat{s}}\widehat{\mathbf{w}}_{\widehat{s}}^{*}-\widehat{\mathbf{w}}_{un}^{*\mathbf{T}}\Sigma\widehat{\mathbf{w}}_{un}^{*}=\hat{D}_{1,\hat{s}}+\hat{D}_{2,\hat{s}}, \tag{1.33}$$

where

$$\begin{aligned}\hat{D}_{1,\hat{s}} &= \widehat{\sigma}_{\widehat{s}}^{2}-\widehat{\sigma}_{un}^{2}, &(1.34)\\ \hat{D}_{2,\hat{s}} &= \widehat{\mathbf{w}}_{\widehat{s}}^{*\mathbf{T}}\Omega_{\widehat{s}\widehat{s}}\left(n\right)\widehat{\mathbf{w}}_{\widehat{s}}^{*}-\widehat{\mathbf{w}}_{un}^{*\mathbf{T}}\Omega\left(n\right)\widehat{\mathbf{w}}_{un}^{*}. &(1.35)\end{aligned}$$

That is, the sum of the difference between their in-sample variances and the difference between the estimation errors.

Given $\hat{S}=S$, Fan et al. (2009) shows that

$$\widehat{\mathbf{w}}_{s}^{*\mathbf{T}}\Omega_{ss}\left(n\right)\widehat{\mathbf{w}}_{s}^{*}\quad\leq\quad\max_{i,j}\left|\omega_{ij}\left(n\right)\right|\left\|\widehat{\mathbf{w}}_{s}^{*}\right\|_{1}^{2}.$$

Therefore if $\left\|\widehat{\mathbf{w}}_{s}^{*}\right\|_{1}^{2}$ is bounded, the estimation error term of the weighted norm mvp is bounded by the maximum$\left|\omega_{ij}\left(n\right)\right|$ scaled by $\left\|\widehat{\mathbf{w}}_{s}^{*}\right\|_{1}^{2}$. Here, size of the optimal weight vector plays an important role in reducing the estimation errors. If $\widehat{\mathbf{w}}_{s}^{*}$ is the optimal no-shortsales weight with $\widehat{\mathbf{w}}_{s}^{*\mathbf{T}}\mathbf{1}_{N}=1$, then $\left\|\widehat{\mathbf{w}}_{s}^{*}\right\|_{1}^{2}=1$. The no-shortsales mvp has the smallest $\left\|\widehat{\mathbf{w}}_{s}^{*}\right\|_{1}^{2}$ over the mvp with the full investment constraint, and it is main reason why it can efficiently eliminate the estimation errors.

The term $\hat{D}_{1,\hat{s}}$ can be further expressed as

$$\left(\widehat{\sigma}_{2,\widehat{s}}^{2}-\widehat{\sigma}_{un}^{2}\right)+\lambda_{2}\left(\left\|\widehat{\mathbf{w}}_{2,s}^{*}\right\|_{2}^{2}-\left\|\widehat{\mathbf{w}}_{s}^{*}\right\|_{2}^{2}\right)+\frac{\lambda_{1}}{2}\left(\mathbf{u}^{\mathbf{T}}\widehat{\delta}_{2,s}-\left\|\widehat{\mathbf{w}}_{s}^{*}\right\|_{1}\right). \tag{1.36}$$

The first term of (1.36) is the difference between the optimal in-sample portfolio variances from (1.6) and (1.7), and it is nonnegative. To see this, at first note that $\widehat{\sigma}_{un}^{2}\leq\widehat{\sigma}_{\widehat{s},un}^{2}$, since the in-sample portfolio variance from optimally choosing a larger set of assets will always be no greater than that from optimally choosing a smaller subset of the same assets. Furthermore, $\widehat{\sigma}_{\widehat{s},un}^{2}\leq\widehat{\sigma}_{2,\widehat{s}}^{2}$, since the later is obtained by restricting the portfolio weights by the l_2 norm on the assets in $\hat{S}$, while the former is obtained without such constraint. For the second and third term of (1.36), as shown in the proof of Lemma 2, is the difference between the optimal in-sample portfolio variances of (1.1) and (1.6), and it is nonnegative. Therefore (1.36) in general is nonnegative.

The term $\hat{D}_{2,\hat{s}}$ is the difference between the estimation errors, and it is determined by the estimation error matrix and its principal submatrix, and the two optimal weight vectors. Given that $\hat{D}_{1,\hat{s}}\geq0$, to make (1.44) nonpositive, $\hat{D}_{2,\hat{s}}$ needs to be large enough to offset the nonnegativity caused by $\hat{D}_{1,\hat{s}}$. Furthermore, we can have the following lemma for the estimation error *omega*$_i$ when the returns are i.i.d. normal and the sample variance and covariance estimators are used.

Lemma 4 *Suppose* $\mathbf{R}_i\overset{iid}{\sim}\mathcal{N}\left(\mu,\Sigma\right)$, $i=1,\ldots,n$, *and every element in the covariance* Σ *is finite. If* $\widehat{\Sigma}$ *is the sample covariance matrix. Then*

$$P\left(\left|\omega_{ij}\left(n\right)\right|\geq v\right)=O\left(n^{-\frac{1}{2}}\exp\left(-a_{ij}v^{2}n\right)\right),$$

where a_{ij} *is positive and dependent on* σ_i^2, σ_j^2 *and* σ_{ij}.

Proof. Note that

$$\begin{aligned} |\omega_{ij}(n)| &= |\widehat{\sigma}_{ij} - \sigma_{ij}| \\ &\le \left| \frac{1}{n-1} \sum_{k=1}^{n} (R_{ik}R_{jk} - \sigma_{ij}) \right| + \left| \frac{1}{n-1} \left(n\overline{R}_i \overline{R}_j - \sigma_{ij} \right) \right|. \end{aligned} \tag{1.37}$$

The first term can be shown to satisfy

$$\left| \frac{1}{n-1} \sum_{k=1}^{n} (R_{ik}R_{jk} - \sigma_{ij}) \right| \le E_{n,1} + E_{n,2},$$

where

$$\begin{aligned} E_{n,1} &= \left| \frac{1}{4(n-1)} \sum_{k=1}^{n} \left((R_{ik} + R_{jk})^2 - \left(2\sigma_{ij} + \sigma_i^2 + \sigma_j^2 \right) \right) \right| \\ E_{n,2} &= \left| \frac{1}{4(n-1)} \sum_{k=1}^{n} \left((R_{ik} - R_{jk})^2 - \left(\sigma_i^2 + \sigma_j^2 - 2\sigma_{ij} \right) \right) \right|. \end{aligned}$$

Let $Var_{ij}^{+} = \sigma_i^2 + \sigma_j^2 + 2\sigma_{ij}$ and $Var_{ij}^{-} = \sigma_i^2 + \sigma_j^2 - 2\sigma_{ij}$. It is known that

$$\begin{aligned} R_{ik} + R_{jk} &\sim \mathcal{N}\left(0, Var_{ij}^{+}\right), \\ R_{ik} - R_{jk} &\sim \mathcal{N}\left(0, Var_{ij}^{-}\right), \end{aligned}$$

Then

$$\begin{aligned} P(E_{n,1} \ge v) &= P\left(\left| \sum_{k=1}^{n} \left(\left(\frac{R_{ik} + R_{jk}}{\sqrt{Var_{ij}^{+}}} \right)^2 - 1 \right) \right| \ge \frac{4(n-1)v}{Var_{ij}^{+}} \right) \\ &= P\left(\chi_n^2 \ge n + \frac{4(n-1)v}{Var_{ij}^{+}} \right) + P\left(\chi_n^2 \le n - \frac{4(n-1)v}{Var_{ij}^{+}} \right), \end{aligned} \tag{1.38}$$

where $v \ge 0$ is a constant. By similar arguments,

$$P(E_{n,2} \ge v) \le P\left(\chi_n^2 \ge n + \frac{4(n-1)v}{Var_{ij}^{-}} \right) + P\left(\chi_n^2 \le n - \frac{4(n-1)v}{Var_{ij}^{-}} \right). \tag{1.39}$$

I first discuss the second terms in (1.38) and (1.46). If

$$v \ge \frac{n}{4(n-1)} \max\left(Var_{ij}^{-}, Var_{ij}^{+} \right),$$

then

$$P\left(\chi_n^2 \le n - \frac{4(n-1)v}{Var_{ij}^{+}} \right) = P\left(\chi_n^2 \le n - \frac{4(n-1)v}{Var_{ij}^{-}} \right) = 0,$$

since χ_n^2 is nonnegative with probability one. Let

$$\theta_{ij} = \frac{4}{\max\left(Var_{ij}^{-}, Var_{ij}^{+} \right)}.$$

Suppose that

$$0 < n - \frac{4(n-1)v}{\max\left(Var_{ij}^{-}, Var_{ij}^{+}\right)} = n - \theta_{ij} v n + \theta_{ij} v \leq n.$$

Then

$$\begin{aligned}
\max\left(P\left(\chi_n^2 \leq n - \frac{4(n-1)v}{Var_{ij}^{-}}\right), P\left(\chi_n^2 \leq n - \frac{4(n-1)v}{Var_{ij}^{+}}\right)\right) &= P\left(\chi_n^2 \leq (1-\theta_{ij} v) n + \theta_{ij} v\right) \\
&\leq P\left(\chi_n^2 \leq (1-\theta_{ij} v) n\right) + P\left(\chi_n^2 \leq \theta_{ij} v\right).
\end{aligned}$$

We know that

$$\begin{aligned}
P\left(\chi_n^2 \leq (1-\theta_{ij} v) n\right) &= \frac{\underline{\Gamma}\left(\frac{n}{2}, \frac{(1-\theta_{ij} v) n}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}, \\
P\left(\chi_n^2 \leq \theta_{ij} v\right) &= \frac{\underline{\Gamma}\left(\frac{n}{2}, \frac{\theta_{ij} v}{2}\right)}{\Gamma\left(\frac{n}{2}\right)},
\end{aligned}$$

where $\underline{\Gamma}(u, z) = \int_0^u x^{z-1} \exp(-x)\, dx$ is the lower incomplete gamma function, and $\Gamma(u)$ is the gamma function. It can be shown that

$$\underline{\Gamma}(u, z) \leq \frac{z^u \exp(-z)}{u - \frac{uz}{u+1}} \leq \frac{z^u \exp(-z)}{u - \frac{uz}{u+1}} = \frac{u+1}{u} \frac{z^u \exp(-z)}{u+1-z}.$$

Then

$$\begin{aligned}
\ln \underline{\Gamma}(u, z) &\leq \ln(u+1) - \ln u + u \ln z - z - \ln(u+1-z) \\
\ln \underline{\Gamma}(u, \tau u) &\leq \ln(u+1) - \ln u + u(\ln \tau + \ln u) - \tau u - \ln(u+1-\tau u).
\end{aligned}$$

If $u > 0$, the logrithm of the gamma function is

$$\begin{aligned}
\ln \Gamma(u) &= \left(u - \frac{1}{2}\right) \ln u - u + \frac{1}{2} \ln 2\pi + 2 \int_0^\infty \left(\frac{\arctan x/u}{\exp(2\pi x) - 1} dx\right) \\
&= \left(u - \frac{1}{2}\right) \ln u - u + \frac{1}{2} \ln 2\pi + O(1).
\end{aligned}$$

Therefore

$$\begin{aligned}
\ln \frac{\underline{\Gamma}(u, z)}{\Gamma(u)} &\leq \ln\left(\frac{u+1}{u}\right) + \ln \frac{u^{\frac{1}{2}}}{u-z} - (\ln u - (\ln z + 1)) u - z - \frac{1}{2} \ln 2\pi - O(1), \\
\ln \frac{\underline{\Gamma}(u, \tau u)}{\Gamma(u)} &\leq \ln\left(\frac{u+1}{u}\right) - \frac{1}{2} \ln u - (\tau - (1 + \ln \tau)) u - \ln(1-\tau) - \frac{1}{2} \ln 2\pi - O(1).
\end{aligned}$$

Note that $\tau > 1 + \ln \tau$ if $\tau \in (0,1)$. Furthermore, if $\tau \in (0,1)$,

$$(1-\tau)^2 \lesssim \tau - (1 + \ln \tau).$$

Then if I set $u = \frac{n}{2}$

$$P\left(\chi_n^2 \le \tau n\right) = \frac{\underline{\Gamma}\left(\frac{n}{2}, \frac{\tau n}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} = O\left(n^{-\frac{1}{2}} \exp\left(-\frac{\left(1-\tau\right)^2 n}{2}\right)\right).$$

Replacing τ with $(1-\theta_{ij}v)$ and z with $n^{-1}\theta_{ij}v$, then

$$\begin{aligned}\frac{\underline{\Gamma}\left(\frac{n}{2}, \frac{(1-\theta_{ij}v)n}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} &= O\left(n^{-\frac{1}{2}} \exp\left(-\frac{\theta_{ij}^2 v^2 n}{2}\right)\right) \\ \frac{\underline{\Gamma}\left(\frac{n}{2}, \frac{\theta_{ij}v}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} &= O\left(n^{-\frac{1}{2}} \exp\left(-\frac{(\ln n)\, n}{2}\right)\right).\end{aligned}$$

Therefore when n goes large

$$\max\left(P\left(\chi_n^2 \le n - \frac{4\,(n-1)\,v}{Var_{ij}^-}\right), P\left(\chi_n^2 \le n - \frac{4\,(n-1)\,v}{Var_{ij}^+}\right)\right) = O\left(n^{-\frac{1}{2}} \exp\left(-\frac{\theta_{ij}^2 v^2 n}{2}\right)\right) \quad (1.40)$$

For the first terms in (1.38) and (1.46), it can be shown that

$$\begin{aligned}\max\left(P\left(\chi_n^2 \ge n + \frac{4\,(n-1)\,v}{Var_{ij}^-}\right), P\left(\chi_n^2 \ge n + \frac{4\,(n-1)\,v}{Var_{ij}^+}\right)\right) &\le P\left(\chi_n^2 \ge (1+\theta_{ij}v)\,n - \theta_{ij}v\right) \\ &= \frac{\overline{\Gamma}\left(\frac{n}{2}, \frac{(1+\theta_{ij}v)n-\theta_{ij}v}{2}\right)}{\Gamma\left(\frac{n}{2}\right)},\end{aligned}$$

where $\overline{\Gamma}(u,z) = \int_z^\infty x^{u-1}\exp(-x)\,dx$ is the upper incomplete gamma function. By similar arguments to prove (1.40), one can derive the upper bound for these terms. In sum, we can conclude that

$$P\left(\left|\frac{1}{n-1}\sum_{k=1}^{n}(R_{ik}R_{jk} - \sigma_{ij})\right| \ge v\right) = O\left(n^{-\frac{1}{2}} \exp\left(-a'_{ij}v^2 n\right)\right), \quad (1.41)$$

where a'_{ij} is positive and dependent on σ_i^2, σ_j^2 and σ_{ij}. As for the second term of (1.37), it can be shown that

$$\left|\frac{1}{n-1}\left(n\overline{R}_i\overline{R}_j - \sigma_{ij}\right)\right| \le E_{3,n} + E_{4,n},$$

where

$$\begin{aligned}E_{3,n} &= \left|\frac{1}{n\,(n-1)}\sum_{k=1}^{n}(R_{ik}R_{jk} - \sigma_{ij})\right| \\ E_{4,n} &= \left|\frac{1}{n-1}\sum_{k=2}^{n}R_{i1}R_{jk}\right|.\end{aligned}$$

Then it follows

$$\begin{aligned} P\left(E_{3,n} \geq v\right) &\leq P\left(\chi_n^2 \geq n + \frac{4\left(n-1\right)nv}{Var_{ij}^{+}}\right) + P\left(\chi_n^2 \leq n - \frac{4\left(n-1\right)nv}{Var_{ij}^{+}}\right) \\ &\quad + P\left(\chi_n^2 \geq n + \frac{4\left(n-1\right)nv}{Var_{ij}^{-}}\right) + P\left(\chi_n^2 \leq n - \frac{4\left(n-1\right)nv}{Var_{ij}^{-}}\right) \end{aligned} \tag{1.42}$$

and

$$P\left(E_{4,n} \geq v\right) \leq 2P\left(\chi_{n-1}^2 \geq n + \frac{4\left(n-1\right)v}{\sigma_i^2 + \sigma_j^2}\right) + 2P\left(\chi_{n-1}^2 \geq n - \frac{4\left(n-1\right)v}{\sigma_i^2 + \sigma_j^2}\right). \tag{1.43}$$

by $cov\left(R_{i1}, R_{jk}\right) = 0$ for $k = 2, \ldots, n$ (i.i.d. of the returns). By similar arguments to prove (1.41) it is not difficult to see (1.42) and (1.43) have the same order as (1.41). Thus in sum

$$P\left(\left|\omega_{ij}\left(n\right)\right| \geq v\right) = O\left(n^{-\frac{1}{2}} \exp\left(-a_{ij} v^2 n\right)\right),$$

where a_{ij} is positive and determined by a'_{ij}, σ_i^2, σ_j^2 and σ_{ij}. ■

1.8.6 Proof of Theorem 1

Proof. Given $S \subseteq \{1, \ldots, N\}$, by Lemma 2, one can have

$$\begin{aligned} \widehat{\mathbf{w}}_s^{*\mathbf{T}} \Sigma_{ss} \widehat{\mathbf{w}}_s^{*} - \widehat{\mathbf{w}}_{un}^{*\mathbf{T}} \Sigma \widehat{\mathbf{w}}_{un}^{*} &= \left(\widehat{\sigma}_s^2 - \widehat{\sigma}_{un}^2\right) - \left(\widehat{\mathbf{w}}_{un}^{*\mathbf{T}} \Omega\left(n\right) \widehat{\mathbf{w}}_{un}^{*} - \widehat{\mathbf{w}}_s^{*\mathbf{T}} \Omega_{ss}\left(n\right) \widehat{\mathbf{w}}_s^{*}\right) \\ &= \left(\hat{c}_{s,1}\left(\hat{c}_{s,1}+1\right)+1\right) \widehat{\sigma}_{un,s}^2 - \hat{\sigma}_{un}^2 - \lambda_2 \left\|\widehat{\mathbf{w}}_s^{*}\right\|_2^2 \\ &\quad + \frac{\lambda_1}{2}\left(\mathbf{u}^{\mathbf{T}} \hat{\delta}_{2,s} - \left\|\widehat{\mathbf{w}}_s^{*}\right\|_1\right) - \left(\widehat{\mathbf{w}}_{un}^{*\mathbf{T}} \Omega\left(n\right) \widehat{\mathbf{w}}_{un}^{*} - \max_{i,j \in S} \left|\omega_{ij}\left(n\right)\right| \left\|\widehat{\mathbf{w}}_s^{*}\right\|_1^2\right) \\ &\leq \hat{c}_{s,2} \widehat{\sigma}_{un,s}^2 - \left(\hat{\sigma}_{un}^2 + \widehat{\mathbf{w}}_{un}^{*\mathbf{T}} \Omega\left(n\right) \widehat{\mathbf{w}}_{un}^{*}\right) + \left(\lambda_1 \hat{c}_{s,3} - \lambda_2 \hat{c}_{s,4}\right) \\ &\quad - \lambda_1 \left\|\widehat{\mathbf{w}}_s^{*}\right\|_1 + \max_{i,j \in S} \left|\omega_{ij}\left(n\right)\right| \left\|\widehat{\mathbf{w}}_s^{*}\right\|_1^2 \end{aligned} \tag{1.44}$$

where $\hat{c}_{s,1} = \lambda_2^{\frac{1}{2}} \left(\phi^{\min}\left(\hat{\Sigma}_{ss}\right)\right)^{-\frac{1}{2}}$, $\hat{c}_{s,2} = \hat{c}_{s,1}\left(\hat{c}_{s,1}+1\right)+1$, $\hat{c}_{s,3} = 2^{-1}\left(\mathbf{u}^{\mathbf{T}} \hat{\delta}_{2,s} + \left\|\widehat{\mathbf{w}}_s^{*}\right\|_1\right)$, and $\hat{c}_{s,4} = \left\|\widehat{\mathbf{w}}_s^{*}\right\|_2^2$. As shown by El Karoui (2009),

$$\widehat{\mathbf{w}}_{un}^{*\mathbf{T}} \Sigma \widehat{\mathbf{w}}_{un}^{*} \cong \frac{1}{1-\rho_N} \sigma_{un}^2,$$

and it follows that $\widehat{\mathbf{w}}_{un}^{*\mathbf{T}} \Omega\left(n\right) \widehat{\mathbf{w}}_{un}^{*} - \rho_N \sigma_{un}^2 = O_p\left(\rho_N \left(1-\rho_N\right)^{-1}\right)$, and it is positive. Therefore the last inequality is bounded by

$$\hat{c}_{s,2} \widehat{\sigma}_{un,s}^2 - \hat{\sigma}_{un}^2 - \rho_N \sigma_{un}^2 + \left(\lambda_1 \hat{c}_{s,3} - \lambda_2 \hat{c}_{s,4}\right) + \lambda_1 \left\|\widehat{\mathbf{w}}_s^{*}\right\|_1 + \max_{i,j \in S} \left|\omega_{ij}\left(n\right)\right| \left\|\widehat{\mathbf{w}}_s^{*}\right\|_1^2$$

Then we have a look of $\hat{c}_{s,3}$. With some algebra, it can be shown that $\lambda_1 \hat{c}_{s,3}$ is the difference between two optimal values of objective functions: the weighted norm and squared l_2 norm optimizations, when $\hat{\Sigma}_{ss}$ is used. Note that squared l_2 norm optimization is a special case of the weighted norm optimization when $\lambda_1 = 0$. By the assumption that the weighted norm optimization is feasible for for $0 \leq \lambda_1, \lambda_2 < \infty$, as n, $N \to \infty$, we can conclude that $\hat{c}_{s,3} = O_p\left(1\right)$. With similar arguments, $\hat{c}_{s,4} = O_p\left(1\right)$, since it is the squared l_2 norm penalty of the objective function of the weighted norm portfolio optimization.

Let

$$\mathcal{A}_{1,s} = \left\{ \hat{c}_{s,2}\widehat{\sigma}^2_{un,s} - \rho_N \sigma^2_{un} + (\lambda_1 \hat{c}_{s,3} - \lambda_2 \hat{c}_{s,4}) + \max_{i,j\in S} |\omega_{ij}(n)| \, \|\widehat{\mathbf{w}}^*_s\|_1^2 \le \hat{\sigma}^2_{un} + \lambda_1 \|\widehat{\mathbf{w}}^*_s\|_1 \right\},$$

$\mathcal{A}_{1,s}$ is a sufficient condition to make (1.44) nonpositive. Define the following three events,

$$\begin{aligned} \mathcal{B}_{1,s} &= \{\lambda_1 \hat{c}_{s,3} - \lambda_2 \hat{c}_{s,4} \le B_{1,n}\}, \\ \mathcal{B}_{2,s} &= \left\{ \max_{i,j\in S} |\omega_{ij}(n)| \, \|\widehat{\mathbf{w}}^*_s\|_1^2 \le \lambda_1 \|\widehat{\mathbf{w}}^*_s\|_1 \right\}, \\ \mathcal{B}_{3,s} &= \left\{ \hat{c}_{s,2}\widehat{\sigma}^2_{un,s} - \rho_N \sigma^2_{un} \le \hat{\sigma}^2_{un} - B_{1,n} \right\}, \end{aligned}$$

where $B_{1,n}$ is a finite nonnegative constant. It follows that $\bigcap_{i=1}^3 \mathcal{B}_{i,s} \subseteq \mathcal{A}_{1,s}$. Therefore

$$\begin{aligned} P\left(\widehat{\mathbf{w}}^{*\mathbf{T}}_s \Sigma_{\hat{s}\hat{s}} \widehat{\mathbf{w}}^*_{\hat{s}} \le \widehat{\mathbf{w}}^{*\mathbf{T}}_{un} \Sigma \widehat{\mathbf{w}}^*_{un} | \hat{S} = S \right) &= P\left(\widehat{\mathbf{w}}^{*\mathbf{T}}_s \Sigma_{ss} \widehat{\mathbf{w}}^*_s \le \widehat{\mathbf{w}}^{*\mathbf{T}}_{un} \Sigma \widehat{\mathbf{w}}^*_{un} \right) \\ &\ge P(\mathcal{A}_{1,s}) \\ &\ge 1 - \sum_{i=1}^3 P\left(\mathcal{B}^c_{i,s}\right). \end{aligned}$$

I set

$$\lambda_1 = \lambda_2 = \lambda_{n,N} = B_0 \sqrt{\frac{2\log N}{a^{\min} n}},$$

$\lambda_2 = o(1)$ and $B_{1,n} = (\log n)^{-\frac{1}{2}}$, where $B_0 > 0$ is a constant such that $\sup_{S\subseteq\{1,\ldots,N\}} P\left(\|\widehat{\mathbf{w}}^*_s\|_{l_1} > B_0 \right) = o(1)$. For $P(\mathcal{B}^c_{1,s})$, we can have

$$P(\mathcal{B}^c_{1,s}) = P\left(\hat{c}_{s,3} - \hat{c}_{s,4} > B_2 \sqrt{\frac{n}{\log n \log N}} \right) \le P\left(\hat{c}_{s,3} - \hat{c}_{s,4} > \frac{B_2\sqrt{n}}{\log n} \right), \tag{1.45}$$

where $B_2 = \sqrt{a^{\min}} / \left(\sqrt{2} B_0\right)$. Since $\hat{c}_{s,3} - \hat{c}_{s,4} = O_p(1)$ and $B_2\sqrt{n}/\log n \to \infty$, (1.45) converges to zero as n, $N \to \infty$. Then we have a look of $P\left(\mathcal{B}^c_{2,s}\right)$. By Lemma 4 and the assumption that $\sup_{S\subseteq\{1,\ldots,N\}} P\left(\|\widehat{\mathbf{w}}^*_s\|_1 > B_0\right) = o(1)$, one can show that

$$\begin{aligned} P\left(\mathcal{B}^c_{2,s}\right) &= P\left(\max_{i,j\in S} |\omega_{ij}(n)| \, \|\widehat{\mathbf{w}}^*_s\|_1 > \lambda_1 \right) \\ &\le P\left(\left\{ \max_{i,j\in S} |\omega_{ij}(n)| B_0 > \lambda_1 \right\} \bigcap \left\{ \|\widehat{\mathbf{w}}^*_s\|_1 \le B_0 \right\} \right) + o(1) \\ &\le P\left(\max_{i,j\in S} |\omega_{ij}(n)| B_0 > \lambda_1 \right) + o(1). \\ &= N^2 O\left(n^{-\frac{1}{2}} \exp\left(-a^{\min} \frac{\lambda_1^2 n}{B_0^2} \right) \right) + o(1). \\ &= O\left(n^{-\frac{1}{2}} \right). \end{aligned}$$

For $P\left(\mathcal{B}^c_{3,s}\right)$, I at first show that

$$\hat{c}_{s,2}\widehat{\sigma}^2_{un,s} \xrightarrow{p} \frac{n - |S| - 1 + k}{n-1} \sigma^2_{un,s}, \tag{1.46}$$

as n and $n-|S|$ go large. Apply Lemma 3 to $\hat{\Sigma}_{ss}$, it can be shown that almost surely,

$$\lim_{n\to\infty} \phi^{\min}\left(\hat{\Sigma}_{ss}\right) \geq \phi^{\min}\left(\Sigma_{ss}\right)\left(1-\sqrt{\rho_s}\right)^2 > 0,$$

where $\rho_s = \lim_{n\to\infty} |S|/n$. Therefore

$$\lim_{n\to\infty} \hat{c}_{s,1} = \lim_{n\to\infty} \sqrt{\frac{\lambda_2}{\phi^{\min}\left(\hat{\Sigma}_{ss}\right)}} \leq \frac{1}{1-\sqrt{\rho_s}}\sqrt{\frac{\lambda_2}{\phi^{\min}\left(\Sigma_{ss}\right),}} = \frac{c_{s,1}}{1-\sqrt{\rho_s}}.$$

Now with $\lambda_2 = \lambda_{n,N}$, $c_{s,1} = o(1)$, thus $\hat{c}_{s,1} = o_p(1)$. Therefore $\hat{c}_{s,2} = \hat{c}_{s,1}\left(\hat{c}_{s,1}+1\right)+1 \xrightarrow{p.} 1$. For $\widehat{\sigma}^2_{un,s}$, it has been shown that given S,

$$\widehat{\sigma}^2_{un,s} = \mathbf{u}^{\mathbf{T}}\left(A_s\hat{\Sigma}_{ss}^{-1}A_s^T\right)^{-1}\mathbf{u} = \mathbf{u}^{\mathbf{T}}\left(A_s\Sigma_{ss}^{-1}A_s^T\right)^{-1}\mathbf{u}\frac{\chi_{n-1-|S|+k}}{n-1} = \sigma^2_{un,s}\frac{\chi^2_{n-1-|S|+k}}{n-1}.$$

Thus as n and $n-|S|$ go large, $\widehat{\sigma}^2_{un,s} \xrightarrow{p.} (n-1)^{-1}(n-|S|-1+k)\,\sigma^2_{un,s}$ by $\chi^2_{n-1-|S|+k} \xrightarrow{p.} n-|S|-1+k$. Combining the results, the claim (1.46) follows. Furthermore, as k is fixed and $k \ll |S|$, we can have

$$\frac{n-|S|-1+k}{n-1}\sigma^2_{un,s} \to (1-\rho_s)\,\sigma^2_{un,s},$$

as n and $|S|$ go large. Thus

$$P\left(\mathcal{B}^c_{3,s}\right) \quad = \quad P\left(\hat{c}_{s,2}\widehat{\sigma}^2_{un,s} - \rho_N\sigma^2_{un} > \hat{\sigma}^2_{un} - B_{1,n}\right) \to P\left((1-\rho_s)\,\sigma^2_{un,s} > \sigma^2_{un}\right),$$

by setting $B_{1,n} = (\log n)^{-\frac{1}{2}}$.

I then prove that if (1.17) hold, $P\left(\widehat{\mathbf{w}}^{*\mathbf{T}}\Sigma\widehat{\mathbf{w}}^* \leq \widehat{\mathbf{w}}^{*\mathbf{T}}_{un}\Sigma\widehat{\mathbf{w}}^*_{un}\right) \to 1$ as $n\to\infty$. We are hoping to construct the lower bound of the unconditional probability by aggregating $P\left(\widehat{\mathbf{w}}^{*\mathbf{T}}_{\hat{s}}\Sigma_{\hat{s}\hat{s}}\widehat{\mathbf{w}}^*_{\hat{s}} \leq \widehat{\mathbf{w}}^{*\mathbf{T}}_{un}\Sigma\widehat{\mathbf{w}}^*_{un}|\hat{S}\right)$ over different realized $\hat{S}$:

$$\begin{aligned}
P\left(\widehat{\mathbf{w}}^{*\mathbf{T}}\Sigma\widehat{\mathbf{w}}^* \leq \widehat{\mathbf{w}}^{*T}_{un}\Sigma\widehat{\mathbf{w}}^*_{un}\right) \quad &= \quad \sum_{S\subseteq\{1,\ldots,N\}} P\left(\widehat{\mathbf{w}}^{*\mathbf{T}}_{\hat{s}}\Sigma_{\hat{s}\hat{s}}\widehat{\mathbf{w}}^*_{\hat{s}} \leq \widehat{\mathbf{w}}^{*\mathbf{T}}_{un}\Sigma\widehat{\mathbf{w}}^*_{un}|\hat{S}=S\right)P\left(\hat{S}=S\right)\\
&\geq \quad \sum_{S\subseteq\{1,\ldots,N\}}\left(1-\sum_{i=1}^{3}P\left(\mathcal{B}^c_{i,s}\right)\right)P\left(\hat{S}=S\right)\\
&= 1- \sum_{i=1}^{3}\sum_{S\subseteq\{1,\ldots,N\}} P\left(\mathcal{B}^c_{i,s}\right)P\left(\hat{S}=S\right)
\end{aligned}$$

First, note that upper bound of the third term $P\left(\mathcal{B}^c_{2,s}\right)$ is independent of $\hat{S}$, thus

$$\sum_{S\subseteq\{1,\ldots,N\}} P\left(\mathcal{B}^c_{2,s}\right)P\left(\hat{S}=S\right) = O\left(n^{-\frac{1}{2}}\right).$$

For the other two terms, we have

$$\begin{aligned}\sum_{S\subseteq\{1,\ldots,N\}} P\left(\mathcal{B}_{1,s}^{c}\right) P\left(\hat{S}=S\right) &= \sum_{S\subseteq\{1,\ldots,N\}} P\left(\lambda_1 \hat{c}_{s,3}-\lambda_2\hat{c}_{s,4} > B_{1,n}\right) P\left(\hat{S}=S\right) \\ &\le \sup_{S\subseteq\{1,\ldots,N\}} P\left(\hat{c}_{s,3}-\hat{c}_{s,4} > \frac{B_2\sqrt{n}}{\log n}\right), \qquad (1.47)\\ \sum_{S\subseteq\{1,\ldots,N\}} P\left(\mathcal{B}_{3,s}^{c}\right) P\left(\hat{S}=S\right) &= \sum_{S\subseteq\{1,\ldots,N\}} P\left(\hat{c}_{s,2}\widehat{\sigma}_{un,s}^{2}-\rho_N\sigma_{un}^2 > \hat{\sigma}_{un}^2 - B_{1,n}\right) P\left(\hat{S}=S\right) \\ &\le \sup_{S\subseteq\{1,\ldots,N\}} P\left(\hat{c}_{s,2}\widehat{\sigma}_{un,s}^{2}-\rho_N\sigma_{un}^2 > \hat{\sigma}_{un}^2 - \frac{1}{\sqrt{\log n}}\right). \qquad (1.48)\end{aligned}$$

By the assumption that the weighted norm optimization is feasible, $\hat{c}_{s,3}-\hat{c}_{s,4}=O_p(1)$ holds for every S as n, $N\to\infty$, thus (1.47) converges to zero as $n\to\infty$. For (1.48), as n, $n-|S|\to\infty$,

$$\begin{aligned}P\left(\hat{c}_{s,2}\widehat{\sigma}_{un,s}^{2}-\rho_N\sigma_{un}^2 > \hat{\sigma}_{un}^2 - B_{1,n}\right) &\to P\left(\left(1-\rho_s\right)\sigma_{un,s}^2 > \sigma_{un}^2\right)\\ &= P\left(\frac{\sigma_{un,s}^2-\sigma_{un}^2}{\sigma_{un,s}^2}-\rho_s > 0\right)\\ &\le P\left(\sup_{S\subseteq\{1,\ldots,N\}}\left(\frac{\sigma_{un,s}^2-\sigma_{un}^2}{\sigma_{un,s}^2}-\rho_s\right) > 0\right)\\ &= 0\end{aligned}$$

for every S, if the MRPV condition holds. Combining the above results, the conclusion follows.

By similar fashion, for comparing $\widehat{\mathbf{w}}_s^{*\mathbf{T}}\Sigma_{ss}\widehat{\mathbf{w}}_s^{*}$with $N^{-2}\mathbf{1}_N^{\mathbf{T}}\Sigma\mathbf{1}_N$, we can have

$$\begin{aligned}\widehat{\mathbf{w}}_s^{*\mathbf{T}}\Sigma_{ss}\widehat{\mathbf{w}}_s^{*}-\sigma_{\frac{1}{N}}^2 &= \widehat{\sigma}_s^2+\widehat{\mathbf{w}}_s^{*\mathbf{T}}\Omega_{ss}(n)\,\widehat{\mathbf{w}}_s^{*}-\sigma_{\frac{1}{N}}^2\\ &\le \hat{c}_{s,2}\widehat{\sigma}_{un,s}^2-\sigma_{\frac{1}{N}}^2+\left(\lambda_1\hat{c}_{s,3}-\lambda_2\hat{c}_{s,4}\right)-\lambda_1\left\|\widehat{\mathbf{w}}_s^{*}\right\|_1+\max_{i,j\in S}\left|\omega_{ij}(n)\right|\left\|\widehat{\mathbf{w}}_s^{*}\right\|_1^2 \qquad (1.49)\end{aligned}$$

Let

$$\mathcal{A}_{1,s}' = \left\{\hat{c}_{s,2}\widehat{\sigma}_{un,s}^2+\left(\lambda_1\hat{c}_{s,3}-\lambda_2\hat{c}_{s,4}\right)+\max_{i,j\in S}\left|\omega_{ij}(n)\right|\left\|\widehat{\mathbf{w}}_s^{*}\right\|_1^2 \le \sigma_{\frac{1}{N}}^2+\lambda_1\left\|\widehat{\mathbf{w}}_s^{*}\right\|_1\right\}.$$

$\mathcal{A}_{1,\widehat{S}_\lambda}'$ is a sufficient condition to make (1.49) non-positive. Define the following event,

$$\begin{aligned}\mathcal{B}_{1,s}' &= \left\{\lambda_1\hat{c}_{s,3}-\lambda_2\hat{c}_{s,4}\le B_{3,n}\right\},\\ \mathcal{B}_{3,s}' &= \left\{\hat{c}_{s,2}\widehat{\sigma}_{un,s}^2\le\sigma_{\frac{1}{N}}^2-B_{3,n}\right\}.\end{aligned}$$

where $B_{3,n}=\underline{\sigma}^2\left(N^{-1}\log N\right)^{\frac{1}{2}-\varepsilon}$. Note that $\sigma_{1/N}^2-B_{3,n}>0$, since by assumption that $0<\underline{\sigma}^2\left(N^{-1}\log N\right)^{\frac{1}{2}-\varepsilon}<\sigma_{\frac{1}{N}}^2$, where $\underline{\sigma^2}$ and $\varepsilon>0$ are two constants. It then can be shown that $\mathcal{B}_{1,s}'\cap\mathcal{B}_{2,s}\cap\mathcal{B}_{3,s}'\subseteq\mathcal{A}_{1,s}'$. For $P(\mathcal{B}_{1,s}'^{c})$, it follows that

$$P(\mathcal{B}_{1,s}'^{c}) = P\left(\hat{c}_{s,3}-\hat{c}_{s,4} > B_2'\left(\frac{n}{\log N}\right)^{\varepsilon}\right) \le P\left(\hat{c}_{s,3}-\hat{c}_{s,4} > B_2'\left(\frac{n}{\log n}\right)^{\varepsilon}\right),$$

where

$$B_2' = \underline{\sigma}^2 B_2\rho_N^{\varepsilon-\frac{1}{2}}$$

is a constant. Since $\hat{c}_{s,3}-\hat{c}_{s,4}=O_p(1)$, and $B_2'\left(n/\log n\right)^{\varepsilon}\to\infty$, $P(\mathcal{B}_{1,s}'^{c})$ converges to zero as n,

$N \to \infty$. By similar fashion for (1.47), as $n \to \infty$

$$\sup_{S \subseteq \{1,\ldots,N\}} P\left(\hat{c}_{s,3} - \hat{c}_{s,4} > B_2'\left(\frac{n}{\log n}\right)^{\varepsilon}\right) \to 0$$

since $\hat{c}_{s,3} - \hat{c}_{s,4} = O_p(1)$ holds for every S as n, $N \to \infty$. We then only need to show that

$$\sum_{S \subseteq \{1,\ldots,N\}} P\left(\mathcal{B}_{3,s}'^{c}\right) P\left(\hat{S} = S\right) \to 0.$$

Following the same strategy used in previous proof and setting $\lambda_1 = \lambda_2 = \lambda_{n,N}$, it can be shown that

$$\begin{aligned}
P\left(\hat{c}_{s,2}\hat{\sigma}_{un,s}^2 > \sigma_{\frac{1}{N}}^2 - B_{3,n}\right) &= P\left(\hat{c}_{s,2}\hat{\sigma}_{un,s}^2 - \sigma_{\frac{1}{N}}^2 > -B_{3,n}\right) \\
&\to P\left((1-\rho_s)\,\sigma_{un,s}^2 - \sigma_{\frac{1}{N}}^2 > 0\right) \\
&= P\left(\frac{\sigma_{un,s}^2 - \sigma_{\frac{1}{N}}^2}{\sigma_{un,s}^2} - \rho_s > 0\right) \\
&\leq P\left(\sup_{S \subseteq \{1,\ldots,N\}}\left(\frac{\sigma_{un,s}^2 - \sigma_{\frac{1}{N}}^2}{\sigma_{un,s}^2} - \rho_s\right) > 0\right) \\
&= 0
\end{aligned}$$

for every S, if the MRPV condition (1.18) holds. Therefore the proof is completed. ■

1.8.7 Derivation of the Coordinate Wise Descent Algorithm

From the KKT conditions, when the linear constraint is $\mathbf{w}^{\mathbf{T}}\mathbf{1}_N = 1$, by fixing w_j, $j = 1, \ldots, N$, $j \neq i$, one can solve w_i as

$$\frac{ST\left(\gamma - z_i, \lambda_1\right)}{2\left(\sigma_i^2 + \lambda_2\right)},$$

where $ST(x,y) = sign(x)\left(|x| - y\right)_+$ is the soft thresholding function and $z_i = 2\sum_{j\neq i}^{N} w_j \sigma_{ij}$. Let $S_+ = \{i : w_i > 0\}$ and $S_- = \{i : w_i < 0\}$. Then we know that

$$\begin{aligned}
\mathbf{w}^{\mathbf{T}}\mathbf{1}_N &= \gamma\left(\sum_{i\in S_+\cup S_-}\frac{1}{2\left(\sigma_i^2+\lambda_2\right)}\right) - \sum_{i\in S_+\cup S_-}\frac{z_i}{2\left(\sigma_i^2+\lambda_2\right)} + \\
&\quad \lambda_1\left(\sum_{i\in S_-}\frac{1}{2\left(\sigma_i^2+\lambda_2\right)} - \sum_{i\in S_+}\frac{1}{2\left(\sigma_i^2+\lambda_2\right)}\right).
\end{aligned}$$

Since $\mathbf{w}^{\mathbf{T}}\mathbf{1}_N = 1$, one can solve for γ as

$$\frac{1 + \sum_{i\in S_+\cup S_-}\frac{z_i}{2\left(\sigma_i^2+\lambda_2\right)} - \lambda_1\left(\sum_{i\in S_-}\frac{1}{2\left(\sigma_i^2+\lambda_2\right)} - \sum_{i\in S_+}\frac{1}{2\left(\sigma_i^2+\lambda_2\right)}\right)}{\left[\sum_{i\in S_+\cup S_-}\frac{1}{2\left(\sigma_i^2+\lambda_2\right)}\right]}.$$

To implement the algorithm, I set initial value of each weight $w_1^{(0)} = w_2^{(0)} = \cdots = w_p^{(0)} = N^{-1}$, and $\gamma^{(0)} > \lambda_1$. The algorithm starts from updating $w_1, w_2, \ldots$, and w_N sequentially, and then use the updated vector $\mathbf{w}$ to update γ. The procedure terminates until $\mathbf{w}$ and γ converge. The algorithm can be summarized as follows.

Algorithm 1 ***Coordinate-wise descent update for the weighted norm mvp optimization with the full investment constraint***

1. *Fix* λ_1 *and* λ_2 *at some constant levels.*
2. *Initialize* $\mathbf{w}^{(0)} = N^{-1}\mathbf{1}_N$ *and* $\gamma^{(0)} > \lambda_1$
3. *For* $i = 1, \ldots, N$*, and* $k > 0$*,*

$$w_i^{(k)} \leftarrow \frac{ST\left(\gamma^{(k-1)} - z_i^{(k)}, \lambda_1\right)}{2\left(\sigma_i^2 + \lambda_2\right)},$$

where

$$z_i^{(k)} = 2(\sum_{j<i} w_j^{(k)} \sigma_{ij} + \sum_{j>i} w_j^{(k-1)} \sigma_{ij}).$$

4. *For* $k > 0$*, update* γ *as*

$$\begin{aligned}\gamma^{(k)} \quad \leftarrow \quad & \left[\sum_{i \in S_+^{(k)} \cup S_-^{(k)}} \frac{1}{2\left(\sigma_i^2 + \lambda_2\left(1-\alpha\right)\right)}\right]^{-1} \times \\ & \left[1 + \sum_{i \in S_+^{(k)} \cup S_-^{(k)}} \frac{z_i^{(k)}}{2\left(\sigma_i^2 + \lambda_2\right)} - \right. \\ & \left. \lambda_1 \left(\sum_{i \in S_-^{(k)}} \frac{1}{2\left(\sigma_i^2 + \lambda_2\right)} - \sum_{i \in S_+^{(k)}} \frac{1}{2\left(\sigma_i^2 + \lambda_2\right)}\right)\right],\end{aligned}$$

where $S_+^{(k)} = \left\{i : w_i^{(k)} > 0\right\}$ *and* $S_- = \left\{i : w_i^{(k)} < 0\right\}$.

5. *Repeat 3 and 4 until* $\mathbf{w}^{(k)}$ *and* $\gamma^{(k)}$ *converge.*

1.8.8 More Discussions on the Stochastic Dominance Test

The formal definitions for the FSD and SSD are as follows.

Definition 2 *Let* $u(.)$ *be an nondecreasing (*$u'(.) \geq 0$ *) von Neumann-Morgenstern utility function, and* $F_1(r)$ *and* $F_2(r)$ *be the cumulative distribution functions (c.d.f.) of random variables* R_1 *and* R_2 *respectively.* R_2 *is first order stochastic dominated by* R_1*, i.e.* $R_1 \succeq_{FSD} R_2$*, if and only if*

$$\mathbb{E}\left(u\left(R_1\right)\right) \geq \mathbb{E}\left(u\left(R_2\right)\right),$$

for all $u(.)$ *and with strict inequality for some* $u(.)$; *or*

$$F_1(r) \leq F_2(r),$$

for all r *and with strict inequality for some* r.

Definition 3 *Let* $u(.)$ *be an nondecreasing* ($u'(.) \geq 0$) *and concave* ($u"(.) \leq 0$) *von Neumann-Morgenstern utility function, and* $F_1(r)$ *and* $F_2(r)$ *be the cumulative distribution functions (c.d.f.) of random variables* R_1 *and* R_2 *respectively.* R_2 *is second order stochastic dominated by* R_1*, i.e.* $R_1 \succeq_{SSD} R_2$*, if and only if*

$$\mathbb{E}(u(R_1)) \geq \mathbb{E}(u(R_2)),$$

for all $u(.)$ *and with strict inequality for some* $u(.)$; *or*

$$\int_{-\infty}^{r} F_1(x)\,dx \leq \int_{-\infty}^{r} F_2(x)\,dx,$$

for all r *and with strict inequality for some* r.

Note that the above definition of SSD does not require the property of equal mean of R_1 and R_2. Does it matter? To see this,

$$\begin{aligned}
\mathbb{E}(u(R_1)) - \mathbb{E}(u(R_2)) &= \int_{-\infty}^{\infty} u(r)\,dF_1(r) - \int_{-\infty}^{\infty} u(r)\,dF_1(r) \\
&= \int_{-\infty}^{\infty} u'(r)(F_2(r) - F_1(r))\,dr \\
&= u'(r)\int_{-\infty}^{r} (F_2(x) - F_1(x))\,dx\Big|_{-\infty}^{\infty} \\
&\quad - \int_{-\infty}^{\infty} u"(r)\left(\int_{-\infty}^{r} (F_2(x) - F_1(x))\,dx\right) dr.
\end{aligned}$$

Clearly, if the second condition in definition 2 holds, and $u(r)$ is nondecreasing and concave, $\mathbb{E}(u(R_1)) \geq \mathbb{E}(u(R_2))$ no matter whether R_1 and R_2 have the same mean or not.

The concepts of FSD and SSD state whether one portfolio strategy can generate higher expected utility than another, and also whether the portfolio strategy can be less risky than another. If $R_1 \succeq_{FSD} R_2$ holds, it is equivalent to saying that every expected utility maximizer will prefer $F_1(r)$ to $F_2(r)$. On the other hand, $F_1(r)$ clearly deliver higher expected utility than $F_2(r)$. If $R_1 \succeq_{SSD} R_2$ holds, it is equivalent to saying that every risk-averse expected utility maximizer will prefer $F_1(r)$ to $F_2(r)$. Or we can say $F_1(r)$ is less risky than $F_2(r)$.

To see whether one random variable first or second order stochastic dominates the other random variable, one can implement some formal statistical tests via comparing functionals of their c.d.f's. We can empirically estimate the c.d.f. by

$$\hat{F}_i(r) = \frac{1}{T}\sum_{t=1}^{T} 1\{R_{it} \leq r\},$$

$i = 1, 2$. Let $\Delta_{1,2}^{(1)}(r) := F_1(r) - F_2(r)$, and $\delta_{(1)}^{*} := \sup_r \left(\Delta_{1,2}^{(1)}(r)\right)$. To test whether $R_1 \succeq_{FSD} R_2$, we can form a null hypothesis as the following

$$H_0 : \delta_{(1)}^{*} \leq 0.$$

The above null hypothesis states that for all r, $F_1(r) \leq F_2(r)$. On the other hand, if we cannot reject the null, there is not enough evidence to say that $R_1 \succeq_{FSD} R_2$ does not hold. The empirical analogue of $\delta^*_{(1)}$ can be

$$\hat{\delta}^*_{(1)} = \sup_r \sqrt{T}\left(\hat{\Delta}^{(1)}_{1,2}(r)\right),$$

where $\hat{\Delta}^{(1)}_{1,2}(r) = \hat{F}_1(r) - \hat{F}_2(r)$.

Let $CF_i(r) := \int_{-\infty}^{r} F_i(x)\,dx$. For testing SSD, at first note that by intergating by parts, given $F_i(-\infty) := 0$, it can be shown that

$$\begin{aligned} CF_i(r) &= F_i(r)\,r - \int_{-\infty}^{r} x dF_i(x) \\ &= \int_{-\infty}^{r} r dF_i(x) - \int_{-\infty}^{r} x dF_i(x) \\ &= \int_{-\infty}^{r} (r - x)\,dF_i(x). \end{aligned}$$

Therefore we can empirically estimate $CF_i(r)$ by

$$\hat{CF}_i(r) = \frac{1}{T}\sum_{t=1}^{T}(r - R_{it})\,1\{R_{it} \leq r\}.$$

Let $\Delta^{(2)}_{1,2}(r) := CF_1(r) - CF_2(r)$, and $\delta^*_{(2)} := \sup_r\left(\Delta^{(2)}_{1,2}(r)\right)$. To test whether $R_1 \succeq_{SSD} R_2$, we can form a null hypothesis as the following

$$H_0 : \delta^*_{(2)} \leq 0.$$

The above null hypothesis states that for all r, $CF_1(r) \leq CF_2(r)$. On the other hand, if we cannot reject the null, there is not enough evidence to say that $R_1 \succeq_{SSD} R_2$ does not hold. The empirical analogue of $\delta^*_{(2)}$ can be

$$\hat{\delta}^*_{(2)} = \sup_r \sqrt{T}\left(\hat{\Delta}^{(2)}_{1,2}(r)\right),$$

where $\hat{\Delta}^{(2)}_{1,2}(r) = \hat{CF}_1(r) - \hat{CF}_2(r)$. Let

$$\begin{aligned} \underline{R} &= \min(R_{11}, \ldots, R_{1T}, R_{21}, \ldots, R_{2T}), \\ \overline{R} &= \max(R_{11}, \ldots, R_{1T}, R_{21}, \ldots, R_{2T}). \end{aligned}$$

To numerically evaluate $\hat{\delta}^*_{(1)}$and $\hat{\delta}^*_{(2)}$, we divide the interval $\left[\underline{R}, \bar{R}\right]$ into 200 equally spaced grids and search the value of $r \in \left[\underline{R}, \bar{R}\right]$ over the grids to maximize$\hat{\Delta}^{(1)}_{1,2}(r)$ or $\hat{\Delta}^{(2)}_{1,2}(r)$

We are interested in is whether the weighted norm mvp first (or second) order stochastic dominates the other three strategies ($1/N$, no-shortsales and GMVP). Formally, it can be stated as $R_\alpha \succeq_{FSD} R_l$ (or $R_\alpha \succeq_{SSD} R_l$), where R_α is return of the weighted norm mvp with parameter value α and R_l is return of strategy l, $l = 1/N$, no-shortsales and GMVP. To empirically construct critical values and p-values, I adopt subsampling method suggested by Linton et al. (2005). I briefly describe the scheme as follows. Suppose we have T realized return observations, $R_1, R_2 \ldots, R_T$. The test statistic, $\sqrt{T}\theta_T(R_1, R_2 \ldots, R_T)$, is a function of the T observations, and is determined by $\sup_r \hat{\Delta}^{(1)}_{1,2}(r)$ for the FSD test and by $\sup_r \hat{\Delta}^{(2)}_{1,2}(r)$ for the SSD test. Let $G_T(x) = P\left(\sqrt{T}\theta_T(R_1, R_2 \ldots, R_T) \leq x\right)$ be the distribution function of the test statistic. Following Linton et al. (2005), I approximate the distribution of $G_T(x)$by

$$\hat{G}_{T,b}(x) = \frac{1}{T-b+1}\sum_{t=1}^{T-b+1} 1\left\{\sqrt{b}\theta_{T,b,t} \le x\right\},$$

where $\theta_{T,b,t} := \theta_b(R_t, R_{t+1}\ldots, R_{t+b-1})$, $t = 1,\ldots,T-b+1$, is just the function of θ evaluated with subsample $R_t, R_{t+1}\ldots, R_{t+b-1}$. Note that each $\sqrt{b}\theta_{T,b,t}$ is the test statistic for the FSD (or SSD test) obtained from b subsamples $R_t, R_{t+1}\ldots, R_{t+b-1}$, and it can be seen that the subsampling scheme is essentially very similar as the rolling window scheme. With this approximation, I then define the subsample critical value at significant level δ as $\hat{g}_{T,b}(\delta) = \inf\left\{x : \delta \ge 1 - \hat{G}_{T,b}(x)\right\}$, and the subsample p-values $\hat{p}_{T,b} = 1 - \hat{G}_{T,b}\left(\sqrt{T}\theta_T\right)$. The decision rule is that we reject the null if $\sqrt{T}\theta_T > \hat{g}_{T,b}(\delta)$ or $\hat{p}_{T,b} < \delta$.

1.8.9 Coordinate-Wise Descent on MVP Penalized by the Generalized l_1 Penalty

The subsection provides a derivation of the coordinate-wise descent algorithm for solving the mvp penalized by the generalized l_1 norm penalty in Section 1.6.7. As the linear constraint is the full investment constraint, the penalized mvp optimization is given by

$$\min_{\mathbf{w}} \mathbf{w}^{\mathbf{T}}\Sigma\mathbf{w} + \lambda_1\|\mathbf{w}\|_1 + \lambda_2\|\mathbf{w}-\mathbf{w}_0\|_1, \text{ subject to } \mathbf{w}^{\mathbf{T}}\mathbf{1} = 1.$$

Let $w_{0,i} \ge 0$, $i = 1,\ldots,N$ denote elements in $\mathbf{w}_0$. At the stationary point, the following subgradient equation should hold

$$2\sigma_i^2 w_i + 2\sum_{j\neq i}^{N}\sigma_{ij}w_j + \lambda_1 sign(w_i) + \lambda_2 sign(w_i - w_{0,i}) - \gamma = 0,$$

for $i = 1,\ldots,N$, and also $\mathbf{w}^{\mathbf{T}}\mathbf{1} = 1$. Again, let $z_i = \sum_{j\neq i}^{N}\sigma_{ij}w_j$. By fixing w_j, $j \neq i$, one can solve w_i as

$$w_i = \begin{cases} \frac{\gamma - z_i - (\lambda_1+\lambda_2)}{2\sigma_i^2} & \text{if } \gamma - z_i > (\lambda_1+\lambda_2) + 2\sigma_i^2 w_{i,0}, \\ w_{i,0} & \text{if } (\lambda_1-\lambda_2) + 2\sigma_i^2 w_{i,0} \le \gamma - z_i \le (\lambda_1+\lambda_2) + 2\sigma_i^2 w_{i,0}, \\ \frac{\gamma - z_i - (\lambda_1-\lambda_2)}{2\sigma_i^2} & \text{if } \lambda_1 - \lambda_2 < \gamma - z_i < (\lambda_1-\lambda_2) + 2\sigma_i^2 w_{i,0}, \\ 0 & \text{if } -\lambda_1 - \lambda_2 \le \gamma - z_i \le \lambda_1 - \lambda_2, \\ \frac{\gamma - z_i + (\lambda_1+\lambda_2)}{2\sigma_i^2} & \text{if } \gamma - z_i < -\lambda_1 - \lambda_2. \end{cases}$$

Let $\Delta_1 = \{i : w_{i,0} < w_i < \infty\}$, $\Delta_2 = \{i : w_i = w_{i,0}\}$, $\Delta_3 = \{i : 0 < w_i < w_{i,0}\}$, and $\Delta_4 = \{i : -\infty < w_i < 0\}$. One can solve γ by using the full investment constraint,

$$\gamma = \frac{1 - \sum_{i\in\Delta_2} w_{i,0} + \sum_{i\in\Delta_1}\frac{z_i+(\lambda_1+\lambda_2)}{2\sigma_i^2} + \sum_{i\in\Delta_3}\frac{z_i+(\lambda_1-\lambda_2)}{2\sigma_i^2} + \sum_{i\in\Delta_4}\frac{z_i-(\lambda_1+\lambda_2)}{2\sigma_i^2}}{\sum_{i\in\Delta_1\cup\Delta_3\cup\Delta_4}\frac{1}{2\sigma_i^2}}.$$

The algorithm can be summarized as follows.

Algorithm 2 ***Coordinate-wise descent update for mvp penalized by the generalized*** l_1 *penalty.*

1. *Fix λ_1 and λ_2 at some constant levels.*

2. *Initialize* $\mathbf{w}^{(0)} = N^{-1}\mathbf{1}_N$ *and* $\gamma^{(0)} > \max(\lambda_1, \lambda_2)$

3. *For* $i = 1, \ldots, N,$ *and* $k > 0,$

$$w_i^{(k)} \leftarrow \begin{cases} \frac{\gamma^{(k-1)} - z_i^{(k)} - (\lambda_1 + \lambda_2)}{2\sigma_i^2} & \text{if } \gamma^{(k-1)} - z_i^{(k)} > (\lambda_1 + \lambda_2) + 2\sigma_i^2 w_{i,0}, \\ w_{i,0} & \text{if } (\lambda_1 - \lambda_2) + 2\sigma_i^2 w_{i,0} \leq \gamma^{(k-1)} - z_i^{(k)} \leq (\lambda_1 + \lambda_2) + 2\sigma_i^2 w_{i,0}, \\ \frac{\gamma^{(k-1)} - z_i^{(k)} - (\lambda_1 - \lambda_2)}{2\sigma_i^2} & \text{if } \lambda_1 - \lambda_2 < \gamma^{(k-1)} - z_i^{(k)} < (\lambda_1 - \lambda_2) + 2\sigma_i^2 w_{i,0}, \\ 0 & \text{if } -\lambda_1 - \lambda_2 \leq \gamma^{(k-1)} - z_i^{(k)} \leq \lambda_1 - \lambda_2, \\ \frac{\gamma^{(k-1)} - z_i^{(k)} + (\lambda_1 + \lambda_2)}{2\sigma_i^2} & \text{if } \gamma^{(k-1)} - z_i^{(k)} < -\lambda_1 - \lambda_2. \end{cases}$$

where

$$z_i^{(k)} = 2\left(\sum_{j<i} w_j^{(k)} \sigma_{ij} + \sum_{j>i} w_j^{(k-1)} \sigma_{ij}\right).$$

4. *For* $k > 0,$ *update* γ *as*

$$\begin{aligned} \gamma^{(k)} \leftarrow & \left[\sum_{i \in \Delta_1^{(k)} \cup \Delta_3^{(k)} \cup \Delta_4^{(k)}} \frac{1}{2\sigma_i^2} \right]^{-1} \times \\ & \left[1 - \sum_{i \in \Delta_2^{(k)}} w_{i,0} + \sum_{i \in \Delta_1^{(k)}} \frac{z_i^{(k)} + (\lambda_1 + \lambda_2)}{2\sigma_i^2} \right. \\ & \left. + \sum_{i \in \Delta_3^{(k)}} \frac{z_i^{(k)} + (\lambda_1 - \lambda_2)}{2\sigma_i^2} + \sum_{i \in \Delta_4^{(k)}} \frac{z_i^{(k)} - (\lambda_1 + \lambda_2)}{2\sigma_i^2} \right], \end{aligned}$$

where

$$\begin{aligned} \mathbf{\Delta}_1^{(k)} &= \left\{ i : \gamma^{(k-1)} - z_i^{(k)} > (\lambda_1 + \lambda_2) + 2\sigma_i^2 w_{i,0} \right\}, \\ \mathbf{\Delta}_2^{(k)} &= \left\{ i : (\lambda_1 - \lambda_2) + 2\sigma_i^2 w_{i,0} \leq \gamma^{(k-1)} - z_i^{(k)} \leq (\lambda_1 + \lambda_2) + 2\sigma_i^2 w_{i,0} \right\}, \\ \mathbf{\Delta}_3^{(k)} &= \left\{ i : -\lambda_1 - \lambda_2 \leq \gamma^{(k-1)} - z_i^{(k)} \leq \lambda_1 - \lambda_2, \right\}, \\ \mathbf{\Delta}_4^{(k)} &= \left\{ i : \gamma^{(k-1)} - z_i^{(k)} < -\lambda_1 - \lambda_2. \right\}, \end{aligned}$$

5. *Repeat 3 and 4 until* $\mathbf{w}^{(k)}$ *and* $\gamma^{(k)}$ *converge.*

1.8.10 The Multistage Portfolio Optimization and the l_0 norm Penalty

In the following, I show that the multistage portfolio optimization in section 1.6.7 can be viewed as using the majorization-minimization method to approximately solve a l_0 norm penalized portfolio optimization. The l_0 norm of $\mathbf{w}$ is defined as $\|\mathbf{w}\|_0 = \sum_{i=1}^N |w_i|^0$, where

$$|w_i|^0 := \begin{cases} 1 & \text{if } w_i \neq 0 \\ 0 & \text{if } w_i = 0 \end{cases}.$$

Therefore to penalize the portfolio weights by the l_0 norm is equivalent to restricting the number of assets included in a portfolio. The penalized mvp optimization is given by

$$\min_{\mathbf{w}} \mathbf{w}^{\mathbf{T}}\Sigma\mathbf{w} + \lambda \left\|\mathbf{w}\right\|_0, \text{ subject to } A\mathbf{w} = \mathbf{u}, \tag{1.50}$$

As mentioned in the beginning of the paper, an optimization problem involved with the l_0 norm penalty in practice is difficult to solve. To make the problem tractable, one may approximate the l_0 norm penalty by

$$AP\left(\mathbf{w}{,}\epsilon\right) = \sum_{i=1}^{N} 1 - \frac{1}{\exp\left(\epsilon\left|w_i\right|\right)}.$$

and then solve the optimization with the approximated penalty $AP\left(\mathbf{w}{,}\epsilon\right)$. It can be seen that $AP\left(\mathbf{w}{,}\epsilon\right)$ converges to $\left\|\mathbf{w}\right\|_0$ as ϵ goes large,

$$\lim_{\epsilon\to\infty} AP\left(\mathbf{w}{,}\epsilon\right) = \left\|\mathbf{w}\right\|_0.$$

Then (1.50) becomes

$$\min_{\mathbf{w}} \left(\lim_{\epsilon\to\infty} \mathbf{w}^{\mathbf{T}}\Sigma\mathbf{w} + \lambda AP\left(\mathbf{w}{,}\epsilon\right)\right), \text{ subject to } A\mathbf{w} = \mathbf{u},. \tag{1.51}$$

However, we now meet another difficulty. Since $AP\left(\mathbf{w}{,}\epsilon\right)$ is concave in w_i, the objective function is not guaranteed to be a convex function of $\mathbf{w}$, thus a coordinate-descent type algorithm is not applicable here. To circumstance this, one can adopt the majorization-minimization approach. A real valued function $f\left(x,y\right)$ is said to majorize a real valued function $g\left(x\right)$ at point y if

$$\begin{aligned} f\left(x,y\right) &\geq g\left(x\right) \text{ for all } x,\ y\in\mathbb{R}, \\ f\left(y,y\right) &= g\left(y\right) \text{ for all } x\in\mathbb{R}. \end{aligned}$$

Suppose $x = y^*$ minimizes $f\left(x,y\right)$, then

$$\begin{aligned} g\left(y^*\right) &= f\left(y^*,y\right) + g\left(y^*\right) - f\left(y^*,y\right) \\ &\leq f\left(y^*,y\right) + g\left(y\right) - f\left(y,y\right) \\ &\leq g\left(y\right). \end{aligned}$$

Now let $y^* = v^{(l+1)}$, and $y = v^{(l)}$, then

$$g\left(v^{(l+1)}\right) \leq g\left(v^{(l)}\right).$$

That says, if one wants to find a sequence of $v^{(l)}$ to decrease the function $g\left(x\right)$, one can achieve this by sequentially minimizing its majorization function $f\left(x,v^{(l)}\right)$ with respect to x,

$$v^{(l+1)} = \arg\min_x f\left(x,v^{(l)}\right),$$

An algorithm which sequentially minimizes a majorization function of a certain objective function in order to find its global minimizer is called the minimization-majorization (MM) algorithm. For solving (1.51), we can try to find a majorization function of the objective function of (1.51), and cast the MM algorithm to find its global minimizer. However, the majorization should be convex for $\mathbf{w}$, otherwise we still will have the same difficulty as we have in solving (1.51).

Since $\mathbf{w^T}\Sigma\mathbf{w}$ is already a convex function of $\mathbf{w}$, one can just find a convex majorization function of $AP\left(\mathbf{w},\epsilon\right)$ to replace $AP\left(\mathbf{w},\epsilon\right)$, and the new objective function will be a convex function of $\mathbf{w}$. Note that for all $x \geq 0$ and $\epsilon > 0$, $1 - \exp\left(-\epsilon x\right)$ is a concave function of x . And it can shown that

$$1 - \exp\left(-\epsilon x\right) \leq 1 - \exp\left(-\epsilon y\right) + \epsilon \exp\left(-\epsilon y\right)\left(x - y\right),$$

for all $x, y \geq 0$ and $\epsilon > 0$. The right hand side of the above inequality is a linear function of x which is tangent to the graph of $1 - \exp\left(-\epsilon x\right)$ at the point y. Let

$$APM\left(\mathbf{w}, \mathbf{w}^{*(l)}, \epsilon\right) = \sum_{i=1}^{N}\left[1 - \frac{1}{\exp\left(\epsilon\left|w_i^{*(l)}\right|\right)} + \frac{\epsilon}{\exp\left(\epsilon\left|w_i^{*(l)}\right|\right)}\left(|w_i| - \left|w_i^{*(l)}\right|\right)\right].$$

It can be seen that $APM\left(\mathbf{w}, \mathbf{w}^{*(l)}, \epsilon\right)$ majorizes $AP\left(\mathbf{w},\epsilon\right)$ at point $\mathbf{w}^{*(l)}$, and it is also a convex function of $\mathbf{w}$. Therefore the global minimizer of $\mathbf{w^T}\Sigma\mathbf{w} + \lambda AP\left(\mathbf{w},\epsilon\right)$ can be obtained by sequentially solving

$$\min_{\mathbf{w}} \mathbf{w^T}\Sigma\mathbf{w} + \lambda APM\left(\mathbf{w}, \mathbf{w}^{*(l)}, \epsilon\right), \text{ subject to } A\mathbf{w} = \mathbf{u},$$

where

$$\mathbf{w}^{*(l)} = \arg\min_{\mathbf{w}, A\mathbf{w}=\mathbf{u}} \mathbf{w^T}\Sigma\mathbf{w} + \lambda APM\left(\mathbf{w}, \mathbf{w}^{*(l)}, \epsilon\right),$$

which is equivalent to solving

$$\min_{\mathbf{w}} \mathbf{w^T}\Sigma\mathbf{w} + \lambda\sum_{i=1}^{N}\frac{\epsilon}{\exp\left(\epsilon\left|w_i^{*(l)}\right|\right)}\left|w_i\right|, \text{ subject to } A\mathbf{w} = \mathbf{u},$$

sequentially. As only the full investment constraint is imposed, the above optimization can be easily solved by using algorithm 1 with $\lambda_1 = \epsilon\exp\left(-\epsilon\left|w_i^{*(l)}\right|\right)$ and $\lambda_2 = 0$.

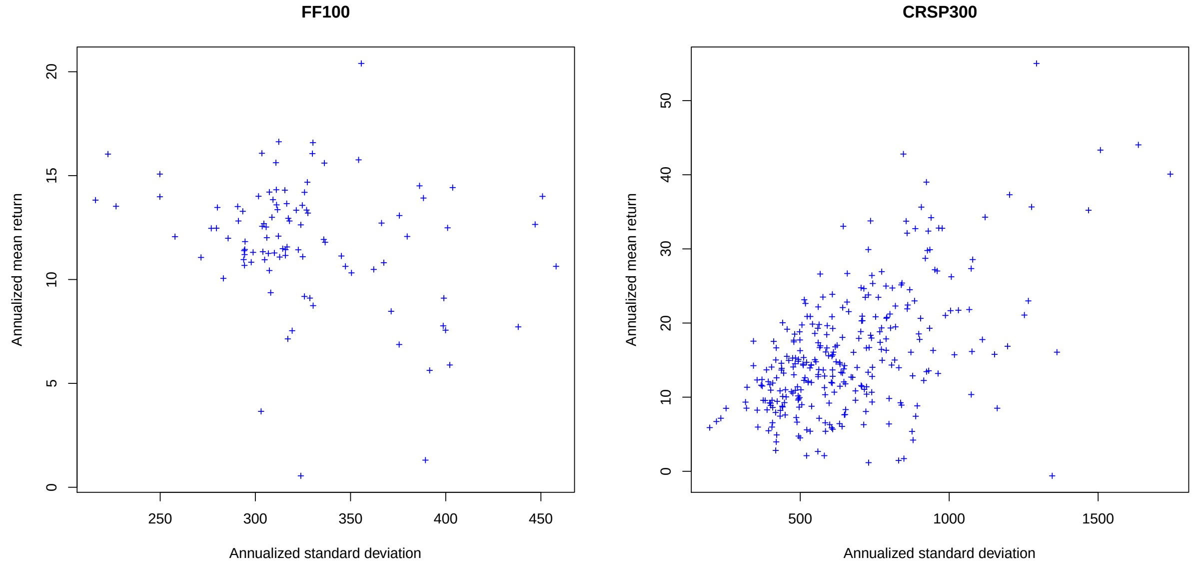

Figure 1.1: The figure shows annualized mean returns (%) and standard deviations (%) of individual assets in FF100 and CRSP300. The mean returns and standard deviations are calculated with daily data over the whole sample period. The sample period for FF100 is from July-12-1987 to Dec-31-2010 (5,415 observations), and for CRSP300 is from July-30-1998 to Dec-31-2010 (3,127 observations).

Table 1.1: The table shows results of performances for daily FF100 data when the full investment constraint $\mathbf{w}^{\mathbf{T}}\mathbf{1} = 1$ is imposed. We estimate Σ by sample covariance matrix with expanding window scheme. Here $N = 100$ and initial window length $\tau_0 = 120$. For the weighted norm constraint, we set $\lambda_1 = \lambda_{1,t}$, $\lambda_1 = \lambda_{2,t}$ and vary α at six different levels. Testing period is from Jan-02-1990 to Dec-31-2010 and $T = 5,295$. SV denotes sample variance of out-of-sample net returns (when the transaction fees are deducted). The sample variance is annualized. The transaction fee we consider is 35 basis points. SR denotes annualized Sharpe ratio, and the yearly risk free rate used is 3.63%. Certainty equivalence is obtained with $\psi = 5$. Column TOR and PAC show average turnover rate and average proportion of active assets, respectively. Column HHI and ANHHI show average values of Herfindahl–Hirschman index, and average values of adjusted normalized Herfindahl–Hirschman index. Column SLR shows average short-long ratio. In the parentheses are the bootstrap standard errors obtained from using stationary bootstrap of Politis and Romano (1994).

	SV(%)	SR	CE(%)	TOR	PAC	HHI	ANHHI	SLR
$\alpha = 0$	**82.9718**	**1.0666**	**0.0451**	**0.0407**	1.0000	0.0202	0.0103	0.5696
	(14.6744)	(0.4704)	(0.0158)	(0.0034)	(0.0000)	(0.0004)	(0.0005)	(0.0037)
$\alpha = 0.2$	**86.6844**	**1.1442**	**0.0485**	**0.0301**	0.6496	0.0323	0.0170	0.4713
	(16.0203)	(0.4721)	(0.0160)	(0.0026)	(0.0077)	(0.0007)	(0.0008)	(0.0050)
$\alpha = 0.4$	**90.3660**	**1.1752**	**0.0502**	**0.0257**	0.5207	0.0422	0.0230	0.4163
	(17.1536)	(0.4709)	(0.0162)	(0.0022)	(0.0091)	(0.0011)	(0.0012)	(0.0054)
$\alpha = 0.6$	**93.4194**	**1.1827**	**0.0509**	**0.0231**	0.4316	0.0515	0.0284	0.3758
	(17.9204)	(0.4697)	(0.0164)	(0.0020)	(0.0083)	(0.0015)	(0.0015)	(0.0057)
$\alpha = 0.8$	**96.1646**	**1.1861**	**0.0514**	**0.0215**	0.3714	0.0608	0.0340	0.3432
	(18.5558)	(0.4682)	(0.0166)	(0.0019)	(0.0078)	(0.0018)	(0.0018)	(0.0061)
$\alpha = 1$	**98.6034**	**1.1847**	**0.0517**	**0.0202**	0.3204	0.0700	0.0389	0.3173
	(19.0446)	(0.4648)	(0.0167)	(0.0019)	(0.0072)	(0.0023)	(0.0023)	(0.0063)
N.S.	176.2114	0.7409	0.0362	0.0082	**0.1049**	0.1650	0.0709	0.0000
	(38.2327)	(0.3494)	(0.0182)	(0.0012)	(0.0034)	(0.0075)	(0.0048)	(0.0000)
$1/N$	311.1636	0.4485	0.0150	**0.0051**	1.0000	0.0100	0.0000	0.0000
	(67.1655)	(0.2488)	(0.0195)	(0.0004)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
GMVP	**77.6752**	0.9444	0.0400	0.0726	1.0000	0.0212	0.0113	0.6263
	(12.0891)	(0.4802)	(0.0160)	(0.0133)	(0.0000)	(0.0005)	(0.0005)	(0.0053)

Table 1.2: The table shows results of performances for daily CRSP300 data when the full investment constraint $\mathbf{w}^{\mathbf{T}}\mathbf{1} = 1$ is imposed. We estimate Σ by sample covariance matrix with expanding window scheme. Here $N = 300$ and initial window length $\tau_0 = 360$. For the weighted norm constraint, we set $\lambda_1 = \lambda_{1,t}$, $\lambda_1 = \lambda_{2,t}$ and vary α at six different levels. Testing period is from Jan-03-2000 to Dec-31-2010 and $T = 2,767$. SV denotes sample variance of out-of-sample net returns (when the transaction fees are deducted). The sample variance is annualized. The transaction fee we consider is 35 basis points. SR denotes annualized Sharpe ratio, and the yearly risk free rate used is 2.50%. Certainty equivalence is obtained with $\psi = 5$. Column TOR and PAC show average turnover rate and average proportion of active assets, respectively. Column HHI and ANHHI show average values of Herfindahl–Hirschman index, and average values of adjusted normalized Herfindahl–Hirschman index. Column SLR shows average short-long ratio. In the parentheses are the bootstrap standard errors obtained from using stationary bootstrap of Politis and Romano (1994).

	SV(%)	SR	CE(%)	TOR	PAC	HHI	ANHHI	SLR
$\alpha = 0$	**50.7208**	**0.5576**	**0.0208**	0.0468	1.0000	0.0132	0.0099	0.3539
	(15.2769)	(0.4650)	(0.0124)	(0.0052)	(0.0000)	(0.0003)	(0.0003)	(0.0084)
$\alpha = 0.2$	**49.7156**	**0.8009**	**0.0276**	0.0320	0.7619	0.0221	0.0178	0.2361
	(16.4991)	(0.4928)	(0.0120)	(0.0030)	(0.0077)	(0.0005)	(0.0005)	(0.0092)
$\alpha = 0.4$	**50.8529**	**0.8865**	**0.0302**	0.0267	0.6417	0.0289	0.0238	0.1795
	(17.5011)	(0.5073)	(0.0121)	(0.0025)	(0.0095)	(0.0008)	(0.0007)	(0.0092)
$\alpha = 0.6$	**51.9853**	**0.9265**	**0.0315**	0.0236	0.5544	0.0344	0.0285	0.1440
	(18.2561)	(0.5122)	(0.0122)	(0.0023)	(0.0096)	(0.0009)	(0.0009)	(0.0086)
$\alpha = 0.8$	**52.9633**	**0.9514**	**0.0324**	0.0215	0.4911	0.0390	0.0324	0.1198
	(18.8245)	(0.5134)	(0.0122)	(0.0020)	(0.0085)	(0.0011)	(0.0010)	(0.0079)
$\alpha = 1$	**53.8325**	**0.9705**	**0.0331**	0.0201	0.4499	0.0431	0.0359	0.1021
	(19.2547)	(0.5152)	(0.0123)	(0.0019)	(0.0076)	(0.0012)	(0.0011)	(0.0074)
N.S.	67.4157	**1.1038**	**0.0395**	0.0126	0.2054	0.0734	0.0574	0.0000
	(24.1749)	(0.5062)	(0.0134)	(0.0010)	(0.0067)	(0.0014)	(0.0010)	(0.0000)
$1/N$	404.4271	0.6587	0.0225	0.0157	1.0000	0.0033	0.0000	0.0000
	(112.3455)	(0.3344)	(0.0293)	(0.0009)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
GMVP	52.3428	0.4147	0.0168	0.0537	1.0000	0.0144	0.0111	0.3673
	(15.3964)	(0.4513)	(0.0127)	(0.0080)	(0.0000)	(0.0003)	(0.0003)	(0.0104)

Table 1.3: The table shows p-values of the stochastic dominance tests proposed by Linton et al. (2005). The data used here is the realized daily net returns (when the transaction fees are deducted) of FF100 as different portfolio strategies are used. The transaction fee we consider is 35 basis points. Testing period is from Jan-02-1990 to Dec-31-2010 and $T = 5,295$. FSD and SSD denote first and second order stochastic dominances, respectively. The p-values are obtained from the subsampling method, and the subsample size is set to 300.

	$1/N$		N.S.		GMVP	
	FSD	SSD	FSD	SSD	FSD	SSD
$\alpha = 0$	0.0000	0.4235	0.0000	0.4582	0.7714	0.1229
$\alpha = 0.2$	0.0000	0.3881	0.0000	0.4023	0.5783	0.1111
$\alpha = 0.4$	0.0000	0.3811	0.0000	0.3805	0.6769	0.1087
$\alpha = 0.6$	0.0000	0.3755	0.0000	0.3765	0.4658	0.1063
$\alpha = 0.8$	0.0000	0.3705	0.0000	0.3717	0.4660	0.1019
$\alpha = 1$	0.0000	0.3639	0.0000	0.3647	0.3004	0.0915

Table 1.4: The table shows p-values of the stochastic dominance tests proposed by Linton et al. (2005). The data used here is the realized daily net returns (when the transaction fees are deducted) of CRSP300 as different portfolio strategies are used. The transaction fee we consider is 35 basis points. Testing period is from Jan-03-2000 to Dec-31-2010 and $T = 2,767$. FSD and SSD denote first and second order stochastic dominances, respectively. The p-values are obtained from the subsampling method, and the subsample size is set to 300.

	$1/N$		N.S.		GMVP	
	FSD	SSD	FSD	SSD	FSD	SSD
$\alpha = 0$	0.0000	0.0539	0.1803	0.0960	0.5985	0.7848
$\alpha = 0.2$	0.0000	0.0636	0.0045	0.0985	0.1864	0.5733
$\alpha = 0.4$	0.0000	0.0827	0.0446	0.1102	0.3165	0.3780
$\alpha = 0.6$	0.0000	0.0879	0.1001	0.1179	0.4652	0.3383
$\alpha = 0.8$	0.0000	0.0891	0.0713	0.1216	0.4344	0.3428
$\alpha = 1$	0.0000	0.0900	0.0928	0.1240	0.9955	0.3408

Table 1.5: The table shows results of performances for weekly FF100 data when the full investment constraint $\mathbf{w}^{\mathbf{T}}\mathbf{1} = 1$ is imposed. We estimate Σ by sample covariance matrix with expanding window scheme. Here $N = 100$ and initial window length $\tau_0 = 120$. For the weighted norm constraint, we set $\lambda_1 = \lambda_{1,t}$, $\lambda_1 = \lambda_{2,t}$ and vary α at six different levels. Testing period is from from first week of 1990 to the last week of 2010 and $T = 1,095$. SV denotes sample variance of out-of-sample net returns (when the transaction fees are deducted). The sample variance is annualized. The transaction fee we consider is 35 basis points. SR denotes annualized Sharpe ratio, and the yearly risk free rate used is 3.63%. Certainty equivalence is obtained with $\psi = 5$. Column TOR and PAC show average turnover rate and average proportion of active assets, respectively. Column HHI and ANHHI show average values of Herfindahl–Hirschman index, and average values of adjusted normalized Herfindahl–Hirschman index. Column SLR shows average short-long ratio. In the parentheses are the bootstrap standard errors obtained from using stationary bootstrap of Politis and Romano (1994).

	SV(%)	SR	CE(%)	TOR	PAC	HHI	ANHHI	SLR
$\alpha = 0$	**122.1502**	**0.9108**	**0.2047**	**0.0934**	1.0000	0.0162	0.0063	0.6038
	(26.1549)	(0.4318)	(0.0850)	(0.0082)	(0.0000)	(0.0003)	(0.0003)	(0.0020)
$\alpha = 0.2$	**128.6224**	**0.9074**	**0.2059**	**0.0686**	0.5393	0.0317	0.0134	0.4608
	(28.5429)	(0.4291)	(0.0865)	(0.0071)	(0.0059)	(0.0011)	(0.0010)	(0.0032)
$\alpha = 0.4$	**135.7170**	**0.9254**	**0.2119**	**0.0588**	0.3781	0.0461	0.0200	0.3801
	(30.5520)	(0.4167)	(0.0859)	(0.0063)	(0.0050)	(0.0018)	(0.0016)	(0.0042)
$\alpha = 0.6$	**141.8157**	**0.9331**	**0.2153**	**0.0558**	0.2904	0.0616	0.0278	0.3205
	(32.0246)	(0.4061)	(0.0856)	(0.0063)	(0.0056)	(0.0026)	(0.0023)	(0.0051)
$\alpha = 0.8$	**146.9892**	**0.9144**	**0.2123**	**0.0534**	0.2278	0.0786	0.0355	0.2758
	(33.2714)	(0.3976)	(0.0857)	(0.0071)	(0.0054)	(0.0038)	(0.0032)	(0.0059)
$\alpha = 1$	**151.2552**	**0.8700**	**0.2029**	**0.0598**	0.1853	0.0992	0.0464	0.2446
	(34.0646)	(0.3896)	(0.0861)	(0.0134)	(0.0056)	(0.0060)	(0.0048)	(0.0068)
N.S.	211.963	0.5994	0.1357	0.0260	**0.0936**	0.1795	0.0756	0.0000
	(50.4148)	(0.3053)	(0.0874)	(0.0033)	(0.0043)	(0.0118)	(0.0082)	(0.0000)
$1/N$	317.9742	0.4641	0.0761	**0.0119**	1.0000	0.0100	0.0000	0.0000
	(77.8459)	(0.2417)	(0.0945)	(0.0011)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
GMVP	**118.647**	0.8010	0.1806	0.3160	1.0000	0.0176	0.0076	0.7558
	(18.5033)	(0.4428)	(0.0892)	(0.0599)	(0.0000)	(0.0004)	(0.0004)	(0.0084)

Table 1.6: The table shows results of performances for monthly FF100 data when the full investment constraint $\mathbf{w^T 1} = 1$ is imposed. We estimate Σ by sample covariance matrix with expanding window scheme. Here $N = 100$ and initial window length $\tau_0 = 120$. For the weighted norm constraint, we set $\lambda_1 = \lambda_{1,t}$, $\lambda_1 = \lambda_{2,t}$ and vary α at six different levels. Testing period is from January 1990 to December 2010 and $T = 252$. SV denotes sample variance of out-of-sample net returns (when the transaction fees are deducted). The sample variance is annualized. The transaction fee we consider is 35 basis points. SR denotes annualized Sharpe ratio, and the yearly risk free rate used is 3.63%. Certainty equivalence is obtained with $\psi = 5$. Column TOR and PAC show average turnover rate and average proportion of active assets, respectively. Column HHI and ANHHI show average values of Herfindahl–Hirschman index, and average values of adjusted normalized Herfindahl–Hirschman index. Column SLR shows average short-long ratio. In the parentheses are the bootstrap standard errors obtained from using stationary bootstrap of Politis and Romano (1994).

	SV(%)	SR	CE(%)	TOR	PAC	HHI	ANHHI	SLR
$\alpha = 0$	**144.469**	**0.7928**	**0.7957**	**0.1990**	1.0000	0.0153	0.0053	0.5760
	(27.1036)	(0.3711)	(0.3508)	(0.0287)	(0.0000)	(0.0001)	(0.0002)	(0.0009)
$\alpha = 0.2$	**157.8873**	**0.7550**	**0.7642**	**0.1514**	0.4785	0.0366	0.0159	0.3847
	(30.3782)	(0.3643)	(0.3630)	(0.0294)	(0.0092)	(0.0005)	(0.0007)	(0.0020)
$\alpha = 0.4$	**167.6569**	**0.7416**	**0.7534**	**0.1150**	0.2926	0.0612	0.0277	0.2844
	(34.3140)	(0.3534)	(0.3635)	(0.0190)	(0.0065)	(0.0015)	(0.0019)	(0.0030)
$\alpha = 0.6$	**174.3271**	**0.7361**	**0.7493**	**0.0973**	0.2050	0.0807	0.0331	0.2270
	(37.5616)	(0.3401)	(0.3578)	(0.0156)	(0.0043)	(0.0022)	(0.0029)	(0.0041)
$\alpha = 0.8$	**180.8716**	**0.7119**	**0.7236**	**0.0911**	0.1629	0.0996	0.0401	0.1892
	(40.6771)	(0.3271)	(0.3536)	(0.0147)	(0.0033)	(0.0042)	(0.0047)	(0.0038)
$\alpha = 1$	**187.1527**	**0.6869**	**0.6957**	**0.0987**	0.1397	0.1245	0.0563	0.1605
	(42.8446)	(0.3148)	(0.3500)	(0.0168)	(0.0033)	(0.0062)	(0.0060)	(0.0041)
N.S.	219.3593	0.5222	0.4900	0.0581	**0.1001**	0.1692	0.0738	0.0000
	(49.5957)	(0.2663)	(0.3453)	(0.0077)	(0.0054)	(0.0101)	(0.0074)	(0.0000)
$1/N$	305.5895	0.5053	0.4019	**0.0287**	1.0000	0.0100	0.0000	0.0000
	(62.2650)	(0.2234)	(0.3597)	(0.0042)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
GMVP	255.2823	0.3886	0.2881	1.2783	1.0000	0.0170	0.0071	0.8379
	(36.0606)	(0.2874)	(0.3937)	(0.2674)	(0.0000)	(0.0002)	(0.0002)	(0.0114)

Table 1.7: The table shows results of performances for weekly CRSP300 data when the full investment constraint $\mathbf{w}^{\mathbf{T}}\mathbf{1} = 1$ is imposed. We estimate Σ by sample covariance matrix with expanding window scheme. Here $N = 300$ and initial window length $\tau_0 = 360$. For the weighted norm constraint, we set $\lambda_1 = \lambda_{1,t}$, $\lambda_1 = \lambda_{2,t}$ and vary α at six different levels. Testing period is from the first week of 2000 to the last week of 2010 and $T = 573$. SV denotes sample variance of out-of-sample net returns (when the transaction fees are deducted). The sample variance is annualized. The transaction fee we consider is 35 basis points. SR denotes annualized Sharpe ratio, and the yearly risk free rate used is 2.50%. Certainty equivalence is obtained with $\psi = 5$. Column TOR and PAC show average turnover rate and average proportion of active assets, respectively. Column HHI and ANHHI show average values of Herfindahl–Hirschman index, and average values of adjusted normalized Herfindahl–Hirschman index. Column SLR shows average short-long ratio. In the parentheses are the bootstrap standard errors obtained from using stationary bootstrap of Politis and Romano (1994).

	SV(%)	SR	CE(%)	TOR	PAC	HHI	ANHHI	SLR
$\alpha = 0$	**75.5405**	**0.4162**	**0.0813**	0.1473	1.0000	0.0082	0.0049	0.4822
	(25.9349)	(0.4583)	(0.0743)	(0.0143)	(0.0000)	(0.0003)	(0.0003)	(0.0059)
$\alpha = 0.2$	**69.3276**	**0.6094**	**0.1123**	0.0884	0.6555	0.0177	0.0126	0.2800
	(27.6839)	(0.5044)	(0.0716)	(0.0095)	(0.0037)	(0.0004)	(0.0004)	(0.0079)
$\alpha = 0.4$	**70.1200**	**0.6767**	**0.1233**	0.0693	0.5160	0.0247	0.0183	0.2038
	(29.3341)	(0.5194)	(0.0717)	(0.0078)	(0.0042)	(0.0005)	(0.0005)	(0.0089)
$\alpha = 0.6$	**71.3702**	**0.7241**	**0.1314**	0.0603	0.4410	0.0300	0.0226	0.1634
	(30.5486)	(0.5275)	(0.0717)	(0.0069)	(0.0048)	(0.0005)	(0.0006)	(0.0089)
$\alpha = 0.8$	**72.6740**	**0.7605**	**0.1378**	0.0550	0.3974	0.0344	0.0262	0.1363
	(31.4910)	(0.5310)	(0.0715)	(0.0063)	(0.0060)	(0.0006)	(0.0006)	(0.0086)
$\alpha = 1$	**74.3669**	**0.7836**	**0.1423**	0.0515	0.3668	0.0386	0.0297	0.1153
	(32.5348)	(0.5310)	(0.0715)	(0.0060)	(0.0073)	(0.0006)	(0.0007)	(0.0084)
N.S.	95.4867	**0.9127**	**0.1737**	0.0340	0.1956	0.0656	0.0491	0.0000
	(43.6905)	(0.4872)	(0.0705)	(0.0039)	(0.0069)	(0.0023)	(0.0018)	(0.0000)
$1/N$	393.5178	0.6621	0.1115	0.0347	1.0000	0.0033	0.0000	0.0000
	(132.7015)	(0.3422)	(0.1421)	(0.0030)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
GMVP	88.2586	0.2131	0.0441	0.2131	1.0000	0.0086	0.0053	0.5492
	(26.6200)	(0.3741)	(0.0709)	(0.0315)	(0.0000)	(0.0003)	(0.0003)	(0.0128)

Table 1.8: The table shows results of performances of weighted norm mvp for daily CRSP300 and FF100 data when the full investment constraint $\mathbf{w}^{\mathbf{T}}\mathbf{1} = 1$ and target return constraint $\mathbf{w}^{\mathbf{T}}\mu = \bar{\mu}$ are imposed. The target return $\bar{\mu}$ shown in the table is annualized. We estimate μ and Σ by sample mean and covariance matrix with expanding window scheme. Here $N = 100$ and 300 for FF100 and CRSP300 respectively, and initial window length $\tau_0 = 1.2N$. For the weighted norm constraint, we set $\lambda_1 = \lambda_{1,t}$, $\lambda_1 = \lambda_{2,t}$ and vary α at three different levels. SV denotes sample variance of out-of-sample net returns (when the transaction fees are deducted). The sample variance is annualized. The transaction fee we consider is 35 basis points. SR denotes annualized Sharpe ratio, and the yearly risk free rates for FF100 and CRSP300 are 3.63% and 2.5%, respectively. Certainty equivalence is obtained with $\psi = 5$. Column PAC shows average proportion of active assets. In the parentheses are the bootstrap standard errors obtained from using stationary bootstrap of Politis and Romano (1994).

FF100, Jan-02-1990 to Dec-31-2010, $T = 5,295$										
	$\bar{\mu} = 10\%$					$\bar{\mu} = 20\%$				
	SV(%)	SR	CE(%)	TOR	PAC	SV(%)	SR	CE(%)	TOR	PAC
$\alpha = 0$	96.6592	0.5857	0.0279	0.0615	1.0000	83.7073	1.0200	0.0435	0.0502	1.0000
	(18.6169)	(0.4286)	(0.0167)	(0.0034)	(0.0000)	(13.8393)	(0.4591)	(0.0157)	(0.0053)	(0.0000)
$\alpha = 0.6$	111.2906	0.6429	0.0305	0.0488	0.4938	94.7331	1.0596	0.0463	0.0382	0.4328
	(22.2266)	(0.4135)	(0.0171)	(0.0027)	(0.0111)	(16.8293)	(0.4509)	(0.0163)	(0.0049)	(0.0075)
$\alpha = 1$	118.5553	0.6170	0.0295	0.0468	0.3801	100.7166	1.0592	0.0470	0.0373	0.3212
	(23.5070)	(0.4017)	(0.0173)	(0.0030)	(0.0096)	(17.9121)	(0.4453)	(0.0166)	(0.0052)	(0.0069)
CRSP300, Jan-03-2000 to Dec-31-2010, $T = 2,767$										
	$\bar{\mu} = 5\%$					$\bar{\mu} = 10\%$				
	SV(%)	SR	CE(%)	TOR	PAC	SV(%)	SR	CE(%)	TOR	PAC
$\alpha = 0$	52.5846	0.5876	0.0218	0.0530	1.0000	50.2325	0.4033	0.0164	0.0474	1.0000
	(16.4380)	(0.4681)	(0.0127)	(0.0049)	(0.0000)	(14.6432)	(0.4542)	(0.0125)	(0.0053)	(0.0000)
$\alpha = 0.6$	53.6010	0.9629	0.0328	0.0320	0.5674	51.0724	0.7185	0.0254	0.0246	0.5585
	(19.2751)	(0.5067)	(0.0123)	(0.0019)	(0.0081)	(17.4733)	(0.4937)	(0.0124)	(0.0024)	(0.0088)
$\alpha = 1$	55.3200	0.9966	0.0341	0.0290	0.4603	52.9987	0.7406	0.0263	0.0213	0.4549
	(20.1596)	(0.5111)	(0.0124)	(0.0015)	(0.0068)	(18.5730)	(0.4948)	(0.0126)	(0.0020)	(0.0069)

Table 1.9: The table shows results of performances for daily FF100 data when three alternative penalties: berhu, generalized l_1 norm, and adaptive penalties are imposed. The linear constraint is the full investment constraint $\mathbf{w^T 1} = 1$. We estimate Σ by sample covariance matrix with expanding window scheme. Here $N = 100$ and initial window length $\tau_0 = 120$. For each penalty, we uniformly set the penalty parameter equal to $\hat{a}_t \hat{B}_t \sqrt{2 \log N/n_t}$. Testing period is from Jan-02-1990 to Dec-31-2010 and $T = 5,295$. SV denotes sample variance of out-of-sample net returns (when the transaction fees are deducted). The sample variance is annualized. The transaction fee we consider is 35 basis points. SR denotes annualized Sharpe ratio, and the yearly risk free rate used is 3.63%. Certainty equivalence is obtained with $\psi = 5$. Column TOR and PAC show average turnover rate and average proportion of active assets, respectively. In the parentheses are the bootstrap standard errors obtained from using stationary bootstrap of Politis and Romano (1994).

			Berhu Penalty		
	SV(%)	SR	CE(%)	TOR	PAC
$\kappa = 0.02$	146.0434	0.8283	0.0400	0.0143	0.5483
	(32.1981)	(0.3924)	(0.0180)	(0.0010)	(0.0049)
$\kappa = 0.05$	119.1196	0.9940	0.0460	0.0178	0.4142
	(25.0352)	(0.4391)	(0.0176)	(0.0013)	(0.0062)
$\kappa = 0.1$	106.9375	1.1123	0.0498	0.0191	0.3531
	(22.1391)	(0.4676)	(0.0175)	(0.0016)	(0.0074)
			Generalized l_1 Norm Penalty		
	SV(%)	SR	CE(%)	TOR	PAC
TWN	104.0272	1.0798	0.0482	0.0221	1.0000
	(20.2157)	(0.4463)	(0.0166)	(0.0018)	(0.0000)
$TWN - l_1$	114.8295	1.0561	0.0483	0.0166	0.6677
	(23.2896)	(0.4355)	(0.0170)	(0.0014)	(0.0047)
$TWNS$	98.5689	1.1792	0.0515	0.0200	0.3090
	(18.9658)	(0.4638)	(0.0166)	(0.0018)	(0.0068)
$TWNS - l_1$	111.5312	1.0922	0.0495	0.0155	0.2196
	(22.5187)	(0.4421)	(0.0169)	(0.0015)	(0.0055)
			Adaptive Penalty		
	SV(%)	SR	CE(%)	TOR	PAC
$\epsilon = 1, l = 1$	96.5597	1.2096	0.0524	0.0208	0.2978
	(18.2046)	(0.4696)	(0.0163)	(0.0019)	(0.0075)
$\epsilon = 2.5, l = 1$	107.2337	1.1244	0.0504	0.0178	0.1778
	(20.2679)	(0.4479)	(0.0166)	(0.0019)	(0.0052)
$\epsilon = 1, l = 2$	96.4462	1.2107	0.0524	0.0210	0.2952
	(18.1246)	(0.4692)	(0.0163)	(0.0020)	(0.0077)
$\epsilon = 2.5, l = 2$	107.8656	1.1139	0.0500	0.0190	0.1519
	(20.0121)	(0.4452)	(0.0166)	(0.0025)	(0.0052)

Table 1.10: The table shows results of performances for daily FF100 data when three alternative penalties: berhu, generalized l_1 norm, and adaptive penalties are imposed. The linear constraint is the full investment constraint $\mathbf{w^T1} = 1$. We estimate Σ by sample covariance matrix with expanding window scheme. Here $N = 100$ and initial window length $\tau_0 = 120$. For each penalty, we uniformly set the penalty parameter equal to $\hat{a}_t\hat{B}_t\sqrt{2\log N/n_t}$. Testing period is from Jan-03-2000 to Dec-31-2010 and $T = 2,767$. SV denotes sample variance of out-of-sample net returns (when the transaction fees are deducted). The sample variance is annualized. The transaction fee we consider is 35 basis points. SR denotes annualized Sharpe ratio, and the yearly risk free rate used is 2.5%. Certainty equivalence is obtained with $\psi = 5$. Column TOR and PAC show average turnover rate and average proportion of active assets, respectively. In the parentheses are the bootstrap standard errors obtained from using stationary bootstrap of Politis and Romano (1994).

	Berhu Penalty				
	SV(%)	SR	CE(%)	TOR	PAC
$\kappa = 0.02$	57.2220	1.0880	0.0372	0.0230	0.4932
	(18.1071)	(0.5389)	(0.0133)	(0.0019)	(0.0054)
$\kappa = 0.05$	54.3572	1.0274	0.0349	0.0210	0.4617
	(18.6517)	(0.5264)	(0.0125)	(0.0018)	(0.0064)
$\kappa = 0.1$	53.7601	0.9926	0.0337	0.0201	0.4517
	(19.1032)	(0.5181)	(0.0122)	(0.0018)	(0.0075)
	Generalized l_1 Norm Penalty				
	SV(%)	SR	CE(%)	TOR	PAC
TWN	54.9798	0.7611	0.0271	0.0286	1.0000
	(18.7627)	(0.4787)	(0.0126)	(0.0023)	(0.0000)
$TWN - l_1$	57.5252	0.9593	0.0333	0.0196	0.6340
	(20.7838)	(0.5066)	(0.0127)	(0.0017)	(0.0070)
$TWNS$	55.2197	0.9643	0.0331	0.0191	0.4346
	(19.5056)	(0.4856)	(0.0119)	(0.0018)	(0.0063)
$TWNS - l_1$	58.2940	1.0160	0.0352	0.0158	0.3240
	(20.7498)	(0.4924)	(0.0123)	(0.0014)	(0.0045)
	Adaptive Penalty				
	SV(%)	SR	CE(%)	TOR	PAC
$\epsilon = 1, l = 1$	53.8823	0.9597	0.0328	0.0200	0.4468
	(18.8170)	(0.4949)	(0.0119)	(0.0018)	(0.0076)
$\epsilon = 2.5, l = 1$	59.0413	0.9925	0.0346	0.0151	0.2851
	(20.8862)	(0.4926)	(0.0123)	(0.0013)	(0.0053)
$\epsilon = 1, l = 2$	53.8860	0.9592	0.0328	0.0200	0.4467
	(18.8191)	(0.4949)	(0.0119)	(0.0018)	(0.0076)
$\epsilon = 2.5, l = 2$	59.2664	0.9842	0.0344	0.0151	0.2808
	(20.9691)	(0.4914)	(0.0124)	(0.0013)	(0.0053)

Chapter 2

Bond Variance Risk Premia (Joint Work with Philippe Mueller and Andrea Vedolin)

Abstract: Using data from 1983 to 2010, we propose a new fear measure for Treasury markets, akin to the VIX for equities, labeled TIV. We show that TIV explains one third of the time variation in funding liquidity and that the spread between the VIX and TIV captures flight to quality. We then construct Treasury bond variance risk premia as the difference between the implied variance and an expected variance estimate using autoregressive models. Bond variance risk premia display pronounced spikes during crisis periods. We show that variance risk premia encompass a broad spectrum of macroeconomic uncertainty. Uncertainty about the nominal and the real side of the economy increase variance risk premia but uncertainty about monetary policy has a strongly negative effect. We document that bond variance risk premia predict excess returns on Treasuries, stocks, corporate bonds and mortgage-backed securities, both in-sample and out-of-sample. Furthermore, this predictability is not subsumed by other standard predictors.
KEYWORDS: Variance risk premium, Treasury implied volatility, predictability, uncertainty, Treasury bond returns, stock returns, corporate bond returns.
JEL Codes: E43, E47, G12, G17.

2.1 Introduction

During the recent financial crisis one sector generated significant profits for the leading investment banks: Volatility arbitrage trading in forex, fixed income, and commodities. According to a BIS (2010) survey on foreign exchange and derivatives markets activity, the interest rate derivatives market has grown by 24% over the last three years to reach an average daily turnover of USD 2.1 trillion. As a consequence, both market and academic interest in equity-index volatility measures and their associated risk premia has grown rapidly. For instance, the VIX index—also dubbed the "investors' fear index"—is believed to be a good proxy of aggregate uncertainty or risk aversion.[1] The VIX is also shown to be a good predictor for the cross-section of stocks (Ang et al., 2006), corporate credit spreads (Collin-Dufresne et al., 2001) and bond excess returns (Baele et al., 2010). Furthermore, the associated variance risk premium extracted from equity markets predicts the equity premium (Drechsler and Yaron, 2011), as well as corporate credit spreads (Wang et al., 2010). Given this extensive literature for equity markets, it is rather surprising that no effort has been undertaken to measure these risk premia in fixed income markets. Filling this gap is one goal of this paper.

The importance of understanding interest rate volatility and the risk premia associated with it is manifested in Figure 2.1, where we plot the Mortgage Bankers Association (MBA) Refinancing index together with the variance risk premium calculated from 30 year Treasury futures. The MBA refinancing index is based on the number of applications for mortgage refinancing. The figure nicely displays that mortgage refinancing is subject to distinct waves such as the peak of the housing boom in May 2003 or the bust during the most recent financial crisis. During these periods, the bond variance risk premium also peaks and it has moved almost in tandem with the MBA index since 2005. The intuition is relatively straightforward. If interest rates drop, the duration of any mortgage-backed security portfolio decreases due to a higher refinancing rate. To hedge the duration, the portfolio manager must either buy Treasury bonds or bond options. This hedging activity not only affects the price of the underlying but also its volatility (Duarte, 2008). Consequently, risk-averse investors demand a premium for the associated risks.

We contribute to the literature in the following ways. First, we construct a new measure of fear for Treasury markets (akin to the VIX for equities), which we label TIV for *Treasury Implied Volatility*.[2] Second, we construct and describe the term structure of bond variance risk premia for 30 year, 10 year and 5 year Treasury futures and investigate the underlying economic drivers of these risk premia. Finally, we document the strong predictive power of bond variance risk premia for excess returns on Treasury bonds, stocks, corporate bonds and mortgage-backed securities. The predictability is in-sample and out-of-sample and robust to

[1] See, e.g., Bollerslev et al. (2011), Korteweg and Polson (2010) and Bekaert et al. (2011), among others.

[2] We calculate both equity and bond implied volatilities using newly available high frequency data. The time-series for the TIV starts in 1983. In addition, we calculate our own VIX measure, also with a start date in 1983, thus considerably longer than the Chicago Board of Option Exchange (CBOE) VIX available since 1991.

the inclusion of other standard predictors.

To construct the TIV, we follow the recent literature and calculate implied variance measures using a model-free approach (Britten-Jones and Neuberger, 2000).[3] Moreover, we are the first to estimate and study a term structure of variance risk premia for the Treasury market. Even though there is ample evidence of priced variance risk in both the index and single stock equity market, we know surprisingly little about the compensation for variance risk in fixed income markets.

The variance risk premium is defined as the difference between the expected risk-neutral and physical variance. While the risk-neutral expectation can be estimated in a completely model-free fashion using a cross-section of options written on the underlying asset, the calculation of the objective expectation requires some mild auxiliary modeling assumptions. A priori, it is not clear, what the best proxy for this objective expectation should be. Andersen et al. (2007a) show that simple autoregressive type models estimated directly for the realized volatility often perform better than parametric approaches designed to forecast the integrated volatility. In calculating our benchmark bond variance risk premium, we thus use the HAR-TCJ model for realized variance proposed by Corsi et al. (2010). We augment the model by including lagged implied variance as additional regressors.[4]

Using data from 1983 to 2010, we find that the implied volatility measures we derive in both equity and bond markets are remarkably similar, which is manifested in the high unconditional correlation of around 60% on average.[5] Increases in the VIX index are often dubbed as an increase in economic uncertainty. We find a similar pattern for the bond market. Implied volatility in bond markets spikes in crisis times and it therefore offers itself as a gauge of fear for fixed income markets. The construction of the TIV measure has an economic merit which goes beyond that of the VIX itself. First, we show that TIV is strongly related to proxies of funding liquidity. A one standard deviation change in the TIV implies more than half a standard deviation change in a funding liquidity proxy and spikes in the TIV can therefore be interpreted as shocks to funding liquidity. This empirical finding relates to the theoretical work of (Brunnermeier and Pedersen, 2009), who show that lower liquidity can lead to higher asset volatility. Moreover, the authors demonstrate that in periods of flight to quality, highly liquid assets are characterized by relatively low volatility. Thus, the spread between low and high volatility assets can explain part of the liquidity spread. In our empirical analysis, we

[3] In addition, we also construct implied variance measures as in Martin (2011), using so called simple variance swaps. Whereas the replication of standard variance swaps relies on the Itô assumption which is violated in case there are jumps, simple variance swaps provide a genuine measure of implied variance under very general assumptions. While this distinction is important from a methodological perspective, the main results are robust to the choice of method.

[4] Recently, Bollerslev et al. (2012) use a simple heterogeneous autoregressive RV model to construct the stock market variance risk premium while Busch et al. (2011) use the augmented HAR-RV model with lagged IV to improve forecasts of realized volatility. In the Online Appendix we show that the HAR-TCJ model with lagged IV performs best in predicting out-of-sample realized variance.

[5] The correlation between the implied volatility measures for 30 year Treasury futures and equities is as high as 69%, whereas the correlation between the implied volatility measures for 5 year Treasury futures and equities is around 53%.

find that the volatility spread between the VIX and the TIV provides a useful measure of flight to quality periods: While the spread is most of the time no more than 10%, it almost triples during the October 1987 crash, the LTCM default in August 1998, and the Lehman bankruptcy in September 2008.

While the co-movement between the time-series is high, the TIV and VIX differ in their magnitude. For our sample period, the average model-free implied volatility of the S&P 500 index is 20% with a standard deviation of 8.4, in contrast, the 30 year Treasury implied volatility is 10% on average with a standard deviation of merely 2.4. Thus, in the case of the S&P 500 index, volatility risk accounts for a much larger proportion of overall risk than in Treasury markets. Despite the high co-movement, we find that variance risk premia in bond and equity markets can behave differently. While the variance risk premium in the equity market is essentially always positive (i.e., it acts as an insurance premium), the variance risk premium in the Treasury market can switch sign. To grasp a better intuition for this behavior, we study macroeconomic determinants of bond versus equity variance risk premia. Proxies of macroeconomic uncertainty from forecast data explain up to 45% of the time variation in bond and equity variance risk premia. Higher uncertainty usually implies a higher variance risk premium. However, uncertainty about short term yields (which can be interpreted as uncertainty about monetary policy actions) has a significant negative impact on bond variance risk premia. Larger uncertainty about the short end makes investors with negative expectations about future interest rates willing to buy long term bonds, because these provide a hedge against the increased duration due to a drop in short term yields. Hence, investors pay a premium for holding these bonds. Inline with recent findings (Joslin, 2010), we also document that the shape of the term structure significantly affects bond variance risk premia.

If bond variance risk premia encompass general macroeconomic uncertainty, do they contain any useful information about asset returns? A principal components analysis of the bond variance risk premia time series allows us to summarize the information in the term structure of bond variance risk premia in a parsimonious way. We show that the three principal components have economically significant predictive power for excess returns across different assets. The results can be summarized as follows: A one standard deviation change in the third principal component of Treasury bond variance risk premia, a curvature factor, induces a 0.17 standard deviation decrease in bond excess returns, while the same kind of shock has an opposite effect of roughly the same magnitude on stock market excess returns. A one standard deviation change in the second principal component, a slope factor, induces a 0.24 standard deviation increase in stock excess returns and up to half a standard deviation change in corporate bond and mortgage-backed securities excess returns. Finally, a one standard deviation shock to the first principal component, a level factor, has a strong positive effect on corporate bond and mortgage-backed securities excess returns. Bond variance risk premia explain roughly 3% of the time variation in Treasury excess returns, around 9% of stock excess returns and up to 35% of corporate and mortgage-backed securities excess returns.

When we add the equity variance risk premium to the regressions, the significance of the bond variance risk premia remains economically and statistically high, whereas the equity variance risk premium adds very little predictive power. We show that the predictability of bond variance risk premia prevails in-sample and out-of-sample and is robust to the inclusion of other standard predictors in the literature. We conclude that bond variance risk premia broadly capture uncertainty about the macroeconomy and monetary policy, as well as additional information about the term structure that is relevant for all asset classes. Moreover, bond variance risk premia have a great advantage over most of the other predictor variables that rely on either macroeconomic fundamentals or forecast data: They can be obtained on a daily basis (or at even higher frequencies) while the other variables are often only available at the monthly frequency at best.

Our paper is related to two strands of the literature. First, it fits into the large body of research that has focused on the stock market variance risk premium and—to a lesser degree—on variance risk premia of individual stocks or commodities.[6] To the best of our knowledge, our paper is the first to study variance risk premia in the Treasury market.

One reason why variance risk in fixed income markets has been neglected in the past could be that standard dynamic term structure models assume that the fixed income market is complete and therefore, interest rate derivatives are redundant assets. Only recently, there is emerging (albeit sometimes mixed) evidence for the existence of unspanned stochastic volatility, the second body of research that is related to our paper.[7]

To summarize, in this paper, we provide new empirical facts about variance risk premia in the fixed income markets. We construct a term structure of Treasury bond variance risk premia and investigate its determinants. In addition, we document the strong predictive power for a wide range of assets. However, we remain agnostic about the form of structural model that could rationalize our findings and leave this for future research.

The rest of the paper is organized as follows. Section 2.2 describes our data set and Section 2.3 describes the econometric methods used to estimate the TIV measure and the variance risk premia. Section 2.4 presents the results of our empirical study and Section 2.5 concludes. To save space, we defer additional data description, alternative methods to estimate implied and realized variance, and robustness checks to the Online Appendix.[8]

[6]For literature on the stock market variance risk premium, see, e.g., Driessen et al. (2009), Bollerslev et al. (2011), Carr and Wu (2009), Cremers et al. (2010) and Todorov (2010), among others. Bakshi and Kapadia (2003) and Vedolin (2010) for example study the variance risk premia of individual stocks and Trolle and Schwartz (2009) investigate variance risk premia in commodity markets.

[7]Joslin (2010) studies the variance risk premium using swap data and proposes a model that under certain restrictions can generate unspanned stochastic volatility. Additionally, see Collin-Dufresne and Goldstein (2002), Heidari and Wu (2003), Casassus et al. (2005), Collin-Dufresne et al. (2008), Bibkov and Chernov (2009), Trolle and Schwartz (2009), Andersen and Benzoni (2010) and Almeida et al. (2010) among others.

[8]The Online Appendix is available on the authors' webpage.

2.2 Data

In this section, we briefly introduce the data used in our analysis. Firstly, we use futures and options data to construct the bond and equity variance risk premia. Secondly, we calculate excess returns for Treasury, stock, corporate bond and mortgage-backed securities portfolios. Finally, we use a large set of macro and forecast data as controls to explore the determinants and the predictive power of the variance risk premia. The summary statistics for excess returns and additional variables are contained in the Online Appendix.

2.2.1 Futures and Options Data

Treasury Futures and Options: To calculate implied and realized variance measures for Treasury bonds, we use futures and options data from the Chicago Mercantile Exchange (CME). We use high-frequency intra-day price data of the 30 year Treasury bond futures, the 10 year and 5 year Treasury notes futures and end-of-day prices of options written on the underlying futures. The data runs from October 1982, May 1985 and May 1990 to June 2010 for the 30 year, 10 year, and 5 year Treasury bond futures and options, respectively. Using a monthly frequency throughout the paper, we have at most 333, 302, and 242 observations available, respectively.

Treasury futures are traded electronically as well as by open outcry. While the quality of electronic trading data is higher, the data only becomes available in August 2000. To maximize our time span, we use data from electronic as well as pit trading sessions. We only consider trades that occur during regular trading hours (07:20–14:00) when the products are traded side-by-side in both markets.[9]

The contract months for the Treasury futures are the first three (30 year Treasury bond futures) or five (10 year and 5 year Treasury notes futures) consecutive contracts in the March, June, September, and December quarterly cycle. This means that at any given point in time, up to five contracts on the same underlying are traded. To get one time series, we roll the futures on the 28^{th} of the month preceding the contract month.

For options, the contract months are the first three consecutive months (two serial expirations and one quarterly expiration) plus the next two (30 year futures) or four (10 year and 5 year futures) months in the March, June, September, and December quarterly cycle. Serials exercise into the first nearby quarterly futures contract, quarterlies exercise into futures contracts of the same delivery period. We roll our options data consistent with the procedure applied to the futures.[10]

S&P 500 Index Futures and Options: Inline with our approach for Treasuries, we calculate the implied and realized variance measures for the stock market using futures and options on the

[9] Liquidity in the after-hours electronic market is significantly smaller than during regular trading hours.

[10] Detailed information about the contract specifications of Treasury futures and options can be found on the CME website, www.cmegroup.com.

S&P 500 index from CME. The sample period is from January 1983 to June 2010.[11]

2.2.2 Excess Returns Data

Treasury Bonds: We use the Fama-Bliss discount bond database from CRSP to calculate annual Treasury bond excess returns for two to five year bonds. We denote the return on a τ-year bond with log price $p_t^{(\tau)}$ by $r_{t+1}^{(\tau)} = p_{t+1}^{(\tau-1)} - p_t^{(\tau)}$. The annual excess bond return is defined as $rx_{t+1}^{(\tau)} \equiv r_{t+1}^{(\tau)} - y_t^{(1)}$, where $y_t^{(1)}$ is the one year yield.
Stocks: To proxy for the market portfolio, we use the value-weighted index from CRSP. The growth and value portfolio returns are constructed using the six portfolios formed on size and book-to-market from Ken French's data library. The respective returns are the average of the returns of the small and big growth and value portfolio, respectively. The three, six and twelve month excess returns are defined as the cumulative return on the respective portfolio minus the Treasury yield.
Corporate Bonds and CMBS: We use corporate bond and commercial mortgage-backed securities (CMBS) indices from Barclays Capital to calculate three, six and twelve month excess returns. We use AAA, BBB, and CCC indices for long and intermediate corporate bonds and AAA, BBB, and B indices for CMBS. CMBS data is available starting in 1997.

2.2.3 Other Data

*Forecasts:*We use forecast data from BlueChip Economic Indicators (BCEI) to calculate proxies of uncertainty about macroeconomic variables. BCEI collects monthly forecasts of twelve key financial and macroeconomic indicators from about fifty professional economists in leading financial and economic advisory firms.[12] The forecasts are made for different time horizons. This data exhibits strong seasonality and thus, we adjust the series using a 12-period ARIMA filter. We use the cross sectional standard deviation of the filtered panel data within each month as the monthly gauge of uncertainty. We calculate the time series of the cross sectional standard deviation using the forecasts for the current and the subsequent calendar year for each forecast variable i. Thus, for each variable we have two time series reflecting the uncertainty of the forecaster. Our uncertainty proxy $\widehat{U}^i$ is the first principal component extracted from

[11] We compare our results to the VIX and VXO measures that are calculated using options on the S&P500 cash index instead of S&P500 index futures. The VIX is the implied volatility calculated using a model-free approach, whereas the VXO is calculated using the (Black and Scholes, 1973) implied volatility. The VIX is available starting in January 1990 and the VXO is available since January 1986. Over the common sample period, the VIX and our implied volatility measure from index futures options using the same methodology have a correlation of over 99.4% and the root mean squared error is below 1%.

[12] The twelve series are the real gross domestic product (RGDP), the GDP chained price index (GDPI), the consumer price index (CPI), industrial production (IP), real disposable personal income (DPI), non-residential investment (NRI), the unemployment rate (UNEM), housing starts (HS), corporate profits (CP), total US auto and truck Sales (AS), the three-month secondary market T-bill rate (SR) and the ten year constant maturity Treasury yield (LR).

these two time series.[13] In our analysis, we use uncertainty about the real (RGDP) and the nominal (CPI) side of the economy, as well as uncertainty about short and long rates (SR and LR), where $\widehat{U}^{SR}$ can also be interpreted as uncertainty about monetary policy. The forecast data is available until December 2009.
Macroeconomic Factors: We compute the eight static macroeconomic factors $\widehat{F}_j, j = 1 \ldots, 8$, from Ludvigson and Ng (2009) and Ludvigson and Ng (2011) for an updated data set through June 2010.[14] We also estimate volatility proxies for inflation and consumption, σ_π and σ_g. We calculate these by estimating a GARCH process for monthly CPI inflation and consumption (non-durables and services). The data is from Global Insight and the Federal Reserve Economic Data base (FRED).
Additional Variables: Using the Fama-Bliss data, we also construct a tent-shaped factor from forward rates, the Cochrane and Piazzesi (2005) factor, CP. Furthermore, we calculate the slope of the term structure as the difference between the ten year and the one month Treasury yield (SLOPE). In addition, we use the log dividend yield (DY), the log earnings/price ratio (E/P), and the net equity expansion (NTIS) from Goyal and Welch (2008), and REF, the Mortgage Bankers Association refinancing index.

2.3 Estimation of Bond Variance Measures and Variance Risk Premia

In this section, we describe the methods used to estimate the expected risk-neutral and objective variance, $\mathbb{E}_t^{\mathbb{Q}}\left(\int_t^T \sigma_u^2 du\right)$ and $\mathbb{E}_t^{\mathbb{P}}\left(\int_t^T \sigma_u^2 du\right)$, and the variance risk premium, defined as the difference between the two.[15]

Moreover, we define a *Treasury Implied Volatility* or TIV measure in the spirit of the well known VIX index that is calculated by CBOE for the S&P500 index. Our proposed TIV measure is the 30 year Treasury bond futures implied volatility, i.e. the square root of the implied variance.[16] We calculate a daily TIV measure going back to October 1982.[17]

[13]As the principal components are latent, we ensure that the first principal component is positively correlated with the two uncertainty proxies.

[14]The original data set was previously used in Stock and Watson (2002). Some of the macroeconomic variables are no longer available after 2007. Consequently, we use 125 instead of 132 macroeconomic time series. In addition, we exclude all stock market and interest rate time series and work with a set of 104 variables. We also use the full data set with 125 variables and the original factors for shorter sample period ending in 2007 as a robustness check. Our results remain unchanged. A detailed description of the macroeconomic data is provided in the Online Appendix.

[15]We present and discuss additional methods to estimate expected variance in the Online Appendix. Overall, our empirical results are robust to using reasonable alternative methods to what is described in this section.

[16]Unlike the 10 year and 5 year instruments, the 30 year Treasury futures and options have the longest available history and they are very liquid even in the 1980s.

[17]The time series for the TIV measure will be made available on the authors' website. As mentioned in Section 2.2, we construct our own VIX measure, which is based on options on S&P500 index futures rather than on the underlying cash index. This allows us to obtain a longer time series as options data on S&P500 index futures date back to the 1980s with high trading volumes, whereas the VIX only starts in 1990.

2.3.1 Implied Variance

As is commonly done, we use options to back out a proxy for the expected variance under the risk-neutral measure, $\mathbb{E}_t^{\mathbb{Q}}\left(\int_t^T \sigma_u^2 du\right)$.[18] We implement a model-free method as proposed by Britten-Jones and Neuberger (2000) that only requires current option prices to calculate the implied variance (subsequently denoted MIV).

One well-established application of the model-free implied variance is the VIX, which is an index of implied volatility (i.e. the square root of the MIV) calculated using options on the S&P500.Neuberger (1994) shows that the VIX corresponds to the quadratic variation of the forward price of the S&P500 index under the risk-neutral measure. One issue with the replication of the variance swap is that it heavily relies on the Itassumption for the underlying process. In the presence of skewness, Carr and Lee (2009) show that the VIX will be upward biased compared to the true risk-neutral quadratic variation. Martin (2011) introduces the simple variance swap for which the realized leg can be computed from simple returns of the underlying index and the index forward. He shows that—just as the VIX—the SVIX can also be approximated as a portfolio of out-of-the-money options and can be constructed under slightly weaker assumptions and in the presence of jumps.[19]

To implement the methods for calculating the implied variance for options on Treasury futures, we treat the American options as European[20]. Furthermore, we assume that the short risk-free rate is non-stochastic (or at least not too volatile) such that the forward and futures prices coincide.[21]

To calculate the model-free implied variance, MIV, we then follow Demeterfi et al. (1999) and Britten-Jones and Neuberger (2000). They show that if the underlying asset price is continuous, the risk-neutral expectation of total return variance is defined as an integral of option prices over an infinite range of strike prices. Since in practice, the number of traded options for any underlying asset is finite, the available strike price series is a finite sequence. Denote $\mathsf{C}(T, K)$ the spot call price with strike price K expiring at time T. Suppose the

[18] The simplest way to calculate the implied variance would be to invert the standard Black (1976) formula (we denote the implied variance from this method BIV). Black's model is often used to value interest rate options. Busch et al. (2011) for example use this measure to study the forecasting power of implied volatility for realized volatility of Treasury bond futures. However, one of the relevant assumptions underlying the model is constant volatility, which is inconsistent with the application to forecasting changes in volatility. Nevertheless, the empirical results are qualitatively robust to using the BIV measure instead of a model-free approach.

[19] For robustness checks, we also implement this method and denote the resulting implied variance measure SIV. Again, results are robust. Summary statistics are provided in the Online Appendix.

[20] Jorion (1995) shows that early exercise premia are small for short maturity at-the-money options on futures, while Overdahl (1988) demonstrates that early exercise of options on Treasury futures happens about 0.1% of the time and happens both with calls and puts but only with options that are significantly in the money. In the empirical implementation, we use only out-of-the money options and thus assume that the early exercise option will not distort the option price.

[21] Similar to the issue with the Black (1976) implied volatility, this is a slight inconsistency in the approach as interest rates are clearly assumed to be stochastic when it comes to calculating the payoff of the option (which is written on a futures contract that is dependent on an underlying interest rate process). However, the assumption we have to make to implement the method concerns the short risk-free rate and not directly the interest rate underlying the Treasury futures option.

available strike prices of the call options belong to $\left[\underline{K}^{\mathsf{c}},\ \overline{K}^{\mathsf{c}}\right]$, where $\overline{K}^{\mathsf{c}} \geq \underline{K}^{\mathsf{c}} \geq 0$. As shown in Jiang and Tian (2005), a truncated version of the integral over the infinite range of strike prices can be used to evaluate the model-free implied volatility. We use the trapezoidal rule to numerically calculate the integral:

$$2\int_{\underline{K}^{\mathsf{c}}}^{\overline{K}^{\mathsf{c}}} \frac{\mathsf{C}\,(T,K) - \max\left(0, F_t - K\right)}{K^2} dK \approx \frac{\overline{K}^{\mathsf{c}} - \underline{K}^{\mathsf{c}}}{m} \sum_{i=1}^{m} \left[g_{t,T}\left(K_i^{\mathsf{c}}\right) + g_{t,T}\left(K_{i-1}^{\mathsf{c}}\right)\right],$$

where

$$g_{t,T}\left(K_i^{\mathsf{c}}\right) = \frac{\mathsf{C}\,(T,K_i^{\mathsf{c}}) - \max\left(0, F_t - K_i^{\mathsf{c}}\right)}{\left(K_i^{\mathsf{c}}\right)^2}, \tag{2.1}$$

F_t is the forward price and K_i^{c} is the i^{th} largest strike price for the call option. To implement the trapezoidal rule, we now need the option prices $\mathsf{C}\,(T, K_i^{\mathsf{c}})$, for $i = 1, \ldots, m$. Since some of these prices are not available, we apply a cubic spline interpolation method as proposed in Forsythe et al. (1977) to obtain the missing values.[22]

Then,

$$MIV_{t,\tau} = \frac{\overline{K}^{\mathsf{c}} - \underline{K}^{\mathsf{c}}}{m} \sum_{i=1}^{m} \left[g_{t,T}\left(K_i^{\mathsf{c}}\right) + g_{t,T}\left(K_{i-1}^{\mathsf{C}}\right)\right], \tag{2.2}$$

where $\tau = T - t$ denotes the time horizon or time to maturity. As mentioned above, we replace F_t in equation (2.1) by the futures price. Since in-the-money options are less liquid, equation (2.2) is evaluated for out-of-the money options whose strike prices are no less than $0.94 \times F_t$ (calls) or no bigger than $1.06 \times F_t$ (puts)[23]. Finally, we set $m = 100$ and restrict $\mathrm{MIV}_{t,\tau} = 0$ when $t = T$.

We estimate the MIV at the end of each month for a $\tau = 30$ day horizon to get our monthly time series, denoted $MIV_t^{(i)}$, where $i = \{30y, 10y, 5y, E\}$ stands for either the 30 year, 10 year of 5 year Treasuries or the equity index [24].

2.3.2 Realized Variance

To estimate $\mathbb{E}_t^{\mathbb{P}}\left(\int_t^T \sigma_u^2 du\right)$, the daily expected variance under the physical measure, we first consider the daily realized variance $RV_{t,oneday}$, which is defined as:

$$RV_{t,oneday} \ = \ \sum_{i=1}^{M} r_{t,i}^2,$$

[22]Jiang and Tian (2005) take a different approach: They first calculate the implied volatilities of available options with the Black and Scholes formula, and then use the interpolation method to obtain the Black and Scholes implied volatilities of the unavailable options. Using these implied volatilities, they use the Black and Scholes formula again to obtain a continuum of option prices. They claim that their method can avoid the nonlinearity problem in the option prices. However, we find a direct use of the interpolation method on the option prices to be more robust.

[23]Following Jiang and Tian (2005) we use out-of-the-money puts to get prices for the in-the-money calls.

[24]Note that we drop the subscript τ as we focus on the monthly horizon.

where $r_{t,i} = \log P(t-1+i/M) - \log P(t-1+(i-1)/M)$ is the intra-daily log return in the i^{th} sub-interval of day t and $P(t-1+i/M)$ is the asset price at time $t-1+i/M$. For each day, we take $r_{t,i}$ between 7:25 and 14:00. Inline with Andersen et al. (2007a), we use five minute intervals to calculate $RV_{t,oneday}$.

The normalized monthly realized variation $RV_{t,mon}$ is defined by the average of the 21 daily measures.[25] The normalized weekly realized variation $RV_{t,week}$ is correspondingly defined by the average of the five daily measures:

$$RV_{t,week} = \frac{1}{5}\sum_{j=0}^{4} RV_{t-j,oneday}, \text{ and } RV_{t,mon} = \frac{1}{21}\sum_{j=0}^{20} RV_{t-j,oneday}.$$

To better capture the long memory behavior of volatility, Corsi (2009) proposes the heterogenous autoregressive model for realized variance using the daily, weekly and monthly realized variance estimates. Andersen, Bollerslev, and Diebold (2007)Andersen et al. (2007a) extend the standard HAR-RV model to show that the predictability for realized variance over different time intervals almost always comes from the continuous component of the total price variation, rather than the discontinuous jump component. Corsi, Pirino, and Ren(2010) introduce the concept of threshold bipower variation and show that it is well suited for estimating models of volatility dynamics where continuous and jump components are used as explanatory variables. They document that jumps can have a highly significant impact on the estimation of future volatility. Their HAR-TCJ model for forecasting daily realized variance is expressed as:

$$RV_{t+1,oneday} = \alpha + \beta_D \widehat{TC}_{t,oneday} + \beta_W \widehat{TC}_{t,week} + \beta_M \widehat{TC}_{t,mon} + \beta_J \widehat{TJ}_{t,oneday} + \varepsilon_{t+1},$$

where the threshold bipower variation measure is used to estimate the jump component, $\widehat{TJ}_{t,oneday} = I_{C-Tz>\Psi_\alpha} \times (RV_{t,oneday} - TBPV_t)^+$ and the continuous part $\widehat{TC}_{t,oneday} = RV_{t,oneday} - \widehat{TJ}_{t,oneday}$.[26]

This simple method avoids some difficulties in long memory time series modeling and the parameters can be consistently estimated by OLS. However, a Newey-West correction is needed to make appropriate statistical inference. Moreover, such a HAR-TCJ type model can be easily modified, for example, by adding extra covariates that contain predictive power.

We aim to obtain the monthly estimates directly, so we replace the daily realized variance $RV_{t+1,oneday}$ by the normalized monthly measure $\mathrm{RV}_{t+21,mon}$. Moreover, we include lagged estimates of implied variance to further improve the realized variance forecasts. Hence, we run

[25] On average, we have 21 trading days per month.

[26] The expression for the threshold bipower variation, $TBPV_t$, is given in Corsi, Pirino, and Ren(2010). We use the confidence level $\alpha = 99.9\%$.

the following OLS regression for the projection:

$$RV_{t+21,mon} = \alpha + \beta_D \widehat{TC}_{t,oneday} + \beta_W \widehat{TC}_{t,week} + \beta_M \widehat{TC}_{t,mon} + \beta_J \widehat{TJ}_{t,oneday} + \beta_{IV}' \mathbf{IV}(\mathbf{L})_t + \varepsilon_{t+21,mon}, \quad (2.3)$$

where $\mathbf{IV}(\mathbf{L})_t$ contains lagged implied variances up to lag L.[27]

We implement this regression using an expanding window. This allows us to obtain real-time forecasts $\widehat{RV}_{t+21,mon}$ for $RV_{t+21,mon}$ without any look ahead bias.[28] As the HAR-TCJ predictor for the one month horizon, denoted $RV_t^{(HARIVJ)}$, we use:

$$RV_t^{(HARIVJ)} = 21 \times \widehat{RV}_{t+21,mon},$$

where $\widehat{RV}_{t+21,mon}$ is the projected value from regression (2.3). Furthermore, we denote the simple realized variance estimator obtained from summing $RV_{t+1,oneday}$ over the past month $RV_t^{(RVs)}$.

The left Panels in Figure 2.2 show time series plots for the annualized monthly implied volatility measures for the 30 year, 10 year, and 5 year Treasury bond futures, respectively as well as the S&P500 index (i.e. we take the square root of the corresponding variance measures to make the magnitudes comparable to the VIX).[29] The right Panels plot the realized volatility measures. The first two Panels in Table 2.1 present summary statistics of implied and realized volatility measures. Again, all numbers shown are annualized and expressed in percent. The implied volatility measures are on average larger than the realized quantities, both for the Treasury and the equity market, implying a positive variance risk premium. The equity implied and realized volatility is notably higher than the measures for the Treasury markets. The magnitudes of Treasury volatilities are increasing with the maturity of the underlying bonds. Moreover, all measures exhibit positive skewness and excess kurtosis. The autocorrelation coefficients range between roughly 70% and 80%. Finally, as previously mentioned, the summary statistics of the implied volatility measure calculated using options on S&P500 index futures are almost identical to the summary statistics of the original VIX.

Remark: In principle, there exist many different measures of realized variance and a priori, it is not clear what measure we should use. In the Online Appendix, we show that the HAR-TCJ model augmented by lagged implied variance terms performs the best when predicting out-of-sample future variance. The results are robust to the different loss functions we use to evaluate the performance.

[27] We choose the lag length to be four using the Akaike and Bayesian information criteria.

[28] We use daily realized variance estimates from the first 222 trading days as the input for initial estimation: Daily realized variances from day 1 to day 200 are used to construct $RV_{t,oneday}$, $RV_{t,week}$, and $RV_{t,mon}$. Daily realized variances from day 22 to day 222 are used to construct $RV_{t+21,mon}$. On day 222, the first out of sample forecast $\widehat{RV}_{t+21,mon}$ from the fitted model is constructed by using $RV_{222,oneday}$, $RV_{222,week}$, and $RV_{222,mon}$ as the input data to the initial fitted model. The same method is applied for day $223, 224, \ldots$ with the corresponding parameters.

[29] As mentioned before, we use options on S&P500 index futures to be consistent with our calculations for the Treasury implied variance measure.

2.3.3 Variance Risk Premia

We define the variance risk premium for horizon τ as follows:

$$VRP_{t,\tau} \equiv \mathbb{E}_t^{\mathbb{Q}} \left(\int_t^T \sigma_u^2 du \right) - \mathbb{E}_t^{\mathbb{P}} \left(\int_t^T \sigma_u^2 du \right),$$

where $\tau = T - t$ denotes the time horizon.[30] Economic theory suggests that the variance risk premium should be positive in order to compensate investors who bear risks from expected price fluctuations. The general positiveness of the variance risk premia can be confirmed empirically from comparing the means of the different volatility measures in Table 2.1.

In Figure 2.3 we plot the annualized variance risk premia (expressed in percent), defined as the difference between the model-free implied variance $MIV^{(i)}$ and the realized variance estimate $RV^{(HARIVJ,i)}$ for the 30 year, 10 year and 5 year Treasury futures ($VRP^{(30y)}$, $VRP^{(10y)}$ and $VRP^{(5y)}$, respectively), and the S&P 500 index futures ($VRP^{(E)}$). As we can see, the three Treasury time series share a lot of co-movement: The unconditional correlations between the 5 year, 10 year and 30 year bond variance risk premia is between 57% and 75%. We also note that $VRP^{(30y)}$ displays the largest volatility, especially during crisis periods indicated by the shaded areas. The bond variance risk premia are positive on average but they change sign. In contrast, the equity variance risk premium $VRP^{(E)}$ is essentially always positive and on average significantly higher in magnitude. The correlations between the bond and the equity variance risk premia range between 44% ($VRP^{(5y)}$) and 66% ($VRP^{(30y)}$). The summary statistics of the annualized variance risk premia expressed in percent are reported in Table 2.1, Panel C.

2.3.4 Treasury Implied Volatility (TIV)

In this section, we introduce a measure for Treasury Implied Volatility in the spirit of the VIX. The TIV measure is the square root of the one month implied variance for futures on 30 year Treasuries, $MIV^{(30y)}$. The top Panel in Figure 2.4 plots the annualized TIV measure and our VIX measure (backed out from options on futures) for the common sample period 1983 to 2010. The unconditional correlation between the two monthly time series is 46%. The unconditional correlation between the TIV measure and the original VIX for the period 1990 to 2010 is 62%.[31]

The construction of the TIV measure has an economic merit, which goes beyond that of the VIX measure alone. First, the TIV measure can be related to funding liquidity in Treasury markets and second, the spread between the VIX and the TIV can be interpreted as a proxy

[30] For notational simplicity, we subsequently drop the subscript τ as we always consider the one month horizon.

[31] The correlation between the TIV and our VIX measure for the same time period is exactly the same, which is not surprising given the near perfect correlation between the original VIX calculated using options on the cash index and our measure calculated using options on futures. We also calculate the implied volatilities using simple returns, STIV and SVIX. The correlation for the full sample period is 49% and the correlation since 1990 is 64%.

for flight to quality.

Theoretical work by Gromb and Vayanos (2002) and Brunnermeier and Pedersen (2009) predicts that higher volatility leads to a tightening of funding constraints for market makers. Moreover, Fontaine and Garcia (2011) establish a robust link between market uncertainty (measured by the VIX) and funding liquidity. In their paper, funding liquidity is estimated through price differentials of Treasuries of different age. Empirically, our TIV measure is strongly related to funding liquidity. When regressing the funding liquidity proxy on the TIV measure, we find a highly significant slope coefficient with a t-statistic larger than five and an R^2 of 30%.[32] The relationship is also economically significant: A one standard deviation change in the TIV implies more than half a standard deviation change in the funding liquidity proxy. Thus, spikes in the TIV can be interpreted as funding shocks, which lead to an increase in the funding liquidity premium.

Brunnermeier and Pedersen (2009) also show that in periods of flight to quality, highly liquid assets are characterized by relatively low volatility. As a consequence, the volatility spread between low and high volatility assets explains part of the liquidity spread. We plot the volatility spread between the VIX and the TIV in Figure 2.4 (middle Panel). Most of the time, the spread is no larger than 10% but sometimes experiences sudden extreme spikes. For example during the October 1987 crash, the LTCM default in August 1998 or the Lehman default in September 2008, the spread tripled within a month. These are periods usually associated with flight to quality (see, e.g. Caballero and Krishnamurthy (2008).

Apart from the TIV measure, there is a Treasury option volatility measure available in the market, the Merrill Lynch Option Volatility Estimate (MOVE) index. The MOVE is a yield curve weighted index of the normalized implied volatility on one month Treasury options, which are weighted on the 2, 5, 10, and 30 year contracts. This index is available since 1988. The bottom Panel of Figure 2.4 plots our TIV measure along with the MOVE index. The correlation for the common time period is 81%. Two main differences between the MOVE and the TIV make the latter a more appealing proxy for volatility risk in the Treasury market. First, the MOVE is calculated using Treasury options while we use options on Treasury futures. Treasury options are options on benchmark Treasury securities, which are not exchange traded and hence are significantly less liquid. In practice, they are thus marked at some fixed spread to swaptions. Second, there is no transparent market for out-of-the money Treasury options, so the MOVE index is calculated using the Black (1976) model to compute implied volatility of at-the-money options.[33] In the next section, we study the determinants of the variance risk premia and document the strong predictive power of the bond variance risk premia for excess returns on Treasury, stock, corporate bond and CMBS portfolios.

[32] We thank Ren Garcia for sharing the data.

[33] Moreover, the MOVE would not suitable to calculate bond variance risk premia as this would require high frequency data on benchmark Treasury securities.

2.4 Empirical Evidence

In this section, we first investigate the economic determinants of bond and equity variance risk premia. Inline with intuition, we find that the variance risk premia are largely driven by uncertainty about real and nominal variables. Secondly, we study the predictive power of the Treasury bond variance risk premia for Treasury, stock, corporate bond and mortgage-backed security excess returns. We do this univariate and multivariate, i.e. we run regressions using only information from the bond variance risk premia measures as regressors before including additional explanatory variables. Due to multicollinearity concerns with regards to the bond variance risk premia, we perform a principal components analysis and use the three principal components as regressors instead of the individual Treasury variance risk premia. We find that estimated coefficients of bond variance risk premia are both economically and statistically significant, in- and out-of-sample, even if we include standard predictors suggested in the literature.

We calculate the variance risk premia using the methods described in the previous section. In our main specification, the variance risk premium is the difference between the model-free implied variance and the augmented HAR-TCJ projection. However, the results in this section are also robust to using other IV or RV measures.[34]

To study the determinants and predictability of bond variance risk premia, we choose July 1991 to June 2010 as the sample period. Starting in mid 1990, we have data for all variance risk premia and we can calculate the principal components. However, since we calculate the variance risk premia using an expanding window to remove any look-ahead bias, we allow for a burn-in period of one year.

2.4.1 What Drives Bond Variance Risk Premia?

It is natural to assume that variance risk premia are associated with higher uncertainty. Options provide investors with a hedge against high variance in the underlying returns and high variance usually occurs when unexpected shocks affect macroeconomic variables. The premium that investors are willing to pay or receive to hedge against such events is related to their uncertainty. Equilibrium models that study variance risk premia focus on the equity market only. Drechsler and Yaron (2011) link the variance risk premium of the stock market index to uncertainty about fundamentals. In particular, time variation in economic uncertainty and a preference for early resolution of uncertainty are required to generate a positive variance premium that is time varying and predicts excess stock market returns. Drechsler (2010) reports a high correlation between the variance risk premium and the dispersion in the forecasts of next quarter's real GDP growth from the Survey of Professional Forecasters.

To examine whether uncertainty affects bond variance risk premia as well, we regress the

[34] We also implement implied variance measures based on simple variance swaps (as in Martin (2011)) and by inverting the Black (1976) formula for ATM options. As for the realized variance measures, the results are for example robust to using the standard HAR-RV projection proposed by Corsi (2009).

monthly variance risk premia measures on uncertainty factors constructed from BlueChip Economic Indicator forecast data. We proxy for uncertainty about the real and nominal side of the economy by the cross sectional standard deviation of the forecasts of CPI ($\widehat{U}^{CPI}$) and real GDP ($\widehat{U}^{RGDP}$) for the current and the next calendar year, respectively. We also construct uncertainty proxies for the three month Treasury bill rate (which can be interpreted as uncertainty about monetary policy) and the ten year Treasury note yield ($\widehat{U}^{SR}$ and $\widehat{U}^{LR}$, respectively).[35] In addition to the uncertainty measures, we include two variables that measure the time-varying volatility of inflation and consumption (σ_π and σ_g, respectively) and two macro factors that can be interpreted as a real ($\widehat{F}_1$) and a nominal (or inflation) factor ($\widehat{F}_2$) in the regression. The macro volatilities are calculated by estimating a GARCH(1,1) process using monthly CPI and per capita consumption (non-durables and services). The macro factors are constructed using the first two principal components of a large set of macro variables as in Ludvigson and Ng (2009, 2011).[36] Finally, we add the slope of the term structure (SLOPE) and the refinancing index from the Mortgage Bankers Association (REF) as additional regressors. Hence, we run the following regression:

$$VRP_t^{(i)} = \beta'^U \widehat{\mathbf{U}}_t + \beta'^F \widehat{\mathbf{F}}_t + \beta'^S \widehat{\mathbf{S}}_t + \epsilon_t^{(i)},$$

where $VRP_t^{(i)}$ is the bond or equity variance risk premium ($i = \{30y, 10y, 5y, E\}$) at time t, $\widehat{\mathbf{U}}_t$ is a vector of the uncertainty measures, $\widehat{\mathbf{F}}_t$ contains the real and nominal macro factors, and $\widehat{\mathbf{S}}_t$ contains the macro volatilities, the slope of the term structure and the refinancing index. $\epsilon_t^{(i)}$ is the error term. All coefficients are estimated with ordinary-least squares and standardized to allow for a straightforward assessment of the economic significance. We report t-statistics that are calculated using Newey and West (1987) standard errors. The sample spans the period from July 1991 to December 2009.[37]

In order to avoid multicollinearity concerns in this regression, we consider the cross-correlations of the potential determinants presented in Table 2.2. The uncertainty proxies for the short and the long end of the yield curve are positively correlated, but the coefficient is merely 42%. Both uncertainty measures also exhibit a positive correlation with uncertainty about real GDP. Apart from this, only $\widehat{F}_1$ and σ_π exhibit sizable correlations with other determinants: the correlation for the two time series is -0.61. Moreover, $\widehat{F}_1$ and σ_π are correlated with $\widehat{U}^{CPI}$ (-0.74 and 0.57, respectively). Overall, correlations are fairly moderate and give no reason for concern.

The determinant regression results are presented in Table 2.3. In summary, the results confirm that uncertainty variables have relevant explanatory power for variance risk premia

[35]See Section 2.2 for details.

[36]Given that the factors are principal components, the economic interpretation is not straightforward. We calculate the marginal correlations of the individual time series with the respective factors for our data set. As in Ludvigson and Ng (2009), it is reasonable to interpret the first factor as a real factor. The second factor can be interpreted as a nominal or inflation factor. See the Online Appendix for additional information.

[37]We only have the BlueChip data available until the end of 2009.

but there are distinct differences between the Treasury maturities and also between Treasury and equity variance risk premia.

The four uncertainty factors alone explain around 45% of the variation in 30 year bond and equity variance risk premia. For 10 year and 5 year bond variance risk premia, this number drops to 21% and 12%, respectively. In general, higher uncertainty is associated with an increase in bond and equity variance risk premia. However, uncertainty about the short rate is a notable exception, as 30 year Treasury and equity variance risk premia significantly decrease given an increase in uncertainty about monetary policy. Intuitively, we can explain this as follows: Larger uncertainty about the short end makes investors with negative expectations more willing to buy long term bonds, because these provide a hedge against an increased duration due to a drop in the short term yields. Hence, investors pay a premium for holding these bonds. It should be noted, however, that interpreting signs with proxies of uncertainty can be difficult. First, the impact of uncertainty on risk premia in equilibrium models usually depends on the amount of pessimists versus optimists, where pessimists are those agents who have a lower than the consensus forecast (see, e.g., Jouini and Napp (2007) for equity markets, and Xiong and Yan (2010), for bond markets). Uncertainty only implies a positive impact on risk premia if wealth-weighted beliefs are dominated by pessimists. Second, it is not mandatory that higher uncertainty is always associated with worsening economic conditions (see, e.g., Patton and Timmermann (2010)).

Uncertainty about inflation has the largest economic impact on the variance risk premia: A one standard deviation change in inflation uncertainty implies on average almost half a standard deviation change in the variance risk premia. Inline with intuition, uncertainty about the long rate predominantly affects 30 year bond variance risk premia and uncertainty about real GDP is only significant for equity variance risk premia. The shape of the term structure, i.e. the slope, affects both bond and equity variance risk premia. A one standard deviation move in the slope moves bond variance risk premia by between 0.2 and 0.35 standard deviations.

These results are robust to adding levels and volatilities of macro variables to the regression. The macro volatilities are at most marginally significant, while the real factor has a negative effect on both bond and equity variance risk premia with an increasing statistical significance in the maturity of the underlying. For bond variance risk premia, the coefficient is around -0.14 and the effect almost doubles for equity variance risk premia. In addition, the price factor has a significantly positive effect on equity variance risk premia. At the same time, adding $\widehat{F}_1$ and $\widehat{F}_2$ to the bond variance risk premia regressions only marginally improves the adjusted R^2, further supporting the notion that variance risk premia are driven predominantly by uncertainty and not by actual macro fundamentals.[38]

As shown in Figure 2.1, the MBA refinancing index is highly correlated with Treasury variance risk premia. In a regression of the individual variance risk premia on the refinancing index, the coefficients are positive and strongly significant. A one standard deviation move in

[38]Adding even more macro factors does not further improve the fit of the regression.

the refinancing index is associated with 0.3 to 0.45 standard deviation moves in the variance risk premia. In the multivariate regressions, however, the effect is muted. The strong statistical significance only remains for the equity variance risk premia, while the uncertainty proxies largely drive out the refinancing index for bond variance risk premia.

2.4.2 Principal Components

Next we want to assess the predictive power of bond variance risk premia for different assets. To this end, we regress excess returns on bond variance risk premia. Since the average pairwise correlation between the individual bond variance risk premia is very high, we calculate the principal components of the bond variance risk premia to circumvent the issue of multicollinearity.[39] The principal components analysis allows us to summarize in a parsimonious way the information in the term structure of bond variance risk premia.

We denote by $VRP^{(PC1)}$, $VRP^{(PC2)}$ and $VRP^{(PC3)}$, the first, second, and third principal component, respectively. The first principal component explains roughly 77% of the variation in the Treasury variance risk premia, while the second and third principal components explain the remaining 15% and 8%, respectively. Table 2.4 reports the factor loadings for the three bond variance risk premia. It seems appropriate to interpret the three principal components in analogy to the term structure literature as level, slope and curvature (see, e.g., Litterman and Scheinkman (1991)).

In interpreting the loadings, one has to keep in mind our setup, where the horizon for the variance risk premia is constant while the underlying bond maturity is changing. To calculate a term structure in the usual sense, we would need longer maturity options for each futures. Due to a lack of liquidity and availability of longer maturity options, this is not possible.[40]

Adding the equity variance risk premium to the principal components analysis does not significantly alter the pattern. The first principal component still explains almost 70% of the total variation and can be interpreted as a level factor for the Treasury variance risk premia (the correlation with $VRP^{(PC1)}$ is 98%). The second factor explains 15% and can be interpreted as a slope factor for bond variance risk premia. Moreover, the equity variance risk premium strongly loads on this factor. The correlation with $VRP^{(PC2)}$ is 53%. The third factor explains 10% and is again a slope factor, exhibiting a correlation with $VRP^{(PC2)}$ of almost 85%. Finally, the last factor is a curvature factor. It explains the remaining 5% of the variation and its correlation with $VRP^{(PC3)}$ is 84%.

In the next section, we use the three principal components instead of the individual variance risk premia to examine the predictive ability of the bond variance risk premia. However, we include the regression results using the original variance risk premia in the Online Appendix. The principal components analysis allows to better understand some of the regression results

[39] The pairwise correlations between the bond variance risk premia range between 57% and 75% for our sample.

[40] Feunou et al. (2011) estimate a term structure of uncertainty using equity options. They use multiple horizons for the same underlying and then perform a principal components analysis. They find that the first three principal components can also be interpreted as level, slope and curvature.

using the actual variance risk premia. It turns out for example that the third PC, the curvature factor, has significant predictive power for Treasury bond excess returns. As shown in Table 2.4, the loadings of the 30 year and the 10 year Treasury variance risk premia on this factor are exactly opposite. Hence, it is not too surprising that the sign of the coefficients for these two variance risk premia is exactly opposite as well.

2.4.3 In-Sample Predictability

Using the principal components for the Treasury variance risk premia, we study the in-sample predictive power of bond variance risk premia for fixed income and equity excess returns. To do this, we run the following type of regression:

$$rx_{t+h}^{(i)} = \beta_h'^{(i)} \mathbf{VRP}_t + \gamma_h'^{(i)} \mathbf{M}_t + \epsilon_{t+h}^{(i)},$$

where $rx_{t+h}^{(i)}$ denotes the h-period excess returns for asset i. VRP_t is a vector containing the principal components of the bond variance risk premia, $VRP^{(PC1)}$, $VRP^{(PC2)}$ and $VRP^{(PC3)}$. $\mathbf{M}_t$ denotes a vector of additional predictor variables and $\epsilon_{t+h}^{(i)}$ is the error term.

We calculate excess returns for two to five year Treasury bonds, the stock market, a growth and value portfolio, corporate bond indices for AAA, BBB and CCC rated securities, and commercial mortgage-backed securities indices for AAA, BBB and B rated securities. For Treasuries, we only calculate annual excess returns, whereas for all other assets we calculate three, six and twelve month excess returns as the difference between the respective portfolio returns and the corresponding Treasury rate.

Note that we always report standardized regression results, meaning that, for all regressors and regressands, we de-mean and divide by the standard deviation. This makes coefficients comparable across different predictors and allows to directly interpret not only the statistical but also the economic significance. We report t-statistics that are calculated using Newey and West (1987) standard errors. The sample period is always from July 1991 to June 2010, except for CMBS excess returns, which are only available starting in January 1997.

For the fixed income excess return regressions (Treasuries, corporates, CMBS), $\mathbf{M}_t$ includes the equity variance risk premium, the Cochrane and Piazzesi (2005) factor, CP, and the eight macro factors from Ludvigson and Ng (2009, 2011), $\widehat{F}_j, j = 1 \ldots, 8$. For the stock portfolio excess return regressions, we include the log dividend yield, DY, the log earnings to price ratio, E/P, and NTIS, the net equity expansion as additional regressors as in Goyal and Welch (2008).[41]

In Tables 2.5 and 2.6 we present the regression results excluding the additional control variables.[42] Panel A in Table 2.5 contains the results for Treasury bond excess returns. The

[41] We exclude some of the additional regressors from Goyal and Welch (2008) as the updated data is not available. We also exclude the book-to-market ratio as it exhibits a correlation of almost 80% with DY over the sample period. Moreover, it is not significant.

[42] Slightly abusing the language, we refer to these results as the univariate regression results.

coefficients for the third PC are significant and negative for all maturities, implying that spikes in this PC lead to lower excess returns. A one standard deviation positive shock roughly results in a 0.18 standard deviation reduction in bond excess returns for all maturities. The third PC is the curvature factor, meaning that an increase in the curvature of bond variance risk premia predicts lower Treasury bond excess returns. The average adjusted R^2 is roughly 3%.

Univariate regression results for stock excess returns are reported in Panel B of Table 2.5. We find predictability in the second and third principal components for the market, growth and value portfolio excess returns for horizons between six and twelve months. The statistical and economic significance for the second PC, the slope factor, is particularly high. A one standard deviation move in the slope factor results in 0.2 to 0.3 standard deviation larger excess returns for longer horizons. The third PC is significant for the market and growth portfolios. However, unlike for Treasury bond excess returns, the coefficient for stock excess returns is now positive and is estimated at 0.17 for six month excess returns. Adjusted R^2 range between 7% and 9% for the market and growth portfolio and reach 11% for the value portfolio.

Table 2.6 presents the regression results for long and intermediate corporate bond excess returns (Panels A and B), and commercial mortgage-backed securities excess returns (Panel C). Overall, the predictability is strong with adjusted R^2 ranging between 5% for three month excess returns on intermediate AAA rated bonds and reaching 38% for twelve month BBB rated CMBS. The predictive power increases with the horizon and is strongest for intermediate rating categories, i.e. BBB rated securities. Predictability for intermediate maturity bonds is slightly higher than for long maturity bonds. Unlike for Treasuries and stocks, the curvature factor does not contain any predictive power. However, the first PC, the level factor, is very strongly significant, both statistically and empirically. For twelve month corporate bond excess returns, the coefficients range between 0.35 for intermediate AAA bonds and 0.52 for intermediate BBB bonds. For CMBS, the first PC is strongly significant for AAA and BBB rated securities and six to twelve month excess returns. The second PC is also almost uniformly strongly significant at all horizons. However, it works less well for high yield corporate bonds and CMBS. As with the first PC, the coefficients almost reach 0.5 for six and twelve month excess returns.

To summarize the univariate regressions, bond variance risk premia have significant predictive power for a wide range of assets at various horizons. It is also worth noting that the predictability is not concentrated in one specific latent factor. Relevant information is contained in the whole "term structure" of bond variance risk premia. In a separate analysis, we also run truly univariate regressions, i.e. using only one variance risk premium at a time. The main results hold, meaning that we do find predictability. However, as the predictability for Treasury excess returns for example is mainly contained in the third PC, it is not always straightforward to pick it up. In addition, the variance risk premia may load with different signs on principal components with predictive power, further complicating the detection and interpretation of the predictability when using only one bond variance risk premium time series at a time.

To check the robustness of our univariate results, we next add different established predictors of bond and equity excess returns to the regression. The results are reported in Tables 2.7 to 2.10.

To summarize, the results from the multivariate regressions with respect to the principal components of Treasury variance risk premia are remarkably robust to the inclusion of a host of control variables. Moreover, Treasury variance risk premia truly seem to pick up information that is relevant for a wide range of asset classes and that goes beyond what is contained in standard macroeconomic variables and the term structure of interest rates. In particular, the predictive power of Treasury variance risk premia is much more general than the documented predictability of the equity market variance risk premium, which predominantly works for stock excess returns.

The coefficients for Treasury bond excess returns presented in Table 2.7 are still significantly negative and economically relevant. Now, all three PCs are significant with the same sign. The coefficients range between -0.14 and -0.27. Thus, an increase in the level, slope or curvature of the term structure of bond variance risk premia results in lower Treasury excess returns. Including the CP factor and the macro factors increases the adjusted R^2 to almost 30% across all maturities for the Treasury bond excess return regressions. As in Ludvigson and Ng (2009), the macro factors explain a significant fraction of the variation in bond excess returns over the sample period, and the CP factor is highly significant as well. Unlike the bond variance risk premia, the equity variance risk premium does not seem to contain any relevant information for forecasting bond excess returns at an annual horizon.[43]

Table 2.8 presents the results for the stock excess return regressions including as additional control variables the equity market variance risk premium, the dividend yield (DY), the earnings to price ratio (E/P) and net equity expansion (NTIS). The equity variance risk premium has significant predictive power for the market and growth portfolios, a result which is inline with the findings in Bollerslev et al. (2010). NTIS is strongly significant for six and twelve month excess returns, while DY only has predictive power for twelve month excess returns on the market and growth portfolios. As in the univariate regressions, the second and third PCs are statistically significant for six to twelve month excess returns while the economic significance remains largely unchanged. Overall, the adjusted R^2 raise to between 30% and 40% for twelve month excess returns.

Not very surprisingly at this stage, the univariate results for corporate bonds and CMBS largely carry over to the multivariate regressions.[44] The second PC, i.e. the slope factor emerges as the strongest and most robust predictor while the first PC seems to be driven out at short horizons by the additional controls, the CP factor in particular.

[43]These findings echo the results in Mueller et al. (2011) who find that the equity variance risk premium heavily loads on short-term bond risk premia but does not predict excess returns at the annual horizon.

[44]We report the regression results for intermediate corporate bond and CMBS excess returns in Tables 2.9 and 2.10. The results for long-term corporate bonds are qualitatively similar but in the interest of space they are deferred to the Online Appendix.

In summary, we find that excess returns on Treasuries, stocks, corporate bonds and CMBS are predictable using bond variance risk premia. Overall, the reported in-sample predictability is strong, both statistically and economically.

2.4.4 Out-of-Sample Predictability

There is ample evidence in the literature for the fact that in-sample predictability does not necessarily imply out-of-sample predictability. For instance, Goyal and Welch (2008) show that a large number of predictors have very little out-of-sample predictive power for stock market excess returns and they attribute the inconsistent out-of-sample performance of individual predictive regression models to structural instability. In this section, we report results on the out-of-sample forecasting performance of the regression models studied in the previous section. We run out-of-sample regressions using the bond variance risk premia directly instead of the principal components, since these are estimated using an expanding window. We therefore use data only through time t for forecasting excess returns at time $t+1$.

We run the out-of-sample test for the Treasury, long corporate bond, and the stock excess returns but not the mortgage-backed securities, as the available sample period for these is too short. We obtain the initial estimates based on the period from July 1991 to July 1999 and study the out-of-sample predictability for the period starting in July 2000 and ending in June 2010.

For the corporate bond and stock excess returns, we compare the out-of-sample forecasting performance of the bond variance risk premia to a constant expected returns benchmark. For the Treasury excess returns, we have two different model specifications. First, we compare the out-of-sample forecasting performance of the bond variance risk premia to a constant expected returns benchmark where, apart from an MA(12) error term, excess returns are unforecastable as in the expectations hypothesis. Second, since the expectations hypothesis is violated in the data, we compare the out-of-sample forecasting performance of a specification that includes bond variance risk premia plus the CP factor to a benchmark model that only includes CP factor and a constant.

To check whether the bond variance premia have out-of-sample predictive power, we consider two different metrics for the evaluation. The first one is the out-of-sample R^2 statistic, denoted by R^2_{OOS} (Campbell and Thompson, 2008). The R^2_{OOS} measures the proportional reduction in the mean squared error for a competing model relative to the benchmark forecast. It is akin to the in-sample R^2 and has the following form:

$$R^2_{\text{OOS}} = 1 - \frac{\frac{1}{T}\sum_{t=1}^{T}\left(y_t - \hat{y}_t^i\right)^2}{\frac{1}{T}\sum_{t=1}^{T}\left(y_t - \hat{y}_t^j\right)^2},$$

where $\hat{y}_t^i$ is the forecast of the variable y_t from the competing model i and $\hat{y}_t^j$ is the forecast of variable y_t from the benchmark model j. Note that both $\hat{y}_t^i$ and $\hat{y}_t^j$ are obtained based on the data up to period $t-1$. If the R^2_{OOS} is positive, the competing predictive regression has a lower average out-of-sample mean-squared prediction error than the benchmark.

The second metric we employ is the ENC-NEW test statistic of Clark and McCracken (2001). The null hypothesis of the ENC-NEW test is that the model with additional variables does not have better predictive power for excess returns than the benchmark or restricted model. The alternative is that these additional variables have additional information and they could be used to obtain a better forecast.[45]

We report the results in Table 2.11. The R^2_{OOS} are positive for all assets, which indicates that a model which includes the bond variance risk premia improves both over the constant expected returns benchmark and a model that includes the CP factor. We draw the same conclusion from the reported ENC-NEW statistics. The test statistics reveal that the improvement in forecasting power is strongly statistically significant.

2.5 Conclusion

In this paper, we construct daily measures of implied and realized variance in fixed income markets. To calculate the implied variance, we use daily options on 30 year, 10 year and 5 year Treasury futures. We use high frequency futures data to calculate the realized variance. Using thirty years of options data, we propose a new measure of fear for Treasury markets, the TIV, calculated as the square root of the implied variance for 30 year Treasury futures. The TIV measure has two interesting properties: First, it is strongly related to a proxy of funding liquidity and second, the spread between the VIX and TIV can be interpreted as a measure of flight to quality. The behavior of the TIV resembles in many ways the one of the VIX, however, while the unconditional correlation is high, the two series differ in their magnitudes. Not very surprisingly, we find that the compensation for variance risk in fixed income markets is considerably smaller than in the equity market.

The bond variance risk premia we derive are akin to the variance risk premium for the S&P 500 index. However, while the variance risk premium for the equity index is essentially always positive, i.e. it acts like an insurance premium, the variance risk premia in Treasury markets can turn negative. To grasp a better intuition of this behavior, we explore the economic determinants of both equity and bond variance risk premia in more detail. We find that both are strongly driven by proxies of macroeconomic uncertainty, however, a proxy of uncertainty about monetary policy strongly reduces the variance risk premia in bond markets, especially for longer maturities. We also find that the shape of the term structure of Treasury yields significantly affects bond variance risk premia. These findings corroborate the intuition that bond variance risk premia encompass general macroeconomic uncertainty together with information from the term structure of yields.

[45] The limiting distribution of ENC-NEW is non-standard. Following Ludvigson and Ng (2009), we base our inference on comparing the calculated test statistics with the corresponding 95th percentile of the asymptotic distribution of the ENC-NEW test statistic. Critical values can be found in Clark and McCracken (2001).

We then study whether bond variance risk premia contain any predictive power for excess returns on Treasury bonds, stocks, corporate bonds and commercial mortgage-backed securities. We find that bond variance risk premia explain a significant proportion of the variation in excess returns both in-sample and out-of-sample. Moreover, the predictive power is very robust to the horizon and in particular the inclusion of standard predictors found in the literature.

We are primarily interested in documenting the facts. Is there a compensation for volatility risk in fixed income markets? If yes, how large is it? How does it compare to equity markets? However, ultimately, our paper remains agnostic about the theoretical underpinnings of the empirical results but instead raises new research questions. First, many papers have argued that in times of uncertainty, there is a so-called flight to quality, i.e. investors move from relatively more risky stocks to relatively less risky bonds. Comparing the two implied volatility measures for equity and bond markets, we find that during certain periods, there is a decoupling of the two time series while during other time periods, the time series almost move in lock step. A next step would be to study the lead and lag relationships and the feedback effects between the equity and bond markets in greater detail using the implied volatility measures we introduce for Treasuries.

Secondly, if bond variance risk premia have predictive power across different asset classes, we would expect that these risk premia pick up more than just information contained in the term structure of interest rates. Since we do not find a similar result for the equity variance risk premium, we conclude that bond variance risk premia capture macroeconomic uncertainty which goes beyond that contained in equity variance risk. Moreover, there is also information contained in the term structure of bond variance risk premia that is orthogonal to what can be learned from standard macroeconomic variables. In particular, our results hint that the term structure of uncertainty, i.e. how uncertainty evolves at any given point in time for different horizons, could have opposing effects on variance risk premia. Exploring the effects of different uncertainty horizons on risk compensation is another interesting topic for future research.

Table 2.1: Summary Statistics. Panels A and B report summary statistics for the implied and realized volatility measures. MIV denotes the model-free implied variance for a one month horizon. $RV^{(TCJ)}$ denotes monthly realized variance sampled at the five minute frequency, and $RV^{(5m)}$ denotes the HAR-TCJ realized variance estimator augmented with lagged implied variance terms. The VIX is obtained from CBOE. All quantities are annualized and expressed in percent. Note that we report the summary statistics for the implied and realized volatilities, which are obtained as the square root of the respective variance measures. Panel C presents summary statistics for bond and equity variance risk premia. The variance risk premia are annualized and expressed in percent. They are calculated as the difference of the model-free implied variance and the projected value from the HAR-TCJ realized variance estimator augmented with lagged implied variance terms. All data is monthly and the sample spans the period from July 1991 to June 2010.

PANEL A: IMPLIED VOLATILITY					
	30y Treasury	10y Treasury	5y Treasury	S&P500	
	MIV	MIV	MIV	MIV	VIX
Mean	10.00	6.86	4.43	20.11	20.22
StDev	2.38	1.65	1.22	8.39	8.13
Min	6.03	3.70	1.95	9.97	10.42
Max	21.96	13.26	9.53	58.46	58.89
Skewness	1.95	0.68	0.78	1.55	1.61
Kurtosis	8.92	3.97	4.15	6.53	6.80
AC(1)	0.83	0.72	0.72	0.86	0.86

PANEL B: REALIZED VOLATILITY								
	30y Treasury		10y Treasury		5y Treasury		S&P500	
	$RV^{(TCJ)}$	$RV^{(5m)}$	$RV^{(TCJ)}$	$RV^{(5m)}$	$RV^{(TCJ)}$	$RV^{(5m)}$	$RV^{(TCJ)}$	$RV^{(5m)}$
Mean	8.43	8.38	5.57	5.52	3.77	3.72	14.65	14.22
StDev	1.42	2.03	0.80	1.45	0.57	1.04	6.10	7.97
Min	6.25	4.77	4.07	2.87	2.72	1.86	7.37	5.04
Max	15.95	18.46	9.24	10.79	6.71	7.55	51.80	73.79
Skewness	1.99	1.22	1.48	0.83	1.50	0.84	2.18	3.13
Kurtosis	8.76	5.79	6.59	3.89	7.69	3.63	11.20	19.45
AC(1)	0.84	0.71	0.78	0.67	0.76	0.62	0.83	0.77

PANEL C: VARIANCE RISK PREMIUM				
	30y Treasury	10y Treasury	5y Treasury	S&P500
	VRP	VRP	VRP	VRP
Mean	0.33	0.18	0.07	2.23
StDev	0.35	0.17	0.08	2.19
Min	-0.07	-0.16	-0.04	-0.01
Max	2.44	1.05	0.50	13.22
Skewness	3.14	1.30	1.88	2.38
Kurtosis	16.21	6.18	8.96	9.82
AC(1)	0.68	0.50	0.57	0.77

Table 2.2: Correlations of Determinants. The table reports correlations between determinants of variance risk premia. The uncertainty variables are defined as the cross sectional standard deviation of the forecasts of the short and long end of the term structure ($\widehat{U}^{(SR)}$ and $\widehat{U}^{(LR)}$), real GDP ($\widehat{U}^{(RGDP)}$), and CPI ($\widehat{U}^{(CPI)}$). SLOPE is the slope of the term structure calculated as the difference between the ten year and the one month Treasury yield. The macro volatilities σ_π and σ_g are estimated using a GARCH(1,1) process for inflation and per capita consumption (non durables and services). The macro variables $\widehat{F}_j, j = 1, 2$ are estimated as the first two principal components from a data set of 104 macroeconomic variables. They can be interpreted as a real and a nominal or inflation factor, respectively. REF is the refinancing index published by the Mortgage Bankers Association. All data is monthly and the sample spans the period from July 1991 to December 2009.

	$\widehat{U}^{LR}$	$\widehat{U}^{RGDP}$	$\widehat{U}^{CPI}$	SLOPE	σ_π	σ_g	$\widehat{F}_1$	$\widehat{F}_2$	REF
$\widehat{U}^{LR}$	0.42	0.42	0.00	-0.04	-0.28	0.47	-0.06	0.03	-0.24
$\widehat{U}^{RGDP}$		0.27	-0.03	0.04	-0.34	0.29	0.07	-0.08	-0.12
$\widehat{U}^{CPI}$			0.33	0.18	-0.07	0.32	-0.37	0.01	-0.06
SLOPE				0.12	0.57	-0.23	-0.74	0.07	0.26
σ_π					0.04	-0.01	-0.18	-0.10	0.09
σ_g						-0.46	-0.61	0.05	0.35
$\widehat{F}_1$							0.16	-0.02	-0.20
$\widehat{F}_2$								0.01	-0.40
REF									-0.09

Table 2.3: Economic Drivers of Bond Variance Risk Premia. The table reports the results from regressing the respective variance risk premia on uncertainty measures $\widehat{\mathbf{U}}_t$ and additional variables $\widehat{\mathbf{F}}_t$ and $\widehat{\mathbf{S}}_t$: $\text{VRP}_t^{(i)} = \beta'^U \widehat{\mathbf{U}}_t + \beta'^F \widehat{\mathbf{F}}_t + \beta'^S \widehat{\mathbf{S}}_t + \epsilon_t^{(i)}$. The uncertainty variables are defined as the cross sectional standard deviation of the forecasts of the short and long end of the term structure ($\widehat{U}^{(SR)}$ and $\widehat{U}^{(LR)}$), real GDP ($\widehat{U}^{(RGDP)}$), and CPI ($\widehat{U}^{(CPI)}$). $\widehat{\mathbf{S}}_t$ includes SLOPE, the slope of the term structure calculated as the difference between the ten year and the one month Treasury yield, the macro volatilities σ_π and σ_g estimated using a GARCH(1,1) process for inflation and per capita consumption (non durables and services) and REF, the MBA refinancing index. The macro variables $\widehat{F}_j, j = 1, 2$ are estimated as the first two principal components from a data set of 104 macroeconomic variables. They can be interpreted as a real and a nominal or inflation factor, respectively. Regressions are standardized, meaning all variables are de-meaned and divided by their standard deviation. Coefficients are estimated with ordinary-least squares, t-statistics are in parentheses and are calculated using Newey and West (1987)Newey and West (1987) standard errors. Data is monthly and the sample spans the period from July 1991 to December 2009.

	$VRP^{(30y)}$			$VRP^{(10y)}$			$VRP^{(5y)}$			$VRP^{(E)}$		
$\widehat{U}^{SR}$	-0.213	-0.160	-0.145	-0.062	-0.048	-0.034	-0.011	0.033	0.053	-0.319	-0.282	-0.252
	(-2.45)	(-2.05)	(-1.87)	(-0.51)	(-0.41)	(-0.29)	(-0.09)	(0.29)	(0.47)	(-4.40)	(-3.81)	(-4.89)
$\widehat{U}^{LR}$	0.225	0.215	0.214	0.160	0.109	0.105	0.070	0.007	0.015	0.123	0.158	0.172
	(3.98)	(3.62)	(3.00)	(1.76)	(1.25)	(1.12)	(0.68)	(0.06)	(0.15)	(1.39)	(1.98)	(2.94)
$\widehat{U}^{RGDP}$	0.025	0.007	-0.013	0.129	0.051	0.028	0.107	0.072	0.052	0.182	0.246	0.214
	(0.25)	(0.09)	(-0.17)	(1.32)	(0.57)	(0.31)	(0.97)	(0.60)	(0.44)	(2.13)	(2.73)	(2.10)
$\widehat{U}^{CPI}$	0.630	0.441	0.370	0.381	0.251	0.182	0.308	0.285	0.188	0.553	0.536	0.376
	(3.76)	(2.94)	(2.28)	(3.00)	(1.77)	(1.18)	(1.99)	(1.37)	(0.92)	(4.13)	(4.54)	(3.89)
SLOPE		0.284	0.256		0.370	0.347		0.239	0.194		-0.125	-0.199
		(3.57)	(3.28)		(4.42)	(4.19)		(2.61)	(2.18)		(-1.30)	(-2.57)
σ_π		0.168	0.090		0.096	0.019		-0.125	-0.228		0.009	-0.158
		(1.61)	(0.88)		(1.02)	(0.20)		(-0.86)	(-1.60)		(0.08)	(-1.86)
σ_g		-0.156	-0.155		-0.079	-0.079		-0.225	-0.221		-0.085	-0.078
		(-1.59)	(-1.60)		(-0.75)	(-0.78)		(-1.89)	(-1.95)		(-1.05)	(-1.07)
$\widehat{F}_1$			-0.137			-0.142			-0.169			-0.273
			(-1.68)			(-1.45)			(-1.39)			(-2.61)
$\widehat{F}_2$			-0.035			-0.061			0.037			0.075
			(-0.61)			(-1.02)			(0.84)			(2.20)
REF			0.114			0.099			0.182			0.294
			(1.58)			(1.09)			(1.58)			(2.79)
R^2	0.45	0.53	0.55	0.21	0.31	0.32	0.12	0.17	0.20	0.44	0.45	0.57

Table 2.4: Principal Components of Treasury Variance Risk Premia. The table reports the loadings for the three principal components of the bond variance risk premia, $VRP^{(PC1)}$, $VRP^{(PC2)}$ and $VRP^{(PC3)}$, respectively. We also report the the percentage of the explained variation. Data is monthly and the sample spans the period from July 1991 to June 2010.

	$VRP^{(PC1)}$	$VRP^{(PC2)}$	$VRP^{(PC3)}$
$VRP^{(30y)}$	0.58	0.57	0.59
$VRP^{(10y)}$	0.60	0.19	-0.77
$VRP^{(5y)}$	0.55	-0.80	0.23
Percent explained	77.33	14.93	7.74
Cum. percent explained	77.33	92.26	100.00

Table 2.5: Excess Bond and Stock Returns. We run the following regression: $rx^{(i)}_{t+h} = \beta'^{(i)}(h)\mathbf{VRP}_t + \epsilon^{(i)}_{t+h}$. For bonds (Panel A), $rx^{(i)}_{t+h}$ is the one year (h = 12 months) excess return for $i = \{24, 36, 48, 60\}$ month Treasury bonds. For stocks (Panel B), $rx^{(i)}_{t+h}$ is the three, six, or twelve month excess return on the market (value-weighted CRSP index), value and growth portfolio (from Ken French's Data Library), respectively. $\mathbf{VRP}_t$ is a vector containing the principal components of the bond variance risk premia, $VRP^{(PC1)}$, $VRP^{(PC2)}$ and $VRP^{(PC3)}$. Regressions are standardized, meaning all variables are de-meaned and divided by their standard deviation. Coefficients are estimated with ordinary-least squares, t-statistics are in parentheses and are calculated using Newey and West (1987) standard errors. Data is monthly and the sample spans the period from July 1991 to June 2010.

PANEL: FAMA BLISS TREASURY BONDS

	2y	3y	4y	5y
$VRP^{(PC1)}$	0.014	0.037	0.032	0.066
	(0.15)	(0.41)	(0.37)	(0.81)
$VRP^{(PC2)}$	-0.080	-0.060	-0.057	-0.045
	(-0.99)	(-0.73)	(-0.68)	(-0.53)
$VRP^{(PC3)}$	-0.166	-0.173	-0.191	-0.189
	(-2.15)	(-2.15)	(-2.42)	(-2.39)
$AdjR^2$	0.03	0.03	0.03	0.03

PANEL B: STOCKS

	MARKET			GROWTH			VALUE		
	3m	6m	12m	3m	6m	12m	3m	6m	12m
$VRP^{(PC1)}$	0.028	0.091	0.150	0.079	0.123	0.188	-0.051	0.053	0.162
	(0.19)	(0.91)	(1.42)	(0.67)	(1.42)	(1.94)	(-0.28)	(0.41)	(1.33)
$VRP^{(PC2)}$	0.118	0.241	0.212	0.058	0.193	0.177	0.138	0.276	0.304
	(1.11)	(2.34)	(1.86)	(0.58)	(2.10)	(1.90)	(1.20)	(2.49)	(2.38)
$VRP^{(PC3)}$	0.120	0.176	0.137	0.090	0.162	0.140	0.069	0.122	0.038
	(1.83)	(2.59)	(1.58)	(1.40)	(2.38)	(1.65)	(0.96)	(1.65)	(0.43)
$AdjR^2$	0.02	0.09	0.08	0.01	0.07	0.08	0.02	0.09	0.11

Table 2.6: Excess Corporate Bond and Commercial Mortgage-Backed Securities Returns. We run the following regression: $rx_{t+h}^{(i)} = \beta'^{(i)}(h)\mathbf{VRP}_t + \epsilon_{t+h}^{(i)}$, where $rx_{t+h}^{(i)}$ is the three, six, or twelve month excess return on long and intermediate maturity corporate bond (Panels A and B) or commercial mortgage-backed securities indices (Panel C), respectively. We use AAA, BBB and CCC indices for corporate bonds and AAA, BBB, and B indices for CMBS. $\mathbf{VRP}_t$ is a vector containing the principal components of the bond variance risk premia, $VRP^{(PC1)}$, $VRP^{(PC2)}$ and $VRP^{(PC3)}$. Regressions are standardized, meaning all variables are de-meaned and divided by their standard deviation. Coefficients are estimated with ordinary-least squares, t-statistics are in parentheses and are calculated using Newey and West (1987) standard errors. Data is monthly and the sample spans the period from July 1991 to June 2010. CMBS data start in January 1997.

PANEL A: LONG CORPORATE BONDS									
	AAA			BBB			CCC		
	3m	6m	12m	3m	6m	12m	3m	6m	12m
$VRP^{(PC1)}$	0.163	0.257	0.423	0.209	0.346	0.508	0.158	0.254	0.397
	(2.04)	(2.81)	(5.32)	(2.05)	(3.67)	(5.48)	(2.00)	(2.84)	(-4.21)
$VRP^{(PC2)}$	0.270	0.245	0.166	0.190	0.285	0.183	0.089	0.115	0.087
	(3.24)	(3.10)	(2.03)	(2.48)	(3.35)	(2.19)	(1.36)	(1.27)	(0.94)
$VRP^{(PC3)}$	-0.046	-0.025	-0.105	0.074	0.093	0.000	0.016	0.074	0.043
	(-0.65)	(-0.30)	(-1.35)	(1.05)	(1.16)	(0.00)	(0.24)	(0.97)	(0.53)
$Adj R^2$	0.090	0.120	0.210	0.080	0.200	0.290	0.020	0.080	0.160
PANEL B: INTERMEDIATE CORPORATE BONDS									
	AAA			BBB			CCC		
	3m	6m	12m	3m	6m	12m	3m	6m	12m
$VRP^{(PC1)}$	0.130	0.185	0.352	0.237	0.400	0.517	0.238	0.363	0.452
	(1.90)	(2.33)	(3.99)	(2.30)	(4.18)	(4.79)	(1.98)	(3.77)	(3.74)
$VRP^{(PC2)}$	0.201	0.216	0.195	0.277	0.340	0.238	0.211	0.261	0.132
	(2.94)	(2.70)	(2.32)	(3.63)	(3.97)	(2.66)	(2.11)	(2.96)	(1.33)
$VRP^{(PC3)}$	-0.073	-0.060	-0.057	0.120	0.132	0.046	0.216	0.195	0.070
	(-1.29)	(-0.86)	(-0.79)	(1.48)	(1.65)	(0.57)	(-0.58)	(-0.20)	(0.80)
$Adj R^2$	0.050	0.080	0.160	0.140	0.290	0.320	0.140	0.230	0.220
PANEL C: COMMERCIAL CORPORATE BONDS									
	AAA			BBB			CCC		
	3m	6m	12m	3m	6m	12m	3m	6m	12m
$VRP^{(PC1)}$	0.161	0.245	0.361	0.051	0.127	0.321	-0.206	-0.133	-0.048
	(1.55)	(2.61)	(2.75)	(0.31)	(0.86)	(2.75)	(-1.28)	(-0.92)	(-0.33)
$VRP^{(PC2)}$	0.287	0.363	0.416	0.351	0.469	0.488	0.197	0.211	0.353
	(3.35)	(3.34)	(2.99)	(2.73)	(3.29)	(3.46)	(1.59)	(1.29)	(2.01)
$VRP^{(PC3)}$	0.172	0.196	0.057	0.032	0.095	0.062	-0.041	-0.014	-0.106
	(1.56)	(1.68)	(0.45)	(0.41)	(1.34)	(0.66)	(-0.48)	(-0.17)	(-0.96)
$Adj R^2$	0.160	0.270	0.340	0.120	0.270	0.380	0.060	0.040	0.110

Table 2.7: Treasury Bonds Excess Return Predictability. We run the following regression: $rx^{(i)}_{t+h} = \beta'^{(i)}(h)\mathbf{VRP}_t + \gamma'^{(i)}(h)\mathbf{M}_t + \epsilon^{(i)}_{t+h}$, where $rx^{(i)}_{t+h}$ is the one year excess return for $i = \{24, 36, 48, 60\}$ month Treasury bonds. $\mathbf{VRP}_t$ is a vector containing the principal components of the bond variance risk premia, $VRP^{(PC1)}$, $VRP^{(PC2)}$ and $VRP^{(PC3)}$. $\mathbf{M}_t$ includes the equity market variance risk premium ($\text{VRP}^{(E)}_t$), the Cochrane and Piazzesi (2005) factor (CP) and the macro factors $\widehat{F}_j, j = 1\ldots, 8$ from Ludvigson and Ng (2009). Regressions are standardized, meaning all variables are de-meaned and divided by their standard deviation. Coefficients are estimated with ordinary-least squares, t-statistics are in parentheses and are calculated using Newey and West (1987) standard errors. Data is monthly and the sample spans the period from July 1991 to June 2010.

	2y	3y	4y	5y
$VRP^{(PC1)}$	-0.265	-0.237	-0.209	-0.142
	(-2.56)	(-2.33)	(-2.22)	(-1.54)
$VRP^{(PC2)}$	-0.169	-0.162	-0.162	-0.149
	(-2.30)	(-2.16)	(-2.14)	(-1.92)
$VRP^{(PC3)}$	-0.174	-0.175	-0.174	-0.158
	(-2.80)	(-2.74)	(-2.72)	(-2.44)
$VRP^{(E)}$	0.102	0.133	0.113	0.104
	(0.73)	(0.93)	(0.78)	(0.70)
CP	0.357	0.359	0.376	0.374
	(3.78)	(3.96)	(4.26)	(4.36)
$\widehat{F}_1$	-0.433	-0.380	-0.331	-0.276
	(-4.30)	(-3.63)	(-3.16)	(-2.58)
$\widehat{F}_2$	0.025	0.024	0.023	0.030
	(0.65)	(0.65)	(0.64)	(0.94)
$\widehat{F}_3$	0.145	0.175	0.226	0.250
	(2.03)	(2.68)	(3.76)	(4.10)
$\widehat{F}_4$	0.028	0.055	0.089	0.123
	(0.43)	(0.79)	(1.25)	(1.68)
$\widehat{F}_5$	0.027	0.042	0.057	0.073
	(0.47)	(0.74)	(1.04)	(1.33)
$\widehat{F}_6$	-0.012	0.040	0.078	0.093
	(-0.15)	(0.49)	(0.94)	(1.13)
$\widehat{F}_7$	0.146	0.192	0.217	0.254
	(1.74)	(2.23)	(2.53)	(2.95)
$\widehat{F}_8$	0.044	0.095	0.126	0.150
	(0.63)	(1.33)	(1.76)	(2.14)
$AdjR^2$	0.280	0.280	0.300	0.310

Table 2.8: Stock Excess Returns Predictability. We run the following regression: $rx_{t+h}^{(i)} = \beta'^{(i)}(h)\mathbf{VRP}_t + \gamma'^{(i)}(h)\mathbf{M}_t + \epsilon_{t+h}^{(i)}$, where $rx_{t+h}^{(i)}$ is the three, six, or twelve month excess return on the market (value-weighted CRSP index), value and growth portfolio (from Ken French's Data Library), respectively. $\mathbf{VRP}_t$ is a vector containing the principal components of the bond variance risk premia, $VRP^{(PC1)}$, $VRP^{(PC2)}$ and $VRP^{(PC3)}$. $\mathbf{M}_t$ includes the equity market variance risk premium ($VRP_t^{(E)}$), the log dividend yield (DY), the log earnings to price ratio (E/P), and the net equity expansion (NTIS) from Goyal and Welch (2008). Regressions are standardized, meaning all variables are de-meaned and divided by their standard deviation. Coefficients are estimated with ordinary-least squares, t-statistics are in parentheses and are calculated using Newey and West (1987) standard errors. Data is monthly and the sample spans the period from July 1991 to June 2010.

	MARKET			GROWTH			VALUE		
	3m	6m	12m	3m	6m	12m	3m	6m	12m
$VRP^{(PC1)}$	-0.029	0.043	0.084	-0.024	0.013	0.054	-0.056	0.064	0.173
	(-0.24)	(0.43)	(0.80)	(-0.23)	(0.14)	(0.57)	(-0.43)	(0.55)	(1.43)
$VRP^{(PC2)}$	0.090	0.219	0.197	0.026	0.162	0.144	0.105	0.251	0.292
	(1.00)	(2.71)	(2.55)	(0.28)	(2.12)	(2.03)	(1.12)	(2.75)	(3.29)
$VRP^{(PC3)}$	0.129	0.203	0.162	0.069	0.153	0.128	0.098	0.173	0.102
	(1.51)	(2.07)	(2.56)	(0.87)	(1.71)	(1.87)	(1.21)	(1.50)	(1.58)
$VRP^{(E)}$	0.207	0.298	0.333	0.267	0.383	0.412	-0.006	0.085	0.190
	(1.70)	(3.42)	(3.99)	(2.47)	(4.16)	(4.71)	(-0.04)	(0.78)	(1.57)
DY	0.146	0.169	0.282	0.149	0.174	0.275	0.085	0.082	0.136
	(1.25)	(1.35)	(2.37)	(1.29)	(1.35)	(2.10)	(0.74)	(0.63)	(1.20)
E/P	0.073	0.153	0.200	0.093	0.172	0.180	-0.091	-0.012	0.083
	(0.43)	(0.89)	(1.38)	(0.64)	(1.06)	(1.24)	(-0.57)	(-0.08)	(0.55)
NTIS	0.285	0.429	0.470	0.213	0.363	0.409	0.193	0.346	0.487
	(1.62)	(2.13)	(2.75)	(1.41)	(2.12)	(2.83)	(1.16)	(1.64)	(2.67)
$Adj R^2$	0.11	0.30	0.39	0.07	0.25	0.33	0.05	0.19	0.34

Table 2.9: Excess Intermediate Corporate Bond Returns Predictability. We run the following regression: $rx_{t+h}^{(i)} = \beta'^{(i)}(h)\mathbf{VRP}_t + \gamma'^{(i)}(h)\mathbf{M}_t + \epsilon_{t+h}^{(i)}$, where $rx_{t+h}^{(i)}$ is the three, six, or twelve month excess return on a intermediate-term corporate bond portfolio for rating classes AAA, BBB and CCC, respectively. $\mathbf{VRP}_t$ is a vector containing the principal components of the bond variance risk premia, $VRP^{(PC1)}$, $VRP^{(PC2)}$ and $VRP^{(PC3)}$. $\mathbf{M}_t$ includes the equity market variance risk premium ($VRP_t^{(E)}$), the Cochrane and Piazzesi (2005) factor (CP) and the macro factors $\widehat{F}_j, j = 1 \ldots, 8$ from Ludvigson and Ng (2009). Regressions are standardized, meaning all variables are de-meaned and divided by their standard deviation. Coefficients are estimated with ordinary-least squares, t-statistics are in parentheses and are calculated using Newey and West (1987) standard errors. Data is monthly and the sample spans the period from July 1991 to June 2010.

	AAA			BBB			CCC		
	3m	6m	12m	3m	6m	12m	3m	6m	12m
$VRP^{(PC1)}$	-0.043	0.014	0.143	0.021	0.181	0.238	0.114	0.214	0.216
	(-0.52)	(0.15)	(1.56)	(0.31)	(2.40)	(2.61)	(1.19)	(2.27)	(2.18)
$VRP^{(PC2)}$	0.135	0.150	0.083	0.200	0.274	0.145	0.179	0.249	0.093
	(1.50)	(2.02)	(1.39)	(3.02)	(2.89)	(1.61)	(2.45)	(2.61)	(0.93)
$VRP^{(PC3)}$	-0.084	-0.053	-0.065	0.069	0.098	-0.002	0.198	0.188	0.051
	(-1.24)	(-0.77)	(-1.09)	(0.95)	(1.42)	(-0.04)	(3.06)	(2.48)	(1.00)
$VRP^{(E)}$	0.174	0.056	0.055	0.220	0.173	0.170	0.149	0.169	0.127
	(1.69)	(0.46)	(0.37)	(1.85)	(1.42)	(1.46)	(0.98)	(1.19)	(1.26)
CP	0.203	0.229	0.202	0.170	0.197	0.185	0.140	0.202	0.205
	(2.34)	(2.41)	(2.34)	(2.26)	(2.25)	(2.56)	(2.05)	(2.33)	(2.50)
$\widehat{F}_1$	-0.121	-0.257	-0.324	-0.128	-0.208	-0.339	-0.044	-0.094	-0.324
	(-1.29)	(-2.66)	(-3.40)	(-0.86)	(-1.39)	(-3.62)	(-0.30)	(-0.62)	(-3.85)
$\widehat{F}_2$	0.003	0.075	0.107	-0.001	0.039	0.092	-0.095	0.004	0.059
	(0.09)	(1.87)	(3.16)	(-0.01)	(1.08)	(2.17)	(-1.27)	(0.100)	(1.65)
$\widehat{F}_3$	0.193	0.192	0.244	0.172	0.111	0.115	0.017	-0.077	-0.107
	(2.04)	(2.01)	(2.99)	(3.01)	(1.71)	(2.18)	(0.28)	(-1.02)	(-1.13)
$\widehat{F}_4$	0.094	0.080	0.106	-0.139	-0.096	-0.027	-0.302	-0.277	-0.243
	(1.16)	(0.93)	(1.50)	(-1.12)	(-0.90)	(-0.32)	(-2.55)	(-2.70)	(-2.55)
$\widehat{F}_5$	-0.012	-0.035	-0.011	-0.012	-0.029	-0.061	-0.025	-0.042	-0.099
	(-0.13)	(-0.40)	(-0.20)	(-0.18)	(-0.4)	(-1.14)	(-0.43)	(-0.84)	(-1.78)
$\widehat{F}_6$	-0.010	0.021	0.132	0.159	0.113	0.100	0.204	0.174	0.188
	(-0.16)	(0.33)	(1.97)	(2.36)	(1.65)	(1.43)	(2.96)	(2.86)	(2.84)
$\widehat{F}_7$	0.133	0.235	0.281	0.088	0.118	0.202	0.016	0.047	0.140
	(2.11)	(3.69)	(3.80)	(1.48)	(1.96)	(3.97)	(0.35)	(0.88)	(2.73)
$\widehat{F}_8$	0.079	0.087	0.189	-0.021	0.009	0.115	-0.013	-0.008	0.077
	(0.99)	(1.41)	(3.82)	(-0.36)	(0.16)	(2.55)	(-0.22)	(-0.13)	(1.79)
$AdjR^2$	0.13	0.23	0.43	0.27	0.40	0.52	0.29	0.38	0.46

Table 2.10: Excess Commercial Mortgage-Backed Securities Returns Predictability. We run the following regression: $rx_{t+h}^{(i)} = \beta'^{(i)}(h)\mathbf{VRP}_t + \gamma'^{(i)}(h)\mathbf{M}_t + \epsilon_{t+h}^{(i)}$, where $rx_{t+h}^{(i)}$ is the three, six, or twelve month excess return on a commercial mortgage-backed securities portfolio for rating classes AAA, BBB and B, respectively. $\mathbf{VRP}_t$ is a vector containing the principal components of the bond variance risk premia, $VRP^{(PC1)}$, $VRP^{(PC2)}$ and $VRP^{(PC3)}$. $\mathbf{M}_t$ includes the equity market variance risk premium ($VRP_t^{(E)}$), the Cochrane and Piazzesi (2005) factor (CP) and the macro factors $\widehat{F}_j, j = 1\dots, 8$ from Ludvigson and Ng (2009). Regressions are standardized, meaning all variables are de-meaned and divided by their standard deviation. Coefficients are estimated with ordinary-least squares, t-statistics are in parentheses and are calculated using Newey and West (1987) standard errors. Data is monthly and the sample spans the period from July 1991 to June 2010.

	AAA			BBB			CCC		
	3m	6m	12m	3m	6m	12m	3m	6m	12m
$VRP^{(PC1)}$	0.081	0.140	0.095	0.119	0.067	0.098	0.087	0.178	0.053
	(1.09)	(1.58)	(0.61)	(1.49)	(0.92)	(0.74)	(0.69)	(1.44)	(0.36)
$VRP^{(PC2)}$	0.241	0.336	0.391	0.257	0.412	0.456	0.212	0.260	0.392
	(3.73)	(2.86)	(2.70)	(3.19)	(3.62)	(3.12)	(2.83)	(2.08)	(2.80)
$VRP^{(PC3)}$	0.201	0.215	0.056	0.044	0.105	0.047	0.080	0.138	0.040
	(1.70)	(1.95)	(0.75)	(0.52)	(1.32)	(0.70)	(0.95)	(1.24)	(0.42)
$VRP^{(E)}$	-0.041	-0.073	-0.025	-0.260	-0.120	0.033	-0.318	-0.285	-0.138
	(-0.76)	(-0.84)	(-0.28)	(-2.14)	(-1.52)	(0.35)	(-1.71)	(-2.16)	(-1.27)
CP	0.125	0.135	0.139	0.028	0.079	0.099	0.170	0.162	0.184
	(2.16)	(1.63)	(1.68)	(0.42)	(1.19)	(1.15)	(2.47)	(2.13)	(2.34)
$\widehat{F}_1$	-0.066	-0.166	-0.402	-0.058	-0.128	-0.269	0.266	0.309	0.120
	(-0.40)	(-0.94)	(-5.02)	(-0.37)	(-0.71)	(-2.53)	(1.61)	(1.62)	(0.81)
$\widehat{F}_2$	-0.141	0.010	0.115	-0.125	-0.005	0.057	-0.154	-0.089	-0.055
	(-1.21)	(0.23)	(2.15)	(-1.65)	(-0.13)	(1.17)	(-1.43)	(-1.52)	(-1.48)
$\widehat{F}_3$	0.245	0.197	0.145	0.360	0.263	0.142	0.259	0.051	-0.017
	(4.53)	(2.83)	(1.75)	(2.98)	(2.88)	(1.36)	(2.58)	(0.34)	(-0.09)
$\widehat{F}_4$	-0.220	-0.217	-0.159	-0.283	-0.255	-0.191	-0.187	-0.292	-0.365
	(-1.59)	(-1.81)	(-1.80)	(-2.26)	(-2.72)	(-2.17)	(-1.22)	(-2.03)	(-2.70)
$\widehat{F}_5$	-0.027	-0.064	-0.123	0.012	-0.074	-0.119	0.028	-0.036	-0.118
	(-0.40)	(-0.93)	(-1.96)	(0.19)	(-1.62)	(-1.91)	(0.37)	(-0.61)	(-1.62)
$\widehat{F}_6$	0.169	0.088	0.021	0.226	0.210	0.073	0.102	0.179	0.233
	(1.74)	(1.09)	(0.31)	(2.66)	(2.49)	(1.14)	(1.11)	(1.95)	(2.66)
$\widehat{F}_7$	0.033	-0.009	0.062	-0.102	-0.127	-0.034	0.019	-0.047	0.001
	(0.64)	(-0.12)	(1.17)	(-1.59)	(-2.06)	(-0.71)	(0.28)	(-0.60)	(0.01)
$\widehat{F}_8$	-0.170	-0.127	-0.041	-0.181	-0.158	-0.084	-0.156	-0.162	-0.107
	(-1.65)	(-1.48)	(-0.72)	(-2.17)	(-1.84)	(-1.50)	(-1.86)	(-1.56)	(-1.23)
$AdjR^2$	0.33	0.38	0.49	0.42	0.48	0.50	0.34	0.34	0.37

Table 2.11: Out-of-Sample Predictability. This table reports the results of the out-of-sample forecast evaluation for one year excess returns for Treasury bonds, the market, growth and value portfolio, and long corporate bonds. $\mathbf{VRP}_t$ is a vector containing the individual bond variance risk premia, $VRP^{(30y)}$, $VRP^{(10y)}$ and $VRP^{(5y)}$. $\mathbf{VRP}_t$ vs const' reports forecast comparisons of an unrestricted model with bond variance risk premia versus a restricted constant expected return benchmark model. $\mathbf{VRP}_t$+CP vs const+CP' reports forecast comparisons of an unrestricted model with bond variance risk premia and the CP factor versus a restricted benchmark model including a constant and the CP factor. R^2_{OOS} denotes the out-of-sample R^2 of Campbell and Thompson (2008). A positive number indicates that the unrestricted model has a lower forecast error than the restricted benchmark model. ENC-NEW denotes the test statistic of Clark and McCracken (2001) for the null hypothesis that the benchmark model encompasses the unrestricted model with additional predictors. The alternative is that the unrestricted model contains information that could be used to improve the benchmark model's forecast. * indicates significance for the ENC-NEW test statistic at minimally the 95% level. We obtain the initial estimates based on the period from July 1991 to July 1999 and study the out-of-sample predictability for the period starting in July 2000 and ending in June 2010.

PANEL A: FAMA BLISS TREASURY BONDS		
$\mathbf{VRP}_t$ v.s. const.	R^2_{OOS}	ENC-NEW
2y Treasury Bond	0.06	7.74*
3y Treasury Bond	0.06	7.63*
4y Treasury Bond	0.07	8.57*
5y Treasury Bond	0.07	8.31*
$\mathbf{VRP}_t$+CP v.s. const.+CP	R^2_{OOS}	ENC-NEW
2y Treasury Bond	0.13	16.72*
3y Treasury Bond	0.14	16.97*
4y Treasury Bond	0.17	20.02*
5y Treasury Bond	0.18	20.30*
PANEL B: STOCKS		
$\mathbf{VRP}_t$ v.s. const.	R^2_{OOS}	ENC-NEW
MARKET	0.12	17.10*
GROWTH	0.13	18.04*
VALUE	0.17	23.78*
PANEL C: LONG CORPORATE BONDS		
$\mathbf{VRP}_t$ v.s. const.	R^2_{OOS}	ENC-NEW
AAA	0.32	38.36*
BBB	0.38	52.72*
CCC	0.24	35.93*

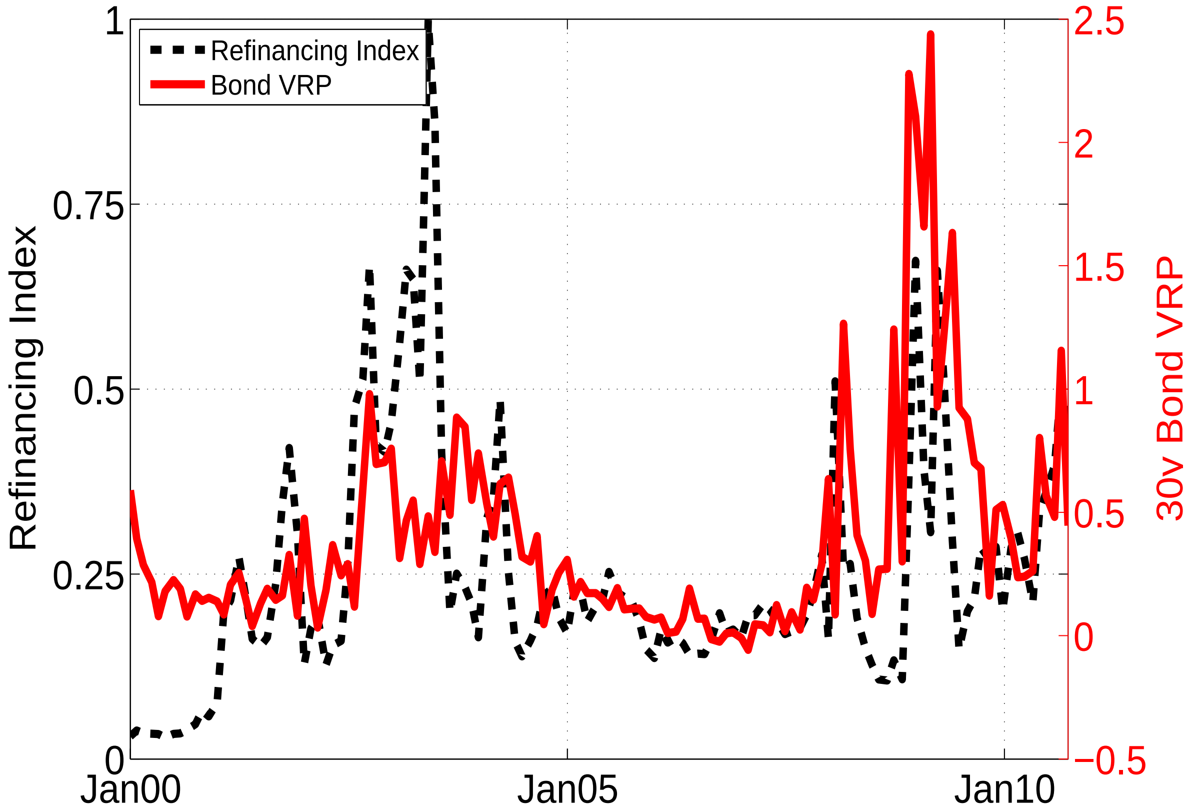

Figure 2.1: MBA Refinancing Index and 30 Year Bond Variance Risk Premium. This figure plots the time-series of the Mortgage Bankers Association (MBA) refinancing index (dashed line) and the bond variance risk premium for 30 year Treasury futures (solid line). The 30 year Treasury variance risk premium is calculated as the difference between the model-free implied variance (MIV) and the expected realized variance using a HAR-TCJ realized variance estimator augmented with lagged implied variance terms ($RV^{(HARIVJ)}$). The MBA refinancing index reflects the number of applications for mortgage refinancing and covers about three quarters of all new residential mortgage loans made. The index is seasonally adjusted and divided by a factor of 1,000. Data is monthly and the sample spans the period from January 2000 to June 2010.

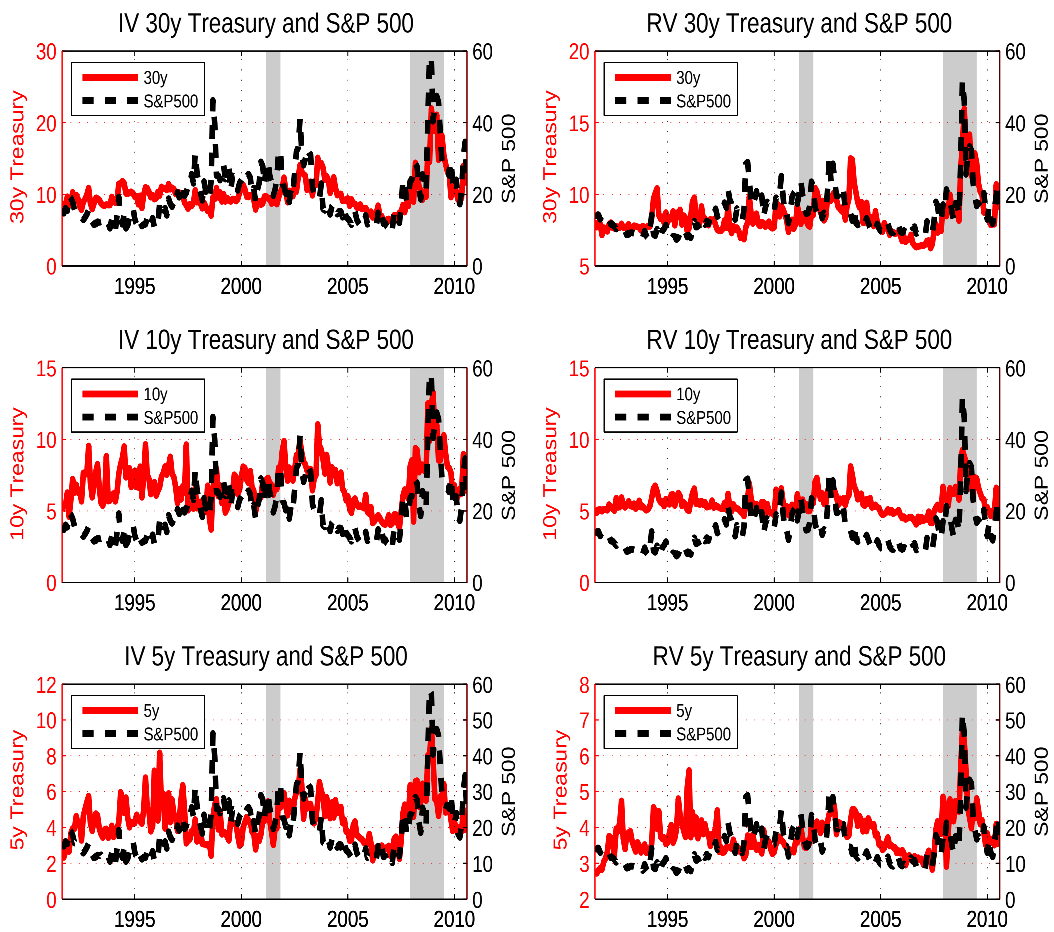

Figure 2.2: Realized and Implied Volatilities of Treasury and equities. In the left Panels we plot implied volatility measures (IV) for the 30 year, 10 year and 5 year Treasury futures (solid lines) together with the implied volatility of the S&P 500 index (dashed lines), all for a one month horizon. The implied volatilities are the square root of the model-free implied variance (MIV) calculated using options on the respective underlying futures. In the right Panels we plot the realized volatility measures (RV), which are the square root of the HAR-TCJ realized variance estimator augmented with lagged implied variance terms ($RV^{(HARIVJ)}$). All numbers are annualized and in percent. Shaded areas correspond to recessions as defined by the NBER. Data is monthly and the sample spans the period from July 1991 to June 2010.

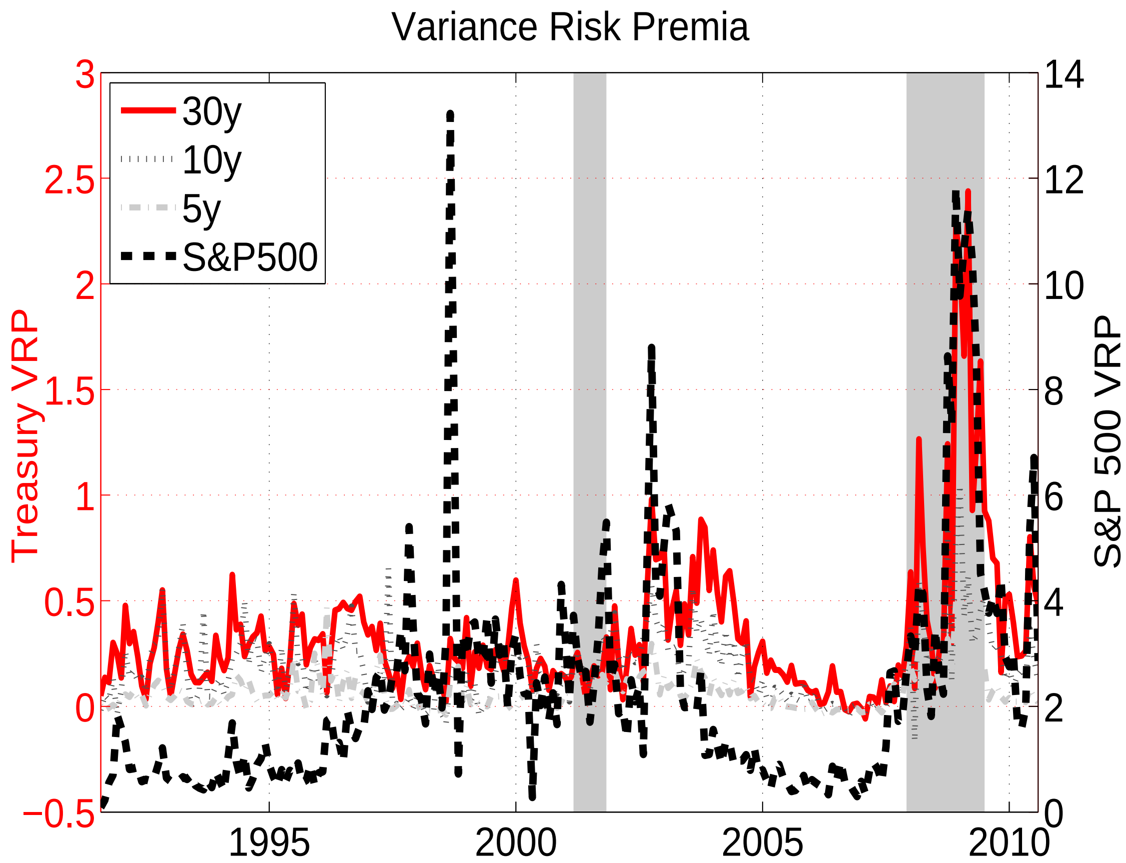

Figure 2.3: Treasury Bond and Equity Variance Risk Premia. This figure plots annualized variance risk premia for the 30 year, 10 year and 5 year Treasury bonds (left axis, solid, dotted and dashed-dotted lines) and the S&P500 index (right axis, bold dashed line). The variance risk premia are calculated as the difference between the model-free implied variance (MIV) and the expected realized variance using a HAR-TCJ realized variance estimator augmented with lagged implied variance terms ($RV^{(HARIVJ)}$). Shaded areas correspond to recessions as defined by the NBER. Data is monthly and the sample spans the period from July 1991 to June 2010.

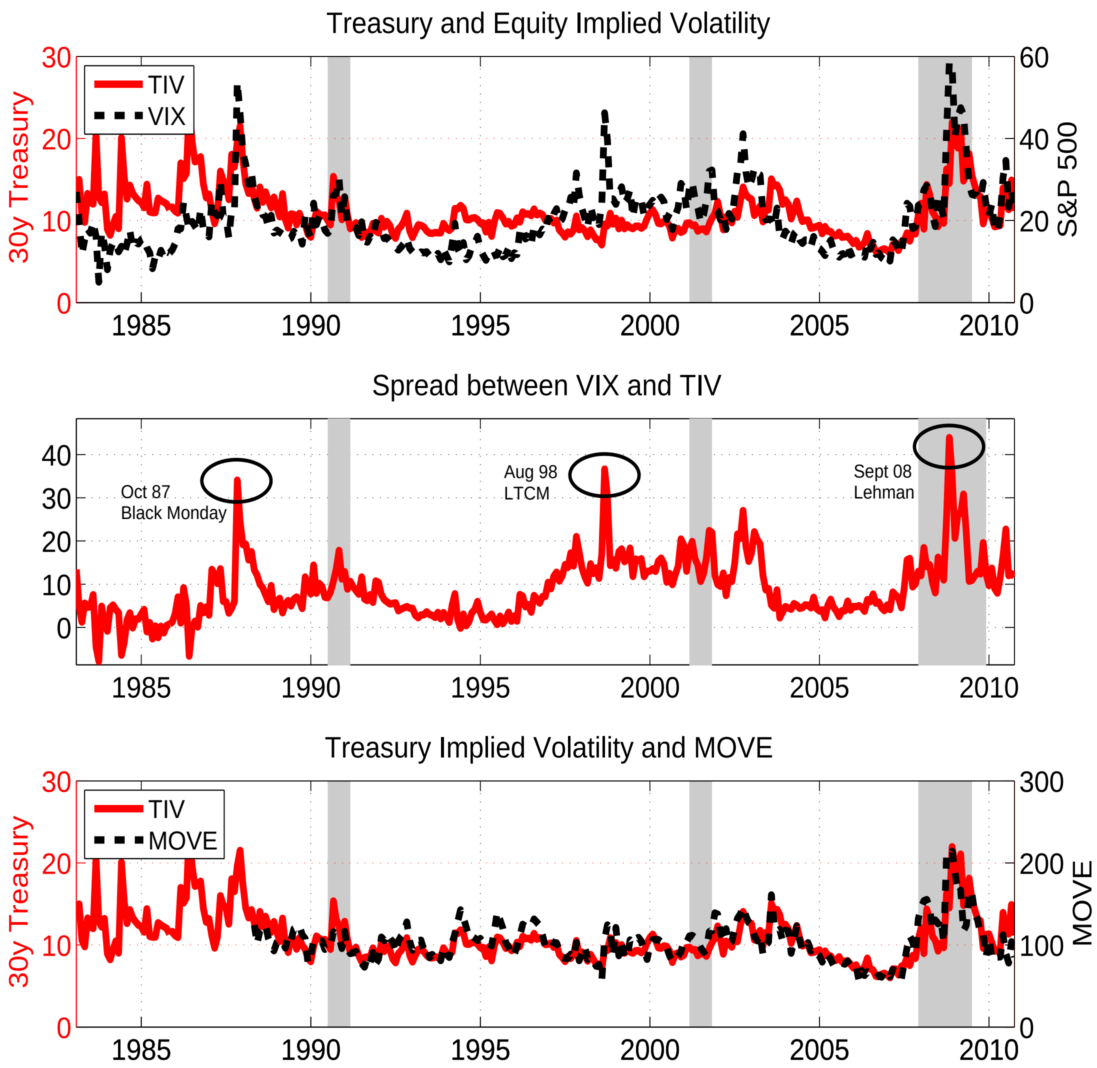

Figure 2.4: Treasury and Equity Implied Volatility. The top Panel plots the Treasury (solid line) and equity (dashed line) implied volatility measures TIV and VIX, respectively. The measures are calculated using options on the 30 year Treasury bond and the S&P 500 index futures, respectively, as the square root of the model-free implied variance MIV. The unconditional correlation between the TIV and the VIX is 46% over the whole sample period and 63% since 1990, the start date of the CBOE VIX. The middle Panel plots the spread between the VIX and the TIV. The bottom Panel plots the TIV (solid line) together with the Merrill Option Volatility Estimate (MOVE) index (dashed line). The MOVE index is a yield curve weighted index of normalized implied volatility on one month Treasury options for 2, 5, 10 and 30 year Treasuries. Shaded areas correspond to recessions as defined by the NBER. Data is monthly and the sample spans the period from January 1983 to June 2010.

Chapter 3

Testing Jumps via False Discovery Rate Control

Abstract: Many recently developed nonparametric jump tests can be viewed as multiple hypothesis testing problems. For such multiple hypothesis tests, it is well known that controlling type I error often unavoidably makes a large proportion of erroneous rejections, and such situation becomes even worse when the jump occurrence is a rare event. To obtain more reliable results, we aim to control the false discovery rate (FDR), an efficient compound error measure for erroneous rejections in multiple testing problems. We perform the test via a nonparametric statistic proposed by Barndorff-Nielsen and Shephard (2006), and control the FDR with a procedure proposed by Benjamini and Hochberg (1995). We provide asymptotic results for the FDR control. From simulations, we examine relevant theoretical results and demonstrate the advantages of controlling FDR. The hybrid approach is then applied to empirical analysis on two benchmark stock indices with high frequency data.
KEYWORDS: False discovery rate, BH procedure, Nonparametric BNS jump test.

3.1 Introduction

Recently, many hypothesis testing procedures have been proposed for detecting asset price jumps (Ait-Sahalia and Jacod, 2009; Bollerslev et al., 2008; Barndorff-Nielsen and Shephard, 2006; Fan and Fan, 2008; Jacod and Todorov, 2009; Jiang and Oomen, 2008; Lee and Mykland, 2008). These procedures use high frequency data to calculate test statistics for a certain period and then use these test statistics to test whether jumps occur in that period. Formally, the null hypothesis for such test at each period i, $i = 1, \ldots, m$, can be stated as

$$H_i^0 : \text{No jump occurs in period } i. \tag{3.1}$$

In addition to know whether there are price jumps, the "one test statistic for one period" approach for testing (3.1) also allows us to extract information about when and how frequently jumps occur in the whole sampling period. Such information is even more important for research on event study, derivative pricing and portfolio management.

If the number of periods m is greater than one, the jump test can be naturally viewed as a multiple hypothesis testing problem. Previous research used different test statistics, but often followed a similar decision procedure: rejecting the null hypothesis if the corresponding p-value is less than the controlled type I error α. Nevertheless, controlling type I error often unavoidably makes a large proportion of erroneous rejections. Such situation becomes even worse when the jump occurrence is a rare event.

To solve the problem described above, one may look for a more sensible compound error rate measure. In this paper we focus on false discovery rate (FDR). We will use the test statistic proposed by Barndorff-Nielsen and Shephard (2006) to obtain a p-value for each single hypothesis test, and then use the procedure proposed by Benjamini and Hochberg (1995) to control the FDR when simultaneously carrying out these hypothesis tests.

Several literatures on jump tests also tried to deal with the multiplicity issue. For example, Lee and Mykland (2008) set the significance level based on the distribution of the extreme value of the test statistic under the null hypothesis. This ensures that the jump test can achieve the probability of global misclassification to zero under some regular conditions. Applying FDR control to identify jump components also has been adopted by Fan and Fan (2008), who used an improved version of the test statistic proposed by Ait-Sahalia and Jacod (2009) to obtain a p-value for each single hypothesis test. The main difference between Fan and Fan (2008) and our paper is that we will give theoretical justifications on performance of the jump test statistic in a multiple hypothesis testing context. We will also conduct a simulation study to support our theoretical results.

The paper is organized as follows. In Section 2, we briefly describe the Benjamini-Hochberg (BH) procedure and the Barndorff-Nielsen-Shephard (BNS) nonparametric test. We then discuss some asymptotic results in Section 3. We focus on the case when p-values are calculated based on the asymptotic distributions of the test statistics. We show that with some appropriate conditions, the FDR can be asymptotically controlled by the BH procedure when p-values are obtained via the asymptotic distributions. In addition, magnitude of the approximate error for the asymptotic FDR control is bounded by a non-decreasing function of the expected number of true null hypotheses. This property indicates that the more the false null hypotheses there are, the better performance the asymptotic FDR control will have. In Section 4, we conduct a simulation study to show that performance of the BNS-BH hybrid procedure is positively related to the number of false hypotheses and sampling frequency of data, and is stable when the number of hypotheses and the required FDR level change. In Section 5, we apply the proposed procedure to analyze high frequency data of S&P500 in cash index and Dow Jones industrial average. Section 6 is the conclusion.

3.2 The Methodology

3.2.1 FDR and the BH procedure

For $i = 1, 2, \ldots m$, let H_i^0 and p_i denote the ith null hypothesis and the corresponding p-value, respectively. Among the m hypotheses, we let m_0 be the number of true hypotheses and $m_1 = m - m_0$ be the number of false hypotheses. Table 1 shows different situations when a multiple testing is performed. The numbers of hypotheses we reject and do not reject are denoted by R and $m - R$, respectively. In addition, U, T, V and S denote the numbers of hypotheses we correctly accept, falsely accept, falsely reject and correctly reject, respectively. The false discovery rate (FDR) is then defined as the expectation of false discovery proportion (FDP), i.e.

$$\text{FDR} = \mathbb{E}\left(\text{FDP}\right),$$

where

$$\text{FDP} = \begin{cases} 0 & \text{if } R = 0 \\ \frac{V}{R} & \text{if } R \neq 0. \end{cases}$$

In testing jump hypotheses, controlling the FDR has several advantages over controlling other compound error rate. First, if the price processes really have no jump component, i.e. all the null hypotheses are true, then controlling the FDR will be equivalent to controlling the familywise error rate $\Pr\left(V \geq 1\right)$. Second, if the intensity of the jump process $\lambda \neq 0$, then as time goes on (m increases), the proportion of false hypotheses among all hypotheses will be a nonzero constant with a high probability. Although such proportion may not be large, one may still expect the more (fewer) rejections there are, the more (fewer) erroneous rejections are allowed to occur; or at least the number of rejections should be proportional to m. In this situation, controlling error measures associated with proportion of erroneous rejections, like the FDR, makes sense. In addition, compound error rates such as the FWER are sometimes too stringent to get rejections when the number of hypotheses becomes large. The FDR criterion is less conservative in this aspect. Also, controlling the FDR currently seems to be more acceptable than controlling other compound error rates in many different research fields (Romano et al., 2008b).

Let $p_{(1)} \leq \ldots \leq p_{(m)}$ be the ordered p_i's and $H^0_{(1)}, \ldots, H^0_{(m)}$ be the corresponding null hypotheses. Benjamini and Hochberg (1995) proposed a stepwise procedure to control the FDR at the required level γ. The BH procedure can be simplified as the following two-step decision rule:

1. Obtain $i^* = \max_{i=1,2,\ldots,m} \left\{ i : p_{(i)} \leq \frac{i}{m}\gamma \right\}$.
2. Reject $H^0_{(i)}$ for all $i \leq i^*$.

Unlike some computational intensive methods, which often need a resampling scheme to construct the rejection region, the BH procedure is far easier to implement. The only computational burden of the BH procedure is to rank the p-values. Such advantage becomes even more obvious when the number of hypotheses goes large.

It can be shown that there is a relationship between type I error α and the FDR. For example, if we reject H^0_i when $p_i \leq \alpha$, then it will be possible to know what level of the FDR is controlled. In addition, if hypotheses are identical and the test statistics are all independent, Storey (2002) proposed the following estimator:

$$\widehat{\mathrm{FDR}}_{\kappa}(\alpha) = \frac{\#\{p_i > \kappa\}\,\alpha}{(1-\kappa)\max\left(\#\{p_i \leq \alpha\}, 1\right)},$$

to estimate the FDR, where κ is a turning parameter.

How the BH procedure performs relies on dependent structure of test statistics. In Benjamini and Yekutieli (2001), they showed that the BH procedure can still control the FDR when the test statistics are not independent, if the positive regression dependency (PRDS) for each test statistic under the true null hypotheses can be satisfied. In addition, simulation study in Romano et al. (2008b) showed that if the PRDS condition is violated, e.g. there exist negative common correlations between test statistics or the covariance matrix has an arbitrary structure, the BH procedure can still provide a satisfactory control of the FDR. Finally, if the test statistics have an arbitrary dependent structure, it can be shown that the BH procedure still guarantees that

$$\mathrm{FDR} \leq \gamma \sum_{k=1}^{m} \frac{1}{k} \approx \gamma \left(\log(m) + \frac{1}{2} \right).$$

A more detailed discussion on the theoretical properties of the FDR and the BH procedure is provided in section 3.

3.2.2 The BNS nonparametric jump test

Barndorff-Nielsen and Shephard (2006) proposed a nonparametric test statistic (henceforth the BNS test statistic), which utilizes realized variance and bi-power variation, to test jump components for the Brownian semimartingale plus jump class. A random variable $X(i)$ is said to belong to the Brownian semimartingale plus jump class if

$$X(i) = \int_0^i \mu(t)\,dt + \int_0^i \sigma(t)\,dW(t) + \sum_{j=1}^{N(i)} D(j),$$

where $\mu(t)$ and $\sigma(t)$ are assumed to be càdlàg, $W(t)$ is a standard Brownian Motion, $D(j)$ is the quantity of j th jump within $(0,i]$, and $N(i)$ is total number of the jumps occurring within $(0,i]$. Here $N(i) - N(i-1)$ is assumed to be finite.

The realized variance and the realized bi-power variation in period i are defined as

$$RV_i = \sum_{h=1}^{M} r_{i,h}^2, \tag{3.2}$$

$$BV_i = \frac{\pi}{2}\left(\frac{M}{M-1}\right)\sum_{h=1}^{M-1} |r_{i,h}|\,|r_{i,h+1}|, \tag{3.3}$$

respectively, where

$$r_{i,h} := \log P\left(i-1+\frac{h}{M}\right) - \log P\left(i-1+\frac{h-1}{M}\right)$$

is the intra-period log return in sub-interval h of period i, and $P(i-1+h/M)$ is the asset price at time point $i-1+h/M$. Assume that for $t \in (i-1,i]$, $\log P(t)$ belongs to the Brownian semimartingale plus jump class. Then it can be shown that under some regular conditions,

$$RV_i \xrightarrow{P} \int_{i-1}^{i} \sigma^2(t)\,dt + \sum_{j=N(i-1)+1,}^{N(i)} D^2(j),$$

$$BV_i \xrightarrow{P} \int_{i-1}^{i} \sigma^2(t)\,dt$$

as $M \to \infty$. Here $BV_i \xrightarrow{P} \int_{i-1}^{i} \sigma^2(t)\,dt$ can hold without any further assumption on the jump process, the joint distribution of the jump process and $\sigma(t)$.

Barndorff-Nielsen and Shephard (2006) showed that

$$JV_i = RV_i - BV_i$$

can consistently estimate $\sum_{j=N(i-1)+1}^{N(i)} D^2(j)$. In addition, if some regular conditions hold, the joint distribution of (3.2) and (3.3) will converge asymptotically to a bivariate normal distribution. Moreover, under the null hypothesis, when no jump presents on period i,

$$\frac{\sqrt{M}\left(RV_i - BV_i\right)}{\sqrt{A\int_{i-1}^{i}\sigma^4\left(t\right)dt}} \xrightarrow{L} \mathcal{N}\left(0,1\right),$$

where $A = (\pi/2)^2 + \pi - 5$. To estimate integrated quarticity $\int_{i-1}^{i}\sigma^4\left(t\right)dt$, we can use the realized tri-power quarticity

$$TP_i = \mu_{\frac{4}{3}}^{-3}\left(\frac{M^2}{M-2}\right)\sum_{h=1}^{M-2}\left(\left|r_{i,h}\right|\left|r_{i,h+1}\right|\left|r_{i,h+2}\right|\right)^{\frac{4}{3}},$$

where $\mu_a = \mathbb{E}\left(|Z|^a\right)$ and $Z \sim \mathcal{N}\left(0,1\right)$.

In the following analysis, we will also use some improved versions of the BNS statistic. The first one is proposed by Barndorff-Nielsen and Shephard (2006) themselves, which uses the log transformation and is defined as

$$Z_{\log,i} = \frac{\sqrt{M}\left(\log\left(RV_i\right) - \log\left(BV_i\right)\right)}{\sqrt{A\max\left(1,B\right)}},$$

where

$$B = \frac{\int_{i-1}^{i}\sigma^4\left(t\right)dt}{\left(\int_{i-1}^{i}\sigma^2\left(t\right)dt\right)^2}.$$

The second one is the Box-Cox transformed test statistic with $\rho = -1.5$, which is defined as

$$Z_{-1.5,i} = \frac{\sqrt{M}\left(\int_{i-1}^{i}\sigma^2\left(t\right)dt\right)^3\left(BV_i^{-1.5} - RV_i^{-1.5}\right)}{1.5\sqrt{A\max\left(1,B\right)}}.$$

Here the Box-Cox transformation for a positive number x is defined as

$$g_\rho\left(x\right) = \begin{cases} \frac{x^\rho - 1}{\rho} & \text{if } \rho \neq 0 \\ \log\left(x\right) & \text{if } \rho = 0. \end{cases}$$

The third one is the ratio type test statistic (Huang and Tauchen, 2005):

$$Z_{ratio,i} = \frac{\sqrt{M}\frac{RV_i - BV_i}{RV_i}}{\sqrt{A\max\left(1,B\right)}}.$$

Under the null hypothesis that there is no jump occurring in period i, the test statistics $Z_{-1.5,i}$, $Z_{\log,i}$ and $Z_{ratio,i}$ will have the standard normal distribution as the limiting joint distribution. When jumps occur in period i, the test statistics will approach to infinity as $M \to \infty$. It can also be shown that given jump variation and the integrated quarticity, as $M \to \infty$, the limiting joint distribution of the test statistic for each hypothesis is a mixed distribution (Veraart, 2010).

3.3 Asymptotically controlling FDR with the BH procedure

3.3.1 Asymptotic results

Let $\{\mathbf{X}_i = (X_{i,1} \ldots X_{i,M}) : i \in \mathbb{N}\}$ be a vector of samples defined on a probability space $(\Omega, \mathcal{F}, \mathcal{P})$. Let $\{\mathcal{F}_{M,i} : i \in \mathbb{N}\}$ be the smallest sub-σ field of $\mathcal{F}$ such that for $j = 1, \ldots M$, $X_{i,j}$ is $\mathcal{F}_{M,i}$ measurable. Let $\mathcal{F}_i = \lim_{M\to\infty} \mathcal{F}_{M,i}$. A test statistics for testing marginal hypothesis i with M samples is a function $\widehat{T}_{M,i} : \mathbf{X}_i \longmapsto \mathbb{R}$, and $\widehat{T}_{M,i}$ is $\mathcal{F}_{M,i}$ measurable. Let $\widehat{\mathbf{T}}_M = \left(\widehat{T}_{M,1}, \ldots, \widehat{T}_{M,m}\right)$ denote the vector of the test statistics for testing m $(m \geq 1)$ hypotheses. Given each $\mathcal{F}_i$, suppose that there exists a vector of random variables $\mathbf{T} = (T_1, \ldots, T_m)$ such that for $i = 1, \ldots m$, T_i is $\mathcal{F}_i$ measurable. Also assume $\widehat{T}_{M,i} \xrightarrow{L} T_i$ for each i as $M \to \infty$. Let Ψ_i be the limiting distribution function of the test statistic under the null hypothesis i. For one-sided test, p-value of the i th one-sided hypothesis is defined by $p_i(x) = 1 - \Psi_i(x)$ ($p_i(x) = 1 - 2\Psi_i(x)$ for two-sided hypothesis). Let $p_i = 1 - \Psi_i(T_i)$. A feasible estimated p-value for hypothesis i is then given by

$$\widehat{p}_{M,i} = p_i\left(\widehat{T}_{M,i}\right) = 1 - \Psi_i\left(\widehat{T}_{M,i}\right).$$

In our case of testing jumps, since our null hypotheses are homogeneous, Ψ_i is the c.d.f. of $\mathcal{N}(0,1)$ for all i.

Let $I_0 = \left\{i : H_i^0 \text{ is true}\right\}$ and $I_1 = \left\{i : H_i^0 \text{ is false}\right\}$. Let $\Psi_{M,i}$ be the exact distribution of $\widehat{T}_{M,i}$ under the null hypothesis i. The p- value under such distribution for hypothesis i is $p_{M,i} = 1 - \Psi_{M,i}\left(\widehat{T}_{M,i}\right)$. For $a \in (0,1)$, as $M \to \infty$, $\Pr(\widehat{p}_{M,i} \leq a) \to \Pr(p_i \leq a)$, if $\widehat{T}_{M,i} \xrightarrow{L} T_i$. If $T_1, \ldots, T_m$ are continuous random variables, then

$$\Pr(p_i \leq a) = a,$$

for $i \in I_0$. If $\widehat{T}_{M,1}, \ldots, \widehat{T}_{M,m}$ are also continuous random variables, then $\Pr(p_{M,i} \leq a) = a$ for $i \in I_0$.

Before we proceed to discuss our main results, we need the following two definitions. Let $\mathcal{B}$ denote the Borel set. Let $(\mathbb{R}^m, \mathcal{B}^m) = \prod_{i=1}^m (\mathbb{R}, \mathcal{B})$ define the $m-$fold products of the real line $\mathbb{R}$ with the Borel sets $\mathcal{B}$. Let

$$I_m = \{\{i_1, \ldots i_m\} : i_k \in \{1, \ldots m\} \text{ for } k = 1, \ldots m,\ i_k \neq i_l \text{ for } k \neq l\}.$$

Let $Q_{\mathbf{m}}$ be a probability measure on $(\mathbb{R}^m, \mathcal{B}^m)$ where $\mathbf{m} \in I_m$.

Definition 4 *A collection of* $\{Q_{\mathbf{m}}\}_{\mathbf{m}\in I_m}$ *is consistent if it satisfies*

- *Let* $\mathbf{m} = \{i_1, \ldots, i_m\}$, *and* $\mathbf{m}' = \{i_{k_1}, \ldots i_{k_m}\} \in I_m$ *but* $\mathbf{m}' \neq \mathbf{m}$. *Then for each* $B_i \in \mathcal{B}$,

$i = 1, \ldots m,$

$$Q_{\mathbf{m}}(B_1 \times \ldots \times B_m) = Q_{\mathbf{m}'}(B_{k_1} \times \ldots \times B_{k_m})$$

- *For each* $B_i \in \mathcal{B}$, $i = 1, \ldots m$,

$$Q_{\mathbf{m}}(B_1 \times \ldots B_{k-1} \times \mathbb{R} \times B_{k+1} \ldots \times B_m) = Q_{\mathbf{m}/i_k}(B_1 \times \ldots B_{k-1} \times B_{k+1} \ldots \times B_m).$$

Definition 5 *Let* $\mathbf{Y} = (Y_1, \ldots, Y_l)$*,* $\mathbf{X} = (X_1, \ldots, X_m)$*, and* I_0 *be a collection of index* $i \in \{1, \ldots, m\}$*. For any decreasing set* $\mathbf{\Theta}$ *and increasing set* $\mathbf{\Lambda}$*, an* $l-$*dimensional random vector* $\mathbf{Y}$ *is said to be positive regression dependency on each one from a subset (PRDS)* I_0 *of a* $m-$*dimensional random vector* $\mathbf{X}$ *is that* $\Pr(\mathbf{Y} \in \mathbf{\Theta} | X_i = x)$ *is non-increasing in* x *or* $\Pr(\mathbf{Y} \in \mathbf{\Lambda} | X_i = x)$ *is non-decreasing in* x *for any* $i \in I_0$*.*

In the above definition, a decreasing set $\mathbf{\Theta}$ is that $\mathbf{X} = (X_1, \ldots, X_m) \in \mathbf{\Theta}$ implies $\mathbf{Z} = (Z_1, \ldots, Z_m) \in \mathbf{\Theta}$ if $Z_i \leq X_i$ for any $i = 1, \ldots, m$. An increasing set $\mathbf{\Lambda}$ is that $\mathbf{Y} = (Y_1, \ldots, Y_l) \in \mathbf{\Lambda}$ implies $\mathbf{Z}' = (Z_1', \ldots, Z_l') \in \mathbf{\Lambda}$ if $Z_i' \geq Y_i$ for any $i = 1, \ldots, l$.

Let $\mathbf{p} = (p_1, \ldots, p_m)$ and $\widehat{\mathbf{p}}_M = (\widehat{p}_{M,1}, \ldots, \widehat{p}_{M,m})$. Suppose we want to control FDR at the level γ with the BH procedure with $\mathbf{p}$. FDR conditional on m_0 true null hypotheses is given by

$$\mathbb{E}\left(\frac{V}{R} \mid \widetilde{m}_0 = m_0\right) = \sum_{s=0}^{m_1} \sum_{v=1}^{m_0} \frac{v}{v+s} \Pr\left(\mathbf{p} \in \mathbf{D}_{m_0}^{v,s}\right). \tag{3.4}$$

$\mathbf{D}_{m_0}^{v,s}$ is a well constructed union of $m-$dimensional cubes such that $\left\{\mathbf{p} \in \mathbf{D}_{m_0}^{v,s}\right\}$ is the event that v true and s false null hypotheses are rejected when the BH procedure is implemented with $\mathbf{p}$. Benjamini and Yekutieli (2001) and Sarkar (2002) showed that if the joint distribution of p_i is PRDS on I_0, then $\mathbb{E}(V/R \mid \widetilde{m}_0 = m_0) \leq m_0\gamma/m$. Since $\widetilde{m}_0$ is bounded by m, we can get

$$\mathbb{E}\left(\frac{V}{R}\right) \leq \frac{\gamma \mathbb{E}(\widetilde{m}_0)}{m} \leq \gamma.$$

If $\widehat{\mathbf{p}}_M$ is used, the analogue of (3.4) is then given by

$$\mathbb{E}_{\widehat{\mathbf{p}}_M}\left(\frac{V}{R} \mid \widetilde{m}_0 = m_0\right) = \sum_{s=0}^{m_1} \sum_{v=1}^{m_0} \frac{v}{v+s} \Pr\left(\widehat{\mathbf{p}}_M \in \mathbf{D}_{m_0}^{v,s}\right),$$

where $\left\{\widehat{\mathbf{p}}_M \in \mathbf{D}_{m_0}^{v,s}\right\}$ is the event that v true and s false hypotheses are rejected when the BH procedure is implemented with $\widehat{\mathbf{p}}_M$.

Ideally, if we know $\Psi_{M,i}$, and the joint distribution of $p_{M,i}$ is PRDS on I_0, we can implement the BH procedure directly with $p_{M,1}, \ldots, p_{M,m}$. However, such information is often unknown, and instead only $\widehat{\mathbf{p}}_M$ is feasible. In the following, we show that under appropriate conditions, FDR can be asymptotically controlled with $\widehat{\mathbf{p}}_M$ under a desired level. Our strategy is to show that under appropriate conditions, $\Pr\left(\widehat{\mathbf{p}}_M \in \mathbf{D}_{m_0}^{v,s}\right) \to \Pr\left(\mathbf{p} \in \mathbf{D}_{m_0}^{v,s}\right)$ as $M \to \infty$ and then to prove

$$\mathbb{E}_{\widehat{\mathbf{p}}_M}\left(\frac{V}{R} \mid \widetilde{m}_0 = m_0\right) \to \mathbb{E}\left(\frac{V}{R} \mid \widetilde{m}_0 = m_0\right).$$

as $M \to \infty$. Therefore implementing the BH procedure with $\widehat{\mathbf{p}}_M$ is asymptotically equivalent to implementing the procedure with $\mathbf{p}$.

The main results are the following two theorems, and their proofs are given in the supplementary materials.

Theorem 3 *Suppose we have m hypotheses to be tested simultaneously. If the following conditions hold,*

1. *The joint distribution of p_i and the joint distribution of $\widehat{p}_{M,i}$ satisfy the consistency for multivariate distribution.*
2. *The joint distribution of p_i is PRDS on p_i for $i \in I_0$ and all $m \geq 1$.*
3. $\Pr(p_i \leq a) \leq a$ *for $i \in I_0$ and $a \in (0,1)$.*
4. $\sup_{1\leq k\leq m} \sup_{i\in I_0} |\Pr(p_i \leq q_k) - \Pr(\widehat{p}_{M,i} \leq q_k)| = O\left(1/M^{\delta}\right)$, *where $q_k = k\gamma/m$ and $\delta > 0$.*
5. *Given m_0 true null hypotheses, let $\mathbf{p}^{(-i)}$ and $\widehat{\mathbf{p}}_M^{(-i)}$ denote the random $m-1$ dimensional vectors obtained by eliminating p_i and $\widehat{p}_{M,i}$ from the $m-$dimensional random vectors $\mathbf{p}$ and $\widehat{\mathbf{p}}_M$ respectively. Let $p_{(1)}^{(-i)} \leq \cdots \leq p_{(m-1)}^{(-i)}$ and $\widehat{p}_{M,(1)}^{(-i)} \leq \cdots \leq \widehat{p}_{M,(m-1)}^{(-i)}$ denote the ordered components of $\mathbf{p}^{(-i)}$ and $\widehat{\mathbf{p}}_M^{(-i)}$ respectively. For every $m \geq 1$,*

$$\sup_{1\leq k\leq m} \sup_{i\in I_0}\left| m \left(\begin{array}{c} \Pr\left(\widehat{p}_{M,i} \leq q_k, \widehat{p}_{M,(k)}^{(-i)} > q_{k+1}, \ldots, \widehat{p}_{M,(m-1)}^{(-i)} > q_m\right) \\ -\Pr\left(p_i \leq q_k, p_{(k)}^{(-i)} > q_{k+1}, \ldots, p_{(m-1)}^{(-i)} > q_m\right) \end{array}\right)\right| = o(1), \quad (3.5)$$

then the BH procedure implemented with the estimated p-values $\widehat{\mathbf{p}}_M$ asymptotically control FDR at the required level γ in the sense that

$$\lim_{M\to\infty} \mathbb{E}_{\widehat{\mathbf{p}}_M}\left(\frac{V}{R}\right) = \mathbb{E}\left(\frac{V}{R}\right) \leq \gamma.$$

Theorem 4 *Suppose we have m hypotheses to be tested simultaneously. If conditions 1 and 4 in Theorem 1 and the following conditions hold,*

1. $\Pr(\widehat{p}_{M,i} \leq a) \leq a$ *for $i \in I_0$ and $a \in (0,1)$*
2. *$T_1, \ldots, T_m$ are mutually independent and continuous random variables.*
3. *$\widehat{T}_{M,1}, \ldots, \widehat{T}_{M,m}$ are mutually independent,*

then the BH procedure implemented with the estimated p-values $\widehat{\mathbf{p}}_M$ asymptotically control FDR at the required level γ in the sense that

$$\lim_{M \rightharpoonup \infty} \mathbb{E}_{\widehat{\mathbf{p}}_M} \left(\frac{V}{R} \right) = \mathbb{E} \left(\frac{V}{R} \right) \leq \gamma.$$

3.3.2 Discussions on the asymptotical results

The two theorems say that under some regular conditions, we can asymptotically control FDR. The results still hold even we let the number of hypotheses $m \to \infty$. A key condition making the two theorems different is the requirement on dependent structure of elements in vector $\mathbf{p}$ and $\widehat{\mathbf{p}}_M$. If dependent structure of p_i satisfies PRDS on I_0, it ensures that $\mathbb{E}(V/R) \leq \gamma$. Here we only require the PRDS should hold on I_0, and the dependent structure of p_i on I_1 can be arbitrary. Marginal distributions of p_i and $\widehat{p}_{M,i}$ converging with the rate $O\left(1/M^{\delta}\right)$ simultaneously for all i is also needed for the consistent control. In addition, we also require the convergence of the joint distribution of the ordered p-values. But as stated in Theorem 2, such condition can be ignored if other conditions hold.

The approximation error $\epsilon = \left|\mathbb{E}_{\widehat{\mathbf{p}}_M}(V/R) - \mathbb{E}(V/R)\right|$ essentially vanishes to zero when $M \rightharpoonup \infty$. Magnitude of ϵ, as shown in our proof, is bounded by a non-decreasing function of $\mathbb{E}(\widetilde{m}_0)$. This property indicates that the more the false null, the better the convergence.

We then have a look of condition 1 in Theorem 1. This is a sufficient condition to ensure that $\Pr(\mathbf{p} \in \mathbf{D}^{v,s})$ (and $\Pr(\widehat{\mathbf{p}}_M \in \mathbf{D}^{v,s})$) exists as $m \to \infty$. It is due to Kolmogorov's extension theorem (Karatzas and Shreve, 1991, pg. 50): an extension of any consistent family of probability measures on $(\mathbb{R}^m, \mathcal{B}^m)$ to a probability measures on $(\mathbb{R}^\infty, \mathcal{B}^\infty) = \prod_{i=1}^{\infty} (\mathbb{R}, \mathcal{B})$ necessarily exists and is unique. Conversely, if we have a probability measure on $(\mathbb{R}^\infty, \mathcal{B}^\infty)$, we can induce a family of finite-dimensional distributions on $(\mathbb{R}^m, \mathcal{B}^m)$, and these induced finite-dimensional distributions all satisfy consistency for multivariate distribution.

Condition 2 in Theorem 1 requires that the joint distribution of the p- values should satisfy PRDS on the subset I_0. It is a sufficient condition for $\mathbb{E}(V/R) \leq \gamma$ when we implement the BH procedure with $\mathbf{p}$. Since our purpose is to control FDR with $\widehat{\mathbf{p}}_M$, if we can guarantee that $\lim_{M \rightharpoonup \infty} \mathbb{E}_{\widehat{\mathbf{p}}_M}(V/R) = \mathbb{E}(V/R)$, only the distribution of $\mathbf{p}$ satisfying the condition is needed. For practically using the BH procedure, Benjamini and Yekutieli (2001) listed many situations when the condition holds. For example, if $\mathbf{T} \sim \mathcal{N}(\mu, \mathbf{\Sigma})$, where $\mu = (\mu_1, \ldots, \mu_m)$ and $\mathbf{\Sigma}$ is a $m \times m$ covariance matrix with element σ_{ij}. Suppose for each $i \in I_0$, and each $j \neq i$, $\sigma_{ij} \geq 0$, then the distribution of $\mathbf{T}$ is PRDS on I_0, regardless what the covariance structure of $i \in I_1$ is. Mutual independence of $T_1, \ldots, T_m$ can be easily seen as a special case of PRDS on I_0. As for the nonparametric jump test in this paper, since the limiting distribution of the test statistics is a multivariate normal with $\sigma_{ij} = 0$ for each $i \in I_0$ and each $j \neq i$, it implies PRDS on I_0.

The condition that $\Pr(p_i \leq a) \leq a$ for $a \in (0,1)$ is called the distribution of p_i is stochastically dominated by the Uniform$(0,1)$. If $\lim_{M \to \infty} \Pr(\widehat{p}_{M,i} \leq a) \leq a$, it is called that the distribution of $\widehat{p}_{M,i}$ is stochastically dominated by the Uniform$(0,1)$ distribution asymptotically. In order to control FDR with the BH method asymptotically, we at least need that $\Pr(p_i \le a) \le a$ for $a \in (0,1)$ and $i \in I_0$. The condition is more liberal than that p_i has the exact Uniform$(0,1)$ distribution for $i \in I_0$, and applies to the case when the test statistics are discrete random variables.

As shown in the proof of Theorem 1,

$$\begin{aligned}
& \left|\mathbb{E}_{\widehat{\mathbf{p}}_M}\left(\frac{V}{R} \mid \widetilde{m}_0 = m_0\right) - \mathbb{E}\left(\frac{V}{R} \mid \widetilde{m}_0 = m_0\right)\right| \\
= & \left|\sum_{i\in I_0}\sum_{k=1}^{m}\frac{1}{k}\left(\begin{array}{c}\Pr\left(\widehat{p}_{M,i} \le q_k, \widehat{p}_{M,(k-1)}^{(-i)} \le q_k, \widehat{p}_{M,(k)}^{(-i)} > q_{k+1}, \ldots, \widehat{p}_{(m-1)}^{(-i)} > q_m\right) \\ -\Pr\left(p_i \le q_k, p_{(k-1)}^{(-i)} \le q_k, p_{(k)}^{(-i)} > q_{k+1}, \ldots, p_{(m-1)}^{(-i)} > q_m\right)\end{array}\right)\right| \quad (3.6) \\
\le & \left|\sum_{i\in I_0}\sum_{k=1}^{m-1}\frac{1}{k(k+1)}\left(\begin{array}{c}\Pr\left(\widehat{p}_{M,i} \le q_k, \widehat{p}_{M,(k)}^{(-i)} > q_{k+1}, \ldots, \widehat{p}_{M,(m-1)}^{(-i)} > q_m\right) - \\ \Pr\left(p_i \le q_k, p_{(k)}^{(-i)} > q_{k+1}, \ldots, p_{(m-1)}^{(-i)} > q_m\right)\end{array}\right)\right| \quad (3.7) \\
& + \left|\sum_{i\in I_0}\frac{1}{m}\left(\Pr\left(\widehat{p}_{M,i} \le q_m\right) - \Pr\left(p_i \le q_m\right)\right)\right|. \quad (3.8)
\end{aligned}$$

In the first equality, $\Pr\left(p_i \le q_k, p_{(k-1)}^{(-i)} \le q_k, p_{(k)}^{(-i)} > q_{k+1}, \ldots p_{(m-1)}^{(-i)} > q_m\right)$ is the probability that in addition to rejecting the hypothesis i, we also reject other $k-1$ hypotheses. If $m_0 = m$, Sarkar (1998) showed that

$$\begin{aligned}
& \sum_{i\in I_0}\sum_{k=1}^{m}\frac{1}{k}\Pr\left(p_i \le q_k, p_{(k-1)}^{(-i)} \le q_k, p_{(k)}^{(-i)} > q_{k+1}, \ldots p_{(m-1)}^{(-i)} > q_m\right) \\
= & \ 1 - \Pr\left(\widehat{p}_{M,(1)} > q_1, \ldots, \widehat{p}_{M,(m)} > q_m\right).
\end{aligned}$$

Therefore if $m_0 = m$, (3.6) becomes

$$\left|\left(1 - \Pr\left(\widehat{p}_{M,(1)} > q_1, \ldots, \widehat{p}_{M,(m)} > q_m\right)\right) - \left(1 - \Pr\left(p_{(1)} > q_1, \ldots, p_{(m)} > q_m\right)\right)\right|. \quad (3.9)$$

(3.9) is the difference between the probability that we at least have one false rejection (or the familywise error rate, FWER), when the BH procedure is implemented with p and $\widehat{\mathbf{p}}_M$. The result is not surprising since when all null hypotheses are true, FDR=FWER.

To make (3.7) vanish as $M \to \infty$, (3.5) in condition 5 of Theorem 1 is one of the sufficient conditions. However, as shown in Theorem 2, such condition is redundant when test statistics are independent and continuous.

We finally have a look of the assumption:

$$\sup_{1\le k\le m}\sup_{i\in I_0}\left|\Pr\left(p_i \le q_k\right) - \Pr\left(\widehat{p}_{M,i} \le q_k\right)\right| = O\left(\frac{1}{M^{\delta}}\right). \quad (3.10)$$

The assumption says that the convergence in law should hold simultaneously at the points q_k for $1 \le k \le m$, and for all $i \in I_0$. Such convergence is reasonable for test statistics with limiting normal distribution if we set $\delta = s/2$, $s = 1, 2, \ldots$. Note that if T_i and $\widehat{T}_{M,i}$ are continuous,

$$\begin{aligned} & \Pr\left(p_i \leq q_k\right) - \Pr\left(\widehat{p}_{M,i} \leq q_k\right) \\ = \ & \Pr\left(\widehat{T}_{M,i} \leq \Psi_i^{-1}\left(1 - q_k\right)\right) - \Pr\left(T_i \leq \Psi_i^{-1}\left(1 - q_k\right)\right). \end{aligned}$$

If $\widehat{T}_{M,i}$ and T_i are asymptotically normal, and satisfy $\widehat{T}_{M,i} = T_i + O_p\left(1/M^{\frac{s}{2}}\right)$ for an integer $s \geq 1$, then by theory of Edgeworth expansion of the distributions of $\widehat{T}_{M,i}$ and T_i (Hall, 1992, pg.76),

$$\Pr\left(\widehat{T}_{M,i} \leq a\right) = \Pr\left(T_i \leq a\right) + O\left(\frac{1}{M^{\frac{s}{2}}}\right).$$

So $\Pr\left(\widehat{p}_{M,i} \leq 1 - \Psi_i\left(a\right)\right)$ can converge to $\Pr\left(p_i \leq 1 - \Psi_i\left(a\right)\right)$ with the rate of $O\left(1/M^{\frac{s}{2}}\right)$.

In our nonparametric jump test, standard normal is used to approximate $\Pr\left(\widehat{T}_{M,i} \leq a\right)$ under the null. There are several methods to improve the approximation, for example, the bootstrap approximation and the Box-Cox transformation. Some theoretical results about how the methods perform have been established. Goncalves and Meddahi (2009) showed that when no jump presents, distribution of the test statistic for standardised realized volatility can be approximated by $\mathcal{N}(0,1)$ with the rate of convergence $O\left(1/\sqrt{M}\right)$. They also documented that under some situations, the bootstrap approximation is better than the standard normal approximation, and the error rate can be reduced to $o_p\left(1/\sqrt{M}\right)$. For the Box-Cox transformation, Goncalves and Meddahi (2011) showed that without jump component, the skewness of the test statistic for realized volatility can be efficiently reduced via optimally choosing the parameter for the Box-Cox transformation.

In practice, the number of samples M within a hypothesis, may be less than the number of hypotheses m. How such the large m, small M (or in statisticians' view: Large p (number of dimensions), small n (number of samples)) situation affects statistical inferences has been intensively studied recently, especially in simultaneously convergence of the test statistics. For example, when the samples are iid, sufficient conditions for $\widehat{p}_{M,i} \xrightarrow{P} p_i$ uniformly for all i already was provided by Kosorok and Ma (2007). Clarke and Hall (2009) documented that the difficulties caused by dependence of test statistics can be alleviated when m grows, but the result subjects to that distributions of test statistics should have light tails such as normal or Student's t. Fan et al. (2007) proved that if normal or Student's t distribution is used to approximate the exact null distribution, the rejection area is accurate when $\log m = o(M^{\frac{1}{3}})$; but if the bootstrap methods are applied, then $\log m = o(\sqrt{M})$ is sufficient to guarantee the asymptotic-level accuracy.

In practice, high frequency returns might not be iid distributed. Instead of assuming that samples have certain distributional properties, here we assume that (3.10) need to hold. However, by jointly restricting growth rates of M and m, and together with other mild conditions, (3.10) also can be achieved. It can be seen in the following proposition.

Proposition 5 *For all* $i \in I_0$ *and every* a*, if there exists some constant* $M_0 > 0$ *such that for* $M \geq M_0$*,* $\Pr\left(\max_{i\in I_0}\left|M^{\delta}\left(T_i - \widehat{T}_{M,i}\right)\right| > a\right) \leq c_1 \exp\left(-c_2 a^p\right)$*, where* c_1 *and* c_2 *are two constants and* $\delta > 0$*, and* $p \geq 1$*. Also* $\left(\log\left(m\right)\right)^{\frac{1}{p}}/M^{\delta} = o\left(1\right)$ *as* $M \to \infty$ *hold, then* $\sup_{1\leq k\leq m}\sup_{i\in I_0}\left|\Pr\left(\widehat{p}_{M,i} \leq q_k\right) - \Pr\left(p_i \leq q_k\right)\right| = o(1)$.

3.4 Simulation study

3.4.1 The model

For the simulation study, we consider the following stochastic volatility with a jump component (SVJ) model:

$$\begin{aligned} d\log P\left(t\right) &= \left(\mu - \frac{1}{2}\sigma^2\left(t\right)\right)dt + \sigma\left(t\right)dW_1\left(t\right) + dJ\left(t\right), \\ d\sigma^2\left(t\right) &= a\left(b - \sigma^2\left(t\right)\right)dt + \omega\sigma\left(t\right)dW_2\left(t\right), \\ J\left(t\right) &= \sum_{j=1}^{N(t)} D\left(t,j\right),\ D\left(t,j\right) \overset{iid}{\sim} \mathcal{N}\left(0,1\right), \\ & N\left(t\right) \overset{iid}{\sim} \text{Poisson}\left(\lambda dt\right), \end{aligned}$$

where $dW_1\left(t\right)$ and $dW_2\left(t\right)$ follow the standard Brownian motion, and $\sigma^2\left(t\right)$ follows the CIR process. $J\left(t\right)$ follows a Compound Poisson Process (CPP) with a constant intensity λdt, and $N\left(t\right)$ is the number of jumps occurring within the small interval $(t - \triangle t, t]$. The leverage effect is not allowed in this model, and the correlation between $dW_1\left(t\right)$ and $dW_2\left(t\right)$ is set to zero. We use the following parameter values for the simulation:

$$\mu = 0.05,\ a = 0.015,\ b = 0.2,\ \text{and}\ \omega = 0.05.$$

In the simulation, the unit of a period is one day. We vary the (daily) jump intensity λ at five different levels: 0, 0.02, 0.05, 0.1, 0.15, and 0.2. Here λ is essentially the expected number of jumps occurring per day. Different values of λ tend to have different numbers of jump days over the whole sampling period, therefore result in different numbers of false null hypotheses. This allows us to see how such differences affect outcomes of the method.

We mimic the US stock market and generate 1-min intradaily log prices over 6.5 hours each day. Thus in our simulation, $M = 6.5 \times 60 = 390$, $dt \approx \triangle t = \frac{1}{M}$, and $\lambda dt \approx \frac{\lambda}{M}$. After obtaining a sample path, the jump test statistics $Z_{-1.5,i}$, $Z_{\log,i}$ and $Z_{ratio,i}$ and their corresponding p-values are calculated. We test hypothesis (3.1) with the test statistics and control the FDR at the level γ with the BH procedure.

3.4.2 Simulation results

We first focus on the case when the FDR control level $\gamma = 0.05$ and the number of null hypotheses $m = 1000$. Figure 3.1 to Figure 3.5 show the plots of average values of relevant quantities from 1000 simulation runs. Figure 3.1 is for performances of the three different test statistics when FDR is controlled with the BH procedure. In the top left panel, ability to satisfy the required FDR level is shown. The solid horizontal line is at the level $\gamma = 0.05$. It can be seen that the realized FDR of $Z_{-1.5,i}$ is almost around or under the required level, while $Z_{\log,i}$ has the largest realized FDR for all different values of λ. Overall, as λ increases, no matter which test statistic we use, the desired level can be achieved.

Let $\widehat{S}$ denote the realized number of correct rejections. We use $\widehat{S}/m_1$ to measure the ability of the test statistics to correctly reject the false hypotheses. As shown in the top right panel of Figure 3.1, the three test statistics have small differences in $\widehat{S}/m_1$. In addition, $\widehat{S}/m_1$ increases slightly as λ increases.

As can be seen in the bottom left panel of Figure 3.1, the significance level $\widehat{i^*}\gamma/m$ obtained from the BH procedure increases as λ increases. As λ goes up, the number of false hypotheses m_1 tends to increase, and we have less possibility that the test statistic will signal a true null as a false one. Consequently, we do not need a more stringent $\widehat{i^*}\gamma/m$ to prevent the false rejections, and more rejections can be obtained.

The average number of rejections $\widehat{S}$ made by the BH procedure is constantly less than the average value of m_1, as shown in the bottom right panel of Figure 3.1. It might be due to that $\gamma = 0.05$ is too restricted to obtain more rejections. A remedy is that we can use a more liberal level ($\gamma = 0.1$ or 0.15), but tolerate more false rejections. One thing worth to note here is that the average values of m_1/m would in general be less than their corresponding λ, since there may be more than one jump on a day, and this becomes even more obvious when λ becomes large.

We then compare performances of the BH procedure with the conventional procedure of controlling type I error in each hypothesis: H_i^0 is rejected if its realized p-value is no greater than α. Here α we specify are two frequently used levels: 0.01 and 0.05. Relevant results are shown in Figure 3.2. As can be seen in the first row, when different test statistics are used, the conventional procedure results in high realized FDR, especially when the jump intensity λ is small (the number of the false null hypotheses tends to be relatively low). An extremely case is that when there is no jump ($\lambda = 0$), rejecting H_i^0 when $\widehat{p}_i \leq 0.01$ (or 0.05) results in 100% false rejections. It says that the probability we at least make one false rejection (the familywise error rate) is one as we follow the conventional procedure. The reason is that when all the null are true and the test statistics for each hypothesis are almost serially independent, if we reject H_i^0 when $\widehat{p}_{i,M} \leq \alpha$, on average we would reject $m\alpha$ hypotheses, and all of these rejections are wrong. However, the BH procedure performs far better in this situation. Even in the worst case, on average it only takes about probability 0.276 to make such an error.

Since the specified $\alpha's$ are on average greater than $\widehat{i^*}\gamma/m$, it is expected that more rejections can be obtained under the conventional procedure than the BH procedure. This can be seen in the second row of Figure 3.2. $\widehat{S}/m_1$ of the conventional procedure tends to be better than that of the BH procedure, but as λ goes up, their gap becomes small.

Figure 3.3 shows performances of the method when lower frequency (5-min, 10-min and 15-min) data is used. $Z_{-1.5,i}$ still has the best ability to satisfy the required FDR levels, but it suffers the greatest loss of $\widehat{S}/m_1$ when the data frequency goes lower. $Z_{\log,i}$ does not perform better than the case when 1-min data is used, no matter in satisfying the required FDR level or $\widehat{S}/m_1$. For $Z_{ratio,i}$, its performance still is in the middle, but overall its performance is more stable than the other two competitors.

We then look at how the method performs when the number of hypotheses changes. We vary m at several different levels, ranging from 50 to 2000 and keep $\gamma = 0.05$. The results are shown in Figure 3.4. It can be seen that when $\lambda \neq 0$ and m is large (no less than 100), the realized FDR and $\widehat{S}/m_1$ are stable over different m.

How does the method perform when FDR is controlled at different required levels? Figure 3.5 shows different required levels γ and the realized FDR. The thick line is a 45-degree line, and the vertical dotted line is for $\gamma = 1/2$. Ideally the realized FDR needs to be equal or below the 45-degree line. For $\lambda = 0.05$ and 0.15, the method performs well, especially when γ goes large. However, when $\lambda = 0$, there is a significant difference between the three test statistics, and the required FDR level becomes difficult to achieve in this situation.

The above results suggest that performances of the hybrid method are positively related to M and λ. Although the BH procedure results in quite stringent rejection criteria, it still can keep $\widehat{S}/m_1$ at a satisfying level. Fixing rejection region at $\alpha = 0.01$ and 0.05 indeed can have better $\widehat{S}/m_1$, but it can suffer far higher false rejections when the number of true null is large. In sum, the simulations show that combining the BNS test with the BH procedure, FDR can be well controlled and the test statistics can also keep substantial ability to correctly identify jump components.

3.5 Real data applications

3.5.1 Summary statistics

The raw data used for the empirical application is the 1-min recorded prices of S&P500 in cash (SPC500) and Dow Jones Industrial Average (DJIA), spanning from January 2003 to December 2007. In order to reduce estimation errors caused by microeconomic structure noise, we use 5-min log returns to estimate RV_i, BV_i and JV_i and the jump test statistics. Some descriptions of the data and discussions on the microstructure issue can be found in the supplementary materials.

Table 2 shows summary statistics of the price variations, different types of $\widehat{T}_{i,M}$, their corresponding $\widehat{p}_{i,M}$ and mutual correlations of these quantities of the two indices. Results of the Ljung-Box test (denote by LB.10) indicate that the price variations are highly serially correlated. However, for $\widehat{T}_{i,M}$ and $\widehat{p}_{i,M}$, the Ljung-Box test instead suggests that they exhibit almost no serial correlation, and the property allows us to use the BH procedure to control the FDR.

The daily test statistics of the two indices have high mutual correlations. This property is quite different from that between individual stocks and the market index. As shown in Bollerslev et al. (2008), the jump test statistics of individual stocks and the market index almost have no mutual correlation, even though their returns are highly correlated. Such low correlation is due to a large amount of idiosyncratic noise in the individual stock returns, which causes a low signal-to-noise ratio in the nonparametric test (the high mutual correlation between the two benchmark indices implies that the idiosyncratic noise of returns is not significant). It suggests that we can have more reliable results when we perform the jump test at the market level.

3.5.2 Common jump days

To measure daily price variation induced by jumps, we use sum of squared intradaily jumps, which can be estimated by the following estimator:

$$JV_{i,\gamma} = JV_i \times \mathbf{1}\left\{\widehat{p}_{i,M} \leq \frac{\widehat{i}^*}{m}\gamma\right\}, \tag{3.11}$$

where $JV_i = RV_i - BV_i$. Table 3 shows summary statistics of $JV_{i,\gamma}$ when FDR is controlled at level $\gamma = 0.01$ and 0.05. The mean and standard deviation of (3.11) shown here are conditional on $\widehat{p}_{i,M} \leq \widehat{i}^*\gamma/m$. The conditional mean is around 0.14 to 0.22 for SPC500 and 0.13 to 0.16 for DJIA. For SPC500 and DJIA, the significant levels $\widehat{i}^*\gamma/m$ for the three statistics are all below 0.006 when the FDR control level $\gamma = 0.05$. Depending on different test statistics, the proportion of identified jump days among all days, is around 1.5% to 11.6% for SP500 and around 2.4% to 8.6% for DJIA.

Common components in two highly correlated asset prices are often one of the most widely studied issues in empirical finance. Here we document some relevant empirical findings. Figure 3.6 shows the time series plots of the identified $JV_{i,\gamma}$ on the common jump days, and Table 4 shows their summary statistics. The term common jump days used here only means that the two indices both have jumps on these days. It does not necessarily mean that the two indices jump exactly at the same time within these days. Since the daily BNS test statistic is obtained by integrated quantities over one day, it cannot tell us how many and what exact time the jumps occur within that day. Nevertheless such test at least let us know what common days they have jumps, and this information is still valuable for further research.

Two different approaches are implemented to identify common jump days. The first approach identifies the jump days of the two indices separately under the same FDR control level, and then find common days among these identified days. Therefore in general we get two different significant levels for the two indices. However, since we take all of the rejections from the two indices together, the separate method cannot guarantee that they satisfy the same FDR level when they are pooled together. Thus the second approach is to pool all the nulls together, and perform the BH procedure to obtain a unified significant level.

It can be seen that the results from the two methods are very similar. When the FDR control level $\gamma = 0.05$, the proportion of common jump days among all jump days is around 41% for SPC500. In addition, this proportion varies from 31% to 51% for DJIA, depending on different test statistics. Comparing magnitudes of the variations in Table 4 with those in Table 3, the two indices tend to have larger jumps on these common days. The result seems to imply that a common shock such as announcements of macroeconomic news, may induce a larger jump than that induced from the shock of the news of individual stocks.

3.5.3 Jump intensity estimation

Jump intensity of an asset price process is a very crucial parameter for evaluating risks of the asset. As shown in Tu and Zhou (2011) and Andersen et al. (2007b), the jump intensity seems to change over time, which implies that clustering of jump variations is time varying. The time varying jump intensity also demonstrates very different dynamic behavior across different assets. In the previous literatures, the time varying jump intensity is estimated via moving average of the number of identified jump days, but the threshold for identifying these jump days is a fixed type I error. Here, rather than controlling the fixed type I error over the whole sampling period, we try to incorporate the FDR control into the rolling window estimation.

The simple moving average (rolling window) intensity estimator for the kth day is defined as

$$\widehat{\lambda}_k^{mov} = \frac{1}{K} \sum_{i=k-K+1}^{k} \mathbf{1} \left\{ \widehat{p}_{i,M} \leq \theta \right\},$$

where θ is a threshold, and K is length of the rolling window. The estimator can serve as a local approximation for the true intensity of the jump process, if we assume that number of jumps occurring at most once per day. In the following analysis, we set $K = 120$, and θ is chosen based on two different ways: The first one is the FDR criterion using the whole $m = 1247$ hypotheses, and the second one is the FDR criterion using the K hypotheses within that window.

While the first method always has θ fixed, the later method leads to an adaptive FDR criterion which may change over time, since including new $\widehat{p}_{i,M}$ sometimes makes a different FDR criterion. As shown in our simulations, a lower value of γ causes an underestimation on the number of jump days, we therefore set a higher $\gamma = 0.15$. Time series plots for the estimations with the three different jump test statistics are illustrated in Figure 3.7. In the left panel are plots for SPC500 and the right panel are plots for DJIA. It can be seen that with $Z_{-1.5,i}$, $\widehat{\lambda}_k^{mov}$ tends to be constantly lower than with the other two test statistics. When θ is chosen adaptively over the whole sampling period, $\widehat{\lambda}_k^{mov}$ is more volatile; and it tends to be higher (lower) when more (less) jump days are identified. This phenomenon holds no matter which test statistic is used. On the other hand, with fixed θ, $\widehat{\lambda}_k^{mov}$ is less sensitive to inform such large price movements. Finally, one should note that adaptively choosing θ is only meaningful if the control procedure can lead to different choices of θ as different information appended, which is possible for the BH procedure but can never be achieved via the conventional type I error control.

3.6 Conclusion

In this paper, we have tested whether a stochastic process has jump components by the BNS nonparametric statistics, and controlled FDR of the multiple testing with the BH procedure.

Theoretical and simulation results are presented to support validity of the hybrid method. Under appropriate conditions, FDR can be asymptotically controlled if the p-values are obtained via the asymptotical distributions. The simulation results show that the transformed BNS test statistics can perform well in satisfying the required FDR level with the BH procedure. Their ability to correctly reject false hypotheses is also improved as the frequency of jumps increases. By controlling FDR, we can have a large chance to avoid any wrong rejection even when the stochastic process does not have any jump component. Overall, our simulation results suggest that performance of the method is positively related to jump intensity and sampling frequency, and is stable over different numbers of hypotheses and the required FDR levels.

As for the empirical results, we find the daily nonparametric test statistics and their corresponding p-values almost have no serial correlation, either for SPC500 or for DJIA. But the test statistics between the two indices are highly mutually dependent. The two indices tend to have larger jumps on the common jump days. We also demonstrate different properties of jump intensity estimations from fixed and adaptive threshold methods. Jump intensity estimated from adaptive threshold method is more sensitive to inform large price movements.

Table 3.1: Number of hypotheses and rejections when a multiple testing is performed

	Test statistic is not significant	Test statistic is significant	Total number
True null hypotheses	U	V	m_0
Non-true null hypotheses	T	S	m_1
Total number	$m-R$	R	m

Table 3.2: The table shows summary statistics of the price variations, different types of $\widehat{T}_{i,M}$, their corresponding $\widehat{p}_{i,M}$ and mutual correlations of these quantities of SPC500 and DJIA. The column LB.10 shows p-values of the Ljung-Box statistic based on autocorrelation coefficients with 10 lagged values. The quantities of price variations shown are all scaled by 10000.

	SPC500			DJIA			
	Mean	Std.	LB.10 p-value	Mean	Std.	LB.10 p-value	Corr.
RV_i	0.4842	0.4817	0.0000	0.4793	0.4505	0.0000	0.9580
BV_i	0.4314	0.4348	0.0000	0.4298	0.4013	0.0000	0.9713
JV_i	0.0368	0.0830	0.0000	0.0326	0.0800	0.0000	0.5311
$\rho=-1.5$							
$\widehat{T}_{i,M}$	0.7782	1.0842	0.5096	0.7034	1.1005	0.4902	0.6925
$\widehat{p}_{i,M}$	0.3024	0.2674	0.8873	0.3222	0.2755	0.6606	0.6767
Log Type							
$\widehat{T}_{i,M}$	0.9390	1.2945	0.3319	0.8568	1.3026	0.4456	0.6814
$\widehat{p}_{i,M}$	0.2916	0.2697	0.8765	0.3113	0.2777	0.6594	0.6741
Ratio Type							
$\widehat{T}_{i,M}$	0.8274	1.1427	0.4499	0.7504	1.1564	0.4772	0.6901
$\widehat{p}_{i,M}$	0.2987	0.2683	0.8837	0.3184	0.2764	0.6598	0.6760

Table 3.3: The table shows summary statistics of significant daily discontinuous quadratic variation $JV_{i,\gamma}$ (sum of squared intradaily jumps) of SPC500 and DJIA. FDR is controlled at level 0.01 and 0.05. The quantities of price variations shown are all scaled by 10000.

	SPC500, $m = 1247$					
	$\gamma = 0.01$			$\gamma = 0.05$		
	$\rho = -1.5$	Log Type	Ratio Type	$\rho = -1.5$	Log Type	Ratio Type
$\frac{\widehat{i^*}}{m}\gamma$	2.41e-05	0.0004	8.82e-05	0.0008	0.0058	0.0023
No. of days	3	55	11	19	144	58
Mean	0.1464	0.1816	0.2232	0.2163	0.1376	0.1744
Std.	0.0727	0.1885	0.2863	0.2311	0.1492	0.1861

	DJIA, $m = 1247$					
	$\gamma = 0.01$			$\gamma = 0.05$		
	$\rho = -1.5$	Log Type	Ratio Type	$\rho = -1.5$	Log Type	Ratio Type
$\frac{\widehat{i^*}}{m}\gamma$	4.01e-05	0.0004	0.0002	0.0012	0.0044	0.0020
No. of days	5	52	22	30	107	51
Mean	0.1458	0.1649	0.1642	0.1662	0.1324	0.1634
Std.	0.0501	0.1198	0.0896	0.1088	0.1053	0.1205

Table 3.4: The table shows summary statistics of significant daily discontinuous quadratic variation $JV_{i,\gamma}$ (sum of squared intradaily jumps) of SPC500 and DJIA on the common jump days by adopting separate and pool methods. The term common jump days used here only means that the two indices both have jumps on these days. The mean and standard deviation of $JV_{i,\gamma}$ are calculated conditional on $\widehat{p}_{i,M} \leq \widehat{i^*}\gamma/m$. The quantities of price variations shown are all scaled by 10000.

	Separate					
	$\gamma = 0.01$			$\gamma = 0.05$		
	$\rho = -1.5$	Log Type	Ratio Type	$\rho = -1.5$	Log Type	Ratio Type
No.of common days	1	22	6	8	55	22
Mean, SPC500	0.2304	0.2289	0.3115	0.2769	0.1690	0.2289
Std., SPC500	N.A.	0.2234	0.3759	0.3244	0.1836	0.2234
Mean, DJIA	0.2101	0.2067	0.2060	0.2021	0.1582	0.2067
Std., DJIA	N.A.	0.1271	0.1459	0.1249	0.1286	0.1271
Corr.	N.A.	0.8503	0.9694	0.9556	0.8889	0.8503

	Pool, $m = 2494$					
	$\gamma = 0.01$			$\gamma = 0.05$		
	$\rho = -1.5$	Log Type	Ratio Type	$\rho = -1.5$	Log Type	Ratio Type
$\frac{\widehat{i^*}}{m}\gamma$	3.21e-05	0.0004	0.0001	0.0010	0.0051	0.0021
No.of common days	1	22	7	9	58	22
No.of days, SPC500	3	55	12	21	140	55
Mean, SPC500	0.2304	0.2289	0.2877	0.2720	0.1647	0.2289
Std., SPC500	N.A.	0.2234	0.3489	0.3038	0.1797	0.2234
No.of days, DJIA	5	52	22	29	114	52
Mean, DJIA	0.2101	0.2067	0.1989	0.2102	0.1535	0.2067
Std., DJIA	N.A.	0.1271	0.1345	0.1193	0.1270	0.1271
Corr.	N.A.	0.8503	0.9694	0.9249	0.8897	0.8503

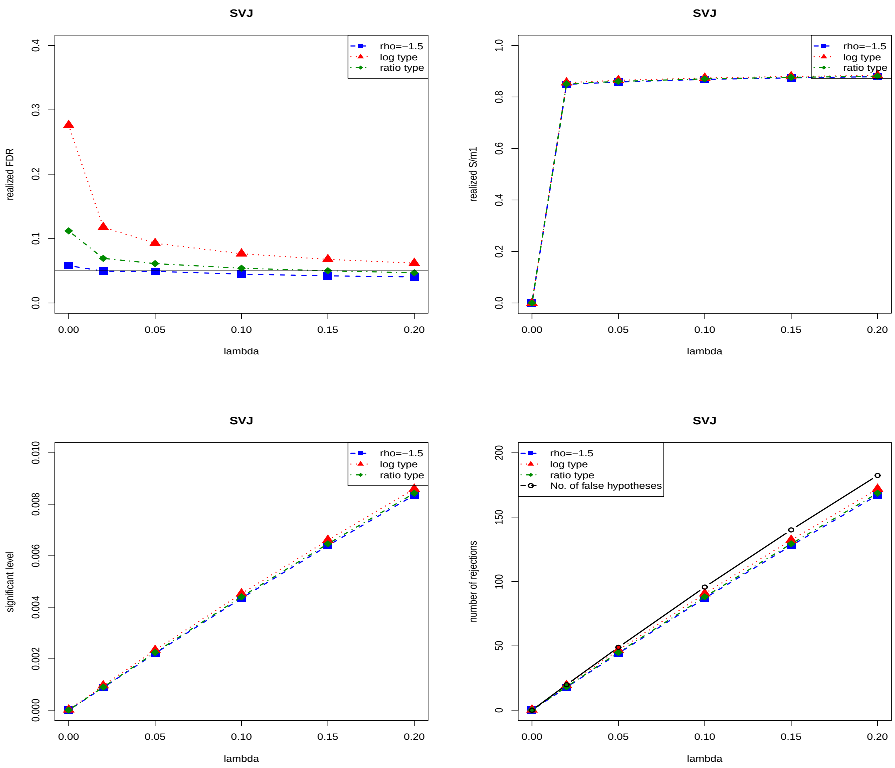

Figure 3.1: Realized FDR, $\hat{S}/m_1$, significance level obtained from the BH procedure and number of rejections. In the graphs, each point is an average value from 1000 simulations.

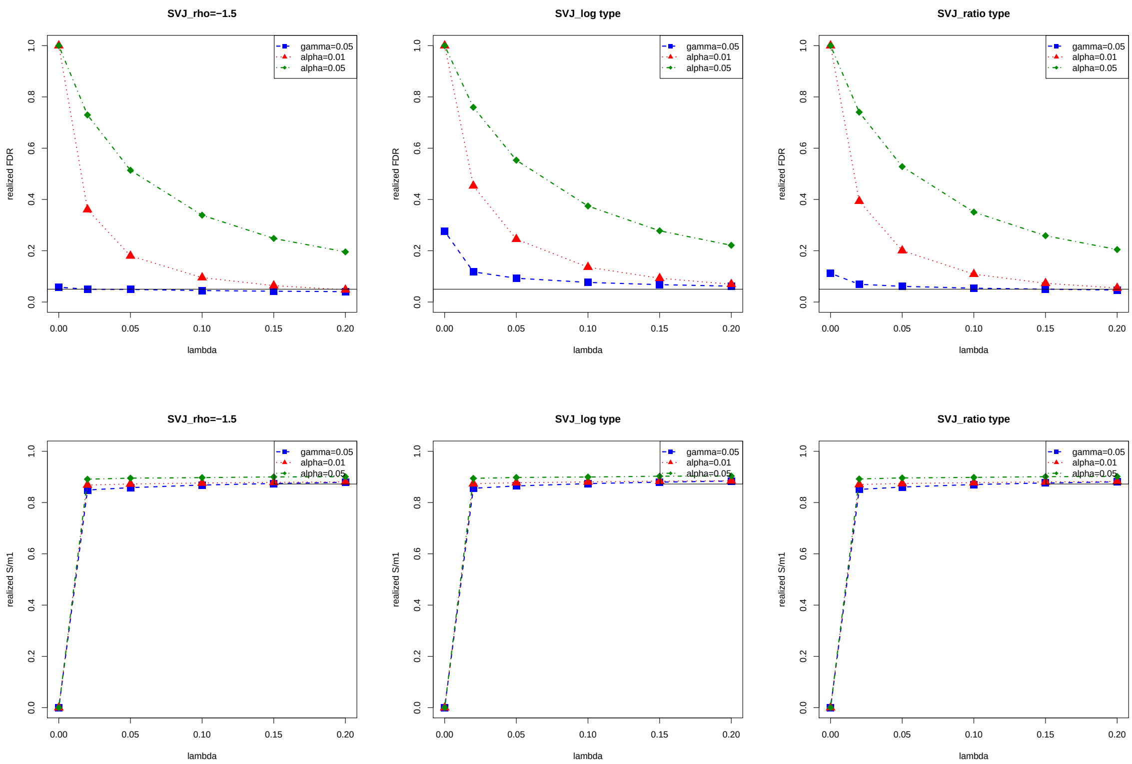

Figure 3.2: Realized FDR and $\hat{S}/m_1$ of the hybrid method and the conventional procedure. In the graphs, each point is an average value from 1000 simulations.

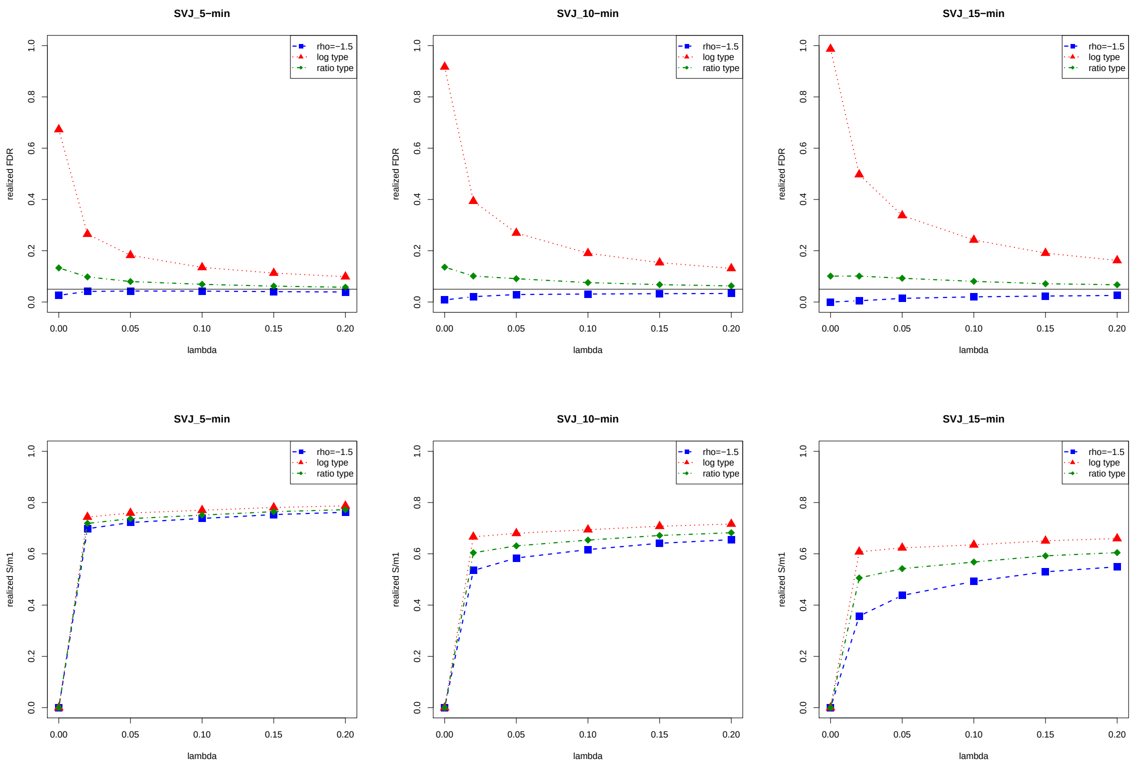

Figure 3.3: Realized FDR and $\hat{S}/m_1$ of the hybrid method with lower frequency data. In the graphs, each point is an average value from 1000 simulations.

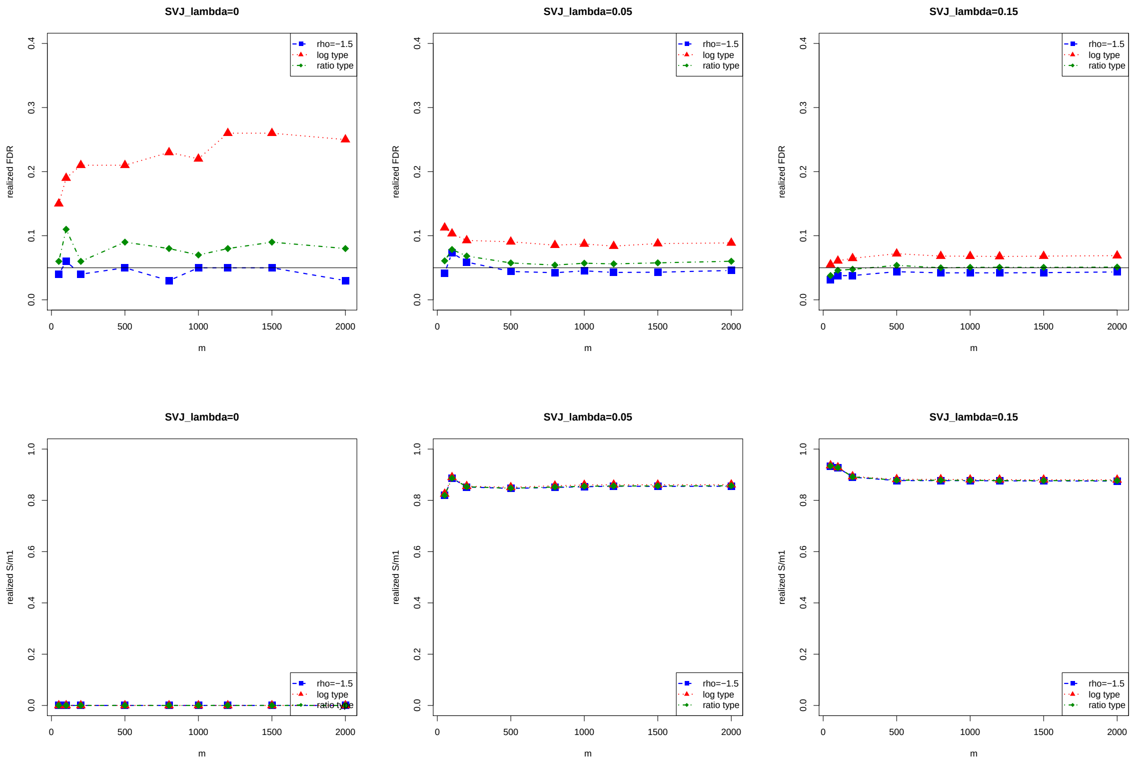

Figure 3.4: Realized FDR and $\hat{S}/m_1$ of the hybrid method when the number of hypotheses varies. Here $m = 50$, 100, 200, 500, 800, 1000, 1200, 1500 and 2000. In the graphs, each point is an average value from 1000 simulations.

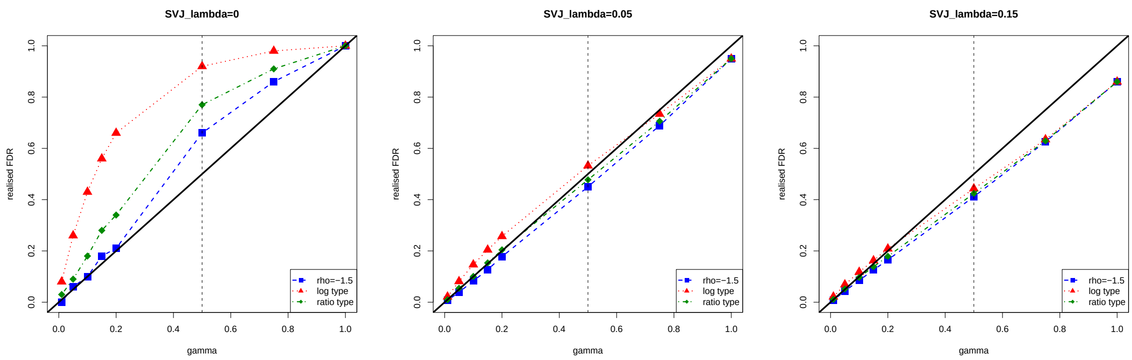

Figure 3.5: Realized FDR of the hybrid method under different required γ. We fix $m = 1000$ in the simulation. In the graphs, each point is an average value from 1000 simulations.

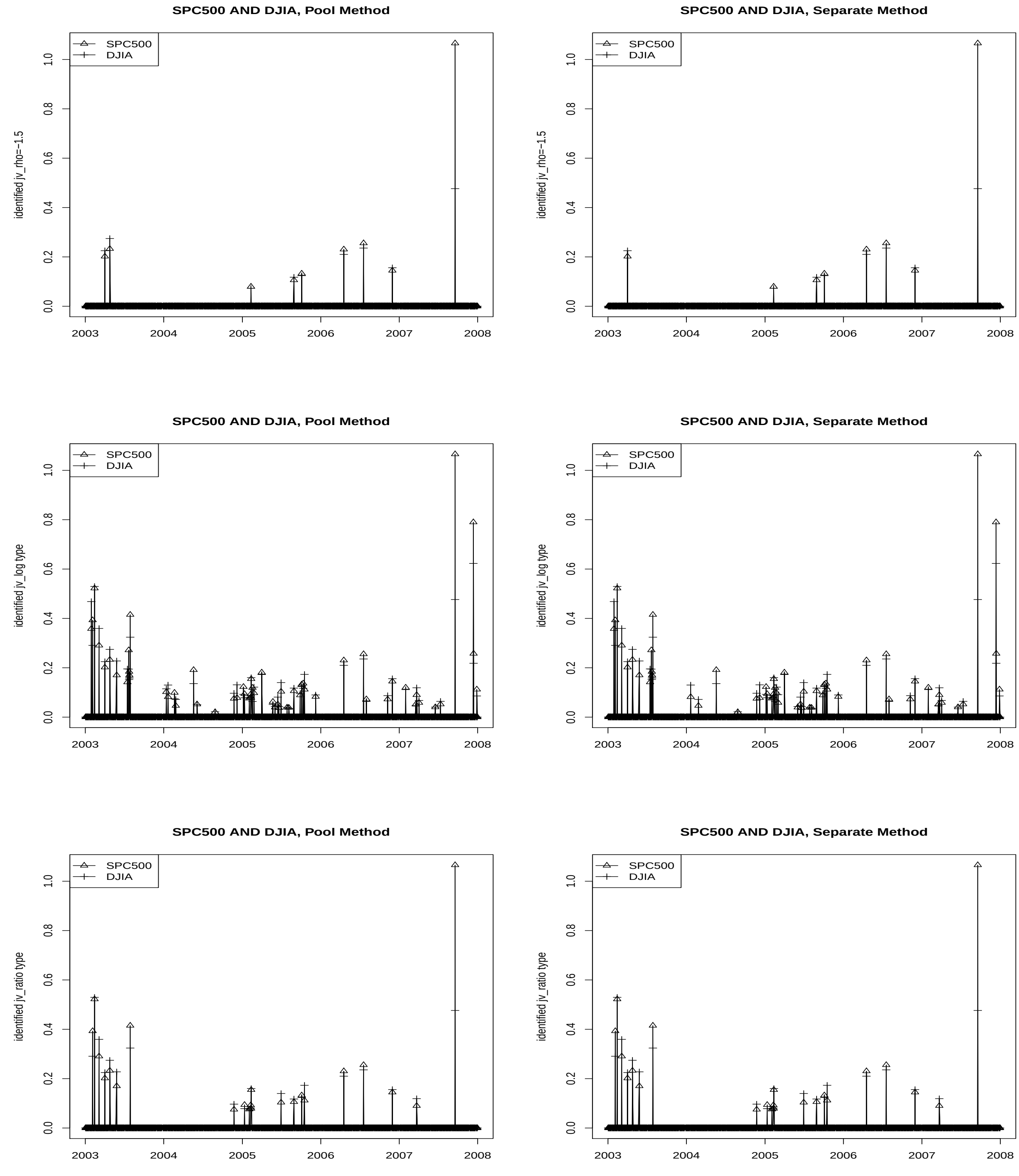

Figure 3.6: Time series plots for identified jump variation on common jump days with the three different jump test statistics. Left: FDR controlled by using the pool method. Right: FDR controlled by using the separate method. The quantities shown here are all scaled by 10000.

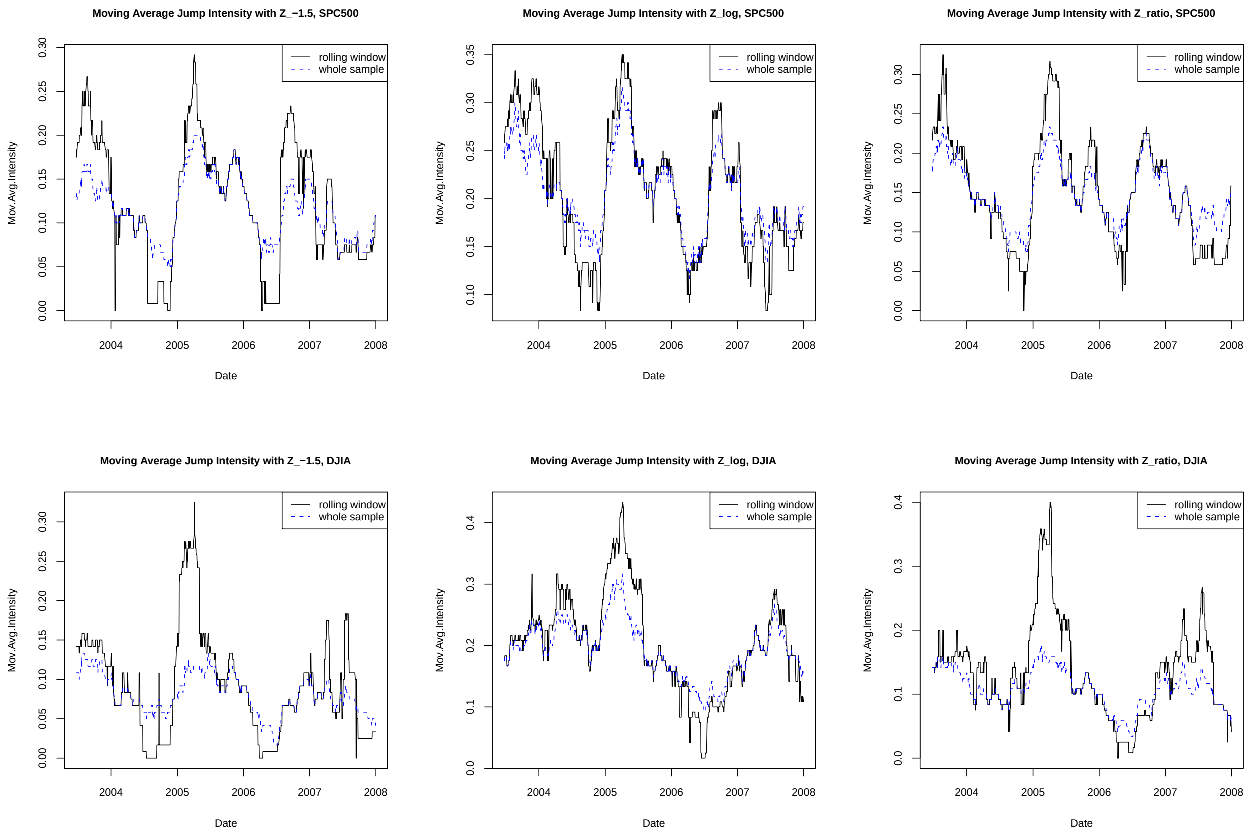

Figure 3.7: Time series plots for jump intensity estimations when different jump test statistics are used.

SUPPLEMENTARY MATERIALS

Supplementary Materials contain the following sections:

7. Some Proofs: This section provides proofs of theoretical results in section 3.

8. The PRDS condition: This section provides a more detailed discussion on the PRDS condition.

9. Simulation with SV1FJ model: This section provides simulation results from another stochastic jump model (SV1FJ).

10. Data descriptions and summary statistics: This section provides descriptions of the real data used in section 5 and summary statistics of realized variance, realized bi-power and jump test statistics derived from the data.

3.7 Proofs of some theoretical results

3.7.1 Proof of Theorem 1

Proof. Let's start our proof from how to construct the $\mathbf{D}_{m_0}^{v,s}$. Without loss of generality, suppose that the first m_0 hypotheses are true, and the rest $m_1 = m - m_0$ hypotheses are false. Now consider events such that we reject the first v true null hypotheses and the first s false hypotheses. Let the optimal significance level selected by the BH procedure $i^*\gamma/m = q_{v+s}$. Then

$$\Pr\begin{pmatrix} \widehat{p}_{M,1} \le q_{v+s}, \dots, \widehat{p}_{M,v} \le q_{v+s}, \widehat{p}_{M,v+1} > q_{v+s+1}, \dots, \widehat{p}_{M,m_0} > q_{m_0+s}, \\ \widehat{p}_{M,m_0+1} \le q_{v+s}, \dots, \widehat{p}_{M,m_0+s} \le q_{v+s}, \widehat{p}_{M,m_0+s+1} > q_{m_0+s+1}, \dots, \widehat{p}_{M,m} > q_m \end{pmatrix}$$

represents probability of one of such events. Note that here $i^* = v + s$, and $q_i = i\gamma/m$ for $i = v + s + 1, \dots, m$ is the crietria corresponding to a hypothesis which is not rejected. Let

$$D_{1,1,m_0}^{v,s} = [0, q_{v+s}]^v \times \prod_{i=v+s+1}^{m_0+s} (q_i, 1] \times [0, q_{v+s}]^s \times \prod_{i=m_0+s+1}^{m} (q_i, 1],$$

and the above probability can be rewritten as $\Pr\left(\widehat{\mathbf{p}}_M \in D_{1,1,m_0}^{v,s}\right)$. Let $\mathbb{E}^m = \prod_{i=1}^{m} [0,1]$ be the $m-$fold products of interval $[0,1]$. Note that joint density of $\widehat{\mathbf{p}}_M$ is integrable over the set $\mathbb{E}^m$. Apparently $D_{1,1,m_0}^{v,s} \subseteq \mathbb{E}^m$, so $\Pr\left(\widehat{\mathbf{p}}_M \in D_{1,1,m_0}^{v,s}\right)$ exists. By suitably varying permutations of intervals $[0, q_{v+s}]$ and $(q_i, 1]$ $(i = v + s + 1, \dots, m)$, we can obtain different $m-$dimensional cubes to construct sets for events of rejecting s false and v true null hypotheses, and the total number of such permutations is $\binom{m_0}{v} \times \binom{m_1}{s} \times (m - s - v)!$.

To see this, at first we focus on the events when $\widehat{p}_{M,1} \le q_{v+s}, \dots, \widehat{p}_{M,v} \le q_{v+s}$ and $\widehat{p}_{M,m_0+1} \le q_{v+s}, \dots, \widehat{p}_{M,m_0+s} \le q_{v+s}$ occur, and the rest p-values are greater than their corresponding significance levels. In this case, there are total $(m-s-v)!$ possible permutations of $(q_i, 1]$ for these non-rejected hypotheses. Let

$$D^{v,s}_{1,m_0} = \bigcup_{j=1}^{(m-s-v)!} D^{v,s}_{1,j,m_0},$$

be union of such events, and also obviously $D^{v,s}_{1,m_0} \subseteq \mathbb{E}^m$. Furthermore, if we vary permutations of the interval $[0, q_{v+s}]$ for the s false (the v true null) hypotheses, there are $\binom{m_1}{s}(\binom{m_0}{v})$ such different permutations. Therefore for the s false and the v true null hypotheses, total number of possible permutations of the interval $[0, q_{v+s}]$ is $\binom{m_0}{v} \times \binom{m_1}{s}$. Let $h^{v,s}_{m_0} = \binom{m_0}{v} \times \binom{m_1}{s}$, and $D^{v,s}_{h,m_0} = \bigcup_{j=1}^{(m-s-v)!} D^{v,s}_{h,j,m_0}$, for $h = 1, \ldots, h^{v,s}_{m_0}$ denote such union of the $m-$dimensional cubes. Finally, let

$$\mathbf{D}^{v,s}_{m_0} = \bigcup_{h=1}^{h^{v,s}_{m_0}} D^{v,s}_{h,m_0} = \bigcup_{h=1}^{h^{v,s}_{m_0}} \bigcup_{j=1}^{(m-s-v)!} D^{v,s}_{h,j,m_0}.$$

$\mathbf{D}^{v,s}_{m_0} \subseteq \mathbb{E}^m$, since all $D^{v,s}_{h,j,m_0} \subseteq \mathbb{E}^m$. When there are m_0 true null hypotheses, the probability of rejecting v true null and s false hypotheses under the BH procedure is thus given by

$$\Pr\left(\bigcup_{h=1}^{h^{v,s}_{m_0}} \bigcup_{j=1}^{(m-s-v)!} \left\{\widehat{\mathbf{p}}_M \in D^{v,s}_{h,j,m_0}\right\}\right) = \Pr\left(\widehat{\mathbf{p}}_M \in \bigcup_{h=1}^{h^{v,s}_{m_0}} \bigcup_{j=1}^{(m-s-v)!} D^{v,s}_{h,j,m_0}\right) = \Pr\left(\widehat{\mathbf{p}}_M \in \mathbf{D}^{v,s}_{m_0}\right).$$

The same approach can be used to construct the probability of rejecting v true null and s false hypotheses when we implement the BH procedure with p, and it is given by $\Pr\left(\mathbf{p} \in \mathbf{D}^{v,s}_{m_0}\right)$. Furthermore, if the consistency for multivariate distribution holds, $\Pr\left(\widehat{\mathbf{p}}_M \in \mathbf{D}^{v,s}_{m_0}\right)$ and $\Pr\left(\mathbf{p} \in \mathbf{D}^{v,s}_{m_0}\right)$ exist when $m \to \infty$.

Then $\mathbb{E}\left(V/R \mid \widetilde{m}_0 = m_0\right)$ and $\mathbb{E}_{\widehat{\mathbf{p}}_M}\left(V/R \mid \widetilde{m}_0 = m_0\right)$ can be expressed as a function of the marginal distributions of p-values. Let us use $\mathbb{E}\left(V/R \mid \widetilde{m}_0 = m_0\right)$ as an example. As shown in Lemma 4.1 of Benjamini and Yekutieli (2001), $\Pr\left(\mathbf{p} \in \mathbf{D}^{v,s}_{m_0}\right)$ can be further expressed as

$$\frac{1}{v} \sum_{i \in I_0} \Pr\left(p_i \leq q_{v+s} \bigcap \left\{\mathbf{p} \in \bigcup_{h=1}^{h^{v,s}_{m_0}} D^{v,s}_{h,m_0}\right\}\right),$$

and therefore

$$
\begin{aligned}
\mathbb{E}\left(\frac{V}{R} \mid \widetilde{m}_0 = m_0\right) &= \sum_{s=0}^{m_1}\sum_{v=1}^{m_0}\left(\frac{v}{v+s}\Pr\left(\mathbf{p}\in\mathbf{D}_{m_0}^{v,s}\right)\right) \\
&= \sum_{s=0}^{m_1}\sum_{v=1}^{m_0}\left(\frac{v}{v+s}\Pr\left(\mathbf{p}\in\bigcup_{h=1}^{h_{m_0}^{v,s}} D_{h,m_0}^{v,s}\right)\right) \\
&= \sum_{s=0}^{m_1}\sum_{v=1}^{m_0}\left(\frac{v}{v+s}\frac{1}{v}\sum_{i\in I_0}\Pr\left(p_i\le q_{v+s}\bigcap\left\{\mathbf{p}\in\bigcup_{h=1}^{h_{m_0}^{v,s}} D_{h,m_0}^{v,s}\right\}\right)\right) \\
&= \sum_{s=0}^{m_1}\sum_{v=1}^{m_0}\sum_{i\in I_0}\frac{1}{v+s}\Pr\left(p_i\le q_{v+s}\bigcap\left\{\mathbf{p}\in\bigcup_{h=1}^{h_{m_0}^{v,s}} D_{h,m_0}^{v,s}\right\}\right).
\end{aligned}
$$

Let $\Lambda_{(i),m_0}^{v,s}$ denote the event that if $p_i \le q_{v+s}$ occurs and then $v-1$ true null and s false hypotheses are rejected. We can see that

$$
\{p_i\le q_{v+s}\}\bigcap\left\{\mathbf{p}\in\bigcup_{h=1}^{h_{m_0}^{v,s}} D_{h,m_0}^{v,s}\right\} = \{p_i\le q_{v+s}\}\bigcap\Lambda_{(i),m_0}^{v,s}.
$$

Also let

$$
q_k = \{q_{v+s} : v+s=k\} = \frac{k}{m}\alpha, \text{ and } \Lambda_{(i),m_0}^{k} = \bigcup\left\{\Lambda_{(i),m_0}^{v,s} : v+s=k\right\}.
$$

Note that $\Lambda_{(i),m_0}^{v,s}$ is mutually disjoint for different v and s. $\Lambda_{(i),m_0}^{k}$ is the event that except H_i^0, we reject the other $k-1$ hypotheses given m_0 true null hypotheses, and it is disjoint for different i. Then

$$
\begin{aligned}
\mathbb{E}\left(\frac{V}{R} \mid \widetilde{m}_0 = m_0\right) &= \sum_{s=0}^{m_1}\sum_{v=1}^{m_0}\sum_{i\in I_0}\frac{1}{v+s}\Pr\left(p_i\le q_{v+s}\bigcap\left\{\mathbf{p}\in\bigcup_{h=1}^{h_{m_0}^{v,s}} D_{h,m_0}^{v,s}\right\}\right) \\
&= \sum_{s=0}^{m_1}\sum_{v=1}^{m_0}\sum_{i\in I_0}\frac{1}{v+s}\Pr\left(p_i\le q_{v+s}\bigcap\Lambda_{(i),m_0}^{v,s}\right) \\
&= \sum_{k=1}^{m}\sum_{i\in I_0}\frac{1}{k}\Pr\left(p_i\le q_k\bigcap\Lambda_{(i),m_0}^{k}\right).
\end{aligned}
$$

Considering $\Pr\left(\widehat{p}_{M,i}\le q_k\bigcap\widehat{\Lambda}_{(i),m_0}^{k}\right)$, an analog of $\Pr\left(p_i\le q_k\bigcap\Lambda_{(i),m_0}^{k}\right)$ when $\widehat{\mathbf{p}}_M$ is used. Following the same way,

$$
\mathbb{E}_{\widehat{\mathbf{p}}_M}\left(\frac{V}{R} \mid \widetilde{m}_0 = m_0\right) = \sum_{k=1}^{m}\sum_{i\in I_0}\frac{1}{k}\Pr\left(\widehat{p}_{M,i}\le q_k\bigcap\widehat{\Lambda}_{(i),m_0}^{k}\right).
$$

Thus

$$
\begin{aligned}
&\left|\mathbb{E}_{\widehat{\mathbf{p}}_M}\left(\frac{V}{R} \mid \widetilde{m}_0 = m_0\right) - \mathbb{E}\left(\frac{V}{R} \mid \widetilde{m}_0 = m_0\right)\right| \\
= &\left|\sum_{k=1}^{m}\sum_{i\in I_0}\frac{1}{k}\left(\Pr\left(\widehat{p}_{M,i} \le q_k \bigcap \widehat{\Lambda}^k_{(i),m_0}\right) - \Pr\left(p_i \le q_k \bigcap \Lambda^k_{(i),m_0}\right)\right)\right|.
\end{aligned}
$$

Note that the consistency for multivariate distribution should hold, then the above joint probability functions exist when $m \to \infty$. $\Pr\left(p_i \le q_k \bigcap \Lambda^k_{(i),m_0}\right)$ is just the probability that if $p_i \le q_k$, then the other $k-1$ hypotheses are rejected. Therefore $\Pr\left(p_i \le q_k \bigcap \Lambda^k_{(i),m_0}\right)$ can be explicitly expressed as

$$
\begin{aligned}
&\Pr\left(p_i \le q_k \bigcap \Lambda^k_{(i),m_0}\right) \\
= &\Pr\left(p_i \le q_k, p^{(-i)}_{(k-1)} \le q_k, p^{(-i)}_{(k)} > q_{k+1}, \ldots p^{(-i)}_{(m-1)} > q_m\right) \\
= &\Pr\left(p_i \le q_k, p^{(-i)}_{(k)} > q_{k+1}, \ldots p^{(-i)}_{(m-1)} > q_m\right) \\
&-\Pr\left(p_i \le q_k, p^{(-i)}_{(k-1)} > q_k, p^{(-i)}_{(k)} > q_{k+1}, \ldots p^{(-i)}_{(m-1)} > q_m\right).
\end{aligned}
$$

Then

$$
\sum_{k=1}^{m}\frac{1}{k}\Pr\left(p_i \le q_k \bigcap \Lambda^k_{(i),m_0}\right) = \sum_{k=1}^{m}\frac{1}{k}\left(\begin{array}{c}\Pr\left(p_i \le q_k, p^{(-i)}_{(k)} > q_{k+1}, \ldots p^{(-i)}_{(m-1)} > q_m\right) \\ -\Pr\left(p_i \le q_k, p^{(-i)}_{(k-1)} > q_k, p^{(-i)}_{(k)} > q_{k+1}, \ldots p^{(-i)}_{(m-1)} > q_m\right)\end{array}\right).
$$

The first term of the above summation ($k=1$) is $\Pr\left(p_i \le q_1, p^{(-i)}_{(1)} > q_2, \ldots p^{(-i)}_{(m-1)} > q_m\right)$, while the last term ($k=m$) is$1/m\left(\Pr\left(p_i \le q_m\right) - \Pr\left(p_i \le q_m, p^{(-i)}_{(m-1)} > q_m\right)\right)$. Summation of the middle $m-2$ terms is

$$
\begin{aligned}
&\sum_{k=2}^{m-1}\frac{1}{k}\left(\begin{array}{c}\Pr\left(p_i \le q_k, p^{(-i)}_{(k)} > q_{k+1}, \ldots p^{(-i)}_{(m-1)} > q_m\right) \\ -\Pr\left(p_i \le q_k, p^{(-i)}_{(k-1)} > q_k, p^{(-i)}_{(k)} > q_{k+1}, \ldots p^{(-i)}_{(m-1)} > q_m\right)\end{array}\right) \\
= &\sum_{k=2}^{m-1}\frac{1}{k}\Pr\left(p_i \le q_k, p^{(-i)}_{(k)} > q_{k+1}, \ldots p^{(-i)}_{(m-1)} > q_m\right) \\
&-\sum_{k=1}^{m-2}\frac{1}{k+1}\Pr\left(p_i \le q_{k+1}, p^{(-i)}_{(k)} > q_{k+1}, p^{(-i)}_{(k+1)} > q_{k+2}, \ldots p^{(-i)}_{(m-1)} > q_m\right) \\
= &\sum_{k=1}^{m-1}\frac{1}{k}\Pr\left(p_i \le q_k, p^{(-i)}_{(k)} > q_{k+1}, \ldots p^{(-i)}_{(m-1)} > q_m\right) - \\
&\sum_{k=1}^{m-1}\frac{1}{k+1}\Pr\left(p_i \le q_{k+1}, p^{(-i)}_{(k)} > q_{k+1}, p^{(-i)}_{(k+1)} > q_{k+2}, \ldots p^{(-i)}_{(m-1)} > q_m\right) - \\
&\Pr\left(p_i \le q_1, p^{(-i)}_{(1)} > q_{2+1}, \ldots p^{(-i)}_{(m-1)} > q_m\right) + \frac{1}{m}\Pr\left(p_i \le q_m, p^{(-i)}_{(m-1)} > q_m\right).
\end{aligned}
$$

Therefore

$$\begin{aligned}
&\sum_{k=1}^{m}\frac{1}{k}\Pr\left(p_i\le q_k\bigcap\Lambda_{(i),m_0}^{k}\right)\\
=&\sum_{k=1}^{m-1}\left(\begin{array}{c}\frac{1}{k}\Pr\left(p_i\le q_k,p_{(k)}^{(-i)}>q_{k+1},\ldots,p_{(m-1)}^{(-i)}>q_m\right)\\ -\frac{1}{k+1}\Pr\left(p_i\le q_{k+1},p_{(k)}^{(-i)}>q_{k+1},\ldots,p_{(m-1)}^{(-i)}>q_m\right)\end{array}\right)+\frac{1}{m}\Pr\left(p_i\le q_m\right)
\end{aligned}$$

By similar way,

$$\begin{aligned}
&\sum_{k=1}^{m}\frac{1}{k}\Pr\left(\widehat{p}_{M,i}\le q_k\bigcap\widehat{\Lambda}_{(i),m_0}^{k}\right)\\
=&\sum_{k=1}^{m-1}\left(\begin{array}{c}\frac{1}{k}\Pr\left(\widehat{p}_{M,i}\le q_k,\widehat{p}_{M,(k)}^{(-i)}>q_{k+1},\ldots,\widehat{p}_{M,(m-1)}^{(-i)}>q_m\right)\\ -\frac{1}{k+1}\Pr\left(\widehat{p}_{M,i}\le q_{k+1},\widehat{p}_{M,(k)}^{(-i)}>q_{k+1},\ldots,\widehat{p}_{M,(m-1)}^{(-i)}>q_m\right)\end{array}\right)\\
&+\frac{1}{m}\Pr\left(\widehat{p}_{M,i}\le q_m\right).
\end{aligned}$$

Note that $q_k=k\gamma/m$, so in general as m goes large,

$$\Pr\left(p_i\le q_k,p_{(k)}^{(-i)}>q_{k+1},\ldots,p_{(m-1)}^{(-i)}>q_m\right)\approx\Pr\left(p_i\le q_{k+1},p_{(k)}^{(-i)}>q_{k+1},\ldots,p_{(m-1)}^{(-i)}>q_m\right).$$

Then

$$\begin{aligned}
&\sum_{k=1}^{m-1}\left(\begin{array}{c}\frac{1}{k}\Pr\left(p_i\le q_k,p_{(k)}^{(-i)}>q_{k+1},\ldots,p_{(m-1)}^{(-i)}>q_m\right)\\ -\frac{1}{k+1}\Pr\left(p_i\le q_{k+1},p_{(k)}^{(-i)}>q_{k+1},\ldots,p_{(m-1)}^{(-i)}>q_m\right)\end{array}\right)\\
\approx&\sum_{k=1}^{m-1}\left(\begin{array}{c}\frac{1}{k}\Pr\left(p_i\le q_k,p_{(k)}^{(-i)}>q_{k+1},\ldots,p_{(m-1)}^{(-i)}>q_m\right)\\ -\frac{1}{k+1}\Pr\left(p_i\le q_k,p_{(k)}^{(-i)}>q_{k+1},\ldots,p_{(m-1)}^{(-i)}>q_m\right)\end{array}\right)\\
=&\sum_{k=1}^{m-1}\frac{1}{k\left(k+1\right)}\Pr\left(p_i\le q_k,p_{(k)}^{(-i)}>q_{k+1},\ldots,p_{(m-1)}^{(-i)}>q_m\right).
\end{aligned}$$

Also

$$\begin{aligned}
&\sum_{k=1}^{m-1}\left(\begin{array}{c}\frac{1}{k}\Pr\left(\widehat{p}_{M,i}\le q_k,\widehat{p}_{M,(k)}^{(-i)}>q_{k+1},\ldots,\widehat{p}_{M,(m-1)}^{(-i)}>q_m\right)\\ -\frac{1}{k+1}\Pr\left(\widehat{p}_{M,i}\le q_{k+1},\widehat{p}_{M,(k)}^{(-i)}>q_{k+1},\ldots,\widehat{p}_{M,(m-1)}^{(-i)}>q_m\right)\end{array}\right)\\
\approx&\sum_{k=1}^{m-1}\frac{1}{k\left(k+1\right)}\Pr\left(\widehat{p}_{M,i}\le q_k,\widehat{p}_{M,(k)}^{(-i)}>q_{k+1},\ldots,\widehat{p}_{M,(m-1)}^{(-i)}>q_m\right).
\end{aligned}$$

Finally

$$
\begin{aligned}
&\left|\mathbb{E}_{\widehat{\mathbf{p}}_M}\left(\frac{V}{R} \mid \widetilde{m}_0=m_0\right)-\mathbb{E}\left(\frac{V}{R} \mid \widetilde{m}_0=m_0\right)\right| \\
= &\left|\sum_{i\in I_0}\sum_{k=1}^{m}\frac{1}{k}\left(\Pr\left(\widehat{p}_{M,i}\leq q_k\bigcap\widehat{\Lambda}^k_{(i),m_0}\right)-\Pr\left(p_i\leq q_k\bigcap\Lambda^k_{(i),m_0}\right)\right)\right| \\
\approx &\left|\begin{array}{c}\sum_{i\in I_0}\sum_{k=1}^{m-1}\frac{1}{k(k+1)}\left(\begin{array}{c}\Pr\left(\widehat{p}_{M,i}\leq q_k,\widehat{p}^{(-i)}_{M,(k)}>q_{k+1},\ldots,\widehat{p}^{(-i)}_{M,(m-1)}>q_m\right)- \\ \Pr\left(p_i\leq q_k,p^{(-i)}_{(k)}>q_{k+1},\ldots,p^{(-i)}_{(m-1)}>q_m\right)\end{array}\right) \\ +\sum_{i\in I_0}\frac{1}{m}\left(\Pr\left(\widehat{p}_{M,i}\leq q_m\right)-\Pr\left(p_i\leq q_m\right)\right)\end{array}\right| \\
\leq &\left|\sum_{i\in I_0}\sum_{k=1}^{m-1}\frac{1}{k\left(k+1\right)}\left(\begin{array}{c}\Pr\left(\widehat{p}_{M,i}\leq q_k,\widehat{p}^{(-i)}_{M,(k)}>q_{k+1},\ldots,\widehat{p}^{(-i)}_{M,(m-1)}>q_m\right)- \\ \Pr\left(p_i\leq q_k,p^{(-i)}_{(k)}>q_{k+1},\ldots,p^{(-i)}_{(m-1)}>q_m\right)\end{array}\right)\right| \\
&+\left|\sum_{i\in I_0}\frac{1}{m}\left(\Pr\left(\widehat{p}_{M,i}\leq q_m\right)-\Pr\left(p_i\leq q_m\right)\right)\right|.
\end{aligned}
$$

If condition 4 holds, the second term of the last inequality is bounded by $O\left(1/M^{\delta}\right)$. If condition 5 hold, the first term of the last inequality becomes

$$
\begin{aligned}
&\left|\sum_{i\in I_0}\sum_{k=1}^{m-1}\frac{1}{k\left(k+1\right)}\left(\begin{array}{c}\Pr\left(\widehat{p}_{M,i}\leq q_k,\widehat{p}^{(-i)}_{M,(k)}>q_{k+1},\ldots,\widehat{p}^{(-i)}_{M,(m-1)}>q_m\right)- \\ \Pr\left(p_i\leq q_k,p^{(-i)}_{(k)}>q_{k+1},\ldots,p^{(-i)}_{(m-1)}>q_m\right)\end{array}\right)\right| \\
\leq\ & m_0\left(1-\frac{1}{m}\right)\frac{m}{m}\times\sup_{1\leq k\leq m}\sup_{i\in I_0}\left|\left(\begin{array}{c}\Pr\left(\widehat{p}_{M,i}\leq q_k,\widehat{p}^{(-i)}_{M,(k)}>q_{k+1},\ldots,\widehat{p}^{(-i)}_{M,(m-1)}>q_m\right) \\ -\Pr\left(p_i\leq q_k,p^{(-i)}_{(k)}>q_{k+1},\ldots,p^{(-i)}_{(m-1)}>q_m\right)\end{array}\right)\right| \\
=\ & \frac{m_0}{m}\left(1-\frac{1}{m}\right)o\left(1\right).
\end{aligned}
$$

We then can conclude that

$$
\left|\mathbb{E}_{\widehat{\mathbf{p}}_M}\left(\frac{V}{R} \mid \widetilde{m}_0=m_0\right)-\mathbb{E}\left(\frac{V}{R} \mid \widetilde{m}_0=m_0\right)\right|\leq\frac{m_0}{m}\left(1-\frac{1}{m}\right)o\left(1\right)+\frac{m_0}{m}O\left(\frac{1}{M^{\delta}}\right).
$$

Then

$$\begin{aligned}
& \left|\mathbb{E}_{\widehat{\mathbf{p}}_M}\left(\frac{V}{R}\right) - \mathbb{E}\left(\frac{V}{R}\right)\right| \\
= & \left| \begin{array}{c} \sum_{m_0=0}^{m} \mathbb{E}_{\widehat{p}_M}\left(\frac{V}{R} \mid \widetilde{m}_0 = m_0\right) \times \Pr\left(\widetilde{m}_0 = m_0\right) \\ -\sum_{m_0=0}^{m} \mathbb{E}\left(\frac{V}{R} \mid \widetilde{m}_0 = m_0\right) \times \Pr\left(\widetilde{m}_0 = m_0\right) \end{array} \right| \\
= & \left| \sum_{m_0=0}^{m} \left(\mathbb{E}_{\widehat{\mathbf{p}}_M}\left(\frac{V}{R} \mid \widetilde{m}_0 = m_0\right) - \mathbb{E}\left(\frac{V}{R} \mid \widetilde{m}_0 = m_0\right)\right) \times \Pr\left(\widetilde{m}_0 = m_0\right) \right| \\
\leq & \sum_{m_0=0}^{m} \left(\frac{m_0}{m}\left(1 - \frac{1}{m}\right) o(1) + \frac{m_0}{m} O\left(\frac{1}{M^\delta}\right)\right) \times \Pr\left(\widetilde{m}_0 = m_0\right) \\
= & \mathbb{E}\left(\widetilde{m}_0\right)\left(\frac{1}{m}\left(1 - \frac{1}{m}\right) o(1) + \frac{1}{m} O\left(\frac{1}{M^\delta}\right)\right) = o(1).
\end{aligned}$$

As shown in the proof of Theorem 1.2 of Benjamini and Yekutieli (2001), if condition 2 holds, then $\sum_{k=1}^{m} \Pr\left(\Lambda_{(i),m_0}^{k} | p_i \leq q_k\right) \leq 1$. By the assumption that $\Pr\left(p_i \leq q_k\right) \leq \frac{k}{m}\gamma$,

$$\Pr\left(\{p_i \leq q_k\} \bigcap \Lambda_{(i),m_0}^{k}\right) \leq \Pr\left(\Lambda_{(i),m_0}^{k} | p_i \leq q_k\right) \frac{k}{m}\gamma.$$

Thus

$$\begin{aligned}
\mathbb{E}\left(\frac{V}{R} \mid \widetilde{m}_0 = m_0\right) &= \sum_{k=1}^{m} \sum_{i \in I_0} \frac{1}{k} \Pr\left(\{p_i \leq q_k\} \bigcap \Lambda_{(i),m_0}^{k}\right) \\
&= \sum_{k=1}^{m} \sum_{i \in I_0} \frac{1}{k} \Pr\left(\Lambda_{(i),m_0}^{k} | p_i \leq q_k\right) \Pr\left(p_i \leq q_k\right) \\
&\leq \sum_{k=1}^{m} \sum_{i \in I_0} \frac{1}{k} \Pr\left(\Lambda_{(i),m_0}^{k} | p_i \leq q_k\right) \frac{k}{m}\gamma \\
&= \frac{\gamma}{m} \sum_{i \in I_0} \sum_{k=1}^{m} \Pr\left(\Lambda_{(i),m_0}^{k} | p_i \leq q_k\right) \leq \frac{m_0 \gamma}{m} \leq \gamma, \\
\mathbb{E}\left(\frac{V}{R}\right) &= \sum_{m_0=0}^{m} \mathbb{E}\left(\frac{V}{R} \mid \widetilde{m}_0 = m_0\right) \times \Pr\left(\widetilde{m}_0 = m_0\right) \\
&\leq \sum_{m_0=0}^{m} \frac{m_0}{m}\gamma \times \Pr\left(\widetilde{m}_0 = m_0\right) = \frac{\mathbb{E}\left(\widetilde{m}_0\right)\gamma}{m} \leq \gamma.
\end{aligned}$$

Finally we can conclude that

$$\lim_{M \to \infty} \mathbb{E}_{\widehat{p}_M}\left(\frac{V}{R}\right) = \mathbb{E}\left(\frac{V}{R}\right) \leq \gamma.$$

■

3.7.2 Proof of Theorem 2

Proof. To start our proof, at first we have a look of the inequality,

$$\Pr\left(\widehat{p}_{M,i} \leq a\right) \leq a,$$

where $a \in (0,1)$ and $i \in I_0$. Suppose that $a = q_k = k\gamma/m$, $k = 1, \ldots, m$, and $\gamma \in (0,1)$, then the above inequality becomes

$$\Pr\left(\widehat{p}_{M,i} \leq q_k\right) \leq \frac{k}{m}\gamma.$$

It implies $m\Pr\left(\widehat{p}_{M,i} \leq q_k\right) \leq k\gamma$ for all $k = 1, \ldots, m$ and and $i \in I_0$. Let

$$\frac{m}{k}\Pr\left(\widehat{p}_{M,i} \leq q_k\right) = F_{\widehat{p}_{i,M}}\left(q_k\right),$$

therefore for $i \in I_0$, $F_{\widehat{p}_{i,M}}\left(q_k\right)$ is bound by γ as $m \to \infty$. Furthermore, since $T_1, \ldots, T_m$ are continuous random variables, $\Pr\left(p_i \leq q_k\right) = k\gamma/m$. Let $F_{p_i}\left(q_k\right) = m\Pr\left(p_i \leq a\right)/k$, then for $i \in I_0$, $F_{p_i}\left(q_k\right)$ is also bounded. Since both $F_{p_i}\left(q_k\right)$ and $F_{\widehat{p}_{i,M}}\left(q_k\right)$ are bound and continuous functions of $\Pr\left(p_i \leq q_k\right)$ and $\Pr\left(\widehat{p}_{M,i} \leq q_k\right)$ respectively, we can conclude that as $M \to \infty$, if

$$\sup_{1\leq k\leq m}\sup_{i\in I_0}\left|\Pr\left(\widehat{p}_{M,i} \leq q_k\right) - \Pr\left(p_i \leq q_k\right)\right| = O\left(\frac{1}{M^\delta}\right),$$

then

$$\begin{aligned} &\sup_{1\leq k\leq m}\sup_{i\in I_0}\left|F_{\widehat{p}_{M,i}}\left(q_k\right) - F_{p_i}\left(q_k\right)\right| \\ = &\sup_{1\leq k\leq m}\sup_{i\in I_0}\left|\frac{m}{k}\Pr\left(\widehat{p}_{M,i} \leq q_k\right) - \frac{m}{k}\Pr\left(p_i \leq q_k\right)\right| = O\left(\frac{1}{M^\delta}\right). \end{aligned}$$

Since $T_1, \ldots, T_m$ are independent, then $p_1, \ldots, p_m$ are also independent. Therefore the event $\Lambda^k_{(i),m_0}$ and $\{p_i \leq q_k\}$ are independent, and $\Pr\left(\Lambda^k_{(i),m_0}|p_i \leq q_k\right) = \Pr\left(\Lambda^k_{(i),m_0}\right)$. Furthermore, by $\Lambda^k_{(i),m_0}$ are mutually exclusive for k and $\bigcup_{k=1}^{m}\Lambda^k_{(i),m_0}$ is the whole space, therefore

$$\sum_{k=1}^{m}\Pr\left(\Lambda^k_{(i),m_0}|p_i \leq q_k\right) = \sum_{k=1}^{m}\Pr\left(\Lambda^k_{(i),m_0}\right) = \Pr\left(\bigcup_{k=1}^{m}\Lambda^k_{(i),m_0}\right) = 1.$$

Since $\widehat{T}_{M,1},\ldots,\widehat{T}_{M,m}$ are also mutually independent, by similar argument as above, $\sum_{k=1}^{m}\Pr\left(\widehat{\Lambda}^k_{(i),m_0}\right)=\Pr\left(\bigcup_{k=1}^{m}\widehat{\Lambda}^k_{(i),m_0}\right)=1$. From proof of Theorem 1, we know that

$$\begin{aligned}
&\left|\mathbb{E}_{\widehat{\mathbf{p}}_M}\left(\frac{V}{R}\mid\widetilde{m}_0=m_0\right)-\mathbb{E}\left(\frac{V}{R}\mid\widetilde{m}_0=m_0\right)\right|\\
=\ &\left|\sum_{k=1}^{m}\sum_{i\in I_0}\frac{1}{k}\left(\Pr\left(\widehat{p}_{M,i}\le q_k\bigcap\widehat{\Lambda}^k_{(i),m_0}\right)-\Pr\left(p_i\le q_k\bigcap\Lambda^k_{(i),m_0}\right)\right)\right|.
\end{aligned}$$

It can be shown that

$$\begin{aligned}
&\Pr\left(\widehat{p}_{M,i}\le q_k\bigcap\widehat{\Lambda}^k_{(i),m_0}\right)-\Pr\left(p_i\le q_k\bigcap\Lambda^k_{(i),m_0}\right)\\
=\ &\Pr\left(\widehat{\Lambda}^k_{(i),m_0}|\widehat{p}_{M,i}\le q_k\right)\Pr\left(\widehat{p}_{M,i}\le q_k\right)-\Pr\left(\widehat{\Lambda}^k_{(i),m_0}|\widehat{p}_{M,i}\le q_k\right)\Pr\left(p_i\le q_k\right)\\
&+\Pr\left(\widehat{\Lambda}^k_{(i),m_0}|\widehat{p}_{M,i}\le q_k\right)\Pr\left(p_i\le q_k\right)-\Pr\left(\Lambda^k_{(i),m_0}|p_i\le q_k\right)\Pr\left(p_i\le q_k\right)\\
=\ &\Pr\left(\widehat{\Lambda}^k_{(i),m_0}|\widehat{p}_{M,i}\le q_k\right)\left(\Pr\left(\widehat{p}_{M,i}\le q_k\right)-\Pr\left(p_i\le q_k\right)\right)\\
&+\left(\Pr\left(\widehat{\Lambda}^k_{(i),m_0}|\widehat{p}_{M,i}\le q_k\right)-\Pr\left(\Lambda^k_{(i),m_0}|p_i\le q_k\right)\right)\Pr\left(p_i\le q_k\right).
\end{aligned}$$

Therefore

$$\begin{aligned}
& \left|\mathbb{E}_{\widehat{\mathbf{p}}_M}\left(\frac{V}{R} \mid \widetilde{m}_0 = m_0\right) - \mathbb{E}\left(\frac{V}{R} \mid \widetilde{m}_0 = m_0\right)\right| \\
= & \left|\sum_{k=1}^{m}\sum_{i\in I_0}\frac{1}{k}\left(\Pr\left(\widehat{p}_{M,i} \le q_k \bigcap \widehat{\Lambda}^k_{(i),m_0}\right) - \Pr\left(p_i \le q_k \bigcap \Lambda^k_{(i),m_0}\right)\right)\right| \\
\le & \left|\sum_{i\in I_0}\sum_{k=1}^{m}\frac{1}{k}\Pr\left(\widehat{\Lambda}^k_{(i),m_0}|\widehat{p}_{M,i} \le q_k\right)\left(\Pr\left(\widehat{p}_{M,i} \le q_k\right) - \Pr\left(p_i \le q_k\right)\right)\right| \\
& + \left|\sum_{i\in I_0}\sum_{k=1}^{m}\frac{1}{k}\left(\Pr\left(\widehat{\Lambda}^k_{(i),m_0}|\widehat{p}_{M,i} \le q_k\right) - \Pr\left(\Lambda^k_{(i),m_0}|p_i \le q_k\right)\right)\Pr\left(p_i \le q_k\right)\right| \\
= & \left|\sum_{i\in I_0}\sum_{k=1}^{m}\frac{1}{k}\Pr\left(\widehat{\Lambda}^k_{(i),m_0}|\widehat{p}_{M,i} \le q_k\right)\frac{k}{m}\frac{m}{k}\left(\Pr\left(\widehat{p}_{M,i} \le q_k\right) - \Pr\left(p_i \le q_k\right)\right)\right| \\
& + \left|\sum_{i\in I_0}\sum_{k=1}^{m}\frac{1}{k}\left(\Pr\left(\widehat{\Lambda}^k_{(i),m_0}|\widehat{p}_{M,i} \le q_k\right) - \Pr\left(\Lambda^k_{(i),m_0}|p_i \le q_k\right)\right)\frac{k}{m}\right| \\
= & \left|\sum_{i\in I_0}\sum_{k=1}^{m}\frac{1}{k}\Pr\left(\widehat{\Lambda}^k_{(i),m_0}\right)\frac{k}{m}\sup_{1\le k\le m}\sup_{i\in I_0}\left|\frac{m}{k}\Pr\left(\widehat{p}_{M,i} \le q_k\right) - \frac{m}{k}\Pr\left(p_i \le q_k\right)\right|\right| \\
& + \left|\sum_{i\in I_0}\sum_{k=1}^{m}\frac{1}{k}\left(\Pr\left(\widehat{\Lambda}^k_{(i),m_0}\right) - \Pr\left(\Lambda^k_{(i),m_0}\right)\right)\frac{k}{m}\gamma\right| \\
\le & \left|\sum_{i\in I_0}\sum_{k=1}^{m}\Pr\left(\widehat{\Lambda}^k_{(i),m_0}\right)\frac{1}{m}\times O\left(\frac{1}{M^\delta}\right)\right| + \left|\frac{\gamma}{m}\sum_{i\in I_0}\sum_{k=1}^{m}\left(\Pr\left(\widehat{\Lambda}^k_{(i),m_0}\right) - \Pr\left(\Lambda^k_{(i),m_0}\right)\right)\right| \\
= & \frac{m_0}{m}O\left(\frac{1}{M^\delta}\right),
\end{aligned}$$

since $\sum_{k=1}^{m}\left(\Pr\left(\widehat{\Lambda}^k_{(i),m_0}\right) - \Pr\left(\Lambda^k_{(i),m_0}\right)\right) = 0$. So

$$\begin{aligned}
& \left|\mathbb{E}_{\widehat{\mathbf{p}}_M}\left(\frac{V}{R}\right) - \mathbb{E}\left(\frac{V}{R}\right)\right| \\
= & \left|\sum_{m_0=0}^{m}\left(\mathbb{E}_{\widehat{\mathbf{p}}_M}\left(\frac{V}{R} \mid \widetilde{m}_0 = m_0\right) - \mathbb{E}\left(\frac{V}{R} \mid \widetilde{m}_0 = m_0\right)\right)\times\Pr\left(\widetilde{m}_0 = m_0\right)\right| \\
\le & \frac{\mathbb{E}\left(\widetilde{m}_0\right)}{m}O\left(\frac{1}{M^\delta}\right) = o\left(1\right).
\end{aligned}$$

Finally, if $T_1, \ldots, T_m$ are mutually independent, their joint distribution is PRDS on the subset of p-values corresponding to true null hypotheses. Thus the conclusion follows. ■

3.7.3 Proof of Proposition 1

Similar as in Kosorok and Ma (2007), we apply Orlicz norm to prove the proposition. The Orlicz norm $\|U\|_{\psi}$is defined by

$$\|U\|_{\psi} = \inf\left\{c_3 > 0 : \mathbb{E}\left(\psi\left(\frac{|U|}{c_3}\right)\right) \le 1\right\},$$

where ψ is a non-decreasing, convex function with $\psi(0) = 0$. As suggested by van der Vaart and Wellner (1996),

$$\psi_p(u) = \exp(u^p) - 1,$$

is particular useful in proving consistency. The corresponding Orlicz norm of $\psi_p(u)$ is called an exponential Orlicz norm. For all nonnegative u, $u^p \le \psi_p(u)$, which implies that

$$\|U\|_p \le \|U\|_{\psi_p}$$

for each p. Thus the L_p-norm is bounded by $\|U\|_{\psi_p}$.

Proof. Let $M^{\delta}\left(\widehat{T}_{M,i} - T_i\right) = U_{i,M}$. With $\psi_p(u) = \exp(u^p) - 1$ and $\psi_p^{-1}(m) = (\log(1+m))^{\frac{1}{p}}$, the proof directly follows from lemma 2.2.1 and 2.2.2 in van der Vaart and Wellner (1996). Given m_0 true null hypotheses, as $M \ge M_0$

$$\begin{aligned}\left\|\max_{i\in I_0}|U_{i,M}|\right\|_p &\le \left\|\max_{i\in I_0}|U_{i,M}|\right\|_{\psi_p} \\ &\le c_5(\log(1+m_0))^{\frac{1}{p}}\max_{i\in I_0}\|U_{i,M}\|_{\psi_p} \\ &\le c_5(\log(1+m))^{\frac{1}{p}}\left(\frac{1+c_1}{c_2}\right)^{\frac{1}{p}} \\ &\le 2c_5(\log(m))^{\frac{1}{p}}\left(\frac{1+c_1}{c_2}\right)^{\frac{1}{p}},\end{aligned}$$

by $\log(1+m) \le 2\log m$. Thus

$$\left\|\max_{i\in I_0}\left|T_i - \widehat{T}_{M,i}\right|\right\|_p = \left\|\max_{i\in I_0}\frac{|U_i|}{M^{\delta}}\right\|_p \le c_6\frac{(\log(m))^{\frac{1}{p}}}{M^{\delta}},$$

where $c_6 = 2c_5\left(\frac{1+c_1}{c_2}\right)^{\frac{1}{p}} < \infty$. Therefore if $\frac{(\log(m))^{\frac{1}{p}}}{M^{\delta}} = o(1)$ as $M \to \infty$, we can conclude that $\widehat{T}_{M,i} \xrightarrow{P} T_i$ for all $i \in I_0$, and $\sup_{1\le k\le m}\sup_{i\in I_0}|\Pr(\widehat{p}_{M,i} \le q_k) - \Pr(p_i \le q_k)| = o(1)$ since convergence in probability implies convergence in law. ■

3.8 More discussions on the PRDS condition

PRDS is a special case of positive regression dependent. Lehmann (1966) defined a random variable Y positive regression dependent on a random variable X as

$$\Pr\left(Y \leq y \mid X = x\right) \text{ is non-increasing in x,} \tag{3.12}$$

while Y is negative regression dependent on X if $\Pr\left(Y \leq y \mid X = x\right)$ is non-decreasing in x. Y positive (negative) regression dependent on X is also called stochastic monotonicity of $\Pr\left(Y \leq y \mid X = x\right)$.

Y positive regression dependent on X also implies that

$$\Pr\left(Y \leq y \mid X \leq x\right) \geq \Pr\left(Y \leq y \mid X \leq x'\right), \tag{3.13}$$

for all $x \leq x'$ and

$$\Pr\left(Y \leq y, X \leq x\right) \geq \Pr\left(Y \leq y\right)\Pr\left(X \leq x\right). \tag{3.14}$$

(3.14) is called X and Y are positively quadrant dependent. It says that the more possibility of X being small (large), the more possibility of Y also being small (large). If we let $x' \to \infty$, then (3.13) becomes (3.14). With simple algebra, it can be shown that (3.12) implies (3.13), and (3.13) implies (3.14). All of the three conditions can be extended to multiple variables. Positive regression dependent of an $l-$dimensional random vector $\mathbf{Y}$ on a $m-$dimensional random vector $\mathbf{X}$ is that

$$\Pr\left(Y_1 \leq y_1, \ldots, Y_l \leq y_l \mid X_1 = x_1, \ldots, X_m = x_m\right) \tag{3.15}$$

is non-increasing in $x_1, \ldots x_m$. Obviously $\mathbf{Y}$ is PRDS on a subset I_0 of $\mathbf{X}$ is less stringent than (3.15).

Another frequently used but more restricted criteria for dependency of multivariate random variables is the multivariate totally positive of order 2 (MTP_2). Karlin and Rinott (1981) defined a $m-$dimensional random vector $\mathbf{X}$ to have an MTP_2 distribution if the corresponding joint density $f_{\mathbf{X}}$ satisfies

$$f_{\mathbf{X}}\left(\mathbf{y} \vee \mathbf{z}\right) f_{\mathbf{X}}\left(\mathbf{y} \wedge \mathbf{z}\right) \geq f_{\mathbf{X}}\left(\mathbf{y}\right) f_{\mathbf{X}}\left(\mathbf{z}\right),$$

where

$$\begin{aligned} \mathbf{y} &= \left(y_1, \ldots, y_m\right), \ \mathbf{z} = \left(z_1, \ldots, z_m\right), \\ \mathbf{y} \vee \mathbf{z} &= \left(\max\left(y_1, z_1\right), \ldots, \max\left(y_m, z_m\right)\right), \\ \mathbf{y} \wedge \mathbf{z} &= \left(\min\left(y_1, z_1\right), \ldots, \min\left(y_m, z_m\right)\right). \end{aligned}$$

The number of dimension m can be extended to infinity or even continuous. MTP_2 implies positive regression dependent, and therefore implies PRDS (Sarkar, 2002). It can be shown that joint density of m random variables X_i satisfying MTP_2 implies $Cov\,(X_i, X_j) \geq 0$ for $i, j = 1, \ldots, m$. Nevertheless, except multivariate normal, PRDS and $Cov\,(X_i, X_j) \geq 0$ may not imply each other (Benjamini and Yekutieli, 2001). In a more general situation, empirically verifying whether data structure satisfies the above conditions may be difficult. But some solutions have been suggested, for example, a nonparametric test for stochastic monotonicity proposed by Lee et al. (2009).

3.9 A Simulation study with SV1FJ

For an additional simulation study, we use the following stochastic volatility with one jump component model (SV1FJ), which also was considered in Huang and Tauchen (2005),

$$\begin{aligned} d \log P\,(t) &= \mu dt + \exp\left(\beta_0 + \beta_1 \sigma\,(t)\right) dW_1\,(t) + dJ\,(t)\,, \\ d\sigma\,(t) &= a\sigma\,(t)\,dt + dW_2\,(t)\,, \\ J\,(t) &= \sum_{j=1}^{N(t)} D\,(t,j)\,,\ D\,(t,j) \overset{iid}{\sim} \mathcal{N}\,(0,1)\,, \\ & \quad N\,(t) \overset{iid}{\sim} \text{Poisson}\,(\lambda dt)\,, \end{aligned}$$

where $dW_1\,(t)$ and $dW_2\,(t)$ follow the standard Brownian motion, and $\sigma^2\,(t)$ follows a simple stochastic process. $J\,(t)$ follows a Compound Poisson Process (CPP) with a constant intensity λdt, and $N\,(t)$ is the number of jumps occurring within the small interval $(t - \triangle t, t]$.

For the simulation, we set the parameter to the following values.

$$\mu = 0.03,\ \beta_0 = 0,\ \beta_1 = 0.125,\ \text{and } a = -0.1.$$

In addition, we also add the leverage effect into the model, and the correlation between $dW_1\,(t)$ and $dW_2\,(t)$ is set to -0.62.

All of the other settings for the simulation are the same as in the SVJ case. Relevant results are shown in Figure 3.8 to Figure 3.12. It can be seen that all the results are qualitatively similar to those of the SVJ case.

3.10 Data descriptions

The raw data used for the empirical application is the 1-min recorded prices of S&P500 in cash (SPC500) and Dow Jones Industrial Average (DJIA) from January 2003 to December 2007. The data sets are provided by Olsen Financial Technologies in Zūrich, Switzerland. During this period, market closed at 1 pm on a few days. Such days were inactive trading days, and we do not include them in our samples. After eliminating these inactive trading days, we have

1247 active trading days for both indices. In section 5, all estimated realized price variations and test statistics are based on the data from the 1247 active trading days.

To estimate the intradaily price variations, we use 5-min log returns but exclude overnight returns. Some issues of microstructure noise are also concerned here. When observed prices contain microstructure noise, realized variations estimated with different sampling frequencies will have different degrees of biasness. Since the two indices are not really traded, their price series would be less likely to suffer distortions from the microstructure noise than those of traded futures. The property of immunizing the microstructure noise can be seen in 3.13, which shows volatility signature plots. The horizontal dashed line in each plot is the average daily realized variance when the 5-min log returns are used. It can be seen that the average values of the realized variances are downward biased when their sampling intervals are small. As the sampling interval becomes moderately large, the average values become stable, and the biasness is mitigated. However, the downward biasness reappears when the sampling interval increases beyond one hour. From the figure, we can see that the realized variances estimated from the 5-min log return data seem to suffer little microstructural effect. This is the reason why the 5-min log return data is used to construct the realized variance estimations.

We then calculate the three different jump test statistics $Z_{-1.5,i}$, $Z_{\log,i}$ and $Z_{ratio,i}$ and their corresponding p-values. To avoid effects of abnormal trades, we omit data of the first five minutes (09:31-09:35) and the last ten minutes (16:01-16:10), so the number of samples for each day equals to 77. This additional step of screening the data makes our estimates reflect intradaily dynamics of the two indices more homogeneously and efficiently. Note that the additional screening step only applies to JV_i and the daily jump test statistics. For RV_i and BV_i, we still keep the 80 samples each day. Figure 3.14 shows time series plots of RV, BV and $JV_{i,0.05}$ for the two indices. It can be seen that the log type statistic have most identified jump days.

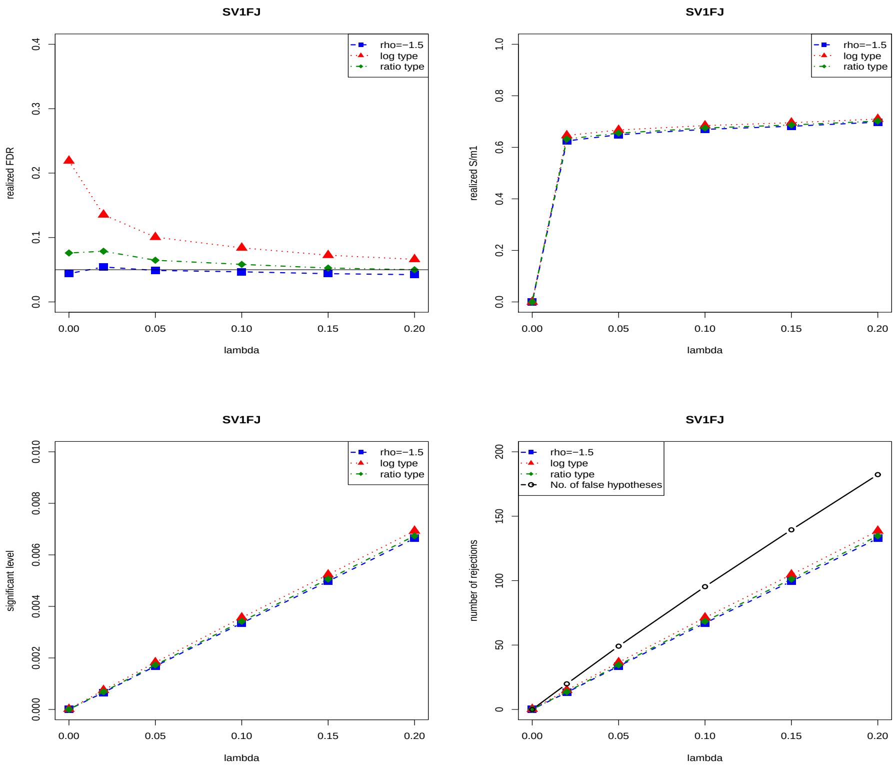

Figure 3.8: Realized FDR, $\hat{S}/m_1$, significance level obtained from the BH procedure and number of rejections. In the graphs, each point is an average value from 1000 simulations.

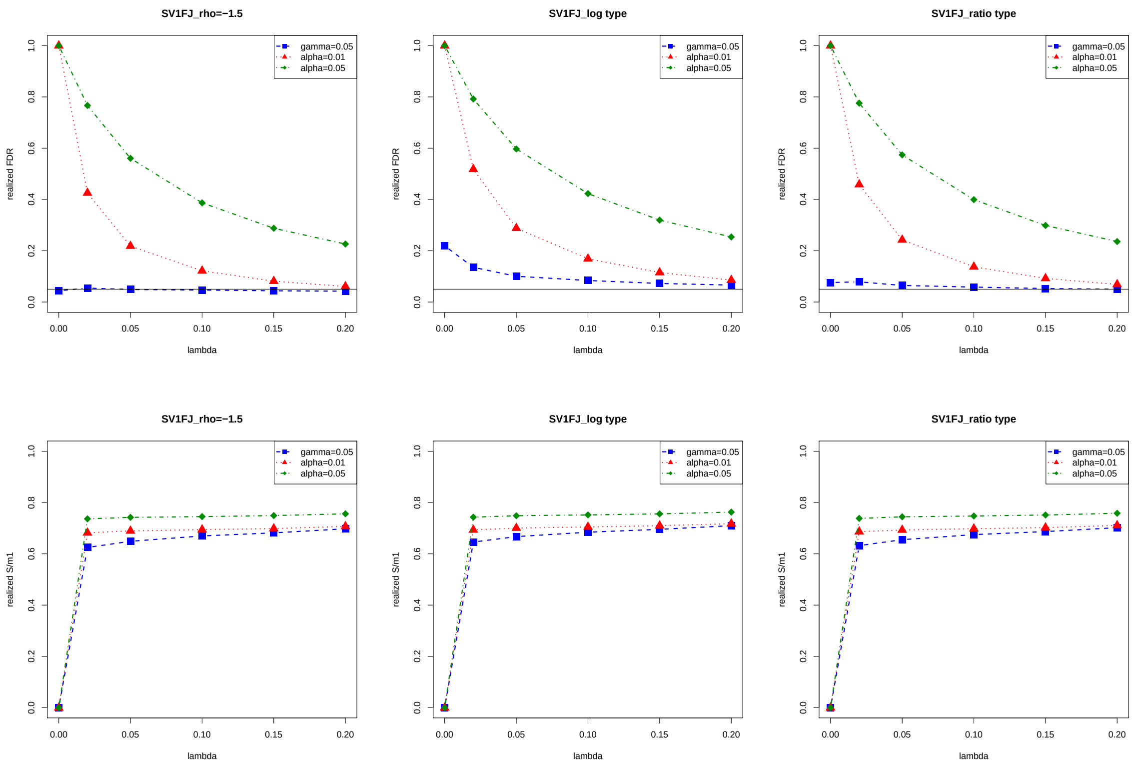

Figure 3.9: Realized FDR and $\hat{S}/m_1$ of the hybrid method and the conventional procedure. In the graphs, each point is an average value from 1000 simulations.

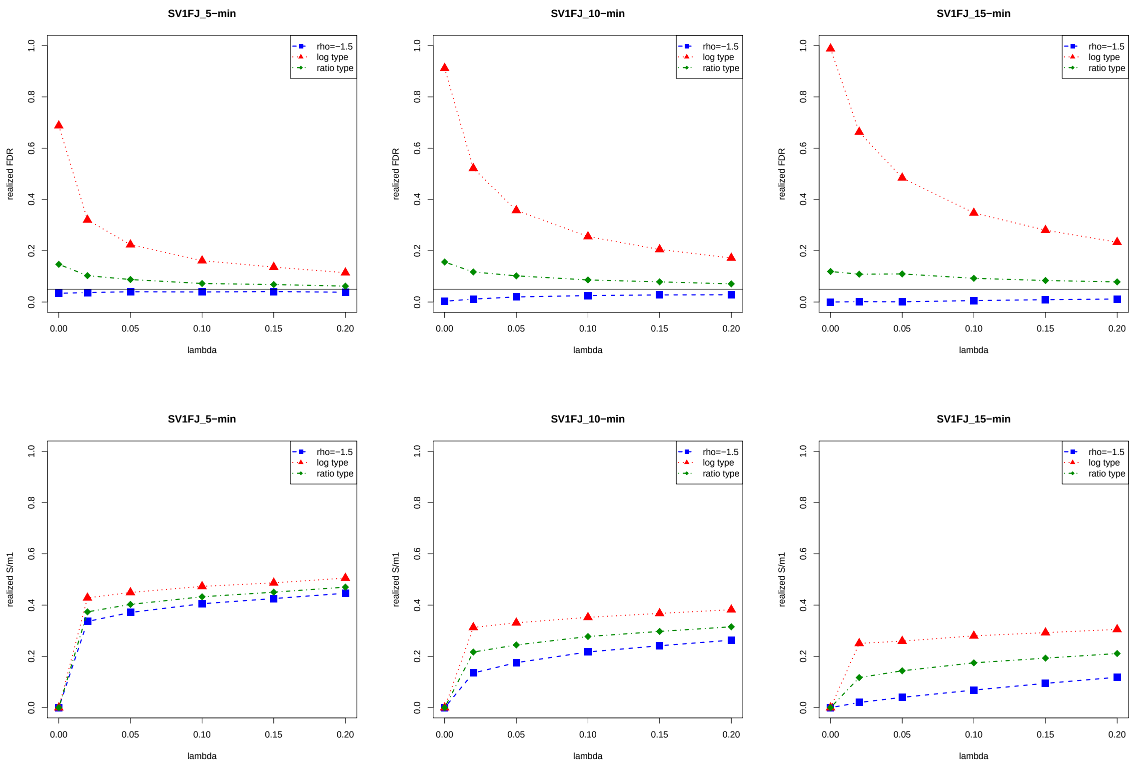

Figure 3.10: Realized FDR and $\hat{S}/m_1$ of the hybrid method with lower frequency data. In the graphs, each point is an average value from 1000 simulations.

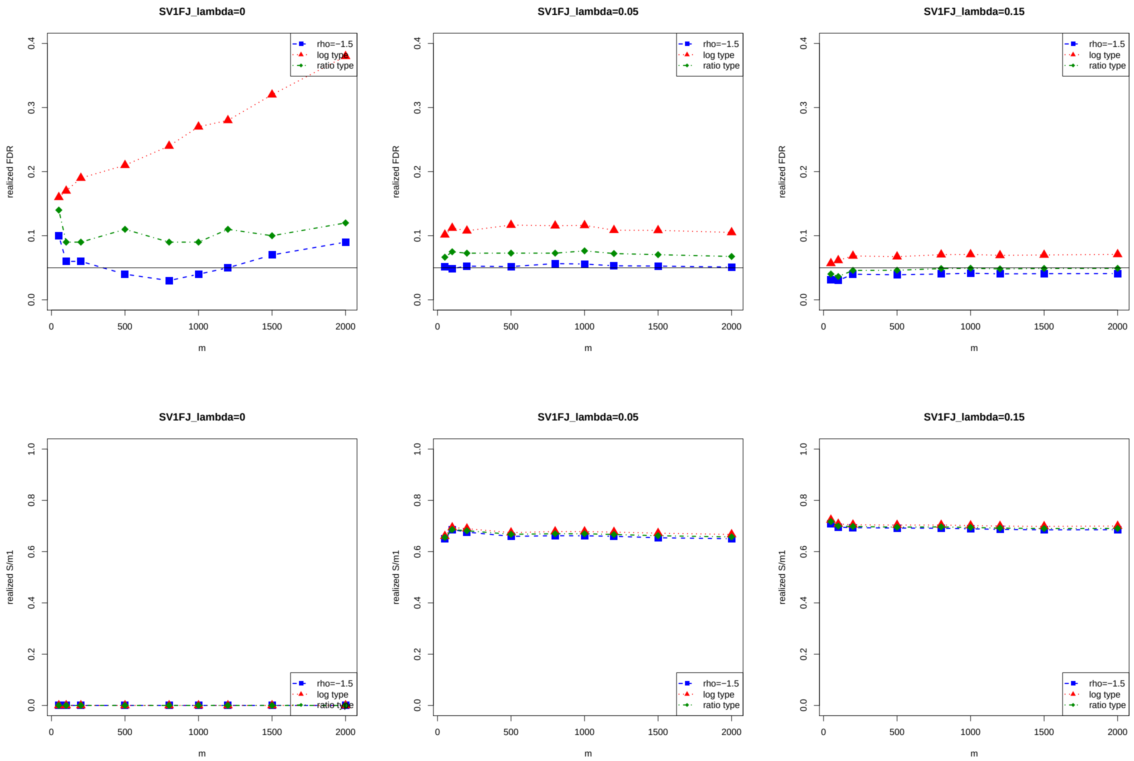

Figure 3.11: Realized FDR and $\hat{S}/m_1$ of the hybrid method when the number of hypotheses varies. Here $m = 50$, 100, 200, 500, 800, 1000, 1200, 1500 and 2000. In the graphs, each point is an average value from 1000 simulations.

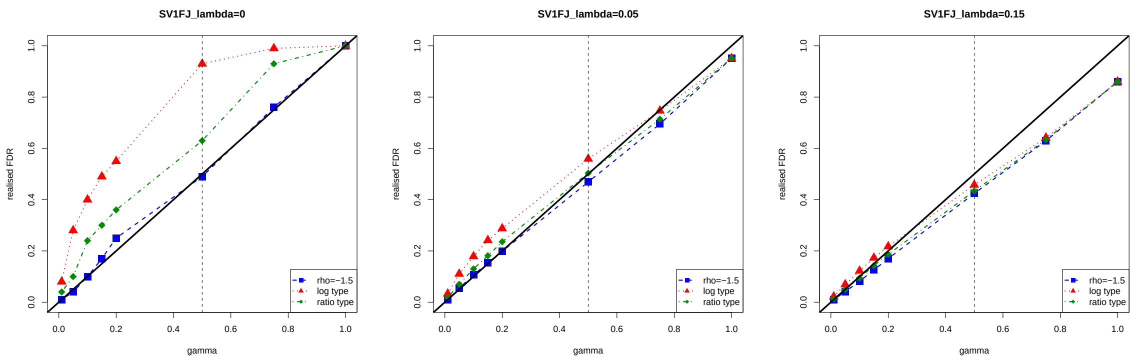

Figure 3.12: Realized FDR of the hybrid method under different required γ. We fix $m = 1000$ in the simulation. In the graphs, each point is an average value from 1000 simulations.

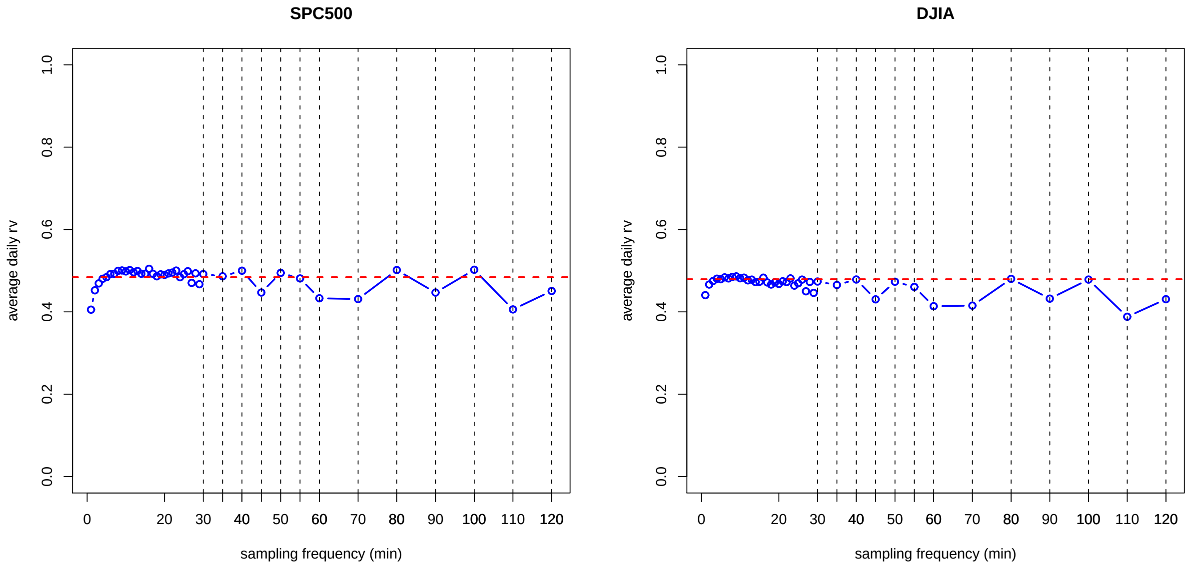

Figure 3.13: Volatility signature plots for the SPC500 and DJIA. The red line in each graph is the average of daily realized variations when sampling frequency is 5 minute.

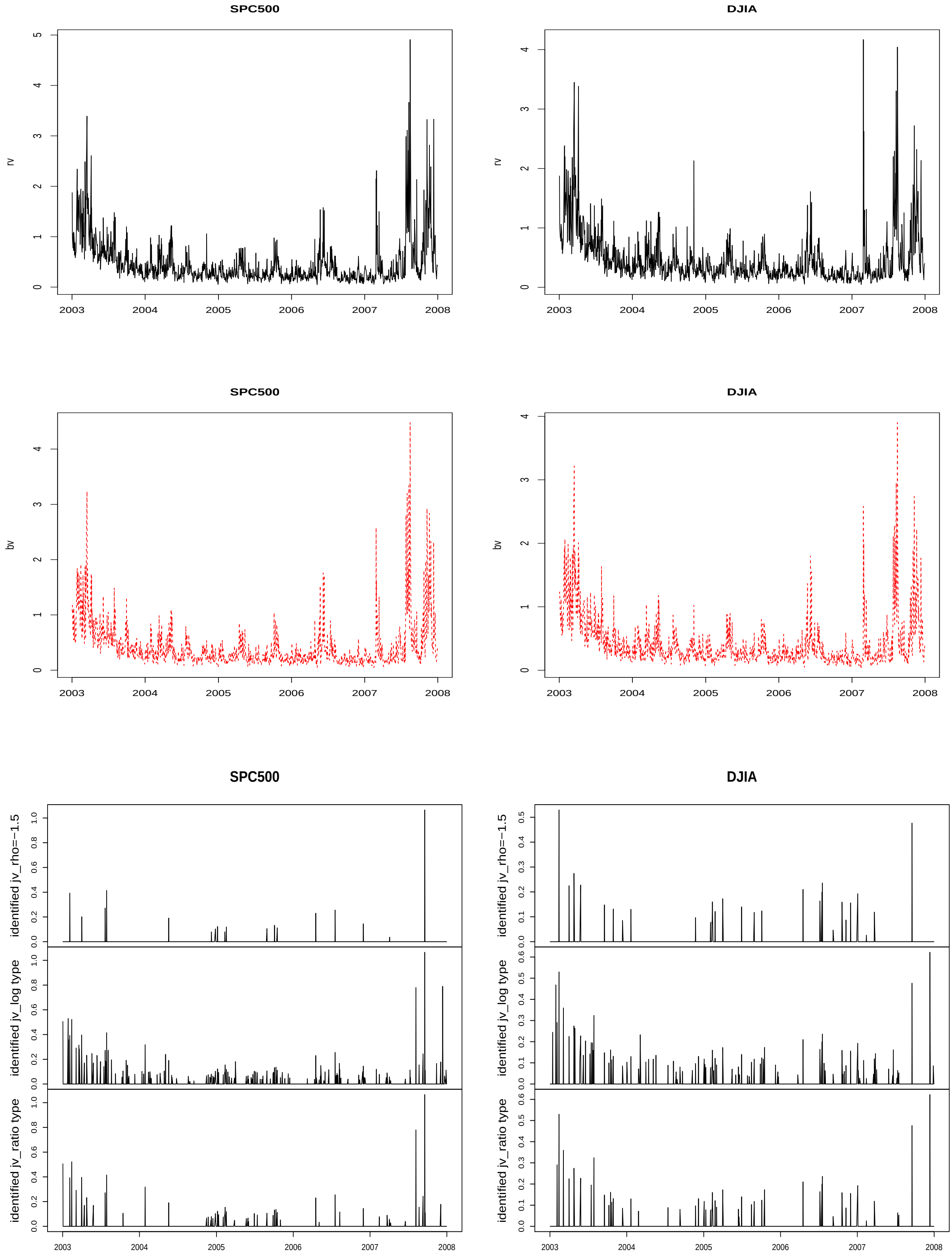

Figure 3.14: Time series plots for 5-min realized variance, realized bi-power variation and identified jump variation with the three different jump test statistics. The quantities shown here are all scaled by 10000.

Bibliography

Abaluck, J. and J. Gruber (2011): "Choice Inconsistencies among the Elderly: Evidence from Plan Choice in the Medicare Part D Program," *American Economic Review*, 101, 1180–1210.

Ait-Sahalia, Y. and J. Jacod (2009): "Testing for jumps in a discretely observed process," *The Annals of Statistics*, 37, 184–222.

Almeida, C., J. Graveline, and S. Joslin (2010): "Do Options Contain Information About Excess Bond Returns?" *forthcoming, Journal of Econometrics.*

Andersen, T. G. and L. Benzoni (2010): "Do Bonds Span Volatility Risk in the U.S. Treasury Market? A Specification Test for Affine Term Structure Models," *Journal of Finance*, 65, p. 603 – 653.

Andersen, T. G., T. Bollerslev, and F. X. Diebold (2007a): "Roughing it Up: Including Jump Components in the Measurement, Modeling, and Forecasting of Return Volatility," *The Review of Economics and Statistics*, 89, p. 701 – 720.

——— (2007b): "Roughing It Up: Including Jump Components in the Measurement, Modeling, and Forecasting of Return Volatility," *The Review of Economics and Statistics*, 89, 701–720.

Ang, A., R. J. Hodrick, Y. Xing, and X. Zhang (2006): "The Cross-Section of Volatility and Expected Returns," *Journal of Finance*, 61, p. 259 – 299.

Baele, L., G. Bekaert, and K. Inghelbrecht (2010): "The Determinants of Stock and Bond Return Comovements," *Review of Financial Studies*, 23, p. 2374 – 2428.

Bai, J. and S. Ng (2008): "Forecasting economic time series using targeted predictors," *Journal of Econometrics*, 146, 304–317.

Bai, Z. and J. W. Silverstein (2010): *Spectral Analysis of Large Dimensional Random Matrices*, Springer, 2nd ed.

Bakshi, G. and N. Kapadia (2003): "Volatility Risk Premiums Embedded in Individual Equity Options: Some New Insights," *Journal of Derivatives*, Fall, p. 45 – 54.

Barndorff-Nielsen, O. E. and N. Shephard (2006): "Econometrics of Testing for Jumps in Financial Economics Using Bipower Variation," *Journal of Financial Econometrics*, 4, 1–30.

Bekaert, G., M. Hoerova, and M. Lo Duca (2011): "Risk, Uncertainty and Monetary Policy," Working Paper, Columbia University.

Benjamini, Y. and Y. Hochberg (1995): "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 57, 289–300.

Benjamini, Y. and D. Yekutieli (2001): "The Control of the False Discovery Rate in Multiple Testing Under Dependency," *The Annals of Statistics*, 29, 1165–1188.

Beshears, J., J. J. Choi, D. Laibson, and B. C. Madrian (2006): "Simplification and Saving," NBER Working Papers 12659, National Bureau of Economic Research, Inc.

Bibkov, R. and M. Chernov (2009): "Unspanned Stochastic Volatility in Affine Models: Evidence from Eurodollar Futures and Options," *Management Science*, 55, p. 1292 – 1305.

BIS (2010): "Triennial Central Bank Survey of Foreign Exchange and Derivatives Market Activity in April 2010," Tech. rep., Bank for International Settlements.

Black, F. (1976): "The Pricing of Commodity Contracts," *Journal of Financial Economics*, 3, p. 16 –179.

Black, F. and M. Scholes (1973): "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy*, 81, p. 637–659.

Bollerslev, T., M. Gibson, and H. Zhou (2011): "Dynamic Estimation of Volatility Risk Premia and Investor Risk Aversion from Option-Implied and Realized Volatilities," *Journal of Econometrics*, 160, p. 235 – 245.

Bollerslev, T., T. H. Law, and G. E. Tauchen (2008): "Risk, Jumps, and Diversification," *Journal of Econometrics*, 144, 234–256.

Bollerslev, T., N. Sizova, and G. Tauchen (2012): "Volatility in Equilibrium: Assymmetries and Dynamic Dependencies," *Review of Finance*, 16, p. 31 – 80.

Bollerslev, T., G. Tauchen, and H. Zhou (2010): "Expected Stock Returns and Variance Risk Premia," *Review of Financial Studies*, 22, p. 4463 – 4492.

Brandt, M. W., P. Santa-Clara, and R. Valkanov (2009): "Covariance Regularization By Parametric Portfolio Policies: Exploiting Characteristics in the Cross-Section of Equity Returns," *The Review of Financial Studies*, 22, 3411–3447.

Britten-Jones, M. and A. Neuberger (2000): "Option Prices, Implied Price Processes, and Stochastic Volatility," *Journal of Finance*, 55, 839–866.

Brodie, J., I. Daubechies, C. D. Mol, D. Giannone, and I. Loris (2009): "Sparse and stable Markowitz portfolios," *Proceedings of the National Academy of Sciences of the United States of America*, 106, 12267–12272.

Brunnermeier, M. K. and L. H. Pedersen (2009): "Market Liquidity and Funding Liquidity," *Review of Financial Studies*, 22, p. 2201 – 2238.

Buhlmann, P. and S. van de Geer (2011): *Statistics for High-Dimensional Data: Methods, Theory and Applications*, Springer.

Busch, T., B. J. Christensen, and M. Ø. Nielsen (2011): "The Role of Implied Volatility in Forecasting Future Realized Volatility and Jumps in Foreign Exchange, Stock, and Bond Markets," *Journal of Econometrics*, 160, 48–57.

Caballero, R. and A. Krishnamurthy (2008): "Collective Risk Management in a Flight to Quality Episode," *Journal of Finance*, 63, p. 2195 – 2236.

Campbell, J. Y. and S. B. Thompson (2008): "Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average?" *Review of Financial Studies*, 21, p. 1509 – 1531.

Carr, P. and R. Lee (2009): "Volatility Derivatives," *Annual Review of Financial Economics*, 1, p. 1 – 21.

Carr, P. and L. Wu (2009): "Variance Risk Premiums," *Review of Financial Studies*, 22, 1311–1341.

Casassus, J., P. Collin-Dufresne, and R. S. Goldstein (2005): "Unspanned Stochastic Volatility and Fixed Income Derivative Pricing," *Journal of Banking and Finance*, 29, p. 2723 – 2749.

Clark, T. E. and M. W. McCracken (2001): "Tests of Forecast Accuracy and Encompassing for Nested Models," *Journal of Econometrics*, 105, p. 85 – 110.

Clarke, S. and P. Hall (2009): "Robustness of Multiple Testing Procedures Against Dependence," *The Annals of Statistics*, 37, 332–358.

Cochrane, J. H. (2011): "Presidential Address: Discount Rates," *Journal of Finance*, 66, 1047–1108.

Cochrane, J. H. and M. Piazzesi (2005): "Bond Risk Premia," *American Economic Review*, 95, p. 138 – 160.

COHEN, R., C. POLK, AND B. SILLI (2009): “Best Ideas,” Tech. rep., London School of Economics and Political Science.

COLLIN-DUFRESNE, P. AND R. S. GOLDSTEIN (2002): “Do Bonds Span the Fixed-Income Markets?” *Journal of Finance*, 57, p. 1685 – 1730.

COLLIN-DUFRESNE, P., R. S. GOLDSTEIN, AND C. S. JONES (2008): “Identification of Maximal Affine Term Structure Models,” *Journal of Finance*, 63, p. 743 – 795.

COLLIN-DUFRESNE, P., R. S. GOLDSTEIN, AND J. S. MARTIN (2001): “The Determinants of Credit Spread Changes,” *Journal of Finance*, 56, p. 2177 – 2207.

CORSI, F. (2009): “A Simple Approximate Long-Memory Model of Realized Volatility,” *Journal of Financial Econometrics*, 7, p. 174 – 196.

CORSI, F., D. PIRINO, AND R. RENO (2010): “Threshold Bipower Variation and the Impact of Jumps on Volatility Forecasting,” *Journal of Econometrics*, 159, p. 276 – 288.

CREMERS, M., M. HALLING, AND D. WEINBAUM (2010): “In Search of Aggregate Jump and Volatility Risk in the Cross-Section of Stock Returns,” Working Paper, Yale University.

DE MOL, C., D. GIANNONE, AND L. REICHLIN (2008): “Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components?” *Journal of Econometrics*, 146, 318–328.

DEMETERFI, K., E. DERMAN, M. KAMAL, AND J. ZHOU (1999): “A Guide to Volatility and Variance Swaps,” *Journal of Derivatives*, 6, p. 9 – 32.

DEMIGUEL, V., L. GARLAPPI, F. J. NOGALES, AND R. UPPAL (2009a): “A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms,” *Management Science*, 55, 798–812.

DEMIGUEL, V., L. GARLAPPI, AND R. UPPAL (2009b): “Optimal versus naive diversification: How inefficient is the *1/N* portfolio strategy?” *Review of Financial Studies*, 22, 1915–1953.

DEMIGUEL, V., F. J. NOGALES, AND R. UPPAL (2010): “Stock Return Serial Dependence and Out-of-Sample Portfolio Performance,” *SSRN eLibrary*.

DRECHSLER, I. (2010): “Uncertainty, Time-Varying Fear, and Asset Prices,” Working Paper, New York University.

DRECHSLER, I. AND A. YARON (2011): “What’s Vol Got to Do With It,” *Review of Financial Studies*, 24, p. 1 – 45.

DRIESSEN, J., P. MAENHOUT, AND G. VILKOV (2009): “The Price of Correlation Risk: Evidence from Equity Options,” *Journal of Finance*, 64, p. 1377 – 1406.

Duarte, J. (2008): "The Causal Effect of Mortgage Refinancing on Interest Rate Volatility: Empirical Evidence and Theoretical Implications," *Review of Financial Studies*, 21, p. 1698 – 1731.

Efron, B., T. Hastie, I. Johnstone, and R. Tibshirani (2004): "Least angle regression," *The Annals of Statistics*, 32, 407–451.

El Karoui, N. (2009): "On the realized risk of high-dimensional Markowitz portfolios," Technical Report 784, Department of Statistics, UC Berkeley.

——— (2010): "High-dimensionality effects in the Markowitz problem and other quadratic programs with linear constraints: Risk underestimation," *The Annals of Statistics*, 38, 3487–3566.

Fan, J., P. Hall, and Q. Yao (2007): "To How Many Simultaneous Hypothesis Tests Can Normal, Student's t or Bootstrap Calibration Be Applied?" *Journal of the American Statistical Association*, 102, 1282–1288.

Fan, J., J. Zhang, and K. Yu (2009): "Asset allocation and risk assessment with gross exposure constraints for vast portfolios," Tech. rep., Princeton University.

Fan, Y. and J. Fan (2008): "Testing and Detecting Jumps Based on a Discretely Observed Process," *SSRN eLibrary*.

Feunou, B., J.-S. Fontaine, A. Taamouti, and R. Tédongap (2011): "The Equity Premium and the Maturity Structure of Uncertainty," Working Paper, Stockholm School of Economics.

Fontaine, J.-S. and R. Garcia (2011): "Bond Liquidity Premia," *forthcoming, Review of Financial Studies*.

Forsythe, G. E., M. A. Malcolm, and C. B. Moler (1977): *Computer Methods for Mathematical Computations*, Prentice-Hall.

Frahm, G. and M. Christoph (2011): "Dominating estimators for minimum-variance portfolios," *Journal of Econometrics*, forthcoming.

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

Gabaix, X. (2011): "A Sparsity-based model of bounded rationality," *NBER Working Paper 16911*.

Garleanu, N. and L. H. Pedersen (2011): "Margin-based Asset Pricing and Deviations from the Law of One Price," *Review of Financial Studies*, 24, 1980–2022.

Goncalves, S. and N. Meddahi (2009): "Bootstrapping Realized Volatility," *Econometrica*, 77, 283–306.

——— (2011): "Box-Cox transforms for realized volatility," *Journal of Econometrics*, 160, 129–144.

Goyal, A. and I. Welch (2008): "A Comprehensive Look at The Empirical Performance of Equity Premium Prediction," *Review of Financial Studies*, 21, p. 1455 – 1508.

Gromb, D. and D. Vayanos (2002): "Equilibrium and Welfare in Markets with Financially Constrained Arbitrageurs," *Journal of Financial Economics*, 66, p. 361–407.

Hall, P. (1992): *The Bootstrap and Edgeworth Expansion*, New York: Springer-Verlag.

Hastie, T., R. Tibshirani, and J. Friedman (2009): *The Elements of Statistical Learning*, Springer, 2nd ed.

Heidari, M. and L. Wu (2003): "Are Interest Rate Derivatives Spanned by the Term Structure of Interest Rates?" *Journal of Fixed Income*, 13, p. 75 – 86.

Huang, X. and G. E. Tauchen (2005): "The Relative Contribution of Jumps to Total Price Variance," *Journal of Financial Econometrics*, 3, 456–499.

Iyengar, S. S. and E. Kamenica (2010): "Choice proliferation, simplicity seeking, and asset allocation," *Journal of Public Economics*, 94, 530–539.

Jacod, J. and V. Todorov (2009): "Testing for Common Arrival of Jumps for Discretely Observed Multidimensional Processes," *The Annals of Statistics*, forthcoming.

Jagannathan, R. and T. Ma (2003): "Risk reduction in large portfolios: Why imposing the wrong constraints helps," *Journal of Finance*, 58, 1651–1684.

Jiang, G. and Y. Tian (2005): "Model-Free Implied Volatility and Its Information Content," *Review of Financial Studies*, 18, p. 1305 – 1342.

Jiang, G. J. and R. C. Oomen (2008): "Testing for Jumps When Asset Prices are Observed with Noise - A Swap Variance Approach," *Journal of Econometrics*, 144, 352–370.

Jorion, P. (1986): "Bayes-Stein estimation for portfolio analysis," *Journal of Financial and Quantitative Analysis*, 21, 279–292.

——— (1995): "Predicting Volatility in the Foreign Exchange Market," *Journal of Finance*, 50, p. 507 – 528.

Joslin, S. (2010): "Pricing and Hedging in Fixed Income Markets," Working Paper, University of Southern California.

Jouini, E. and C. Napp (2007): "Consensus Consumer and Intertemporal Asset Pricing with Heterogeneous Beliefs," *Review of Economic Studies*, 74, p. 1149 – 1174.

Kan, R. and G. F. Zhou (2007): "Optimal portfolio choice with parameter uncertainty," *Journal of Financial and Quantitative Analysis*, 42, 621–656.

Karatzas, I. and S. E. Shreve (1991): *Brownian Motion and Stochastic Calculus*, New York: Springer, 2nd ed.

Karlin, S. and Y. Rinott (1981): "Total Positivity Properties of Absolute Value Multinormal Variables with Applications to Confidence Interval Estimates and Related Probabilistic Inequalities," *The Annals of Statistics*, 9, 1035–1049.

Kirby, C. and B. Ostdiek (2011): "Optimal Active Portfolio Management with Unconditional Mean-Variance Risk Preferences," *SSRN eLibrary*.

Korteweg, A. and N. Polson (2010): "Corporate Credit Spreads under Parameter Uncertainty," Working Paper, Stanford University.

Kosorok, M. R. and S. Ma (2007): "Marginal Asymptotics for the "Large p, Small n Paradigm" : With Applications to Microarray Data," *The Annals of Statistics*, 35, 1456–1486.

Lai, T. L., H. Xing, and Z. Chen (2011): "Mean-variance portfolio optimization when means and covariances are unknow," *The Annals of Applied Statistics*, forthcoming.

Ledoit, O. and M. Wolf (2003): "Improved estimation of the covariance matrix of stock returns with an application to portfolio selection," *Journal of Empirical Finance*, 10, 603–621.

Lee, S., O. Linton, and Y.-J. Whang (2009): "Testing for Stochastic Monotonicity," *Econometrica*, 77, 585–602.

Lee, S. S. and P. A. Mykland (2008): "Jumps in Financial Markets: A New Nonparametric Test and Jump Dynamics," *Review of Financial Studies*, 21, 2535–2563.

Lehmann, E. L. (1966): "Some Concepts of Dependence," *The Annals of Mathematical Statistics*, 37, 1137–1153.

Linton, O., E. Maasoumi, and Y.-J. Whang (2005): "Consistent Testing for Stochastic Dominance under General Sampling Schemes," *Review of Economic Studies*, 72, 735–765.

Litterman, R. and J. Scheinkman (1991): "Common Factors Affecting Bond Returns," *Journal of Fixed Income*, 1, p. 54 – 61.

Ludvigson, S. C. and S. Ng (2009): "Macro Factors in Bond Risk Premia," *Review of Financial Studies*, 22, p. 5027 – 5067.

——— (2011): *A Factor Analysis of Bond Risk Premia*, Chapman and Hall, chap. Handbook of Empirical Economics and Finance, p. 313 – 372.

Markowitz, H. (1952): "Portfolio selection," *The Journal of Finance*, 7, 77–91.

Martin, I. (2011): "Simple Variance Swaps," Working Paper, Stanford University.

Mueller, P., A. Vedolin, and H. Zhou (2011): "Short-Run Bond Risk Premia," Working Paper, London School of Economics.

Neuberger, A. (1994): "Hedging Volatility - the Case for a New Contract," *Journal of Portfolio Management*, 20, p. 74 – 80.

Newey, W. K. and K. D. West (1987): "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, p. 703 – 708.

Nieuwerburgh, S. V. and L. Veldkamp (2009): "Information Immobility and the Home Bias Puzzle," *Journal of Finance*, 64, 1187–1215.

Overdahl, J. A. (1988): "The Early Exercise of Options on Treasury Bond Futures," *Journal of Financial and Quantitative Analysis*, 23, p. 437 – 449.

Owen, A. B. (2007): "A robust hybrid of lasso and ridge regression," *Contemporary Mathematics*, 443, 59–72.

Patton, A. J. and A. Timmermann (2010): "Why Do Forecasters Disagree? Lessons from the Term-Structure of Cross-Sectional Dispersion," *Journal of Monetary Economics*, 57, p. 803 – 820.

Politis, D. N. and J. P. Romano (1994): "The Stationary Bootstrap," *Journal of the American Statistical Association*, 89, 1303–1313.

Romano, J. P., A. M. Shaikh, and M. Wolf (2008b): "Control of the False Discovery Rate under Dependence using the Bootstrap and Subsampling," *TEST*, 17, 417–442.

Saad, Y. (2003): *Iterative Methods for Sparse Linear Systems*, Society for Industrial and Applied Mathematics, 2nd ed.

Sarkar, S. K. (1998): "Some Probability Inequalities for Ordered MTP_2 Random Variables: A Proof of The Simes Conjecture," *The Annals of Statistics*, 26, 494–504.

——— (2002): "Some Results on False Discovery Rate in Stepwise Multiple Testing Procedures," *The Annals of Statistics*, 30, 239–257.

Stock, J. H. and M. W. Watson (2002): "Macroeconomic Forecasting unsing Diffusion Indexes," *Journal of Business and Economic Statistics*, 20, p. 147 – 162.

Storey, J. D. (2002): "A Direct Approach to False Discovery Rates," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64, 479–498.

Tibshirani, R. (1996): "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 58, 267–288.

Todorov, V. (2010): "Variance Risk Premium Dynamics: The Role of Jumps," *Review of Financial Studies*, 23, p. 345 – 383.

Trolle, A. B. and E. S. Schwartz (2009): "Unspanned Stochastic Volatility and the Pricing of Commodity Derivatives," *Review of Financial Studies*, 22, p. 4423 – 4461.

Tseng, P. (2001): "Convergence of a block coordinate descent method for nondifferentiable minimization," *Journal of Optimization Theory and Applications*, 109, 475–494.

Tu, J. and G. Zhou (2011): "Markowitz meets Talmud: A combination of sophisticated and naive diversification strategies," *Journal of Financial Economics*, 99, 204–215.

van der Vaart, A. and J. Wellner (1996): *Weak Convergence and Empirical Processes: With Applications to Statistics*, New York: Springer-Verlag.

Vedolin, A. (2010): "Uncertainty and Leveraged Lucas Trees: The Cross Section of Equilibrium Volatility Risk Premia," Working Paper, London School of Economics.

Veraart, A. E. (2010): "Inference For The Jump Part Of Quadratic Variation Of Ito Semimartingales," *Econometric Theory*, 26, 331–368.

Wang, H., H. Zhou, and Y. Zhou (2010): "Credit Default Swap Spreads and Variance Risk Premia," Working Paper.

Welsch, R. E. and X. Zhou (2007): "Application of robust statistics to asset allocation models," *Revstat*, 5, 97–114.

Xiong, W. and H. Yan (2010): "Heterogeneous Expectations and Bond Markets," *Review of Financial Studies*, 23, p. 1433 – 1466.

Yen, Y. and T. Yen (2011): "Solving Norm Constrained Portfolio Optimizations via Coordinate-Wise Descent Algorithms," *Working paper*.

Zou, H. and T. Hastie (2005): "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67, 301–320.