

©Copyright 2013

Xiaolin Sun

Essays on Applications of the Factor Model

Xiaolin Sun

A dissertation

submitted in partial fulfillment of the  
requirements for the degree of

Doctor of Philosophy

University of Washington

2013

Reading Committee:

Eric Zivot, Chair

Yu-chin Chen

Douglas Martin

Program Authorized to Offer Degree:

Economics

University of Washington

**Abstract**

Essays on Applications of the Factor Model

Xiaolin Sun

Chair of the Supervisory Committee:

Professor Eric Zivot

Economics

Estimating the volatilities and correlations of asset returns plays an important role in portfolio and risk management. As of late, interest in the estimation of the covariance matrix of large dimensional portfolios has increased. Estimating large dimensional covariance poses a challenge in that the cross-sectional dimension is often similar to or bigger than the number of observations available. Simple estimators are often poorly conditioned with some small eigenvalues, and so are unsuitable for many real world applications, including portfolio optimization and tracking error minimization.

The first chapter introduces our two large dimensional covariance matrix estimators. We estimate the large dimensional realized covariance matrix by using the methods of asymptotic principal components analysis based factor modeling and singular value decomposition.

In the second chapter, we show through simulation that our proposed estimators are closer to the true covariance matrix than the current popular shrinkage estimator. We also simulate conducting the out sample portfolio performance tests and find that the portfolios constructed based on our proposed estimators have lower risk than portfolios constructed using the shrinkage

matrix. Using S&P 500 stocks from 1926 to 2011, we back test our proposed covariance matrix. In addition, the portfolios constructed based on our proposed estimators exhibit lower risk than portfolios constructed using the shrinkage matrix.

The third chapter proposes a new volatility index—a cross-sectional volatility index of residuals using factor model. The cross-sectional volatility index moves closely with the VIX for the S&P 500 stock universe. It is a non-parametric, model-free volatility index, which could be estimated at any frequency for any region, sector, and style of world equity market and also does not depend on any option pricing. We provide some interpretation of the cross-sectional volatility index of residuals as a proxy for aggregate economic uncertainty, and show a high correlation between the VIX index and the corresponding cross-sectional volatility index of residuals based on the S&P 500 universe. Our results show that the portfolio hedged based on the cross-sectional volatility index of residuals has a much higher Sharpe ratio than the portfolio without hedge. Overall, these findings suggest that the cross-sectional volatility index of residuals can be used as a reliable proxy for volatility when volatility measures are not available.

## Table of Contents

List of Figures .....	iii
List of Tables .....	iv
Chapter 1 Estimation High Dimensional Covariance Matrix Using a Factor Model and Singular Value Decomposition.....	1
1.1 Introduction.....	1
1.2 Realized Covariance Matrix.....	5
1.3 Factor Models.....	10
1.4 Covariance Estimators.....	13
1.5 Conclusion.....	18
Chapter 2 Empirical Study of Estimating High Dimensional Covariance Matrix: Through Simulation and Backtesting.....	19
2.1 Literature on Comparing Covariance Estimators.....	20
2.2 Simulation.....	24
2.3 Empirical Studies.....	29
2.4 Conclusion.....	32
Chapter 3 A New Volatility Index using a Factor Model.....	38
3.1 Volatility Index Background.....	38
3.2 Cross-Sectional Volatility.....	43
3.3 Empirical Studies.....	47
3.4 Conclusion.....	51

Bibliography ..... 58

## List of Figures

Figure Number	Page
2.1 PRIAL of Three Covariance Estimators.....	34
2.2 Global Minimum Portfolio Variance .....	35
2.3 Mean Minimum Variance Portfolio with Annual Target Return 8%.....	35
2.4 PRIAL of Factor Model Based Covariance Estimators.....	36
2.5 PRIAL of Singular Value Decomposition Based Covariance Estimator.....	36
2.6 Global Minimum Portfolio Volatility for S&P 500 Stocks.....	37
2.7 Mean Minimum Variance Portfolio Volatility for S&P 500 Stocks.....	37
3.1 Time Series Volatility of the S&P 500 vs. the VIX, 1990-2011.....	54
3.2 Time Series Plot of the Return of the S&P 500 vs. the VIX, 1990-2011.....	55
3.3 Average Correlation, Average Volatility and the Cross-Sectional Volatility.....	56
3.4 Time Series Plot of the CSV vs. the VIX, 1990-2011.....	57

## **List of Tables**

Figure Number	Page
2.1 Relative Estimation Errors of Eigenvalues for Monthly Estimation.....	33
2.2 Relative Estimation Errors of Eigenvalues for Daily Estimation.....	33
3.1 Currently Available Volatility Indices .....	53
3.2 Daily Return and Volatility of the S&P 500 Index, VIX Portfolio and CSV Portfolio.....	53

## **Acknowledgements**

I would like to express my deepest gratitude to Chair of my dissertation committee, Eric Zivot, for his guidance and encouragement at various stages of writing this dissertation. Aside from providing advice on this dissertation, he has also shown me what it means to be a true great scholar. I am very fortunate to have him as my mentor.

I am indebted to my committee members, Yu-chin Chen and Douglas Martin, for their help and guidance during my dissertation research. I am grateful to their insightful comments and constructive suggestions on the dissertation drafts. I thank Ka Kit Tung for his services as Graduate School Representative.

Finally I would like to extend my most sincere and heartfelt thanks to my family for their love and support.

## **Dedication**

To my parents

## Chapter 1

### Estimating High Dimensional Covariance Matrix

#### Using a Factor Model and Singular Value Decomposition

Estimating and forecasting the volatilities and correlations of asset returns plays an important role in portfolio and risk management. As of late, interest in the estimation of the covariance matrix of large dimensional portfolios has increased. Estimating large dimensional covariance is a challenging problem since the cross-sectional dimension is often similar to or bigger than the number of observations available. Simple estimators are often poorly conditioned with some small eigenvalues, and so are unsuitable for many real world applications, including portfolio optimization and tracking error minimization. This chapter proposes a new method to estimate the large dimensional covariance matrix. In this work, we estimate the large dimensional realized covariance matrix by using the asymptotic principal components analysis based factor model method and the singular value decomposition method.

The first section of this chapter is dedicated to the literature review of the covariance matrix estimation. The second section reviews the realized covariance matrix. The third section of this chapter reviews the factor model and the fourth section of this chapter proposes two covariance matrix estimators. We conclude in the last section.

#### 1.1. Introduction

A covariance matrix of asset returns plays an important role in modern portfolio analysis and risk management. Markowitz's (1952) mean-variance portfolio optimization theory shows that we can construct optimal portfolios if accurate estimation of expected returns, variance and

covariance of every asset could be obtained. This section provides a brief review of the literature on estimating a return covariance matrix.

The literature on estimating a return covariance matrix is quite extensive. We focus our review on methods that use mostly historical stock return data and we do not consider the methods that require non-price data.

### 1.1.1. Sample Historical Covariance Matrix

The traditional estimator is the sample covariance matrix. If we want to calculate sample covariance matrix for the stocks in the Russell 1000 index using daily data for a year,  $N$  is 1000 stocks and  $T$  is around 250 trading days. If we denote the number of stocks as  $N$  and the  $N \times 1$  vector of daily stock returns on date  $t$  as  $R_t$ , generally we assume that stock returns are independent and identically distributed (iid) with population mean  $\mu$  and population covariance matrix  $\Sigma$  through time. The sample covariance estimator  $S$  is then,

$$S = \frac{1}{T} \sum_{t=1}^T (R_t - \bar{R})(R_t - \bar{R})', \quad (1.1)$$

where  $T$  is the number of the days in a year and  $\bar{R}$  is the sample mean of daily return.

Although the sample covariance matrix based on historical return data contains some useful information about future covariances, many studies find that it contains a lot of noise and under-performs other methods as a forecast for the future covariance matrix.

There are two main problems with using this conventional sample covariance estimator. One is that we are assuming the covariance estimator is constant throughout the year, while in practice, the covariance estimator varies through time. Another problem is that the sample covariance matrix is singular when the number of observations  $T$  is less than or equal to the

number of assets  $N$ . Also, it is often close to singular when the number of observations exceeds the number of assets if the number of assets is not small, even if the true covariance matrix is known to be non-singular. This is particularly a problem when we want to use these covariance estimators to construct optimized portfolios. This is so because optimization calls for the inverse of covariance matrix as input, while the sample covariance estimators are not invertible when  $N > T$ .

As market conditions change over time, the covariance structure of returns changes as well. A conventional method to characterize the time-varying covariance structure of returns is to employ a multivariate GARCH model or EWMA model. The major problem with these models is that they require a large number of parameters to be estimated and are hence computationally feasible for only a small number of stocks. For example, if the number of stocks is 1000, there are more than 500,000 parameters in the covariance matrix. By taking the daily data of the past year, we roughly have  $T=250$ . Without imposing any structure, it is unrealistic to estimate the large dimensional covariance matrix with so few data points. As we are interested in estimating the return covariance matrix for a reasonably large amount of stocks, these models are not very helpful. This problem is tackled from a number of aspects. One method is to impose some structure on the covariance matrix to reduce the estimation parameters. Although this introduces some specification errors in the estimation process, it could improve the overall performance of the estimated covariance matrix. Another method is to use high frequency data to increase the observations for estimation. Studies have found that using high frequency data often incurs other problems related to market microstructure. In addition, as portfolios are usually managed on a relatively long term basis (as compared to stocks which may be traded on a more frequent minute

or even second basis), we cannot rely on using high frequency data to improve the estimation of a covariance matrix for portfolio management purposes.

Another issue is that when we have lots of stocks, the sample covariance matrix is singular. Assuming we have 1000 stocks, with daily sampling we get about 20 data points per month, which is much less than the number of stocks. The existing solution to this problem is to shrink the sample covariance matrix towards some target, as is done by Ledoit and Wolf (2003).

### **1.1.2.Index Models**

The most well known structural model is the Sharpe (1963)'s single-index model. It assumes that stocks move together only because of their common responses to the market index. Elton, Gruber, Brown and Goetzmann (2007) find that the single-index covariance estimator outperforms the sample historical correlation matrix. This indicates that a large part of the observed covariance structure between securities, not captured by the single-index model, represents random noise with respect to forecasting. In single-index model, beta measures the sensitivity of a stock's return to the return of the market. The single-index model is only as good as the estimates of betas.

### **1.1.3.Shrinkage Model**

Ledoit and Wolf (2003) introduce a shrinkage method as an alternative way to impose a factor structure. First, they use a weighted average of the sample covariance matrix and the identity matrix. The full rank of the identity matrix guarantees the full rank of the covariance estimator. Then they consider the identity matrix multiplied by the average of the diagonal elements of the sample covariance matrix.

$$\hat{\Sigma} = \rho S + (1 - \rho)I \quad (1.2)$$

$$\hat{\Sigma} = \rho S + (1 - \rho)\mu I \quad (1.3)$$

This is sensible in the context of their work because their objective is to obtain the estimator that minimizes the Frobenius norm of the difference with the true covariance matrix. Ledoit and Wolf show that this estimator has smaller risk and is better conditioned even when the dimension of the covariance matrix is large compared to the sample size. The Ledoit and Wolf prove that their estimator is consistent, but this is a large sample property and may not be very useful as the large T is not realistic in practice.

To solve these two problems, we would like to have a covariance matrix estimator that allows the time variation non-parametrically and is full rank when N is greater than T.

## 1.2. Realized Covariance Matrix

The realized covariance methods employ high-frequency data to enable precise estimation of the daily covariance of the underlying assets, thus making it effectively observable. Realized covariance estimators are introduced by Anderson, Bollerslev, Diebold and Labys (2003) and Barndorff-Nielsen and Shephard (2005) and look very similar to the conventional sample covariance estimator. Under the assumption that prices follow an arbitrage free semi-martingale, Barndorff-Nielsen and Shephard (2004) show that the realized covariance estimator is a consistent estimator of the integrated covariance.

### 1.2.1. Construction of Realized Covariance

A daily realized covariance is obtained by summing the products of intraday returns. Once such daily measures have been obtained, they can be modeled, e.g. for a prediction purpose.

A nice feature of this approach is that unlike MGARCH and multivariate stochastic volatility models, the  $N(N-1)/2$  covariance components of the conditional variance matrix (or, rather, the components of its Choleski decomposition) can be forecasted independently, using as many univariate models.

The time- $t$  realized  $h$ -period covariance for asset  $i$  and asset  $j$  ( $i, j = 1, \dots, n$  and  $i \neq j$ ) is defined as

$$\text{cov}_{ij,h}(t; m) = \sum_{i=1, \dots, mh} r_{k,(m)}(t-h+(i/m)) r_{j,(m)}(t-h+(i/m)) \quad (1.4)$$

For any fixed sampling frequency  $m$ , the realized covariance is directly observable.

### 1.2.2. Latent Factor Structure in Volatility

Andersen, Bollerslev, Diebold and Ebens (2001) explore the properties of realized volatility in the context of a simple multivariate model with an explicit factor structure. They focus on three of the empirical results noted above: the tendency for volatilities to move together, the tendency for correlations to be high when the corresponding volatilities are high, and the tendency for an arbitrary correlation to be high when other correlations are also high.

$$\begin{aligned} r_{it} &= \lambda_i f_t + v_{it}, v_{it} \sim (0, \omega_i^2), \text{cov}(v_{it}, v_{jt}') = 0, \forall i \neq j, t \neq t', \\ &\text{where } i, j = 1, \dots, N, \text{ and } t = 1, \dots, T. \\ f_t | I_t &\sim (0, h_t) \end{aligned} \quad (1.5)$$

It is readily established that volatilities tend to move together in such a factor model. Concretely, the  $i$ th and  $j$ th time- $t$  conditional variances, for arbitrary  $i$  and  $j$ , are

$$h_{it} = \lambda_i^2 h_t + \omega_i^2 \quad (1.6)$$

$$h_{jt} = \lambda_j^2 h_t + \omega_j^2 \quad (1.7)$$

Note in particular that the conditional variance  $s_t$ , which are themselves covariance stationary stochastic processes, are linear functions of latent volatility  $h_t$  and are therefore driven entirely by movements in volatility. The unconditional covariance between  $h_{it}$  and  $h_{jt}$  is

$$\begin{aligned} & \text{cov}(h_{it}, h_{jt}) \\ &= E[((\lambda_i^2 h_t + \omega_i^2) - (\lambda_i^2 E(h_t) + \omega_i^2))((\lambda_j^2 h_t + \omega_j^2) - (\lambda_j^2 E(h_t) + \omega_j^2))] \\ &= \lambda_i^2 \lambda_j^2 (h_t - E(h_t))^2 \end{aligned} \quad (1.8)$$

which is unambiguously positive. Hence the unconditional correlation between  $h_{it}$  and  $h_{jt}$  is also unambiguously positive.

It is also readily seen why a factor structure induces high correlations in situations of high volatility. The  $ij$ th time- $t$  conditional covariance is

$$\text{cov}_{ijt} = \lambda_i \lambda_j h_t \quad (1.9)$$

so the conditional correlation is

$$\text{corr}_{ijt} = \frac{\lambda_i \lambda_j h_t}{\sqrt{\lambda_i^2 h_t + \omega_i^2} \sqrt{\lambda_j^2 h_t + \omega_j^2}} \quad (1.10)$$

### 1.2.3. Empirical Analysis of Equity Returns

Andersen, Bollerslev, Diebold and Ebens (2001) study the properties of realized covariance and correlation measures for the 30 DJIA stocks from January 2, 1993 until May 29, 1998, for a total of 1,366 trading days. They rely on artificially constructed five-minute returns. With the daily transaction record extending from 9:30 EST until 16:05 EST, there are a total of 79 five-minute returns for each day. The five-minute horizon is short enough that the accuracy of the continuous record asymptotics underlying their realized volatility measures work well, and

long enough that the confounding influences from market microstructure frictions are not overwhelming.

Andersen, Bollerslev, Diebold and Ebens (2001) find that the unconditional distributions of realized variances and covariances are highly right-skewed, while the realized logarithmic standard deviations and correlations are approximately Gaussian, as are the distributions of the returns scaled by realized standard deviations. Realized volatilities and correlations show strong temporal dependence and appear to be well described by long-memory processes. Finally, there is strong evidence that realized volatilities and correlations move together in a manner broadly consistent with latent factor structure.

Many key economic and financial questions depend upon the perceived commonality in volatility movements across assets and markets. Most of the existing evidence concerning the extent of such co-movements relies on very specific parametric volatility models. The realized volatility measures allow for a direct assessment of the relationship between the individual standard deviations and correlations.

#### **1.2.4. Modeling and Forecasting Multivariate Realized Volatility**

Typical econometric approaches for multivariate volatility modeling include multivariate GARCH models stochastic volatility models and realized covariance measures. While in the GARCH and stochastic volatility framework the volatility process is latent, the realized covariance methods employ high-frequency data to enable precise estimation of the daily covariance of the underlying assets, thus making it effectively observable. The existing literature has typically focused on univariate analysis of realized volatilities or single realized covariance (correlation) series. Andersen Bollerslev, Diebold and Labys (2003) model log-realized

volatilities and realized correlations with univariate ARFIMA models, while Corsi (2009) and Audrino and Corsi (2007) develop heterogeneous autoregressive (HAR) models to capture the strong persistence through a hierarchical autoregressive structure.

### 1.2.5. Monthly Realized Covariance Matrix

As shown by Andersen Bollerslev, Diebold and Labys (2003), although the use of the realized covariance matrix facilitates rigorous measurement of conditional volatility in much higher dimensions than is feasible with MGARCH and multivariate SV models, it does not allow the dimensionality to become arbitrarily large. Indeed, to ensure the positive definiteness of the realized covariance matrix, the number of assets ( $N$ ) cannot exceed the number of intraday returns for each trading day. The main drawback is that intraday data remain relatively costly and are not readily available for all assets. Furthermore, a large amount of data handling and computer programming is usually needed to retrieve the intraday returns from the raw data files supplied by the exchanges or data vendors. On the contrary, working with daily data is relatively simple and the data are broadly available.

In this study, I assume the covariance matrix is constant over a month, and then I estimate the monthly covariance matrix using daily demeaned log-returns. Assume  $r_t$  is the  $N \times 1$  vector of demeaned stock returns at day  $t$ , then the realized covariance estimator over a month is

$$RCOV = \sum_{t=1}^m r_t r_t' \quad (1.11)$$

where  $m$  is the number of days over a given month. So the realized covariance matrix is a non-parametric estimator, and it will achieve consistency when the sample number of intra-month observations goes to infinity.

### **1.3. Factor Models**

The objective of modern portfolio theory is to provide investment managers with an optimal portfolio according to the investor's requirements. The manager's objectives are measured in terms of expected return at minimal risk in the face of an infinite number of possible securities. Consequently, the investor needs a statistical model that describes how the return on a security is produced, referred to as the return-generating process. Factor models can be used to predict returns, to generate estimates of return, and also to estimate the variability and covariability of returns. Additionally, factor models are simple as well as intuitive, and they offer the researcher parsimony. They are a popular and widely used approach to modeling security returns.

Factor models assume that the return on a security is sensitive to the movements of various factors. In the sense of the return-generating process, a factor model attempts to capture factors and their impact that systematically move prices of securities. These common factors are often interpreted as fundamental risk components. The model itself isolates the assets' sensitivities to these risk factors. Therefore, a primary goal of security analysis is to determine these factors and the sensitivities of the security to the determined factors.

Multifactor models of security returns can be divided into three types: macroeconomic, fundamental and statistical factor models. Macroeconomic factor models use observable economic time series, such as inflation and interest rates, as measures of the pervasive shocks to security returns. Fundamental factor models use the returns to portfolios associated with observed security attributes such as dividend yield, the book-to-market ratio, and industry identifiers. Statistical factor models derive their pervasive factors from the factor analysis of the panel dataset of security returns.

### 1.3.1 Factor Model Framework

Each type of multifactor model for asset returns has the general form:

$$R_{it} = \alpha_i + \beta_{1i}f_{1t} + \beta_{2i}f_{2t} + \cdots + \beta_{ki}f_{kt} + \varepsilon_{it} = \alpha_i + \beta_i' f_t + \varepsilon_{it} \quad (1.12)$$

Where:

$R_{it}$  Return on asset  $i$  ( $i=1, \dots, N$ ) in time period  $t$  ( $t=1, \dots, T$ )

$\alpha_i$  The intercept ( $i=1, \dots, N$ )

$f_{kt}$  The  $k^{th}$  common factor ( $k=1, \dots, K$ ) in time period  $t$  and  $f_t = (f_{1t}, \dots, f_{kt})'$

$\beta_{ki}$  The factor loading or factor beta for asset  $i$  on  $k^{th}$  factor and  $\beta_i = (\beta_{1i}, \dots, \beta_{ki})'$

$\varepsilon_{it}$  The specific factor of asset  $i$  in time period  $t$

Assumption:

1. The factors are stationary with unconditional mean and variance:

$$E[f_t] = \mu_f \quad (1.13)$$

$$\text{cov}(f_t) = E[(f_t - \mu_f)(f_t - \mu_f)'] = \Omega_f \quad (1.14)$$

2. The asset specific error terms  $\varepsilon_{it}$  are uncorrelated with each of the common factors  $f_{kt}$ . This means that the outcome of the factor has no bearing on the random error term, so that:

$$\text{cov}(f_{kt}, \varepsilon_{it}) = 0, \text{ for all } k, i \text{ and } t \quad (1.15)$$

3. It is also assumed that the error terms  $\varepsilon_{it}$  are serially uncorrelated and contemporaneously uncorrelated across assets. This means that the outcome of the random error term of one security has no bearing on the outcome of the random error term of any other security.

$$\text{cov}(\varepsilon_{it}, \varepsilon_{js}) = \sigma_i^2, \text{ for all } i=j \text{ and } t=s \quad (1.16)$$

$$\text{cov}(\varepsilon_{it}, \varepsilon_{js}) = 0, \text{ otherwise} \quad (1.17)$$

$$\text{So } \text{cov}(\varepsilon) = \begin{bmatrix} \sigma_{\varepsilon_{11}}^2 & 0 & \cdots & 0 \\ 0 & \sigma_{\varepsilon_{22}}^2 & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & & \sigma_{\varepsilon_{NN}}^2 \end{bmatrix} \quad (1.18)$$

The multifactor model (1.12) may be rewritten as a cross-sectional regression model at time  $t$  by stacking the equations for each asset to give

$$\underset{(NX1)}{R_t} = \underset{(NX1)}{\alpha} + \underset{(NXk)}{B} \underset{(KX1)}{f_t} + \underset{(NX1)}{\varepsilon_t}, t = 1, \dots, T. \quad (1.19)$$

### 1.3.2. Asymptotic Principal Components

To estimate the statistical factor model, we can use either the maximum likelihood method or the standard principal components method. But neither method is well configured for asset returns where the cross-section tends to be very large. Connor and Korajczyk (1986) develop an alternative method called asymptotic principal components, building on the approximate factor model theory of Chamberlain and Rothschild (1983). Connor and Korajczyk analyze the eigenvector decomposition of the  $T \times T$  cross product matrix of returns rather than of the  $n \times n$  covariance matrix of returns. They show that given a large cross-section, the first  $k$  eigenvectors of this cross-product matrix provide consistent estimates of the  $k \times T$  matrix of factor returns. Connor and Korajczyk (1988) extend the procedure to account for cross-sectional heteroskedasticity and Jones (2001) extends the procedure to account for time-series heteroskedasticity.

Since the factors are eigenvectors, they have statistical properties:

$$E[f_t] = 0 \quad (1.20)$$

$$\text{cov}(f_t) = E[(f_t - \mu_f)(f_t - \mu_f)'] = I \quad (1.21)$$

### 1.3.3. Optimal Number of Factors

Determining the number of factors is a crucial step in the identification of the factor structure. There is unique number of common factors for which the data satisfy the assumptions of the approximate factor model and the common and idiosyncratic component can be identified and consistently estimated.

Bai and Ng (2002) pioneered the literature by proposing an information criterion aimed at minimizing the variance of the idiosyncratic components. They modify the usual AIC and BIC taking into account the double asymptotic framework when penalizing the criterion in order to avoid overparametrization.

Kapetanios (2010) considers the limit of the empirical distribution of the eigenvalues of the sample covariance matrix. The idea is that the number of eigenvalues diverging as  $N$  diverges is equal to the number of factors driving the dataset.

Onatski (2009) tests the null hypothesis of  $r_0$  static factors against the alternative of  $r_1$  static factors. The test is based on the few largest eigenvalues of the covariance matrix of a complex-valued sample derived from the original dataset, which asymptotically distribute as a Tracy-Widom.

## 1.4. Covariance Estimators

In this study, we analyze securities returns using statistical factor model. Given a demeaned return matrix  $R_{n \times T}$ , it contains the returns for  $N$  assets at  $T$  time intervals. We use realized covariance matrix as sample covariance matrix. So

$$RCOV_{NXN} = RR' \quad (1.22)$$

$$RCOV_{TXT} = R'R \quad (1.23)$$

### 1.4.1 Asymptotic Principal Components Based Factor Model Covariance Estimator

Using the asymptotic principal components method, we can first extract factors from T by T matrix  $R'R$ . We then regress the return on the factors. For each asset, we have a time series regression as follows:

$$R_i = F' B_i + \varepsilon_i, i = 1, \dots, N.$$

$$\begin{bmatrix} r_{i1} \\ \vdots \\ r_{iT} \end{bmatrix} = \begin{bmatrix} f_{11} & \dots & f_{k1} \\ \vdots & \ddots & \vdots \\ f_{1T} & \dots & f_{kT} \end{bmatrix} \begin{bmatrix} \beta_{i1} \\ \vdots \\ \beta_{ik} \end{bmatrix} + \begin{bmatrix} \varepsilon_{i1} \\ \vdots \\ \varepsilon_{iT} \end{bmatrix} \quad (1.24)$$

Then for all assets, we can get the functional form:

$$R = B F + E \quad (1.25)$$

$N \times T$     $N \times k$   $k \times T$     $N \times T$

Given the functional form of asset returns, we can calculate the covariance estimator:

$$RCOV = R * R' = \begin{bmatrix} r_{11} & \dots & r_{1T} \\ \vdots & \ddots & \vdots \\ r_{N1} & \dots & r_{NT} \end{bmatrix} \begin{bmatrix} r_{11} & \dots & r_{N1} \\ \vdots & \ddots & \vdots \\ r_{1T} & \dots & r_{NT} \end{bmatrix}$$

$$= B F F' B' + E E'$$

$$= \begin{bmatrix} \beta_{11} & \dots & \beta_{1k} \\ \vdots & \ddots & \vdots \\ \beta_{N1} & \dots & \beta_{Nk} \end{bmatrix} \begin{bmatrix} f_{11} & \dots & f_{1T} \\ \vdots & \ddots & \vdots \\ f_{k1} & \dots & f_{kT} \end{bmatrix} \begin{bmatrix} f_{11} & \dots & f_{k1} \\ \vdots & \ddots & \vdots \\ f_{1T} & \dots & f_{kT} \end{bmatrix} \begin{bmatrix} \beta_{11} & \dots & \beta_{N1} \\ \vdots & \ddots & \vdots \\ \beta_{1k} & \dots & \beta_{Nk} \end{bmatrix} \quad (1.26)$$

$$+ \begin{bmatrix} \varepsilon_{11} & \dots & \varepsilon_{1T} \\ \vdots & \ddots & \vdots \\ \varepsilon_{N1} & \dots & \varepsilon_{NT} \end{bmatrix} \begin{bmatrix} \varepsilon_{11} & \dots & \varepsilon_{N1} \\ \vdots & \ddots & \vdots \\ \varepsilon_{1T} & \dots & \varepsilon_{NT} \end{bmatrix}$$

Given the properties of the statistical factors, we know that

$$\text{cov}(f_t) = I \quad (1.27)$$

$$\text{cov}(\varepsilon) = \begin{bmatrix} \sigma_{\varepsilon_{11}}^2 & 0 & \dots & 0 \\ 0 & \sigma_{\varepsilon_{22}}^2 & & \vdots \\ \vdots & & \ddots & \\ 0 & \dots & & \sigma_{\varepsilon_{NN}}^2 \end{bmatrix} \quad (1.28)$$

I propose my first covariance estimator:

$$RCOV_{FM}^{NXN} = \begin{bmatrix} \hat{\beta}_{11} & & \hat{\beta}_{1k} \\ & \ddots & \\ \hat{\beta}_{N1} & & \hat{\beta}_{Nk} \end{bmatrix} \begin{bmatrix} \hat{\beta}_{11} & & \hat{\beta}_{N1} \\ & \ddots & \\ \hat{\beta}_{1k} & & \hat{\beta}_{Nk} \end{bmatrix} + \begin{bmatrix} \sum_{t=1}^T \hat{\varepsilon}_{it}^2 & & 0 \\ & \ddots & \\ 0 & & \sum_{t=1}^T \hat{\varepsilon}_{Nt}^2 \end{bmatrix} \quad (1.29)$$

The rank of the first part is less than or equal to k, the number of factors. The rank of the second part is N. So  $RCOV_{FM}$  is full rank.

#### 1.4.2 Singular Value Decomposition Covariance Estimator

I also propose another covariance estimator based on singular value decomposition.

$$\begin{aligned} RCOV_{FM}^{NXN} &= \begin{bmatrix} e_{11} & & e_{N1} \\ & \ddots & \\ e_{1N} & & e_{NN} \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_N} \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_N} \end{bmatrix} \begin{bmatrix} e_{11} & & e_{1N} \\ & \ddots & \\ e_{N1} & & e_{NN} \end{bmatrix} \\ &= \begin{bmatrix} P_1 & P_2 & \dots & P_N \end{bmatrix} \begin{bmatrix} \Lambda_1 & & \\ & \ddots & \\ & & 0 \end{bmatrix} \begin{bmatrix} \Lambda_1 & & \\ & \ddots & \\ & & 0 \end{bmatrix} \begin{bmatrix} P_1' \\ P_2' \\ \vdots \\ P_N' \end{bmatrix} \\ &= P_{k1} \Lambda_1^2 P_{k1}' \end{aligned} \quad (1.30)$$

$$\lambda_1 > \dots > \lambda_{k1}, \lambda_{k1} = \dots = \lambda_N = 0.$$

Here  $\Lambda_1$  is a  $k1$  by  $k1$  matrix with the non-zero eigenvalues.

$$\Lambda_1 = \begin{bmatrix} \sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_{k1}} \end{bmatrix} \quad (1.31)$$

$$P_i = \begin{bmatrix} e_{i1} \\ \vdots \\ e_{iN} \end{bmatrix} \quad (1.32)$$

Similarly,

$$\begin{aligned}
RCOV_{TXT} &= \begin{bmatrix} q_{11} & & q_{T1} \\ & \ddots & \\ q_{1T} & & q_{TT} \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_T} \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_T} \end{bmatrix} \begin{bmatrix} q_{11} & & q_{1T} \\ & \ddots & \\ q_{T1} & & q_{TT} \end{bmatrix} \\
&= \begin{bmatrix} Q_1 & Q_2 & \cdots & Q_T \end{bmatrix} \begin{bmatrix} \Lambda_1 & & \\ & \ddots & \\ & & 0 \end{bmatrix} \begin{bmatrix} \Lambda_1 & & \\ & \ddots & \\ & & 0 \end{bmatrix} \begin{bmatrix} Q_1' \\ Q_2' \\ \vdots \\ Q_T' \end{bmatrix} \\
&= Q_{k1} \Lambda_1^2 Q_{k1}'
\end{aligned} \tag{1.33}$$

For each asset, we can do time series regression. And for all securities,

$$R = BQ_{k1}' + D \tag{1.34}$$

Then

$$B = RQ_{k1} (Q_{k1}' Q_{k1})^{-1} = RQ_{k1} \tag{1.35}$$

From singular value decomposition (SVD), suppose  $R = U\Lambda V'$ . Here  $U$  is  $N$  by  $N$  matrix,  $V$  is  $T$  by  $T$  matrix and  $\Lambda$  is a  $N$  by  $T$  matrix.

$$\Lambda = \begin{bmatrix} \Lambda_1 & 0 \\ 0 & 0 \end{bmatrix} \tag{1.36}$$

Let  $U = [U_1 \ U_2]$ ,  $V = [V_1 \ V_2]$ ,  $U_1$  is  $N$  by  $k_1$  and  $U_2$  is  $N$  by  $(N-k_1)$ .  $V_1$  is  $T$  by  $k_1$  and  $V_2$  is  $T$  by  $(T-k_1)$ .

$$R = U\Lambda V' = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Lambda_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1' \\ V_2' \end{bmatrix} = U_1 \Lambda_1 V_1' \tag{1.37}$$

$$RR'_{NXN} = U\Lambda_1 V' V \Lambda_1 U' = U\Lambda_1^2 U' = P_{k1} \Lambda_1^2 P_{k1}' \tag{1.38}$$

$$R' R_{TXT} = V \Lambda_1 U' U \Lambda_1 V' = V \Lambda_1^2 V' = Q_{k1} \Lambda_1^2 Q_{k1}' \tag{1.39}$$

$U$  and  $V$  are all eigenvectors corresponding to eigenvalue  $\Lambda_1^2$ , and therefore  $U_1 = P_{k1}$ ,  $V_1 = Q_{k1}$ . The eigenvectors of  $RR'$  are called the left singular vectors ( $P_{k1}$ ) while the eigenvectors of

$R$ 's are the right singular vectors ( $Q_{k1}$ ). By retaining the nonzero eigenvalues  $k1$ , a singular value decomposition (SVD) can be constructed. That is

$$R = P_{k1} \Lambda_1 Q_{k1}' \quad (1.40)$$

So

$$B = R Q_{k1} = P_{k1} \Lambda_1 Q_{k1}' Q_{k1} = P_{k1} \Lambda_1 \quad (1.41)$$

If we choose  $k$  factors,

$$R = P_k \Lambda_k Q_k' + D, D = P_{k2} \Lambda_{k2} Q_{k2}' \quad (1.42)$$

Where  $k2 = k1 - k$ .

And then time series regression,

$$R_{N \times T} = B_{N \times k} Q_k' + D_{N \times T} \quad (1.43)$$

$$\begin{bmatrix} r_{11} & \dots & r_{1T} \\ \vdots & & \vdots \\ r_{N1} & \dots & r_{NT} \end{bmatrix} = \begin{bmatrix} e_{11} & \dots & e_{k1} \\ \vdots & & \vdots \\ e_{1N} & \dots & e_{kN} \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} & & \\ & \dots & \\ & & \sqrt{\lambda_k} \end{bmatrix} \begin{bmatrix} q_{11} & \dots & q_{T1} \\ \vdots & & \vdots \\ q_{1k} & \dots & q_{Tk} \end{bmatrix} + \begin{bmatrix} d_{11} & \dots & d_{1T} \\ \vdots & & \vdots \\ d_{N1} & \dots & d_{NT} \end{bmatrix}$$

$$RCOV_{N \times N} = R_{N \times T} * R'_{N \times T} = BB' + DD' \quad (1.44)$$

To estimate the realized covariance matrix, we obtain the estimation of error matrix from matrix subtraction. The matrix of the error is the return matrix minus  $BB'$ , and we estimate the variance of the error using the realized variance. We retain only the diagonal part and ignore off-diagonal 0 since we assume the errors of the assets are not correlated. Now we have the second covariance estimator:

$$RCOV_{SVD}^{NXN} = \begin{bmatrix} e_{11} & & e_{k1} \\ & \ddots & \\ e_{1N} & & e_{kN} \end{bmatrix} \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_k \end{bmatrix} \begin{bmatrix} e_{11} & & e_{1N} \\ & \ddots & \\ e_{k1} & & e_{kN} \end{bmatrix} + \begin{bmatrix} \sum_{t=1}^T d_{it}^2 & & 0 \\ & \ddots & \\ 0 & & \sum_{t=1}^T d_{Nt}^2 \end{bmatrix} \quad (1.45)$$

Originally, as the error matrix  $D$  is  $N$  by  $T$  matrix,  $DD'$  is singular. The covariance matrix as well is singular. Here, the diagonal error matrix causes the covariance matrix to be invertible.

### 1.5. Conclusion

This chapter proposes two new methods to estimate the large dimensional covariance matrix. The purpose of the new estimators is to constructing the time varying large dimensional covariance matrix that is full rank.

## Chapter 2

### **Empirical Study of Estimating High Dimensional Covariance Matrix: Through Simulation and Backtesting**

There are two major drawbacks of the existing covariance matrix estimators, especially for large dimensional covariance matrix estimators. One is that there are too many parameters need to be estimated and make the covariance matrix estimators not time varying. The second one is that the covariance matrix estimators are singular and we could not apply the inverse of the covariance matrix in the optimization problem. Chapter one proposed two high dimensional covariance matrix estimators. One is using asymptotic principal components analysis based factor model method and the other one based on the singular value decomposition method. These two estimators are time varying and invertible.

This chapter conducts a comprehensive empirical analysis of the new methods of estimating the large dimensional covariance matrix using the standard comparison criteria. We first do the comparison through the simulation. And then we use the real stock returns from S&P 500 stocks in past fifty years to do backtesting. Through simulation, we show that our proposed estimators are closer to the true covariance matrix than the current popular shrinkage estimator. I also conduct the out sample portfolio performance tests and find that the portfolios constructed based on my proposed estimators have lower risk than portfolios constructed using the shrinkage matrix. Through backtesting, we show that the portfolio consisting S&P 500 stocks and constructed based on our proposed covariance estimators has lower volatility. And this applies to both global minimum variance portfolio and the mean variance portfolio with a target return.

The first section shows the criteria to compare the covariance matrix estimators. Section

two shows the simulation results. Backtesting results follows in section three. And section four concludes.

## **2.1. Literature on Comparing Covariance Estimators**

In this section, we first review the statistical and economic evaluation criteria that have been used in the literature to compare alternative covariance estimators. And then list the criteria we use to evaluate our estimators. Elton and Gruber (1973) suggest that both statistical and economic criteria should be used to compare the accuracy of alternative covariance estimation techniques. Most existing studies have followed this dual comparison approach to compare the performance of covariance estimators.

### **2.1.1. Statistical Criteria**

The statistical criteria quantify the difference between values implied by an estimator and the true values of the quantity being estimated. For our study, the statistical criteria measure the ability of different covariance estimators to estimate accurate pair-wise covariances. The popular statistical measures include mean square error (MSE), and other forms from the same family such as root mean square error (RMSE) and mean absolute error (MAE).

In statistics, the mean squared error (MSE) of an estimator is a risk function, corresponding to the expected value of the squared error loss or quadratic loss. MSE measures the average of the squares of the "errors." The error is the amount by which the value implied by the estimator differs from the quantity to be estimated. The difference occurs because of randomness or because the estimator doesn't account for information that could produce a more accurate estimate.

The MSE is the second moment of the error, and thus incorporates both the variance of the estimator and its bias. For an unbiased estimator, the MSE is the variance of the estimator. Like the variance, MSE has the same units of measurement as the square of the quantity being estimated. In an analogy to standard deviation, taking the square root of MSE yields the root-mean-square error or root-mean-square deviation (RMSE or RMSD), which has the same units as the quantity being estimated; for an unbiased estimator, the RMSE is the square root of the variance, known as the standard deviation.

Both MSE and RMSE serve to aggregate the magnitudes of the errors in predictions for various times into a single measure of predictive power. A smaller MSE/RMSE indicates a smaller estimation deviation from the actual observation. Root mean square error (RMSE) is an even better measure because RMSE has the same unit as the measured subject, and therefore is easier to interpret. The RMSE is a good measure of accuracy, but only to compare forecasting errors of different models for a particular variable and not between variables, as it is scale-dependent.

The problem with MSE (and RMSE) measure is that it measures the average pair-wise covariance estimation errors. By breaking the covariance matrix to the element-by-element level estimates, the information contained in the structure of the covariance matrix is lost. As a result, it is purely a statistical measure and does not provide any implications regarding the use of the estimation of the covariance matrix.

### **2.1.2. Economic Criteria**

The economic criteria measure the ability of covariance estimators in producing efficient out-of-sample portfolios. There are a few methods proposed to establish the economic

significance of different covariance estimators. Cohen and Pogue (1967) compare the performances of different covariance structures by comparing the locations of the mean-variance efficient frontiers constructed according to these different covariance estimators. Subsequently, Elton and Gruber (1973) compare the risk and return relationships of portfolios on the different efficient frontiers over a number of pre-specified risk levels. Chan, Karceski and Lakonishok (1999) use the performance of the minimum variance portfolio (MVP) to compare alternative covariance estimators. The MVP is constructed by minimizing the portfolio variance without the constraint on the target level of returns. It is the only portfolio on the efficient frontier whose weights do not depend on expected returns. Comparing the performance of the MVPs therefore helps to focus on the effect of the estimation of covariances than expected returns.

### 2.1.3. Comparison Criteria

In our study, we use both statistical and economic criteria to compare our proposed estimator with other existing ones.

For the statistical criteria, we follow Ledoit and Wolf (2004) and define the Percentage Relative Improvement in Average Loss of covariance estimator as:

$$PRIAL = 1 - \frac{E[\|\hat{\Sigma} - \Sigma\|^2]}{E[\|S - \Sigma\|^2]} \quad (2.1)$$

Here  $\hat{\Sigma}$  is covariance estimator,  $\Sigma$  is true covariance matrix, and  $S$  is the sample covariance matrix.

Given  $E[\|S - \hat{\Sigma}\|^2] = E[\|S - \Sigma\|^2] + E[\|\Sigma - \hat{\Sigma}\|^2]$ , we know that  $0 \leq PRIAL \leq 1$ . When the covariance estimator is equal to the real covariance, PRIAL is one. And when the covariance estimator is equal to the sample covariance estimator, PRIAL is zero. The higher PRIAL is, the

better. We apply this criteria in our simulation process.

For the economic criteria, we follow Markowitz (1952) framework. Consider a universe of  $N$  stocks whose returns are distributed with mean vector  $\mu$  and covariance matrix  $\Sigma$ . I compare then two kinds of portfolios. Markowitz (1952) defines the problem of portfolio selection as:

$$\begin{aligned} \min_w w' \Sigma w \\ \text{st. } w' \mathbf{1} = 1 \\ w' u = p \end{aligned} \tag{2.2}$$

This is referred to as mean variance efficient portfolio with target return. Additionally, if we don't set the return target, we could get the global minimum variance portfolio, which is:

$$\begin{aligned} \min_w w' \Sigma w \\ \text{st. } w' \mathbf{1} = 1 \end{aligned} \tag{2.3}$$

In practice, the covariance matrix is estimated from historical data available up to a given date, optimal portfolio weights are computed from this estimate, and then the portfolio is formed on that date and held until the next rebalancing occurs. The performance of a covariance matrix estimator is measured by the variance of this optimal portfolio after it is formed. It is a measure of out-of-sample performance or of predictive ability. An estimator that overfits in-sample data can turn out to work very poorly for portfolio selection, which is why imposing some structure is beneficial. We evaluate our covariance matrix estimator using these economic criteria in both simulation and backtesting process.

## 2.2. Simulation

In this section I use a simulation study to illustrate and augment the theoretical results and to verify the finite-sample performance of the estimator  $\hat{\Sigma}$  as well as  $\hat{\Sigma}^{-1}$ . Chapter one proposed two high dimensional covariance matrix estimators. One is using asymptotic principal components analysis based factor model method and the other one based on the singular value decomposition method. When we use all nonzero eigenvalues, these two estimators are identical. When we choose first  $k$  important eigenvalues, they are slightly different.

### 2.2.1. The Relative Estimation Errors

When  $N$  and  $T$  are very close, the largest eigenvalue of the covariance matrix is overestimated and the smallest eigenvalue of the covariance matrix is underestimated. The goal then is to find an appropriate  $T$  corresponding to  $N$  to make our estimation more accurate. Gittens and Tropp (2011) estimates bounds of eigenvalues and finds that if

$$N = \varepsilon^{-2} \max[(T - k + 1) \log(T - k + 1), k \log k], \quad (2.4)$$

then with high probability  $\hat{\lambda}_k$  estimates  $\lambda_k$  to within a relative error of  $\varepsilon$ .

Here, I adjust the monthly realized covariance matrix using daily data. Supposing that there are 20 trading days within a month, I can calculate the bias of our eigenvalue estimation given different number of the assets. Also I derive the bias for biweekly realized covariance matrix. Table 2.1 shows the relative error of the estimations given  $N=50, 100, 500, 1000$ . I choose 3 factors and 5 factors separately. We can see that when we have more assets in our estimation, the bias is smaller. I also check the bias of eigenvalues in using intraday high frequency data. De Pooter, Martens and van Dijk (2005) investigates the merits of high-frequency intraday data when forming mean-variance efficient stock portfolios with daily

rebalancing from the individual constituents of the S&P 100 index. They find that the optimal sampling frequency ranges between 30 and 65 minutes, which is considerably lower than the popular five-minute frequency.

Following their sampling frequency, if we sample every 30 minutes, we get 13 data points each trading day. And if we sample every 65 minutes, we get 7 data points each trading day. I calculate the bias of our eigenvalue estimation given different number of the assets. Table 2.2 shows the relative error of the estimations given  $N=50, 100, 500, 1000$ . I choose 3 factors and 5 factors separately.

### 2.2.2. Percentage Relative Improvement in Average Loss of Covariance Estimator

We first focus on the Percentage Relative Improvement in Average Loss of covariance estimator. For simplicity, I fix the number of factors at five in my study, and as a robustness check, three factors give similar results. For each simulation, we carry out the following steps:

- We first generate a random sample of factors  $f = (f_1, f_2, f_3, f_4, f_5)'$  with size  $T=20$  from the multivariate standard normal distribution  $N(0, I_5)$ . And then we normalize each factor such that the norm of each factor equals 1.
- Then for each dimensionality  $N$  increasing from 20 to 1000 with increment 20, we do the following:
- Generate  $N$  factor loading vectors  $b_1, \dots, b_N$  as a random sample of size  $n$  from the multivariate standard normal distribution  $N(0, I_5)$ .
- Generate a random sample of  $\varepsilon = (\varepsilon_1, \dots, \varepsilon_N)'$  with size  $T=13$  from multivariate standard normal distribution  $N(0, I_N)$ .

- Then from the model, we get a random sample of  $R = (R_1, \dots, R_N)'$  with the size N by T, T is 20 here.
- Finally, I compute the factor model based realized covariance matrix estimator  $RCOV_{FM}$  and singular value decomposition based realized covariance matrix estimator  $RCOV_{SVD}$ , as well as sample covariance matrix S.

We repeat the above simulation 500 times. We calculate the Frobenius norm of the difference between the covariance estimator and real covariance, and we calculate the PRIAL for two estimators separately. We also compare them with shrinkage estimators. The results are in Figure 2.1. The X-axis represents the number of assets, and the Y-axis represents PRIAL. The dashed line represents the singular value decomposition estimator, the dotted line represents the asymptotic principal components based factor model covariance estimator and the solid line represents shrinkage estimator. In the data generation process and the factor model here, we use five factors, and our sampling frequency is 20.

We can see that when the number of assets is small, the shrinkage estimator is the best. When the number of assets exceeds 200, the singular value decomposition estimator and the asymptotic principal components estimator are better, with the singular value decomposition estimator outperforming the asymptotic principal components estimator. We can conclude then, that when we need to estimate the covariance matrixes for a large set of assets, for example S&P 500 or Russell 1000 index, these two estimators are better.

### 2.2.3. Performance Tests

To check the economic criteria, we follow Markowitz (1952)'s performance tests, comparing the ability of covariance estimators to select portfolios of stocks with low out-of-sample variance.

In our simulation, we suppose there are 20 trading days in each month and generate data for 5 years, giving me 60 months and 1200 data points. We continue to use 5 factors, and as a robustness check, 3 factors again give similar results. We calculate the covariance estimator based on every 20 data points and the monthly return based on 20 daily returns. After constructing a portfolio based on the covariance estimators and calculating the return on it for the next month, we calculate the covariance estimator and rebalance it. As a final step, we calculate the variance of the portfolio. For each dimensionality,  $N$  increases from 20 to 1000 with an increment of 20. We do 500 simulations for each  $N$ , and the results are in Figure 2 and Figure 3.

Figure 2.2 shows the variance of the global minimum portfolio. The X-axis represents the number of assets, and the Y-axis represents the portfolio variance. The dashed line represents the singular value decomposition estimator, the dotted line the asymptotic principal components covariance estimator, and the solid line the shrinkage estimator. Here in the data generation process and the factor model, we use 5 factors, and my sampling frequency is 20. We can see that our covariance estimators have lower variance.

Figure 2.3 shows the variance of mean minimum variance portfolio with the target return at 8%. We still see that my covariance estimators have lower variance.

As a robustness check, I generate the intraday high frequency data using the half hour sampling and maintain similar results. My proposed realized covariance estimators are closer to

the true covariance matrix, and the portfolio rebalanced based on my realized covariance estimators has lower risk.

#### 2.2.4. Shrinkage Estimator

We propose two shrinkage estimators using a combination of a sample covariance matrix and my proposed covariance estimators. One is the factor model based shrinkage estimator, and another is singular value decomposition based shrinkage estimator.

$$\hat{\Sigma}_{FM} = \rho S + (1 - \rho)RCOV_{FM} \quad (2.5)$$

$$\hat{\Sigma}_{SVD} = \rho S + (1 - \rho)RCOV_{SVD} \quad (2.6)$$

We repeat the simulation process to compare the factor model based estimator with the factor model based shrinkage estimator. Also we compare the singular value decomposition based estimator with the singular value decomposition based shrinkage estimator. The results are in Figure 2.4 and Figure 2.5. We can see that if the sample covariance matrix is shrinkage to factor model based estimators, it is better than traditional shrinkage estimator. But it is slightly worse than the factor model based estimator itself. The same results hold for the singular value decomposition based estimator.

The factor model based shrinkage estimator continues to be better than the traditional shrinkage estimator, but it is a little bit worse than the factor model based estimator itself.

In chapter one of this study, we constructed two realized covariance estimator models for a large dimensional covariance matrix. One of them is based on the factor model using the asymptotic principle component method. Another one is based on the singular value decomposition method. They are both full rank, which can be used in portfolio construction. In the simulation, we show that compared to the shrinkage estimator, our proposed covariance

estimators are closer to the real covariance. In the performance tests, our covariance estimators have a lower variance compared to the shrinkage estimator.

### **2.3. Empirical Studies**

Our covariance matrix estimators perform well in simulation studies. In this section, we use the real historical stock returns to check the performance of our covariance matrix estimators through backtesting. We still follow Markowitz (1952)'s framework to construct the global minimum variance portfolio and the minimum variance portfolio with the target return. Instead of using the simulated data as we did in the previous section, we use the historical stock returns in this section.

#### **2.3.1. Data**

Stock returns were extracted from the Center for Research in Security Prices (CRSP) daily and monthly database. The same procedure is repeated for every month from January 1926 to December 2011. We use daily stock return data at the end of month  $t$  to calculate the covariance matrix. We get the optimal weight based on the covariance matrix at the end of month  $t$ , which is also the beginning of month  $t+1$ . And we rebalance the portfolio based on the optimal weights. We hold the portfolio for one month. At the end of month  $t+1$ , we could calculate the portfolio return for month  $t+1$  and we could rebalance the portfolio again based on month  $t+1$ 's daily return. Thus the in-sample period is month  $t$  and the out of sample period is month  $t+1$ . The main quantity of interest is the out-of-sample standard deviation of this investment strategy.

### **2.3.2. S&P 500 Market Index**

In our empirical study, we consider the common stocks in the Standard & Poor's 500 or S&P 500 market index. S&P 500 is based on common stocks of 500 top publicly traded US companies. It is one of the most commonly followed indices and many consider it the best representation of the market and a bellwether for the U.S. economy. Standard & Poor's introduced its first stock index in 1923. Before 1957 its primary daily stock market index was the S&P 90, a value-weighted index based on 90 stocks. The S&P 500 index in its present form began on March 4, 1957. The S&P 500 Index measures the performance of the large-cap segment of the U.S. equity universe. It is constructed to provide a comprehensive and unbiased barometer for the large-cap segment and is completely reconstituted annually at the end of June to ensure new and growing equities are reflected.

We choose to use common stocks in S&P 500 index for several reasons. The advantage of our estimators is that we could apply to large dimensional covariance matrix. S&P 500 index gives us a large enough universe to test our estimators. The stocks in S&P 500 index are highly liquid and this reduces the estimation errors.

### **2.3.3. Empirical Results**

We study the out-of-sample performance of our shrinkage estimator, using historical stock market data. We use the stocks on Standard and Poor 500 index. We have the daily stock return data from January 1926 to December 2011. At the end of each month, we use the daily return of stocks from the given month to calculate the covariance matrix estimators. We determine the optimal number of factors dynamically each month. And then we calculate the weights of the stocks based on the covariance matrix estimators. We rebalance the portfolio

using the calculated weights and hold the portfolio for a month. And then we can calculate the portfolio return over the month. Using the monthly portfolio return from January 1926 to December 2011, we could calculate the volatility of the portfolio. We use three covariance estimators: factor model based covariance matrix estimator, eigenvalue decomposition based covariance matrix estimator and the shrinkage covariance estimator that shrinkage to the average volatility. And then I compare the portfolio risks.

We consider two minimum variance portfolios: the global minimum variance portfolio and the portfolio with minimum variance under the constraint of having a targeted expected return. In both cases short sales are allowed, and no additional restriction is imposed (except that weights sum up to one). For expected returns, we just take the average realized return of S&P 500 index over the last two years. This may or may not be a good predictor of future expected returns, but our goal is not to predict expected returns: it is only to show what kind of reduction in out-of-sample variance our method yields under a fairly reasonable linear constraint. We also look at the minimum variance portfolio with 8% expected return for a robustness check. We calculate the one-year rolling volatility of the portfolio based on monthly return. The results are in Figure 2.6 and Figure 2.7.

Figure 2.6 shows the volatility of the global minimum portfolio. The X-axis represents the time, and the Y-axis represents the portfolio volatility. The dashed line represents the singular value decomposition estimator, the dotted line represents the asymptotic principal components covariance estimator, and the solid line is the shrinkage estimator. We decide the optimal number of factor each month dynamically. Figure 2.7 shows the volatility of the mean variance efficient portfolio with 8% target return. For both cases, we could see that our covariance estimators have lower variance.

## 2.4. Conclusion

In this chapter, we test the performance of our proposed covariance matrix through both simulation and backtesting.

Through simulation, we show that our proposed estimators are closer to the true covariance matrix than the current popular shrinkage estimator and the portfolios constructed based on our proposed estimators have lower risk than portfolios constructed using the shrinkage matrix.

Through backtesting, we show that the portfolio consisting S&P 500 stocks and constructed based on our proposed covariance estimators has lower volatility. And this applies to both global minimum variance portfolio and the mean variance portfolio with a target return.

Table 2.1: Relative Estimation Errors of Eigenvalues for Monthly Estimation

T=20			T=10		
N	k=3	k=5	N	k=3	k=5
1000	0.228094	0.210622	1000	0.128979	0.103685
500	0.322573	0.297864	500	0.182404	0.146633
100	0.721295	0.666044	100	0.407867	0.32788
50	1.020066	0.941928	50	0.576811	0.463693

Table 2.2: Relative Estimation Errors of Eigenvalues for Daily Estimation

T=13			T=7		
N	k=3	k=5	N	k=3	k=5
1000	0.16241	0.140624	1000	0.089706	0.057409
500	0.229682	0.198872	500	0.126864	0.081189
100	0.513584	0.444691	100	0.283676	0.181544
50	0.726317	0.628888	50	0.401178	0.256743

Figure 2.1: PRIAL of Three Covariance Estimators

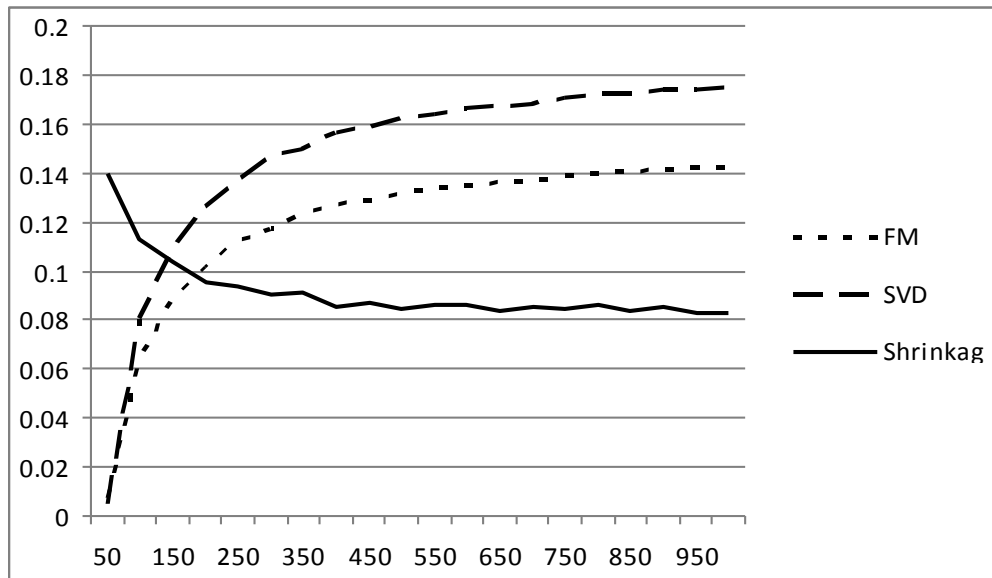


Figure 2.2: Global Minimum Portfolio Variance

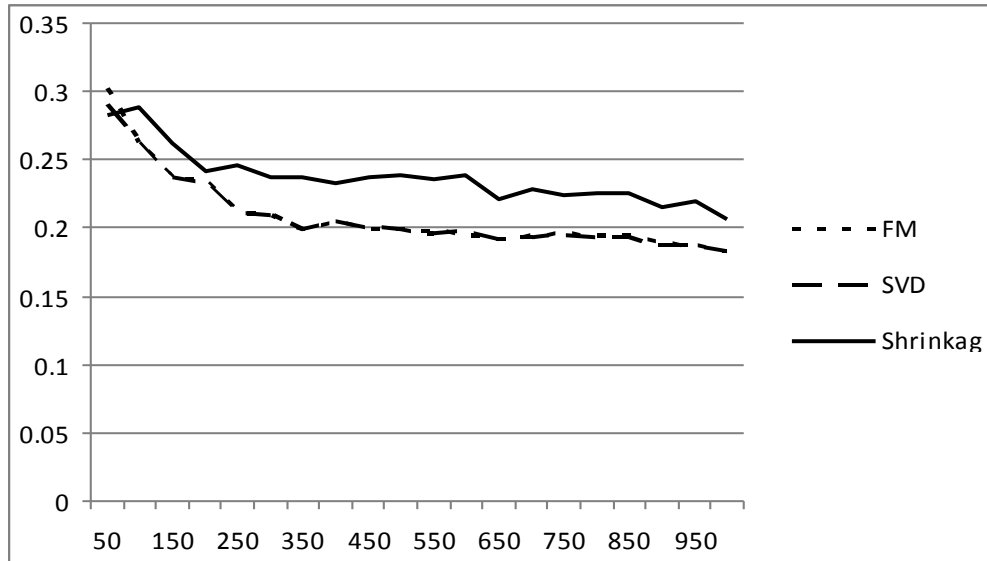


Figure 2.3: Mean Minimum Variance Portfolio with Annual Target Return 8%

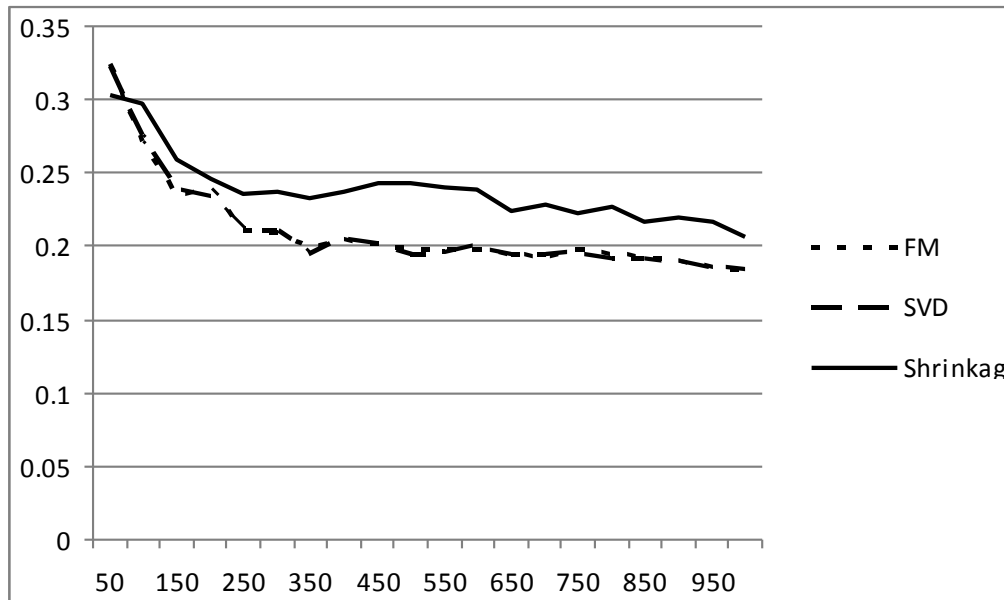


Figure 2.4: PRIAL of Factor Model Based Covariance Estimators

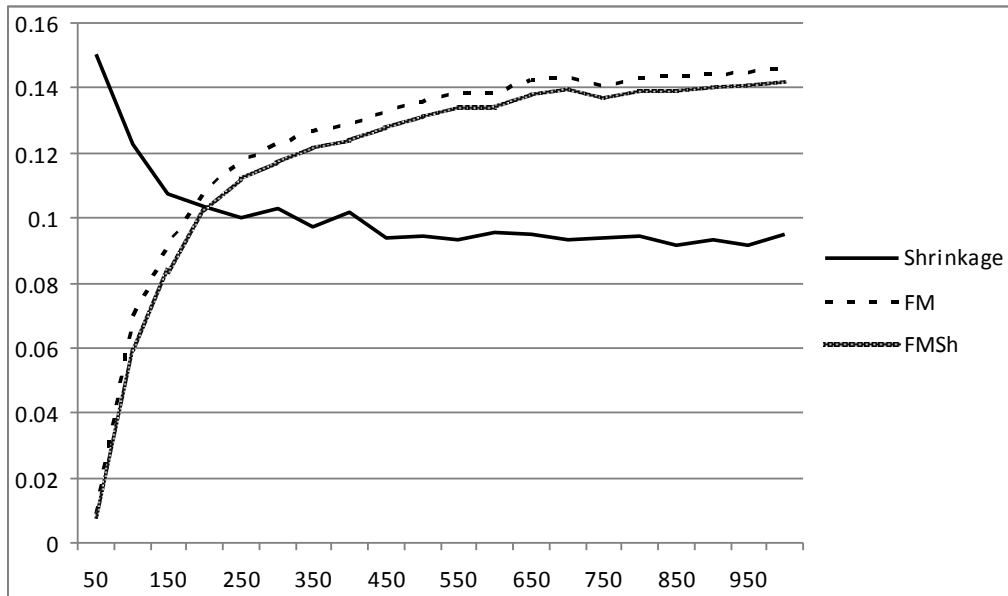


Figure 2.5: PRIAL of Singular Value Decomposition Based Covariance Estimator

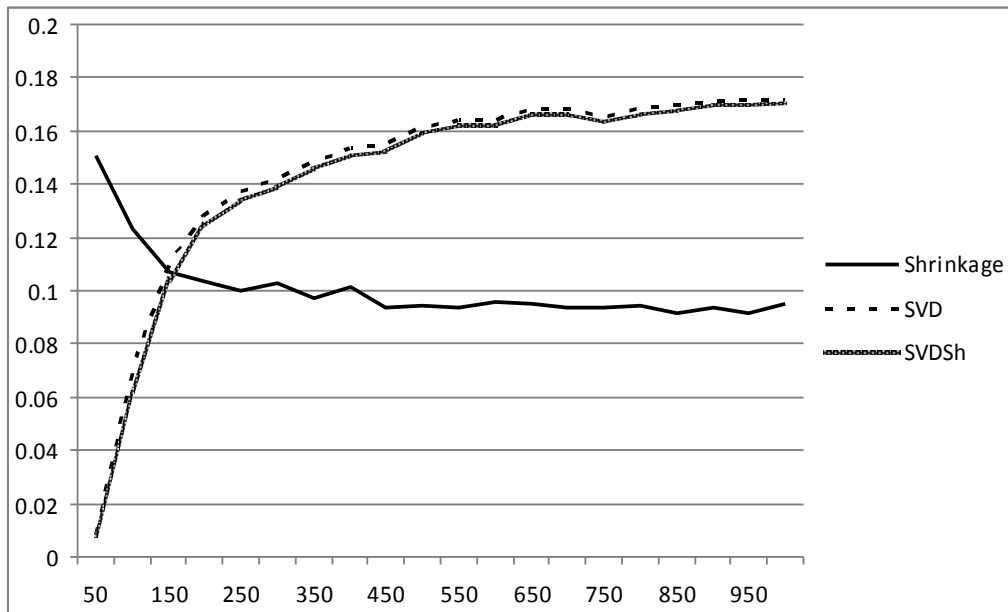


Figure 2.6: Global Minimum Portfolio Volatility for S&amp;P 500 Stocks

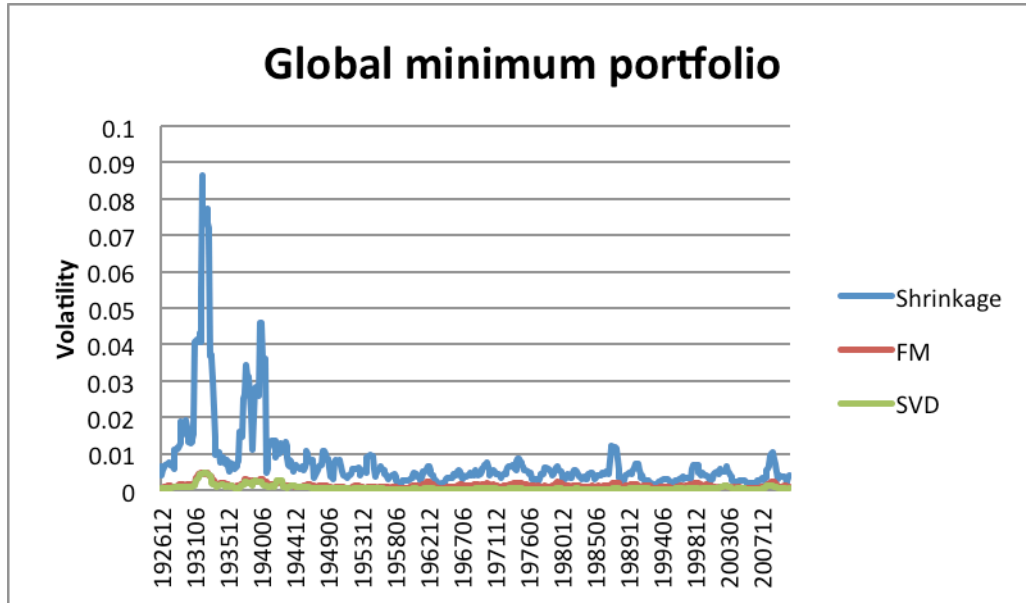
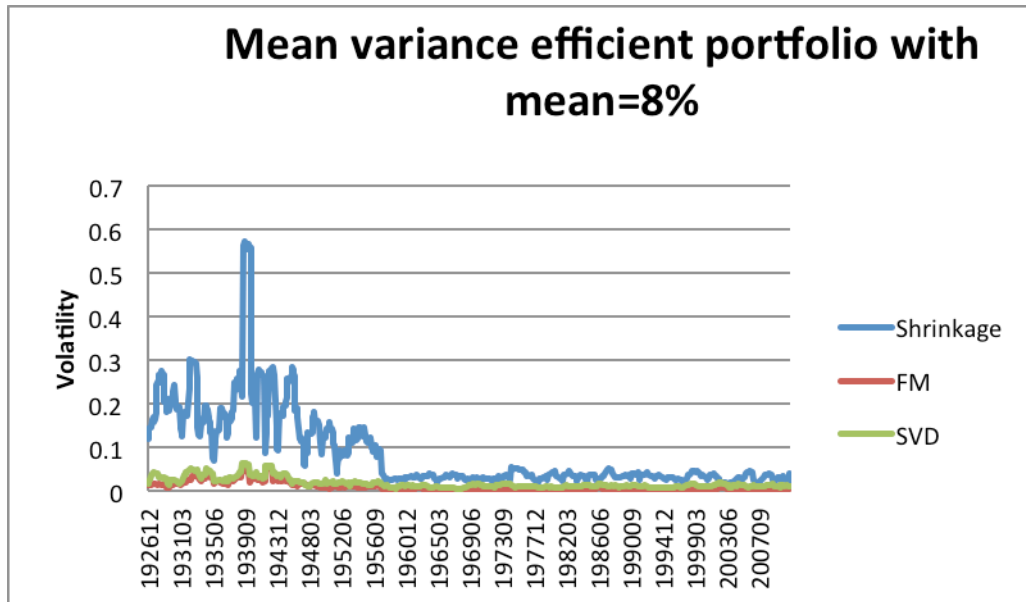


Figure 2.7: Mean Minimum Variance Portfolio Volatility for S&amp;P 500 Stocks



## Chapter 3

### A New Volatility Index using a Factor Model

In this chapter, we introduce a new form of volatility index, the idiosyncratic cross-sectional volatility index. We show that the cross-sectional dispersion of stock returns residuals could represent the movement of the market. This non-parametric estimator of market volatility is model free and could be estimated at any frequency for any region, sector, and style of world equity markets. It is also not anchored to any option pricing. Inspired by the covariance matrix estimator we propose in chapter one, we estimate this volatility index utilizing factor model.

The first section of this chapter reviews the literature on volatility indexing. The second section introduces cross-sectional volatility and proposes a volatility index—of residuals, which is cross-sectional and moves closely with the VIX for S&P 500 stocks universe. We review some volatility related empirical studies in section three and we conclude in section four.

#### 3.1. Volatility Index Background

Volatility is a statistical measure of the dispersion of returns for a given security or market index. In essence, this measure tracks the amount of uncertainty or risk concerning the degree of change in an underlying security or index value. A higher level of volatility indicates that the underlying value could span a larger range of values, thus signaling greater risk for investors holding the security or index. High volatility does not necessarily dictate a loss in holdings, but rather signals that the price of the security can change dramatically in either direction over a given interval of time. The converse is indicated by lower volatility; changes in a security or index value are predicted to occur at a steadier pace over a period of time.

Crucially, volatility does not predict the direction in the change, only the degree of change. Additionally, these indices, which investors find very important, are the measure of the market aggregate volatilities.

### **3.1.1. The Implied Volatility Index**

In 1993, the CBOE launched the first volatility index based on the implied volatilities of a set of S&P 100 options. Known initially as the VIX and as VXO after a methodology update in 2003, the index is computed as a weighted average of the implied volatilities. The implied values are inferred from the combined prices of near-the-money call and put options on the two nearest monthly expiries (above and below) to the fixed 30 calendar days maturity. In this way, the VXO represents the implied volatility of a synthetic ATM option with 30 calendar days to expiration. Other indices have since been adopted which consider longer terms to maturity. Deutsche Börse's German volatility index VDAX looks ahead 45 days, and the Marché des Options négociables de Paris (MONEP) calculates the VX1 for 31 calendar days and the VX6, for 185.

This method is well-known for suffering from two flaws: it is not model-free and it lacks a theoretical interpretation for practical implications. Volatility indices are calculated by using the implied volatilities derived from option market prices. These prices are in turn based on an option pricing model, which may be misspecified given the assumptions that underlie its derivation. Indeed, it is a well-documented fact that the implied volatility of options with the same time to maturity varies across different strike prices, giving rise to volatility smiles and skews (Rubinstein, 1994; Dumas, Fleming and Whaley, 1998; Foresi and Wu, 2005). In contrast, the Black-Scholes model assumes constant volatility. Moreover, the practical implications inferred from the theoretical interpretation of the derived implied volatility index are not

straightforward. After defending this is assertion, we will demonstrate how our new algorithm overcomes these two problems.

### **3.1.2. Model Free Volatility Index**

The CBOE implemented new methodology for its 2003 launch of the VIX, retooled to reference options on the primary US stock market benchmark S&P 500 instead of S&P 100. One innovation for this VIX was basing the construction algorithm on the concept of the fair delivery value of the future realized variance of a variance swap, as suggested by Demeterfi, Derman, Kamal and Zou (1999). A variance swap is a forward contract on the realized variance of the underlying index over the life of the contract. At maturity, the holder of the contract pays a fixed variance rate (the variance swap rate), determined at the contract initiation, and in exchange receives the realized variance rate of the underlying index. The difference between these two rates is multiplied by the notional amount of the swap. Since the contract has zero value at the time of entry, by no-arbitrage, the variance swap rate equals the risk-neutral expected value of the realized variance. Thus, the variance swap rate is the fair delivery value of future realized variance.

Demeterfi, Derman, Kamal and Zou (1999) derive a formula to replicate the variance swap rate by means of two positions: a dynamic one in futures trading and a static one in a portfolio of European call and put options over a wide range of strikes (i.e. not just near-the-money options). As it involves integrating option prices over an infinite range of strike prices, the CBOE in reality uses a discrete approximation of this formula. First, two variance swap rates are computed by using near-term (less than 30 days) and next-term (more than 30 days) S&P 500 options. The VIX is then calculated as the squared root of the weighted average of these rates.

The volatility index has an applicable theoretical interpretation under this new algorithm in that, once it is squared, it approximates the variance swap rate. Furthermore, basing the VIX on a wide range of option prices allows for it to represent a market consensus view of the expected volatility of the S&P 500. An additional theoretical interpretation for VIX provided by Carr and Lee (2009) shows that ATM implied volatility approximates the volatility swap rate of a volatility swap. Unlike variance swaps, however, the payoff on a volatility swap has been shown to be difficult to replicate.

A natural continuation of this new methodology, given the explicit meaning of the volatility indices constructed with it and its direct link to a portfolio of options, is the launch of products derivative of volatility indices (i.e. futures and options). Futures based on the VIX began to be traded on the CBOE Futures Exchange (CFE) in 2004 with the introduction of VIX options on the CBOE two years later. As Psychoyios and Skiadopoulos (2006) found, these implied volatility derivatives make it possible to to speculate in turn on future volatility or hedging volatility risk. Indeed, investor interest in these products has increased in tandem with the increase in volatility that the financial markets have experienced during the recent financial crisis. According to the 2011 annual report from the CBOE, VIX options and futures are among the most actively traded contracts at the CBOE and the CFE.

Since the inception of the VIX, the number of volatility indices calculated with a formula similar to that employed by the CBOE has noticeably increased. Table 3.1 lists the currently available volatility indices.

### 3.1.3. Stylized Facts about Volatility Indices

The relationship between stock returns and volatility has inspired a great deal of research, the pioneers of which are Markowitz (1952) and Tobin (1958). While they established the basis of modern portfolio theory by relating the expected return of a portfolio with its standard deviation Sharpe (1964) is also notable for having formalized the celebrated capital asset pricing model (CAPM). Subsequent studies by Black (1976) and Christie (1982) document a negative relationship between stock market returns and changes in ex post calculated future volatility. Moreover, Black (1976) and Schwert (1989, 1990) find that the relationship between returns and expected volatility is asymmetric. They explain that the negative return associated with an increase in volatility is larger than the positive return associated with a decrease in volatility.

There has also been extensive documentation of the negative, often asymmetric contemporaneous association between changes in volatility indices and the underlying stock index returns. This association has earned volatility indices the status of gauges of investor fear (Whaley, 2000). Whaley (2000) and Giot (2005) document the asymmetric negative correlation for the VXO in the US market, a finding confirmed for the VXN by Simon (2003) and for the VIX by Whaley (2009). With respect to European volatility indices, Siriopoulos and Fassas (2008) note a negative and asymmetric relationship between VFTSE changes and FTSE-100 returns. Similarly, González and Novales (2009) provide evidence of a strong negative contemporaneous relationship between changes in VDAX-NEW, VSMI and VIBEX-NEW and their respective stock indices returns, although there is no evidence of asymmetry. For the Greek market specifically, Siriopoulos and Fassas (2012) reports an inverse and asymmetric correlation based on GRIV.

In a recent study on the role of the VIX as a gauge of investor fear, Sarwar (2012) analyzes investor attitudes toward foreign equity markets. The focus of this study is the relationship between VIX and stock market returns in Brazil, Russia, India and China in particular. It documents a contemporaneous negative correlation for China, Brazil and India, and reports evidence of asymmetry the two former markets. Not only do these results have implications for investors with positions in stocks and for portfolio diversification in general, they also encourage further research on markets outside of the US and Europe.

### **3.2. Cross-Sectional Volatility**

The existing volatility indices suffer from a number of shortcomings. On the one hand, volatility indices are not available for an extensive set of markets because they require the presence of a liquid option market. To name a few examples, no volatility index exists for small cap stocks, growth/value stocks, or various sectors for developed markets. In most emerging markets volatility indices do not exist even at the broad market level. On the other hand, the implied volatility estimates that do exist are plagued by option-market problems that have little to do with underlying equity markets. In this section we review the existing cross-sectional volatility of returns and propose a new volatility index built on cross-sectional volatility of residuals.

#### **3.2.1. Cross-Sectional Volatility of Returns**

While volatility is generally defined as the standard deviation of returns over a period, cross-sectional volatility is a metric that indicates the available opportunity set within an investable universe. This ultimately affects the outcomes of active management. Essentially, it

measures the dispersion of a set of asset returns at a given time. The level of dispersion provides a good indication of the potential outperformance in the market for active managers to exploit. It is customarily measured using a time-series variation in returns. Cross-sectional volatility then should be monitored over a long enough period to notice any significant trends.

Active managers illustrate their skill by either overweighting or underweighting shares against a benchmark they have views on. It can be inferred that when active managers try to predict the performance of shares, they are in essence forecasting the cross-sectional dispersion (or standard deviation) of returns. This is simply a more formal term for the future distribution of relative winners and losers. As cross-sectional volatility rises, the likelihood of outperforming and underperforming the market or benchmark also increases. This rise in cross-sectional volatility also implies that the variance between the best- and worst-performing shares is large. If the cross-sectional volatility is zero, this implies all shares would generate the same return. This information is beneficial to active management, and, as such, cross-sectional volatility should be regarded as an important metric in active portfolio management.

Cross-Sectional Volatility measures the dispersion of stock returns within a market on a given day or month. The formula for calculating cross-sectional volatility (CSV) is:

$$CSV(r) = \sigma_x = \sqrt{\sum_i w_i (r_i - R)^2} \quad (3.1)$$

Where:

- $w_i$  = the beginning of period, float-adjusted capitalization weight of stock i
- $r_i$  = the total return of stock i for the period
- $R$  = the published return of the relevant market index for the period (a day or a month)

Gorman, Sapra and Weigand (2010) show that the cross-sectional dispersion of returns is related to time series volatility and time series pairwise correlations. Assume that all stocks have a variance equal to  $\sigma^2$  and a covariance with all other assets equal to  $\rho\sigma^2$ . If further assume that all the stocks have the same weight, we can get

$$\sigma_x = \sigma\sqrt{1-\rho} \quad (3.2)$$

For the general case, we may expect the approximation holds, which gives

$$\sigma_x \cong \sigma\sqrt{1-\rho} \quad (3.3)$$

To properly measure the combined impact of correlation and volatility on the portfolio, the cross-sectional volatility should be used.

Cross-sectional volatility of stock returns is also a consistent and asymptotically efficient estimator for idiosyncratic volatility. Through central limit arguments, Garcia, Mantilla- Garcia, and Martellini (2012) provide the formal conditions under which the cross-sectional variance (CSV) of stock returns asymptotically converges towards the average idiosyncratic variance. Garcia, Mantilla- Garcia, and Martellini (2013) introduce the cross-sectional volatility index as a new form of volatility index.

### 3.2.2. Cross-Sectional Volatility of Residuals

We have so far focused on the cross-sectional volatility of returns. Inspired by the application of the factor model, we here apply factor model on returns. We get that

$$R_{it} = \beta_i' f_t + \varepsilon_{it} \quad (3.4)$$

And then we apply the cross-sectional volatility formula on both sides. Since the factor is the same to all asset at a give time and we could get

$$CSV(R_{it}) = CSV(\beta_i)' f_t + CSV(\varepsilon_{it}) \quad (3.5)$$

We know the factor loadings are persistent, so the changes of the dispersion of the returns result more from the changes of the dispersion of the residuals. Garcia, Mantilla- Garcia, and Martellini (2012) assume the homogeneous betas and residual variances across stocks and show that the cross-sectional variance of stock returns is a consistent and asymptotically efficient estimator for idiosyncratic volatility. We apply the factor model on stock returns and we focus on the cross-sectional volatility of residuals. If we assume homogeneous residual variances across stocks, we could also derive that cross-sectional volatility of residuals is a consistent and asymptotically efficient estimator for idiosyncratic volatility as in below.

$$\begin{aligned} CSV(\varepsilon_{it}) &= \sum_{i=1}^N \omega_{it} (\varepsilon_{it} - \sum_{j=1}^N \omega_{jt} \varepsilon_{jt})^2 \\ &= \sum_{i=1}^N \omega_{it} \varepsilon_{it}^2 + (\sum_{j=1}^N \omega_{jt} \varepsilon_{jt})^2 - 2 \sum_{i=1}^N \sum_{j=1}^N \omega_{it} \omega_{jt} \varepsilon_{it} \varepsilon_{jt} \\ &= \sum_{i=1}^N \omega_{it} \varepsilon_{it}^2 - (\sum_{j=1}^N \omega_{jt} \varepsilon_{jt})^2 \end{aligned} \quad (3.6)$$

Following Garcia, Mantilla- Garcia, and Martellini (2012)'s results, we know that  $\sum_{j=1}^N \omega_{jt} \varepsilon_{jt} \rightarrow 0$

almost surely and  $\sum_{i=1}^N \omega_{it} \varepsilon_{it}^2 \rightarrow \sigma_\varepsilon^2$  almost surely. So  $CSV(\varepsilon_{it}) \rightarrow \sigma_\varepsilon^2$  almost surely.

At this juncture we introduce the cross-sectional volatility of residuals as a volatility index. It provides several advantages over the indices reviewed above. For one, it is a non-parametric, model-free estimator. In addition, it could be estimated at any frequency for any region, sector, and style of world equity markets. Most importantly, it does not depend on any option pricing, and therefore can be applied to stocks that do not have options available.

When all the stocks are the same, the cross-sectional volatility of residuals and the cross-sectional volatility of returns are all zero.

### **3.3. Empirical Studies**

#### **3.3.1. Data**

Our sample consists of all stocks included in the S&P 500 index from January 1990 to December 2011, although S&P 500 stocks have longer data. Daily stock return and stock price are obtained from the Center for Research in Securities Prices database (CRSP). We start our study from 1990 to match the volatility index data, and we obtained a daily time series of the VIX from 1990-2011 from Chicago Board Options Exchange website. As we only have we have VOX data for before 2003, which is the VIX using the old methodology, we combine this with the post-2003 VIX data to get a longer time series.

#### **3.3.2. Empirical Studies of VIX**

We first look at some properties of VIX. Figure 3.1, panel A compares the volatility of the S&P 500 and the VIX from 1990-2011. VIX is the ticker symbol for the Chicago Board Options Exchange Market Volatility Index, a popular measure of the implied volatility of S&P 500 index options. Often referred to as the fear index or the fear gauge, it represents one measure of the market's expectation of stock market volatility over the next 30 day period. To compare, the volatility of S&P 500 is the volatility over a year and it is calculated on a rolling basis. Since there is only one data point change each time, the volatility of S&P 500 index is much smoother. The correlation between the volatility of the S&P 500 and the VIX is a high 0.68. Using the

volatility of S&P 500 for the past 30 days, the generates an even higher correlation at 0.79. This relation is shown in panel B of figure 3.1.

We further investigate how return is related to volatility. Figure 3.2, Panel A shows the time series plot of S&P 500 returns and the changes of VIX. There's a clear negative correlation at -0.49. This is expected as volatility increases when the market is falling and vice versa. If the VIX, or its ETF the VXX, were to be at a raised level, there's a strong chance that the S&P 500 index would be very low. Panel B shows a slight negative correlation between volatility levels and stock returns. If we choose a future date such that it is past volatility change and future return, there's no correlation. Thus it seems to be mainly a contemporaneous pattern applied to changes in volatility. The rise in volatility is not predicting price moves, just reflecting them. This is exemplified by the high volatility correlated with both the large downturn and the large rebound in the latter part of 2009.

### **3.3.3. Empirical Studies of CSV**

We examine next the properties of the cross-sectional volatility, beginning with the correlations. Active asset managers currently use the average correlation as an indicator for investment opportunity. In a low correlation environment, active asset management is provided with more opportunities, and diversification is one way to reduce investment risk. Diversification relies on the lack of a tight positive relationship among the assets' returns.

There are some problems with the average correlation. First, it does not evolve the weights of the stocks in the index in the calculation, giving all the stocks the same importance. Second, it is estimated on a rolling basis. As each new data point is added, most of the data is identical to the prior observation.

We first look at how correlation changes through time. We calculate correlations for each pair of stocks in each month, and then calculate the average of these pairwise correlations. Figure 3.3, panel A shows the average monthly correlations and monthly volatility. We can see that the average correlations and average volatility vary with time and both the average correlations and the average volatilities have increased recently.

We show that the cross-sectional volatility only as a function of two time series based variables: the average monthly time series volatility of each stock, as measured by  $\sigma$ , and the average monthly pairwise correlations between stocks, as measured by  $\rho$ . The cross-sectional volatility is positively related then to volatility and inversely related to the correlation. Figure 3.3 panel B compares the monthly cross-sectional volatility with  $\sigma\sqrt{1-\rho}$ . This illustrates that they have similar pattern, and correlation between them is as high as 0.85.

Up to this point, we have discussed the correlations between total returns. Here, we would like to impose the correlations between common factors and the correlations between residual returns. We will extract common factors from stock returns first and then calculate the cross-sectional volatility for the common returns and the cross sectional volatility for the residual returns.

From the formula, we see that the cross-sectional volatility of the common return depends on the cross-sectional volatility of the betas, which are factor loadings from the time series regressions of stock returns on factors. We know these factor loadings are persistent over time so that the cross-sectional volatility of beta does not change much over time. The cross-sectional volatility of the common return also depends on the common factors, the first of which is highly correlated with the market return. So, opportunities for active portfolio managers to manage the portfolio come from the dispersion of the return. These dispersions in turn mainly originate from

the cross-sectional volatility of the residual return. A high cross-sectional volatility of the residual return provides money managers with a great opportunity of alphas.

We now investigate the relationship between VIX and the cross-sectional volatility. Figure 3.4 compares the VIX and the cross-sectional volatility of the returns and the cross-sectional volatility of residuals. The correlation between VIX and the cross-sectional volatility of returns is 0.18. The correlation between VIX and the cross-sectional volatility of residuals is 0.62. We see then that the cross-sectional volatility of residuals is more representative to the dispersion of the stocks and can use it then as a volatility index.

#### **3.3.4. An Application of the Cross-Sectional Volatility of Residuals**

The goal of an asset manager is to find a combination of assets that reduces risk without significantly affecting portfolio returns. Combining assets that are negatively correlated with one another can provide tremendous diversification benefits, as that ultimately reduces the volatility of the portfolio returns. In the search for passive investments that reduce the risk of the overall portfolio without significantly affecting returns, it appears that volatility indices may offer a solution.

Volatility indices are a tradable asset class. Numerous studies have found that adding volatility as a separate asset class to an S&P 500 portfolio reduces risk without significantly affecting return. The VIX, for example, can serve as a very powerful diversification tool and risk management aid. Moran and Dash (2007) compared a portfolio made up of 100% equities to one composed of 95% equities and 5% VIX from 1990 through March 2007. They discovered that the 5% VIX allocation lowered the overall volatility of the portfolio by 92 basis points and increased the Sharpe ratio. Total return was only reduced by six basis points. Daigler and Rossi

(2006) also examined the risk-return profile of an S&P 500-volatility portfolio compared to that of an S&P 500 only portfolio. They analyzed data from 1992 through 2002 and concluded that adding VIX to the equity-only portfolio significantly reduced risk without a consequential affect on return.

In this study, we use the S&P 500 index return as a benchmark and construct two more portfolios. The VIX portfolio invests 95% on S&P 500 index and 5% on VIX, while the CSV portfolio invests 95% on S&P 500 index and 5% on the cross-sectional volatility index of residuals. The results in table 3.2 show that the average daily return for S&P 500 index is 0.089% and the average daily return for the VIX portfolio is 0.094%. The average daily return for the CSV portfolio is 0.231%. The VIX portfolio does not only have higher return than S&P 500 Index, it also has lower volatility, giving the VIX portfolio a higher Sharpe ratio in addition. The CSV portfolio has much higher return than both S&P 500 index and VIX portfolio and it has slightly higher volatility. Therefore, the CSV portfolio has the highest Sharpe ratio.

This gives us an application for the cross-sectional volatility index of residuals. Crucially, we could form this volatility index for any universe of stocks and at any frequency. This new application is a great tool for diversification and risk management.

### **3.4. Conclusion**

In this chapter, we propose a new volatility index, a cross-sectional one of residuals. The cross-sectional volatility index moves closely with VIX for the S&P 500 stock universe. It is a non-parametric, model free volatility index, can be estimated at any frequency for any region, sector, and style of world equity markets, and does not depend on any option pricing.

We provide some interpretation of the cross-sectional volatility index of residuals as a proxy for aggregate economic uncertainty. Our report of the high correlation between the VIX index and the corresponding cross-sectional volatility index of residuals based on the S&P 500 universe confirms this intuition. Our findings also indicate that the portfolio that hedged based on the cross-sectional volatility index of residuals has a much higher Sharpe ratio than the portfolio without hedge. These results suggest that the cross-sectional volatility index of residuals can be used as a reliable proxy for volatility when volatility measures are not available.

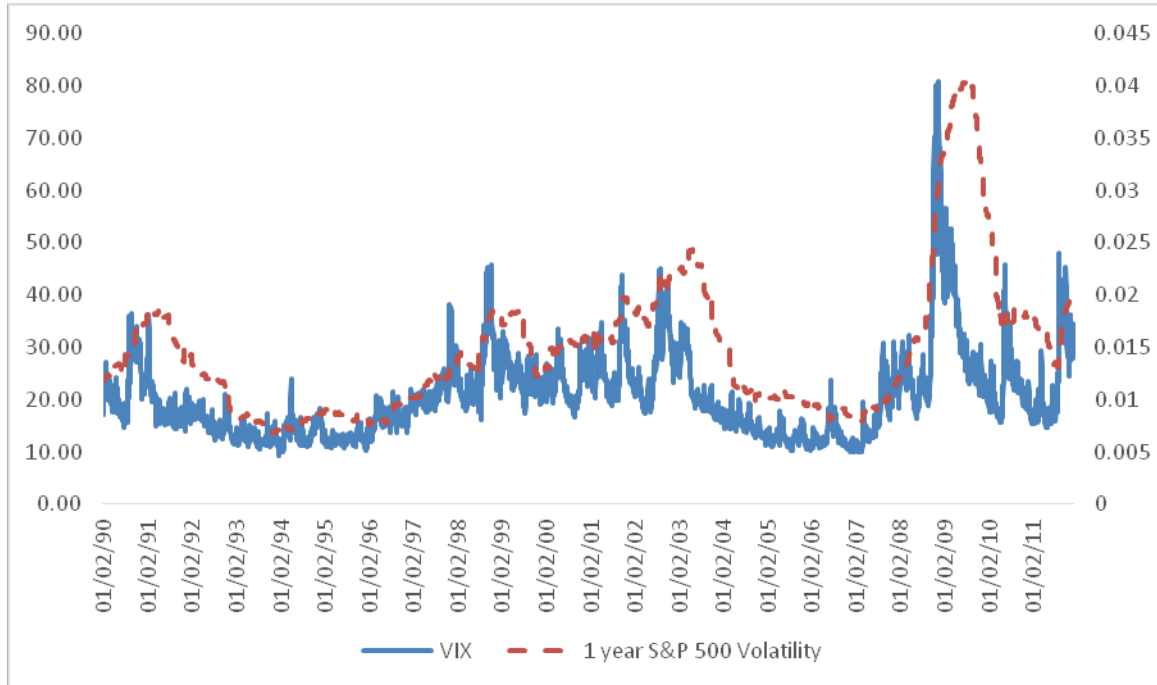
Table 3.1: Currently Available Volatility Indices

Index Name	Underlying Asset	Equity Market	Index Provider
VIX	S&P 500	US	CBOE
VXD	DJIA	US	CBOE
R VX	Russell 2000	US	CBOE
VXN	Nasdaq 100	US	CBOE
VXV	S&P 500	US	CBOE
S&P/TSX 60 VIX	S&P/TSX 60	Canada	TMX group
VDAX-NEW	DAX	Germany	Deutsche Börse
VSMI	SMI	Switzerland	SIX Swiss Exchange
VSTOXX	Dow Jones EURO STOXX 50	Eurozone	STOXX Ltd
VAEX	AEX	Netherlands	NYSE Euronext
VBEL	BEL 20	Belgium	NYSE Euronext
VCAC	CAC 40	France	NYSE Euronext
VFTSE	FTSE 100	UK	NYSE Euronext
VIBEX-NEW	IBEX 35	SPAIN	
GRIV	FTSE/ATHEX-20	GREECE	
VKOSPI	KOSPI 200	Korea	Korea Exchange KRX
India VIX	NIFTY	India	National Stock Exchange of India
VHSI	HSI	Hong Kong	Hang Seng Indexes Company
VXJ	Nikkei 225	Japan	
S&P/ASX 200 VIX	S&P/ASX 200	Australia	Australian Securities Exchange
New SAVI	FTSE/JSE Top40	South Africa	Johannesburg Stock Exchange

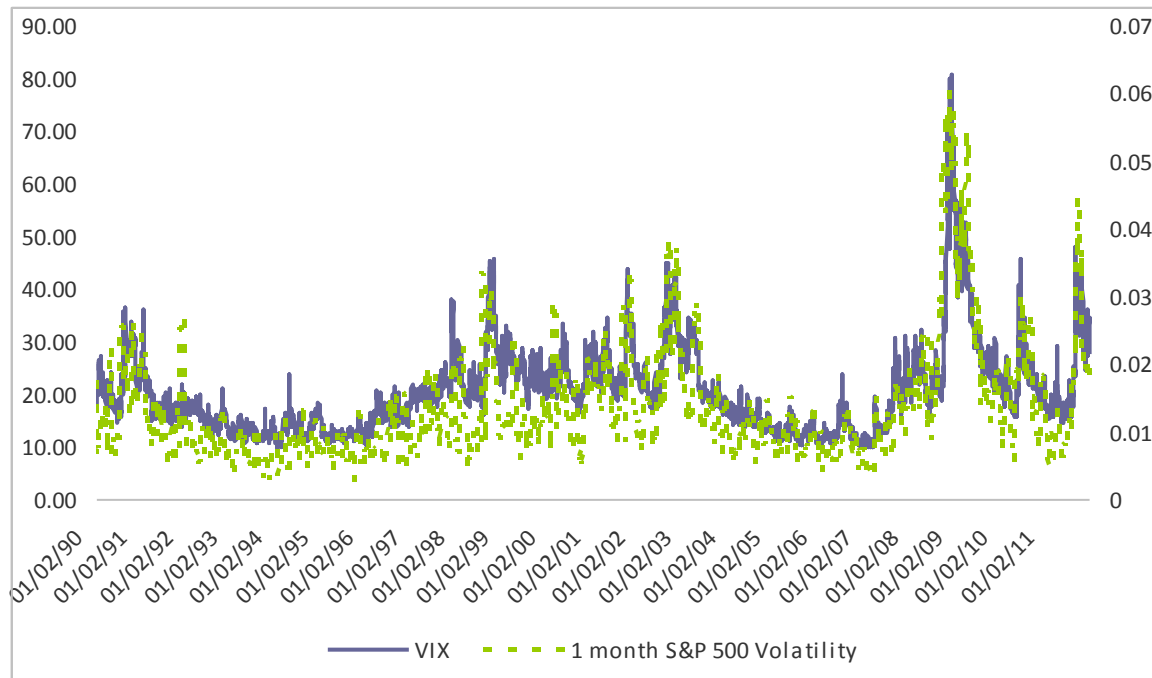
Table 3.2: Daily Return and Volatility of the S&amp;P 500 Index, VIX Portfolio and CSV Portfolio

	Return	Volatility	Sharpe Ratio
S&P 500 Index	0.089%	1.64%	5.39%
VIX Portfolio	0.094%	1.43%	6.54%
CSV Portfolio	0.231%	1.93%	11.94%

Figure 3.1: Time Series Volatility of the S&P 500 vs. the VIX, 1990-2011.

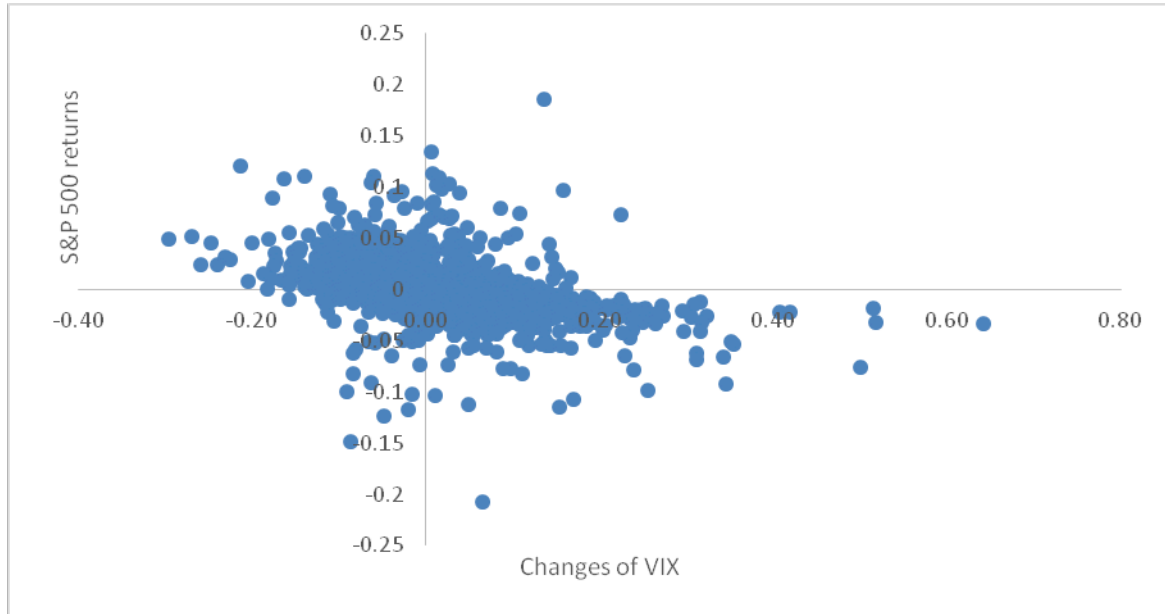


Panel A: 1 year rolling Volatility of the S&P 500

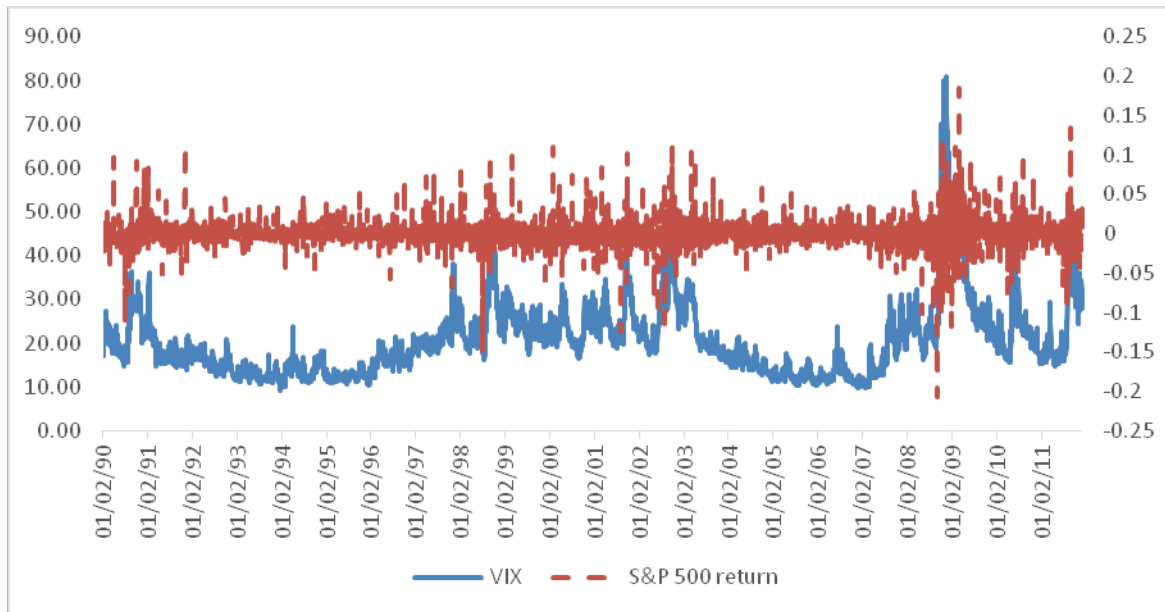


Panel B: 30 days rolling Volatility of the S&P 500

Figure 3.2: Time Series Plot of the Return of the S&amp;P 500 vs. the VIX, 1990-2011.

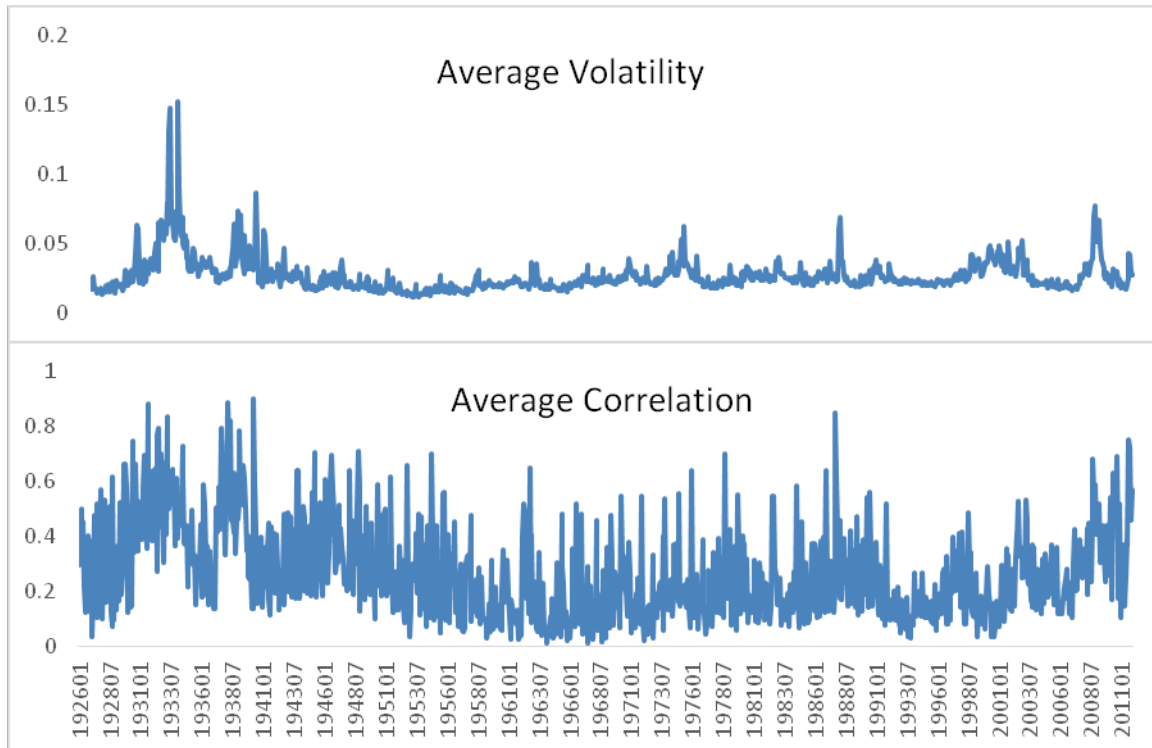


Panel A: Time Series Plot of the Return of the S&amp;P 500 vs. the changes of VIX, 1990-2011.

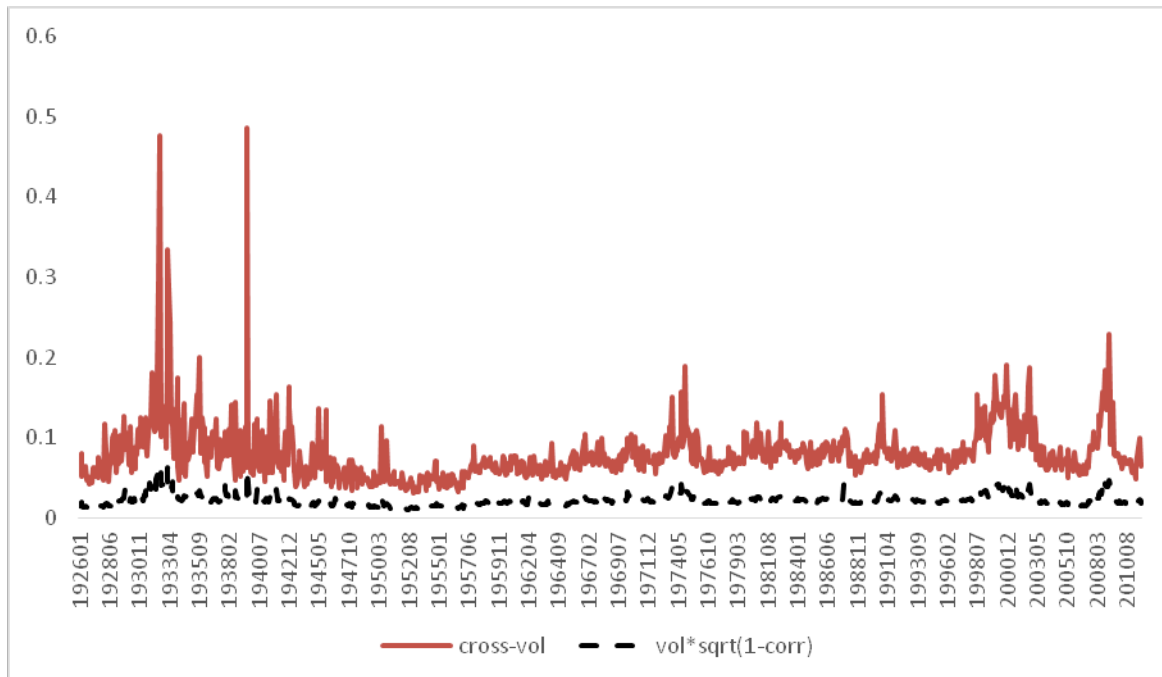


Panel B: Time Series Plot of the Return of the S&amp;P 500 vs. the levels of VIX, 1990-2011.

Figure 3.3: Average Correlation, Average Volatility and the Cross-Sectional Volatility from Jan. 1926 to Nov. 2011.

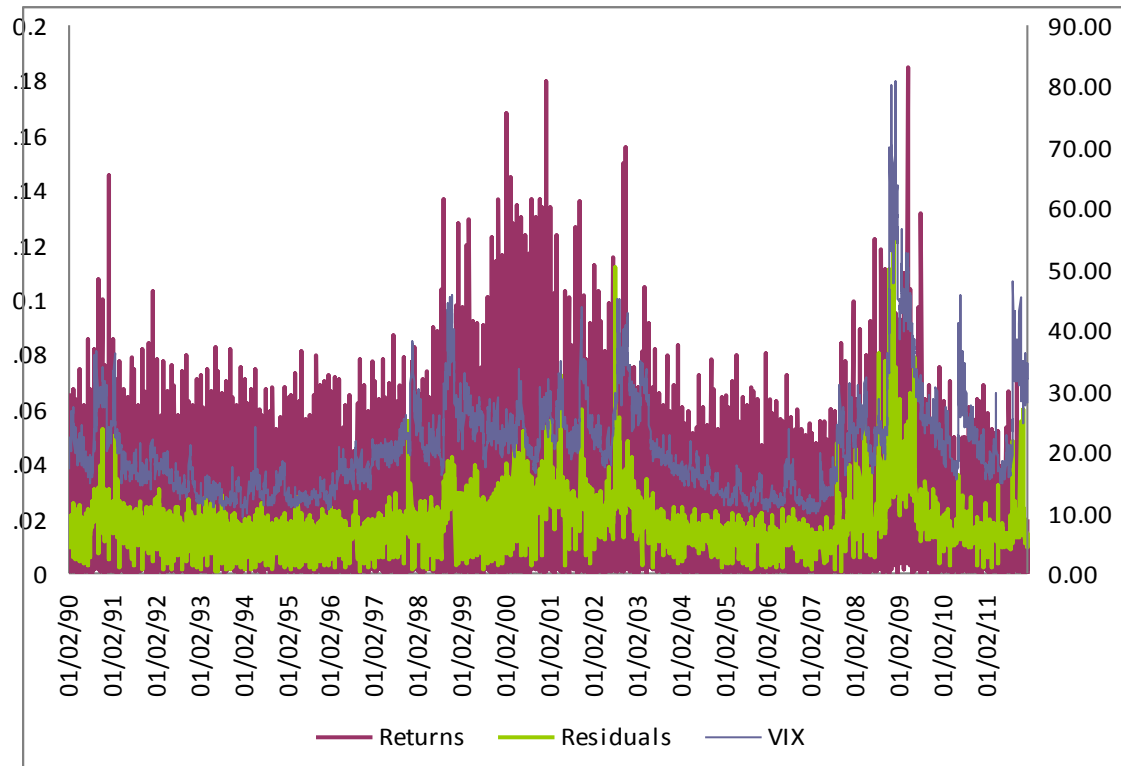


Panel A: Average monthly correlations and average monthly volatility



Panel B: Cross-Sectional Volatility and the average correlation and average volatility

Figure 3.4: Time Series Plot of the CSV vs. the VIX, 1990-2011.



### Bibliography

1. Andersen, T., T. Bollerslev, F. Diebold and H. Ebens (2001). "The Distribution of Realized Stock Return Volatility," *The Journal of Financial Economics*, 61, 43-76.
2. Andersen, T., T. Bollerslev, F. Diebold and P. Labys (2003). "Modeling and Forecasting Realized Volatility," *Econometrica*, 71(2), 579-626.
3. Audrino, F., and Corsi, F. (2007). "Realized Correlations Tick-By-Tick," Working paper, University of St. Gallen.
4. Bai, J. and S. Ng (2002). "Determining the number of factors in approximate factor models," *Econometrica*, 70(1), 191 - 221.
5. Barndorff-Nielsen, O.E., and N. Shephard (2004). "Econometric Analysis of Realized Covariation: High Frequency Based Covariance, Regression, and Correlation in Financial Economics," *Econometrica*, 73(3), 885-926.
6. Barndorff-Nielsen, O. E. and N. Shephard (2005). "Power variation and time change," *Theory of Probability and Its Applications*, 50, 1-15.
7. Black, F. (1976). "The pricing of commodity contracts," *The Journal of Financial Economics*, 3(1), 167-179.
8. Carr, P. and R. Lee (2009). "Volatility Derivatives," *The Annual Review of Financial Economics*, 1, 1-21.
9. Chamberlain, G. and Rothschild, M. (1983). "Arbitrage, factor structure and mean-variance analysis in large asset markets," *Econometrica*, 51, 1305-1324.
10. Chan, L., J. Karceski, and I. Lakonishok (1999). "On portfolio optimization: forecasting covariance and choosing the risk model," *Review of Financial Studies*, 12, 937-974.
11. Christie, A. (1982). "The stochastic behavior of common stock variances: value, leverage and interest rate effects," *The Journal of Financial Economics*, 10(4):407-432.
12. Cohen, K. and J. Pogue (1967). "An empirical evaluation of alternative portfolio section models," *The Journal of Business*, 40, 166-193.
13. Connor, G. and R. Korajczyk (1986). "Performance measurement with the arbitrage pricing theory: A new framework for analysis," *The Journal of Financial Economics*, 15, 373-394.
14. Connor, G. and R. Korajczyk (1988). "Risk and return in an equilibrium APT: Application of a new test methodology," *The Journal of Financial Economics*, 21, 255-289.

15. Corsi, F. (2009). "A Simple Approximate Long-memory Model of Realized Volatility," *The Journal of Financial Econometrics*, 7, 174–196.
16. Daigler, R. and L. Rossi (2006). "A Portfolio of Stocks and Volatility," *The Journal of Investing*, 15(2), 99-108.
17. de Pooter, M., M. Martens and D. van Dijk (2005). "Predicting the Daily Covariance Matrix for S&P 100 Stocks Using Intraday Data - But Which Frequency to Use?," Tinbergen Institute Discussion Papers.
18. Demeterfi, K., E. Derman, M. Kamal and J. Zou (1999). "More than you ever wanted to know about volatility swaps," Goldman Sachs Quantitative Strategies Research Notes.
19. Dumas, B., J. Fleming J and R. Whaley (1998). "Implied volatility functions: Empirical tests," *The Journal of Finance*, 53(6), 2059-2106.
20. Elton, E. and M. Gruber (1973). "Estimating the dependence structure of share prices-implications for portfolio selection," *The Journal of Finance*, 28, 1203-1232.
21. Elton, E., M. Martin, S. Brown, and W. Goetzmann (2007). "Modern portfolio theory and investment analysis," *Wiley*, 7th edition.
22. Garcia, R., D. Mantilla-Garcia, and L. Martellini (2012). "Aggregate Idiosyncratic Volatility and the Cross-Sectional Variance of Stock Returns," EDHEC working paper.
23. Giot, P. (2005). "Relationships between implied volatility indexes and stock index returns," *The Journal of Portfolio Management*, 31(3), 92-100.
24. A. Gittens and J. A. Tropp (2011). "Tail bounds for all eigenvalues of a sum of random matrices," working paper.
25. González, M. and A. Novales (2009). "Are volatility indices in international stock markets forward looking?," *RACSAM - Revista de la Real Academia de Ciencias Exactas, Fisicas y Naturales. Serie A. Matematicas*, 103(2), 339-352.
26. Gorman, L., S. Sapra and R. Weigand (2010). "The Cross-Sectional Dispersion of Stock Returns, Alpha, and the Information Ratio," *The Journal of Investing*, 19(3), 113-127.
27. Jones, C. S. (2001). "Extracting factors from heteroskedastic asset returns," *The Journal of Financial Economics*, 62, 293-325.
28. Kapetanios, G. (2010). "A Testing Procedure for Determining the Number of Factors in Approximate Factor Models With Large Datasets," *The Journal of Business & Economic Statistics*, 28(3), 397-409.
29. Ledoit, O. and M. Wolf. (2003). "Improved estimation of the covariance matrix of stock returns with an application to portfolio selection," *The Journal of Empirical Finance*, 10, 603-621.

30. Ledoit, O. and M. Wolf. (2004). "Honey, I Shrunk the Sample Covariance Matrix," *The Journal of Portfolio Management*, 30(4), 110-119.
31. Markowitz, H. (1952). "Portfolio Selection," *Journal of Finance*, 7, 177-91.
32. Moran, M. and S. Dash (2007). "VIX Futures and Options: Pricing and Using Volatility Products to Manage Downside Risk and Improve Efficiency in Equity Portfolios," *The Journal of Trading*, 2(3), 96-105.
33. Onatski, A. (2009). "Testing Hypotheses About the Number of Factors in Large Factor Models," *Econometrica*, 77 (5), 1447-1479.
34. Psychoyios, D. and G. Skiadopoulos (2006). "Volatility options: Hedging effectiveness, pricing, and model error," *The Journal of Future Market*, 26(1), 1-31.
35. Rubinstein, M. (1994). "Implied binomial trees," *The Journal of Finance*, 49(3), 771-818.
36. Sarwar, G. (2012). "Is VIX an investor fear gauge in BRIC equity markets?," *The Journal of Multinational Financial Management*, 22(3), 55-65.
37. Schwert, G. (1989). "Why does stock market volatility change over time?," *The Journal of Finance*, 44(5), 1115-1154.
38. Schwert, G. (1990). "Stock volatility and the crash of 1987," *Review of Financial Studies*, 3(1), 77-102.
39. Sharpe, W. (1963). "A simplified model for portfolio analysis," *Management Science*, 9, 277-293.
40. Simon, D. (2003). "The Nasdaq volatility index during and after the bubble," *The Journal of Derivative*, 11(2), 9-24.
41. Siriopoulos, C. and A. Fassas (2008). "The information content of VFTSE," Working paper (available at SSRN: <http://ssrn.com/abstract=1307702>).
42. Siriopoulos, C. and A. Fassas (2012). "An investor sentiment barometer-Greek implied volatility index (GRIV)," *Global Finance Journal*, 23(2), 77-93.
43. Tobin, J. (1958). "Liquidity preference as behavior towards risk," *Review of Economic Studies*, 67, 65-86.
44. Whaley, R. (2000). "The investor fear gauge," *The Journal of Portfolio Management*, 26(3), 12-17.
45. Whaley, R. (2009). "Understanding the VIX," *The Journal of Portfolio Management*, 35(3), 98-105.

**Vita**

Xiaolin Sun received a Bachelor of Science in Mathematics from University of Science and Technology of China and a Master of Science in Statistics from Syracuse University. She also holds a Master of Arts in Economics and a Master of Business Administration from University of Washington. She is expected to obtain her Ph.D. in Economics in December 2013, with a focus on Financial Economics.