Modeling Systemic Risk in The Options Market

by

Doris Dobi

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Mathematics

New York University

May 2014

Professor Marco Avellaneda

To my family, old and new

Acknowledgements

Isaac Newton once said "If I have seen further than others, it is by standing upon the shoulders of giants". This has resonated a profound truth with me, and does even more so at this moment. This thesis would not have been possible without the guidance of my advisor Marco Avellaneda. Thank you Marco, but not only for the technical tools and mathematics, but more so for being so open with me regarding all facets of academic life and life in general. You are one of a kind, and it has truly been my privilege to work under your guidance.

I would like to thank the Courant mathematics department for all of the freedom to explore it offers and for the high-caliber people to be found here. Thank you for keeping your doors open and for sharing your thoughts. Thank you to all of the friends I've made. You've made the ride so much more enjoyable, and it's been an honor to have known you all; godspeed ahead.

Thank you Mom and Dad for the absolute unconditional love and support. You are true Giants in so many ways. Thank you Kledin for all of the hard and timeless lessons. Thank you to my new family: Igor, Luda, and Anton for your kindness and support. Thank you to my husband Oleg; you remind me every day of what we are working towards, and I could not have asked for a better companion for the journey.

Abstract

This thesis provides a cross-sectional classification of optionable equities in U.S. Markets based on implied volatility data. For each security in the OptionMetrics database for which there is data available over the period from August 2004 to August 2013 we model its implied volatility surface (IVS). We then use the spectrum of the IVS, in particular the leading eigenvalue, to classify options into those carrying mostly systemic risk and into those carrying mostly idiosyncratic risk.

We use implied volatility data across 13 different deltas and 4 expiration dates, hence our data on the options market is 52 times bigger than that of the equities market. By employing methods from principal component analysis (PCA), and results from random matrix theory (RMT), we classify the significant eigenvalues and conclude that usually three principal components suffice to reproduce the IVS. In this way we reduce dimensionality without loosing any meaningful information.

Using these results, we formulate an explicit model which can be used to model the dynamics of the IVS, yet is compact and computationally feasible. Out of the original 52 implied volatility pivots we narrow our model to use only 9-pivots, offering roughly 6 times variable reduction from the original 52-pivots, and over 14 times variable reduction from the 130 implied volatility returns available in OptionMetrics.

We conclude by performing a PCA of the entire equities and options market as made available by OptionMetrics. We begin with the markets as determined by constituents of SPX and their options. In order to determine the significant part of the spectrum, in addition to PCA and Marchenko-Pastur, we also employ the Tracy-Widom Law. Before we do so, we first test the applicability of both Marchenko-Pastur as well as Tracy-Widom on random matrixes with the same

underlying distribution as our empirical data.

The results from using such random matrixes are in excellent agreement with those predicted by both Marchenko-Pastur and by Tracy-Widom. In particular, the underlying distribution of our empirical data[1] forms an ensemble class on which Tracy-Widom applies. Utilizing all of the aforementioned tools, we classify the number of significant factors driving the U.S. equities and options market as follows:

1. Equities in SPX: **15** significant factors (account for 55% of variance).

2. Options with underlying in SPX: **84** significant factors (account for 55% of variance).

3. All assets in OptionMetrics: **20** significant factors (account for 24% of variance).

4. All options with underlying asset in OptionMetrics: **108** significant factors (account for 50% of variance).

[1]Our underlying data exhibits a student-t distributions with degrees of freedom between 4 and 5.

Contents

Dedication iii

Acknowledgements iv

Abstract v

List of Figures xiii

List of Tables xv

Introduction 1

1 Principal Component Analysis of Implied Volatility Surfaces 3

1.1 Quick Overview of Implied Volatility and Implied Volatility Surfaces 3

1.2 Implied Volatility 6

1.3 Preliminary Analysis of Implied Volatility Coupled with the Underlying Asset 15

1.4 Secondary Analysis of Implied Volatility Coupled with the Underlying Asset 19

1.5 Systemic vs. Idiosyncratic Risk 26

1.6 S&P500 constituents by industry sector and other groupings 43

1.7 Analysis of options on the remaining underlying assets 48

1.8 Conclusion 52

2 The Pivot Method and Determining Significant Components via Random Matrix Theory **54**

2.1 Overview of results and goal of following sections 54

2.2 Stability of the first eigenvalue across time 56

2.3 Pivot Method Description and Statistical Analysis of Pivot Method Quality 62

2.4 PCA of the entire options market: tying things together 74

2.5 Conclusion 100

A Principal Component Analysis of Implied Volatility Surfaces **103**

B The Pivot Method and Determining Significant Components via Random Matrix Theory **116**

Bibliography **118**

List of Figures

1.1 Reconstructed Implied Volatility Surface for SPX 5

1.2 Example of Implied Volatility Surface Dynamics for SPX 7

1.3 Analysis of the spectrum of the correlation matrix for SPX 9

1.4 Analysis of the eigenvectors of the correlation matrix for SPX . . . 10

1.5 Analysis of the spectrum of the correlation matrix for SPX 11

1.6 Analysis of the eigenvectors of the correlation matrix for SPX . . . 12

1.7 Principal Component Surfaces for SPX 14

1.8 Spectrum of the correlation matrix for the implied volatilities of SPX coupled with the underlying asset 17

1.9 Eigenvectors of the correlation matrix for the implied volatilities of SPX coupled with the underlying asset 18

1.10 Principal components for SPX, standardizes and coupled with SPX 20

1.11 Principal components for IWM, standardizes and coupled with IWM 21

1.12 Principal components for AAPL, standardizes and coupled with AAPL 22

1.13 Principal components for VXX, standardizes and coupled with VXX 23

1.14 Significant eigenvalues á la Random Matrix Theory 27

1.15 First eigenvalue for the top 20 most liquid ETFs 28

1.16 First three eigenvalue for the top 20 most liquid ETFs 29

1.17 Number of eigenvalues greater than the MP-threshold for the top 20 most liquid ETFs 30
1.18 First eigenvalue for 6 volatility ETFs 31
1.19 Top three eigenvalues for 6 volatility ETFs 32
1.20 Number of eigenvalues greater than the MP-threshold for 6 volatility ETFs 33
1.21 First and second eigenvalues of the constituents of the S&P500 . . . 35
1.22 Percent of variation explained by the first three principal components of the constituents of the S&P500 36
1.23 Number of eigenvalues exceeding the MP upper limit of the spectrum for the constituents of the S&P500 37
1.24 Correlations 38
1.25 Correlation continued 47
1.26 First and second eigenvalues of 3300 remaining underlying assets in the OptionMetrics 49
1.27 Percent of variation explained by the first three principal components of 3300 remaining underlying assets in the OptionMetrics . . 50
1.28 Number of eigenvalues exceeding the MP upper limit of the spectrum for all assets in the OptionMetrics 51

2.1 Dynamic PCA Results for GS 57
2.2 Dynamic PCA Results for JPM 58
2.3 Dynamic PCA Results for WEC 59
2.4 Dynamic PCA Results for POM 60
2.5 Dynamic PCA Results for the SPX index 61

2.6 Example Schema of linear interpolation amongst the pivots via projection onto δ-space and time-to-expiration-space. In this schema we use 9-pivots, and the point we wish to replicate is denoted by an X. 64

2.7 Difference between the first four eigenvalues of the 2-pivot model and those of the original spectra. Note that the x-axis is the list of all constituents of SPX in increasing order of their first eigenvalue as computed from the original data using all 52 implied volatility returns. 66

2.8 Difference between the first four eigenvalues of the 6-pivot model and those of the original spectra. Note that the x-axis is the list of all constituents of SPX in increasing order of their first eigenvalue as computed from the original data using all 52 implied volatility returns. 67

2.9 Difference between the first four eigenvalues of the 9-pivot model and those of the original spectra. Note that the x-axis is the list of all constituents of SPX in increasing order of their first eigenvalue as computed from the original data using all 52 implied volatility returns. 68

2.10 Difference between the first four eigenvalues of the 12-pivot model and those of the original spectra. Note that the x-axis is the list of all constituents of SPX in increasing order of their first eigenvalue as computed from the original data using all 52 implied volatility returns. 69

2.11 Original distribution using all 52-pivots. 71

2.12 Original distribution using all 52-pivots vs. the distribution produced by each model indicated. 73
2.13 Top 20 eigvals SPX. 76
2.14 Top 20 eigvals SPX. 78
2.15 Random spectrum SPX. 79
2.16 Random eigenvalues SPX. 80
2.17 Random eigenvalues SPX. 81
2.18 random eigenvalues SPX. 82
2.19 Top 20 eigvals options on SPX. 84
2.20 Top 6 eigvecs options on SPX. 85
2.21 Top 20 eigvals options on SPX. 86
2.22 Random eigenvalues options on SPX. 87
2.23 Random eigenvalues options SPX. 88
2.24 Random eigenvalues options on SPX. 89
2.25 Top 25 eigvals all equities. 91
2.26 Random eigenvalues all equities. 92
2.27 Random eigenvalues all equities. 93
2.28 Random eigenvalues all equities. 94
2.29 Top 25 eigvals all options. 96
2.30 Random eigenvalues all options. 97
2.31 Random maximum eigenvalues all options. 98
2.32 TW vs. empirical entire options. 99

A.1 VXX PCs. 105
A.2 VXZ PCs. 106
A.3 AAPL PCs. 107

A.4 NFLX PCs. 108
A.5 BIIB PCs. 109
A.6 CVS PCs. 110
A.7 HOT PCs. 111
A.8 HD PCs. 112
A.9 FB PCs. 113
A.10 DAL PCs. 114
A.11 GLD PCs. 115

B.1 Original distribution using all 52-pivots vs. the distribution produced by each model indicated. 117

List of Tables

1.1 Correlation Matrix of the constituents of the S&P500 39

1.2 Top 15 constituents and bottom 15 constituents by first eigenvalue. 42

1.3 Top 15 constituents and bottom 15 constituents by first eigenvalue. 42

1.4 Results for the constituents of the S&P500 partitioned by industry sector 44

1.5 Results for the constituents of the S&P500 partitioned by industry sector 44

1.6 Top 10 constituents of the S&P500 representing value, growth and index weight 45

1.7 Correlation matrix of all remaining remaining assets (around 3300) 46

1.8 Average values across all assets and those of the bottom and top 1 percent 48

1.9 Results for all remaining assets partitioned by industry sector . . . 52

1.10 Results for all remaining assets partitioned by industry sector . . . 52

2.1 Implied Volatilities Used as Pivots for Each Model 63

2.2 Significance of Eigenvalues in the S&P500 Equities Market a la TW 83

2.3 Significance of Eigenvalues for Options with an underlying in S&P500 90

2.4 Significance of Eigenvalues in the entire Equities Market a la TW . 95

2.5 Significance of Eigenvalues in the entire Options Market a la TW . 100

Introduction

There have been many studies which seek to classify the number of factors needed to model the equities market. Principal component analysis shows that around 15 components can be used to capture the vast majority of market movements. On the other hand, not as much work has been done on classifying the important driving factors underlying the change of option prices. We show that the implied volatility surface of most individual option contracts can be described using four components. In addition, the leading eigenvalue can be used to characterize the type of risk intrinsic to a specific option contract. Furthermore, we build a model which can be used to reproduce the implied volatility surface of an option while providing dimensionality reduction in the number of variables needed.

The options market is forward-looking in time and thus options traders face the risk of known unknowns as well as unknown unknowns. One can hedge a position against known unknowns with appropriate strategies, but it is unclear how to risk manage a portfolio against unknown unknowns. In order to better understand the risk intrinsic to an option position and to better hedge the unknowns, we classify options into those that carry mostly systemic risk and into those that carry mostly idiosyncratic risk.

Due to the fact that there are many available strikes and expiration dates for a given option contract, the option market is considerably larger than the equities market. Because of this, a large part of our work consists of data-mining and working with and manipulating big-data. Our data on the options market is roughly 52 times bigger than that of the equities market. We seek to reduce dimensionality in a meaningful way without loosing information. In Chapter 1, we begin to accomplish this by classifying the "significant" part of the spectrum

using principal component analysis (PCA) and results from random matrix theory (RMT). Studying the distribution of these resulting eigenvalues allows us to classify an asset according to the type of risk it carries. In particular, we discover that the leading eigenvalue of the correlation matrix plays a critical role in our classification scheme. In this way, RMT and PCA allow us to build the basis of a dimensionality reduction model.

In Chapter 2, we extend the work of Chapter 1 by using dynamic PCA with a moving window of 252-days to see how the significant part of the spectrum varies across time. In addition, we also perform a dynamic analysis of the number of eigenvalues which exceed the MP-bound across time. In the last section we construct, test, and recommend various dimensionality-reducing models for the implied volatility surface.

Using these results, we give two ways of capturing the movement of the implied volatility surface; one using its corresponding principal component surfaces (Chapter 1), and another using what we refer to as the "pivot method" (Chapter 2). Our choice for a pivot model uses only 9 out of the 52 implied volatility points and interpolates between these points to generate the original data. We test these methods by examining how well they preserve the significant part of the spectrum. We further test each pivot model by examining how well each model preserves the original distribution across different risk-classes. Both methods offer significant dimensionality reduction when modeling the implied volatility surface. Furthermore, the pivot model provides a firm basis for a risk-management system for portfolios of options.

Chapter 1

Principal Component Analysis of Implied Volatility Surfaces

In this chapter, after a quick review of implied volatility and implied volatility surfaces, we produce the principal component surfaces for various assets. By analysing the resulting spectrum we classify the significance of these surfaces in describing the movement of the implied volatility surface. Utilizing the Marchenko-Pastur distribution from random matrix theory, we further classify the spectrum into significant components and use principal component analysis to determine the explanatory power of each significant component. The tools we develop along the way are used to classify the type of risk inherent in the options of a given asset.

1.1 Quick Overview of Implied Volatility and Implied Volatility Surfaces

An option contract is usually specified by

- The underlying asset (S)
- Type (put or call)
- Strike price (K)
- Notional amount
- Time to expiration (or maturity, T)
- Style (European or American)
- Settlement (cash or physical)

One of the most successful option-pricing models is the Black-Scholes (BS) model which gives the value of the option as a function of its strike, spot, maturity, risk-free rate of return, implied dividend yield, and finally the implied volatility of the underlying asset: BS(K,S,T,r,q,σ).

The implied volatility of an option contract is the value of the volatility of the underlying stock which makes a theoretical option-valuation model (e.g., Black-Scholes) equal to the observed market price for the option contract. For many market practitioners the market price of an option and its implied volatility are interchangeable; many quote the option price in terms of its implied volatility.

It has long since the original formulation of the Black-Scholes model been observed that the implied volatility $\sigma(K,T)$ of an option contract depends on both K and T, and hence is not constant in (K,T)-space as the model predicts. For example, consider the implied volatility surface below for SPX constructed from empirical data for call options.

Clearly the volatility surface is not constant in neither K nor T. This is not just a peculiar feature of call options, it is also apparent for put options. Furthermore,

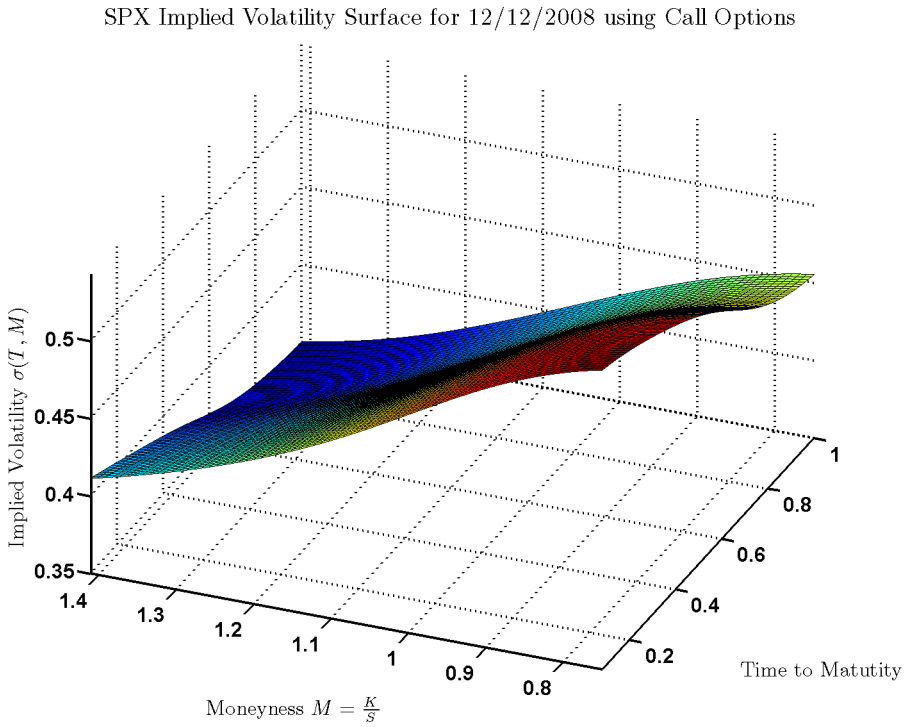


Figure 1.1: Implied Volatility Surface for SPX using call options, December 12 2008.

the implied volatility surface (IVS) of a given asset moves in time. This dynamic feature of the IVS makes it even more difficult to accurately model it, but it is precisely this feature of the IVS that is crucial in managing the risk of a portfolio of options. I have included below the implied volatility surfaces across time for S&P500, again for call options.

1.2 Implied Volatility

In the previous section we discussed the shortcomings of the Black-Scholes pricing model in capturing either the (K,T)-dependence of option prices or its dynamics across time. The empirical evidence against Black-Scholes is vast. In this section we use empirical results as a firm stepping stone for better understanding the behaviour of implied volatility.

The main question we seek to answer is how to capture the change in the shape of the implied volatility surface. To begin with, we fix four maturities, namely the 30, 91, 182 and 365 days. Next, we consider all call options on SPX from August 31, 2008 until August 31, 2013, and across all available deltas, namely the 20, 25, $\cdots$, 75 and 80 delta. With this data[14] we form four matrices of price data for each day (1258 days in total) for each of the 13 different deltas for each fixed maturity. We use these matrices to compute corresponding 1257×13 log-returns matrices. One of our goals is to characterize the dynamics of implied volatility across time to expiration and delta. An analogous study has been done by [Cont, Fonseca] among others. Our results are in line with previously published work [see bibliography] although our analysis are on a previously unexamined time-period and for a much

[14]All data was obtained from OptionMetrics on Wharton Research Data Services (WRDS).

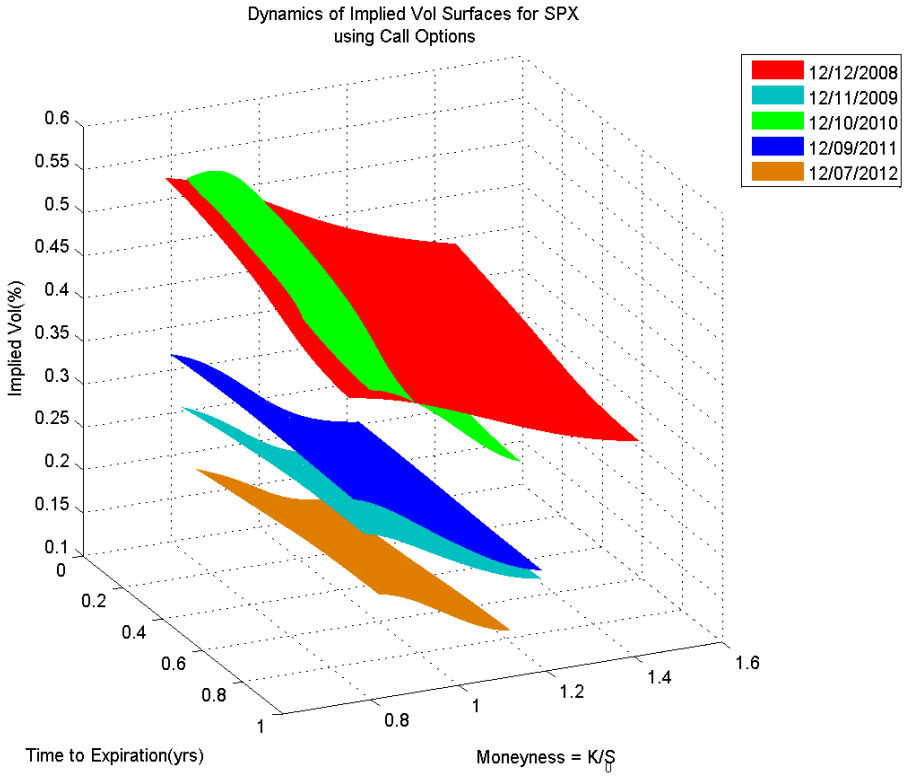


Figure 1.2: We can see that the IVS moves in time, so the dynamic nature of the surface is important.

larger class of assets.

For the 30-day maturity call options we perform a principal component analysis on the correlation matrix of the log-return data for the full five years. We find that the first principal component explains the vast majority of variation in the data at a level of 97 percent. The first three eigenvalues alone account for almost all of the variation and are sufficient for reconstructing the correlation matrix of the implied volatilities. Note that the first eigenvalue can also be used as a very good approximation to the average correlation among the different deltas across all expirations [see apprendix for a proof]. This high level of correlation among the variables gives us hope that we may be able to model the IVS using only a few factors.

When we examine the distribution of values across each of the first three principal components, we find that the first principal component is indicative of a "market" vector with roughly equal allocation across each of the 13 implied volatilities. The next two principal components account for much less of the variation. The second eigenvector displays some sort of a "slopping" effect with increasing weights across the volatilities, and the third eigenvector displays the same pattern as the second, but comes back to higher values eventually. As expected, the volatilities of the assets (volatility of volatilities so to say) increase for more in-the-money options.

For each of the other maturity options, i.e., the 91-day, the 182-days, and the 365-days, we observe close to identical findings to the 30-day maturity options. We will not include the graphs here for the sake of conciseness. In Figure 3 we plot all four maturities together. The result is a prosaic of the individual maturity options except for the second eigenvector where we observe four distinct nodes

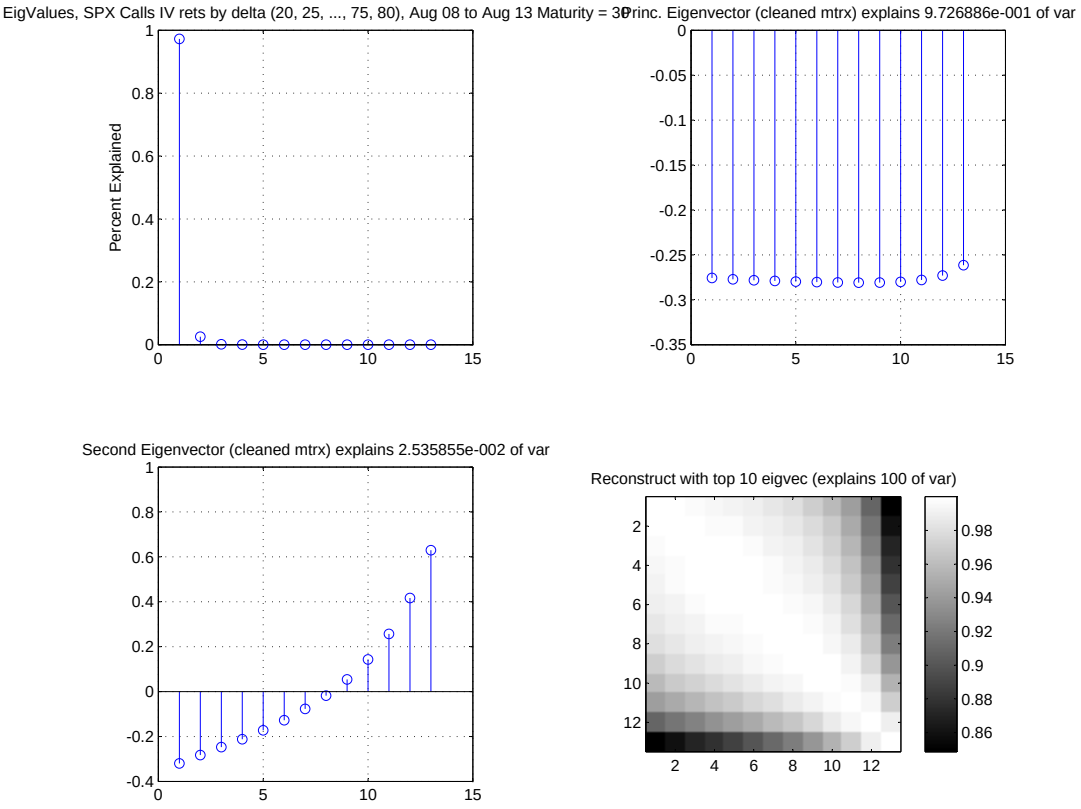


Figure 1.3: The leading eigenvalue accounts for the majority of variation. We see that the market eigenvector accounts for the vast majority of variation, and assigns roughly equal weight to each volatility. The second eigenvector accounts for roughly 2.5 percent of the variation and exhibits clear delta-skew. We can reproduce the correlation matrix using the first 4 principal components. We fix maturity at 30 days.

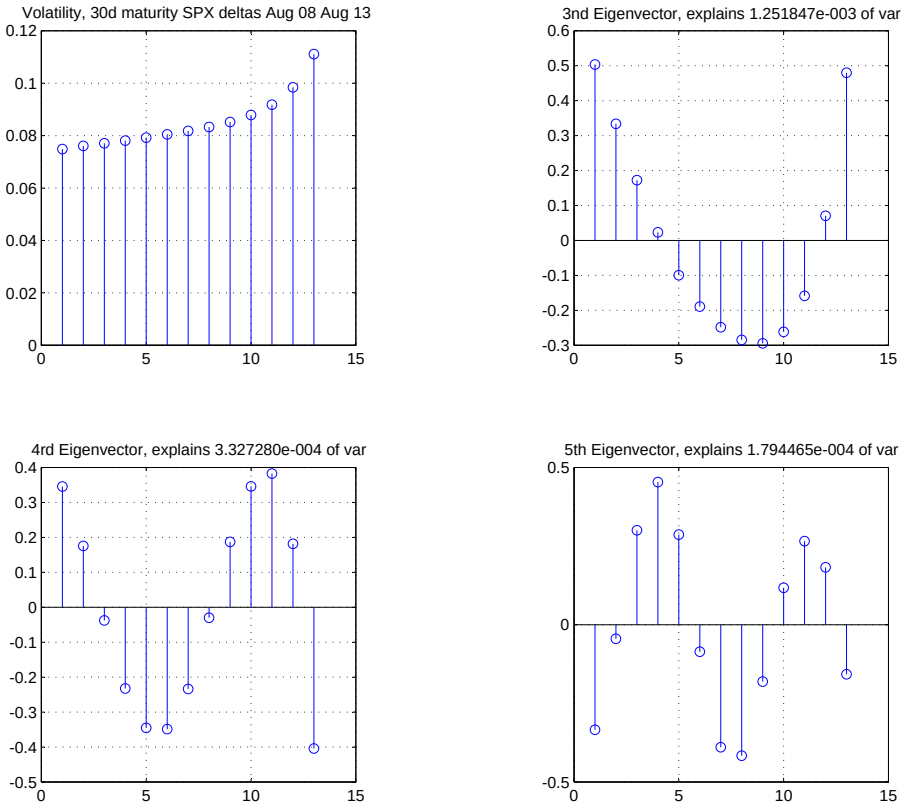


Figure 1.4: The next few eigenvectors carry little explanatory power compared to the first. The volatilities of the variables increase with moneyness. The fourth and the fifth components carry almost no explanatory power. We fix maturity at 30 days.

(representing each maturity) with each node decreasing with delta. The principal eigenvalue in this aggregate case also largely predominates. The volatility displays the same pattern for each node as in the individual maturity cases with slightly higher volatility for deeper in the money call options.

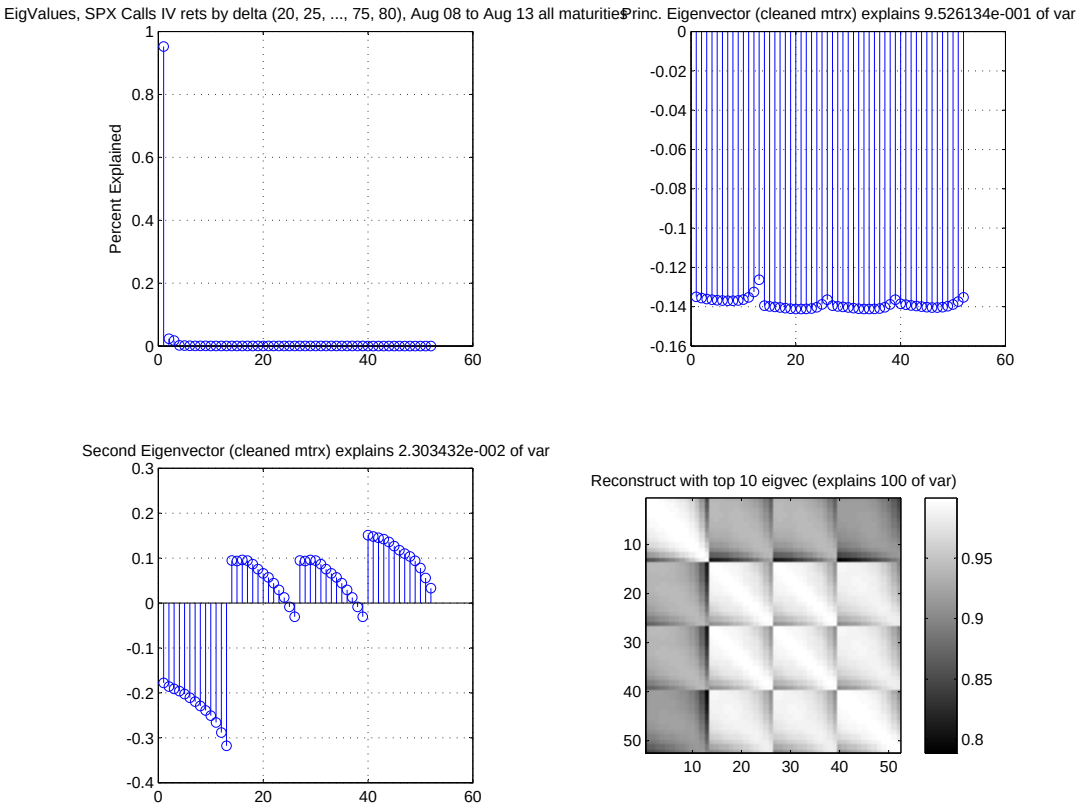


Figure 1.5: The principal eigenvalue accounts for the vast majority of variation, with the other eigenvalues trailing off very quickly. Here we juxtapose all 4 expiration dates together and use the corresponding 1257×52 log-returns matrix. The results indicate four distinct nodes, each corresponding to a fixed maturity.

From this analysis we have learned that the market-mode eigenvector which allocates roughly equal weight to each volatility accounts for the majority of vari-

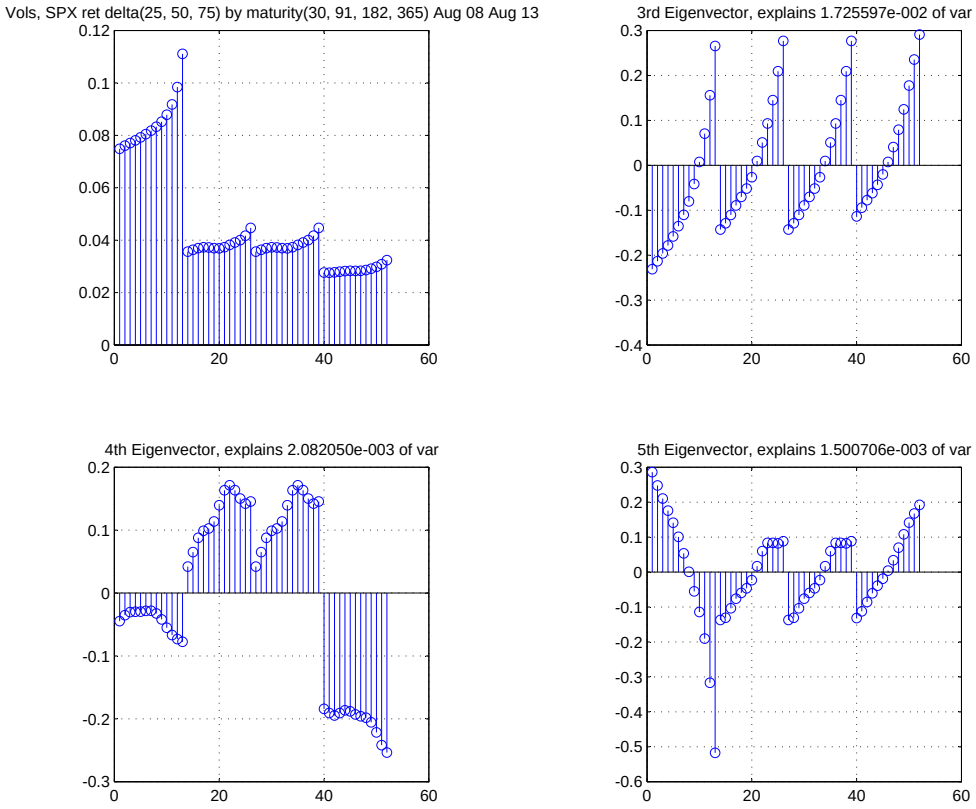


Figure 1.6: The next few eigenvectors have much less explanatory power, and the volatility of each variable is increasing in moneyness and decreasing in time-to-expiration. Here we juxtapose all 4 expiration dates together and use the corresponding 1257×52 log-returns matrix. The results depict four distinct nodes, each corresponding to an individual fixed maturity.

ation of implied volatilities. We could also use the next two eigenvectors for a complete description. These eigenvectors give us more of a slope in our weight allocations. Furthermore, the first three eigenvalues are sufficient to explain the vast majority of variation across volatilities, hence our model will need at most three of the principal components in capturing the change of the IVS.

In order to get closer to a full description of the movement of the implied volatility surface, and how it varies across different asset classes, we examine the first three principal components of the surface across multiple asset classes such as indexes, individual stocks, exchange traded funds/notes, and volatility products. As a natural continuation from what we have already discussed we begin with the S&P500. We find the following results for the first four principal components (the fourth component is included for completion as it usually has very little explanatory power).

Since the principal eigenvalue explains 95 percent of the variation in movement, and this value is also approximately the average correlation[1] of the implied volatility returns across maturity and strike, it is plausible to expect that we may model the dynamics of the surface using just a few factors.

Note that the scale on the first principal component is on the order of 10^{-3} hence the surface is actually almost constant when viewed in the scale of the second or the third principal components. Since combined these three components account for almost all of the variation in the IVS, we conjecture that the IVS can be described as a randomly fluctuating manifold which is very closely approximated by these three principal component factors.

The shape of the first principal component surface is largely flat and it accounts

[1]See appendix for a proof.

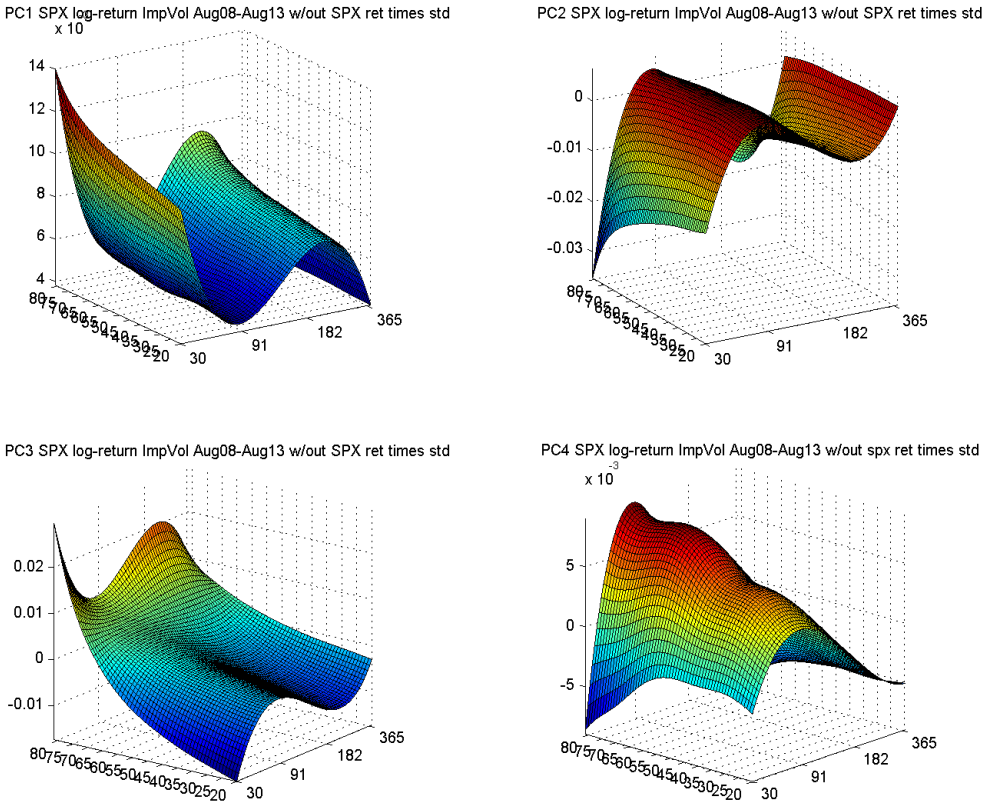


Figure 1.7: First four principal component surfaces (standardized) of the implied volatility surface for SPX. Percent of variation in the change of the original surface each component accounts for is 95, 2.3, 1.7, and .2 percent respectively. We use 52 implied volatility returns and linearly interpolate between them to generate the surface.

for 95 percent of the change in the surface. We interpret this to mean that the most predominant change in the surface is a level effect. The second and third components indicate changes to the surface arising from a time-to-maturity skew a delta-skew respectively. We will discuss these effects in more depth later on.

In the appendix we include similar principal component surface decomposition diagrams for the volatility products VXX and VXZ, and the single-name stocks AAPL, NFLX, BIIB, CVS, HOT, HD, FB, DAL, as well as for the exchange traded funds GLD and IWM. One can see very similar results for these products with very similar explanatory power for the first three principal components. These empirical observations further support being able to characterize the movement of the implied volatility surface using just the first three principal components of the surface.

1.3 Preliminary Analysis of Implied Volatility Coupled with the Underlying Asset

Since implied volatility is a function of the underlying asset, the two are surely related and we seek to examine how the addition of the underlying asset in our implied volatility returns matrix effects the principal components of the implied volatility surface for the underlying asset. We begin by extracting the first four principal components of the implied volatility surface where we couple the implied volatility with the underlying stock. Our first variable is the underlying stock returns and each subsequent variable is the implied volatility returns across the maturities and deltas mentioned previously. As a reminder, the implied volatility returns are computed for a specific call option contract on each given day in our

sample.

The information we expect to extract from the principal components is a determination of how the implied volatility surface changes in time. For example, even in the coupled case we see that the first principal component usually explains about 90 percent of the variation in the IVS. Just as in the previous section, the first component is usually "flat". This information tells us that around 90 percent of the change in the shape of the implied volatility surface is due to a parallel shift of the entire surface, even when we couple with the underlying asset.

The second and third components are either flat and very similar to the first component, or they exhibit either time-skew and/or delta-skew. In the case where there is skew, we see that the weights of the second principal component usually increase with time-to-maturity for fixed delta, and the weights of the third principal component usually increase with delta for a fixed maturity. So the second and third components imply that the remaining change in the shape of the surface is skew. The fourth component exhibits a convexity-effect and usually has very little explanatory power for how the surface varies, but is included for completeness.

The average correlation of the underlying SPX index with all implied volatility returns across deltas and maturities is about -.65. This tells us that the underlying asset and its implied volatilities move in opposite direction. As can be seen from the charts of the eigenvalues, the average correlation is still very high, this is due to the very high correlation among the different implied volatilities as we observed in the previous section. The negative correlation of the stock and its implied volatility is also reflected in the first principal component where we see that the weight allocated to the stock component is of opposite sign than the weight allocated to its implied volatilities.

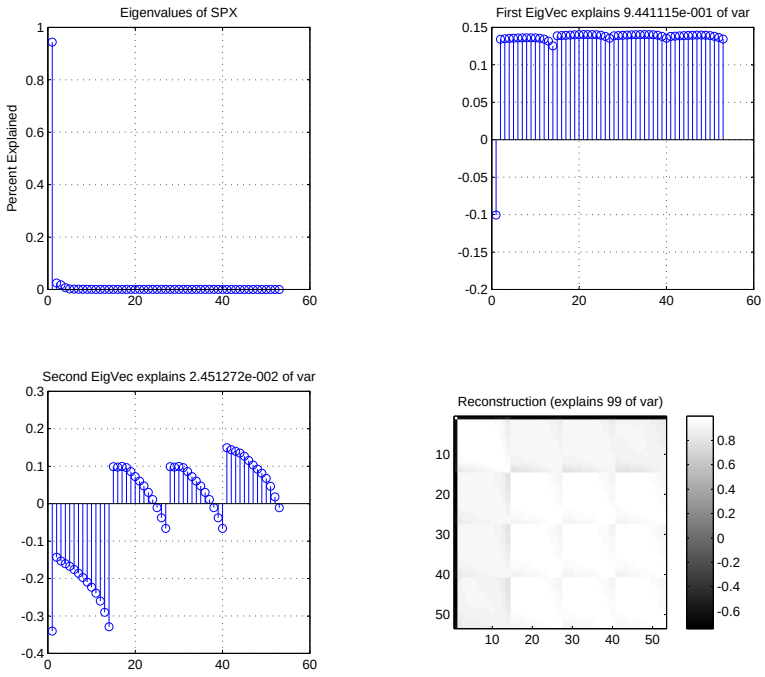


Figure 1.8: Results based on PCA decomposition of the correlation matrix of log-returns of SPX coupled with the log-returns of the implied volatilities for SPX. The results are very similar to those in the previous section without coupling, but with an extra level at the beginning representing the underling asset used to couple. It is important to note that coupling with the underlying asset does not change the results once it is accounted for as compared to not coupling with the underlying asset.

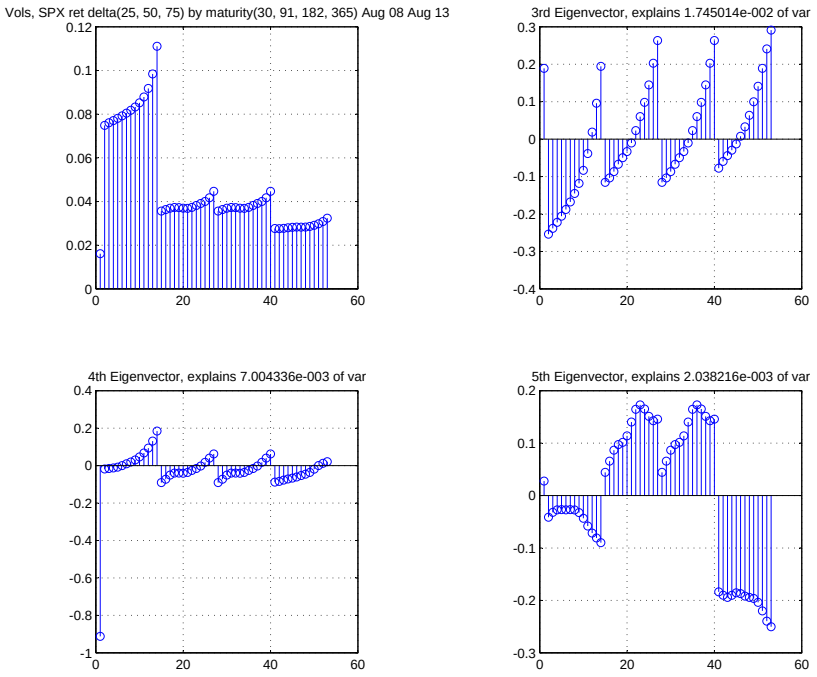


Figure 1.9: Volatilities and remaining significant eigenvectors. Results based on PCA decomposition of the correlation matrix of log-returns of SPX coupled with the log-returns of the implied volatilities for SPX. The results are very similar to those in the previous section without coupling, with an extra level representing the underling asset.

The second and third eigenvectors display very similar behaviour to those corresponding eigenvectors in the study of implied volatilities of the previous section, but with the exception of one extra component added in the front representing the allocation to the underlying asset. Just as in the implied volatility case, when we couple the stock with its implied volatilities we find that the first three principal components account for the majority of variation (about 99 percent).

As examples we now include the principal component surfaces for the index SPX, for the ETF IWM, for the single stock AAPL, and for the volatility product VXX (the principal component surfaces of the other assets mentioned in the last paragraph of the previous section can be found in the appendix).

1.4 Secondary Analysis of Implied Volatility Coupled with the Underlying Asset

To further aid our study of classifying the significant factors responsible for the change of the IVS we expand our toolbox by incorporating results from random matrix theory to our PC analysis in the previous sections. We review the relevant parts of random matrix theory as they pertain to our work. In particular, we focus on the Marchenko-Pastur distribution of random matrixes as described below:

- Let X denote an $M \times N$ random matrix whose entries are i.i.d. with mean 0 and variance $\sigma^2 < \infty$. Denote the correlation matrix and the spectrum (viewed as random variables) by:

$$Y_N = \frac{1}{N} XX' \text{ and } \{\lambda_1, \lambda_2, \cdots, \lambda_n\}.$$

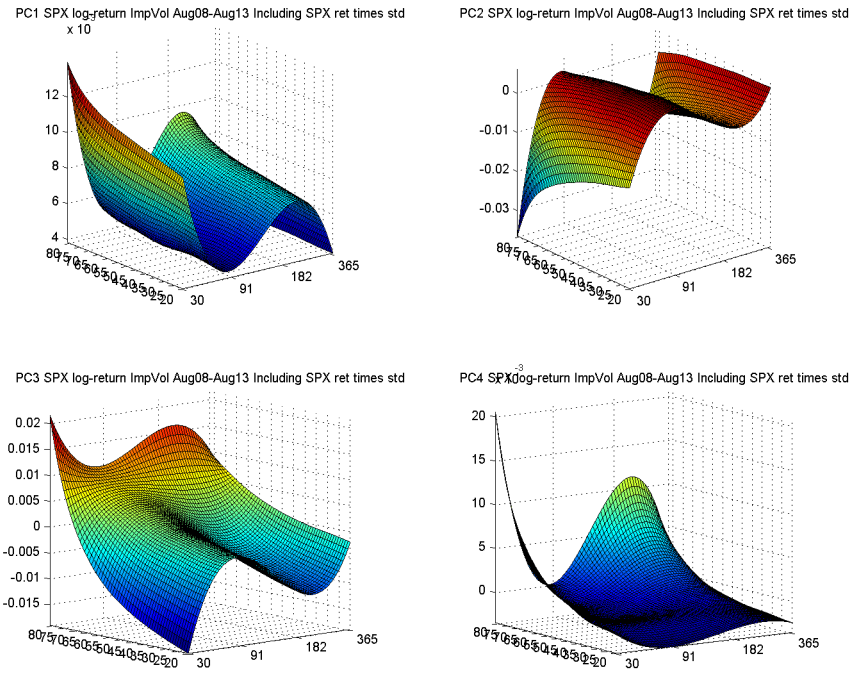


Figure 1.10: Results based on PCA decomposition of the correlation matrix of log-returns of SPX coupled with the log-returns of the implied volatilities for SPX call options. Displayed are the first four principal components which account for roughly 94, 3, 2, and .7 percent of the variation of the implied volatility surface for SPX respectively.

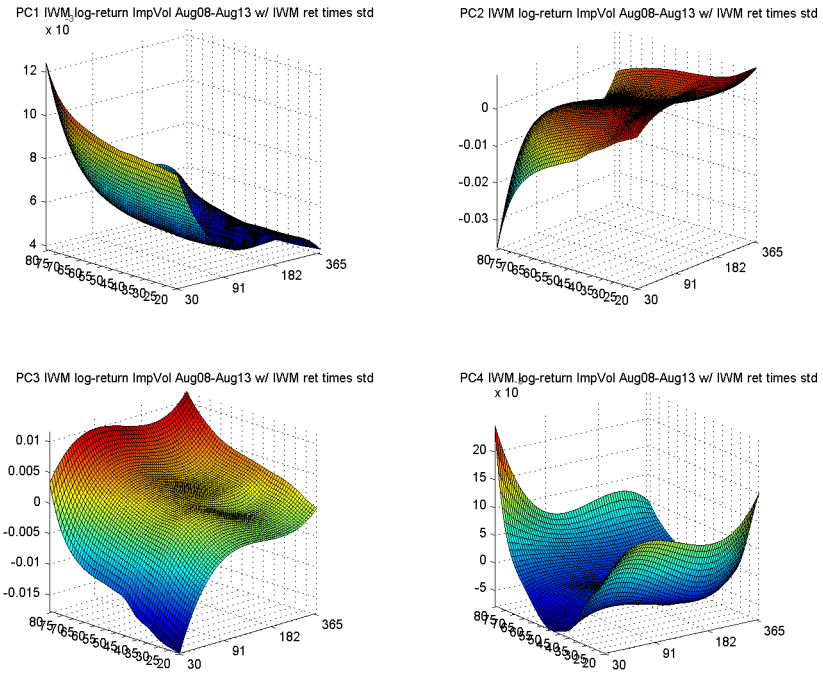


Figure 1.11: Results based on PCA decomposition of the correlation matrix of the log-returns of IWM coupled with the implied volatilities log-returns for IWM options. Displayed are the first four principal components which account for roughly 88, 4, 2, and 1 percent of the variation of the implied volatility surface for IWM respectively.

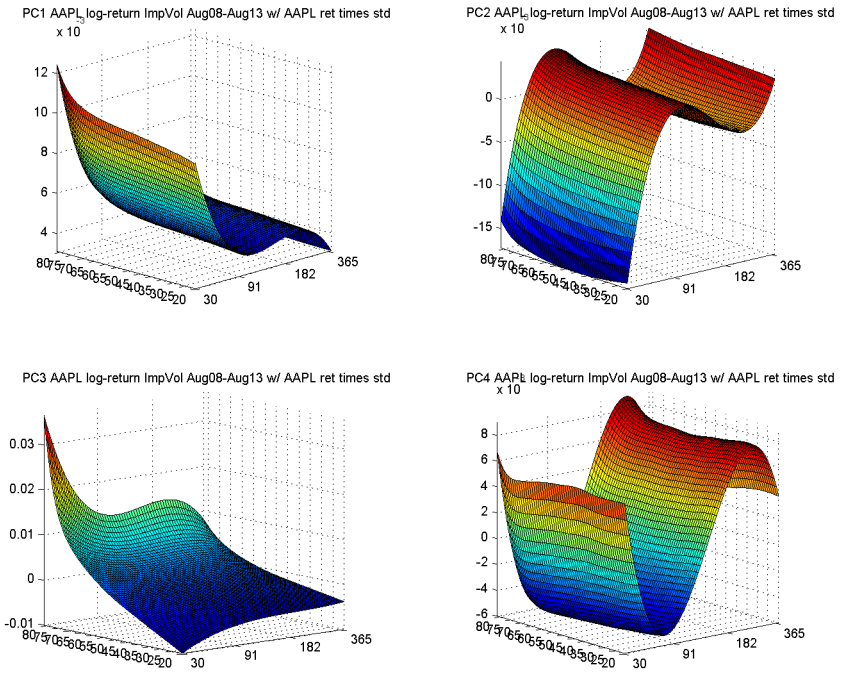


Figure 1.12: Results based on PCA decomposition of the correlation matrix of the log-returns of AAPL coupled with the implied volatilities log-returns for AAPl options. Displayed are the first four principal components which account for roughly 86, 7, 2, and 1 percent of the variation of the implied volatility surface for AAPL respectively.

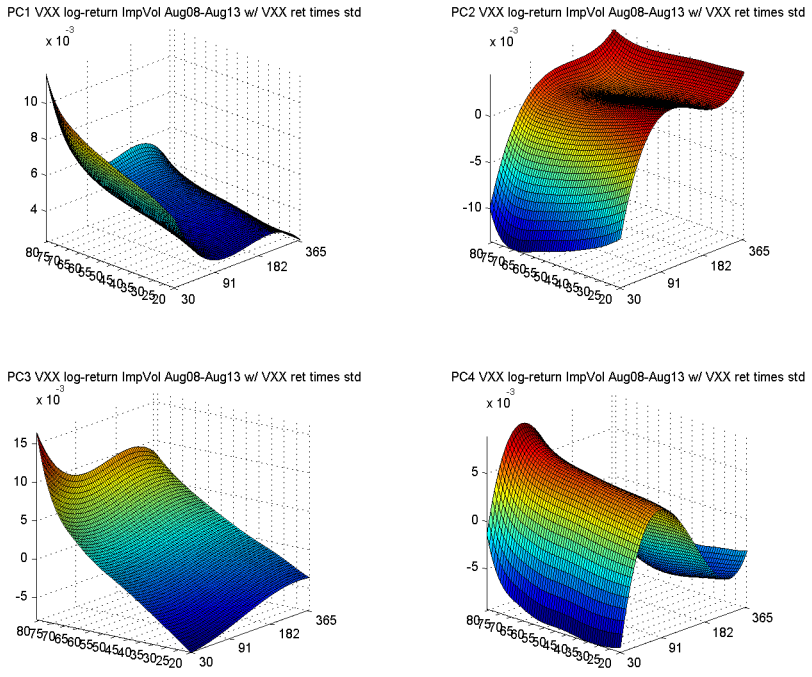


Figure 1.13: Results based on PCA decomposition of the correlation matrix of the log-returns of VXX coupled with the implied volatilities log-returns for VXX options. Displayed are the first four principal components which account for roughly 66, 15, 5, and 4 percent of the variation of the implied volatility surface for VXX respectively. We find that the latter eigenvectors carry more explanatory power for this volatility product when compared to the instruments above. Volatility products require more eigenvalues and corresponding principal components to account for variation. Usually the top 4 eigenvalues still account for at least 90 percent of variation. The surfaces indicate mainly a "flat" effect.

- Consider the density of states (dos):

$$\mu_M(A) := \frac{1}{M}\#\{\lambda_j \in A\}, A \subset \mathbb{R}.$$

- *Marchenko-Pastur distribution*: Assume $M, N \to \infty$ s.t. $\frac{M}{N} \to \Lambda \in (0, \infty)$. Then $\mu_N \to \mu$ in distribution, where

$$\mu(A) = \begin{cases} (1 - \frac{1}{\Lambda})\mathbb{1}_{0 \in A} + \nu(A), & \text{if } \Lambda > 1 \\ \nu(A), & \text{if } 0 \leq \Lambda \leq 1 \end{cases}$$

- and

$$d\nu(x) := \frac{1}{2\pi\sigma^2}\frac{\sqrt{(\lambda_+ - x)(x - \lambda_-)}}{x\Lambda}\mathbb{1}_{[\lambda_-, \lambda_+]}dx.$$

- Where $\lambda_\pm = \sigma^2(1 \pm \sqrt{\Lambda})^2$.

- In our case, we define the *MP-threshold* to be the upper bound λ_+: $(1+\sqrt{\frac{N}{T}})^2$. Where $N = 53$ for the number of variables we use (1 for the underlying asset, and 52 implied volatility returns), and $T = \max$(asset creation date, August 2004).

From empirical results, we can see that the vast majority of the variation in the implied volatility surface is almost always explained by the first three or four principal components. An application of random matrix theory in the style of Bouchaud and Potters [2] confirms these results. Following the theory and notation

developed in their paper and discussed above, if we denote by X our $T \times N$ matrix of standardized returns, and by C the Pearson estimator of the correlation matrix which is given by

$$C = \frac{1}{T} X'X,$$

then the upper-bound for the domain of the Marcenko-Pastur density for the distribution of eigenvalues of a random matrix whose entries are standard Gaussian random variables is given by

$$\lambda_+ = \left(1 + \sqrt{\frac{N}{T}}\right)^2.$$

In our case $\lambda_+ = 1.0862$. We denote by "significant" all eigenvalues greater than λ_+. These eigenvalues give us some information about the "true" [2] correlation matrix and hence are useful in separating signal from noise. In our study, theory confirms empirical evidence. We find that across the 13 assets mentioned usually three or four of the eigenvalues and their corresponding eigenvectors lie outside of the Marcenko-Pastur upper bound of 1.0862. This gives further evidence that the top three or four eigenvalues and their corresponding surfaces suffice in describing the behaviour of the IVS.

As further support that we can use the Marchenko-Pastur distribution as an indicator of where to separate noise from signal, i.e., that the underlying assumptions of the MP-distribution hold for our data, we perform the following experiment:

1. For each option in OptionMetrics we generate a random matrix R with the same underlying distribution as the empirical data by permuting the time-series for each of the 53-variables.

[2] Please refer to the Bouchaud and Poters paper for a more detailed description.

2. We compute the correlation matrix C of R, and compute the spectrum of C.

3. We compute the empirical CDF $F_n(x) := \frac{1}{n}\sum_{i=1}^{n}(\mathbb{1}_{X_i \leq x})$ where X_i represents the ith eigenvalue.

4. Using Kolmogorv-Smirnov, we calculate the test statistic $D_n = \sqrt{n}\sup_x |F_n(x) - CMP(x)|$ where $CMP(x)$ the cumulative MP distribution.

Using this framework, we test the null hypothesis that the empirical data is generated from the corresponding MP-distribution. We find that for 100 percent of the option contracts in OptionMetrics we cannot reject the null hypothesis at the 1 percent significance level.

1.5 Systemic vs. Idiosyncratic Risk

In the previous section we saw for the assets studied that three factors suffice in capturing the majority of variation in the IVS. In this section we further examine to what extent three factors suffice in explaining the dynamics of the IVS.

We accomplish this by producing the first four IVS principal components, the spectrum of eigenvalues, the correlation matrix, and other relevant statistics for as many asset classes as possible. This will provide us with a comprehensive collection of results on which we can formulate and test our model. For this purpose we build a collection of matlab and python scripts which automate the collection of data and the generation of statistical results for each asset. We collect our data from OptionMetrics via WRDS from August 31, 2004 until August 31, 2013, unless the asset does not have data that far back, in which case we use its full history of data until August 31, 2013. All of the relevant code and the database of results is ready

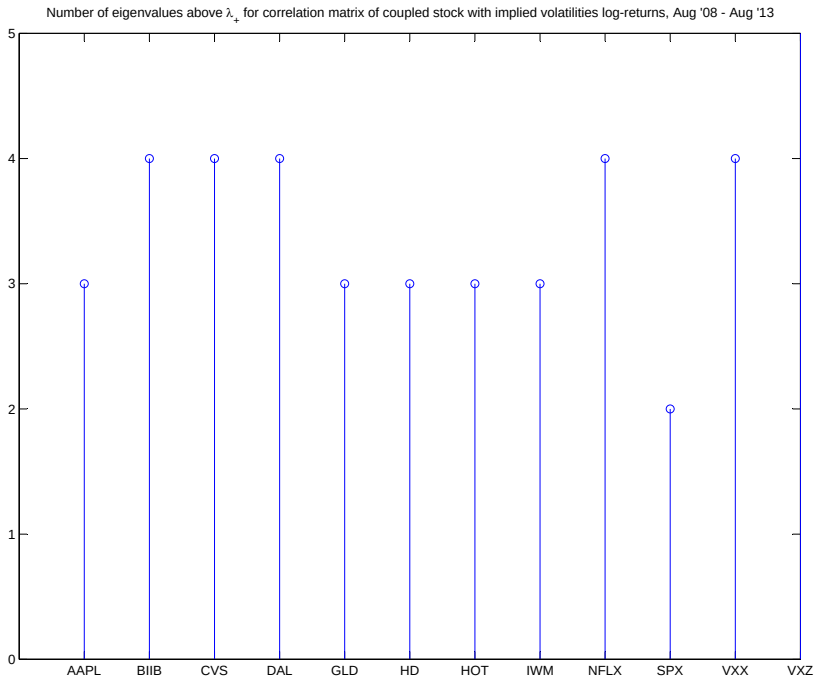


Figure 1.14: Results indicate that around three or four of the top eigenvalues are usually significant in the sense that projecting along these corresponding eigenvectors can be used to distinguish signal from noise in the correlation matrix.

upon request.

We begin by examining the spectrum of the top 20 most liquid ETFs, where liquidity is measured as average daily volume traded. The percent variation explained by the first three components is given below. We can see that for the majority of these ETFs the percent explained by the first three eigenvectors is above 90. The first eigenvalue is also shown, and it usually accounts for 70 and 90 percent of variation.

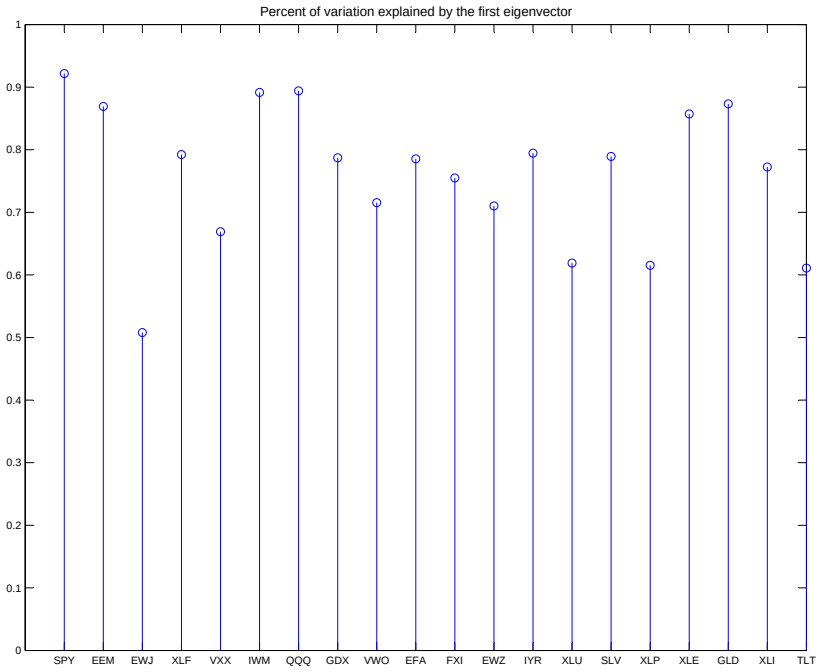


Figure 1.15: Percent of variation explained by the first component for the 20 most liquid ETFs listed. Data from August 2004 (or creation of asset) until August 2013.

In addition, we determine the number of eigenvalues which exceed the Marcenko-Pastur (MP) upper bound as discussed in the previous section. We display the

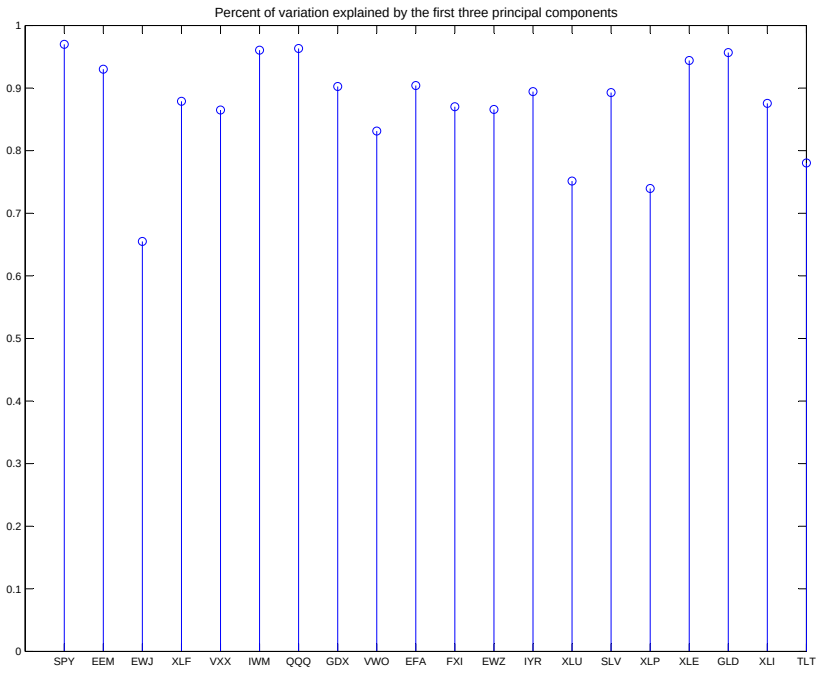


Figure 1.16: Percent of variation explained by the first three components for the 20 most liquid ETFs. Data from August 2004 (or creation of asset) until August 2013.

results below. From this analysis, we find that the majority of the ETFs have around three or four eigenvalues which exceed the MP upper bound; again confirming that the topmost three eigenvalues can be used to separate signal from noise in the correlation matrix.

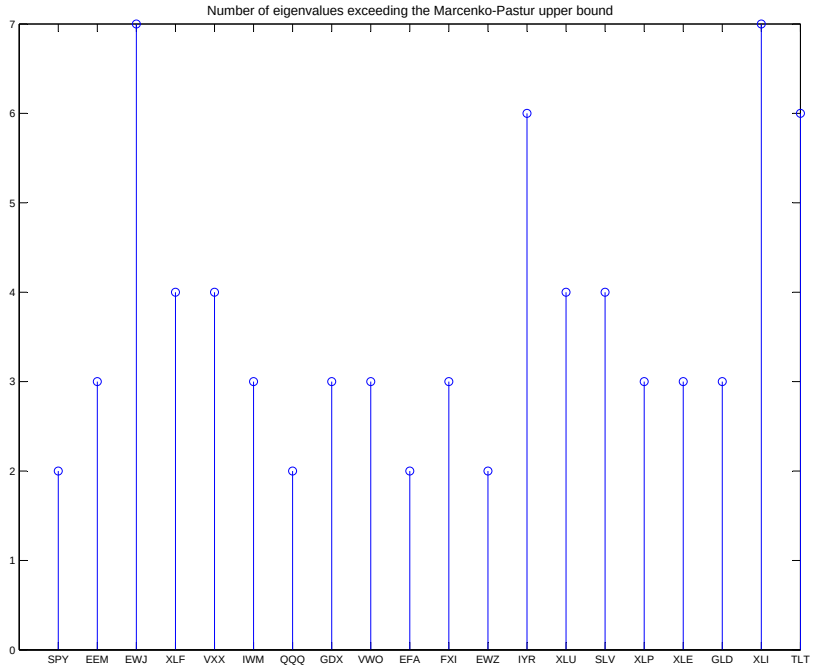


Figure 1.17: Number of eigenvalues exceeding the MP upper limit of the spectrum for the 20 most liquid ETFs.

We examine the same information as above for the some of the most liquid volatility products; note that VXX is present in both asset classes.

Next, let us compare the results of the 20 most liquid ETFs (lETFs) to those of the 6 volatility ETFs (vETFs) above. The first component usually explains slightly more of the variation for lETFs than for the vETFs with an average value of 76 percent versus 70 percent respectively. The opposite conclusion holds for

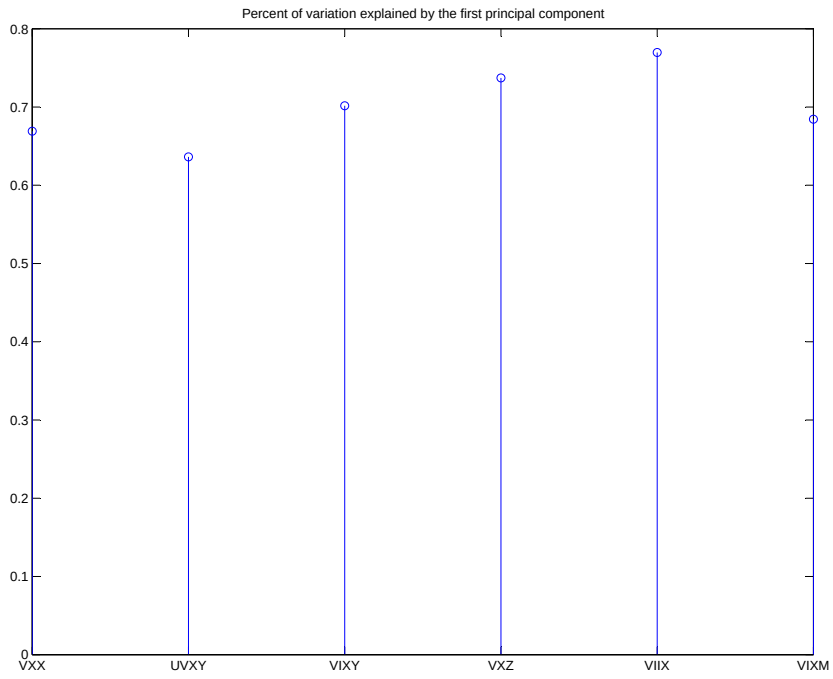


Figure 1.18: Percent of variation explained by the first component for the 6 volatility ETFs listed. Data from August 2004 (or creation of asset) until August 2013.

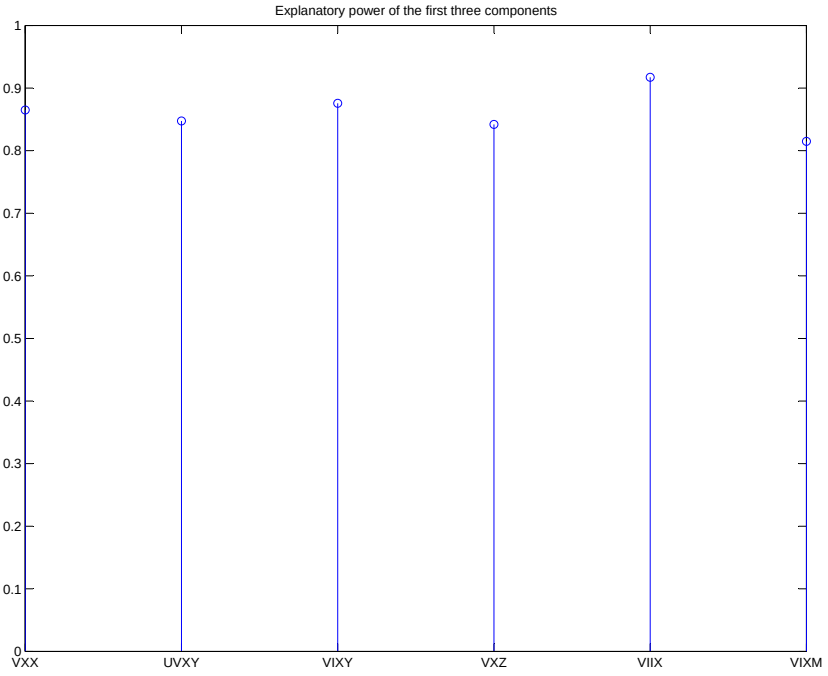


Figure 1.19: Percent of variation explained by the first three components for the 6 volatility ETFs listed. Data from August 2004 (or creation of asset) until August 2013.

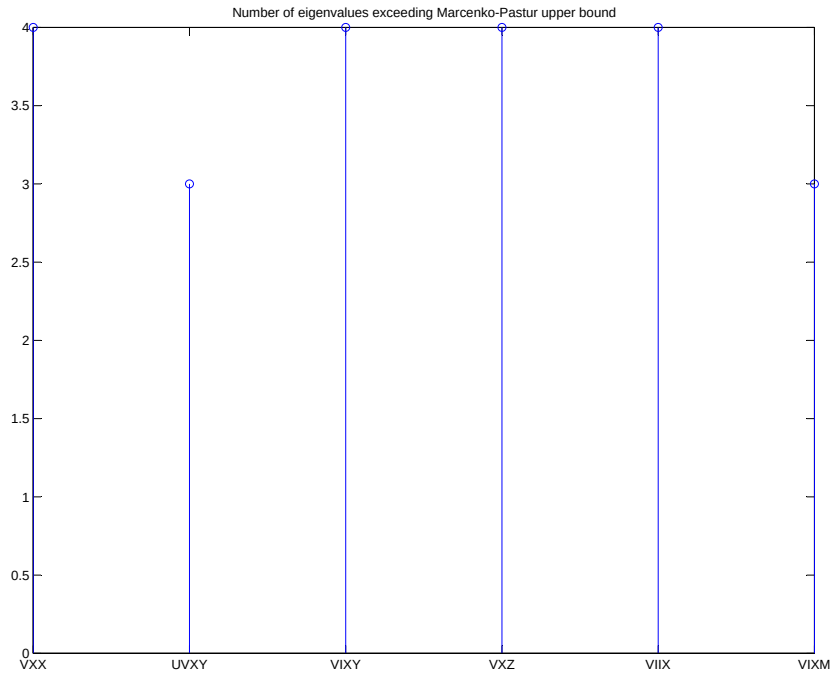


Figure 1.20: Number of eigenvalues exceeding the MP upper limit of the spectrum for the 6 volatility ETFs listed. The greater the systemic risk faced, the smaller the number of eigenvalues exceeding the MP upper limit.

the second and third factors respectively: on average, they explain 6.6 and 4.4 percent for lETFs and 11 and 4.7 percent for vETFs. The percent explained by the first three components is roughly 87 percent on average for both classes, as is the number of eigenvalues above the Marcenko-Pastur threshold at roughly 3.7. The conclusion we draw from these results is that lETFs and vETFs are comparable, but vETFs carry more what we refer to as idiosyncratic risk which is captured in the slightly lower principal eigenvalue (i.e., the eigenvalue describing the significance of the market-mode) and redistributed to the second, third, and to the fourth eigenvalues.

In order to gain more insight into the differences the type of risk options across equities, volatility products, and other ETFs carry, we examine the spectrum of the various products in each asset class. For equities, we begin with the constituents for the S&P500. The data is collected from August 31, 2004 until August 31, 2013, again via WRDS.

On average, the results for the individual stocks in the S&P500 index are comparable to those of vETFs and lETFs. Thus far indicating no statistically significant difference between different asset classes. We use the magnitude of the leading eigenvalue as an indicator of systemic risk. The higher the principal eigenvalue, the stronger the level effect in describing the movement of the IVS. This effect can be interpreted as a roughly equal change across all volatility products, indicating systemic risk. On the other side of the coin, the size of the second and third eigenvalues indicate idiosyncratic risk. These eigenvalues capture non-uniform movements in the IVS.

It is interesting to see whether the eigenvalues are correlated to other properties

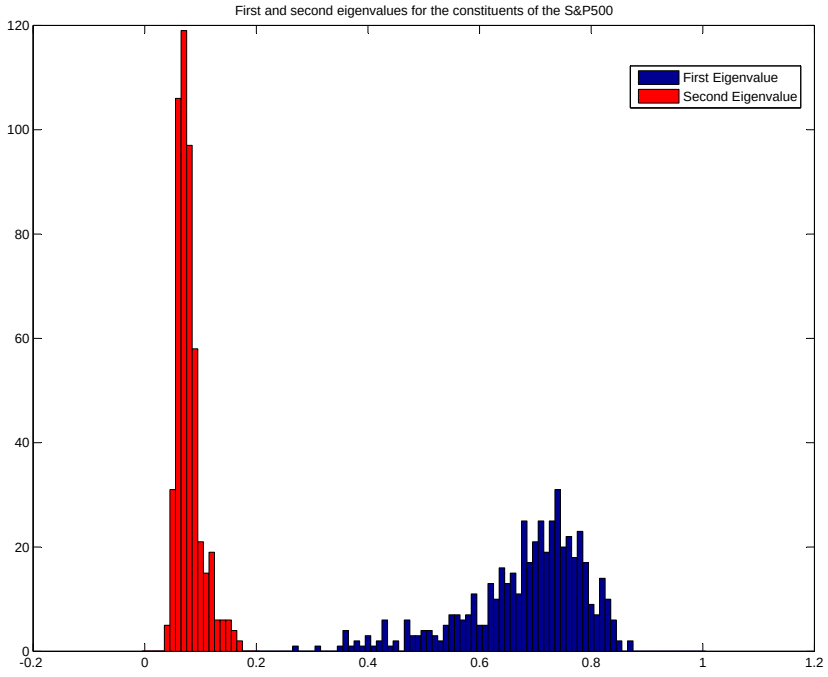


Figure 1.21: First and second eigenvalues of the constituents of the S&P500 from August 2004 to August 2013. Average value is 68 percent and 7.9 percent respectively.

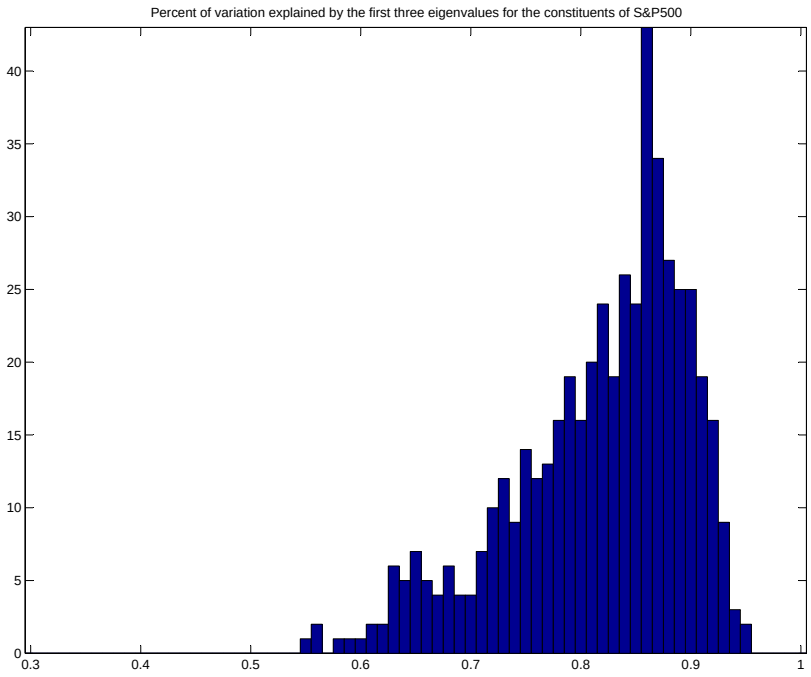


Figure 1.22: Percent of variation explained by the first three principal components of the constituents of the S&P500 from August 2004 to August 2013. Average value is 82 percent.

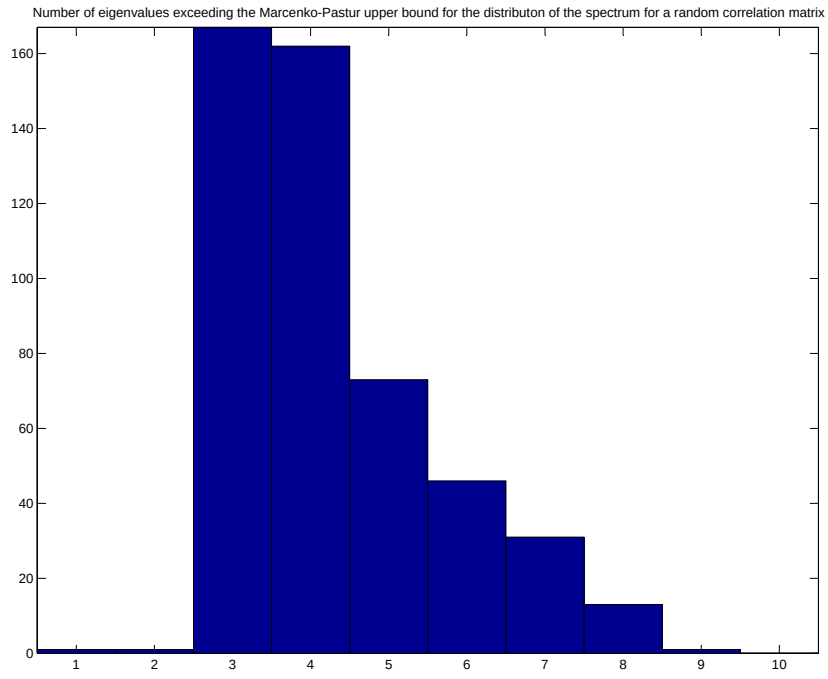


Figure 1.23: Number of eigenvalues exceeding the MP upper limit of the spectrum for the constituents of the S&P500 from August 2004 to August 2013. Average value is 4.3 eigenvalues.

of the stocks such as average daily volume traded [3], to the daily volatility of the stock, or to the correlation of the individual stock to the index itself (i.e., the market index). We present the results of these inquiries below.

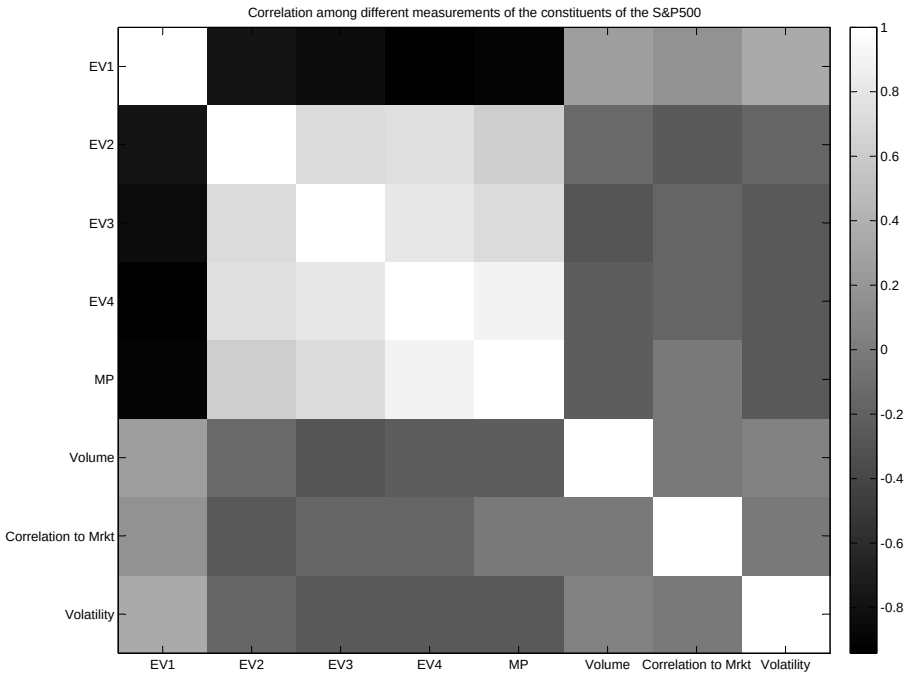


Figure 1.24: Correlation among the first, second, third, and fourth eigenvalues of options on the constituents of the S&P500 to each other and to the number of eigenvalues above the Marcenko-Pastur-threshold, to the average daily volume traded and to the correlation to the market index. The average correlation to the market is 57 percent. All results are computed on daily data from August 2004 to August 2013. Light colors signify positive correlation, and dark colors signify negative correlation.

[3]All data is computed from August 2004 to August 2013 on a daily basis

Table 1.1: Correlation Matrix of the constituents of the S&P500

EV1	EV2	EV3	EV4	# > MP	Volatility	Corr. to Mkt.	Volume
1.00	-0.77	-0.85	-0.94	-0.89	0.26	0.15	0.33
-0.77	1.00	0.72	0.74	0.62	-0.15	-0.25	-0.17
-0.85	0.72	1.00	0.81	0.72	-0.29	-0.17	-0.25
-0.94	0.74	0.81	1.00	0.88	-0.22	-0.16	-0.24
-0.89	0.62	0.72	0.88	1.00	-0.23	-0.00	-0.26
0.26	-0.15	-0.29	-0.22	-0.23	1.00	-0.00	0.04
0.15	-0.25	-0.17	-0.16	-0.00	-0.00	1.00	-0.03
0.33	-0.17	-0.25	-0.24	-0.26	0.04	-0.03	1.00

From the correlations matrix we see that there is a strong negative correlation between the first eigenvalue with the second and the third eigenvalues, and a very strong negative correlation with the fourth eigenvalue. These results can be explained by the fact that the sum of the eigenvalues must be 1; so the higher the leading eigenvalue, the lower the remaining eigenvalues. Better yet, the results can also be explained in terms of idiosyncratic risk; the higher the leading eigenvalue the less idiosyncratic risk, and hence the lower the remaining eigenvalues.

If we use the remaining eigenvalues as a measure of idiosyncratic risk, then our claim is further supported by the strong negative correlation between the number of eigenvalues which exceed the MP-threshold and the leading eigenvalue, and the strong positive correlation between the remaining eigenvalues and the MP-threshold. The conclusion is that the higher the idiosyncratic risk faced by options on a company, the lower the leading eigenvalue, and the more weight is carried by the remaining eigenvalues, and in turn, and the higher the number of eigenvalues exceeding the MP-threshold.

These results indicate that the higher the leading eigenvalue, the less idiosyncratic risk the option has, and the more"randomly" price fluctuations behave. This is also indicated by the fact that the leading eigenvalue is strongly negatively correlated to the number of eigenvalues above the MP-threshold, thus indicating more "random" behavior for high leading first eigenvalue option contracts since fewer eigenvalue exceed the MP upper-bound. The opposite effect results for option contracts with higher values of the remaining eigenvalues. These eigenvalues give an indication of the strength other factors intrinsic to the underlying stock have on the option returns (i.e., idiosyncratic risk).

We compare the correlation of each eigenvalue with the underlying stock's correlation to the market we see that there is a small positive correlation with the leading eigenvalue, and negative correlations with the remaining eigenvalues. This ties in with our interpretation above; the higher the principal eigenvalue the less idiosyncratic risk the option carries, and the more like the overall market the it behaves.

The opposite conclusion holds for the remaining eigenvalues, especially the second, which gives a stronger indication of idiosyncratic risk. Furthermore, in part due to the positive correlation to the market, we expect to see that stocks with less idiosyncratic risk and more systemic risk move more on average as they are more influenced by various news and speculation on a daily basis. This conclusion is also reflected by our correlation matrix, which shows positive correlation of volatility to the first eigenvalue, and negative correlation of volatility to the latter eigenvalues.

Let's now take a look at the top 15 underlying constituents in S&P500 whose options market exhibit the highest leading eigenvalue, as well as the bottom 15 underlying constituents whose options market exhibits the lowest leading eigenvalue,

and see how our story of idiosyncratic versus systemic risk fits in.

Table 1.2: Top 15 underlying constituents and bottom 15 underlying constituents by first eigenvalue.

Bottom 15 Constituents	EV1	Top 15 Constituents	EV1
KMI	0.27272	GS	0.87176
POM	0.31406	JPM	0.87092
WEC	0.34958	BAC	0.85115
PNW	0.35611	SLB	0.84669
HCBK	0.35995	CAT	0.84219
NLSN	0.36338	AAPL	0.84126
TE	0.36342	XOM	0.84109
NU	0.3667	NOV	0.83737

Table 1.3: Remaining top 15 underlying constituents and bottom 15 underlying constituents by first eigenvalue.

Bottom 15 Constituents	EV1	Top 15 Constituents	EV1
BMS	0.37675	CME	0.83639
XEL	0.3828	MA	0.8358
WIN	0.38664	MS	0.83471
RSG	0.39779	APA	0.83081
FTR	0.40043	GOOG	0.83011
MKC	0.40475	HIG	0.8301
XYL	0.40596	HES	0.82965

The top 15 constituents in terms of the leading eigenvalue predominantly come from the financials, energy, and information technology sectors. The bottom 15 constituents predominantly come from the utilities, industrials, with one outlier in each of consumer staples, energy, financials and telecommunication services.

A few things stand out from Figures 1.2 and 1.3. The leading eigenvalue explains drastically more of the variation in the IVS for the top 15 stocks than for the bottom 15. Options on Goldman Sachs (GS) for example, have leading eigenvalue with explanatory power as much as over three times that of options on Kinder Morgan (KMI). Another observation which is readily apparent is that the bottom 15 names are small-cap growth stocks or newly-formed companies. The opposite observation holds for the top 15 names: they are large, widely recognized, popular

in the media, and well-established companies. This observation further supports our claim that idiosyncratic risk in negatively correlated to the principal eigenvalue. The bottom 15 underlying stocks carry a lot of idiosyncratic risk which is intimately connected to the nature of their business. The top 15 underlying companies are much more dependent on the state of the overall market and carry much less idiosyncratic risk and much more systemic risk.

Furthermore, we have that the top 15 underlying stocks above have an average correlation to the market of 63 percent (overall average is 60 percent), an average daily volatility of 2.8 percent (overall average is 2.3 percent), an average number of eigenvalues exceeding MP-threshold of 3 (overall average is 4.3), and an average daily trading volume of three times that of the market. On the other hand, the bottom 15 underlying stocks above have an average correlation to the market of 47 percent (overall average is 60 percent), an average daily volatility of 1.5 percent (overall average is 2.3 percent), an average number of eigenvalues exceeding MP-threshold of 7.3 (overall average is 4.3), and an average daily trading volume of three times less than that of the market.

1.6 S&P500 constituents by industry sector and other groupings

We partition the constituents of S&P500 into 10 industry sectors: Energy, Materials, Industrials, Consumer Discretionary, Consumer Staples, Health Care, Financials, Information Technology, Telecommunication Services, and Utilities. We perform the same analysis as in the previous section for these industries. The results are given in Table 1.3.

Table 1.4: Results for options on the constituents of the S&P500 partitioned by industry sector

	Enrg	Mat.	Indust.	CD	CS
EV1	.73	0.69	0.69	.68	0.65
EV2	.078	.077	0.076	0.08	.077
EV3	.053	0.056	0.056	0.058	.059
EV4	.027	0.032	0.03	0.03	.034
# > MP-bound	3.7	4.48	4.05	4.14	4.7
Volume	5.2M	2.9M	2.8M	4.3M	4.4M
Corr to Mrkt	.59	.63	.65	.52	.47
Volatility	.029	.024	.02	.024	0.016

[1] Enrg, Mat., Indust., CD, CS, HC, Fin, IT, TS, Util. are short for Energy, Materials, Industrials, Consumer Discretionary, Consumer Staples, Health Care, Financials, Information Technology, Telecommunication Services, and Utilities respectively.

[2] All results are on average for given sector. Data is collected on a daily basis from August 2004 until August 2013.

Table 1.5: Results for options on the constituents of the S&P500 partitioned by industry sector

	HC	Fin	IT	TS	Util.
EV1	.69	0.71	.71	.54	.51
EV2	.081	.067	.08	.089	.103
EV3	.058	.05	0.05	.072	.072
EV4	.034	.032	0.029	.029	.049
# > MP-bound	4.19	4.1	3.97	5.83	6.3
Volume	4.4M	3.6M	11.1M	6.8M	1.8M
Corr to Mrkt	.50	.64	0.554	.528	.632
Volatility	.019	.029	.024	.017	.015

[1] Enrg, Mat., Indust., CD, CS, HC, Fin, IT, TS, Util. are short for Energy, Materials, Industrials, Consumer Discretionary, Consumer Staples, Health Care, Financials, Information Technology, Telecommunication Services, and Utilities respectively.

[2] All results are on average for given sector. Data is collected on a daily basis from August 2004 until August 2013.

Across industry sectors we see that with the exception of telecommunication services and utilities all other sectors have a first eigenvalue very comparable to the overall options market which is 68 percent. This discrepancy with options in the telecommunication services and utilities sectors is further reflected by their larger-

than-average second, and third eigenvalues, as well as the larger-than-average number of eigenvalues exceeding the Marcenko-Pastur bound. These results imply that options on the telecommunication services and utilities sectors carry the most idiosyncratic risk. This is especially true for options on the utilities sector which is at the extreme end in this analysis.

We perform the same study for options with underlying assets in the top ten constituents representing pure value stocks, pure growth stocks, and by stock index weight as allocated to S&P500. We display and analyse the results below.

Table 1.6: Options with underlying constituents in the S&P500 representing value, growth and market index weight

	EV1	EV2	EV3	EV4	# EVs > MP	Volume	Corr	Vol.
Value	0.71	0.073	0.054	0.029	4	5.4M	0.51	0.028
Growth	0.70	0.098	0.05	0.033	3.8	8.3M	0.27	0.031
Weight	0.81	0.059	0.035	0.018	3.1	22,1M	0.69	0.02
Average	0.68	0.079	0.056	0.032	4.3	4.9M	0.57	0.023

[1] Results based on the top ten constituents of the S&P500 representing each of pure value, pure growth, and by their index weight. The last row represents the market average across all constituents.

From Table 4 we see that the first four eigenvalues are very comparable for both options on pure value and pure growth stocks. Furthermore, the first eigenvalue is slightly larger and the remaining eigenvalues are slightly lower for options on pure value stocks than for those of pure growth stocks. This can again be understood in terms of our framework of systemic risk vs. idiosyncratic risk; the pure growth stocks carry more paradigm changes, hence they carry slightly more idiosyncratic risk which is in turn reflected in the slightly higher values for the second, and third eigenvalues. This is also reflected in the much lower correlation to the market for pure growth stocks versus pure value stocks, which is in turn reflected in slightly lower systemic risk for pure growth stocks. Overall options on the two asset classes

are comparable and carry similar risk.

For the top ten stocks by index weight constituting the S&P500 we naturally see a stronger correlation to the market than for either growth or value stocks as well as much higher daily volume. Furthermore, these stocks have very large principal eigenvalue and lower remaining eigenvalues thus reflecting less idiosyncratic risk and more systemic risk inherent in options on these stocks. The lower number of eigenvalues exceeding the Marcenko-Pastur bound for index weight stocks further supports this analysis.

Now we examine the same statistics as for the constituents of the S&P500 for the all remaining assets found in the WRDS OptionMetrics database (about 3300 remaining assets). We exclude the assets already studied from this list. The corresponding correlation matrix for these remaining assets is given below.

Table 1.7: Correlation matrix of all remaining remaining assets (around 3300)

EV1	EV2	EV3	EV4	# EVs > MP	Vol	Corr.	Daily Volume
1.00	-0.64	-0.83	-0.89	-0.83	0.01	-0.03	0.07
-0.64	1.00	0.59	0.47	0.29	0.03	-0.15	-0.06
-0.83	0.59	1.00	0.71	0.64	0.05	0.02	-0.06
-0.89	0.47	0.71	1.00	0.78	-0.018	-0.01	-0.06
-0.83	0.29	0.64	0.78	1.00	0.01	0.20	-0.05
0.01	0.03	0.05	-0.01	0.01	1.00	-0.12	-0.02
-0.03	-0.15	0.02	-0.01	0.20	-0.12	1.00	0.07
0.07	-0.06	-0.06	-0.06	-0.05	-0.02	0.07	1.00

[1] Corr. refers to the correlation with the market index.

We now show the results for the average values of the variables discussed across all remaining assets and those for the top and bottom 1 percent (which amounts to around 33 assets each).

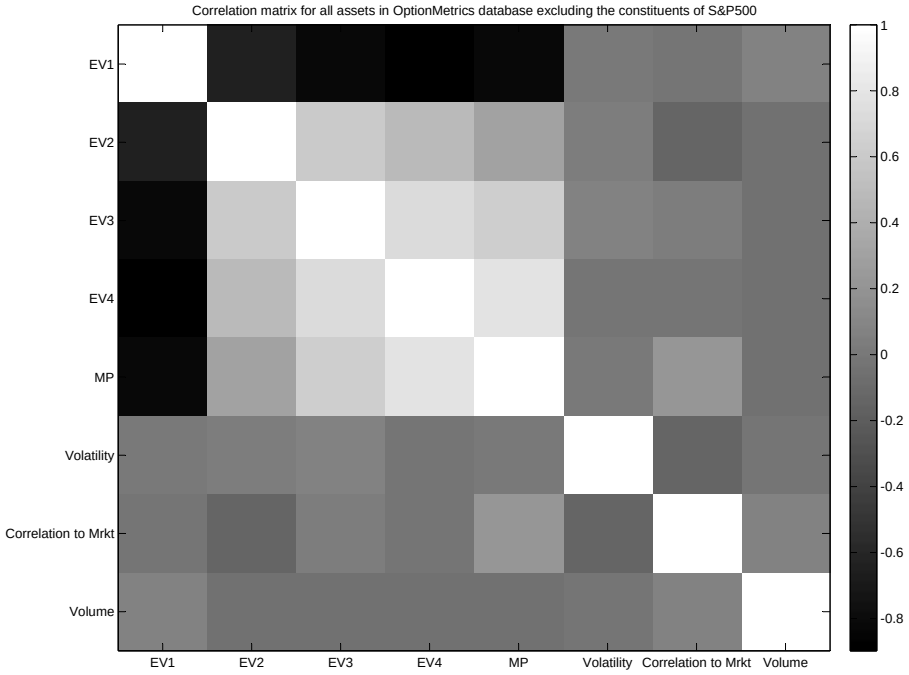


Figure 1.25: Correlation among the first, second, third, and fourth eigenvalues of all remaining remaining assets (around 3300) to each other and to the number of eigenvalues above the Marcenko-Pastur-threshold, the average daily volume traded (of the underlying asset) and the correlation to the market index. All results are computed on daily data from August 2004 to August 2013.

Table 1.8: Average values across all remaining remaining assets (around 3300) and those of the bottom and top 1 percent

	EV1	EV2	EV3	EV4	# EVs > MP	Vol.	Corr.	Volume
Top 1%	0.921	0.031	0.017	0.012	1.617	0.030	0.208	168M
Bottom 1%	0.322	0.166	0.114	0.074	7.30	0.023	0.159	.65M
Mrkt Avg.	0.590	0.115	0.077	0.042	5.08	0.031	0.278	4.5M

The bottom 1 percent of all remaining assets in terms of the leading eigenvalue come from many sectors, with many from the financial sector representing retail investments, other investments, or asset managements. These are unheard of firms, unpopular with the media and their options carry idiosyncratic risk. The top 1 percent of assets come from many sectors including the financial sector. Most of these company are very popular and their options carry mostly systemic risk.

For completion, we also include histograms of each of the first and second eigenvalues, as well as the percent of variation explained by the top three eigenvalues, and the number of eigenvalues which exceed the MP-threshold. Results are displayed below.

1.7 Analysis of options on the remaining underlying assets

From the cross-sector statistics we see that on average, the statistics are very similar across different sectors. Since the average leading eigenvalue of the industrials sector is slightly lower, and the remaining eigenvalues are slightly higher we associate more idiosyncratic risk to this sector.

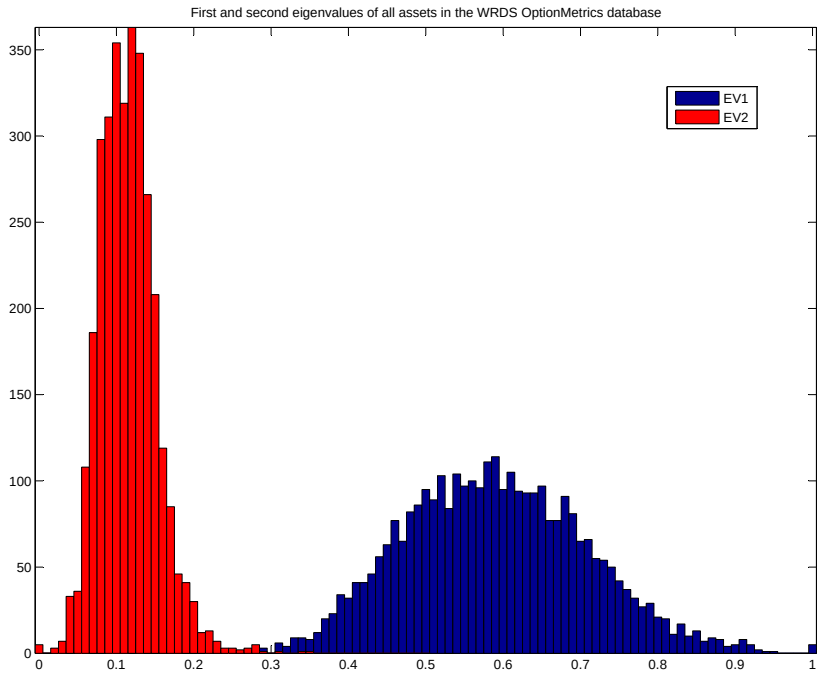


Figure 1.26: First and second eigenvalues of all assets in the OptionMetrics database from August 2004 to August 2013. Average value is 59 percent and 11.5 percent respectively.

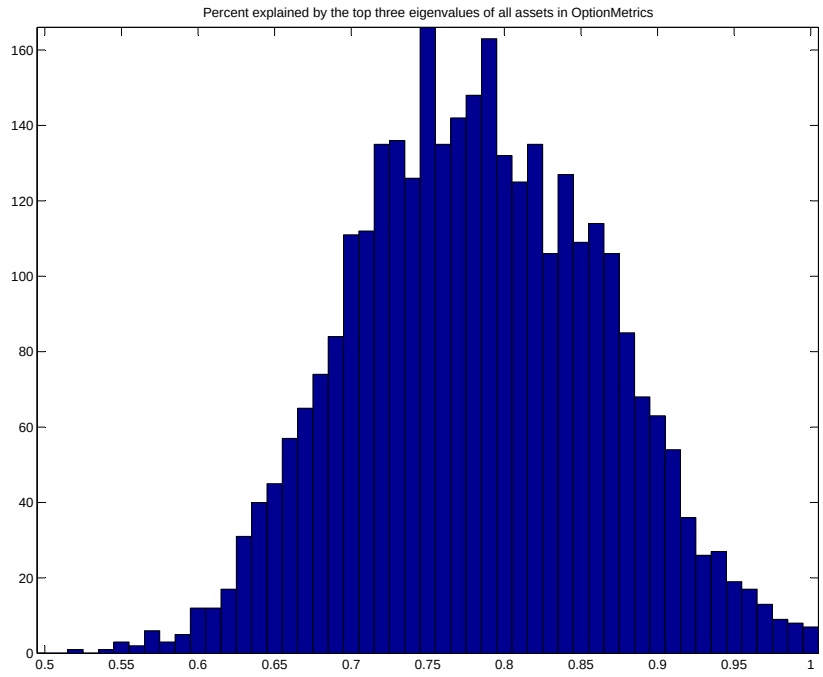


Figure 1.27: Percent of variation explained by the first three principal components of the 3300 remaining underlying assets in the OptionMetrics database from August 2004 to August 2013. Average value is 78 percent.

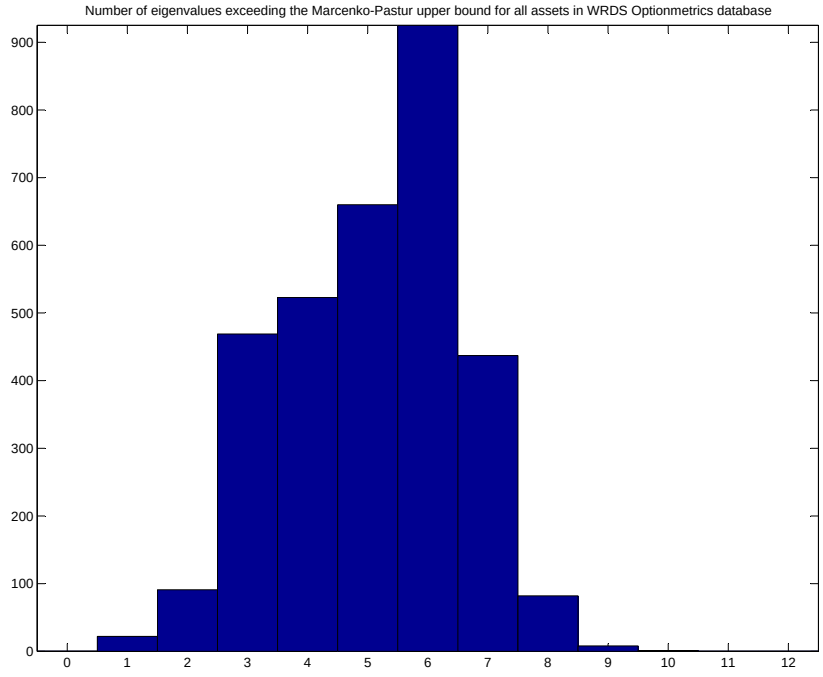


Figure 1.28: Number of eigenvalues exceeding the MP upper limit of the spectrum for the 3300 remaining underlying assets in the OptionMetrics database from August 2004 to August 2013. Average value is 5.1 eigenvalues.

Table 1.9: Results for all remaining remaining assets (around 3300) partitioned by industry sector

	EV1	EV2	EV3	EV4	# > MP-bound	Vol.	Corr.	Volume	MrkCap
Enrg	0.56	0.11	0.08	0.045	5.24	0.032	0.25	.98M	4385M
Mat.	0.59	0.10	0.07	0.042	5.09	0.038	0.27	1.1M	2433M
Indust.	0.58	0.11	0.07	0.04	5.24	0.032	0.356	.58M	1985M
CD	0.59	0.11	0.075	0.042	5.09	0.033	0.315	1.1M	3277M
All Assets	0.59	0.115	0.077	0.042	5.08	0.031	0.27	4.54M	2815M

[1] Enrg, Mat., Indust., CD, CS, HC, Fin, IT, TS, Util. are short for Energy, Materials, Industrials, Consumer Discretionary, Consumer Staples, Health Care, Financials, Information Technology, Telecommunication Services, and Utilities respectively.

[2] All results are on average for given sector. Data is collected on a daily basis from August 2004 until August 2013. About 2600 out of 3300 of all remaining assets are characterized into one of the sectors above.

[3] If we filter the data only for those stocks with market capitalization of at least $500M the results remain roughly unchanged.

Table 1.10: Results for all remaining assets partitioned by industry sector

	EV1	EV2	EV3	EV4	# > MP-bound	Vol.	Corr.	Volume	MrkCap
CS	0.57	0.11	0.078	0.043	5.12	0.029	0.29	.84M	3153M
HC	0.58	0.119	0.082	0.045	5.23	0.040	0.21	.56M	1835M
Fin	0.55	0.12	0.083	0.046	5.43	0.027	0.36	.69M	2989M
IT	0.59	0.118	0.078	0.043	5.15	0.035	0.268	1.2M	2202M
TS	0.53	0.12	0.085	0.049	5.73	0.030	0.347	1.88M	6709M
Util	0.50	0.127	0.089	0.051	6.07	0.031	0.431	.69M	4888M
All Assets	0.59	0.115	0.077	0.042	5.08	0.031	0.27	4.54M	2815M

[1] Enrg, Mat., Indust., CD, CS, HC, Fin, IT, TS, Util. are short for Energy, Materials, Industrials, Consumer Discretionary, Consumer Staples, Health Care, Financials, Information Technology, Telecommunication Services, and Utilities respectively.

[2] All results are on average for given sector. Data is collected on a daily basis from August 2004 until August 2013. About 2600 out of 3300 of all remaining assets are characterized into one of the sectors above.

[3] If we filter the data only for those stocks with market capitalization of at least $500M the results remain roughly unchanged.

1.8 Conclusion

In this chapter we gain a better understanding of which components drive changes in the IVS. We also characterize the type of risk carried by options. For each option contract in OptionMetrics with available data (about 3900 assets in total) we compute the PC surfaces as well as the corresponding spectrum of the

correlation matrix across 52 implied volatilities. We use tools from PCA as well as RMT to determine which part of the spectrum is significant. It turns out that we rarely need more than three or four of the leading components in order capture most of the change in the IVS. We also determine that the leading eigenvalue is an important indicator of the type of risk an option contract carries.

We classify option contracts into two classes: those carrying mostly systemic risk, and those carrying mostly idiosyncratic risk. Systemic risk increases with the first eigenvalue, while idiosyncratic risk decreases with the first eigenvalue, and increases with the second and third eigenvalues. The magnitude of the leading eigenvalue gives the percent of variation explained by the first principal component surface. We find that the first principal component surface is usually flat, and hence represents a parallel shift of the IVS. The larger the first eigenvalue, the greater the contribution of the parallel shift to the change of the IVS. This parallel shift is reflected in a roughly equal change across all implied volatility points.

On the other side of the risk-coin we have idiosyncratic risk. The higher the second and third eigenvalues, the more the variation of the IVS is explained by delta-skew and time-to-expiration-skew. In this case, we have unequal changes across the implied volatility points, thus reflecting unequal expectations of future asset prices in the options market.

In this chapter we also perform cross-sectional studies of the type of risk found in various types of assets, industries and asset-groupings. The results are summarized across different tables and figures.

Chapter 2

The Pivot Method and Determining Significant Components via Random Matrix Theory

2.1 Overview of results and goal of following sections

We have developed a systematic way of identifying the risk intrinsic to the options of an asset according to whether that asset faces mainly idiosyncratic or systemic risk. We saw that the leading eigenvalue corresponding to a PCA of the correlation matrix of log-return options data can be used to identify systemic risk. Likewise, the second, third, and (possibly) the fourth eigenvalues, as well as the number of eigenvalues exceeding the MP-bound can be used to identify

idiosyncratic risk. Furthermore, we characterized skew captured by the second and third principal components and determined if these factors are significant in capturing change in the IVS.

Options are forward-looking in time, so the idiosyncratic or systemic component of the underlying stock effects the option market via a forward-looking expectation of future news about the underlying asset. In effect, this is where the idiosyncratic or systemic component plays a large role. If we are faced with options on a market-index like SPY or QQQ, or a similar ETF representative of the economy, the leading eigenvalue is usually very large, hence we are dealing with systemic risk and must risk-manage accordingly. If we are faced with a sector-specific asset or a very opaque and unpopular one, we are faced with more idiosyncratic risk which is pertinent to the nature of the business itself, and this is the risk we face when trading its options.

We have constructed our implied volatility surfaces, as well as determined the principal components by using the 52 implied volatility pivots across the four expirations and thirteen deltas given in Chapter 1. In the next few sections, we examine to what extent we can reduce the number of implied volatility returns in modeling the change of the IVS. We refer to each individual implied volatility return as a pivot. We analyse how many pivots are necessary to capture the movement of the surface while keeping in mind that a large reduction in dimensionality is desirable due to the decrease in computation it provides.

In order to decrease dimensionality we utilize what we call a "pivot" model. As mentioned in the previous paragraph, pivots are just specific implied volatility returns which we use to generate all other implied volatilities. We accomplish this via time and delta projections and then linear interpolation. For example,

the implied volatility return corresponding to a 25-delta and 182-day expiration option could constitute one pivot. In chapter one we used 52 pivots to model our principal component surface. We can model the first few principal components of the surface via 52-pivots as we've already done so, but we seek to reduce the number of pivots used (i.e. reduce dimensionality) and still preserve the original structure of the IVS.

In the last section we employ the Tracy-Widdom distribution, as well as the Marchenko-Pastur distribution to determine the structure and the significant eigenvalues and eigenvectors of the entire options market[1]. We begin with the equities market as determined by the constituents of the S&P500, then move on the options market on the constituents of the S&P500. We end with a study of the options market as determined by those 3100 underlying assets with at least 500 days of data.

2.2 Stability of the first eigenvalue across time

It is interesting to know how the IVS and the spectrum, in particular the first eigenvalue, change during high-volatility and turbulent periods. So far we have used a static-in-time PCA approach. In order to see the evolution of the spectrum, particularly the first eigenvalue and the number of eigenvalues exceeding the MP-threshold we perform a dynamic i.e., a moving window across time PCA analysis. Beginning with August 31, 2004 we use a 252-day moving window to recompute the correlation matrix for options on each constituent of S&P500. We also perform this dynamic analysis on options of each of the liquid ETFs discussed in Chapter 1.

[1] As made available by the OptionsMetrics database.

We use these dynamic correlation matrixes to perform PCA and Marchenko-Pastur analysis across time.

We give the results for the two topmost (by leading eigenvalue) constituents of SPX and the two bottommost constituents of SPX: namely GS, JPM, and POM and WEC respectively. Also, we include the dynamic PCA results for the S&P500 index below.

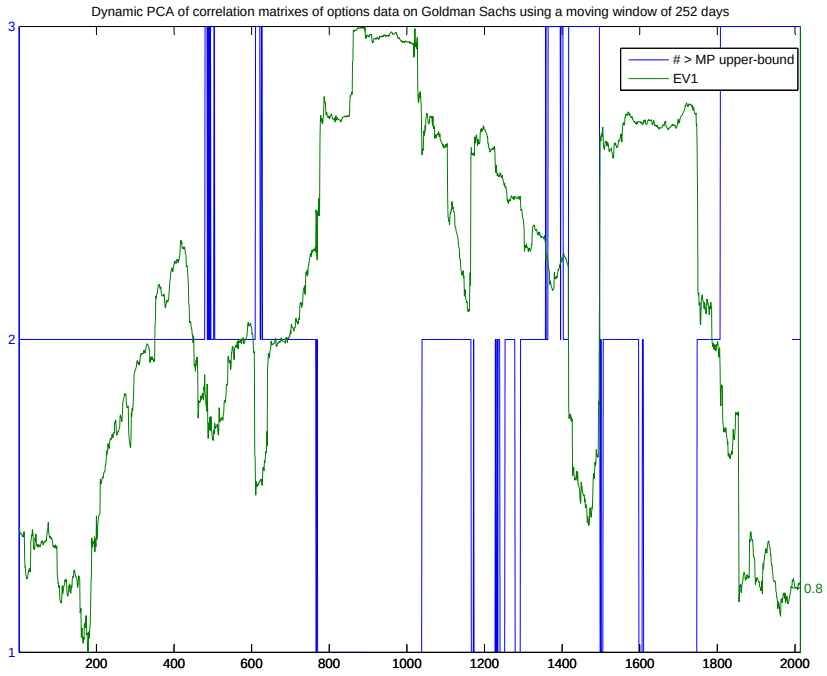


Figure 2.1: Dynamic PCA Results for GS using a moving window of 252-days from August 2004 to August 2013.

From these dynamic plots the inverse relationship between the leading eigenvalue and the number of eigenvalues exceeding the MP-threshold is confirmed. More importantly, in times of crises and very high volatility, such as October 2008 (Lehman Brothers bankruptcy), May 2010 (2010 Flash Crash), and November 2011

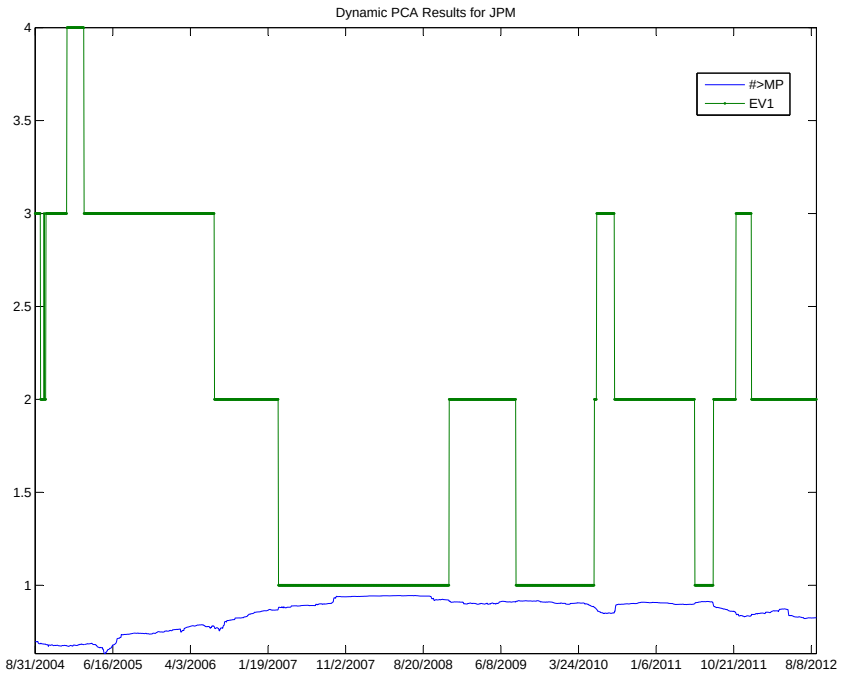


Figure 2.2: Dynamic PCA Results for JPM using a moving window of 252-days from August 2004 to August 2013.

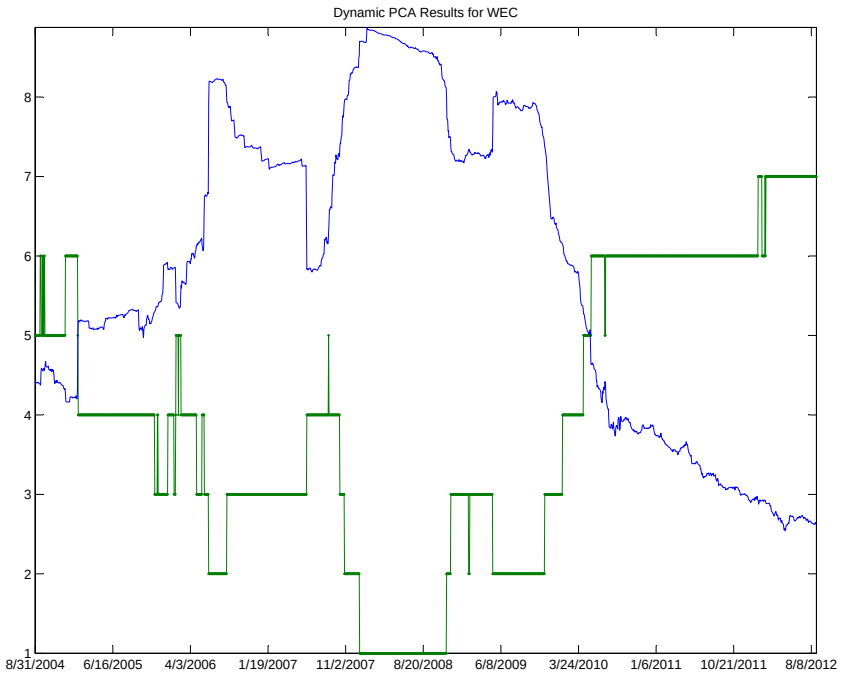


Figure 2.3: Dynamic PCA Results for WEC using a moving window of 252-days from August 2004 to August 2013. The percent of variation accounted for by the leading eigenvalue is multiplied by 10 for better comparison

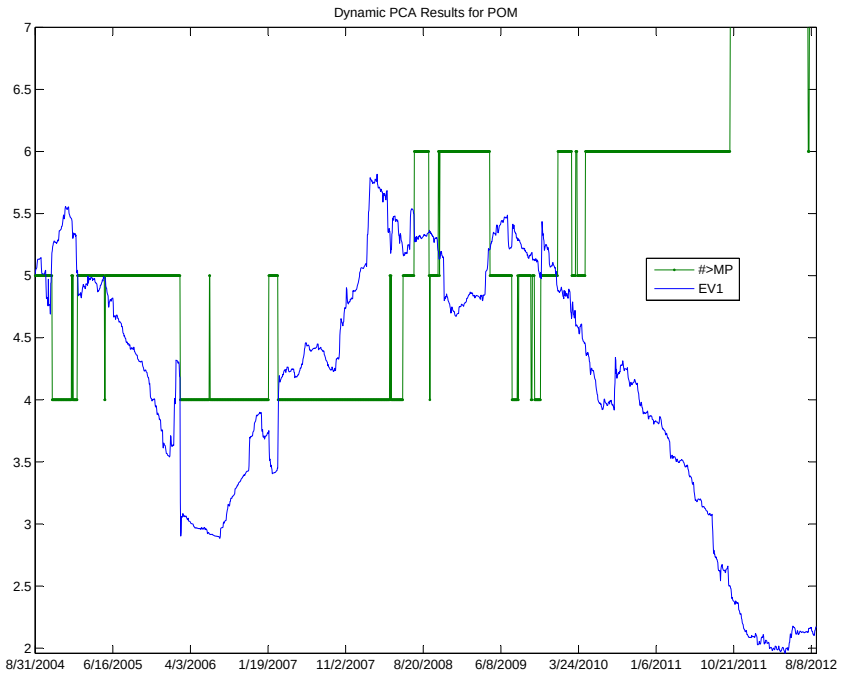


Figure 2.4: Dynamic PCA Results for POM using a moving window of 252-days from August 2004 to August 2013. The percent of variation accounted for by the leading eigenvalue is multiplied by 10 for better comparison

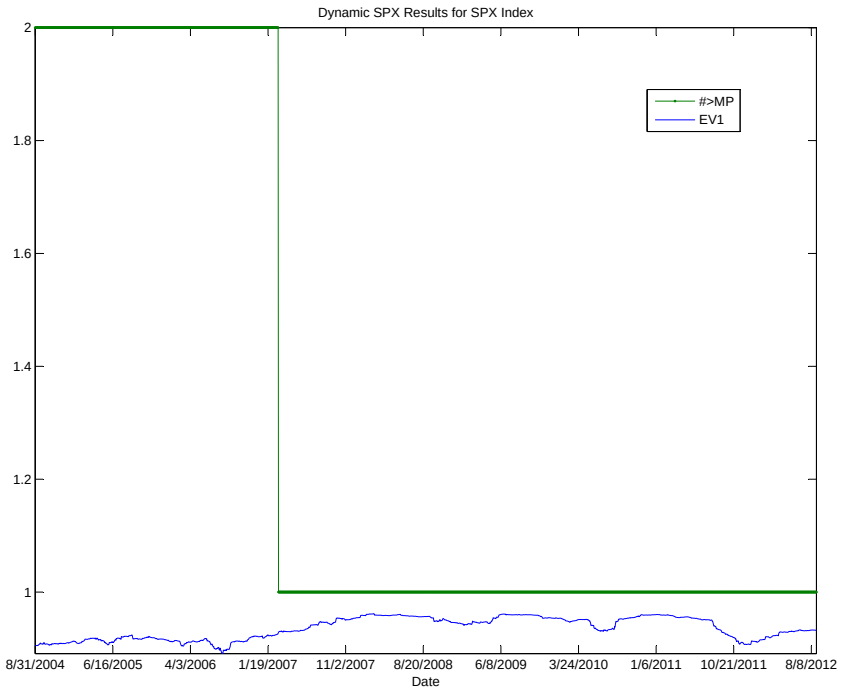


Figure 2.5: Dynamic PCA Results for the SPX index using a moving window of 252-days from August 2004 to August 2013.

(U.S. federal government credit-rating downgrades), the leading eigenvalue tends to increase, indicating the temporary increase in systemic risk associated with the option's market, i.e., every underlying asset begins to behave more like the market index.

From a risk-management perspective, and in the framework of our pivot model, the implications of these results are that in high-volatility periods, we actually need less pivots to model the dynamics of the IVS. Another way to see this is that since the leading eigenvalue captures the average correlation among all implied volatilities [see Appendix for proof], and since our dynamic analysis shows systematic increase in the first eigenvalue during high-volatility periods, this indicates an overall increase in the average correlation among all implied volatilities of the options market, hence less pivots are necessary for interpolation.

2.3 Pivot Method Description and Statistical Analysis of Pivot Method Quality

In this section we examine the results of various pivot models. We choose certain implied volatilities which we then use as pivots to get all other implied volatilities via interpolation. In this way we model the entire IVS by using only a few implied volatility returns. For example, if we choose n pivots, v_i for $i \in \{1, 2, ..., n\}$ then we seek to replicate any implied volatility return, say v_k for $k \notin \{1, 2, ..., n\}$ as

$$v_k = \sum_{i=1}^{n} \alpha_i v_i.$$

We formulate 2-pivot, 4-pivot, 5-pivot, 6-pivot, 7-pivot, 9-pivot and 12-pivot

models and test how well they preserve the spectrum of the correlation matrix of the original data. We also test how well each model preserves the distribution across various risk classes captured by the degree of systemic risk and idiosyncratic risk options face. The pivots used for each model are summarized in table 2.1.

Table 2.1: Implied Volatilities Used as Pivots for Each Model

	2-Pivots	4-Pivots	5-Pivots	6-Pivots	7-Pivot	9-Pivots	12-Pivots
25δ 30				YES	YES	YES	YES
50δ 30	YES	YES	YES			YES	YES
75δ 30				YES	YES	YES	YES
25δ 91				YES	YES		YES
50δ 91					YES		YES
75δ 91				YES	YES		YES
25δ 182		YES	YES			YES	YES
50δ 182			YES			YES	YES
75δ 182		YES	YES			YES	YES
25δ 365				YES	YES	YES	YES
50δ 365	YES	YES	YES			YES	YES
75δ 365				YES	YES	YES	YES

The right hand side indicates the implied volatility used, e.g., 25δ 365 refers to an option expiring in 365 days with strike at 25δ. Each pivot model has a YES indicating that it uses the pivot in that row.

Any of the above pivot methods provides a great reduction in the number of implied volatilities we have to use to generate our IVS. Once we have the pivots, we obtain any other point in the IVS by:

1. Locating the grid the point is located in.

2. Interpolating amongst the pivots delineating this grid.

In order to interpolate amongst the pivots, we project onto delta-space and time-to-expiration-space. So that if we let α be the projection along the time-to-maturity axis and β the projection along the δ axis, and we seek the value of the implied volatility return iv_{ret} at time t for some δ and k, $\text{iv}_{\text{ret}}(t, \delta, k)$, then we

obtain this via interpolation by projecting onto the nearest delta and the nearest time-to-expiration axis available:

$$\begin{aligned}\mathrm{iv}_{\mathrm{ret}}(t,\delta,k) = \beta(\alpha\mathrm{iv}_{\mathrm{ret}}(t,\delta_+,k_+) + (1-\alpha)\mathrm{iv}_{\mathrm{ret}}(t,\delta_+,k_-)) \\ + (1-\beta)(\alpha\mathrm{iv}_{\mathrm{ret}}(t,\delta_-,k_+) + (1-\alpha)\mathrm{iv}_{\mathrm{ret}}(t,\delta_-,k_-))\end{aligned}$$

Where $\alpha = \frac{k}{k_+ - k_-}$ and $\beta = \frac{\delta}{\delta_+ - \delta_-}$. Here $\delta_+ \geq \delta \geq \delta_-$ and $k_+ \geq k \geq k_-$ and the interpolation is done with pivots available in the specified model. In addition, for any other pivots (δ_1, k_1) we have $\delta_1 \geq \delta_+$ or $\delta_1 \leq \delta_-$, likewise $k_1 \geq k_+$ or $k_1 \leq k_-$. In other words, (δ_+, δ_-) gives the "tightest" delta pivot enclosure and (k_+, k_-) gives the "tightest" time-to-expiration pivot enclosure. If $\delta \geq 75$ then we simply interpolate via projection in time-to-expiration along the $\delta = 75$ axis. Likewise for $\delta \leq 25$.

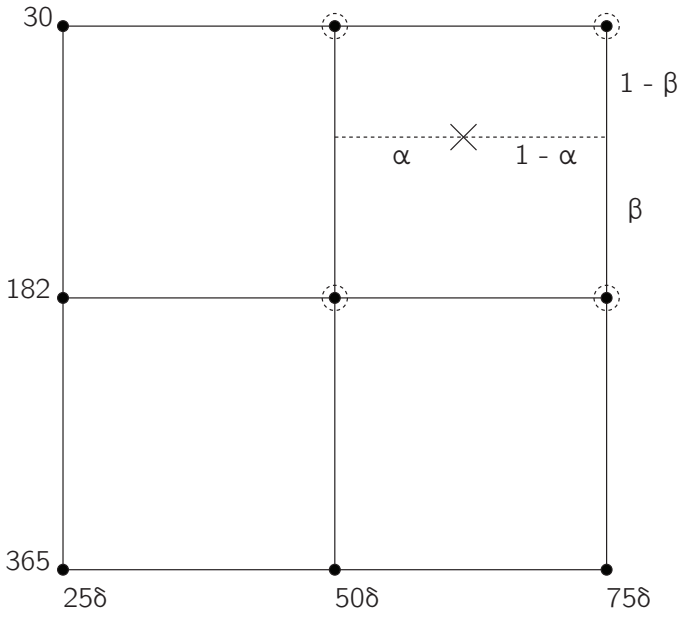


Figure 2.6: Example Schema of linear interpolation amongst the pivots via projection onto δ-space and time-to-expiration-space. In this schema we use 9-pivots, and the point we wish to replicate is denoted by an X.

In the first chapter we saw that the leading eigenvalues can be used to determine idiosyncratic or systemic risk of an option. In particular, the magnitude of the top three or four eigenvalues are very important indicators. Hence it is important that the magnitude of the top three or four eigenvalues of the correlation matrix is preserved when using any pivot model.

We interpolate the original return data matrix for each constituent of SPX using each of 2-pivots, 4-pivots, 5-pivots, 6-pivots, 7-pivots, 9-pivots and 12-pivots as shown in table 2.1. For each model, we use the interpolated data matrix to generate its corresponding correlation matrix, and then perform a PCA on this correlation matrix which generates the spectrum as well. We then compare the spectrum of the interpolated data correlation matrix with that of the original correlation matrix. We use the degree of agreement between the two spectra as one measure of the success of the model. We repeat this procedure for the topmost twenty liquid ETFs as well.

The graphs that follow display the difference between the first four eigenvalues generated by the pivot models and the original eigenvalues. As expected, the difference generally goes to zero with the number of pivots used, i.e., more pivots give better agreement with the original spectra.

For the constituents of SPX, there is little difference between the 4-pivot model and the 5-pivot model. The 6-pivot model performs better than both and is comparable to the 7-pivot model, but slightly outperforms it. The 9-pivot model does better than any of the previous models. The 12-pivot model performs best of all.

The same results hold for the twenty most liquid ETFs: the 2-pivot model is insufficient, the 4-pivot model and the 5-pivot model are very similar, and are both surpassed by the 6-pivot model. The 6-pivot model is almost indistinguishable from

the 7-pivot model. The 9-pivot model does better yet. The 12-pivot model does best of all and is in very good agreement with the original spectra.

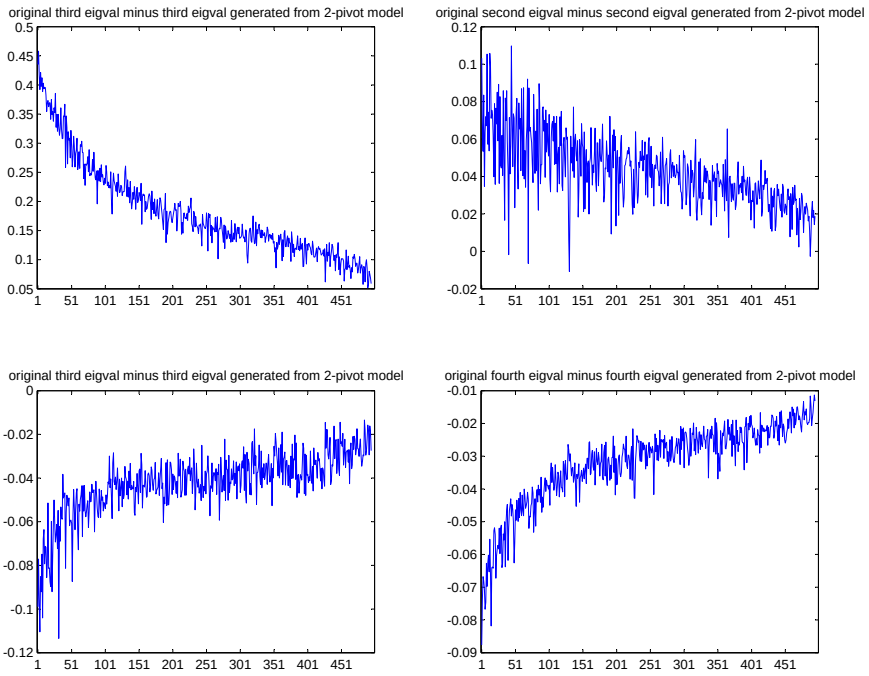


Figure 2.7: Difference between the first four eigenvalues of the 2-pivot model and those of the original spectra. Note that the x-axis is the list of all constituents of SPX in increasing order of their first eigenvalue as computed from the original data using all 52 implied volatility returns.

It is important to note that since the constituent stocks are ordered by increasing original leading eigenvalue, and all differences tend to zero, we see that the number of pivots used in our model is less important for options which carry more systemic risk and more important for those which carry more idiosyncratic risk. Otherwise said, a model with many pivots becomes more important for modeling the IVS of those options carrying mainly idiosyncratic risk. In particular, we saw

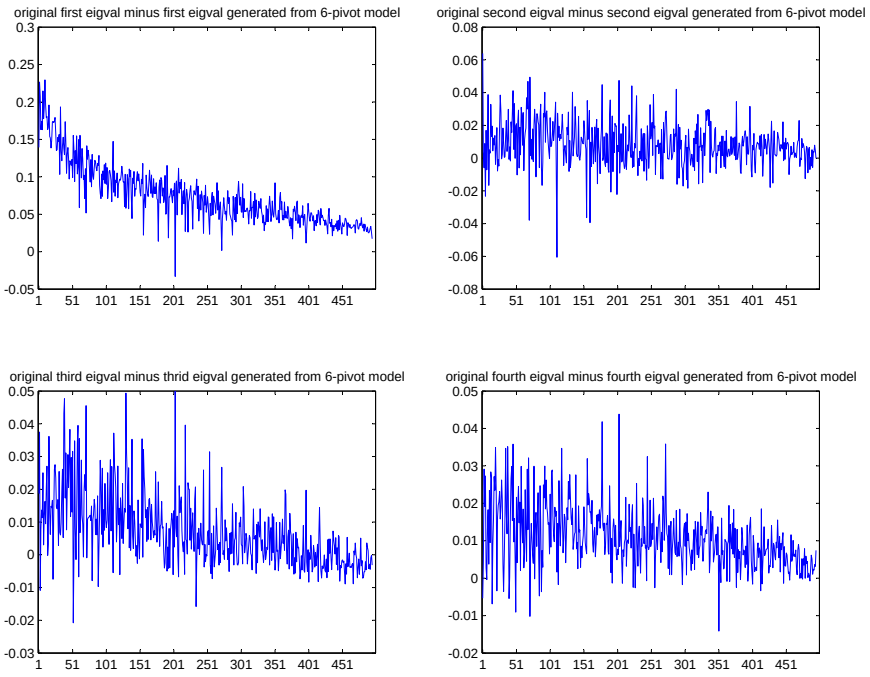


Figure 2.8: Difference between the first four eigenvalues of the 6-pivot model and those of the original spectra. Note that the x-axis is the list of all constituents of SPX in increasing order of their first eigenvalue as computed from the original data using all 52 implied volatility returns.

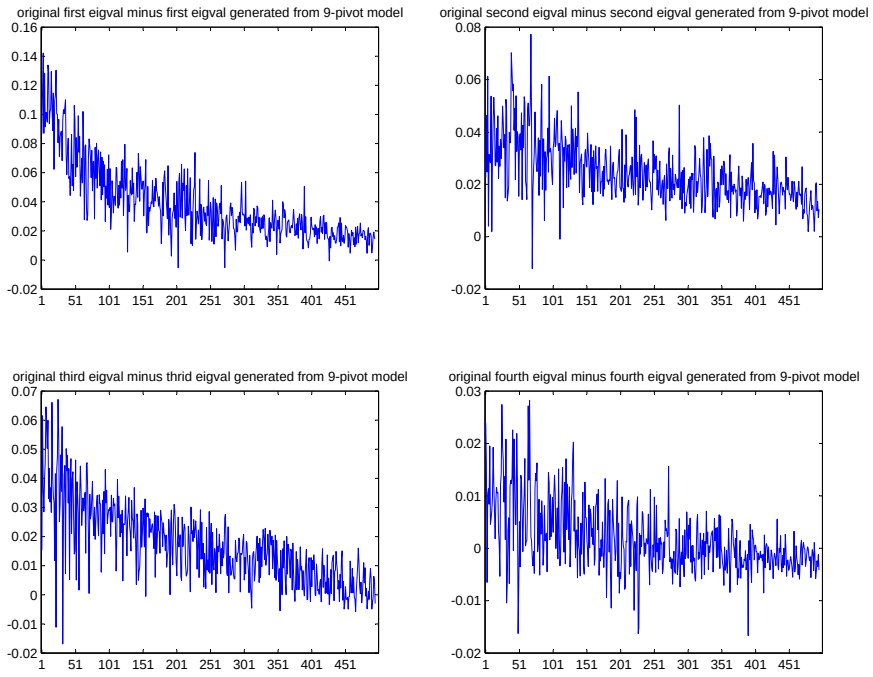


Figure 2.9: Difference between the first four eigenvalues of the 9-pivot model and those of the original spectra. Note that the x-axis is the list of all constituents of SPX in increasing order of their first eigenvalue as computed from the original data using all 52 implied volatility returns.

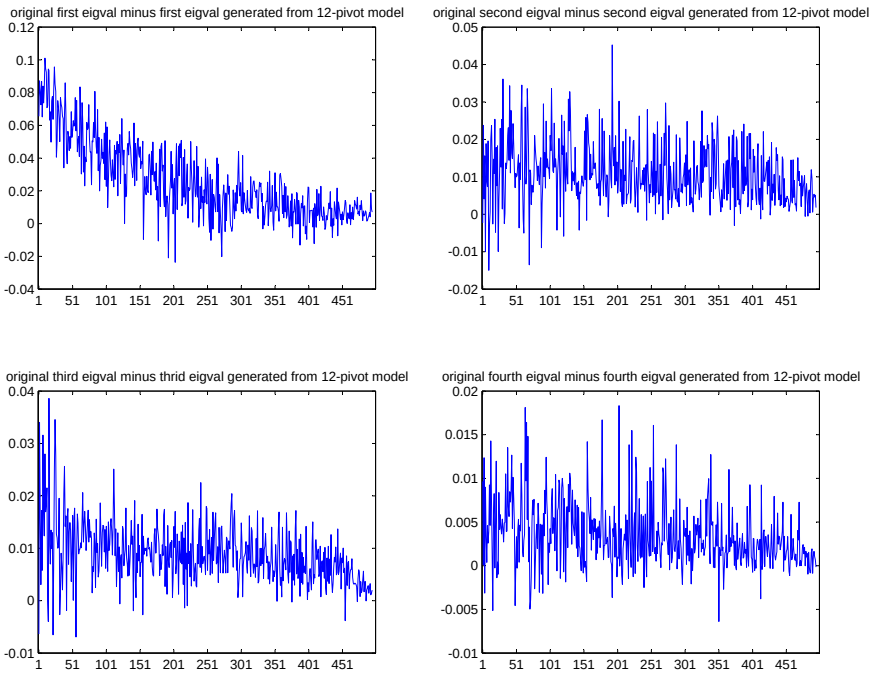


Figure 2.10: Difference between the first four eigenvalues of the 12-pivot model and those of the original spectra. Note that the x-axis is the list of all constituents of SPX in increasing order of their first eigenvalue as computed from the original data using all 52 implied volatility returns.

in section 2 that during very volatile periods the leading eigenvalue tends to increase indicating overall higher systemic risk. Hence the performance of any model improves during such periods.

In addition to studying how well a model preserves the critical eigenvalues of the spectrum, it is interesting to see how well it preserves the original distribution of stocks across different risk-classes. We classify the different risk-classes as follows:

- **Very Idiosyncratic**: first eigenvalue is more than two standard deviations less than the mean.
- **Idiosyncratic**: first eigenvalue is between one and two standard deviations less than the mean.
- **Somewhat Idiosyncratic**: first eigenvalue is between zero and one standard deviations less than the mean.
- **Somewhat Systemic**: first eigenvalue is between zero and one standard deviations more than the mean.
- **Systemic**: first eigenvalue is between one and two standard deviations more than the mean.
- **Very Systemic**: first eigenvalue is more than two standard deviations more than the mean.

We analyse how the distribution across these risk-classes changes with each model for the constituents of the S&P500. The original distribution across the six different risk classes is computed using all 52 implied volatility returns. The original distribution is given in figure 2.10.

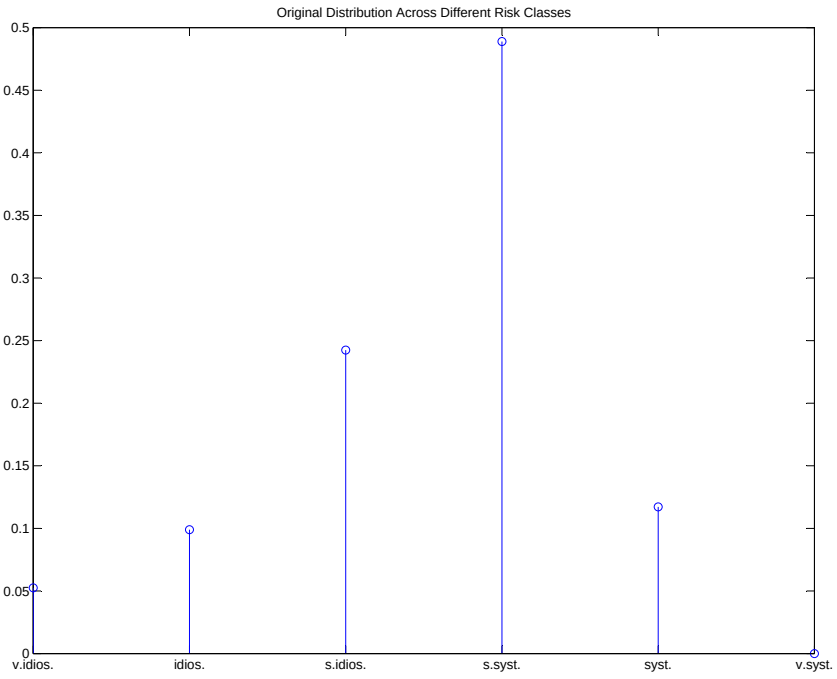


Figure 2.11: Original distribution across the various risk-classes using all 52-pivots for options on the constituents of S&P500.

From figure 2.10 we see that the vast majority of options whose underlying is a constituent of the S&P500 carry slight systemic risk, the next largest risk class is made up of those options with slight idiosyncratic risk. We perform the same analysis for each pivot model described in table 2.1, and display the results in figure 2.11. We compare how well each model preserves the initial distribution by overlaying the original distribution on top of each model distribution. We find that the distributions produced by the 4-pivot and the 5-pivot model produce very similar results, and both underestimate the original distribution more than the 6-pivot model. The 7-pivot model actually underestimates the original distribution more than the 6-pivot model. We include the 4-pivot, 5-pivot, and 7-pivot model in the appendix.

From Figure 2.11 we see that the 2-pivot model does a very bad job of preserving the original distribution across the various risk classes; it places most options under the systemic-risk class. The 6-pivot model performs considerably better than the 2-pivot model, but still deviates considerably from the original distribution. The 9-pivot and the 12-pivot models are very comparable, and perform very well in terms of agreeing with the original distribution.

The results of section 2 indicate that during high-volatile periods, there is an overall increase in systemic risk, and thus an overall increase in the first eigenvalue. From this section we know (Figures 2.6-2.9) that the discrepancy of any of the pivot models decreases with the first eigenvalue, hence any one of these models will perform even better during such periods when risk-management become even more crucial. In light of these results, we recommend the 9-pivot model for simulating fluctuations in the IVS. We have seen that the 9-pivot and 12-pivot models are very comparable, yet the 9-pivot model has the benefit of an additional 33 percent

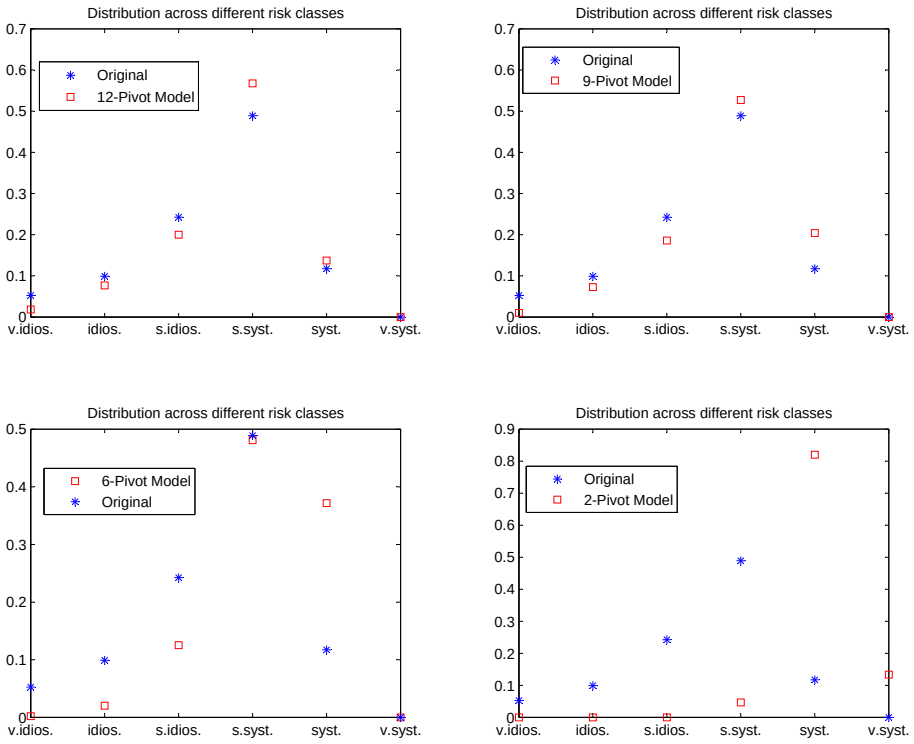


Figure 2.12: Original distribution across the various risk-classes using all 52-pivots for options on the constituents of S&P500 vs. the distribution produced by each model indicated.

dimensionality reduction.

2.4 PCA of the entire options market: tying things together

In chapter one we used PCA and random matrix theory on each individual option available in OptionMetrics. This allowed us to characterize the ways in which the IVS changes across time, as well as to determine the type of risk inherent to a specific option contract. We further used the spectrum of the correlation matrix across all implied volatility returns for a given option to determine those eigenvalues significant in capturing risk and in classifying risk.

In this chapter, we develop, test and recommend the 9-pivot model for simulating the IVS. We use all of these tools for our next goal: to perform PCA on the overall equities and options market. In other words, for each of roughly 3100 [2] options in OptionMetrics, we collect the historical data for the 9 implied volatilities used as pivots in our 9-pivot model. We use our model to generate all 52-implied volatility returns for each contract, and juxtapose this data across all option contracts together. If we used all original 130 implied volatility returns, the size of our data would be over 14 times larger; this would prove much more inefficient in terms of space and time required for execution.

The time period we use dates from August 31, 2004 until August 31, 2013. We make the restriction that each name and each contract should have at least 500 days of data. This results in 3141 possible underlying assets. The number of

[2] We only use those underlying assets with at least 500 days of available data

variables in the resulting options data matrix is approximately 10×3141[3] or 31,410.

As in chapter 1, we begin our analysis with the underlying constituents of S&P500. We first examine the equities market as determined by these constituents, and then move onto their options market. The number of observations we use is 2026. The number of variables in the resulting options data matrix is $10 \times 440 = 4400$, since we couple with the underlying asset, and use the 9-pivot model to generate all other implied volatility returns.

In this case, we find that 16 of the eigenvalues (out of the initial 440) are greater than the Marchenko-Pastur upper bound $\lambda_+ = 2.15$, and they account for roughly 55 percent of the variation in the overall surface. We give the top 20 eigenvalues below in Figure 2.13.

We have seen that for a single option, the top three or four eigenvalues are very distinguished, and thus it is easy to identify the significant components. In this case, it is not so apparent where we can draw the boundary between significant and non-significant components. In particular, we test whether we can use the statement and conclusion of the Marchenko-Pastur Law to determine this boundary. In addition to the Marchenko-Pastur distribution we test whether we can apply the Tracy-Widom Law to our data. Before we move on, we give the statement of Tracy-Widom as we apply it to our analysis, with the goal that we wish to distinguish signal from noise as determined by the spectrum. In other words, where does the random bulk of eigenvalues end, and where does the signal-carrying portion begin?

[3]We still couple with the underlying asset: 1+9 = 10, the first column corresponds to the underlying asset returns and the next 9 to the implied volatility returns used as pivots.

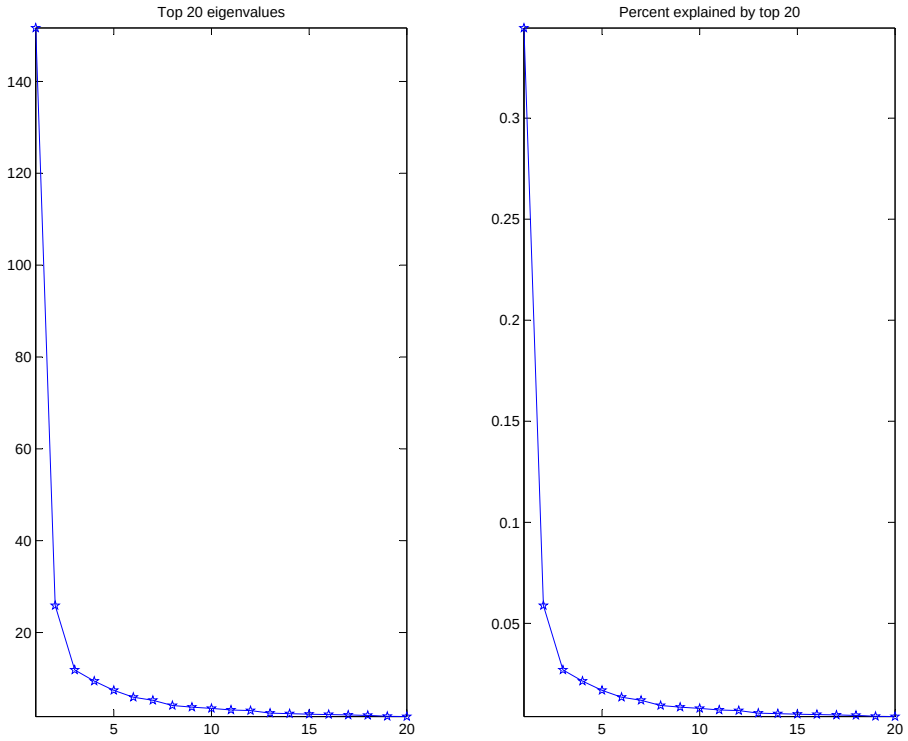


Figure 2.13: Top 20 eigenvalues of the equities market as determined by S&P500 stocks and their explanatory power.

- Tracy-Widom: The distribution of the largest eigenvalue, $\lambda_{\max}$, of a random correlation matrix is given by:

$$Pr(T\lambda_{\max} < \mu_{TN} + s\sigma_{TN}) = F_1(s)$$

- with

$$\mu_{TN} = (\sqrt{T - .5} + \sqrt{N - .5})^2$$

and

$$\sigma_{TN} = (\sqrt{T - .5} + \sqrt{N - .5})\left(\frac{1}{\sqrt{T - .5}} + \frac{1}{\sqrt{N - .5}}\right)^{\frac{1}{3}}$$

- Tracy-Widom holds for specific $\beta-$ensembles: Gaussian Orthogonal Ensemble: $\beta = 1$, Gaussian Unitrary Ensemble: $\beta = 2$, and Gaussian Symplectic Ensemble: $\beta = 4$. We use $\beta = 1$.

$F_1(s)$ is stated in terms of the Painleve II differential equation, and other equivalent formulations. We leave that part of the theory out as it is not directly relevant. We use the $s \to F_1(s)$ table from Michael's Prahofer's website: http://www-m5.ma.tum.de/KPZ. In figure 2.14 we plot the CDF $s, F_1(s).For a more thorough treatment of this stl$

Tracy-Widom holds only for specific distributions of the underlying data. In particular, it has been shown [Bouchaud et. al.] that fat-tails can massively increase the maximum eigenvalue in the theoretical limiting spectrum of the random matrix. It would be of interest to know how applicable both Marchenko-Pastur (MP) and Tracy-Widom (TW) are when we consider a random matrix with the same underlying distribution as our original data.

We test the performance of MP and TW by taking the original data, and for each variable (i.e., implied volatility return or underlying asset return) we permute

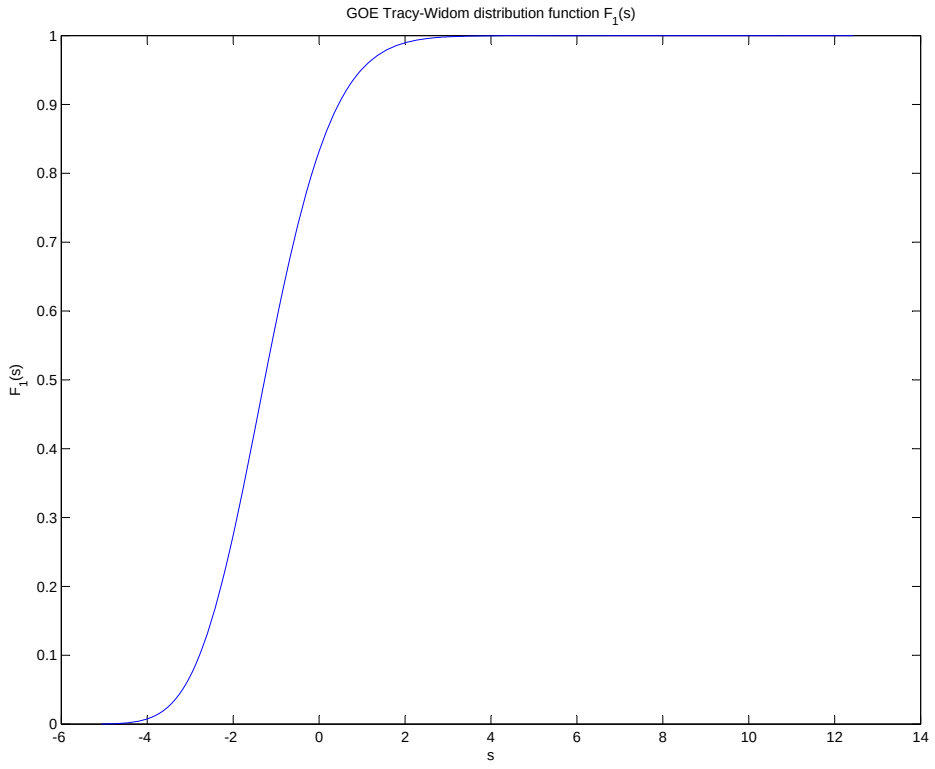


Figure 2.14: Tracy-Widom CDF for the $\beta = 1$ gaussian orthogonal ensemble case. 100 percentile $\rightarrow$ s=13.63, 99 percentile $\rightarrow$ s=2.06, and 95 percentile $\rightarrow$ s=1.

its time series independently of the time-series of any other variable. We then test how the results compare to those predicted by MP and TW.

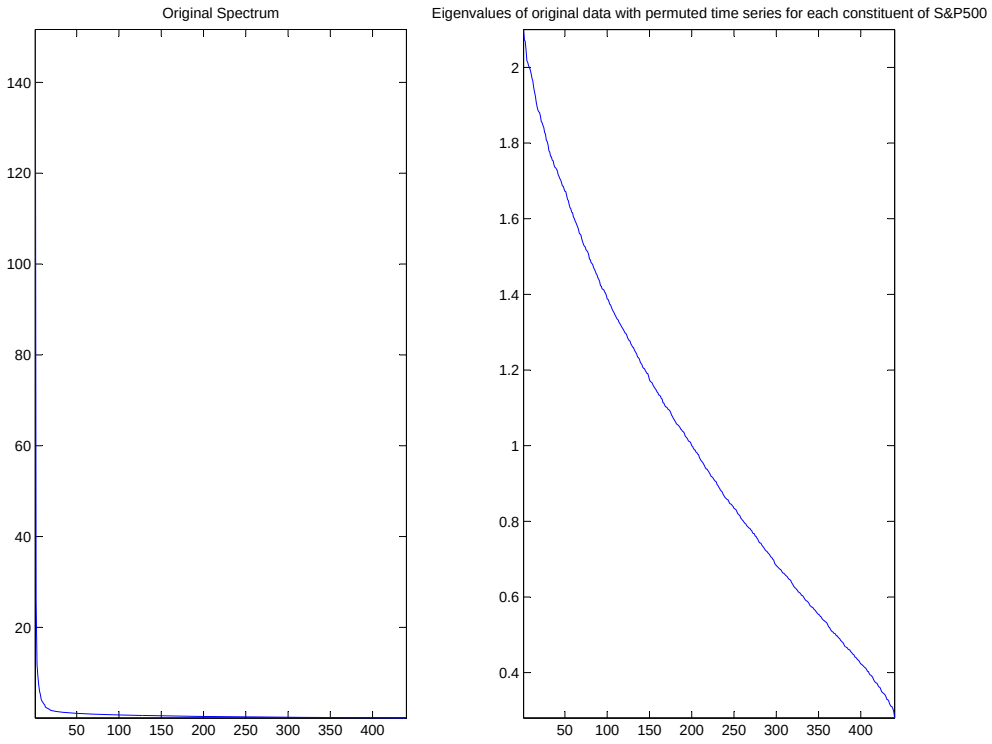


Figure 2.15: Eigenvalues of the correlation matrix of original data, and those of the correlation matrix of time-series permuted data. Based on equities market as determined by S&P500 stocks. 2026 observations and 440 variables. The bulk of the random spectrum lie in the theoretical range $(\lambda_-, \lambda_+) = (.29, 2.15)$ according to MP.

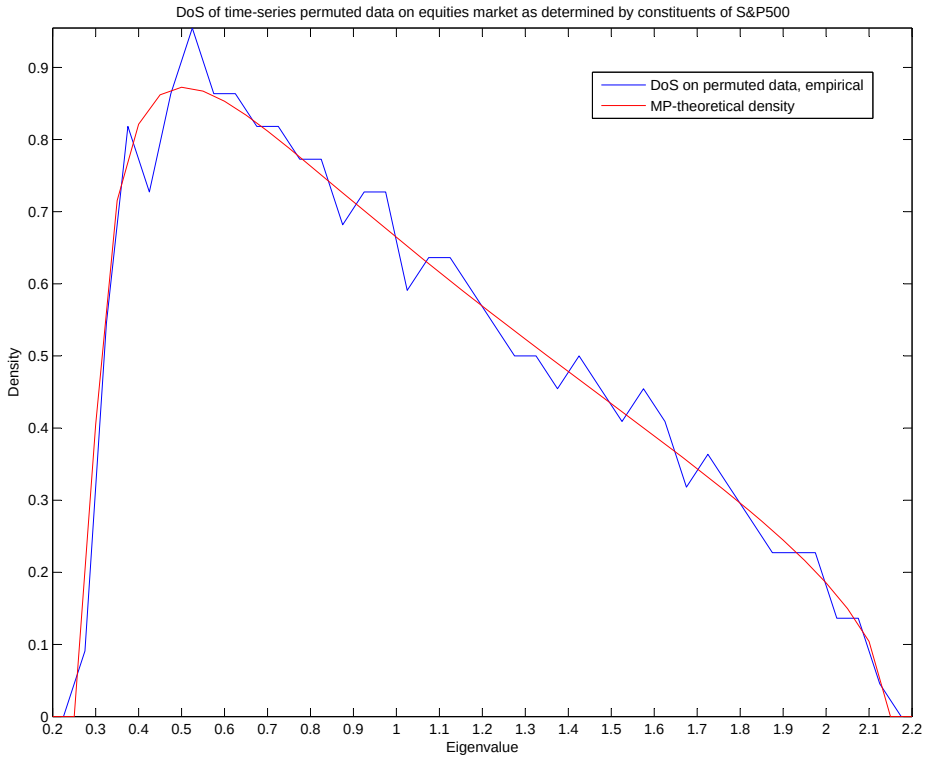


Figure 2.16: Empirical density of all eigenvalues of correlation matrix of randomly permuted time-series on original data versus Marchenko-Pastur distribution. We cannot reject the null hypothesis that the two distributions are the same based on Kolmogorov-Smirnov 2-series test at the 1 percent significance level. Based on equities market as determined by S&P500 stocks. 2026 observations and 440 variables.

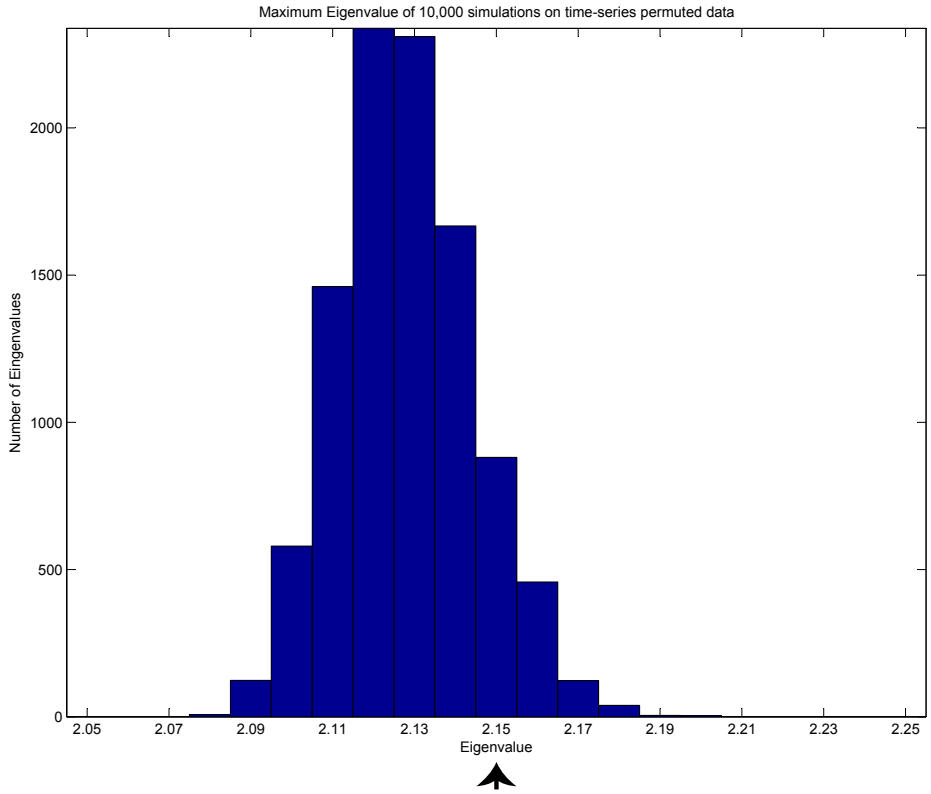


Figure 2.17: Largest eigenvalue of the correlation matrix of time-series permuted data. Arrow points to the theoretical value of $\lambda_+ = 2.15$ from Marchenko-Pastur theory. Ninety one percent of all 10,000 simulations lie below λ_+, and all lie in the range $(2.08, 2.22)$. Based on equities market as determined by S&P500 stocks. 2026 observations and 440 variables.

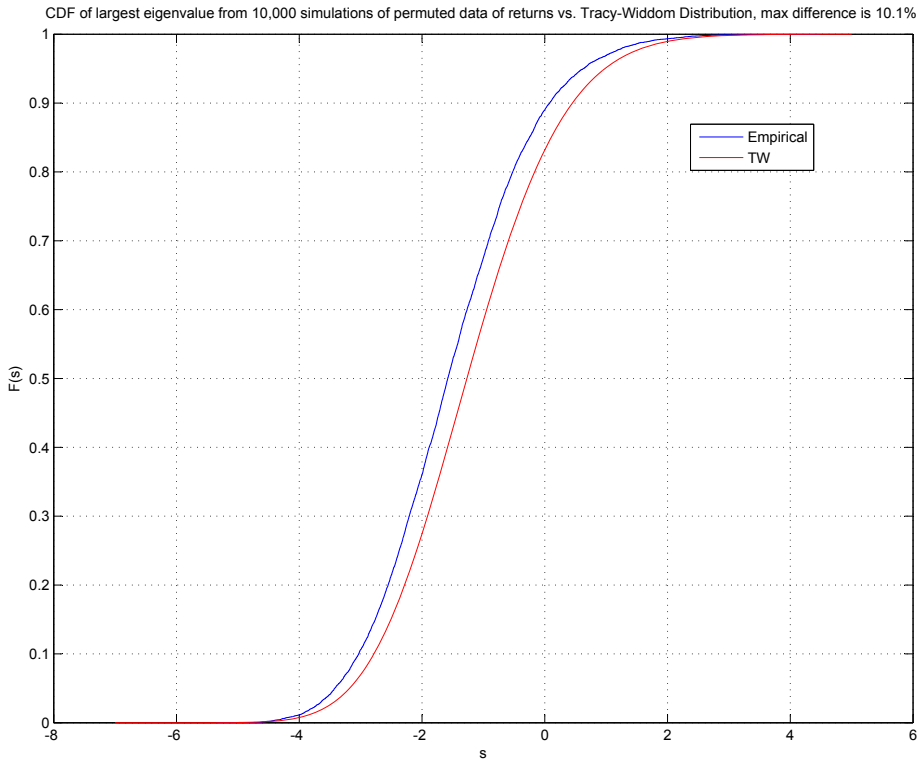


Figure 2.18: Cumulative density of maximum eigenvalue of randomly permuted time-series on original data versus Tracy-Widom. We cannot reject the null hypothesis that the two distributions are the same based on Kolmogorov-Smirnov one-series test at the 1 percent significance level. Based on equities market as determined by S&P500 stocks. 2026 observations and 440 variables.

Table 2.2: Significance of Eigenvalues in the S&P500 Equities Market a la TW

Top 16 Eigenvalues	s-value	$F_1(s)$
$\lambda_1 = 151.63$	11,144	1
$\lambda_2 = 25.88$	1,769	1
$\lambda_3 = 11.85$	724	1
$\lambda_4 = 9.42$	543	1
$\lambda_5 = 7.41$	393	1
$\lambda_{10} = 3.48$	100	1
$\lambda_{11} = 3.14$	74	1
$\lambda_{12} = 2.99$	63	1
$\lambda_{15} = 2.21$	5	1
$\lambda_{16} = 2.15$	.15	.86

- 440 assets and 2026 days used.
- Corresponding $\lambda_+ = 2.15$, 16 eigenvalues exceed it and account for 55 percent of variation.
- 15 eigenvalues and their corresponding eigenvectors are deemed significant by Tracy-Widdom.
- We interpret this to mean the S&P500 equities market is driven by 15 factors.

We now present the results on the same inquiries as above for the options market with underlying a constituent of S&P500. The number of eigenvalues which exceed the MP threshold $\lambda_+ = 6.12$ is 84, and these eigenvalues account for 55 percent of the variation.

It is interesting to note that the more insignificant eigenvectors are distributed according to the maximum entropy distribution (i.e., standard gaussian in this case), whereas the corresponding top eigenvectors deviate from this distribution indicating more structure. We show this for options on stocks in S&P500, but the same phenomena holds for the equities market on S&P500, the overall equities market, and the overall options market.

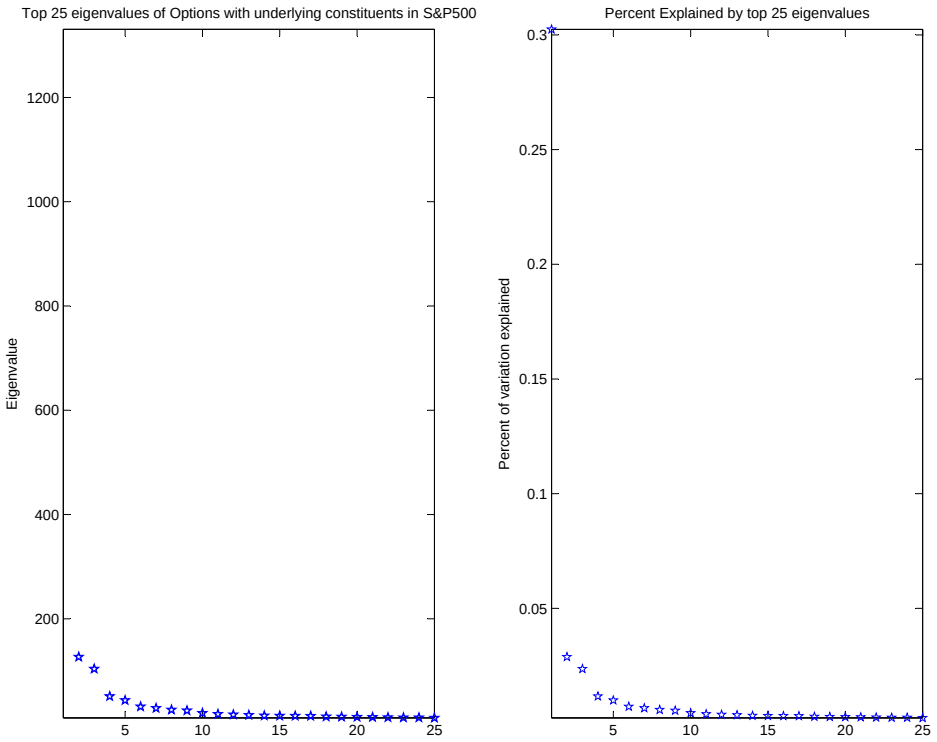


Figure 2.19: Top 25 eigenvalues of the options market as determined by those options with an underlying stock in S&P500. They explain 45 percent of the variation in the IVS. Data uses 2026 sample days and 4400 variables.

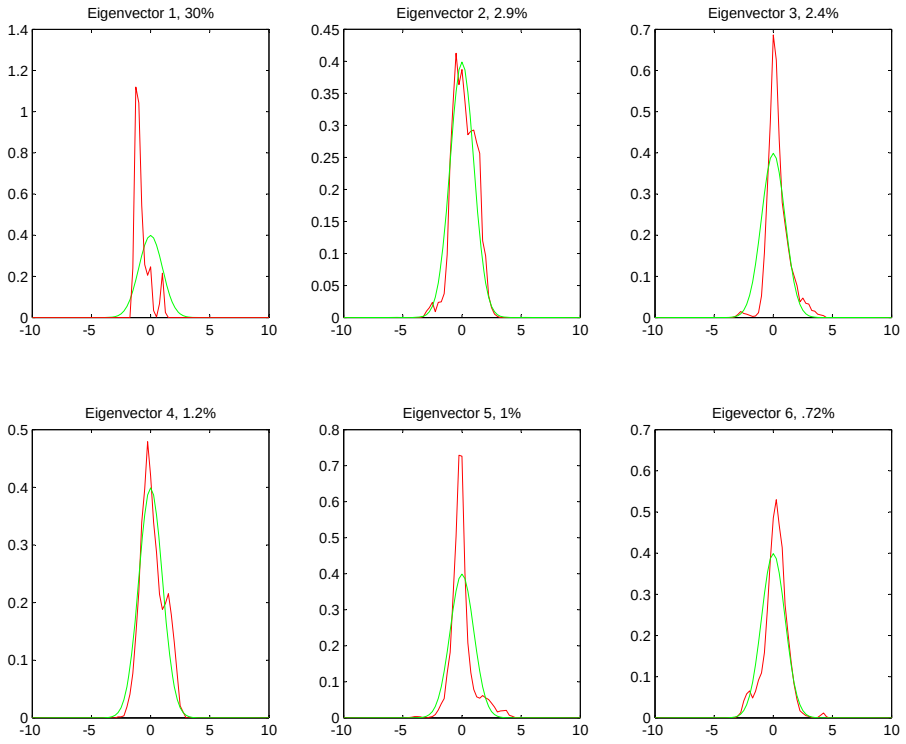


Figure 2.20: Top 6 eigenvectors and the percent of variation explained by each superimposed with the standard Gaussian distribution. Data uses 2026 sample days and 4400 variables.

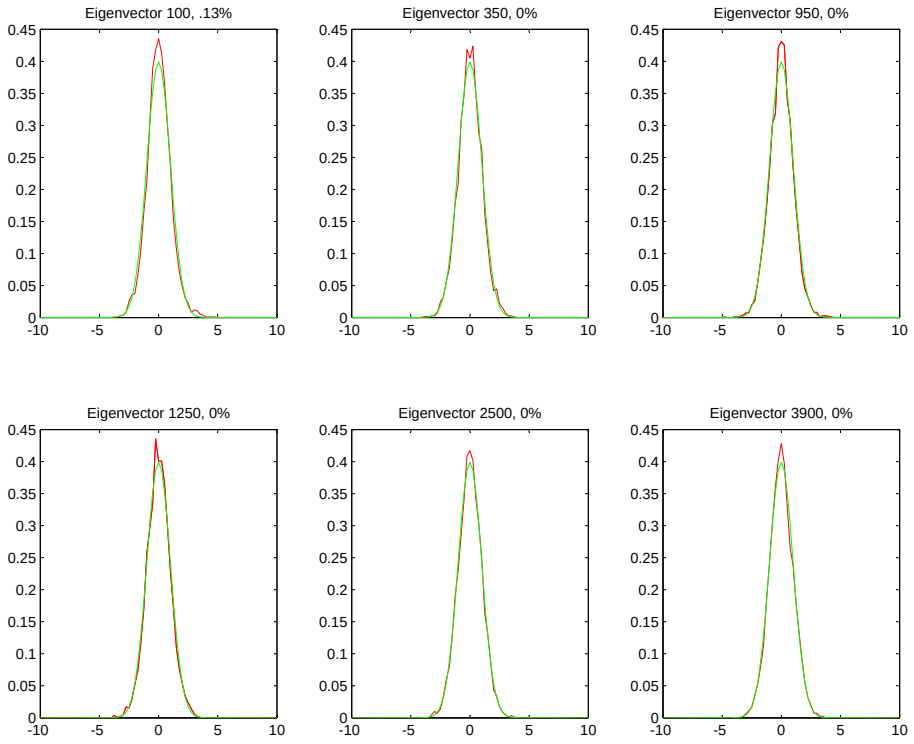


Figure 2.21: Random 6 eigenvectors from the bulk and the percent of variation explained by each superimposed with the standard Gaussian distribution. Data uses 2026 sample days and 4400 variables.

We now test how Marchenko-Pastur and Tracy-Widom perform when we consider a random matrix with the same underlying distribution as our options data. As before, we take the original data, and for each variable we permute its time series independently of the time-series of any other variable. We present some relevant results below.

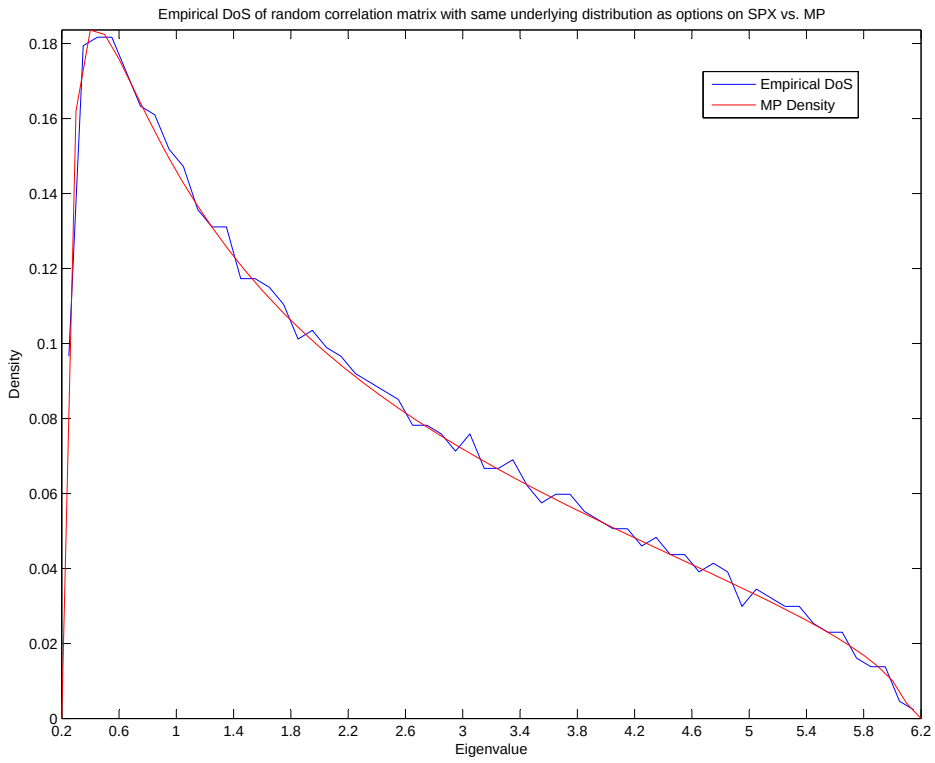


Figure 2.22: Empirical density of all eigenvalues of correlation matrix of randomly permuted time-series on original options data on S&P500 constituents versus Marchenko-Pastur distribution. 4400 variables each over 2026 observations are used. For both distributions there is a point mass of weight 54 percent at zero. We cannot reject the null hypothesis that the two distributions are the same based on Kolmogorov-Smirnov 2-series test at the 1 percent significance level.

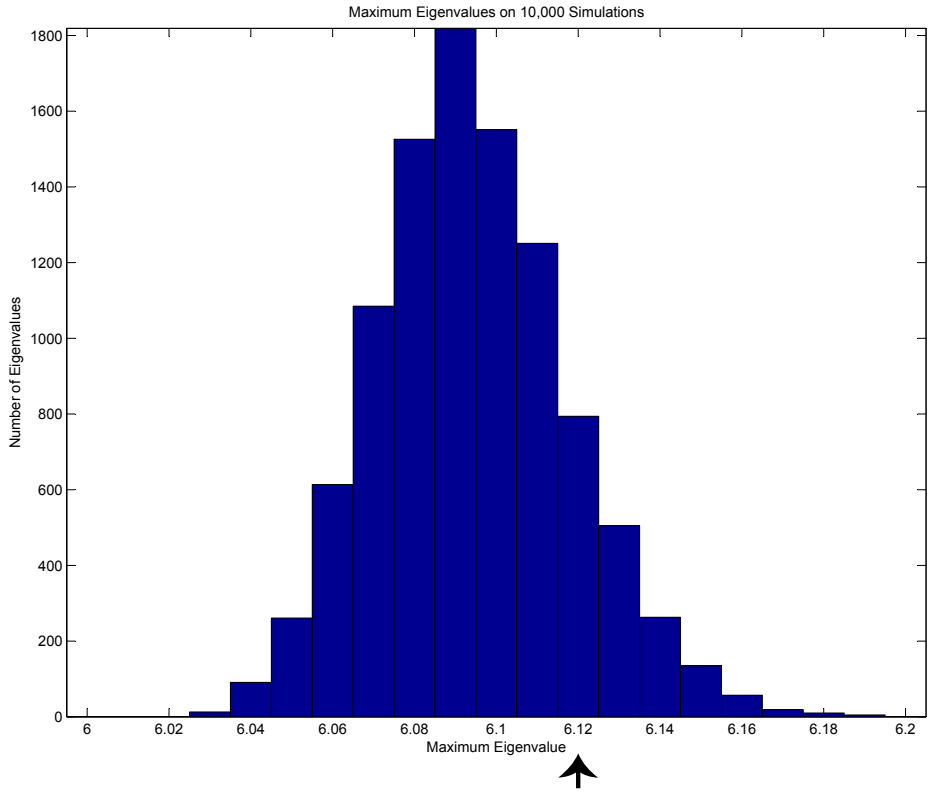


Figure 2.23: Largest eigenvalue of the correlation matrix of time-series permuted data on options with underlying asset in the S&P500. Arrow points to the theoretical value of $\lambda_+ = 6.12$ from Marchenko-Pastur theory. Eighty six percent of all 10,000 simulations lie below λ_+, and all lie in the range $(6.03, 6.19)$. Uses 2026 observations and 4400 variables.

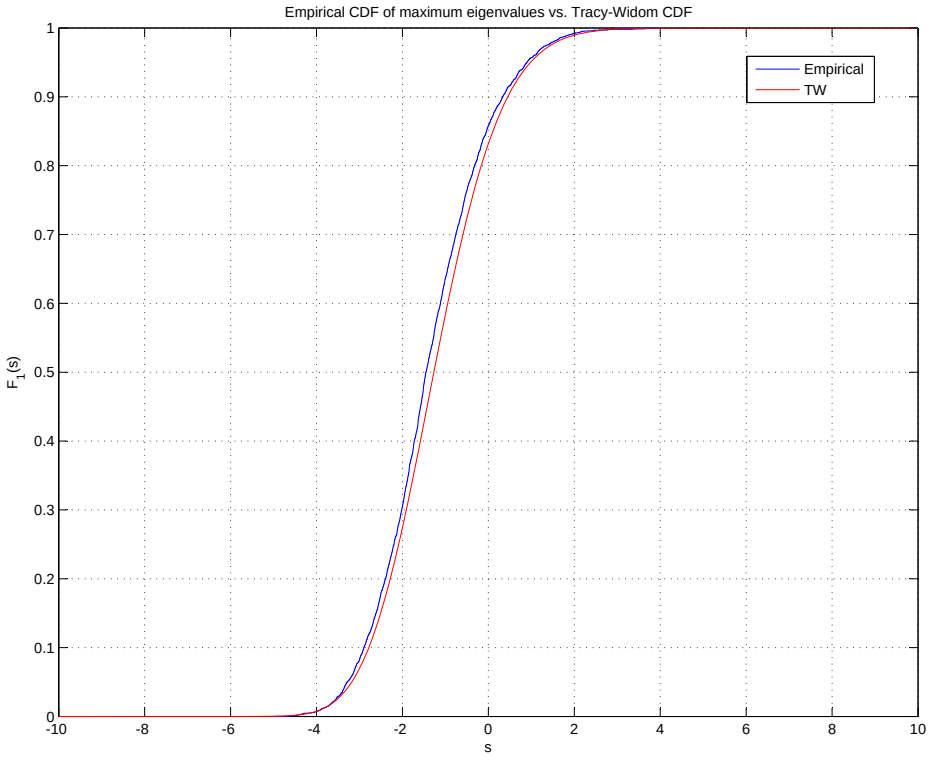


Figure 2.24: Cumulative density of maximum eigenvalue of randomly permuted time-series on original data of options on S&P500 constituents versus Tracy-Widom. We cannot reject the null hypothesis that the two distributions are the same based on Kolmogorov-Smirnov one-series test at the 1 percent significance level. We use 2026 observations and 4400 variables. Overall 10,000 simulations were performed and the corresponding eigenvalue in each case was computed.

Table 2.3: Significance of Eigenvalues for Options with an underlying in S&P500

Top 85 Eigenvalues	s-value	$F_1(s)$
$\lambda_1 = 1331$	72143	1
$\lambda_5 = 44$	2042	1
$\lambda_{10} = 19.2$	713	1
$\lambda_{20} = 11.5$	295	1
$\lambda_{40} = 8.2$	115	1
$\lambda_{50} = 7.6$	79	1
$\lambda_{60} = 7.1$	53	1
$\lambda_{70} = 6.4$	28	1
$\lambda_{80} = 6.3$	8.4	.99
$\lambda_{84} = 6.1$	.99	.95
$\lambda_{85} = 6.09$	-1.53	.41

- 4400 assets and 2026 days used.
- Corresponding $\lambda_+ = 6.12$, 84 eigenvalues exceed it and account for 55 percent of variation.
- 84 eigenvalues and their corresponding eigenvectors are deemed significant by Tracy-Widdom.
- We can interpret this to mean the S&P500 options market is driven by 84 components.

We now present results on the same inquiries as above for the equities market as a whole as determined by available data in OptionMetrics, and for those assets with at least 500 days of data. This amounts to 3141 variables and 500 observations. The number of eigenvalues which exceed the MP threshold $\lambda_+ = 12.30$ is 20, and these eigenvalues account for 24 percent of the variation.

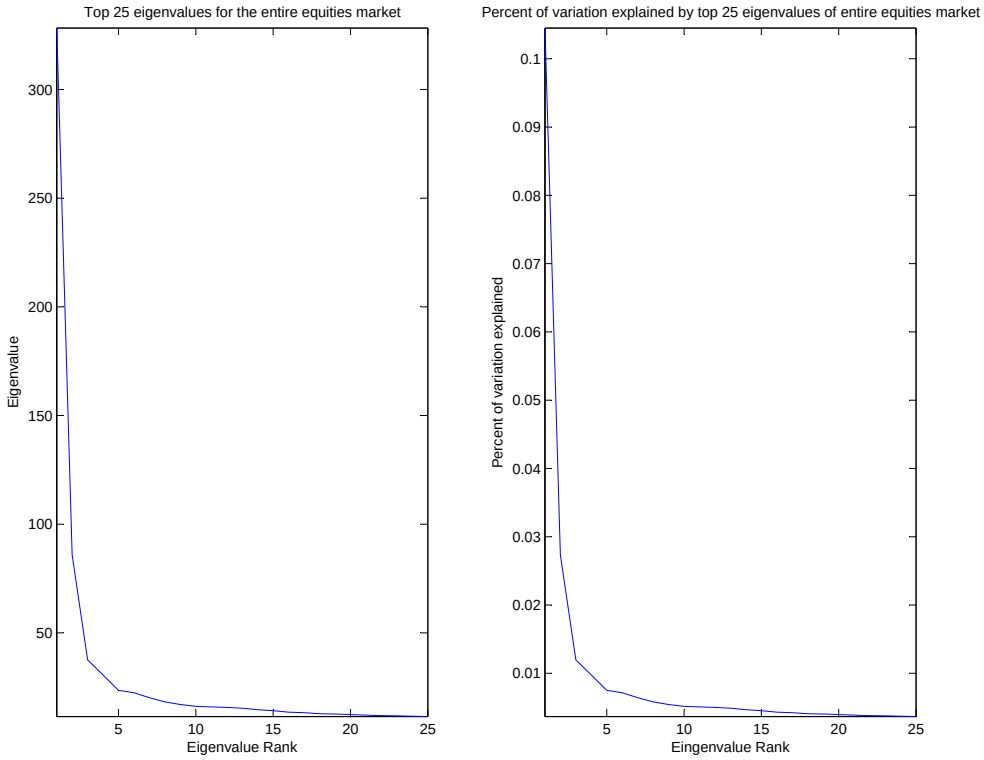


Figure 2.25: Top 25 eigenvalues of the equities market. They explain 25 percent of the variation in the IVS. Data uses 500 sample days and 3141 variables.

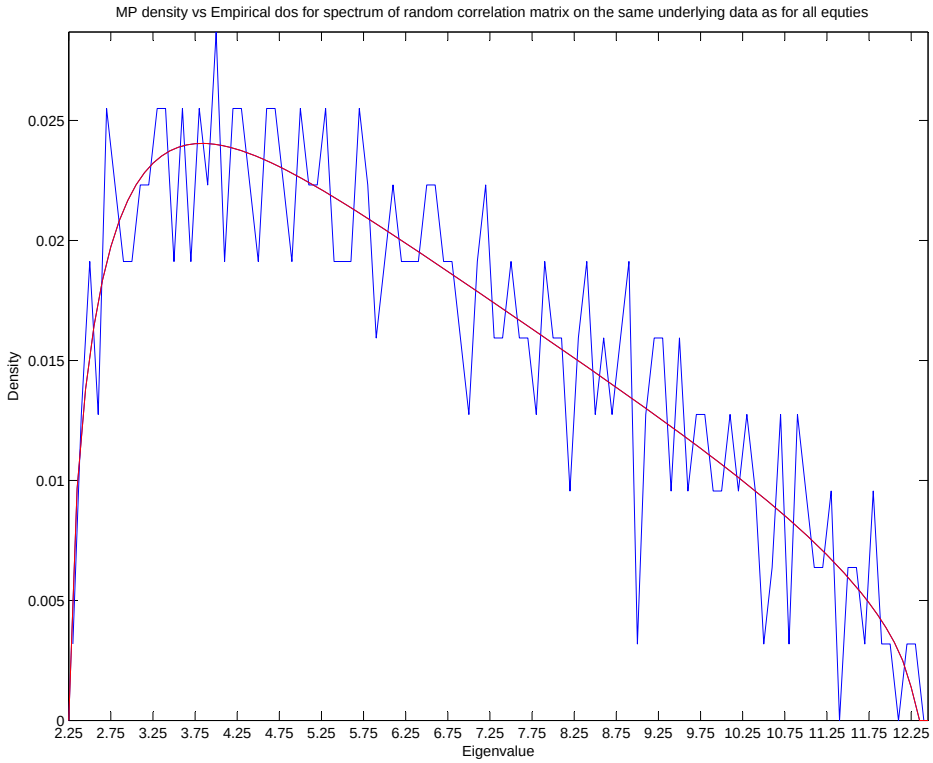


Figure 2.26: Empirical density of all eigenvalues of correlation matrix of randomly permuted time-series on all original equities data versus Marchenko-Pastur distribution. For both distributions there is a point mass of weight 84 percent at zero. 3141 variable used each over 500 observations. We cannot reject the null hypothesis that the two distributions are the same based on Kolmogorov-Smirnov 2-series test at the 1 percent significance level.

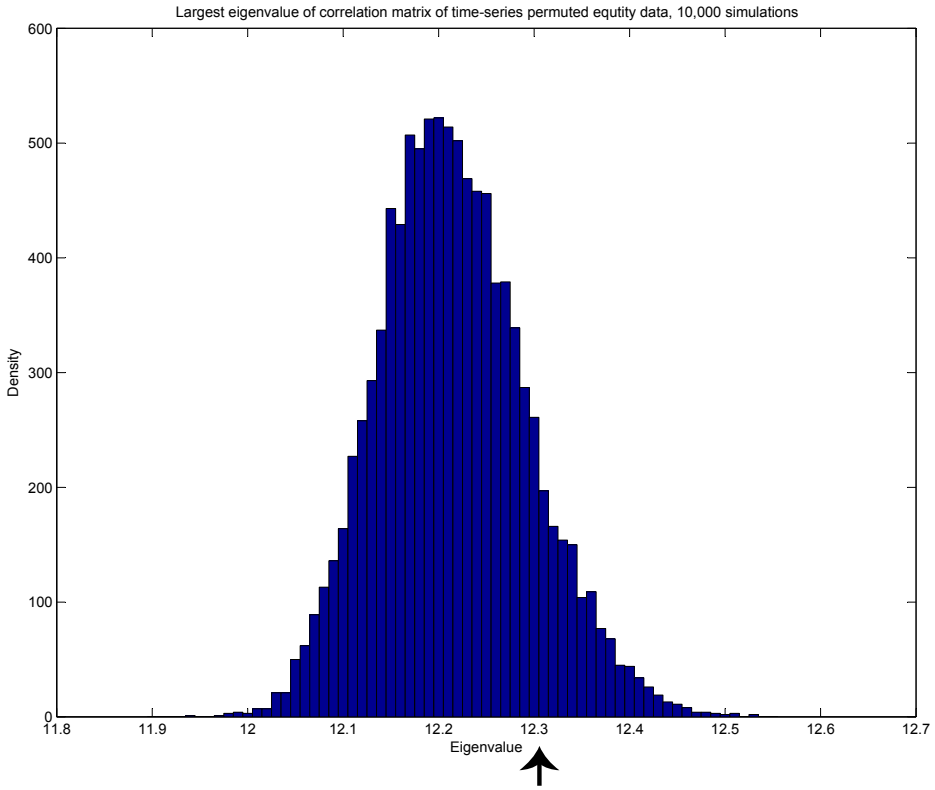


Figure 2.27: Largest eigenvalue of the correlation matrix of time-series permuted data on all equity data. Arrow points to the theoretical value of $\lambda_+ = 12.29$ from Marchenko-Pastur theory. Eighty five percent of all 10,000 simulations lie below λ_+, and all lie in the range $(11.94, 12.53)$. Uses 2026 observations and 4400 variables.

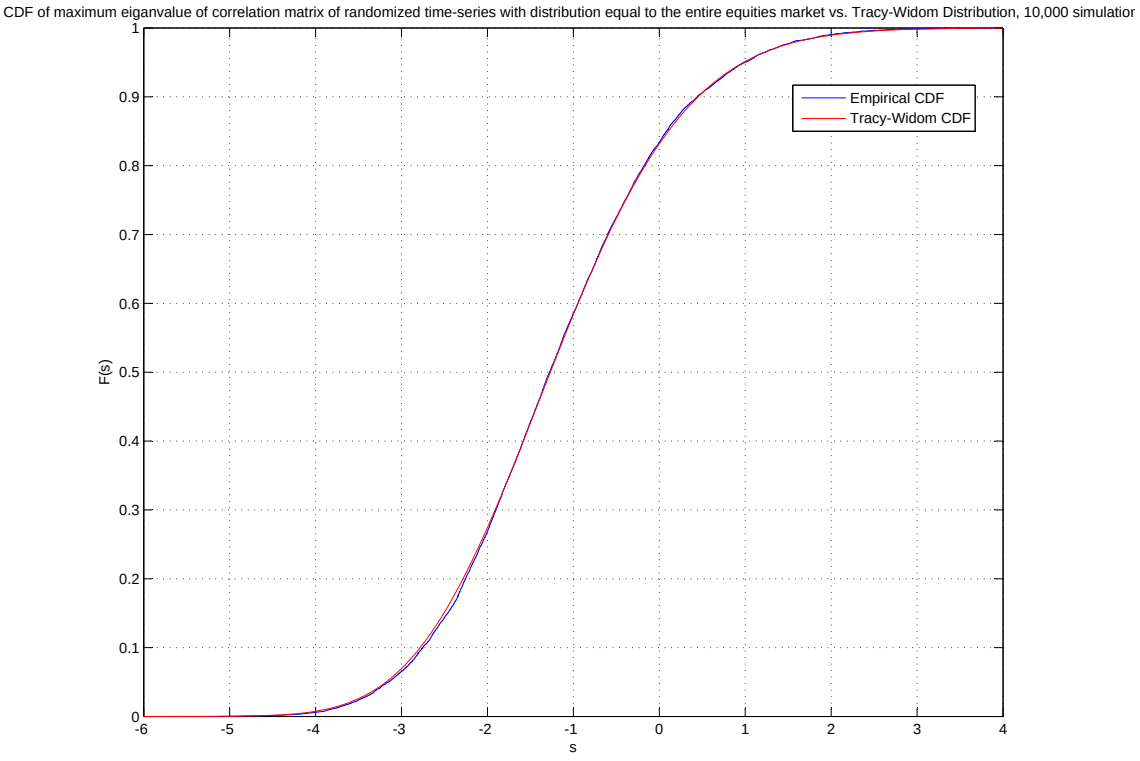


Figure 2.28: Cumulative density of maximum eigenvalue of randomly permuted time-series on original data of all equities versus Tracy-Widom. We cannot reject the null hypothesis that the two distributions are the same based on Kolmogorov-Smirnov one-series test at the 1 percent significance level. We use 500 observations and 3141 variables. Overall 10,000 simulations were performed and the corresponding eigenvalue in each case was computed.

Table 2.4: Significance of Eigenvalues in the entire Equities Market a la TW

Top 25 Eigenvalues	s-value	$F_1(s)$
$\lambda_1 = 328.25$	5076	1
$\lambda_5 = 23.58$	181	1
$\lambda_{10} = 16.22$	63	1
$\lambda_{15} = 14.22$	31	1
$\lambda_{20} = 12.42$	2.18	.9924
$\lambda_{21} = 12.18$	-1.64	0.38
$\lambda_{22} = 11.89$	-6.39	3.22e-07
$\lambda_{23} = 11.81$	-7.59	7.65e-11
$\lambda_{24} = 11.67$	-9.93	7.36e-22
$\lambda_{25} = 11.52$	-12.26	1.49e-38

- 3141 assets and 500 days used.
- Corresponding $\lambda_+ = 12.295$.
- 20 eigenvalues exceed it and account for 24 percent of variation.
- All twenty are deemed significant by Tracy-Widdom.

We now present results on the same inquiries as above for the options market as a whole as determined by available data in OptionMetrics, and for those assets with at least 500 days of data. This amounts to 31410 variables and 500 observations. The number of eigenvalues which exceed the MP threshold $\lambda_+ = 79.67$ is 108, and these eigenvalues account for 50 percent of the variation.

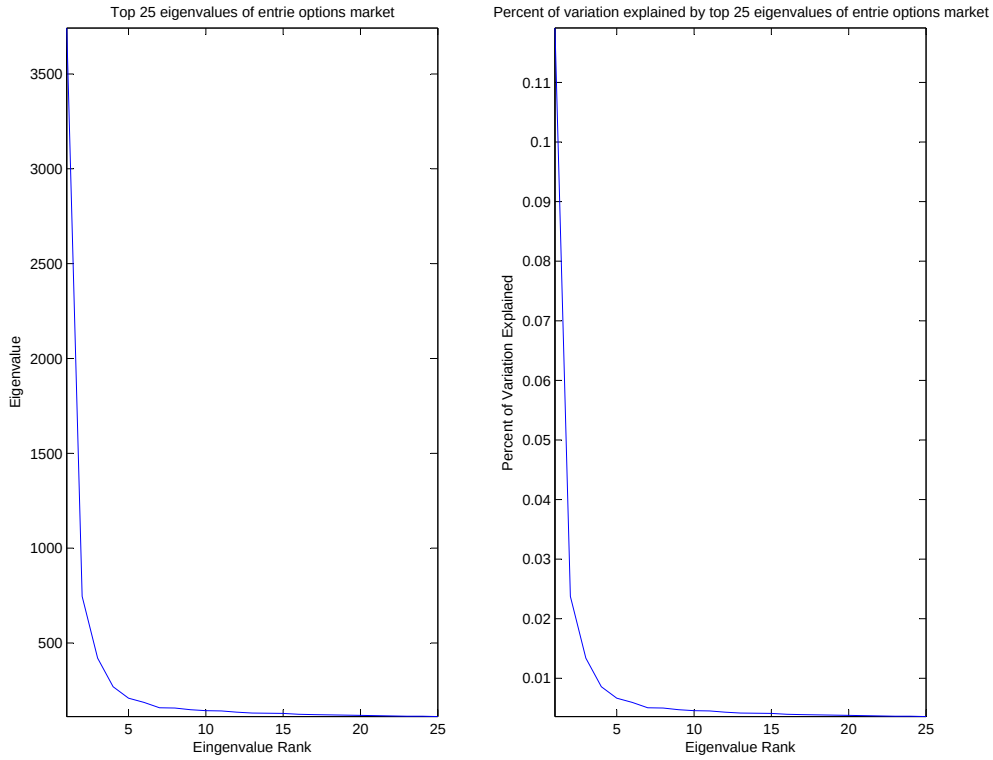


Figure 2.29: Top 25 eigenvalues of the options market. They explain 26 percent of the variation in the IVS. Data uses 500 sample days and 31410 variables.

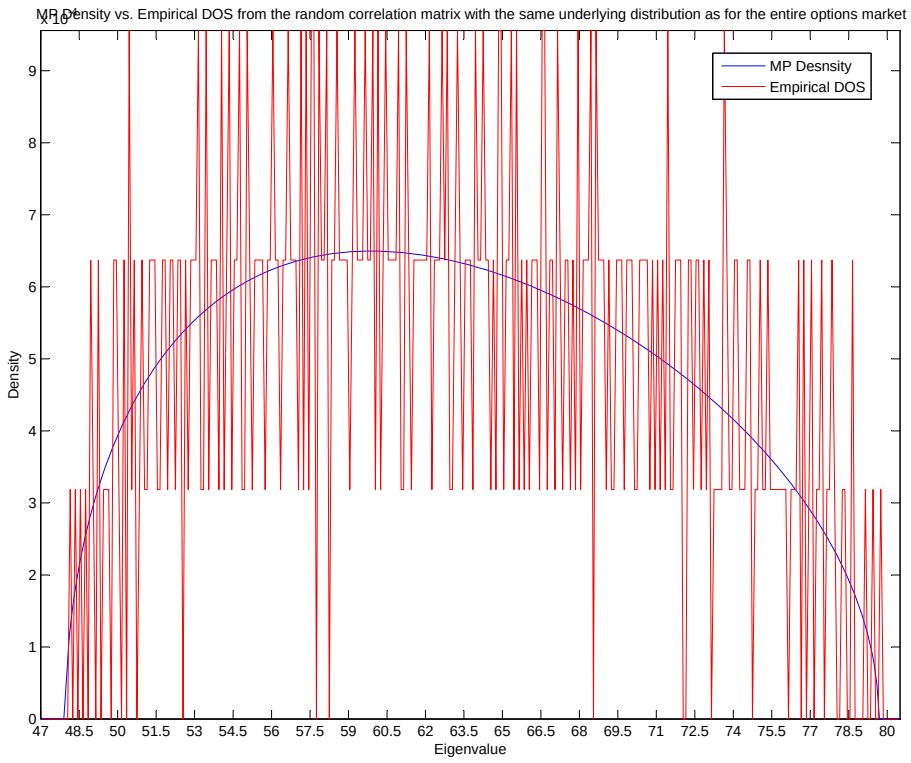


Figure 2.30: Empirical density of all eigenvalues of correlation matrix of randomly permuted time-series on entire options data versus Marchenko-Pastur distribution: $\lambda_+ = 79.67$ and $\lambda_- = 47.97$. For both distributions there is a point mass of weight 98.41 percent at zero. 31410 variable used each over 500 observations.

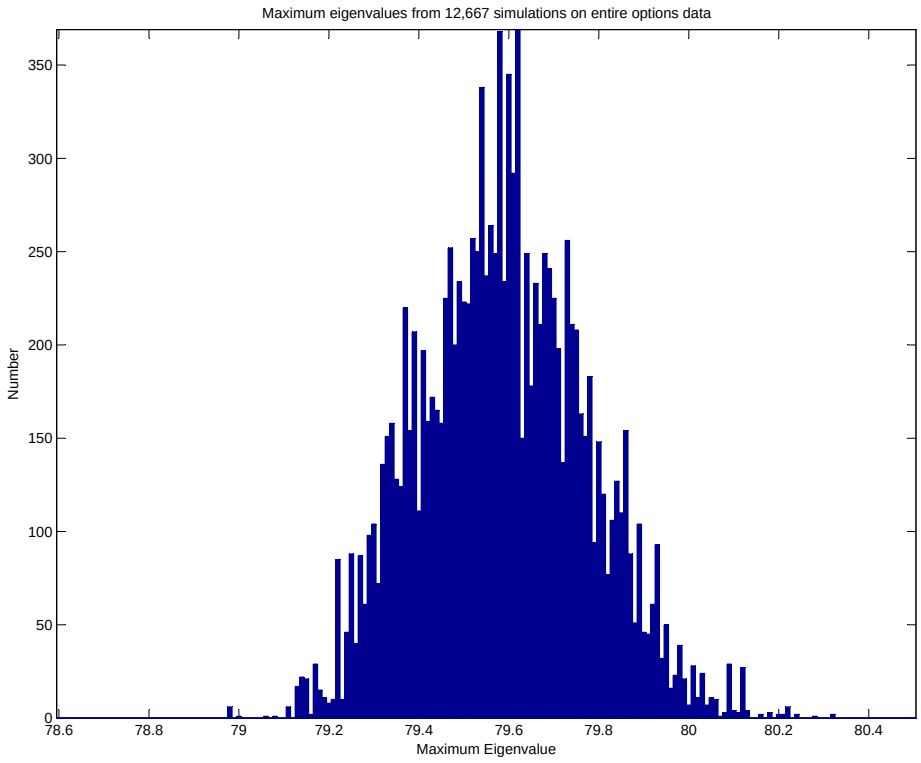


Figure 2.31: Largest eigenvalue of the correlation matrix of time-series permuted data on the entire options market. About 70 percent of all 12,667 simulations lie below $\lambda_+ = 79.67$, the theoretical value from Marchenko-Pastur theory. and all lie in the range $(78.98, 80.32)$. Uses 500 observations and 31410 variables.

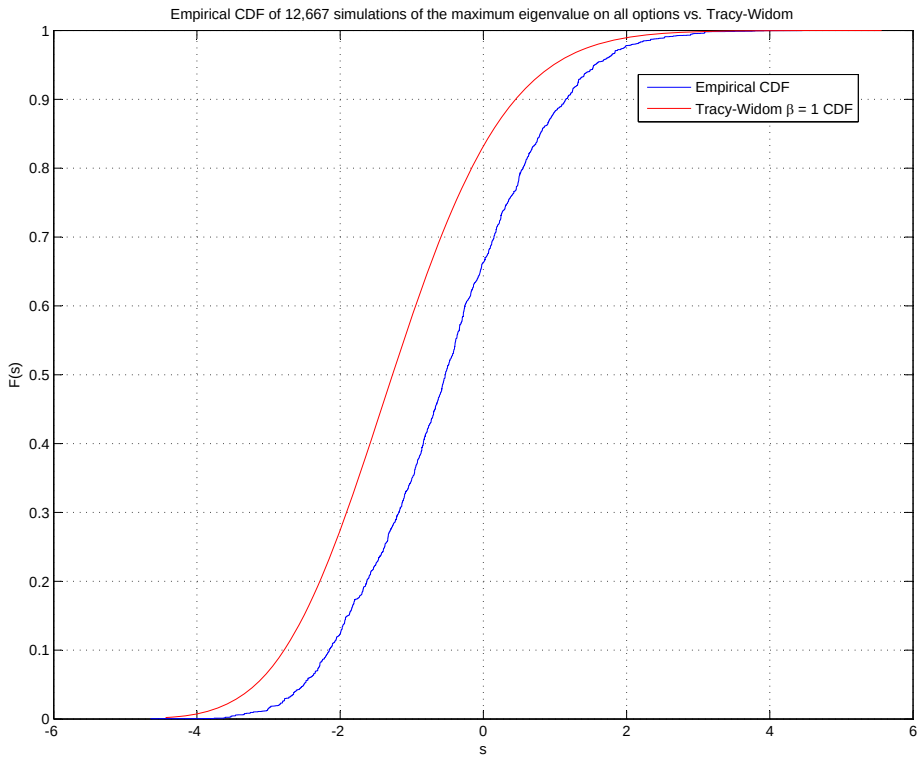


Figure 2.32: CDF of largest eigenvalue of the correlation matrix of time-series permuted data on the entire options market versus Tracy-Widom $\beta = 1$ CDF. We use 500 observations and 31410 variables. Overall 12,667 simulations were performed and the corresponding maximum eigenvalue in each case was computed.

Table 2.5: Significance of Eigenvalues in the entire Options Market a la TW

Top 110 Eigenvalues	s-value	$F_1(s)$
$\lambda_1 = 3742$	24843	1
$\lambda_5 = 209.27$	879.14	1
$\lambda_{10} = 143.5$	433.04	1
$\lambda_{20} = 118.19$	261.32	1
$\lambda_{40} = 102.62$	155.74	1
$\lambda_{50} = 97.40$	120.35	1
$\lambda_{70} = 90.48$	73.35	1
$\lambda_{90} = 84.56$	33.21	1
$\lambda_{107} = 80.21$	3.70	.9996
$\lambda_{108} = 80.04$	2.60	.996
$\lambda_{109} = 79.65$	-.10	.80
$\lambda_{110} = 79.41$	-1.71	.35

- 31410 assets and 500 days used.
- Corresponding $\lambda_+ = 79.672$.
- 108 eigenvalues exceed it and account for 50 percent of variation.
- All 108 are deemed significant by Tracy-Widdom.

2.5 Conclusion

In the second section of this chapter we performed dynamic PCA with a 252-day moving window and observed that the leading eigenvalue across different asset classes is stable across time. In particular, during times of high-volatility and general market turmoil, the leading eigenvalue increases and the number of eigenvalues exceeding the Marchenko-Pastur threshold decreases. Specifically, the magnitude of the leading eigenvalue and the number of eigenvalues exceeding the Marchenko-Pastur threshold are strongly negatively correlated. This phenomena supports our observations in the first chapter, and indicates higher systemic risk for all options

during such periods.

In the third section we formulate various pivot models which we can use to reduce the number of variables needed to model the IVS. Of the seven pivot models formulated, we test each model by comparing how well it preserves the original spectrum. We also test how well it preserves the distribution across various types of risk faced by options whose underlying stock is a constituent of the S&P500.

Our results from this section indicate that the 6-pivot model, the 9-pivot model, and the 12-pivot model are acceptable in terms of how well they replicate the original statistics. Naturally, we find that the 12-pivot model is best, followed by the 9-pivot model, and last the 6-pivot model. In terms of overall efficacy, we recommend and use the 9-pivot model.

For risk-management purposes, and in light of these results, any one of the three models mentioned in the previous paragraph is a very good candidate for modeling the IVS while offering considerable dimensionality reduction. We study the performance of these models based on a statistical point of view, and thus it is crucial to backtest against historic option portfolios before implementing any model in practice.

We conclude by performing a PCA of the entire equities and options market as made available by OptionMetrics. We begin with the markets as determined by constituents of S&P500 and their options. To better distinguish signal from noise in these markets in their entirety, in addition to PCA and Marchenko-Pastur, we also employ the Tracy-Widom Law in order to gauge the significant part of the spectrum.

To determine the applicability of both Marchenko-Pastur and Tracy-Widom on our data, we first test both Marchenko-Pastur and Tracy-Widom on random

matrixes with the same underlying distribution as our empirical data. The results from using such random matrixes are in excellent agreement with those predicted by both Marchenko-Pastur and by Tracy-Widom. In particular, the underlying distribution of our empirical data forms an ensemble class on which Tracy-Widom applies. Utilizing all of the aforementioned tools, we classify the number of significant factors driving the U.S. equities and options market as follows:

1. Equities in SPX: **15** significant factors (account for 55% of variance).

2. Options with underlying in SPX: **84** significant factors (account for 55% of variance).

3. All assets in OptionMetrics: **20** significant factors (account for 24% of variance).

4. All options with underlying asset in OptionMetrics: **108** significant factors (account for 50% of variance).

Appendix A

Principal Component Analysis of Implied Volatility Surfaces

For the sake of conciseness we do not include the code nor all of our results with this thesis. All code used to generate the results of Chapter 1 is ready upon request. All databases of results for all assets in OptionMetrics with available data is also ready upon request.

In Chapter 1 we made the assertion that the first eigenvalue λ_1 of a correlation matrix can be used as an approximation for the average correlation $< \rho >$, we give the proof below. Let C be the correlation matrix, and V the matrix of eigenvectors in the spectral decomposition of C. We can write:

$$\begin{aligned} \lambda_1 &= V^{(1)T} C V^1 = \sum_{i=1}^{N} (V_i^1)^2 + \sum_{i \neq j} V_i^1 V_j^1 \rho_{ij} = 1 + \sum_{i \neq j} V_i^1 V_j^1 \rho_{ij} \\ &= 1 + \sum_{i \neq j} V_i^1 V_j^1 \frac{\sum_{i \neq j} V_i^1 V_j^1 \rho_{ij}}{\sum_{i \neq j} V_i^1 V_j^1} \end{aligned}$$

Rearranging gives,

$$\frac{\lambda_1 - 1}{\sum_{i \neq j} V_i^1 V_j^1} = \frac{\sum_{i \neq j} V_i^1 V_j^1 \rho_{ij}}{\sum_{i \neq j} V_i^1 V_j^1}$$

$$\therefore \text{Since } V_i^1 \approx \frac{1}{\sqrt{N}} \text{ then } \sum_{i \neq j} V_i^1 V_j^1 \approx N \frac{N-1}{N} = N - 1$$

$$\text{hence, } \frac{\lambda_1 - 1}{N - 1} \approx \rho, \text{so that, } \rho \approx \frac{\lambda_1}{N}.$$

We now include the principal component surfaces for VXX, VXZ, AAPL, NFLX, BIIB, CVS, HOT, HD, BF, DAL, and GLD.

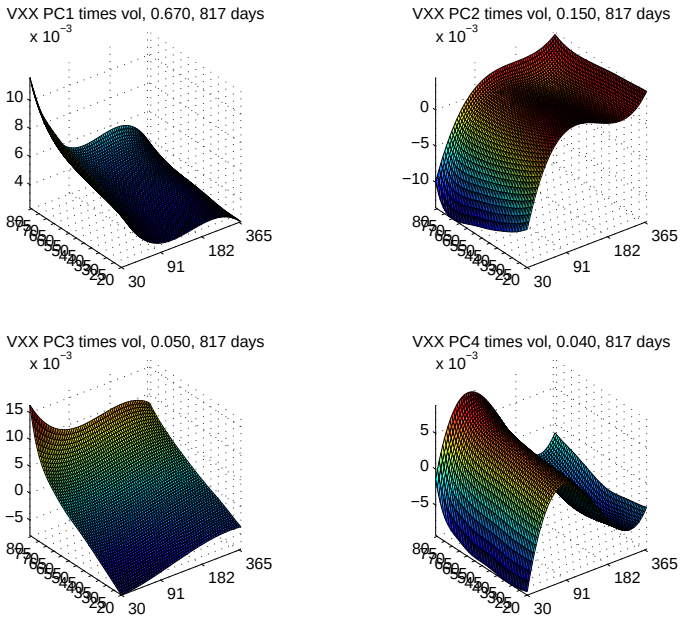


Figure A.1: First four PCs for VXX. Options on volatility products tend to carry more idiosyncratic risk as can be seen from the magnitude of the top four eigenvalues and principal components given above. Data series ends on August 31, 2013 and goes back as far as possible, but no earlier than August 31, 2004.

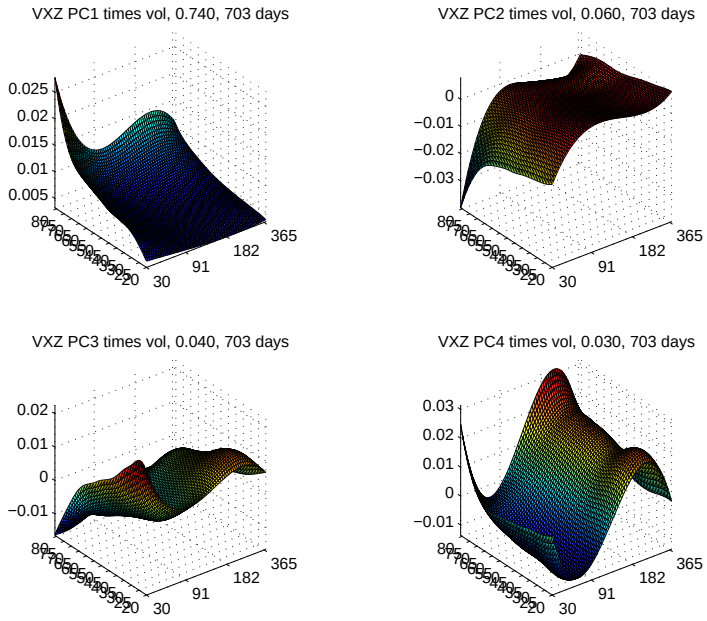


Figure A.2: First four PCs for VXZ. Options on volatility products tend to carry more idiosyncratic risk as can be seen from the magnitude of the top four eigenvalues and principal components given above. Data series ends on August 31, 2013 and goes back as far as possible, but no earlier than August 31, 2004.

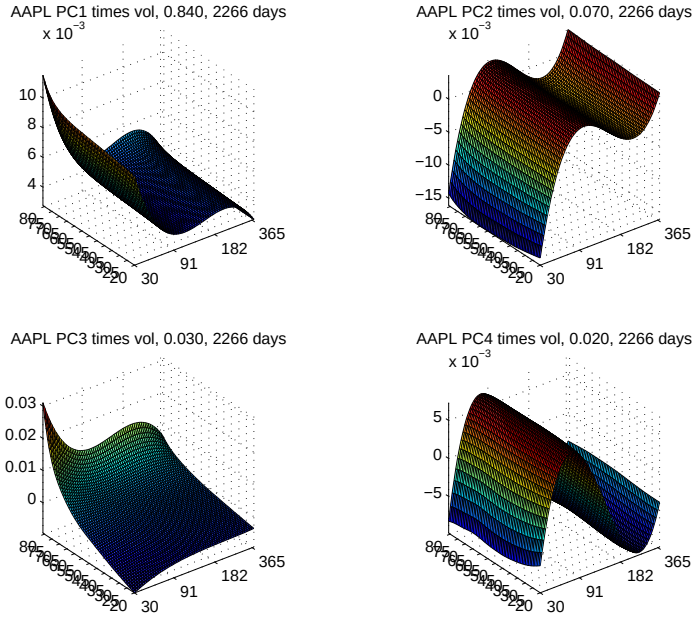


Figure A.3: First four PCs for AAPL. Options on large, popular, well-established stocks like aaple carry mostly systemic risk as can be seen from the magnitude of the eigenvalues as well as the shape of the principal components. Data series ends on August 31, 2013 and goes back as far as possible, but no earlier than August 31, 2004.

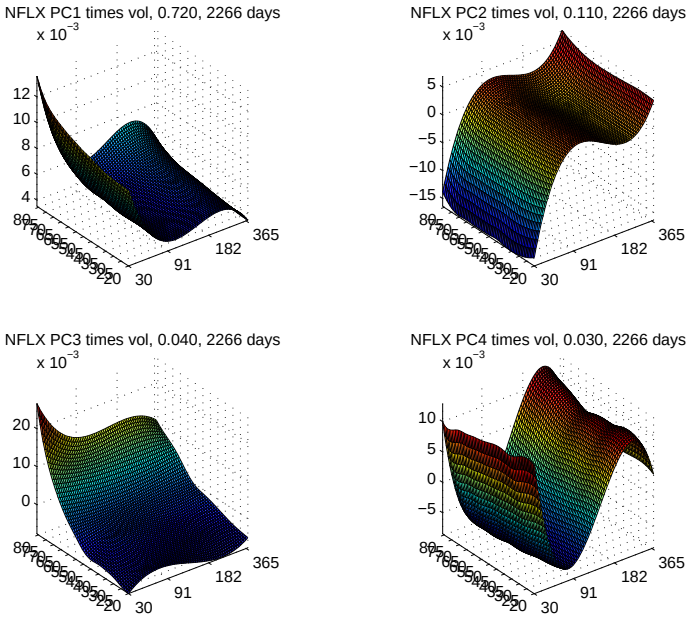


Figure A.4: First four PCs for NFLX. Data series ends on August 31, 2013 and goes back as far as possible, but no earlier than August 31, 2004.

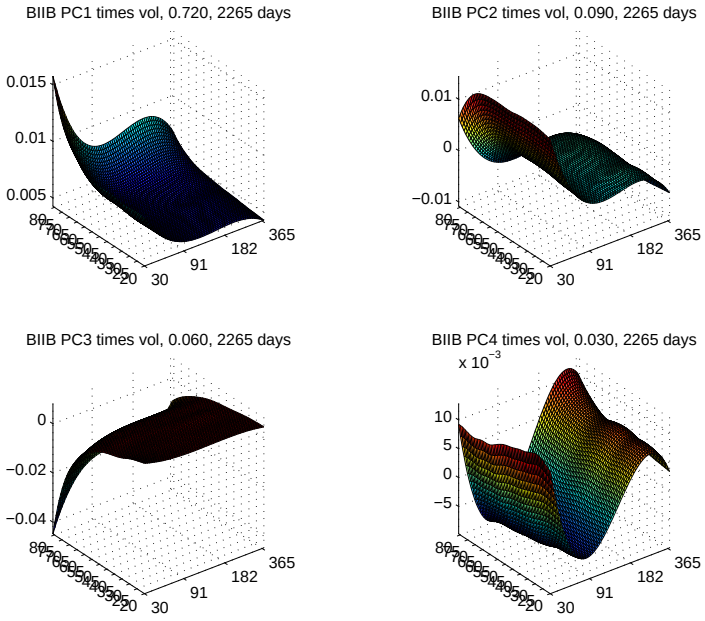


Figure A.5: First four PCs for BIIB. Data series ends on August 31, 2013 and goes back as far as possible, but no earlier than August 31, 2004.

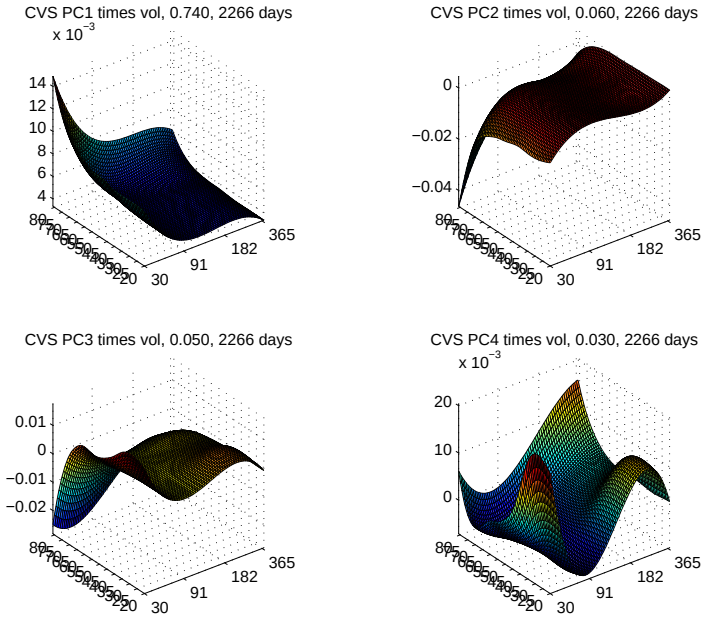


Figure A.6: First four PCs for CVS. Data series ends on August 31, 2013 and goes back as far as possible, but no earlier than August 31, 2004.

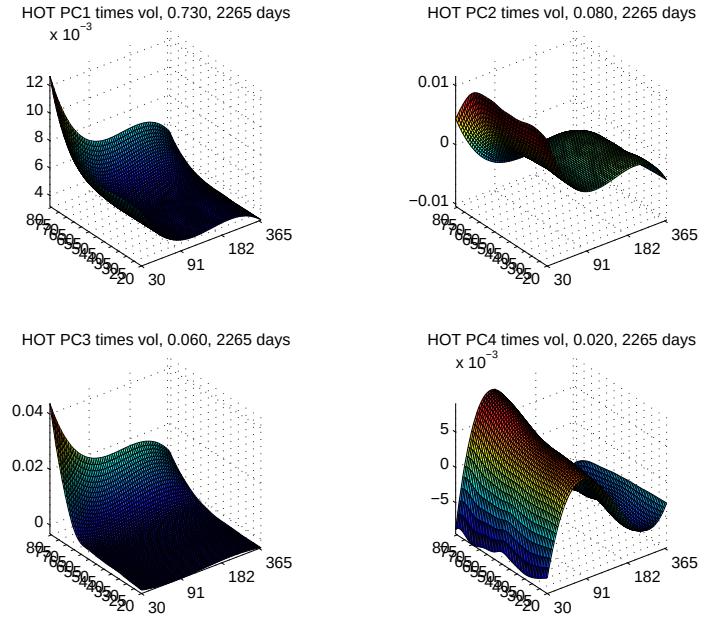


Figure A.7: First four PCs for HOT. Data series ends on August 31, 2013 and goes back as far as possible, but no earlier than August 31, 2004.

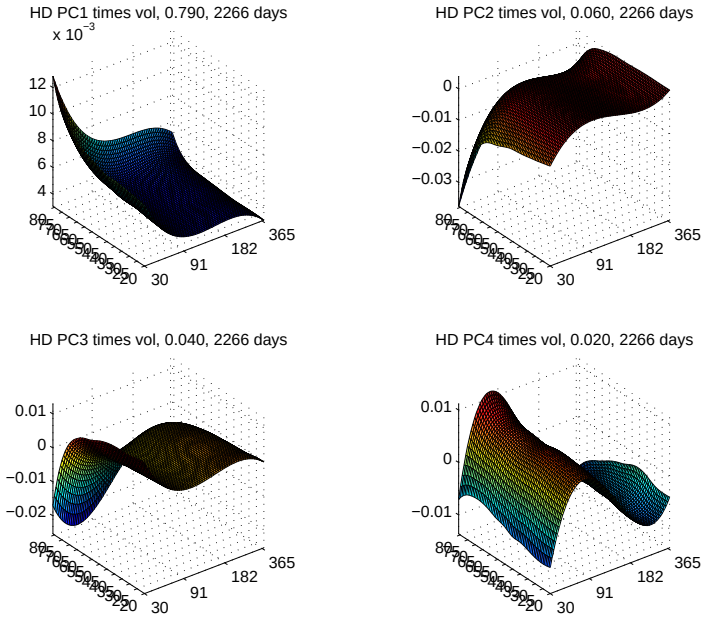


Figure A.8: First four PCs for HD. Data series ends on August 31, 2013 and goes back as far as possible, but no earlier than August 31, 2004.

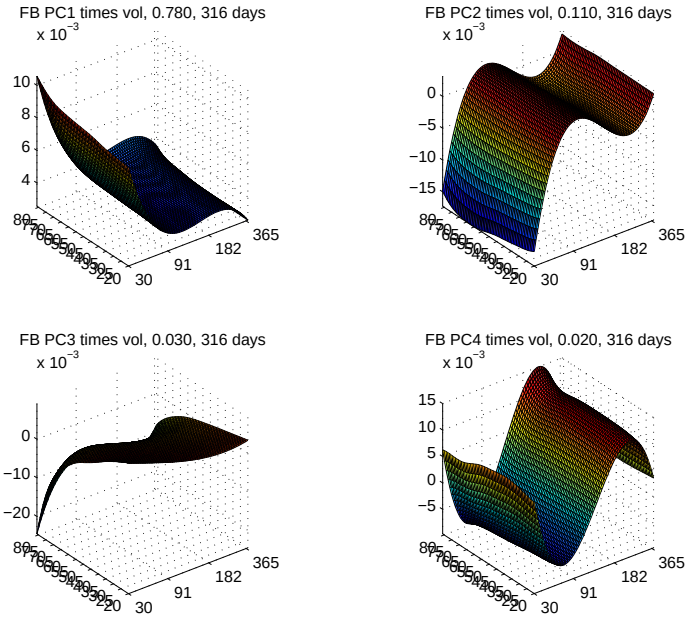


Figure A.9: First four PCs for FB. Data series ends on August 31, 2013 and goes back as far as possible, but no earlier than August 31, 2004.

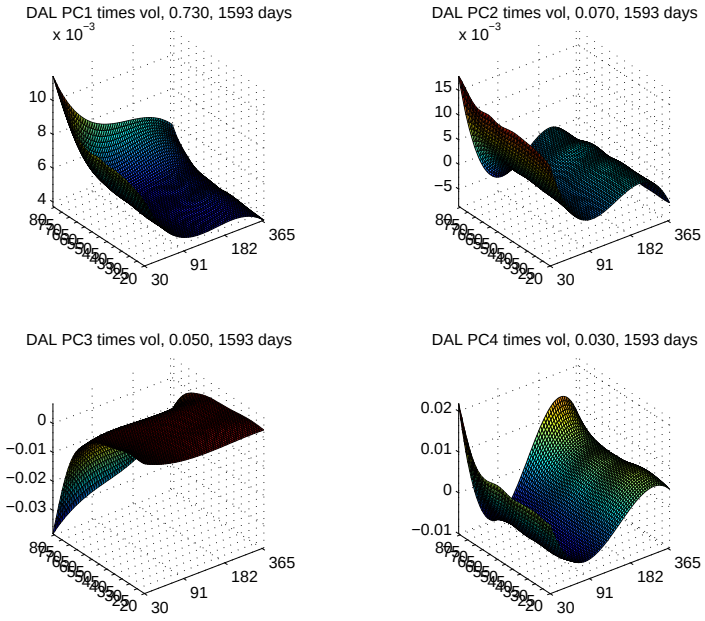


Figure A.10: First four PCs for DAL. Data series ends on August 31, 2013 and goes back as far as possible, but no earlier than August 31, 2004.

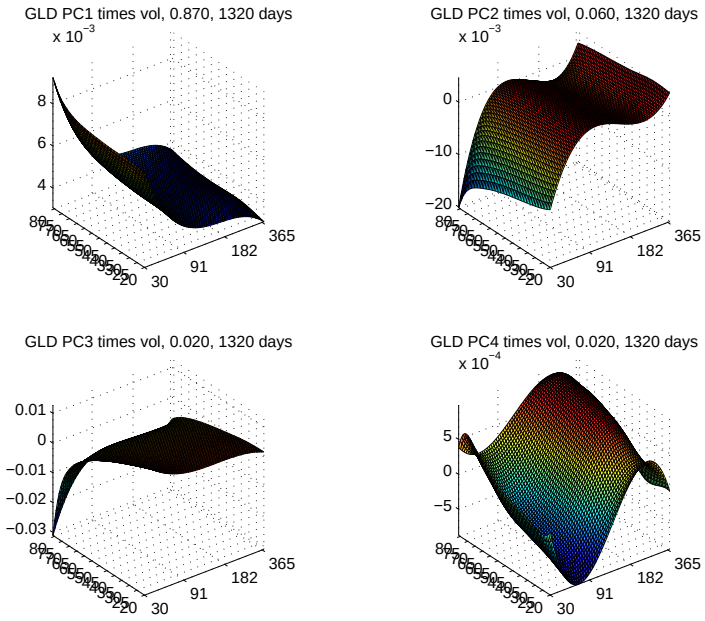


Figure A.11: First four PCs for GLD. Options on this gold ETF carry mostly systemic risk since gold is dependent on the overall market and state of the economy. Data series ends on August 31, 2013 and goes back as far as possible, but no earlier than August 31, 2004.

Appendix B

The Pivot Method and Determining Significant Components via Random Matrix Theory

In section 3 we excluded the distribution across different risk classes for the 4-pivot, 5-pivot, and 7-pivot model. We now include these results below. Note that the 4-pivot and 5-pivot model are indistinguishable, and under-perform the 7-pivot model results. Analogously to spectrum replication, the 7-pivot model performs worst than the 6-pivot model in this context as well.

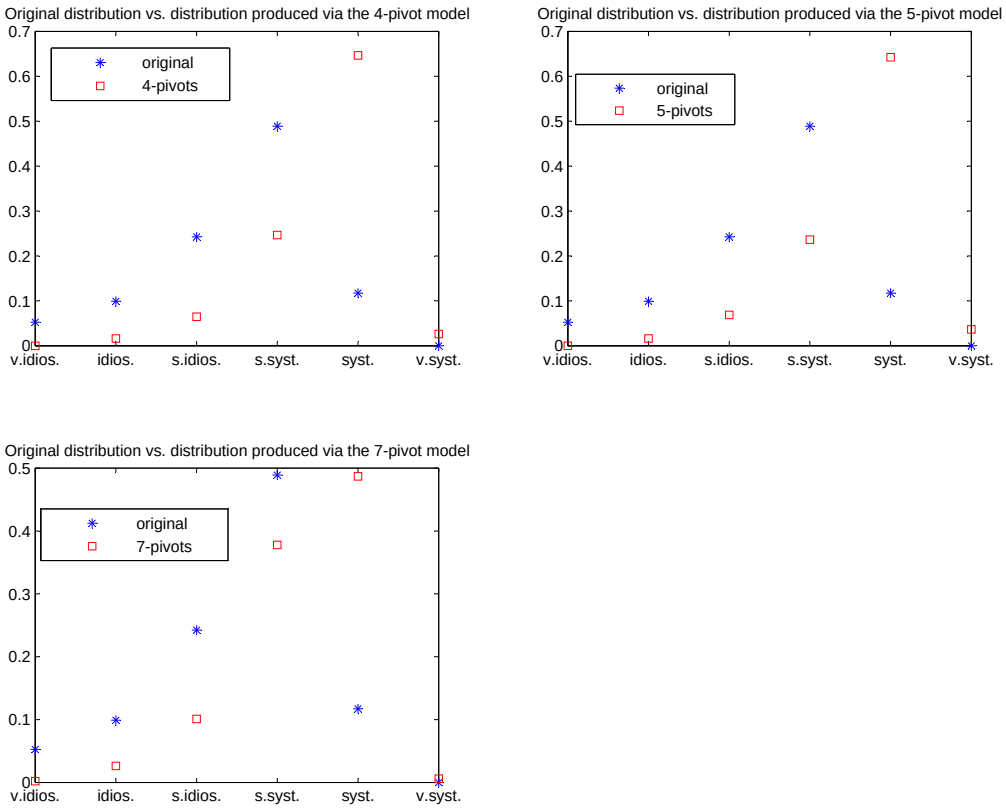


Figure B.1: Original distribution using all 52-pivots vs. the distribution produced by each model indicated for the constituents of S&P500.

Bibliography

[1] Avellaneda, M., Cont, R., (2002) *Special issue on volatility modeling.*, Quantitative Finance.

[2] Biroli, G., Bouchaud, J-P., Potters, M., (2009) *On the top eigenvalue of heavy-tailed random matrices.* http://arxiv.org/abs/cond-mat/0609070.

[3] Bouchaud, J-P., Potters, M., (2009) *Financial Application of Random matrix Theory: a short review.* http://arxiv.org/abs/0910.1205.

[4] Cont, R., Fonseca, J., (2002) *Dynamics of implied volatility surfaces.*, Journal of Investment Management.

[5] Dupire, Bruno (1994) *Pricing With a Smile.*,Risk.

[6] Tao, T., (2012) *Topics in Random Matrix Theory.*