Modeling Volatility Risk in Equity Options Market: A Statistical Approach

Doris Dobi∗ Marco Avellaneda†

June 17, 2014

Abstract

This paper provides a cross-sectional analysis of U.S. option markets based on implied volatility data from August 2004 to August 2013. We analyse the implied volatility surface (IVS) for each security in the OptionMetrics database. We use implied volatility data across 13 deltas and 4 expiration dates. Employing methods from principal component analysis (PCA), and results from random matrix theory (RMT), we identify the significant eigenvalues of the correlation matrix of implied volatilities and conclude that, usually, three principal components suffice to reproduce the IVS. In this way we reduce dimensionality of the options market without loosing meaningful information. From this analysis we classify equities into those carrying mostly “systemic” risk and into those carrying mostly “idiosyncratic” risk.

Based on the PCA results, we formulate a model which can be used to describe the dynamics of the joint statistics of the IVS of all U.S. options, yet is compact and computationally feasible. Using 9 volatility points to represents each IVS, the model offers significant dimension reduction for each asset as well as for all assets in aggregate.

∗Courant Institute of Mathematical Sciences, 251 Mercer Street, New York, N.Y. 10012.
e-mail: dobi@cims.nyu.edu
e-mail: avellaneda@courant.nyu.edu
We conclude with a PCA study of the correlation matrix of the entire cross-section equities and options market\(^1\). We find that the number of significant factors driving the U.S. equities and options market are as follow:

1. Equities in SPX: 15 significant factors (account for 55% of variance).
2. Equities and options with underlying in SPX: 84 significant factors (account for 55% of variance).
3. Equities in OptionMetrics: 20 significant factors (account for 24% of variance).
4. All equities and options with underlying asset in OptionMetrics: 108 significant factors (account for 50% of variance).

\(^1\) As made available by the OptionMetrics database.
## Contents

1. Introduction ................................................. 4

2. Principal Component Analysis of the Correlation Matrix of the Implied Volatility Surfaces ........................... 6

3. Implied Volatility Surfaces and Random Matrix Theory (RMT) .................................................. 11

4. Classification of Optionable Stocks into “Systemic” and “Idiosyncratic” ........................................ 14

5. First Dimensional Reduction: the Pivot Method .............................................................................. 21


7. Conclusion ....................................................... 42

A. Appendix ......................................................... 44

References .................................................................. 48
1 Introduction

There have been many studies which seek to classify the “number of factors” needed to econometrically model the market [3][4][6][10]. Principal component analysis of correlation matrices shows that around 15 components can be used to capture the vast majority of systemic movement in the equity market. These 15 factors account for 55 percent of the variation in returns. This represents dramatic dimensional reduction, and is the analogue for equities of Litterman and Scheinkman [10] who find that 82 percent of the variation in returns of all Treasury bond correlations can be explained in terms of only three factors (or principal components).

On the other hand, not as much work has been done on classifying factors underlying option returns. Avellaneda and Cont [1], Carmona and Nadtochiy [5], Cont and Fonseca [6], Dupire [7], and Schweizer and Wissel [12] are among some of the researchers who delve into this topic. Cont and Fonseca [6] find that the variance in returns of the correlation matrix of implied volatility of DAX options can be explained by roughly three principal components.

In this paper, we show that the implied volatility surface (IVS) of most equity option contracts can be described using four principal components, and usually three components suffice. We suggest that the size of the leading eigenvalue can be used to characterize the type of risk intrinsic to options on a specific underlying asset. Furthermore, we propose a parsimonious framework which can be used to reproduce the IVS of any option.

In order to better understand the volatility of option positions, we classify options into those that carry mostly systemic risk and into those that carry mostly idiosyncratic risk. Systemic risk gives an indication of exposure to or vulnerability to aggregate market risk. Idiosyncratic risk, on the other hand, measures underlying-asset-specific risk, or the risk tied with the specific nature of the underlying and uncorrelated to overall-market
risk. Idiosyncratic risk can be thought of as corresponding to corporate events, such as takeovers, company-specific news releases, earnings, and so forth.

We propose to analyze the volatility risk of options by studying the fluctuations of the IVS. 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. Original data on the options market is 130 times bigger than that of the equities market\(^2\).

We characterize the significant part of the spectrum of IVS fluctuations using principal component analysis (PCA) and results from random matrix theory (RMT). We standardize the data by looking as the correlation matrix of implied volatility returns. Studying the distribution of the resulting eigenvalues allows us to classify an asset according to the type of risk it carries. We find that the leading eigenvalue of the correlation matrix plays a critical role in our classification scheme.

We propose various dimensionality-reducing models for replicating the IVS. We refer to these models as pivot models to indicate the use of a few implied volatility returns (the pivots) in generating the entire IVS via linear interpolation. We keep in mind certain constraints any such model must adhere to; the most crucial being efficiency in the number of pivots used. In particular, our pivot model must reduce dimensionality, and abide by limitations in computing power and resources.

To measure the quality of each model, we test how well each pivot scheme preserves the spectrum of the original correlation matrix. In addition, we also compare how well the replicated data preserves original risk classification. Both of these tests are used to determine how well the interpolated IVS replicates the original IVS.

Our work shows that a 9 pivot model yields excellent results while offering very significant dimensionality reduction of over 14 times the original data size. The 9 pivot model provides a firm basis for a risk-management system for portfolios of options. Based on these results, we perform a PCA of the correlation matrix of the entire equities and

\(^2\)As made available by the OptionMetrics database. In our statistical analysis we use 52 of these 130 points to generate the original IVS; see below.
equity options market. The dimensionality reduction gained from the 9 pivot model makes large PCA computations feasible on the market as a whole.

We conclude by determining the number of significant components in the entire options and equity markets. In order to accomplish this, we employ the Marchenko-Pastur distribution and the Tracy-Widom law from RMT [3][4][9][13]. Results are given in the last section.

2 Principal Component Analysis of the Correlation Matrix of the Implied Volatility Surfaces

Consider the implied volatility surface for SPX given in Figure 1, clearly the implied volatility surface is not constant in neither strike (K) nor maturity (T). Furthermore, the implied volatility surface (IVS) fluctuates; it is precisely this feature of the IVS that is crucial in managing the risk of a portfolio of options [see Figure 2].

We use empirical results as a stepping stone for better understanding the behaviour of implied volatility. We begin by extracting the first four principal components of the correlation matrix of the implied volatility surface for SPX where we couple the implied volatility with the underlying asset.

There are 130 available\(^3\) implied volatilities across each of the 10 maturities \(\in\{30, 60, 91, 122, 152, 182, 273, 365, 547, 730\}\) and each of the 13 deltas \(\in\{20, 25, \cdots , 75, 80\}\). For our analysis, we use a subset of 52 of these 130 points: the implied volatility return across each of the 4 maturities \(\in\{30, 60, 91, 182\}\) and across all 13 available deltas \(\in\{20, 25, \cdots , 75, 80\}\). This choice is somewhat arbitrary and is made to reduce computation while preserving available information. Figure 3 depicts this selection. The data matrix on which the correlation matrix is computed is constructed as follows: our first variable (i.e., the first column) is the underlying stock returns and each subsequent

\(^3\)Via OptionMetrics.
Figure 1: Implied Volatility Surface for SPX using call options, December 12 2008.

\[ \text{Moneyness } M = \frac{K}{F} \]

Time to Maturity

Implied Volatility \( \sigma(T, M) \)
Figure 2: We can see that the IVS moves in time, so the dynamic nature of the surface is important.
variable (i.e., each subsequent column) is one of the 52 implied volatility returns\(^4\).

\[
\begin{array}{cccccccccccc}
T(\text{days})
\end{array}
\]

\[
\begin{array}{cccccccccccc}
730 & 547 & 365 & 273 & 182 & 152 & 122 & 91 & 60 & 30
\end{array}
\]

Figure 3: Depiction of the 52 implied volatility points selected (in bold) out of the original 130 implied volatilities available via OptionMetrics.

From the principal components we seek to determine how the implied volatility surface changes in time. For example, we see that the first principal component usually explains about 90 percent of the variation in the IVS, and is usually “flat”: this 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.

The second and third components are either flat and very similar to the first component, or they exhibit time-skew and/or delta-skew, in which case the remaining change in the shape of the surface is attributed to skew. The fourth component exhibits a convexity-effect and usually has very little explanatory power for how the surface varies, but is nonetheless included for completeness. See the appendix for some additional examples of top principal component surfaces for various major underlying assets.

\(^4\)As a reminder, the implied volatility returns are computed for a specific call option contract for each
Figure 4: 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. Time period dates from August 31, 2008 until August 31, 2013.
3 Implied Volatility Surfaces and Random Matrix Theory (RMT)

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:

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

Consider the density of states (DOS):

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

We will make use of

Proposition 1 (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

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

and

$$d\nu(x) := \frac{1}{2\pi\sigma^2} \sqrt{\frac{(\lambda_+ - x)(x - \lambda_-)}{x\Lambda}} \frac{1}{1}\left|\lambda_- \lambda_+\right| dx.$$ 

Where $\lambda_\pm = 4(1 \pm \sqrt{\Lambda})^2$.

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

We observe that the vast majority of the variation in the implied volatility surface is 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.

given day in our sample.
All eigenvalues greater than the MP-threshold $\lambda_+$ will be referred to as *significant*. These eigenvalues give us information about the true correlation matrix of the market\(^5\) and hence are useful in separating signal from noise. In our study, we find that across all IVSs\(^6\) studied usually three or four of the eigenvalues and their corresponding eigenvectors lie above of their corresponding Marcenko-Pastur upper bound.

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

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

4. Using the Kolmogorov-Smirnov test, 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 98 percent of the option contracts in OptionMetrics we cannot reject the null hypothesis at the 1 percent significance level. In addition, we find that for 100 percent of the option contracts with an underlying in SPX we cannot reject the null hypothesis at the 1 percent significance

---


\(^6\)As made available by the OptionsMetrics database.
level. Thus, we can apply Marchenko-Pastur to matrices with the same underlying distribution as our empirical data. Below we include the number of eigenvalues exceeding the MP-threshold for various major underlying assets.

Figure 5: Results indicate that around three or four of the top eigenvalues are usually significant as determined by exceeding the MP-threshold $\lambda_{+}$. By projecting along these corresponding eigenvectors we may distinguish signal from noise in the correlation matrix.
4 Classification of Optionable Stocks into “Systemic” and “Idiosyncratic”

In Figure 6 we display the percent of variation explained by the first principal component for 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 in Figure 7 for these same ETFs. 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 between 70 and 90 percent of variation.

![Percent of variation explained by the first principal component for the top 20 most liquid ETFs by market volume](image)

Figure 6: 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 order of increasing systemic risk.

In addition, we determine the number of eigenvalues (before normalization) which exceed the Marcenko-Pastur (MP) threshold in Figure 8. We display the results below.
Figure 7: 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.
From this analysis, we find that the majority of the ETFs have around three or four eigenvalues which exceed the MP threshold; again confirming that the topmost three eigenvalues can be used to separate signal from noise in the correlation matrix.

![Number of eigenvalues exceeding the Marchenko-Pastur upper bound for the topmost 20 liquid ETFs by market volume](image)

Figure 8: Number of eigenvalues exceeding the MP upper limit of the spectrum for the 20 most liquid ETFs.

We propose using the magnitude of the leading eigenvalue as an indicator of the equity’s exposure to 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, the magnitude of the second and third eigenvalues indicate idiosyncratic risk. These eigenvalues capture non-uniform movements in the IVS, i.e., the “volatility” of time-skew and delta-skew.

It is interesting to see whether the eigenvalues are correlated to other properties of
the stocks such as average daily volume traded \(^7\), 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 for the constituents of S&P500 in Table 1.

### Table 1: Correlation Matrix of the constituents of the S&P500

<table>
<thead>
<tr>
<th>EV1</th>
<th>EV2</th>
<th>EV3</th>
<th>EV4</th>
<th># &gt; MP</th>
<th>Volatility</th>
<th>Corr. to Mkt.</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>-0.77</td>
<td>-0.85</td>
<td>-0.94</td>
<td>-0.89</td>
<td>0.26</td>
<td>0.15</td>
<td>0.33</td>
</tr>
<tr>
<td>-0.77</td>
<td>1.00</td>
<td>0.72</td>
<td>0.74</td>
<td>0.62</td>
<td>-0.15</td>
<td>-0.25</td>
<td>-0.17</td>
</tr>
<tr>
<td>-0.85</td>
<td>0.72</td>
<td>1.00</td>
<td>0.81</td>
<td>0.72</td>
<td>-0.29</td>
<td>-0.17</td>
<td>-0.25</td>
</tr>
<tr>
<td>-0.94</td>
<td>0.74</td>
<td>0.81</td>
<td>1.00</td>
<td>0.88</td>
<td>-0.22</td>
<td>-0.16</td>
<td>-0.24</td>
</tr>
<tr>
<td>-0.89</td>
<td>0.62</td>
<td>0.72</td>
<td>0.88</td>
<td>1.00</td>
<td>-0.23</td>
<td>-0.00</td>
<td>-0.26</td>
</tr>
<tr>
<td>0.26</td>
<td>-0.15</td>
<td>-0.29</td>
<td>-0.22</td>
<td>-0.23</td>
<td>1.00</td>
<td>-0.00</td>
<td>0.04</td>
</tr>
<tr>
<td>0.15</td>
<td>-0.25</td>
<td>-0.17</td>
<td>-0.16</td>
<td>-0.00</td>
<td>-0.00</td>
<td>1.00</td>
<td>-0.03</td>
</tr>
<tr>
<td>0.33</td>
<td>-0.17</td>
<td>-0.25</td>
<td>-0.24</td>
<td>-0.26</td>
<td>0.04</td>
<td>-0.03</td>
<td>1.00</td>
</tr>
</tbody>
</table>

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\(^8\); 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.

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 underlying stock carries, and the more it behaves like the overall market.

The opposite conclusion holds for the remaining eigenvalues, especially the second, which gives a stronger indication of idiosyncratic risk (skew fluctuations). Furthermore,

---

\(^7\) All data is computed from August 2004 to August 2013 on a daily basis

\(^8\) We normalize each eigenvalue by dividing by the sum of all eigenvalues, which are all positive since the correlation matrix is positive definite.
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 economic or “macro” 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 op-
tions 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 2: Top 15 underlying constituents and bottom 15 underlying constituents by first 
eigenvalue.

<table>
<thead>
<tr>
<th>Bottom 15 Constituents</th>
<th>EV1</th>
<th>Top 15 Constituents</th>
<th>EV1</th>
</tr>
</thead>
<tbody>
<tr>
<td>KMI</td>
<td>0.27272</td>
<td>GS</td>
<td>0.87176</td>
</tr>
<tr>
<td>POM</td>
<td>0.31406</td>
<td>JPM</td>
<td>0.87092</td>
</tr>
<tr>
<td>WEC</td>
<td>0.34958</td>
<td>BAC</td>
<td>0.85115</td>
</tr>
<tr>
<td>PNW</td>
<td>0.35611</td>
<td>SLB</td>
<td>0.84669</td>
</tr>
<tr>
<td>HCBK</td>
<td>0.35995</td>
<td>CAT</td>
<td>0.84219</td>
</tr>
<tr>
<td>NLSN</td>
<td>0.36338</td>
<td>AAPL</td>
<td>0.84126</td>
</tr>
<tr>
<td>TE</td>
<td>0.36342</td>
<td>XOM</td>
<td>0.84109</td>
</tr>
<tr>
<td>NU</td>
<td>0.3667</td>
<td>NOV</td>
<td>0.83737</td>
</tr>
<tr>
<td>BMS</td>
<td>0.37675</td>
<td>CME</td>
<td>0.83639</td>
</tr>
<tr>
<td>XEL</td>
<td>0.3828</td>
<td>MA</td>
<td>0.8358</td>
</tr>
<tr>
<td>WIN</td>
<td>0.38664</td>
<td>MS</td>
<td>0.83471</td>
</tr>
<tr>
<td>RSG</td>
<td>0.39779</td>
<td>APA</td>
<td>0.83081</td>
</tr>
<tr>
<td>FTR</td>
<td>0.40043</td>
<td>GOOG</td>
<td>0.83011</td>
</tr>
<tr>
<td>MKC</td>
<td>0.40475</td>
<td>HIG</td>
<td>0.8301</td>
</tr>
<tr>
<td>XYL</td>
<td>0.40596</td>
<td>HES</td>
<td>0.82965</td>
</tr>
</tbody>
</table>

The leading eigenvalue explains 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
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 is negatively correlated to the principal eigenvalue. The bottom 15 underlying stocks carry a lot of idiosyncratic risk which is intimately connected to the specific nature of their business. The top 15 underlying assets are much more dependent on the state of the overall market and carry less idiosyncratic risk and 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. In Tables 3 and 4 we display the relevant information for all available remaining assets.

### Table 3: Correlation matrix of all remaining remaining assets (around 3300)

<table>
<thead>
<tr>
<th>EV1</th>
<th>EV2</th>
<th>EV3</th>
<th>EV4</th>
<th># EVs &gt; MP</th>
<th>Vol</th>
<th>Corr</th>
<th>Daily Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>-0.64</td>
<td>-0.83</td>
<td>-0.89</td>
<td>-0.83</td>
<td>0.01</td>
<td>-0.03</td>
<td>0.07</td>
</tr>
<tr>
<td>-0.64</td>
<td>1.00</td>
<td>0.59</td>
<td>0.47</td>
<td>0.29</td>
<td>0.03</td>
<td>-0.15</td>
<td>-0.06</td>
</tr>
<tr>
<td>-0.83</td>
<td>0.59</td>
<td>1.00</td>
<td>0.71</td>
<td>0.64</td>
<td>0.05</td>
<td>0.02</td>
<td>-0.06</td>
</tr>
<tr>
<td>-0.89</td>
<td>0.47</td>
<td>0.71</td>
<td>1.00</td>
<td>0.78</td>
<td>-0.018</td>
<td>-0.01</td>
<td>-0.06</td>
</tr>
<tr>
<td>-0.83</td>
<td>0.29</td>
<td>0.64</td>
<td>0.78</td>
<td>1.00</td>
<td>0.01</td>
<td>0.20</td>
<td>-0.05</td>
</tr>
<tr>
<td>0.01</td>
<td>0.03</td>
<td>0.05</td>
<td>-0.01</td>
<td>0.01</td>
<td>1.00</td>
<td>-0.12</td>
<td>-0.02</td>
</tr>
<tr>
<td>-0.03</td>
<td>-0.15</td>
<td>0.02</td>
<td>-0.01</td>
<td>0.20</td>
<td>-0.12</td>
<td>1.00</td>
<td>0.07</td>
</tr>
<tr>
<td>0.07</td>
<td>-0.06</td>
<td>-0.06</td>
<td>-0.06</td>
<td>-0.05</td>
<td>-0.02</td>
<td>0.07</td>
<td>1.00</td>
</tr>
</tbody>
</table>

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).

Table 4: Average values across all remaining remaining assets (around 3300) and those of the bottom and top 1 percent

<table>
<thead>
<tr>
<th></th>
<th>EV1</th>
<th>EV2</th>
<th>EV3</th>
<th>EV4</th>
<th># EVs &gt; MP</th>
<th>Vol.</th>
<th>Corr.</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top 1%</td>
<td>0.921</td>
<td>0.031</td>
<td>0.017</td>
<td>0.012</td>
<td>1.617</td>
<td>0.030</td>
<td>0.208</td>
<td>168M</td>
</tr>
<tr>
<td>Bottom 1%</td>
<td>0.322</td>
<td>0.166</td>
<td>0.114</td>
<td>0.074</td>
<td>7.30</td>
<td>0.023</td>
<td>0.159</td>
<td>.65M</td>
</tr>
<tr>
<td>Mrkt Avg.</td>
<td>0.590</td>
<td>0.115</td>
<td>0.077</td>
<td>0.042</td>
<td>5.08</td>
<td>0.031</td>
<td>0.278</td>
<td>4.5M</td>
</tr>
</tbody>
</table>

For each option contract in OptionMetrics with available data (about 3800 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 this picture 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.

5 First Dimensional Reduction: the Pivot Method

Options are forward-looking in time, so the idiosyncratic or systemic components of the underlying stock affect the option market via a forward-looking expectation of future news about the underlying asset. This is where the distinction between idiosyncratic or systemic component plays a significant role. If we are faced with options on a market-index like SPY or QQQ, or a similar broad-market ETF, 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 less popular 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.

So far, we have constructed our implied volatility surfaces and determined the principal components by using the 52 implied volatility points determined by expiration $\in \{30, 91, 182, 365\}$ and $\delta \in \{20, 25, ..., 75, 80\}$. In the next few sections, we examine to what extent we can further reduce the number of implied volatility points 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.

Pivots are specific implied volatility returns which we use to generate all other implied volatilities via time and delta projections and linear interpolation as demonstrated in Figure 9. For example, the implied volatility return corresponding to a 25-delta and 182-day expiration option could constitute a pivot. In chapter one we used the 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 52-pivot IVS.

From our analysis we find that in times of crises and very high volatility, such as October 2008 (Lehman Brothers bankruptcy), May 2010 (2010 Flash Crash), and November 2011 (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 would 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, 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.

We analyse 2 pivot, 4 pivot, 5 pivot, 6 pivot, 7 pivot, 9 pivot and 12 pivot models in terms of how well they preserve the spectrum of the correlation matrix of the original data. The pivots used for each model are summarized in Table 5.
Table 5: Implied Volatilities Used as Pivots for Each Model

<table>
<thead>
<tr>
<th>2 Pivots</th>
<th>4 Pivots</th>
<th>5 Pivots</th>
<th>6 Pivots</th>
<th>7 Pivot</th>
<th>9 Pivots</th>
<th>12 Pivots</th>
</tr>
</thead>
<tbody>
<tr>
<td>25 δ 30</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>50 δ 30</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>75 δ 30</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>25 δ 91</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>50 δ 91</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>75 δ 91</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>25 δ 182</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>50 δ 182</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>75 δ 182</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>25 δ 365</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>50 δ 365</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>75 δ 365</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
</tbody>
</table>

The right hand side indicates the implied volatility used, e.g., 25 δ 365 refers to an option expiring in 365 days with strike at 25 δ. 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 \( \alpha \) be the projection along the time-to-maturity axis and \( \beta \) the projection along the \( \delta \) axis, and we seek the value of the implied volatility \( \text{iv}_{\text{ret}} \) at time \( t \) for some \( \delta \) 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:

\[
\text{iv}_{\text{ret}}(t, \delta, k) = \beta (\alpha \text{iv}_{\text{ret}}(t, \delta_+, k_+) + (1 - \alpha)\text{iv}_{\text{ret}}(t, \delta_+, k_-)) + (1 - \beta)(\alpha \text{iv}_{\text{ret}}(t, \delta_-, k_+) + (1 - \alpha)\text{iv}_{\text{ret}}(t, \delta_-, k_-))
\]

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 $(\delta_1, k_1)$ we have $\delta_1 \geq \delta_+ \text{ or } \delta_1 \leq \delta_-$, likewise $k_1 \geq k_+ \text{ or } k_1 \leq k_-$. In other words, $(\delta_+, \delta_-)$ 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 $\delta \leq 25$.

![Diagram of linear interpolation](image)

Figure 9: Example schema of linear interpolation amongst the pivots via projection onto $\delta$-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.

We interpolate the original return data matrix for each constituent of SPX using each of the pivot models as shown in Table 5. 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 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 a 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 top two eigenvalues generated by the pivot models (the difference for the remaining eigenvalues goes to zero) and the original eigenvalues for each constituent of SPX. As expected, the difference
diminishes as the number of pivots increases, 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. For demonstration we show the performance of the 5, 9, and 12 pivot models.

![Graph showing the performance of different pivot models](image)

Figure 10: Difference between the top two eigenvalues of the 5 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, 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 necessary for modeling the IVS of those options carrying mainly idiosyncratic risk.

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.
Figure 11: Difference between the top two 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 12: Difference between the top two 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 13: Original distribution across the various risk-classes using all 52 pivots for options on the constituents of S&P500.
From Figure 13, 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 5, and display the results in Figure 14. 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.

From Figure 14 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.

As previously mentioned, 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 see (Figures 10-12) 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 dimensionality reduction.
Figure 14: 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.
6 Second Dimensional Reduction: Cross-Sectional Analysis of Correlations via RMT

Based on the 9 pivot model, we perform PCA on the correlation matrix of all equity returns and implied volatility returns of the U.S. options market. In other words, for each of 3141\(^9\) options in OptionMetrics, we collect the historical data for the 9 implied volatilities used as pivots in our 9 pivot model, and organize them into a large data matrix of 31,410 columns.

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. Using this data we perform different experiments involving the 3141 equities as well as the constituents of the S&P500 (to be able to compare with Bouchaud and Potters [3]).

We begin our analysis with the underlying constituents of S&P500. We first examine the equities market as determined by these constituents, without the implied volatilities (like Bouchaud and Potters [3]). 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.

The boundary between significant and insignificant nodes is not easy to determine in practice. With the goal of better distinguishing fluctuations in the spectrum, we recall the Tracy-Widom law as we apply it to our analysis. Using the same set-up as in the Marchenko-Pastur case, the Tracy-Widom Law states:

**Proposition 2 (Tracy-Widom Law)** The distribution of the largest eigenvalue, \(\lambda_{\text{max}}\), of a random correlation matrix is given by:

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

---

\(\text{We only use those underlying assets with at least 500 days of available data}\)
with
\[ \mu_{TN} = (\sqrt{T - .5} + \sqrt{N - .5})^2 \]
and
\[ \sigma_{TN} = (\sqrt{T - .5} + \sqrt{N - .5}) \left( \frac{1}{\sqrt{T - .5}} + \frac{1}{\sqrt{N - .5}} \right)^{\frac{1}{2}} \]

Tracy-Widom holds for specific \( \beta \)-ensembles: Gaussian Orthogonal Ensemble: \( \beta = 1 \), Gaussian Unitary 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\(^\text{10}\).

We test the performance of MP and TW by taking the original data, and for each variable (i.e., for each implied volatility return or underlying asset return) we permute 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. We perform this randomization experiment for each of four markets: equities as determined by the constituents of the S&P500, options with underlying in the S&P500, all assets available in OptionMetrics, and all options with an underlying asset in OptionMetrics. The results for the former two markets can be found in the appendix. We display the results for the latter two markets below.

We 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.

\(^{10}\)We use the \( s \to F_1(s) \) table from Michael’s Prahofer’s website: http://www-m5.ma.tum.de/KPZ. In Figure 15 we plot the CDF \( F_1(s) \). For a more thorough treatment of this theory please refer to *Topics in Random Matrix Theory*, by Terence Tao [6].
Figure 15: 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 $\rightarrow s=1$. 
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.

We present results 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. Eighty five percent of the maximum eigenvalues of 10,000 randomized simulations lie below $\lambda_+ = 12.29$, and all maximum eigenvalues lie in the range (11.94, 12.53). For this market, we use 2026 observations and 3140 variables.

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. About 70 percent of all maximum eigenvalues for each of 12,667 randomized simulations lie below $\lambda_+$, and all maximum eigenvalues lie in the range (78.98, 80.32). For this market we use 500 observations and 31410 variables.
Figure 16: 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 17: 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.
Figure 18: 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 variables used each over 500 observations.
Figure 19: 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. Note that in this case the discrepancy between the theoretical and the empirical distribution is greater than for the other markets, but this does not in anyway alter our conclusion that we may approximate the largest eigenvalue of a random correlation matrix with the same underlying distribution as our empirical data via Tracy-Widom. This is because we care about large values of $s$ (as demonstrated in the tables that follow), and the two distributions converge for $s > 3$. 
The results above allow us to conclude that we may apply the Marchenko-Pastur distribution and the Tracy-Widom law in order to determine the significant eigenvalues is our original correlation matrix. Below we show the significant eigenvalues for each market considered.

Table 6: Significance of Eigenvalues in the entire Equities Market a la TW

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

- 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-Widom.
Table 7: Significance of Eigenvalues in the entire Options Market a la TW

<table>
<thead>
<tr>
<th>Top 110 Eigenvalues</th>
<th>s-value</th>
<th>$F_1(s)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_1 = 3742$</td>
<td>24843</td>
<td>1</td>
</tr>
<tr>
<td>$\lambda_5 = 209.27$</td>
<td>879.14</td>
<td>1</td>
</tr>
<tr>
<td>$\lambda_{10} = 143.5$</td>
<td>433.04</td>
<td>1</td>
</tr>
<tr>
<td>$\lambda_{20} = 118.19$</td>
<td>261.32</td>
<td>1</td>
</tr>
<tr>
<td>$\lambda_{40} = 102.62$</td>
<td>155.74</td>
<td>1</td>
</tr>
<tr>
<td>$\lambda_{50} = 97.40$</td>
<td>120.35</td>
<td>1</td>
</tr>
<tr>
<td>$\lambda_{70} = 90.48$</td>
<td>73.35</td>
<td>1</td>
</tr>
<tr>
<td>$\lambda_{90} = 84.56$</td>
<td>33.21</td>
<td>1</td>
</tr>
<tr>
<td>$\lambda_{107} = 80.21$</td>
<td>3.70</td>
<td>.9996</td>
</tr>
<tr>
<td>$\lambda_{108} = 80.04$</td>
<td>2.60</td>
<td>.996</td>
</tr>
<tr>
<td>$\lambda_{109} = 79.65$</td>
<td>-1.10</td>
<td>.80</td>
</tr>
<tr>
<td>$\lambda_{110} = 79.41$</td>
<td>-1.71</td>
<td>.35</td>
</tr>
</tbody>
</table>

- 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.

To summarize, we have determined that we may reduce dimension greatly in each of the four markets considered. Specifically markets on the underlying assets are determined by 15 and by 20 factors out of the original 440 and 3141 variables used (refer to appendix). The options markets on these underlying assets are determined by a comparable number of factors: 84 and 108 out of an original 4400 and 31,410 variable respectively.

Finally, we compare the correlation of the residuals of all assets in OptionMetrics and its corresponding Marchenko-Pastur distribution. We have determined that this market consists of 20 significant factors. Hence, we perform a multi-linear regression of the original data matrix on these 20 factors, determine the matrix of residuals, and compute the correlation of the residuals matrix. Since we wish to consider all information as co-movements of variables regardless of their variance, and in order to apply Marchenko-
Pastur, we normalize the correlation of the residuals matrix so that each column has variance equal to 1. Then we compute the spectrum of the correlation of residuals and plot the corresponding density of states (dos). The result can be seen in Figure 20.

Figure 20: Spectrum of correlations of residuals after projecting onto the top 20 eigenportfolios for all assets in OptionMetrics. Point mass of weight 84% for both distributions. We project onto the first 20 eigenportfolios (accounts for 24% of var.). Twenty eigenvalues lie outside of $\lambda_+ = 12.30$ and are deemed significant by Tracy-Widom. We cannot reject the null hypothesis that the two densities are the same based on Kolmogorov-Smirnov test at the 5 percent significance level.
“Idiosyncratic” assets have more activity on higher-order modes, whereas “systemic” assets’ IVSs are driven primarily by the first mode. 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 phenomenon indicates higher systemic risk for all options during high-volatility periods.

We formulate various pivot models which we 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.

We conclude by performing a PCA of the entire equities and options market as made available by OptionMetrics using a 9 pivot description of the IVSs. 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 determine the significant part of the spectrum.

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 ruling out “fat-tail” effects [2]. 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. Equities and options with underlying in SPX: 84 significant factors (account for 55% of variance).

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

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

Results 1 and 3 confirm those of Bouchaud and Potters [3]. Results 2 and 4 provide us with very large dimension reduction in the option-market space (84 significant factors from 4400, and 108 significant factors from 31,410 respectively).
A Appendix

For the sake of conciseness we do not include the code nor all of our results with this thesis. All code and databases of results are also ready upon request.

We made the assertion that the first eigenvalue $\lambda_1$ 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:

$$
\lambda_1 = V^{(1)^T}CV^1 = \sum_{i=1}^{N}(V^1_i)^2 + \sum_{i\neq j} V^1_i V^1_j \rho_{ij} = 1 + \sum_{i\neq j} V^1_i V^1_j \rho_{ij}
$$

$$
= 1 + \sum_{i\neq j} V^1_i V^1_j \frac{\sum_{i\neq j} V^1_i V^1_j \rho_{ij}}{\sum_{i\neq j} V^1_i V^1_j}
$$

Rearranging gives,

$$
\frac{\lambda_1 - 1}{\sum_{i\neq j} V^1_i V^1_j} = \frac{\sum_{i\neq j} V^1_i V^1_j \rho_{ij}}{\sum_{i\neq j} V^1_i V^1_j}
$$

$.\,$ Since $V^1_i \approx \frac{1}{\sqrt{N}}$ then $\sum_{i\neq j} V^1_i V^1_j \approx N \frac{N - 1}{N} = N - 1$

hence, $\frac{\lambda_1 - 1}{N - 1} \approx \rho$, so that, $\rho \approx \frac{\lambda_1}{N}$.
Figure 21: 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. Time period dates from August 31, 2008 until August 31, 2013.
Figure 22: 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. Time period dates from August 31, 2008 until August 31, 2013.
Figure 23: 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. Time period dates from August 31, 2008 until August 31, 2013.
References


