Implied Volatility Term Structure and Vega Hedging in the Q-Alpha-Sigma Model

Matthew Dixon, Paul Oreto, David Starr, Chen Zheng

June 8, 2008

Abstract

We examine the implied volatility surface using an alternative to the Black-Scholes model known as the q-alpha-sigma model. This model captures much of the smile and so reduces the volatility surface to a one-dimensional term structure curve. We study the dynamics of this curve using principal component analysis and a GARCH model. This allows us to implement a simple vega-hedging strategy.

1 Introduction

It is well known that the Black-Scholes model fails to accurately predict option prices. This failure can be parametrized by the so-called volatility surface. To do this one observes the option price, strike, stock price, time to maturity, and interest rate and then inverts the Black-Scholes equation to compute the *implied volatility*. The Black-Scholes model assumes that the volatility is a fixed property of the underlying asset. If this assumption were correct then the volatility surface would be a constant across options and static in time. However, implied volatilities are observed to vary both with strike and time to maturity. Furthermore, the surface fluctuates in time.

The cross-sections of the volatility surface that vary with strike are known as *smiles* or *smirks*, because of their characteristic shape. The cross-setions that vary with maturity are known as the *term structure*. The smile is believed to be a consequence of market participants adjusting for large stock movements, particularly negative ones. In this view, out-of-the-money options are uncharacteristically expensive because traders wish to insure themselves against steep stock declines or bet on large price increases. There is increased demand for these out-of-the-money options, and so they trade as if large stock movements were more likely, and this increases the implied volatility. This intuition agrees with the observation that the log returns of stocks are not normally distributed, as Black-Scholes assumes, but rather have fatter tails.

Understanding the volatility surface is equivalent to understanding the prices of options, and so great effort has been made to model its shape and dynamics. One common approach is to simply modify the volatility so that it varies across strike and maturity. While this method does a reasonable job of modeling the surface at one instant, it inadequately captures the dynamics of the surface. More sophisticated methods, known as stochastic volatility models, endow the volatility with stochastic dynamics of its own. Yet another class of models replaces the Brownian motion that governs the dynamics of the underlying asset with a different stochastic process. The q-alpha-sigma is one such model.

The q-alpha-sigma model replaces the Brownian motion with a different stochastic process, one which incorporates feedback between large and small scales. This models the fact that small-scale traders react directly to large-scale stock motions. The resulting stock returns have fatter tails and more accurately represent observed prices. Consequently, the q-alpha-sigma model does a respectable job of modeling the volatility smile. If one uses this model to invert q-alpha-sigma implied volatilities from observed option prices, then the resulting volatility surface is nearly flat across strikes. The two-dimensional volatility surface is effectively reduced to a one-dimensional term structure.

It is also crucial to understand the dynamics of the volatility surface. If one is managing a large book of options then it is important to hedge against all sources of risk. Fluctuations of the volatility surface represent one such source of risk.[1] The derivative of an option price with respect to volatility is known as *vega*, and the associated risk is known as *vega risk*. Effectively hedging vega risk is an area of active research.

We study both the shape and dynamics of the implied volatility term structure in the q-alpha-sigma model. We apply principal component analysis (PCA) and a GARCH model to the volatility fluctuations. The results of the PCA are then used to implement a simple vega-hedging strategy.

2 The Q-Alpha-Sigma Model

The q-alpha-sigma model was created to address a well-known deficiency of the famous Black-Scholes model. Black-Scholes assumes that the underlying stock process obeys geometric brownian motion with constant volatility,

$$dS = \mu S dt + \sigma S dW(t). \tag{1}$$

A consequence of demanding that the underlying obeys geometric brownian motion is that the log returns are gaussian distributed,

[1] One might naïvely argue that fluctuations of the volatility surface are precisely equivalent to fluctuations in price, and consequently volatility risk represents all of one's portfolio risk. However, the implied volatilities are inverted from the stock price, time to maturity, and interest rate, as well as the observed option prices. The contributions of these auxiliary factors to the changes in option price are therefore effectively removed when studying the volatility surface.

$$\log\left(\frac{S_i}{S_o}\right) = (\mu - \frac{1}{2}\sigma^2)T + \sigma W(T). \tag{2}$$

Addressing the problems with Black-Scholes is difficult because the assumptions of Black-Scholes are parsimonious and allow a simple closed form solution. Any loosening of the assumptions threatens to produce a option pricing formula that is too complicated and computationally intractable for use by the everyday trader.

In [Borland, PRL (2002), Borland L. and Bouchard, J.P. (2004)], Borland produced an option pricing formula which relaxes the Black-Scholes assumption of gaussian log returns but which can has a closed-form solution (i.e. reduces to deterministic numeric integration as opposed to Monte Carlo).

This model assumes that the underlying obeys the following distribution,

$$d(\log\frac{S(t)}{S_o}) = \mu dt + \sigma S_o^{1-\alpha} S^{\alpha} d\Omega, \tag{3}$$

where

$$d\Omega = P(\Omega)^{(1-q)/2} dW. \tag{4}$$

The probability function $P(\Omega, t)$ is given by,

$$P(\Omega, t) = \frac{1}{Z(t)}(1 - \beta(t)(1-q)(\Omega(t) - \Omega(0)))^{1/(1-q)}, \tag{5}$$

where $\beta(t)$ and $Z(t)$ are constants chosen such that

$$P_q(\Omega(t)) = \delta(\Omega(t) - \Omega(0)). \tag{6}$$

The form of $P(\Omega, t)$ is often called a "Tsallis" distribution. The constant q paramatrizes the deviation from gaussianity. Studies across various liquid market securities [Borland, Quant. Fin. (2002)] have found in general $q \approx 1.3 - 1.5$. Using this form for the underlying gives an expression for a plain-vanilla European call (which we do not reproduce here, for brevity, see [Borland, PRL (2002)]) that is more complicated than Black-Scholes, but which is still numerically tractable.

The q-alpha-sigma model appears to eliminate the smile (i.e. the variation in the volatility surface in the moneyness direction). Figure 2 in Borland's paper shows how in inverting with Black-Scholes the q-alpha-sigma model option prices with constant volatility, one obtains a smile.

3 The Data: An Overview

Our data set tracks the behavior of various stocks and their options during the period from March 3, 1997 to May 31, 2006. The relevant stock symbols are ALL, AMGN, AXP, CSO, DIS, PG, T, UIS, and XRX. These stocks were chosen because of the quality of data available for them, in particular the consistency in the number of option series traded on them: six

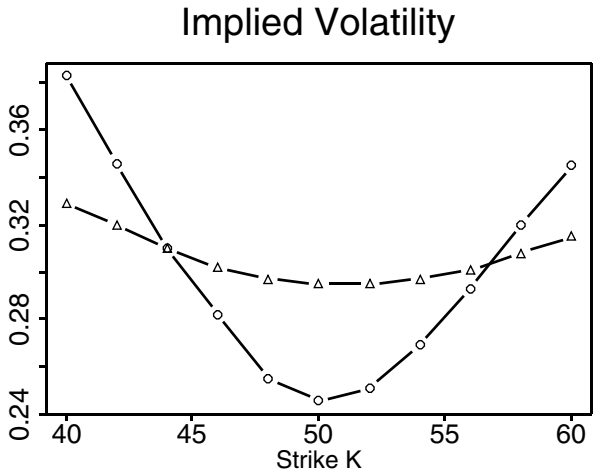

Figure 1: This figure is taken from [Borland, PRL (2002)]. Call option prices are generated with $q = 1.5$, $\sigma = .3$, $S_o = 50$, $r = .06$. The line with triangles is the BS implied volatility calculated using these option prices with T = .01 (the circles) and T = .4 (the triangles).

series over the period studied. The series traded were 1-, 2-, 5-, 8-, 20-, and 32-month options. Throughout this paper all options are at-the-money (ATM), American-style call options.

EvA provided us with data on the daily implied q-alpha-sigma volatility of each option and the interest rate appropriate for investment over the lifetime of the option. We collected daily stock prices (the daily close) from Yahoo! Finance. Additionally, we used the OptionMetrics database to determine the strike of each day's at-the-money option and the price of that option (which we took to be the average of the best closing bid and offer). Finally, we computed the q-alpha-sigma delta and vega for each option using C++ libraries supplied to us by EvA. All statistical analysis was performed in R.

4 Principal Component Analysis

4.1 Background

Principal Component Analysis (PCA) is a statistical technique used to dimensionally reduce a complicated data set by finding the uncorrelated factors that contribute most to the variation of that data. We are interested in fluctuations of the implied volatility surface, and so PCA will determine for us the surface deformations that dominate those fluctuations.

Consider an n-dimensional time series $X_i(t)$. Here i labels the component of the observation (and so ranges from 1 to n), and t is the time of observation. The covariance matrix C_{ij} of this time series is defined by

$$C_{ij} = \mathbb{E}_t \left[\left(X_i - \overline{X_i} \right) \left(X_j - \overline{X_j} \right) \right], \tag{7}$$

where the mean $\overline{X_i}$ and expectation $\mathbb{E}_t$ are over time. The eigenvectors of this matrix are known as the PCA components; by definition they have zero covariance. These eigenvectors represent a new basis for performing measurements, different from the basis (labeled by i) in which measurements were actually made. Measurements made along different axes in this new basis are uncorrelated. The eigenvalues are, by definition, the variance of the sample along the corresponding axes.

The eigenvectors with the largest eigenvalues account for the greatest variation in the sample. In many cases the first few eigenvalues are much larger than the rest. In such situations the corresponding eigenvectors might represent important physical quantities, while the rest are negligible. One may then truncate the space of observations to only those PCA components with large eigenvalues, a process known as dimensional reduction.

4.2 The Term Structure Curve

For each stock we construct a volatility term structure curve from the observed implied volatilities of traded options. This is done by linear interpolation between observed volatilities (with horizontal interpolation outside of the observed maturity range). A separate term structure curve is generated for each day of trading. We also explored the possibility of interpolation using various spline techniques, but we could not justify the additional structure that was introduced. See Figure 2 for a sample term structure curve.

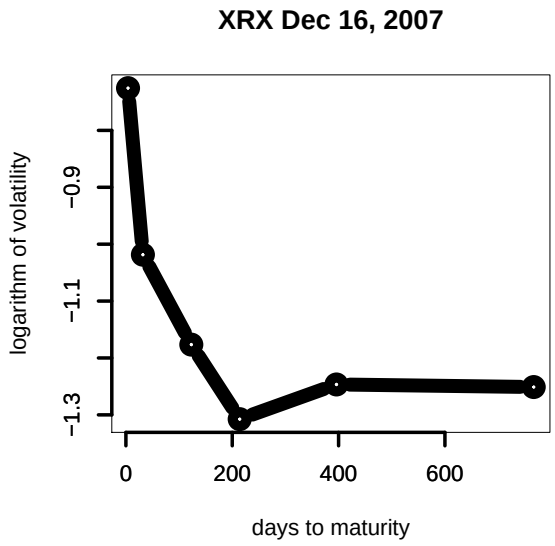

Figure 2: A linearly interpolated volatility term structure curve

It is important to note that, for example, an option that is released with a maturity of 8 months will eventually become a 7-month option and then a 6-month option, until it

becomes the 5-month option and a new 8-month option is released. If one bins the maturity axis in months then this leads to discontinuous motion of traded options along the axis. Such motion would be convoluted with authentic shifts in the volatility surface, and it would be difficult to disentangle the two effects.

There are two possible resolutions. One is to view each option series as static: the 8-month option might be considered a "7-month" option throughout its lifetime, even when it is trading at 8 or 6 months maturity. While this diminishes the frequency with which traded options switch maturity bins, the effect still exists and, when it does exist, would be greatly exaggerated. Instead, we bin the maturity axis in days. In this way the options move relatively smoothly along the volatility surface. It also most accurately samples the idealized, smooth term structure curve.

The vertical axis of the term structure curve is the logarithm of the implied volatility. While this in principle makes no difference if we were to study the term structure itself, it is important because we are performing PCA on the time series of fluctuations, and PCA assumes that the data is normally distributed. We find that the fluctuations of the logarithms are more normally distributed than the fluctuations of the volatilities (see section 5). Additionally, it is possible (when modeling the fluctuations) for a large shift in one PCA component to push the term structure curve negative. By working with the logarithms one can guarantee that the resulting volatilities (gotten through exponentiation) are indeed positive, as one would like.

4.3 Term Structure Fluctuations

The time series $X_i(t)$ we study is the series of volatility surface fluctuations, where the index i labels the points at which the term structure is sampled. In principle one could study the term structure as a fully continuous object, in which case the index i would be a continuous label along the maturity axis. For our study we sample the term structure curve at a discrete set of maturities labeled by i.

As previously mentioned in section 4.2 and discussed later in section 5, we study the term structure in terms of the logarithms of the implied volatilities. This means that the time series may be written as

$$X_i(t) = \Delta \log \sigma_i(t) = \log \sigma_i(t) - \log \sigma_i(t+1) \tag{8}$$

where the σ_i are the implied volatilities at maturity i.

The choice of sample points i is somewhat tricky, as the sampling method can affect the results of the PCA. We will return to this point later. To make progress, we shall for definiteness choose a particular sampling method and return later to the question of how this affects the PCA results.

4.4 Principal Component Analysis

To start we shall try sampling the term structure curve at every month i from 1 to 32 months. The term structure curve is measured in days, so we shall define a month to span

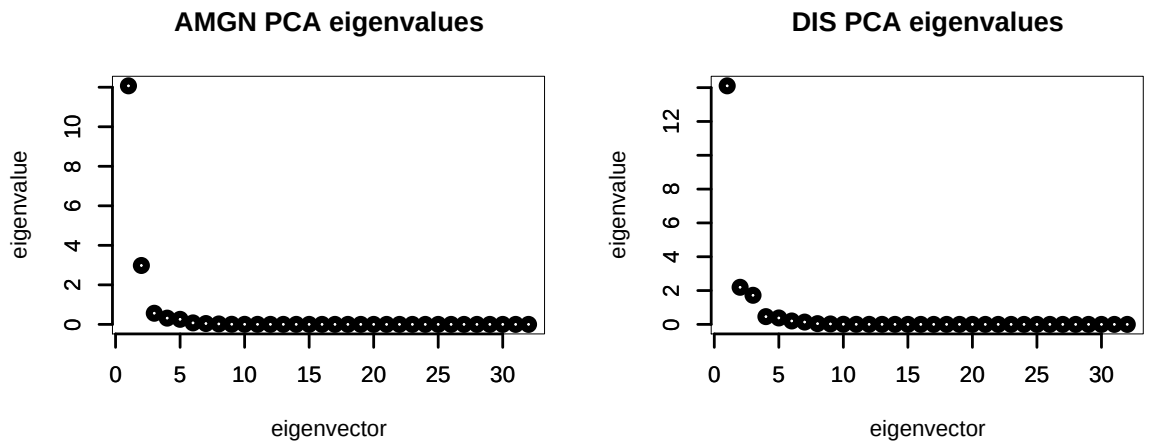

Figure 3: PCA eigenvalues for the AMGN and DIS symbols, using monthly sampling. Each is dominated by the first two or three eigenvectors.

precisely 30 days. Thus we sample the term structure at 30 days, 60 days, . . . , 960 days. The resulting time series is 32 dimensional and consists of the daily fluctuations in the logarithms of the implied volatilities. From this time series we compute the covariance matrix (7) and diagonalize to obtain the PCA components.

The eigenvalues for AMGN and DIS can be seen in Figure 3. Clearly the first two or three PCA components dominate over the others. It can be useful to look at the PCA eigenvectors to understand what the relevant deformations of the volatility surface look like. The eigenvectors for AMGN and DIS can be seen in Figures 4 and 5.

One is immediately aware of a significant problem. The AMGN PCA eigenvectors are peaked at particular maturities. Recall that the traded options have maturities of 1, 2, 5, 8, 20, and 32 months. The first PCA eigenvector is clearly peaked at the 32-month option, the second at the 20-month option, and the third at the 8-month option. This is a problem because it means that the implied volatilities of traded options move independently of each other. If one wishes to use the PCA eigenvectors to hedge an option portfolio, then one must exploit correlations across maturities. This appears to be absent. The situation is similar for the DIS eigenvectors, although in that case the third eigenvector is more broadly peaked at the short maturities.

One might argue that this behavior is a consequence of our sampling method. In particular, we sample the term structure curve at every month, from 1 to 32 months. We linearly interpolate between the implied volatilities of traded maturities, and these traded maturities become increasingly sparse as one looks at later maturities. For example, if the 32-month option will expire in 26 months, then the motion of this option alone is responsible for all 7 sample points from 26 to 32 months. Similarly, if the 20-month option will expire in 14 moths, then the 32-month and 20-month options alone are responsible for the motion of all

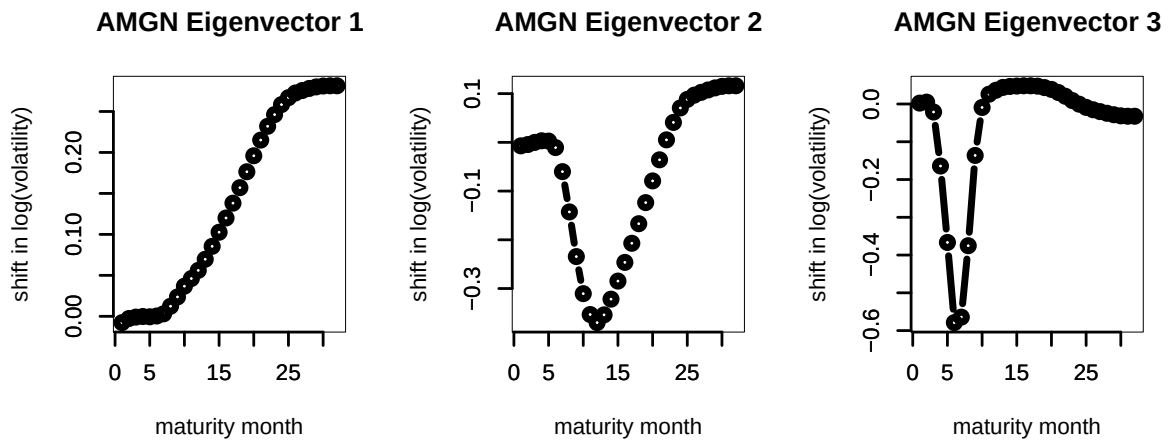

Figure 4: First three PCA eigenvectors for the AMGN symbol, using monthly sampling. These are strongly peaked at specific, late-maturity options.

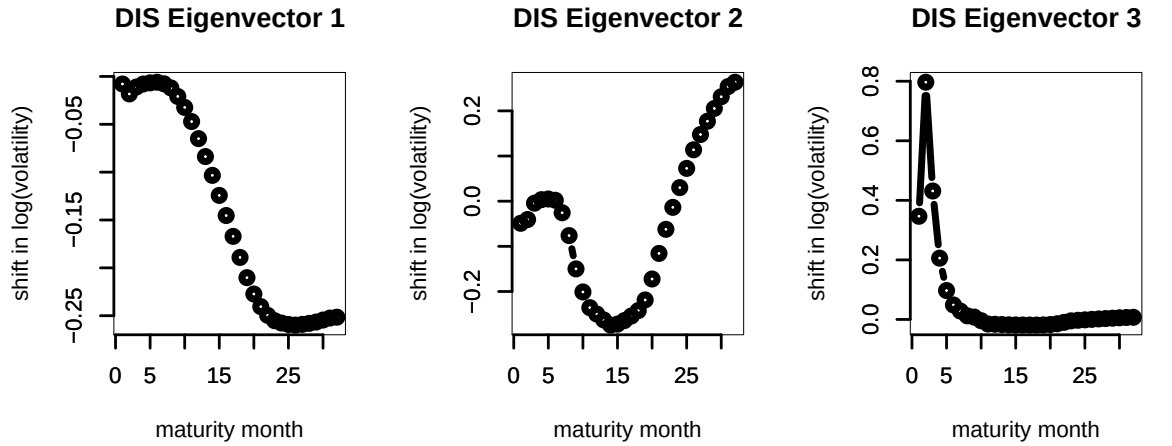

Figure 5: First three PCA eigenvectors for the DIS symbol, using monthly sampling. The third eigenvector is more generally peaked at young maturities than for AMGN.

sample points between 14 and 26 months. Thus these PCA eigenvectors would appear to be a simple consequence of our sampling scheme.

To test this hypothesis we try a second sampling method. If we sample the data at 1, 2, 4, 6, 12, and 24 months then the eigenvectors are as in Figures 6 and 7. The eigenvectors are still peaked, though slightly differently than before. In particular, the relative importance of eigenvectors has shifted for DIS. Also, the DIS eigenvectors have changed their shape somewhat, though the dominant two are still quite peaked. The eigenvalues, as seen in Figure 8, indicate that the first few eigenvectors do still dominate, though less strongly than before. This agrees with our previous intuition, that the monthly sampling weighted attributed too much change to certain maturities.

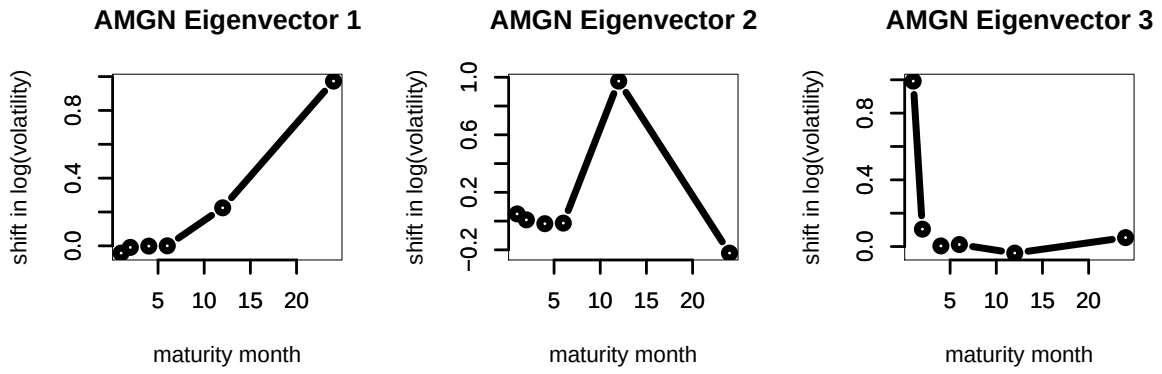

Figure 6: PCA eigenvectors for the AMGN symbol, using custom sampling at 1, 2, 4, 6, 12, and 24 months. The dominant eigenvectors have not changed much.

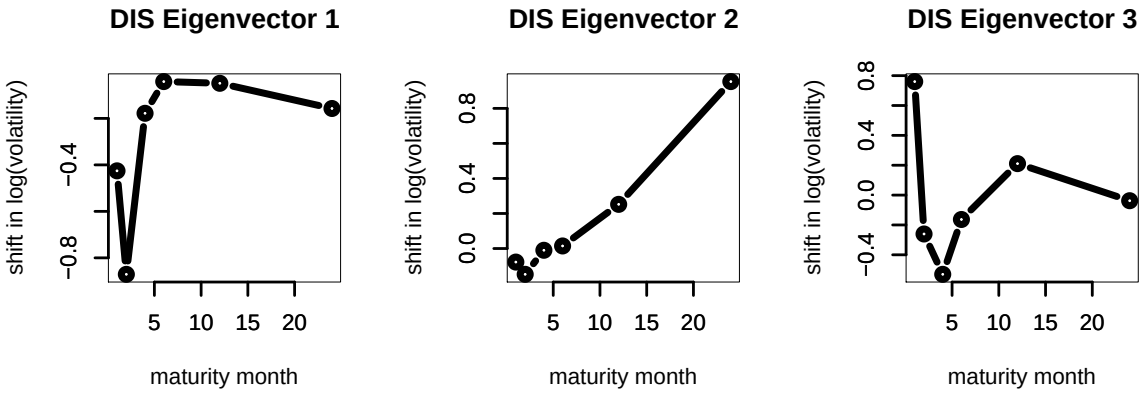

Figure 7: PCA eigenvectors for the DIS symbol, using custom sampling at 1, 2, 4, 6, 12, and 24 months. The dominant eigenvectors have shifted in relative importance, and the shapes have changed somewhat.

This introduces a new problem. What constitutes the 'correct' sampling method? One can sample the term structure curve in an arbitrary manner, with different weights applied to different maturities. Our first sample method approximates an even weight distribution over the entire curve, while the second approximates delta-function weighting at the traded maturities. This ambiguity corresponds to a functional degree of freedom that we were unable to address systematically. Perhaps the ultimate solution is to tailor the weighting so as to reflect the characteristics of the portfolio one is trying to hedge. Throughout the rest of the PCA analysis and discussion of vega hedging, we will use the second, customized weighting scheme.

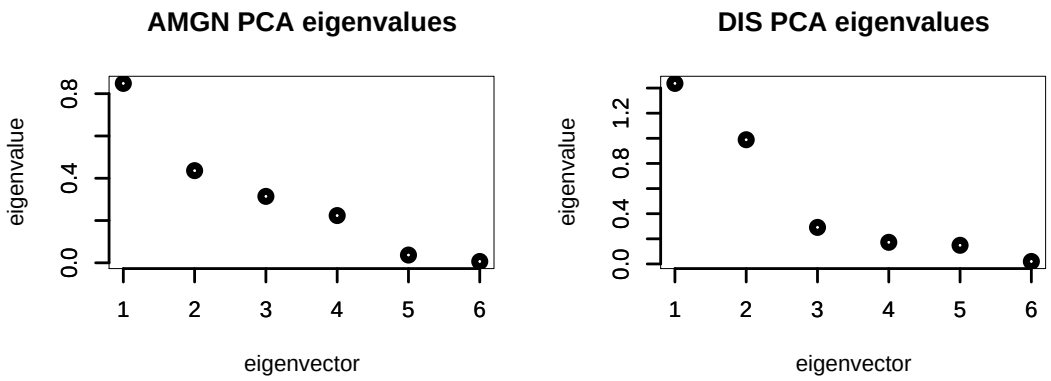

Figure 8: PCA eigenvalues for the AMGN and DIS symbols, using custom sampling at 1, 2, 4, 6, 12, and 24 months. The first few eigenvectors still dominate, though not as strongly as before.

We would like to understand how useful the dominant principal components are for explaining the fluctuations of the volatility surface. How many PCA components are required to capture most of the daily fluctuations? We project each daily fluctuation $X_i(t)$ onto the PCA eigenvectors $e_i(j)$, where j labels the eigenvectors. Clearly one can reproduce the daily fluctuation in its entirety by projecting onto all of the eigenvectors:

$$\vec{X}(t) = \sum_j \left(\vec{X}(t) \cdot \vec{e}(j)\right) \vec{e}(j). \tag{9}$$

However, one might truncate this sum by projecting onto only the first n eigenvectors:

$$\vec{X}^n(t) = \sum_{j=1}^{n} \left(\vec{X}(t) \cdot \vec{e}(j)\right) \vec{e}(j). \tag{10}$$

If one normalizes the size of the truncated projection $||\vec{X}^n(t)||$ by the size of the full fluctuation $||\vec{X}(t)||$, this gives a measure of how much change is captured by the first n eigenvectors:

$$\text{amount of change explained} = \frac{||\vec{X}^n(t)||}{||\vec{X}(t)||} \tag{11}$$

For a fixed number of eigenvectors n, one can average the amount of change explained over time to gather an estimate of how well those eigenvectors capture the change in the sample.

This analysis can be seen in Figure 9. Including only the three most dominant eigenvectors explains approximately 80% of the daily change, on average. Of course this analysis is somewhat unfair, as the PCA eigenvectors were computed from the same data on which the eigenvectors are tested. A more appropriate test is to construct PCA eigenvectors from

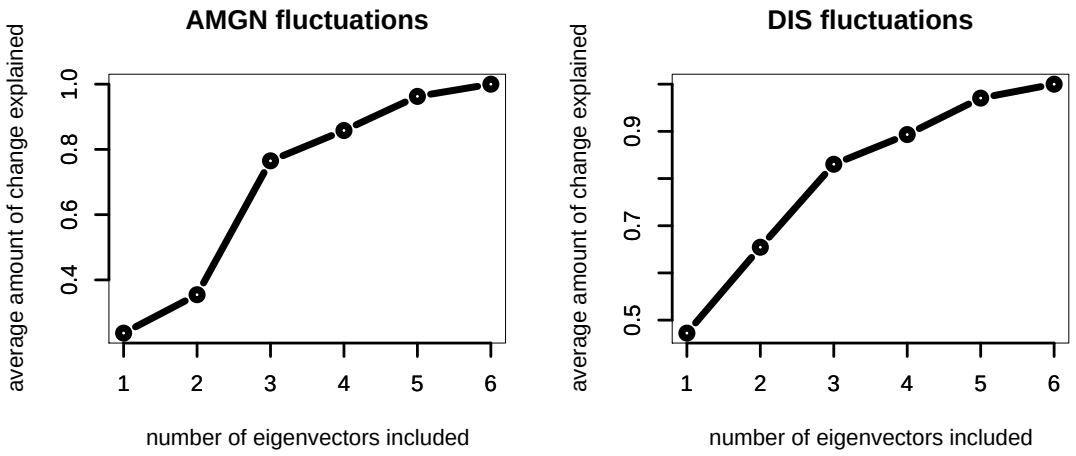

Figure 9: Average amount of change explained, using definition 11. Custom sampling at 1, 2, 4, 6, 12, and 24 months

the first half of the data sample and see how much of the change they explain when applied to the second half of the data sample. Such out-of-sample tests can be seen in Figure 10. The explanatory power is reduced somewhat, but three eigenvectors still account for about 70-85% of the daily change.

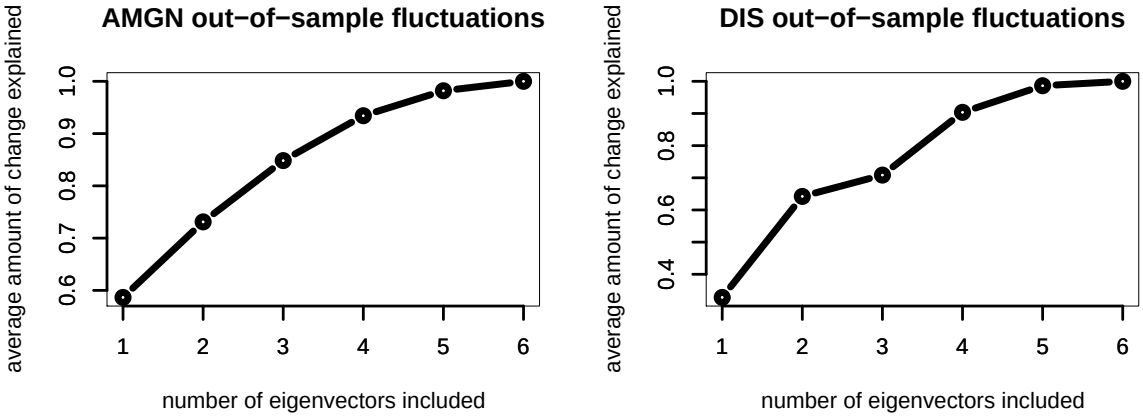

Figure 10: Average out-of-sample change explained, using definition 11. Custom sampling at 1, 2, 4, 6, 12, and 24 months

5 Autocorrelation & Normality Test

5.1 Introduction

The times series of the volatilities usually exhibit high autocorrelation and also considerable skewness. Therefore we want to study the time series of the logarithms of their daily differences instead. We will perform normality tests on this time series.

5.2 Methods

5.2.1 Autocorrelation

The autocorrelation function $R(t,s)$ of a random process X_t with mean μ and variance σ^2 is defined as

$$R(t,s) = \frac{E[(X_t - \mu)(X_s - \mu)]}{\sigma^2}. \tag{12}$$

We can also adopt the Ljung-Box test to confirm whether the autocorrelations of the time series are statistically different from zero. Instead of testing randomness at each distinct lag, it tests the overall randomness based on a number of lags. The test statistic is defined as,

$$Q = n(n+2)\sum_{j=1}^{h}\frac{\rho_j^2}{n-j}, \tag{13}$$

where n is the sample size and ρ_j is the sample autocorrelation at lag j, and h is the number of lags being tested. For significance level α, the hypothesis of randomness is rejected if $Q > \chi^2_{1-\alpha,h}$.

5.2.2 Normality

We will use the Shapiro-Wilk test and make quantile-quantile plots and histograms to study the normality of the log volatility time series.

The test statistics for the Shapiro-Wilk test is

$$W = \frac{(\sum_{i=1}^{n} a_i x_{(i)})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}, \tag{14}$$

where $x_{(i)}$ is the ith order statistic, $\bar{x}$ is the mean, the constants a_i are given by $(a_1, ..., a_n) = \frac{m^T V^{-1}}{(m^T V^{-1} V^{-1} m)^{1/2}}$, m are the expected values of the order statistics of independent and identically-distributed random variables, and V is the covariance matrix of those order statistics.

The Q-Q plot is a graphical method to diagnose differences between the probability distribution of a statistical group from which a random sample has been taken and a comparison distribution. If the population distribution is the same as the proposed theoretical distribution, there should be a straight line, especially near the center. The significant deviation from the line indicates the statistician rejects the null hypothesis that the two samples come

from the same underlying distribution. We make Q-Q plots to compare the probability distribution of log vol with normal distribution.

Finally, we draw histograms directly to look at the true distributions and visually confirm whether they resemble a normal distribution.

5.3 Results

We chose a subgroup of our data - five tickers (XRX, UIS, AMGN, DIS, PG), and test the time series of the mean term structures in various ways. The results of the autocorrelation, Q-Q plot, and histogram are assembled in Figure 11.

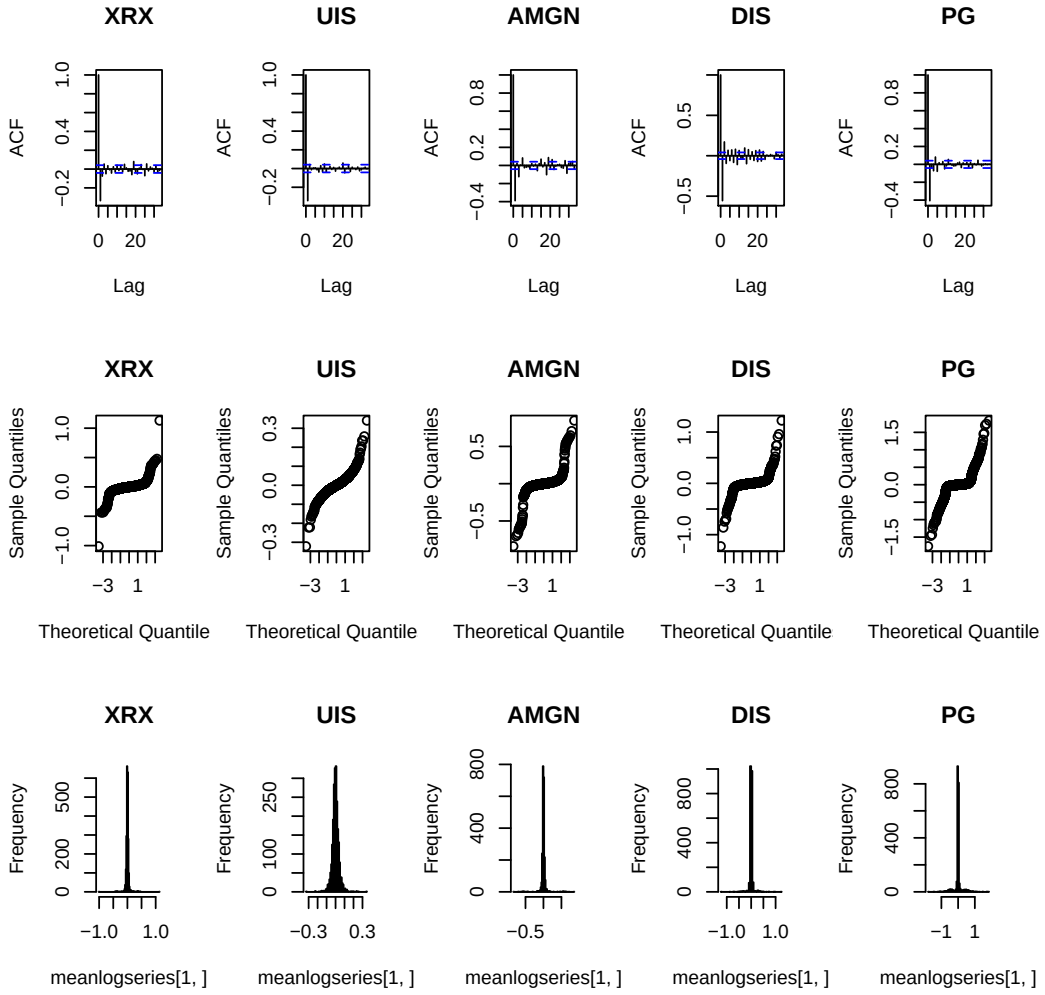

Figure 11: The first row shows the autocorrelation of the time series of log volatility term structure; the second row is a set of Q-Q plot; the third row is the histogram of the data.

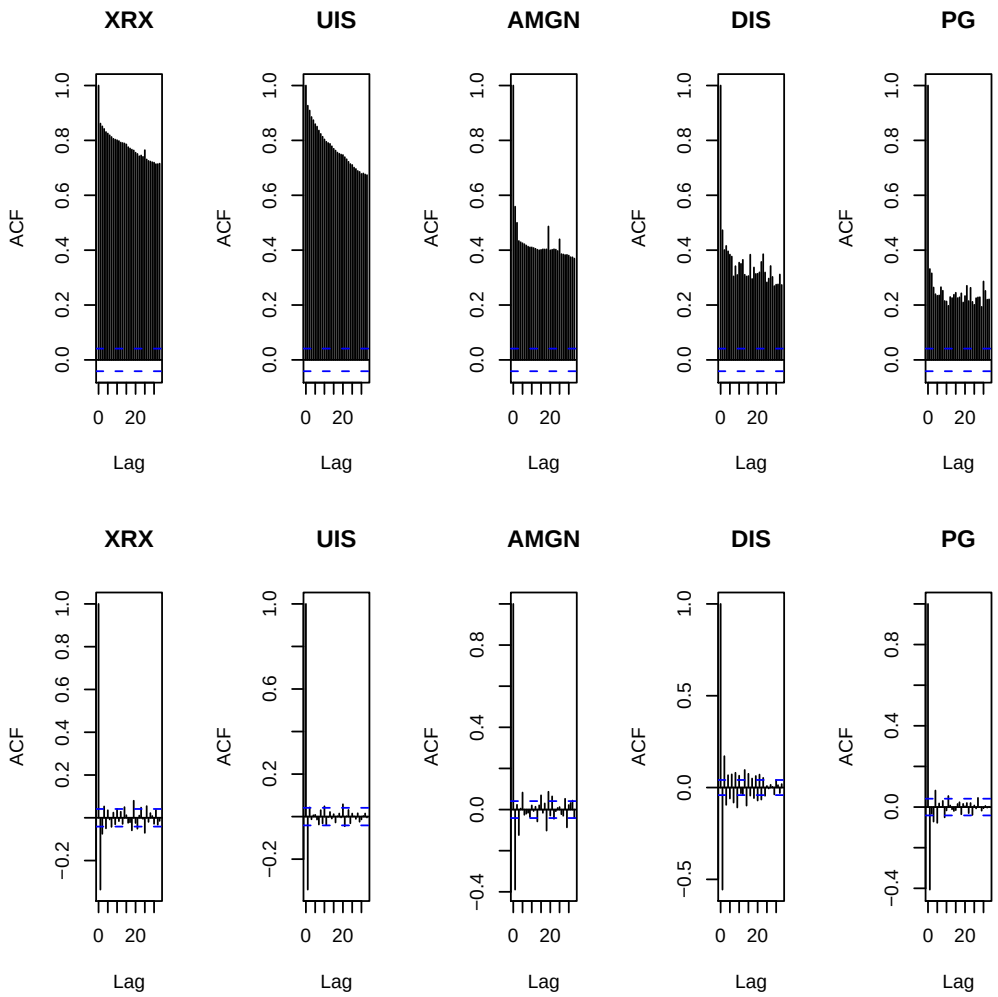

Figure 12: The time series of volatilities usually show high autocorrelation while the time series of daily differences of log volatility show very little autocorrelation.

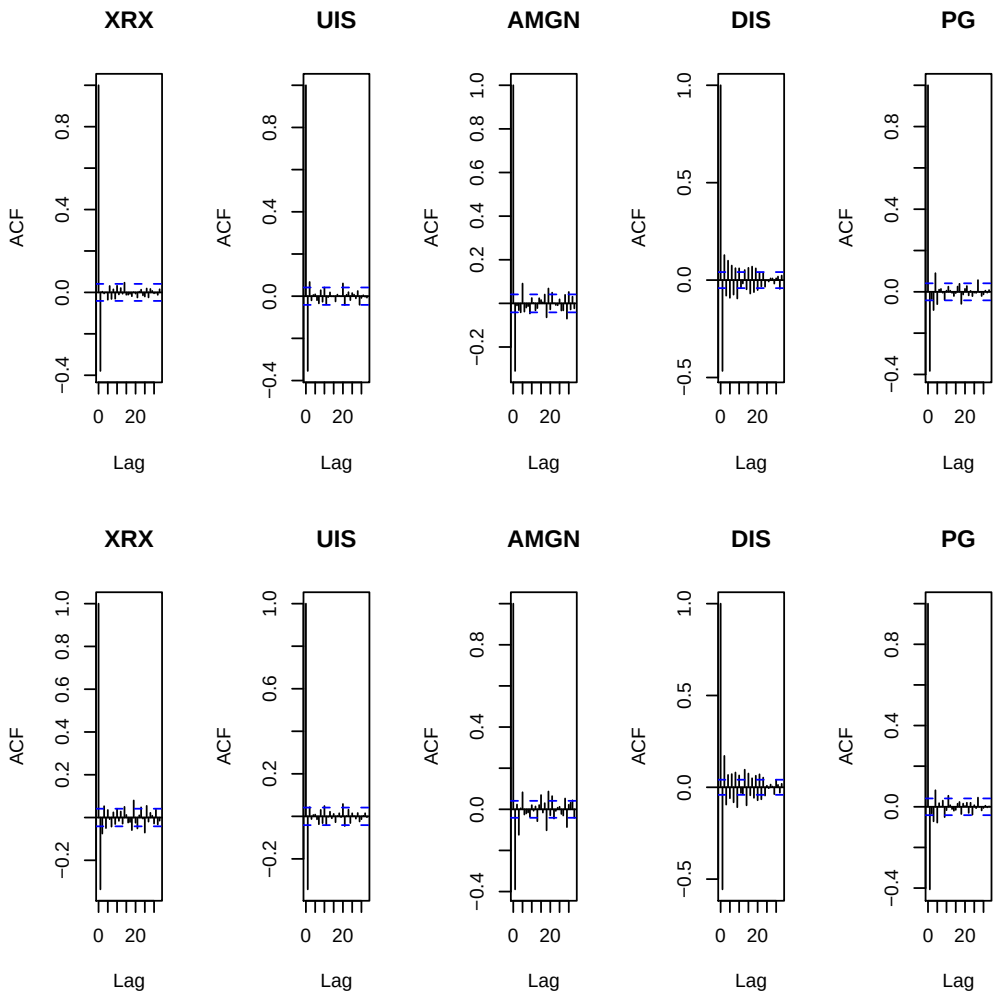

Figure 13: The top panel shows the autocorrelation existing in the time series of volatility daily differences while the bottom shows the autocorrelation of the time series of log volatility daily differences.

As we can see, the time series of the daily differences of log volatility term structure has little autocorrelation, compared to the original volatility time series (see Figure 12). The small values of of q (roughly 2×10^{-16}) from the Box-Ljung tests also confirm that conclusion. In order to understand whether such reduction of autocorrelation comes from the study of daily differences or the log function or both, we compare the autocorrelation plots of the time series of the daily differences of the volatility term structure with those of the time series of the daily differences of the log volatility term structure. These two sets of plots (see Figure 13) actually have very little difference, indicating the log function does not help to reduce the autocorrelation significantly in the time series study.

The Q-Q plots and the flat-tail shape of the histograms indicate a certain degree of non-normality, same as the test results from Shapiro-Wilk (p values are extremely small). Also we can see from Figure 14 and Figure fig:histcomp that the log function helps to improve the normality but the improvement is not significant. However, the study of the daily differences dramatically improves the normality of the data (see Figure 16 and Figure 17).

5.4 Discussions & ICA

The test results support our previous normality assumption about PCA analysis. We can see the study of the daily differences of the log volatility time series improve the quality of those assumptions significantly. On the other hand, the certain degree of the non-normality generated from the fat tail of the distribution (or the spikes on the volatility surface, and we have not cut them out in our PCA analysis) may bring the limitations of such analysis.

As we know, PCA finds a set of projections in a low-dimensional subspace that maximizes the likelihood of the data points, assuming the noise of the data is Gaussian. When this Gaussian assumption is violated, the dimensionality reduction schemes could be substantially different from PCA. The projections that lie in a linear subspace will typically correspond to a nonlinear surface in the space of the data. In other words, PCA simply de-correlates the axes but not guarantees the independency of the components for non-normal data, although such independency of the components is crucial for modeling each contribution to the volatility separately.

We thus try an alternative way - Independent Component Analysis, to decompose the volatility surface. The basic goal of ICA is to find a transformation of the data in which the components are statistically as independent from each other as possible. This method is similar as a classical PCA but it takes into account the clustering and independence of the components in the data set and could simplify the explanation of each individual contributions. It starts to be a new method to study the time series of volatility term structure recently [Ane & Labidi (2001), Wu & Yu (2005)].

We adopt the fastICA algorithm (developed by Marchini, Heaton, and Ripley) by choosing three independent components based on the PCA analysis and perform it on the same data set. Then we compare the first three ICA projection components with the corresponding PCA projection components (see Figure 18). The strong condensation of the ICA projection components on PCA projection, especially for ticker AMGN and XRX, confirms that our PCA analysis does find the real contribution, although the maximum variance components

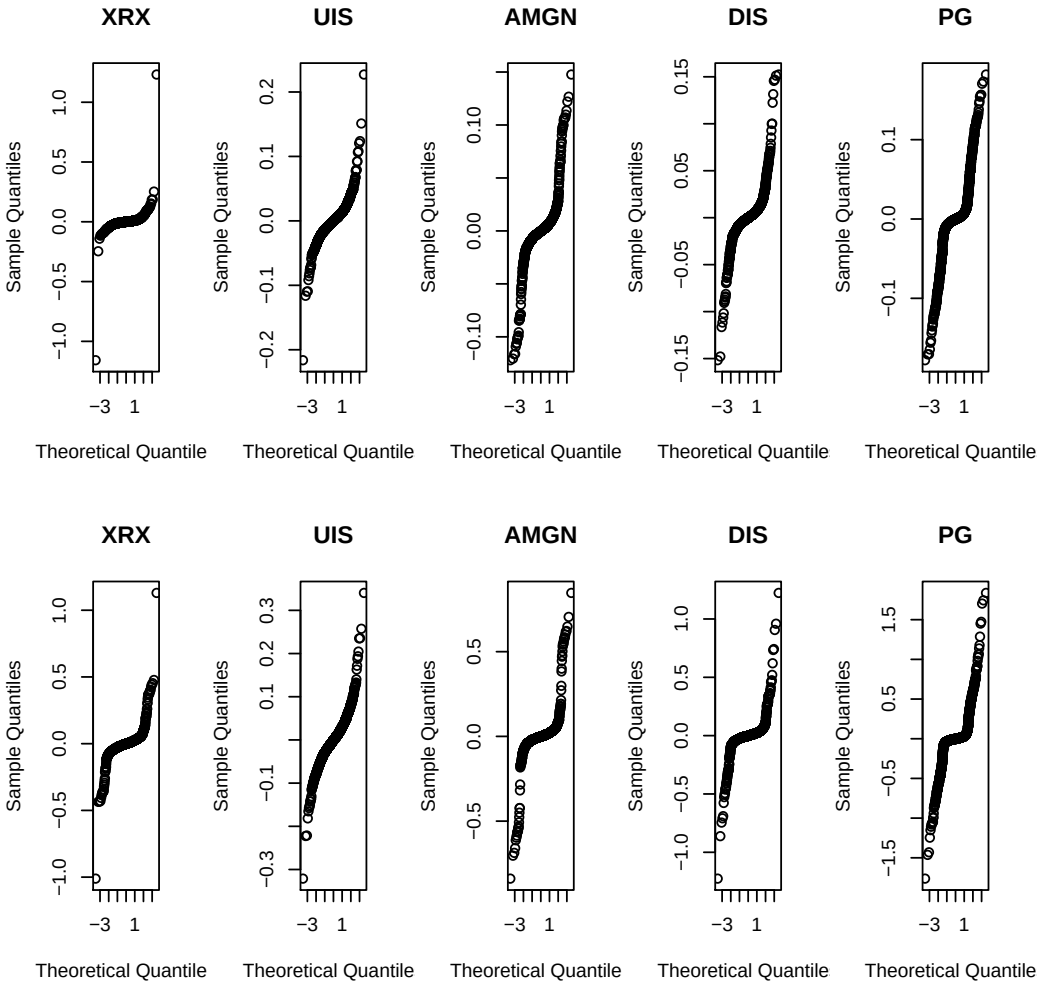

Figure 14: The first panel shows the Q-Q plots for the time series of volatility daily differences while the second one shows the Q-Q plots for the time series of log volatility daily differences.

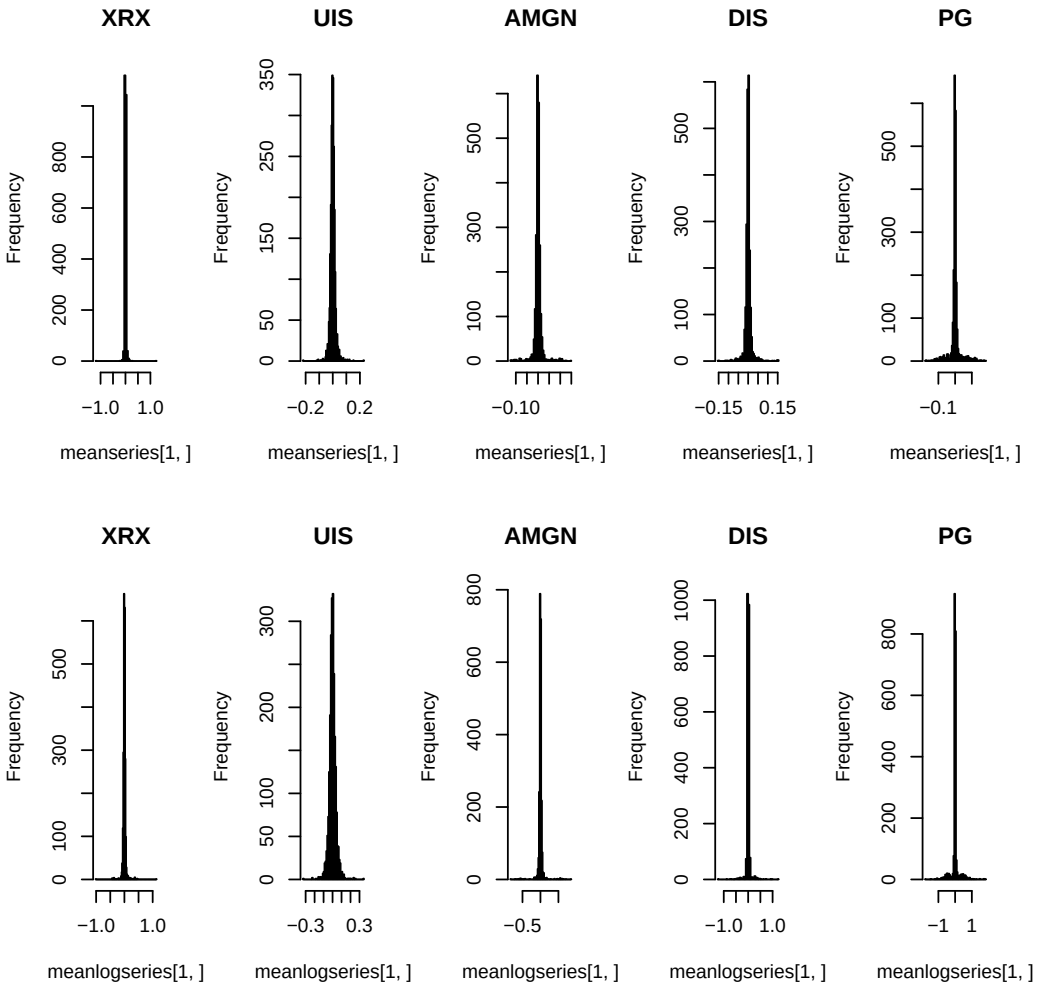

Figure 15: The first panel shows the histograms for the time series of volatility daily differences while the second panel shows the histograms for the time series of log volatility daily differences.

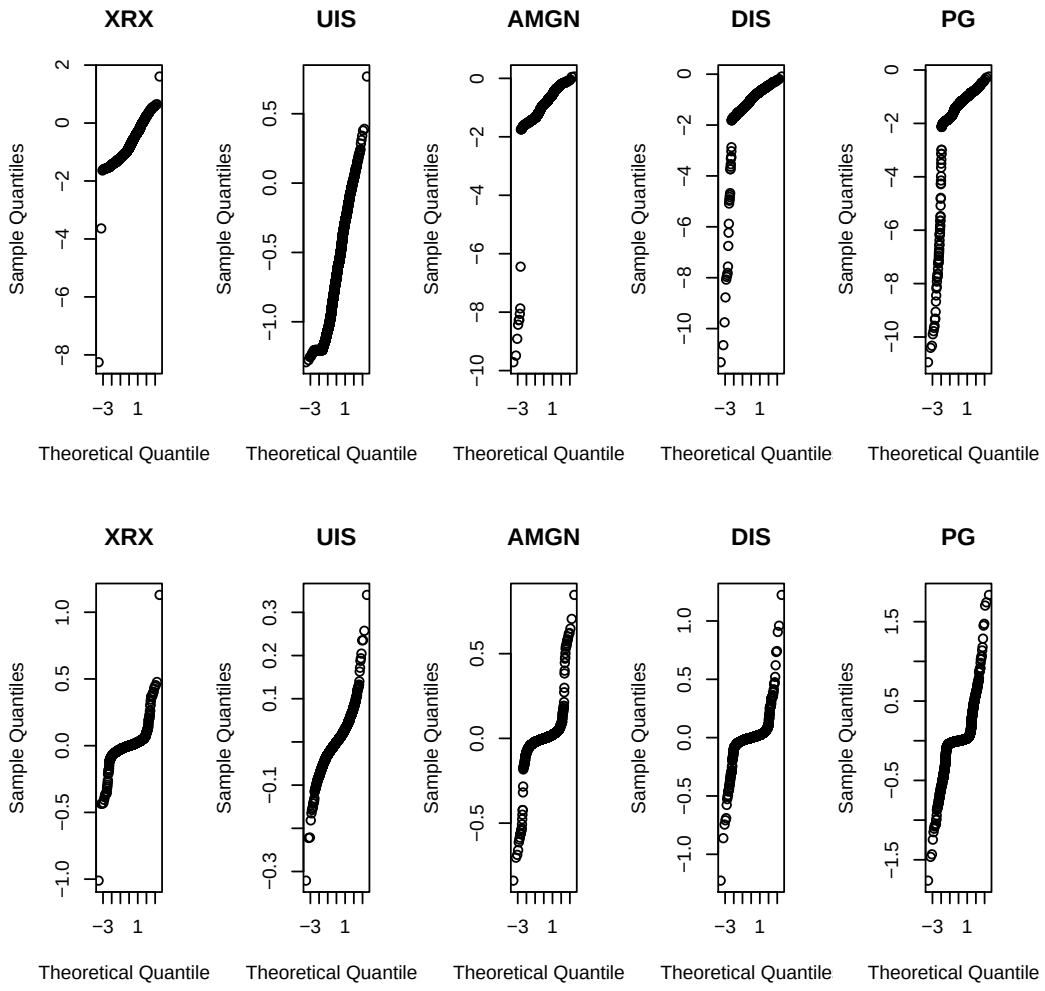

Figure 16: The first panel shows the Q-Q plots for the time series of volatility while the second one shows the Q-Q plots for the time series of log volatility daily differences.

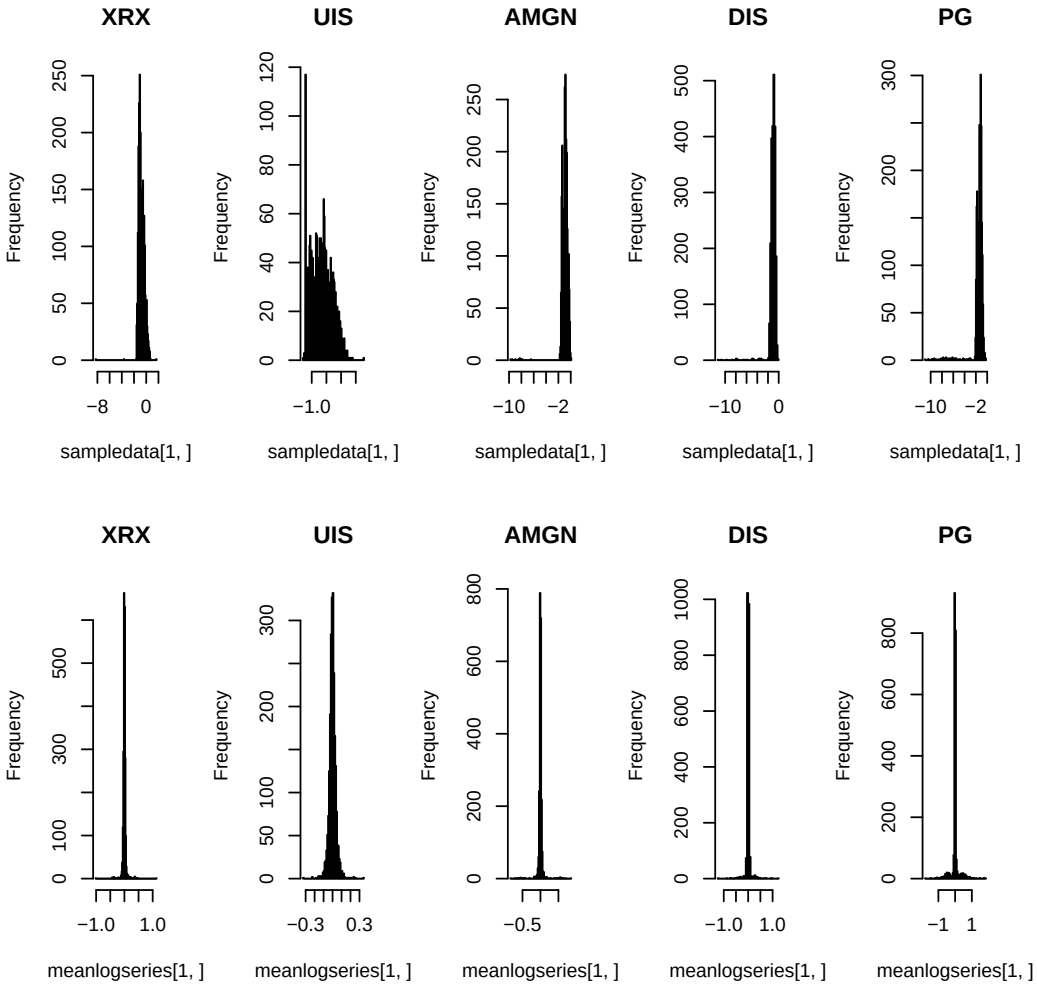

Figure 17: The first panel shows the histograms for the time series of volatility while the second panel shows the histograms for the time series of log volatility daily differences.

are not statistically independent modes. We also look at the ICA vectors (see Figure 19) and compare them with PCA eigenvectors (see Figure 20). The ICA vectors still have sharp peaks. Thus it seems not able to remove the same difficulty for the vega hedging as PCA analysis.

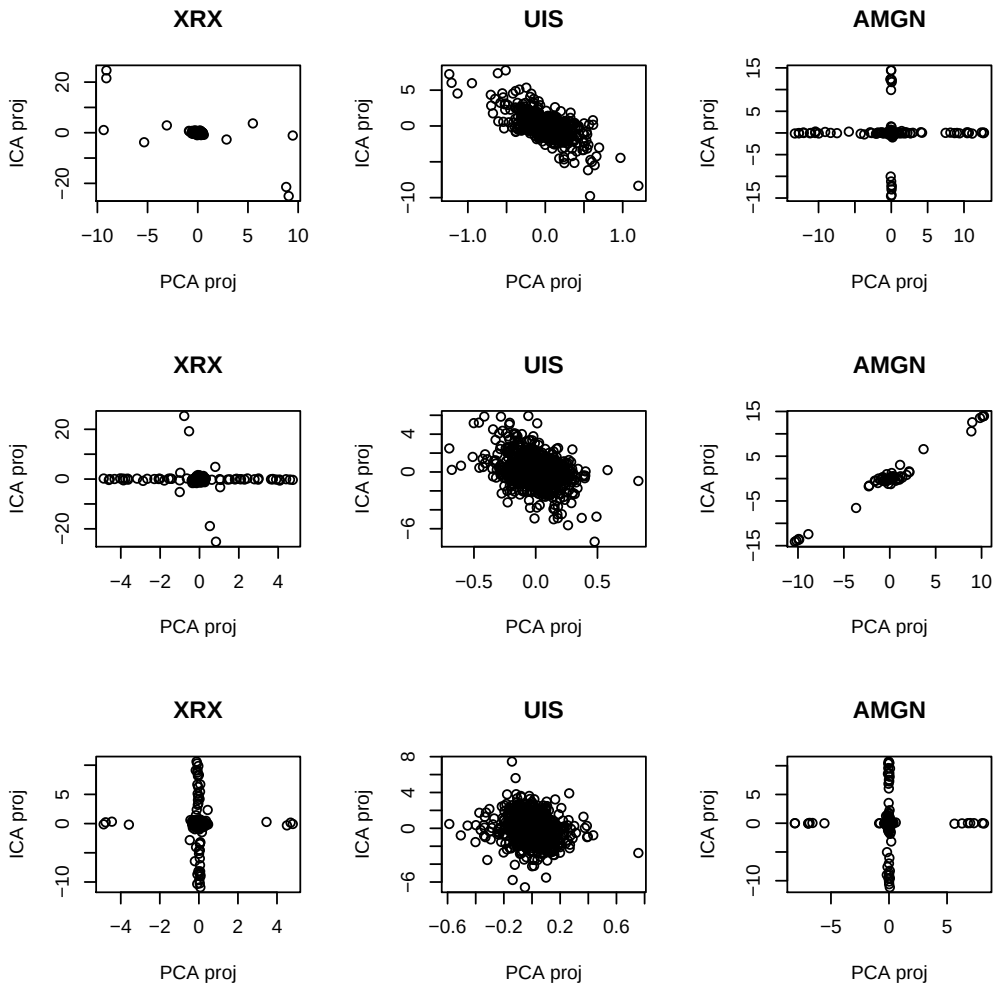

Figure 18: The comparison between the first three ICA projection components and PCA projection components.

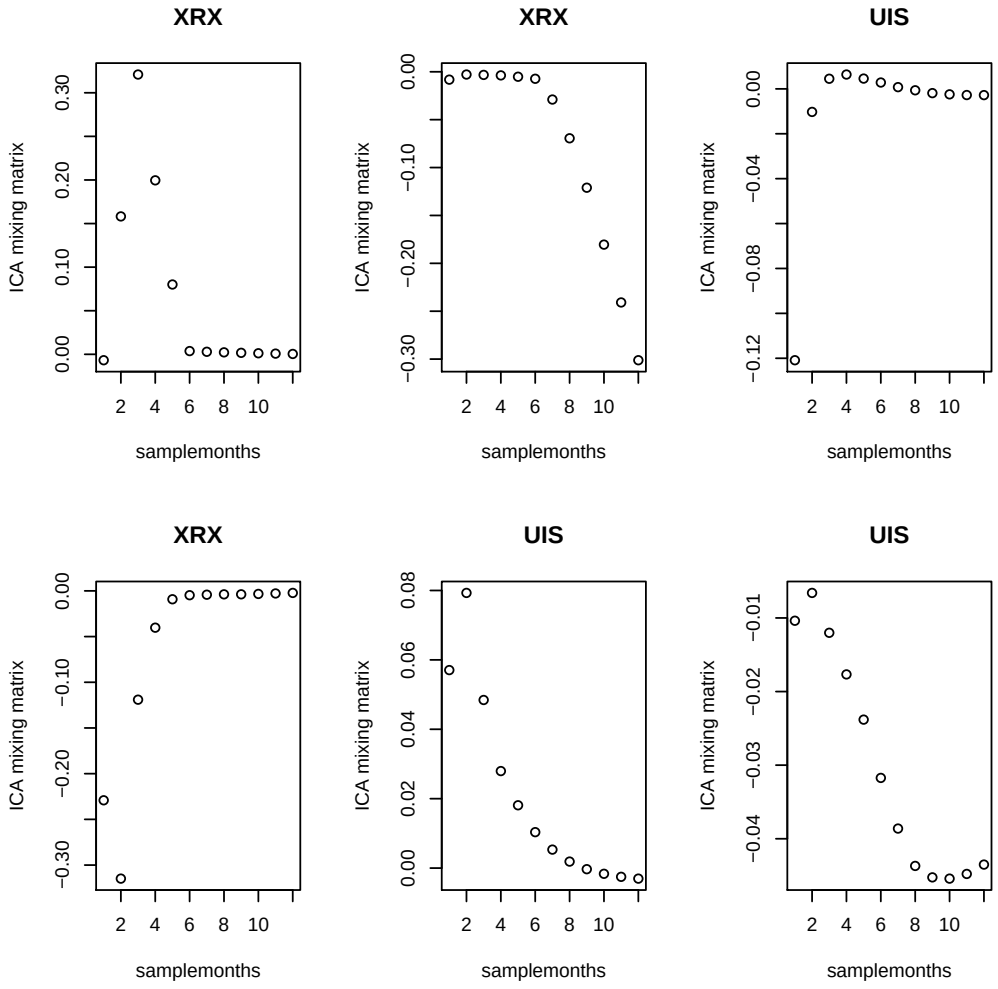

Figure 19: In the order of from top to bottom and then from left to right, the plot shows the first three vectors of ICA analysis.

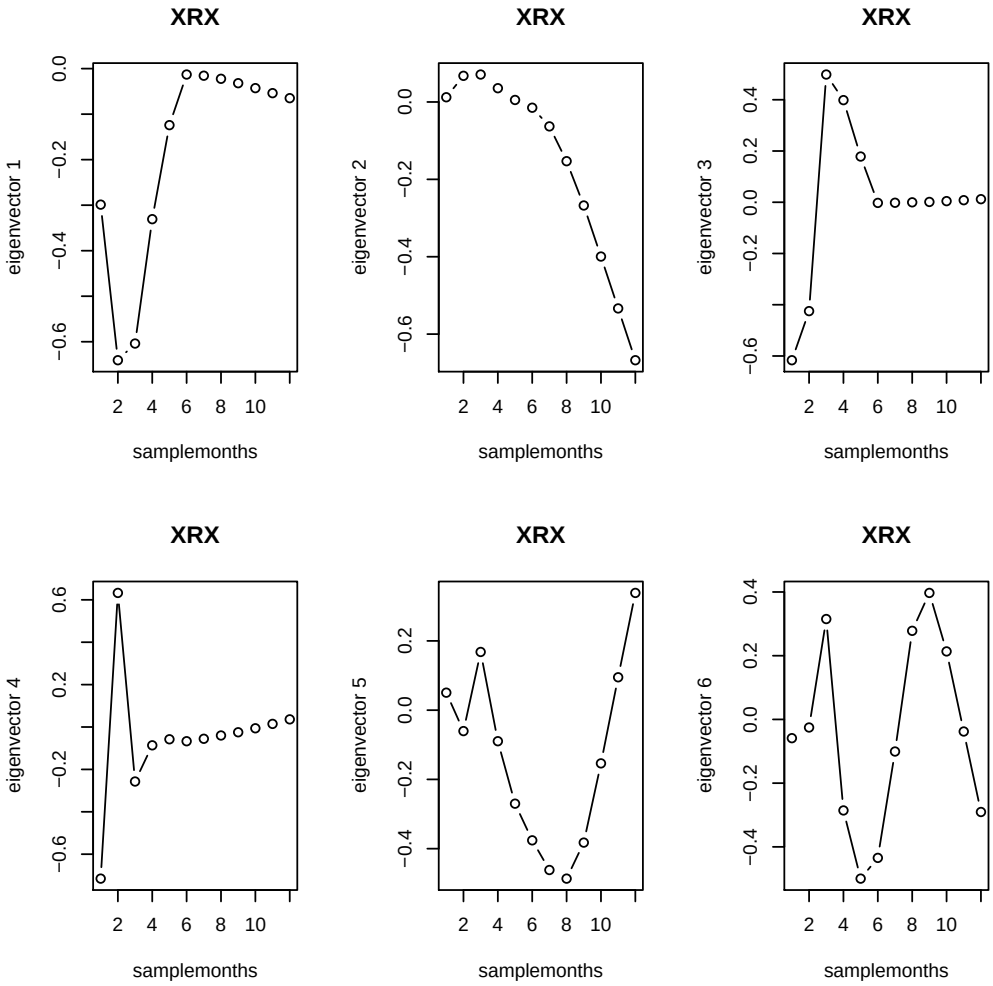

Figure 20: In the order of from left to right and then from top to bottom, the plot shows the first six eigenvectors of PCA analysis.

6 A Multi-Factor Implied Volatility Term-Structure Model

Introduction We compare two related approaches for modeling the *reduced* dynamics of the implied volatility term-structure. In each case we choose to fix the number of factors n of the model. This number is determined from the PCA of the historical implied volatility times series and corresponds to the number of sources of uncorrelated randomness. The dynamics are reduced when the number of factors is less than the number of maturity dates N.

The key idea uses the property that the uncorrelated n^{th} principal component corresponds to a change in the $(n-1)^{th}$ derivative of the implied volatility term-structure. So, in principal, the Taylor series truncation of the implied volatilities corresponds to the principal component series of the log returns of the implied volatility. The overall objective of this section is to assess the extent to which the PCA model can be fitted in comparison with the simpler Taylor series approach.

Example Suppose that only the first two principal components ($n = 2$) of the covariance matrix of the log returns are non-zero and that the log returns are normally distributed. Then, the dynamics of the implied volatility term structure are driven by a parallel shift and a change of slope.

Our first model follows [Cont, R and da Fonseca, J] in so far as it is based on projecting the log returns onto the principal components. Each projected component is then modeled as an Ornstein-Uhlenbeck stochastic process with uncorrelated Gaussian noise.

The second model is a simple alternative to this approach and is based on a Taylor series expansion of the term-structure about the mid-point of the term-structure $T/2$. It is primarily intended as a conceptually simpler version of model (1) for the case when $n \leq 3$. We emphasize that this second model does *not* use the principal components themselves, only the number n of principal components with non-neglible variance.

Definition 6.0.1 (Reduced Model (1): Term-Structure of Principal Components) *The estimated term-structure of implied volatility log returns $T_j \to x_j(t_k)$ at time t_k is given by*

$$x_j(t_k) := x(t_k, \tau_j) := ln \frac{\hat{\sigma}(t_k, \tau_j)}{\hat{\sigma}(t_{k-1}, \tau_j)} = \sum_{i=0}^{n-1} \hat{x}_i(t_k) F_{ij}, \qquad \hat{x}_i(t_k) = x_j(t_k) F_{ji} \tag{15}$$

or in vector form

$$\mathbf{x}(t_k) = \hat{\mathbf{x}}(t_k) F, \qquad \hat{\mathbf{x}}(t_k) = \mathbf{x}(t_k) F^T, \tag{16}$$

where F is a $N \times n$ orthonormal matrix whose columns represent the principal component vectors, $\mathbf{x}(t_k)$ is a N-vector representing the term-structure of log returns at time t_k and $\hat{\mathbf{x}}(t_k)$ is an n-vector of log returns in the principal component basis.

Definition 6.0.2 (Reduced Model (2): Taylor Series Implied Volatility Term-Structure) *The estimated implied volatility term-structure $T_j \rightarrow \hat{\sigma}(t, T_j)$ at time t_k is given by the Taylor series expansion*

$$ln\hat{\sigma}(t_k, \tau_j) := \sum_{i=0}^{n-1} \frac{\partial^i}{\partial \tau^i}[ln\sigma(t_k, T/2)] \frac{(\tau_j - T/2)^i}{(i+1)!} \tag{17}$$

about $T/2$ where $\sigma(t, T/2)$ is the historical mid-point implied volatility.

Analogously to interest rate modeling, specific functional forms of $\sigma(t_k, \tau)$ may be chosen apriori. For example, under a Vasicek type formulation, the implied volatility is

$$\sigma(t_k, \tau) = \bar{\sigma}(t) exp\{-\beta(t)(\tau - T/2)\}, \tag{18}$$

where $n = 2$ and $\frac{\partial}{\partial \tau} ln\sigma(t_k, \tau) = -\beta(t)$.

6.1 GARCH Models

Labelling the i^{th} component in each reduced model of the log return implied volatility at time t_k as $x_k^{[i]}$ and its conditional variance $v_k^{[i]} := var(x_k^{[i]} | x_{k-1}^{[i]})$, a GARCH(1,1) model of the form

$$x_k^{[i]} = \mu^{[i]} + \epsilon_k^{[i]}, \qquad v_k^{[i]} = \alpha_0^{[i]} + \alpha_1^{[i]} (\epsilon_k^{[i]})^2 + \beta_0^{[i]} v_{k-1}^{[i]}, i := 1 \rightarrow n, \tag{19}$$

is found to fit the historical time series under the assumption of normalility of the distribution of the log returns. This assumption is imposed only for simplicity and is not intrinsic to the auto-regressive model.

Model (2) example For the case that $n = 2$

$$ln\hat{\sigma}(t, \tau) = ln\sigma(t, T/2) + (\tau - T/2)\Delta ln\sigma(t, T/2). \tag{20}$$

For robustness, we replace the first term in the above equation with $ln\bar{\sigma}(t)$ so that it is more representative of parallel shifts to the entire term-structure. This of course introduces some error as the mid-point will not generally coincide with the mean. Using more explicit notation, the first and second components of the log returns are

$$\bar{x}_k := ln\left(\frac{\bar{\sigma}_{k+1}}{\bar{\sigma}_k}\right) \tag{21}$$

and

$$x_{\Delta,k} := ln\left(\frac{\Delta\sigma_{k+1}}{\Delta\sigma_k}\right), \tag{22}$$

with respective conditional variances $\bar{v}_k := var(\bar{x}_k | \bar{x}_{k-1})$ and $v_{\Delta,k} := var(x_{\Delta,k} | x_{\Delta,k-1})$. The two Garch(1,1) models for the mean and slope are

$$\begin{aligned} \bar{x}_k &= \bar{\mu} + \bar{\epsilon}_k, \qquad & \bar{v}_k &= \bar{\alpha}_0 + \bar{\alpha}_1 \bar{\epsilon}_k^2 + \bar{\beta}_0 \bar{v}_{k-1}, \\ x_{\Delta,k} &= \mu_\Delta + \epsilon_{\Delta,k}, \qquad & v_{\Delta,k} &= \alpha_{\Delta,0} + \alpha_{\Delta,1} \epsilon_{\Delta,k}^2 + \beta_{\Delta,0} v_{\Delta,k-1}. \end{aligned} \tag{23}$$

6.2 Mean reverting processes

The GARCH(1,1) model provides a predictive model of the log returns in discrete time. For completeness, we briefly comment that in the continuous time limit, each of the GARCH(1,1) sequences define a mean-reverting stochastic process for the conditional variance $v^{[i]}(t)$ of the form

$$dv^{[i]}(t) = \gamma^{[i]}(\bar{v}^{[i]} - v^{[i]}(t))dt + \alpha_1^{[i]} 2^{\frac{1}{2}} v^{[i]}(t) dZ^{[i]}(t), \tag{24}$$

where the mean reversion rate $\gamma^{[i]} = 1 - \alpha_1^{[i]} - \beta_0^{[i]}$ and the equilibrium level $\bar{v}^{[i]} = \frac{\alpha_0^{[i]}}{\gamma^{[i]}}$.

Model (1) The mean-reversion process for the variance of the log returns takes the form

$$dv(t, \tau_j) = \sum_{i=0}^{n-1} [\gamma^{[i]}(\bar{v}^{[i]} - v^{[i]}(t))dt + \alpha_1^{[i]} 2^{\frac{1}{2}} v^{[i]}(t) dZ^{[i]}(t)] F_{ij}. \tag{25}$$

Model (2) Similarly, the variance process is given by

$$dv(t, \tau_j) = \sum_{i=0}^{n-1} [\gamma^{[i]}(\bar{v}^{[i]} - v^{[i]}(t))dt + \alpha_1^{[i]} 2^{\frac{1}{2}} v^{[i]}(t) dZ^{[i]}(t)] \frac{(\tau_j - T/2)^i}{(i+1)!}. \tag{26}$$

In the next section, we compare each reduced model by fitting the GARCH sequences to the log returns of the implied volatility data from the $q - \alpha - \sigma$ model.

7 Results

The results in this section were obtained using `R` on the Xerox time series, provided in the VIX data of the $q-\alpha-\sigma$ volatilities. This ticker was chosen because it had almost 2000 days of continguous data over almost all of the maturities and required minimal interpolation to obtain missing implied volatilties.

7.1 Experiment 1

GARCH models were fitted to the projected log return time series on each of the first ten principal components. GARCH(1,1) models were found to provide the best fit by trial and error under the assumption of normality of the log returns. Table 1, reports the estimates and standard errors in the parameters for the first two principal components. Figure 1 also

shows the mean reversion rates and equilibria of the first ten components. As expected, the equilibria conditional variances generally decrease with the order of the principal component. The mean reversion rates, show less of a distinct trend in the order of the principal component and indicate a very low rate of mean reversion in the first few components.

Figure 2 compares the conditional variances of the fitted GARCH(1,1) models in the first four principal components with rolling 30-day variance estimations on each projected time series. The GARCH(1,1) models tend to consistently underestimate the variance compared to the rolling 30-day variance estimations of the projected time series in each of the four principal components, but capture the overall heteroskedacity well.

Figure 3 shows the normalized log-likelihoods of the fitted GARCH(1,1) series in the first ten principal components which are observed to decrease with the order of the component. The normalized error in the fitted GARCH(1,1) model decreases with the order of the principal component. This is to be expected as the higher order principal components fluctuate the least and are easier to fit a GARCH(1,1) model too.

	$\hat{x}_1(t_k)$		$\hat{x}_2(t_k)$	
	Estimate	Std. Error	Estimate	Std. Error
μ	1.88×10^{-3}	1.72×10^{-2}	-3.83×10^{-5}	2.80×10^{-3}
α_0	9.85×10^{-2}	6.14×10^{-3}	1.14×10^{-4}	1.45×10^{-5}
α_1	1.75×10^{-1}	2.43×10^{-2}	2.67×10^{-2}	1.54×10^{-3}
β_0	8.26×10^{-1}	9.39×10^{-3}	9.69×10^{-1}	8.36×10^{-4}

Table 1: (Model 1:) This table shows the estimates and standard errors of the GARCH(1,1) parameters for the log returns of the first (left) and second (right) principal component of the implied volatility term-structure.

7.2 Experiment 2

GARCH(1,1) models were fitted to the series for the log returns of the mean and slope of the implied volatility term-structure, under the assumption of normality of the log returns. Table 2, reports the estimates and standard errors in the parameters of each GARCH model. These values correpond to mean reversion rates $\gamma v = -0.3074$ and $\gamma_\Delta = -0.0050$ and the equilibrium values are $\bar{v} = -2.7668e-04$ and $\bar{v}_\Delta = -11.4200$.

Figure 4 compares the projected time series of the log returns in the first two principal components with the time series of the log returns of the mean and slope of the implied volatility term-structure. We observe that the behaviour of the first component is similar to that of the mean up to a scaling factor. The behaviour of the second component and the slope is less comparable.

This observation is also supported by the measures of fit. The normalized log-likelihoods for the fitted GARCH (1,1) series of the mean and slopes of the log volatility term-structure are 1.336021 and 1.072633. The normalized log-likelihood of the fitted GARCH (1,1) series

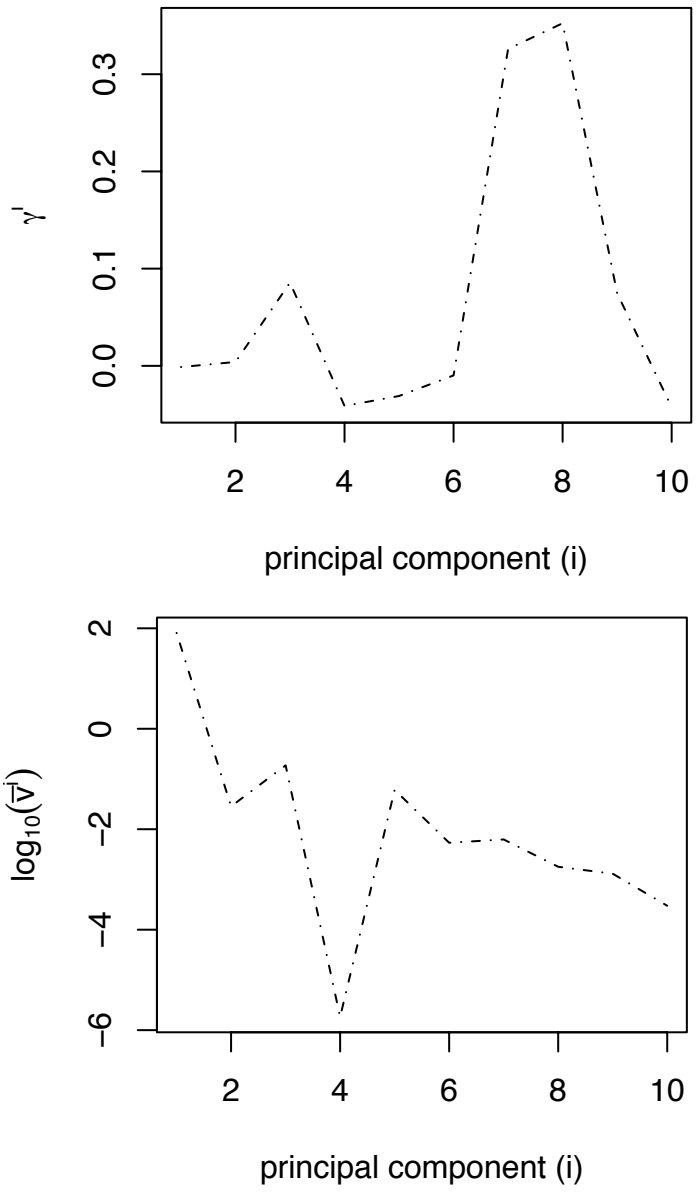

Figure 21: (Model 1:) These graphs show the estimated mean reversion rate (left) and equilibria (right) in the GARCH(1,1) models of each of the first ten principal components of the implied volatility log returns. The equilibria are observed to generally decrease with the order of the principal components, but the mean reversion rate doesn't exhibit a distinct trend other than that it is generally greater in the higher order principal components.

of the first principal component has a comparable value of 1.357841. However, the second principal component has a considerably smaller value of -0.04277524, indicating that it fits

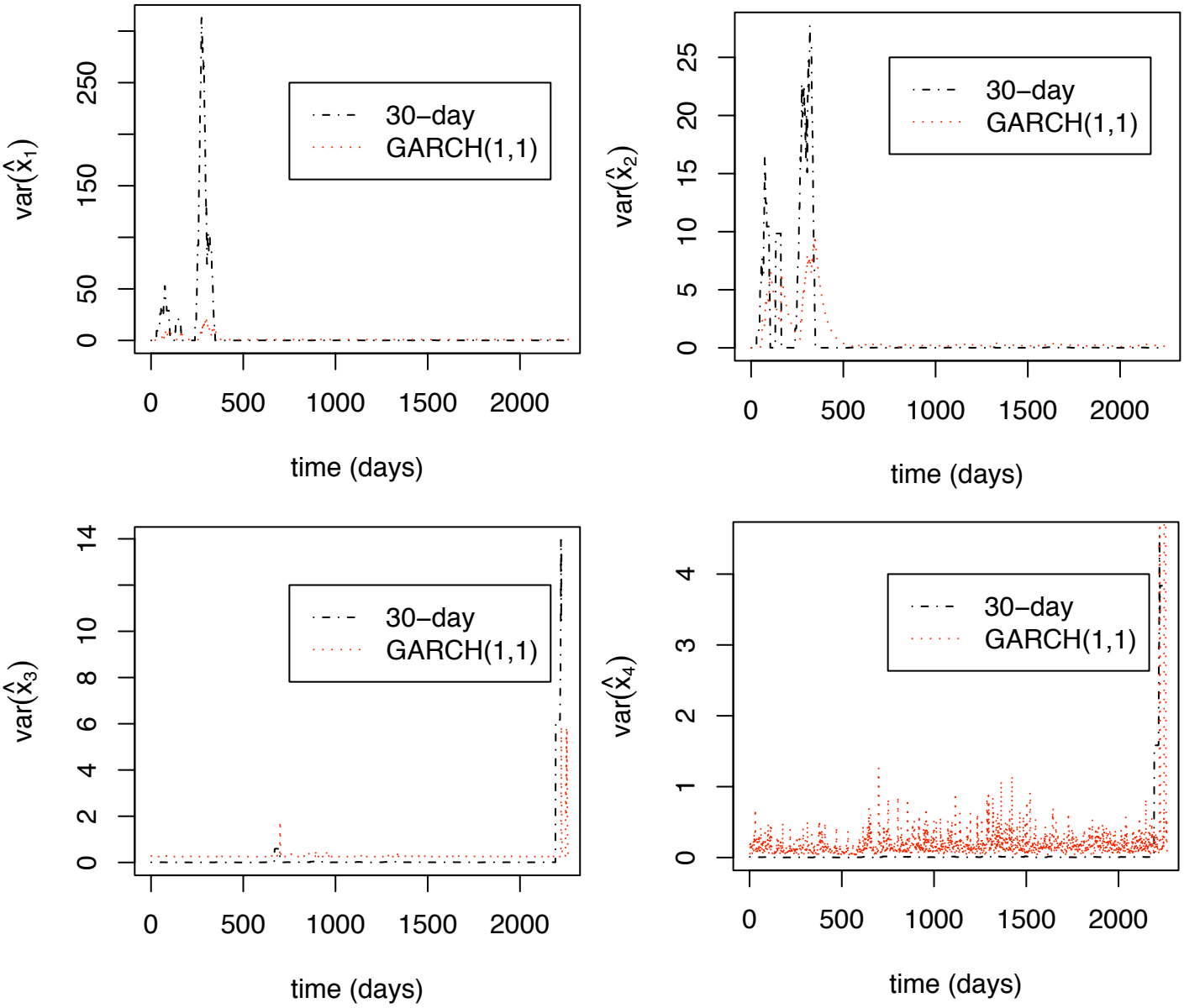

Figure 22: (Model 1:) These graphs compare the conditional variance of the fitted GARCH(1,1) series of log returns with the rolling 30-day variance estimation of the data. Each graph corresponds to one of the first four principal components.

the data much better than the slope.

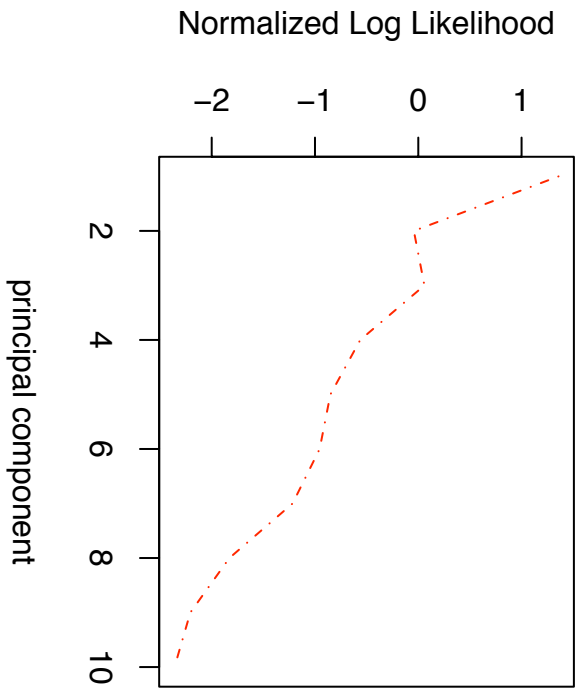

Figure 23: (Model 1:) This graph shows the normalized log-likelihood of the first ten fitted principal component GARCH(1,1) series. The normalized log-likelihood provides a measure of fit of the series to the data. We observe that the log-likehoods decrease with the order of the principal components.

	$\bar{x}_k$		$\bar{x}_{\Delta,k}$	
	Estimate	Std. Error	Estimate	Std. Error
μ	-3.60×10^{-4}	9.20×10^{-4}	-3.56×10^{-3}	1.18×10^{-2}
α_0	8.51×10^{-5}	2.28×10^{-5}	5.71×10^{-2}	8.54×10^{-3}
α_1	4.98×10^{-1}	4.38×10^{-2}	3.54×10^{-1}	4.52×10^{-2}
β_0	8.07×10^{-1}	6.49×10^{-3}	6.51×10^{-1}	3.20×10^{-2}

Table 2: (Model 2:) This table shows the estimates and standard errors of the GARCH(1,1) parameters for the log returns of the mean (left) and slope (right) of the implied volatility term-structure.

7.3 Summary

In this section, a GARCH(1,1) model is fitted to the components of the reduced dynamics of the implied volatility log returns. This model

- Assumes stationarity of the time series of the log returns of the implied volatility

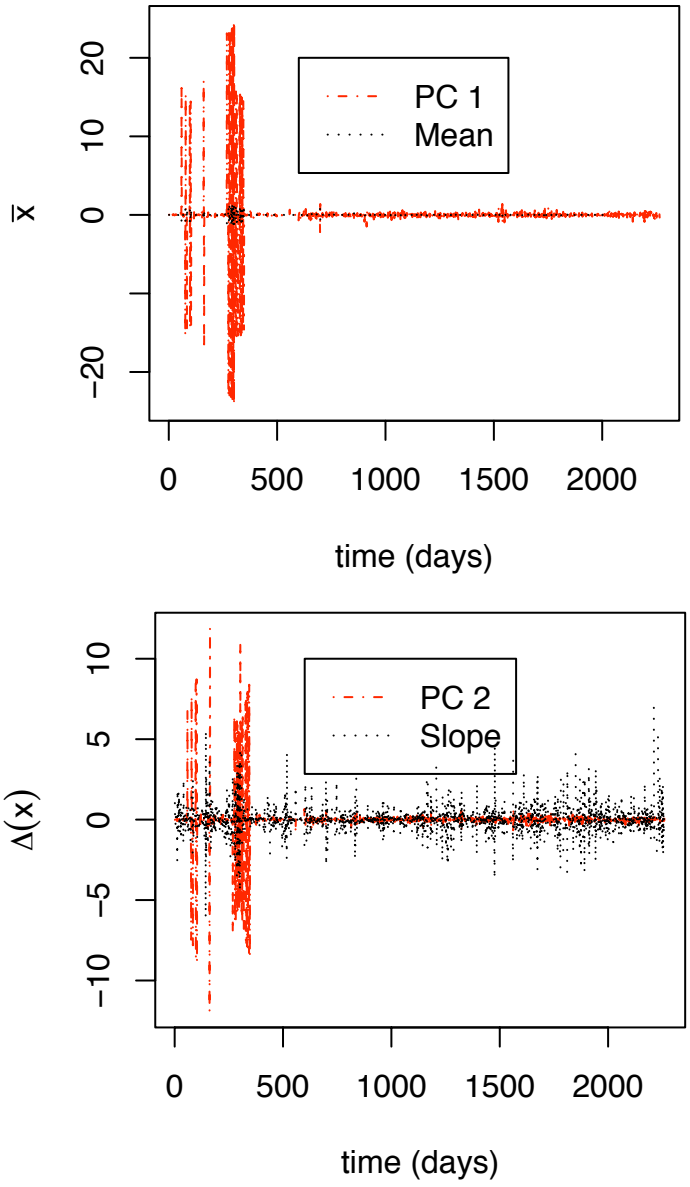

Figure 24: (Left) This graph compares the series of the log returns of the mean of implied volatility term-structure with the series of the first principal component of the log returns. (Right) This graph compares the series of the log returns of the slope of implied volatility term-structure with the series of the second principal component of the log returns.

- Captures observed heteroskedacity
- Decomposes dynamics into those attributed to time-to-maturity independent effects on the implied volatility of the option contract and those which are time-to-maturity

dependent.

The overall objective of this section was to see how well PCA performs against a simpler and, in principle, equivalent approach for reducing the sources of randomness. This simpler approach is based on modeling the mean and the slope of the implied volatility term-structure.

Our results suggest that whilst the mean may be a useful proxy, comparable with the first principal component, the slope deviates considerably from the second principal component and does not fit the data as well either. In fact, the PCA is consistently convergent in its order - in other words, the GARCH model fits the projected time series on higher order components and in general exhibits both less variance and a stronger rate of mean reversion.

8 Vega Hedging

8.1 Basic Idea

Fluctuations of the volatility surface contribute to changes in option price in much the same way as do fluctuations of the stock price or interest rates. The associated risk is known as *vega risk*. In order to hedge against this vega risk one needs a model for the fluctuations. Once a model has been established it is possible to compute a portfolio's exposure to those fluctuations and buy securities to offset the risk.

We will take the perspective of a trader who trades call options on a single underlying stock. We would like to hedge our vega risk as cheaply as possible while preserving certain desirable features of our portfolio, for example delta neutrality. It is possible to extend our analysis to portfolios that contain options on multiple stocks, but we will avoid this further complication.

Our hedging strategy is built upon the PCA analysis of Section 4. We assume that the first two or three PCA eigenvectors dominate the fluctuations of the volatility surface, and we purchase options to cancel the associated risk. For simplicity we will concentrate only on the first PCA eigenvector and ignore issues such as delta neutrality. Later we will discuss how to extend the hedging method to more eigenvectors and delta-neutral portfolios.

Consider a portfolio that contains a number Π_m of at-the-money call options with values C_m, where m denotes the maturity. Later we will extend our analysis to portfolios with out-of-the-money options. The total value Π of the portfolio is then

$$\Pi = \sum_m \Pi_m C_m. \tag{27}$$

In order to understand our vega risk we must compute the exposure of our portfolio to the dominant PCA eigenvector E_m. That is, we wish to know the change in value $d\Pi$ that our portfolio experiences with a typical shift in E. This is easily obtained through differentials:

$$d\Pi = \sum_m \Pi_m dC_m = \sum_m \Pi_m \frac{\partial C_m}{\partial \sigma_m} \frac{\partial \sigma_m}{\partial \log \sigma_m} d\log \sigma_m = \sum_m \Pi_m \text{vega}_m \sigma_m E_m \cdot \alpha \tag{28}$$

Here α is a factor needed to account for the normalization of the PCA eigenvector E_m. On any given day the shift in the volatility surface has projections onto each of the PCA eigenvectors. We take α to be the average size of the projections onto the first eigenvector. In the notation of Equation 9, it is $\mathbb{E}[\vec{X} \cdot \vec{e}(1)]$.

To hedge this risk one simply needs to purchase appropriate options to cancel the shift $d\Pi$. To minimize the cost, one can compare the values of dC_m/C_m for each available option. The option with the highest ratio is most cost effective for eliminating vega risk. The hedge is now easily accomplished by purchasing a quantity $-d\Pi/dC_m$ of the most cost-effective options.

8.2 Implementation and Analysis

We implemented the above hedging strategy for random portfolios of at-the-money (ATM) options on the nine stocks we studied. These were then tested against the time series of option prices available in the data set. We found that our strategy was inadequate for hedging against changes in volatility. To understand why this is the case, one must get a feel for the kinds of numbers involved. For simplicity we shall focus on portfolios of XRX stock throughout this discussion.

XRX had options from six maturity series available for purchase. Our portfolios contained random quantities of options from each series, drawn from a normal distribution with mean 0 and standard deviation 100. The portfolios had mean absolute value approximately \$2000. On average our exposure to the first PCA eigenvector, as predicted by the model in Equation (28), was about \$200 or 10% of the portfolio value. Thus we were attempting to hedge against changes on the order of \$200.

The portfolios contain hundreds of options and so, barring any accidental cancellations, the exposure $d\Pi$ of the portfolio is generally a hundred times the exposure dC_m of each option separately. Thus we generally had to purchase approximately a hundred options to cancel our vega risk. As the options typically cost about \$10, this means that we were spending on the order of \$1000 to hedge our risk. In fact, we spent on average 35% of our portfolio value on hedging.[2]

One might wonder whether, despite the high cost of initially constructing a hedge, it is relatively inexpensive to maintain that hedge over time. A portfolio that is initially hedged will, over time, become less well hedged as market conditions change. We calculated how expensive it would be to maintain a portfolio hedge on a daily basis. In fact each day we would spend on average 8% of our portfolio maintaining the hedge.

The above calculations were performed in the context of the model (28), which assumed that exposure to the first eigenvector dominated the portfolio vega risk. It is interesting to investigate this hypothesis further. In the language of Equations (9) and (10) we would like to know how well portfolio exposure to $\vec{X}^1(t)$ approximates exposure to $\vec{X}(t)$. To

[2]All averages in this section are average magnitudes, where positive and negative numbers contribute with the same sign. This is because we are interested in the size of the overall fluctuations. Including the signs would lead to deceptively small numbers.

do this we may calculate the portfolio exposure to the successively representative shifts $\vec{X}^1(t), \vec{X}^2(t), \ldots, \vec{X}^6(t)$ and see how that exposure progresses with increasingly complete projections.[3]

This calculation can be seen in Figure (25) for two specific dates. The behavior of the portfolio exposure is in stark contrast to Figure (9). It is clear that all of the PCA eigenvectors contribute significantly to the portfolio's vega exposure, rather than quickly converging to a final value. In particular, the fluctuations in the portfolio exposure are on the same order as the exposure itself. We will discuss this in greater depth in the next section, but for the moment we are content to note that this means our model will not hedge shifts in the volatility surface effectively unless it hedges against shifts of the other PCA eigenvectors, as well.

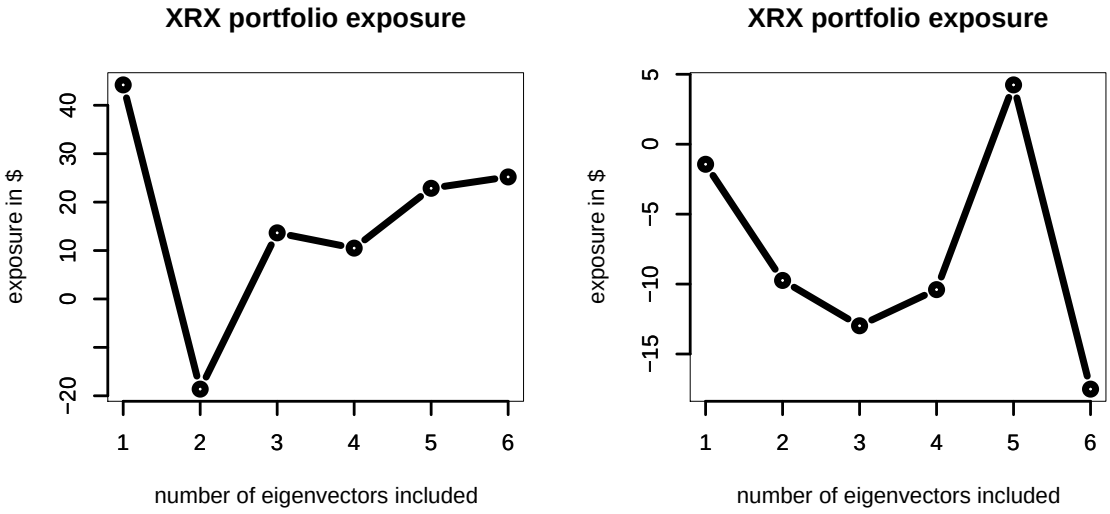

Figure 25: Portfolio exposure on two specific dates, using increasingly many PCA eigenvectors to capture the true interday volatility shift. The exposure does not asymptote.

8.3 Discussion

We have seen that our simple vega hedging strategy is expensive to initiate and to maintain. Given the high cost of hedging relative to the predicted exposure, we cannot recommend that this strategy be implemented in the real world. However, it is interesting to speculate as to why the strategy, which seems to straightforward, fails so miserably. This is linked to

[3]Of course, the portfolio exposure is still a model-dependent quantity. But exposure to $\vec{X}^6(t)$ represents exposure to the true shift in the volatility surface, whereas exposure to $\vec{X}^1(t)$ represents exposure to an additionally-smodel-dependent projection.

the question raised in Figure 25: namely, why doesn't the first eigenvector capture most of the portfolio exposure?

We believe that this results from two contributing effects. First, the PCA eigenvectors, as we saw in Figures 6 and 7 are peaked. This means that the various maturities behave rather independently. If we focus on portfolio exposure to a specific eigenvector then our hands are tied when we try to choose which option would be most cost effective at hedging the risk. It limits our ability to select cheap options with high exposure.

Secondly, and perhaps more importantly, the portfolio exposure is determined by more than just the shape of the dominant eigenvector and its relative importance over the others in explaining fluctuations of the volatility surface. Rather, from Equation (28), we see that many other factors contribute, as well: the specific option holdings of the portfolio, the various option vegas, and even the option volatilities. All of these factors are convoluted together with the eigenvector to determine the portfolio exposure. Apparently, when all of these factors are included, all six of the PCA eigenvectors becomes equally important.

These observations point us to a potential resolution, which ties together with our observations in Section 4.4 about sampling. Since we are interested in portfolio exposure, where all of these market factors are convoluted together with fluctuations of the volatility surface, then we should study fluctuations of the convoluted surface, itself. That is, we should do a PCA analysis on the surface of option volatility exposures.

We could take this a step further and envision a PCA analysis tailored to the specific option portfolio that one is trying to hedge. The sampling of the term structure would be weighted by the distribution of options in the portfolio. This is clearly a direction for future study.

8.4 Vega Appendix: Extensions

In our discussion of vega hedging we considered portfolios that contain at-the-money options only. However, it is straightforward to extend our discussion to portfolios that contain options of all strikes and maturities. The implied volatility surface derived from the q-alpha-sigma model is relatively flat in the direction of strike, which is why we were able to study the term structure alone. The PCA eigenvectors that represent shifts in this term structure implicitly extend flat into the strike dimension. The expression (28) for the portfolio and option exposure is then simply extended to include out-of-the-money strikes by summing over an additional index.

We also ignored the question of how to hedge additional PCA eigenvectors or preserve a portfolio statistic, such as delta neutrality. This is easily accomplished by using additional options to hedge the additional sources of risk. One simply purchases appropriate quantities of n different options, where n is the number of sources of risk plus the number of constraints. The quantities can be chosen so as to hedge the risks and satisfy the constraints. Furthermore, this linear combination can be optimized to minimize price or satisfy some other requirement.

9 Conclusions and Future Directions

The q-alpha-sigma model offers a promising alternative to the Black-Scholes as a method for pricing options. This model captures the volatility smile and so reduces the implied volatility surface to a term structure alone. We studied the statistics and dynamics of this volatility term structure using principal component analysis and a GARCH model and found both techniques to be successful. We implemented a simple vega hedging strategy, using knowledge gleaned from the PCA analysis, but found that the method employed was ineffective. We have indicated how one could perhaps extend our strategy to develop one that is more successful.

10 Acknowledgements

We would like to thank the teaching staff Kay Giesecke and Ben Armbruster. Ben in particular was incredibly helpful and gave us a lot of his time. We would also like to thank the people at EvA: Christian Silva, Lisa Borland and Jeremy Evnine. They were all extremely generous with their help at various stages of the project and also allowed us to use their codes to compute q-alpha-sigma prices.

References

[Ane & Labidi (2001)] Ane, T. & Labidi, C. 2001, Journal of Economics and Finance, 25, 3

[Borland, PRL (2002)] Borland, L. 2002, Phys. Rev. Lett. 89, 098701

[Borland L. and Bouchard, J.P. (2004)] Borland, L. & Bouchard, J.P., arXiv:cond-mat/0403022v2

[Borland, Quant. Fin. (2002)] Borland, L., Quant. Fin. Vol 2 (2002) 415

[Cont, R and da Fonseca, J] Cont, R & da Fonseca, J, Quant. Fin. Vol 2 (2002) 45

[Wu & Yu (2005)] Wu, E.H.C. & Yu, P.L.H. 2005, IDEAL 2005, LNCS 3578, 571