PRINCIPAL COMPONENT ANALYSIS

GRAEME WEST, FINANCIAL MODELLING AGENCY

Principal component analysis is a vector space transform used to reduce multidimensional data sets to lower dimensions, in an attempt to make the analysis of the correlation or covariance structure easier. PCA involves the calculation of the eigenvalue decomposition of the correlation or covariance matrix of a data set.
Here we try to give a fairly complete explanation of the mathematics involved in this procedure. We then give some applications. The one theme of applications is using PCA as a summary tool and scenario generation tool. The other theme of applications is to use PCA in portfolio selection.

1. The expected return and risk of a portfolio of assets

Suppose we have a portfolio with n assets, the i^{th} of which delivers a return $R_{t,i}$ at time t. The number of asset i that we have in the portfolio will be denoted W_i.

1.1. What is return?

Definition 1.1. The simple return on a financial instrument $P_.$ is $R_t = \frac{P_t - P_{t-1}}{P_{t-1}}$.

This definition has a number of caveats:

- The time from $t-1$ to t is one business day. Thus it is the daily return. We could also be interested in monthly returns, annual returns, etc.
- P_t is a price. Sometimes a conversion needs to be made to the raw data in order to achieve this. For example, if we start with bond yields y_t, it doesn't make much sense to focus on the return as formulated above. Why?
- We need to worry about other income sources, such as dividends, coupons, etc.
- The continuous return is $\ln \frac{P_t}{P_{t-1}}$. This has better mathematical and modelling properties than the simple return above. For example, it is what occurs in all financial modelling and (hence) is what we use for calibrating a historical volatility calculator.

Let

$$w_{t,i} = \frac{P_{t,i} W_{t,i}}{\sum_{i \in I} P_{t,i} W_{t,i}} \tag{1}$$

be the proportion of the wealth of the portfolio at time t that is comprised by stock i. Note that $\sum_{i=1}^{n} w_{t,i} = 1$. Negative weights are possible because of short sales. We close our eyes to the possibility that the denominator might be (close to) 0, which could occur in the setting of hedge funds, for example.

Date: August 25, 2009.
Thanks to Paulius Jakubenas, Johnny Whitamore and Dennis Stoller for useful suggestions on this document.

In order to calculate performance measures of the portfolio as a function of the performance measures of the stocks, we proceed as follows:

$$\begin{aligned}
R_{t,P} &= \frac{\sum_{i=1}^n W_i P_{t,i} - \sum_{i=1}^n W_i P_{t-1,i}}{\sum_{j=1}^n W_j P_{t-1,j}} \\
&= \frac{\sum_{i=1}^n W_i (P_{t,i} - P_{t-1,i})}{\sum_{j=1}^n W_j P_{t-1,j}} \\
&= \frac{\sum_{i=1}^n W_i P_{t-1,i} \frac{P_{t,i} - P_{t-1,i}}{P_{t-1,i}}}{\sum_{j=1}^n W_j P_{t-1,j}} \\
&= \frac{\sum_{i=1}^n W_i P_{t-1,i} R_{t,i}}{\sum_{j=1}^n W_j P_{t-1,j}} \\
&= \sum_{i=1}^n \left[\frac{W_i P_{t-1,i}}{\sum_{j=1}^n W_j P_{t-1,j}} \right] R_{t,i} \\
&= \sum_{i=1}^n w_{t-1,i} R_{t,i}
\end{aligned}$$

These are simple daily returns. A similar calculation involving continuous returns would not work because the log of a sum is not the sum of the logs. Nevertheless, volatility calculations are usually made using this assumption, which is valid to first order. See [J.P.Morgan and Reuters, December 18, 1996, §4.1] for additional information.

1.2. **What is the expected return and variance?** Suppose we have a portfolio with n assets, the i^{th} of which delivers a return $R_{t,i}$ at time t. This return has a mean $\mu_{t,i}$ and a variance $\sigma^2_{t,i}$. Suppose the proportion of the value of the portfolio that asset i makes up is w_i (so $\sum_{i=1}^n w_i = 1$).

What is the mean and standard deviation of the return R of the portfolio? All known values are assumed to be known at time t, and the t will be implicit in what follows. We can suppress the subscript t as long as we understand that all of the parameters are dynamic and we need to refresh the estimates on a daily basis. Then

$$\mu := \mathbb{E}\left[R\right] = \mathbb{E}\left[\sum_{i=1}^n w_i R_i\right] = \sum_{i=1}^n w_i \mathbb{E}\left[R_i\right] = \sum_{i=1}^n w_i \mu_i \tag{2}$$

and

$$\begin{aligned}
\sigma^2(R) &= \mathbb{E}\left[(R-\mu)^2\right] \\
&= \mathbb{E}\left[(\sum_{i=1}^n w_i (R_i - \mu_i))^2\right] \\
&= \mathbb{E}\left[\sum_{i=1}^n \sum_{j=1}^n w_i w_j (R_i - \mu_i)(R_j - \mu_j)\right] \\
&= \sum_{i=1}^n \sum_{j=1}^n w_i w_j \mathrm{covar}(R_i, R_j) \\
&= \sum_{i=1}^n \sum_{j=1}^n w_i w_j \sigma_{i,j} \\
&= w' \Sigma w
\end{aligned}$$

where $w = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix}$ and $\Sigma = [\sigma_{i,j}] = \begin{bmatrix} \sigma_{11} & \cdots & \cdots & \cdots & \sigma_{1n} \\ \vdots & \ddots & & & \vdots \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ \sigma_{n1} & \cdots & \cdots & \cdots & \sigma_{nn} \end{bmatrix}$. This is called the covariance matrix. So, the return on the portfolio has

$$\begin{aligned} \mathbb{E}[R] &= w'\mu \\ \sigma(R) &= \sqrt{w'\Sigma w}. \end{aligned}$$

Note that

- σ_{ij} is the covariance between R_i the return on asset i and R_j the return on asset j.
- $\sigma_i^2 = \sigma_{ii}$ is the variance of R_i.
- $\rho_{ij} = \frac{\sigma_{ij}}{\sigma_i \sigma_j}$ is the correlation of R_i and R_j.

We will denote

- the covariance matrix by Σ;
- the correlation matrix $[\rho_{ij}] = \begin{bmatrix} \rho_{11} & \cdots & \cdots & \cdots & \rho_{1n} \\ \vdots & \ddots & & & \vdots \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ \rho_{n1} & \cdots & \cdots & \cdots & \rho_{nn} \end{bmatrix}$ by $\mathbb{P}$;
- the diagonal matrix of standard deviations $\begin{bmatrix} \sigma_1 & 0 & \cdots & \cdots & 0 \\ 0 & \sigma_2 & & & \vdots \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & \sigma_n \end{bmatrix}$ by S.

Then

$$\Sigma = S\mathbb{P}S \tag{3}$$

and so

$$\sigma(R) = \sqrt{w'S\mathbb{P}Sw} \tag{4}$$

2. Definitions and facts about matrices

All the matrices we consider are $n \times n$.

A matrix E is orthonormal if $EE' = I_n$. In this case E' is a right inverse of E and so is the inverse: $E^{-1} = E'$; $E'E = I_n = EE'$.

A matrix A for which $A = A'$ is said to be symmetric (we ignore the phenomenon of matrices with complex entries). Such a matrix can be expressed in the form $A = EDE'$ where E is an orthonormal matrix and D is

a diagonal matrix. This is equivalent to the matrix equation $AE = ED (\neq DE)$ which shows that the columns of E are the eigenvectors of A and the diagonal elements of D are the eigenvalues.
Let the eigenvectors be $e_1,\ e_2,\ \ldots,\ e_n$ and the corresponding eigenvalues (singular values) be $\lambda_1,\ \lambda_2,\ \ldots,\ \lambda_n$. Since the eigenvectors form an orthonormal basis for $\mathbb{R}^n$, we have that for any $x \in \mathbb{R}^n$, $x = \sum_{i=1}^n \langle x, e_i\rangle e_i$.
We show that $A = \sum_{i=1}^n \lambda_i e_i e_i'$: for any $x \in \mathbb{R}^n$ we have

$$\begin{aligned} Ax &= A\left(\sum_{i=1}^n \langle x, e_i\rangle e_i\right) \\ &= \sum_{i=1}^n \langle x, e_i\rangle A e_i \\ &= \sum_{i=1}^n \langle x, e_i\rangle \lambda_i e_i \\ &= \sum_{i=1}^n \lambda_i e_i e_i' x \end{aligned}$$

The eigenvalues are real. An interesting case is where the eigenvalues are all (strictly) positive. If they are all strictly positive, the matrix A is said to be positive definite; if the possibility that 0 occurs is not excluded then the matrix A is said to be positive semi-definite. A correlation (or covariance matrix) is theoretically positive definite, and in practice will be at least positive semi-definite.
We will, by permuting the indices, assume that the λ_i are decreasing. In a sense that we can make precise, this ordering is in order of importance: we can show that by putting $A_m = \sum_{i=1}^m \lambda_i e_i e_i'$, we have $\|A - A_m\|_\infty \leq \lambda_{m+1}$. So, suppose $x \in \mathbb{R}^n$, then

$$\begin{aligned} \|(A - A_m)x\|_2^2 &= \left\| \sum_{i=m+1}^n \lambda_i e_i e_i' x \right\|^2 \\ &= \left\| \sum_{i=m+1}^n \lambda_i e_i e_i' \left(\sum_{j=1}^n \langle x, e_j\rangle e_j \right) \right\|^2 \\ &= \left\| \sum_{i=m+1}^n \lambda_i e_i \langle x, e_i\rangle \right\|^2 \\ &\leq \sum_{i=m+1}^n \lambda_i^2 |\langle x, e_i\rangle|^2 \\ &\leq \lambda_{m+1}^2 \sum_{i=m+1}^n |\langle x, e_i\rangle|^2 \\ &\leq \lambda_{m+1}^2 \|x\|_2^2 \end{aligned}$$

by the Bessel inequality. Thus, if the λ_i get small sufficiently quickly, then the partial sum is a good approximation for A.

3. Analysis of the Correlation Matrix $\mathbb{P}$

Suppose $x_1,\ x_2,\ \ldots,\ x_n$ are variables with mean 0 and standard deviation 1, with correlation matrix (equivalently, covariance matrix) $\mathbb{P}$. If I form the combination variable $a'x = \sum_{i=1}^{n} a_i x_i$ then the same calculation as before shows that the variance of this variable is $a'\mathbb{P}a$.

Remember that e_i has an eigenvalue λ_i. Let us form the variable $e_i'x$. Then

$$\operatorname{var}(e_i'x) = e_i'\mathbb{P}e_i = e_i'EDE'e_i = \lambda_i$$

and this shows that $\lambda_i \geq 0$.

3.1. **Facts about the correlation matrix.** The covariance matrix is in theory positive definite as long as the variables are truly different i.e. we do not have the situation that one is a linear combination of the others (so that there is some combination which gives the 0 entry). If there are more assets in the matrix than number of historical data points the matrix will be rank-deficient and so only positive semi-definite. Moreover, in practice because all parameters are estimated, and in a large matrix there will be some assets which are nearly linear combinations of others, and also taking into account numerical roundoff, the matrix may not be positive semi-definite at all [Dowd, 1998, §2.3.4]. However, this problem has recently been completely solved Higham [2002], by mathematically finding the (semi-definite) correlation matrix which is closest (in an appropriate norm) to a given matrix, in particular, to our mis-estimated matrix.

3.2. **The principal components.** We want to find the a that gives us the maximal variance. As stated we could scale a up arbitrarily, so we impose the standard normalisation condition on a, that is, $\sum_{i=1}^{n} a_i^2 = 1$. First let us put $a = e_1$. (Recall that the eigenvalues have been arranged in decreasing order.) We have that $\operatorname{var}(e_1'x) = \lambda_1$.

If a is any vector with the stated normalisation condition, then

$$\begin{aligned}
\operatorname{var}(a'x) &= a'\mathbb{P}a \\
&= a'\sum_{i=1}^{n} \lambda_i e_i e_i' a \\
&= \sum_{i=1}^{n} \lambda_i a' e_i e_i' a \\
&= \sum_{i=1}^{n} \lambda_i |\langle a, e_i\rangle|^2 \\
&\leq \lambda_1 \sum_{i=1}^{n} |\langle a, e_i\rangle|^2 \\
&= \lambda_1 \|a\|_2^2 \\
&= \lambda_1
\end{aligned}$$

by Parseval's theorem.

It follows that the maximisation is achieved by putting $a = e_1$; the maximum is then λ_1.

Having done this, we want to find a second combination b so that $b \perp e_1$, $b'x$ is uncorrelated to $e_1'x$, has maximal variance amongst all possible such combinations, and has a similar size condition.

It should not surprise to find that $b = e_2$ will achieve this: as before, $e_2'\mathbb{P}e_2 = \lambda_2$, and

$$\operatorname{covar}(e_1 x, e_2 x) = e_1'\mathbb{P}e_2 = e_1'EDE'e_2 = 0$$

For the inequality, much as before,

$$\begin{aligned}
\operatorname{var}(b'x) &= b'\mathbb{P}b \\
&= b'\sum_{i=1}^{n}\lambda_i e_i e_i' b \\
&= b'\sum_{i=2}^{n}\lambda_i e_i e_i' b \\
&= \sum_{i=2}^{n}\lambda_i |\langle b, e_i\rangle|^2 \\
&\le \lambda_2 \sum_{i=2}^{n} |\langle b, e_i\rangle|^2 \\
&= \lambda_2 \|b\|_2^2 \\
&= \lambda_2
\end{aligned}$$

We proceed in this manner for $1,\ 2,\ \ldots,\ n$.

Let us see what we have achieved. The variables $x_1,\ x_2,\ \ldots,\ x_n$ have been replaced with variables $e_1'x$, $e_2'x$, $\ldots$, $e_n'x$ which have the following properties:

- The variables are uncorrelated.
- The variance of $e_i'x$ is equal to λ_i.

Note that from Lidskii's trace theorem,

$$n = \operatorname{trace}(A) = \lambda_1 + \lambda_2 + \cdots + \lambda_n \tag{5}$$

so each of these variables explains $\frac{\lambda_i}{n}$ of the total variability of the data, and the first m describe $\frac{1}{n}\sum_{i=1}^{m}\lambda_i$ of the data. Thus, we now come to a fundamental idea: choosing some small n we might get a large percentage of the variability of the data, and be safe in ignoring the residual. This is the dimensionality reduction we have sought.

4. Factor scores

Let a history of the standardised variables $x_1,\ x_2,\ \ldots,\ x_n$ for time $1 \le t \le T$ be represented by a $T \times n$ matrix, which we call X. Let $Z = XE$; these are the uncorrelated variables, with the variance of z_i being λ_i; that is, the covariance matrix of Z is equal to D.

Put $Y = XED^{-1/2}$, then the covariance of Y is I_n. Y is known as the **factor scores**. More (fairly arbitrary) terminology is that the columns of $ED^{1/2}$ are known as the **factor loadings**, while the columns of $F = ED^{-1/2}$ are known as the **factor score coefficients**.

We now discuss the inverse process. Note that $X = YD^{1/2}E'$.

We can imagine that it will be a valid procedure to discard the eigenvectors corresponding to the small eigenvalues, and hope that these are the less important ones. For example in the case of yield curve data, we typically retain only the first three principal components, which explain $\frac{\lambda_1+\lambda_2+\lambda_3}{n}$ of the variability of the data, which is close to the total amount. This corresponds to discarding all but the first three columns of Y and of E.

This 'discarding' work can be done entirely by modifying the matrix D. We introduce a diagonal matrix D_m which has diagonal elements $\lambda_1,\ \lambda_2,\ \ldots,\ \lambda_m$ and subsequently 0's, and a diagonal matrix $D_{>m}$ which are

exactly those discarded terms i.e. the matrix with m-many 0's on the diagonal and then $\lambda_{m+1}, \ldots, \lambda_n$. For what will follow we define I_m and $I_{>m}$ similarly: diagonal matrices with 1's and 0's on the diagonal.
We then define $X_m = Y D_m^{1/2} E'$. This is an approximate reconstruction of the actual X that occurred (and is $T \times n$). Note that in effect we have discarded all but the first m columns of Y.
How good is the approximation? For this, we find the covariance matrix for $X - X_m$, and verify that the entries are 'small'. Note that

$$X_m = XED^{-1/2} D_m^{1/2} E' = XEI_m E'$$

and so

$$X - X_m = XEI_{>m} E'$$

Thus

$$\begin{aligned} \operatorname{covar}(X - X_m) &= EI_{>m}E' \operatorname{covar}(X) EI_{>m}E' \\ &= EI_{>m}E' \mathbb{P} EI_{>m}E' \\ &= EI_{>m}DI_{>m}E' \\ &= ED_{>m}E' \end{aligned}$$

Since E is unitary, this shows that $\|\operatorname{covar}(X - X_m)\|_\infty \leq \lambda_{m+1}$.
Note that the covariance matrix of X_m is given by

$$\operatorname{covar}(X_m) = ED_m^{1/2} I D_m^{1/2} E' = ED_m E' = \sum_{i=1}^{m} \lambda_i e_i e_i' \tag{6}$$

In summary, we have the decomposition:

$$\operatorname{covar}(X) = EDE' \tag{7}$$

$$\operatorname{covar}(X_m) = ED_m E' \tag{8}$$

$$\operatorname{covar}(X - X_m) = ED_{>m}E' \tag{9}$$

5. Why do we aim for dimensionality reduction?

What has all this achieved?

(1) Rather than perform analysis and hypothesis testing on n variables, we perform it on m.
(2) The m variables can have a high degree of intuitive meaning, and can be used to estimate and calibrate statistical models, such as yield curve models. Typically in yield curve modelling three components used. The first component represents level changes, the second slope changes, and the third curvature changes. Here, one is using the spot rates of the yield curve, these clearly form a highly correlated system. If one used forward rates they would be less highly correlated and the PCA might be less satisfactory. As we will see, PCA also works better if the curve is a modelled curve (such as Nelson and Siegel [1987] or Svensson [1994]) rather than bootstrapped. See also Lekkos [2000].
(3) We can sample from m variables rather than n, and then artificially use those m variables to generate a sample of all n variables. This would be a structured Monte Carlo method, and will be materially faster than sampling for all n variables. The sampling is done from the (modelled or actual) distribution of Y_m.

For example, we can analyse the empirical distributions of $Y_1,\ Y_2,\ \ldots,\ Y_m$ and fit distributions to each of these, and then sample independently from each distribution. If we fail to fit a distribution, we might simply sample from the empirical distributions for random t. Either way, a very large set of data from which to sample has been bootstrapped.

As discussed in Jamshidian and Zhu [1996], one needs fewer and fewer samples from the Y_i as i increases. They give an example where they are performing a VaR calculation which involves movements in the yield curve. For this they choose 7 sample scores for the first component, 5 for the second component, and 3 for the third, with all higher components being discarded. A yield curve scenario is then any combination of these moves; since the components are by construction uncorrelated we have $7 \cdot 5 \cdot 3 = 105$ equiprobable scenarios which can be used in the calculation.

Suppose y_i are sampled from Y_i for $i = 1,\ 2,\ \ldots,\ m$. To infer the sample $x_1,\ x_2,\ \ldots,\ x_n$ we calculate the matrix product

$$\hat{x}_m = [(y_1,\ y_2,\ \ldots,\ y_m,\ 0,\ \ldots, 0)]\, D_m^{1/2} E'.$$

However, care needs to be taken here in the context of time series. If we consider our current principal component analysis, then the historical factor scores are indeed uncorrelated. However, in the past, the component decomposition was not static in time, and so the factor scores that would have been measured in the past would not be the same as if those moves were measured against today's components, because the components have changed. In fact, to measure past movements against today's components seems rather moot. Thus, what we should rather do, is measure the factor scores through time, against each component decomposition through time, and record those factor scores. We can then analyse the empirical distribution of factor scores and fit a parametric distribution to each one, normalising as necessary. Then, in the Monte Carlo experiments to follow, we can draw from those parametric distributions, randomly and independently for each one.

6. A synopsis of the process, and one last twist not considered so far

In reality, our n variables are not standardised. So, we have variables $v_1,\ v_2,\ \ldots,\ v_n$ with standard deviations $s_1,\ s_2,\ \ldots,\ s_n$. We remove any drift and standardise in the usual way to form variables $x_1,\ x_2,\ \ldots,\ x_n$, so $x_i = \frac{v_i - \overline{v}_i}{s_i}$. (In financial applications, in particular for yield curves, there is often no loss in supposing a priori that $\overline{v}_i = 0$. Often the reasoning is that there is mean reversion of interest rates, for example.)

Given our variables X we find the correlation (= covariance) matrix $\mathbb{P}$ and perform the PCA eigenvalue-eigenvector decomposition $\mathbb{P} = EDE'$. We then consider the variables X_m which have the covariance matrix ED_mE'.

Of course, this is not a correlation matrix. So, when we restore the variables by multiplying by the standard deviations, so $v_i^m = s_i x_i^m + \overline{v}_i$, we have a set of variables with covariance matrix $SED_mE'S$. In particular, remembering (6), we have that the variance of the i^{th} entry is

$$s_i^2 \sum_{j=1}^{m} \lambda_j e_{ij}^2$$

which is less than s_i^2. (Note that $\sum_{j=1}^{n} \lambda_j e_{ij}^2 = 1$.) Thus, in order to produce experiments with the correct variance, we should scale the experiments by the required correction factor. (This is obviously important in finance where we need the correct volatility in pricing derivatives.) Let C_m be the diagonal matrix which has

ii^{th} element $\sum_{j=1}^{m} \lambda_j e_{ij}^2$. Then the matrix $C_m^{-1/2} E D_m E' C_m^{-1/2}$ is indeed a correlation matrix and if we rather define

$$\hat{x}_m = [(y_1,\ y_2,\ \ldots,\ y_m,\ 0,\ \ldots, 0)]\, D_m^{1/2} E' C_m^{-1} \tag{10}$$

then $\hat{x}_m$ has the correct variance. Note that this matrix has entries

$$[\mathbb{P}_m]_{ij} = \sum_{k=1}^{m} \frac{1}{\sqrt{c_{ii} c_{jj}}} \lambda_k e_{ik} e_{jk} \tag{11}$$

This 'normalised principal components' method is conceptually attractive and easy to implement [Rebonato and Jäckel, 1999, §3], Brigo [2002]. However, and perhaps slightly surprisingly, this correlation matrix is not the closest correlation matrix of a given reduced rank to the original correlation matrix in the Frobenius norm. Alternative procedures which do find this matrix (and this is a much more difficult problem) are discussed in Rebonato and Jäckel [1999], Brigo [2002], Brigo et al. [2002], Weigel [2004], Morini and Webber [2004], Zhang and Wu [2003] and Wu [2002].[1]

7. Applications: data analysis

We now consider some applications of PCA. We will see a clear distinction in those cases where PCA is appropriate and those where it is not. PCA is designed for systems which have a high degree of built in correlation. Thus, it works well for futures curves, but not well for stock index constituents. The case of yield curves is more interesting: the results can go either way, depending upon the construction method.

7.1. **Analysis of yield curve evolution.**

7.1.1. *Modelled curves.* We consider GBP interbank curves.

Data available at `http://www.bankofengland.co.uk/statistics/yieldcurve/index.htm`.

This data consists of modelled zero rates. The Svensson procedure has been applied Svensson [1994], Svensson [1995].

First we consider the 1y, 2y through to 30y zero rates. The first three factors explained 98% of the variation. See Figure 1.

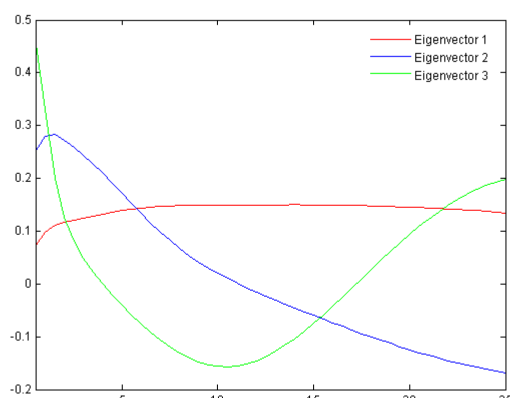

Figure 1. The first three rates components from modelled curve inputs

[1]Thanks to Nevena Šelić for this information. See Šelić [2006].

We extract the 6m rate, the 6x12, 12x18 etc. FRA rates from this curve. These rates are appropriate for use in a LIBOR Market Model - in the UK, swaps settle against a 6 month fixing. We then perform a PCA analysis of those FRA rates. Again, this would be appropriate for calibration of the LMM.

The results are satisfactory, with the first three factors having expected characteristics. See Figure 2. 90% of variation is modelled with the first three factors.

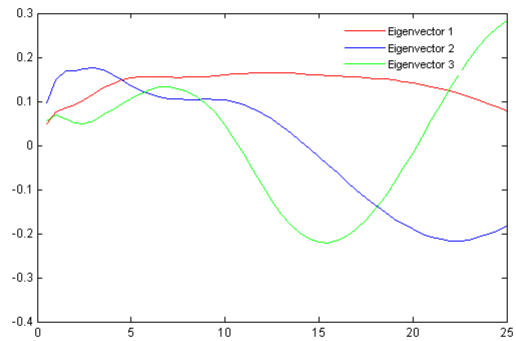

FIGURE 2. The first three FRA components from modelled curve inputs

These are modelled curves: they are generated by some parametric model, and consequentially have significant built in smoothness associated. In fact, the Svensson method is tailor made for PCA - it is a parametric method, with the entire curve described by a single set of parameters representing the long-run level of interest rates, the slope of the curve and humps in the curve. These curves are not guaranteed to price back the inputs that were available in the market at those time.

7.1.2. *Bootstrapped curves.* We consider GBP interbank curves, bootstrapped using LIBOR and swap data available on Bloomberg.

First we consider zero rates as before. The first three factors explained 95% of the variation. See Figure 3.

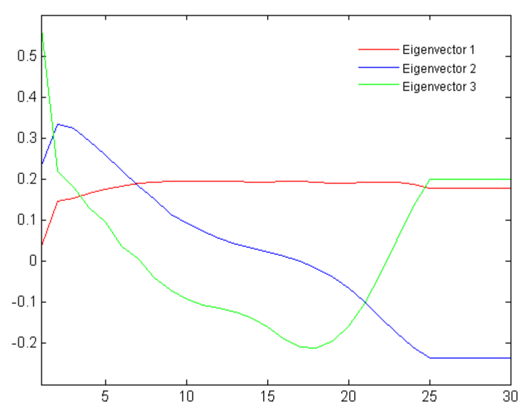

FIGURE 3. The first three rates components from bootstrapped curve inputs: monotone convex interpolation

Now we consider the FRA rates as before. One will find that the results are exceptionally dirty. The factors do not have fundamental economic interpretations. Moreover, it is highly dependent on the interpolation method.

For a discussion of interpolation methods see Hagan and West [2006], Hagan and West [2008]. Under monotone convex interpolation, only 67% of variation is modelled by the first three factors. See Figure 4. Under raw interpolation, the percentage is 73%; see Figure 5.

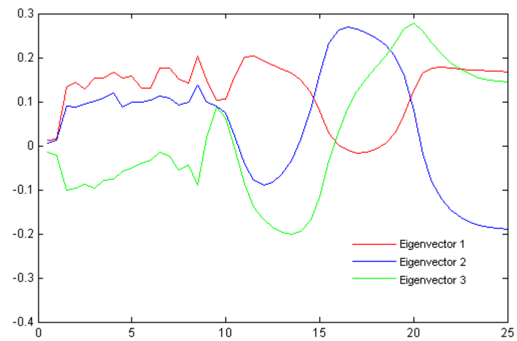

FIGURE 4. The first three FRA components from bootstrapped curve inputs: monotone convex interpolation

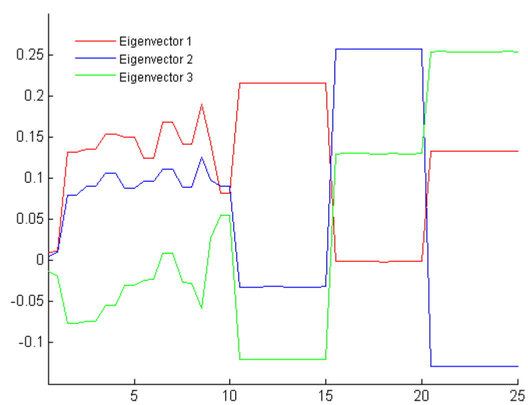

FIGURE 5. The first three FRA components from bootstrapped curve inputs: raw interpolation

7.2. **Stress scenarios of yield curves.** We can use this method to calibrate yield curve stresses that have occurred historically. The idea is to recalibrate to current conditions the first principal component shift that has occurred historically.

Suppose on some historic date t we have a correlation matrix $\mathbb{P}^t$, eigenvalues λ_i^t and corresponding eigenvectors e_i^t. Note that the first factor score for that day is ${x^t}' e_1^t {\lambda_1^t}^{-1/2} = {\lambda_1^t}^{-1/2}\langle x^t, e_1^t\rangle$. We could take this factor score and apply it as a factor score for today, which we denote by T. Thus, the stress movement in the yield curve is $\sqrt{\frac{\lambda_1^T}{\lambda_1^t}}\langle x^t, e_1^t\rangle {e_1^t}'$. This provides us with an appropriate mechanism for updating historic stress events to be applicable to today's market regime.

An alternative requirement will be for multi-day stresses. One approach would simply be to build a correlation matrix of multi-day returns and proceed as before. Of course, a problem with this approach is (as return periods need to be disjoint, otherwise spurious autocorrelation is introduced) that for longer multi-day returns (in practice, as soon as we go beyond a week or so) there will be insufficient data to have built a reliable correlation matrix.

So, suppose the s-day return that occurred is x_{t+s}. Since $e_{t,1}, \dots e_{t,m}$ form an orthonormal basis for $\mathbb{R}^t$, $x_{t+s} = \sum_{i=1}^{m} \langle x_{t+s}, e_{t,i} \rangle e_{t,i}$. The first component approximation to the shift in yield curve is $\langle x_{t+s}, e_{t,1} \rangle e_{t,1}(\tau)$, where τ denotes the term along the yield curve. This quantity $s_t = \langle x_{t+1}, e_{t,1} \rangle$ can be stored and the corresponding stress event at time T could be given by $s_t e_{1,T}(\tau)$. Alternatively, the normalisation by the ratio of the square root of the current first eigenvalue to some sort of average value of the same at time t could be used. As always with stress tests, there is a need to be creative, and not abide by fixed rules, or hope for statistical tests of the validity of processes, which by their very nature, do not allow for testing.

7.3. **Asynchronous updates to the at-the-money term structure.** We see that in the SABR model, the at-the-money volatility parameter is of the utmost importance. While the other parameters only update occasionally, it is critical this parameter is updated in pace with the market.

The issue then arises as what to do when the market trades for one or a couple of expiries and not for the others. It would be suitable to have a model to update those expiries for mark to market purposes.

As an exchange traded option approaches its expiry date, the term to maturity declines with each day. Thus, in order to analyse the dynamic of the at the money volatility, we need to remodel the data as having fixed terms to maturity. We took the mark to market data published at SAFEX SAFEX [2009] and recorded the times to maturity of each of the options. Then by using raw interpolation on the variances (the squares of the volatilities) we obtained volatilities for terms of 45, 136, 227, 318, 409, 500, 591, 682, 773 days to expiry. (This sequence starts at 1.5 months, which is the average time to expiry of the near contract, and increases in steps of 3 months.)

In order to identify common features the use of principal component analysis is suggested Alexander [2001]. We performed a PCA on the correlation matrix of differences in ATM implied volatility. The volatilities and correlations were calculated using an EWMA weighting scheme, in order to lend more weight to the more recent data. The data analysed here are the changes in volatility, not price returns.

The analysis produces expected features. The first component can be understood as a parallel shift to volatilities and accounts for 69.1% of the variation in the data. The second component can be understood as a tilt in the term structure and accounts for 14.8

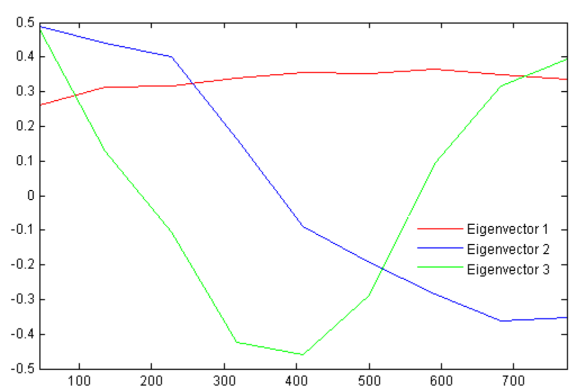

FIGURE 6. The first three components of the PCA analysis on the ATM volatilities

The scree plot appears in Figure 7

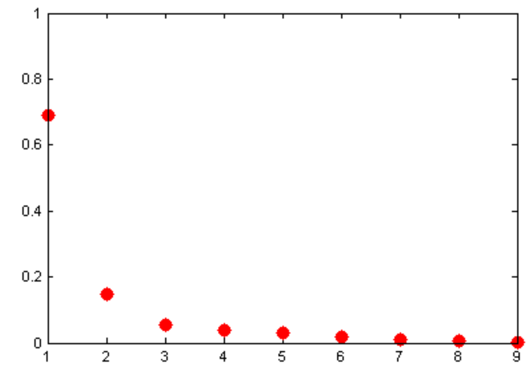

FIGURE 7. The scree plot of the PCA analysis on the ATM volatilities

Taking only the first factor, we see that

$$\Delta\sigma_{\text{atm}} \approx f_1 \underline{e}_1$$

where f_1 is the so-called factor score, and $\underline{e}_1$ is the first eigenvector as diagrammed. Thus, given any reliable change to an at the money volatility, we can model the changes to all the other at the money volatilities. Similarly, taking the first two factors, we see that

$$\Delta\sigma_{\text{atm}} \approx f_1 \underline{e}_1 + f_2 \underline{e}_2$$

Thus, given two changes to at the money volatility, we can model the changes to all the other volatilities to the second order.

For example, in the graph above, we have the eigenvectors in Table 1.

	$\underline{e}_1$	$\underline{e}_2$	$\underline{e}_2$
45	0.2601	0.4888	0.4833
136	0.3149	0.4406	0.1278
227	0.3161	0.4005	-0.1049
318	0.3379	0.1637	-0.4233
409	0.3568	-0.0907	-0.4609
500	0.3518	-0.1934	-0.2885
591	0.3665	-0.2837	0.0922
682	0.3480	-0.3611	0.3152
773	0.3355	-0.3510	0.3955

TABLE 1. The first three eigenvectors from the volatility analysis

Suppose the 227 day rate changes by 0.04 and the 591 day rate by 0.02. Then we have the system

$$f_1 0.3161 + f_2 0.4005 = 0.04$$

$$f_1 0.3665 - f_2 0.2837 = 0.02$$

which solves (for example, just use Cramer's rule) as $f_1 = 0.0819$, $f_2 = 0.0353$. Thus in this example the 45 day change is 0.0385.

7.4. **Returns of stock index constituents.** We consider the South African equity market.

At the time of writing this section (July 2006) the TOPI consisted of 41 stocks, the youngest of which was listed on 4 March 2003. So we take the closes of those stocks from that date until 7 July 2006.[2] We calculate the log price ratios and compute the correlation matrix of these log price ratios.

We then extract the principal components. The first component is roughly a parallel move: the entire market moving up or down with general economic trends. The second component might be split into a resources component and an (opposite) financial component, although even that might be a stretch, as LBT and OML would be placed in the resources component. In fact, PCA in its entirety is probably dubious here: there is no inherent correlation structure. One can expect a significant first factor, it would be akin to the CAP-M factor. However, the first component only explains 29% of the data; even 10 factors only explain 60% of the data.

8. Applications: portfolio hedging

This is a variation of [Back, 2005, §11.6]. Suppose we have a position P_0 (or indeed, portfolio) which has sensitivity to yield curve shifts.

We wish to hedge this instrument again movements in the yield curve. Within the PCA framework, we decide that our objective is to hedge the portfolio against any variation in the first m principle components, where we might take as usual $m = 3$. For this, we need m other instruments which have sensitivity to the yield curve. One assumes that the remaining components have trivial p&l impact; perhaps mathematically we take the view that these other components are fixed.

Thus, we choose instruments $P_1, P_2, \ldots, P_m$. We seek a portfolio $P = \sum_{i=0}^m \alpha_i P_i$ (where $\alpha_0 = 1$, the α_i are the amounts of the instruments held) so that this portfolio does not change value when the first m components experience arbitrary moves.

Let P_i also represent the value of the i^{th} portfolio, decomposed into cash flows C_{ij} at times t_{ij}. We assume that these times t_{ij} coincide with the tenors in the PCA, otherwise, some sort of 'RiskMetrics-type' mapping needs to be undertaken. Thus

$$P_i = \sum_{j=1}^n Z(0, t_{ij}) C_{ij}$$

and

$$\Delta P_i \approx - \sum_{j=1}^n Z(0, t_{ij}) t_{ij} C_{ij} \Delta r_{ij}$$

Now, our requirement is that the portfolio is immune to the values of the first m factor scores. From (10) we see that

$$\Delta r_j = \sum_{k=1}^m e_{jk} z_k$$

[2] One could for each pair of stocks take the closes as far back as the younger of the pair has been listed, and calculate the correlation based on this data. This approach is advisable if there is one or more of the stocks which are indeed quite young. However, the matrix that arises might not be a genuine correlation matrix, so the correction algorithm of Higham [2002] would be required. In this case, all the shares are over three years old, so we just truncate the data as described.

where r_j is the j^{th} rate in the PCA, $e_1,\ e_2,\ \ldots,\ e_m$ are the largest eigenvectors, and z_k is a random factor. Then

$$\begin{aligned}\Delta P_i &\approx -\sum_{j=1}^{n} Z(0,t_{ij})t_{ij}C_{ij}\sum_{k=1}^{m} e_{jk}z_k \\ &= -\sum_{k=1}^{m}\left(\sum_{j=1}^{n} Z(0,t_{ij})t_{ij}C_{ij}e_{jk}\right)z_k \\ &:= \sum_{k=1}^{m} w_{ki}z_k\end{aligned}$$

Thus

$$\begin{aligned}\Delta P &\approx \sum_{i=0}^{m}\sum_{k=1}^{m}\alpha_i w_{ki}z_k \\ &= \sum_{k=1}^{m}\left(\sum_{i=0}^{m}\alpha_i w_{ki}\right)z_k\end{aligned}$$

and so our task is achieved if we set

$$\sum_{i=0}^{m}\alpha_i w_{ki} = 0$$

for $k = 1,\ 2,\ \ldots,\ m$.

We find $\alpha_1,\ \alpha_2,\ \ldots,\ \alpha_m$ by linear inversion, thus

$$\begin{bmatrix} w_{11} & \cdots & \cdots & \cdots & w_{1m} \\ \vdots & \ddots & & & \vdots \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ w_{m1} & \cdots & \cdots & \cdots & w_{mm}\end{bmatrix}\begin{bmatrix}\alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_m\end{bmatrix} = \begin{bmatrix}-w_{10} \\ -w_{20} \\ \vdots \\ -w_{m0}\end{bmatrix}$$

9. Applications: exposure to factors

A swap butterfly is a combination of short- and long term plain vanilla swaps (called the wings) and a medium term swap (called the body). In a receiver swap butterfly, the investor receives the fixed leg of the body and pays the fixed leg of the wings.

A similar definition will apply to a bond butterfly, where the 'term' concept will probably be replaced by duration. The advantage of swap butterflies is that they are cash-neutral. [3]

By convention, the size of the body is unity, the sizes of the wings are then solved for using the PCA analysis similar to §8. A swap butterfly is made insensitive to changes in the first two PCA factors, retaining exposure only to the third factor.

To perform the PCA analysis, we can model the swap as a fixed coupon bond whose coupons are the swap rate, that is, we remember that a receive fixed swap is a long position in a fixed coupon bond and a short position in a floating rate note. The floating rate note has no exposure to movements in the yield curve.

[3] Fabozzi et al. [2005] construct swap butterflies by hedging again movements in the β_0 and β_1 factors in the swap curve as modelled by Nelson and Siegel [1987]. They refer to Martellini et al. [2003] for the construction of bond butterflies.

For simplicity assume that we have a PCA analysis with annual nodes of the swap curve, and that the swaps we consider have annual coupons. Let the wings be denoted with the indices -1 and 1, and the body with index 0. Suppose

- the swap rates are $R_\cdot$,
- the tenors (in years) are $t_\cdot$,
- the values of the corresponding unit bonds are $P_\cdot$,
- the nominals are $\alpha_\cdot$, so $\alpha_1 = 1$, $\alpha_{-1}, \alpha_1 < 1$.

The portfolio is $P = \sum_{i=-1}^{1} \alpha_i P_i$. Then

$$P_i = \sum_{j=1}^{t_i} Z(0,j)R_i + Z(0,t_i)$$

and

$$\Delta P_i \approx -\sum_{j=1}^{t_i} jZ(0,j)R_i \Delta r_j - t_i Z(0,t_i)\Delta r_{t_i}$$

We are looking for a portfolio, that is, we must solve for $\alpha_\pm$, which is invariant under yield curve moves of the following form:

$$\Delta r_j = e_{j1}z_1 + e_{j2}z_2$$

Then

$$\begin{aligned}\Delta P_i &\approx -\sum_{j=1}^{t_i} jZ(0,j)R_i \left[e_{j1}z_1 + e_{j2}z_2\right] - t_i Z(0,t_i)\left[e_{t_j1}z_1 + e_{t_j2}z_2\right] \\ &:= w_{1i}z_1 + w_{2i}z_2\end{aligned}$$

where

$$\begin{aligned}w_{1i} &= -R_i \sum_{j=1}^{t_i} jZ(0,j)e_{j1} - t_i Z(0,t_i)e_{t_j1} \\ w_{2i} &= -R_i \sum_{j=1}^{t_i} jZ(0,j)e_{j2} - t_i Z(0,t_i)e_{t_j2}\end{aligned}$$

Then

$$\Delta P \approx \sum_{i=-1}^{1} \alpha_i w_{1i} z_1 + \alpha_i w_{2i} z_2$$

and so our task is achieved if we set

$$\begin{aligned}\sum_{i=-1}^{1} \alpha_i w_{1i} &= 0 \\ \sum_{i=-1}^{1} \alpha_i w_{2i} &= 0\end{aligned}$$

which is the solution to

$$\begin{bmatrix} w_{1,-1} & w_{1,1} \\ w_{2,-1} & w_{2,1} \end{bmatrix} \begin{bmatrix} \alpha_{-1} \\ \alpha_1 \end{bmatrix} \begin{bmatrix} -w_{1,0} \\ -w_{2,0} \end{bmatrix}$$

Bibliography

Carol Alexander. *Market Models: A Guide to Financial Data Analysis*. John Wiley and Sons, Ltd, 2001.

Kerry Back. *A Course in Derivative Securities: Introduction to Theory and Computation*. Springer, 2005.

Damiano Brigo. A note on correlation and rank reduction. Banca IMI Working Paper, 2002. URL http://www.damianobrigo.it/.

Damiano Brigo, Fabio Mercurio, and Francesco Rapisarda. Parameterizing correlations: a geometric interpretation. Product and Business Development Group, Banca IMI Working Paper, 2002. URL http://www.fabiomercurio.it/papers.html.

Kevin Dowd. *Beyond Value at Risk: The new science of risk management*. Wiley series in Frontiers of Finance. John Wiley and Sons, 1998.

Frank J. Fabozzi, Lionel Martellini, and Philippe Priaulet. Predictability in the shape of the term structure of interest rates. *Journal of Fixed Income*, 15(1):40–53, 2005.

Patrick S. Hagan and Graeme West. Interpolation methods for curve construction. *Applied Mathematical Finance*, 13(2):89–129, 2006.

Patrick S. Hagan and Graeme West. Methods for constructing a yield curve. *WILMOTT Magazine*, May: 70–81, 2008. URL http://www.finmod.co.za/interpreview.pdf.

Nicholas J. Higham. Computing the nearest correlation matrix - a problem from finance. *IMA Journal of Numerical Analysis*, 22:329–343, 2002. URL http://www.ma.man.ac.uk/~nareports/narep369.pdf.

F. Jamshidian and Y. Zhu. Scenario simulation model for risk management. *Capital Market Strategies*, 12 (December):26–30, 1996.

J.P.Morgan and Reuters. *RiskMetrics - Technical Document*. J.P.Morgan and Reuters, New York, fourth edition, December 18, 1996. URL http://www.riskmetrics.com.

Ilias Lekkos. A critique of factor analysis of interest rates. *Journal of Derivatives*, 8(1):72–83, 2000.

Lionel Martellini, Philippe Priaulet, and Stéphane Priaulet. *Fixed-Income Securities: Valuation, Risk Management and Portfolio Strategies*. The Wiley Finance Series, 2003.

Massimo Morini and Nick Webber. An EZI method to reduce the rank of a correlation matrix. Warwick Finance WF04-210, 2004. URL http://www2.warwick.ac.uk/fac/soc/wbs/research/wfri/wpaperseries/.

C.R. Nelson and A.F. Siegel. Parsimonious modelling of yield curves. *Journal of Business*, 60(4), 1987.

Riccardo Rebonato and Peter Jäckel. A methodology for creating a valid correlation matrix for risk management and option pricing purposes. Quantitative Research Centre, The Royal Bank of Scotland, 1999. URL http://www.quarchome.org/.

SAFEX, 2009. URL http://www.safex.co.za.

L.E. Svensson. Estimating and interpreting forward interest rates: Sweden 1992-1994, 1994. URL http://www.nber.org/papers/W4871. IMF working paper.

L.E. Svensson. Estimating forward interest rates with the extended Nelson and Siegel method. *Sveriges Riksbank Quarterly Review*, 3, 1995.

Nevena Šelić. The LIBOR market model. Master's thesis, University of the Witwatersrand, 2006. URL http://www.cam.wits.ac.za/mfinance/papers/NevenaMSc.pdf.

Peter Weigel. A note on optimal calibration of the LIBOR market model to the correlations. *The Journal of Derivatives*, 12(2):43–50, 2004.

Lixin Wu. Fast at-the-money calibration of the Libor market model using Lagrange multipliers. *Journal of Computational Finance*, 6(2):39–78, 2002.

Zhenyue Zhang and Lixin Wu. Optimal low-rank approximation to a correlation matrix. *Linear Algebra and its Applications*, 364:161–187, 2003.

Financial Modelling Agency, 19 First Ave East, Parktown North, 2193, South Africa.
E-mail address: graeme@finmod.co.za