



## RESEARCH REPORT AUGUST 2014

INSURANCE ERS

David Redfern PhD  
Douglas McLean PhD  
Moody's Analytics Research

### Contact Us

Americas  
+1.212.553.1658  
clientservices@moodys.com

Europe  
+44.20.7772.5454  
clientservices.emea@moodys.com

Asia (Excluding Japan)  
+85 2 2916 1121  
clientservices.asia@moodys.com

Japan  
+81 3 5408 4100  
clientservices.japan@moodys.com

# Principal Component Analysis for Yield Curve Modelling

## Reproduction of out-of-sample yield curves

### Overview

» This note describes principal component analysis (PCA) and our method for using it to model yield curve dynamics. This has particular application to risk drivers representing interest rate movements in proxy functions, as generated using the B&H Proxy Generator.

» The theoretical basis of PCA is explained, along with its relation to model reduction. The main calculation methods for obtaining the principal components (PCs) are presented and the role of transformations, including centering and scaling as well as more specific transformations, is outlined. The key question of how many PCs (or factors) should be retained in the reduced model is addressed.

» The inverse problem of determining the PCs magnitudes that best represent a given yield curve is studied in detail and a simple analytical solution is presented.

» We consider the trade-off between the number of PCs that are retained and the ability to reproduce yield curves to a given accuracy. Of particular concern is the reproduction of out-of-sample yield curves (i.e. those that were not used in the analysis to derive the PCs) which often arise during stress testing. This is shown to potentially require many more PCs than we might first think. A simple method for artificially extending the range of yield curves is presented, which makes the PCs more robust to common yield curve perturbations and/or stresses.

» In summary, it is not always safe to assume that a two-factor interest rate model means that only two PCs are required or that a three-factor interest rate model means that only three PCs are required. The actual number of PCs required depends on the desired usage of the reduced model in terms of what yield curves are to be reproduced. The danger is that two or three PCs are assumed and people rely on these being sufficient to accurately model yield curves that were not used in the analysis, such as a 100bp parallel stress.

---

## Table of Contents

<b>Introduction</b>	<b>3</b>
Document structure	
<b>Preliminaries</b>	<b>4</b>
What does a principal component look like?	4
Intuitive interpretation	4
PCA as a model reduction technique	6
The reduced model in perspective	6
What is the population?	7
How well does the sample set represent the population?	7
What if the population changes?	8
The inverse problem	8
<b>Steps to generate the principal components</b>	<b>9</b>
Using the covariance matrix	9
Using singular value decomposition on the raw data matrix	10
Comparison	10
<b>Data transformations</b>	<b>11</b>
Mean centring and scaling	11
More general transformations	11
An example	12
<b>How many PCs to include in a reduced model</b>	<b>13</b>
Scores	14
An example	14
<b>The inverse problem: Decomposing a yield curve into principal components</b>	<b>15</b>
An example	15
<b>Expanding the universe of yield curves</b>	<b>16</b>
The 100 basis point parallel stress	16
Incorporating stresses into the reduced model	18
An example	18
<b>Conclusions</b>	<b>21</b>
<b>Appendix A Eigenvalues, eigenvectors and the eigenvalue decomposition</b>	<b>22</b>
An example	22
<b>Appendix B Singular value decomposition</b>	<b>24</b>
<b>Appendix C Link between SVD and eigenvalue decomposition</b>	<b>24</b>
SVD applied to a real symmetric matrix	25
<b>Further Reading</b>	<b>26</b>

## 1. Introduction

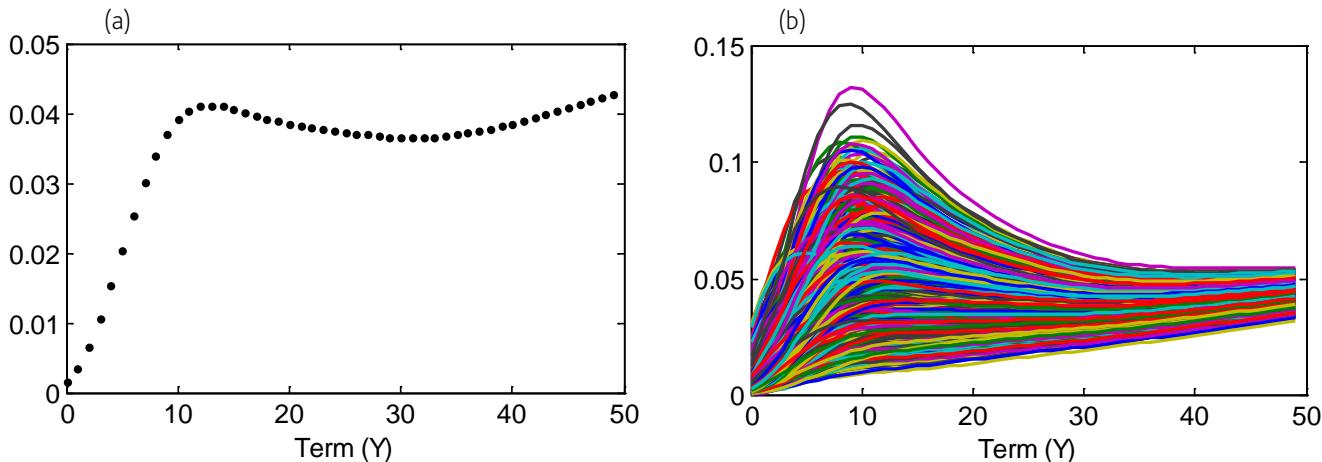
This document describes the method of principal component analysis (PCA) and its application to the selection of risk drivers for capital modelling purposes. There are a multitude of other uses which revolve around common themes of dimension reduction, system simplification, etc. The main idea behind PCA is that a high dimensional system can be approximated to a reasonable degree of accuracy by a system with a smaller number of dimensions by exploiting correlations between the system variables.

One of the main applications in finance is to the modelling of yield curve dynamics. The yield curve could be portrayed in a number of formats and we may apply PCA to any of these formats, but we focus on forward rates here. If we model the yield curve as a series of points representing forward rates at various terms as presented in Figure 1(a), we might have 50 values describing the yield curve. Typically a yield curve model is developed and calibrated so as to produce realistic yield curves at future times. As the yield curve evolves over time, each forward rate can change. It is well understood that adjacent points on the yield curve do not move independently. PCA is a method for identifying the dominant ways in which various points on the yield curve move together.

Using a structural model for yield curve evolution such as the two factor Black-Karasinski model, LMM or LMM Plus, realistic samples of the yield curve at a particular time horizon (typically one year) can be obtained. By *structural*, we mean a stochastic yield curve model that has a structure of yield curve evolution over time that is defined by stochastic differential equations. As an example, consider the two factor Black-Karasinski model calibrated for the USD economy at the end of December 2012. We use 50 forward rates with annual tenor to represent the yield curves. The structural model and its calibration define the yield curve dynamics and hence the set of possible yield curves that may be produced in a scenario set, such as those shown in Figure 1(b). Note that this set of possible yield curves is usually limited by the structure of the model, often restricted to realistic yield curves that obey no arbitrage conditions or the like. Alternatively, the same PCA methods can be applied to a set of historical yield curves.

PCA allows us to take such a set of yield curves (or related quantity), process them using standard mathematical methods, and then define a reduced form model for the yield curve (or related quantity). Typically, this reduced form model retains only a small number of principal components (PCs) but can reproduce the vast majority of yield curves that the full structural model could. Hence this reduced model has many fewer sources of uncertainty (i.e. dimensions) than if the 50 points of the yield curve were modelled independently. Note that PCA and the reduced model itself have no notion of time evolution. The PCs and the output of the reduced model represent the same quantity at the data set analysed, being yield curves (or related quantity) in this study.

**Figure 1** Representation of a yield curve as 50 forward rates with 1-year tenor. This is for the USD economy at the end of December 2012. Subfigure (b) shows 10000 yield curves (with lines joining the dots for clarity) at a projection horizon of  $t = 1$ , obtained from the two factor Black-Karasinski model.



Firstly we deal with some preliminary and background details, including an introduction to principal components and the intuition behind the associated reduced model. In Section 3, we summarise the theory behind PCA and the mathematical methods that

can be employed. Some useful data transformations are discussed in Section 4, and then Section 5 addresses the problem of how many PCs should be retained in a reduced model.

Section 6 considers the inverse problem in which we are given a specific yield curve and are required to determine the magnitudes of the (small number of) principal components that best reproduces the yield curve. This is arguably the most important section as it highlights the limitations of trying to reproduce out-of-sample yield curves, which often arise during stress testing.

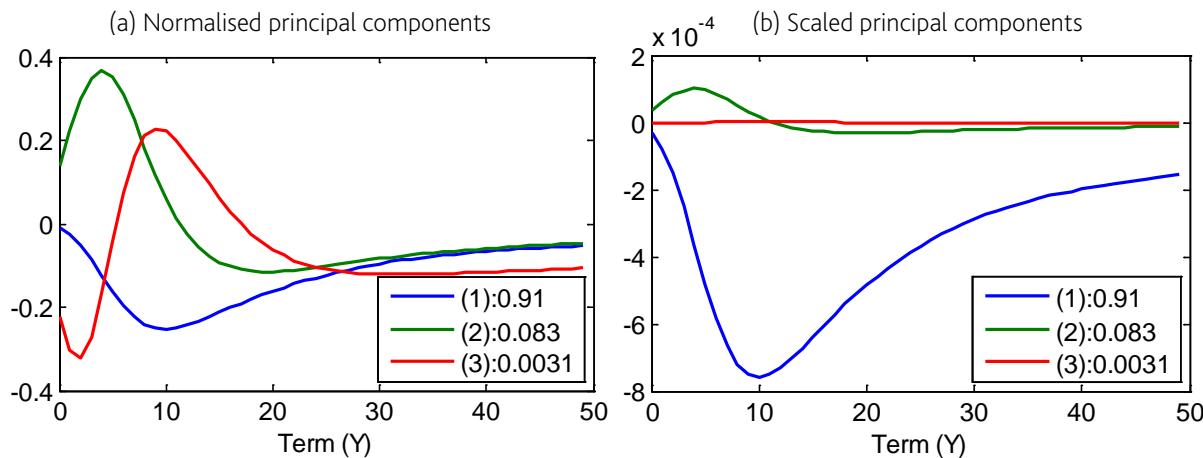
Finally, Section 7 considers some approaches that can be used to expand the universe of possible yield curves to achieve better reproduction of out-of-sample yield curves.

## 2. Preliminaries

### What does a principal component look like?

In mathematical terms, a principal component (PC) is a vector, one component for each variable. In general these variables can represent anything, but for the purposes of this document we consider that the variables are the individual forward rates defining a yield curve. Figure 2 shows some of the PCs obtained by applying PCA to the set of yield curves shown in Figure 1(b). These PCs are represented by 50 points, but they are shown with lines connecting the points. The PCs are ordered so that the first PC is the most important in capturing variability in the yield curves, the second PC is next most important, and so on. When used in practice, PCs are scaled by some real number and so the sign of the actual PC is not important.

**Figure 2** Principal components, represented by 50 forward rates of yield curve, but shown with lines connecting the points. Subfigure (a) shows the normalised form of the three most significant PCs, whilst subfigure (b) shows these scaled according to their significance (explained in more detail later). The legend shows the proportion of variability attributed to each of the PCs.



### Intuitive interpretation

The most intuitive way of obtaining PCs is via eigenvalue decomposition of a covariance matrix (described in detail in a later section). This means we are in the world of variance and covariance, which are measures of central tendency. Thus we are really talking about deviation from the mean. Intuitively, PCs represent ways in which the forward rates making up a yield curve can deviate from their mean levels. With reference to the yield curves in Figure 1(b), the scaled PCs portrayed in Figure 2(b) can be interpreted as follows.

The first PC (in blue in Figure 2) represents the situation that all forward rates in the yield curve move in the same direction but points around the 10 year term move more than points at the shorter or longer parts of the yield curve. This corresponds to a

general rise (or fall) of all of the forward rates in the yield curve, but in no way can this be called a uniform or parallel shift. The impact of the first PC can be easily observed amongst the yield curves in Figure 1(b).

The second PC (in green in Figure 2) represents situations in which the short end of the yield curve moves up at the same time as the long end moves down, or vice versa. This is often described as a *tilt* in the yield curve, although in practice there is more subtle definition to the shape. This reflects the particular yield curves that were used for the analysis, as well as the structural model and calibration that were used to create them. The influence of the second PC is much less easy to observe amongst the yield curves in Figure 1(b), where it accounts for only about 9% of the variability in the yield curves.

The third PC is further interpreted as a higher order buckling in which the short end and long end move up at the same time as a region of medium term rates move down, or vice versa. In this particular example, this type of movement is only responsible for about 0.31% of the variability. This means that it might be appropriate to use two PCs in a reduced model as these should cover around 99% of the yield curve variability.

Figure 3 shows the net effect of the most important PC in explaining shifts of the forward rates in the yield curve about their mean values. The solid black line is the mean yield curve at the  $t = 1$  projection horizon. The broken curves result from addition or subtraction of an arbitrary scaling of the PC. That is, the individual forward rates,  $F_j$  (for  $j = 0, \dots, 49$ , representing the forward rate applying for terms between  $j$  and  $j + 1$  years) are:

$$F_j = \mu_j + \alpha_1 v_{1,j},$$

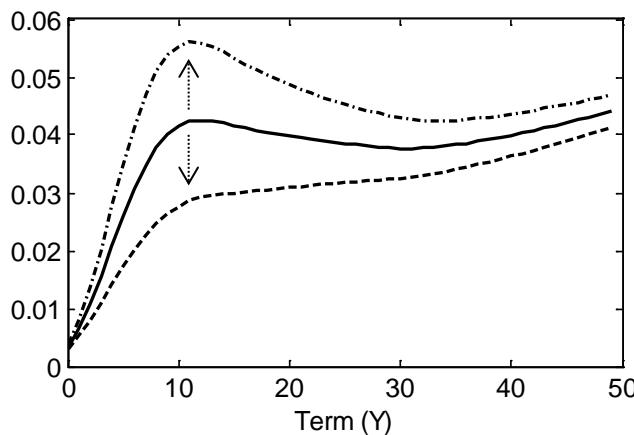
where  $\mu_j$  is the mean value of the  $j$ th forward rate at the  $t = 1$  projection horizon,  $v_{1,j}$  is the  $j$ th element of the first PC, and  $\alpha_1$  is a real value used to scale the first PC. In vector notation this is:

$$\mathbf{F} = \boldsymbol{\mu} + \alpha_1 \mathbf{v}_1,$$

where  $\mathbf{F}$  represents the entire vector of 50 forward rates making up the yield curve,  $\boldsymbol{\mu}$  is the entire vector of mean values of the forward rates at the  $t = 1$  projection horizon, and  $\mathbf{v}_1$  represents the entire vector that is the first PC. In effect this is a very simple reduced form model for the yield curves at the specific  $t = 1$  projection horizon. Since the first PC is responsible for 91% of the variability of the set of yield curves used in the analysis, this reduced model can reproduce a good proportion of the yield curves of the full structural model. That is, suitable values for  $\alpha_1$  will produce yield curves that reasonably match many (but not all) of the yield curves of the full structural model. The appropriate distribution of values of  $\alpha_1$  is a byproduct of the PCA method, as will be shown later.

Note that the universe of possible yield curves is a result of the shape of this first PC as well as the mean yield curve. To expand the universe of possible yield curves we can simply include more PCs, as described next.

**Figure 3** Yield curves produced by a very simple reduced form model composed of just a single PC. The yield curves are defined as a deviation about the mean yield curve, which is the solid black line. The deviation is by the addition of a scaled version of the first principal component to produce yield curves like those shown as broken curves.



### PCA as a model reduction technique

It should be noted that if we are modelling the yield curves as 50 forward rates, then the covariance matrix will be 50x50 and there will be 50 PCs. These are the 50 eigenvectors produced by the eigenvalue decomposition, with the corresponding eigenvalues indicating the relative significance of the PCs. A model including all of the PCs turns out to be able to reproduce any yield curve, realistic or not. This concept is explored in more detail in later sections. In particular, it can perfectly reproduce all of the yield curves on which the PCA analysis was performed. Such a model should be called a full statistical model, and can be represented as:

$$F = \mu + \sum_{i=1}^n \alpha_i v_i,$$

where  $n$  is the number of forward rates, equal to 50 in the present example. This can be expressed in matrix notation as:

$$F = \mu + V\alpha,$$

where  $V = [v_1, v_2, \dots, v_n]$  is a square  $n \times n$  matrix whose columns are the PCs (i.e. the eigenvectors of the covariance matrix), and  $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_n]^T$  is a vector of magnitudes of the PCs.

It has been mentioned that the full structural model has a limited universe of yield curves, whereas the full statistical model presented here can reproduce any yield curve. This means that the full statistical model is unnecessarily flexible. As alluded to earlier, PCA provides a natural way of simplifying the statistical model whilst maintaining its ability to reproduce the majority of yield curves that can be produced by the structural model, by simply reducing the number of PCs that are used. This reduced model includes only the quantity of PCs that are deemed important, as indicated by the corresponding eigenvalues. In particular, we only retain those PCs that are associated with the largest eigenvalues. The reduced model can be expressed as:

$$F = \mu + \sum_{i=1}^k \alpha_i v_i,$$

where  $k$  is the number of PCs retained. This means that the yield curves are expressed as a linear combination of the  $k$  PCs. This can be expressed in matrix notation as:

$$F = \mu + V^{(k)}\alpha,$$

where  $V^{(k)} = [v_1, v_2, \dots, v_k]$  is a rectangular  $n \times k$  matrix whose columns are the  $k$  PCs corresponding to the largest  $k$  eigenvalues, and  $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_k]^T$  is a vector of magnitudes of the PCs.

In a reduced model, the corresponding population or universe of possible yield curves is also reduced. If the number of PCs is large enough the majority of the yield curves used in the analysis can be reproduced to a high level of accuracy. In the modelling of yield curves it often suffices to retain just two or three PCs, but the appropriate number can be deduced by considering the eigenvalues. This is explored in more detail in a later section.

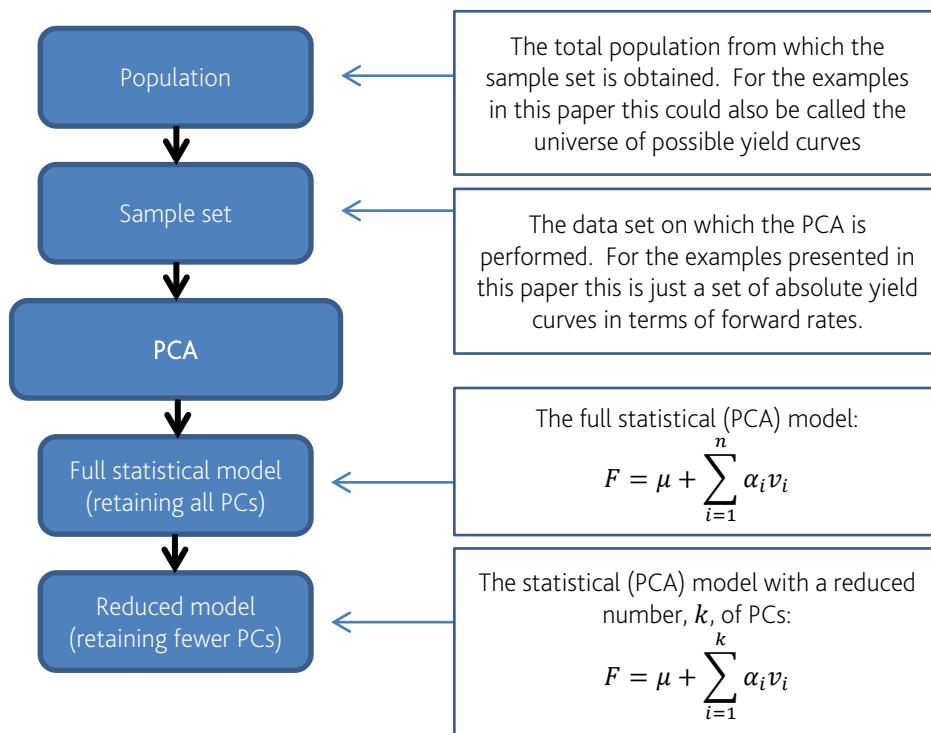
The PCs in such a reduced model are also sometimes called factors. A reduced model retaining two PCs is then called a two-factor model, and a reduced model with three PCs is then called a three-factor model, and so on.

In a more advanced setting, it may be advantageous to include various data transformations (such as taking logs of the forward rates) in the PC analysis. The corresponding reduced model must then incorporate these data transformations. In particular, full details of the transformations are required in order to properly deal with the inverse problem of decomposing a given yield curve into magnitudes of the PCs. Such transformations are dealt with in a later section.

### The reduced model in perspective

It is vital that the limitations of the reduced model are fully understood. Figure 4 seeks to put the whole of the process of determining the reduced model into perspective. Often, the sample set that is analysed in PCA is thought of as the start of the process, and then the reduced model is thought of as producing yield curves that "resemble those in the sample set". However, it is important to appreciate the bigger picture. More likely is that the aim of the reduced model is to be able to produce samples from the population (or universe) from which the sample set was obtained. It is important to appreciate that the population is not the same as the sample set on which the PCA was performed.

Figure 4 The entire process of determining the reduced model.



#### WHAT IS THE POPULATION?

The population governs what can be contained in the sample set. In the case of yield curves, the population may also be called the universe of possible yield curve and could be:

- » Synthetically modelled. This includes output from a structural model representing yield curve dynamics, such as the output produced by a particular calibration of a scenario generator. These could be yield curves produced at a given time horizon (such as t=1 yield curves), yield curves produced at all time horizons, changes in yield curves over particular time intervals, etc. It could be a population of yield curves in terms of forward rates, yield curves in terms of spot rates, or any other way. Most of the examples presented in this document fall into this category. This type of population is likely to be infinite (or exceedingly large), meaning that different random number seeding can be used to produce a different sample sets from the same population. Note that a different calibration of the models would result in a different population.
- » Historically inferred. The population here could be absolute yield curves for any particular historical frequency for a given economy or a given set of economies. It could also be changes in the yield curve over a particular period of time for one or more economies, etc. The population could also be considered to be infinite here, despite the fact that there are only a finite and relatively small number of yield curves (or changes in yield curves, etc) that have been observed. It is often forgotten that these observed yield curves are just a sample set from the population of yield curves that could have occurred. Thinking of this population of yield curves that might have occurred is a much more subjective task, but it is one that must be addressed if we are to use the reduced model to predict future yield curves in a responsible way.

#### HOW WELL DOES THE SAMPLE SET REPRESENT THE POPULATION?

For synthetic sample sets this is often thought of as a task of including more and more samples from the structural model. This should reduce sampling error and is usually a sufficient strategy if the population is defined to be the set of yield curves that can be produced by the structural model, given its particular calibration. If the population is further widened to include any realistic yield curves then we should be aware of the ways in which the structural model and its calibration may deviate from reality.

For historical sample sets, a certain amount of expert judgement should be used when considering just how well the observed set of yield curves matches the population of yield curves that could have occurred (or could occur in the future, if this is relevant). A number of economic and financial theories could come into play, including no arbitrage principles, term premium assumptions, yield curve extrapolation, etc.

There are also artificial concepts that might be worth considering such as stress testing mechanisms that regulators or providers of best practice risk management solutions might dictate.

There is always likely to be a difference in coverage between the population and the sample set on which the PCA is performed. Understanding this gap is key to understanding the limitations of what the reduced model can and cannot be used for. This can also be used to inform the construction of the reduced model, concerning the number of PCs that are retained in the reduced model (considered in more detail in Section 5), or through selection of the sample set itself, which is the subject of Section 7.

#### WHAT IF THE POPULATION CHANGES?

If the sample set of yield curves is obtained from a structural model then the population of yield curves could easily be changed by a standard recalibration. If the population of yield curves (or related quantity) does change then, in theory, the PCA should be repeated. A new sample set should be obtained that represents the new population, the PCA calculations should be repeated and a new reduced model obtained. In practice, the degree to which this is essential will depend on the extent to which the population has changed. It is possible that the same PCs might be appropriate if the change in the population is only minor or if sufficient flexibility had already been build into the reduced model. This is discussed further in Section 7.

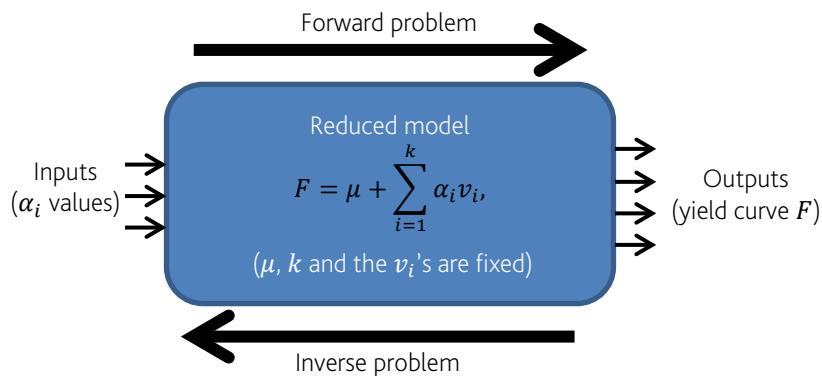
#### The inverse problem

Mathematical literature often talks of the inverse problem as being the task of reversing the normal calculation steps. That is, determining the inputs from a given output (or set of outputs). These can be notoriously difficult problems, often due to the forward problem not being one-to-one. In some cases, specific values of the output may not be achievable for any combination of values of the inputs. The problem is then usually considered to be that of determining the values of the inputs that most closely produce the desired output values. In other cases, the outputs may be identically the same (or the same to the level of numerical precision) for more than one set of input values. The problem is then either to choose one such set of input values according to some criteria or to reduce the ambiguity in possible input values as much as possible.

In the current setting, the inverse problem can be defined as the task of working backward from a given yield curve to determine the inputs that could be used to create it. In terms of inputs, we can reasonably consider that the model structure (that is, the form of the structural model or the number of PCs in the reduced model) and calibration parameters are known and fixed. Hence, it is only the random samples in the structural models or the PC magnitudes of the statistical models that are to be determined. Figure 5 shows this schematically.

There are a number of things that need to be considered when dealing with the inverse problem. Firstly, we need to understand the universe of yield curves that can possibly be produced by the particular model. Secondly, we need to consider the difficulty (in terms of speed and robustness) of any calculation methods. On both of these counts it will be shown that the statistical/reduced models that make use of PCA have distinct advantages. This is explored in more detail in Sections 6 and 7.

*Figure 5 Schematic of the forward problem and the inverse problem.*



### 3. Steps to generate the principal components

This section describes how to determine a set of PCs that are appropriate to represent a given structural model. This section also explains how to determine the number of PCs that should be retained when constructing a reduced model.

In terms of calculations there are two main ways that PCs can be determined: using eigenvalue decomposition; and using singular value decomposition. In this section we concentrate on the simple case of applying PCA to yield curves that have not been transformed in any way, but mention where extensions can be made. These extensions are considered in more detail in later sections.

#### Using the covariance matrix

This method requires us to determine a covariance matrix describing the movement in the points in the yield curve and then performing eigenvalue decomposition (diagonalisation) or singular value decomposition of this covariance matrix.

1. Generate a set of yield curves (in terms of forward rates) from the structural model. Construct a data matrix of forward rates:

$$\mathbf{F} = [F_{ij}],$$

where  $F_{ij}$  is the  $j^{\text{th}}$  forward rate of the  $i^{\text{th}}$  yield curve. That is, each row of the matrix  $\mathbf{F}$  is a yield curve and each column is a set of values of a particular forward rate.

2. Transform the values in the data matrix as desired (all optional). See later description. Note that the transformations here need to be taken into account when the inverse problem of determining PC magnitudes that best match a yield curve is tackled.
3. Centre and scale the data matrix if desired:
  - a. Subtract the mean of each of the forward rates to obtain data that has zero mean.
  - b. Subtract the mean and then divide by the standard deviation of the forward rates, to obtain data that has zero mean and standard deviation of 1.0.
4. Calculate a matrix of covariances between the (possibly transformed, centred and/or scaled) forward rates

$$\mathbf{C} = \text{Cov}(\mathbf{F}),$$

where this is interpreted as a matrix of covariances of pairs of columns. Note that if the data has mean of zero and standard deviation of 1.0 this will also be the correlation matrix.

5. Compute the eigenvalue decomposition or singular value decomposition, with eigenvalues or singular values in decreasing order. The eigenvalue decomposition is:

$$\mathbf{C} = \mathbf{V}\Lambda\mathbf{V}^T,$$

where  $\mathbf{V}$  is an orthonormal matrix and  $\Lambda$  is a diagonal matrix. See 0 for more details of the eigenvalue decomposition.

The singular value decomposition is:

$$\mathbf{C} = \mathbf{U}\mathbf{S}\mathbf{V}^T,$$

where  $\mathbf{S}$  is a diagonal matrix and  $\mathbf{U}$  and  $\mathbf{V}$  are orthonormal matrices that will be identical (to machine precision) since the covariance matrix is positive semidefinite. See Appendix B and Appendix C for more details of the singular value decomposition.

In both cases, the columns of  $\mathbf{V}$  are normalised eigenvectors  $\mathbf{v}_i$ . These are the principal components that we seek.

6. The significance,  $\psi_i$ , of each principal component is given by:

$$\psi_i = \frac{\lambda_i}{\sum_{j=1}^n \lambda_j}, \quad \text{or} \quad \psi_i = \frac{s_i}{\sum_{j=1}^n s_j}.$$

where  $\lambda_i$  are the diagonal elements of the diagonal matrix  $\Lambda$ , and  $s_i$  are the diagonal elements of the diagonal matrix  $\mathbf{S}$ . This provides information as to how many PCs are required in order to capture the range of yield curves that the underlying model can produce. See Section 5 for more details.

7. (Optional) Compute the mean of each forward rate. This mean does not affect the covariance matrix since, by definition, the covariance is a measure of the joint deviation from the mean values. Hence it will not affect the PCs that are obtained. However, this vector of mean values is needed to be known if the inverse problem is to be solved. That is, if a yield curve is to be decomposed into principal components. Any drift governed by the underlying model structure is captured in this mean, as is the roll forward of the yield curve due to time evolution.

### Using singular value decomposition on the raw data matrix

The method outlined here applies singular value decomposition (SVD) to the raw data value matrix (not a covariance or correlation matrix). There is no need to explicitly determine the covariance/correlation matrix if this is not otherwise required, but this also means that mean shifting is significant here.

The steps to obtain the PCs using singular value decomposition (SVD) are similar to those outlined above.

1. Follow step 1 from the eigenvalue decomposition approach outlined above.
2. Follow step 2 from the eigenvalue decomposition approach outlined above.
3. Centre the data and scale if desired:
  - a. Subtract the mean of each of the forward rates to obtain data that has zero mean. Note this was not required when using decomposition of the covariance matrix, but makes a material difference here.
  - b. Subtract the mean and then divide by the standard deviation of the forward rates, to obtain data that has zero mean and standard deviation of 1.0.
4. Compute the singular value decomposition on the (possibly transformed, centred and scaled) matrix  $\mathbf{F}$ :

$$\mathbf{F} = \mathbf{U}\mathbf{S}\mathbf{V}^T$$

See Appendix B for more details of the singular value decomposition. As before, the columns of  $\mathbf{V}$  are the principal components in normalised form. Here these are the eigenvectors  $v_i$  of  $\mathbf{F}^T\mathbf{F}$ . Note that when the columns of  $\mathbf{F}$  have mean of zero then  $\mathbf{F}^T\mathbf{F} = \text{Cov}(\mathbf{F})$  and hence the principal components will be the same as before.

5. The proportion of variance explained by each PC,  $\psi_i$ , is given by

$$\psi_i = \frac{\sqrt{s_i}}{\sum_{j=1}^n \sqrt{s_j}}$$

This provides information as to how many PCs are required in order to capture the range of yield curves that the underlying model can produce.

### Comparison

There are more details on the ins and outs of eigenvalue decomposition and singular value decomposition in 0, Appendix B and Appendix C. All approaches will produce the same results in many cases. Using the eigenvalue decomposition is usually the most intuitive of the approaches and is easier computationally, despite the need to compute the covariance matrix. One advantage of SVD is that it is more accurate in difficult numerical cases and is the preferred approach for PCA tools implemented in some commercial numerical packages such as matlab.

The application of SVD directly to the data matrix,  $\mathbf{F}$ , is computationally the most expensive but it may also be the most accurate. This matrix may have many more rows than columns, since it is common to perform the analysis on 1000 or 10000 or more yield curves, which may each be modelled with around 50 forward rates. Eigenvalue decomposition performed on a 50x50 covariance matrix is usually much quicker than SVD performed on a 1000x50 matrix.

In many cases, the covariance matrix may have a very large condition number, indicating that the matrix is nearly singular. In the examples presented in this document the covariance matrix often had a condition number of  $\sim 10^{18}$ . If the covariance matrix is

going to end up in this condition, it might be best to use SVD directly on the data matrix, especially if PCs of lesser importance are to be obtained with good accuracy.

Another point to note is that mean centring should always be performed when using SVD directly on the data matrix  $\mathbf{F}$ , whereas this is not necessary when covariance matrix methods are used.

## 4. Data transformations

### Mean centring and scaling

Some text book descriptions include a mean centring of the data and then a scaling of the data before attempting PCA. Depending on the nature of the problem and the way it is to be solved, these may or may not be essential.

Mean centring is literally the subtraction of the average (of each forward rate) from the data values so that the new mean value is zero. As mentioned earlier, mean centring is crucial when the SVD approach is used but is not necessary when eigenvalue decomposition is used.

Scaling then involves dividing the data by the standard deviation (of each forward rate) so that the standard deviation of the new data is 1.0. Scaling is particularly useful if the variables have quite different orders of magnitude, although this is not likely to be the case for yield curves. If both centring and scaling are performed then the covariance matrix is the same as the correlation matrix. Note that we then need to reverse these transformations (i.e. scale by standard deviation of the original data and add on the mean of the original data) in order to recreate any of the original yield curves.

### More general transformations

Since PCA is a purely statistical technique, you can transform the data in almost any way and still successfully apply the technique. The inverse transformation is required for the inverse problem of determining the magnitudes of PCs that produce the best fit to a given yield curve. Some transforms may mean that the reduced model allows negative interest rates or other undesirable features.

Let  $T(\cdot)$  be a transformation function that can be applied to the forward rates in the data matrix. PCA is then performed on the transformed data and the final equation for the reduced model is as follows:

$$\mathbf{F} = T^{-1}\left(\boldsymbol{\mu} + \sum_i \alpha_i \mathbf{v}_i\right),$$

where  $\mathbf{F}$  represents the entire vector of transformed data points,  $\boldsymbol{\mu}$  is the entire vector of mean values of the transformed data,  $\mathbf{v}_i$  represents the entire vector of the  $i^{\text{th}}$  PC, and  $\alpha_i$  is an appropriate scaling of the  $i^{\text{th}}$  PC. In matrix notation this is:

$$\mathbf{F} = T^{-1}(\boldsymbol{\mu} + \mathbf{V}^{(k)} \boldsymbol{\alpha}),$$

where  $\mathbf{V}^{(k)} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k]$  is a rectangular  $n \times k$  matrix whose columns are the  $k$  PCs corresponding to the largest  $k$  eigenvalues, and  $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \dots, \alpha_k]^T$  is a vector of magnitudes of the PCs.

Some useful transformations are:

- » General scaling. If the raw data is a set of possible yield curves at some time horizon, it might be meaningful to scale by the initial ( $t = 0$ ) yield curve to get relative movements. The PCs would then represent proportional changes in the yield curve over the given time interval.

$$T(\mathbf{F}) = \frac{\mathbf{F}}{\mathbf{F}_0},$$

where  $\mathbf{F}_0$  is the vector of forward rates representing the initial yield curve ( $t = 0$ ). The reduced model is then:

$$\mathbf{F} = \mathbf{F}_0 \cdot (\boldsymbol{\mu} + \mathbf{V}^{(k)} \boldsymbol{\alpha})$$

- » Logarithmic transformation. Some structural models for nominal yield curves, such as the two factor Black-Karasinski and LMM models, have a lognormal structure in order to prevent negative interest rates. Without special treatment, there is nothing preventing some of the forward rates from a reduced model being negative, since the scale parameters  $\boldsymbol{\alpha}$  can take any real value, including dramatically negative values. One common way to prevent negative interest rates being produced from a reduced model is to use a logarithmic transformation of the raw yield curve data before performing PCA. The

transformed data can take any real value, but when transformed back to actual forward rates it is theoretically impossible to get negative rates. The transformation function is:

$$T(F) = \ln(F).$$

This means that the reduced model would be:

$$F = \exp(\mu + V^{(k)}\alpha)$$

Clearly, the forward rates that result from this reduced model can never be negative, for any real PCs and for any real values of  $\alpha$ .

- » Displaced logarithmic transformation. This is another useful transformation that is specifically suited to the LMM Plus model. The LMM Plus model features a displacement of the forward rates which has the effect of shifting the yield curve downwards. This means that the resulting forward rates can be as negative as the (positive) displacement parameter, and so a plain log transform will break down. However, the displaced forward rates from the LMM Plus models are lognormally distributed, and so the following transform can be used:

$$T(F) = \ln(F + \delta),$$

with associated reduced model:

$$F = -\delta + \exp(\mu + V^{(k)}\alpha).$$

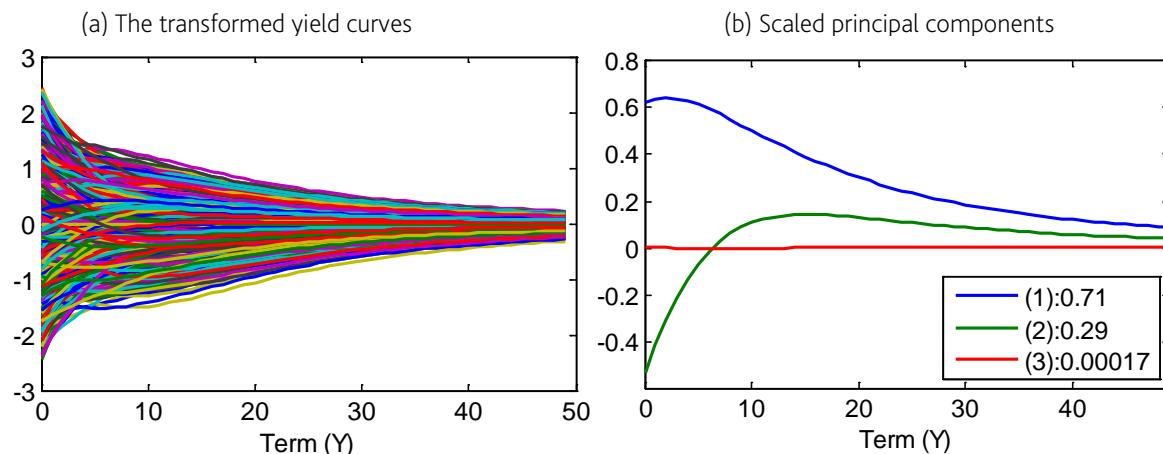
Clearly, the forward rates that result from this reduced model can never be more negative than  $-\delta$ , for any real PCs and for any real values of  $\alpha$ .

### An example

The following example illustrates some of the benefits of using transformations. Consider the use of a logarithmic transformation and mean centring during the PCA on the same set of yield curves as appear in Figure 1(b). The transformed yield curves are shown in Figure 6(a). Note that the general shape of the data is more generic and more symmetric, and hence less resembles the lognormal distribution usually associated with the two factor Black-Karasinski model. The three largest associated PCs are shown in Figure 6(b), including the details of the proportion of variability attributable to each of the PCs.

Compared with the non-transformed PCs shown in Figure 2(b), the transformations have shifted some importance from the first PC to the second PC. Now only around 71% of the variability can be obtained using a reduced model with only a single PC. However the important thing here is that the importance of the third PC has decreased from 0.31% to 0.017%. This means that a greater proportion of the variability in the yield curves can be explained by using a reduced model with two PCs, something over 99.9%.

**Figure 6** The transformed yield curves corresponding to the example using the two factor Black-Karasinski model presented in Figure 1(b). The corresponding (scaled) PCs are also shown, along with proportion of variability attributable to each PC, which can be directly compared with the results in Figure 2(b).



## 5. How many PCs to include in a reduced model

The question of just how many PCs should be retained in a reduced model is common. It has been mentioned before that if we retain all of the PCs then we can exactly reproduce all of the yield curves on which the PCA was performed. Note that this set of yield curves is only a sample set obtained from the infinite population of yield curves that the structural model can produce. This means that it is possible that the full structural model can produce a yield curve (that was not in the set of yield curves on which PCA was applied) that the reduced model cannot exactly reproduce, even with all of the PCs.

The more PCs are retained the more accurately the reduced model will be able to reproduce the yield curves on which the PCA was performed. In some cases it may be important to allow flexibility to be able to fit yield curves that were not in the set of yield curves on which the PCA was performed. This might also include perturbed or stressed yield curves that cannot even be produced by the full structural model. Hence we may decide to retain a relatively large number of PCs. However, with more PCs comes more complexity which may cause problems in systems in which the reduced models are to be used. In some cases it is desirable to retain the smallest number of PCs that can reasonably approximate most of the yield curves. It is important to be aware of the trade-off between the number of PCs that are retained and the ability to reproduce yield curves to a given accuracy.

As mentioned in Section 3, the proportion of variance explained by each principal component is given by:

$$\psi_i = \frac{\lambda_i}{\sum_{j=1}^n \lambda_j} \quad \text{or} \quad \psi_i = \frac{s_i}{\sum_{j=1}^n s_j},$$

when eigenvalue or singular value decomposition is performed on the covariance matrix. Alternatively,

$$\psi_i = \frac{\sqrt{s_i}}{\sum_{j=1}^n \sqrt{s_j}},$$

when singular value decomposition is applied to the raw data matrix. These expressions come about because of the properties of the matrix decompositions that are used. In the case of the eigenvalue decomposition approach, the decomposition is performed on the covariance matrix. Since the eigenvectors are orthogonal and normalised, the eigenvalues are then measures of the variance of the corresponding forward rate in a coordinate system in which the transformed forward rates are independent (See 0 for more details of the interpretation of the eigenvalue decomposition). This means that the denominator in the expressions above represents the total variance (i.e. the sum of the variances of independent variables), and then  $\psi_i$  is the proportion of the total variance attributed to the  $i^{\text{th}}$  PC.

The variance of the reduced model as a proportion of the variance of the set of yield curves that were analysed is given by the cumulative sum of the expressions above for each PC in the model. That is:

$$\sum_{i=1}^k \psi_i,$$

where  $k$  is the number of PCs included in the reduced model, keeping in mind that the PCs are included in the reduced model in order of decreasing eigenvalue (or singular value). That is, PCs associated with the largest eigenvalues are included before PCs corresponding to smaller eigenvalues. Hence, based on this measure, there is diminishing worth in retaining more and more PCs in the reduced model.

The number of PCs in the reduced model should be such that the expression above is close to 1.0, meaning that almost 100% of the variance is covered. In some cases it might be sufficient that 95% of the variance is covered, but in other applications a higher degree of coverage might be required.

It is worth remembering that these measures of variability (and hence  $\psi_i$ ) are only with-respect-to the set of yield curves on which the PCA was performed. If this set does not sufficiently represent the population or universe of yield curves of interest (as outlined in Section 2), then using the criteria outlined above may not be appropriate. This is particularly evident when the sample set consists of historically observed yield curves, since the notion of population is a bit more subjective in this case. But it is also an issue for sample sets produced from structural models when the population changes or when we wish the reduced model to be able to reproduce yield curves that are vastly different from those in the sample set. This is the subject of Section 7.

Another way to consider how many PCs should be retained in the reduced model is to consider scores.

## SCORES

Scores are the magnitudes of the PCs that are required to fit each yield curve in a particular set. Most often this set is the original set of yield curves on which the PCA is performed, but the concept can be extended to slightly different sets. For a given set of yield curves, the scores can be determined by solving the inverse problem. Since the PCs are orthogonal this is not a difficult task, and is described in detail in Section 6.

The following example includes some distributions of scores, and the process is further examined in Section 7 using another specific example.

### An example

Continuing the example of the yield curves obtained from the two factor Black-Karasinski model, the proportion of the variance attributable to each of the first four PCs is shown in Table 1. PCs obtained from ordinary PCA (as portrayed in Figure 2(b)), and PCs using log transformed and mean centred PCA (as portrayed in Figure 6(b)) are included separately. The cumulative proportion of variance for a reduced model containing the specified number of PCs is also shown for the two cases. In both cases, a great proportion of the variance is reproducible if two PCs are included in the reduced model. This is expected since the full structural model is a two factor model. The proportion covered by the log transformed PCs is larger (except for the reduced model with only a single PC), highlighting the benefits of the log transform in this example. Note that, in terms of variance coverage for this particular set of yield curves, it would even be better to use the log transformed model with two PCs (99.98%) than the non-transformed model with three PCs (99.93%).

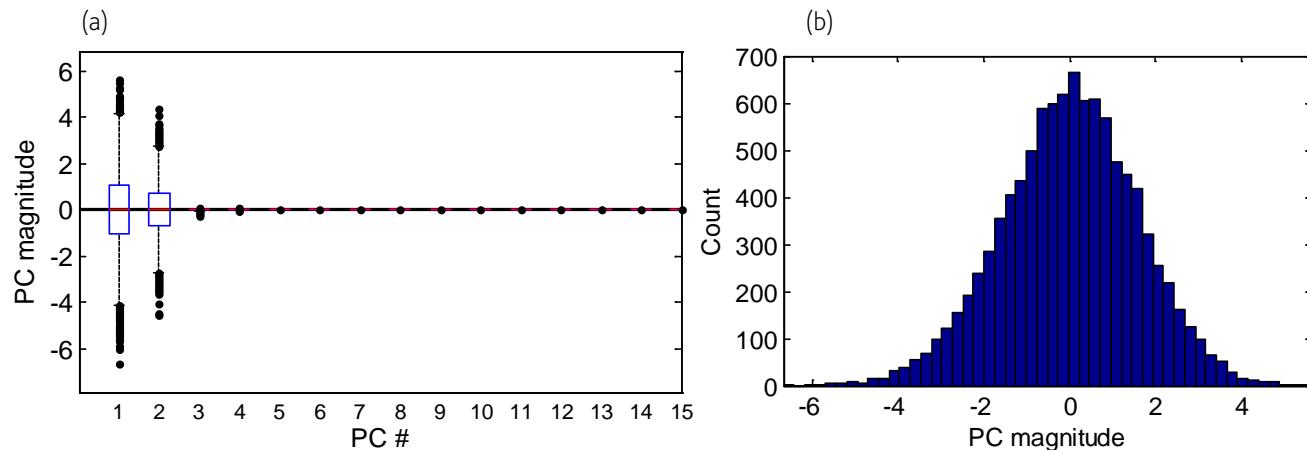
0(a) shows the distribution of scores for the first 15 PCs for the log transformed case. For each PC, the edges of the box signify the 25<sup>th</sup> percentile and 75<sup>th</sup> percentile of the distribution and the red central line is the median. The whiskers extend 1.5 times the interquartile range either side of the box, and outliers (marked with black dots) are any points lying outside the whiskers. This tells us that there is significant requirement for the first two PCs and only very tiny need for any other PC. 0(b) shows a histogram of the distribution of magnitudes of the first PC, giving a more complete view of the distributional form.

As has been previously mentioned, care should be taken when using these criteria to determine the number of PCs to retain since they are only measured against the set of yield curves on which the PCA was performed, rather than the full population or universe of yield curves that the full structural model can produce. The indicative number of PCs should only be considered as a minimum. To ensure that the indicated number of PCs here is more meaningful the set of yield curves should be quite large. Additional PCs can be included in the reduce model, in order to build in some flexibility to reproduce a wider universe of yield curves if desired. The ideas behind this concept are extended in the Section 7.

**Table 1** The proportion of the variability that can be explained by the first four PCs, for the example of the yield curves obtained from the two factor Black-Karasinski model. PCs obtained from ordinary PCA (as portrayed in Figure 2(b)), and PCs using log transformed and mean centred PCA (as portrayed in Figure 6(b)) are included.

PC #	Ordinary PCA		Log transformed and mean centred PCA	
	Proportion of variance attributable to this PC	Cumulative proportion of variance including this PC	Proportion of variance attributable to this PC	Cumulative proportion of variance including this PC
1	91.36%	91.36%	71.26%	71.26%
2	8.25%	99.61%	28.72%	99.98%
3	0.31%	99.93%	0.017%	99.998%
4	0.0659%	99.996%	0.00146%	99.9998%

**Figure 7** Scores of the log transformed and mean centred PCA for the example of the yield curves obtained from the two factor Black-Karasinski model. Subfigure (a) portrays a box plot showing the distribution of scores of the first 15 PCs. For each PC, the edges of the box signify the 25<sup>th</sup> percentile and 75<sup>th</sup> percentile of the distribution and the red central line is the median. The whiskers extend 1.5 times the interquartile range either side of the box, and outliers (marked with black dots) are any points lying outside the whiskers. Subfigure (b) shows a histogram of the distribution of magnitudes for PC #1.



## 6. The inverse problem: Decomposing a yield curve into principal components

This section deals with the problem of decomposing a given yield curve into an approximate weighted sum of a given set of PCs. This is a fairly common requirement that comes about when a reduced form model is used in practice. For example, when risk drivers for a proxy function are chosen to be magnitudes of a set of PCs, the magnitudes of these PCs need to be determined from a specific yield curve when the proxy function is to be evaluated. These magnitudes are also known as scores.

It was mentioned earlier that the full statistical model with all PCs can reproduce any yield curve. The magnitudes of the PCs that can be used to exactly match a given yield curve can be determined by a simple matrix multiplication. When the model has been reduced to only utilise a smaller number of PCs the universe of possible yield curves is also reduced. This means that there are some yield curves that cannot be reproduced exactly. In this case we are really seeking the magnitudes of the retained PCs that best fit the yield curve in some sense. The same matrix multiplication naturally provides the best fit in the transformed domain in a least squares sense. Note that no numerical optimisation is necessary to find the best fit for any yield curve.

The task is to find the vector  $\alpha$  that solves the equation  $F = \mu + \mathbf{V}^{(k)}\alpha$ . This can be obtained analytically as follows:

$$\alpha = \mathbf{V}^{(k)^{-1}} \cdot (F - \mu) = \mathbf{V}^{(k)^T} \cdot (F - \mu),$$

since the orthogonality property of the eigenvalue matrix means that  $\mathbf{V}^{(k)^{-1}} = \mathbf{V}^{(k)^T}$ .

When a data transformation is being used the solution becomes:

$$\alpha = \mathbf{V}^{(k)^T} \cdot (T(F) - \mu),$$

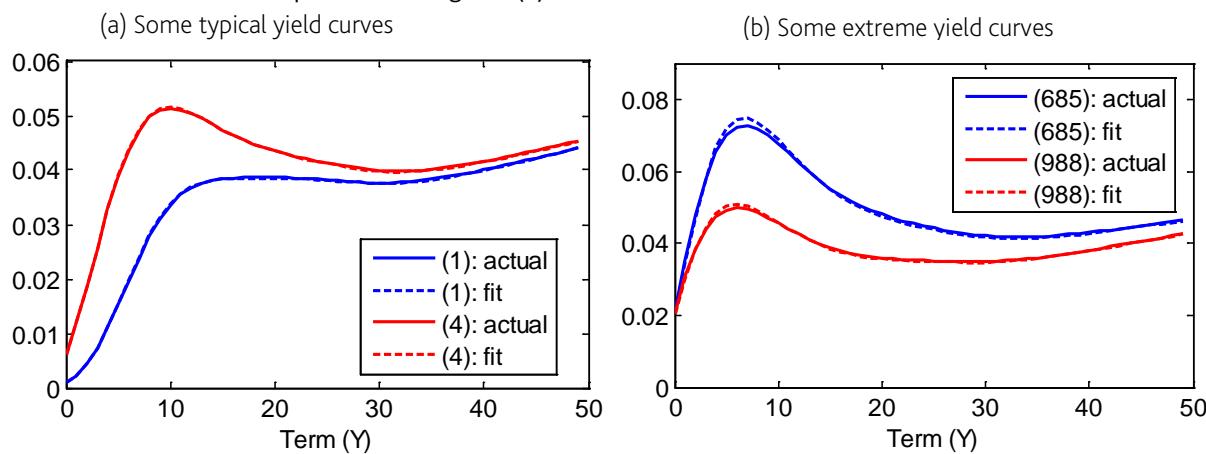
where  $\mu$  is to be interpreted as the mean value of the transformed data rather than the raw data.

### An example

In terms of an example, consider the reproduction of yield curves using the reduced model obtained from the 10000 yield curves obtained from the two factor Black-Karasinski model. The PCA is used with a log transform and mean centring and so the PCs are those shown in Figure 6(b). Figure 8 shows the resulting fit to some of the yield curves using the reduced model with two PCs. the first and fourth yield curves are shown in Figure 8(a) and are typical of the fit to the majority of the yield curves that were used to determine the PCs. However, since the proportion of variance covered by these two PCs is only 99.98% (see Table 1), the fit to some of the more extreme yield curves is not so quite so good. Yield curves 685 and 988 were identified as particularly extreme

examples, and their reproduction is shown in Figure 8(b). Note the slight discrepancy occurring between the 5 year and 10 year points for these particular yield curves.

**Figure 8** Reproduction of yield curves using a reduced model with two PCs. The majority of the yield curves are reproduced with very good accuracy, such as yield curves 1 and 4 in subfigure(a). However there are a few of the more extreme yield curves in which there is a significant deviation such as yield curves 685 and 988 from the set of 10000 yield curves from the two factor Black-Karasinski model. The PCs used here are those obtained from the log transformation presented in Figure 6(b).



## 7. Expanding the universe of yield curves

In this section we consider an example based on a simple yield curve stress that highlights one important aspect of the tradeoff between the number of PCs in the reduced model and the entire population or universe of yield curves that can be reproduced. The reproduction of out-of-sample yield curves (i.e. those that were not used in the analysis to derive the PCs) is shown to require many more PCs than we might first think. It is then shown that if we can anticipate the stresses that might commonly be applied in practice, there is a simple way to obtain PCs that expand the universe of yield curves to include these stresses. This approach is shown to reduce the number of PCs required to reproduce these stressed yield curves.

### The 100 basis point parallel stress

Consider the problem of reproducing a yield curve that has been stressed by the addition of 100 basis points uniformly across the entire term structure. To be specific, we will consider the first yield curve generated by the two factor Black-Karasinski model as shown in Figure 8(a). We use the same reduced model with two PCs as was used in the previous section, using the PCs shown in Figure 6(b). Note that this stressed yield curve was not in the original set of yield curves that the (transformed) PCA was performed on. In this regard, this stressed yield curve should be considered to be an *out-of-sample* yield curve.

Figure 9(a) shows the best fit using the reduced model with two PCs. Clearly this is a very poor reproduction. This comes about because the stressed yield curve is not in the universe of yield curves that the reduced model with two PCs can reproduce. It is then natural to consider expanding the reduced model to include another PC, in the hope that the expanded universe of yield curves includes one that closely matches the stressed yield curve we are targeting.

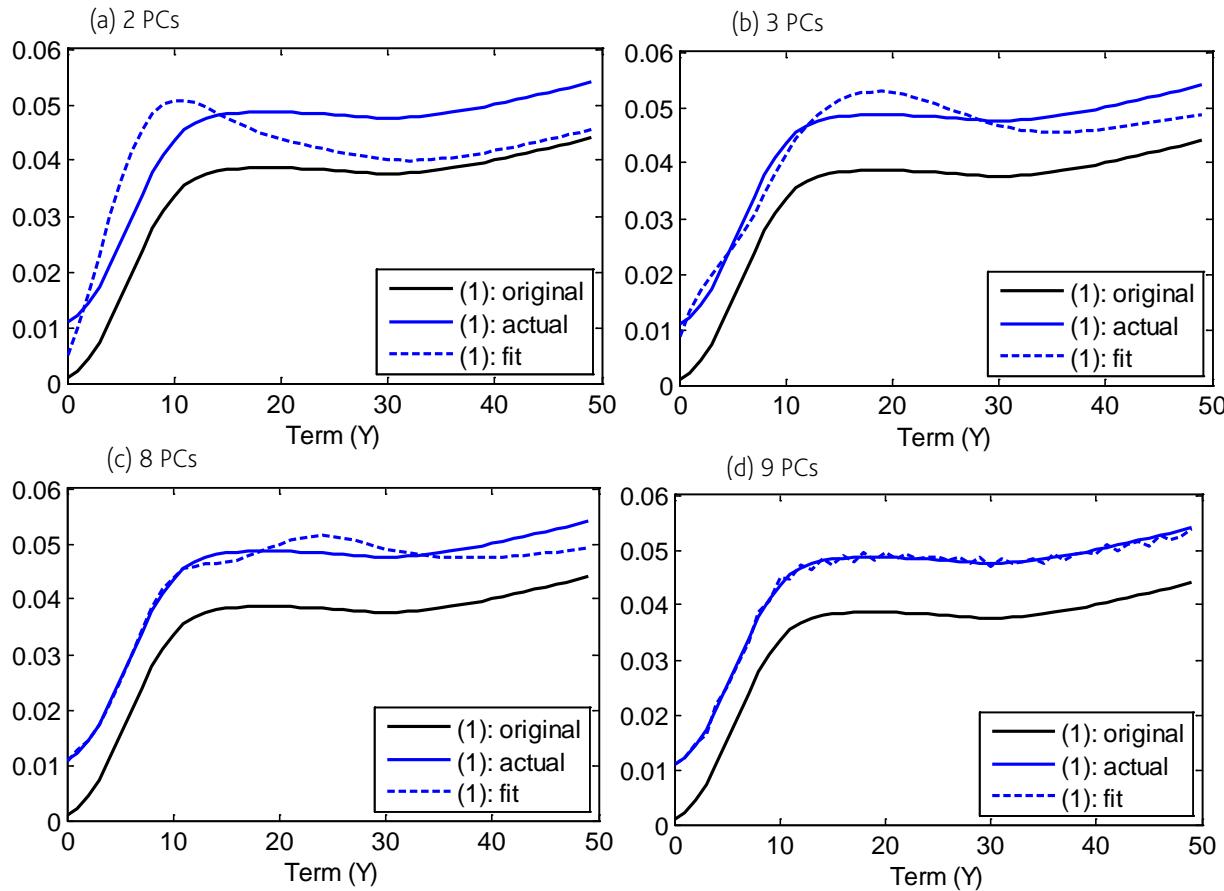
With three PCs we see a dramatic improvement to the fit, as shown in Figure 9(b), but this would still not reasonably be considered good enough for many applications. Adding more and more PCs to the reduced model progressively helps the fit, but there are still significant deviations when even eight PCs are used (Figure 9(c)). A reasonable fit across the entire term structure is only achieved when nine PCs are retained in the reduced model, as shown in Figure 9(d).

Figure 10(a) shows the magnitudes of the PCs that are required to best fit this stressed yield curve. Figure 10(b) shows the same information for 100 basis point stresses of each of the 10000 original yield curves. Clearly the majority of these stressed yield curves require nine PCs in the reduced model for good reproduction.

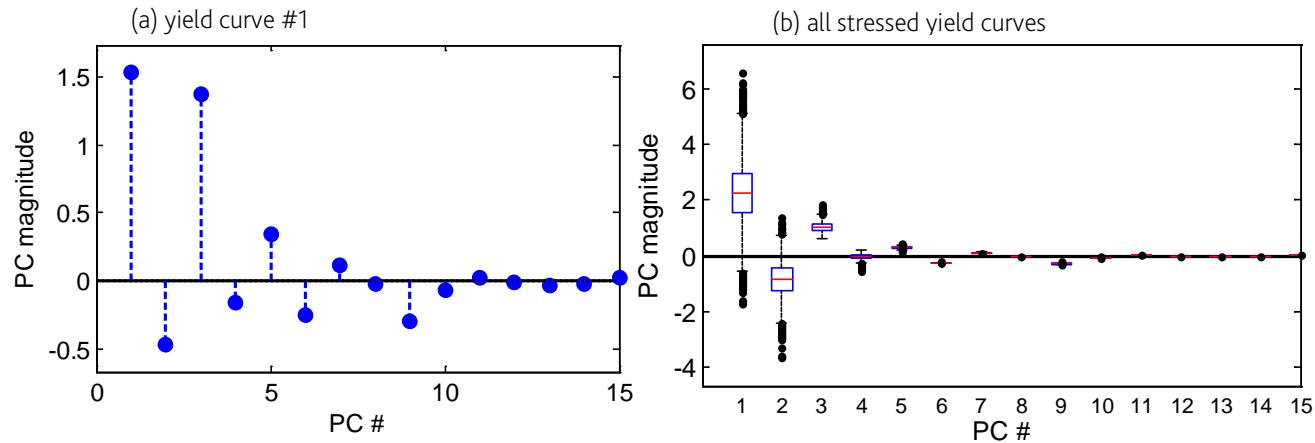
The prime reason that as many as nine PCs are required to reproduce the stressed yield curves is that the PCs were obtained via an analysis method that had no knowledge of such stresses. That is, these PCs are not fit for the purpose of reproducing yield

curves stressed in this way. One way to improve the situation might be to provide such information when the PCs are being derived.

**Figure 9** Reproduction of yield curves that have had a 100 basis point (additive) parallel stress applied, using a reduced model with different quantities of PCs. The PCs use here are those obtained from the log transformation presented in Figure 6(b).



**Figure 10** The magnitudes of the PCs that are required to best fit 100 basis point uniform stressed to the yield curves studied earlier. For the box plot in subfigure (b), the edges of the boxes signify the 25<sup>th</sup> percentile and 75<sup>th</sup> percentile of the distribution and the red central line is the median. The whiskers extend 1.5 times the interquartile range either side of the box, and outliers (marked with black dots) are any points lying outside the whiskers.



### Incorporating stresses into the reduced model

If the general nature of the perturbations or stresses that might be required in practice can be anticipated, it is possible to include this information into the PCs. Depending on the actual form of the perturbation, this could be done either:

- » treating it as being separate from the PCA, which would usually mean that it is defined in the original yield curve domain; or
- » as a part of the PCA, either in the original yield curve domain or in the transformed domain.

The simple case of a parallel stress could be dealt with either as part of the PCA or as being separate from the PCA. Treating it separate from the PCA, a simple additive shift in the reduced model will achieve the desired effect. For example, for the log transformed reduced model, might look like the following:

$$F = \exp(\mu + V^{(k)}\alpha) + \delta,$$

where  $\delta$  is a scalar. This would mean that the universe of possible yield curves would be that of the (transformed) reduced model, governed by the PCs that it retains, as well as any parallel shift of such yield curves. Determination of magnitudes of the PCs and  $\delta$  for a given yield curve might then be more difficult and might best be done via a numerical optimiser. Note also that this approach might detract from some of the advantages of the log transformation, such as the assurance that the entire yield curve cannot go negative.

In many cases, stresses might more naturally be incorporated directly into the PCA. This is as simple as applying the anticipated stresses to the original yield curves and performing PCA on the expanded set of yield curves. This does not require much more calculation when the eigenvalue decomposition approach is used since the dimensions of the covariance matrix will still just be  $n \times n$  (reflecting the granularity with which we are representing the yield curves). The PCs obtained will be able to reproduce the original yield curves as well as the stressed yield curves using an optimal number of PCs.

It turns out that the stresses need only to be defined in quite general terms. For example, the PCs obtained using a notional parallel stress of 100 basis points will serve for a reasonably wide range of parallel stress magnitudes.

Whilst we have only considered a very simple stress here, the technique for expanding the universe of yield curves can be generalised to more complex perturbations such as "short end down, long end up", "short end up, long end down", etc.

### An example

The original set of yield curves shown in Figure 1(b) have been put together with the set of corresponding yield curves stressed by 100 basis points to create a super set. PCA has been performed on this super set to obtain new PCs. These new PCs are shown in Figure 11 in both normalized and scaled form. Note that these PCs are slightly different from those obtained earlier (see Figure 6(b)). There are distinct similarities in the first two PCs, but the importance of the third PC has increased. Note that since there are now at least three PCs of significance, it is likely that there has been a sacrifice to the fits of the original yield curves. That is, it should be appreciated that two PCs will no longer be sufficient for a good reproduction of the original yield curves.

Reproduction of some of the original yield curves is shown in Figure 12, allowing us to consider the extent to which the new PCs have degraded the fits. As expected, reproduction of the typical yield curves in Figure 12(a) with two PCs is far from satisfactory. The same two typical yield curves are shown with a more reasonable fit using three PCs in Figure 12(b). Hence, we would expect that at least three of these new PCs would be required to fit the more extreme of the original yield curves. The two extreme yield curves examined earlier (Figure 8(b)) are considered again in Figure 12(c) and (d), along with fits using three and four PCs. The fit using three PCs is not so bad but might be considered unreasonable for some applications. The fit using four PCs is much better but, as in Figure 8(b), there is still a small part of the term structure over which there is a slight discrepancy.

Reproduction of the stressed yield curve considered earlier is examined in Figure 13. The fit using three PCs is not so bad but there are still small parts of the term structure over which there is a slight discrepancy. The fit using four PCs is much better. These can be compared with those shown in Figure 9, where it was seen that as many as nine PCs were required.

Widening the scope to include all of the yield curves in the super set, Figure 14 shows the PC magnitudes that are required to reproduce all of the yield curves, original or stressed. Note that there are very few yield curves requiring a significant magnitude of the fourth PC, and even less requiring a fifth. This is a reflection of the proportion of variability explained by the fourth and fifth being quite insignificant, as shown in Figure 11(b), and in stark contrast to the nine PCs required for the out-of-sample stressed yield curve examined in Figure 9 and Figure 10.

Figure 11 Principal components obtained from a super set containing the original yield curves as well as a stressed copy of the yield curves. The stress applied here is a 100bp (additive) parallel shift.

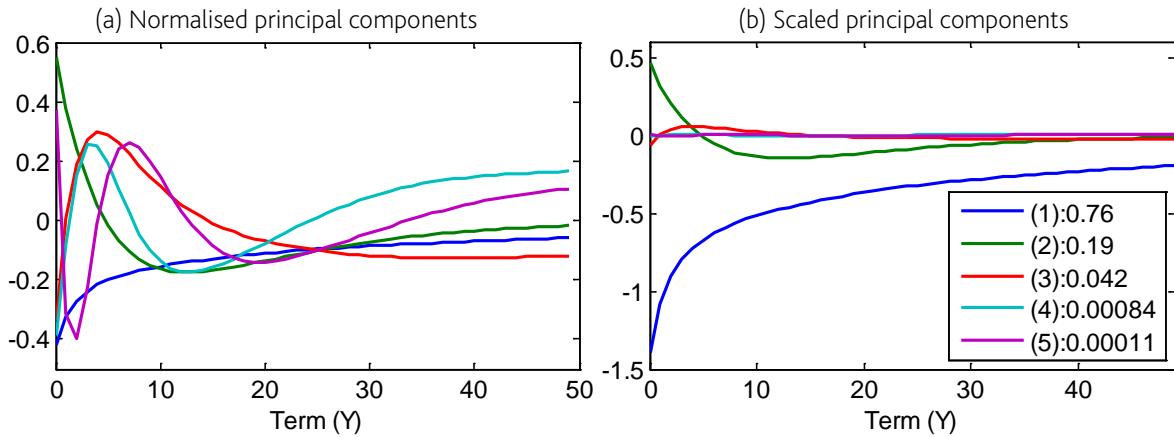


Figure 12 Reproduction of original yield curves using new PCs obtained from the super set of original and stressed yield curves. These can be compared with those shown in Figure 8.

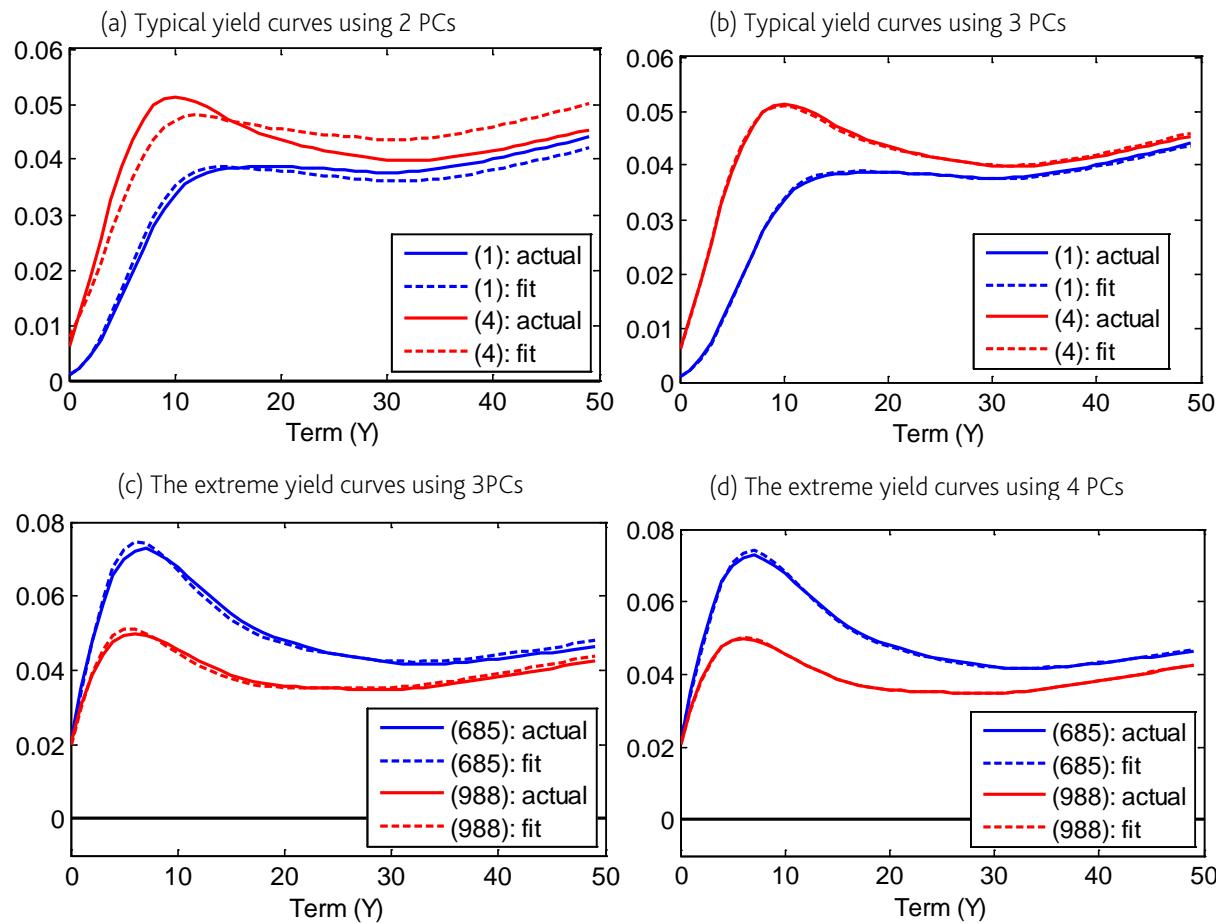


Figure 13 Reproduction of a stressed yield curve using new PCs obtained from the super set of original and stressed yield curves. These can be compared with those shown in Figure 9.

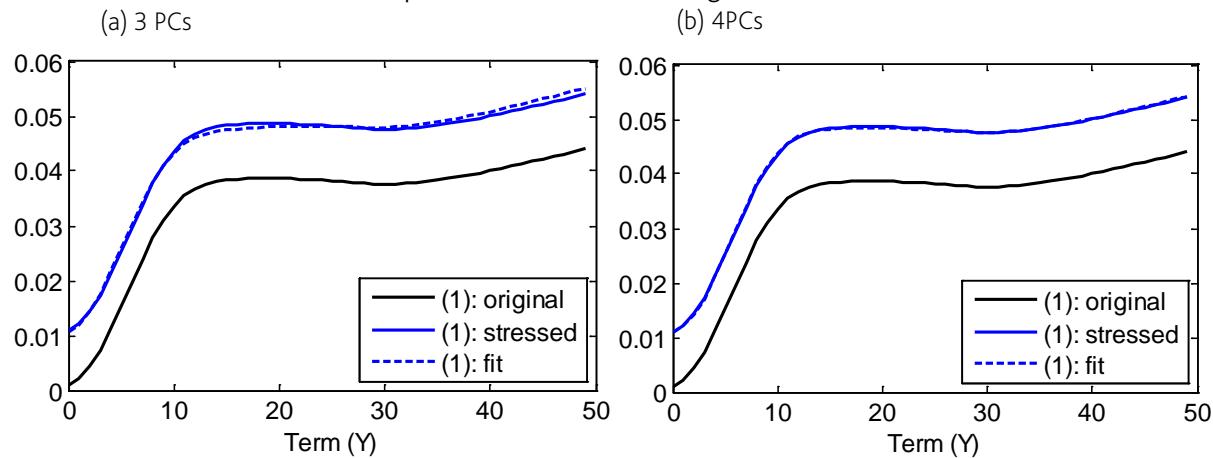
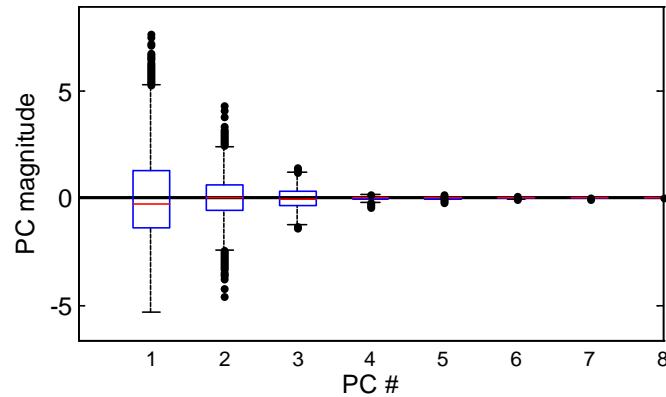


Figure 14 Box plot showing the magnitudes of the new PCs that are required to best fit each yield curve in the superset of original and stressed yield curves. For each PC, the edges of the box signify the 25<sup>th</sup> percentile and 75<sup>th</sup> percentile of the distribution and the red central line is the median. The whiskers extend 1.5 times the interquartile range either side of the box, and outliers (marked with black dots) are any points lying outside the whiskers. This can be compared to Figure 10(b).



---

## 8. Conclusions

This document has presented the basics of PCA as well as some more advanced concepts. PCA has been presented as a tool through which a full structural model can be approximated by a simpler statistical model referred to as a reduced model. We have strived to be clear about the association between PCs, the associated reduced model, and the universe of yield curves that it can reproduce.

Specific techniques for PCA have been explored including eigenvalue decomposition and singular value decomposition, as well as transformations that are useful in some cases, such as for preventing negative interest rates.

The problem of how to choose an appropriate number of PCs for a reduced model has been specifically addressed and, in the naïve use case, it has been shown to depend on the sizes of the associated eigenvalues of the covariance matrix. More correctly it depends on the desired use of the reduced model, meaning the population or universe of yield curves that we wish to be able to reproduce.

One important conclusion is that there is a direct relationship between the number of PCs retained in the reduced model and the quality of yield curve reproduction. But often the whole purpose of using PCA is to reduce the model to be as simple as possible and often there are strong disadvantages in having to deal with more PCs. This means there is a tradeoff between the number of PCs in the reduced model and the universe of yield curves that can be reproduced.

Another important conclusion is that there is a simple analytic expression for solving the inverse problem of determining the PC magnitudes that best represent a given yield curve. This means that numerical optimisation is not required, either for yield curves that were used to derive the PCs or for any out-of-sample yield curve.

The reproduction of out-of-sample yield curves (to reasonable accuracy over the entire term structure) can sometimes require many more PCs than yield curves from which the PCs were obtained. The simple case of a 100 basis point parallel (additive) stress of a yield curve was examined in detail and shown to require nine PCs for sufficient reproduction, whereas most of the in-sample yield curves only required two PCs. A few methods were proposed to lessen the burden of using many more PCs to reproduce out-of-sample yield curves. The most general technique involved anticipating the types of out-of-sample stresses/perturbations that might be required, and expanding the set of yield curves on which the PCA was performed to include stressed yield curves. In the case of the 100 basis point additive stress, this approach was shown to reduce the number of PCs required to reproduce the number of PCs from nine down to just three or four, depending on the accuracy required.

In summary, it is not always safe to assume that a two-factor interest rate model means that only two PCs are required or that a three-factor interest rate model means that only three PCs are required. The actual number of PCs required depends on the desired usage of the reduced model in terms of what yield curves can be reproduced. The danger is that two or three PCs are used in a model and people rely on these being sufficient to accurately model yield curve stresses, such as a 100bps parallel shift.

## Appendix A Eigenvalues, eigenvectors and the eigenvalue decomposition

If  $A$  is a square matrix, the eigenvectors of  $A$  are the non-zero vectors  $v$  that satisfy:

$$Av = \lambda v,$$

where  $\lambda$  is a scalar associated with  $v$ . That is, matrix multiplication of the eigenvector is equivalent to a scalar multiplication, with the scale factor being called the eigenvalue. An eigenvector corresponds to a particular eigenvalue, and the pair are specific to the matrix  $A$ . If  $A$  is a  $n \times n$  matrix then it has  $n$  eigenvectors (and corresponding eigenvalues) and these eigenvectors are orthonormal, meaning that they are orthogonal and have unit length. In the special case in which the matrix  $A$  is real and symmetric, the eigenvalues are all real.

Augmenting the  $n$  eigenvectors into a column matrix,  $V$ , and putting the corresponding eigenvalues on the diagonal of a diagonal matrix  $\Lambda$  we get the following matrix equation:

$$AV = V\Lambda$$

The orthogonality of the eigenvectors means that  $V^T V = VV^T = I$  and hence  $V^{-1} = V^T$ . When  $V$  is invertible this leads to the eigenvalue decomposition:

$$A = V\Lambda V^T.$$

This can be expanded to express the matrix  $A$  as a summation as follows:

$$A = \sum_{i=1}^n \lambda_i v_i v_i^T$$

This summation proves to be particularly useful representation in the context of PCA as it clearly indicates that the effect of eigenvectors diminishes with the size of the eigenvalue. The basis of PCA is that the matrix  $A$  can be approximated using only those eigenvectors corresponding to significantly large eigenvalues. In some situations this can mean that just a small number of eigenvectors can be used to represent a system with many tens or hundreds of dimensions.

In terms of interpretation, the eigenvector matrix  $V$  represents a transformation onto orthogonal dimensions that explain maximum variability. That is, the eigenvector,  $v_1$ , corresponding to the largest eigenvalue,  $\lambda_1$ , will point in the direction of greatest variability. The second eigenvector,  $v_2$ , will point in the direction of greatest variability subject to the condition that it must be orthogonal to the  $v_1$ . The third eigenvector,  $v_3$ , will point in the direction of greatest variability subject to being orthogonal to both  $v_1$  and  $v_2$ , and so on. This is explained in more detail in the simple example below.

This transformation (change of basis, or change of coordinates) so that all the points are represented as a linear combination of the eigenvectors  $v_1$ ,  $v_2$  and  $v_3$  can be performed by (left) multiplying the data points by  $V^T$ . That is, if an existing point is  $x = [x_1, x_2, \dots, x_n]^T$  then its representation in the new coordinate system is  $\bar{x} = V^T x$ . The values in this new coordinate system are the scores. That is, they are the magnitudes of the eigenvectors that reproduce the original data.

In matrix form this is

$$\bar{X} = V^T X$$

The elements of  $\bar{x}$  (in the new coordinate system) are independent. The covariance matrix is  $\Lambda$  and hence is diagonal. This means that all covariances are zero and the variance of each variable is  $\lambda_i$ . The total variance is the sum of the eigenvalues and this is also the total variance of the original data.

### An example

Consider the simple example of three variables with normally distributed random samples that are correlated according to the following correlation matrix:

$$C = \begin{bmatrix} 1 & 0.7 & 0.3 \\ 0.7 & 1 & 0 \\ 0.3 & 0 & 1 \end{bmatrix}.$$

The correlation of the variables means that there will be greater variability in some three-dimensional directions, but these dimensions will not generally align with the axes or other convenient directions.

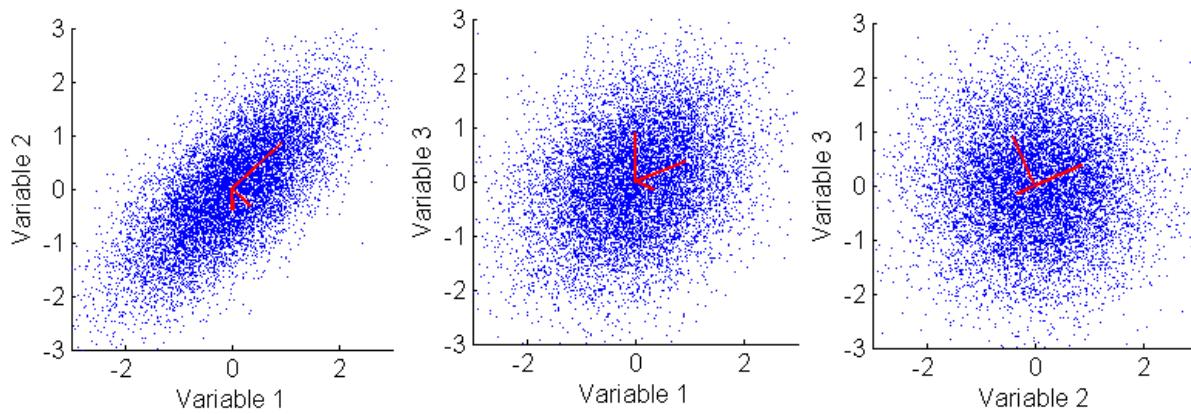
The eigenvalue decomposition of  $C$  gives the following (to three decimal places):

$$V = [v_1 \ v_2 \ v_3] = \begin{bmatrix} 0.707 & 0 & -0.707 \\ 0.650 & -0.394 & 0.650 \\ 0.279 & 0.919 & 0.279 \end{bmatrix} \quad \text{and} \quad \Lambda = \begin{bmatrix} 1.762 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0.238 \end{bmatrix}.$$

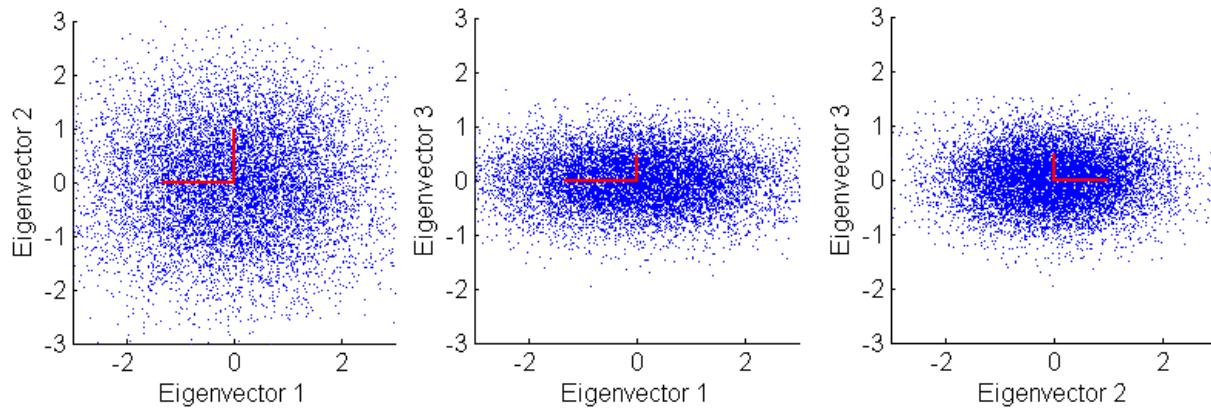
Scatter plots of the points are shown in Figure 15 in the original coordinate system. That is, using regular Cartesian coordinates as the basis:  $[1,0,0]^T$ ,  $[0,1,0]^T$  and  $[0,0,1]^T$ . Each point is represented as a linear combination of these orthonormal basis vectors. Also shown in Figure 15 (in red) are the eigenvectors scaled by the square root of the corresponding eigenvalues. The largest of these eigenvectors points in the direction of maximum variability, which is the eigenvector corresponding to the largest eigenvalue. This is the direction  $v_1 = [0.707, 0.650, 0.279]^T$ . This can be seen as the longest red line in the scatter plot of variable 1 versus variable 2, as these are the most highly correlated pair of variables. This same direction (as well as the other directions) is less obvious in the other subfigures as the eigenvectors do not align nicely with the axes.

Applying a change of basis (change of coordinate system) so that all the points are represented as a linear combination of the eigenvectors  $v_1$ ,  $v_2$  and  $v_3$  can be performed by (left) multiplying the data points by  $V^T$ . The points in the new coordinate system are plotted in 0. This gives us a clearer view of the system as the direction of greatest variability is (by definition) aligned with the axes. The eigenvectors of the data points in the new coordinate system are also shown in 0 in red. The eigenvector corresponding to the largest eigenvalue is in the direction of the "Eigenvector 1" axis. This gives the direction of greatest variability. The direction of the eigenvector corresponding to the second largest eigenvalue points in the direction of greatest variability subject to the restriction that it must be orthogonal to  $v_1$ . This is the axis denoted "Eigenvector 2". Finally, the direction corresponding to the last eigenvector is completely defined by the need to be orthogonal to both of the other eigenvectors.

**Figure 15** Scatter plots in terms of the points in axes of the original variables. Also shown in red is indication of the eigenvectors for the covariance matrix of the points in this coordinate system, which are scaled by the square root of the corresponding eigenvalue. Note that the eigenvectors point in the direction of largest variability. This can be observed in the scatter plot between variables 1 and 2 (as these are the most highly correlated), but is less clear in the other subfigures because of the three-dimensional nature of the system.



**Figure 16** The same points as in Figure 15, but transformed into coordinates of the eigenvectors of the covariance matrix of the original points. In this coordinate system the points represent the magnitudes of the eigenvalues that are required to reproduce the original points. These are sometimes known as the “scores”. Also shown in red is an indication of the eigenvectors for the covariance matrix of the points in this new coordinate system, which are scaled by the square root of the corresponding eigenvalue. Note that these eigenvectors always point in the direction (positive or negative) of the axes, confirming that these new axes are independent.



## Appendix B Singular value decomposition

SVD reduces a real  $m \times n$  matrix  $M$  to three components as follows:

$$M = USV^T.$$

Here,  $U$  is an  $m \times n$  orthonormal matrix,  $S$  is an  $n \times n$  diagonal matrix containing the “singular values”, and  $V$  is an  $n \times n$  orthonormal matrix.

The columns of the matrix  $U$  are eigenvalues of the matrix  $MM^T$ , and the columns of the matrix  $V$  are eigenvectors of the matrix  $M^TM$ . The singular values (diagonal entries of the matrix  $S$ ) are all non-negative, being the positive square root of the eigenvalues of the matrix  $M^TM$ .

The matrix produced by the multiplication  $US$  is known as the scores matrix and represents the magnitudes of the columns of  $V$  that produce the raw data in  $M$ .

In theory, SVD is a very stable decomposition as it is always perfectly conditioned. Hence it is often used instead of eigenvalue decomposition for ill conditioned or rank deficient matrices. Note that correlation matrices can be very highly ill-conditioned.

## Appendix C Link between SVD and eigenvalue decomposition

If the original data matrix,  $M$ , has mean zero, then the pairwise covariance matrix,  $C$ , is

$$C = \frac{1}{n-1} M^T M,$$

which is a symmetric  $n \times n$  matrix. Using the SVD of  $M$  we have

$$M^T M = (USV^T)^T (USV^T) = VS^T U^T USV^T = VS^2 V^T,$$

and hence

$$C = \frac{1}{n-1} VS^2 V^T.$$

Directly applying eigenvalue decomposition to the correlation matrix gives

$$\mathcal{C} = V\Lambda V^T.$$

Equating these two decompositions of  $\mathcal{C}$  means that:

$$\frac{1}{n-1}VS^2V^T = V\Lambda V^T.$$

Since the matrix  $V$  is the same in each case, we obtain an equation linking the eigenvalues and singular values:

$$\Lambda = \frac{1}{n-1}S^2.$$

As both  $\Lambda$  and  $S$  are diagonal, this equation also holds elementwise.

### Application to a real symmetric matrix

Now consider the application of SVD to a real symmetric matrix,  $A$ , such as a covariance or correlation matrix. This means that

$$A = USV^T.$$

Since  $A$  is symmetric, then  $USV^T$  must also be symmetric. This means that  $|U_{ij}| = |V_{ij}|$ , meaning that the elements can only differ by a sign change. Relating this to the eigenvalue decomposition, this also implies that the eigenvalues and singular values can also only differ by a sign change. It turns out that the differences in sign are associated with negative eigenvalues and so  $U = V$  whenever  $A$  is positive semidefinite. In this case there is no difference between SVD and eigenvalue decomposition apart from arbitrary sign changes in the eigenvectors, and numerical accuracy effects when the matrix is ill conditioned.

For any singular value corresponding to a negative eigenvalue, the corresponding column of  $U$  is multiplied through by -1. This is basically moving the negative from the eigenvalue to the corresponding column of  $U$ , so as to make the singular value positive. When SVD is used, negative eigenvalues can be detected by looking for differences in sign between columns of  $U$  and  $V$ .

For the purposes of PCA either method can be used. They both produce the same PCs (eigenvectors) up to a sign change, and they both produce the same eigenvalues/singular values up to a sign change.

---

## Acknowledgements

The authors wish to acknowledge the helpful comments and suggestions of Moody's Analytics colleagues Sandy Sharp, Brian Robinson, Nick Jessop, and Harry Hibbert.

---

## Further Reading

For additional details, please read the following resources.

- » Press, Teukolsky, Vetterling and Flannery: "Numerical Recipes in Fortran 77, The Art of Scientific Computing" (2nd ed.) Cambridge University Press, 1992.

© 2014 Moody's Corporation, Moody's Investors Service, Inc., Moody's Analytics, Inc. and/or their licensors and affiliates (collectively, "MOODY'S"). All rights reserved.

CREDIT RATINGS ISSUED BY MOODY'S INVESTORS SERVICE, INC. ("MIS") AND ITS AFFILIATES ARE MOODY'S CURRENT OPINIONS OF THE RELATIVE FUTURE CREDIT RISK OF ENTITIES, CREDIT COMMITMENTS, OR DEBT OR DEBT-LIKE SECURITIES, AND CREDIT RATINGS AND RESEARCH PUBLICATIONS PUBLISHED BY MOODY'S ("MOODY'S PUBLICATIONS") MAY INCLUDE MOODY'S CURRENT OPINIONS OF THE RELATIVE FUTURE CREDIT RISK OF ENTITIES, CREDIT COMMITMENTS, OR DEBT OR DEBT-LIKE SECURITIES. MOODY'S DEFINES CREDIT RISK AS THE RISK THAT AN ENTITY MAY NOT MEET ITS CONTRACTUAL, FINANCIAL OBLIGATIONS AS THEY COME DUE AND ANY ESTIMATED FINANCIAL LOSS IN THE EVENT OF DEFAULT. CREDIT RATINGS DO NOT ADDRESS ANY OTHER RISK, INCLUDING BUT NOT LIMITED TO: LIQUIDITY RISK, MARKET VALUE RISK, OR PRICE VOLATILITY. CREDIT RATINGS AND MOODY'S OPINIONS INCLUDED IN MOODY'S PUBLICATIONS ARE NOT STATEMENTS OF CURRENT OR HISTORICAL FACT. MOODY'S PUBLICATIONS MAY ALSO INCLUDE QUANTITATIVE MODEL-BASED ESTIMATES OF CREDIT RISK AND RELATED OPINIONS OR COMMENTARY PUBLISHED BY MOODY'S ANALYTICS, INC. CREDIT RATINGS AND MOODY'S PUBLICATIONS DO NOT CONSTITUTE OR PROVIDE INVESTMENT OR FINANCIAL ADVICE, AND CREDIT RATINGS AND MOODY'S PUBLICATIONS ARE NOT AND DO NOT PROVIDE RECOMMENDATIONS TO PURCHASE, SELL, OR HOLD PARTICULAR SECURITIES. NEITHER CREDIT RATINGS NOR MOODY'S PUBLICATIONS COMMENT ON THE SUITABILITY OF AN INVESTMENT FOR ANY PARTICULAR INVESTOR. MOODY'S ISSUES ITS CREDIT RATINGS AND PUBLISHES MOODY'S PUBLICATIONS WITH THE EXPECTATION AND UNDERSTANDING THAT EACH INVESTOR WILL, WITH DUE CARE, MAKE ITS OWN STUDY AND EVALUATION OF EACH SECURITY THAT IS UNDER CONSIDERATION FOR PURCHASE, HOLDING, OR SALE.

MOODY'S CREDIT RATINGS AND MOODY'S PUBLICATIONS ARE NOT INTENDED FOR USE BY RETAIL INVESTORS AND IT WOULD BE RECKLESS FOR RETAIL INVESTORS TO CONSIDER MOODY'S CREDIT RATINGS OR MOODY'S PUBLICATIONS IN MAKING ANY INVESTMENT DECISION. IF IN DOUBT YOU SHOULD CONTACT YOUR FINANCIAL OR OTHER PROFESSIONAL ADVISER.

ALL INFORMATION CONTAINED HEREIN IS PROTECTED BY LAW, INCLUDING BUT NOT LIMITED TO, COPYRIGHT LAW, AND NONE OF SUCH INFORMATION MAY BE COPIED OR OTHERWISE REPRODUCED, REPACKAGED, FURTHER TRANSMITTED, TRANSFERRED, DISSEMINATED, REDISTRIBUTED OR RESOLD, OR STORED FOR SUBSEQUENT USE FOR ANY SUCH PURPOSE, IN WHOLE OR IN PART, IN ANY FORM OR MANNER OR BY ANY MEANS WHATSOEVER, BY ANY PERSON WITHOUT MOODY'S PRIOR WRITTEN CONSENT.

All information contained herein is obtained by MOODY'S from sources believed by it to be accurate and reliable. Because of the possibility of human or mechanical error as well as other factors, however, all information contained herein is provided "AS IS" without warranty of any kind. MOODY'S adopts all necessary measures so that the information it uses in assigning a credit rating is of sufficient quality and from sources MOODY'S considers to be reliable including, when appropriate, independent third-party sources. However, MOODY'S is not an auditor and cannot in every instance independently verify or validate information received in the rating process or in preparing the Moody's Publications.

To the extent permitted by law, MOODY'S and its directors, officers, employees, agents, representatives, licensors and suppliers disclaim liability to any person or entity for any indirect, special, consequential, or incidental losses or damages whatsoever arising from or in connection with the information contained herein or the use of or inability to use any such information, even if MOODY'S or any of its directors, officers, employees, agents, representatives, licensors or suppliers is advised in advance of the possibility of such losses or damages, including but not limited to: (a) any loss of present or prospective profits or (b) any loss or damage arising where the relevant financial instrument is not the subject of a particular credit rating assigned by MOODY'S.

To the extent permitted by law, MOODY'S and its directors, officers, employees, agents, representatives, licensors and suppliers disclaim liability for any direct or compensatory losses or damages caused to any person or entity, including but not limited to by any negligence (but excluding fraud, willful misconduct or any other type of liability that, for the avoidance of doubt, by law cannot be excluded) on the part of, or any contingency within or beyond the control of, MOODY'S or any of its directors, officers, employees, agents, representatives, licensors or suppliers, arising from or in connection with the information contained herein or the use of or inability to use any such information.

NO WARRANTY, EXPRESS OR IMPLIED, AS TO THE ACCURACY, TIMELINESS, COMPLETENESS, MERCHANTABILITY OR FITNESS FOR ANY PARTICULAR PURPOSE OF ANY SUCH RATING OR OTHER OPINION OR INFORMATION IS GIVEN OR MADE BY MOODY'S IN ANY FORM OR MANNER WHATSOEVER.

MIS, a wholly-owned credit rating agency subsidiary of Moody's Corporation ("MCO"), hereby discloses that most issuers of debt securities (including corporate and municipal bonds, debentures, notes and commercial paper) and preferred stock rated by MIS have, prior to assignment of any rating, agreed to pay to MIS for appraisal and rating services rendered by it fees ranging from \$1,500 to approximately \$2,500,000. MCO and MIS also maintain policies and procedures to address the independence of MIS's ratings and rating processes. Information regarding certain affiliations that may exist between directors of MCO and rated entities, and between entities who hold ratings from MIS and have also publicly reported to the SEC an ownership interest in MCO of more than 5%, is posted annually at [www.moodys.com](http://www.moodys.com) under the heading "Shareholder Relations — Corporate Governance — Director and Shareholder Affiliation Policy."

For Australia only: Any publication into Australia of this document is pursuant to the Australian Financial Services License of MOODY'S affiliate, Moody's Investors Service Pty Limited ABN 61 003 399 657 AFSL 336969 and/or Moody's Analytics Australia Pty Ltd ABN 94 105 136 972 AFSL 383569 (as applicable). This document is intended to be provided only to "wholesale clients" within the meaning of section 761G of the Corporations Act 2001. By continuing to access this document from within Australia, you represent to MOODY'S that you are, or are accessing the document as a representative of, a "wholesale client" and that neither you nor the entity you represent will directly or indirectly disseminate this document or its contents to "retail clients" within the meaning of section 761G of the Corporations Act 2001. MOODY'S credit rating is an opinion as to the creditworthiness of a debt obligation of the issuer, not on the equity securities of the issuer or any form of security that is available to retail clients. It would be dangerous for "retail clients" to make any investment decision based on MOODY'S credit rating. If in doubt you should contact your financial or other professional adviser.