Components of Yield Curve Movements
Illustrative Worked Examples

An Interim Report of
The Stress Test Working Party

Finance, Investment and
Risk Management Convention

June 2008

Working party members:
Ralph Frankland (Chairman)
Enrico Biffis
David Dullaway
Seth Eshun
Andy Holtham
Andrew D Smith
Elliot Varnell
Tim Wilkins.

Summary

This working paper is an extract from a larger forthcoming paper from the Stress Tests working party. We may modify this material prior to final publication, and we welcome readers' suggested improvements.

The yield curve on a particular date describes variations in interest rates as a function of the term for which a deposit is committed. Movements in the yield curve from one period to the next are an important driver of profits and losses for most financial institutions.

So-called "full models" of yield curves, treat every point of the curve as a random variable in its own right. However, most firms prefer to use "component models". For example, the first component might be the level of the yield curve; the second component might be its slope and the third component its curvature. A firm's sensitivities to changes in each of these components are combined to assess the firm's resilience to yield curve movements.

This note considers several ways to decompose a full model into components, including polynomials, principal components and variance matching.

1. Introduction

1.1. Variables Modelled

We consider the movement in the yield curve from now until a point in 1 year's time. We denote this movement is denoted by X, a random vector. For the purposes of this note, it does not matter whether these are spot yields, par yields or forward yields.

We use a deliberately unsophisticated model, so we can easily interpret outputs and so readers can readily replicate our results. Suppose X has zero mean, so the expected yield curve in a year's time is exactly where it is now. We suppose X is modelled at certain *key maturities*, these being maturities: {1,2,3,5,10,20}.

1.2. Variance-Covariance Matrices

At each key rate, we assume the standard deviation of X is 1%. The correlation matrix between yields at different maturities is assumed to take the following simple form:

Correlations	t=1	t=2	t=3	t=5	t=10	t=20
t=1	1	0.9	0.8	0.7	0.6	0.5
t=2	0.9	1	0.9	0.8	0.7	0.6
t=3	0.8	0.9	1	0.9	0.8	0.7
t=5	0.7	0.8	0.9	1	0.9	0.8
t=10	0.6	0.7	0.8	0.9	1	0.9
t=20	0.5	0.6	0.7	0.8	0.9	1

The standard deviations are all 1%. Therefore, measured in percentage terms, the covariance matrix is the same as the correlation matrix, which is convenient for our example. We denote this matrix by V.

The appropriate assumptions to use for a particular yield curve, a particular financial entity and a particular point in time are all subject to debate. Our assumptions for the worked example are not derived from any particular data set, but they are broadly representative of moves in yields for developed economies in recent times. By choosing simple assumptions we aim to make it easy for anyone else to verify our numbers.

1.3. What are components?

Sometimes we want to break down a yield curve into components. Each component is independent (or at least uncorrelated with) the other components. Traditionally, the components are ordered so that the first represents the level of the curve, the second component the slope and the third component the curvature.

It is natural to construct components as polynomials. To avoid explosions at infinity, we do not use polynomials in t itself, but in A^t for some $0<A<1$. For example, taking A

= 0.8, we find the following polynomial components. The nth component, that is the coefficient of Z_n in the n^{th} column, is a polynomial of order n-1 in A^t.

Polynomial Components						
	Z1	Z2	Z3	Z4	Z5	Z6
t=1	0.866025	-0.482214	0.017380	0.065366	-0.113533	-0.002384
t=2	0.866025	-0.286508	-0.284323	-0.159815	-0.238404	-0.068529
t=3	0.866025	-0.129943	-0.417666	-0.067763	0.003679	0.232518
t=5	0.866025	0.095510	-0.440996	0.196032	0.080885	-0.037812
t=10	0.866025	0.364980	-0.207677	0.201473	-0.147034	0.107004
t=20	0.866025	0.482214	-0.017380	-0.065366	0.113533	0.002384

In our examples, components are multiplied by standard normal variables Z_1, Z_2, …. Z_6, then added together to produce the yield curve shift. We will verify that this replicates the desired correlation matrix.

Polynomials are not the only approach to decompositions. Another approach is to use *principal component analysis*, or PCA (Anderson, 1957), devised so that the early components (Z_1 and Z_2 for example) explain as much as possible of the variability in the rates, minimising the role of later components Z_3 to Z_6. For example, taking the first two components of the yield at t=1, the variance of the modelled yield is the sum of the squares of the first two elements of the table's first row. If we are interested not only in convergence at t=1 but for all t, then we might add together the corresponding variances for each of the six modelled time points. PCA then equates to maximising the sum of the squares of elements in the first two columns. The table shows principal components of what we will later call "Model 6". We will later discuss a number of related PCA algorithms.

Principal Components: Model 6						
	Z1	Z2	Z3	Z4	Z5	Z6
t=1	0.833278	-0.481806	0.222651	0.129099	0.077942	0.034592
t=2	0.909996	-0.352706	0.014383	-0.129099	-0.147311	-0.094507
t=3	0.949163	-0.129099	-0.209257	-0.129099	0.072806	0.129099
t=5	0.949163	0.129099	-0.209257	0.129099	0.072806	-0.129099
t=10	0.909996	0.352706	0.014383	0.129099	-0.147311	0.094507
t=20	0.833278	0.481806	0.222651	-0.129099	0.077942	-0.034592

It is often helpful to construct models with a small number of components. In that case, we can only approximate the desired correlation matrix. One technique for constructing low component models is variance matching. The table shows a three-component example of variance-matching:

Variance matching components			
	Z1	Z2	Z3
t=1	0.974679	-0.223607	0.000000
t=2	0.974679	0.223607	0.000000
t=3	0.872082	0.223607	-0.435286
t=5	0.872082	-0.223607	-0.435286
t=10	0.974679	-0.223607	0.000000
t=20	0.974679	0.223607	0.000000

1.4. Verifying Components

How can we verify a model re-expressed as components? We have expressed our models in the form:

$X = BZ$

Here, X is the vector of yields varying by maturity. Z is a vector of independent normal variables with mean zero and unit variance. The matrix B holds the components, with each column of B corresponding to one component. The element B_{ij} is the value of the j^{th} component evaluated at the i^{th} key maturity.

From standard matrix theory (see, for example, the book by Anderson), the variance-covariance matrix of X is BB^T, where a superscript T denotes a matrix transpose. We ideally want the decomposition to reproduce the variance-covariance matrix of X. This means that

$BB^T = V$

It is easy to verify numerically that this is satisfied exactly for the polynomial components and for the principal components tabulated above. As we have demonstrated two different decompositions, we can deduce that a decomposition is not necessarily unique. This raises the question of whether one decomposition is "better" than other.

On the other hand, if we constrain the number of columns of B, it may not be possible to find any matrix B to satisfy $BB^T = V$. In that case, we want to find B so that BB^T is as close as possible to V. For example, using the variance matching components, we find the following implied correlation matrix:

Correlation matrix implied by Variance Matching Components						
	t=1	t=2	t=3	t=5	t=10	t=20
t=1	1	0.9	0.8	0.9	1	0.9
t=2	0.9	1	0.9	0.8	0.9	1
t=3	0.8	0.9	1	0.9	0.8	0.9
t=5	0.9	0.8	0.9	1	0.9	0.8
t=10	1	0.9	0.8	0.9	1	0.9
t=20	0.9	1	0.9	0.8	0.9	1

We can see that the shaded elements: the main diagonal and two above and below it, are replicated exactly. However, the variance matching components overstate the elements on the bottom left and top right of the matrix. In particular, the variance matching components force the rates at t=1 and t=10 to be equal. It also forces equality between rates at t=2 and t=20. This is an undesirable side-effect of using only three components when full replication of the correlation matrix requires six components.

The remainder of this paper examines each of these decomposition tools in more detail.

2. Cholesky Methods

2.1. Cholesky decomposition

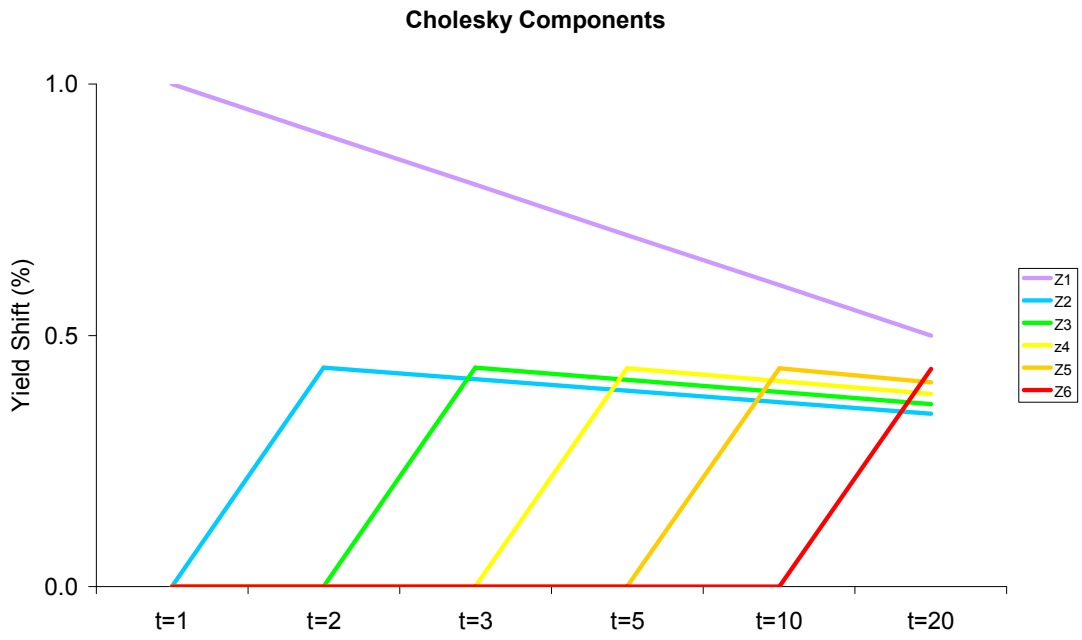
A *Cholesky decomposition* is the text-book solution to finding a matrix B such that $BB^T = V$. The decomposition works provided that V is symmetric and positive semi-definite. These are precisely the conditions that V is a valid variance-covariance matrix.

The equation $BB^T = V$ does not determine the matrix B uniquely. The Cholesky method produces a matrix B which is *lower triangular*, that is, so that all elements above and to the right of the main diagonal are zero. It also ensures that the diagonal elements of B are non-negative. Within these constraints, B is uniquely specified.

The table shows a Cholesky decomposition for our correlation matrix:

Cholesky Decomposition						
	Z1	Z2	Z3	Z4	Z5	Z6
t=1	1.0	0	0	0	0	0
t=2	0.9	0.435890	0	0	0	0
t=3	0.8	0.412948	0.435286	0	0	0
t=5	0.7	0.390007	0.411103	0.434613	0	0
t=10	0.6	0.367065	0.386921	0.409048	0.433861	0
t=20	0.5	0.344124	0.362738	0.383482	0.406745	0.433013

We can see that the first component captures the first column of the correlation matrix. The second component starts at zero for t=1. This is because the first component already explains the yield shift at t=1. And so the pattern continues. The chart shows the components:



Although these components do reproduce the desired correlation matrix, they do not appear natural or intuitive. For example, if we recomputed the Cholesky calculation with the time points in reverse order, starting from t=20, the components look different – in fact, given the symmetry of the correlation matrix, reversing the time axis has the effect of a vertical reflection of the whole chart.

This lack of intuition is not a problem for some applications. For example, Cholesky decomposition is the preferred algorithm in Monte Carlo work, on account of its simplicity. On the other hand, for communicating capital requirements, more intuitive decompositions are helpful. We now develop some of these.

2.2. Solution Rotation

Given the multiplicity of solutions B to the equation $BB^T = V$, we can ask how the different solutions are related.

The answer lies in rotations. We have the model $X = BZ$. Here, the random vector Z consists of independent identically distributed N(0,1) variables. Contours of equal density are concentric hyper-spheres around the origin.

Rotations can be written as matrix multiplication, by some matrix Ω. A matrix Ω corresponds to a rotation if and only if $\Omega^{-1} = \Omega^T$. If Z is a vector of independent identically distributed N(0,1) variables, then so is ΩZ. Thus, a model $X = BZ$ produces the same distribution for X as a model $B\Omega Z$. This means that if B is one decomposition, then $B\Omega$ is another.

In our example, we can convert a Cholesky decomposition to polynomial components using the following rotation matrix:

Rotation to convert Cholesky Decomposition to Polynomial Components						
	original Z1	original Z2	original Z3	original Z4	original Z5	original Z6
New Z1	0.866025	-0.482214	0.017380	0.065366	-0.113533	-0.002384
New Z2	0.198680	0.338353	-0.688167	-0.501604	-0.312521	-0.152294
New Z3	0.209427	0.266734	-0.338611	0.200056	0.513594	0.683033
New Z4	0.221404	0.440493	-0.104847	0.606656	0.163600	-0.592584
New Z5	0.234834	0.568667	0.480340	0.047983	-0.529753	0.328335
New Z6	0.250000	0.253819	0.412004	-0.577723	0.564144	-0.226510

2.3. *Orthonormal Polynomials*

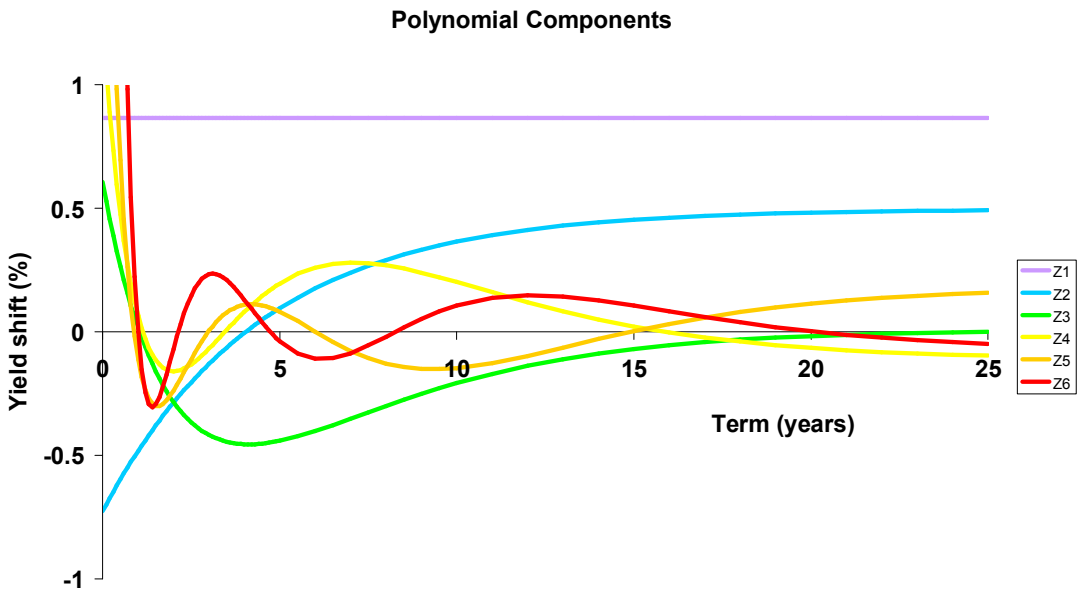
Until now, our analysis has treated the yields at $t = \{1,2,3,5,10,20\}$ as six distinct random variables. We have not used the associated time values in our models.

However, it is seen empirically that yield curves are often smooth functions of time t. This suggests we look to smooth sample functions to build up components. One possible choice is the family of polynomials in t. A disadvantages of polynomials is their tendency to infinity for large t, while yield curves in practice tend to flatten out. A solution to this problem is use polynomials, not in t itself, but in A^t for some $0<A<1$. In our examples, we have selected $A = 0.8$.

The table shows the coefficients of the polynomial decomposition:

Polynomial Coefficients						
	Z1	Z2	Z3	Z4	Z5	Z6
1	0.8660	0.4963	0.0091	-0.1128	0.1816	-0.0776
A^t	0	-1.2232	-2.3339	4.2749	-6.3338	7.9003
A^{2t}	0	0	2.9302	-13.7102	37.9863	-87.5140
A^{3t}	0	0	0	10.8064	-74.8113	328.9030
A^{4t}	0	0	0	0	45.8107	-493.1470
A^{5t}	0	0	0	0	0	254.3899

The upper diagonal form gives a hint that these coefficients are also obtained from a form of Cholesky decomposition. One advantage of using polynomials is that they interpolate naturally for other values of t. The chart shows the polynomial components for a range of t values. We see that successive components become smaller but also more wiggly – a well-known feature of higher order polynomials.



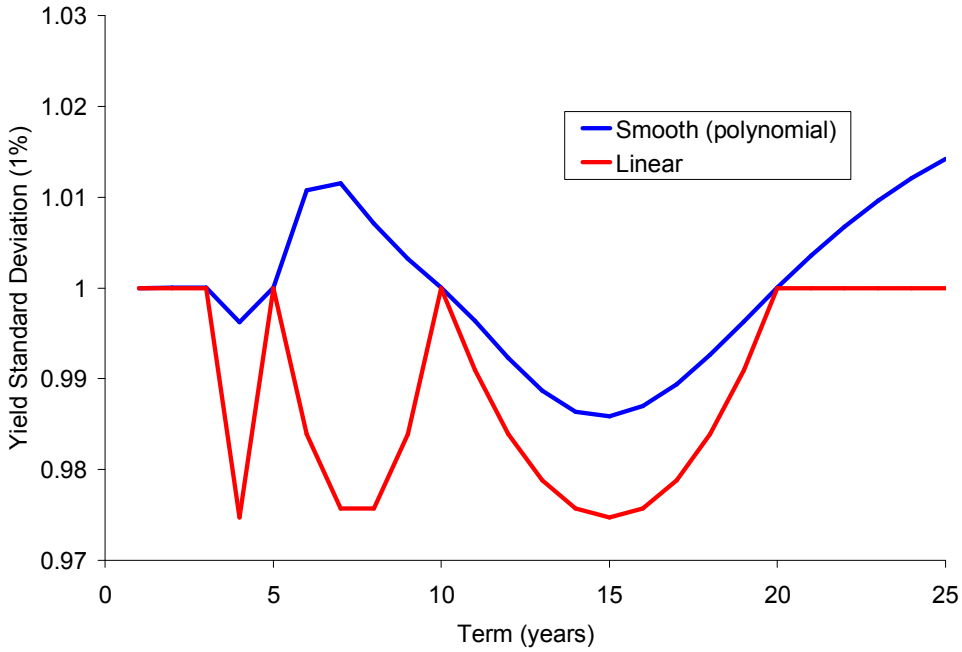
The behaviour for t <1 is a possible cause for concern. It reflects the tendency of polynomials to wiggle. Our choice of A = 0.8 produces only minor wiggle problems with our chosen correlation structure. An informal test of examining charts by eye, suggests that most other choices of A produce more pronounced wiggles, which is how we chose A = 0.8 in this example.

2.4. Interest Rate Standard Deviation.

We have described a yield curve model using polynomials to fit yield curves. Our original model had yield standard deviations of 1% at all terms. However, these standard deviations do not necessarily apply at other, interpolated terms. The chart shows the interest rate standard deviation by term for our polynomial model. Note the vertical axis – the standard deviations are close to 1, in fact, as near to 1 as makes little practical difference.

Polynomials are not the only way to interpolate yield curves. An alternative, and indeed simpler algorithm, is to use linear interpolation between observed points and constant extrapolation at other points. This always gives a standard deviation of 1% or less, with the reduction due to diversification in the interpolation between points.

The chart compares standard deviations for both smooth and linear interpolation.

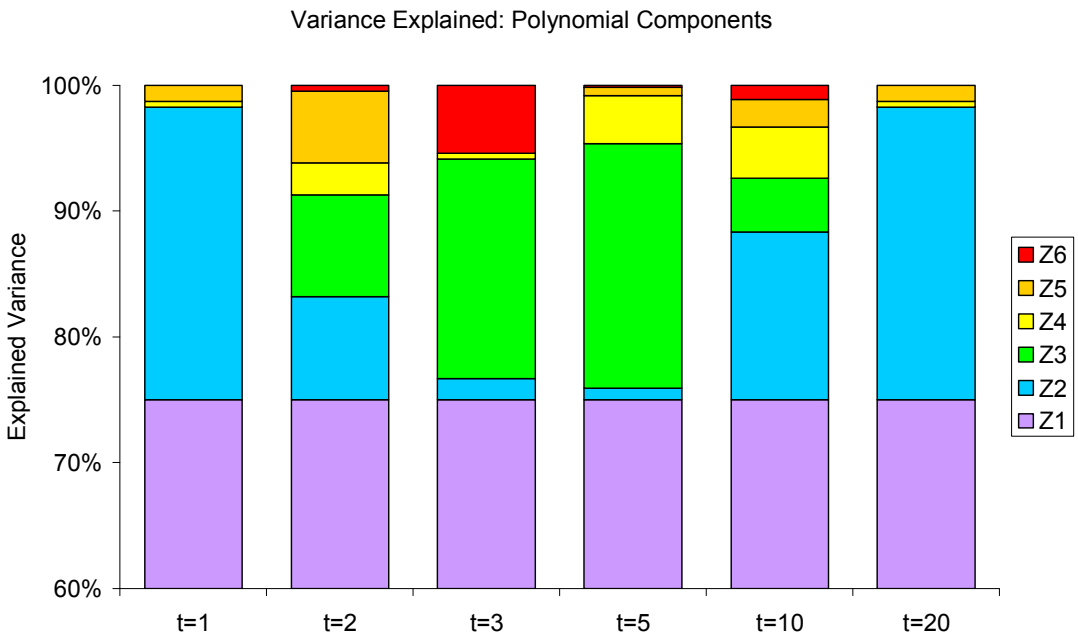


3. Principal Components

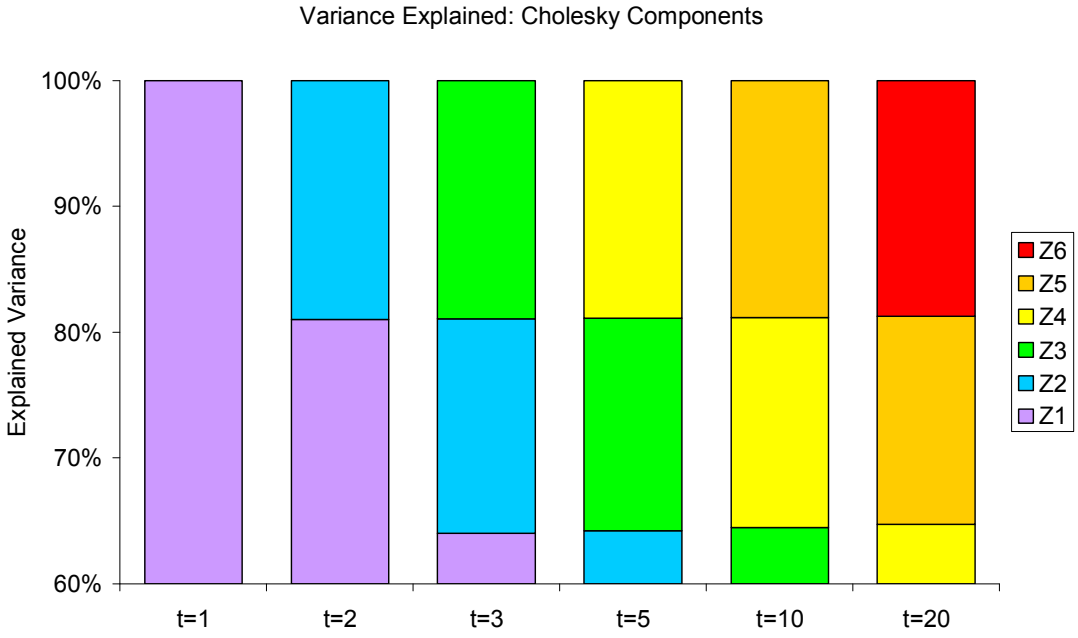
3.1. Proportions Explained

Let us consider the variance of the rate at a particular term as components are added.

The first component necessarily understates the variance. For the yield curve as a whole, the contribution of the first component can be measured as the sum of the squares of the first column, which is the sum of the squares of the first noise term $Z1$. The inclusion of each new component increases the variance, until all the components are incorporated and the sum of the squares of the entire matrix is 1. We can investigate the proportion of variance explained by each component, split according to the time point, that is, by rows of the original matrix. The chart shows the results for polynomial components. In order to make the higher components more visible, the vertical axis starts at 60%.

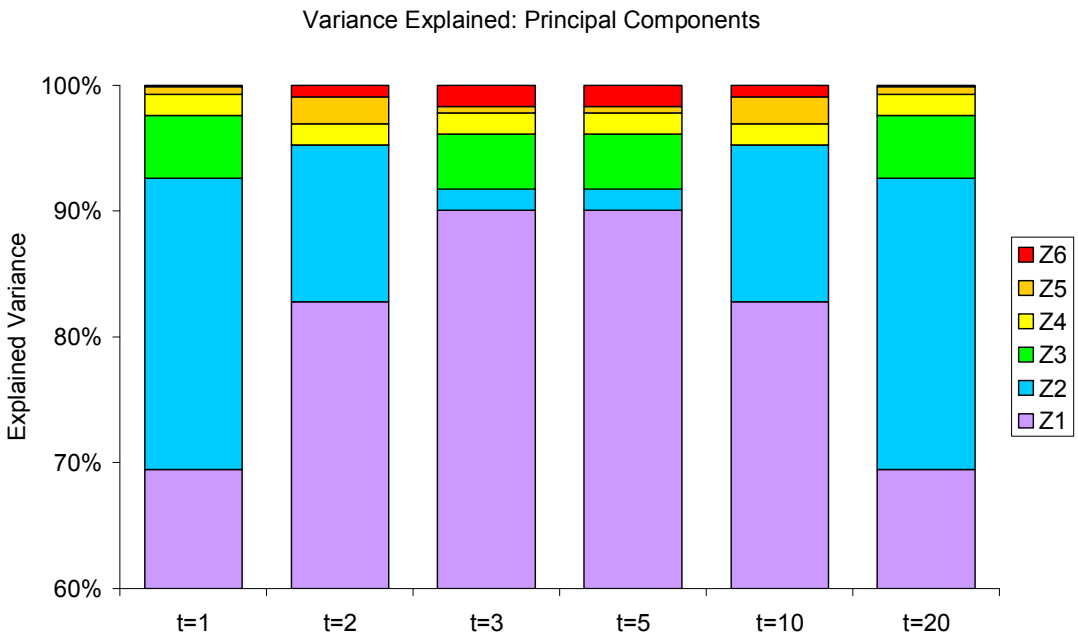


We also show the corresponding figure for Cholesky decomposition:



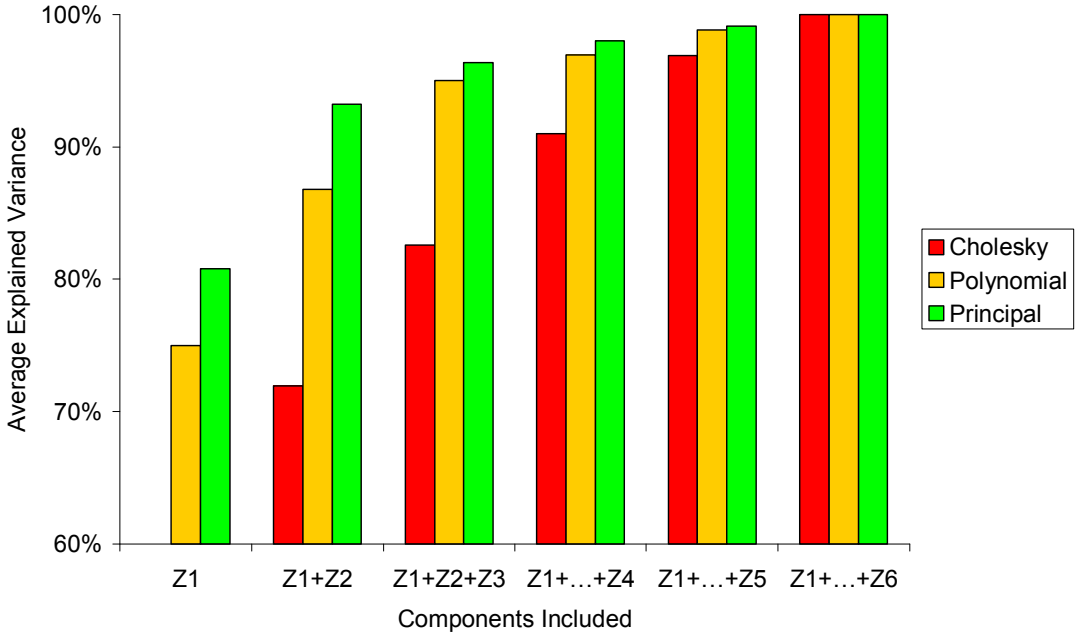
Here we see that the first component explains all of the variance for the 1 year rate. However, convergence is much slower for longer term rates, and indeed, is so bad for $t \geq 5$ as to fall below our vertical scale.

We can show the same chart for principal components:



We can consider how to choose the components to maximise the variance that the early components explain. There is clearly a trade-off here. Cholesky decomposition does a good job at t=1 but a terrible job at t=20. Polynomials and principal components are more consistently convergent across a range of terms.

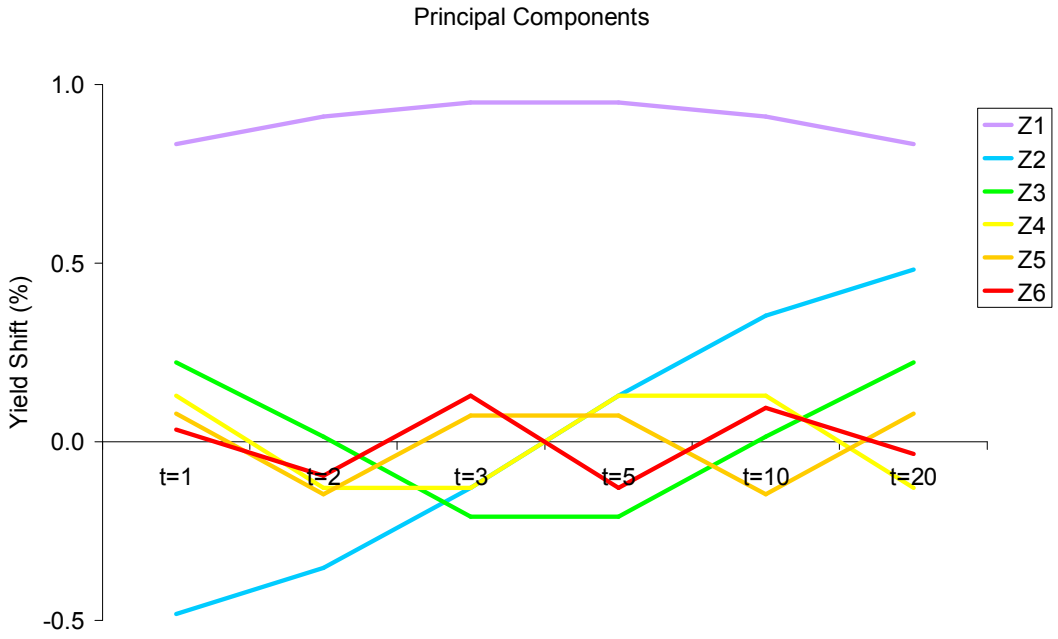
Taking the average across the six terms considered, we can measure the speed of convergence as more components are added. The bar chart shows the results (note that in order to make the differences more visible, the vertical axis starts at 60% explanation and not 0%).



We can see that the Cholesky approach converges slowest of all. The polynomial decomposition is much better, but not quite as fast as principal components analysis.

In fact, principal components are defined in order to maximise the speed of convergence, averaged across the key terms, that is to maximise the height of each green bar in the chart above. Therefore, the principal components represent the best possible convergence outcome, which cannot be beaten. This is encouraging, because it gives us a basis for claiming that some decompositions are "better" than others, and in particular that principal components analysis is "best possible". Furthermore, this analysis has assumed nothing about the underlying financial business. This raises the hope of component analysis that can be performed once and is then valid for multiple applications.

The chart shows the principal components resulting from the analysis:



We can see that the first three components capture aspects of level, slope and curvature, with higher components being wigglier. This is similar to the situation for polynomial components.

3.2. Why Reduce the Number of Components?

When all components are considered, any yield curve decomposition explains 100% of the variance. If some components are excluded, then principal components analysis converges best, in terms of average variance explained.

Why, then might we want to truncate the decomposition, and exclude some components? The most important reason arises in the context of Value-at-Risk calculations.

Let us suppose the net assets of a firm are a smooth function of the yield curve shift X. Then, for small X, the net assets are approximately of the form $a_0 + g^{\mathrm{T}}X$, where a_0 is the starting net assets before the yield shift and g is the vector gradient. With a variance-covariance matrix V, the variance of net assets is then $g^{\mathrm{T}}Vg$. Assuming

multivariate normal distributions and a specified confidence level α (for example α=0.995), the value at risk is $\sqrt{g^T V g}\Phi^{-1}(\alpha)$.

To evaluate this expression, we need to compute the gradient *g*. This is commonly estimated using central finite difference methods. With our model, based on 6 points of the yield curve, the estimation of *g* requires 12 net asset calculations, also known as "stress tests".

The calculation of stress tests may be an easy task, for example if all future cash flows are fixed. But financial firms' cash flows are typically variable, depending not only on market conditions but also the actions of customers and management. In this case, net asset calculation may be an onerous task. A request to recalculate 12 stress tests could have important operational implications. The burden gets worse if more than 6 points on the curve are modelled.

A constraint on the number of components can reduce the effort required in value-at-risk calculations. As an intermediate step, we need to compute the variance $g^{\mathrm{T}} V g$. Now suppose we can approximate $V \approx BB^{\mathrm{T}}$. Then we can approximate the variance $g^{\mathrm{T}} V g \approx g^{\mathrm{T}} B B^{\mathrm{T}} g = (B^{\mathrm{T}} g)^{\mathrm{T}} (B^T g)$. A saving arises because we can compute $B^T g$ with fewer stress tests than to estimate *g*. The number of stress tests required is twice the number of components.

This is how to estimate $B^T g$ without knowing *g*. Let us consider the first column of *B*, that is, the first component, a vector, *b* say. Then the first element of $B^T g$ is estimated as

$(2h)^{-1}$[net assets(X=hb) – net assets(X=-hb)]

We need one of these calculations for each component, not for each point on the yield curve.

Other contexts may also show an advantage from needing fewer components. For example, Monte Carlo work involves simulating the components of *Z* and computing the matrix product *BZ*. Both of these operations are quicker if the number of components is reduced.

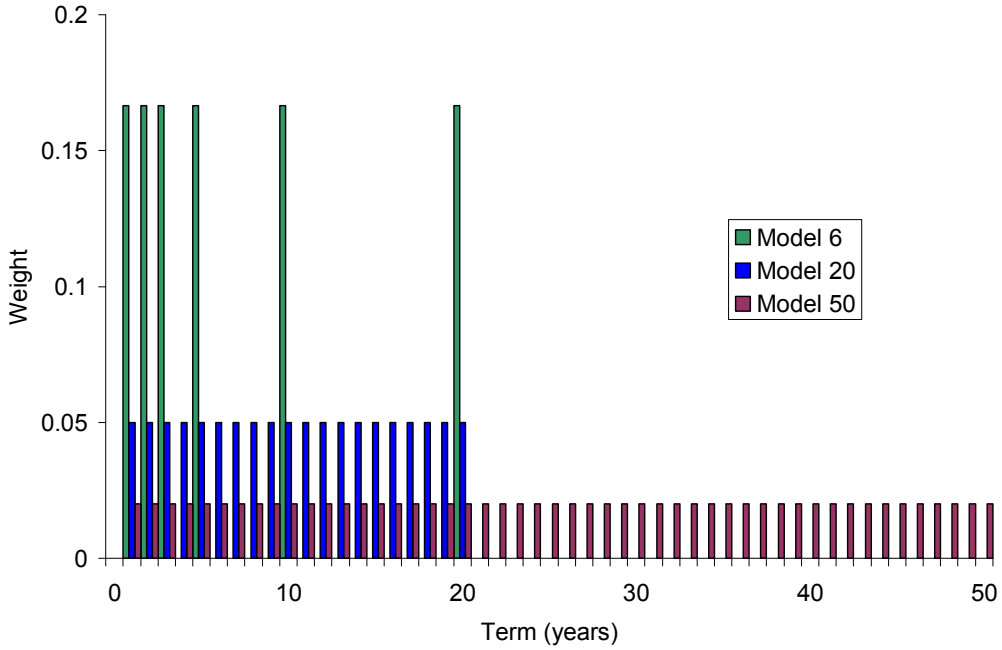
The pricing of some financial products involves optimal stopping problems. Examples include estimating the best time to pre-pay a fixed-rate loan or to cash in a fixed-rate deposit. The analysis of such products involves movable boundary problems, whose solution is only straightforward for low dimensional problems. For this reason, consideration of such products is usually in the context of 1 or 2-component models.

In each of these cases, neglecting higher components introduces errors, because the correlation structure is modelled inaccurately. The errors may still be considered a price worth paying in order to benefit from the run time advantages of a low dimensional interest rate model.

3.3. Alternative Choices of Weights

We have analysed principal components based on speed of convergence at key time points t={1,2,3,5,10,20}.

We could consider alternative weights. For example, a "Model 20" that gives equal weight to all time points between t=1 and t=20 inclusive. Or a "Model 50" based on a 50-year yield curve. The respective weights are show in the chart below:



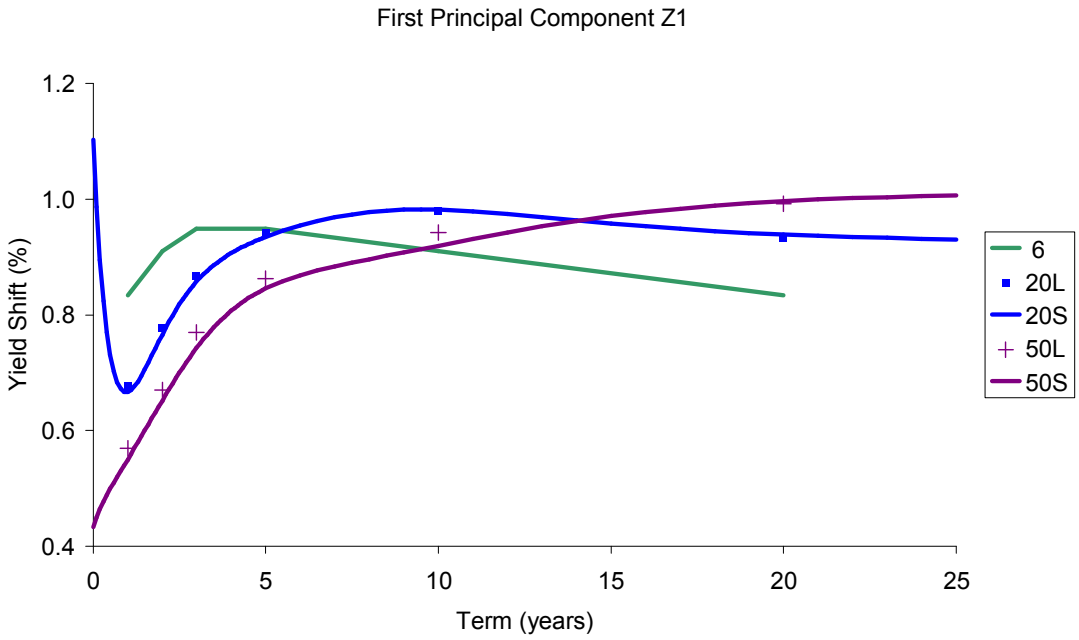
As we shall see, it is an unhelpful feature of principal components analysis that the choice of weights is a major determinant of the calculated components. The need to choose weights does not arise for our other methods of component construction: orthogonal portfolios and variance matching. When using principal components analysis, we might (for example) construct a system of weights to reflect the relative size of cash flows at different terms. The ideal choice of weights might therefore vary from one business to another. On the other hand, it may be better to use a common compromise set of weights for all businesses in order to simplify aggregation calculations across multiple lines of business.

If we are to use intermediate time points between those originally modelled, we must also specify a model for yields at those points. In this note we consider two alternatives. The linear model "L" uses linear interpolation, with flat extrapolation beyond the 20 year point. The smooth model "S" fits a polynomial of order 5 to the six observed points.

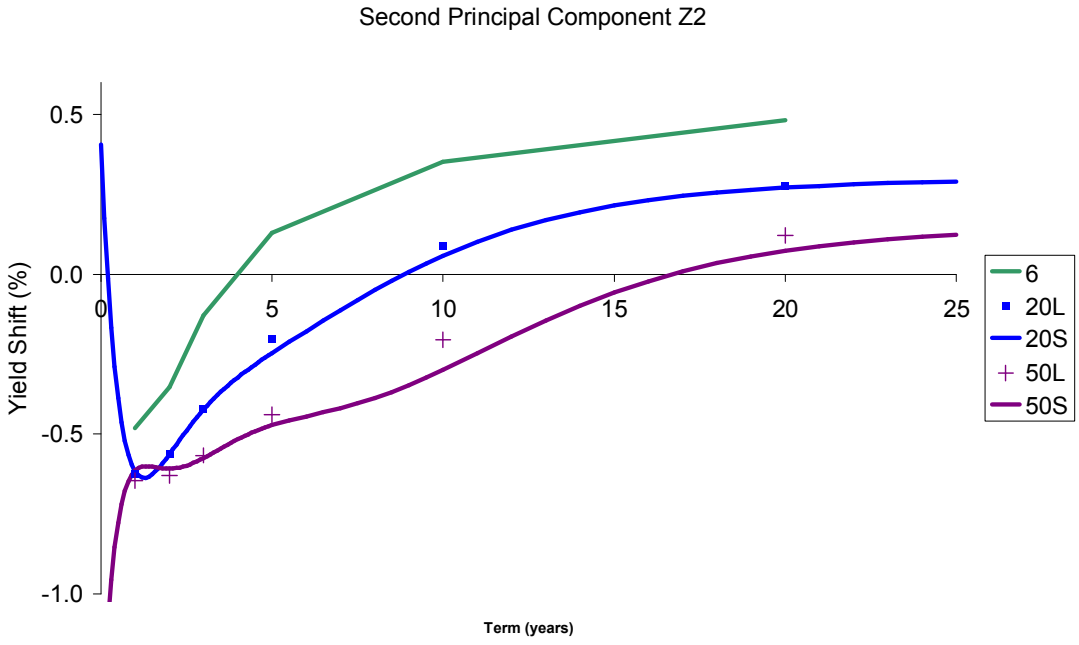
The possible combinations of weights and interpolations give us five models, which we will denote as Model 6, 20L, 20S, 50L and 50S.

3.4. Principal Component Comparisons

We now compare principal components under our five models: 6, 20L, 20S, 50L and 50S.



In each case, the first principal component can reasonably be interpreted as a measure of "level". We see differences according to where the weights lie. The first component is largest where the weights are largest. Thus, model 6 produces a maximum at t=3, while model 50 has its maximum at t = 25. The choice of interpolation method (L or S) has little effect.

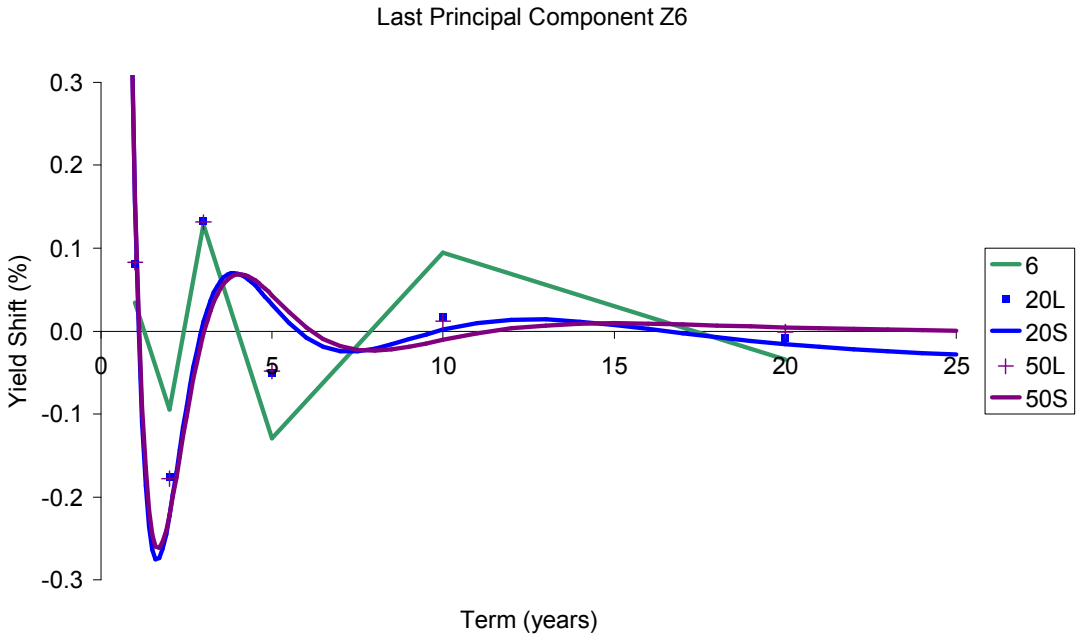


The second component captures the slope, or twist, of a yield curve move. One important question in value at risk calculation is the choice of pivot, that is, the term of interest rates which is unchanged by the twist stress test. This corresponds to the Y intercept of the second principal component.

Our example shows that the pivot is not an inherent property of past yield curve moves. On the contrary, it depends on the weights chosen for principal components

analysis. If the weights extend a long way into the future, then the pivot occurs at a large t value. The choice of weights is not a purely technical decision with limited impact. On the contrary, a consideration of the second principal component reveals the importance of the choice of weights. Given the potential difference in calculated value at risk, a rigorous motivation for the choice of weights is important, in place of the heuristic reasoning we are able here to provide. The choice of smoothing method, however, remains unimportant.

We now jump to the last principal component.



It is unlikely that this component is much used in practice. However, the comparison with previous figures is interesting. We see a modest effect of the choice of weights, with the 20 and 50 year models looking very similar. What the last component picks up is the smoothing algorithm. The wiggles from the "S" models come through clearly, in contrast to the simpler shape under linear interpolation.

3.5. Implicit Data Enhancement

Ideally, interest rate data is observed from actual trades or bid/offer quotes in a deep liquid market. In many markets, however, trades may be infrequent or bid-offer spreads wide, especially for long dated cash flows.

Given these difficulties, we might expect consequences for yield curve data quality and availability. Online sources, however, provide apparently complete information extending far back in time. They can do this because of substantial investment in data cleaning. Cleaning methods include interpolation and extrapolation to infer missing data points or to adjust out-of-date price information. The data collection may be a many stage process: individual banks apply their own cleaning algorithms to data made even cleaner by commercial data vendors' systems.

Clean data does not easily reveal which data points are real market information and which are filled in by algorithm. However, principal component analysis may reveal this information. And you could test this in a real market by trying to transact at prices posted on the system.

For an example of the power of PCA, suppose a user downloads yield curve data out to 50 years maturity. Theoretically, this data set may require 50 components for a full explanation of the correlation structure. Maybe, analysis of the data shows 6 significant components, with the higher components accounting for a negligible proportion of total variability.

One possible explanation is that there really are only 6 components in the economy. Another explanation is that we are dealing with a 50L or 50S model, with 6 real data points and 44 points constructed by interpolation. The shape of the 6^{th} component reveals more about the type of interpolation used. In other words, the higher components tell us about the process of data collection and cleaning, rather then about risk in the financial markets.

3.6. Value-at-Risk for Hedging

Economic capital is an important application of yield curve models. Economic capital is often regarded as having a cost, and so firms try to trade to minimise stated capital requirements.

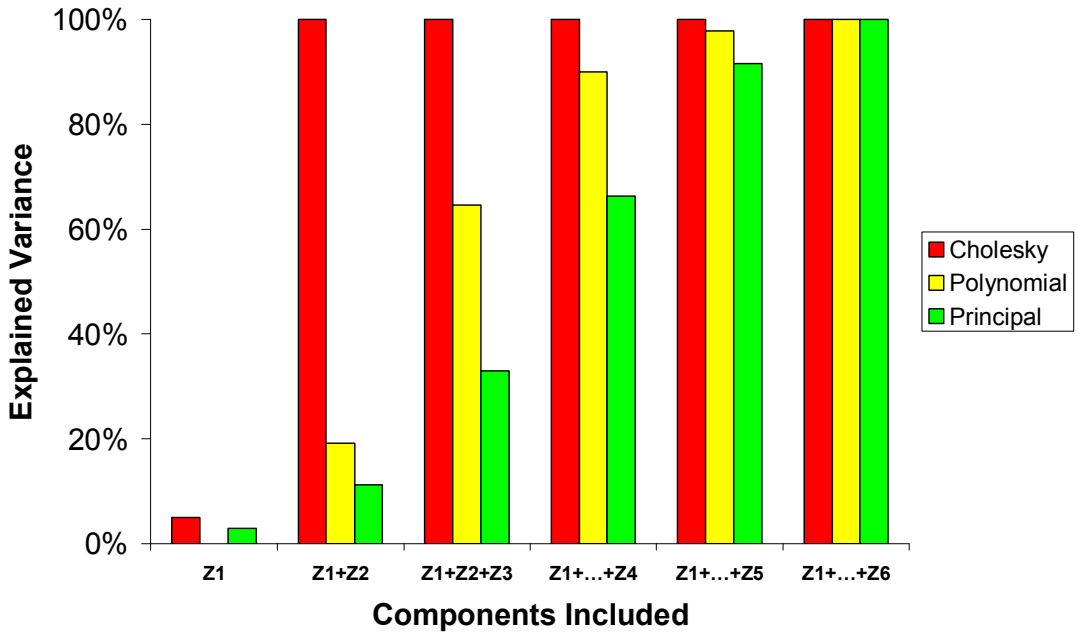
One important tool for this is *hedging*. For example, if a firm has a long exposure to an 11 year interest rate, they may trade in the interest rate markets to acquire a corresponding short exposure to the same interest rate. The net effect is immunisation – the firm is protected against moves in either direction, at least against small moves.

Hedging is often less exact than this. A firm may decide to hedge the 11-year exposure with a 10-year trade, for example because the 10-year instrument is more liquid than an exact match. This hedge should still be effective – because the 10 year rate and 11 year rate are strongly correlated, but not quite as good as hedging the same rate as the original exposure.

In this case, PCA converge more slowly than we hope. PCA finds components that well explain movements in the yield curve, but that is not the same as explaining movements in my portfolio. PCA may ensure that the variances of 1 and 2 year rates are well explained. But to assess hedge effectiveness you also need to know the correlation between them. PCA solves a particular, objective function, but that objective function takes no account of how fast correlations converge. Technically, we could include covariances in the weights for evaluating convergence, but the optimisation would then trade off covariances against variances elsewhere.

For example, let us suppose an investor has an exposure of £1 per 1% move in 1 year rates, which they have partially hedged with an equal and opposite exposure to the two year rate. From our assumed correlation matrix, we can calculate the variance of profit to be 1+1-2*0.9 = 0.2. We can investigate how quickly the different

decompositions converge to the true value. The principal components calculation refers to Model 6.



As we would expect, the Cholesky method scores well, as it builds up the volatility by iteration starting at the short end of the curve. With cash flows only at t=1 and t=2 in our example, the Cholesky method has captured all the variance in the first two terms.

The principal components method performs surprisingly badly. Section 3.1 suggests that the first three principal components explain more than 95% of the yield curve variability. However, in our example, only 33% is explained by the first three components. Arguably, this is an unfair comparison; had we known that the cash flows stopped at time 2, we would have applied more weight to the early years when calculating principal components, and so obtain faster convergence for those flows.

In this particular example, the Cholesky method gives good convergence, while Principal Components has the slowest convergence. This will not always be the case. We could construct examples illustrating any of the six possible orderings. Instead, our example is intended to illustrate the potential gap (in either direction) between an advertised "percentage explained" for the yield curve as a whole, compared to the actual explanatory power for a particular set of cash flows. By construction, PCA maximises the advertised percentage explained, and therefore carries the greatest potential to disappoint.

In general, the problem of slow convergence is particularly acute if yield decomposition into components is used for constructing a hedge in the first place. For example, given a particular definition of level, slope and curvature, it is easy to find a hedge portfolio that immunises all three. Tested against that decomposition, it appears that risk is eliminated. What has really happened is that risk is concentrated in the fourth and higher components, which have been discarded on order to speed up the value-at-risk calculation. There are two solutions to this problem. One is to use more components for analysing risk than are used for building the hedge in the first place.

The second solution is to use a full model, including all components rather than truncating.

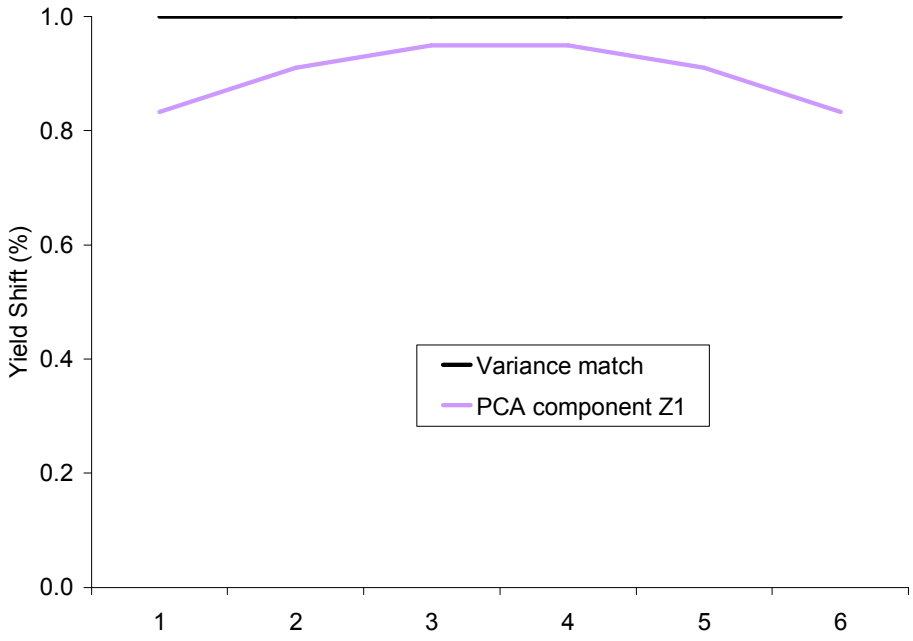
4. Variance matching

4.1. A One-Factor Model

Suppose we are constrained to use only a single component. Then interest rates moves at different terms must be 100% correlated.

A naïve solution is to set the interest rate standard deviation to 1% at all the modelled terms. We then have a solution that explains 100% of the variance for each key rate, although clearly the correlations are overstated. We call this "variance matching". This is essentially the test which CEIOPS have calibrated for solvency II, based on the volatility of yields at various terms.

A theoretically more sophisticated solution is to use the first component from PCA. The chart shows a comparison:



The PCA is supposed to maximise the variance explained by each component. In our example, the first component explains 80.8% of the yield curve variance.

Yet, the variance match explains 100% of the variances. The correlations are equally wrong in both one-component models. So it is difficult to describe any sense in which the theoretical superior PCA is better in practice. Indeed, we might wonder how we ever convinced ourselves that 80.8% is the best possible, given that variance matching explains 100% of the variance.

The answer to these points is subtle. The advantage of PCA is that the first component is one of a series. By adding more and more terms, we can get closer to the true yield

curve distribution. This is useful if we can somehow test convergence, optionally adding higher terms only when necessary.

In contrast, the variance match might produce a good first guess, but we cannot refine that guess by adding more components. If we add further components, we might get the correlations more accurate, but those extra turns will also increase the variances, making them too large.

If we know at the outset that we might want to explore further stress tests, then the PCA makes sense. The first component is one step on a longer path. On the other hand, if more than one stress is excluded, for example by computation costs, then the variance match could be a better approach.

4.2. Three Factor Variance Match

We have discussed a variance match for a single factor model. We now consider extending this idea to three factors.

We assume the importance of capturing the variance of yields at each term. With three factors, we are able to impose additional constraints. Knowing that hedging often involves offsetting risks at adjacent terms, we can ask that our three factors correctly replicate the correlations between adjacent terms. This is equivalent to reproducing the variance of the yield curve slope between terms.

These are still too few constraints to determine a three factor model. We can insist also on capturing the correlations between rates that are next-but-one to each other. Equivalently, we reproduce the variance of the second differences in yield slope. These differences are relevant to "barbell" hedging strategies that (for example) seek to hedge a 2 year exposure with an average of 1 and 3 year exposures.

We have articulated some constraints that are relevant to common business strategies. Now all we need is to solve the equations. The solution is as given in the introduction:

Variance matching components			
	Z1	Z2	Z3
t=1	0.974679	-0.223607	0.000000
t=2	0.974679	0.223607	0.000000
t=3	0.872082	0.223607	-0.435286
t=5	0.872082	-0.223607	-0.435286
t=10	0.974679	-0.223607	0.000000
t=20	0.974679	0.223607	0.000000

We can use these three components to reconstruct the following implied correlation matrix:

Correlation matrix implied by Variance Matching Components						
	t=1	t=2	t=3	t=5	t=10	t=20
t=1	1	0.9	0.8	0.9	1	0.9
t=2	0.9	1	0.9	0.8	0.9	1
t=3	0.8	0.9	1	0.9	0.8	0.9
t=5	0.9	0.8	0.9	1	0.9	0.8
t=10	1	0.9	0.8	0.9	1	0.9
t=20	0.9	1	0.9	0.8	0.9	1

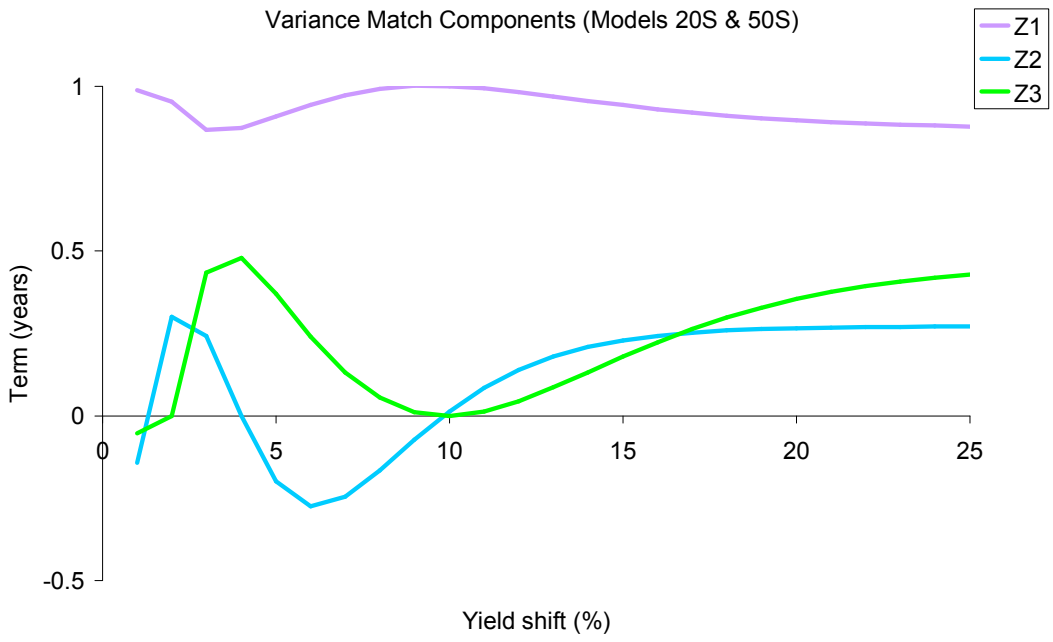
We can see that the shaded elements: the main diagonal and two above and below it, are replicated exactly, as required. However, the variance matching components overstate the elements on the bottom left and top right of the matrix. In particular, the variance matching components force the rates at t=1 and t=10 to be equal. It also forces equality between rates at t=2 and t=20. This is an undesirable side-effect of using only three components when six are required.

4.3. Three Factor Models: Alternative Weightings

We have constructed a three factor variance-match solution for a 6-point model. We now consider how that solution extends to models with more points on the yield curve.

The variance match for model 6 also works for model 20L and 50L. The reason is that the variance match fits the variance of the value and the first and second differences. The second differences of a linearly interpolated curve, are zero, except at the data points. This means that solving for the three components commutes with interpolation. It does not matter whether we interpolate first, then construct components, or if we solve first for the components and then interpolate.

The situation is different for the models 20S and 50S. The non-linear interpolation means that fitting the variance of second differences is not trivial. Nevertheless, it can be done. Furthermore, the solution for model 20S is simply the decomposition for 50S, restricted to the first 20 years.



As with the original decomposition, these solutions are unique only up to rotation of the underlying normal variables. In this example, we chose a rotation so that the

second component vanishes at t=4, while the third vanishes at t=2 and t=10. These properties are then shared with the Model 6 components.

The variance matching solution has many advantages, including replication of yield variances, and also the variances of first and second differences, as well as avoiding the need to specify time weights. There are also a few disadvantages. One disadvantage is the inability to build on this solution by adding more factors – if we want to add a fourth factor we have to go back to the beginning and build all four from scratch. In that sense, the PCA approach is better, as fourth factor can be added without disrupting the previous three. The second disadvantage of the variance match approach is the tendency to wiggle. This is necessary in order to capture the variability of yield curve slopes, but at the same time negates the intuitive appeal of a second factor relating to "slope" and a third to "curvature".

5. Bibliography

Anderson, T W (1957). Multivariate Analysis, Wiley. 3rd edition 2003.

Gibson, Michael S and Matthew Pritsker (2000). Improving Grid-based Methods for Estimating Value at Risk of Fixed-Income Portfolios
http://www.federalreserve.gov/Pubs/feds/2000/200025/200025pap.pdf

Jamshidian, Farshid and Yu Zhu, 1997. Scenario simulation model: Theory and methodology. Finance and Stochastics 1:1.

Loretan., Mico (1997). Generating market risk scenarios using principal components analysis: methodological and practical considerations
http://www.bis.org/publ/ecsc07c.pdf

Matzner-Løber, Eric and Villa, Christophe (2004). Functional Principal Component Analysis of the Yield Curve
http://www.u-cergy.fr/AFFI_2004/IMG/pdf/MATZNER.pdf

Thomas, M and E Maré (2007). Long Term Forecasting and Hedging of the South African Yield Curve.
http://www.actuarialsociety.org.za/scripts/file_build.asp?id=100000794

Tolmasky, C F (2007). Principal Components Analysis in Yield-Curve Modeling (slides)
http://www.math.umn.edu/finmath/seminar/Materials/tolmasky1.pdf