

## **Interpolating the South African Yield Curve**

**James Maitland**

Department of Statistics and Actuarial Science

University of the Witwatersrand

Johannesburg, South Africa

Telephone: +27 11 486 1946

Facsimile: +27 11 339 6640

Email: [maitland@stats.wits.ac.za](mailto:maitland@stats.wits.ac.za)

Email: [james\\_maitland@hotmail.com](mailto:james_maitland@hotmail.com)

### **Abstract**

A principal components analysis of the South African Yield Curve suggests that two factors explain most of the variability in both yield levels and changes in yields. This result is then used to select which interest rates to model and how to use these rates to reproduce the entire curve.

**Keywords:** Principal components, yield curves, interpolation, South Africa.

### *Introduction*

In many actuarial and investment applications such as asset-liability management, solvency testing and value-at-risk calculations, realistic estimates of the distribution of future yields are required. Since it is not practical to model the entire yield curve, Tilley (1992; 527) suggests modelling 8 key yields and interpolating in order to reduce the dimension of the model at the expense of producing non-key yields that are not arbitrage free. In order to reduce the dimension of the problem further, Sherris (1994) suggests factor analysis to determine the dimension of randomness in the yield curve. In this paper, Principal Components Analysis (PCA) is used to determine the number of points,  $n$ , required to adequately model the South African yield curve. The subset consisting of the first  $n$  principal components is then used to reproduce the entire yield curve, given a specific subset of  $n$  points along the curve. These maturities can then be modelled as part of a larger set of variables including other asset categories and economic variables for the purposes of modelling the assets and liabilities of a financial institution.

### *Background*

Prior to 1982, there was virtually no active secondary market in bonds. Prescribed asset legislation forced pension and provident funds and insurance companies to hold a certain percentage of their assets in respect of liabilities in government bonds, cash and other approved bonds. In the 1970's, insurance companies and pension funds held on average 41% of the long-term domestic marketable stock debt of the central government (compared with 47% by the Public Investment Commissioners); and 70% of local authorities' stock, (Falkena et al., 1984; 129).

In the early 1980's, an active secondary market in South African bonds began developing and has been growing rapidly ever since (McLeod, 1990). In 1986, the Johannesburg Stock Exchange (JSE) instituted a bond clearing-house and although the majority of bond trading was "over-the-counter" (OTC), a small number of trades were recorded on the JSE. Since some trades were recorded at each available maturity and

since these trades would have reflected yields traded OTC, the JSE-Actuaries Yield Curve can be considered to be a fair estimate of market yields prevailing at the time. In 1996, the bond exchange opened and the Financial Markets Control Act now requires all bond trades to be recorded by a recognised exchange.

#### *Principal components analysis*

A number of discrete-time models suggested by academic researchers and practitioners conclude that the short rate is non-stationary. A partial listing of these authors include Stock and Watson (1988), Mills (1994; 68), Ang and Moore (1994), Johansen and Juselius (1992), Juselius (1995), and Pesaran and Shin (1996). In contrast, continuous-time models of the short-term interest rate often include a mean reversion term (see, for example, Vasicek (1977), Brennan and Schwartz (1982) and Cox, Ingersol and Ross (1985)). In this paper, a principal components analysis (PCA) of both the levels and first differences of the South African yield is presented.

Let  $\mathbf{x}$  be a random  $d$ -vector with mean  $\mu$  and covariance matrix  $\Sigma$ , and let  $\mathbf{T} = (t_1, t_2, \dots, t_d)$  be an orthogonal matrix such that  $\mathbf{T}'\Sigma\mathbf{T} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_d)$ , where  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d$  are the eigenvalues of  $\Sigma$ . If  $\mathbf{y} = \mathbf{T}'(\mathbf{x} - \mu)$ , then  $y_j = t_j'(\mathbf{x} - \mu)$  is called the  $j^{\text{th}}$  principal component score of  $\mathbf{x}$  and is the orthogonal projection of  $\mathbf{x} - \mu$  in the direction  $t_j$  (Seber, 1984; 176). Principal components analysis explains the variance-covariance structure of the original variables through an orthogonal rotation of  $\mathbf{x}$  such that the first principal component gives the direction of maximum variation, the second principal component gives the next largest direction of maximum variability orthogonal to the first principal component, and so on.  $d$  principal components are required to reproduce the total system variability completely but much fewer principal components may explain a reasonable proportion of the total variability and hence reduce the dimension of the model with only a small loss of information.

We define the yields for annual terms from 0 and 25 years along the JSE-Actuaries Yield Curve (with the INET (1998) codes *JAYC00*, *JAYC01*... *JAYC25*) to be our 26

dimensional random vector. If yields are stationary with structural breaks, the moments of the level yields exist. Table 1 provides summary statistics for key yields at annual maturities from 0 to 25 years using monthly data from January 1986 to December 1998 while Figure 1 illustrates the yield curve over this period.

	Mean	Median	Standard deviation	Minimum	Maximum
<i>JAYC00</i>	14.55	14.58	3.31	8.47	22.99
<i>JAYC01</i>	14.38	14.59	2.72	9.24	2.31
<i>JAYC02</i>	14.28	14.55	2.25	10.01	1.64
<i>JAYC03</i>	14.37	14.72	1.94	10.36	1.06
<i>JAYC04</i>	14.59	14.90	1.74	10.61	0.59
<i>JAYC05</i>	14.84	15.21	1.59	10.91	0.24
<i>JAYC07</i>	15.26	15.67	1.41	11.39	9.84
<i>JAYC10</i>	15.58	15.80	1.33	11.75	9.72
<i>JAYC15</i>	15.69	15.94	1.29	11.96	9.49
<i>JAYC20</i>	15.65	15.77	1.25	12.02	9.34

Table 1 JSE-Actuaries Yield Curve (1986-1998)

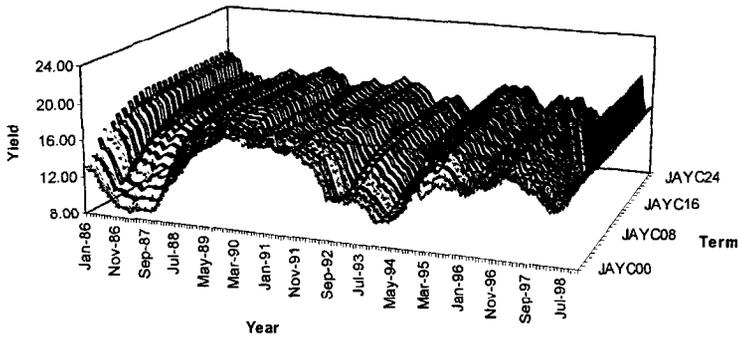


Figure 1 JSE-Actuaries Yield Curve (1986-1998)

A PCA on the covariance matrix reveals that the first principal component explains 77.1% of the total variability in the yield curve, the first two principal components together explain 98.4% and the first three principal components together explain 99.4% of the total variability in the yield curve. Figure 2 illustrates the coefficients of each of the first three principal components by term to maturity.

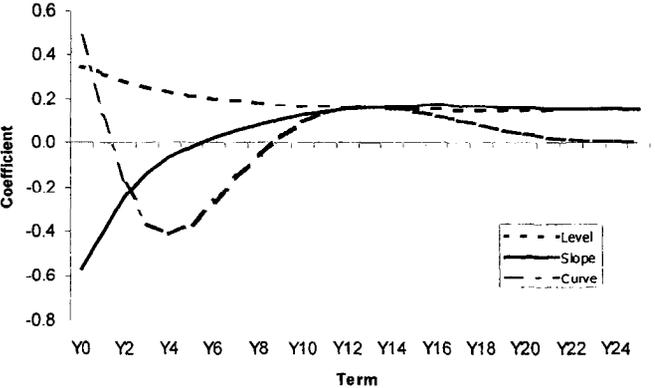


Figure 2 Coefficients for the first three principal components of yield levels

The coefficients for the first principal component are all positive so that an increase in the score of the first principal component results in an increase in all yields. The first principal component can therefore be regarded as a level factor. Since the coefficients are not all equal, a change in the score of the first principal component does not result in a parallel shift; instead, the short end of the curve moves more than the long end.

The coefficients for the second factor are negative at the short end and monotonically increase to a positive value at the long end. Hence, a change in the score of the second principal component results in an opposite effect on the two ends of the yield curve and this factor can be viewed as causing a change in the slope of the yield curve. The third principal component has a negative effect on medium yields and a positive effect

on short and long-term yields and hence can be interpreted as a curvature factor. Figure 3 illustrates the principal component scores for the first three principal components from January 1986 to December 1998.

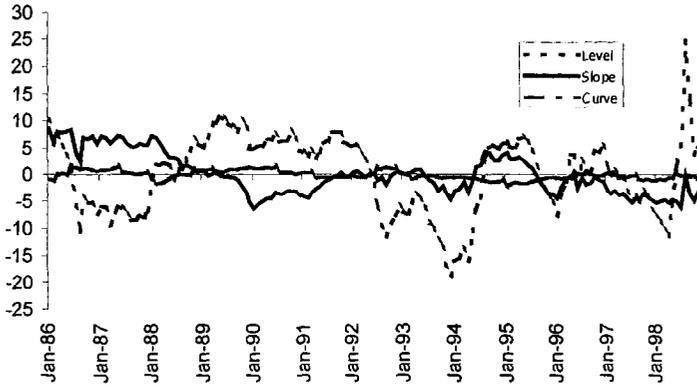


Figure 3 Principal component scores for yield levels (1986 – 1998)

The third principal component accounts for only 1% of the total variability and the remaining 23 principal components account for about 0.5% of the total variability. Hence, two principal components appear to capture most of the variability in the yield curve. The next section discusses how these principal components can be used to reconstruct the entire yield curve.

So far, we have considered level yields. If yields are non-stationary, then the population moments of the level yields do not exist. We now define the changes in yields for annual terms from 0 and 25 years along the JSE-Actuaries Yield Curve to be our 26 dimensional random vector and again use monthly yield data from January 1986 to December 1998.

A PCA on the covariance matrix of changes in yields reveals that the first principal component alone explains 92.8% of the total variability, the first two principal components together explain 97.3% and the first three principal components together explain 98.4% of the total variability. Hence, two principal components again appear to capture most of the variability in yield curve changes. Figure 4 illustrates the coefficients of the first three principal components by term to maturity.

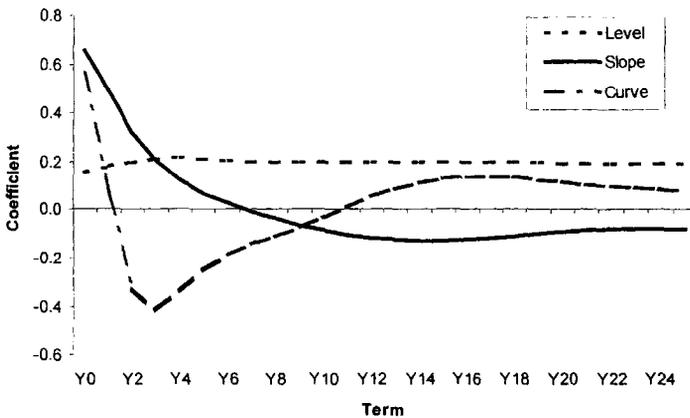


Figure 4 Coefficients for the first three principal components of yield changes

The first principal component affects all maturities by similar amounts and in the same direction. It can be interpreted as a level shift factor but not as a parallel shift factor since the coefficients are unequal. Unlike the levels PCA, the short end of the curve moves less than the long end in response to the score of the first principal component. The second factor has an opposite effect on short and long yields and can be viewed as a slope change factor. The third principal component has a negative effect on medium yields and a positive effect on short and long-term yields and hence can be interpreted as a curvature factor. Figure 5 illustrates the principal component scores for the first

three principal components of yield curve changes for the period January 1986 to December 1998.

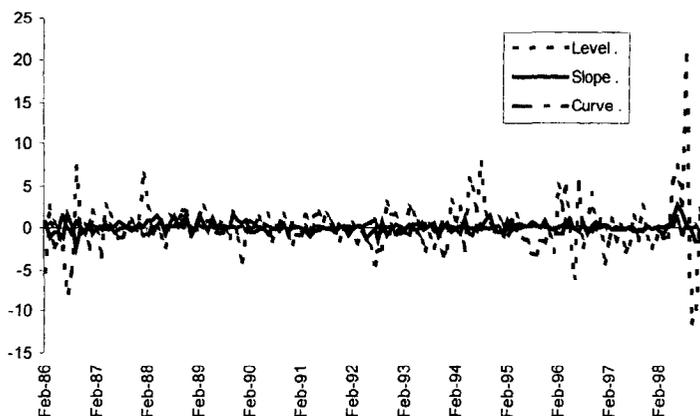


Figure 5 Principal component scores for yield changes (1986 – 1998)

The traditional theory of immunisation as developed by Redington (1952) immunises a portfolio against parallel shifts in the yield curve. Parallel shifts imply the existence of arbitrage opportunities (see Boyle, 1978) and it is important to note that the first principal component does not represent an entirely parallel shift. However, for terms greater than 5 years, the first principal component does seem to represent a parallel shift and for terms greater than 12 years, the second principal component also seems to represent virtually parallel shifts. Hence, the first two principal components, which represent 97.3% of the total variability, appear to indicate the regular occurrence of parallel shifts and hence arbitrage opportunities at the long end of the curve.

Estimates of variances, covariances and correlations can be very sensitive to outliers and so we can expect principal components to have the same sensitivity. The extreme scores for the first principal component between August and October 1998 shown in

Figure 5 and the corresponding large changes in the level of the yield curve evident from Figure 1 suggest the need for a PCA for sub-periods of the data. For the sub-period 1986-1997, the proportion of the variability explained by the first principal component of yield curve changes decreases from 92.8% (for the period 1986-1998) to 90.0%.

For level yields, the proportion of the variability explained by the first principal component reduces from 77.1% (for the period 1986-1998) to 76.1% for the sub-period 1986-1997. In both the level and the differenced yield sub-period analyses, the principal components remain relatively unchanged suggesting that the full period analysis is relatively robust to the outliers from August to October 1998. A number of alternative sub-periods have been considered and the results of the full period appear to be relatively robust to the choice of sub-period.

In the above analyses, principal components have been derived from the covariance matrix. If the variables in a PCA are measured on scales with widely differing ranges, it is preferable to use the correlation matrix (see Seber, 1984). Although the higher volatility of short rates compared with long rates results in an increased loading of the short rate on the first few factors, a PCA for both the levels yields and yield differences using the correlation matrix gives principal components and variability proportions that are similar to those obtained using the covariance matrices. Hence, the results of the PCA on the covariance matrix appear to be relatively robust to the lack of scaling. This is not too surprising given that the standard deviations of short and long yields are of the same order of magnitude.

One further point worth considering is the effect that the mathematical formulation of the JSE-Actuaries Yield Curve may have on the principal components analysis. The curve is constructed in two steps (see McLeod (1990)):

- 1) Using a form of cluster analysis, five cluster points are estimated and bonds are assigned to each cluster. The bonds in each cluster are then used to determine a weighted average term to maturity and a weighted average yield for their respective clusters. A sixth cluster with a maturity of 30 years and yield equal to the weighted average yield of the cluster with the highest weighted average yield is also determined.
- 2) Using these six cluster points, intermediate points along the curve are estimated using cubic spline interpolation.

Since the yield value of the sixth cluster is derived directly from one of the existing five cluster points, there are effectively five independent points along the curve. Hence, it is unlikely that more than five principal components would be required to reproduce most of the variability of the yield curve. The fact that two principal components capture most of the variability is a strong indication that the PCA is not constrained by the mathematical formulation of the yield curve.

*Reconstructing the yield curve*

Using the principal components,  $\mathbf{T}$ , and the principal component scores at time  $t$ ,  $\mathbf{y}_t$ , of the level yields (change in yields), the level yields (change in yields) at time  $t$ ,  $\mathbf{x}_t$ , can be reconstructed as  $\mathbf{x}_t = \mathbf{T} \cdot \mathbf{y}_t + \mu$ . Since the first two principal components capture most of the variability in  $\mathbf{x}$  for a PCA of both the levels and first differences,  $\mathbf{x}_t \approx y_{1,t} \cdot \mathbf{t}_1 + y_{2,t} \cdot \mathbf{t}_2 + \mu$ .

In order to model the evolution of  $\mathbf{x}_t$  over time, the time series properties of  $y_{1,t}$  and  $y_{2,t}$  must be modelled. However, since the link between these scores and the other variables in a full model will depend on the eigenvectors  $\mathbf{t}_1$  and  $\mathbf{t}_2$ , the resulting model may be difficult to interpret. Alternatively, if any two yields (change in yields),  $x_{a,t}$  and  $x_{b,t}$ , are modelled stochastically, these can be used to estimate  $y_{1,t}$  and  $y_{2,t}$ , from which can be derived the full yield curve as explained above. More formally:

Let 
$$\mathbf{z}_t = \begin{bmatrix} z_{a,t} \\ z_{b,t} \end{bmatrix} = \begin{bmatrix} x_{a,t} \\ x_{b,t} \end{bmatrix} - \begin{bmatrix} \mu_{a,t} \\ \mu_{b,t} \end{bmatrix}. \quad (\text{i})$$

Then 
$$\mathbf{z}_t = \begin{bmatrix} t_{1,a} & t_{2,a} \\ t_{1,b} & t_{2,b} \end{bmatrix} \cdot \begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \mathfrak{I} \cdot \tilde{y}_t, \text{ say,} \quad (\text{ii})$$

so 
$$\tilde{y}_t = \mathfrak{I}^{-1} \cdot \mathbf{z}_t \quad (\text{iii})$$

Hence 
$$\mathbf{x}_t = [\mathbf{t}_1 \quad \mathbf{t}_2] \cdot \mathfrak{I}^{-1} \mathbf{z}_t + \boldsymbol{\mu}_t. \quad (\text{iv})$$

Monthly data for the JSE-Actuaries Yield Curve for annual terms between 0 and 25 years is available from January 1986 onwards. Prior to this, only yields on 3- and 20-year bonds are available as well as the Alexander Forbes Money Market Index, from which can be derived a proxy for the short rate. These three series are available from 1960 onwards. Hence, if data for the full period from 1960 to 1998 is required for modelling purposes, it is only possible to model these three points on the yield curve.

If most of the variability in the yield curve could be explained by one principal component, the correlation between yields at different terms would be close to one and the yield at any term would be sufficient to reproduce the entire yield curve. Since two principal components are required to explain most of the variability in the yield curve, we require two terms,  $a$  and  $b$ , to reproduce the entire yield curve. These two terms should be chosen so that the correlation between them is as small as possible in order to minimise the error in estimating  $y_{1,t}$  and  $y_{2,t}$ . The correlation matrix for *JAYC00*, *JAYC03* and *JAYC20* is presented in Table 2.

	JAYC00	JAYC03	JAYC20
JAYC00	1	0.74	0.67
JAYC03	0.87	1	0.92
JAYC20	0.44	0.75	1

Table 2 Yield correlations (1986 – 1998)  
(Level yield correlations below the diagonal  
and differenced yield correlations above)

The correlations between *JAYC00* and *JAYC20* in Table 2 are less than the other correlations suggesting that *JAYC03* can be dropped from the set of model variables. Figures 2 and 4 confirm this suggestion since the greatest differences between the coefficients of the first and second principal components are at the short and long maturities. Further, for most months between January 1986 and December 1998, *JAYC03* lies between *JAYC00* and *JAYC20*. Since the difference in term between *JAYC00* and *JAYC20* is the largest, errors in forecasting *JAYC00* and *JAYC20* have a smaller effect on the forecast error for *JAYC03* than any other pair of yields might have on the remaining yield.

### *Conclusion*

From a statistical perspective, the short rate and the long-bond yield should be used to reconstruct the par yield curve, given the first and second principal components. Hence, in forecasting the yield curve, one need only forecast the short rate and the long-bond yield. If these variables are modelled as non-stationary variables, the yield curve can be reconstructed given forecast changes together with the yield curve at time zero. Otherwise, the yield curve can be reconstructed directly using equations (i) – (iv). A number of other reasons exist for modelling the long-bond yield and the short rate as part of a larger set of variables but a discussion of this is beyond the scope of this paper.

## References

1. Ang, A. and Moore, D. (1994) "The Australian yield curve," *unpublished working paper*, Macquarie University, Sydney, Australia.
2. Brennan, M.J. and Schwartz, E.S. (1982) "An equilibrium model of bond pricing and a test of market efficiency," *Journal of Financial and Quantitative Analysis*, **17**, 75-100.
3. Boyle, P.P. (1978) "Immunsation under stochastic models of the term structure," *J.I.A.*, **105**, 177-187.
4. Cox, J.C., Ingersol, J.E. and Ross, S.A. (1985) "A theory of the term structure of interest rates," *Econometrica*, **53**, 385-407.
5. Falkena, H.B., Fourie, L.J. and Kok, W.J. (1984) *The Mechanics of the South African Financial System*, MacMillan South Africa, 2<sup>nd</sup> edition.
6. INET Graphics and Database Service (1998). Intelligent Network (Pty) Ltd.
7. Johansen, S. and Juselius, K. (1992) "Testing structural hypotheses in a multivariate cointegration analysis of the PPP and UIP for UK," *Journal of Econometrics*, **53**, 211-244.
8. Juselius, K. (1995) "Do PPP and UIP hold in the long-run," *Journal of Econometrics*, **69**, 211-240.
9. McLeod, H. D. (1990) "The development of a market yield curve: the South African solution," *1<sup>st</sup> AFIR International Colloquium*, **2**, 197-212.
10. Mills, T.C. (1994) *The econometric modelling of financial time series*. Cambridge University Press.
11. Pesaran, M.H. and Shin, Y. (1996) "Cointegration and speed of convergence to equilibrium," *Journal of Econometrica*, **71**, 117-143.
12. Redington, F.M. (1952) "Review of the principals of life office valuations," *J.I.A.*, **78**, 286-315.
13. Seber, G.A.F. (1984) *Multivariate Observations*, John Wiley & Sons.
14. Sherris, M. (1995) "Interest rate risk factors in the Australian bond market," *Proceedings of the 1995 AFIR Colloquium*.

15. Stock, J.H. and Watson, M. (1988) "Testing for common trends," *Journal of the American Statistical Association*, **83**, 1097-1107.
16. Tilley, J.A. (1992) "An actuarial layman's guide to building stochastic interest rate generators," *Transactions of the Society of Actuaries*, **44**, 509-564.
17. Vasicek, O. (1977) "An equilibrium characterisation of the term structure," *Journal of Financial Economics*, **5**, 177-188.