Florida International University
FIU Digital Commons

FIU Electronic Theses and Dissertations | University Graduate School

3-30-2004

Principal component and second generation wavelet analysis of treasury yield curve evolution

Mark L. Copper
Florida International University

DOI: 10.25148/etd.FI14061507
Follow this and additional works at: https://digitalcommons.fiu.edu/etd

Part of the Business Administration, Management, and Operations Commons

Recommended Citation
Copper, Mark L., "Principal component and second generation wavelet analysis of treasury yield curve evolution" (2004). *FIU Electronic Theses and Dissertations*. 2521.
https://digitalcommons.fiu.edu/etd/2521

This work is brought to you for free and open access by the University Graduate School at FIU Digital Commons. It has been accepted for inclusion in FIU Electronic Theses and Dissertations by an authorized administrator of FIU Digital Commons. For more information, please contact dcc@fiu.edu.

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

PRINCIPAL COMPONENT AND SECOND GENERATION WAVELET ANALYSIS OF TREASURY YIELD CURVE EVOLUTION

A dissertation submitted in partial fulfillment of the

requirements for the degree of

DOCTOR OF PHILOSOPHY

in

BUSINESS ADMINISTRATION

by

Mark L. Copper

2004

To: Dean Joyce Elam
College of Business Administration

This dissertation, written by Mark L. Copper, and entitled Principal Component and Second Generation Wavelet Analysis of Treasury Yield Curve Evolution, having been approved in respect to style and intellectual content, is referred to you for judgment.

We have read this dissertation and recommend that it be approved.

Gary A. Anderson

Brice Dupoyet

Ali M. Parhizgari

Joel R. Barber, Major Professor

Date of Defense: March 30, 2004

The dissertation of Mark L. Copper is approved.

Dean Joyce Elam
College of Business Administration

Dean Douglas Wartzok
University Graduate School

Florida International University, 2004

ABSTRACT OF THE DISSERTATION

PRINCIPAL COMPONENT AND SECOND GENERATION WAVELET ANALYSIS OF TREASURY YIELD CURVE EVOLUTION

by

Mark L. Copper

Florida International University, 2004

Miami, Florida

Professor Joel R. Barber, Major Professor

Prices of U.S. Treasury securities vary over time and across maturities. When the market in Treasurys is sufficiently complete and frictionless, these prices may be modeled by a function time and maturity. A cross-section of this function for time held fixed is called the yield curve; the aggregate of these sections is the evolution of the yield curve. This dissertation studies aspects of this evolution.

There are two complementary approaches to the study of yield curve evolution here. The first is principal components analysis; the second is wavelet analysis. In both approaches both the time and maturity variables are discretized. In principal components analysis the vectors of yield curve shifts are viewed as observations of a multivariate normal distribution. The resulting covariance matrix is diagonalized; the resulting eigenvalues and eigenvectors (the principal components) are used to draw inferences about the yield curve evolution.

In wavelet analysis, the vectors of shifts are resolved into hierarchies of localized fundamental shifts (wavelets) that leave specified global properties invariant (average change and duration change). The hierarchies relate to the degree of localization with movements restricted to a single maturity at the base and general movements at the apex. Second

generation wavelet techniques allow better adaptation of the model to economic observables. Statistically, the wavelet approach is inherently nonparametric while the wavelets themselves are better adapted to describing a complete market.

Principal components analysis provides information on the dimension of the yield curve process. While there is no clear demarkation between operative factors and noise, the top six principal components pick up 99% of total interest rate variation 95% of the time. An economically justified basis of this process is hard to find; for example a simple linear model will not suffice for the first principal component and the shape of this component is nonstationary.

Wavelet analysis works more directly with yield curve observations than principal components analysis. In fact the complete process from bond data to multiresolution is presented, including the dedicated Perl programs and the details of the portfolio metrics and specially adapted wavelet construction. The result is more robust statistics which provide balance to the more fragile principal components analysis.

TABLE OF CONTENTS

CHAPTER		PAGE
1	Introduction	1
2	Principal Components	9
2.1	Portfolio interest rate risk	10
2.2	Principal components and immunization	12
2.3	Estimating principal components from historical data	13
2.4	Tests of some statistical hypotheses	17
2.5	Possible shortcomings of principal component analysis	27
3	Second Generation Wavelets for Yield Curve Data	29
3.1	The lazy transform	29
3.2	The prediction step	30
3.3	The update step	32
4	Data and Data Preparation	37
5	The Wavelet Transform of Yield Curve Shifts	40
5.1	Update details	42
5.2	The wavelet transform	47
6	Wavelet transforms of Treasury yield curve shifts	53
6.1	The linear model	53
6.2	The distribution of detail coefficients	55
6.3	Power-law distributions	56
6.4	Nonparametric tests	57
6.5	Mean absolute deviation of detail coefficients	60
6.6	Adaptive thresholding	61
7	Conclusion	70
Bibliography		75
Appendix		78
Vita		104

LIST OF TABLES

TABLE PAGE

2.1 Common yield curve nodes, in years, used for principal component analysis. The discount function is estimated each month with a cubic spline; the spline is then interpolated at these nodes. 14

2.2 The twelve largest eigenvalues of the sample covariance matrix for Treasury securities 1992–2001 at nodes listed in Table 2.1. 14

2.3 Test for equality of the smallest $q = m - k$ sample covariance matrix eigenvalues. The statistic P_k has an asymptotically χ^2 distribution. The p-value shows there is no support for equality at any level. However the standard deviations, $\sqrt{l_q}$, of the smaller principal components are small compared to measurement uncertainty. 20

2.4 Confidence intervals and actual observed values of proportions of variation captured by the first 6 principal components. 21

2.5 Values of the statistic W for linear first principal components with slopes as given in the first column. W is minimal for linear vectors with slope near -0.008, but the upper 1/2% point of $\chi^2_{47} = 75.7$ and the hypothesis of a linear first principal component is rejected for all slopes. The second column shows how it is the projection of the linear vector along the principal components with small eigenvalue that overwhelm W. 25

5.1 Wavelet transform (multiresolution analysis) of the yield curve shift in January, 1992. Read this table from left to right, then top to bottom. That is, with zero offset, coefficients whose indices are divisible by 8 are found in the first column. 51

6.1 The average thresholding parameter λ determined for each level by the SURE and Ogden-Parzen methods 65

LIST OF FIGURES

FIGURE PAGE

2.1 Principal components derived from monthly McCulloch yield curve estimate changes. Parameters derived by R. Bliss from 1992–2001 CRSP Treasury data using McCulloch's programs. 15

2.2 Principal components derived from monthly McCulloch yield curve estimate changes. Parameters derived by R. Bliss from 1992–2001 CRSP Treasury data using McCulloch's programs. 16

2.3 p-values for the explanatory power of the 4 largest principal components. 90% confidence intervals are indicated by the vertical dotted lines. The vertical axis is scaled logarithmically for better visualization. 22

2.4 Is the first principal component stationary? The principal components corresponding to the largest covariance matrix eigenvalues are shown. Ten years' data from 1991 through 1992 are broken into two successive five year windows, and the principal components for each interval are computed. A statistical test for equality of the principal components corresponding to the largest variances is negative unless data corresponding to the 42 smallest variances is neglected. 26

3.1 Schematic for the "lazy" transform, the first step in the wavelet transform. y_0 is the input sequence, L is the transform, and u_{-1} and l_{-1} are the output sequences. 30

3.2 A pictorial representation of the prediction step of the wavelet transform. The height of the vertical line above x_{2k+1} is the prediction of $u_{-1,k}$ which is obtained by linearly interpolating $l_{-1,k}$ and $l_{-1,k+1}$. The difference between $u_{-1,k}$ itself and its predicted value is the difference coefficient, $d_{-1,j}$. 31

3.3 Schematic of the prediction step of the wavelet transform. Transform L splits y_0 into two signals as in Figure 3.1. The lower signal, l_{-1}, is output as before, but now it is also used to form a prediction of the upper signal as well. This prediction is subtracted from the true upper signal, u_{-1}, to form a signal of differences, d_{-1}, which is then output alongside l_{-1}. 31

3.4 Hatched area and its moment represent global information lost in the prediction step. 33

3.5 Pictorial representation of the update step; information lost in prediction (Figure 3.4 is reincorporated in such a way as to preserve both area and moment of area. 34

3.6 Scheme of the entire wavelet transform through lifting 35

3.7 Multiresolution analysis of y_0. Repeated application of L, P, and U generates a series of sequences of detail coefficients $d_{-1}, \ldots, d_{-j}$ and the coarsest possible approximation y_{-j}. The process is invertible, so y_0 can be recovered from the d_k and y_{-j}; that is, no information is lost in multiresolution analysis. 36

5.1 Beginning from the coarse approximation of the yield curve shift obtained by the wavelet transform, level -8, one can work back to the raw data by adding the details back in. One simply multiplies the wavelets for January 1992 (see appendix) by the coefficients from Table 5.1 and adds the resulting functions to the previous level's approximation. 52

6.1 Changes in ten year rates are plotted against changes in one day rates for 119 months. The best linear fit is superposed. Correlation between rate changes is positive but weak with $R^2 = 3.3\%$. The slope of the regression line is .30, significant at the 5% level. 54

6.2 Cumulative distribution of coarsest scaling coefficients (squares) and all detail coefficients (circles). About 70% of scaling coefficients are larger than 10bp while only 20% of detail coefficients are. 54

6.3 The detail coefficients are ranked in one hundred quantiles according to size. The median value from each is plotted against the middle of the corresponding quantile of the standard normal distribution. The extent to which the graph departs from linear is a measure of the extent to which the aggregate of detail coefficients fails to be normal. 55

6.4 This plots the fraction of detail coefficients greater than the argument on a log-log scale. If a power-law distribution provided a good model, the graph would be linear. 57

6.5 The mean absolute deviation (MAD) of each of the 255 detail coefficients is computed and sorted into the corresponding multiresolution levels. Data is taken from monthly observations during 1992–2001. Levels are -8 to -1 top to bottom. A common scale is used throughout; the tallest bar is about 42 basis points. 62

6.6 Scatter plot of 119 months of 128 level -1 detail coefficients before and after thresholding using the Ogden-Parzen parameter in Table 6.1. 68

6.7 Scatter plot of 119 months of 64 level -2 detail coefficients before and after thresholding using the Ogden-Parzen parameter in Table 6.1. 69

1 Introduction

The following study examines and compares two approaches to interest rate risk. The first approach centers around the principal component analysis of sample covariance matrices derived from the history of the term structure of interest rates. This approach was first used in [1] and is closely related to the use of common factor analysis by Litterman and Scheinkman in [20]. Moreover, this approach has continued to be of interest to researchers, both as a simple model for term structure evolution [8] [14] [30] and for other, closely related interest rate risk questions [3] [7] [13] [19]. Here I update the principal component analysis of [1] for more recent data, I perform statistical tests of hypotheses occasioned by principal component analysis, and I consider some of the limitations of the principal component approach.

In the second approach term structure shifts are decomposed into atoms called "wavelets". In contrast to principal component analysis where the basis best adapted to the data is chosen, wavelet bases are formed subject to certain structural constraints. Although there is a loss of efficiency in the choice of a more structured basis, the structure provides an increased ability to capture local trends and fluctuations transparently.

Although the term "wavelet" itself appears to be fairly new, the mathematical development it pertains to arises from mainstream mathematical analysis. See Meyer's historical perspective [24]. And like evolutionary convergence essentially the same ideas developed independently in several different engineering and scientific fields. See Sweldens' and Schröder's overview in the introduction to [32]. It is no surprise then that these techniques

of "signal processing" would be applied to financial time series as well. In fact, there are now book-length treatments of just this topic available [12].

To my knowledge, however, there has been no attempt to apply wavelet analysis to term structure evolution. Heretofore, the barrier to a using wavelet analysis as described in [24] is quite basic; the key ingredient in the development of the theory was a family of basis functions that remained invariant under both translations and dilations. Although this made the full arsenal of Fourier analysis available, it did not match up well with term structure data. The term structure, as a construct, derives from a family of fixed income securities where maturities are clustered toward the short term in no neat pattern, where all maturities are constrained to an interval, and where translation invariant measure does not appear to have a reasonable economic interpretation.

On the other hand, these types of constraints are commonly encountered in many applications, not just finance and economics, and several authors addressed them. A line of development particularly promising for our purpose is due to Wim Sweldon, whose own work was inspired by the earlier investigations of Donoho [5] and Lounsbery [21].

"Second generation" wavelet analysis [31], as Sweldens has called it, obtained central elements of wavelet analysis without the Fourier transform and in such a way that the analysis could be extended easily to irregularly spaced observations on a finite interval equipped with any one of many possible metrics. Moreover, Sweldens's approach provides a way to exploit the latitude in the choice of wavelet basis so that the optimal wavelet basis for modeling yield curve shifts might be used.

Armed with this innovation, I returned to term structure analysis with two goals. First, I wanted to see if a more rational framework for approximation than used for principal

components analysis would provide insights into interest rate movement. Specifically, the yardstick in principal components analysis is finite dimensional euclidean measure. Prior to analysis, a yield curve is abstracted from securities prices; the yield curve is evaluated at a vector of (usually hypothetical and certainly arbitrary) maturities. This is done at two successive times, and the magnitude of the yield curve change during the intervening time is the sum of squares of differences between between the yield curve values along the vector, no matter how the vector of maturities is distributed. This, it appeared to me, is disconnected from the economics; not only do sums of squares have no economic meaning, but the location of a yield change along the spectrum of maturities certainly does. It is true, as shown in Sections 2.1 and 2.2, that we can reconnect to the economics further down the line, but it was still a question to me whether a more direct approach might be fruitful.

A more compelling alternative measure of the magnitude of an interest rate change would be the impact of a yield curve shift on the value of a portfolio. Using Swelden's approach, we may choose a portfolio, the market portfolio, say, and construct wavelets which, as term structure shifts, have no effect on portfolio value. Thus, when wavelet analysis begins with the raw data of a yield curve shift and rewrites it as a sum of a simple smooth curve and wavelets, it follows that the smooth curve represents exactly the same impact on the value of the portfolio as the input data. One may then study the joint distribution of the curve parameters, or one may study the wavelets, but in either case, one maintains an economic connection to the original data.

The euclidean nature of principal components analysis has another important facet. Principal components analysis "diagonalizes" the covariance matrix. That is, there is a reformulation of our observation vector that greatly simplifies the structure of the multi-

variate random variable that our yield curve shifts comprise. Our "raw" data for yield curve changes consists of observations of interest rate changes for an array of individual zero coupon bonds. Now naive observation indicates that there is significant correlation of interest rate changes among these bonds, especially among bonds with similar maturities. Principal components analysis extends this intuition by extracting patterns of interest rate movements.

Each pattern extracted by PCA comprises a vector of perfectly correlated interest rate movements among the bonds, there is no correlation among the different patterns themselves, and, taken together, the patterns may be used to construct any individual observation by scaling and superposition. Moreover, these patterns are arranged into a hierarchy of variances which measures the contribution each pattern makes to the whole. In fact, PCA may capture real simplicity when it might not be apparent otherwise. That is, the effective rank of the covariance matrix may be very small; nearly all the observed interest rate movement may be accounted for by of a handful of patterns. It is, perhaps, an interesting occurence that interest rate movements do exhibit such simplicity; this is explored in Section 2.4.2.

Like a dog catching the proverbial bus, however, it is a little unclear what to make of these simplifying patterns once we have them, and it is this problem that underlies the statistical investigation that follows. To begin with, it would be interesting to relate the principal components to economic factors, much as Litterman and Scheinkman [20] try to do. Failing that, perhaps there are simple rules that can reproduce the top principal components, and these simple rules, if not economically justified *a priori* may be justified *a posteriori* by being useful. Unfortunately, the statistical tests we perform seem to cast more shadow than light. There is no support for the simplest economic hypothesis, that interest

rates are all essentially the same type of thing and respond in unision to economic changes. In the tests that follow (Section 2.4), we see that the dominant principal component is not parallel translation, or even a line at all. And whatever it is, it does not persist; that is, we need more than just chance to explain the differences between the dominant principal components over different epochs.

This is all vaguely disconcerting. The dominant principal component *looks* as if it could reasonably be approximated by a nearly flat line, and it *looks* as if the dominant principal component persists over successive five year periods. Furthermore, it seems odd that we must jump all the way from interest rate changes on single bonds to patterns of perfectly correlated changes all across the yield curve, that in order to incorporate exceptional behaviour in a neighborhood of a single maturity, we must involve the entire yield curve.

Certainly, it is possible to indicate the gradation of correlation of rate movements among bonds with closer maturities more highly correlated than more distant maturities, but this is done by superposition of patterns, each of which involves the entire yield curve. In fact, it is quite informative to see how articulation is introduced as one incorporates principal components with lower variances. Looking at the principal components in Figures 2.1 and 2.2 and looking at the variances in Table 2.2, one can "see" how interest rate movements are predominantly positively correlated across the yield curve with secondary decouplings brought in by various yield curve twists associated to the principal components beyond the first.

Nevertheless, there is something stiff and unforgiving about PCA. I do not know whether a more careful formulation would alleviate this jarring against intuition or whether distributional assumptions underlying the statistical tests should be revised or whether intuition

should simply submit to the statistical results as they stand. It would be nice to have an independent take on the phenomena, one that could transparently represent phenomena at all scales from local to global. Thus my second goal in exploring the wavelet analysis of term structure fluctuations has been to provide a more flexible basis for modeling term structure shifts.

The data principal components analysis begins with, as presented in Chapter 2, is not strictly raw data but rather a yield curve approximation at each time of observation. In reality, choosing the appropriate approximation is not a simple matter as Bliss [2] has shown. I have used Bliss's implementation of McCulloch's [22] method,[1] in which cubic splines are regressed against raw data, the nodes of which splines are chosen adaptively. Once the spline approximations to the yield curve have been made at each observation time, all the splines are evaluated at the same fixed set of maturities. This generates the multivariate sample to which PCA is applied.

Wavelet analysis, often called "multiresolution analysis" in this context, begins with a linear spline interpolating a raw zero-coupon yield curve, coupon Treasurys having been stripped with a bootstrap method.

Multiresolution analysis then generates a cascade of splines that approximate an interest rate movement. The approximations are reminisent of, but quite distinct from, OLS regression. The cascade consists of a series of splines with progressively fewer nodes, each one a smoother but coarser approximation of the original data. Our splines[2] all retain the same portfolio weighted average change and duration as the input shift.

[1]Another formulation question that would be interesting to check is whether PCA is substantially affected by the choice yield curve approximation method.

[2]Although our splines are linear here, higher (and lower) order splines may be used in much the same way; see [32] for an indication.

At each step of the cascade, the information that is lost in going from a finer to a coarser approximation is stored in a wavelet. Thus, at each step, the magnitude of the wavelet indicates the quality of the approximation at the next level of smoothness and the support of the wavelet indicates the scale at which the adjustment is being made.

At the end of the multiresolution process, one is left with a simple polynomial approximation (i.e. a spline on one interval) of the yield curve and a collection of wavelets. At each level one has separated the yield curve shift into locally disassociated movement, captured by the wavelet, and coherent movement, captured by the yield curve shift approximation at the next level, and one has repeated this process as long as it was possible to do so. No information has been lost, and input data can be recovered completely by reversing the process.

With wavelet analysis we have a way to model term structure shifts with changes restricted to segments of the term structure on a full gamut of scales of resolution. Simple order statistics of the wavelet coefficients indicate where introducing more detail into the yield curve shift approximations will most efficiently improve the model. Thus, just as principal components with smaller and smaller associated variances may be introduced to refine a PCA model of yield curve shifts, the location of wavelet coefficients indicate where new nodes might be most effectively added when refining the multiresolution model.

Since I know of no exposition of second generation wavelet analysis in the finance literature, I have included an introduction to the simple techniques I use here in Chapter 3. The devil being in the details, I have also included in the appendix the program listings which I developed for extracting, preparing and transforming data from the CRSP bond file, and algorithm details are exposed in Chapter 5.

Gross results seem to indicate phenomena similar to those from PCA. Comparing Figures 2.1 and 6.5, there appears to be a correspondence between the fact that the first wavelet coefficient at each level is the largest and the fact that first principal component is convex for the first five years along the yield curve. That there should be no good linear approximation to the first principal component might be related to the apparent independence of short and long term rate changes displayed in Figure 6.1. Finally, the linear approximation generated by multiresolution cascade only rarely fails to provide a good fit to the observed change (Figure 6.2); this could be related to the rapidity with which the principal component variances decay (Figure 2.3).

This last observation raises a question about the extremes of the wavelet coefficient distribution. Might they be larger than would be expected if the deviation from 0 where simply a Gaussian noise term? Figure 6.3 shows large wavelet coefficients occur more frequently than would be compatible with the large number of very small terms if all terms were just noise. This suggests a possible signal-noise model—perhaps coefficients are normally distributed with small variance most of the time, but when an infrequent signal arrives from the economy the coefficients are normally distributed with much larger variance—but it forces one to use nonparametric methods for further statistical examination of the wavelet coefficients. I did perform nonparametric tests to confirm the appearance that wavelet coefficients at larger scales tend to be larger, and I concluded Chapter 6 with a look at two methods that attempt to separate signal from noise, most especially with the object of resizing wavelet coefficients by level in order to refine the apparent relative contributions of the wavelet coefficients to fitting yield curve shifts.

The goal of principal component analysis is to clarify the correlations among the components of a multivariate random variable. We use these relations to detail the interest rate risk of a fixed income portfolio.

Suppose that X is a multivariate random variable, and suppose that Σ is its covariance matrix. Since Σ is a positive symmetric matrix, Σ may be factored

$$\Sigma = H\Lambda H' \tag{2.1}$$

where H is a rotation matrix, $HH' = H'H = I, \det(H) = 1$, and Λ is a positive diagonal matrix, $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_m)$, with $\lambda_j \geq 0, j = 1, \ldots, m$. It is convenient to assume, as we may, that the λ_j are arranged in decreasing order, $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_m$. With this assumption, the columns of H are unique up to sign. Let U_j denote the jth column of H. The U_j are known as the principal components of X. See [25] for details.

In order to fix ideas, let us assume that the probability distribution of X is multivariate normal, $N(\mu, \Sigma)$. In this case the density function of X is

$$f_X(x) = (2\pi)^{m/2}(\det \Sigma)^{-1/2} \exp\left[-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right] \tag{2.2}$$

Now substitute $H\Lambda H'$ for Σ and write $Y = H'X$. Then the density (2.2) factorizes as

$$f_Y(y) = \prod_{j=1}^{m} \frac{1}{\sqrt{2\pi\lambda_j}} e^{-\frac{(y-H'\mu)_j^2}{2\lambda_j}} \tag{2.3}$$

That is, simply by rotating the axes appropriately, we may view the distribution of X as the joint distribution of m independent univariate normal random variables.

In our application, X will denote a change in the yield curve for U.S. Treasurys. Then the factorization (2.3) provides an alternative view of the way yields change on the family

of Treasury securities. The basis in which X is first presented derives from observations of bond prices, and X_i denotes the yield change for zero-coupon bonds that mature in t_i units of time from the present. The alternate view recombines yield movements on individual but correlated bonds into uncorrelated patterns of movement. The first pattern is the unitary combination of yield movements with maximum variance; the second is the unitary combination with maximum variance subject to the constraint of being uncorrelated to the first pattern, and so on. Thus, in the alternate basis X_1 is the contribution to X made by the pattern whose contribution is expected to be largest among all unitary patterns. We sacrifice simplicity of representation in order to achieve simplicity of interrelation.

Geometrically, the level surfaces of the function k_X, defined on $\mathcal{R}^m$ by

$$k_X(x) = (x - \mu)'\Sigma(x - \mu),$$

form an ellipsoid, and the components of $Z = H'X$ lie along the semi-axes of the ellipsoid. Knowing these semi-axes, the analyst may identify levels of interest rate risk in a family of portfolios.

2.1 Portfolio interest rate risk

Let A denote a portfolio whose risk we examine. That is, with a schedule of cash flow dates fixed, A_i is the cash flow scheduled for portfolio A at time $t_i, i = 1, \ldots, m$. Similarly, let d_i be the present market value of a \$1 payment scheduled for time t_i. As a notational convenience let T and D denote the diagonal matrices $T_{ii} = t_i$ and $D_{ii} = d_i$ respectively. It

is also convenient to let $B = DA$ and to let ι the $m \times 1$ column vector of ones. Thus, for example, the market value of A is $V = \iota' B$.

If the discount function D should change in such a way that

$$T^{-1}\log(D\iota) \to T^{-1}\log(D\iota) + X$$

then, clearly, the portfolio value $V = \iota' B$ is a non-linear function of X. However, V is differentiable and the best approximation to the change in the value of V by a linear function is given by the differential applied to X:

$$\nabla V(0)X = X'TB. \tag{2.4}$$

We therefore regard the variance of $X'TB$ as the approximate variance of portfolio value $\mathrm{var}(V)$. Since we are assuming that X is $N(\mu, \Sigma)$, this approximate variance is

$$B'T\Sigma TB. \tag{2.5}$$

Using (2.5) as our measure of risk, then, a portfolio manager might ask for the portfolio with present value \$1 and minimal interest rate risk. It is straightforward to see, using Langrange multipliers for instance, that this portfolio must be

$$\frac{T^{-1}\Sigma^{-1}T^{-1}\iota}{\iota' T^{-1}\Sigma^{-1}T^{-1}\iota}.$$

Unfortunately, this mathematical eloquence may not translate to a practical alternative for the portfolio manager. After all, the zero coupon bonds maturing at times t_i were purely hypothetical constructions. Principal component analysis may be a more flexible tool for moderating interest rate risk in conjunction with market realities and manager goals.

Substituting the principal component decomposition of Σ (2.1) into our working measure of portfolio interest rate risk (2.5), we obtain

$$B'T\Sigma TB = \sum_i \lambda_i (H'TB)_i^2. \tag{2.6}$$

Recalling that $\lambda_1 > \lambda_2 > \ldots > \lambda_m$, we see the most volatile component of our portfolio is $(H'TB)_1$. Since we have that

$$TB = \sum_i (H'TB)_i U_i,$$

we interpret $(H'TB)_1$ as the projection of TB along the principal component U_1. Thus one step in reducing interest rate risk that may be practical is simply to select a portfolio A such that the projection of TDA along U_1 is minimal. A conservative estimate of the benefit of such a step may be obtained as follows. From (2.6) and the Cauchy inequality it follows that if the value of the portfolio A equals 1, $B\iota = 1$, then our risk measure (2.5) satisfies

$$B'T\Sigma TB \leq \lambda_1 \max_{B\iota=1} \sum_i t_i^2 B_i^2.$$

If, in addition, it is required that $U_1'TB = 0$, then our risk measure satisfies

$$B'T\Sigma TB \leq \lambda_2 \max_{\substack{B\iota = 1 \\ U_1'TB = 0}} \sum_i t_i^2 B_i^2.$$

In the historical data below, we find that λ_2 is less than 1/3 the size of λ_1.

If TDA contains no component of U_1, then we say that A is immunized along U_1.[1] Similarly, should interest rate risk reduction be high enough a priority, then TDA may be

[1]Although we have vectorized notation, this use of the term "immunize" is completely classical in the case $U_1 = \iota/\sqrt{m}$

chosen to minimize its projection along several of the most volatile principal components. Moreover, since these constraints, $-\epsilon \leq (H'TDA)_i \leq \epsilon$ for example, are linear, they can easily be integrated with other practical constraints in a linear program.

Since immunization is based on the gradient (2.4), it is strictly a "local" technique, properly applied only when the shock considered transpires over a "small" interval of time. Just as the gradient method is used to integrate differential equations (cf. [10], p. 263), so the stochastic calculus provides processes that integrate "infinitesimal" conditions on random variables. The description of these processes is greatly simplified by their diagonalization, and they are presented as such in Heath, Jarrow and Morton [15], and, indeed, Jarrow [17] uses principal components to construct discretized models of approximate solutions to their stochastic differential equations.[2]

2.3 Estimating principal components from historical data

Using historical data to estimate principal components in the evolution of the yield curve requires two levels of abstraction. First a zero-coupon yield curve must be inferred from a universe of securities, many of which are coupon bearing. Second, to define and compare yield curve changes, every constructed zero-coupon yield curve must be interpolated at a common set of nodes. For the estimates presented below, we use Bliss's unfiltered imple-

[2]The reader may wonder why local methods would be of any interest at all when global methods exist. One difficulty is simply that the global methods are considerably more complicated and for all that remain deficient in ways. For example, as Rogers in [28] points out, the Heath-Jarrow-Morton model does not provide workable formulae in cases where the instantaneous spot rate is constrained to be positive. Rogers indicates this problem may be inherent to their approach, beginning, as it does, "by trying to model *derived* quantities (the forward rates) instead of the *fundamental* quantity (the spot rate)."

0.083	0.167	0.250	0.333	0.417	0.500	0.583	0.667
0.750	0.833	0.917	1.000	1.083	1.167	1.250	1.333
1.417	1.500	1.750	2.000	2.500	3.000	4.000	5.000
6.000	7.000	8.000	9.000	10.000	11.000	12.000	13.000
14.000	15.000	16.000	17.000	18.000	19.000	20.000	21.000
22.000	23.000	24.000	25.000	26.000	27.000	28.000	29.000

Table 2.1: Common yield curve nodes, in years, used for principal component analysis. The discount function is estimated each month with a cubic spline; the spline is then interpolated at these nodes.

0.0247224	0.0067709	0.0012182	0.0005729	0.0002075	0.0001106
0.0000804	0.0000599	0.0000294	0.0000160	0.0000121	0.0000063

Table 2.2: The twelve largest eigenvalues of the sample covariance matrix for Treasury securities 1992–2001 at nodes listed in Table 2.1.

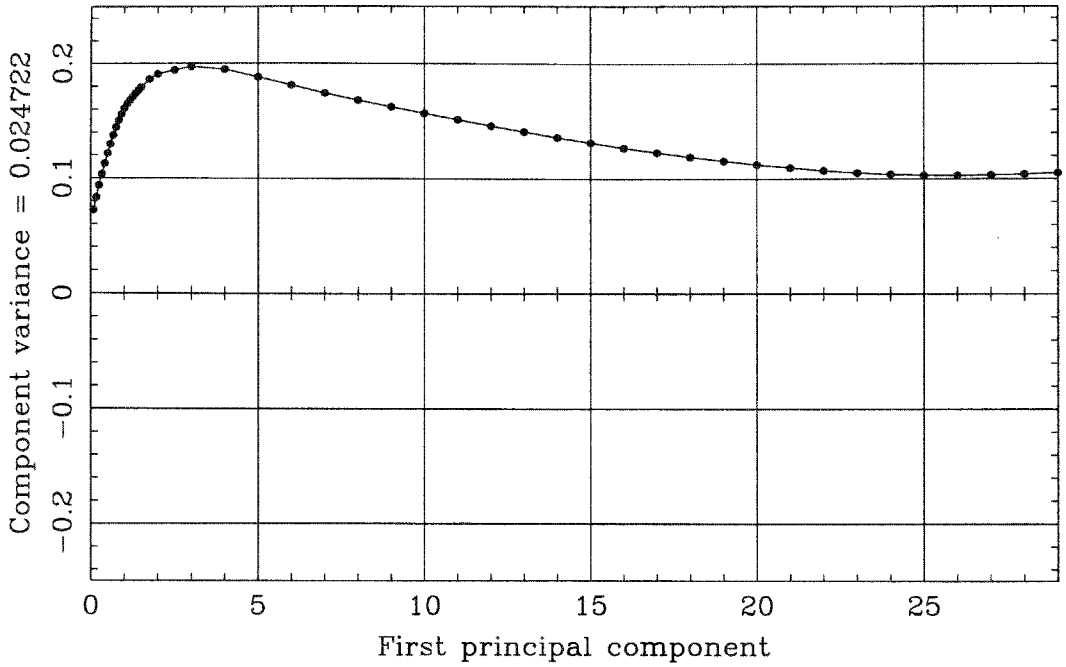
mentation of McCulloch's cubic spline regression for the zero-coupon yield curve.[3] For the common set of nodes we use McCulloch's own selection in data published with Kwon [23]. These nodes are given in years in Table 2.1.

A Perl program, listed in the appendix as `mcmunge.pl`, selects ten years' data and formats it for input to the second Perl program, listed as `csinterpol.pl`. This second program computes the sample covariance matrix and extracts the principal components and their eigenvalues from it. Table 2.2 lists the twelve largest eigenvalues. The remaining eigenvalues are small, but as our tests show, still significantly different from both zero and each other. Graphs of the first four principal components are given in Figures 2.1 and 2.2.

There seems to be some persistance in the shape of the principal component graphs. See Barber and Copper[4] [1] and even Litterman and Scheinkman [20] who use factor analysis

[3]Details are given in Bliss's paper [2]; Mr. Bliss kindly sent me updated data.

[4]A graph of the fourth principal component for McCulloch and Kwon's 8/85–2/91 data is available from the authors



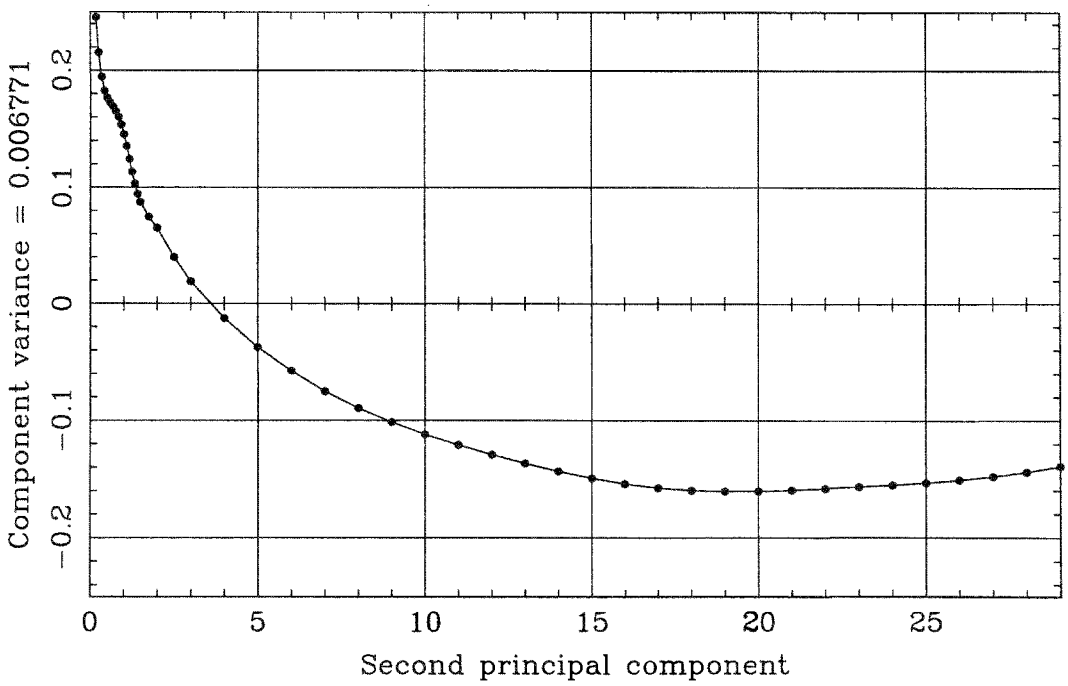
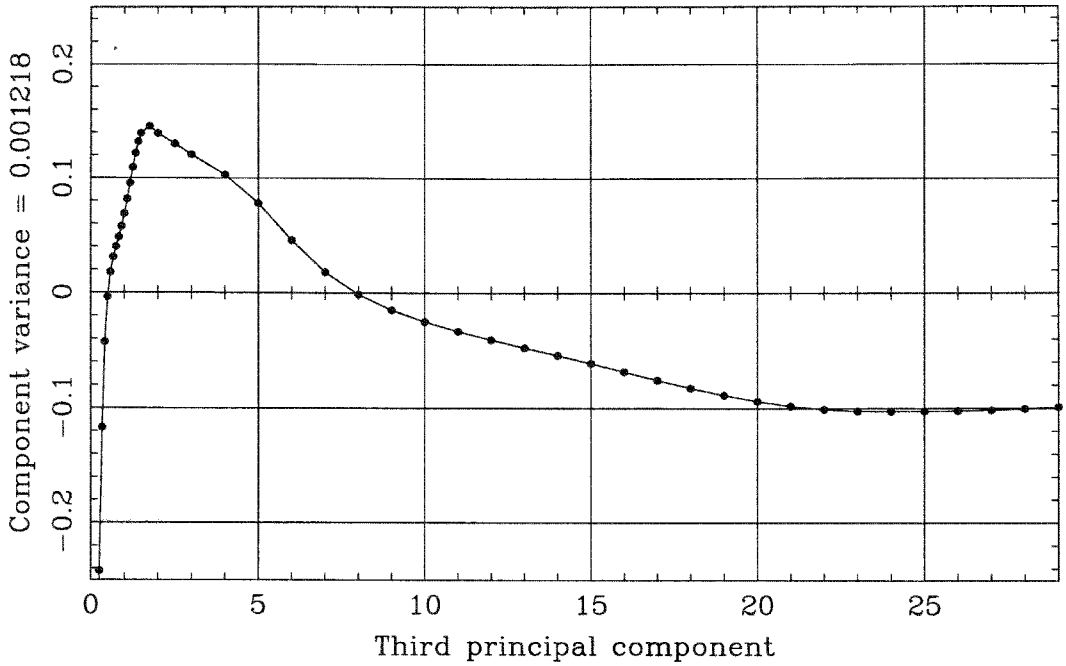


Figure 2.1: Principal components derived from monthly McCulloch yield curve estimate changes. Parameters derived by R. Bliss from 1992–2001 CRSP Treasury data using McCulloch's programs.



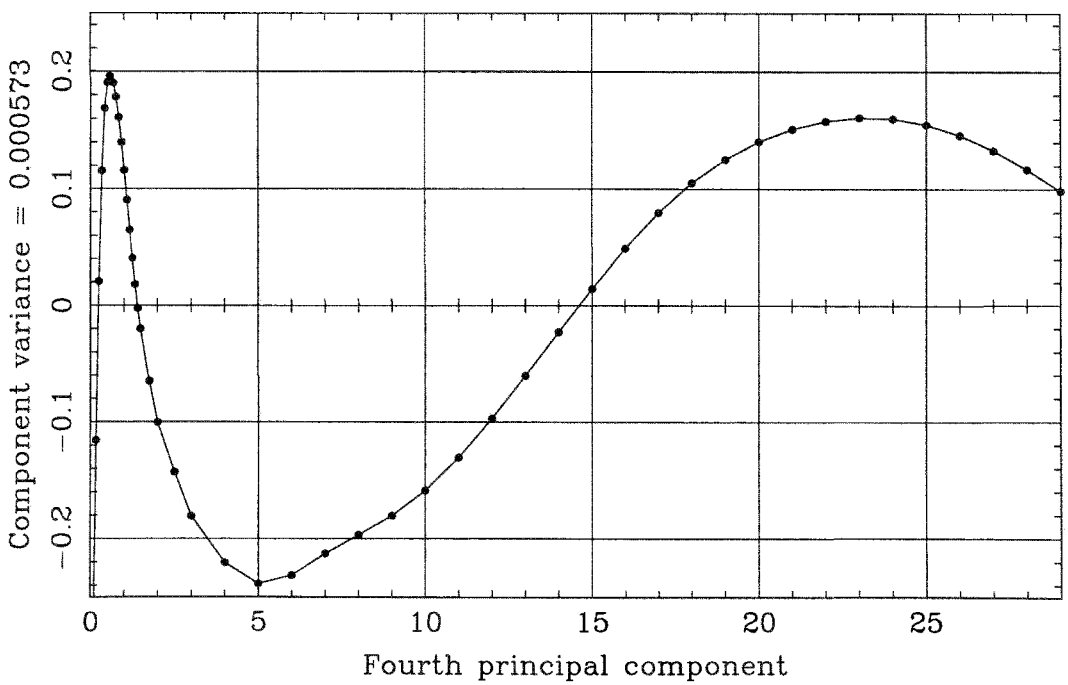


Figure 2.2: Principal components derived from monthly McCulloch yield curve estimate changes. Parameters derived by R. Bliss from 1992–2001 CRSP Treasury data using McCulloch's programs.

applied to weekly returns for 1/84–6/88 and unspecfied interpolation technique.[5] However, some caution in drawing conclusions directly from the graphs seems also to be warranted. For example, Golub and Tilman's graphs ([14], p. 94) don't seem to line up quite as well. Perhaps more quantitative analysis will aid the intuition conveyed by the graphs.

2.4 Tests of some statistical hypotheses

The tests described in this section were developed by several statisticians, Lawley, James, Anderson, to name, perhaps, the principals. We rely on the exposition by Muirhead [25]. All tests depend critically on the assumption that the random variable is multivariate normal. All three tests on our sample data were implemented by the Perl program listed as `stats.pl` in the appendix.

2.4.1 Equality of eigenvalues

First we test hypotheses that the smallest eigenvalues of the principal components are equal to one another:

$$H_k : \lambda_{k+1} = \cdots = \lambda_m$$

for $k = 0, 1, \ldots, m-2$.

In a sense this is a test for randomness. For to the extent that the eigenvalues λ_j are equal, the principal components are not unique. In the extreme case, if the hypothesis H_0 were true, the distribution of X would be spherical, no direction of interest rate movements

[5] See [18] for more detailed descriptions of their techniques; see [4], Section 6.4.1, for some comparison between factor and principal component analysis.

being any more likely than another. That is, there would be no coherence whatsoever among movements at the different maturities.

On the other hand, H_k can also be related to the dimension of the space needed to model yield curve shifts. To illustrate, suppose that H_1 were accepted, and suppose it were determined that the eigenvalues $\{\lambda_2, \ldots, \lambda_m\}$ were statistically indistinguishable from 0. That would essentially be the classical case of parallel translation; all yield curve shifts being dilations of a single prototype. More generally, if H_k is acceptable for k "small" and the common value $\lambda = \lambda_{k+1} = \cdots = \lambda_m$ is also "small", we will see evidence for a k-factor interest rate evolution with observed fluctuations in directions associated to the eigenvalue λ due merely to measurement error.

In order to formulate the statistic used to test H_k, we set

$$\tilde{l}_q = \frac{1}{q} \sum_{i=k+1}^{m} l_i,$$

where the l_i are the eigenvalues of the sample covariance matrix S and, for notational convenience, $q = m - k$, the number of hypothetically equal eigenvalues. Then set[6]

$$V_k = \frac{\prod_{i=k+1}^{m} l_i}{\tilde{l}_q^q}.$$

When H_k is true, the distribution of $-n \log V_k$ is asymptotically $\chi^2_{(q+2)(q-1)/2}$ as $n \to \infty$. However, in our test cases where $n = 119$ for 10 years of monthly data and $n = 59$ for 5 years, we employ the more accurate statistic developed by Bartlett and Lawley,

$$P_k = -\left[n - k - \frac{2q^2 + q + 2}{6q} + \sum_{i=1}^{k} \frac{\tilde{l}_q^2}{(l_i - \tilde{l}_q)^2}\right] \log V_k,$$

[6]Note that $0 \leq V_k \leq 1$ since the l_i are eigenvalues of a positive matrix and the geometric mean of these values is always less than the arithmetic mean.

having the same asymptotic distribution but faster convergence as $n \to \infty$. See Table 2.3 for the results of this test.

We see that the data are quite highly structured and there is no case at all that the smallest q eigenvalues are equal for any q, $q = 48, \ldots, 2$. We obtain, but do not display, similar conclusions using five year data and using coarser meshes. It would appear there is no "natural" dimension for interest rate fluctuations; however, there is an impracticality to such a conclusion. For if we do not conclude, say, that the smallest two eigenvalues are equal, we are left assuming the second smallest eigenvalue is nonzero. In particular, variation in the direction of the corresponding principal component is thus assumed to be about 10^{-7}. whereas one basis point is only 10^{-4}. We examine the case for a "practical" dimension in the next section.

2.4.2 Proportion of variation

In our first test we did not see any inherent limit to the dimensionality of the space of observed yield curve fluctuations. Eventhough we cannot dismiss any of the eigenvalues as insignificantly different from 0, we can still test whether the fraction of portfolio variance explained by the last k principal components is a relatively small amount, h, say. Formally, we consider the hypothesis,

$$H_k^* : \frac{\sum_{i=k+1}^{m} \lambda_i}{\sum_{k=1}^{m} \lambda_i} = h.$$

These hypotheses can be tested with the statistics

$$M_k = -h \sum_{i=1}^{k} l_i + (1-h) \sum_{i=k+1}^{m} l_i.$$

For if H_k^* is true, then (see [25], p. 416) $\sqrt{n} M_k$ is asymptotically $N(0, \tau^2)$ as $n \to \infty$ where

$$\tau^2 = 2h^2 \sum_{i=1}^{k} \lambda_i^2 + 2(1-h)^2 \sum_{i=k+1}^{m} \lambda_i^2.$$

q	P_k	p-value	$\sqrt{l_q}$
48	101286.16	0.000000e+00	1.572335e-01
47	51808.85	0.000000e+00	8.228546e-02
46	45557.80	0.000000e+00	3.490286e-02
45	42303.59	0.000000e+00	2.393510e-02
44	39209.77	0.000000e+00	1.440518e-02
43	37206.09	0.000000e+00	1.051532e-02
42	35579.12	0.000000e+00	8.968597e-03
41	33788.61	0.000000e+00	7.742445e-03
40	31570.91	0.000000e+00	5.426519e-03
39	29724.92	0.000000e+00	4.004516e-03
38	28130.74	0.000000e+00	3.472371e-03
37	26126.87	0.000000e+00	2.516362e-03
36	24281.56	0.000000e+00	1.754004e-03
35	22809.32	0.000000e+00	1.368457e-03
34	21394.59	0.000000e+00	1.163865e-03
33	19647.14	0.000000e+00	9.312330e-04
32	17397.35	0.000000e+00	5.901655e-04
31	15501.61	0.000000e+00	3.613146e-04
30	14244.15	0.000000e+00	2.805237e-04
29	12990.39	0.000000e+00	2.150908e-04
28	11744.50	0.000000e+00	1.831290e-04
27	9930.76	0.000000e+00	1.034678e-04
26	8853.57	0.000000e+00	7.523517e-05
25	7938.86	0.000000e+00	5.830019e-05
24	7063.72	0.000000e+00	3.963088e-05
23	6500.31	0.000000e+00	3.652810e-05
22	5751.86	0.000000e+00	2.789453e-05
21	5079.29	0.000000e+00	1.959196e-05
20	4629.85	0.000000e+00	1.426484e-05
19	4351.82	0.000000e+00	1.294142e-05
18	4053.58	0.000000e+00	1.192338e-05
17	3698.73	0.000000e+00	9.670846e-06
16	3391.38	0.000000e+00	9.324346e-06
15	2920.66	0.000000e+00	7.464477e-06
14	2401.97	0.000000e+00	4.630039e-06
13	2137.11	0.000000e+00	4.099755e-06
12	1803.86	0.000000e+00	3.398030e-06
11	1390.28	0.000000e+00	2.108780e-06
10	1169.35	0.000000e+00	1.713246e-06
9	941.08	0.000000e+00	1.420020e-06
8	642.31	0.000000e+00	8.370498e-07
7	510.39	0.000000e+00	6.257029e-07
6	419.42	0.000000e+00	5.224277e-07
5	320.53	6.105600e-58	4.226790e-07
4	195.63	2.727400e-35	2.840302e-07
3	85.50	5.920500e-15	1.653242e-07
2	25.99	2.270300e-04	9.708477e-08

Table 2.3: Test for equality of the smallest $q = m - k$ sample covariance matrix eigenvalues. The statistic P_k has an asymptotically χ^2 distribution. The p-value shows there is no support for equality at any level. However the standard deviations, $\sqrt{l_q}$, of the smaller principal components are small compared to measurement uncertainty.

	1 p.c.	2 p.c.'s	3 p.c.'s	4 p.c.'s	5 p.c.'s	6 p.c.'s
Lower 5%	67%	92%	96 %	98.1%	98.8%	99.2%
Observed	73%	93%	96.7%	98.4%	99.0%	99.4%
Upper 5%	78%	94%	97 %	98.7%	99.2%	99.5%

Table 2.4: Confidence intervals and actual observed values of proportions of variation captured by the first 6 principal components.

Thus, replacing λ_i by l_i, $i = 1, \ldots, m$, in this expression for τ^2, we obtain an approximate test of H_k^* for our sample data. By computing the choices of h and k that make H_k^* acceptable, we derive confidence intervals.

These computations provide a "practical" measure of the dimension of the interest rate fluctuation space. Figure 2.3 displays graphically the fraction of total variation captured by the first four principal components. Table 2.4 tabulates the same information for the first six principal components and, in addition, includes the proportion actually observed. For example, our data show that the first principal component represented 73% of the observed interest rate fluctuation, and, if our model is correct, we can be 90% sure the first principal component will capture between 66% and 78% of the total variation in future observations. The actual machinery for these computations is laid out in the `stats.pl` listing in the appendix.

The confidence intervals provide a basis for comparison with other measurements of proportional variation. The 90% confidence interval for the first component alone includes neither the 80% proportion found in [1] nor any of the proportions for the various maturities found through factor analysis by Litterman and Scheinkman in [20], Table 2. These earlier studies were conducted using different data; we may be seeing evidence of non-stationarity. Also compare to Soto [30], note 16, for Spanish bonds.

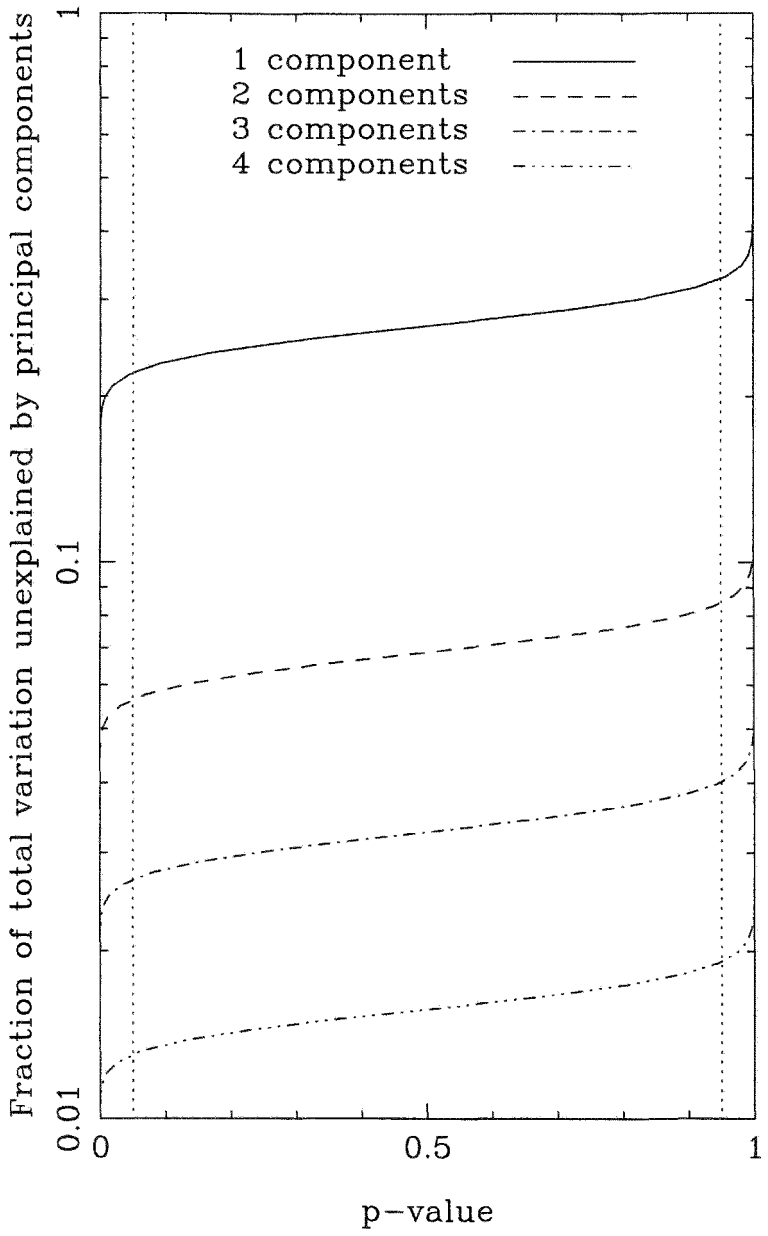


Figure 2.3: p-values for the explanatory power of the 4 largest principal components. 90% confidence intervals are indicated by the vertical dotted lines. The vertical axis is scaled logarithmically for better visualization.

2.4.3 Shape of first principal component

Linearity. Early work in interest rate risk assumes, essentially, that only one principal component has a non-zero eigenvalue and that the principal component equals ι, suitably normalized.[7] Even with the advent of multi-factor models we find the primary engine of interest attributed to parallel translation. This quote from Litterman and Scheinkman [20] seems representative of received opinion:[8]

> [T]he yield changes caused by the first factor are basically constant across maturities. That is, the first factor represents essentially a parallel change in yields.... Thus, hedging against Factor 1 is close to duration hedging.... The impact of Factor 1 on yield levels leads us to name it the *level* factor.

Though it must be noted that Litterman and Scheinkman themselves note the slight difference between their Factor 1 and literal parallel translation.

All this begs the question, is the first principal component statistically different from parallel translation? or, more generally, linear translation? Statistical tests for equality between our sample first principal component and any linear model are unquestionably negative.

In more detail, let H^{**} be the null hypothesis that the vector of coefficients U_1 of the first principal component is equal to a specified vector U_1^0 of unit length,

[7] See the proof of the immunization theorem, Appendix B, in Fisher-Weil [9]. This paper also provides references to the still earlier work in interest rate risk by Macaulay, Reddington and others.

[8] Along the same lines, see Willner [34] and Soto, Note 17 [30].

$$H^{**} : U_1 = U_1^0, \qquad U_1^{0\prime} U_1^0 = 1.$$

T. W. Anderson (See [25], section 9.6.) showed that if H^{**} is true then the two statistics

$$W_1 = n l_1 U^{0\prime} S^{-1} U_1^0 - n \quad \text{and} \quad W_2 = \frac{n}{l_1} U_1^{0\prime} S U_1^0 - n$$

sum to a single statistic, $W = W_1 + W_2$, having a limiting distribution of χ^2_{m-1}. Here, as before, S is the $m \times m$ sample covariance matrix, and l_1 is the eigenvalue of the first principal component of S. Thus, a test of H^{**} of asymptotic size α is to reject H^{**} if $W > c(\alpha; m-1)$, where $c(\alpha; m-1)$ is the upper $100\alpha\%$ point of the χ^2_{m-1} distribution. Note that if $U_1 = U_1^0$ then $W_1 = W_2 = 0$, and that as U_1^0 rotates away from U_1, W_2 becomes negative, W_1 positive. For matrices S having large condition numbers, as ours do, very small projections along U_k for k large can result in large contributions to the statistic W.

This test applied to our sample data indicate that no linear first principal is remotely acceptable. Table 2.5 illustrates the value of the statistic W in a neighborhood of the "best" linear approximation.[9] The smallest value of W obtained was 305.8, but the 1/2% critical point of χ^2_{m-1}, $m = 48$ is 75.7. Note that W is dominated by W_1, the part generated from the lengths of the projections of the test vector along the principal components divided by the eigenvalues.

Nonetheless, these results do not seem to be numerical artifacts since the same conclusion is reached both with coarser meshes and with the smallest eigenvalues eliminated from

[9] W is a quadratic form so there are no hidden extreme points.

Slope	W_1	W_2	W
0.000	372.7892	-6.1334	366.6558
-0.002	348.3617	-5.6868	342.6749
-0.004	329.3381	-5.2919	324.0462
-0.006	316.4551	-4.9565	311.4986
-0.008	310.5076	-4.6896	305.8179
-0.010	312.3468	-4.5006	307.8461
-0.012	322.8766	-4.3997	318.4770
-0.014	343.0478	-4.3976	338.6503
-0.016	373.8486	-4.5056	369.3429

Table 2.5: Values of the statistic W for linear first principal components with slopes as given in the first column. W is minimal for linear vectors with slope near -0.008, but the upper 1/2% point of $\chi^2_{47} = 75.7$ and the hypothesis of a linear first principal component is rejected for all slopes. The second column shows how it is the projection of the linear vector along the principal components with small eigenvalue that overwhelm W.

consideration. Similar results were also obtained for the yield curve data analyzed in [1] and for more recent data.

Persistence. If we can't justify viewing the first principal component as a linear shift, perhaps the shape of this shift, whatever it is, persists over time. Visually, this seems plausible, as illustrated in Figure 2.4. We can check this with the same test applied previously to the linear principal component hypothesis (cf. `splitstats.pl` listed in the appendix). We find

$$W_1 = 659.67, \quad W_2 = -0.41, \quad W = 659.26, \quad c(.05; 47) = 64.00.$$

with W as before and $c(.05; 47)$ denoting the upper 5% point of the χ^2_{47} distribution. This compares to $W = 586$ for the best linear fit to the five-year first principal component, and it appears that the previous five-year first principal component is an even worse candidate for the most recent five-year first principal component.

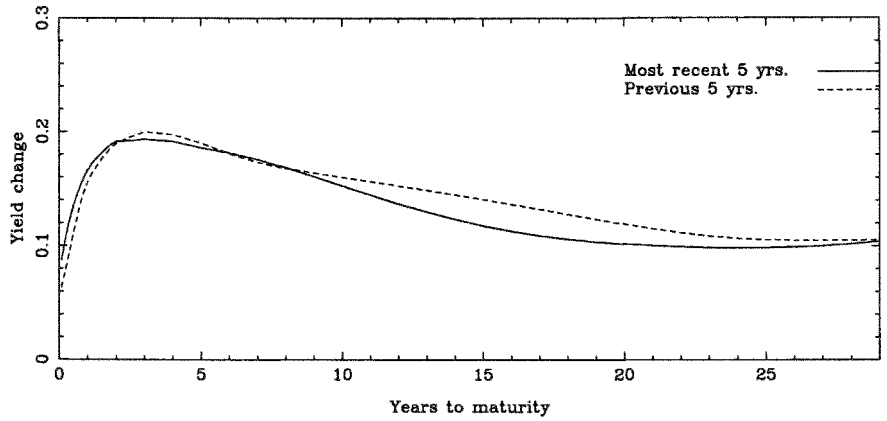


Figure 2.4: Is the first principal component stationary? The principal components corresponding to the largest covariance matrix eigenvalues are shown. Ten years' data from 1991 through 1992 are broken into two successive five year windows, and the principal components for each interval are computed. A statistical test for equality of the principal components corresponding to the largest variances is negative unless data corresponding to the 42 smallest variances is neglected.

On the other hand, as we have seen, the first six principal components alone account for over 99% of total interest rate fluctuation. If we eliminate the smallest $48 - 6 = 42$ principal components as "noise", and recompute our statistics, we obtain

$$W_1 = 3.71, \quad W_2 = -0.41, \quad W = 3.30, \quad c(.05; 47) = 11.07.$$

We can't use this calculation to support a claim that the shape of the first principal persists over time, not without some more work at any rate, but it does show once again how our statistics aren't lining up with our intuition. The last 42 principal components appear insignificant to the eye, but they play a commanding role in the analysis of the sample covariance matrix. It is also worth noting that we also obtain a diminution of W in the linear case when the last 42 principal components are eliminated, but we don't get close to the range where the hypothesis might be accepted.

2.5 Possible shortcomings of principal component analysis

An econometric difficulty with applying principal component analysis to yield curve evolution is that the Euclidean norm, upon which the analysis is so dependent, seems to have no economic meaning. What does it mean for the sum of squares of key rate movements to equal one? How *does* one devise a Richter scale for yield curve movements? Speaking of key rates, how does one choose them rationally? We have followed McCulloch-Kwon without offering any rationale, but certainly different discretizations of an interpolated yield curve will result in principal component differences, however slight, and we have seen how apparent slight differences can be statistically significant.

After atomizing the universe of bonds into hypothetically rationally priced zero-coupon bonds, principal component analysis flips to the opposite extreme: all components are whole-yield curve movements; any movement attributable to a principal component involves movements of all zero-coupon bonds and these movements are perfectly correlated. Moreover, this model is very "stiff"; regardless of any visual impression, the first principal component is statistically remote from any linear vector. Checking to see whether the difference between the observed first principal component and a linear vector is simply noise is characteristic of the analytical process itself: to resolve a complex object to a comprehensible relation among elementary objects. The benefit of principal component analysis is the simple relation it provides; the weakness is that the elementary objects may not have any intrinsic significance. Sometimes insight into the object of study provides the significance; Golub and Tilman, [14], Section 3.2.3, advance the intriguing theory that the first principal component is closely related to the term structure of volatility of changes in U.S. spot rates. Lacking such an understanding, one has no way to distinguish between data mining and the

discovery of persistent mechanisms. In such cases the analyst is willing to sacrifice some efficiency of representation in exchange for better understood elements. This is our goal in the next section where we analyze yield curve data in terms of adapted wavelets.

The assumption that yield curve shocks are normally distributed proved important for drawing statistical inferences. Even in the slightly larger class of elliptical distributions, a diagonal covariace matrix guarantees independent components only when the distribution is normal ([25], Theorem 1.5.3). We also encounter the usual bugbear of gaussian processes; viz. the positive probability of negative rates. Moreover, we have assumed rate changes are governed by a multivariate normal distribution without offering any supporting empirical evidence. This difficulty is accentuated by evidence of non-normality presented in the next section.

Underlying our entire analysis has been an implicit assumption that yield curve shocks are temporally independent identically distributed. Although we do not address it here, this is where we part company with the stochastic process approach to yield curve evolution. For example, if the Heath-Jarrow-Morton process really caught the essence of interest rate movements, then, not only does the probability distribution of the yield curve change depend on the yield curve, but it also depends on its history. Testing such theories with available data seems challenging.

Assume our input data is a series of observed yield curve shifts. A given observation, y_0, is represented by a vector $\{y_{0,k}\}_{k=0}^{n}$ which records the values of y_0 at maturities $\{x_k\}_{k=0}^{n}$.

We are going to rewrite this series of shifts as a pair of series. One is "smoother" than the original observation while the other records the information lost in the process. The first reflects coherent movement, movement experienced by all maturities in a locale; the second, by contrast, reflects idiosyncratic movement around a particular maturity.

3.1 The lazy transform

The first step in this process is to split the "signal" y_0 into two separate signals with the "lazy" transform. That is, the result of applying the lazy transform, L, to y_0 is the pair of signals, the "upper" signal of "forgotten" samples and the "lower" signal of "subsamples." Symbolically,

$$L : y_0 \mapsto (l_{-1}, u_{-1}). \tag{3.1}$$

where

$$l_{-1,k} = y_{0,2k}, \quad k = 0, 1, 2, \ldots \tag{3.2}$$

$$u_{-1,k} = y_{0,2k+1}, \quad k = 0, 1, 2, \ldots. \tag{3.3}$$

The lazy transform is illustrated schematically in Figure 3.1.

Note that y_0 may be recovered from l_{-1} and u_{-1} simply by interleaving the two sequences. A scheme for such a recovery could be obtained from Figure 3.1 simply by reversing the direction of the arrows.

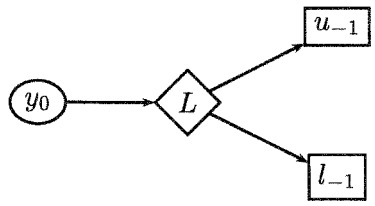


Figure 3.1: Schematic for the "lazy" transform, the first step in the wavelet transform. y_0 is the input sequence, L is the transform, and u_{-1} and l_{-1} are the output sequences.

3.2 The prediction step

The second, or prediction, step affects the upper signal, u_{-1}. The subsamples in l_{-1} are used to predict the data at the points x_{2j+1}. The differences between these predictions and the values of u_{-1} are then recorded as d_{-1}, the difference coefficients. The idea is that if the yields at the various maturities move together, then the data $\{l_{-1}\}$ should form the basis for an accurate prediction and the coefficients $\{d_{-1}\}$ will tend to be small.

There are several ways of formulating a prediction that various authors have investigated (cf. Sweldens and Schröder [32]). The simplest method assumes the least "structure", and merely uses the neighboring even values to predict the odd. We are going to assume there is a modicum of structure, that the yield curve is at least continuous. Therefore we expect to obtain a significantly more efficient encoding of the yield curve fluctuation by using linear interpolation of the even values to predict the odd. Thus

$$d_{-1,k} = u_{-1,k} - \frac{(x_{2k+2} - x_{2k+1})l_{-1,k} + (x_{2k+1} - x_{2k})l_{-1,k+1}}{x_{2k+2} - x_{2k}}. \tag{3.4}$$

and we write

$$P : (l_{-1}, u_{-1}) \mapsto d_{-1}.$$

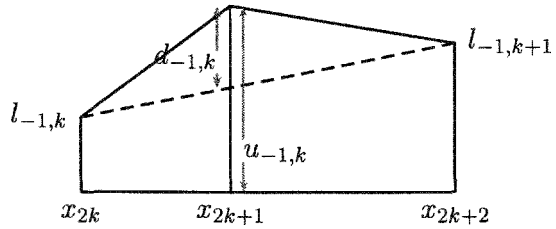


Figure 3.2: A pictorial representation of the prediction step of the wavelet transform. The height of the vertical line above x_{2k+1} is the prediction of $u_{-1,k}$ which is obtained by linearly interpolating $l_{-1,k}$ and $l_{-1,k+1}$. The difference between $u_{-1,k}$ itself and its predicted value is the difference coefficient, $d_{-1,j}$.

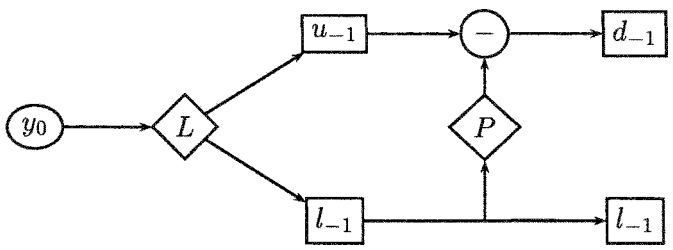


Figure 3.3: Schematic of the prediction step of the wavelet transform. Transform L splits y_0 into two signals as in Figure 3.1. The lower signal, l_{-1}, is output as before, but now it is also used to form a prediction of the upper signal as well. This prediction is subtracted from the true upper signal, u_{-1}, to form a signal of differences, d_{-1}, which is then output alongside l_{-1}.

A picture of this step is provided in Figure 3.2, and a schematic dagram extending Figure 3.1 is given in Figure 3.3. It shows the composition of the lazy transform and the prediction step. Since the signal u_{-1} can be recovered from d_{-1} and l_{-1} by interpolating l_{-1} and adding d_{-1}, it follows that y_0 may be recovered from d_{-1} and l_{-1}. A scheme for this inversion may be derived from Figure 3.3 by toggling "$-$" to "$+$" and reversing the direction of right pointing arrows (the vertical arrows remaining as they are).

3.3 The update step

The last function, the update step, is applied to the lower signal, l_{-1}. As we have explained, if yield fluctuations correlate strongly across maturities which are close together, the difference, or detail, coefficients d_{-1} be small, and consequently the lower signal, though coarser, will remain a good estimator of the entire yield curve shift. This is the strategy for obtaining a simpler approximation which minimizes the loss of local information. On the other hand there may be (and are!) global properties of the observed signal that we want a simplified approximation to retain. The update step is designed to recoup information from the detail coefficients in such a way that, by minimal changes to $l_{-1,k}$ induced by neighboring $d_{-1,k}$, the updated signal retains specified global properties of y_0.

In our application, the global quantity of a yield curve shift we want a simplification to share with its original is the amount by which the value of a portfolio changes. Suppose, for purposes of illustration, that our portfolio is a continuous perpetuity that pays interest on a single unit of currency. If the continously compounded interest rate at a time x is $r(x)$, then the value of the portfolio at time 0 would be

$$\int_0^\infty e^{-r(x)x} dx.$$

Next, suppose that the yield curve undergoes a shift, $r \to r + \Delta r$. Then we are going to regard Δr as our signal, and the quantities we will use the update step to preserve are

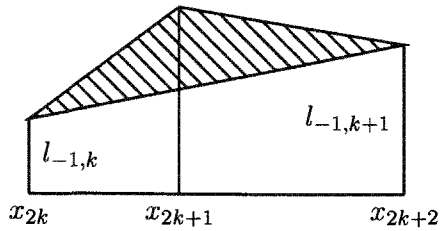


Figure 3.4: Hatched area and its moment represent global information lost in the prediction step.

$$\int_0^\infty \Delta r(x) e^{-r(x)x} dx, \tag{3.5}$$

which we regard as the mean value of the shift (weighted by the discount function) and

$$\int_0^\infty x \Delta r(x) e^{-r(x)x} dx, \tag{3.6}$$

which we regard as an approximation to the change in value the portfolio undergoes.

Returning to our general discrete case (using y_0 instead of Δr), the global quantities we wish to preserve correspond to areas and moments. The first quantity, the average change, corresponds to the integral (3.5). It is the area under the linear interpolation of y_0 (divided by the length of the interval). The second quantity, the first moment of change, corresponds to the integral (3.6) and is the "moment of inertia" of the area under the interpolated y_0.

The hatched area in Figure 3.4 represents how the area would change if the prediction step were not followed by an update. Figure 3.5 illustrates how the $l_{-1,k}$ are updated to a new sequence $y_{-1,k}$. The new sequence y_{-1} has the same average and same first moment as y_0. This is indicated by arranging for the hatched areas in Figures 3.4 and 3.5 to have the same area and the same first moment.

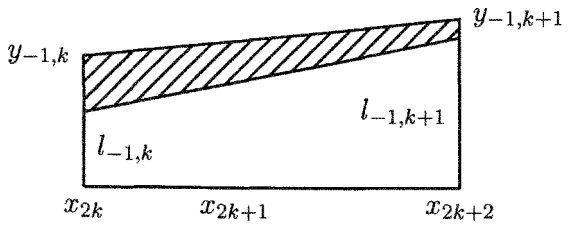


Figure 3.5: Pictorial representation of the update step; information lost in prediction (Figure 3.4 is reincorporated in such a way as to preserve both area and moment of area.

The hatched area in Figure 3.4 is simply $\frac{1}{2}d_{-1,k}(x_{2k+2} - x_{2k})$; the hatched area in Figure 3.5 is $\frac{1}{2}(y_{-1,k} + y_{-1,k+1})(x_{2k+2} - x_{2k})$. Setting these quantities equal to each other and simplifying, we have the linear equation

$$d_{-1,k} = y_{-1,k} + y_{-1,k+1}. \tag{3.7}$$

It's a little messier, but the moment of the hatched area in Figure 3.4 is

$$\frac{d_{-1,k}}{6}(-x_{2k}x_{2k+1} - x_{2k}^2 + x_{2k+2}^2 + x_{2k+2}x_{2k+1})$$

while the moment of the hatched area in Figure 3.5 is

$$(-\frac{4}{3}x_{2k}^2 - \frac{5}{6}x_{2k}x_{2k+2} + \frac{1}{6}x_{2k+2}^2)y_{-1,k} + (\frac{5}{6}x_{2k}^2 + \frac{5}{6}x_{2k}x_{2k+2} + \frac{1}{3}x_{2k+2}^2)y_{-1,k+1}.$$

Setting the moments equal gives us a second linear equation in $y_{-1,k}$ and $y_{-1,k+1}$. Putting this equation together with (3.7) we obtain a linear system in $y_{-1,k}$ and $y_{-1,k+1}$. The determinant of coefficients of $y_{-1,k}$ and $y_{-1,k+1}$ from this system may be calculated to be

$$13x_{2k}^2 + 10x_{2k}x_{2k+2} + x_{2k+2}^2 \tag{3.8}$$

which is positive whenever $0 \leq x_{2k} < x_{2k+2}$ as we do in fact assume. Thus, we can always find $y_{-1,k}$ and $y_{-1,k+1}$ in Figure 3.5 to fit our requirements.

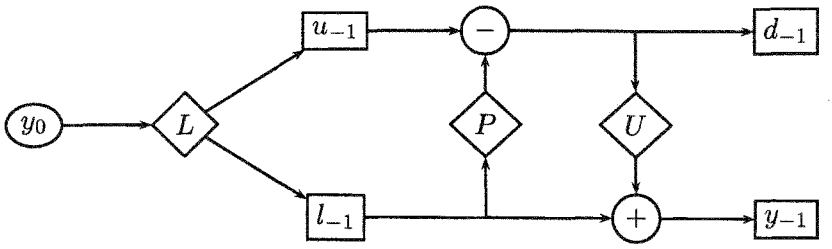


Figure 3.6: Scheme of the entire wavelet transform through lifting

The general case involves one more complication in that there are many points x_{2k+1} so the updates at interior points x_{2k} must be coordinated (see Section 5.1), but this will do for our introductory sketch. We symbolize the update by

$$U : (l_{-1}, d_{-1}) \mapsto y_{-1}.$$

The entire wavelet transform is schematized in Figure 3.6.

The final point to make concerning the update step is that the diagram can be inverted. The update step is invertible because the determinant (3.8) is nonzero, and we have already observed the previous steps were invertible. Again the inversion may be diagramed by reversing the right pointing arrows and toggling the + and − signs. The final result is that the original signal y_0 can be recovered from the coarse approximation y_{-1} and the details d_{-1} that the coarse approximation leaves out.

The process that takes the input sequence y_0 to the pair of output sequences (d_{-1}, y_{-1}) can, in turn, be applied to the sequence y_{-1} itself. That is, the same process, suitably modified, would take y_{-1} as input and produce a new pair of sequences d_{-2}, y_{-2} as output. Thus, with our specific prediction and update routines, we would begin with a sequence y_0 defined on a mesh of points $\{x_k\}_{k=0}^{n}$ where n would be a perfect power of 2, say $n = 2^j$. We would then iterate our process j times and from y_0 produce a series of detail coefficients,

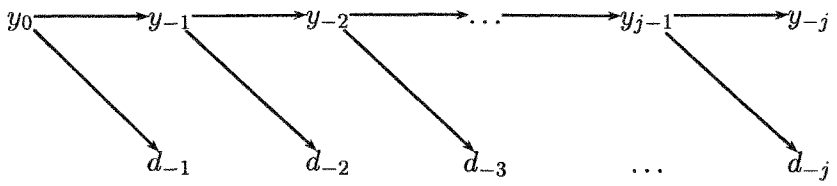


Figure 3.7: Multiresolution analysis of y_0. Repeated application of L, P, and U generates a series of sequences of detail coefficients $d_{-1}, \ldots, d_{-j}$ and the coarsest possible approximation y_{-j}. The process is invertible, so y_0 can be recovered from the d_k and y_{-j}; that is, no information is lost in multiresolution analysis.

$\{d_{-1}, d_{-2}, \ldots, d_{-j}\}$ and one last approximation y_{-j}, defined only at points x_0 and x_n. We view the totality of these iterations as a single transformation, calling it the wavelet transform or multiresolution analysis. We have illustrated the multiresolution process in Figure 3.7.

Since each of the steps

$$y_{-j} \mapsto (d_{-j-1}, y_{-j-1})$$

is invertible, it follows that perfect recovery of the orginal sequence y_0 is possible if one should know only the details of the transforms, the nodes x_k, the detail coefficients d_{-k} and the final approximation y_{-j}. In other words, this is just a fancy change of basis for the vector y_0; nevertheless, it is hoped that this change of basis, like the principal components analysis of the previous section, will help illuminate the process of interest rate evolution.

The data used to illustrate the wavelet analysis comes from the CRSP 2001 Monthly Treasury Database. Data from the CRSP files were extracted and processed using the free software, Larry Wall's Perl (portable extraction and report language) and the Perl Data Language package.

The program, `mbxmunge.pl` in the appendix, extracts 120 months of data and uses the unfiltered Fama-Bliss [2], or "bootstrap" method to construct a piecewise linear yield curve, nodes at the maturity dates of outstanding issues at the time of observation. When several issues matured on the same date, yields to maturity of the various issues were averaged to obtain the yield for that particular date.

The term "bootstrap" is used only to mean that the yield curve determined by shorter term securities was used to "strip" longer term coupon bearing securities before using these latter securities to extend the yield curve. Coupon prices were obtained by linearly interpolating the yields at neighboring maturities whenever no security was scheduled to mature on a coupon payment date. The simple (continuously compounded) yield to maturity was then used as the yield curve value at that maturity. When the yield curve did not extend far enough to strip the last few coupons, a modified internal rate of return was used to estimate the yield. The modification was simply that the yields at the coupon dates would lie on the line determined by the previous latest maturity and the new one.

The output of this extraction program is then a table corresponding to each observation month. Each such table comprises a schedule of unique maturity dates and a yield corresponding to each such date.

The yield curve shifts. In order to estimate the yield curve shift from one month to the next, the smallest common refinement of the two successive months' maturity tables was taken. Linear interpolation was then used to estimate the yield curve at the nodes added by the refinement. The instantaneous spot rate was always taken equal to the shortest observed yield, and difference curves were computed over intervals shorter than the longest maturity of any observed yield curve. Thus the common refinement provided a "common denominator" between successive months, and differences in yields were computed at the nodes of this refinement.

Finally differences for all pairs of successive months were put on the same footing by linear interpolation at a dyadic mesh of nodes. Thus the input for the wavelet transform of each month's yield curve shift was a table of yield differences computed on the minimum possible mesh and then interpolated to a common set of nodes used for all yield curve shifts in the data set under study. Commensurate with the typical number of maturity dates for outstanding Treasurys, the number of nodes in the common mesh was taken to be 257; that is, one plus a power of 2 (since the mesh is dyadic), 8 deemed the most appropriate power.

Moreover, since maturities are highly concentrated towards the short end of the yield curve, we did not use equally spaced intervals for our mesh points. This is a significant departure from the format for "first generation" wavelets (basically those defined by translation and dilation of a single "mother wavelet" or, equivalently, through special properties of the wavelet Fourier transform). Our definition and use of wavelets follows Wim Sweldens and Peter Schröder [32]. The idea is that by adapting our analysis of the yield curve to the information available about it, (much as McCulloch chose spline knots), we will isolate the

salient sections of yield curve movement more effectively. The actual mesh, in days to cash flow, may be found in the appendix in `ycmesh.dat`.

5 The Wavelet Transform of Yield Curve Shifts

Having estimated all yield curve shifts in the interval being studied at a common set of nodes, we proceed to compute a wavelet transform of each of the difference curves. The particular wavelet transfrom used here is an adaptation of the biorthogonal (2,2) Cohen-Daubechies-Feauveau transform. This adaptation uses "second generation wavelets" or the "lifting scheme". This adaptation is used because it extends the wavelet transform to bounded intervals, irregular samples, and weighted inner products.

All of these extensions are important in the study of the yield curve. The yield curve is necessarily bounded at zero maturity and practically bounded at 10 to 30 years on the far end. The irregularity of maturity dates of outstanding securities at any given time brings in irregular samples. Weighted inner products are related to the magnitudes of the yield curve shifts, as we will explain next.

As we noted in our overview of the lifting scheme, the wavelet transform tries to analyze the yield curve shift into shifts that are localized to maturities in a short interval and broader shifts that involve a range of maturities. This is quite familiar to market participants, of course; as this is written (summer 2003) pundits discuss why short-term and long-term rates should be moving in opposite directions. In order to analyze a movement into local and global components, it is necessary to specify exactly what we mean by a global movement. Imagine, for example, a shift where the shortest and longest term rates remain constant while the yield curve, as a whole flattens out. How should such a movement be resolved into local and global movements?

Our criterion for global movements will be stated in terms of the effect the movement has on the value of a porfolio. The global component of a movement will have the same effect while the local component will have no effect. Specifically, denote a portfolio by a sum of delta functions

$$\sum_i b_i \delta_i \tag{5.1}$$

corresponding to a payment schedule $\{x_i\}$. The function δ_i is the point mass concentrated at x_i, and b_i is the current market value of the payment expected at time x_i. Thus the value of the entire portfolio may be expressed

$$\int_0^T \sum_i b_i \delta_i(x) dx,$$

where T is an upper bound for all maturities x_i. Thinking of expression (5.1) as a weight function for the integral inner product and letting Δr denote the yield curve shift, we regard

$$\frac{1}{T} \int (\Delta r)(x) \sum_i b_i \delta_i(x) dx \tag{5.2}$$

as the average shift. Suitably normalized, the first moment of this same integral,

$$\int x(\Delta r)(x) \sum_i b_i \delta_i(x) dx \tag{5.3}$$

is the duration, or coefficient of the best linear approximation to the portfolio change in value brought on by Δr. In terms of these measures, then, a yield curve shift will be resolved into shifts that are local and shifts that are flatter while retaining the same average and duration as the original. The portfolio we work with is the market portfolio of Treasurys, the face value of which is reported by the U.S. Government and recorded in the CRSP files.

It is important to note that the concepts of local and global expand as we work from finer to coarser scales. For example if a yield curve shift were to consist of a single "chirp" at

a single maturity, the successive separations into local and global from the finest to coarsest levels would spread the effect of the chirp to the entire curve, much as a single Hershey's Kiss might slowly melt in a baking tin.

5.1 Update details

The estimation of the global component of a given yield curve fluctuation is obtained first by subsampling and second by updating the subsamples so that, serving as nodes of a linearly interpolated approximation to the originally observed yield curve shift, they determine a function with the same moments (5.2) and (5.3) as the original. We describe how this is done.

We begin with our fundamental functions

$$\phi_{i,j}, \qquad i = 0, -1, \ldots, -n, \quad j = 0, 1, \ldots, 2^{n+i}.$$

The finest level of nodes is chosen, as we have described, comensurate with the level of observation. In particular, as seen in `ycmesh.dat` in the appendix, we start with the interval 1 to 3713 days (just over 10 years) and subdivide it into 256 subintervals, shorter intervals for shorter maturities, longer intervals for longer maturities. For the first level, level $i = 0$, the functions $\phi_{0,j}$ are piecewise linear functions interpolating the delta funtions on the 257 points of the observation mesh. For example, $\phi_{0,0}$ is the piecewise linear function on the interval $[1, 3713]$ which equals 1 at 1, and 0 at all other observation points.

The functions $\phi_{i,j}$ are called fundamental because all observations and transforms of observations may be expressed in terms of them. For example the "observed" yield curve

shift for January of 1992 was 29, 29, 31, 34 and 38 basis points at 1, 4, 7, 10 and 13 days, and so on. Thus the January yield curve shift corresponds to the combination,

$$29\phi_{0,0} + 29\phi_{0,1} + 31\phi_{0,2} + 34\phi_{0,3} + 38\phi_{0,4} + \ldots.$$

In general, when there are $2^n + 1$ original observation points, and the observed values are $\{s_j\}$, then the original (level 0) linearly interpolated approximation is the function

$$\sum_{j=0}^{2^n} s_{0,j}\phi_{0,j}.$$

The level -1 approximation lives on the even indexed points, $0, 2, 4, \ldots$, having the general form

$$\sum_{j=0}^{2^{n-1}} s_{-1,j}\phi_{-1,j}.$$

The lazy transform provides the basis for the coarser estimate simply by subsampling: $29, 31, 38$, and so on at the even indexed points. Interpolation predicts 30 and 34.5 bp at the omitted points, so the first detail coefficients are $d_{-1,0} = -1$ and $d_{-1,1} = -.5$ bp.

The integral of the January 1992 yield curve shift against the weight of market portfolio cash flows is

$$\int_0^T (29\phi_{0,0}(x) + 29\phi_{0,1}(x) + 31\phi_{0,2}(x) + \ldots) \sum_i b_i\delta_i(x)dx, \tag{5.4}$$

as in (5.2). Note that, by linearity, it is enough to compute the integrals of the functions $\phi_{i,j}$ alone. Thus, set

$$M_{i,j}^{(0)} = \int_0^T \phi_{i,j}(x) \sum_k b_k\delta_k(x)dx,$$

and (5.4) becomes a sum of these precomputed integrals,

$$29M_{0,0}^{(0)} + 29M_{0,1}^{(0)} + 31M_{0,2}^{(0)} + \ldots.$$

Similarly, the first moment of January 1992's movement may be expressed as a linear combination of the moments of the fundamental functions,

$$29M_{0,0}^{(1)} + 29M_{0,1}^{(1)} + 31M_{0,2}^{(1)} + \dots.$$

where

$$M_{i,j}^{(1)} = \int_0^T x\phi_{i,j}(x) \sum_k b_k \delta_k(x) dx.$$

In order to retain the integral and moment for the coarser levels, it is required that the new coefficients $s_{-1,j}$ satisfy

$$s_{-1,0}M_{-1,0}^{(0)} + s_{-1,1}M_{-1,1}^{(0)} + \dots = 29M_{0,0}^{(0)} + 29M_{0,1}^{(0)} + 31M_{0,2}^{(0)} + \dots \tag{5.5}$$

$$s_{-1,0}M_{-1,0}^{(1)} + s_{-1,1}M_{-1,1}^{(1)} + \dots = 29M_{0,0}^{(1)} + 29M_{0,1}^{(1)} + 31M_{0,2}^{(1)} + \dots \tag{5.6}$$

and we want the method used to derive the the $s_{-1,j}$ to be the same, regardless of particular shift at hand.

Consider, for example the level -1 fundamental functions, $\phi_{-1,j}$. Each of these is also a perfectly valid yield curve shift in itself; in particular, each may be expressed as a combination of level 0 fundamental functions:

$$\phi_{-1,j} = \sum_k h_{-1,j,k}\phi_{0,k}$$

It is easy to work out the coefficients $h_{-1,j,k}$ in partcular situations, but one knows that they may always be found from the observation that functions which are piecewise linear with nodes on a subset of `ycmesh.dat` are also piecewise linear with nodes on `ycmesh.dat` itself.

Moreover, since the nodes of $\phi_{-1,j}$ are contained in level -1 nodes, the prediction step from level 0 to level -1 is perfect and the detail coefficients are all zero for these functions.

Since the functions themselves are unchanged by the successive application of the lazy transform and the prediction step, their integrals and first moments are unchanged as well. Consequently, whenever the details are 0, the coefficients $s_{-1,j} = s_{0,2j}$. The general form would need to be

$$s_{-1,j} = s_{0,2j} + (\textit{something depending on the detail coefficients})$$

The simplest dependence would be linear; the least distortion of local information would require that only details close to $s_{0,2j}$ be used. Hence, the relation

$$s_{-1,j} = s_{0,2j} + A_{-1,j-1}d_{-1,j-1} + B_{-1,j}d_{-1,j} \tag{5.7}$$

is proposed.

Having gleaned information from hypothetical shifts which were all global and no local (the $\phi_{-1,j}$), we investigate shifts at level 0 that disappear at level -1. That is, we investigate whether an update step of the form (5.7) can provide us with a function which is 0 at level -1, which has a single non-zero detail coefficient, and which satisfies (5.5) and (5.6).

If lazy-predict-update applied to a yield curve shift did result in a function with $s_{-1,j} = 0$ for all j with the single detail coefficient $d_{-1,k} = 1$ and the other detail coefficients were equal to 0, then only two even coefficients of the original shift could possibly be non-zero:

$$\begin{aligned} s_{0,2k} &= s_{-1,k} - B_{-1,k} \\ s_{0,2k+2} &= s_{-1,k+1} - A_{-1,k}. \end{aligned}$$

In fact, since we assume the $s_{-1,j}$ all equal zero, we must have

$$s_{0,2k} = -B_{-1,k}$$

$$s_{0,2k+2} = -A_{-1,k}.$$

Finally, in order to get the detail coefficients right, the shift at level 0 must be of the form

$$-B_{-1,k}\phi_{-1,k} + \phi_{0,2k+1} - A_{-1,k}\phi_{-1,k+1}. \tag{5.8}$$

Since (5.8) maps to zero at level -1, its integral and first moment must also be 0. This leads to the linear equation

$$\begin{bmatrix} M^{(0)}_{-1,k} & M^{(0)}_{-1,k+1} \\ M^{(1)}_{-1,k} & M^{(1)}_{-1,k+1} \end{bmatrix} \begin{bmatrix} B_{-1,k} \\ A_{-1,k} \end{bmatrix} = \begin{bmatrix} M^{(0)}_{0,2k+1} \\ M^{(1)}_{0,2k+1} \end{bmatrix} \tag{5.9}$$

for each k. Thus the integrals and moments of the fundamental functions determine the coefficients $B_{-1,k}$ and $A_{-1,k}$, and these coefficients are used to calculate the update transformation of any yield curve shift.

Thus, so long as the system (5.9) has a solution, the update map may be taken in the simple form (5.7). The function (5.8) is known as a wavelet.

Returning to the January, 1992, shift, our first two equations corresponding to $k = 0$ and $k = 1$ in (5.9) are

$$\begin{bmatrix} 3507 & 17536 \\ 21043 & 105217 \end{bmatrix} \begin{bmatrix} B_{-1,0} \\ A_{-1,0} \end{bmatrix} = \begin{bmatrix} 7014 \\ 42087 \end{bmatrix}$$

$$\begin{bmatrix} 17536 & 64211 \\ 105217 & 896006 \end{bmatrix} \begin{bmatrix} B_{-1,1} \\ A_{-1,1} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

for which we obtain

$$\begin{bmatrix} B_{-1,0} \\ A_{-1,0} \end{bmatrix} = \begin{bmatrix} .07692 \\ .38462 \end{bmatrix}$$

$$\begin{bmatrix} B_{-1,1} \\ A_{-1,1} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

and, referring to our model (5.7),

$$s_{-1,0} = 29 + .07692 \cdot (-1) \quad \text{and} \quad s_{-1,1} = 31 + .38462 \cdot (-1)$$

in basis points.

The scaling function integrals $M_{i,j}^{(0)}$ and first moments $M_{i,j}^{(1)}$ are computed by the routine in `wghtkern.pl`, listed in the appendix, which sends its output to `wsclfcnint.dat`, available from the author on request. The update coefficients are computed by the routine `wupdatecoef.pl`, listed in the appendix, which, in turn sends its output to `wupdatecoef.dat`, also available from the author on request. The appendix also contains images of the first 12 level -1 wavelets. The asymmetry of these wavelets is due to the irregularity of the point-mass measure of total Treasury payments for the weighted inner product. All the wavelets have mean change and duration change equal to zero; this explains the "wave", the oscillation above and below 0, at least where the inner product is nonzero.

5.2 The wavelet transform

The process described above takes the yield curve described at $2^n + 1$ points and maps it to a coarser representation based on $2^{n-1} + 1$ points. It separates the yield curve shift into wider scale movements and localized wavelets of movement. This separation involves no loss of information since the finer scale representation can be reconstructed from the coarser scale representation and the wavelets. This same process is repeated on the coarse representation to generate an even coarser representation and wavelets scaled to intervals twice as large as

the first. In fact, this process is repeated until we end up with a representation based only on the end points of the maturity interval we started with and wavelets of ever widening scale. The appendix contains images of the first 12 wavelets at level -2 as well as all the wavelets from levels $-5, -6, -7, -8$. These wavelets also have zero integral and zero first moment when computed against the weight (5.1). Additional asymmetry is induced by the irregularity of the cash flow schedule. The wavelets all have the same basic shape; only the scale changes from level to level.

First generation wavelets are strictly dilations and translations of a single generating "mother" wavelet. Second generation wavelets are more flexible. The price to pay is the loss of some nice properties such as orthogonality, but just as the second generation wavelets have "similar" shapes, other important properties of wavelets are sufficiently robust to hold for second generation wavelets as well.

The final result is called a "multiresolution" analysis of our finest scale representation. It begins with $\sum s_{0,j}\phi_{0,j}$ and generates successively coarser and smoother approximations to the yield curve shift, finally stopping at a simple line, $s_{-n,0}\phi_{-n,0} + s_{-n,1}\phi{-n}, 1$. Moreover, this process can be reversed because all the information lost in a coarser approximation has been saved in the detail coefficients.

To formulate this resolution, let $\psi_{i,j}$ denote the special function with the following two properties.

1. $\psi_{i,j}$ is a function in level $i+1$ that multiresolution takes to zero in level i; that is, the approximation of $\psi_{i,j}$ on level i is 0.

2. All details of this map are 0 except $d_{i,j}$ and that detail is 1.

These $\psi_{i,j}$ are called wavelets, and each step in the multiresolution process amounts to rewriting the approximation at level $i+1$ as the sum of the approximation at level i and wavelets:

$$\sum s_{i+1,j}\phi_{i+1,j} = \sum s_{i,j}\phi_{i,j} + \sum d_{i,j}\psi_{i,j}.$$

Thus, our fine scale function, which we regard as level 0, was first written

$$\sum_{i=0}^{2^n} s_{0,i}\phi_{0,i}$$

$n = 8$ in our example run and transformed to a function written

$$\sum_{i=0}^{2^{n-1}} s_{-1,i}\phi_{-1,i} + \sum_{i=0}^{2^{n-1}-1} d_{-1,i}\psi_{-1,i}.$$

Repeating until we arrive at level $-n$, we end with

$$s_{-n,0}\phi_{-n,0} + s_{-n,1}\phi_{-n,1} + \sum_{j=1}^{n}\sum_{i=0}^{2^{n-j}-1} d_{-j,i}\psi_{-j,i}. \tag{5.10}$$

The first two terms of (5.10) constitute the linear curve shift having the same mean and duration change as the observed shift. That is, measured in terms of the mean and duration change described in (5.2) and (5.3), these two terms describe the best linear approximation to the yield curve shift. For example, if the yield curve shift were the paradigmatic parallel shift, then all the detail coefficients would be zero, and the shift would be entirely captured by the two coefficients $s_{-n,0}$ and $s_{-n,1}$ of (5.10).

The detail or wavelet coefficients indicate not only the extent to which the yield curve shift fails to be linear, but it also hints at the nature of the failure. In particular if a simple piecewise linear model accurately depicts yield curve movements, we expect most detail coefficients to be essentially zero.

We call the map that begins with the "observed" yield curve shift $\{s_{0,j}\}$ and resolves it to (5.10) the wavelet transform of the yield curve shft. We propose to study the wavelet transform of historical yield curve shifts for insights into the evolution of the yield curve.

Once the yield curve shifts had been approximated and the update coefficients computed, the routine listed in `wwaveycdif.pl` was used to compute the wavelet transform of the 119 monthly yield curve shifts observed in the years 1992 through 2001. The listing is in the appendix. The wavelet transforms are written to the file `wycwavtran.dat` and is available from the author. The wavelet transform of just the January 1992 shift is given in Table 5.1. All entries in Table 5.1 are detail coefficients except the first and last. The first and last coefficients indicate the linear yield curve shift with the same mean shift and same duration change. This linear shift comprises 8bp at the short end and 3bp at the long end. The odd columns in Table 5.1 (assuming 0 off-set) all comprise the first pass (level -1) detail coefficients. Columns 2 and 6 comprise the second pass details, column 4 the third, etc.

The wavelet transform of the January 1992 shift detailed in Table 5.1 resolves the finest approximation of the yield curve shift to the coarsest approximation and a sum of wavelets. Inverting this process, we may start with the coarsest approximation and add the details for each level back in, one level at a time, obtaining at each level a finer approximation to the observed yield curve shift. This can be read off from the formula in (5.10 and is indicated graphically in Figure 5.1.

0	1	2	3	4	5	6	7
0.000825	-0.000058	0.001227	0.000000	-0.002739	-0.009000	-0.004357	0.001046
-0.001444	-0.000467	-0.003219	0.001636	-0.000077	0.000092	-0.001128	-0.001198
0.000847	0.000000	0.001178	-0.000074	0.000139	-0.000108	0.000901	0.000413
-0.000597	0.000302	0.000752	-0.000089	0.000211	-0.000056	0.000132	0.000205
-0.000066	0.000000	-0.000071	-0.000274	-0.000087	0.000028	0.000655	0.000384
-0.001118	0.000166	0.000777	0.000036	0.000718	-0.000203	-0.000026	-0.000000
0.000005	0.000016	0.000019	0.000257	0.000433	-0.000147	0.000459	0.000720
-0.001362	0.000119	0.000327	0.000000	-0.000490	-0.000250	-0.001051	-0.000000
0.000383	0.000678	-0.001538	0.000120	-0.000549	0.000175	-0.000582	-0.000145
-0.001179	-0.000199	-0.000883	-0.000553	-0.001293	-0.000143	-0.000280	-0.000441
-0.000148	-0.000024	0.000157	-0.000953	-0.000675	-0.000003	0.001181	-0.000069
0.000053	-0.000016	-0.000163	-0.000013	-0.000002	0.000187	0.000360	-0.000056
-0.000854	-0.000065	-0.000005	-0.000003	-0.000174	0.000026	-0.000191	0.000009
-0.001068	0.000524	0.000778	-0.001712	-0.004497	0.000967	0.003279	0.000006
-0.000985	-0.000032	0.000056	0.000103	0.000303	0.000085	-0.000339	-0.000052
-0.000098	-0.000038	0.000365	0.000048	-0.000140	-0.000125	0.000177	0.000268
0.000481	-0.000098	0.000332	0.000034	-0.001822	-0.000402	0.000669	0.000378
0.000123	-0.000336	-0.000284	0.000066	-0.000291	-0.000083	0.000211	0.000136
-0.000671	-0.000034	-0.000404	-0.000235	0.000655	0.000712	-0.001537	-0.000352
-0.000772	0.000034	0.000326	0.000025	0.000870	-0.000013	-0.000304	0.000000
0.000064	0.000003	0.000189	-0.000004	0.000041	-0.000002	-0.000048	0.000001
0.000102	-0.000000	0.000346	0.000047	-0.000448	-0.000524	0.000282	-0.000020
0.000119	0.000047	0.000182	-0.000296	-0.000249	0.000102	0.000098	-0.000001
0.000186	0.000055	0.000112	0.000015	0.000071	0.000009	0.000382	-0.000207
0.000670	-0.001048	0.000450	-0.000193	0.001373	0.000041	-0.000381	-0.000046
0.000787	0.000038	0.000086	0.000003	-0.000027	-0.000124	0.000218	0.000085
-0.000203	-0.000199	-0.000600	0.000010	-0.000384	-0.000132	-0.001945	-0.010058
0.000248	0.003665	0.003025	-0.000619	0.001791	0.001014	-0.000625	-0.000081
0.000436	0.000587	0.001787	0.000053	0.000728	0.000000	0.000154	0.000075
0.000502	-0.000028	0.000085	0.000012	-0.000389	-0.000199	-0.000044	0.000012
-0.000651	-0.000026	0.000038	0.000144	-0.000124	-0.000011	-0.000272	0.000020
0.000564	0.000068	-0.000061	-0.000026	0.000054	-0.000038	-0.000182	-0.000004
0.000267							

Table 5.1: Wavelet transform (multiresolution analysis) of the yield curve shift in January, 1992. Read this table from left to right, then top to bottom. That is, with zero offset, coefficients whose indices are divisible by 8 are found in the first column.

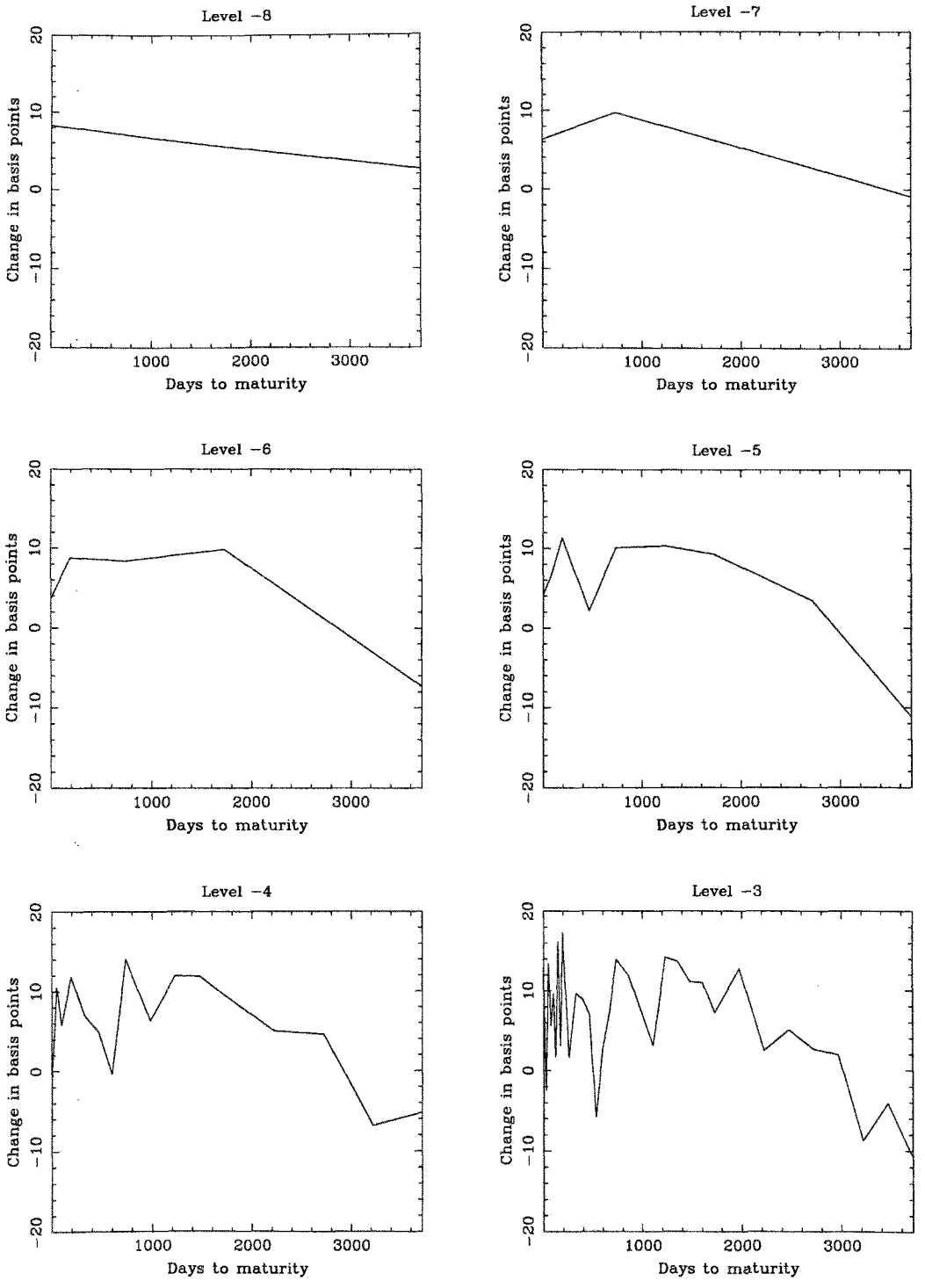


Figure 5.1: Beginning from the coarse approximation of the yield curve shift obtained by the wavelet transform, level −8, one can work back to the raw data by adding the details back in. One simply multiplies the wavelets for January 1992 (see appendix) by the coefficients from Table 5.1 and adds the resulting functions to the previous level's approximation.

6 Wavelet transforms of Treasury yield curve shifts

Having transformed yield curve shifts over 10 years, we look for clues to the way the yield curve moves in the array of wavelet transform coefficients.

6.1 The linear model

If yield curve movements were simply parallel translations except for statistical "noise" we would expect to see the level -8 scaling coefficients (coefficients $s_{-n,0}$ and $s_{-n,1}$ in equation (5.10)) highly correlated with each other and of similar magnitude. We would also expect these coefficients to be of significantly larger magnitude than the detail coefficients. That is, it is only noise that keeps the detail coefficients from being exactly 0 and only noise that keeps the scaling coefficients from being exactly equal to each other.

On the contrary, the data tabulated in Figure 6.1 indicates that, even though there is some correlation between one day and ten year rate movements, there is no support for a hypothesis that the closest linear approximations to yield curve movements is any fixed slope, much less slope 0. On the other hand, Figure 6.2 shows good separation between the coarse scaling coefficients and the detail coefficients. In fact, we see that while 70% of the scaling coefficients exceed 10bp in absolute value, only 20% of the detail coefficients are so large. While it would be hard to argue that the detail coefficients are just noise, it does appear that a two factor linear model obtained from wavelet analysis offers a good rough model.

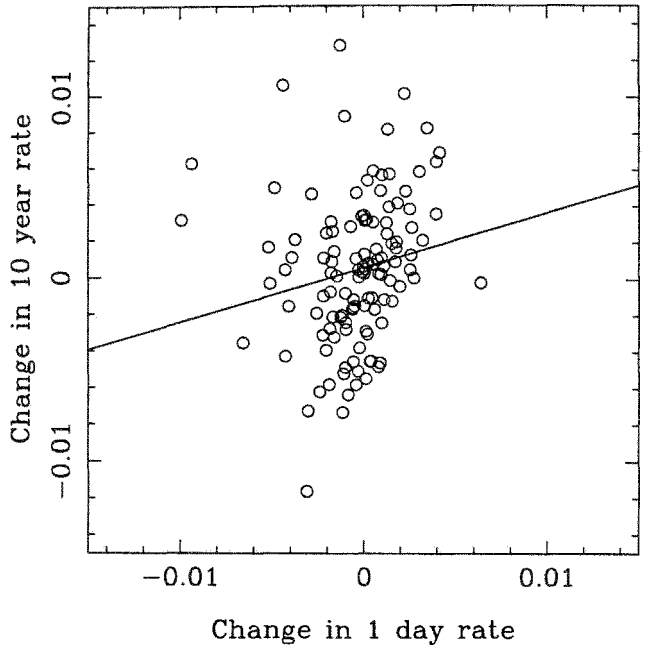


Figure 6.1: Changes in ten year rates are plotted against changes in one day rates for 119 months. The best linear fit is superposed. Correlation between rate changes is positive but weak with $R^2 = 3.3\%$. The slope of the regression line is .30, significant at the 5% level.

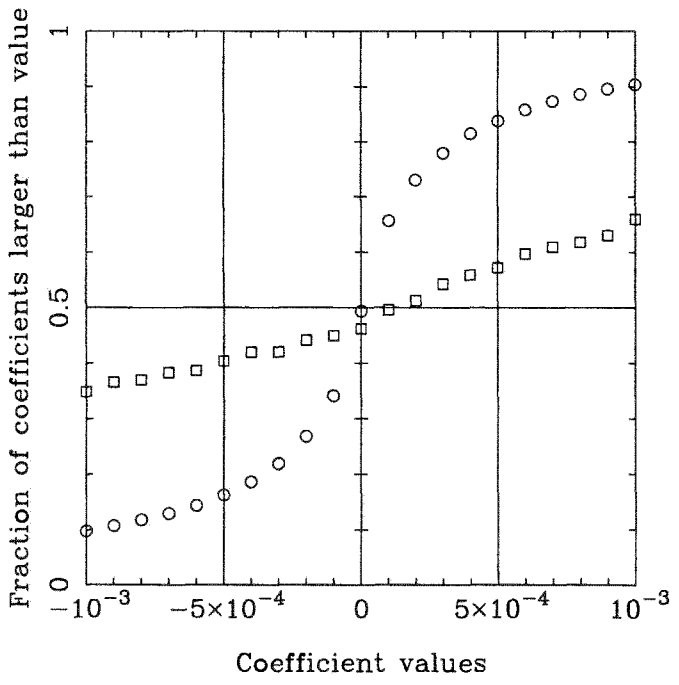


Figure 6.2: Cumulative distribution of coarsest scaling coefficients (squares) and all detail coefficients (circles). About 70% of scaling coefficients are larger than 10bp while only 20% of detail coefficients are.

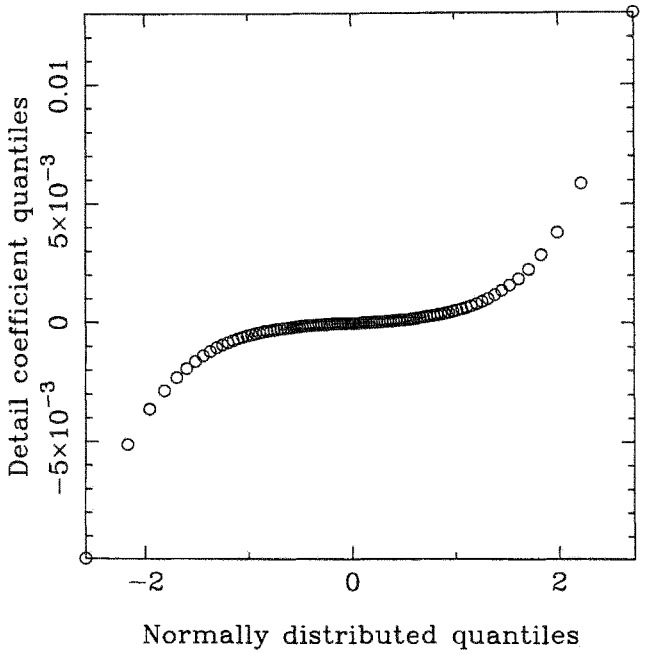


Figure 6.3: The detail coefficients are ranked in one hundred quantiles according to size. The median value from each is plotted against the middle of the corresponding quantile of the standard normal distribution. The extent to which the graph departs from linear is a measure of the extent to which the aggregate of detail coefficients fails to be normal.

6.2 The distribution of detail coefficients

If 2bp were prososed as the noise threshold, roughly 50% of the detail coefficients would import yield curve changes. Perhaps there is useful information hidden in these coefficients; perhaps we can use this information to refine the two factor linear model. First, however, let us consider whether the detail coefficients exhibit the characteristics of simple "white noise". Most naively, we consider whether the $119 \times 255 = 30345$ detail coefficients might themselves be normally distributed. This possibility is quashed by a glance at the "quantile-quantile" plot in Figure 6.3. For if the detail coefficients were normally distributed, a linear transformation would bring their distribution to the standard normal distribution; that is, the elements of the plot would lie along a straight line. On the contrary, we see too narrow

a head matched with too broad shoulders for the distribution to be normal. Or again, the details are normally essentially zero and spike suddenly when rate change information comes through.

Non-normality for the aggregate of detail coefficients does not imply that the detail coefficients are individually non-normal. However, with 119 samples for each detail coefficient, there is enough data to conduct a reasonable test for the normal distribution of detail coefficients at each position at each level. Using the **R** statistical package,[1] we ran the Shapiro-Wilk test [29] on all detail coefficients, level -2 and lower, 127 in all[2]. Of these 127 coefficients, the Shapiro-Wilk test rejects normality for all but 8 at the 5% significance level. Actual statistic and p-values are available from the author. We will proceed under the assumption that the detail coefficients are not normally distributed.

6.3 Power-law distributions

Assuming that the detail coefficients are not normally distributed, it is natural to look for another simple model. Recently, power-laws distributions have proved useful for modeling quantitative changes in the financial markets [11]. Figure 6.4 illustrates the extent to which a power-law distribution might successfully model the distribution of detail coefficients. The graph, though not so concave as would arise from a normal distribution, is nonetheless definitely concave, rendering a power-law model for the distribution of detail coefficients undesirable.

[1] `http://www.r-project.org.`

[2] We regard the information content of level -1 coefficients as somewhat dubious, arising as they do from intervals smaller than actually observed, but rather necessitated by the requirement of plotting successive months on the same grid.

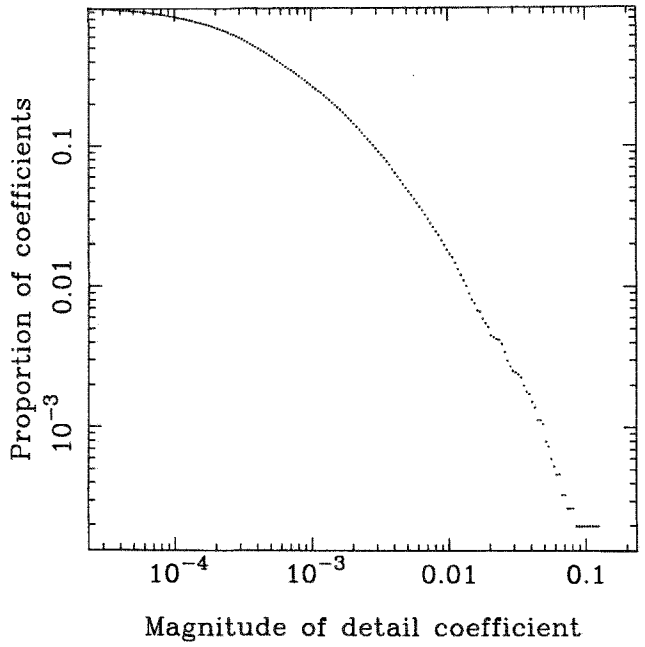


Figure 6.4: This plots the fraction of detail coefficients greater than the argument on a log-log scale. If a power-law distribution provided a good model, the graph would be linear.

6.4 Nonparametric tests

Absent a model for the distribution of detail coefficients, we may still examine the data nonparametrically, and we might also consider methods developed to denoise signals. We first consider two nonparametric tests analogous to analysis of variance. Our question here is whether or not the level of the coefficient is related to its magnitude. To the extent that interest rates move together, we expect to see the lower level (lower frequency) detail coefficients significantly larger than higher level (higher frequency) detail coefficients.

6.4.1 Kruskal-Wallis

We use the Kruskal-Wallis test (cf. [16], Ch. 6.1) to help us determine whether level correlates to coefficient size. Specifically, we set

$$x_{ij} = |d_{ij} - \text{median}_j(d_{ij})|$$

where the d_{ij} denote the detail coefficients obtained in the multiresolution analysis (5.10). The $i = -1, -2, \ldots, -8$ are the levels of the coefficients, and $j = 0, 1, 2, \ldots, 2^{8+i}$ are the indices of the coefficients within the levels starting from the shortest term maturities.

As with parametric one-way ANOVA, we propose to model the magnitudes by

$$x_{ij} = \mu + \tau_i + e_{ij},$$

where the mean μ is chosen so that the "treatment" effects τ_i average to 0, and it is assumed that the e_{ij} are independent identically distributed. The hypothesis tested is

$$H_0 : \tau_{-1} = \tau_{-2} = \ldots = \tau_{-8}$$

against the alternative hypothesis that not all τ_i are equal.

The statistic used to test H_0 is obtained by replacing each detail coefficient by its overall ranking among all detail coefficients. Letting R_i denote the average rank on level i and R stand for the average overall rank, $R = \frac{255 \cdot 119 + 1}{2}$, then the statistic for this test may be written

$$H = \frac{6}{R(2R-1)} \sum_{i=-1}^{-8} 2^{8+i}(R_i - R)^2.$$

Now, if H_0 is true, then H is approximately χ^2 with 6 degrees of freedom. Feeding our data to the Kruskal-Wallis test in the statistical package[3] **R**, we obtain that $H = 3752.72$,

[3] `http://www.r-project.org`

so we would reject H_0 at any significance level. It is consistent with our expectations that level would be related to magnitude somehow; nevertheless, with so much data and so many levels, it is probably a weak result to claim there is some treatment effect. The next statistic tests the same null hypothesis (6.4.1) against a particular ordering of the treatment effects τ_i.

6.4.2 Jonckheere-Terpstra

In our second test, the Jonckheere-Terpstra test (cf. [16], Ch. 6.2), relative rank supplants overall ranking. The alternative hypothesis is

$$H_a : \tau_{-1} \leq \tau_{-2} \leq \ldots \leq \tau_{-7}$$

with at least one strict inequality. Our statistic is

$$J = \sum_{v<u} \sum_{i} \sum_{j} I_{(0,\infty)}(x_{vj} - x_{ui}).$$

where $I_{(0,\infty)}$ is the indicator function on the positive real line. This is simply the number of times a detail coefficient at a lower level exceeds a coefficient at a higher level. Thus the larger J is, the more inclined we are towards the alternative hypothesis H_a. On the other hand, if the null holds, J is approximately normal. Let J^* denote the standardized J,

$$J^* = \frac{J - E[J]}{\sqrt{\mathrm{var}(J)}}.$$

The proceedure `jonktest.pl`, listed in the appendix, computes this statistic to be in excess of 62. Consequently, we find it far more likely that coefficients should tend to be larger at lower levels than at higher levels.

Practically, this means that, should one be in need of a more refined yield curve shift model than the two parameter linear one, then one should first avail oneself of the three

parameter model comprising the coarsest scaling coefficients together with the lowest level detail coefficient. A yet finer model would include the additional two level -7 detail coefficients, and so on.

Although still somewhat crude, this test suggests there is no simple cut-off, there is no level where detail coefficients clearly transit from including rare large shifts to comprising simple white noise. Nevertheless, the benefit we look to wavelets for is a means of measuring not only the level of resolution we need for our purposes but also which coefficients within a level are of greatest importance.

6.5 Mean absolute deviation of detail coefficients

Not only may there be significant differences in detail coefficient magnitude among the various levels, there may also be useful differences among coefficients at different maturities within a particular level. We would like to conduct a two-way analysis of variance, but the fact that the number of detail coefficients double when we move up a level gives us a pyramid rather than a rectangle as a cross-classified design matrix, which precludes any simple application of such an analysis. This would be interesting to come back to.

Lacking a systematic analysis of variance, there is still some value in laying out a table of the scales for each of the detail coefficients. Given the apparent nonnormality of the detail coefficient distribution, the standard deviation of the detail coefficients may not be the best indicator of how the bulk of coefficients are disbursed. Indeed, in a model where the detail coefficients at a given level and maturity are zero except for noise and rare, large signals from the economy, a more resistant measure of scale is appropriate. For in such models standard deviation can be distorted by a single out-sized coefficient. Even in a

model where the detail coefficients are just Gaussian noise, nonparametric scale estimators are still acceptable, if less efficient. See [33], Chapter 5.5, for more discussion and further references.

One commonly used measurement of scale is the "mean absolute deviation" (MAD):

$$\text{MAD} = \text{median}\,\{|d_{i,j} - \text{median}(d_{i,j})|\}$$

where the medians are taken over the number of observations in the sample, 119 in our ten year case.

The mean absolute deviation of each coefficient over the 119 monthly observations in the ten year interval 1992–2001 is plotted in Figure 6.5. Note how on every level the median magnitude is greatest at the short end of the yield curve. Thus, for example, a more effective strategy for constructing a 9 parameter model than that suggested in Section 6.4 would be to include the first detail coefficient from each of levels -2 through -8 than simply to include all coefficients from levels -6, -7 and -8.

6.6 Adaptive thresholding

On the other hand, it should not be just the likely magnitude of a detail coefficient that concerns us; somehow we need to balance both the magnitude and the breadth of a detail coefficient. The wavelets at lower levels correspond to shifts involving broader ranges of maturities and may therefore reflect the impact of macroeconomic events on interest rates while wavelets of the same magnitude at the more transient higher levels may be more reflective of market microstructure idiosyncracies. Indeed, the lower level coefficients

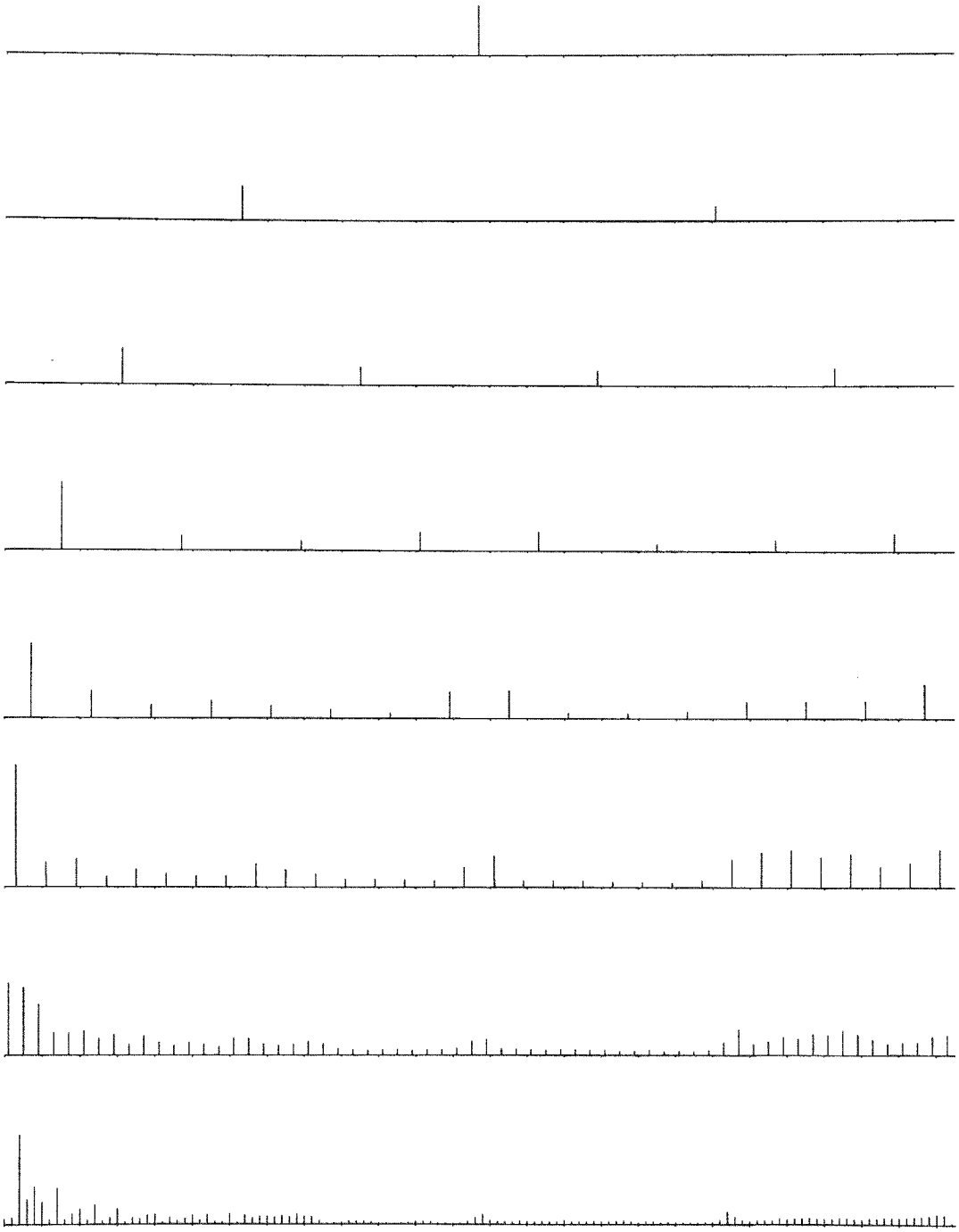


Figure 6.5: The mean absolute deviation (MAD) of each of the 255 detail coefficients is computed and sorted into the corresponding multiresolution levels. Data is taken from monthly observations during 1992–2001. Levels are -8 to -1 top to bottom. A common scale is used throughout; the tallest bar is about 42 basis points.

incorporate yield changes across several issues and can be regarded as exhibiting less market "noise".

To be more systematic in selecting the "significant" detail coefficients, we use adaptive thresholding techniques to separate the coefficients into "signal" and "noise". The basic strategy is to take the non-normal interest rate change process and separate the center from the tails. The center is discarded and the unusually large coefficients at each level are used to reconstruct yield curve shifts which, it is hoped, are better indicators of the effect of events in the ambient economy.

We consider two approaches. The first has solid theoretical underpinnings and is widely used, but violations of technical conditions in our context of biorthogonal wavelet basis and irregular mesh lead us to search for corroborative evidence. The second method turns a test for zero mean normally distributed data on its head; we separate out all data that keeps the null hypothesis from being accepted and discard the rest. Either way, the interpretation of the fat-tailed observed distribution is a superposition of low amplitude white noise on a process of larger, rarer shifts.

6.6.1 Using SURE for threshold selection

The first technique applies Stein's unbiased risk estimate (SURE) to wavelet analysis. First examined by Donoho and Johnstone [6] (cf. [26], Chapter 8, for a discussion and related references), this method chooses a coefficient threshold that minimizes an estimate of loss. Specifically, a one-parameter family of threshold functions is proposed:

$$\delta_\lambda(x) = \begin{cases} x - \lambda, & \text{if} \quad x > \lambda, \\ 0, & \text{if} \quad |x| \leq \lambda, \\ x + \lambda, & \text{if} \quad x < -\lambda. \end{cases}$$

Applied to a set of coefficients, δ_λ reduces small coefficients to zero and reduces the magnitude of all others by λ. Supposing that $\{x_i\}_{i=1}^d$ denote detail coefficients at a given level scaled so that the "noise" component has unit variance,[4] we estimate the "signal" contained in the scaled detail coefficient to be $\delta_\lambda(x_i)$. Now, if $\{y_i\}_{i=1}^d$ denoted the true denoised normalized coefficients, then a measure of our estimation error, the L^2 loss, would be

$$\sum_{i=1}^{d} (\delta_\lambda(x_i) - y_i)^2 ,$$

and our goal is to minimize the expected loss, or risk. Stein's estimator of the L^2 loss is

$$\mathrm{SURE}(\lambda) = d - 2 \cdot \#\{i : |x_i| \leq \lambda\} + \sum_{i=1}^{d} (\min(|x_i|, \lambda))^2 .$$

Here # denotes the number of elements in a set. Now the only undetermined input to this estimator is the threshold coefficient λ and the one we choose at each level is the one which minimizes SURE.

We compute this parameter for each month at each level and average the parameters across all months. The program for this is listed as `SURE.pl` in the appendix, and the results are tabulated in the second column of Table 6.1. Note how λ actually increases as one moves from level -1 to level -4. This may not be surprising in view of the mean absolute deviation (MAD) of the detail coefficients displayed in Figure 6.5; however, it does not correspond well to our intuition that the narrower based detail coefficients of levels -1 and -2 need to overcome a higher threshold to be considered significant; hence our look next at Ogden and Parzen's thresholding procedure next.

[4]Although it is problematical to assume a variance of noise before we determine which component of the observation is noise, it is by regarding the statistical outliers as the "signal" that we may reasonably take the MAD over 0.6745 as the noise variance.

Level	SURE	Ogden-Parzen
−1	0.00000280	0.00080079
−2	0.00000721	0.00074951
−3	0.00001737	0.00069807
−4	0.00003898	0.00064385
−5	0.00003866	0.00057573
−6	0.00001119	0.00048471
−7	0.00000074	0.00042075
−8	0	0.00017274

Table 6.1: The average thresholding parameter λ determined for each level by the SURE and Ogden-Parzen methods

6.6.2 Using recursive hypothesis testing for threshold selection

The approach separating "noise" from "signal" developed by Ogden and Parzen [27] follows along the same lines as the Donoho-Johnstone algorithm just described. That is, the method begins with the consideration of detail coefficients for an observation at a fixed level, then the thresholding function δ_λ is applied to each coefficient. Moreover, the parameter λ is determined by the coefficients themselves; the difference arises in the method by which λ is chosen.

Specifically, the Ogden-Parzen method starts with detail coefficients $\{x_i\}_{i=1}^d$ normalized as before. These coefficients are tested to see if they could reasonably be Gaussian white noise. If the test should fail, however, it is the test rather than the hypothesis that is rejected. The largest coefficient is removed and the remaining coefficients are tested again. Like a salesman who only understands "yes", the analyst repeats this cycle until a test shows it is reasonable to assume the remaining set of coefficients constitute white noise, and λ is set equal to the largest remaining coefficient.

The test statistic is simply

$$\max_{i \in I}(x_i^2) \tag{6.1}$$

where I is the set of indices of coefficients which have not yet been rejected as outliers to white noise. The hypothesis is that (6.1) is the largest observation in $d = \#\{I\}$ samples from a χ^2 distribution with one degree of freedom. The analyst chooses a significance level α which determines the α-critical point to be

$$\lambda_I^2 = \left\{ \Phi^{-1} \left[\frac{(1-\alpha)^{\frac{1}{d}} + 1}{2} \right] \right\}^2 .$$

Here Φ is the cumulative distribution function for the standard normal distribution.

In the application to interest rates, we impose a stiffer standard for narrower based detail coefficients to be regarded as "signal". This means we choose α smaller for higher levels of detail coefficients. In particular, we took $\alpha = .5$ for level -8 and halved it each time we moved to a higher level. With α determined, we computed λ_I^2 for each month for successively smaller sets I until we obtained

$$\max_{i \in I}(|x_i|) < \lambda_I,$$

at which point we set the threshold for that level and month to $\max(|x_i|)$. Averaging the thresholds over the 119 months observations, we determined a global threshold for that level. The programs which executed this process were called `OPtestparam.pl` and `OP.pl` and are listed in the appendix. The global thresholds for each level obtained with the Ogden-Parzen method are given in the third column of Table 6.1.

In constrast to the SURE method, we obtained larger thresholds and these thresholds shrink as we descend to lower levels. This implements our intuition that lower level, broader based detail coefficients should contain less "noise".

Having determined thresholding parameters, we apply the thresholding function δ_λ to each coefficient corresponding to the level of λ. The results of this operation is illustrated

for levels -1 and -2 in Figures 6.6 and 6.7 respectively. These are scatter plots where the horizontal coordinate corresponds to the level and location of a detail coefficient while the vertical coordinate corresponds to the value of the coefficient. The effect of applying the thresholding function is to remove a horizontal strip of breadth λ from the center of the plot and glue the remaining pieces back together. This is quite visible; however, it is harder to obtain a sense of proportion. For example, λ for level -1 is roughly quadruple the MAD for level -1, and hence thresholding at level -1 removes the vast majority of the 15,232 data points. Moreover, referring to Figure 6.5, thresholds of 8 and 7 bp for levels -1 and -2 bring the apparent significance of higher level short-term maturity volatility into better parity with the lower levels.

Graphs 6.6 and 6.7 exhibit a barbell form which is accentuated by thresholding, and this persists through lower levels as suggested by Figure 6.5. Even with the larger Ogden-Parzen thresholds, the story here appears to be that yield curve volatility is dominated by the short-maturity detail coefficients.

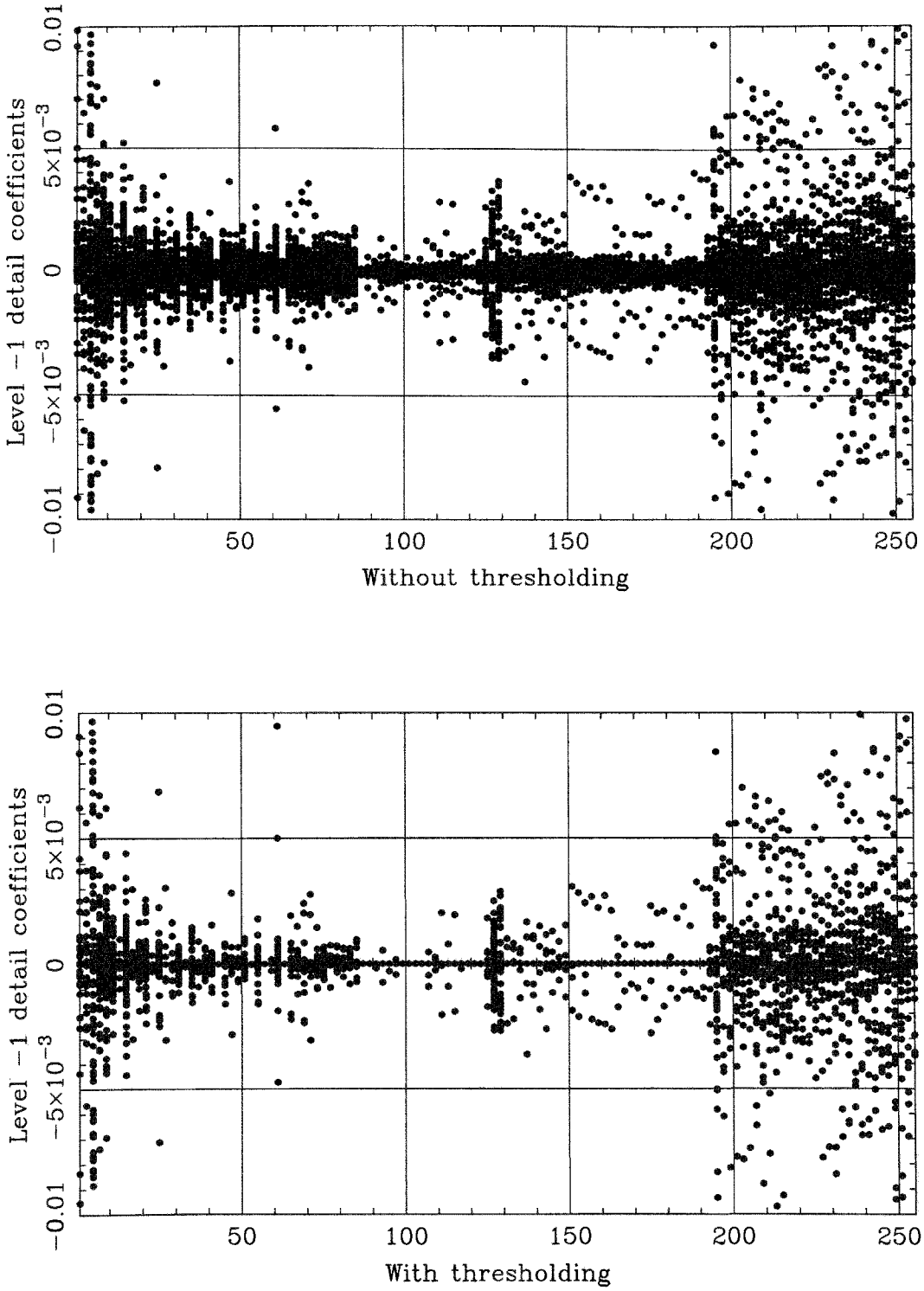


Figure 6.6: Scatter plot of 119 months of 128 level −1 detail coefficients before and after thresholding using the Ogden-Parzen parameter in Table 6.1.

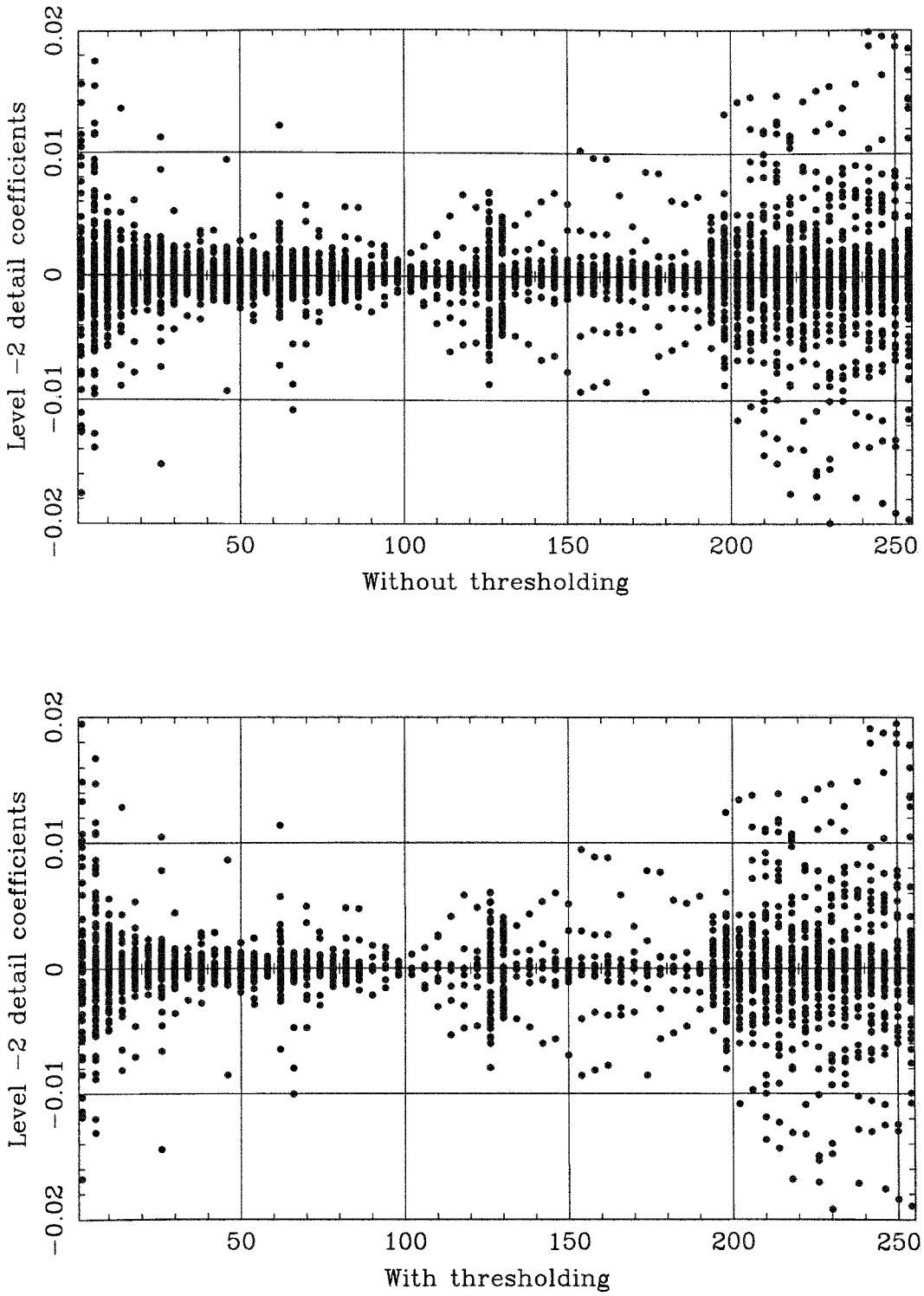


Figure 6.7: Scatter plot of 119 months of 64 level −2 detail coefficients before and after thresholding using the Ogden-Parzen parameter in Table 6.1.

Techniques from principal components analysis (PCA) and from wavelet anaysis have been used to examine movements in the Treasury yield curve. It is hoped that together they provide a stereographic view of these movements. PCA generates an optimal basis for efficient representation of term structure movement, but it is difficult to provide an economic interpretation for the elements of this basis. For example, statistical tests in Chapter 2 keep one from interpreting the first principal component as linear. Multiresolution analysis of the type used here, on the other hand, always provides a linear approximation. Wim Sweldens' second generation wavelet construction adapted to term structure analysis insures that the linear approximation obtained retains two global properties of the input data, the average change and the duration change. Still, the statistical analysis of wavelet data still lacks the precision of the results of Chapter 2.4.

Although the statistics related to PCA are quite sophisticated, I have been concerned that their application in the current context may be compromised by the preprocessing of the data required to apply them. Wavelet analysis works more directly from observation. I have begun the PCA with yield curve approximations like those in McCulloch and Kwon [23]. These approximations involve sampling cubic splines; the splines have no more nodes than the square root of the number of bonds observed. In other words, data are considerably smoothed before PCA is applied, and this may affect the results of sensitive tests. Wavelet analysis, by contrast, begins with a linear interpolation of a yield curve. This yield curve is obtained by stripping coupons from Treasurys with a straightforward bootstrap technique. No other preparation, filtering or smoothing of the data is made. The idea is to make

filtering decisions based on the wavelet analysis; to let multiresolution analysis do the smoothing. Since I also wrote a bootstrapping routine, I have developed an independent complement to PCA over which I have complete control. If it is deemed a certain type of security should not be included in the analysis, I have only to change a few lines of code; if it is desired to perform multiresolution with respect to a different metric than market portfolio cash flows, I can do that too.

Thus multiresolution analysis results in a readily interpreted linear approximation, an approximation that preserves specified economic properties, but nonetheless an approximation that lacks the optimality of the first principal component. Part of what the linear approximation generated by multiresolution analysis lacks in optimality is compensated for by the extra flexibility resulting from the second parameter defining the approximating line. One may see from the apparent near independence of the short and long term scaling coefficients (see Figure 6.1) that one cannot really hope that a single, fixed slope line would ever serve as a dominant principal component.

In fact, we might view multiresolution analysis as a type of regression. Like OLS regression, the output of the algorithm is a line, appropriate or not. The R^2 in OLS regression is essentially the same thing as the variances obtained in our PCA analysis, Table 2.2, and it may also be compared, though not so rigorously, with the scale statistics presented in Figure 6.5, especially when this figure has been modified appropriately with SURE threshold selection or recursive hypothesis testing to eliminate noise terms as described in Chapter 6.6 (See Figures 6.6 and 6.7).

Proceding towards more refined, higher parameter models, Figure 2.3 shows how including more principal components can refine the resulting approximation. As indicated

in Table 2.4 the statistical tests here show one can be quite confident that the first four principal components capture 98% of total yield curve variation. This is consistent with the rapid decay of the eigenvalues in Table 2.2. Still, nagging doubts crop up. Not only do the higher principal components seem even harder to interpret economically (see Figures 2.1 and 2.2) than the first, they also seem to be even less robust. In particular, each principal component involves the entire yield curve.

Turning to multiresolution analysis, one finds again a less efficient representation, but nonetheless a representation that is adaptive and one that localizes the changes to be made if the input data itself suffers only localized changes. For example, if a yield curve shift is observed in which the average change and the duration change are both zero, and if the yield curve shift is limited to maturities at the short end of the yield curve, then multiresolution analysis will resolve the shift completely into a sum of wavelets like those in the appendix but which also have support in the same locality of the yield curve as the observed shift, no fancy cancellation of curves which are nonzero throughout the yield curve required. One practical implication is seen, if one restricts the window of maturities that one analyzes the term structure in. The approximations and wavelets in that window stay the same as for the larger window; this is quite distinct from PCA. Moreover, if there is a loss of efficiency, Figures 6.2 and 6.3 demonstrate that large wavelet coefficients are relatively rare and consequently that multiresolution generated linear approximations tend to be quite good without involving many parameters.

Where PCA statistics gave very precise estimates of the contribution made to total variation by the various principal components, we have only very general, if satisfying, statistical tests of the relative contributions made by the various wavelets. The Kruskal-Wallis

and Jonckheere-Terpstra tests exposed in Chapter 6.4 give evidence that the magnitudes of the wavelet coefficients are differentiated by level and, in particular, that they are larger for lower levels. Since the number of wavelet coefficients decreases by half as one descends from one level to the next, we are unable to apply a standard two-way ANOVA test; however, the scale statistics in Figure 6.5 clearly show there are also differences among coefficients in the same level. In particular, short maturity coefficients are larger at all levels, and consequently, any piecewise linear model for yield curve shifts needs nodes in geometric progression for the first two years[1] maturities. This phenomenon is in rough correspondence to both the convexity of the first principal component and, to a smaller extent, to the magnitude of the second principal component in the very short term.

The following list will wrap matters up. It provides an overview of what I have done in this dissertation.

Principal component results

1. Principal component variances decay rapidly but no two are equal; in particular, none but the last may be considered 0.

2. Most variation of yield curve is explained by 4–6 principal components; approximate confidence intervals are derived for the proportions of variation explained by the largest principal components.

3. The first principal component is not linear, nor does it appear to persist.

[1]For example, start with a short maturity, 3 days, say, and double it repeatedly

Wavelet analysis results

1. Second generation wavelet analysis can be adapted directly to data. Economically meaningful metrics can be substituted for euclidean metrics.

2. In the linear approximations of yield curve shifts derived by multiresolution analysis, the first scaling coefficient is a poor predictor of the second; therefore, at least a two parameter model should be used for yield curve shift approximation.

3. Detail coefficients are not normally distributed; not in aggregate nor in separate positions. Nonparametric methods are required to study wavelet coefficients.

4. Even though the aggregate distribution has tails too fat to be normal, the distribution does not appear to be given by a power law.

5. A Kruskal-Wallis test shows significant difference in scale among levels of wavelet coefficients.

6. A Jonckheere-Terpstra test shows lower level wavelet coefficients are larger than higher level coefficients.

7. Absent a two-way ANOVA test, a scale diagram using mean absolute deviation to define scale shows intralevel trends. Short maturity coefficients are largest. Good piecewise linear yield curve shift approximations should have nodes clustered at the short maturities.

8. SURE threshold selection and recursive hypothesis testing provide methods to filter noise wavelet coefficients. This can refine the scale diagram.

BIBLIOGRAPHY

[1] Joel R. Barber and Mark L. Copper. Immunization using principal component analysis. *Journal of Portfolio Management*, 23(1):99–105, Fall 1996.

[2] Robert R. Bliss. Testing term structure estimation methods. *Advances in Futures and Options Research*, 9:197–231, 1997.

[3] R. Brummelhuis, A. Cordoba, M. Quintanilla, and L. Seco. Principal component value at risk. *Mathematical Finance*, 12(1), January 2002.

[4] John Y. Campbell, Andrew W. Lo, and A. Craig MacKinlay. *The Econometrics of Financial Markets.* Princeton University Press, Princeton, New Jersey, 1997.

[5] David L. Donoho. Interpolating wavelet transforms. Preprint, Department of Statistics, Stanford University. Available on-line at `http://www-stat.stanford.edu/ donoho/reports/1992/interpol.pdf`, 1992.

[6] David L. Donoho and Iain M. Johnstone. Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association*, 90(432):1200–1224, December 1995.

[7] Mardi Dungey, Vance L. Martin, and Adrian R Pagan. A multivariate latent factor decomposition of international bond yield spreads. *Journal of Applied Econometrics*, 15(6), Nov/Dec 2000.

[8] E. Falkenstein and J. Hanweck. Minimizing basis risk from non-parallel shifts in the yield curve. part ii: Principal components. *Journal of Fixed Income*, 7:85–90, 1997.

[9] L. Fisher and R. L. Weil. Coping with the risk of interest rate fluctuations: Returns to bondholders from naive and optimal strategies. *Journal of Business*, Oct 1971.

[10] W. Fleming. *Functions of Several Variables.* Springer-Verlag, 1977.

[11] Xavier Gabaix, Parameswaran Gopikrishnan, Vasiliki Plerou, and H. Eugene Stanley. A theory of power-law distributions in financial market fluctuations. *Nature*, 423:267–270, May 2003.

[12] Ramazan Gençay, Faruk Selçuk, and Brandon Whitcher. *An Introduction to Wavelets and Other Filtering Methods in Finance and Economics.* Academic Press, 2002.

[13] Alois L. J. Geyer and Stefan Pichler. A state-space approach to estimate and test multifactor cox-ingersoll-ross models of the term structure. *The Journal of Financial Research*, 22(1), Spring 1999.

[14] Bennett W. Golub and Leo M. Tillman. *Risk Management: Approaches for Fixed Income Markets.* John Wiley & Sons, New York, 2000.

[15] David Heath, Robert Jarrow, and Andrew Morton. Bond pricing and the term structure of interest rates: a new methodology for contingent claims valuation. *Econometrica*, 60(1):77–105, January 1992.

[16] Myles Hollander and Douglas A. Wolfe. *Nonparametric Statistical Methods.* John Wiley & Sons, New York, 1973.

[17] Robert A. Jarrow. *Modelling Fixed Income Securities and Interest Rate Options.* McGraw-Hill, New York, 1996.

[18] Robert Litterman Knez, Peter J. and Jose Scheinkman. Explorations into factors explaining money market returns. *Journal of Finance*, 49(5):1861–1882, December 1994.

[19] Ilias Lekkos. Factor models and the correlation structure of interest rates: Some evidence for usd, gbp, dem and jpy. *Journal of Banking & Finance*, 25(8), August 2001.

[20] R. Litterman and J. Scheinkman. Common factors affecting bond returns. *The Journal of Fixed Income*, 1, 1991.

[21] Michael Lounsbery, Tony D. DeRose, and Joe Warren. Multiresolution analysis for surfaces of arbitrary topological type. *ACM Transactions on Graphics*, 16(1), January 1997.

[22] J. Huston McCulloch. The tax-adjusted yield curve. *The Journal of Finance*, 30:811–830, June 1975.

[23] J. Huston McCulloch and Heon-Chul Kwon. U.S. term structure data, 1947-1991. Working Paper 93-6, Ohio State University, March 1993.

[24] Yves Meyer. *Wavelets: algorithms and applications.* Society for Industrial and Applied Mathematics, Philadelphia, 1993.

[25] R. J. Muirhead. *Aspects of Multivariate Statistical Theory.* John Wiley, New York, 1982.

[26] R. Todd Ogden. *Essential Wavelets for Statistical Applications and Data Analysis.* Birkhäuser, Boston, 1997.

[27] R. Todd Ogden and Emanuel Parzen. Data dependent wavelet thresholding in non-parametric regression with change-point applications. *Computational Statistics and Data Analysis*, 22:53–70, 1996.

[28] L. C. G. Rogers. Which model for term-structure of interest rates should one use? In Mark H. A. Davis, Darrell Duffie, Wendell H. Fleming, and Steven E. Shreve, editors, *Mathematical Finance*, volume 65 of *The IMA Volumes in Mathematics and its Applications*, pages 93–115. Springer-Verlag, New York, 1995.

[29] S. S. Shapiro and M. B. Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, 52:591–611, 1965.

[30] Gloria M. Soto. Duration models and IRR management: A question of dimensions? *Journal of Banking & Finance*, 2003.

[31] Wim Sweldens. The lifting scheme: A custom-design construction of biorthogonal wavelets. *Applied and Computational Harmonic Analysis*, 3:186–200, 1996.

[32] Wim Sweldens and Peter Schröder. Building your wavelets at home. In *Wavelets in Computer Graphics*, pages 15–87. ACM SIGGRAPH, 1996.

[33] W. N. Venables and B. D. Ripley. *Modern Applied Statistice with S*. Springer-Verlag, fourth edition, 2002.

[34] Ram Willner. A new tool for portfolio managers: Level, slope, and curvature durations. *Journal of Fixed Income*, 6(1):48–59, June 1997.

Program Listings

mcmunge.pl

```
#!/usr/bin/perl -w

use strict;

open DATA, "</home/mark/treas/mc_2fullnew.dat"
   or die "Cannot open mc_2fullnew.dat: $!";
open MUNGE, ">/home/mark/perl/mcmunge.dat"
   or die "Cannot create mcmunge.dat: $!";

my $coef;
my $flag;

while(<DATA>)
{
# Selecting dates for which to extract data:
# Bliss handles Y2K by denoting year 2000 as 100, 1999 having been denoted
# 99.
# Set $flag if we find a date we want.
#
$flag = 1 if (/^\s*9[2-9]|^\s*10[01]/);
if ($flag)
   {
   # The date and the number of spline coefficients are printed at the
   # beginning of each new line.  The remaining entries are nodes and
   # spline coefficients interleaved--except that there is one more spline
   # coefficient than node in McCulloch's scheme.  Bliss sets the last node
   # position to 0.
   #
   chomp;
   print MUNGE $_;
   # Capture the number of spline coefficients into $coef
   if (/\d{6,7}\s+(\d{1,2})/)
      {
      $coef = $1;
      }
   else # that is, if the line comprises node and spline coefficient
      {
      $coef--;
      unless ($coef)
         {
         # After nodes and spline coefficients for a date have been printed
         # to the line, end the line and condition continuation on the next
```

```
        # date.
        print MUNGE "\n";
        $flag = 0;
        }
      }
   }
}
```

csinterpol.pl

```
#!/usr/bin/perl
#
# Takes formated McCulloch term structure data and computes
#  (1) discount function at an independent mesh of points
#  (2) corresponding yields on same mesh
#  (3) resulting matrix of yield shifts
#  (4) singular value decomposition of shift matrix
#
# Thus input is matrix of data and vector of nodes;
# program sends each of the SVD factors to a separate file.

use strict;
use PDL;
use PDL::NiceSlice;

sub csmat
{
# This matrix encodes linear conditions on coefficients of cubic spline;
# it is the matrix M in the equation MC=B where C is the vector of
# nonconstant coefficients of the cubic polynomial in each interval.
# Rows = 0 (mod 3) encode conditions on polynomial functions
# (interpolation requirements); rows = 1 and 2 (mod 3) encode smoothness
# conditions.
#
# Input should be a 1-D pdl of nodes in increasing order.
my $x = shift;
my $h = $x(0:-2)-$x(1:-1);
   # lengths of intervals between nodes times -1; negative because we are
   # orienting the Taylor expansion around the right-hand endpoint of
   # interval.
my $n = $h->getdim(0);
my $m = zeroes(3*$n,3*$n);

# defines main diagonal of $m
my $d = pdl(1, 2, 6);
$d = transpose($h) x $d; # $n-by-3 matrix; same as outer($d,$h)
$d = $d->flat; # rows concatenated into single row of 3*$n entries.
$m->diagonal(0,1) .= $d;
```

```
# defines 1st upper diagonal of $m
$d = pdl(0, 1, 3);
my $dd = $h**2;
$d = transpose($dd) x $d;
$d = $d->flat;
$m->(1:-1,0:-2)->diagonal(0,1) .= $d(1:-1);

# defines 2nd upper diagonal of $m
$d = pdl(0, 0, 1);
$dd = $h**3;
$d = transpose($dd) x $d;
$d = $d->flat;
$m->(2:-1,0:-3)->diagonal(0,1) .= $d(2:-1);

# defines 1st lower diagonal of $m
$d = pdl(0, 1, 2);
$d = transpose(ones($h)) x $d;
$d = $d->flat;
$m->(0:-2,1:-1)->diagonal(0,1) .= $d(1:-1);

# defines 4th lower diagonal of $m
$d = pdl(0, -1, -2);
$d = transpose(ones($h)) x $d;
$d = $d->flat;
$m->(0:-5,4:-1)->diagonal(0,1) .= $d(4:-1);

return $m;
}

sub cseval
{
# Expecting 3 pdl arguments: eval points, knots, polynomial coefficients
# for Taylor expansion around right-hand endpoint of each interval, resp.
#
my ($x, $x0, $coef) = @_;
my $n = $x->getdim(0);
my $y = zeroes($n); # Variable for values of spline evaluated at $x
for (my $i = 0; $i < $n; $i ++)
   {
   # This all assumes argument is inside interplation interval.
   # Get the index of the first node greater than evaluation argument.
   my $cr = which( $x0 >= $x(($i)) )->((0));
   # Pull the coefficients of the interpolating cubic polynomial
   my $c = $coef(,$cr-1;-);
   my $h = $x(($i))-$x0(($cr));
```

```
   $y($i) .= $c((0)) + $h * ($c((1)) + $h * ($c((2)) + $h * $c((3))));
   }
return $y;
}

open DATA, "<$ARGV[0]" or die "Unable to open $ARGV[0]: $!.";
open NET,  "<$ARGV[1]" or die "Unable to open $ARGV[1]: $!.";
open MCPC, ">$ARGV[2]" or die "Unable to open $ARGV[2]: $!.";
open MCSV, ">$ARGV[3]" or die "Unable to open $ARGV[3]: $!.";
open MCRM, ">$ARGV[4]" or die "Unable to open $ARGV[4]: $!.";

my $tsnet = rcols *NET;
my $nnodes = $tsnet->getdim(0);
my $tsmat = transpose $tsnet;
# a device for sizing the matrix of interpolated values

while (<DATA>)
{
# This routine expects data from Bliss's term structure files for
# McCulloch's cubic spline regression, except that all data for a given
# month has been concatenated into a single line; viz. date, number of
# coefficients, then nodes and coefficients interleaved--the last
# coefficient slot is filled with a dummy "0" since there is always one
# more coefficient than node.
#
chomp;
my $mc = pdl split;
my $date = $mc((0));
my $nnode = $mc((1))-1;
my $x = $mc(2:-4:2); # knots for the month's cubic spline
my $a = $mc(3:-1:2); # coefficients for McCulloch's spline basis
my $f = pdl -> zeroes( $nnode, $nnode );
   # This matrix will contain the values f_i(x_j) of McCulloch's ith basis
   # spline f_i at the jth node x_j.

   {  # code block for scoping $h
   # f_i(x_i), as in eqn. A.3, McCulloch except we have 0 offset.
   my $h = $x(1:-1)-$x(0:-2); # lengths of internodal intervals
   $h *= $h;
   $h /= 6;
   $f->diagonal(0,1)->(1:-1) .= $h;
   }

   {  # code block for scoping $h
   # 1st col of $f: f_1(x_j), j > 1; cf. eqn. A.5
   my $h = $x(1:-1)-$x(1);
```

```
   $h /= 2;
   $h += $x(1)/3;
   $f((0),1:-1) .= $x(1)*$h;
   }

for ( my $i = 1; $i < $nnode-1; $i ++)
   {
   # Working through the columns of $f after the first.
   # f_i(d_j), j > 2
   my $h = $x($i+1:-1)-$x($i+1);
   $h /= 2;
   $h += (2*$x($i+1)-$x($i)-$x($i-1))/6;
   $h *= $x($i+1) - $x($i-1);
   $f(($i),$i+1:-1) .= $h;
   }

my $y = 1 + inner($f,$a(0:-2)) + $a(-1)*$x; # vector of values at nodes $x

my $m = csmat($x);

my $b = pdl->zeroes(3*$nnode-3);
$b(0:-3:3) .= $y(0:-2)-$y(1:-1); # continuity condition
$b((1)) .= $a((-1)); # 1st derivative of interpolant at left endpoint
$b((2)) .= $a((0)); # 2nd derivative of interpolant at left endpoint

my $coefs = inv($m) x transpose($b);
$coefs -> reshape(3,$nnode-1);
$coefs = append(transpose($y(1:-1)),$coefs);

my $tsval = cseval($tsnet,$x,$coefs); # prices at mesh points for $1 cf
$tsval->inplace->log;
$tsval /= -$tsnet; # yields on zero coupon bonds payable at mesh points
$tsmat = $tsmat->append(transpose($tsval));
} #end while loop through data file.

$tsmat = ($tsmat(1:-2)-$tsmat(2:-1))->sever;
# yield changes at nonzero maturities
# eliminating "dummy" first col
my $nobs = $tsmat->getdim(0);
my $aux = ones(1,$nobs)/$nobs;
$tsmat -= $tsmat x $aux; # subtracting the avg of each row from each row
my ($q, $s, $r) = svd($tsmat);
$q = $q(0:$nnodes-1;|);
# the svd function for rectangular matrix gives too many columns!
$s = $s(0:$nnodes-1;|);
$r = $r(0:$nnodes-1;|);
```

```
wcols dog($q), *MCPC;  # principal components
wcols $s, *MCSV;  # singular values
wcols dog($r), *MCRM;  # rotation matrix
```

stats.pl

```
#!/usr/bin/perl

use strict;
use PDL;
use PDL::NiceSlice;
use Statistics::Distributions;

for my $period ("", "2")
   {
   for (1..4)
      {
      stats (
      mesh => "tsnet_".$_,
      pc => "pc$period"."_".$_,
      sv => "sv$period"."_".$_,
      rm => "rm$period"."_".$_,
      report => "hyptst$period"."_".$_,
      graph => "varprop$period"."_".$_,
      table => "eqeigentbl$period"."_".$_,
      )
      }
   }

sub stats
{
my %arg = @_;

open MCNET, "<", $arg{mesh} or die "Unable to open $arg{mesh}: $!";
open MCPC, "<", $arg{pc} or die "Unable to open $arg{pc}: $!";
open MCSV, "<", $arg{sv} or die "Unable to open $arg{sv}: $!";
open MCRM, "<", $arg{rm} or die "Unable to open $arg{rm}: $!";
open MCSTAT, ">", $arg{report} or die "Unable to open $arg{report}: $!";
open MCGRAPH, ">", $arg{graph} or die "Unable to open $arg{graph}: $!";

print MCSTAT "This output of stats generated with arguments:\n".
"@_\n";

my $u = cat rcols *MCPC;
my $r = cat rcols *MCRM;
my $s = rcols *MCSV;
my $h = rcols *MCNET;
```

```
my $real = which $s > 5*10**-14;
$s = $s($real)->sever;
$u = $u($real)->sever;
$r = $r($real)->sever;

my $n = $r->getdim(1); # number of observations
my $m = $s->getdim(0); # dimension of each observation
my $q = $m - sequence($m-1);
my $l = $s**2;
my $tvarprop=zeroes($s);
for (my $j=0; $j<$m; $j++)
   {
   $tvarprop($j).=inner($s(0:$j),$s(0:$j))/inner($s,$s)
   }

my $cond = log($l(0)) - log($l(-1));
$cond /= log(10);

wcols "%16.16f", $l, $s, $tvarprop, *MCSTAT,
{HEADER=>"\n# covar mat ev's      sing values     prop var"};
printf MCSTAT "base 10 log covariance matrix condition number:\n %f \n",
   $cond->sclr;

my $lt = zeroes($m);
for (my $k = 0; $k < $m; $k++)
   {
   # Arith mean of smallest m-k eigenvalues; l tilda in Muirhead, p. 408
   $lt($k) .= davg($l($k:-1));
   }
my $llt = $lt->log;

my $lv = zeroes($m-1); # log V_k in Muirhead, p. 408
my $ll = $s->log;
$ll *= 2;
for (my $k = 0; $k < $m-1; $k ++)
   {
   $lv($k) .= dsum($ll($k:-1)-$llt($k));
   }

my $fr = (2+$q*(1+2*$q))/(6*$q);

my $p = $n - sequence($m-1) - $fr;

for (my $k = 0; $k < $m - 1; $k ++)
   {
   my $aux = $l(0:$k-1)/$lt(($k))-1;
```

```
   $aux = $aux**-2;
   $p($k) += dsum($aux);
   }

$p *= -$lv;


my $chi= zeroes($m-1);
for (my $k = 0; $k < $m-1; $k ++)
   {
   my $free = (($q+2)*($q-1)/2)->($k)->sclr;
   $chi($k) .=Statistics::Distributions::chisqrprob($free,$p($k)->sclr);
   }

# upper 5% point of st normal cumulative distn function
my $z = Statistics::Distributions::udistr(.05);
my $fact = 1-$z*sqrt(2/$n/$q);
my $cintvl = $lt(-1:1)/$fact;

print MCSTAT "Testing (a) whether the smallest k covariance matrix".
   "eigenvalues are equal and (b) whether this value is".
   "insignificantly small.\n\n";
wcols "%12.5f", $q, $p, 100*$chi, 10000*$cintvl, $s(0:-2), *MCSTAT,
{HEADER=>"\n sv's:  k         P_k       p-value (%)    95% conf int (bp)".
"largest s.v. in test set"};

open MCEQTBL, ">", $arg{table} or die "Unable to open $arg{table}: $!";
wcols $q, $p, $chi, $s(0:-2), *MCEQTBL;
close MCEQTBL;
open MCEQTBL, "<", $arg{table} or die "Unable to open $arg{table}: $!";
open NEWEQTBL, ">", "$arg{table}.tex"
   or die "Unable to open $arg{table}.tex: $!";
while(<MCEQTBL>)
   {
   chomp;
   my @cells = split;
   printf NEWEQTBL "%d & %7.2f & %e & %e\\\\\n",
   $cells[0], $cells[1], 100*$cells[2], $cells[3];
   }
close MCEQTBL;
unlink $arg{table};

my $lpv = zeroes($m);
my $ltv = dsum($l);
for (my $k = 0; $k < $m; $k ++)
   {
```

```
   $lpv($k) .= dsum($l($k:-1))/$ltv;
   }
print MCSTAT "Max likelihood fraction of variation not explained by 1st k".
"principal components:\n";

my $top = 6;
my $lsq = $l**2;
my $perc = .02*( pdl (1..120));
$perc = .1**$perc->sever;
my $M = zeroes($perc->getdim(0),$top);
my $tau = zeroes($perc->getdim(0),$top);
for (my $j = 0; $j < $perc->getdim(0); $j++ )
   {
   for (my $k = 0; $k < $top; $k++)
      {
      $M($j,$k) .= -$perc(($j))*dsum($l(0:$k));
      $M($j,$k) += (1-$perc(($j)))*dsum($l($k+1:-1));
      }
   $M($j) *=sqrt($n);
   for (my $k = 0; $k < $top; $k ++)
      {
      $tau($j,$k) .= 2*$perc(($j))**2*dsum($lsq(0:$k));
      $tau($j,$k) += 2*(1-$perc(($j)))**2*dsum($lsq($k+1:-1));
      }
   $tau($j)->inplace->sqrt;
   $M($j) /= $tau($j);
   for (my $k = 0; $k < $top; $k ++)
      {
      $tau($j,$k) .= Statistics::Distributions::uprob($M($j,$k;-));
      }
   }
$tau = cat(1-$perc,dog($tau));
wcols "%4.4f", dog($tau), *MCSTAT,
{HEADER=>"\n# %    1 pc    2 pc    3 pc    4 pc    5 pc    6pc"};
wcols dog($tau), *MCGRAPH;

print MCSTAT "\nTest to see if 1st principal component is a line.\n";
# Using Statistic W from Muirhead, p. 419.
my $numpts = 17;
my $inc = .002;
for (my $j=$m-1; $j>0; $j--)
   {
   my $v = $u(0:$j) * $s(0:$j) / $s((0));
   my $vi = $u(0:$j) * $s((0)) / $s(0:$j);
   my $w = zeroes($numpts,4);
   for (my $i = 0; $i < $numpts; $i++)
```

```
      {
      my $h0 = -$i*$inc*$h;
      $h0 += ones($h);
      $h0 /= sqrt(inner($h0,$h0));
      $w($i,0) .= -$i*$inc;
      $w($i,1) .= $n*(inner($h0 x $vi,$h0 x $vi)-1);
      $w($i,2) .= $n*(inner($h0 x $v,$h0 x $v)-1);
      $w($i,3) .= $w($i,1)+$w($i,2);
      }
   my $chi2= zeroes($numpts);
   for (my $i = 0; $i < $numpts; $i ++)
      {
      $chi2($i) .=Statistics::Distributions::chisqrprob($j,$w($i,3)->sclr);
      }
   print MCSTAT "$j\n";
   wcols "%8.4f", dog($w), $chi2, *MCSTAT,
   {HEADER=>"\n  slope   W_S^1   W_S   W    p-value"};
   }
}
```

splitstats.pl

```
#!/usr/bin/perl

use strict;
use PDL;
use PDL::NiceSlice;
use Statistics::Distributions;

open MCPC1, "<pc1_1" or die "Unable to open pc1_1: $!";
open MCPC2, "<pc2_1" or die "Unable to open pc2_1: $!";
open MCSV2, "<sv2_1" or die "Unable to open sv2_1: $!";
open MCRM2, "<rm2_1" or die "Unable to open rm2_1: $!";
open MCNET, "<tsnet_1" or die "Unable to open tsnet_1: $!";
open MCSTAT, ">splithyptest" or die "Unable to open splithyptest";

my $u1 = cat rcols *MCPC1;
my $h0 = $u1(0;-);
my $u = cat rcols *MCPC2;
my $r = cat rcols *MCRM2;
my $n = $r->getdim(1);
my $s = rcols *MCSV2;
my $m = $s->getdim(0);
my $w = zeroes(3);

# Test to see if 1st principal components computed from successive
# intervals are equal.
for (my $j=$m-1; $j>0; $j--)
```

```
   {
   print MCSTAT "\n",$j+1,"\n";
   my $v = $u(0:$j) * $s(0:$j) / $s((0));
   my $vi = $u(0:$j) * $s((0)) / $s(0:$j);
   $w(0) .= $n*(inner($h0 x $vi,$h0 x $vi)-1);
   $w(1) .= $n*(inner($h0 x $v,$h0 x $v)-1);
   $w(2) .= $w(0)+$w(1);
   wcols $w, *MCSTAT;

   my $chisq = Statistics::Distributions::chisqrdistr($j,.05);
   printf MCSTAT "Upper 5%% chi-square boundary: %6.3f\n", $chisq;
   }
```

mbxmunge.pl

```
#!/usr/bin/perl

use strict;
use PDL; # Pearl Data Language
use PDL::NiceSlice; # cleaner indexing routines
use Date::Calc qw(:all);
# documentation: http://theoryx5.uwinnipeg.ca/CPAN/data/Date-Calc/Calc.html

sub decode_crsp_date
{
if ($_[0] =~ /(\d{4})(\d{2})(\d{2})/)
   {
   ($1, $2, $3);
   }#endif
}

sub six_mo_bak
{
# shift insures 6 mo from mo end is at mo end
($_[0], $_[1], $_[2]) = Add_Delta_Days($_[0], $_[1], $_[2], 1);
($_[0], $_[1], $_[2]) = Add_Delta_YM($_[0], $_[1], $_[2], 0, -6);
Add_Delta_Days($_[0], $_[1], $_[2], -1);
}

sub rate
{
my ($r_0, $a, $t_0, $t) = @_;
$r_0 + $a * ( $t - $t_0 );
}

sub discount
{
my ($r, $t) = @_;
```

```
exp -$t*$r;
}

sub newton_step
{
my($t_0, $t, $r_0, $c, $p, $a) = @_;
my $n = $p - inner($c, discount($t, rate($r_0, $a, $t_0, $t)));
my $d = inner($c, discount($t, rate($r_0, $a, $t_0, $t))*$t*($t-$t_0));
$a - $n/$d;
}

sub newton_loop
{
my ($t_0,$t,$r_0,$c,$p,$a,$eps) = @_;
my $count = 1;
while (1)
   {
   my $anew = newton_step($t_0,$t,$r_0,$c,$p,$a);
   if ( abs ( $anew - $a ) < $eps )
      {
      return $anew;
      }
   else
      {
      $a = $anew;
      }
   die "Newton's method did not converge to IRR" if ( $count > 5 );
   $count ++;
   }
}

my $eps = 10**(-9);

open DATA, "/home/mark/treas/mbx.dat" or die "Cannot open mbx.dat: $!";
open YIELDCURVE, ">pcwlinyc" or die "Cannot create yieldcurve file: $!";

my ($obsdates, $matnir, $bid, $ask, $accint);
($obsdates, $matnir, $bid, $ask, $accint) = rcols *DATA, 0,1,2,3,4,
   {INCLUDE => '/^ (200|199[23456789])/'};
$matnir = rint (10**6*$matnir);
my $intrate = $matnir%10**6;
my $maturity = ($matnir-$intrate)/10**6;
$intrate %= 10**5;
$intrate = $intrate - $intrate%10;
$intrate /=10**5;
my ($temp1,$temp2,$temp3);
```

```
rle($obsdates,$temp1=null,$temp2=null); # Slices "run-length encode"
$temp3 = cumusumover($temp1($temp1;?)); # NiceSlice "where" modifier
$obsdates = $temp2(which($temp1));
# Now $obsdates contains one copy of each observation date
my $price = .5 * ($bid + $ask) + $accint;
$maturity = cat $maturity, $intrate, $price;

my @monthly;
push @monthly, $maturity(0:$temp3(0)-1);
for (my $i=1; $i < nelem($obsdates); $i++)
   {
   push @monthly, $maturity($temp3($i-1):$temp3($i)-1);
   }

# @monthly is an array of pdl's. Each pdl contains the maturity, coupon
# interest rate, and estimated price of all outstanding securities, one
# column for each security.
for (my $i = 0; $i < nelem($obsdates); $i++)
   {# step through each month in data set
   my @dim = dims $monthly[$i];
   my $sched = zeroes($dim[0]+1,62);
   # $sched is the matrix of cash flow maturities for bonds outstanding
   # at $obsdates.  Each column corresponds to a treasury, earlier
   # maturities to the left, later to the right.  A column consists of
   # number of days to a cash flow, interest or principle.  Longest term
   # cf's are listed first, closer in cf's are listed below.  Columns are
   # padded below with 0's.
   #
   # We pad on left with column of 0's for ease of yield curve definition.
   # It appears T-bond issued 2/2000 matured 11/2000 w/ 61 coupon pmts, so
   # 62 rows are used.

   my ($y1, $m1, $d1) = decode_crsp_date($obsdates($i));
   my $no_cfs = zeroes($dim[0]);

   for (my $j = 0; $j < $dim[0]; $j++)
      { # cash flow dates in each column of sched, in reverse order
      my ($y2, $m2, $d2) = decode_crsp_date($monthly[$i]->($j,0));
      my $delta = Delta_Days($y1,$m1,$d1,$y2,$m2,$d2);
      do
         {
         # filling columns of $sched with # days to coupon pmts
         $sched($j+1,$no_cfs($j)) .= $delta;
         $no_cfs($j)++;
         ($y2, $m2, $d2) = &six_mo_bak($y2, $m2, $d2);
         $delta = Delta_Days($y1,$m1,$d1,$y2,$m2,$d2);
```

```
      } while $delta > 0;
   $no_cfs($j)--;
   } # end for (j)
my $yc = $sched(:,(0))->copy;
# $yc will be month's yield curve
rle($yc,$temp1=null,$temp2=null);
$yc = $temp2(which($temp1));
# remove multiple occurences of same date
$yc = cat $yc, zeroes( $yc->getdim(0)), zeroes( $yc->getdim(0));
# $yc initialized with unique maturity dates and zeros
my $y = $monthly[$i]->(:,(1));
# aux yield array; init as coupon rates
my $z = $monthly[$i]->(:,(2));
# price of final pmt; init as security price
my $k = 0;
# $j - $k will be number of unique maturity dates.

$y(0) .= (log(100) + log(1+$y(0)/2) - log( $z(0) ))/$yc(1,0);
$yc(0,1) .= $y(0);
$yc(1,1) .= $y(0);
my $el = 0;
for (my $j=1; $j<$dim[0] ; $j++)
   { # $j steps through each treasury outstanding at obs date.
   # look back to see which treasurys mature earlier:
   my $ndx = which($monthly[$i]->(:,0) < $monthly[$i]->($j,0));
   if ($j > $ndx((-1)) + 1 || $ndx -> isempty)
      {
      $k++;
      $el++; # $el is the number reps of this maturity; init below
      }
   else
      {
      $el = 0;
      }
   # $ndx is empty when multiple securities mature at shortest term

   my $m = 0;
   # using $m to index coupon pmt dates; init with latest
   if ( $yc($j-$k,0) < $sched($j+1,$m+1))
      {
      $m++;
      # indicates whether coup pmt date lies after previous maturity
      }
   if ($yc($j-$k,0) < $sched($j+1,$m+1))
      {
      $m++;
```

```
         }
      die "More than an 18 month gap in maturities."
         if ($yc($j-$k,0) < $sched($j+1,$m+1));

      if ( $sched($j+1,$m+1) )
         { # stripping out the present value of coupon payments...
         # first interolate the discount rate...
         my $r = interpol($sched(($j+1),$m+1:$no_cfs($j)),$yc(0:$j-$k,(0)),
            $yc(0:$j-$k,(1)));
         # next compute the value of coupon pmts...
         $z($j) -= 50 * $y($j) * sumover(discount($r, $sched(($j+1),
            $m+1:$no_cfs($j))));
         die "Can't have negative security prices" if $z($j)<=0;
         }
      if ( $m )
         {# use approx method to estimate $yc
         my $t0 = $yc(($j-$k),(0));
         my $t = $sched(($j+1),0:$m);
         my $r0 = $yc(($j-$k),(1));
         my $c = ones($m+1)*50*$y(($j));
         $c(-1) += 100;
         my $a = 0;
         $a = newton_loop($t0,$t,$r0,$c,$z($j),$a,$eps);
         $yc($j-$k+1, 1) .= $r0 + $a * ($t((-1)) - $t0);
         }
      else
         {
         $y($j) .= log(100) + log(1+$y($j)/2) - log( $z($j) );
         $y($j) /= $sched($j+1,0);
         $yc($j-$k+1, 1) .= $el/($el+1)*$yc($j-$k+1,1)+1/($el+1)*$y($j);
         } # end if/elsif/else
      } # end for (j)
   $yc(:,(1)) *= 365;
   wcols $yc(,2;-)+$obsdates(($i)), $yc(,0;-), $yc(,1;-), *YIELDCURVE;
   } # end for (i)
```

wghtkern.pl

```
#!/usr/bin/perl

use strict;
use PDL; # Pearl Data Language
use PDL::NiceSlice; # cleaner indexing routines

open CFSCH, "<cfsched.dat" or die "Cannot open cfsched.dat: $!";
my ($date, $dtcf, $cf) = rcols *CFSCH, 0,1,2;
# contains cash flow schedule for each observation date, observation
# date in the 1st col, days to cash flow in 2nd, amount of cf in 3rd
```

```
open YC, "<pcwlinyc" or die "Cannot open pcwlinyc: $!";
my ($ndate, $node, $rate) = rcols *YC, 0,1,2;
# contains piecewise linear bootstrap yield curve for each observation
# date, observation date in the 1st col, yield curve node in 2nd, interest
# rate in 3rd.

open MSH, "<ycmesh" or die "Cannot open ycmesh: $!";
my $mesh = rcols *MSH;
my $numnodes = $mesh->dim(0);
my $depth = sclr rint(log($numnodes-1)/log(2)); # assuming mesh is dyadic
# Our scaling functions start out as interpolated delta functions on a
# mesh; succeeding scaling functions are defined on submeshes

open SCL, ">wsclfcnint.dat" or die "Cannot create wsclfcnint.dat: $!";


my ($temp1,$temp2) = rle($date); # Slices "run-length encode"
$date = $temp2($temp1;?); # pdl of unique observ dates
# pdl of number of cf's per observation:
my $ptspd = cumusumover($temp1($temp1;?));
$ptspd = append(0,$ptspd);

rle($ndate,$temp1=null,$temp2=null);
$ndate = $temp2($temp1;?); # pdl of unique observ dates
my $ndspd = cumusumover($temp1($temp1;?)); # pdl of yc nodes per observ
$ndspd = append(0,$ndspd);

for (my $i=0; $i<$date->dim(0); $i++)
   {
   # cash flow sched
   my $x = $dtcf($ptspd(($i)):$ptspd(($i+1))-1;|);
   # yield curve nodes
   my $x0 = $node($ndspd(($i)):$ndspd(($i+1))-1;|);
   # yield curve values at nodes
   my $y0 = $rate($ndspd(($i)):$ndspd(($i+1))-1;|);
   # yield curve values interpolated at cf dates
   my ($y,$idx) = interpolate($x,$x0,$y0);
   # yc data was stored as annualized--need daily rate
   $y *= -$x/365;
   $y->inplace->exp; # discount function
   $y *= $cf($ptspd(($i)):$ptspd(($i+1))-1); # PV of cf's
      # Next we restrict the discounted cf function to the interval where
      # the mesh is defined.
   my $ul = $mesh((-1));
   $y = $y($x<=$ul;?)->sever;
```

```
   $x = $x($x<=$ul;?)->sever;

      # Allocating a pdl for scaling function "integrals":
   my $m = zeroes(2*($depth+1),$numnodes);
   for (my $j=0; $j<=$depth; $j++)
      {
      # Defining both scaling function and integral
      my $phi = append ($mesh(0:2**$j:2**$j),10**5);
      $m($j,0) .= inner($y,interpol($x,$phi,pdl(1,0,0)));
      $m($j+$depth+1,0) .= inner($x*$y,interpol($x,$phi,pdl(1,0,0)));
      for (my $k=1; $k<($numnodes-1)/2**$j; $k++)
         {
         my $phi = $mesh(2**$j*($k-1):2**$j*($k+1):2**$j);
         $phi = append(0, $phi)->sever;
         $phi = append($phi, 10**5)->sever;
         $m($j,$k) .= inner($y,interpol($x,$phi,pdl(0,0,1,0,0)));
         $m($j+$depth+1,$k).=inner($x*$y,interpol($x,$phi,pdl(0,0,1,0,0)));
         } # end for (k)
      $phi = append (0,$mesh(-1-2**$j:-1:2**$j));
      $m($j,($numnodes-1)/2**$j) .= inner($y,interpol($x,$phi,pdl(0,0,1)));
      $m($j+$depth+1,($numnodes-1)/2**$j) .= inner($x*$y,
         interpol($x,$phi,pdl(0,0,1)));
      } # end for (j)

   wcols dog($m), *SCL;

   } # end for (i)
```

wupdatecoef.pl

```
#!/usr/bin/perl -w

use Carp;
use strict;
use PDL;
use PDL::NiceSlice;

my $eps = 10**-8;
my $depth = 8;
my $idx = pdl (0,$depth+1);
my $months = 120;

open WSI, "<wsclfcnint.dat" or die "cannot open wsclfcnint.dat: $!";
my $mom = cat rcols *WSI;

open UPDC, ">wupdatecoef.dat" or die "cannot create wupdatecoef.dat: $!";

for (my $i=0; $i<$months; $i++)
```

```
   {
   my $uc = zeroes(2**($depth-1),2*$depth); # update coefficient vector;
   for (my $k=1; $k <= $depth; $k++)
      {
      for (my $j=1; $j <= 2**($depth-$k); $j++)
         {
         my $subm = transpose $mom($idx+$k+$i*($depth+1)*2,$j-1:$j);
         my ($r,$s,$q) = svd($subm);
         my $b = transpose $mom($idx+$k-1+$i*($depth+1)*2,2*$j-1);
         if (all $s > $eps)
            {
            $uc($j-1,2*$k-2:2*$k-1) .= $q * $s**-1 x transpose($r) x $b;
            }
         elsif (all $s <= $eps)
            {
            $uc($j-1,2*$k-2:2*$k-1) .= 0;
            }
         else
            {
            $r = $r(0;|);
            $s = $s(0;|);
            $q = $q(0;|);
            $uc($j-1,2*$k-2:2*$k-1) .= $q * $s**-1 x transpose($r) x $b;
            } # end if/elsif/else
         } # end for($j)
      } # end for($k)
   wcols dog($uc), *UPDC;
   } # end for($i)
```

joncktest.pl

```
#!/usr/bin/perl -w

use strict;
use PDL;
use PDL::NiceSlice;

my $w = cat rcols 'wycwavtran.dat';
$w->reshape(257,119);
my @level;
for (my $i=1; $i<9; $i++)
   {
   my $tmp = $w(1:-2:2;|);
   $tmp->reshape(2**(8-$i)*119);
   my $median = (stats($tmp))[2];
   $tmp -= $median;
   $tmp->inplace->abs;
   push @level, $tmp;
```

```
   $w = $w(0:-1:2;|);
   }

my $u = 0;
for (my $i = 1; $i < 8; $i++)
   {
   for (my $j = 0; $j < nelem($level[$i]); $j++)
      {
      for (my $k = 0; $k < $i; $k ++)
         {
         $u += which($level[$k]<$level[$i]->(($j)))->nelem;
         }
      }
   }
print $u, "\n";

my $N = 255*119;
my $n = 119*2**sequence(8)->(-1:0);
my $num = $N**2 - sum($n**2);
$num /= 4;
my $den = $N**2*(2*$N+3) - sum( $n**2*(2*$n+3) );
$den /= 72;

my $J = ($u - $num)/sqrt($den);
print $J, "\n";
```

SURE.pl

```
#!/usr/bin/perl -w

use strict;
use PDL;
use PDL::NiceSlice;

my $w = cat rcols 'wycwavtran.dat';
$w->reshape(257,119);
my $z = $w -> copy;
my @level;
for (my $i=1; $i<9; $i++)
   {
   my $tmp = $z(1:-2:2);
   $tmp->inplace->abs;
   $tmp = qsort $tmp;
   push @level, $tmp;
   $z = $z(0:-1:2;|);
   }

my $sig = $level[0]->flat;
```

```
for (my $i = 1; $i < 8; $i ++)
   {
   $sig = append($sig,$level[$i]->flat)
   }
$sig -= median $sig;
$sig -> inplace -> abs;
$sig = median $sig;
$sig /= 0.6745;

my $lam = zeroes(8);

for (my $i = 0; $i < 8; $i ++)
   {
   my $x = pop @level;
   $x *= sqrt(2**8+1);
   $x /= $sig;
   my $min = zeroes(119);
   for (my $j = 0; $j < 119; $j ++)
      {
      my $tmp = zeroes(2**$i);
      for (my $k = 0; $k < 2**$i; $k ++)
         {
         $tmp($k) .= inner($x(0:$k,$j),$x(0:$k,$j));
         $tmp($k) += $x($k,$j;-)**2*(2**$i-1-$k);
         $tmp($k) += 2**$i - 2*$k - 2;
         } # end for($k)
      $min($j) .= $x(minimum_ind($tmp),$j;-);
      $min($j) .= 0 if $min($j) > 2**$i;
      $min($j) .= sqrt(2*$i)
         if inner($x(,$j),$x(,$j)) <= 1 + sqrt($i)**3/sqrt(2**$i);
      } # end for($j)
   $lam($i) .= davg $min;
   } # end for($i)

$lam *= $sig;
$lam /= sqrt(2**8+1);
wcols $lam, 'SUREthresh.dat';
```

OPtestparam.pl

```
#!/usr/bin/perl -w

use strict;
use PDL;
use PDL::NiceSlice;

my $d = 119*2**7;
my $cut =zeroes($d,8);
```

```
for (my $i = $d; $i >0; $i --)
   {
   for (my $j = 8; $j > 0; $j --)
      {
      my $inter = ((1-2**-$j)**(1/$i)+1)/2;
      $cut($d-$i,$j-1) .= (ndtri($inter))**2;
      }
   }
wcols dog($cut), 'OPtestparam.dat';
```

OP.pl

```
#!/usr/bin/perl -w

use strict;
use PDL;
use PDL::NiceSlice;

my $w = cat rcols 'wycwavtran.dat';
$w->reshape(257,119);
my $z = $w -> copy;
my @level;
for (my $i=1; $i<9; $i++)
   {
   my $tmp = $z(1:-2:2) -> flat;
   push @level, $tmp;
   $z = $z(0:-1:2;|);
   }

my $sig = zeroes(8);
for (my $i = 0; $i < 8; $i ++)
   {
   my $dat = $level[$i];
   $dat -= median $dat;
   $dat -> inplace -> abs;
   $sig($i) .= median $dat;
   }
$sig /= 0.6745;

my $cut = cat rcols 'OPtestparam.dat';
my $opthresh = zeroes(8);
for (my $i = 0; $i < 8; $i ++)
   {
   my $x = shift @level;
   $x *= 2**4/$sig(($i));
   $x *= $x;
   $x = qsort $x;
   $x = $x(-1:0);
```

```
   my $idx = (which $x < $cut(119*(2**7-2**(7-$i)):-1,7-$i;-))->(0);
   $opthresh($i) .= sqrt($x(($idx)))*$sig*2**-4;
   }
wcols $opthresh, 'OPthresh.dat';
```

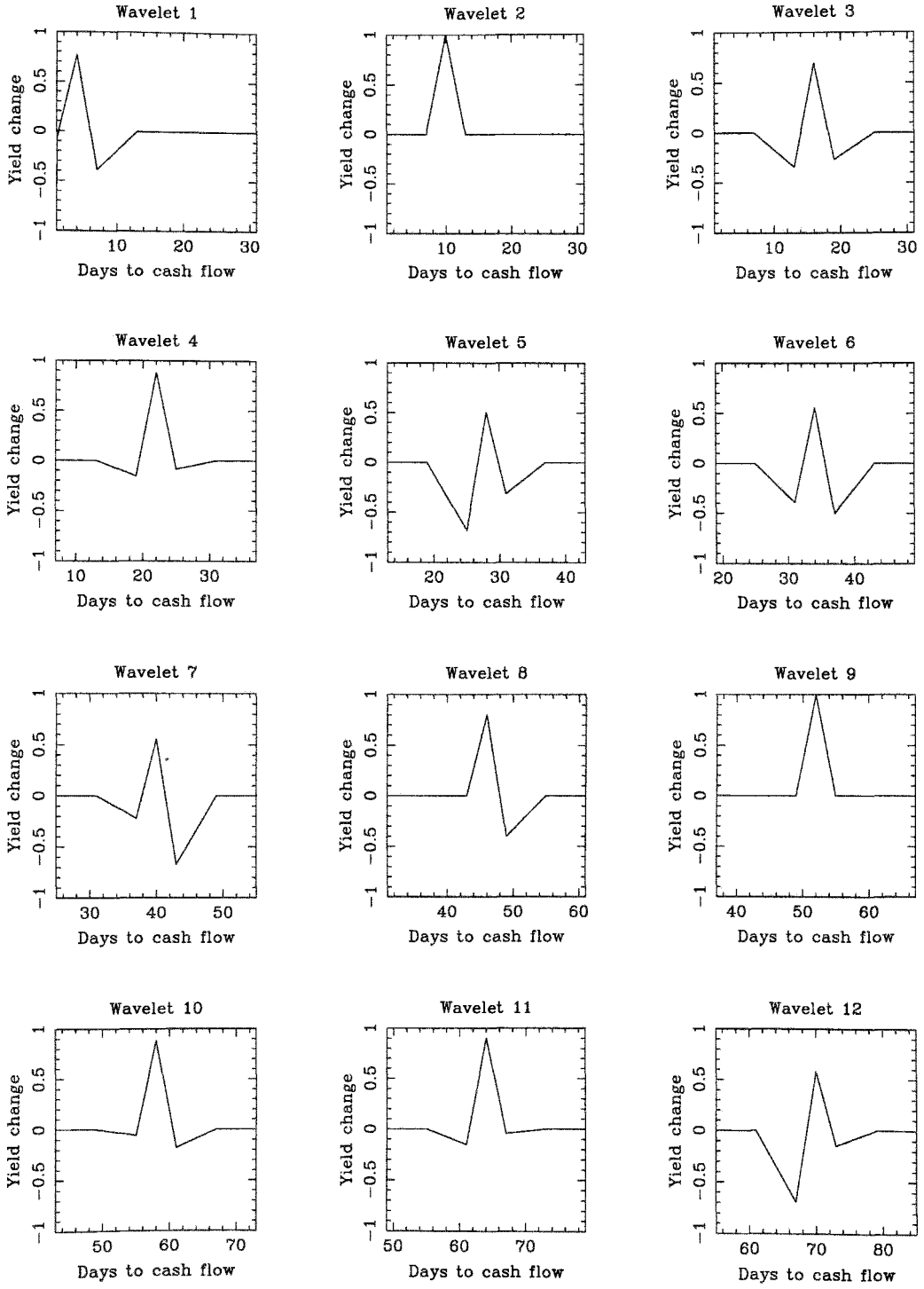
Yield Curve Mesh

ycmesh.dat

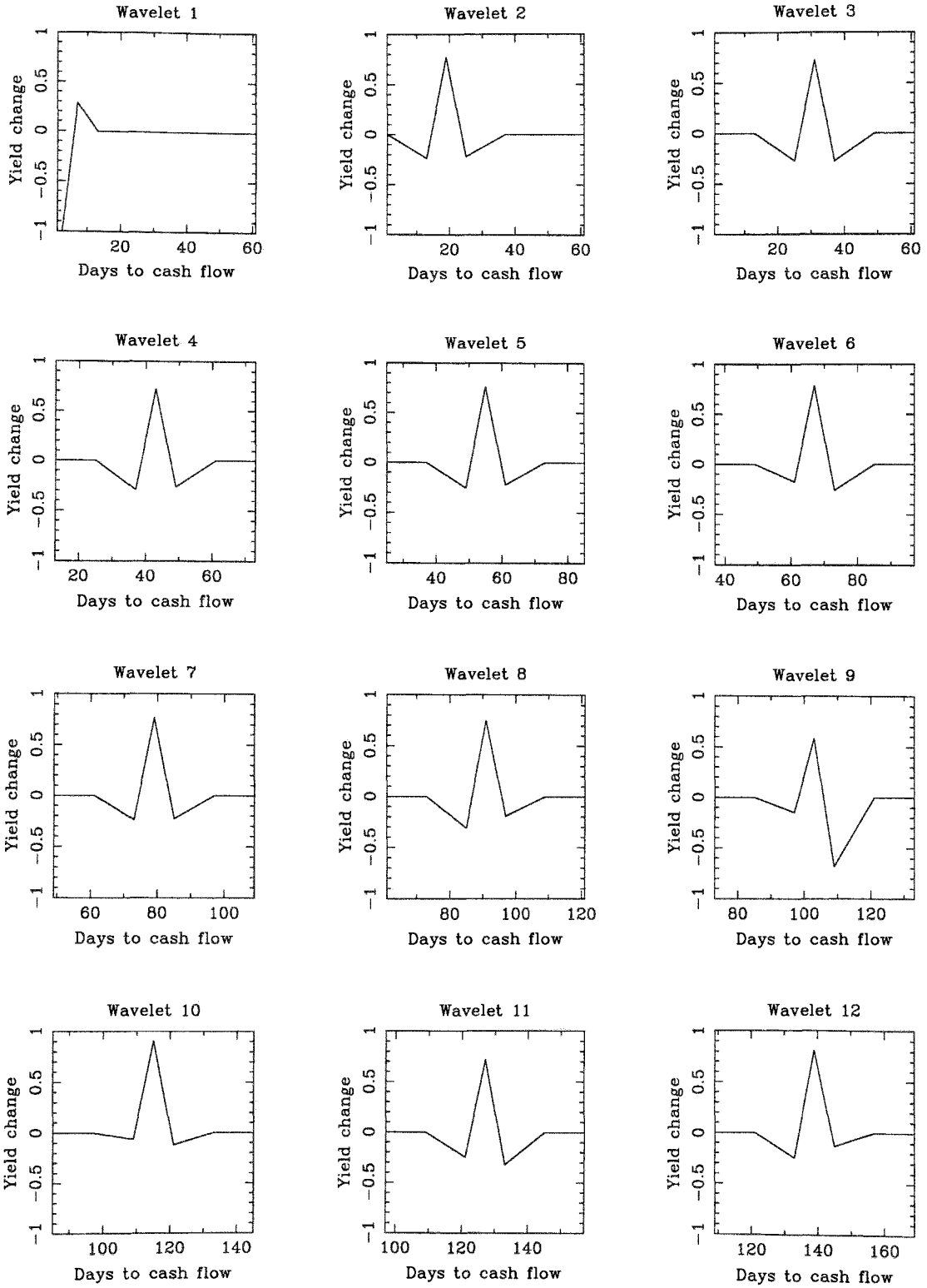
1	4	7	10	13	16	19	22
25	28	31	34	37	40	43	46
49	52	55	58	61	64	67	70
73	76	79	82	85	88	91	94
97	100	103	106	109	112	115	118
121	124	127	130	133	136	139	142
145	148	151	154	157	160	163	166
169	172	175	178	181	184	187	190
193	201.5	210	218.5	227	235.5	244	252.5
261	269.5	278	286.5	295	303.5	312	320.5
329	337.5	346	354.5	363	371.5	380	388.5
397	405.5	414	422.5	431	439.5	448	456.5
465	473.5	482	490.5	499	507.5	516	524.5
533	541.5	550	558.5	567	575.5	584	592.5
601	609.5	618	626.5	635	643.5	652	660.5
669	677.5	686	694.5	703	711.5	720	728.5
737	752.5	768	783.5	799	814.5	830	845.5
861	876.5	892	907.5	923	938.5	954	969.5
985	1000.5	1016	1031.5	1047	1062.5	1078	1093.5
1109	1124.5	1140	1155.5	1171	1186.5	1202	1217.5
1233	1248.5	1264	1279.5	1295	1310.5	1326	1341.5
1357	1372.5	1388	1403.5	1419	1434.5	1450	1465.5
1481	1496.5	1512	1527.5	1543	1558.5	1574	1589.5
1605	1620.5	1636	1651.5	1667	1682.5	1698	1713.5
1729	1760	1791	1822	1853	1884	1915	1946
1977	2008	2039	2070	2101	2132	2163	2194
2225	2256	2287	2318	2349	2380	2411	2442
2473	2504	2535	2566	2597	2628	2659	2690
2721	2752	2783	2814	2845	2876	2907	2938
2969	3000	3031	3062	3093	3124	3155	3186
3217	3248	3279	3310	3341	3372	3403	3434
3465	3496	3527	3558	3589	3620	3651	3682
3713							

Wavelet Graphs

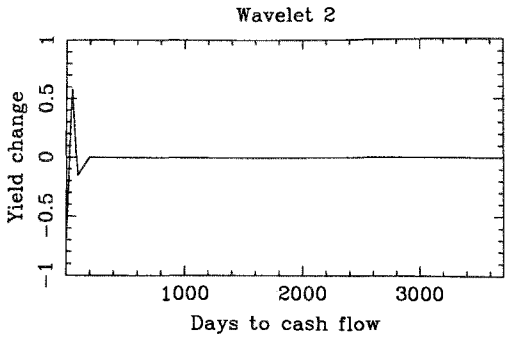
Level −1: The first twelve wavelets of 128, January 1992

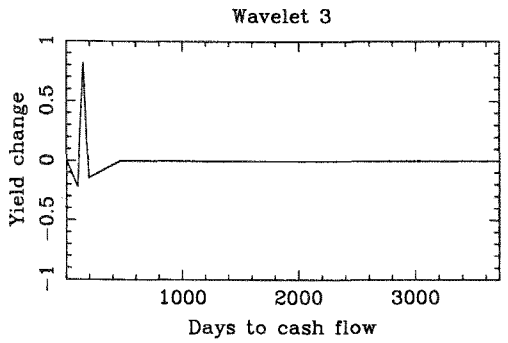


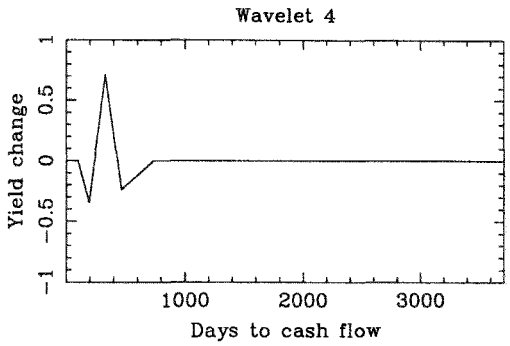
Level −2: The first twelve wavelets of 64, January 1992

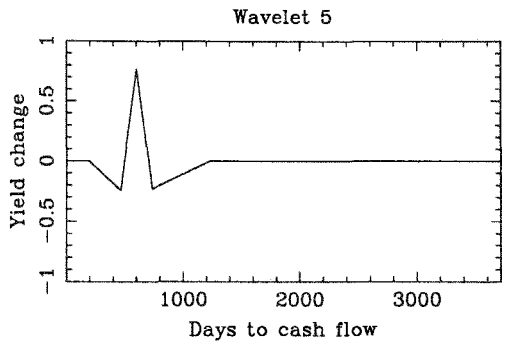


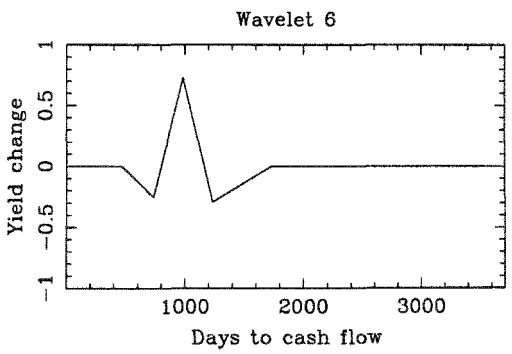
Level −5: The eight wavelets, January 1992

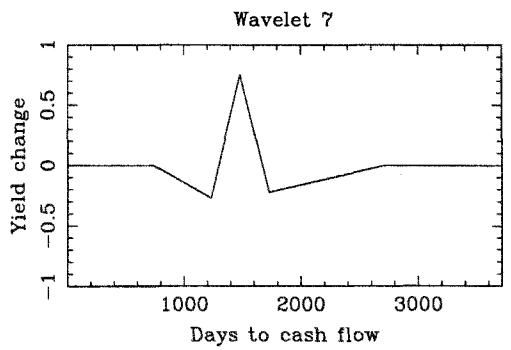


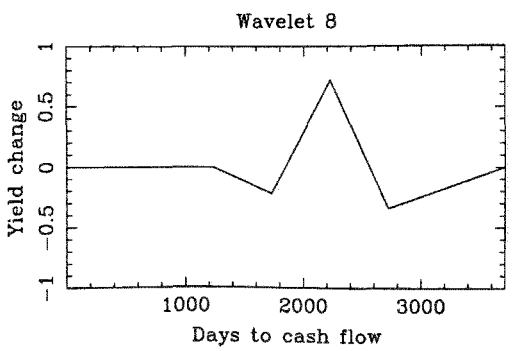


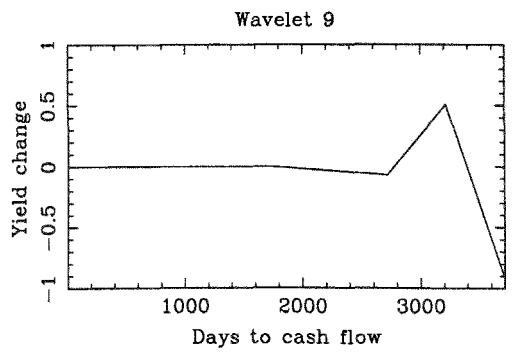




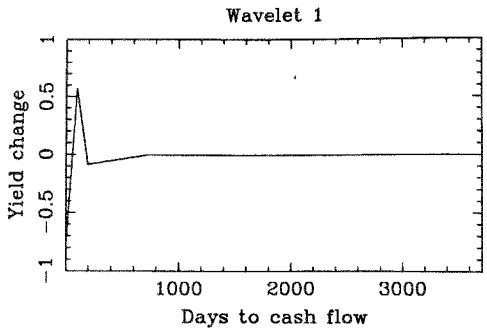


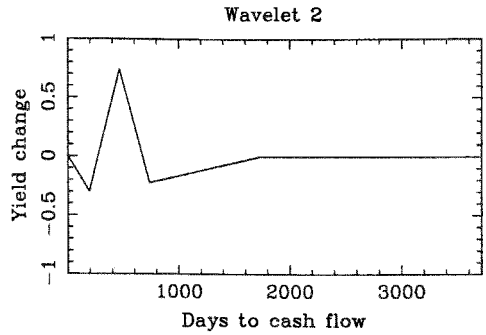


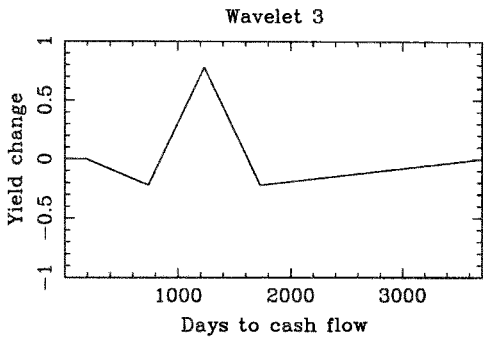


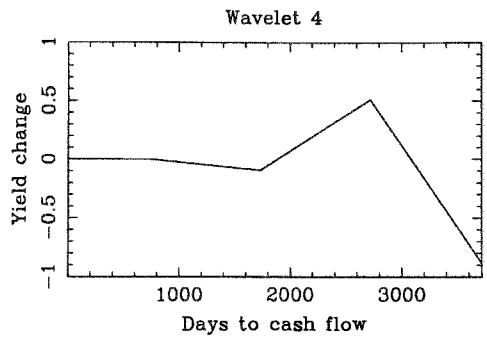


Level −6: The four wavelets, January 1992

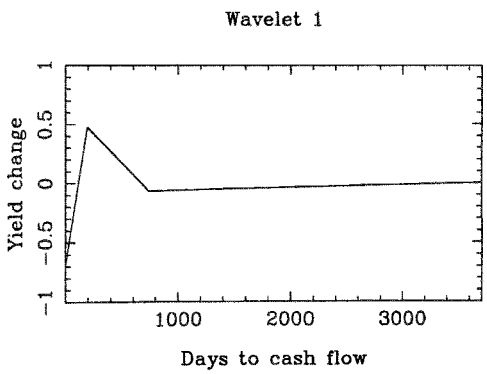


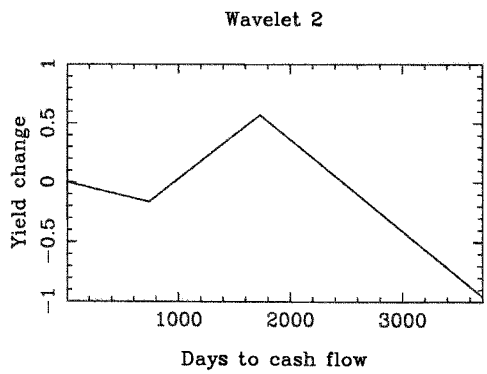




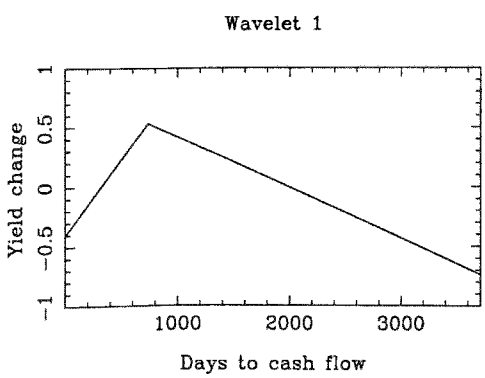


Level −7: The two wavelets, January 1992





Level −8: The only wavelet, January 1992



VITA

MARK L. COPPER

1976	B.A., Liberal Arts St. John's College Santa Fe, New Mexico
1981	M.S., Mathematics University of California Irvine, California
1988	Ph.D., Mathematics University of Georgia Athens, Georgia
1988–1996	Assistant Professor, Mathematics Florida International University Miami, Florida
1998–2003	Senior Lecturer, Finance Wayne State University Detroit, Michigan