Capturing Non-Linearity in the Term Structure of Interest Rates: A Fuzzy Logic Method of Estimating the Yield Curve

Richard Taylor

and

David E.A. Giles

Department of Economics
University of Victoria

(This Draft, June 2003)

Abstract:

This paper provides a new solution to the problem of estimating the yield curve from a sample of coupon bonds. Our approach builds on the work of McMulloch (1971, 1975) by using a new approach to formulate and estimate the discount function for US Government Issue Coupon Bonds. This new method uses fuzzy regression to estimate an arbitrary nonlinear function, rather than estimating a polynomial specification, after using fuzzy clustering to break the bond sample into various clusters based on term to maturity. Estimates from the fuzzy regression are used to calculate the discount function, yield curve and the theoretical prices for the in-sample securities. Finally, estimates of the standard errors of the yield curve, discount function, price estimates, and forward curves can be obtained by bootstrapping. Our fuzzy analysis provides a very flexible way of dealing with the inherent nonlinearities in the problem, without imposing an arbitrary functional form and it provides some encouraging results for the data that have been analyzed to date.

Author Contact: Professor David Giles, Department of Economics, University of Victoria, P.O. Box 1700 STN CSC, Victoria, BC, CANADA V8W 2Y2
FAX: (250) 721-6214; Voice: (250) 721-8540; email: dgiles@uvic.ca

1. Introduction:

The “term structure of interest rates” serves many applied purposes in the world of finance and economics. For a government financier, it can serve as a guide for issuing new debt with the least cost. For the bond dealer, it can direct the pricing of a stripped bond issue, aid in pricing government issues and identifying arbitrage opportunities. For the economist, it can be used to provide insight into the path of future interest rates or help to identify credit conditions in the economy. These are but a few of the many applications of the yield curve whether on the street or in academia. Accordingly, two vital questions that emerge are, “what is the yield curve” and “how is it derived?”

Simply put, the yield curve is a schedule of returns on zero coupon bonds of increasing terms to maturity. More importantly, those yields are taken from government issued securities so it is more accurately termed the *riskless zero coupon* yield schedule. The yield curve can take many shapes but is more often than not upward sloping indicating a premium investors demand on long term investments.

This highlights the first hurdle facing the practitioner who attempts to estimate the yield curve: any model used to capture the relationship between yield and term to maturity must be flexible enough to incorporate all of the possible shapes that the curve can take. Furthermore, this flexibility must be balanced cautiously with the precision of the model as these characteristics usually work at odds with each other.

Given that there is no liquid market for zero coupon bonds of varying maturities, the yield curve must be estimated from a sample of coupon bearing instruments. Most industrialized economies regularly issue coupon bearing bonds in the open market to finance the financial obligations of the government. This creates a very liquid market for coupon bonds of varying degrees of maturity, and in principle the information about these can be used to estimate the yield curve.

There has been a great deal of research and debate into methods of estimating the yield curve because the underlying model specification that is chosen can greatly affect its shape. Not only is there a need for precise, flexible models, but also we require criteria under which those models can be compared with each other. This paper attempts to address both of these issues by providing a new model for estimating the yield curve, and comparing it with a naive base case. The analysis is restricted to a cross-section of data at one point in time, but it could be extended easily to facilitate a time series estimation of the yield curve.

The comparison of models will include many factors, but in this study we focus on deriving a discount function, and in turn a yield curve, that accurately prices the bonds in sample. However, this precision in bond pricing must not come at the cost of unrealistic estimates for the yield curve and forward curves.

The next section of the paper provides a brief outline of the approach to estimating the yield curve that is proposed by McCulloch (1971, 1975), and there we suggest the use of fuzzy clustering and fuzzy regression as a useful approach. Section 3 provides the background details of this fuzzy analysis. The formulations of our fuzzy discount functions, yield curves, forward curves, and price estimates are presented in section 4. An empirical application is provided in section 5, and some concluding comments and directions for extending this research appear in the last section.

2. Estimating the Yield Curve: A General Approach

2.1 General Theory

It is always the case in finance that the price of a security is valued as the sum of the present values of all the future cashflows endowed to the owner. In the world of fixed income securities with no options, the amount and timing of future cash flows are known apriori and in the case of government issued securities, certainty of payment is assured. The price of the *ith* bond in sample will be:

$$P_i = 1000PV(M_i) + \sum_{m=1}^{M_i} C_i PV(m)$$

Where \$1000 is the face value, C_i is the coupon payment, and M_i is the final term to maturity of each bond. PV(M) stands for present value. This representation was used because it illustrates that the price of a bond is simply a weighted sum of the cash flows where the more recent cash flows take on greater weight than those not scheduled for some time in the future. Specifying PV(M) as a general function of M and assuming that coupon payments arrive continuously, the theoretical price of a bond becomes:

$$P_i = 100\delta(M_i) + c_i \int_0^{M_i} \delta(m)dm$$

Where the function $\delta(M)$ is the discount function that provides the proper weight on each classify to appropriately price the bond. It is a smooth function of time and continuously differentiable. $\delta(M)$ can be broken into two parts, the first is the specified component and the second is the estimable component. Least Squares estimation requires that the parameters of $\delta(M)$ enter the function in a linear fashion. This entails that $\delta(M)$ will have to a linear combination of continuously differentiable functions f_j:

$$\delta(m) = \alpha_0 + \sum_{j=1}^{k} \alpha_j f_j(m)$$

Furthermore, given that the present value of a cashflow received today is itself, $\alpha_0 = 1$ and $f_j(0)=0$, we can further specify the discount function as:

$$\delta(m) = 1 + \sum_{j=1}^{k} \alpha_j f_j(m)$$

Substituting the general discount function into the theoretical price of each bond yields:

$$\hat{P}_i = 100\left(1 + \sum_{j=1}^{k} \alpha_j f_j(m)\right) + c_i \int_0^{M_i} \left(1 + \sum_{j=1}^{k} \alpha_j f_j(m)\right) dm$$

McCulloch (1971) provides a manipulation of the general discount function into the following format that can then be estimated by Least Squares.

$$y_i = \sum_{j=1}^{k} \alpha_j x_{ij} + \varepsilon_i$$

where:

$$y_i = p_i - 100 - c_i M_i$$

and:

$$x_{ij} = \sum_{j=1}^{k} \left(100 f_j(M_i) + c_i \int_0^{M_i} f_j(m) dm \right)$$

McCulloch suggests using a weighted least squares regression where the variance of the residuals is proportional to the mean bid ask spread of each bond.

$$var(\varepsilon_i) = \sigma^2 v_i^2$$

where:

$$v_i = \frac{(p_i^a - p_i^b)}{2}$$

Once the α_j ‘s have been estimated, the discount function is approximated as:

$$\hat{\delta}(m) = 1 + \sum_{j=1}^{k} \hat{\alpha}_j f_j(m)$$

The discount function is an exponential decay function and its rate of decay at any point in its range is given by the instantaneous forward rate $\rho(m)$. It is defined as:

$$\rho(m) = -\frac{\delta'(m)}{\delta(m)}$$

The forward rate function can be approximated by:

$$\hat{\rho}(m) = -\frac{\sum_{j=1}^{k} \hat{\alpha}_j f'_j(m)}{1 + \sum_{j=1}^{k} \hat{\alpha}_j f_j(m)}$$

The instantaneous forward rate is the rate of decay at every point on the yield curve, the yield curve is then the average rate of decay over an interval of the forward curve. If the coupon payments were to be received continuously, the yield curve $\eta(m)$ would be calculated as:

$$\eta(m) = -\frac{1}{m}\log(\hat{\delta}(m))$$

Finally, the estimates of the prices of each bond is calculated as follows:

$$P_i = 100 + c_i M_i + \sum_{j=1}^{k} \alpha_j \left(100 f_j + c_i \int_0^M f_j(m) dm\right)$$

A natural choice for f_j is a polynomial because it meets the restriction that $f_j(0)=0$ and it is relatively straightforward to estimate. Another choice suggested by McCulloch is to make it a piecewise polynomial approximation or spline function. Spline functions are convenient for this application because they add an element of flexibility to the model while not violating any of the essential assumptions. However, there is are some added complexities to fitting splines in that the number of knots and the location of those knots greatly affects the shape of the yield and forward curves. For further details regarding spline analysis in general, see Poirier (1976).

Examples of other forms of the discount function used include, Vasicek and Fongs (1982) exponential spline and Nelson and Siegal's (1987) highly non-linear function of the instantaneous forward rate. Shea (1984, 1985) pointed out some of the drawbacks of using both exponential and cubic splines.

In the application reported in section 5 below, an 8th degree polynomial was chosen as the base specification of the discount function. As an alternative approach, we incorporated a fuzzy partitioning algorithm to split the data into fuzzy clusters on the basis of the term to maturity of each bond. A low-order polynomial was then incorporated over each cluster to provide estimates of the discount function, the yield curve, the forward curve and the bond prices.

2.2 Motivation for a Fuzzy Logic Model

Examining the expression for the price of the ith bond in the sample, it is apparent that it is a function of the term to maturity and coupon rates for each bond. To motivate a fuzzy logic formulation, two bonds were taken from an actual bond sample. They have much different terms to maturity but the same coupon rates. Table 1 below illustrates that as the amount of time to the face value repayment increases, its weight in the price of the bond lessons and the weight of coupon stream increases.

Table 1

TTM	Coupon Rate	Price	Yield to Maturity	PV of Coupon Stream	Percent of price attributed to Coupon Stream
2.25	5.75	104.23	3.76	122.95	12%
30.75	5.75	104.37	5.45	853.12	51%

A global (eg, polynomial) approximation of the discount function uses the same parameter estimates to calculate the discount function no matter what the term to maturity of the bond is. There are many bonds of term to maturity less than, say, ten years and very few comparatively of greater than ten years. Accordingly, global approximations tend to over-fit the data at the short end of the term structure, and under-fit at the long end. It would be advantageous to split the discount function over sub-sections of the term structure and ensure that when approximating the discount function over different terms to maturity, information from the relevant term to maturity section is weighted most heavily. Spline functions provide a means of addressing this problem. However, there are no hard and fast rules of specifying the number of knots or the locations of those knots when applying this type of analysis.

By using the fuzzy logic approach that is discussed in the next section, not only are the sub-sections of the yield curve produced endogenously, but the methodology itself ensures that if one is calculating a zero-coupon yield of (say) 22 years, the parameter estimates from that period are having the largest impact on the calculation.

3 Fuzzy Modelling

In this section we discuss some basic notions from fuzzy set theory and fuzzy logic, as originally introduced by Zadeh (1965, 1967). These ideas are then used to introduce the fuzzy regression analysis (Shepherd and Shin, 1998; Giles and Draeseke, 2003; Giles and Mosk, 2003; Giles and Stroomer, 2003) that was foreshadowed in the last section. The aim of fuzzy clustering is to partition a sample into subsets, where every point of each subset has similar characteristics. Methods vary in terms of what criteria are used to assign membership of each point to each subset. The data may be multi-dimensional, though in our case we will be classifying the data on the basis of the characteristics of just a single variable, namely the term to maturity of each bond.

3.1 Fuzzy Sets

A fuzzy set or fuzzy cluster is one whose borders are vague, rather than crisp. So, given a universal set and its elements, the latter generally will be associated with every set in the universe, with some degree of membership. The degree of membership is a measure between zero and unity that reflects how well an element fits into a cluster. Every element has a (usually different) degree of membership with respect to each fuzzy set. To illustrate this idea, consider an example. If there are three people that we want to assign to two conventional (crisp) sets on the basis of height, we might proceed as in Table 2:

Table 2

Person	Height	Set
One	5’3’’	short
Two	6’5’’	tall
Three	6’2’’	tall

In this example, person One belongs to the set associated with short people and persons Two and Three to the set associated with tall people. Another way of putting that is that person One has a degree of membership with the Short set of 1, and a degree of membership of 0 with the Tall set. Conversely, persons Two and Three each have a degree of membership with the Tall set of 1, and 0 with the Short set.

In the context of fuzzy sets, person One would have a degree of membership (between 0 and 1) with both the Short and the Tall fuzzy sets. Obviously, the degree of membership with the Short fuzzy set would be larger than the degree of membership with the Tall fuzzy set. As for person Two, the degree of membership would be greater for the Tall set than the Short fuzzy set. However, person Two's degree of membership with the Tall fuzzy set would be smaller than person Three's, because person Three is taller than person Two. An example of how the membership values might be assigned is shown in Table 3:

Table 3

Person	Degree of Membership with Short Set	Degree of Membership with Tall Set
One	0.8	0.2
Two	0.1	0.9
Three	0.3	0.7

It can be seen that the degrees of membership associated with a particular person sum to unity across the fuzzy sets in Table 3. Although we will be using a computational algorithm that enforces this property in our own analysis below, in general this restriction does not necessarily have to be introduced. In the context of the yield curve model, bonds can be grouped together loosely on the basis of their term to maturity and coupon rates. Under this framework, bonds with high, medium and low coupon rates and terms to maturity are grouped together. Once these clusters have been established, a model can be estimated over each fuzzy cluster and then the results can be combined over the entire sample in a natural weighted average manner. The details of this procedure are presented below. First, let us consider the process that is used to partition the sample data into fuzzy clusters.

3.2 A Fuzzy Clustering Algorithm

Given that the number of clusters (c) has been chosen *a priori*, a procedure known as the Fuzzy c-means (FCM) algorithm (e.g., Ruspini, 1970) can be employed to partition the n data-points into c fuzzy clusters, while simultaneously determining the locations of the clusters. Locations is a loose term for cluster centres as it will be explained below that the choice of cluster centre dictates the nature of the partition. The metric used to calculate the distance between a point in the sample and a cluster centre is squared difference.

Giles and Draeseke (2003) outline the FCM algorithm in the following terms:

"Let x_k be the k^{th} (possibly vector) vector data-point (k = 1, 2, ..., n). Let v_i be the center of the i^{th}. (fuzzy) cluster (i = 1, 2, ..., c). Let $d_{ik} = \| x_k - v_i \|$ be the distance between x_k and v_i , and let u_{ik} be the degree of membership of data-point k in cluster i , where :

$$\sum_{i=1}^{c} u_{ik} = 1$$

The objective is to partition the data-points into the c clusters, and simultaneously locate those clusters and determine the associated degrees of membership , so as to minimize the functional

$$J(U, v) = \sum_{i=1}^{c} \sum_{k=1}^{n} (u_{ik})^m (d_{ik})^2$$

There is no prescribed manner for choosing the exponent parameter, m , which must satisfy $1 < m < \infty$. In practice, $m = 2$ is a common choice. In the case of crisp (hard) memberships, $m = 1$.

In broad terms, the FCM algorithm involves the following steps:

1. Select the initial location for the cluster centres.
2. Generate a (new) partition of the data by assigning each data-point to its closest cluster centre.
3. Calculate new cluster centres as the centroids of the clusters.
4. If the cluster partition is stable then stop. Otherwise go to step 2 above.

In the case of fuzzy memberships, the Lagrange multiplier technique generates the following expression for the membership values to be used at step 2 above:"

$$u_{ik} = \left\{ 1 \Big/ \left(\sum_{j=1}^{n} [(d_{ik})^2 / (d_{jk})^2] \right) \right\}$$

There are three products from the minimization of *J(U,v)*, the first are the cluster centroids but more importantly is the partition for the data into c clusters where each point is allocated to the-cluster whose centroid it is closest to. Finally, given that partition, the degrees of membership for each point with each cluster are calculated.

3.3 Fuzzy Regression

To see how this methodology can be incorporated into an econometric framework, consider a simple regression model with a dependant variable, *Y*, and a single regressor, *X*. One would partition the sample into fuzzy clusters by applying the FCM algorithm to the *X* data, and then estimate the relationship separately over each cluster. This yields *c* sets of parameter estimates that are then combined, using the membership values as the weights, to model the conditional mean of the dependant variable as:

$$\hat{Y}_j = \sum_{i=1}^{c} (\beta_{i0} + \beta_{i1} X_j) u_{ij}$$

where $j = 1 \ldots n$ and n is the sample size.

Each data point for the independent variable in the sample has a degree of membership with each fuzzy cluster, so the parameter estimates obtained from every cluster are used to estimate the conditional mean of the model. Note that for any given point in the sample, the parameter estimates for the cluster that the point best belongs to has the largest impact on the calculation of the conditional mean. Because the conditional mean is constructed using weights that vary continuously throughout the sample, this fuzzy regression analysis allows us to capture highly nonlinear relationships in a very flexible manner. Giles and Draeseke (2003), Giles and Mosk (2003) and Giles and Stroomer (2003) illustrate this point in a number of empirical applications, and they demonstrate its superiority over conventional nonparametric kernel regression.

There is one remaining issue that needs to be discussed. When one wishes to calculate predicted values out of the sample, the degrees of membership for the out-of-sample independent data point must be calculated. The solution to this problem can be best illustrated by using an example.

Table 4

Y	X	u_1	u_2
5	2	0.9	0.1
10	6	0.2	0.8
12	7	0.1	0.9
6	3	0.6	0.4

Table 4 provides information about a hypothetical data set where X is the independent data, u_1 and u_2 represent the degrees of membership for X in the case of two fuzzy clusters. Given the sample of data, the conditional mean is easily modelled in the manner just outlined. However, what if one is interested in the conditional mean of Y given a value of $X=5$(which is not a sample value)? In order to derived the unknown degrees of membership, the new data point is added to the X series which can then be re-partitioned using the FCM algorithm, resulting in 'updated' membership values for *all* of the data points. As that there is no Y value available to be matched with the extra X value, we cannot update the cluster regressions, so the best that can be done is to use the *original* parameter estimates from the original dataset to construct a prediction of the conditional mean as shown above, when $X=5$. The slight approximation associated with this approach my be problematic if the new value of X is an outlier relative to the original sample, but this is not an issue in its application here.

4. The Fuzzy Logic Yield Curve Model

Using the fuzzy clustering/fuzzy regression analysis, the polynomial specification that forms part of the yield curve and other functions, outlined in section 2 need not change. However, one major benefit of using our fuzzy regression analysis is that the specification need not actually be a polynomial over each fuzzy cluster. It could be any parametric function, or even a non-parametric representation. The underlying nonlinearity that is captured by a single polynomial over the full sample in the conventional analysis is dealt with in our case by fitting separate models over each cluster, and then combining the results with continuously changing weights. Indeed, in the fuzzy regression application discussed by Giles and Draeseke (2003), Giles and Mosk (2003), and Giles and Stroomer (2003), the separate regression over the individual clusters are based on *linear* models.

Starting with the discount function, the original formulation given in section 2 is modified as follows:

$$\hat{\delta}(m) = 1 + \sum_{i=1}^{c'} u_{im} \sum_{j=1}^{k} \hat{\alpha}_{ij} f_j(m)$$

Intuitively, when one is attempting to present value a dollar three years from now, information from the sample on a thirty year bond is of little value whereas bonds with term to maturities of three years or less are quite valuable. By incorporating the degrees of membership in the discount function, this prioritization is made possible.

Substituting the new fuzzy discount function into the price of the ith bond yields:

$$\hat{P}_i = 100\left(1 + \sum_{i=1}^{c'} u_{im} \sum_{j=1}^{k} \hat{\alpha}_{ij} f_j(m)\right) + c_i \int_0^{M_i} \left(1 + \sum_{i=1}^{c'} u_{im} \sum_{j=1}^{k} \hat{\alpha}_{ij} f_j(m)\right) dm$$

Factoring out the parameters yields:

$$\hat{P}_i = 100 + 100 \sum_{i=1}^{c'} u_{im} \sum_{j=1}^{k} \hat{\alpha}_{ij} f_j(m) + c_i M_i + c_i \int_0^{M_i} \sum_{i=1}^{c'} u_{im} \sum_{j=1}^{k} \hat{\alpha}_{ij} f_j(m) dm$$

Which reduces to:

$$\hat{P}_i - 100 - c_i M_i = \sum_{i=1}^{c'} u_{im} \sum_{j=1}^{k} \hat{\alpha}_{ij} \left(100 f_j(M_i) + c_i \int_0^{M_i} f_j(m)\right)$$

The yield curve itself is taken directly from the discount function so it needs no adjustments made to it. However, the forward rate function must be adjusted slightly as follows:

$$\hat{\rho}(m) = -\frac{\sum_{i=1}^{c'} u_{im} \sum_{j=1}^{k} \hat{\alpha}_{ij} f'_j(m)}{1 + \sum_{i=1}^{c'} u_{im} \sum_{j=1}^{k} \hat{\alpha}_{ij} f_j(m)}$$

Finally, fuzzy price estimates are arrived at as follows:

$$P_i = 100 + c_i M_i + \sum_{i=1}^{c'} u_{im} \sum_{j=1}^{k} \alpha_j \left(100 f_j + c_i \int_0^{M_i} f_j(m) dm \right)$$

In order to construct the constant maturity yield curve, the degrees of membership for those constant maturities would be calculated using the same methodology as was outlined in the previous section. The data set containing all term to maturities would be augmented by a constant maturity vector and re-partitioned. Those constant maturity degrees of membership would then be incorporated in the above calculation.

5. An Application

5.1 The Data

The data set used in this study was most kindly provided by Dr. Hu McCulloch. It comprises monthly bond and bill quotes of US government securities from January 1983 to December 1991. For estimation purposes, the data associated with Treasury Bill along and with bonds with no coupon payments remaining were omitted. This is consistent with other studies such as Dobson (1978). Bonds that were callable were included in the estimation however, the term to maturity was adjusted so that the term to call was used for bonds selling above par and the term maturity was used to bonds selling below par.

The fuzzy partitioning algorithm and bootstrapp procedure was coded into a matlab package that is available by request. The authors would like to thank Dr. James P. LeSage for the use of his econometrics toolbox for matlab.

For illustrative purposes, the estimation was taken for one point in time, namely December 1985. After the above mentioned bonds were omitted, a sample of 161 coupon bonds were remaining. Work in progress includes estimation with additional data, as the real test of the model is how it performs over time.

5.2 Results

The results of our illustrative application appear in Figure1. There the dark solid black line is the yield curve for a six cluster partition of the data and a second-order polynomial specification over each cluster. It can be seen that as more clusters are added, the yield curve loses the two hump at the long end. The fuzzy estimation varied from a first-order polynomial and three to six clusters, to a second-order polynomial and three to six clusters. All first-order approximations led to erratic yield curves and negative valued discount functions. As can be seen, the results based on the second-order polynomial are quite acceptable.

The dark solid black line is the yield curve for a six cluster partition of the data and a second order polynomial. You can see that as more clusters are added, the yield curve loses the two humps at the long end. The fuzzy estimation varied from a first order polynomial and three to six clusters to a second order polynomial and three to six clusters. All first order approximations led to erratic yield curves and negative discount function. The second order polynomial worked much better.

The conventional polynomial approximation, as outlined in section 2, provides a yield curve that is quite smooth for the first eighteen years but then spikes to over 14% for thirty years before dropping to negative interest rates afterwards. The negative interest rates are not of great concern because our analysis is confined to the range of maturities that are found in the bond sample. In this case, this is just over thirty years. The results from the six cluster fuzzy regression and the conventional polynomial specification are quite similar throughout the term structure, but they diverge at the long end of the term structure. The humps apparent with the polynomial model are much smoother with all of the fuzzy cluster models.

Figure 2 presents the corresponding results for the discount function. As can be seen there, the shape of the discount function for the fuzzy regression model is not particularly sensitive to the choice of the number of clusters. As this number increases, the estimated discount function becomes smoother. Although the fuzzy discount function based on only four clusters does slope up sharply after twenty five years, in general the fuzzy modelling approach produces more satisfactory results for longer terms to maturity than does the conventional polynomial model.

Shea (1984) also observed upward sloping discount function at the long end of the term structure using data for Nippon Telegraph and Telephone zero coupon issues and a piece-wise polynomial approximation. The fact that this model suffers from this problem is of some concern. Since Shea (1984, 1985) a number of authors including Fisher, Nychka, and Zervos (1995) and Waggoner (1997) have implemented smoothing splines to address this problem. Smoothing splines attempt to strike a balance between fitting the in sample and out of sample bond prices while producing a discount function that is negatively sloped across the entire term structure. This is an avenue that is currently being explored. However, the purpose of our work here is to introduce the fuzzy logic model and to show that it produces precise estimates of bond prices.

5.3 Model Evaluation

McCulloch (1971) provides exact and approximate standard errors for the estimates of the yield curve, forward, estimated prices, and discount function. This is one approach that can be taken when evaluating one model versus another. However, in the case of a fuzzy logic model, the derivations of the standard errors of the discount function and price estimates are adjusted slightly.

The expressions for the standard errors of the discount function, yield curve, forward curve and price estimates outlined by McCulloch (1971) are:

$$v\hat{a}r(\hat{\delta}(m)) = (z^T Cz)$$

where:

$$z_j = f_j(m)$$

and C is the estimated variance covariance matrix for the estimated parameters taken from the-Weighted Least Squares (WLS) regression.

Similarly,

$$v\hat{a}r(\hat{\eta}) = \frac{v\hat{a}r(\hat{\delta}(m))}{[m(\hat{\delta}(m))]^2}$$

$$v\hat{a}r(\hat{\rho}(m)) = \hat{\rho}(m)^2(z^T C z)$$

where:

$$z_j = [f'_j(m) / \hat{\delta}'(m)] - [f'_j(m) / \hat{\delta}(m)]$$

Finally:

$$v\hat{a}r(\hat{p}_i) = (z^T C z)$$

where:

$$z_j = \left(100 f_j(M_i) + c_i \int_0^{M_i} f_j(m) dm \right)$$

The expressions for the variance of the discount function and price estimates require the covariance matrix for the parameters. However, those expressions use the variance-covariance matrix from a single WLS regression. In the fuzzy regression analysis, matters are more complicated. Given that we have a number of different regressions, one for each of the fuzzy clusters, there will be a number of covariance matrices that somehow need to incorporated into the model evaluation. This requires the re-formulation of the formulae for the standard errors.

This reformulation addresses the issue that there will not only be covariance between parameter estimates within the same cluster, but also *across* clusters. These cross cluster covariances must be somehow estimated and in order to do this, the regressions across clusters must be estimated as a system of unbalanced Seemingly Unrelated Regressions (SUR). The unbalanced aspect of the estimation (*e.g*, see Srivastava and Giles, 1987, pp. 339-346) arises because it is most unlikely that the FCM algorithm will result in fuzzy clusters with equal numbers of observations in each one. Once method of estimating an unbalanced SUR model is to incorporate intercept and slope dummy variables that select the appropriate clusters in the sample. Giles and Draeseke (2003) illustrate these calculations.

The results are parameter estimates over each cluster and an asymptotic variance-covariance matrix. The formula for the variance of the discount function is the now the quadratic form:

$$v\hat{a}r(\hat{\delta}(m|u)) = (z^T Cz)$$

where:

$$z_{jm} = f_j(m)u_{im}$$

And C is a $(kc \times kc)$ matrix where k is the degree of polynomial over each cluster and c is the number of clusters. z_{jm} denotes the $j*m^{th}$ entry in the row vector z where $j = 1 \ldots k$ and $m = 1 \ldots c$.

This same asymptotic covariance matrix is used for the estimates of the standard errors of the price estimates as follows:

$$v\hat{a}r(\hat{p}_i) = (z^T Cz)$$

where:

$$z_{jm} = \left(100 f_j(M_i) + c_i \int_0^{M_i} f_j(m)dm\right)u_{im}$$

The standard errors that result from this have only asymptotic validity, and so we would obtain only asymptotic confidence intervals for the various curves that we are estimating.

Alternatively, all of the standard errors and confidence intervals can be computed using the bootstrapping method. This has the merit of generating results that are specific to our particular sample size and sample values, and this is the approach that we have adopted in our application here, both for the conventional polynomial modelling and the fuzzy modelling. Our bootstrap algorithm can be summarized as follows:

(i) Estimate the α_j's .

(ii) Sample with replacement from the vector of residuals produced in (i).

(iii) Calculate new y_i's using the independent variables, original parameter estimates and the residuals.

(iv) Use the new y_i's along with the independent data to estimate new $\alpha_j's$.

(v) Use the new $\alpha_j's$ to calculate the discount function, yield curve, forward curves, and estimated bond prices.

(vi) Go back to (ii) above, and repeat (say) 1000 times.

The result would be a large set of discount function, yields, parameter estimates, forward curves, and price estimates. The standard deviations from these sets of estimates are then be used as estimates for the true standard errors. This approach has to be altered slightly again for the fuzzy logic model as follows:

(i) Partition the data and estimate each α_{ij} over each cluster, where $i = 1 \ldots c$ clusters and $j = 1,2$ for a second order polynomial estimated over each cluster.

Over each cluster do the following:

(ii) Sample with replacement from the vector of residuals for the cluster at hand. This will produce a new vector of residuals over that cluster.

(iii) Calculate new y_i's over the cluster at hand using the independent variables for that cluster, the set of parameter estimates and the residuals produced in (ii).

(iv) Use the new y_i's with the independent data to estimate a new α_{ij} for the current cluster.

(v) Take the new α_{ij} the $\alpha_{ij}'s$ that do not correspond with the cluster at hand along with the degrees of membership for that cluster to calculate the discount function, yield curve, forward curves, and estimated bond prices.

(vi) Go back to (ii) and repeat (say) 1000 times.

To illustrate these procedures, we have applied them to the estimated, theoretical prices of the bonds in our sample. The evaluation of the models based on theoretical price estimates was chosen because the theoretical prices are in-sample estimates whereas the discount function, yield curve, and forward curves are all predicted values of term to maturities that do no lie in the sample. Figure 3 is a scatter plot of two standard deviations from zero for both the polynomial and five cluster fuzzy model, based on the bootstrap procedures outlined above. The price of bonds in dollars is represented on the horizontal axis. It clearly shows the advantage of the fuzzy logic model over the polynomial approximation, in that the fuzzy logic model accurately prices bonds selling well above and below par. These bootstrapped estimates were calculated using 10000 iterations for the polynomial approach and 1000 iterations over each cluster for the five cluster model, totalling 5000 iterations for the fuzzy model. Although these results are quite preliminary, they are certainly encouraging and they strongly suggest that the methodology that we have proposed in this paper deserves serious consideration and further investigation.

Figure 1

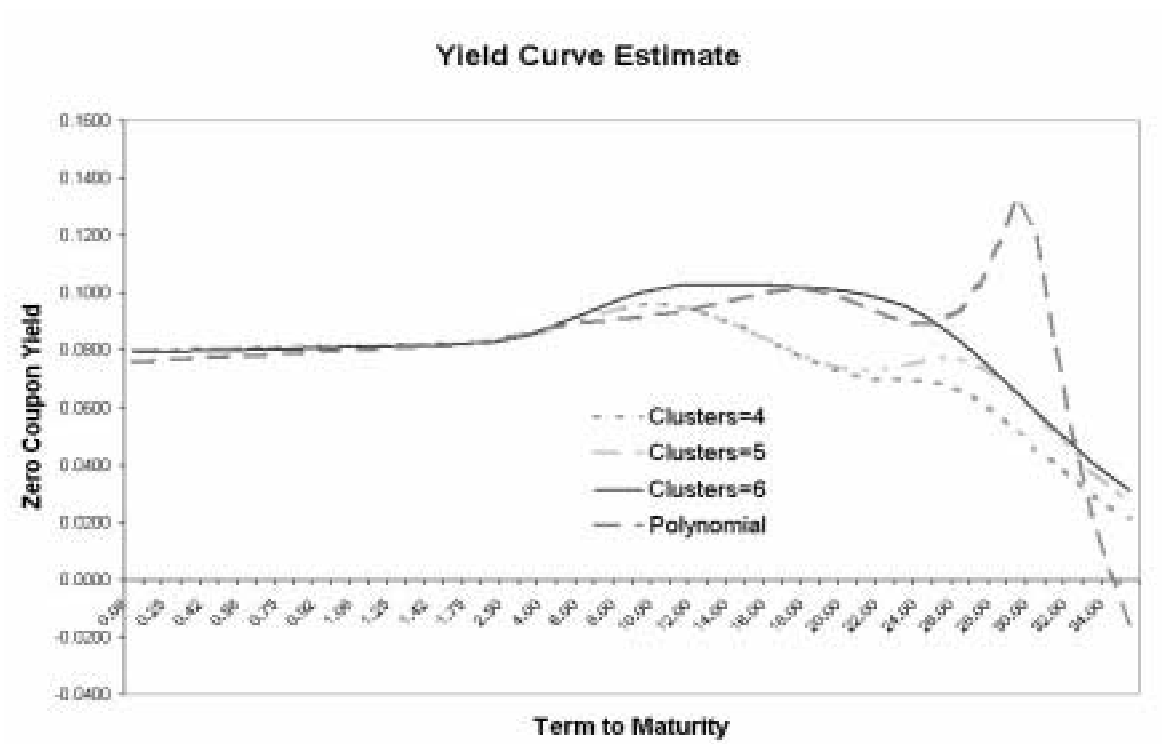


Figure 2

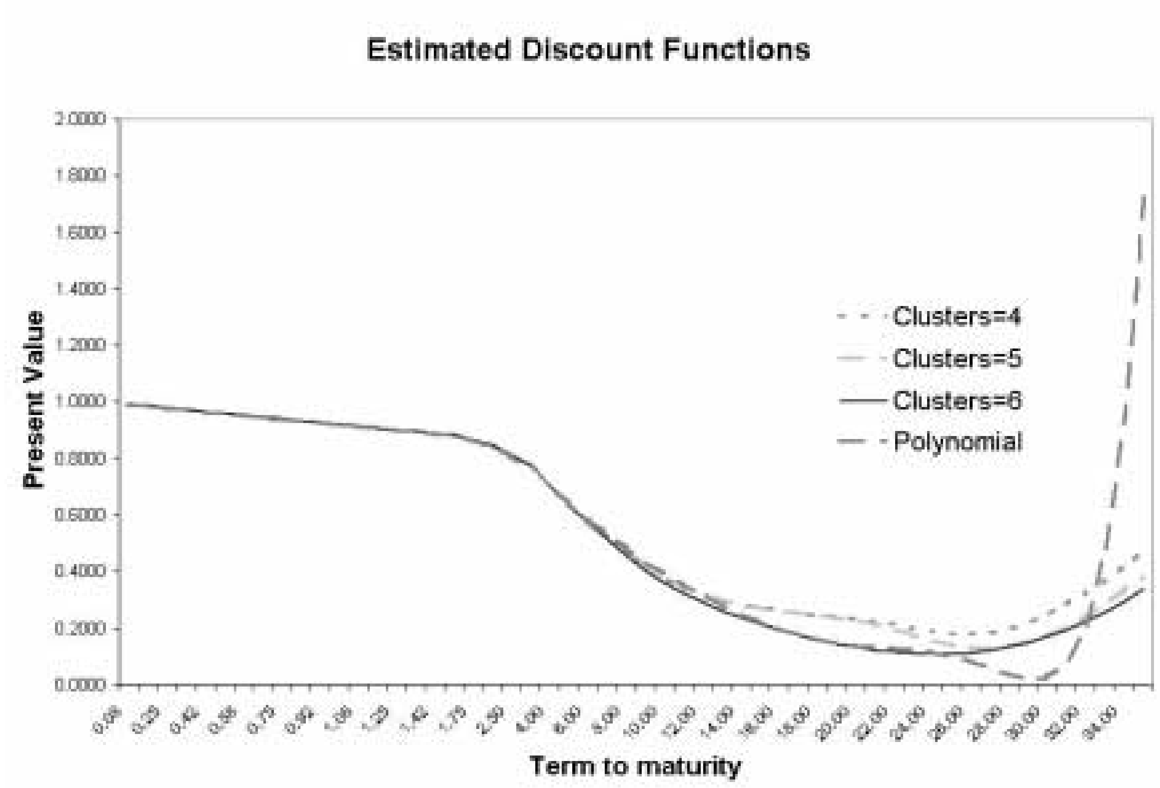
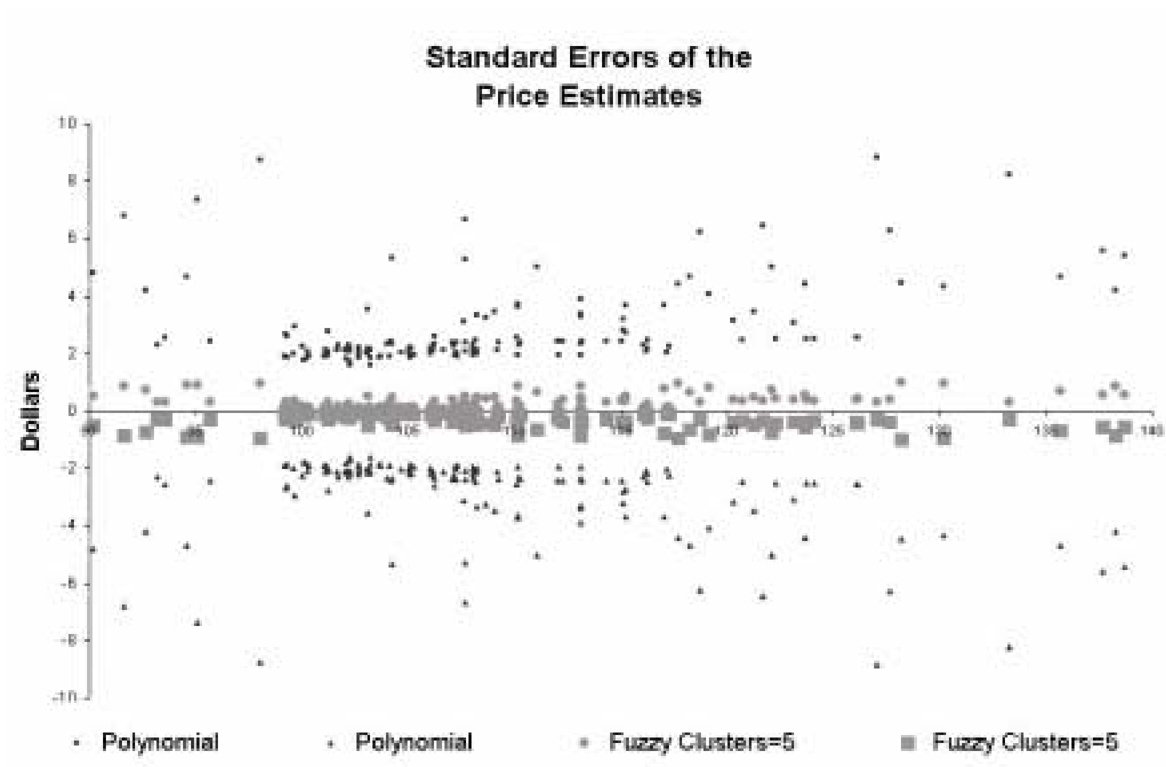


Figure 3



6. Concluding Remarks

There have been many advances in estimating the yield curve since the original work of Durand (1942) and later McCulloch (1971, 1975). This is a fascinating problem in finance and econometrics whose solution is beneficial to both public institutions and private firms. It is clear that the conventional modelling approach, which involves using a simple polynomial approximation at a key stage of the analysis, can place a major degree of constraint on the shape of the yield curve. Given the importance of the associated results, this is undesirable, and in this paper we have also shown that it is unnecessary. The purpose of this paper was to shed light on a relatively new method that could be employed in this application. Our fuzzy clustering/fuzzy regression methodology addresses some of the difficulties inherent in using a polynomial approximation over the full sample when estimating the yield curve. This method is trivially extendable to exponential or even non-parametric representations of the discount function.

As encouraging as our preliminary results are, it must be emphasized that they relate to only one set of data. Our current research extends the analysis in this respect. While it is obvious that there is still a great deal of research to be undertaken on this 'solution' to the problem of estimating the yield curve, recent related research into the use of fuzzy regression analysis has shown that it is a powerful tool that offers very flexible solutions to non-linear modelling problems in a wide range of circumstances. We anticipate that the same will be true, in general, in the case of the yield curve.

References

Bolder, D. and D. Streliski (1999), "Yield Curve Modelling at the Bank of Canada", Bank of Canada Technical Report #84.

Dobson, S. W. (1978), "Estimating Term Structure Equations with Individual Bond Data", *Journal of Finance*, 33, 75-92.

Durand, D. (1942), "Basic Yields of Corporate Bonds, 1900-1942", Technical Paper 3, National Bureau of Economic Research, New York.

Fisher, M., Nychka, D., and Zervos, D. (1995), "Fitting the Term Structure of Interest Rates with Smoothing Splines", Federal Reserve Board, Finance and Economics Discussion Series 95-1.

Giles, D. E. A and R. Draeseke (2003), "Econometric Modelling Based on Pattern Recognition *via* the Fuzzy c-Means Clustering Algorithm", in D. E. A. Giles (ed.), *Computer-Aided Econometrics*, Marcel Dekker, New York, 407-450.

Giles, D. E. A. and C. A. Mosk (2003), "Ruminant Eructation and a Long-Run Environmental Kuznets' Curve for Enteric Methane in New Zealand: Conventional and Fuzzy Regression Analysis", Econometrics Working Paper EWP0306, Department of Economics, University of Victoria.

Giles, D. E. A. and C. Stroomer (2003), "Does Trade Openness Affect the Speed of Income Convergence?", Econometrics Working Paper EWP0307, Department of Economics, University of Victoria.

LeSage, James P. (1999), "Applied Econometrics using MATLAB" Department of Economics University of Toledo 8-22

McCulloch, J. H. (1971), "Measuring the Term Structure of Interest Rates", *Journal of Business*, 44, 19-31

McCulloch, J. H. (1975), "The Tax Adjusted Yield Curve", *Journal of Finance*, 30, 811-830.

Nelson C. R. and A. F. Siegel (1987), "Parsimonious Modeling of Yield Curves", *Journal of Business*, 60, 473-489.

Poirer, D. J. (1976), *The Econometrics of Structural Change, with Special Emphasis on Spline*

Functions, Elsevier, New York.

Ruspini, E. (1970), “Numerical Methods for Fuzzy Clustering”, *Information Science*, 2, 319-350.

Shea, G. S. (1984), “Pitfalls in Smoothing Interest Rate term Structure Data: Equilibrium Models and Spline Approximations”, *Journal of Financial and Quantitative Analysis*, 19, 253-269.

Shea G. S. (1985), “Interest Rate Term Structure Estimation with Exponential Splines: A Note”, *Journal of Business*, 40, 319-325

Shepherd, D. and F. K. C. Shi (1998), “Economic Modelling With Fuzzy Logic”, paper presented at the CEFES ’98 Conference, Cambridge, U.K..

Srivastava, V. K. and D. E. A. Giles (1987), *Seemingly Unrelated Regression Equations Models: Estimation and Inference*, Marcel Dekker, New York.

Waggoner, D. F. (1997), “Spline Methods for Extracting Interest Rate Curves from Coupon Bond Prices”, Federal Reserve Bank of Atlanta, Working Paper 97-10

Zadeh, L. A. (1965), “Fuzzy Sets”, *Information and Control*, 8, 338-353.

Zadeh, L. A. (1987), *Fuzzy Sets and Applications: Selected Papers*, Wiley, New York.