Estimating the Term Structure of Interest Rates using Penalized Splines

Tatyana Krivobokova* Universität Bielefeld

Göran Kauermann Universität Bielefeld

Theofanis Archontakis† Goethe Universität Frankfurt am Main

10th November 2004

We analyse the term structure of interest rates extracted from US Treasury STRIPS data. There is a potential interest from a scientific and economic point of view to look at short and long term bonds simultaneously. In terms of modelling this means to look at smooth functions over time describing the observed term structure. This is the approach pursued in this paper, where penalized spline fitting is employed as smoothing technique. Smoothing is thereby carried out with the respect to both, calendar time and time left to maturity. While the first reveals long term trends, smoothing with respect to the time left to maturity can conceptionally be interpreted as interpolation. Since term structure models have implications for both, the time series and cross-section dimension of yields, estimation techniques involving both dimensions simultaneously are preferred over one-dimensional techniques. Numerical parsimony is applied to fit the large data set and smoothing parameter selection is pursued by building up parallels to linear mixed models.

*Department of Economics, University of Bielefeld, Postfach 100131, 33501 Bielefeld, Germany

†Department of Economics, Goethe-University Frankfurt am Main, Uni-PF77, 60054 Frankfurt am Main, Germany

1 Introduction

Modeling the term structure of interest rates has become an active field of research in finance in the last years. Based on historical developments, a primary area of application is pricing and hedging of different contracts and options written on bonds. Numerous approaches have been proposed concerning the underlying time structure and the stochastic framework of term structure models. Furthermore, different perspectives on term structure modeling have stimulated the development of an enormous variety of models and methods used to study them.

Generally, term structure models can be divided into two main categories: *equilibrium* and *arbitrage-free*[1] models. Within the first category a state variable that determines the term structure is identified and both, the yield curve and the dynamic behaviour of interest rates are determined endogenously. Therefore, one has to estimate or choose parameter values to approximate the average yield curve as well as the short rate. Pioneering models of this kind are Vasicek (1977) and Cox, Ingersoll & Ross (1985). In the second category, the currently observed yield curve is used as an input to model the changes of the term structure over time. The basic model here is proposed in Ho & Lee (1986). Their approach differs from Vasicek in so far as it contains additional time dependent adjustment parameters to calibrate the initial yield curve with the goal to match the observed yield curve exactly (see Backus, Foresi & Zin, 1998).

Despite the widespread use and application of these "theoretical" models, the extraction or estimation of the complete term structure of interest rates from empirically observed bond prices is of less common use. In statistical terms this corresponds to an exploratory analysis of the term structure. In practice it is not possible to obtain the values of the term structure for all horizons since their number exceeds the number of available bonds. To overcome this problem one

[1] Note that both kinds of models are constructed under the assumption of no-arbitrage, therefore the term "arbitrage-free" may be a little misleading.

may use smoothing as an interpolation technique. This is the task of a different stream in the term structure literature and emphasis of the following paper.

In recent work, Ioannides (2003) compares seven estimation methods for the term structure applied to UK data. The methods used in his paper can roughly be categorized as (a) parametric and (b) nonparametric. For the first, a low dimensional basis is used for fitting the term structure to observed data. This approach traces back to McCulloch (1971) and is further explored and discussed for instance in Chambers, Carleton & Waldman (1984) or Nelson & Siegel (1987). The latter paper, in contrast to McCulloch, uses a parsimonious parametric function, with only a small number of unknown parameters, that is flexible enough to represent the shapes generally associated with yield curves (see also Steeley, 1990). In nonparametric estimation, a restrictive parametric term structure modelling is abandoned and replaced by unspecified, unknown functions. The idea is that the functional form should be estimated from the data and not prespecified in advance. This approach was pursued in Fisher, Nychka & Zervos (1995) and is further employed in this paper. Since the term structure has implications both for the cross section and time series dimension of yields, we use a two-dimensional smoothing that leads to a more efficient estimation.

Nonparametric fitting in general has seen a considerable amount of research in the last two decades. Nonetheless, it has been just recently that nonparametric techniques have found their way to term structure modeling. Linton, Mammen, Nielsen & Tanggaard (2000) and Jeffrey, Linton & Nguyen (2001) concentrate on kernel smoothing while Jarrow, Ruppert & Yu (2004) employ penalized spline estimation (P-spline). Comparing the two fitting routines, P-spline smoothing features a considerably reduced numerical effort. This is an important issue, in particular if the number of observations is large, about 125.000 in our application. In P-spline smoothing the unknown term structure is replaced by a high dimensional basis (30-

200 dimensional) which is then fitted in a penalized manner, that is coefficients are shrunk towards zero. This guarantees a smooth fit by prevailing all necessary structure in the function.

The term structure thereby depends on two components, the time left to maturity m and the calendar time t. We take this into account by denoting the term structure function as $g(t, m)$. To explore the term structure at a given time-point, one can fix t at some specific value t_0, say, and fit $g(t_0, m)$ as a function of m only. This is the approach used in the above cited papers. Fixing now the time left to maturity m to m_0, say, the development of $g(t, m_0)$ for a given m_0 is traditionally understood as a stochastic process. This is useful if the focus is on prediction of the yield based on data (and history) available at the current data point. From an exploratory point of view one might however also be interested in describing or visualising the smooth trend in $g(t, m_0)$. This is what could be interpreted as long term development, which is visual from the raw data in crude way only. In previous applied work this approach is mostly done for the short-rate, since it can be seen as an important state variable for the term structure (see e.g. Chan, Karolyi, Longstaff & Sanders, 1992).

Surprisingly, it has been only quite recently that papers in both streams of the term structure literature, i.e. theoretical modeling and smoothing techniques, try to model the complete panel of yield data simultaneously (see e.g Brandt & Yaron, 2003 or Diebold & Li, 2003). In this paper we combine the two approaches by fitting $g(t, m)$ simultaneously as a function of both covariates, time t and time left to maturity m. This means we are using smoothing in two ways. First as interpolation tool for showing $g(t, m)$ for a fixed time as function of m. Secondly with smoothing we visualise long term trends in $g(t, m)$ taken as function of time for fixed m. This allows to explore the term structure and its temporal variation simultaneously.

There are two challenges arising in this modeling exercise. First, one is faced with additional numerical effort, as the dataset has more than 126000 points. Using the link of penalized splines to the linear mixed models and thus fitting our data with standard linear mixed models software makes it possible to overcome this problem. The second challenge occurs since bond prices are correlated over time which has to be taken into account. Ignoring correlation among observations typically leads to serious undersmoothing, that is overfitting, as demonstrated in Opsomer, Wang & Yang (2001). Research on smoothing correlated errors has nearly exclusively discussed univariate or spatial correlation. We here however observe correlation only along yield price development over time of single bonds. To handle this problem we offer a simple procedure, based on accounting for correlation of single yield strips. This means the smoothing parameter is chosen using one-dimensional estimates of correlation structure. The paper is organized as follows. In Section 2 we describe the data at hand. Section 3 presents our estimation routine before providing estimation results in Section 4. A discussion concludes the paper and an appendix provides technical details.

2 Data

Our investigation is based on daily ask quotations of US Treasury STRIPS (Separate Trading of Registered Interest and Principal of Securities). These are securities and synthetic zero-coupon bonds, which are constructed from coupon bearing Treasury bonds and issued by the US Federal Reserve Bank. The sample runs from July 1998 to July 2003 and contains 107 different US Treasury Strip coupon securities with maturities from one month to 30 years, a total of 126251 observations. The data are collected on July 11, 2003 using the Reuters 3000 Xtra information service and include bond prices with maturity dates from August 2003 to May 2033. Figure 1 shows the observed time point and maturity pairs. Table 1 shows some specific properties of the data set. To provide a useful representation, the daily quotes are summarized in classes of different years to maturity. It should be clear that the average yield curve has an increasing, concave shape.

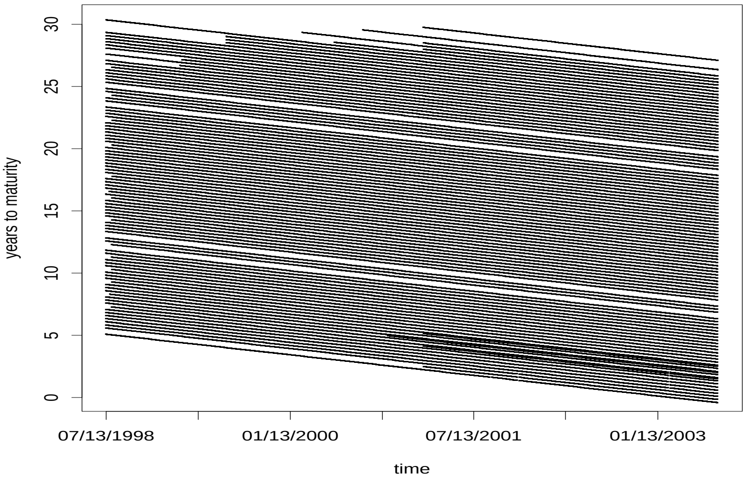

Figure 1: Design of independent covariates time t and maturity m

Maturity	Obs	Mean	St Dev	Autocorr
0.0986-0.25	36	1.0064	0.0365	0.0316
0.25-0.5	95	0.9866	0.1062	0.9202
0.5-1	376	1.1212	0.1419	0.9430
1-3	4255	2.5199	1.1757	0.9878
3-6	13882	4.4737	1.3132	0.9938
6-9	14301	5.2139	0.9423	0.9932
9-12	13457	5.4037	0.7329	0.9907
12-15	14655	5.6697	0.5945	0.9882
15-20	24614	5.8301	0.4677	0.9873
20-25	22776	5.8194	0.3920	0.9873
25-30.3616	17894	5.6992	0.3630	0.9863

Table 1: Properties of US Treasury STRIPS Yields

3 Spline Models for the Term Structure

3.1 Bivariate spline smoothing

We denote with $P_{t,m}$ the price of a zero bond at time point t with m years left to maturity. We consider the continuously compounded yield obtained as

$$y_{t,m} = -\frac{log(P_{t,m})}{m}.$$

Let us first assume that we are interested in modelling the term structure of a zero bond at a particular time t_0, say. This is done by employing a nonparametric function such that

$$y_{t_0,m} = g_{t_0}(m) + \epsilon_{t_0,m}, \tag{1}$$

where $g_{t_0}(m)$ is the term structure as the function of time and $\epsilon_{t_0,m}$ is the residual, for simplicity assumed to be normal and having a constant variance throughout the fit. Variance heterogeneity can be taken into account e.g. by weighting observations, which is not further explored in this paper. For prediction intervals produced later we give up the assumption of variance homogeneity. Estimation of $g_{t_0}(m)$ is done using penalized spline smoothing (P-spline, see Eilers & Marx, 1996). We therefore replace $g_{t_0}(m)$ by the parametric form

$$g_{t_0}(m) = X_m\beta_m + Z_m b_m, \tag{2}$$

where X_m is a low dimensional basis in m (e.g. $X_m = [1, m]$) and Z_m in contrast is a high dimensional basis, linearly independent of X_m. A convenient choice for Z_m are truncated polynomials, e.g. $Z_m = [(m - \mu_1)_+, \ldots, (m - \mu_k)_+]$, where $(.)_+$ gives the positive part, that is $(x)_+$ equals x, if x is positive and zero otherwise. The knots μ_k cover the support of m in some equidistant or appropriately chosen way and the dimension k is chosen generously, i.e between 30 and 100. In principle model (2) is a parametric model, but due to the large dimension of the parameters standard parametric fitting would show unsatisfactory behaviour and estimates would be highly variable leading to a wiggled curve estimate of $g_{t_0}(m)$. Therefore, coefficients b_m need to be penalized in a ridge regression manner. That is the likelihood is maximized subject to the penalty constant $\alpha_m b_m^T D_m b_m$, with α_m as penalty parameter and D_m as some penalty matrix. Using truncated polynomials for Z_m a convenient choice for D_m is the identity matrix (see also Ruppert, Wand & Carroll, 2003 for more justification for this choice).

The P-spline fitting can be easily generalized to accommodate bivariate smoothing as well. We consider now $g_{t_0}(m)$ in (1) as a function of both, time and years left to maturity, so that model (1) generalizes to

$$y_{t,m} = g(t,m) + \epsilon_{t,m}. \tag{3}$$

For estimation we again replace $g(t,m)$ in a high dimensional fashion. A natural extension from one dimensional smoothing towards two dimensional is to use a tensor product for the corresponding one dimensional bases, i.e

$$g(t,m) = (X_t, Z_t) \otimes (X_m, Z_m)\theta.$$

Here X_t is a low dimensional basis in time t and Z_t is the high dimensional supplement like above. For notational simplicity we rearrange the basis matrix to

$$C = [X_t \otimes X_m, X_t \otimes Z_m, Z_t \otimes X_m, Z_t \otimes Z_m],$$

with θ decomposing to (β, b_m, b_t, b_c). The different components in C and θ capture different aspects of the function. Coefficient β is the overall parametric fit, b_m models the dependence on maturity, while b_t mirrors the temporal variation. Finally, b_c captures the interactive influence of t and m. The disadvantage of tensor product matrices is that their dimension increases rapidly. For instance if Z_t and Z_m are 30 dimensional, say, $Z_t \otimes Z_m$ is 900 dimensional, which is at the limit of numerical applicability when it comes to matrix inversion. It is therefore advisable to replace the last component in C by $Z_c = \tilde{Z}_t \otimes \tilde{Z}_m$, where $\tilde{Z}_t$ and $\tilde{Z}_m$ are of lower dimension.

Denoting now with Y the n-dimensional vector of observations $y_{t,m}$, we write the penalized likelihood to be maximized as

$$-\frac{1}{2}(Y - C\theta)^T(Y - C\theta) - \frac{1}{2}\alpha_t b_t^T D_t b_t - \frac{1}{2}\alpha_m b_m^T D_m b_m - \frac{1}{2}\alpha_c b_c^T D_c b_c$$

or, jointly:

$$-\frac{1}{2}(Y - C\theta)^T(Y - C\theta) - \frac{1}{2}\theta^T D(\alpha_t, \alpha_m, \alpha_c)\theta,$$

where $Y = (y_{t,m} : t = 1, \ldots, T; m = 1, \ldots, M)$ and α_t, α_m, α_c are the penalty parameters for the corresponding coefficients.
Penalized Spline Smoothing shows interesting links to linear Mixed Models as generally discussed in Wand (2003). This becomes obvious if the penalty is comprehended as "a priori" distribution for the spline coefficients, that is we assume

$$b_m \sim N(0, \sigma_m^2 D_m^-),\ b_t \sim N(0, \sigma_t^2 D_t^-),\ b_c \sim N(0, \sigma_c^2 D_c^-),$$

with superscript $-$ denotes the generalized inverse here. The smoothing parameter can now be written as variance ratio in the sense $\alpha_m = \sigma_\varepsilon^2/\sigma_m^2$ for instance, and the log-likelihood based on the mixed model results to

$$2l(\beta, \alpha_m, \alpha_t, \alpha_c, \sigma_\varepsilon^2) = -\log(\sigma_\varepsilon^{2n}|V|) - \frac{(Y - X\beta)^T V^{-1}(Y - X\beta)}{\sigma_\varepsilon^2} \quad (4)$$

with $X = X_t \otimes X_m$ and covariance matrix $V = I_n + C\tilde{D}^- C^T$ with $\tilde{D} = \mathrm{diag}(0, \alpha_m D_m, \alpha_t D_t, \alpha_c D_c)$. The fitted values of Y with predicted values for the random effects take now the form $\hat{Y} = C\hat{\theta}$ with $\hat{\theta} = (C^T C + \tilde{D})^{-1} C^T Y$. Apparently, log-likelihood (4) invites for maximization not only with respect to β but also with respect to the smoothing parameters α.

3.2 Spline Smoothing with correlated errors

The penalty parameters $\alpha = (\alpha_t, \alpha_m, \alpha_c)$ steer the amount of penalization and therewith the smoothness of the fit. In fact setting $\alpha \to \infty$ leads to the low dimensional parametric fit $g(t, m) = (X_t \otimes X_m)\beta$, since coefficients b_m, b_t, b_c are penalized to zero. If in contrast $\alpha \to 0$ one ends up with an unpenalized high dimensional parametric fit for $g(t, m) = C\theta$, which is highly variable and hence undesirable. In linear mixed models framework the smoothing parameter α is a ratio of variances, and thus can be estimated using Maximum Likelihood or Restricted Maximum Likelihood (REML) (see Harville, 1977). It is well-known (see Opsomer, Wang & Yang, 2001) that smoothing parameter selection with any data driven method (cross validation, AIC or (RE)ML) fails in case of correlated errors and typically leads to serious undersmoothing, that is overfitting of the data. In the linear mixed models framework, however, once the correlation structure is specified, estimation of regression and correlation parameters can be carried out simultaneously. This routine is implemented in standard linear mixed models software such as the lme() procedure in Splus. We extend this idea in the following way. Let $Y \sim N(C\theta, \sigma_\varepsilon^2 R)$ with correlation matrix R assumed to be known. This yields to $Y^* = R^{-1/2}Y$ as uncorrelated observations and with $C^* = R^{-1/2}C$, one gets that $Y^* \sim N(C^*\theta, \sigma_\epsilon^2 I)$. Hence, knowing the correlation structure we can simplify the estimation to uncorrelated residuals. The idea is now as follows. We will develop a rough estimate for the correlation structure considering data along calendar time only. The estimate is then used to derive Y^*. Even though the estimated correlation might not equal the true correlation exactly, it has been shown in Krivobokova & Kauermann (2004) that the REML estimate still provides reasonable variance estimates even if the correlation is moderately misspecified.

The correlation structure of Y is however not standard. We observe correlation along calendar time t, but for given t it is not reasonable to assume that residuals $\varepsilon_{t,m}$ along maturity m are correlated. Figure 2 shows the yield development of a bond with 6 years left to maturity on the July 1998. The dashed line shows a penalized spline fit if autocorrelation is ignored, the solid line shows the fit if residuals are assumed to have an AR(1) correlation structure. For the latter we plot in Figure 2 the autocorrelation structure of the residuals. The AR(1) assumption seems plausible. Both fits are univariate smoothers and are calculated using the standard linear mixed models

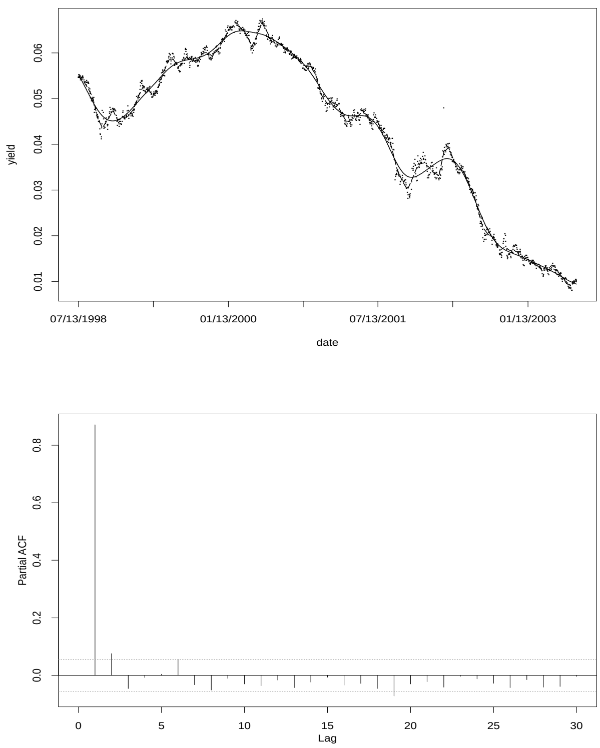

Figure 2: Yield development of a 6 years bond, smoothed with (solid line) and without (dashed line) accounting for correlation, with the corresponding partial autocorrelation function of residuals.

software, as described above without any modification. To account for correlation in the two dimensional fit (3), we now employ the idea of standardising the response variable in the following way. As visible in Figure 1 our data are observed in time series $Y_{t,m_c-(t-t_0)}$ where m_c is the maturity at t_0 in July 1998. All together there are 107 series, one of which is shown in Figure 2. Fitting these series in the same line as that in Figure 2 provides autocorrelation estimates ranging from about 0.8 to 0.9. We therefore use an autocorrelation of 0.85 and let R_{m_c} denote the corresponding correlation matrix. Rearranging Y as $(Y^T_{t,m_1-(t-t_0)}, \ldots, Y^T_{t,m_{107}-(t-t_0)})$ with t taking all observed time points allows to get $Y^* = R^{-1/2}Y$ with R as block diagonal matrix built from the 107 matrices R_{m_c}. Accordingly we get after rearrangement our matrix C^*.

It remains to fit a linear mixed model for independent errors $Y^* \sim N(C^*\theta, \sigma^2_{\epsilon_{t,m}} I_n)$, $b_m \sim N(0, \sigma^2_{b_m} D_m^{-1})$, $b_t \sim N(0, \sigma^2_{b_t} D_t^{-1})$ and $b_c \sim N(0, \sigma^2_{b_c} D_c^{-1})$. Even though the dimension is large, due to independence and the lush but finite dimension of θ linear mixed models software can be applied to obtain the estimates $\hat{\theta} = (\hat{\beta}, \hat{b}_m, \hat{b}_t, \hat{b}_c)^T$ and the resulting fitted response $\hat{Y} = C\hat{\theta}$. This is a numerically handy version to cope with the complex correlation structure in the data. Based on this fit we also analyzed the residuals of the 107 series but did not find obvious violations from the model. There was a slight indication of heteroskedasticity which however was not too evident and for simplicity is ignored.

4 Empirical Results

In Figure 3 we show the bivariate fit of the term structure. The fit is performed using matrix C constructed from truncated squared bases with equidistant knots. The dimensions are thereby chosen as $|Z_t| = 40$, $|Z_m| = 10$ and $|Z_c| = 10 \times 10$. There are two things visible from the plot. First, the functional complexity over time is clearly more exposed than the functional complexity over maturity. Secondly, time and maturity have an interactive effect on the bond price, that is b_c can not be penalized to zero. To better understand the interactive effect we consider a number of plots by slicing the bivariate fit in Figure 3 time-wise and maturity-wise.

Figure 4 shows the estimated term structure for different time points. Beside the fit we have included prediction intervals based on $\pm 2\hat{\sigma}$. Prediction intervals are more useful than confidence intervals in this setting, since the latter are due to the large number of observations so small that they are visually indistinguishable from the fitted curve. From the plots there is a clear dynamics visible over time. The term structure during July 1998 shows as a typical “flat” yield curve, that

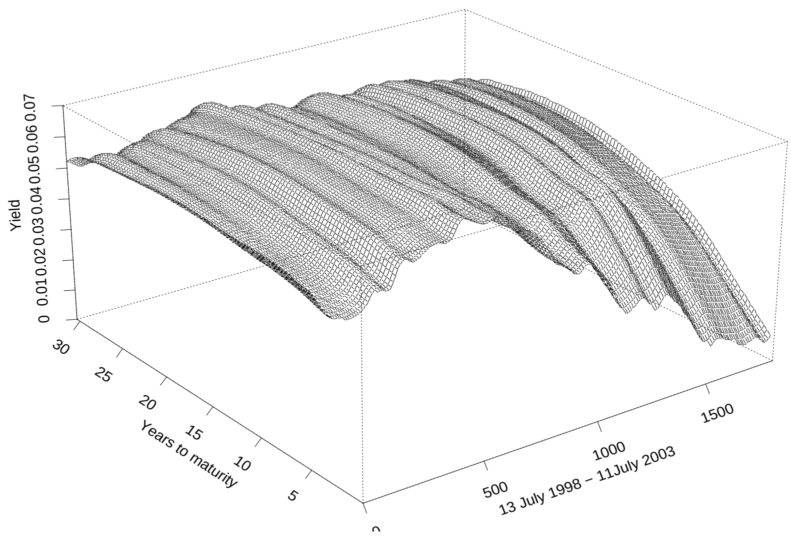

Figure 3: Bivariate fit of the term structure

is the case when the interest rates are about on the same level for different types of bonds. Two years later, in July 2000, after the stock market crash in USA (March 2000) the fitted term structure demonstrates a fully "flat" shape on a high level of about 0.06. Subsequently, the term structure gets curved again, showing a "normal" shape with higher yield for a longer lending time and with more expressed differences between shorter and longer term yields in the most recent years. It is also obvious that long term bonds remain on an interest rate of about 0.06 while yield of short term bonds decreases with time.

Next, we consider the yield development for a given maturity over time. This is shown in Figure 5. The figures visualize the stock market crash in 2000 in that yields for all maturities increase until about spring 2000 and decrease afterwards. In this respect we see that the yield for short maturity bonds is decreasing more rapidly than that for long term bonds.

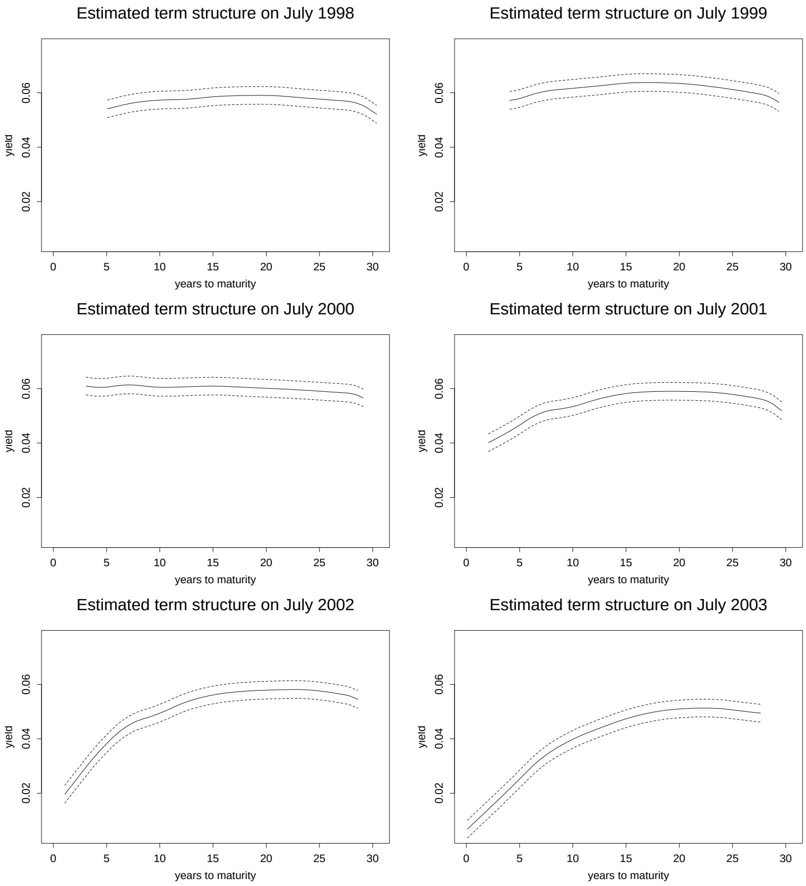

Figure 4: Estimated term structure with prediction intervals

5 Discussion

In the paper we pursued the exercise of fitting zero bonds yield as a bivariate function over time and maturity. We demonstrated the numerical efficiency of penalized spline smoothing as smoothing technique. The modelling exercise allowed to look in the term structure function and to study the dynamic effects. The approach exposed an interesting pattern in the term structure during the stock exchange crash. The empirical and exploratory approach can confirm theoretical investigations and stimulate new insights.

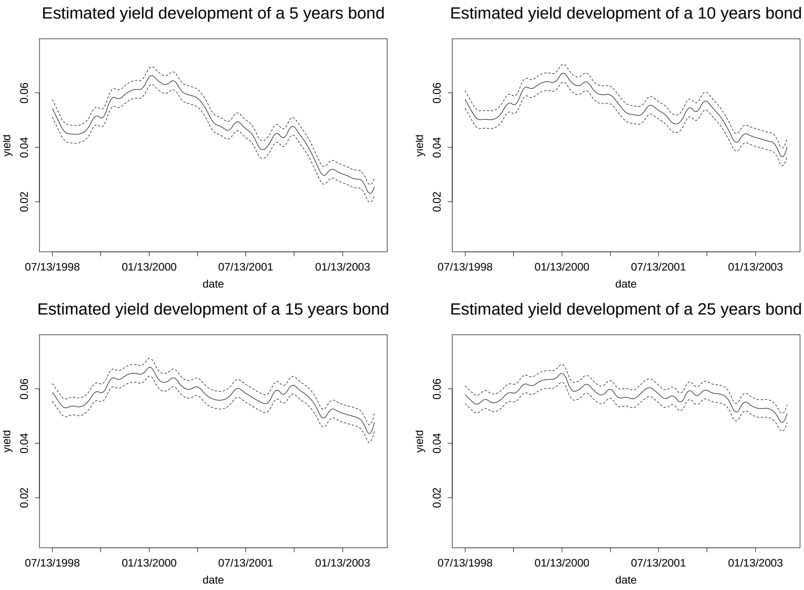

Figure 5: Estimated yield development with prediction intervals

References

Backus, D., Foresi, S., and Zin, S. (1998). Discrete-time models of bond pricing. *NBER Working Paper* **w6736**, www.nber.org/papers/w6736.

Brandt, M. and Yaron, A. (2003). Time-consistent no-arbitrage models of the term structure. *NBER Working Paper* **w9458**, www.nber.org/papers/w9458.

Chambers, D., Carleton, W., and Waldman, D. (1984). A new approach to estimation of the term structure of interest rates. *Journal of Finance and Quantitative Analysis* **19**, 233 – 253.

Chan, K., Karolyi, G. A., Longstaff, F., and Sanders, A. B. (1992). An empirical comparison of alternative models of the short-term interest rate. *The Journal of Finance 47*(3).

Cox, J., Ingersoll, J., and Ross, S. (1985). A theory of the term structure of interest rates. *Econometrica* **53**, 385–408.

Diebold, F. and Li, C. (2003). Forecasting the term structure of government bond yields. *NBER Working Paper* **w10048**, www.nber.org/papers/w10048.

Eilers, P. and Marx, B. (1996). Flexible smoothing with b-splines and penalties. *Stat. Science* **11**, 89 – 121.

Fisher, M., Nychka, D., and Zervos, D. (1995). Fitting the term structure of interest rates with smoothing splines. 1.

Harville, D. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association.* **72**, 320–338.

Ho, T. and Lee, S.-B. (1986). Term structure movements and pricing interest rate contingent claims. *Journal of Finance* **41 (5)**, 1011 – 1029.

Ioannides, M. (2003). A comparison of yield curve estimation techniques using UK data. *Journal of Banking & Finance* **27**, 1 – 26.

Jarrow, R., Ruppert, D., and Yu, Y. (2004). Estimating the term structure of corporate debt with a semiparametric penalized spline model. *Journal of the American Statistical Association.* **99**, 57 – 66.

Jeffrey, A., Linton, O., and Nguyen, T. (2001). Flexible term structure estimation: Which method is preferred. Technical Report, London School of Economics.

Krivobokova, T. and Kauermann, G. (2004). A short note on penalized spline smoothing with correlated errors. Technical Report.

Linton, O., Mammen, M., Nielsen, J., and Tanggaard, C. (2000). Estimating yield curves by kernel smoothing methods. *Journal of Econometrics* **105**, 185 – 223.

McCulloch, J. (1971). Measuring the term structure of interest rates. *Journal of Business* **44**, 19 – 31.

Nelson, C. and Siegel, A. (1987). Parsimonious modelling of yield curves. *Journal of Business 60*(4), 473 – 489.

Opsomer, J., Wang, Y., and Yang, Y. (2001). Nonparametric regression with correlated errors. *Statistical Science* **16**, 134–153.

Ruppert, R., Wand, M., and Carroll, R. (2003). *Semiparametric Regression*. Cambridge University Press.

Steeley, J. (1990). Modelling the dynamics of the term structure interest rates. *The Economic and Social Review 21*(4), 337 – 361.

Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics* **5**, 177–188.

Wand, M. (2003). Smoothing and mixed models. *Computational Statistics* **18**, 223–249.