KYBERNETIKA — VOLUME *33* (1997), NUMBER 6, PAGES 659 – 673

ON ESTIMATING THE YIELD AND VOLATILITY CURVES

Jitka Dupačová, Jozsef Abaffy, Marida Bertocchi and Marie Hušková

Yield curve and yield volatilities are important inputs for pricing interest rate derivatives, for generation of interest rate scenarios, etc. Nonanticipated errors in their estimates may essentially influence the resulting prices, yields and risks. In this paper we explore and compare several types of parametric and nonparametric regression models suitable for estimation of the two curves. In contrast to purely numerical fitting procedures, these methods provide also an information about the precision of the fitted curves and a test of the goodness-of-fit of the postulated parametric model. The parametric models of yield curves are represented by the nonlinear and linearized Bradley–Crane model which is compared with Nadaraya–Watson and Priestley–Chao nonparametric estimators and with cubic splines. The reported numerical experience is based on data from the Italian bond market.

1. INTRODUCTION

Term structure of interest rates provides a characterization of interest rates as a function of maturity. The spot interest rate of a given maturity is defined as the yield on a zero coupon government bond of that maturity. Such bonds are rare in the market and have to be replaced by synthetic zero coupon bonds whose yields correspond to yields of fixed coupon government bonds that do not contain any special provision such as call or put options. The *yield curve* based on these yields is used for analysis of prices and yields of riskier assets, for valuation of interest rate derivatives, it provides a picture about market's current expectations and serves as an input for bond portfolio management models; see for instance [21, 22, 26, 27, 33]. An additional important characteristics whose changes affect the pricing of fixed-income securities and derivatives and the outcome of portfolio management as well is the volatility of interest rates or yields, cf. [24, 25]. Again, it is dependent, inter alia, on maturity; hence, the term *volatility curve*.

Our interest in estimating the yield curve and the volatility curve comes from our involvement in building and solving real life bond portfolio management models, see e. g. [9, 12]. The models are based on interest rate scenarios and the numerical values of the coefficients of the resulting large scale mathematical program result from the choice of the considered bonds, their characteristics (initial prices and future cashflows) and initial holdings, from the scheduled stream of liabilities, transaction costs and spread and from the way how the scenarios of future interest rates are generated and sampled; see [3].

There are various models of evolution of interest rates; we consider interest rate scenarios sampled from the binomial lattice obtained according to Black–Derman–Toy model [4]. It is a one-factor model which assumes a log-normal process of short rates; we refer to [27] for detailed discussions about characteristic properties of this model. To fit the binomial lattice one needs the initial *term structure* which consists of the *yield curve* and the *volatility curve*, i. e., of the yields and standard deviations of logarithms of yields of zero coupon government bonds of all maturities covered by the horizon T of the designed bond portfolio management model; for monthly steps, T is a large number greater or equal maturity (in months) of the longest considered bond. Moreover, we are interested not only in estimates of the coefficients of the chosen parametric form of the yield curve or of the values on the estimated yield curve at the given date but we are also concerned in *precision* of the estimated yield curve and volatility curve at least for selected maturities, as the precision of the estimated initial term structure influences essentially the precision of the fitted binomial lattice and of the scenarios of future short term interest rates. Accordingly, we rely on statistical inference and we propose to use *parametric and nonparametric regression* techniques, see Sections 2 and 3.

The discussion, numerical results and comparisons of the considered techniques reported in Section 4 are related to numerical experience based on real life data from the Italian bond market.

2. THE ESTIMATED YIELD CURVE

The term structure of interest rates consists of yields and log-yield volatilities valid for zero coupon government bonds of *all maturities*. In this section we concentrate on estimation of the yield curve from the existing market data on yields of traded fixed coupon government bonds. Instead of yields one could obviously use the corresponding prices of these bonds as the input, see for instance [2], the discussion in [10] or the approach recommended in Risk Metrics [28]. Regarding the assumption of homoscedasticity commonly present in regression models we decided to use yields; see discussion in [33].

Let the market information at the chosen date consist of the yields y_i, $i = 1, \dots, n$ of various fixed coupon government bonds (without option) characterized by their maturities t_i. The postulated theoretical model

$$y_i = g(t_i; \theta) + e_i, \quad i = 1, \dots, n \tag{1}$$

includes the yield curve $g(t; \theta)$ of a prespecified parametric form where t is usually expressed in years, y is the annualized yield to maturity and $\theta \in \Theta$ is a p-dimensional vector of parameters to be estimated.

Given the market data and the theoretical model of yields, the parameters θ are estimated by the least squares method. It means that the estimate $\hat{\theta}$ of the true parameter vector θ^* is obtained as a solution of

$$\min_{\theta \in \Theta} S(\theta) := \sum_{i=1}^{n} (y_i - g(t_i; \theta))^2. \tag{2}$$

The common assumption is that the residuals e_i in

$$y_i = g(t_i; \theta) + e_i, \quad i = 1, \ldots, n$$

are independent, with zero mean values and an equal unknown variance σ^2 (assumption of homoscedasticity) which is estimated by

$$s^2 = S(\hat{\theta})/(n-p) \approx S(\hat{\theta})/n$$

for large n.

In the case of nonlinear regression one assumes, inter alia, that the function g is twice continuously differentiable with respect to the parameters, that the rank of the (n, p) – matrix $\mathbf{G}(\theta)$ of gradients $\nabla_\theta g(t_i; \theta),\ i = 1, \ldots, n$, equals p for all $\theta \in \Theta$, so called *rank qualification*, that the true parameter value θ^* is an interior point of the convex set Θ and an *identification condition* holds true; see, e. g., [14], [30].

For to get statistical properties of the estimates one relies either on asymptotic results or assumes normal distribution of the residuals and solves a linearized system of normal equations. In the both mentioned cases, the estimates $\hat{\theta}$ are approximately normal, with the mean value equal to θ^* and the covariance matrix $\sigma^2 \Sigma^{-1},\ \Sigma = \mathbf{G}(\hat{\theta})^\top \mathbf{G}(\hat{\theta})$ where σ^2 is estimated by s^2; see, e. g., [30] for details. This allows to construct approximate confidence intervals for components of the true θ^* and an approximate distribution for $g(t; \hat{\theta})$: This distribution is again approximately normal with the mean value $g(t; \theta^*)$ and variance $\sigma^2 Q^2(t)$, where

$$Q^2(t) = \nabla_\theta g(t; \hat{\theta})^\top \Sigma^{-1} \nabla_\theta g(t; \hat{\theta}). \tag{3}$$

The goal is to estimate the yields of zero coupon bonds of all required maturities which are not directly observable. Hence, for each $\tilde{t} \neq t_i$ we replace the unobservable yield by the corresponding value $g(\tilde{t}; \hat{\theta})$ on the estimated yield curve. Such estimates are subject to additional error.

We assume that the yield $\tilde{y} = y(\tilde{t})$ of a zero coupon government bond with maturity $\tilde{t}$ equals

$$\tilde{y} = g(\tilde{t}; \theta^*) + \tilde{e}$$

with $\tilde{e}$ independent of $e_i,\ i = 1, \ldots, n,\ E\tilde{e} = 0,\ \operatorname{var} \tilde{e} = \sigma^2$. Then the mean value of $\tilde{y}$ is approximately $g(\tilde{t}; \hat{\theta})$ and its variance consists of two components

$$\operatorname{var} \tilde{y} = \sigma^2 Q^2(\tilde{t}) + \sigma^2$$

which correspond to the error in the regression model and that of the individual value of the yield for maturity $t = \tilde{t}$.

Under additional assumption of normally distributed errors $\tilde{e}$ and $e_i \,\forall\, i$, $\tilde{y}$ is approximately normal and

$$\tilde{y} - g(\tilde{t}; \hat{\theta}) \sim \mathcal{N}(0, \sigma^2(1 + Q^2(\tilde{t}))) \tag{4}$$

where $Q^2(\tilde{t})$ is computed according to (3). The corresponding approximate $100(1-\alpha)\,\%$ confidence interval for the yield $\tilde{y} = y(\tilde{t})$ for a *fixed* maturity $\tilde{t} \neq t_i,\ i = 1, \ldots, n$ is

$$g(\tilde{t}; \hat{\theta}) \pm s(1 + Q^2(\tilde{t}))^{1/2} t_{n-p}(1 - \alpha/2) \tag{5}$$

and $t_{n-p}(1 - \alpha/2)$ is the corresponding quantile of the t distribution with $n - p$ degrees of freedom.

The assumed parametric form of the yield model (1) allows for a detailed statistical analysis and inference what is welcome for further sensitivity analysis of prices and risks or for the results of the bond portfolio management model. On the other hand, the inference is linked to a specific parametric form of the regression model and it can be misleading if the model is not true (at least approximately). An alternative is to try to fit the yield curve without assuming any specific structure of the data, by *nonparametric regression*, cf. [17, 32]. Two possibilities are given below:

Considering again y_i the yield to maturity of the ith bond and t_i its related maturity, Priestley-Chao estimator is given (see [32]) by the following formula:

$$\hat{g}(t) = \frac{\sum_{i=1}^{n} (t_i - t_{i-1}) K(\frac{t - t_i}{h}) y_i}{h} \tag{6}$$

while Nadaraya–Watson estimator is given (see [6]) by the expression:

$$\hat{g}(t) = \frac{\sum_{i=1}^{n} y_i K(\frac{t - t_i}{h})}{\sum_{i=1}^{n} K(\frac{t - t_i}{h})} \tag{7}$$

where $K(\cdot)$ is a kernel function and h the bandwith parameter which controls the smoothness of the estimator.

Of course, one cannot expect strong results, such as the approximate confidence intervals (5). Various simulation and resampling techniques have been suggested to get approximate confidence intervals, approximate distributions of the test statistics, etc.; see [17, 32] and references therein. However, for nonparametric regression, the number of available observations required for the inference based on asymptotic results is rather large. Besides of designing special simulation and resampling schemes, an open possibility is to use the two-stage approach suggested in [5]. Further development of these techniques will be a subject of our future research.

Still, even graphical comparisons of parametric and nonparametric fit help to support or to reject conjectures about the parametric form of the yield curve and in some cases, it is possible to construct goodness-of-fit tests for parametric regression based on the nonparametric one; see for instance [1, 13, 18, 32].

In the numerical studies presented in this paper we shall mostly restrict ourselves to simple graphical comparisons of cubic splines or kernel estimators and the assumed form of parametric regression and we shall apply the test of the goodness-of-fit of the linearized regression by the method of Eubank and Hart [13].

This method is suitable for fixed design regression models and it requires scaling the points t_i to interval $[0,1]$. Hence, we use $\tau_i = \frac{1}{T} t_i,\ 0 < \tau_1 < \tau_2 \cdots < \tau_n \le 1$. The observed data (τ_i, y_i) are assumed to obey the regression model

$$y_i = g(\tau_i) + e_i, \quad i = 1, \ldots, n$$

where e_i are i.i.d. random variables with zero expectation, a positive variance σ^2 and $Ee_i^4 < +\infty$.

Under null hypothesis, we assume that the function

$$g(\tau) = g(\tau;\theta) = \sum_{j=1}^{p} \theta_j g_j(\tau) \quad \forall\, \tau \in (0,1] \tag{8}$$

where $g_j,\ j = 1, \ldots, p$, are known functions such that the (n,p)-matrix $\mathbf{G} = \{g_j(\tau_i)\}$ is of full column rank, $\theta_j,\ j = 1, \ldots, p$, are unknown constants estimated by least squares estimates $\hat{\theta}_j,\ j = 1, \ldots, p$.

The alternative hypothesis is

$$g(\tau) = \sum_{j=1}^{p} \theta_j g_j(\tau) + f(\tau) \quad \forall\, \tau \in (0,1]$$

where f is a function which cannot be expressed on $(0,1]$ as a linear combination of $g_j,\ j = 1, \ldots, p$. The idea is to replace f by an approximation based on a truncated basis of $0 \le k \le n-p$ functions, say, $u_{jn},\ j = 1, \ldots, k$, which satisfy orthonormality conditions at the design points $\tau_i,\ i = 1, \ldots, n$, and are orthogonal to $\sum_{j=1}^{p} \hat{\theta}_j g_j(\tau)$ at the points $\tau_i \,\forall\, i$. With $a_j, j = 1,\ \ldots, k$, the coefficients of the function f in the chosen truncated basis, the null hypothesis (8) is equivalent to $a_j = 0,\ \forall\, j$.

The test of the null hypothesis (8) requires (recursive) computation of coefficients

$$a_{jn} = \frac{1}{n} \sum_{i=1}^{n} u_{jn}(\tau_i)\, y_i \quad \text{for} \quad k = 1, \ldots, n-p$$

and maximization of the criterion

$$r(k) = \sum_{j=1}^{k} a_{jn}^2 - \frac{c_\alpha \hat{\sigma}^2 k}{n} \tag{9}$$

with respect to $k \in \{1, \ldots, n-p\}$. If the optimal value $\max_k r(k) > 0$, the parametric form of the regression is rejected at the significance level α which is related to the choice of c_α in (9). In principle, any consistent estimate $\hat{\sigma}^2$ of σ^2 can be used in (9).

In contrast to parametric goodness-of-fit tests, this nonparametric test does not assume any prespecified form of the alternative hypothesis but it exploits the homoscedasticity assumption. Heteroscedastic regression models

$$y_i = g(t_i;\theta) + e_i \quad \text{or} \quad y_i = g(t_i) + e_i, \quad i = 1, \ldots, n$$

assume $Ee_i = 0$, $\operatorname{var} e_i = v(t_i)$ and $Ee_i^4 < +\infty \,\forall\, i$. We refer to [29] for a nonparametric method suitable for estimation of the variance function v.

3. THE ESTIMATED VOLATILITY CURVE

The techniques for obtaining volatilities of the yields or log-yields are less obvious. Most of the authors work with an ad hoc fixed constant volatility, say $V(t) = V$; see discussion in [19, 20]. In case of a constant volatility, however, the model does not display the desirable mean reversion, cf. [27].

The volatility curve can be estimated from historical data, see e. g. [23]. Risk Metrics [28] provides *historical volatilities* for 14 major bond markets, including the Italian one; these volatilities are computed daily for several main maturities ranging from 1 year, 2, 3, 4, 5, 7, 9, 10, 15, 20 and 30 years. The proposal is to get the missing yields by linear interpolation and to use the volatilities and correlations of the reported yields to compute the approximate values of yield volatilities for these nonincluded maturities. Clearly, it is not enough data for fitting the volatility curve by a regression model. We plan to explore if the two-stage nonparametric approach by [5] gives reasonable results.

The next source of information is *implied volatilities* computed from quoted bond option prices [25]. At a given day, this provides a set of annualized volatilities related to several different maturities. The next step is to get a volatility curve from these "observed" data. Evidently, the discussion concerning an appropriate parametric or nonparametric estimation procedure appears once more, including the plausible parametric form of the curve and the problem of a small number of available data. A suggestion is to regress the implied bond volatility on the lagged one obtained one period before; cf. [26].

In contrast to the volatility curves obtained independently on the yield curve model one could get *approximate* standard deviations of $\lg y(t)$ from a parametric model of the yield curve provided that the errors in the applied regression model are normally distributed. The first possibility is to construct a regression model directly for log-yields and to use the variance coming from (4), the other possibility is based on a regression model for yields and on theorems about asymptotic distribution of smooth functions of asymptotically normal vectors (cf. [31]). This provides

$$\lg y(\tilde{t}) - \lg g(\tilde{t}; \hat{\theta}) \sim \mathcal{N}(0, \sigma^2 g(\tilde{t}; \hat{\theta})^{-2}(1 + Q^2(\tilde{t})) \tag{10}$$

so that for sufficiently large sample sizes n, the fitted volatility of log-yield for maturity $t = \tilde{t}$ is estimated as

$$V(\tilde{t}) = \frac{s}{g(\tilde{t}; \hat{\theta})} \left(1 + \nabla_\theta g(\tilde{t}; \hat{\theta})^\top \Sigma^{-1} \nabla_\theta g(\tilde{t}; \hat{\theta})\right)^{\frac{1}{2}}. \tag{11}$$

4. NUMERICAL TESTING

Having tried different parametric nonlinear models, as reported in [10], we chose to use a simple form of the yield curve applied already by Bradley and Crane [7]

$$g(t;\theta) = \alpha t^{\beta} e^{\gamma t}. \tag{12}$$

Table 1 reports selected results related to this yield curve model applied for different dates in 1991 − 1996 using prices from the Italian BTP market to estimate the parameters of the yield curve (12); BTPs are government bonds with fixed coupons, without options and with different maturities (3, 5, 10 and 30 years) issued two times per month. The mean values of residuals can be found under heading "means". The results were extended by inclusion of a recent trading day, April 17th, 1997. Notice the change in the form of the yield curve due to a new stance of the monetary policy aimed at reducing interest burden on public debt through a reduction in nominal short-term rates.

The condition number of Σ is of order 2–4, meaning that the matrix is well-conditioned.

Table 1.

Date	n	α	β	γ	means	s^2
Jan 15 '91	16	.117	.066	-.0045	1.e-08	1.e-07
Jun 25 '91	15	.107	.090	-.0130	-2.e-08	3.e-07
Jan 16 '92	22	.108	.032	-.0056	5.e-09	2.e-07
Jun 24 '92	28	.123	-.004	-.0053	4.e-08	2.e-06
Jan 18 '93	33	.116	.023	-.0036	1.e-08	2.e-06
Jun 03 '93	34	.102	.011	.0038	6.e-09	2.e-06
Jan 17 '94	44	.067	.053	.0001	-8.e-08	1.e-06
Jun 13 '94	47	.077	.135	-.0099	1.e-06	2.e-06
Sep 01 '94	49	.097	.137	-.0229	2.e-07	2.e-06
Jan 16 '95	35	.096	.078	-.0050	1.e-06	4.e-06
Jun 26 '95	24	.102	.044	-.0019	-7.e-08	4.e-06
Jan 15 '96	57	.088	-.042	.0105	8.e-07	1.e-05
Jun 24 '96	57	.073	-.027	.0126	4.e-06	2.e-05
Apr 17 '97	60	.057	-.017	.0108	3.e-06	2.e-05

We also applied the *linearized* version of Bradley and Crane's model (12) using logarithms of the already computed yields to maturity as the input and estimating the parameters $\lg \alpha, \beta, \gamma$ by the least squares method. The results are reported in Table 2; the estimated values of α are obtained from estimates of their logarithms.

The obtained estimates of parameters reported in Tables 1 and 2 are comparable and the plots of estimated yields /logarithms of yields versus squares of estimated residuals do not indicate any linear trend in the plot neither for the nonlinear nor for the linearized regression for the dates starting with 1992, i. e., for cases based on a sufficiently large number of observations. Both models seem to repeat the same pattern in the plots and the same outliers can be identified.

Table 2.

Date	n	α	β	γ	R^2	s^2
Jan 15 '91	16	.117	.066	-.0046	.991	8.e-06
Jun 25 '91	15	.107	.090	-.0129	.995	2.e-05
Jan 16 '92	22	.108	.032	-.0056	.850	1.e-05
Jun 24 '92	28	.123	-.004	-.0055	.738	1.e-04
Jan 18 '93	33	.116	.023	-.0037	.303	2.e-04
Jun 03 '93	34	.102	.011	.0040	.589	2.e-04
Jan 17 '94	44	.067	.052	.0003	.827	2.e-04
Jun 13 '94	47	.076	.137	-.0101	.918	3.e-04
Sep 01 '94	49	.097	.137	-.0235	.816	2.e-04
Jan 16 '95	35	.095	.081	-.0053	.837	4.e-04
Jun 26 '95	24	.102	.043	-.0019	.792	4.e-04
Jan 15 '96	57	.088	-.041	.0110	.380	2.e-03
Jun 24 '96	57	.073	-.029	.0144	.369	3.e-03
Apr 17 '97	60	.057	-.017	.0116	.351	4.e-03

Table 3.

Bond	Maturity	Yield	Bond	Maturity	Yield
BTP12669	16/06/1997	0.079485	BTP12673	01/11/1997	0.058892
BTP12675	01/01/1998	0.056453	BTP12677	01/03/2001	0.056524
BTP12678	19/03/1998	0.058943	BTP12679	01/06/2001	0.057963
BTP12681	20/06/1998	0.055527	BTP12683	01/09/2001	0.057953
BTP12684	18/09/1998	0.056223	BTP12687	01/01/2002	0.058764
BTP12688	17/01/1999	0.057112	BTP36605	01/05/2002	0.059394
BTP109236	15/02/2000	0.180284	BTP36607	18/05/1999	0.056257
BTP36613	01/09/1997	0.055174	BTP36614	01/09/2002	0.058663
BTP36622	01/01/1998	0.055694	BTP36623	01/01/2003	0.059265
BTP36631	01/03/1998	0.054822	BTP36632	01/03/2003	0.060516
BTP36635	01/05/1998	0.055086	BTP36641	01/06/1998	0.055920
BTP36642	01/06/2003	0.062019	BTP36650	01/08/1998	0.057311
BTP36651	01/08/2003	0.062621	BTP36659	01/10/1998	0.057647
BTP36660	01/10/2003	0.064148	BTP36665	01/11/2023	0.071364
BTP36675	01/01/1999	0.057960	BTP36676	01/01/2004	0.064629
BTP36683	01/04/1999	0.057671	BTP36684	01/04/2004	0.064649
BTP36691	01/08/1997	0.060402	BTP36692	01/08/1999	0.056666
BTP36693	01/08/2004	0.063835	BTP36708	01/08/1999	0.053531
BTP36709	01/01/2001	0.045493	BTP36715	15/04/1998	0.054643
BTP36716	01/04/2000	0.055658	BTP36717	01/04/2005	0.062987
BTP36727	15/07/1998	0.055686	BTP36728	15/07/2000	0.056920
BTP36731	01/09/2005	0.063115	BTP36740	01/11/1998	0.056147
BTP36741	01/11/2000	0.057696	BTP36747	01/02/1999	0.056928
BTP36748	01/02/2001	0.0595161	BTP36749	01/02/2006	0.064814
BTP36760	15/04/1999	0.057005	BTP36761	01/05/2001	0.059534
BTP36766	01/07/1999	0.057934	BTP36767	01/07/2001	0.060080
BTP36768	01/07/2006	0.065342	BTP36777	01/10/1999	0.057950
BTP36778	15/09/2001	0.060343	BTP36781	01/11/2006	0.065084
BTP36784	01/01/2002	0.060809	BTP36785	01/01/2000	0.058901
BTP108655	01/02/2007	0.066106	BTP108656	01/11/2026	0.068833

Since now, we shall analyze the methods for data of April 17th, 1997; for this date, we were able to collect some of implied volatilities. The government bonds

(with fixed coupons and without options), BTPs, traded on April 17th, 1997, their maturities and yields are listed in Table 3; we have excluded BTP36606 maturing in two weeks horizon (on May 1st, 1997).
The yield curves estimated according to the Bradley–Crane model (12) and according to its linearized version are plotted in Figure 1.

Fig. 1. Term Structure on April 17, 1997; Full Data Set.

Naturally, the fit is sensitive to data; compare Figure 1 with Figure 2 that contains the two fitted yield curves based on 58 observations which remain after deleting the two long bonds, BTP36665 maturing in 2023 and BTP108656 maturing in 2026. The maturities of all the included bonds are less than 10 years.

Fig. 2. Term Structure on April 17, 1997; Long Bonds Omitted.

The plots in Figures 3 and 4 indicate the differences between the fitted parametric form of the yield curve and the nonparametric ones for the full set of 60 data and for the problem with the two long bonds omitted, respectively. In the both cases, the results are plotted only for maturities less than 10 years. The nonparametric regression curves were obtained by means of cubic splines (see [8]) and by Nadaraya–Watson kernel estimator (7). We chose Epaneshnikov kernel (see [32]):

$$K(u) = \begin{cases} \frac{3(1-u^2)}{4}, & |u| \leq 1 \\ 0, & |u| > 1 \end{cases} \tag{13}$$

and the bandwith was selected iteratively by a repeated use of cross-validation (see [32]).

Fig. 3.

Fig. 4.

Cross-validation technique allows to compute the optimal value of the bandwith h by solving the following problem:

$$\min_h \frac{\sum_{i=1}^{n}(y_i - \hat{g}_i(t_i))^2}{n} \tag{14}$$

where $\hat{g}_i(t_i)$ is the kernel estimator computed without the pair (t_i, y_i) and evaluated in t_i.

For Nadaraya–Watson estimator (7), the bandwith values coming from cross-validation were 0.54 After different trials, only one knot at 17.0 and 5.7 was used for the cubic spline.

Priestley–Chao estimator (6) is not appropriate for estimation of the yield curve as the design points t_i are not equidistant. The remaining considered fitting methods are comparable within the range of maturities up to 10 years (see Figures 3 and 4); notice that the different bond yields cover a very small range. For the full data set including maturities up to thirty years, the picture is different, see Figure 5. A similar behaviour has been observed for all the dates we have considered (Figure 5).

Fig. 5.

The test of the parametric form of the regression was done for the linearized Bradley and Crane model (12) by means of the method of Eubank and Hart [13] described briefly in Section 2. We used a polynomial basis, $c_\alpha = 3.22$ (corresponds to the significance level $\alpha = 0.1$) and the variance in (9) was estimated according to [16] as

$$\hat{\sigma}^2 = (n-2)^{-1} \sum_{k=1}^{n-2} \left(\sum_j d_j \lg y_{j+k} \right)^2$$

where $d_0 = 0.8090,\ d_1 = -0.5,\ d_2 = -0.3090$ and $d_j = 0$ otherwise. Its numerical value $\hat{\sigma}^2 = 0.0026$ is comparable with the least squares estimate s^2 from Table 2.

On this significance level, the linearized parametric model (12) of Bradley and Crane was not rejected.

The estimated approximate volatility curves are of a similar character both for the nonlinear and the linearized model, see Figures 6 and 7 related to Figures 1 and 2, respectively. Out of the two models the nonlinear one gives rather low volatilities; these volatilities are comparable with the overall standard deviation of the log-yields which come from the market prices of the traded bonds at the given day. The magnitude of volatilities obtained for the linearized Bradley–Crane model according to (4) is comparable with the magnitude of the implied volatilities. For comparison, we report the annualized historical daily volatilities obtained from Risk Metrics for April 11, 1997. They are relatively high, decreasing from 0.495 for maturity of two years to 0.289 for maturity of 30 years.

Fig. 6. Volatility Structure on April 17, 1997; Full Data Set.

Fig. 7. Volatility Structure on April 17, 1997; Long Bonds Omitted.

5. CONCLUSIONS

Linearized and nonlinear parametric regression models and the nonparametric ones were successfully applied to estimation of the yield curve for the Italian fixed-income market. According to their graphical performance and a formal test, the linearized parametric model (12) gave a good approximation to the data. The assumption of approximately normal distribution of errors in the linearized model is in line with the log-normal distribution of yields postulated in the Black–Derman–Toy model. This in the sequel allows to quantify the precision of the estimated yields and to construct approximate estimates of the log-yield volatilities. General recommendations as to suitable methodology for obtaining adequate volatility curve are not yet available. This will require a further, deep analysis of the existing approaches.

Government fixed-income securities represent more than 85 % of the Italian fixed-income market which is the 4th fixed-income market in the world so that enough data were at disposal. An application of the above estimation methods for thin and emerging markets is limited by the quality and availability of data. For instance, if all Czech government bonds were traded at a given date, it would provide only 10 observations of maturities between 6 months 5 years. The lack of data can be substituted to a certain extent by an indication of a suitable parametric form of the yield curve with a few parameters or of some of its qualitative properties by an expert; our preliminary tests do not support the linearized Bradley and Crane model (12) for the Czech bond market. The two-stage nonparametric regression approach [5] applied to possibly homogeneous data coming from typical trading dates of several subsequent weeks could give some ideas about the evolution and the shape of the yield curve.

Concerning yield volatilities, neither historical volatilities nor implied ones have been reported for the Czech bond market. The only possibility seems to use an ad hoc estimate of a constant volatility and to provide a sensitivity analysis of the subsequent results with respect to this input constant.

ACKNOWLEDGEMENT

This is an extended version of the paper [11] presented at the 20th EWGFM Meeting in Dubrovnik. The research was partially supported by the Grant Agency of the Czech Republic under grants No. 201/96/0230, 201/97/1163 and 402/96/0420, by CNR grant No. 94.00538.ct11, by the 1996 MURST 40 % grant and through the contract "HPC–Finance" (no. 951139) of the INCO'95 project funded by Directorate General III (Industry) of the European Commission. We thank Credito Bergamasco–Crédit Lyonnais for providing implied volatility data.

(Received June 30, 1997.)

REFERENCES

[1] A. Azzalini and A. Bowman: On the use of nonparametric regression for checking linear relationships. J. Roy. Statist. Soc. Ser. B *55* (1993), 549–557.

[2] E. Barone, D. Cuoco and E. Zautzik: Term structure estimation using the Cox, Ingersoll, and Ross model: The case of Italian treasury bonds. J. Fixed Income *1* (1991), 87–95.
[3] M. Bertocchi, J. Dupačová and V. Moriggia: Sensitivity analysis on inputs for a bond portfolio management model. In: Aktuarielle Ansätze für Finanz–Risiken AFIR 1996, Proc. of the VIth AFIR Colloquium, Nuremberg (P. Albrecht, ed.), VVW Karlsruhe 1996, pp. 783–793.
[4] F. Black, E. Derman and W. Toy: A one–factor model of interest rates and its application to treasury bond options. Financial Analysts J. (1990), 33–39.
[5] J. Boularan, L. Ferré and P. Vieu: Growth curves: a two–stage nonparametric approach. J. Statist. Plann. Inference *38* (1994), 327–350.
[6] A. Bowman and S. Young: Graphical comparison of nonparametric curves. Appl. Statist. *45* (1996), 83–98.
[7] S. P. Bradley and D. B. Crane: A dynamic model for bond portfolio management. Management Sci. *19* (1972), 139–151.
[8] P. Dierckx: An algorithm for smoothing, differentiation and integration of experimental data using spline functions. J. Comput. Appl. Math. *1* (1975), 165–184.
[9] J. Dupačová and M. Bertocchi: Management of bond portfolios via stochastic programming – postoptimality and sensitivity analysis. In: System Modelling and Optimization, Proc. of the 17th IFIP TC7 Conference, Prague 1995 (J. Doležal and J. Fidler, eds.), Chapman & Hall, London 1996, pp. 574–582.
[10] J. Dupačová, M. Bertocchi and J. Abaffy: Input Analysis for a Bond Portfolio Management Model. Technical Report 24, University of Bergamo 1996.
[11] J. Dupačová, M. Bertocchi and J. Abaffy: Term structure and volatility curves – A comparison of different approaches. Paper presented at the 20th EWGFM Meeting, Dubrovnik 1997. Technical Report No. 15, University of Bergamo 1997.
[12] J. Dupačová, M. Bertocchi and V. Moriggia: Postoptimality for scenario based financial planning models with an application to bond portfolio management. In: World Wide Asset and Liability Management (W. Ziemba and J. Mulvey, eds.), Cambridge Univ. Press, Cambridge 1997 (to appear).
[13] R. L. Eubank and J. D. Hart: Testing goodness–of–fit in regression via order selection criteria. Ann. Statist. *20* (1992), 1412–1425.
[14] A. R. Gallant: Nonlinear Statistical Models. Wiley, New York 1987.
[15] Guida al mercato obbligazionario Italiano. SIGECO Report 1994.
[16] P. Hall, J. W. Kay and D. M. Titterington: Asymptotically optimal difference based estimation of variance in nonparametric regression. Biometrika *77* (1990), 521–528.
[17] W. Härdle: Smoothing Techniques with Implementation in S. Springer, Berlin 1996.
[18] W. Härdle and E. Mammen: Comparing nonparametric versus parametric regression fits. Ann. Statist. *21* (1993), 1926–1947.
[19] D. Heath et al: Easier done than said. Risk *5* (1992), 9.
[20] J. Hull and A. White: New ways with the yield curve. Risk *3* (1990), 13–15.
[21] J. Hull and W. White: In the common interest. Risk *5* (1992), 3.
[22] R. A. Jarrow: Modelling Fixed Income Securities and Interest Rate Options. Mc Graw Hill, New York 1996.
[23] R. N. Kahn: Fixed income risk modelling. In: The Handbook of Fixed Income Securities, Third edition (F. Fabozzi, ed.), Irwin 1991, pp. 1307–1319.
[24] M. Koenigsberg, J. Showers and J. Streit: The term structure of volatility and bond option valuation. J. Fixed Income *1* (1991), 19–35.
[25] R. C. Kuberek: Predicting interest rate volatility: A conditional heteroscedastic model of interest rate movements. J. Fixed Income *1* (1992), 21–27.
[26] R. Litterman, J. Scheinkman and L. Weiss: Volatility and the yield curve. J. Fixed Income *1* (1991), 49–53.

[27] R. Rebonato: Interest–rate Option Model. Wiley, New York 1996.
[28] Risk Metrics – Technical Document. Third edition. J. P. Morgan, New York 1995.
[29] D. Ruppert, M. Wand, U. Holst and O. Hössjer: Local polynomial variance function estimation. Technometrics (submitted).
[30] G. A. F. Seber and C. J. Wild: Nonlinear Regression. Wiley, New York 1988.
[31] R. J. Serfling, R. J.: Approximation Theorems in Mathematical Statistics. Wiley, New York 1980.
[32] J. S. Simonoff: Smoothing Methods in Statistics. Springer, Berlin 1996.
[33] O. A. Vašíček and H. G. Fong: Term structure modeling using exponential splines. J. Finance *XXXVII* (1982), 339–348.

Prof. RNDr. Jitka Dupačová, DrSc. and Doc. RNDr. Marie Hušková, DrSc., katedra pravděpodobnosti a matematické statistiky, Matematicko–fyzikální fakulta Univerzity Karlovy (Department of Probability and Mathematical Statistics, Faculty of Mathematics and Physics, Charles University), Sokolovská 83, 186 00 Praha 8. Czech Republic.

Prof. Jozsef Abaffy, PhD., Department of Computer Science, University of Economics, Kinizsi u. 1–6, H–1093 Budapest. Hungary.

Prof. Marida Bertocchi, Dr., Department of Mathematics, University of Bergamo, Piazza Rosate 2, I–24129 Bergamo. Italy.