Dynamic polynomial models for the term structure of interest rates [1]

Sonia Petrone, Francesco Corielli

Istituto di Metodi Quantitativi. Università Bocconi, Milano, Italy. (e-mail: {sonia.petrone, francesco.corielli}@uni-bocconi.it).

Abstract. In this paper we discuss a problem quite debated in the literature, namely how to obtain dynamic versions of cross-sectional models for the term structure of interest rates which satisfies some desirable requirements: (a) providing a good reconstruction of market data; (b) having theoretical control of the dynamic no-arbitrage conditions. These requirements are often conflicting: many empirical models in the literature satisfy (a) but not (b), and many no-arbitrage models which control (b) are not satisfactory with respect to (a).

We consider a cross-sectional polynomial model for the logarithm of the discount function, expressed in time-to-maturity, and we study two kinds of extensions: a dynamic, arbitrage-free polynomial model and a dynamic, *quasi-arbitrage-free* polynomial model. We argue that the quasi-arbitrage-free model can give a better reconstruction of the market data (addressing requirement (a)), while allowing to control the approximation error (addressing (b)).

The reasons for this behavior are twofold. One one hand, the model might suffer of a specification error if the true curve is not a polynomial. On the other hand, even if the model were correctly specified, the observed prices might only approximately follow an arbitrage-free evolution, due to faulty observations or imperfection in the markets. Therefore, the (polynomial) statistical model has to be regarded as an approximation of the theoretical, arbitrage-free curve, and the problem becomes to study how the arbitrage-free evolution of the theoretical curve reflects on its polynomial approximation. We are able to give a simple answer to this question by using constructive polynomials, namely Bernstein polynomials, and we show two properties of quasi-no-arbitrage for the proposed model. A discrete-time version of the model can be written in a state-space form, and estimation can be obtained by filtering procedures. The performance of our model is illustrated in a study with real data.

[1] This work was partially supported by grants from Università Bocconi, Ricerca di Base. Francesco Corielli was supported by the Progetto Strategico C.N.R. "Mathematical Modelling of Economic Phenomena"

Key words: Approximation of stochastic processes, Bernstein polynomials, consistency, no-arbitrage, nonparametric regression, state-space models.

JEL Classification: E43, C14

Mathematics Subject Classification (2000): 62P05, 93E11

1 Introduction

In this paper we study a dynamic polynomial model for inference on the term structure of interest rates, with the aim of satisfying the requirements of flexibility, in order to closely reconstruct the daily market data, and simplicity, in order to have control of the dynamic no arbitrage constraints and to allow for efficient statistical estimation. These two requirements are often conflicting and hardly achieved together in the literature. On one hand flexible cross sectional models are proposed for smoothing the zero rates derived from quoted prices and estimating the current discount function (see e.g. Anderson et al, 1996, for a review). These models are static and do not take into account the dynamics of the factors driving the yield curve. Sometimes cross sectional flexibility is achieved at the price of parametric complexity; often the parameters of the models have no natural economic interpretation so that introducing dynamic assumptions on them is not an easy task. Recently, dynamic versions of some popular cross-sectional models have been studied (Diebold and Li (2003)), yet these models do not explicitly take into account the theoretical no-arbitrage conditions. One proposal for a no-arbitrage affine vector model is in Ang and Piazzesi (2003).

On the other hand, in the currently most popular class of dynamic models, the Heath Jarrow and Morton (1992) family, the discount function at time $t = 0$ is taken as an input. However this input is not directly observable as only a finite set of prices are quoted, hence some kind of cross sectional model for interpolating the term structure from observed data is needed. For time $t > 0$, an Heath Jarrow and Morton model fully specifies the shape of the observable term structure; yet, the simplicity requirement greatly restricts the choice of the dynamic parametrization. Consequently, the set of allowed shapes for the term structure is too narrow for reproducing the market data and, in applied work, this drawback leads practitioners to restart afresh their models each day from a new estimate of the term structure. In doing this, practitioners violate their modelling assumptions, using, for daily interpolation of observed prices, cross sectional models which are inconsistent with the dynamic model (as initialized at $t = 0$). Many examples of this behavior can be found in the literature (see e.g. Backus, Foresi and Zin (1998) and Corielli and Petrone (1998)) and a full theoretical analysis was developed in Björk and Christensen (1999) and Filipović (1999, 2001). Dynamic models which produce cross sectional solutions sufficiently flexible to adapt, without inconsistencies, to observed data, seem to require the introduction of a great number of risk factors but this violates the requirement of simplicity and implies relevant statistical problems in the estimation of the model.

A possible way out of the dilemma is to admit that any actually implemented model is only an approximation of the underlying dynamics. Moreover, the underlying market dynamics could in fact be observable only with an error due to frictions such as transaction costs and observation errors (see Gombani and Runggaldier, 2001). Both arguments imply that it is reasonable to allow for some small deviations between the statistical model and the theoretical, arbitrage-free process, while it is quite relevant to conceive the model in such a way that these deviations can be explicitly controlled and, if necessary, reduced by a refining of the approximation.

In this paper we discuss a dynamic polynomial model for the logarithm of the discount function expressed in time to maturity which is both flexible and fairly simple in its parametric form. We show that this model can be made to approximate, with arbitrary precision, any form of the term structures and we show how the approximation error can be controlled.

Our approach underlines a modelling problem whose solution we believe relevant for the practical application of any term structure model. Our basic remark is that a model is just an approximation to the real dynamics of the interest rates (in other words, we explicitly admit a possible specification error). We could impose the constraint of no arbitrage to the dynamics of the approximating model; yet, being the model only an approximation, this would imply arbitrage opportunities in the dynamics of the true model. Alternatively, we could impose to the approximating model the constraints implied by being the model a polynomial approximation of a true arbitrage free process (see Longstaff, 1995, for a discussion of a similar problem in the option valuation context). The two sets of constraints do not coincide so a choice has to be made.

We begin by showing how to build a *consistent* (in the sense of Björk and Christensen, 1999) and *arbitrage free* dynamic polynomial model. A model of this kind has been studied by Cohen (1999), in his unpublished PhD thesis. However, we argue that the model might suffer of a specification error if the true curve is not a polynomial, and, on the other hand, the market prices themselves might only approximately follow a theoretical arbitrage-free process. Consequently, exact no-arbitrage restrictions might be too severe and an exactly arbitrage-free model might fail in closely reproducing the real data.

Consequently, we explore a different approach. Rather than imposing exact no-arbitrage restrictions on the (possibly misspecified) polynomial process, we study how the arbitrage-free evolution of the theoretical curve reflects on its polynomial approximation. Our idea for addressing this question is to use *constructive* polynomials, namely Bernstein polynomials. Bernstein polynomials are constructive since their coefficients are related, in a simple way, to the curve to be approximated. This allows to easily relate the dynamics of the curve and that of its polynomial approximation. We show two properties of quasi-no-arbitrage for the proposed model, in terms of approximation of the stochastic process which describes arbitrage-free bond prices and in terms of option valuation. In particular, we show that, under fairly mild assumptions, the polynomial process converges (in quadratic mean) to the arbitrage-free process, so that, roughly speaking, no-arbitrage holds up to an error term which goes to zero as the order of the polynomial goes to infinity (so that there is no specification error).

A nice feature of our model is that it can be written in a state-space form, and estimation can be obtained by filtering procedures.

The paper is organized as follows. Section 2 contains some preliminaries on the financial processes of interest and on Bernstein polynomials. In section 3 we provide a brief overview of cross-sectional statistical models for estimating the term structure at a given time; then we discuss dynamic extensions of these models, with the restrictions required by consistency and no-arbitrage. We suggest a different approach in section 4. Our model in presented in section 5; we first describe our Bernstein polynomial model for the current discount function, and then its dynamic version. We also study approximation properties. In section 6, the estimation procedure is described. A discretized version of the model can be written in a state-space form, so that estimation can be obtained by maximum likelihood and Kalman filter. Finally, in section 7 we provide an empirical study with the aim of comparing the exactly- versus the approximately-arbitrage-free dynamic models.

2 Preliminaries

2.1 Notation and no-arbitrage conditions

Let $(\Omega, \mathcal{F}, P)$ a complete probability space, $W(t)$ be a d-dimensional Brownian motion defined on $(\Omega, \mathcal{F}, P)$ and $\mathcal{F}_t, t \geq 0$ be the P-augmentation of the natural filtration of $W(t)$. Assume that all processes are $\mathcal{F}_t$-measurable. Let $P(t,T)$ be the price, at date $t \leq T$, of a zero-coupon bond paying one euro at maturity date T and let $f(t,T)$ be the forward rate at date $t \leq T$ for an instantaneous loan at date T. Assuming $f(t,T)$ exists, $P(t,T)$ and $f(t,T)$ are related by

$$P(t,T) = e^{-\int_t^T f(t,s)ds}, \quad f(t,T) = -\frac{\partial lnP(t,T)}{\partial T}.$$

In some circumstances it is easier to model other processes which are in one-to-one correspondence with $P(t,T)$ or $f(t,T)$. In particular, we consider the non-time averaged yield on bond $P(t,T)$ defined by $F(t,T) \equiv -lnP(t,T)$.

Now, $P(t,T), f(t,T)$ and $F(t,T)$ are all *fixed maturity date* processes meaning that they all give information about a bond or loan at a fixed maturity date T. However in dynamic modelling it is more appropriate to work with *fixed time-to-maturity* processes. Therefore we define the fixed time-to-maturity counterparts of the above processes as

$$r(t,x) \equiv f(t,t+x), \quad R(t,x) \equiv F(t,t+x), \quad Q(t,x) \equiv P(t,t+x).$$

Hence, $Q(t,x)$ is the price of a bond at date t paying one euro in x time units from date t, $R(t,x)$ is the non-time averaged yield on $Q(t,x)$ and $r(t,x)$ is the forward rate at date t for an instantaneous loan at x time units from date t.

The functions $r(t,x), R(t,x), Q(t,x)$ are defined for t and x in a bounded closed interval, that we fix as $[0,1]$. For a careful discussion of the domain of the above processes, see Cohen (1999, section 1.4).

The notation $R(t,x)$ underlines the integral/derivative relationship between $r(t,x)$ and $R(t,x)$, namely,

$$R(t,x) = \int_0^x r(t,s)ds, \quad r(t,x) = \frac{\partial R(t,x)}{\partial x}.$$

The spot rate can be defined in terms of any of these processes and we have

$$r_0(t) = \lim_{T \to t} f(t,T) = \lim_{x \to 0} r(t,x)$$

Heath, Jarrow and Morton (1992) give the conditions that the dynamics of the process $P(t,T)$ must satisfy in order to be arbitrage-free. Since there is a one-to-one correspondence between the processes expressed in time to maturity and $P(t,T)$, their evolution is related. We summarize the arbitrage-free restrictions for the processes $P(t,T)$ and $R(t,x)$ (Brace *et al.* (1997); Cohen (1999)). We denote by $(\cdot)'$ the transpose of a vector.

Proposition 1. *Let* $v(t,T) : (\Omega \times \{(t,T)|0 \leq t \leq T\}) \to \Re^d$ *be the* $(d \times 1)$ *vector of the* f*-volatilities. Suppose that*

$$V(t,T) = \int_t^T v(t,s)ds, \quad s(t,x) = v(t,t+x), \quad S(t,x) = V(t,t+x)$$

and for all M

$$E\{\int_0^M \int_0^M \sum_{i=1}^d s_i(t,x)^2 dtdx\} < \infty, \quad E\{\int_0^M \int_0^M \sum_{i=1}^d \frac{\partial s_i(t,x)}{\partial x}^2 dtdx\} < \infty.$$

Then the following conditions are equivalent

1. $P(t,T) > 0, P(t,\cdot)$ *is* $C^2, P(T,T) = 1$ *and* $P(t,T)$ *satisfies*

$$dP(t,T) = (r_0(t) - V(t,T)'\lambda)P(t,T)dt - P(t,T)V(t,T)'dW(t)$$

   for some $(d \times 1)$ *vector* λ.

2. $R(t,\cdot)$ *is* C^2, $R(t,0) = 0$ *and* $R(t,x)$ *satisfies*

$$dR(t,x) = [\frac{1}{2}S(t,x)'S(t,x) + r(t,x) - r_0(t) + S(t,x)'\lambda]dt + S(t,x)'dW(t) \qquad (1)$$

The proposition shows that, in principle, it does not matter which process is modelled. However, in terms of actual model development and implementation some processes might be easier to use. In the following we will work with the process $R(t,x)$, and will assume that it satisfies the following

<u>R-conditions</u>: $R(t,\cdot)$ is a nondecreasing function in C^2, $R(t,0) = 0$, $R(t,1) < \infty$ and the process $\{R(t,\cdot), 0 \leq t \leq 1\}$ satisfies (1).

The R-conditions guarantee that the process $\{R(t,\cdot), 0 \leq t \leq 1\}$ has an arbitrage-free evolution.

2.2 Bernstein polynomials

In this paragraph we review some relevant results about Bernstein polynomials (see Lorentz; 1953), which will be our basic tool in modelling the term structure. These results deal with the polynomial approximation of a given (deterministic) function, but note that in the following section we will be interested in the approximation of a *stochastic process*; for a study of Bernstein polynomials approximation of a random function, see Petrone (1999).

For a function F defined on the closed interval $[0,1]$, the Bernstein polynomial of order k of F is defined by

$$B(x;k,F) = \sum_{j=0}^{k} F(\frac{j}{k})\; q_j(x) \quad ,$$

where $q_j(x) = \binom{k}{j} x^j (1-x)^{k-j}$ are binomial weights; more precisely

$$\begin{cases} q_j(x) = \binom{k}{j} x^j (1-x)^{k-j} & j = 0,\dots,k, x \in (0,1) \\ q_0(0) = 1, \quad q_j(0) = 0 & j \neq 0 \\ q_k(1) = 1, \quad q_j(1) = 0 & j \neq k \end{cases} \tag{2}$$

The following approximation properties are of interest.

Proposition 2. *If the function $F(x)$ is bounded on $[0,1]$, then we have*

$$lim_{k\to\infty} B(x;k,F) = F(x)$$

at each point of continuity x of F; the convergence is uniform on $[0,1]$ if F is continuous on this interval.

The derivative of $B(x;k,F)$, for $0 < x < 1$, is:

$$b(x;k,F) = \frac{d}{dx} B(x;k,F) = \sum_{j=1}^{k} \theta_j \; p_j(x) \tag{3}$$

where $\theta_j = F(\frac{j}{k}) - F(\frac{j-1}{k}),\ j = 1,\dots,k$ and the weights $p_j(x)$ are given by

$$p_j(x) = k \binom{k-1}{j-1} x^{j-1} (1-x)^{k-j} \quad , \; j = 1,\dots,k, \tag{4}$$

with $p_1(0) = k, p_j(0) = 0$ for $j = 1,\dots,k$ and $p_k(1) = k, p_j(1) = 0$ for $j = 1,\dots,k-1$.

The basis functions $p_j(x)$ which form $b(x;k,F)$ are beta densities with parameters $(j, k-j+1)$; the parameter j is related with location and k is a smoothing parameter.

An interesting property is that the r-th derivative $b^{(r)}(\cdot;k,F)$ of the Bernstein polynomial of a function F converges towards the r-th derivative $f^{(r)}$ of F whenever this derivative exists (Lorentz, 1953, p.26).

Proposition 3. *If F is bounded on $[0,1]$ and its r-th derivative $f^{(r)}$ exists at a point $x_0 \in (0,1)$, then $b^{(r)}(x_0;k,F) \to f^{(r)}(x_0)$.*

The following results show the degree of approximation of Bernstein polynomials. For a function F on $[0,1]$, let $\omega_F(\delta)$ denote its modulus of continuity, i.e.

$$\omega_F(\delta) = max_{\{(x,y)\in[0,1]^2:|x-y|<\delta\}}|F(x) - F(y)|.$$

Clearly $\omega_F(\delta)$ decreases to zero if F is continuous.

Proposition 4. *If F is continuous, then*

$$sup_x|F(x) - B(x;k,F)| \leq \frac{5}{4}\,\omega_F(\frac{1}{\sqrt{k}}).$$

A better result holds if F has a continuous derivative f in $[0,1]$ (Lorentz, 1953, p.21).

Proposition 5. *Let $\omega_f(\delta)$ be the modulus of continuity of f; then*

$$sup_x|F(x) - B(x;k,F)| \leq \frac{3}{4}\,\frac{1}{\sqrt{k}}\,\omega_f(\frac{1}{\sqrt{k}}).$$

3 Models for the term structure

3.1 Static models based on basis-functions expansions

Many cross-sectional models proposed in the literature for the term structure at a given time, say $t = 0$, are based on an expansion of the discount function (or of a one-to-one transformation of it) in terms of given basis functions (see e.g. Anderson *et al.*, 1996, chapter 2).

For $i = 1, \ldots, n$, let $P_i(0)$ be the current price of a zero-coupon bond, which has redemption payment C_i at the maturity T_i. Then we have

$$P_i(0) = C_i P(0,T_i)\ , i = 1, \ldots, n. \qquad (5)$$

Since the discount function is infinite-dimensional, it cannot "be observed on the market", but has to be interpolated from the observed prices $P_i(0), i = 1, \ldots, n$. Therefore we need a statistical model for the curve $P(0,T)$. A general approach assumes that it can be defined as a linear combination of a set of basis functions, as

$$P(0,T) = 1 + \sum_{j=1}^{k} a_j\ h_j(T) \qquad (6)$$

where h_j is the j-th basis function and a_j is the corresponding coefficient ($j = 1, \ldots, k$). There are a number of functional forms that the basis functions h_j can take to produce a sensible discount function, such as polynomial splines (McCulloch, 1971; Mastronikola, 1991), exponential splines (Vasicek and Fong, 1982), B-splines (Steeley, 1991), Bernstein polynomials (Schaefer, 1981), kernel functions (Tanggaard, 1997). More parsimonious parametric forms have been proposed by Nelson and Siegel (1987). Modern equilibrium theories suggest an exponential form for the discount function, thus exponential yields models can be preferred (e.g. Langetieg and Smoot, 1989), where

$$P(0,T) = exp\{-T \sum_{j=1}^{k} a_j \; h_j(T)\}. \tag{7}$$

Substituting the expression (6) in (5) gives a regression model

$$P_i(0) = C_i \, (1 + \sum_{j=1}^{k} a_j \; h_j(T_i)) + \epsilon_i \; , i = 1, \ldots, n \; ,$$

whose unknown parameters a_j are estimated from the observed prices, usually by least squares.

3.2 Consistency and no-arbitrage

Static models discussed in the previous section are used for reconstructing the *current* discount function from the observed prices. The common practice is to use the estimated curve as an input in a dynamic model of the Heath, Jarrow and Morton type. The procedure is then repeated over time, i.e. the static model is estimated again the day after and the dynamic model is "recalibrated" with the new estimates, and so on. This procedure does nor guarantee that the sequence of estimates so obtained satisfies the no-arbitrage condition.

More precisely, the family of curves used for modelling the current discount function might be *inconsistent* with the class of curves produced by the dynamic structure (Filipović, 2001). Roughly speaking, this means that on one hand the cross-sectional model assumes that the curve belongs to a family $\mathcal{G}$ (a rich family, in order to have a good fit of the market data) yet, on the other hand, the solution of the dynamic constraints lies in a (simpler) class $\mathcal{G}^*$ and cannot be in $\mathcal{G}$.

One way of proceeding, in order to overcome both the above limitations, is to look for dynamic extensions of cross-sectional models which incorporate the no-arbitrage constraints. One might start with a model of the kind (6) for the discount function, and then introduce dynamic assumptions on the coefficients of the basis expansion, such that the resulting evolution of the discount function is arbitrage-free. This issue is discussed in Cohen (1999), for several choices of the basis functions. Here we focus on the case of a polynomial basis and give an elementary account of Cohen's results. We work with time-to-maturity $x = T - t$ and we model the curve $R(t,x)$.

Let us suppose that

$$R(t,x) = \sum_{j=0}^{k} a_j(t) x^j \; , \tag{8}$$

where, in order to have $R(t,0) = 0$, we must let $a_0(t) = 0$ for any t. Additional constraints must be introduced for $R(t,x)$ to have the correct shape. Then assume that the dynamics of the coefficients is described by

$$da_j(t) = \mu_j(t)dt + v_j(t)' dW(t) \quad , j = 1, \ldots, k, \tag{9}$$

where $v_j(t) = [v_{1,j}(t), \ldots, v_{d,j}(t)]'$ and $W(t)$ is a d-dimensional Brownian motion. From (8) and (9) it follows that

$$dR(t,x) = \sum_{j=1}^{k} da_j(t)\, x^j = \sum_{j=1}^{k} \mu_j(t) x^j dt + S(t,x)' dW(t)$$

where $S(t,x) = [S_1(t,x), \ldots, S_d(t,x)]'$ and

$$S_i(t,x) = \sum_{j=1}^{k} v_{i,j}(t) x^j,\ i = 1, \ldots, d. \tag{10}$$

The no-arbitrage requirement (1) imposes

$$\sum_{j=1}^{k} \mu_j(t) x^j = \frac{1}{2} S(t,x)' S(t,x) + r(t,x) - r_0(t) + S(t,x)'\lambda\ , \tag{11}$$

where

$$r(t,x) = \frac{\partial R(t,x)}{\partial x} = \sum_{j=1}^{k} j a_j(t) x^{j-1}.$$

Now, the left hand side of equation (11) is a polynomial of order k while the right hand side is a polynomial of order $2k$. Therefore, for (11) to hold, we must impose that k is even and that $v_j(t)$ is a vector of zeros for $j = (k/2+1), \ldots, k$. These restrictions are called *balancing conditions* (Cohen, 1999). They imply that (9) has to be restricted to

$$\begin{aligned} da_j(t) = &\quad \mu_j(t)\, dt + v_j(t)'\, dW(t) \quad && j = 1, \ldots, \frac{k}{2} \\ da_j(t) = &\quad \mu_j(t)\, dt && j = \frac{k}{2}+1, \ldots, k. \end{aligned} \tag{12}$$

with the constraint (11) on the drift.

This construction gives an arbitrage-free polynomial model for $R(t,x)$, that can be estimated recursively from the data (once the terms $v_{i,j}(t)$ are specified). However, the system (12) underlines how the balancing conditions required for no-arbitrage strongly restrict the dynamic evolution of the polynomial coefficients of order $j > d$, and consequently reduce the flexibility of the model over time.

We present an empirical study of the performance of the arbitrage-free polynomial model in section 7. The results are fairly good, nevertheless they suggest that exact no-arbitrage restrictions can impoverish the quality of fit. These empirical results might be explained by the fact that, due to market frictions, the price process is only approximately arbitrage-free (see e.g. Longstaff (1995)) so that exact no-arbitrage constraints might be too severe, resulting in a worse fit of the market data. We discuss a different approach to the problem in the following section.

4 A different approach

Let consider the no-arbitrage conditions reviewed in subsection 2.1 and let $\{R(t,\cdot), 0 \leq t \leq 1\}$ be the stochastic process which satisfies (1) (note that the process is infinite-dimensional, in the sense that its realizations, at a given t, are the curves $R(t,\cdot)$, defined on $(0,1)$). A basic point of our approach is the distinction between the process $R(t,\cdot)$ and its polynomial approximation, which we denote by $R_k(t,\cdot)$, i.e. $R_k(t,x) = \sum_{j=1}^{k} a_j(t)\, x^j$. There are two main reasons why a polynomial approximation is of interest, and why the distinction between the two processes $R(t,\cdot)$ and $R_k(t,\cdot)$ is relevant.

In perfect markets, the price process is described by the process $R(t,x)$ which is solution of (1) (once the volatility $S(t,x)$ has been specified). In fact, the volatility specification might be incorrect; but even when the volatility is correctly specified, (1) might be hard to solve, or the analytical form of the solution might be quite complicated, and it might depend on unknown parameters which are consequently difficult to estimate. In this case it seems convenient to consider a simple polynomial approximation $R_k(t,\cdot)$ of the solution of (1) (which we regard as a *statistical model* for $R(t,\cdot)$), which results easier to estimate. Note that an advantage is that the infinite-dimensional process $R(t,\cdot)$ is approximated by a parametric model $R_k(t,\cdot)$, described by the k-dimensional vector of the polynomial coefficients. Of course, if the *true* curve $R(t,\cdot)$ is not a polynomial, the statistical model $R_k(t,\cdot)$ suffers of a specification error, which however can go to zero as the degree k of the polynomial goes to infinity.

On the other hand, in imperfect markets the dynamics of the price process might be only approximately arbitrage-free. From this viewpoint, $R(t,\cdot)$ represents a theoretical curve that the market tends to follow, yet the observed prices are described by a process $R_k(t,\cdot)$ which is close, but not equivalent, to $R(t,\cdot)$.

Both perspectives suggest that the question of interest is no longer what conditions must be imposed on the dynamics of the process $R_k(t,\cdot)$ for it to be arbitrage-free (as discussed in the previous section), but rather how the (arbitrage-free) evolution of the process $R(t,\cdot)$ reflects on its polynomial approximation $R_k(t,\cdot)$. The question is: if the infinite-dimensional process $R(t,x)$ satisfies (1), what is the stochastic differential of its polynomial approximating process $R_k(t,x) = \sum_{j=0}^{k} a_j(t)x^j$?

The answer to this question is not obvious because the polynomial expansion $\sum_{j=0}^{k} a_j(t)x^j$ is not constructive, i.e. the coefficients $a_j(t)$ are not directly related to the curve $R(t,x)$ to be approximated. Therefore, our idea for solving the problem is to rearrange the polynomial's terms in a constructive form, namely in the form of Bernstein polynomials.

Theorem 1. *If the process $\{R(t,\cdot), 0 \leq t \leq T\}$ satisfies (1), then the stochastic differential of its polynomial approximation $R_k(t,x) = \sum_{j=0}^{k} a_j(t)x^j$ is given by*

$$dR_k(t,x) = \mu_k(t,x)dt + \sigma_k(t,x)'dW(t)\ , \tag{13}$$

where

$$\mu_k(t,x) = \sum_{j=1}^{k} \mu(t, \frac{j}{k})\, q_j(x)$$

is the Bernstein polynomial of order k which approximates the drift

$$\mu(t,x) = \frac{1}{2}S(t,x)'S(t,x) + r(t,x) - r_0(t) + S(t,x)'\lambda$$

of (1), the $q_j(x)$*'s being defined by (2), and the i-th component of the vector* $\sigma_k(t,x) = [\sigma_{1,k}(t,x),\dots,\sigma_{d,k}(t,x)]'$ *is the Bernstein polynomial of order k for the i-th component of the volatility in (1), that is*

$$\sigma_{i,k}(t,x) = \sum_{j=1}^{k} S_i(t,j/k)\, q_j(x) \quad , \quad i = 1,\dots,d. \tag{14}$$

Proof. The proof is simple if we constructively relate the polynomial expansion to the function $R(t,x)$. To this aim, let us rearrange the polynomial terms in $R_k(t,x) = \sum_{j=1}^{k} a_j(t)\, x^j$ in the form of Bernstein polynomials (cfr. subsection 2.2):

$$R_k(t,x) = \sum_{j=0}^{k} b_j(t) q_j(x)$$

via the relationship (see Lorentz (1953) p.13)

$$a_j(t) = \binom{k}{j} \sum_{\ell=0}^{j} \gamma_{j,\ell}\, b_\ell(t) \tag{15}$$

where $\gamma_{j,\ell} = (-1)^{j-\ell}\binom{j}{j-\ell}$. In the Bernstein polynomial form, the coefficients $b_j(t)$ are constructively related to $R(t,x)$ by

$$b_j(t) = R(t,\frac{j}{k}) \quad , \quad j = 0,\dots,k\ . \tag{16}$$

Note that $R(t,0) \equiv 0$ so that $b_0(t) \equiv 0$ and $a_0(t) \equiv 0$. Therefore, if $R(t,x)$ satisfies (1), that is

$$dR(t,x) = \mu(t,x)dt + S(t,x)'\, dW(t)$$

where we let $\mu(t,x) = \frac{1}{2}S(t,x)'S(t,x) + r(t,x) - r_0(t) + S(t,x)'\lambda$, then we have that

$$db_j(t) = dR(t,\frac{j}{k}) = \mu(t,\frac{j}{k})dt + S(t,\frac{j}{k})dW(t)\ . \tag{17}$$

Consequently, letting $S_{i,j}(t) = S_i(t,\frac{j}{k}), i = 1,\dots,d; j = 1,\dots,k$, we have

$$\begin{aligned} dR_k(t,x) &= \sum_{j=1}^{k} db_j(t)\ q_j(x) \\ &= \sum_{j=1}^{k} \mu(t,\frac{j}{k})\, q_j(x)dt + \sum_{i=1}^{d}\sum_{j=1}^{k} S_{i,j}(t) q_j(x)\, dW_i(t) \\ &= \mu_k(t,x)dt + \sum_{i=1}^{d} \sigma_{i,k}(t,x)'dW_i(t)\ , \\ &= \mu_k(t,x)dt + \sigma_k(t,x)'dW(t)\ , \end{aligned}$$

as we wanted to show. □

The above result describes the dynamics of the polynomial approximation $R_k(t,x)$ of the arbitrage-free process $R(t,x)$. Clearly, with no further assumptions, the dynamics of $R_k(t,x) = \sum_{j=1}^{k} a_j(t)\, x^j$ is no longer arbitrage-free. In fact, from (15) and (17) we have

$$da_j(t) \quad = \quad \binom{k}{j} \sum_{\ell=1}^{j} \gamma_\ell \, db_\ell(t) = \binom{k}{j} \sum_{\ell=1}^{j} \gamma_\ell \, \mu(t, \frac{\ell}{k})\, dt + \binom{k}{j} \sum_{\ell=1}^{j} \gamma_\ell \, S(t, \frac{\ell}{k})' \, dW(t) \; . \tag{18}$$

Now, the above equation is of the form (9), but clearly the drift does not satisfy the constraint (11). The point is that (17) does not impose the extra *balancing conditions* discussed in subsection 3.2, the only restrictions being those which follow from the assumption (1) on the evolution of $R(t,x)$ and the relationship $b_j(t) = R(t, \frac{j}{k})$.

However, while not arbitrage-free, we can expect the polynomial approximation $R_k(t,x)$ to be close to the arbitrage-free process $R(t,x)$ for k large enough, so that no-arbitrage holds up to an approximation error which goes to zero as k increases to infinity. We provide some result in this sense in the following subsection.

4.1 Approximation properties

As defined in section 2.1, $\{R(t,\cdot) : (0,1) \to \Re, 0 \le t \le 1\}$, and consequently $\{R_k(t,\cdot) : (0,1) \to \Re, 0 \le t \le 1\}$, are $\mathcal{F}_t$-measurable stochastic processes defined on a complete probability space $(\Omega, \mathcal{F}, P)$.

Let us first consider the behavior of $R_k(t,\cdot) = R_k(t,\cdot)(\omega)$ for a fixed t. Assume that, for any ω in a set A with $P(A) = 1$, $R(t,\cdot)$ is continuous and bounded on $(0,1)$ and admits a continuous and bounded derivative $r(t,\cdot)$ with respect to x. Then, from Propositions 2 and 3, for any $\omega \in A$ we have that

$$\lim_{k\to\infty} \sup_{x\in[0,1]} \mid R_k(t,x) - R(t,x) \mid = 0$$

and

$$\lim_{k\to\infty} \sup_{x\in[0,1]} |r_k(t,x) - r(t,x)| = 0 \; .$$

Therefore, for k large enough, $R_k(t,x)$ and $r_k(t,x)$ can be arbitrarily uniformly close to the "true" functions $R(t,x)$ and $r(t,x)$. With further assumptions on the smoothness of $r(t,x)$, we can give a measure of the degree of approximation. Let $\omega_{r_t}(\delta)$ be the modulus of continuity of $r(t,\cdot)$. Then, from proposition 5, we have that

$$\sup_{x\in[0,1]} |R(t,x) - R_k(t,x)| \le \frac{3}{4} \, \frac{1}{\sqrt{k}} \, \omega_{r_t}(\frac{1}{\sqrt{k}}).$$

The above results hold for a fixed t. Now we consider time in the analysis, and assume that the process $R(t,\cdot)$ is the solution of (1). Then the stochastic differential of its polynomial approximation $R_k(t,\cdot)$ is given by (13). We want to show that the sequence of stochastic processes $\{R_k(t,\cdot), k = 1,2,\ldots\}$ converges, in some sense, to the process $R(t,\cdot)$, as k goes to infinity. This problem can be regarded as a problem of stability of the solution of (1) under perturbation of the drift and variance (see e.g. Gikhman and Skorokhod, 1965, ch. 9 and Mahkno, 1991). The peculiar feature of our case is that $R(t,\cdot)$, and the drift and variance, are infinite dimensional, depending on the continuous parameter $x \in [0,1]$.

Theorem 2. *Suppose that the process $R(t,x)$ satisfies the R-conditions of section 2.1. Furthermore, assume that $R(t,x)$ and its derivative $r(t,x) = \frac{\partial R(t,x)}{\partial x}$ are bounded in both their arguments and that the volatilities $S_i(t,x), i = 1,\ldots,d$ are continuous in x for any $t \in [0,1]$, and bounded in (t,x). Then*

$$\lim_{k\to\infty} \sup_{t\in[0,1]} \sup_{x\in[0,1]} E(\{R_k(t,x) - R(t,x)\}^2) = 0.$$

The proof of theorem 2 is provided in the Appendix.

Theorem 2 implies that for any $m \geq 1$ and for any choice of $(x_1,\ldots,x_m,t_1,\ldots,t_m)$,

$$\lim_{k\to\infty} E(\sum_{j=1}^{m}[R_k(t_j,x_j) - R(t_j,x_j)]^2) = 0,$$

and consequently the random vector $[R_k(t_1,x_1),\ldots,R_k(t_m,x_m)]$ converges to $[R(t_1,x_1),\ldots,R(t_m,x_m)]$ in distribution. Therefore the following corollary is proved.

Corollary 1. *For any $m \geq 1$ and for any choice of $(x_1,\ldots,x_m)$, $(t_1,\ldots,t_m)$,*

$$\lim_{k\to\infty} E[g(R_k(t_1,x_1),\ldots,R_k(t_m,x_m))] = E[g(R(t_1,x_1),\ldots,R(t_m,x_m))]\,,$$

for any bounded continuous function $g : \Re^m \to \Re$.

This result is of interest when the model $R_k(t,x)$ is used for option valuation. In fact, the value of an option with terminal pay-off $g(R(t_1,x_1),\ldots,R(t_m,x_m))$ is computed as the expectation of g under the equivalent martingale measure. The above corollary implies that, for k large enough, the polynomial model gives a good approximation of the option value.

5 Bernstein polynomial model

Based on the result of the previous section, we propose a *dynamic Bernstein model* for the term structure of interest rates, which is quite flexible, providing a good fit of the market data, and, at the same time, allows to control the no-arbitrage restrictions.

5.1 Cross-sectional Bernstein model for $R(t,x)$

For given t, the Bernstein polynomial model for ($-\ln$ of) the discount function in time-to-maturity (see section 2) is

$$R_k(t,x) = \sum_{j=0}^{k} b_j(t) q_j(x) \tag{19}$$

This model is a special case of the class of cross-sectional models discussed in subsection 3.1 (equation (7)). However, compared with models based on other expansions, it has the advantage of having a simple constructive structure, with naturally interpretable parameters. Indeed, the parameters of the Bernstein model,

$$b_j(t) = R(t, \frac{j}{k}) \ ,$$

are the $(-\ln)$ prices, at time t, that one would expect in perfect markets for a zero-coupon bond which gives one euro at time-to-maturity $x = j/k$, $j = 1, \ldots, k$ (remember that we are assuming that $x \in (0,1)$).

The shape imposed by the R-conditions (subsection 2.1) on the theoretical curve $R(t, \cdot)$ impose constraints on the coefficients $b_j(t)$. Since $R(t, \cdot)$ is non-decreasing, with $R(t, 0) = 0$ and $R(t, 1) < \infty$, from (16) it follows that

$$0 = b_0(t) \leq b_1(t) \leq \cdots \leq b_k(t). \tag{20}$$

Let us now consider the derivative $r_k(t,x) = \frac{\partial R_k(t,x)}{\partial x}$. From (3) we have

$$r_k(t,x) = \sum_{j=1}^{k} \theta_j(t) p_j(x), \tag{21}$$

where

$$\theta_j(t) = b_j(t) - b_{j-1}(t) \ , \ j = 1, \ldots, k \tag{22}$$

and $p_j(x)$ are the beta basis-functions (4). Note that

$$r_{0,k}(t) = r_k(t, 0) = \lim_{x \to 0} r_k(t,x) = k\theta_1(t) = k(b_1(t) - b_0(t)) = kR(t, 1/k).$$

The constraints (20) induce corresponding restrictions on the θ's parameters

$$\theta_j(t) \geq 0, \quad j = 1, \ldots, k. \tag{23}$$

Since the constraints on the θ's are simpler to handle, for estimation it will be convenient to express the model in the θ parameters (see subsection 6.1).

5.2 Dynamic assumptions

For estimating the Bernstein model (19) over time, we must introduce dynamic assumptions on the parameters $b_j(t)$. A dynamic, arbitrage-free version of the Bernstein model can be obtained along the lines discussed in section 3.2. One can rewrite (19) in powers of x as $R_k(t,x) = \sum_{j=0}^{k} a_j(t) x^j$ via the relationship (15). Then the dynamic assumptions (12) can be imposed to the a_j's parameters (and consequently on the b_j's coefficients). We call model (19) with the dynamics (12) the *arbitrage-free polynomial model*, and we will discuss its performance in the empirical study of section 7.

However, we suggest a different approach. If we regard $R_k(t,x)$ as an approximation of the curve $R(t,x)$ which is solution (1), then the dynamics of $R_k(t,x)$ is not exogenous, but it is induced by that of $R(t,x)$. Indeed, given $R(t,x)$, the stochastic differential of $R_k(t,x)$ is provided by (13) of theorem 1.

In fact, the coefficients $r(t, j/k)$, $j = 1, \ldots, k$ which appear in the drift of (13) are in general not observable. In order to overcome this difficulty, for estimation purposes one can approximate $r(t, x)$ by $r_k(t, x)$, given by (21). In this case, equation (17) in the proof of theorem 1 is replaced by

$$db_j(t) = \mu_j(t)\,dt + S(t, \frac{j}{k})' dW(t) \;, \tag{24}$$

where

$$\mu_j(t) = \frac{1}{2} S(t, \frac{j}{k})' S(t, \frac{j}{k}) + r_k(t, \frac{j}{k}) - r_k(t, 0) + S(t, \frac{j}{k})' \lambda$$

is used in place of the drift $\mu(t, j/k)$. Consequently, the exact stochastic differential (13) is replaced by

$$\begin{aligned} dR_k(t,x) &= \sum_{j=1}^{k} db_j(t)\, q_j(t) \\ &= \sum_{j=1}^{k} \mu_j(t)\, q_j(t)\, dt + \sigma_k(t,x)' dW(t) \\ &= m_k(t,x)\, dt + \sigma_k(t,x)' dW(t) \;. \end{aligned} \tag{25}$$

We call *dynamic Bernstein model* the model (19), with the dynamic assumption (25). (In fact, we should introduce a new notation for the process described by (25), since $R_k(t, x)$ denoted the polynomial approximation of $R(t, x)$ having stochastic differential (13) and not (25); but there should be no confusion in maintaining the same notation for both cases).

There are two consequences of our approach. As underlined in section 4.1, the evolution of $R_k(t, x)$ is not arbitrage-free, even if $R(t, x)$ is so. This remains true if (25) is used in place of (13). However, we showed that the presence of arbitrage is to some extent negligible, since for large k the process $R_k(t, x)$ is close to the arbitrage-free process $R(t, x)$ (theorem 2). It can be easily proved that the result in theorem 2 remains true even if (13) is replaced by (25). In the following subsection we also discuss the amount of arbitrage opportunities in option pricing with our dynamic Bernstein model.

On the other hand, in imperfect markets the balancing conditions imposed by no-arbitrage might be too strong to closely reproduce the observed prices, so that our unrestricted model can give a better performance than its constrained version. This point will be illustrated in the empirical application of section 7.

5.3 Approximate no-arbitrage and option valuation

The aim of this section is to discuss the implications of the lack of exact no arbitrage in the *dynamic Bernstein model* for option valuation.

Let us consider an European option with payoff $g(P(T_1, T_2))$ at time T_1, where $T_2 > T_1$ and $P(T_1, T_2)$ is the price at time T_1 of the zero-coupon bond with maturity T_2 on which the option is written.

Under a no-arbitrage model, the value of the option at time $t < T_1$, paid at time T_1, is given by

$$C(K, T_1, t) = \tilde{E}^t\{g(P(T_1, T_2))\} \qquad (26)$$

where $\tilde{E}^t(\cdot)$ denotes the expectation at time t with respect to the distribution of $(\cdot)$ under the forward risk neutral measure, which is the measure such that the forward price $P(t, T_1, T_2) = P(t, T_2)/P(t, T_1)$ is a martingale.

Now, the unconstrained dynamics of $P(t, T)$ is

$$dP(t, T)/P(t, T) = \alpha(t, T)dt - V(t, T)'dW(t) \;, \qquad (27)$$

where no-arbitrage imposes (see Proposition 1)

$$\alpha(t, T) = r_0(t) - V(t, T)'\,\lambda \;. \qquad (28)$$

The forward price dynamics can be derived from (27), as

$$dP(t, T_1, T_2)/P(t, T_1, T_2) =$$

$$= \{[\alpha(t, T_2) - \alpha(t, T_1)] + [V(t, T_2) - V(t, T_1)]'V(t, T_1)\}dt - [V(t, T_2) - V(t, T_1)]'dW(t) \;.$$

Under no-arbitrage,

$$\alpha(t, T_2) - \alpha(t, T_1) = -[V(t, T_2) - V(t, T_1)]'\lambda \,,$$

and the dynamics under the forward risk neutral measure is obtained by setting

$$\lambda = V(t, T_1) \;.$$

Since in (26), $P(T_1, T_2) = P(t, T_1, T_2)$ for $t = T_1$, the above formulas imply that option valuation only requires the knowledge of $V(t, T)$. In particular, if $V(t, T)$ is a deterministic function, the forward price process $P(t, T_1, T_2)$ is a log-normal martingale under the forward risk neutral measure, and, given $V(t, T)$, the call option price can be computed by the celebrated Black formula (see e.g. Musiela and Rutkowski, 1998, ch.15).

However the dynamics of the approximation $P_k(t, T) = \exp(-R_k(t, T - t))$ of $P(t, T)$ is only approximately arbitrage-free, even if the process $P(t, T)$ is arbitrage-free. This is a consequence of the fact that, as discussed in the previous sections, the dynamics (25) of the process $R_k(t, x)$ is not arbitrage-free. Note that the drift in (25) is

$$\frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{k} S_{i,j}(t)^2 q_j(x) + \sum_{j=1}^{k} r_k(t, \frac{j}{k})q_j(x) - r_k(t, 0) + \sigma_k(t, x)'\lambda \;.$$

Comparing (25) with condition (1) of Proposition 1, we see that, in order to have no-arbitrage, the drift of $R_k(t, T)$ should include the terms

- $\sigma_i(t, x)^2 = (\sum_{j=1}^{k} S_{i,j}(t)q_j(x))^2$ in place of $\sum_{j=1}^{k} S_{i,j}(t)^2 q_j(x)$ (note that the first is the square of the Bernstein polynomial approximation of the variance component $S_i(t, x)$ for $R(t, x)$, while the latter is the Bernstein approximation of $S_i(t, x)^2$);

- $r_k(t,x)$ in place of its Bernstein polynomial approximation $\sum_{j=1}^{k} r_k(t;j/k)q_j(x)$.

Therefore we can write

$$dR_k(t,x) = (\tilde{\mu}_k(t,x) + \epsilon_k(t,x))dt + \sigma_k(t,x)dW(t) \tag{29}$$

where $\tilde{\mu}_k(t,x)$ is the drift that $dR_k(t,x)$ should have in order to be arbitrage-free, and

$$\epsilon_k(t,x) = \sum_{i=1}^{d}(\sum_{j=1}^{k} S_{i,j}(t)^2 q_j(x) - \sigma_{i,k}(t,x)^2) + \sum_{j=1}^{k} r_k(t,\frac{j}{k})q_j(x) - r_k(t,x)\ .$$

Were $\epsilon_k(t,x) = 0$, the evolution of $P_k(t,x)$ would be as described in Proposition 1, that is

$$\frac{dP_k(t,T)}{P_k(t,T)} = \{r_{0,k}(t) - V_k(t,T)'\lambda\}dt - V_k(t,T)'dW(t)\ ,$$

where $V_k(t,T) = \sigma_k(t,T-t)$ and $r_{0,k}(t) = \lim_{x\to 0} r_k(t,x)$.

Because of the ϵ term, the dynamics implied by (29) for the process $P_k(t,T)$ is

$$\frac{dP_k(t,T)}{P_k(t,T)} = \{r_{0,k}(t) - V_k(t,T)'\lambda - \epsilon_k(t,T)\}dt - V_k(t,T)'dW(t)\ ,$$

which does not satisfy the no-arbitrage restriction (28). As a consequence, the forward price process $P_k(t,T_1,T_2) = P_k(t,T_2)/P_k(t,T_1)$ will also be only approximately arbitrage-free. We have

$$\begin{aligned} dP_k(t,T_1,T_2)/P_k(t,T_1,T_2) &= \{-(V_k(t,T_2) - V_k(t,T_1))'\,(\lambda - V_k(t,T_1)) \\ &+\epsilon_k(t,T_1) - \epsilon_k(t,T_2)\}dt - (V_k(t,T_2) - V_k(t,T_1))'dW(t)\ . \end{aligned} \tag{30}$$

Due to the presence of the ϵ_k terms, under the condition $\lambda = V_k(t,T_1)$ the drift is nonzero and the process is not a martingale. (It would be a martingale if $\epsilon_k(t,T_1) = \epsilon_k(t,T_2) = 0$. In principle, one might set the drift equal to zero, but then the model would be inconsistent, in the sense that the solution of (29) would be no longer a polynomial. A related issue is discussed in Gombani and Runggaldier (2001)).

The option value can be computed (possibly via a Montecarlo approximation) as the expectation

$$C(K,T_1,t) = \tilde{E}^t\{g(P_k(T_1,T_2))\} \tag{31}$$

with respect to the measure such that $\lambda = V_k(t,T_1)$, yet this measure is not the forward risk neutral measure.

Clearly, ϵ_k goes to zero as $k\to\infty$. We study empirically the size of ϵ_k in section 7.2, where we compare option values computed by the Black formula with a Montecarlo approximation of the expected value (31). We will show that the arbitrage opportunity due to the presence of the ϵ_k term is negligible, in the sense that the differences among the two prices are smaller than the transaction costs.

6 Estimation

We discuss now a discrete-time version of our model, which can be written in a state-space form, so that estimation can be obtained by filtering techniques.

6.1 Cross-sectional estimation

Let us first consider estimation of our model at a given time t. The data are the prices $(P_1(t), \ldots, P_n(t))$ of n zero-coupon bonds, where the i-th bond has redemptions payment C_i at the time-to-maturity x_i, $i = 1, \ldots, n$. In principle, we have

$$P_i(t) = C_i\, Q(t, x_i) \quad , i = 1, \ldots, n,$$

which can be written as

$$Y_i(t) = -\ln Q(t, x_i) = R(t, x_i),$$

letting $Y_i(t) = -\ln P_i(t) + \ln C_i$. We model the non-observable curve $R(t, \cdot)$ by $R_k(t, \cdot)$ and taking also into account observational errors the model to be estimated results

$$Y_i(t) = R_k(t, x_i) + \epsilon_i(t) = \sum_{j=0}^{k} b_j(t) q_j(x_i) + \epsilon_i(t) \quad , i = 1, \ldots, n,$$

(with the constraints (20) on the parameters). It is convenient to express the model in the parameters $\theta_j(t) = b_j(t) - b_{j-1}(t)$. Simple algebra gives

$$Y_i(t) = \sum_{j=1}^{k} \theta_j(t) z_{i,j} + \epsilon_i(t) \quad , i = 1, \ldots, n, \tag{32}$$

where $z_{i,j} = \sum_{\ell=j}^{k} q_\ell(x_i)$ and the parameters must satisfy the constraints (23). In matrix notation, the observation equation (32) can be written as

$$Y(t) = Z\,\theta(t) + \epsilon(t) \tag{33}$$

where $Y(t) = (Y_1(t), \ldots, Y_n(t))'$, $\theta(t) = (\theta_1(t), \ldots, \theta_k(t))'$, $\epsilon(t) = (\epsilon_1(t), \ldots, \epsilon_n(t))'$ and Z is a $(n \times k)$ matrix with elements $z_{i,j}$. The parameters of this regression model can be estimated by constrained least squares.

6.2 Dynamic estimation

We now introduce time in the analysis, completing the model with a state equation describing the temporal evolution of the state parameters $\theta_j(t)$.

In continuous time, the states' dynamics can be derived by (24). For $j = 1, \ldots, k$ we have

$$\begin{aligned} d\theta_j(t) &= db_j(t) - db_{j-1}(t) \\ &= [\frac{1}{2} s_j(t)^2 + r_k(t; \frac{j}{k}) - r_k(t; \frac{j-1}{k}) + s_j(t)'\lambda]dt + s_j(t)'dW(t), \end{aligned}$$

where $r_k(t,x)$ is given by (21), $r_k(t,0) = kb_1(t)$,

$$s_j(t) = S(t, \tfrac{j}{k}) - S(t, \tfrac{j-1}{k}) \,,$$
$$s_j(t)^2 = S(t, \tfrac{j}{k})' S(t, \tfrac{j}{k}) - S(t, \tfrac{j-1}{k})' S(t, \tfrac{j-1}{k}) \,,\ j = 1, \ldots, k$$

and $S(t,0)$ is a vector of zeros. For estimation, we consider the simple Eulero-Mayurama discrete-time approximation of the process $\theta(t)$ (see e.g. Kloeden and Platen, 1980). Assuming that observations are taken at time points $(t_0 = 0, t_1, \ldots, t_N)$, equally spaced by δ, the approximating process $\theta_j^*(t)$, $j = 1, \ldots, k$, is described by

$$\begin{aligned}
& \theta_j^*(t_{\gamma+1}) \\
= \quad & \theta_j^*(t_\gamma) + [\frac{1}{2} s_j^2(t_\gamma) + r_k(t_\gamma; \frac{j}{k}) - r_k(t_\gamma; \frac{j-1}{k}) + s_j(t_\gamma)' \lambda] \, \delta + s_j(t_\gamma)' W^*(t_\gamma) \\
= \quad & [\frac{1}{2} s_j^2(t_\gamma) + s_j(t_\gamma)' \lambda + \sum_{\ell=1}^{k} \theta_\ell^*(t_\gamma) \pi_{j,\ell} \,] \, \delta + s_j(t_\gamma)' W^*(t_\gamma) \,, \qquad (34)
\end{aligned}$$

where

$$\pi_{j,\ell} = \begin{cases} (p_\ell(\frac{j}{k}) - p_\ell(\frac{j-1}{k})) & \ell \neq j \\ (p_\ell(\frac{j}{k}) - p_\ell(\frac{j-1}{k})) + \frac{1}{\delta} & \ell = j \,, \end{cases}$$

$$W^*(t_\gamma) = [W_1^*(t_\gamma), \ldots, W_d^*(t_\gamma)]',$$

and

$$W_\ell^*(t_\gamma) = W_\ell(t_{\gamma+1}) - W_\ell(t_\gamma) \ \ , \ \ \ell = 1, \ldots, d$$

are independent random variables, with $W_\ell^*(t_\gamma) \sim N(0, \delta)$.

Let $\mu^*(t)$ be a $(k \times 1)$ vector with elements $(\frac{1}{2} s_j^2(t) + s_j(t)' \lambda), j = 1, \ldots, k$; F a $(k \times k)$ matrix with (i,j)-th element $\pi_{i,j}$, and $\Sigma(t)$ be a $(k \times d)$ matrix with j-th row $s_j(t)'$. Then the *state equation* (34) can be written in matrix notation as

$$\theta^*(t_{\gamma+1}) = [\mu^*(t_\gamma) + F \, \theta^*(t_\gamma)] \, \delta + \Sigma(t_\gamma) W^*(t_\gamma) \qquad (35)$$

with $\gamma = 0, 1, \ldots, N$, and $W^*(t) \sim N_d(\mathbf{0}, \delta I_d)$.

At the times $(t_\gamma, \gamma = 0, \ldots, N)$ one observes the (transformed) prices $Y(t_\gamma) = (Y_1(t_\gamma), \ldots, Y_n(t_\gamma))'$, which, from (33), are described by the *observation equation*

$$Y(t_\gamma) = Z \theta^*(t_\gamma) + \epsilon(t_\gamma) \,, \qquad (36)$$

where we assume that

$$\epsilon(t_\gamma) = (\epsilon_1(t_\gamma), \ldots, \epsilon_n(t_\gamma))' \sim N_n(\mathbf{0}, \, u(t_\gamma) I_n)$$

and the $\epsilon(t_\gamma)$'s are serially non correlated, and uncorrelated with the $W^*(t_\gamma)$.

We have a *state-space model* described by the state equation (35) and the observation equations (36). The matrices F and Z are known, while $\mu^*(t)$ and the covariance matrices depend on the unknown *dynamic parameters* $\{\lambda; u(t); S(t, \frac{j}{k}), j = 1, \ldots, k\}$. Given the dynamic parameters, the conditional distribution of $\theta^*(t)$ given the observation up to time t can be updated by the Kalman filter. For simplicity, in the next section we will assume that the parameters $\{\lambda; u(t); S(t, \frac{j}{k}), j = 1, \ldots, k\}$ do not vary over time and depend on an unknown vector of constants ϕ; consequently, ϕ will be estimated by maximum-likelihood and the Kalman filter will then be used for updating the distribution of θ^* given the data and the maximum likelihood estimates $\hat{\phi}$.

7 Empirical study.

The aim of this section is to compare, on a real data set, the performance of the *arbitrage-free polynomial model* and of the *dynamic Bernstein model* of the same degree. As discussed in subsection 5.2, the arbitrage-free polynomial model assumes that the coefficients evolve according to (11), with the balancing restrictions required for no-arbitrage, while the dynamic Bernstein model is defined as an approximation of the theoretical, arbitrage-free price process which the market only tend to follow. As argued before, the latter model might give a better reconstruction of the observed prices.

The data we consider are daily data for zero-coupon interest rates from 1 to 10 years to the expiry. Precisely, the series are derived by money market and swap rates quoted on the eurolira market (data source: Datastream). Swap rates are "stripped" (Miron and Swanell (1992)) in order to derive zero coupon rates $(I_1(t), \ldots, I_{10}(t), t = 1, \ldots, 905)$, and then prices are computed as $P_i(t) = \exp(-iI_i(t)), i = 1, \ldots, 10$. We normalized time to maturity into $[0, 1]$ by dividing the actual times to maturity by 10.

The time period is from July 3rd 1995 to January 2nd 1999 (the birth of the euro). During the 1995 to 1999 period the interest rates for Italy declined while convergence toward the German rates was taking place with a view to the birth of the euro. Figure 1 shows the observed values of the $-\ln$ prices $(Y_1(t), \ldots, Y_{10}(t), t = 1, \ldots, 905)$, and gives an idea of the evolution of the curve $R(t, \cdot)$.

The data appear fairly stable over time, so that we can expect a small interpolation error for both models; yet, since the balancing restrictions introduce more rigidity, we will see that the dynamic Bernstein model gives a better fit of the data in the central part of the time interval, where the curve is more unstable.

7.1 Arbitrage-free versus quasi-arbitrage-free dynamic models

The discrete time version of the *dynamic Bernstein model* is described by the state-space model discussed in section 6.2, with state equation (35) and observation equation (36). We consider a two-factor model ($d = 2$) and we fix $k = 2d = 4$. Alternatively, the degree k of the polynomial might be chosen by model selection criteria (such as AIC). Note however that, as discussed in section 3.2, no-arbitrage imposes that k is even.

For simplicity, the variance of the observational error in (36) is supposed to be a known constant u. It is reasonable to assume that interest rate data are measured with very high precision, so that u is small; sensible values for u can be in the range of 0.001 to 0.00001. We considered several values in this range, obtaining very similar results; here we report the results with $u = 0.0001$.

A more critical step is to specify the volatility components $S_{i,j}(t)$ in the state equation. Note that we are assuming that the stochastic differential $dR_k(t,x)$ is given by equation (25), where the volatility components $\sigma_{i,k}(t,x)$ are given by (14) or, equivalently, expressed as powers of x,

$$\sigma_{i,k}(t,x) = \sum_{j=1}^{k} v_{i,j}(t) x^j \ , \tag{37}$$

the $v_{i,j}$ and $S_{i,j}$'s coefficients being related by (15). This polynomial specification is in principle quite flexible, if the coefficients $v_{i,j}$'s are free to vary over time. Yet, for simplicity, here we suppose that the volatility is constant over time, while it depends on time-to-maturity. The assumption of a non random volatility is clearly restrictive and it could be removed, however it suffices to our aim, which is focussed on a comparison between the quasi-arbitrage-free Bernstein model and the arbitrage-free dynamic polynomial model.

More specifically, we assume that the volatility $S(t,x)$ of the "theoretical curve" $R(t,x)$ has the following parametric form

$$S(t,x) = [\sigma_1 \ x, \ \sigma_2(1 - e^{-\alpha \ x})]' \ , \tag{38}$$

so that

$$\begin{aligned} S_{1,j}(t) &= S_1(t, \frac{j}{k}) = \sigma_1 \ \frac{j}{k} \\ S_{2,j}(t) &= S_2(t, \frac{j}{k}) = \sigma_2(1 - e^{-\alpha \ \frac{j}{k}}) \ , \end{aligned} \tag{39}$$

for $j = 1, \ldots, k$. The specification (38) is analogous to the two-factor volatility function suggested in Heath, Jarrow and Morton (1992). Note, however, that in our model it is sufficient to assume (38) at the points $x = j/k, j = 1, \ldots, k$. Under these assumptions the unknown structural parameters of the state-space model (35) and (36) are $\phi = \{\sigma_1, \sigma_2, \alpha, \lambda_1, \lambda_2\}$. Estimation is obtained by an adaptive Kalman filter. First, the model states $\theta_j^*(t)$ are filtered, for any t, using a set of initial values for the structural parameters ϕ. Given the filtered values, the likelihood of the parameters ϕ is computed and maximized to obtain new estimates $\hat{\phi}$. The final estimates of the model parameters $\theta_j^*(t)$ are computed via the Kalman filter. Smoothed estimated can be obtained, too, from the conditional distribution of $\theta_j^*(t)$ given $\hat{\phi}$ and the data. We used the algorithm described in Koopman, Shephard, and Doornik (1999) as implemented in *E-views 4.0*.

The results are compared with those obtained with an *arbitrage-free polynomial model*. To this aim, it is convenient to rewrite the model in powers of x, that is

$$R_k(t,x) = \sum_{j=1}^{k} b_j(t) q_j(x) = \sum_{j=1}^{k} a_j(t) \ x^j \ .$$

The relationship between the b_j and the a_j's coefficients is given by (15), which for $k = 4$ gives

$$\begin{aligned} a_1(t) &= 4b_1(t) \qquad & a_2(t) &= 6(b_2(t) - 2b_1(t)) \\ a_3(t) &= 4(b_3(t) - 3b_2(t) + 3b_1(t)) \qquad & a_4(t) &= b_4(t) - 4b_3(t) + 6b_2(t) - 4b_1(t). \end{aligned} \tag{40}$$

Now we assume the dynamics (9) for the a_j's coefficients, with the no-arbitrage constraint (11), which imposes $v_{i,3}(t) = v_{i,4}(t) \equiv 0$, $i = 1, 2$ and

$$\begin{aligned} da_1(t) &= (2a_2(t) + v_1(t)'\lambda)\, dt + v_1(t)'\, dW(t) \\ da_2(t) &= (0.5v_1(t)'v_1(t) + 3a_3(t) + v_2(t)'\lambda)\, dt + v_2(t)\, dW(t) \\ da_3(t) &= (v_1(t)'v_2(t) + 4a_4(t))\, dt \\ da_4(t) &= 0.5v_2(t)'v_2(t)\, dt\,. \end{aligned} \tag{41}$$

Again, we consider for simplicity constant volatilities, that is $v_{i,j}(t) = v_{i,j}, i = 1, 2$, and examine two different specifications. The first (*arbitrage-free polynomial model (a)*) is consistent with assumption (39) on the volatility of the b_j's. From equation (40) (or equivalently from (37)), there is a relationship between the $v_{i,j}$ and the $S_{i,j}$'s coefficients, namely $v_{i,1} = 4S_{i,1}$ and $v_{i,2} = 6(S_{i,2} - 2S_{i,1})$, $i = 1, 2$. Therefore (39) gives

$$\begin{aligned} v_{1,1} &= \sigma_1 \quad & v_{2,1} &= 4\sigma_2(1 - \exp(-\tfrac{1}{4}\alpha)) \\ v_{1,2} &= 0 \quad & v_{2,2} &= 6\sigma_2(-1 + 2\exp(-\tfrac{1}{4}\alpha) - \exp(-\tfrac{1}{2}\alpha))\,. \end{aligned} \tag{42}$$

Alternatively (*model (b)*) we can ignore the relationships (40) and (42) and just consider $v_{i,j}$ as unknown constants. The models to be estimated have again observation equation (36), but the state equation is now given by a discrete-time version of (41). Estimation is again obtained by adaptive Kalman filter.

Tables 1 and 2 show the maximum likelihood estimates of the structural parameters, for the dynamic Bernstein model and for the arbitrage-free polynomial models (a) and (b).

Table 1: MLE's of the structural parameters $(\sigma_1, \sigma_2, \alpha, \lambda_1, \lambda_2)$ for the dynamic Bernstein model and for the arbitrage-free polynomial model (a) (in parenthesis is the standard error)

	Bernstein model	arbitrage-free model (a)
σ_1	0.07779 (0.00940)	0.13437 (0.00968)
σ_2	0.03321 (0.01379)	0.01644 (0.00347)
α	3.41461 (1.01879)	-1.26652 (0.27759)
λ_1	-1.98967 (0.36003)	-0.03135 (0.01163)
λ_2	-1.47331 (0.39919)	2.01492 (0.74697)

Figure 2 shows the estimates of the state parameters $a_j(t), j = 1, \ldots, 4$ for the three models under study (the estimates of the parameters $b_j(t)$ for the Bernstein models are transformed into estimates of the parameters $a_j(t)$ via the relationship

Table 2: MLE's of the structural parameters $(v_{1,1}, v_{1,2}, v_{2,1}, v_{2,2}, \lambda_1, \lambda_2)$ for the arbitrage-free polynomial model (b) (in parenthesis is the standard error)

	estimates	standard error
$v_{1,1}$	-0.10725	(1.07532)
$v_{1,2}$	0.00233	(0.02452)
$v_{2,1}$	0.00428	(0.01173)
$v_{2,2}$	0.02450	(0.01127)
λ_1	3.55527	(12.32840)
λ_2	0.89297	(3.04953)

(40)). Note that, for both the arbitrage-free polynomial models (a) and (b), $a_3(t)$ is approximately constant (zero) and $a_4(t)$ is approximately zero. This result is a consequence of the restrictions imposed by no-arbitrage. What happens is particularly clear in the case of constant volatilities $v_{i,j}$'s, since in that case the third and fourth equations in (40) can be solved, and the solutions are, respectively,

$$\begin{aligned} a_3(t) &= a_3(0) + (v_{2,1}v_{2,2} + 4a_4(0))\, t + 2v_{2,2}^2\, t^2 \\ a_4(t) &= a_4(0) + v_{2,2}^2 t \end{aligned}$$

Therefore $a_3(t)$ and $a_4(t)$ will explode as t increases, unless $v_{2,2}$ and $a_4(0)$ are close to zero. This implies that, for sensible results, a_4 is close to zero and a_3 is approximately constant. The only term with a fairly "free" dynamics is a_1. This dynamic behavior is what we actually get for the estimates of the coefficients $a_j(t)$'s shown in figure 2 for model (b). Due to the no-arbitrage constraint, the no-arbitrage polynomial model behaves in fact as a polynomial of degree two. On the other hand, the dynamic Bernstein model displays a considerable variation of the coefficients of the terms of degree 3 and 4. This suggests that, at least when the curve is more unstable, the quality of fit for the no arbitrage models shall be unsatisfactory. This is indeed the case, as we see in Figure 3, which shows the residual sum of squares (RSS) at each date (at weakly intervals), computed using the smoothed state values for the three models. As we can see, the dynamic Bernstein model is fairly better than the other two models, which behave in a very similar way, especially in the interval between the 40th and 100th weeks.

In Figure 4 we plot the differences between the RSS for each arbitrage-free polynomial model and the dynamic Bernstein model, and we compare these values with the difference between the RSS of a cross sectional interpolation using a degree 2 and a degree 4 polynomial model. As we see, the three plotted differences are very similar, and this confirms the fact that the loss of fit for the no-arbitrage models w.r.t. the dynamic Bernstein model is due to the loss of degree implied by the no-arbitrage conditions.

7.2 Option valuation

In the previous section we showed that the Bernstein model can give a good fit of the market data. This result suggests that, assuming that the market evolves according an (unknown) arbitrage-free dynamics, our polynomial model, despite allowing for arbitrage, is close to the true dynamics, and consequently should price derivatives in a very similar way. In the following example we value a set of options using our model and the market benchmark Black's log-normal model (Black, 1976). Our aim is to show that the differences of the two prices are negligible, in the sense that they are inside the bid-ask spread and so cannot be exploited for arbitrage sake.

We consider a call option that gives a payoff $\max(P(T_1, T_2) - K, 0)$ at time T_1, where $T_2 > T_1$ and $P(T_1, T_2)$ is the price at time T_1 of the zero-coupon bond with maturity T_2 on which the option is written.

For valuating an option, we have to preliminary estimate the volatility function $V_k(t, T)$. We obtain such estimate applying the Bernstein model and the Kalman filter procedure described in the previous section.

Then, we compare two different option valuations. The first uses Black's formula with the estimated $V_k(t, T)$. The second uses a Monte Carlo approximation of the expected value (31) (the Monte Carlo sample is obtained by generating sample paths of $P_k(t, T)$ from equation (34), where $\lambda = 0$ and the $W^*(t_\gamma)$'s are i.i.d. $\sim N(0, \delta)$).

If we assume that the market is arbitrage-free, the differences in the results provide a measure of the ability of the Bernstein model in adapting to the underlying unknown dynamics.

We consider nine options with a one year time-to-expiry (i.e., $T_1 = 1$). Each option is written on a different bond, with $T_2 = 2, \ldots, 10$ respectively. The face value of each bond is 100. The options are at the money forward option, that is option strikes are the forward prices for the underlying bonds at the option's expiry date.

Table 3: Value ($\times 100$) of one year to expiry at the money forward call options on bonds with time to maturity from 1 to 9 years.

Years to maturity	1	2	3	4	5	6	7	8	9
Bernstein model	29.6	57.3	80.6	105	125	143	157	173	185
Black model	29.4	56.9	83.1	104	130	146	161	175	191
Differences	0.2	0.6	-2.5	0.8	-5	-3	-4	-2	-6
Spread	4.0	5.6	6.9	8.0	9.0	9.8	10.6	10.6	11.3

The option values obtained with the Bernstein model and with Black formula are reported in Table 3. The last line of the table shows the absolute value of the marked bid-ask spread for similar options as of August 28th, 2002 in the OTC Euribor market. Since historically spreads show a descending trend due to increasing market efficiency and liquidity, these spreads can be considered to understimate the spreads in the sample period. The results show that the differences between the two valuation procedures are well inside the bid-ask spread. The amount of arbitrage in the dynamic Bernstein model is therefore negligible. The advantage of the Bernstein model with respect to the Black model is that it can be easily applied for evaluating options for which Black-like formulas are not available, such as multi-asset options.

8 Final remarks

We discussed a model for inference on the term structure of interest rates, aimed to provide a good cross-sectional fit to market data while controlling for the dynamic no-arbitrage conditions. Our basic point is that, in presence of specification errors, or if the observed market prices only imperfectly follow an arbitrage-fee evolution, the statistical model for interpolation must be regarded as an approximation of the theoretical, arbitrage-free curve. In this case it is essential to keep track of the approximation error, which we do by using constructive Bernstein polynomials, showing that, by refining the approximation, the model can be arbitrarily close to a no-arbitrage model.

We analyzed the performance of the proposed model in comparison with an exactly-arbitrage-free dynamic model, adopting a rather simple specification of the volatility. However, it is possible to extend the analysis to the case of stochastic volatility and to include relevant covariates. This will be the object of further work.

Acknowledgments. We thank Wolfgang Runggaldier and Roberto Silvotti for useful discussions, and Anna Battauz and Fulvio Ortu for comments on a previous version of the paper.

9 Appendix

Proof of theorem 2. We prove the result for $d = 1$. The extension to the case $d > 1$ is straightforward. Let us write

$$R(t,x) = R(0,x) + \int_0^t \mu(s,x)ds + \int_0^t S(t,x)dW(s)$$

and

$$R_k(t,x) = R_k(0,x) + \int_0^t \mu_k(s,x)ds + \int_0^t \sigma_k(t,x)dW(s) \ ,$$

where

$$R_k(0,x) = \sum_{j=0}^{k} R(0, \frac{j}{k})q_j(x) \ .$$

The R-conditions and the assumptions on $r(t,x)$ and on the volatility vector $S(t,x)$ imply that $\mu(t,x)$ is continuous in x and bounded in both the arguments. The same properties hold for $\mu_k(t,x)$ and $\sigma_k(t,x)$.

From proposition 2 it follows that

$$\lim_{k\to\infty} \sup_{x\in[0,1]} |\mu_k(t,x) - \mu(t,x)| = 0$$

and

$$\lim_{k\to\infty}\sup_{x\in[0,1]}|\sigma_k(t,x)-S(t,x)|=0\ .$$

We have

$$\sup_t\sup_x E[(R(t,x)-R_k(t,x))^2]\le\sup_t\sup_x 3\{E[(R(0,x)-R_k(0,x))^2]+ \tag{43}$$

$$E[\{\int_0^t(\mu(s,x)-\mu_k(s,x))ds\}^2]+E[\{\int_0^t(S(t,x)-\sigma_k(t,x))dW(s)\}^2]\}$$

Now, $\sup_x|R(0,x)-R_k(0,x)|\to 0$ as $k\to\infty$, therefore given $\epsilon>0$, the first term in parenthesis on the left hand side can be made $<\frac{\epsilon}{9}$ for k large enough.

By the Cauchy-Schwartz inequality,

$$E[\{\int_0^t(\mu(s,x)-\mu_k(s,x))ds\}^2]\le t\,E(\int_0^t(\mu(s,x)-\mu_k(s,x))^2ds),$$

so that

$$\begin{aligned}\lim_{k\to\infty}\sup_{t\in[0,1]}\sup_{x\in[0,1]}tE(\int_0^t(\mu(s,x)-\mu_k(s,x))^2ds &\le \lim_{k\to\infty}\sup_x E(\int_0^1(\mu(s,x)-\mu_k(s,x))^2ds)\\ &\le \lim_{k\to\infty}E(\int_0^1\sup_x(\mu(s,x)-\mu_k(s,x))^2ds)\end{aligned}$$

and by Lebesgue dominated convergence theorem

$$\lim_{k\to\infty}E(\int_0^1\sup_x(\mu(s,x)-\mu_k(s,x))^2ds)=E(\int_0^1\lim_{k\to\infty}\sup_x(\mu(s,x)-\mu_k(s,x))^2ds)=0.$$

Therefore the second term in parenthesis on the left hand side of (43) can be made $<\frac{\epsilon}{9}$ for k large enough. Finally, we have

$$\begin{aligned}&\sup_{t\in[0,1]}\sup_{x\in[0,1]}E(\{\int_0^t(S(s,x)-\sigma_k(s,x)dW(s)\}^2)\\ =&\sup_{t\in[0,1]}\sup_{x\in[0,1]}\int_0^tE(\{S(s,x)-\sigma_k(s,x)\}^2)ds\\ \le&\sup_{x\in[0,1]}\int_0^1E(\{S(s,x)-\sigma_k(s,x)\}^2)ds\\ \le&\int_0^1E(\sup_{x\in[0,1]}\{S(s,x)-\sigma_k(s,x)\}^2)ds\ .\end{aligned}$$

By Lebesgue dominated convergence theorem

$$\lim_{k\to\infty}\int_0^1E[\{S(s,x)-\sigma_k(s,x)\}^2]ds=\int_0^1E[\lim_{k\to\infty}(S(s,x)-\sigma_k(s,x))^2]ds=0\,,$$

therefore the third term of (43) can be made $\frac{\epsilon}{9}$ for k large enough. The thesis follows. $\diamond$

References

[1] Anderson, N., Breedon, F., Deacon, M., Derry, A. and Murphy, G. (1996). *Estimating and interpreting the yield curve*, Wiley, Chichester.

[2] Ang, A. and Piazzesi, M. (2003). A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. *Journal of Monetary Economics*, **50**, 745-787.

[3] Backus, D.K., Foresi, S. and Zin, S. (1998). Arbitrage opportunities in arbitrage-free models of bond prices. *Journal of Business and Economic Statistics*, **16**, 13-26.

[4] Björk, T. and Christensen, B.J. (1999). Interest rate dynamics and consistent forward rate curves. *Mathematical Finance*, **9**, 323-348.

[5] Black, F. (1976). The pricing of commodity contracts. *Journal of Financial Economics*, **3**, 167-179.

[6] Brace, A., Gatarek, D. and Musiela, M. (1997) The market model of interest rate dynamics. *Mathematical Finance*, **7**, 127-155.

[7] Campagnoli, P., Muliere, P. and Petrone, S. (2001), Generalized dynamic linear models for financial time series. *Applied Stochastic Models in Business and Industry*, **17**, 27-39.

[8] Cohen, J.I.(1999) Markov modeling of the term structure in the Heath-Jarrow-Morton framework. *PhD dissertation*, Cornell University.

[9] Corielli, F. and Petrone, S. (1998). Polinomi di Bernstein e modelli di non arbitraggio per la curva dei rendimenti. (In Italian). *Atti del XXI Convegno Annuale A.M.A.S.E.S.*, 115-129. Roma.

[10] Diebold, F.X. and Li, C. (2003). Forecasting the Term Structure of Government Bond Yields. Avaliable at http://www.ssc.upenn.edu/ fdiebold/papers/papers.html, University of Pennsylvania.

[11] Filipović, D. (1999). A note on the Nelson and Siegel family. *Mathematical Finance*, **9**, 349-359.

[12] Filipović, D. (2001). *Consistency problems for heath-Jarrow-Morton interest rate models* Lecture Notes in Mathematics Vol. 1760. Springer Verlag, Berlin.

[13] Gikhman, I.I. and Skorokhod, A.V. (1965). *Introduction to the theory of random processes.* reprint 1997 by Dover Books on Mathematics, N.Y.

[14] Gombani, A. and Runggaldier, W.J. (2001). A filtering approach to pricing in multifactor term structure models. *International Journal of Theoretical and Applied Finance*, **4**, 303-320.

[15] Kloden, P.E. and Platen, E. (1980). *Numerical Solutions of Stochastic Differential Equations*. Springer-Verlag, Berlin Heidelberg.

[16] Koopman, S.J., Shephard, N. and Doornik, J.A. (1999). Statistical algorithms for models in state space using SsfPack 2.2. *Econometrics Journal*, **2**, 107-160.

[17] Heath, D., Jarrow, R. and Morton, A. (1992). Bond pricing and the term structure of inteerest rates: a new methodology for contingent claims valuation. *Econometrica*, **60**, 77-105.

[18] Langetieg, T.C. and Smoot, J.S. (1989). Estimation of the term structure of interets rates. In *Research in financial services, I*, G.G. Kaufman Ed., JAI Press Inch., Greenwich, Connecticut.

[19] Longstaff, F.A. (1995). Option pricing and the martingale restriction. *The Review of Financial Studies*, **8**, 1091-1124.

[20] Lorentz, G.G. (1953). *Bernstein polynomials*. University of Toronto Press, Toronto.

[21] Makhno, S. (1991). On convergence of solutions of stochastic equations. In: *New Trends in Probability and Statistics*, V.V.Sazonov and T. Shervashidze Eds., 474-484. VSP - Ramadze Math. Institute. Mokslas, Vilnius, Lithuania.

[22] Mastronikola, K. (1991). Yield curves for gilt hedged stocks: a new model. *Discussion paper 49*, Bank of England.

[23] McCulloch, J.H. (1971). Measuring the term structure of interest rates, *Journal of Business*, **44**, 19-31.

[24] Miron, P. and Swanell, P. (1992). *Pricing and hedging swaps*. Euromoney Publications PLC, London.

[25] Musiela, M. and Rutkowski, M. (1998). *Martingale Methods in Financial Modelling*, Springer Verlag, Berlin Heidelberg.

[26] Nelson, C.R. and Siegel, A.F. (1987). Parsimonious modelling of yield curves. *Journal of Business*, **60**, 4, 473-489.

[27] Petrone, S. (1999). Random Bernstein polynomials. *Scandinavian Journal of Statistics*, **26**, 373-393.

[28] Schaefer, S. M. (1981). Tax induced clientele effects in the market for British government securities. *Journal of Financial Economics*, **10**, 121-159.

[29] Steeley, J. (1991). Estimating the gilt hedged term structure: basis splines and confidence intervals. *Journal of Business, Finance and Accounting*, **18**, 513-529.

[30] Tanggaard, C. (1997). Nonparametric smoothing of yield curves. *Review of Quantitative Finance and Accounting*, **9**, 251-267.

[31] Vasicek, O.A. and Fong, H.G. (1982). Term structure modelling using exponential splines. *Journal of Finance*, **37**, 339-356.

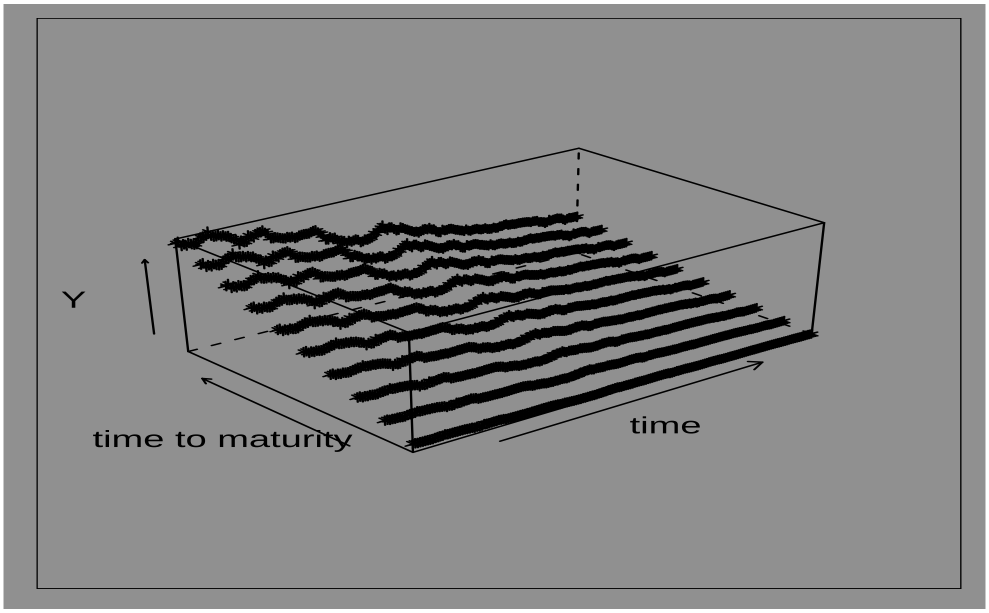

Figure 1: Observed values of the $-\ln$ prices $Y_1, \ldots, Y_{10}$. Daily data from 7/3/1995 to 2/1/1999, plotted every 5 days.

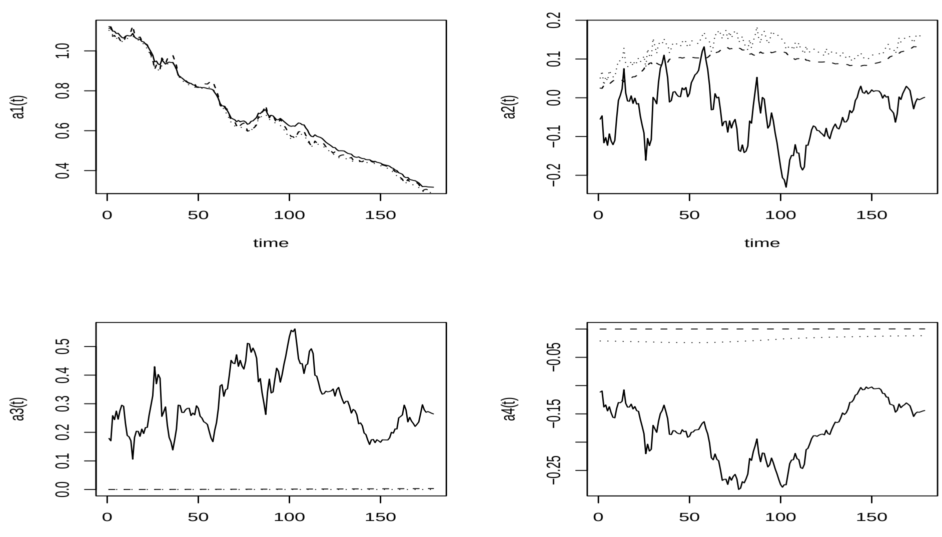

Figure 2: Estimates of the states $a_j(t), j = 1, \ldots, 4$ for the dynamic Bernstein model (solid line) and for the arbitrage-free polynomial models (dashed for model (a) and dotted for model (b))

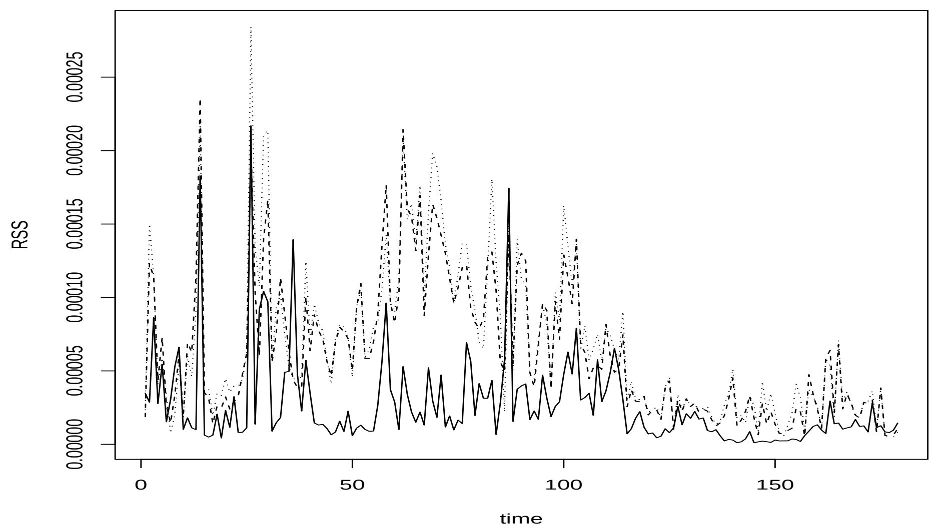

Figure 3: Residual sum of squares (for the ten maturities) for the dynamic Bernstein model (solid line) and for the arbitrage-free polynomial models (dashed for model (a) and dotted for model (b))

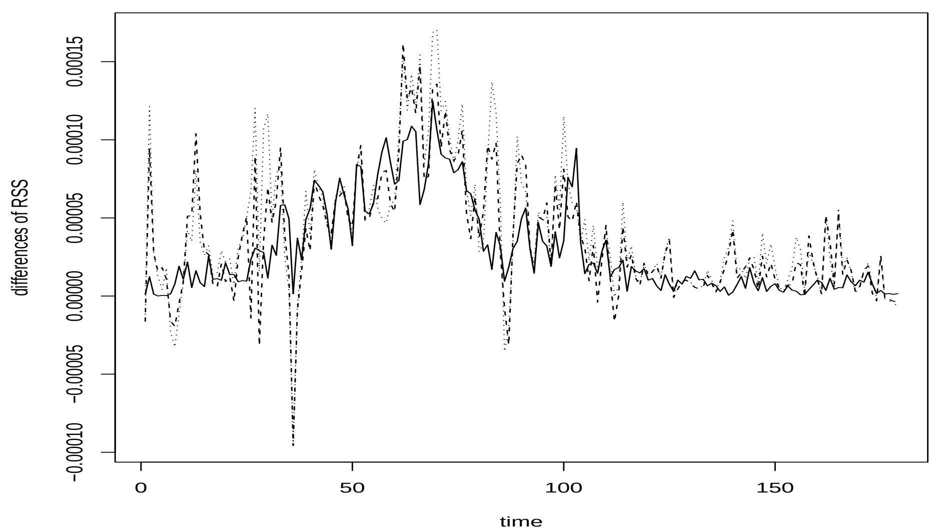

Figure 4: Difference between the residual sums of squares for the arbitrage-free polynomial model and for the Bernstein model (the dashed line is RSS(model (a)) - RSS(Bernstein); the dotted line is RSS(model (b)) - RSS(Bernstein)). The solid line is the difference between the RSS of cross-sectional estimates of a polynomial model of order 2 and the RSS for a cross sectional polynomial model of order 4.