A Semi-Nonparametric Model of the Pricing Kernel and Bond Yields: Univariate and Multivariate Analysis*

Yuriy Kitsul

Department of Economics
Andrew Young School of Policy Studies
Georgia State University
E-mail: ykitsul@gsu.edu

March 13, 2007

Abstract

This study asks whether a sufficiently flexible diffusion framework may explain the interest rates data. We use the eigenfunctions of the infinitesimal generator to model the pricing kernel semi-nonparametrically. Thus, we impose a flexible, but coherent, structure on the short-term risk-free interest rate and on the market price of risk and obtain the closed-form solutions for bond prices. We estimate various versions of the model with a single and multiple independent Gaussian latent factors using the EMM-MMC methodology and US Treasury zero-coupon bond yields data. Our results suggest that flexible nonlinear transformations of the underlying Gaussian factor are sufficient to explain the univariate bond yield dynamics. For the multivariate data, the functional form flexibility does not appear to compensate for the lack of factor correlation.

JEL Codes: C14, G12

*I am deeply indebted to my advisors Dong-Hyun Ahn and A. Ronald Gallant for their suggestions and encouragement. I thank Evan Anderson, Alexander Kovalenkov, Eric Renault, William Parke for their insightful comments. I am also grateful to all the numerous seminar participants for their valuable feedback. All errors and omissions are mine.

1 Introduction

The literature on interest rates has developed in two major directions. Research in the first branch has focused on understanding the time-series dynamics of the short-term risk-free interest rate. The second branch of the literature has concentrated on dynamic models of the term structure of interest rates, which describe how bond yields depend on time to maturity and how the form of this relationship varies over time. Dynamic models of the term structure of interest rates have developed to a significant extent in the continuous time framework and evolved around alternative assumptions regarding the form of the stochastic processes that the short-term risk-free interest rate and the market price of risk follow. Historically, the short-term risk-free interest rate, also called the *instantaneous* risk-free interest rate in the continuous time set-up, was first assumed to be a one-dimensional diffusion process of various types. Then, in attempts to better explain the data, multi-factor, alternative functional form, jump-diffusion and regime-shift extensions were introduced[1]. The subset of extensions providing readily available closed-form solutions for bond prices is limited. Even today the availability of analytic solutions is still a crucial computational issue.

Models with state discontinuities, such as jumps, have a certain intuitive appeal and have become popular in empirical finance literature in general, and in the short-term interest rate and term structure literatures in particular. However, for some purposes their use may be problematic[2]. Therefore, it may be reasonable to ask whether there is a pure diffusion framework that is flexible enough to explain the bond yields data.

We address this question using a semi-nonparametric modelling methodology. This methodology combines flexibility in describing the data, similarly to non-parametric methods, and tractability in the sense of having closed form expressions for the stochastic processes of interest, similarly to parametric models.

In this study we assume that the pricing kernel is an unknown non-linear square-integrable function of unobservable underlying state variables (factors). We model this

[1] See, for example, Dai and Singleton(2003) for a comprehensive literature review. The next section also discusses some representative references.

[2] Jones (2003) reviews the econometric analysis and hedging problems of jump-diffusion models in the context of equity and options.

function semi-nonparametrically[3]. Modeling the pricing kernel is equivalent to joint modeling of the short-term risk-free interest rate and the market price of risk.

We attempt to recover the unknown functional form describing the price deflator by estimating the weights of its expansion on some orthogonal terms. As such orthogonal terms, we use eigenfunctions of the infinitesimal generator of state variables driving the bond prices. The specific form, which the eigenfunctions take, depends on assumptions about the dynamics of the underlying state variables[4]. In the empirical part of this work we assume that factors are mean-reverting and Gaussian.

The properties of the eigenfunctions enable us to conveniently obtain closed-form bond prices and concentrate on the following empirical questions:

1) Is there such a functional form of the relationship between the underlying diffusion state variables and the pricing kernel that would fit the bond yields data? In other words, can we explain the data while still remaining in a pure diffusion framework, which, besides, is analytically tractable?

2) All the square-integrable functions of the underlying state variables considered as candidates for the pricing kernel[5] constitute a large subset of all possible diffusion processes. Nevertheless, is a single unobservable state variable sufficient to explain the data? Otherwise, how many state variables are needed? If the the data fit is still not satisfiable, what are the model-inherent restrictions that cause this to happen?

The semi-nonparametric model by itself does not guarantee that the "no-arbitrage" condition holds. We propose two methods to enforce this condition and explore the implications that the framework bears for the dynamics of the short-term risk-free interest rate and the market price of risk.

There are two alternative paths available to implement empirically the theoretical and methodological issues addressed in this paper. The first path is to estimate a full fledged term-structure model, i.e. to estimate the joint dynamics of several bond yields. The second is to estimate the time series dynamics of a single bond yield following the tradition

[3]To be more precise, the object that we model semi-nonparametrically is often called the price deflator (e.g. Duffie (2001)) or the "discount factor in continuous time" (e.g. Cochrane (2001)). The price deflator, in turn defines the conventional pricing kernel.

[4]Meddahi(2001-a, b) uses eigenfunctions to model stochastic volatility and describes their interesting expectation properties, which we build on later in the paper.

[5]At least theoretically, with an infinite number of terms in the expansion.

of the literature on the dynamics of the short-term interest rate, but at the same time to address the problematic issues inherent in this literature, such as non-instantaneous time to maturity of the bond in question[6]. We follow both directions in this study and utilize both univariate and multivariate data.

We estimate the model using the Efficient Method of Moments (EMM) of Gallant and Tauchen (1996) augmented with the Markov Chain Monte Carlo methodology as described in Chernozhukov and Hong (2003), Gallant (2003) and Gallant and Tauchen (2004-a). In application to our model, EMM enables us to avoid an ad-hoc selection of moment conditions and work with the unobservable variables, such as the underlying factors and the short-term risk-free interest rate. The Markov Chain Monte Carlo methodology provides a convenient extension to frequentist extremum estimation procedures in highly non-linear and computationally formidable settings like the one considered in this paper.

We estimate the model with the univariate time-series data on "6-month-to-maturity" treasury yields, we are not able to reject the model. When trying to fit the joint evolution of three yields at a time, we are less successful. We attribute this result mostly to our assumption that the underlying latent factors are independent. Our findings suggest the univariate yield dynamic can be readily accommodated with a flexible diffusion process. In the multivariate context, our estimation results seem to support the discussion in the literature about the importance of co-dependence among the underlying factors and appear to suggest that flexible functional forms per se are not able to compensate for the lack of factor correlation.

This paper is organized as follows: in part two we discuss the current modeling and empirical issues and relation of our approach to the existing literature. In part three we review the eigenfunction framework and discuss how we use eigenfunctions to model the stochastic discount factor. In part four we discuss bond pricing and implications that our framework bears for the instantaneous risk-free interest rate and the market price of risk. In part five we discuss the data and estimation methodology. Next, we discuss the results and, finally, conclude.

[6]Chapman, Long and Pearson (1997) discuss the cases where ignoring non-zero time to maturity becomes problematic.

2 Current modeling and empirical issues, and related literature.

The literature on the term structure of interest rates has been developing in great part around the fact that in a continuous time framework, a price, $B_t(T)$, at time t of a bond that matures and pays one dollar at time $T > t$ is equal to

$$B_t(T) = E_t^Q\{e^{-\int_t^T r_s ds} \cdot 1\} \tag{1}$$

where r_s is the instantaneous risk-free interest rate, which is also sometimes referred to as a short-term risk-free interest rate, and the expectation E_t^Q is conditional on information available at time t and taken with respect a risk-neutral probability measure[7], Q.

The valuation equation (1) shows that knowing the parameters of the process that drives r_t under the risk-neutral measure should suffice to derive a bond-pricing formula. However, from the time-series data of the observed bond yields, one can estimate parameters of r_t under objective measure, P. The risk-neutral probability measure, Q, is different from the objective probability measure, P, which governs the observed real-world time-series behavior of interest rates. To switch between the two measures, one has to specify the so-called market price of risk and use Girsanov theorem. Using Girsanov theorem, one can obtain:

$$B_t(T) = E_t^P\{e^{-\int_t^T r_s ds}\xi_{t,T} \cdot 1\} \tag{2}$$

where $\xi_{t,T}$ is the Radon-Nikodym derivative, which in turn is a functional of the market price of risk λ_s. Therefore, one also needs to model the market price of risk to be able to switch between the risk-neutral and objective measures. A test of a term structure model, which would use time series data of bond prices, would be a joint test of hypotheses about the dynamics of r_t and the market price of risk.

A subset of specifications for r_t, which produce closed form solutions of the expectation (1) for bond prices is limited. The vast body of literature avoids this problem and focuses on estimation of the short-term risk-free interest rate dynamics, r_t, under physical, or

[7] Harrison and Kreps(1979) is the seminal paper that discusses the issues related to the risk-neutral, or equivalent martingale, measure and Girsanov Theorem in a multi-period setting.

real-world, measure. Chapman and Pearson (2001) provide a comprehensive survey. A few representative references[8] are Aït-Sahalia (1996), Durham (2003), Eraker (2001) and Zhou (2003).

The extra allowed flexibility of r_t, which this literature enjoys, does not come at no price, however. Chapman and Pearson (2000) discuss the so-called proxy biases, which arise because of using longer maturity interest rates as a proxy for the instantaneous interest rate. They find that these biases are especially evident for more sophisticated, non-linear models.

Studies, which do concentrate on the term structure of interest rates, often rely on assuming relatively simple specifications for the short-term interest rate and the market price of risk to obtain closed-form solutions for bond prices and relieve the computational burden. Such analytic tractability explains, for example, the popularity of affine term-structure models. However, the affine term structure models do not appear to be supported by the data. As Dai and Singleton (2000) discuss, there is a trade-off between having a flexible correlation structure of the underlying factors and having flexible dynamics of the underlying factors' volatility. In addition, the market price of risk has a constant sign in the affine models.

In response, various extensions have been offered. For example, Ahn, Dittmar, Gallant (2002), Ahn, Dittmar, Gallant and Gao (2003) avoid the trade-offs inherent in affine models and described above, by assuming a quadratic relationship between underlying state variables and instantaneous risk-free interest rate. Another example is regime shift models of Bansal and Zhou (2002), Bansal, Tauchen and Zhou(2003) and Dai, Singleton and Yang (2006) and others[9]. Duffie, Pan and Singleton(2000) is an example of a work that studies jump-diffusions.

A different way of looking at a valuation equation (2), is through a prism of the stochastic discount factor[10], $M_{t,T}$, which assigns prices at time t to payoffs at different states of nature at time $T > t$. A price at time t of a zero-coupon bond that pays one dollar at time T:

$$B_t(T) = E_t^P\{M_{t,T} \cdot 1\} \tag{3}$$

[8]This list is by no means exhaustive.

[9]A significant number of regime shifts studies operate in a discrete time framework.

[10]Hansen and Richard (1987) is an earlier work that studies the nominal stochastic discount factor.

In continuous time framework, it is more rigorous, however, to talk about the price deflator M_t, that that defines the pricing kernel as $M_{t,T} = \frac{M_T}{M_t}$ and enters the valuation equation as:

$$M_t B_t(T) = E_t^P\{M_T \cdot 1\} \tag{4}$$

Studies on the term-structure of interest normally assumes some specific form of the stochastic discount factor, either directly or by specifying some particular dynamics for the short-term risk-free interest rate and the market price of risk.

In contrast, in this study we examine a question about a functional form of the price deflator, and thus the stochastic discount factor, by utilizing a flexible semi-nonparametric framework.

By doing so, we continue a tradition of empirical and non-linear pricing kernels literature, e.g. Bansal, Hsieh and Vishwanathan (1993) and Bansal and Vishwanathan (1993), Chapman (1997), Dittmar (2002), Rosenberg and Engle (2002), Chernov (2003). Most of these studies try to avoid assumptions on exact form of preferences and the pricing kernel by using various kinds of polynomials for semi-nonparametric modeling[11],[12]. In this study, we use eigenfunctions, which, in some cases, also take a form of orthogonal polynomials. Interesting properties of the eigenfunctions allow us to obtain a bond pricing formula and focus on the dynamics of the short-term risk-free interest rate and term structure. This is one difference of our study from most of the pricing kernels literature. The other difference is that in our framework the state variables are unobservable. This is important, because features captured empirically may provide guidance for preference based, general equilibrium models, which may otherwise be very difficult to test, given quality and limited availability of, for example, consumption or aggregate wealth data.

Some of the examples of use of eigenfunctions for asset pricing are Lewis (1998), Linetsky (2002), Davydov and Linetsky (2003), Gorovoi and Linetsky (2003) and Goldstein and Keirstead (1998). In contrast to our study, these papers consider expansions of payoffs of securities on eigenfunctions of a pricing operator directly with the weight of each eigen-

[11] Exceptions are Bansal and Vishwanathan (1993) who use neural network approximations, and Chernov (2003), who assumes the specific objective and risk-neutral asset prices dynamics and the specific form of the market price of risk and tries to recover the pricing kernel from asset prices.

[12] Brandt and Yaron (2003) use a Hermite polynomial expansion of the pricing kernel to focus on time-consistency issues in no-arbitrage, or market term-structure modeling.

function depending in a specific way on the parameters of the underlying state variable. Also, these studies are theoretical and are not concerned with econometric estimation.

In addition, Rogers (1997) and Rogers and Zane (1997) model the stochastic discount factor using the *potential* approach. In his theoretical paper, Rogers(1997) mentions that one of the generic approaches to construction of the *potential* examples is to use eigenfunctions, and presents the bond pricing formula for such a case.[13] We obtain analogous bond pricing formula using the expectation properties of the eigenfunctions. The discussion of the eigenfunction approach in Rogers (1997) is in generic terms and is not concerned with the practical issues such as the implementation of a multi-factor case or how to impose the positivity restriction on the stochastic discount factor.

Some examples of the papers that use eigenfunctions in econometrics are Hansen and Scheinkman (1995), Florens, Renault and Touzi (1998), Hansen, Scheinkman and Touzi (1998), Chen, Hansen and Scheinkman (2000) and Meddahi (2001-a,b). Although these studies use eigenfunctions in a different context, they discuss the general theory of eigenfunctions and provide the relevant theoretical background, which we build on and discuss in the next section.

3 Modeling Framework

3.1 Eigenfunctions

In this section we review the definition and properties of eigenfunctions[14]. The discussion will be concerned with scalar diffusions. Later we will describe an extension to the case independent multiple factors, which is based on a discussion in Meddahi(2001 a,b).

Let us consider a stationary scalar diffusion described by the following stochastic differential equation

$$dx_t = \mu(x_t)dt + \sigma(x_t)dW_t \tag{5}$$

where $t > 0$, $x \in (l; r)$, $\sigma(x) > 0$, W_t - one-dimensional Brownian motion on a filtration $\{\mathcal{F}_t\}$.

[13] Rogers and Zane (1997) do not use eigenfunctions.

[14] We follow the notation and exposition of Hansen at al.(1998) and Meddahi(2001-a,b) to a certain degree, while describing a general framework for eigenfunctions and spectrum.

Another way to describe this diffusion is by using an infinitesimal generator. For the functions ϕ that satisfy some conditions (C_B^2, twice continuously differentiable and bounded):

$$\mathcal{A}\phi(x) = \mu(x)\phi(x)' + 0.5\sigma(x)\phi(x)'' \tag{6}$$

Let us define scale function and speed density, $S(x)$ and $m(x)$, respectively (see, for example, Karlin and Taylor (1981) or Hansen et al. (1998)).

$$S(x) = \int^x s(\xi)d\xi \tag{7}$$

where

$$s(x) = exp(-\int^x \frac{2\mu(\xi)}{\sigma^2(\xi)}d\xi) \tag{8}$$

Also,

$$m(x) = \frac{1}{s(x)\sigma^2(x)} \tag{9}$$

The density of a stationary distribution of measure Q[15] can be represented as

$$q(x) = \frac{m(x)}{\int_l^r m(\xi)d\xi} \tag{10}$$

assuming finite denominator.

Let us assume for a moment that all the conditions, like time reversibility, appropriate boundary conditions and so on, that ensure existence and discreteness of the spectrum of the infinitesimal generator, hold. As a matter of fact, these conditions will hold for the forms of the stochastic processes governing the dynamics of the latent factors x_t that we use in the empirical work. The spectrum is

$$\mathcal{A}\phi = \sum_{i=0}^{\infty} -\delta_i(\phi_i|\phi)\phi_i \tag{11}$$

where $(\phi_i|\phi)$ is an inner product, i.e.

$$(\phi_i|\phi) = \int_l^r \phi_i^*(x)\phi(x)dQ = \int_l^r \phi_i^*(x)\phi(x)q(x)dx \tag{12}$$

[15]This is not the risk-neutral measure Q discussed in the previous section.

and ϕ_i is an i-th eigenfunction and δ_i is a corresponding eigenvalue, which solve the following equation

$$\mathcal{A}\phi_i = -\delta_i \phi_i \tag{13}$$

which is the same as

$$\mu(x)\phi_i(x)' + \frac{1}{2}\sigma^2(x)\phi_i(x)'' + \delta_i \phi_i(x) = 0 \tag{14}$$

The following property about conditional expectation follows from the definition of eigenfunctions (13) and will be important for derivation of bond prices:

$$E(\phi_i(x_T)|\mathcal{F}_t) = e^{-\delta_i(T-t)}\phi_i(x_t) \tag{15}$$

Eigenfunctions may take various forms depending on the form of underlying diffusion process. For example, let us consider an Ornstein-Uhlenbeck process of the following standardized form:

$$dx_t = -\kappa x_t dt + \sqrt{2\kappa} dW_t \tag{16}$$

One can show that in this case a stationary density, which is normal, and eigenfunctions are Hermite polynomials $H_i(x)$, which are orthogonal with respect to $e^{-\frac{x^2}{2}}$. The eigenvalues are $\delta_i = \kappa i$. Hermite polynomials can be computed using the following relationships:

$$H_0(x_t) = 1, H_1(x_t) = x_t, H_i(x_t) = \frac{1}{\sqrt{i}}(x_t H_{i-1} - \sqrt{i-1} H_{i-2}(x_t)) \tag{17}$$

An Ornstein-Uhlenbeck processes, y_t, of a general form

$$dy_t = \kappa(\theta - y_t)dt + \eta dW_t \tag{18}$$

is a linear transformation of the process x_t described by (16):

$$y_t = \frac{\eta}{\sqrt{2\kappa}} x_t + \theta \tag{19}$$

Then, the eigenfunctions will be of the same form, but with a different argument, i.e.

the eigenfunctions will be equal to the Hermite polynomials of the same form,$H_i(y) = H_i(\frac{\eta}{\sqrt{2\kappa}}x + \theta)$.

There are several other stochastic processes which are characterized by infinitesimal generators whose eigenfunctions are orthogonal polynomials. For example, eigenfunctions of an infinitesimal generator of a square-root process are generalized Laguerre polynomials and the eigenvalues are $\delta_i = \kappa i$. In other words, if x_t follows:

$$dx_t = \kappa(\alpha + 1 - x_t)dt + \sqrt{2\kappa}\sqrt{x_t}dW_t \tag{20}$$

then eigenfunctions are:

$$\begin{aligned}
L_0^\alpha(x_t) &= 1, L_1^\alpha(x_t) = \frac{1 + \alpha - x_t}{\sqrt{1 + \alpha}}, \\
\sqrt{1 + \alpha} i L_i^\alpha(x_t) &= \sqrt{i - 1 + \alpha}(-x_t + 2i + \alpha - 1)L_{i-1}^\alpha(x_t) \\
&+ \sqrt{i - 2 + \alpha}(i + \alpha - 1)L_{i-2}^\alpha(x_t)
\end{aligned}$$

The extension to square-root processes of a general form is of the same nature as the extension in the Ornstein-Uhlenbeck case.

With help of the eigenfunctions any square-integrable function $\psi(x)$, i.e. any function $\phi(x)$ for which the following inequality holds

$$\int_l^r \psi(x)^2 dQ < \infty \tag{21}$$

may be represented as

$$\psi(x) \doteq \sum_{i=0}^{\infty} (\phi_i|\psi)\phi_i(x) \tag{22}$$

where $(\doteq)$ means convergence in mean-square. Denoting $(\phi_i|\psi) = a_i$, we rewrite (22) as

$$\psi(x) \doteq \sum_{i=0}^{\infty} a_i\phi_i(x) \tag{23}$$

In we the next sub-section we use (23) to model the stochastic discount factor.

3.2 Price Deflator

We use the relationship (23) to model stochastic discount factor as a linear combination of the eigenfunctions of the infinitesimal generator of the underlying state variable.

Let us assume that the instantaneous stochastic discount factor M_t, which assigns prices at time t to payoffs at time $t + dt$, is some unknown square-integrable function ψ of the latent factor x_t, i.e.

$$M_t = \psi(x_t) \tag{24}$$

As in (23), the function $\psi(x_t)$ can be expanded on eigenfunctions of the infinitesimal generator of x_t, $\phi_i(x_t)$:

$$M_t = \sum_{i=0}^{\infty} a_i \phi_i(x_t) \tag{25}$$

The next step, which is discussed in the next section, is to derive the bond-pricing formula. Than, we assume some particular forms of the dynamics of the state variable x_t, we truncate the number of terms in the summation (25) to some $n < \infty$ and, since the functional form $\psi(x_t)$ is unknown, we estimate a_i and the parameters of the underlying state process, x_t, from time-series of the observed bond prices.

3.3 Positivity of the Price Deflator

The "no-arbitrage" condition holds if and only if the pricing kernel is positive[16]. We offer two methods that can be used to implement the positivity constraint.

When eigenfunctions take the form of orthogonal polynomials, it is possible to impose the positivity by finding the map between the weights of eigenfunctions and the coefficients of the regular polynomials which only have complex roots.

For example, let us consider the special case, where the stochastic discount factor is modeled as a linear combination of the first two Hermite polynomials $M_t = a_0 + a_1 H_1(x_t) + a_2 H_2(x_t)$.

The next step is to find numerically the map between the coefficients (a_0,a_1, a_2) of this combination of Hermite polynomials and coefficients (b, c) of a regular polynomial of

[16] See, for example, Hansen and Richard (1987).

the second order that has only complex roots and, thus, never crosses zero:

$$a_0 + a_1 H_1(x) + a_2 H_2(x) = (x - b - ic)(x - b + ic) \tag{26}$$

In the process of estimation, one would start with a candidate for (b, c) and map into (a_0, a_1, a_2). Thus, $M_t = a_0 + a_1 H_1(x_t) + a_2 H_2(x_t)$ is either positive or negative on the whole state space of x_t. One only needs to evaluate M_t at any point of the state space of x_t and if M_t is negative, one has to multiply it by -1.

Another way to incorporate the positivity restriction is to consider it as a part of the "prior information" at the estimation stage. This is what we do in this work, since we use the Markov Chain Monte Carlo methodology.

3.4 Multi-Factor Extension

Let us consider the independent factors, x_{1t} and x_{2t}, with eigenfunctions ϕ_{1i} and ϕ_{2j}, respectively, and the stochastic discount factor M_t of the form:

$$M_t = \sum_{i=0}^{p_1} \sum_{j=0}^{p_2} a_{ij} \phi_{1i}(x_{1t}) \phi_{2j}(x_{2t}) \tag{27}$$

As Meddahi (2001-a) discusses, one can define multi-factor functions, $\phi_{ij}(x_t)$, where $x_t = [x_{1t}, x_{2t}]$, as a product of two individual eigenfunctions,

$$\phi_{ij}(x_t) = \phi_{1i}(x_{1t}) \phi_{2j}(x_{2t}), \tag{28}$$

Then, the following property holds:

$$E_t\{\phi_{ij}(x_T)\} = e^{-\delta_{ij}(T-t)} \phi_{ij}(x_t) \tag{29}$$

where $\delta_{ij} = \delta_{1i} + \delta_{2j}$.

Therefore, all the bond pricing results, discussed in the next section, are applicable. In addition, the same principle can be can be used to extend the framework by introducing more than two independent state variables.

3.5 Generality of Eigenfunctions Approach.

In this subsection we explore the generality of the semi-nonparametric eigenfunctions approach. Given a state variable x_t, it in theory allows to model any square integrable function of x_t. The question is how to describe the set of all possible diffusion processes for m_t that we are able to model with an eigenfunction approach once we assumed some specific form for a factor x_t.

For example, if we assume that an underlying latent factor x_t is governed by Ornestein-Uhlenbeck process, then, with the help of Hermite polynomials, we should be able to model the set of all the diffusion processes for m_t that can be represented as any square integrable function of the Ornestein-Uhlenbeck process. Let us call this set of diffusion processes for m_t as set I.

On the other hand, if we assume that an underlying latent factor x_t is governed by a square root process, then, with the help of Laguerre polynomials, we should be able to model the set of all the diffusion processes for m_t that can be represented as any square integrable function of the square-root process. Let us call this set of diffusion processes for m_t as set II.

The question of interest is whether set I and set II coincide. If not, do these two sets intersect? Is it possible to describe formally which processes belong to the intersection of these two sets and which do not?

Consider a latent state variable x_t:

$$dx_t = \mu_x(x_t)dt + \sigma_x(x_t)dW_t \tag{30}$$

Using the eigenfuntions of the infinitesimal generator of the process x_t, we model a function $m_{1t} = m_1(x_t)$, which can be any L^2 of x_t:

$$m_{1t} = \sum_{i=0}^{\infty} a_i E_i^x(x_t) \tag{31}$$

Now consider an alternative state variable y_t, which is a stochastic process different from

x_t, and an L^2 function of y_t, $m_{2t} = m_2(y_t)$:

$$dy_t = \mu_y(y_t)dt + \sigma_y(y_t)dW_t \tag{32}$$

$$m_{2t} = \sum_{i=0}^{\infty} a_i E_i^y(y_t) \tag{33}$$

The same Brownian motion, W_t, drives both x_t and y_t. We would like to ask a question whether for a given function $m_1(x_t)$ there exists such a function $m_2(y_t)$ that $m_1(x_t) = m_2(y_t)$ sample-wise for all t.

After introducing additional restrictions on $m_2(y_t)$ (invertibility), the question can be reformulated to whether there is a function $y_t = m(x_t) = m_2^{-1} m_1(x_t)$.

Using Ito's lemma we obtain a system of differential equations:

$$\mu_y(m(x_t)) = \mu_x(x_t)m'(x_t) + 0.5\sigma_x^2(x_t)m''(x_t) \tag{34}$$

$$\sigma_y(m(x_t)) = \sigma_x(x_t)m'(x_t) \tag{35}$$

The system reduces to the following differential equation:

$$0.5\sigma_x^2(x_t)m''(x_t) = \mu_y(m(x_t)) - \mu_x(x_t)\frac{\sigma_y(m(x_t))}{\sigma_x(x_t)} \tag{36}$$

Therefore, the sets of diffusion processes I and II intersect only if for a given L^2 function, $m_1(x_t)$, there exists a function $m(x_t) = m_2^{-1} m_1(x_t)$ that solves the ordinary differential equation (36), which is not guaranteed in the general case.

4 Bond prices, short-term risk-free interest rate and market price of risk

In this section we derive a price at time t of a zero-coupon bond that matures and pays one dollar at time T, $B_t(T)$. Recall the pricing equation $M_t B_t(T) = E_t(M_T \cdot 1)$. Substitution of the expression for the pricing kernel, (25), into this equation produces:

$$\left(\sum_{i=0}^{\infty} a_i \phi_i(x_t)\right) B_t(T) = E_t\left(\sum_{i=0}^{\infty} a_i \phi_i(x_T)\right) \tag{37}$$

Next, we use the expectation property of eigenfunctions (15), which is

$$E_t(\phi_i(x_T)) = e^{-\delta_i(T-t)}\phi_i(x_t) \tag{38}$$

and obtain:

$$\begin{aligned} B_t(T) &= \frac{\sum_{i=0}^{\infty} a_i e^{-\delta_i(T-t)}\phi_i(x_t)}{\sum_{i=0}^{\infty} a_i\phi_i(x_t)} \\ &= \sum_{i=0}^{\infty} \frac{a_i\phi_i(x_t)}{\sum_{j=0}^{\infty} a_j\phi_j(x_t)} e^{-\delta_i(T-t)} \end{aligned} \tag{39}$$

The obtained bond price is a weighted average of $e^{-\delta_i(T-t)}$, with the normalized orthogonal components of the pricing kernel serving as the weights and summing up to one.

An expression for the corresponding instantaneous risk-free interest rate can be easily obtained from the bond pricing formula (39) using the L'Hopital's rule:

$$\begin{aligned} r_t &= \lim_{T-t\to 0} -\frac{ln(B_t(T))}{T-t} \\ &= \lim_{T-t\to 0} \frac{\sum_{i=0}^{\infty} a_i\delta_i e^{-\delta_i(T-t)}\phi_i(x_t)}{\sum_{i=0}^{\infty} a_i e^{-\delta_i(T-t)}\phi_i(x_t)} \end{aligned} \tag{40}$$

Thus, the instantaneous risk-free interest rate, r_t, becomes

$$r_t = \frac{\sum_{i=0}^{\infty} a_i\delta_i\phi_i(x_t)}{\sum_{i=0}^{\infty} a_i\phi_i(x_t)} \tag{41}$$

The instantaneous risk-free interest rate in equation (41) is a weighted average of the eigenvalues, δ_i, with the normalized orthogonal components of the pricing kernel serving as the weights and summing up to one. It follows from (41) that a rate of change of the instantaneous risk-free interest rates with respect to an eigenvalue, δ_i, is equal to

$$\frac{\partial r_t}{\partial \delta_i} = \frac{a_i\phi_i(x_t)}{\sum_{i=0}^{\infty} a_i\phi_i(x_t)} \tag{42}$$

i.e. it is equal to a weighting term in a bond pricing and instantaneous risk-free interest rate formulas.

Next, we discuss the implications of our approach for the market price of risk. Applying the text book exposition in Duffie (2001) to our case, we assume appropriate regularity conditions and consider some arbitrary security with a price S_t following an Ito process $dS_t = \mu(S_t)dt + \sigma(S_t)dW_t$ and the pricing kernel, M_t, following some Ito process $dM_t = \mu(M_t)dt + \sigma(M_t)dW_t$. Then, the cumulative-return process of this security is defined as

$$dR_t = \mu(R_t)dt + \sigma(R_t)dW_t = \frac{\mu(S_t)}{S_t}dt + \frac{\sigma(S_t)}{S_t}dW_t \tag{43}$$

Then, the expected excess return is expressed as

$$\mu(R_t) - r_t = -\frac{1}{M_t}\sigma(R_t)\sigma(M_t) \tag{44}$$

where $r_t = -\mu(M_t)/M_t$ is equal to a short-term riskless process. One can see that $-\frac{\sigma(M_t)}{M_t}$ is a Sharpe ratio, or a market price of risk, which we denote by $-\lambda_t$.

To derive the market price of risk in our framework, we use Ito lemma to obtain the diffusion process for the instantaneous stochastic discount factor.

$$dM_t = \{\mathcal{A}(\sum_{i=0}^{\infty} a_i\phi_i(x_t))\}dt + \sigma(f_t)(\sum_{i=0}^{\infty} a_i\phi_i(x_t))'dW_t \tag{45}$$

Using the definition of the eigenfunctions:

$$dM_t = -(\sum_{i=0}^{\infty} a_i\delta_i\phi_i(x_t))dt + \sigma(x_t)(\sum_{i=0}^{\infty} a_i\phi_i(x_t)')dW_t \tag{46}$$

Therefore,

$$\lambda_t = \frac{\sigma(x_t)(\sum_{i=0}^{\infty} a_i\phi_i(x_t)')}{\sum_{i=0}^{\infty} a_i\phi_i(x_t)} \tag{47}$$

and

$$r_t = \frac{\sum_{i=0}^{\infty} a_i\delta_i\phi_i(x_t)}{\sum_{i=0}^{\infty} a_i\phi_i(x_t)} \tag{48}$$

The expression (48) for the risk-free interest rate, r_t, coincides with the expression (41), which is derived directly from the bond pricing formula.

Equations (47) and (48) demonstrate that by modeling the pricing kernel in the semi-nonparametric way we allow for general and flexible specifications of the market price of risk and the short-term risk-free interest rate. The rest of the paper is concerned with these specifications' empirical performance.

5 Estimation Methodology

The basis of the estimation methodology is the Efficient Method of Moment (EMM) developed by Gallant and Tauchen(1996). However, the approach followed in this work differs from the original EMM. Instead of using numeric optimization, we apply the Monte Carlo Markov Chain methods along the lines of Chernozhukov and Hong(2003), Gallant (2003) and Gallant and Tauchen(2004).

We first summarize the ideas behind the traditional Efficient Method of Moments (EMM), using notation of Chernov, Gallant, Ghysels and Tauchen(2001). The logic of EMM methodology is related to the ideas underlying the Simulated Method of Moments of Duffie and Singleton (1993) and the Indirect Inference Method of Gourieroux, Monfort and Renault(1993).

The task is to estimate the parameters, ρ, of a structural model, the estimation of which with the maximum likelihood methods is either not feasible or practical. Let us call it the main model. The key idea is to introduce an auxiliary model, which is misspecified, but approximates the data sufficiently well and has a readily computable likelihood function, $f(\cdot)$, in a closed form. The next step is to estimate the parameters, θ, of the auxiliary model with quasi-maximum likelihood and using the observed data, $\tilde{y}_t$, $\tilde{x}_{t-1}$:

$$\tilde{\theta}_n = \arg\max_{\theta\in\Theta} \frac{1}{\mathrm{n}} \sum_{t=1}^{n} log[f(\tilde{y}_t|\tilde{x}_{t-1},\theta)] \tag{49}$$

The obtained score vector is used to generate moment conditions by simulation from the model of interest, i.e. the main model:

$$m(\rho,\theta) = \frac{1}{N}\sum_{t=1}^{N} \frac{\partial}{\partial\theta} log[f(\hat{y}_t|\hat{x}_{t-1},\theta)] \tag{50}$$

If the main model is indeed true, then, by construction, there should exist such values, $\hat{\rho}$, of its parameters that the expectation of the score vector is zero when evaluated at the data simulated using these parameter values. Thus, we choose the parameters of the main model, ρ, such that, in effect, the simulated data resembles the real data as close as possible. We achieve this by making a quadratic form of the expected score close to zero. Thus, we obtain the traditional EMM estimator by:

$$\hat{\rho}_n = \arg\min_{\rho} m'(\rho, \tilde{\theta}_n)(\tilde{I}_n)^{-1} m(\rho, \tilde{\theta}_n) \tag{51}$$

where $\tilde{I}_n$ is a quasimaximum likelihood information matrix.

Gallant and Long (1997) demonstrate that a score of an auxiliary model has to span a score of a true density asymptotically for EMM to be as efficient as Maximum Likelihood. When compared to GMM, EMM allows to avoid both the explicit derivation of moments and ad hoc selection of moment conditions.

A critical task is to choose the auxiliary model that approximates the true conditional density of the process closely enough. For this purpose, Semi-Nonparametric (SNP) density, proposed by Gallant and Nychka (1987), Gallant and Tauchen (1989), is often used.

The semi-nonparametric density function of innovation z_t is represented as

$$h_K(z|x) = \frac{[P_K(z,x)]^2 \phi(z)}{\int [P_K(z,x)]^2 \phi(z) du} \tag{52}$$

where $P(z,x)$ is usually expressed as a polynomial of degree K_z and each coefficient is, in turn, a polynomial of degree K_x in x. The leading term of the expansion, $\phi(z)$ is the normal density. The remaining terms are aimed to capture departures from normality.

Next, to obtain the approximation for a transition density that governs the data, $f(y_t|x_{t-1})$, where $x = (y_{t-1}, y_{t-2}, ...)$, one performs the location-scale transformation:

$$y = \Sigma_x z + \mu_x \tag{53}$$

Then, the conditional density of the data is proportional to the normal density with the first two moments, μ_x and Σ_x, respectively:

$$f(y|x,\theta) \propto [P_K(z,x)]^2 n_M(y|\mu_x, \Sigma_x) \tag{54}$$

In order to nest VAR, GARCH, "level in volatility" and leverage effects, one can also impose additional structure on Σ_x and μ_x and assume that they depend on the past observations, $x = (y_{t-1}, y_{t-2}, ...)$. The number of lags in the GARCH part of conditional volatility function are denoted L_g and L_r. The number of lags in the "level-in-volatility" and leverage parts of conditional volatility are denoted L_w and L_v, respectively. The number of lags in conditional mean is denoted L_u.

In practice, a researcher starts with the least sophisticated and the most parsimonious semi-nonparametric model, and keeps expanding and estimating the models until the optimal one is found based on various information criteria, such as AIC or BIC. The scores of the preferred model, $m(\rho, \tilde{\theta}_n)$, then serve as the moment conditions, as described in equations (50) and (51).

In this work, instead of finding $\hat{\rho}$ by traditional numeric minimization we construct a Markov Chain of parameters, ρ. Let us denote the objective function that we need to minimize as s_n:

$$s_n(\rho) = m'(\rho, \tilde{\theta}_n)(\tilde{I}_n)^{-1} m(\rho, \tilde{\theta}_n) \tag{55}$$

Following Chernozhukov and Hong (2003) and Gallant and Tauchen (2004), we use the objective function $s_n(\rho)$ to construct a function $L(\rho)$ described below, which can be considered as an analog to the likelihood in the Bayesian Markov Chain Monte Carlo (MCMC) methods. The form of $L(\rho)$ motivates the name that Chernozhukov and Hong (2003) use for their estimators, which is "Laplace type estimators":

$$L(\rho) = e^{-n s_n(\rho)} \tag{56}$$

One of the advantages of using "Bayesian-like" methodology is that we can impose a prior, $\pi(\rho, \psi)$ on the models parameters, ρ and/or some functionals of these parameters ψ. In this particular work we impose the positivity restriction on the stochastic discount factor as a prior.

Next, the standard Metropolis-Hastings MCMC algorithm is used to construct a Markov Chain of parameters, ρ. The algorithm consists of the following steps.

1. A researcher comes up with a proposal density, $q(\rho_{new}|\rho_{old})$, which, among other things, should be easy to simulate from.

2. A candidate for the next value in the chain, ρ_{new} is drawn from a proposal density $q(\rho_{new}|\rho_{old})$.

3. Using the candidate value, ρ_{new} the data simulation of length N is obtained and the objective function, $s_n(\rho_{new})$, the functional of the parameters, ψ_{new}, the prior, $\pi(\rho_{new}, \rho_{new})$, and the likelihood, $L(\rho_{new}) = e^{-ns_n(\rho_{new})}$ are computed.

4. The chain moves to ρ_{new} with the probability

$$\min\{\frac{L(\rho_{new})\pi(\rho_{new}, \psi_{new})q(\rho_{new}|\rho_{old})}{L(\rho_{old})\pi(\rho_{old}, \psi_{old})q(\rho_{old}|\rho_{new})}, 1\} \tag{57}$$

By repeating the steps 2-4, we simulate a Markov Chain of the parameters of interest,

$$\{\rho^{(1)}, ..., \rho^{(N_{ch})}\}.$$

where N_{ch} is the number of simulations in the chain.

Intuitively, if the chain is constructed properly, the vector of the parameters, ρ, will visit all the parts of its support, i.e. will "mix", with the relevant frequencies. These frequencies are determined by the marginal distribution of the resulting Markov Chain. The marginal distribution is approximately equal to what Chernozhukov and Hong (2003) refer to as the quasi-posterior distribution. The quasi-posterior distribution, in turn, is proportional to the product of the likelihood, $L(\rho)$ and the prior, $\pi(\rho, \psi)$.

One of the possible Laplace type estimators is the mean with respect to quasi-posterior distribution, which in sample is equal to

$$\hat{\rho} = \frac{1}{N_{ch}} \sum_{i=1}^{N_{ch}} \rho^{(i)} \tag{58}$$

Another estimator is the mode of the Markov Chain.

Chernozhukov and Hong (2003) demonstrate that under the appropriate regularity conditions one of the ways to construct the confidence intervals for the parameter estimates is to use the quantiles of the quasi-posterior distribution, i.e. the quantiles of a sequence that forms the Markov Chain.

Once we obtain the estimate $\hat{\rho}$, we use the traditional EMM model adequacy diagnostic tools, which are based on the fact that

$$ns_n(\hat{\rho}) \sim \chi^2_{dim(\rho)-dim(\theta)}. \tag{59}$$

One can not is not able to reject the null hypothesis that the model generated the data, if the χ^2 criterion is small enough.

An additional diagnostic tool is a reprojection, discussed in detail in Gallant and Tauchen (1998). Briefly and in application to our problem, the methodology consists of comparing conditional SNP density estimated using the data simulated from the estimated structural/main model, and the conditional SNP density estimated (projected) using the observed data. In the next section, we apply the described estimation methodology to the data.

6 Results and Discussion

6.1 Data and Auxiliary Models

The data set we utilize is the same data set used in Ahn et al. (2003) and combines the data sets of McCulloch and Kwon (1993) and Daniel Waggoner from the Federal Reserve Bank of Atlanta[17]. The data set contains the zero-coupon yields of Treasury bills and bonds, and covers the period of January 1952 through December 1999. In the subsequent subsections we perform estimation using both univariate data on the yields with the time to maturity of 6 months and the multivariate data on the yields with the time to maturity of 6 months, 3 years and 10 years. The data is displayed in figure 1.

For the purposes of univariate analysis we select and utilize the SNP model that incorporates semi-nonparametric VAR, GARCH, "level in volatility" and non-linearity effects:[18] $(L_u = 1, L_g = 1, L_r = 1, L_v = 0, L_w = 1, L_p = 1, K_z = 4, I_z = 0, K_x = 0, I_x = 0)$. For the purposes of joint estimation of dynamics of three yields, we select and utilize the similar multivariate score generator: $(L_u = 1, L_g = 1, L_r = 1, L_v = 0, L_w = 0, L_p = 1, K_z = 4, I_z = 0, K_x = 0, I_x = 0)$.

[17]This part of sample is obtained using the methods described in Bliss (1997)

[18]The previous section describes in more detail how to decode SNP specification. We estimate the SNP specification using the methodology, the detailed implementation o f which is discussed in Gallant and Tauchen (2004-b).

6.2 Estimation Results for Models with One Underlying Gaussian Factors Using Univariate Data.

In this subsection, we present the estimation results for the model in which the price deflator is a semi-nonparametric function of a single Gaussian factor. These results are obtained using the time-series yields data for the 6-months-to-maturity bond yields from the data set used in Ahn, Dittmar, Gallant and Gao (2003). The existence of a vast body of literature that studies the time-series dynamics of the short-term interest rate suggests that such an exercise is interesting in its own right. Also, the estimates received using time-series data for just one yield may serve as reasonable starting values for the estimation of joint time-series dynamics of several yields. The objective of this empirical exercise is to to establish whether the data can be explained in a flexible diffusion framework. Table 1 presents the estimation results of the models in which a single Gaussian factor, x_t

$$dx_t = \kappa(\theta - x_t)dt + \sigma dW_t$$

is underlying the dynamics of the price deflator, which in turn is modeled using various numbers of Hermite polynomials of x_t. For example, the model $H(n)$ includes the first n Hermite polynomials. The observational equation for a yield of a zero-coupon bond that matures at time T, $y_t(T)$, implied by the $H(n)$ model is as follows:

$$y_t(T) = -\frac{1}{T-t}\{ln(\sum_{i=0}^{n} a_i e^{-\kappa i(T-t)} H_i(x_t)) - ln(\sum_{i=0}^{n} a_i H_i(x_t))\} \tag{60}$$

The vector of estimated parameters $\{\kappa, \theta, \sigma, a_0, ..., a_n\}$ consists of the parameters of the stochastic process of the underlying state variable and the weights of the corresponding Hermite polynomial expansion. Table 1 contains the results for $n = 2, 3, 4$. The weights a_0 of the Hermite polynomials of the 0th order and a_1 of the Hermite polynomial of the 1st order are set equal to 1 for the econometric identification purposes.

In addition, we fit the CIR process

$$dx_t = \kappa(\theta - x_t)dt + \sigma dW_t$$

as a benchmark directly to the the data in the tradition of the short-term risk-free interest rate literature.

We evaluate the models' goodness of fit using a χ^2 criterion and a corresponding p-value. The first column represents the estimation results for the CIR process. The p-value is 0.003, which means that the data rejects the model at the 1% significance level. The subsequent columns represent the estimation results for the models containing various numbers of Hermite polynomials.

As expected the fit improves as we add semi-nonparametric terms in the form of Hermite polynomials. However, the number of degrees of freedom decreases as we add parameters. We use the corresponding p-values to decide whether the models are rejected. The χ^2 statistic for the model with two Hermite polynomials is equal to 21.7 and the corresponding p-value of 0.001 suggests that the model is rejected by the data even at the 1% significance level. The p-value for the model with three Hermite polynomials is 0.076 and, thus, the data is not able to reject this model at the 7% significance level. Adding the fourth order Hermite polynomial results in an increase of the p-value to 0.12. Thus, the model containing the first four Hermite polynomials can not be rejected by the data at any conventional significance level.

Next, we present some additional diagnostics. Table 2 represents t-ratios of EMM scores produced by the best fitting model. The EMM scores demonstrate the goodness of fit of various empirical features of the data as summarized by the derivatives of the likelihood of the auxiliary model (SNP). All the presented t-ratios are smaller than two in magnitude suggesting that the model with four Hermite polynomials is able to accommodate relevant empirical features of the data, such as non-linearities, volatility persistence, "level in volatility" effect and so on.

Figures 2, 3 and 4 provide some additional insights into our application of Markov Chain Monte Carlo estimation technology on the example of estimation of a model with one Gaussian factor and two Hermite polynomials. Figure 2 presents the chains of the model parameters. The last panel is the chain of the values that the objective function takes for each of the combination of model parameters. The presented results are obtained from the restarted run, where starting values were already "good". Normally, with less reasonable starting values and before the chain stabilizes, one initially observes hill-climbing of the objective function. Figure 3 presents the sample autocorrelations of the parameters obtained from the chain. A good mixing chain is expected to be only mildly serially correlated. Figure 4 contains kernel density estimates of the model parameters from the chain. Ideally, we are interested in the shape to be close to that of the normal density. All these criteria are probably more important in the Bayesian framework when inferences are made directly from the obtained chain. We, on the other hand, use the MCMC chain as the global optimizing mechanism in the spirit of simulated annealing method.

Figures 5 and 6 contain reprojection results and compare the first and the second reprojected (estimated imposing parametric restrictions implied by the model) moments with the projected (unrestricted) ones. The first moments appear to be very close. The second moments are more distinct, but are still reasonably close. This is an additional evidence that we obtained a reasonably good fit of the moment conditions.

The presented results support a statement that it is possible to fit the time-series dynamics of the 6-months to maturity yields at conventional significance levels in a diffusion framework that is flexible enough, but not overly parameterized. This is interesting in the light of the fact that we use a history of the single bond yield to concentrate on the data's time-series properties, which could potentially be alternatively explained by various types of discontinuities such as jumps and others.

6.3 Estimation Results for Gaussian Factor Models Using Multivariate Data.

In this section we discuss estimation results obtained with multivariate bond yield data of maturities 0.5, 3 and 10 years from the data set. Table 3 contains estimation results for semi-nonparametric models building on one Gaussian factor and for a benchmark model (CIR). Results suggest that one-factor Gaussian models with several Hermite polynomials do fit the moment conditions better than the benchmark models, but are apparently not flexible enough. Chi-squared statistics are very high and all the considered models are rejected.

In table 4 we present the results of estimation of the models where the price deflator is a linear combination of multivariate Hermite polynomials of three independent Gaussian factors. Two versions of each model are presented - full and diagonal. Full models include all the interactions among Hermite polynomial of different state variables. Diagonal models do not include any interaction terms. Introduction of additional factors do improve the overall fit of moment conditions dramatically when compared to one factor models. However, all of the considered models are rejected. The p-value are all zero. Therefore, we examine z-values to obtain a better idea about the extent of each model's overall fit of moment conditions. The best fit is provided by full version of the model with three Hermite polynomials, with a z-statistic being around 47.5.

Table 5 contains additional EMM diagnostics in the form of derivatives of the likelihood of the auxiliary model. Comparing the t-ratios produced by diagonal version of the model with two multivariate Hermite polynomials (the least successful overall fit), with those produced by the full version of the model with three Hermite multivariate polynomials (the best overall fit), we can see that the latter model does fit most of moment conditions better. In particular, some of the moment conditions associated with the GARCH (persistence in volatility) features of the data become insignificantly different from zero. However, a significant part of the moment conditions remain to be nonzero.

One of the possible explanations for the insufficient fit of moment conditions is the assumption we made about the statistical properties of the underlying factors, namely that they are Gaussian and independent. The empirical studies up to date have documented that these assumptions are too restrictive[19]. The importance of having flexible correlation structure among the underlying factors was discussed in Dai and Singleton (2000) and Ahn et al. (2001). Apparently, even the introduction of a several Hermite polynomials is not able to save a model with independent Gaussian factors. Insufficient number of terms in the polynomial expansion may be another reason, although to us it seems to be less likely.

[19] Although Ahn et al. (2001) did have some success with Gaussian quadratic model. However, the factors were not independent in their work.

7 Conclusion

In this study, we model the price deflator, and thus the pricing kernel, in a flexible semi-nonparametric diffusion framework, which can be made arbitrarily sophisticated. It is crucial to be able to extract the relevant information about the pricing kernel contained in the bond prices, because it can be used later to either price other securities, e.g. interest-rate derivatives, or study investors' preferences towards risk. The expectation properties of the eigenfunctions of the infinitesimal generator, which we are using as the semi-nonparametric terms, enable us to obtain the closed-form bond prices. In addition, any asset pricing model has to procure the "no-arbitrage" condition. We offer and discuss two different ways of imposing this restriction in our framework.

An important empirical question is to what extent a sufficiently flexible diffusion framework can describe the yield data which is characterized by certain empirical features that may be arguably to jump-diffusion and regime-shift data-generating processes. One pragmatic argument is that diffusion models have better hedging abilities than models with some kind of "discontinuities", for example jumps. This question is relevant both in the context of the short-term risk free interest rate literature and in the context of the dynamic term structure of interest rates literature. The empirical part of this paper follows the tradition of both literatures.

We estimated the model with the underlying Gaussian factors using both univariate and multivariate data for bond yields. While estimating the time-series dynamics of a single yield, we employ the obtained bond pricing formula to control for the fact that the time to maturity is not equal to an instant. As a consequence, in our framework the short-term risk-free interest rate remains truly unobservable and instantaneous. The presented results suggest that the model with a single Gaussian factor and a sufficient, but not excessive, number of semi-nonparametric terms can not be rejected at the conventional significance levels by the data on 6-months-to-maturity treasury yields.

However, the joint data of yields of three different maturities convincingly rejected the one Gaussian factor model prompting us to enrich the model with additional underlying factors. We build our model on three independent Gaussian factors. We estimate a set of semi-nonparametric models with up to three multivariate Hermite polynomials. The data rejects even the most comprehensive model we consider. The explanation appears to lie in the nature of our assumptions about the underlying factors, namely that they are Gaussian and independent.

There are several avenue along which this work can be generalized and extended. The first one is to estimate the model with non-Gaussian underlying factors. The second avenue is to relax the independence assumption on the underlying factors. Finally, filtering out the unobserved stochastic discount factor would produce interesting insights into the nature of investor risk preferences and their evolution over time.

References

Ahn, D.-H., Dittmar, R. F., Gallant, A.R., 2001. Quadratic term structure models: Theory and evidence. Review of Financial Studies 15, 243-288.

Ahn, D.-H., Dittmar, R. F., Gallant, A.R., Gao, B., 2003. Purebred or hybrid?: Reproducing the volatility in term structure dynamics. Journal of Econometrics 116, 147-180.

Aït-Sahalia, Y., 1996. Testing continuous-time models of the spot interest rate. Review of Financial Studies 9, 385-426.

Bansal, R., Hseih, D.A., Viswanathan, S., 1993. A new approach to international arbitrage pricing. Journal of Finance 48, 1719-1747.

Bansal, R., Viswanathan, S., 1993. No arbitrage and arbitrage pricing: a new approach. Journal of Finance 48, 1231-1262.

Bansal, R., Zhou H., 2002. Term structure of interest rates with regime shifts. Journal of Finance 57, 1997-2043.

Bansal, R., Tauchen, G., Zhou H., 2004. Regime shifts, risk premium in the term structure, and the business cycle. Journal of Business and Economic Statistics 22, 396-409.

Bliss, R., 1997. Testing term structure estimation methods. Advances in Futures and Options Research 9, 197 - 231.

Brandt, M., Yaron, A., 2003. Time-consistent no-arbitrage models of the term structure. Duke University, working paper.

Chen, X., Hansen, L.P., Scheinkman, J.A., 2000. Nonlinearity and temporal dependence. University of Chicago, manuscript.

Chapman, D. A., 1997. Approximating the asset pricing kernel. Journal of Finance 52, 1383-1410.

Chapman, D. A., Long, J., Pearson, N.D., 1997. Using proxies for the short rate: when are three months like an instant?", Review of Financial Studies, 12, 763-806.

Chapman, D. A., Pearson, N.D., 2001. What can be learned in recent advances in estimating models of the term structure? Financial Analysts Journal, July/August, 77-95.

Chernov, M., 2003. Empirical reverse engineering of the pricing kernel. Journal of Econometrics 116, 329-364.

Chernov, M., Gallant, R. A., Ghysels, E., Tauchen, G. Alternative models for stock price dynamics 116, 225-257.

Chernozhukov, V., Hong, H., 2003. An MCMC approach to classical estimation. Journal of Econometrics 115, 293-346.

Cochrane, J., 2001. Asset Pricing. Princeton, N.J. : Princeton University Press.

Dai, Q., Singleton, K. J., 2000. Specification Analysis of affine term structure models. Journal of Finance 55, 1943-1978.

Dai, Q., Singleton, K. J., 2003. Term structure dynamics in theory and reality. Review of Financial Studies 16, 631-678.

Dai, Q., Singleton, K. J., and Yang, W., 2006. Regime Shifts in a Dynamic Term Structure Model of the U.S. Treasury Bond Yields. Review of Financial Studies. Forthcoming.

Davydov D., Linetsky V., 2003. Pricing options on scalar diffusions: an eigenfunction expansion approach. Operations Research 51, 185-209

Dittmar, R. F., 2002. Nonlinear pricing kernels, kurtosis preference, and the crosssection of equity returns. Journal of Finance 57, 369-403.

Duffie, D., 2001. Dynamic asset pricing theory. Princeton, NJ: Princeton University Press.

Duffie, D., Singleton, K., 1993. Simulated Moments Estimation of Markov Models of Asset Prices. Econometrica, 61, 929-952.

Durham, G.B., 2003. Likelihood - based specification analysis of continuous time models of the short-term interest rate. Journal of Financial Economics, 70, 463-487.

Eraker, B., 2001. MCMC analysis of diffusion models with application to finance. Journal of Business and Economic Statistics, 19-2, 177-191.

Florens, J.P., Renault, E., Touzi, N., 1998. Testing for embeddability by stationary reversible continuous-time Markov processes. Econometric Theory, 14, 744-769.

Gallant, A.R., Long, J.R., 1997, Estimating stochastic differential equations efficiently by minimum chi-squared, Biometrika 84, 125-141.

Gallant, A.R. Nychka, D.W., 1987, Seminonparametric Maximum Likelihood Estimation, Econometrica 55, 363 - 390.

Gallant, A.G., Tauchen, G., 1989, Seminonparametric Estimation of Conditionally Constrained Heterogeneous Processes: Asset Pricing Applications, Econometrica 57, 1091-1120.

Gallant, A.R., Tauchen, G., 1996. Which moments to match? Econometric Theory 12, 657-681.

Gallant, A.R., Tauchen, G., 1998. Reprojecting partially observed systems with application to interest rate diffusions. Journal of American Statistical Association 93, 10-24.

Gallant, A.R., Tauchen, G., 2004-a. EMM: A program for efficient method of moments estimation; version 2.0; user guide. Duke University.

Gallant, A.R., Tauchen, G., 2004-b. SNP: A program for nonparametric time series analysis; version 2.0; user guide. Duke University.

Goldstein, R., Keirstead, W. P., 1997. On the term structure of interest rates in the presence of reflecting and absorbing boundaries. Ohio State University, working paper.

Gorovoi, V., Linetsky, V. 2003. Black's model of interest rates as options, area functionals, eigenfunction expansions, and Japanese interest rates. Mathematical Finance, forthcoming.

Gourieroux,C. Monfort,A., Renault, E., 1993. Indirect inference. Journal of Applied Econometrics 8, 85-118.

Hansen, L.P., Richard, S., 1987. The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models. Econometrica 55, 587-614.

Hansen, L.P., Scheinkman, J.A.,1995. Back to the future: generating moment implications for continuous-time Markov processes. Econometrica 63, 767-804.

Hansen, L.P., Scheinkman, J.A., Touzi, N., 1998. Spectral methods for identifying scalar diffusions. Journal of Econometrics, 86, 1-32.

Harrison, M., Kreps, D. 1979. Martingales and arbitrage in multiperiod securities market. Journal of Economic Theory 20, 381-408.

Jones, C., 2003. The dynamics of stochastic volatility: evidence from underlying and options markets. Journal of Econometrics 116, 181224.

Karlin, S., Taylor, H.M., 1981. A second course in stochastic processes. New York, NY: Academic Press.

Lewis, A., 1998. Applications of eigenfunction expansions in continuous-time finance. Mathematical Finance 8, 349-383.

Linetsky, V., 2002, The Spectral Decomposition of the Option Value. Northwestern University, working paper.

McCulloch, J., Kwon, H., 1993. U.S. term structure data, 1947-1991. Ohio State University, working paper.

Meddahi, N., 2001-a. An eigenfunction approach for volatility modeling. CIRANO Working Paper, 2001s-70.

Meddahi, N., 2001-b. Moments of Continuous Time Stochastic Volatility Models. University de Montreal, working paper.

Pan, J., 2002. The jump-risk premia implicit in options: evidence from an integrated time-series study. Journal of Financial Economics 63, 3-50.

Rogers, L.C.G., 1997. The potential approach to the term structure of interest rates and foreign exchange rates. Mathematical Finance 7, 157-176.

Rogers, L.C.G., Zane, O., 1997. Fitting potential models to interest rates and foreign exchange rates. Vasicek and beyond, ed. L. P. Hughston, RISK Publications, London, 327-342.

Rosenberg, J., Engle, R., 2002. Empirical pricing kernels. Journal of Financial Economics 64, 341-372.

Vasicek, O., 1977. An equilibrium characterization of the term structure. Journal of Financial Economics 5, 177188.

Zhou, H., 2003. Ito conditional moment generator and the estimation of short rate processes. Journal of Financial Econometrics 1, 250-271.

A Tables and Figures

Table 1: 1 Gaussian factor, Hermite polynomials, Score: 11s1s0s1s1400000

Coefficient/Model	CIR	H(2)	H(3)	H(4)
κ	0.12	0.41	0.74	0.60
	(0.07)	(0.048)	(0.05)	(0.02)
θ	5.1	0.08	0.20	-0.15
	(0.09)	(0.036)	(0.05)	(0.03)
σ	0.197	0.17	0.29	0.17
	(0.008)	(0.02)	(0.02)	(0.01)
a_0		1, fixed	1, fixed	1, fixed
a_1		1, fixed	1, fixed	1, fixed
a_2		-0.24	-0.11	-0.24
		(0.042)	(0.011)	(0.002)
a_3			0.31	0.37
			(0.02)	(0.002)
a_4				-0.12
				(0.002)
χ^2	21.3	21.7	9.97	7.3
$p - value$	0.003	0.001	0.076	0.121
df	7	6	5	4

Table 2: EMM diagnostics of the model with 4 Hermite polynomials. The table presents diagnostics for EMM scores (t-ratios). The score generating model is 11s1s0s1s1400000. Coefficients $a0[k]$ denote the weights of the Hermite polynomials in the SNP expansion, $b0$ and $B(k)$ denote VAR terms of the SNP expansion, $R0, P(k)$ and $Q(k)$) are GARCH terms of the SNP expansion, W(k) represents the "level in volatility" effect.

Coefficient/Model	H(3)	H(4)
$a0[1]$	0.82122	-0.14169
$a0[2]$	1.15741	-1.62795
$a0[3]$	0.50755	0.51216
$a0[4]$	-1.56840	-1.29474
A	0.00000	0.00000
$b0$	0.73823	0.08364
$B(1)$	-0.15197	-0.80028
$R0$	1.33657	-0.84556
$P(1)$	-2.16230	1.12631
$Q(1)$	-2.27475	1.32500
$W(1)$	-1.36818	0.58912

Table 3: 1 Gaussian factor, Hermite polynomials, 3 yields

Coefficient/Model	CIR	H(2)	H(4)
θ	0.073	0 (fixed)	0 (fixed)
κ	0.132	0.0247	0.0248
σ	0.033		
λ	0.033		
a_0		0.958	0.975
a_1		0.676	0.691
a_2		0.226	0.227
a_3			-1.297*
a_4			0.229*
χ^2	6653	2669	2257
z	892	314	273
df	36	35	33

$^* \times 10^{-4}$

Table 4: 3 Gaussian factor, Hermite polynomials, 3 yields

Coefficient/Model	H(2)	H(3)	H(2)d	H(3)d
$\theta_{1,2,3}$	0 (fixed)	0 (fixed)	0 (fixed)	0 (fixed)
κ_1	0.516	0.418	0.546	0.652
κ_2	3.4	2.551	1.308	3.07
κ_3	0.042	0.04	0.046	0.043
a_{000}	0.247	0.238	0.28	0.231
a_{001}	0.031	-0.08	-0.02	0.017
a_{002}	0.094	0.094	0.09	0.093
a_{003}		0.072*		0.012
a_{010}	0.049*	0.014*	0.038*	-0.915*
a_{011}	0.19*	-0.188*		
a_{012}		-0.0046*		
a_{020}	-0.1095*	0.108*	-0.64*	-0.222*
a_{021}		-0.048*		
a_{030}		0.0007*		-0.786*
a_{100}	0.033	0.019	0.039	0.024
a_{101}	-0.017*	-0.235*		
a_{102}		0.172*		
a_{110}	-0.024*	-0.14*		
a_{111}		-0.157*		
a_{120}		-0.063*		
a_{200}	0.375*	0.497*	0.006	0.392*
a_{201}		-0.132*		
a_{210}		-0.112*		
a_{300}		0.151*		0.107*
χ^2	373.41	284.77	552.32	391.79
z	48.59	47.51	68.72	50.73
df	26	16	29	26

$^* \times 10^{-2}$

Table 5: EMM diagnostics of the Models with 3 Gaussian factors, and Hermite polynomials, 3 yields. The models H2d and H2 are diagonal and full versions, respectively, of the model with two Hermite Polynomials. The models H3d and H3 are diagonal and full versions, respectively, of the model with three Hermite Polynomials. The coefficients are as follows: a() coefficients of polynomial expansion; b[] and B() are VAR terms in conditional mean; the rest of the coefficients are ARCH, GARCH, "level-in-volatility" and leverage terms in conditional volatility

t-ratio/model	H2d	H3d	H2	H3
a(001)	1.14561	1.28345	-1.76703	-0.7603
a(002)	-1.91892	-1.27282	-2.17165	-0.80028
a(003)	-0.74574	1.25533	0.14716	0.98773
a(004)	1.94109	1.21672	-1.08947	-0.23999
a(010)	-3.48619	-2.66802	-2.52494	-2.31814
a(020)	1.54831	1.08003	1.53658	1.44732
a(030)	1.78861	1.35591	1.52904	0.09513
a(040)	-0.95856	-1.0055	-2.18202	-1.07637
a(100)	-4.19297	-4.34541	-4.11906	-4.23122
a(200)	-0.45423	-0.27014	-0.75011	-0.26519
a(300)	-0.45887	-0.66277	-0.77003	-0.47026
a(400)	-1.91375	-1.41115	-1.7247	-1.34642
A(1,1)	0	0	0	0
b0[1]	-0.08893	0.10631	-0.19525	-0.32299
b0[2]	-2.89369	-2.2744	-1.94503	-1.31902
b0[3]	-0.49583	-0.39078	-2.47727	-1.53362
B(1,1)	0.31226	0.3931	0.45278	0.65301
B(2,1)	-0.27096	-0.61131	0.7182	1.44031
B(3,1)	-2.40572	-0.73565	0.6262	1.39749
B(1,2)	2.62368	2.67894	2.73346	2.94052
B(2,2)	5.35926	4.31353	4.10612	2.58745
B(3,2)	-1.51983	0.81842	0.53847	1.84353
B(1,3)	1.00841	0.95267	0.93374	0.78887
B(2,3)	2.36126	2.6398	2.44285	1.62182
B(3,3)	3.90336	3.50049	2.26386	3.82997
R0[1]	0.13086	-0.016	-0.19718	-0.0434
R0[2]	-0.83531	-1.74079	-1.06368	-0.94096
R0[3]	-0.21029	-1.21865	-0.05404	0.6307
R0[4]	0.65376	0.3513	0.9389	0.32103
R0[5]	-0.35001	-0.91603	-2.33278	-1.52294
R0[6]	-7.17194	-4.60828	-3.72357	-2.47679
P(1,1)	-1.29189	-0.43725	-0.6631	0.64026
P(2,1)	5.28172	4.45886	4.03658	3.02218
P(3,1)	-0.00097	0.19557	-0.70846	-0.3969
Q(1,1)	-0.79173	-0.09058	-0.47846	0.78592
Q(2,1)	4.52204	3.62819	4.08072	3.0646
Q(3,1)	-3.07584	-1.36811	-1.42938	-1.31739
W(1,1)	-3.27865	-1.38764	-3.11358	-2.65223
W(2,1)	4.94352	4.48599	5.1826	3.41861
W(3,1)	4.07728	2.71037	3.28695	2.5014

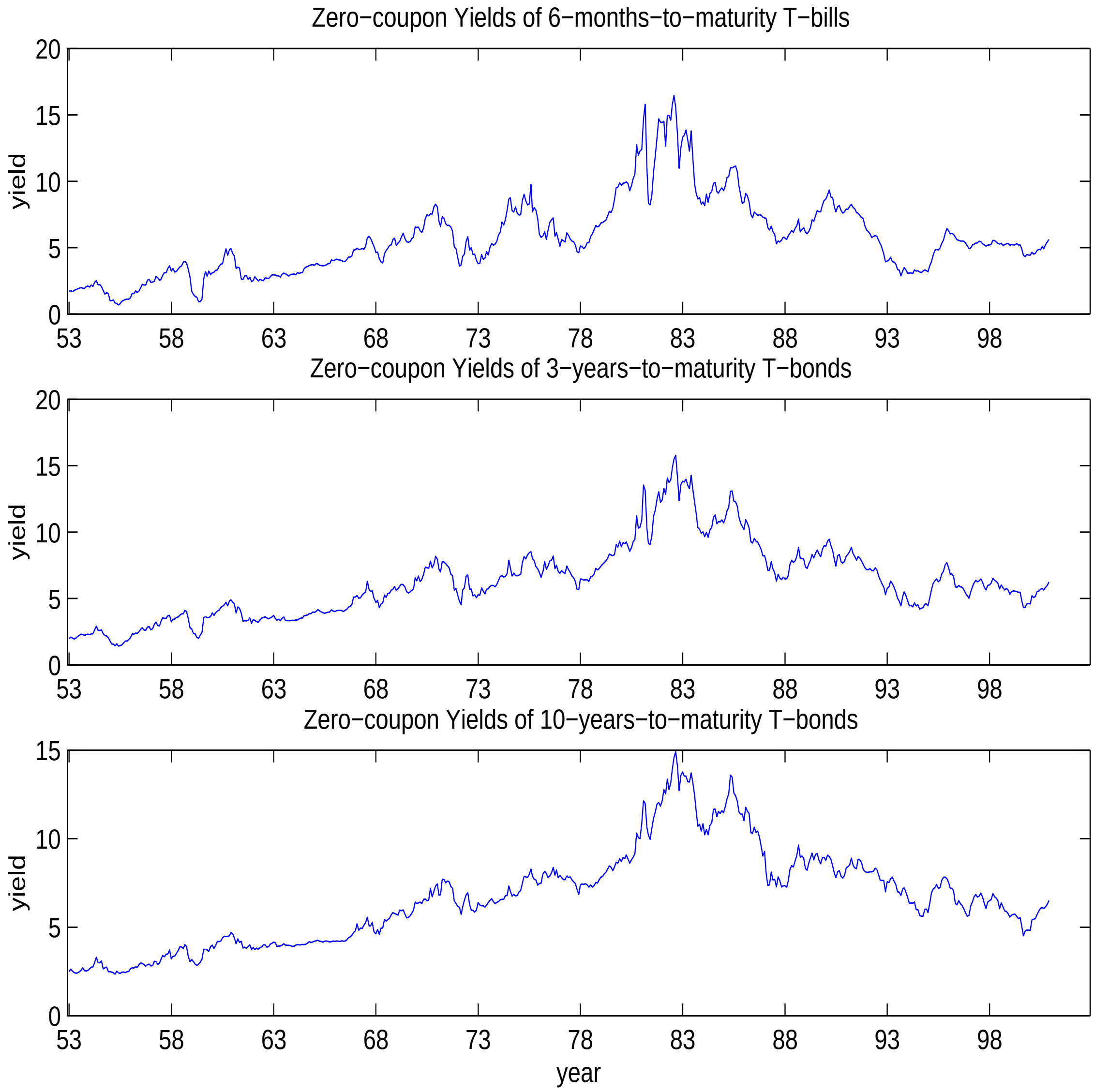

Figure 1: Data Sample of Zero-coupon Yields

.

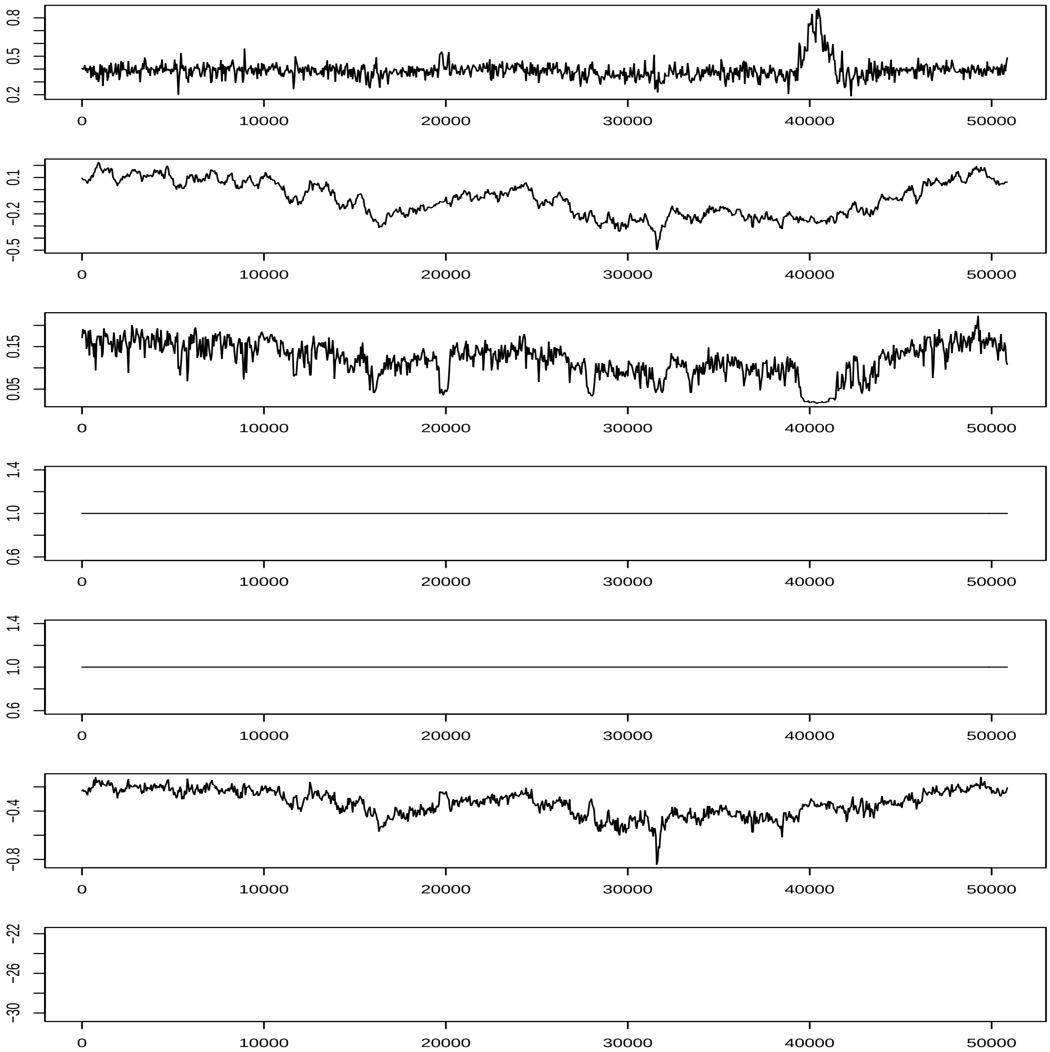

Figure 2: Markov Chain for each of the parameters, $\{\kappa, \theta, \sigma, a_0, ..., a_2\}$ of "1 Gausssian factor - 2 Hermite polynomials" model, estimated using 11s1s0s1s1400000 SNP score. Note: the first two parameters of Hermite expansion, a_0 and a_1 are fixed for the identification purposes. The last panel of the graph is the chain of the values that the objective function takes. Every 50th point is plotted.

.

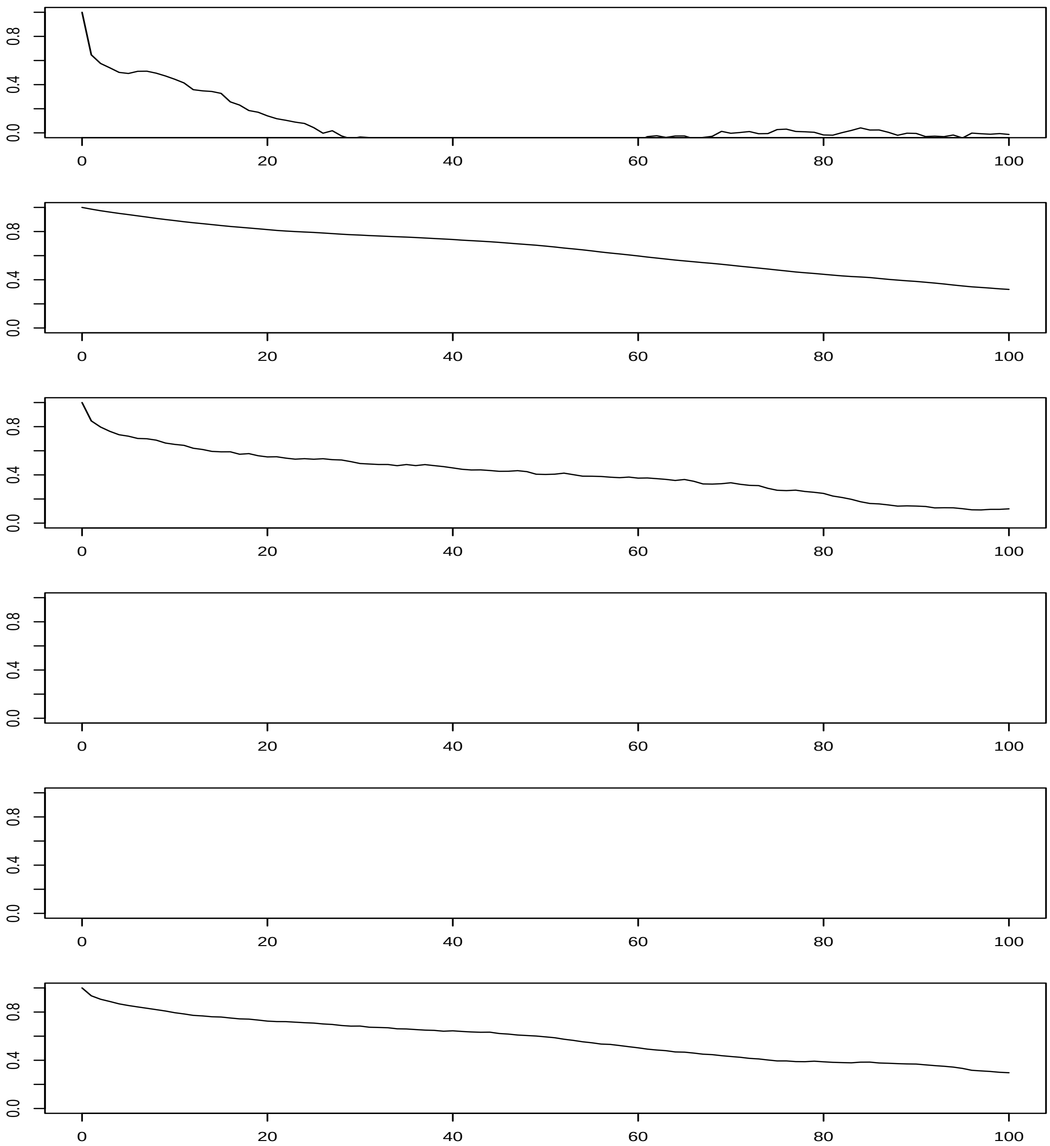

Figure 3: Autocorrelation function for parameters, $\{\kappa, \theta, \sigma, a_0, ..., a_2\}$ of "1 Gausssian factor - 3 Hermite polynomials" model, estimated using 11s1s0s1s1400000 SNP score. Note: the first two parameters of Hermite expansion, a_0 and a_1 are fixed for the identification purposes.

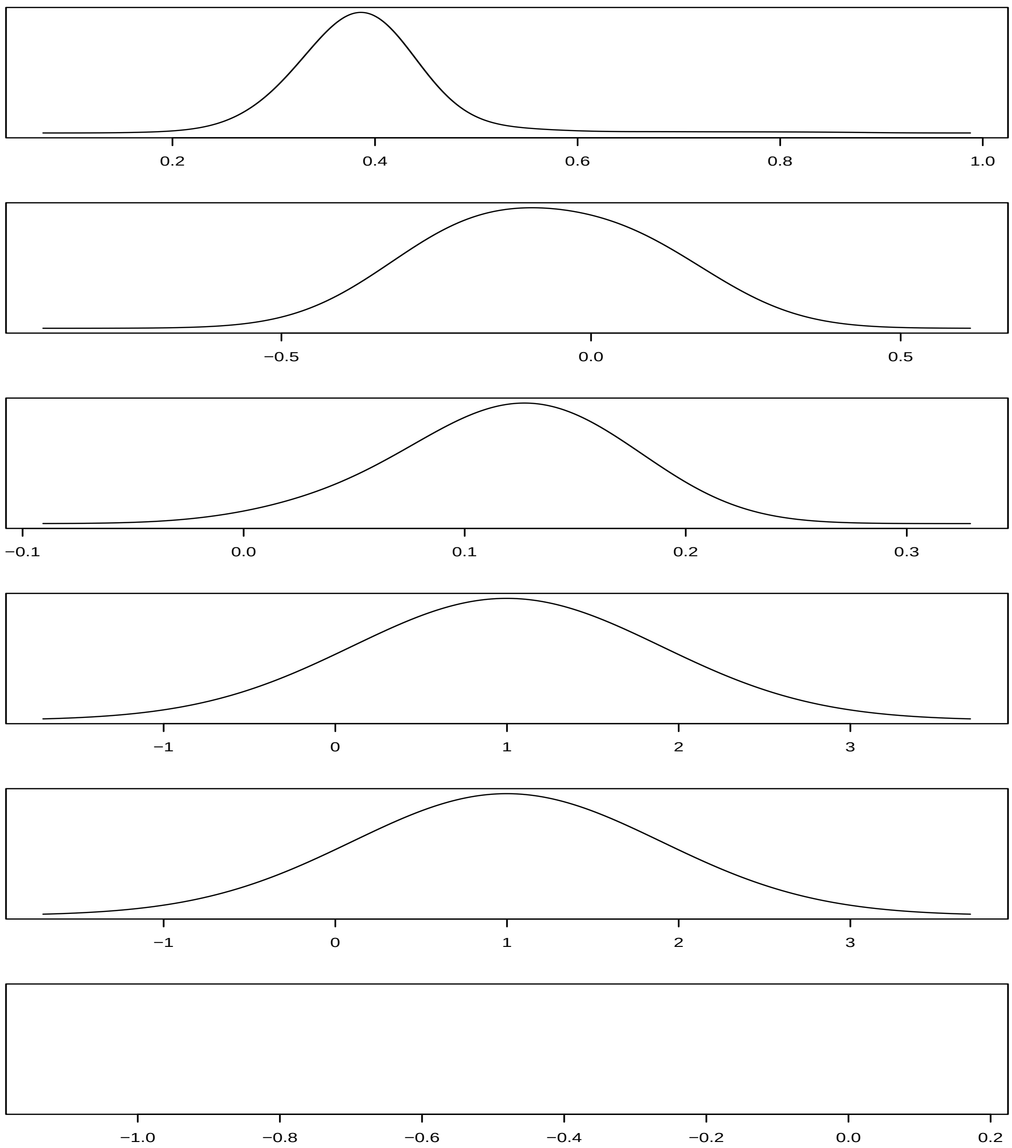

Figure 4: Kernel density estimates from chain of parameters, $\{\kappa, \theta, \sigma, a_0, ..., a_2\}$ of "1 Gausssian factor - 2 Hermite polynomials" model, estimated using 11s1s0s1s1400000 SNP score. Note: the first two parameters of Hermite expansion, a_0 and a_1 are fixed for the identification purposes.

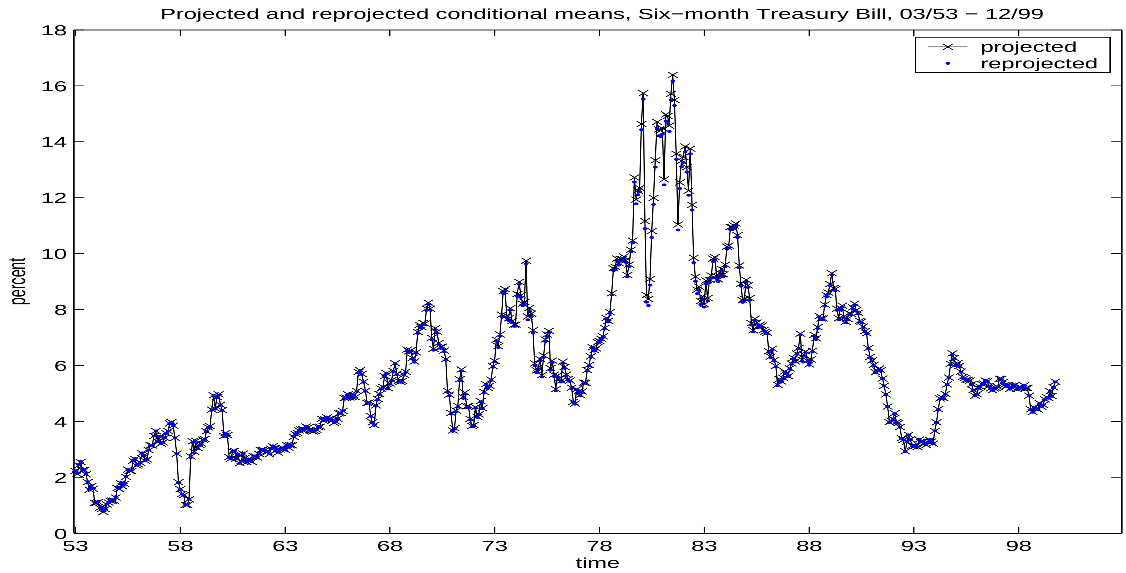

Figure 5: Conditional projected (unrestricted) and reprojected (estimated imposing parametric restrictions implied by the model) first moments; "1 Gausssian factor - 4 Hermite polynomials" model, estimated using 11s1s0s1s1400000 SNP score.

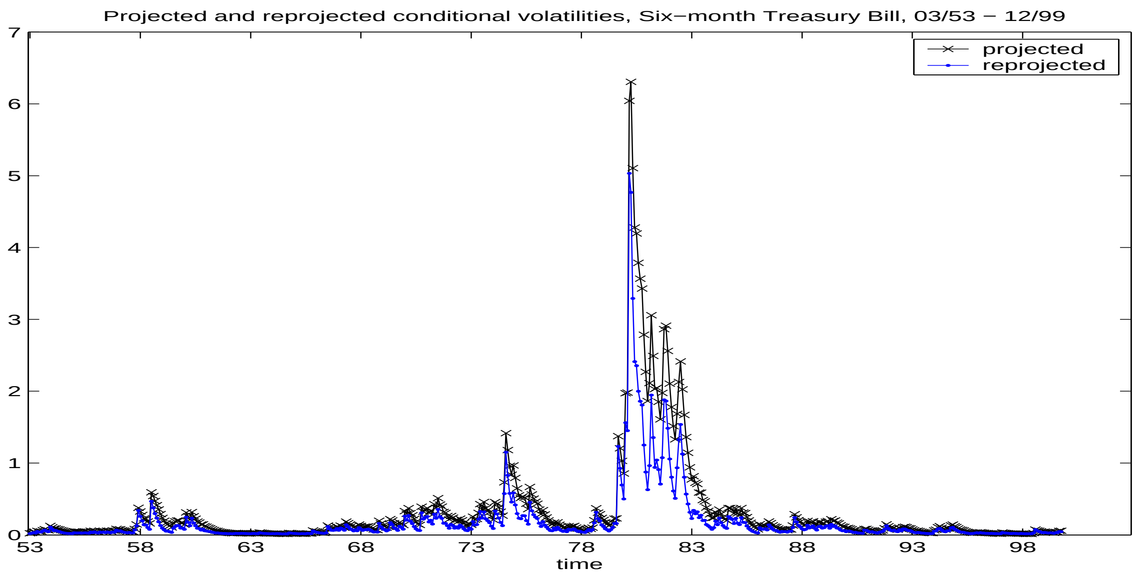

Figure 6: Conditional projected (unrestricted) and reprojected (estimated imposing parametric restrictions implied by the model) second moments; "1 Gausssian factor - 4 Hermite polynomials" model, estimated using 11s1s0s1s1400000 SNP score.