The Dynamics of the Forward Interest Rate Curve: A Formulation with State Variables*

Frank de Jong†
Tilburg University

Pedro Santa-Clara‡
University of California, Los Angeles

January 23, 1998

Abstract

This paper provides a general arbitrage-free model of interest rates in the spirit of Heath, Jarrow and Morton (1992). A characterization with an additional state variable is given, such that its joint process with the short rate is Markovian. This new state variable captures all the information in the history of interest rates that is relevant for pricing. For the models in this class, bond prices are obtained as a function of the two state variables. The Markovian character of this class of models greatly enhances their applicability for pricing of derivatives with numeric methods. Several parametric examples are given that fit stylized facts known about interest rate dynamics. A parametric version of the model is estimated in such a way that the state variables can be extracted. Finally, the empirical performance of the model is compared with the performance of traditional term structure models.

---

*Thanks are due to Mark Grinblatt, Lars Nielsen, Jesús Saa-Requejo, and Eduardo Schwartz for helpful comments and suggestions. Part of the work on this paper was performed while de Jong was visiting UCLA. This paper supersedes Santa-Clara (1995).

†Department of Econometrics, Tilburg University, PO Box 90153, 5000 LE, Tilburg, The Netherlands. Tel: +31-13-4662911. E-mail: f.dejong@kub.nl.

‡The Anderson Graduate School of Management, UCLA, Westwood Plaza, Los Angeles, CA 90095-1481. Tel: (310) 206-6077. E-mail: pedro.santa-clar@anderson.ucla.edu.
The Dynamics of the Forward Interest Rate Curve: A Formulation with State Variables*

Abstract

This paper provides a general arbitrage-free model of interest rates in the spirit of Heath, Jarrow and Morton (1992). A characterization with an additional state variable is given, such that its joint process with the short rate is Markovian. This new state variable captures all the information in the history of interest rates that is relevant for pricing. For the models in this class, bond prices are obtained as a function of the two state variables. The Markovian character of this class of models greatly enhances their applicability for pricing of derivatives with numeric methods. Several parametric examples are given that fit stylized facts known about interest rate dynamics. A parametric version of the model is estimated in such a way that the state variables can be extracted. Finally, the empirical performance of the model is compared with the performance of traditional term structure models.
1 Introduction

Traditional term structure models such as Cox, Ingersoll and Ross (CIR, 1985) and Vasicek (1977) take as given a short rate process and market prices of risk and then price interest rate sensitive securities, either in closed form or using numerical methods like binomial discretizations, simulations or numerical solutions of a PDE. The approach can be extended to include multiple sources of uncertainty by making the dynamics of the yield curve depend on a number of state variables that summarize the information available in the economy. In any case, the basic building blocks of these models are state variables that are Markovian by construction. It is this characteristic that makes the models tractable for pricing, by, for instance, formulating a PDE that must be satisfied by all interest rate derivatives.

There is however, an important practical drawback of this approach: the models can only be made to fit the current observed term structure by making the spot rate dynamics depend on complex functions of time.

A new approach to interest rate modelling started with Ho and Lee (1986) and Heath, Jarrow and Morton (HJM, 1992). These models take as given some initial term structures of interest rates and forward rate volatilities - which are thus automatically fitted into the model - and, by imposing absence of arbitrage opportunities, work out their implications to yield curve dynamics. The relevant parameters of the model can in general be obtained by “inverting” bond option prices, much in the same way implied volatilities are extracted from stock option prices with the Black-Scholes model.

Unfortunately, this approach usually leads to a non-Markovian spot interest rate\(^1\), which makes it difficult to obtain closed form pricing formulas in terms of a reduced set of state variables or even to use numerical pricing methods. The only models in this class known (until recently) to be Markovian have deterministic forward rate volatilities, thus leading to Gaussian interest rates - which is not very appealing.

The general non-Markovian nature of no-arbitrage models makes it impossible to obtain a PDE for pricing. The only method thus available for pricing derivatives is the construction of binomial trees. But, even with these, the possibilities are restricted. The binomial discretization method proposed by Nelson and Ramaswamy (1990) cannot be applied to non-Markovian processes, and the method of Amin and Bodurtha (1995) is computationally feasible only for a small number of time steps\(^2\), thus making it little suited for the pricing of American claims or long maturity securities.

There is therefore a tradeoff in choosing between the two approaches to model interest rates. Our model resolves this tradeoff.

We start with two building blocks: no arbitrage - that is, we recognize that the currently observed term structure\(^3\) has strong implications for the admissible dynamics of interest rates - and a set of state variables that synthesizes the information in the economy. We then obtain conditions on

---

\(^1\)This means that, in general, no finite dimensional set of state variables exists that captures the information necessary for pricing and that can be input into numerical pricing methods.

\(^2\)The binomial tree that is constructed is non-recombiant and thus the number of nodes grows exponentially with the number of time intervals in the discretization. The tree approximates well the state space but not the time line.

\(^3\)We retain the capability of fitting the initial term structure of previous no-arbitrage models.
the shape of the term structure of volatilities that lead to Markovian interest rates. The interest rates are made Markovian by expanding the state space with a particular state variable that has the role of summarizing the information in the path of the term structure. This variable and the spot interest rate are then sufficient to price bonds. For all the models in this class, bond prices are a function only of these two state variables, as opposed to depending on an infinite number of forward rates, which is the usual case in HJM. The Markovian nature of the state variables makes it possible to derive the PDE that all interest rate derivatives must satisfy and easily price derivatives by numerical methods.

It is an important characteristic of our model that the bond pricing equation does not depend on the parameters of the state variable processes. Even if we make our two jointly Markovian state variables dependent on additional state variables, the latter will not show up in bond pricing formulas. Thus, whatever the complexity of their dynamics, the current values of the two state variables are sufficient statistics of the uncertainty relevant for bond pricing.

Although the restrictions that have to be imposed on the volatility of forward rates in order to obtain Markovian interest rates are quite strong, we still obtain a very large class of possible term structure dynamics. This means that we can have very rich dynamics for bond prices while still being able to price bonds with the same (simple) formula. This makes this class of models particularly interesting for empirical work.

We estimate a parsimonious version of our model from a panel data set of bond prices. We put the model in state space form and use the Kalman filter to simultaneously estimate the parameters of the model and filter the unobservable state variables. The estimation is done by pseudo maximum likelihood, using the true conditional moments of the state variables’ processes. We compare the empirical performance of the Markovian HJM model with traditional one-factor term structure models, such as the Vasicek (1977) model and the Cox, Ingersoll and Ross (1985) model.

The paper is organized as follows. In section 2 presents the model and contains the condition for Markovian term structure dynamics and the pricing PDE. Section 3 gives several interesting parametric examples of models in our class. Section 4 presents the empirical results. Section 5 concludes the paper.

2 Markovian Arbitrage-Free Models of Forward Rates

For completeness, we first present the general Markovian Heath, Jarrow and Morton (1992) model. Then, we specialize the term structure of volatilities such that a Markovian characterization of the model is possible. Some parametric examples will be presented in section 3.

We assume that at any time \( t \) a (floating-rate) bank account and riskless discount bonds of all maturity dates \( s \geq t \) trade in this economy. Let \( P(t, s) \) denote the time \( t \) price of the \( s \) maturity bond\(^4\) and the value at time \( t \), of a unit investment at time \( 0 \) in the bank account that is continuously reinvested, be given by

\[
B(t) = B(0) \exp \left( \int_0^t r(s)ds \right),
\]

\(^4\)We require that \( P(s, s) = 1 \), that \( P(t, s) > 0 \) and that \( \partial P(t, s) / \partial s \) exists.
where \( r(t) \) is the instantaneous nominal interest rate.

The instantaneous forward rates at time \( t \) for all dates \( s > t \), \( f(t, s) \), are defined by

\[
f(t, s) = -\frac{\partial \log P(t, s)}{\partial s}
\]

which is the rate that can be contracted at time \( t \) for instantaneous riskless borrowing or lending at time \( s \). From the knowledge of the instantaneous forward rates for all maturities between time \( t \) and time \( s \), the price at time \( t \) of a bond with maturity \( s \) can be obtained by

\[
P(t, s) = \exp \left\{ -\int_t^s f(t, y) dy \right\}
\]

The spot interest rate at time \( t \), \( r(t) \), is the instantaneous forward rate at time \( t \) for date \( t \),

\[
r(t) = f(t, t)
\]

We follow HJM in describing the dynamics of the term structure of interest rates by a family of stochastic processes representing forward rate movements. For all \( s \) in \([0, T]\), let

\[
f(t, s) = f(0, s) + \int_0^t \alpha(v, s) dv + \int_0^t \sigma(v, s) \, dW(v)
\]

with \( \alpha \) and \( \sigma \) are adapted processes in \( \mathbb{R} \) and \( \mathbb{R}^N \) such that the forward rate processes are well defined as Ito processes. The dynamics of the instantaneous interest rate are then given by

\[
r(t) = f(t, t) = f(0, t) + \int_0^t \alpha(v, t) dv + \int_0^t \sigma(v, t) \, dW(v)
\]

HJM show that arbitrage-free instantaneous forward rate processes must verify a constraint on the drift, such that

\[
f(t, s) = f(0, s) + \int_0^t \sigma(v, s) \left( \int_v^s \sigma(v, y) dy \right) dv
\]

\[
- \int_0^t \phi(v) \sigma(v, s) dv + \int_0^t \sigma(v, s) \, dW(v)
\]

or, in differential form,

\[
df(t, s) = \left[ \sigma(t, s) \left( \int_t^s \sigma(t, y) dy \right) - \phi(t) \sigma(t, s) \right] dt
\]

\[
+ \sigma(t, s) \, dW(t)
\]

The instantaneous interest rate process under the no-arbitrage condition is

\[
r(t) = f(0, t) + \int_0^t \sigma(v, t) \left( \int_v^t \sigma(v, y) dy \right) dv
\]

\[
- \int_0^t \phi(v) \sigma(v, t) dv + \int_0^t \sigma(v, t) \, dW(v)
\]
or, in differential form,

\[
\begin{align*}
\frac{dr(t)}{dt} &= df(t, s)_{s=t} + \frac{\partial f(t, s)}{\partial s}_{s=t} dt \\
&= \left[ f_2(0, t) + \int_0^t \sigma_2(v, t) \left( \int_v^t \sigma(v, y) dy \right) dv \\
&\quad + \int_0^t \sigma(v, t) \sigma(v, t) dv - \int_0^t \phi(v) \sigma_2(v, t) dv \\
&\quad - \phi(t) \sigma(t, t) + \int_0^t \sigma_2(v, t) dW(v) \right] dt \\
&\quad + \sigma(t, t)'dW(t)
\end{align*}
\]

(6)

where the subscript 2 denotes a partial derivative with respect to the second argument of the function.

Note that the first term of the drift is the slope of the initial forward curve, the second and third terms depend on the history of the volatility process and the last term depends jointly on the history of the volatility process and the history of the Brownian motion. The two remaining terms depend on the market price of risk vector. It is obvious that, in general, the instantaneous interest rate process will not be Markovian.

We now investigate a Markovian characterization of the economy based on a reduced set of state variables that contain all the information relevant for the term structure of interest rates. In this way, we can obtain a variety of dynamics for interest rates that are consistent with the initial term structure and under which we are able to price bonds in closed form.

Let us assume that the volatility function of the forward rate processes is such that

\[
\sigma_2(t, s) = k(s) \sigma(t, s)
\]

where \( k \) is an arbitrary deterministic scalar function. This condition defines an ODE with fundamental solution

\[
\sigma(t, s) = \sigma(t, t) \exp\left\{ - \int_t^s k(u) du \right\}
\]

thus implying that \( \sigma(t, s) = x(t) l(s) \), for any \( N \) dimensional process \( r \) and deterministic scalar function \( l \).\(^5\) This is a special case of arbitrage-free dynamics of the spot interest rate such that the information in the path of the term structure that is relevant is contained in the state variable \( \varphi \).

Under this condition, from (5) and (6), the dynamics of the instantaneous interest rate can be written as

\[
\frac{dr(t)}{dt} = \left[ f_2(0, t) + k(t) r(t) - k(t) f(0, t) - \phi(t) \sigma(t, t) + \varphi(t) \right] dt + \sigma(t, t)'dW(t)
\]

where

\[
\varphi(t) = \int_0^t \sigma(v, t) \sigma(v, t) dv
\]

\(^5\)The condition on the volatilities of forward rates and the extension of the state space have been independently obtained by Carverhill (1994) for the case of deterministic \( \sigma \) and Ritchken and Sankarasubramanian (1995) for the case of \( x \) dependent only on the short rate \( r \). We see in the examples provided in the next section that we can obtain models that are significantly more general than these.
with dynamics given by
\[ d\varphi(t) = [\sigma(t, t)\sigma(t, t) + 2k(t)\varphi(t)]\, dt \]

Therefore, \((r, \varphi)\) are jointly Markovian if \(\sigma\) depends only on \(r\). If \(\sigma\) depends also on additional state variables, then \((r, \varphi)\) will be jointly Markovian with them.

Note that the first term of the drift of the spot rate is the slope of the initial forward rate curve, the second and third terms reflect some kind of mean reversion to the initial forward rate, then there is a term reflecting the market price of risk, and finally, there is the variable that reflects the path information. It is interesting to consider the case of a flat initial yield curve. This implies that \(f_2(0, t)\) is zero and \(f(0, t)\) is constant. If, in addition, the forward volatility curve is exponentially dampened, \(k\) is constant, and the spot rate process displays the usual mean reversion. If the forward volatility curve is flat, \(k\) is zero, and the drift of the spot rate depends only on the market price of risk and the path variable.

We can now obtain bond prices in closed form.\(^6\)

\[
P(t, s) = \frac{P(0, s)}{P(0, t)} \exp \left\{ -\int_t^s \sigma(v)\, dv \right\} \varphi(t)
\]

(7)

Although we obtain a large class of models, there are important economic implications of the Markovian restriction. The term structure of volatilities is such that \(\sigma_i(t, s)/\sigma_i(t, t) = \varphi(s)/\varphi(t)\) for any \(i = 1, \ldots, n\). In particular, for time homogeneity, there are only two possible cases. The first case has \(\varphi(s)\) constant, and thus all forward rates, and the spot rate, have the same volatility at a given time. In this case,

\[
P(t, s) = \frac{P(0, s)}{P(0, t)} \exp \left\{ -(s - t) (r(t) - f(0, t)) - \frac{(s - t)^2}{2}\varphi(t) \right\}
\]

(8)

\(^6\)Replacing \(\sigma(t, s) = x(t)\varphi(s)\) in (3) and rearranging slightly,

\[
(f(t, s) - f(0, s)) = x(s) \int_0^t x(v)\varphi(v) \left( \int_v^s \varphi(v)\, dv \right) \varphi(t) + l(s) \int_0^t x(v)\varphi(v)\, dv + l(s) \int_0^t x(v)\, dv
\]

We can thus write

\[
\frac{1}{l(s)} (f(t, s) - f(0, s)) - \int_0^t x(v)\varphi(v) \left( \int_v^s \varphi(v)\, dv \right) \varphi(t) = \frac{1}{l(t)} (f(t, t) - f(0, t)) - \int_0^t x(v)\varphi(v) \left( \int_v^s \varphi(v)\, dv \right) \varphi(t)
\]

and, making use of the fact that \(\int_v^s l(v)\, dv = \int_v^t l(v)\, dv = \int_v^t l(t)\, dv\),

\[
\frac{1}{l(s)} (f(t, s) - f(0, s)) = \frac{1}{l(t)} (r(t) - f(0, t)) + \left( \int_t^s l(v)\, dv \right) \left( \int_v^t x(v)\varphi(v)\, dv \right)
\]

Replacing \(\varphi(t) = l(t)^2 \int_0^t x(v)\varphi(v)\, dv\), we obtain

\[
f(t, s) = f(0, s) + \frac{l(s)}{l(t)} (r(t) - f(0, t)) + \frac{l(s)}{l(t)^2} \left( \int_t^s l(v)\, dv \right) \varphi(t)
\]
The second case has \( l(s) = e^{-as} \), where \( a \) is a constant, and we thus have exponential decay of volatilities along the maturity line. Here,

\[
P(t, s) = \frac{P(0, s)}{P(0, t)} \exp \left\{ -\frac{1 - e^{-a(s-t)}}{a} (r(t) - f(0, t)) - \frac{(1 - e^{-a(s-t)})^2}{2a^2} \varphi(t) \right\}
\]

(9)

There is one important case studied in the literature that is incompatible with the constraints we impose on the volatility curve. This is the case of forward rates defined by an SDE - where the volatility of a given forward rate depends on the level of the forward - described in HJM and Amin and Morton (1994). Note however that we can still make the volatility dependent on some number of forward rates appropriately defined.

3 Examples

The following are some examples of the class of models covered by our framework. Since it is what matters for pricing, we will only consider dynamics under the risk adjusted probability measure. We denote by \( Z \) the Brownian motion under this equivalent probability measure. There should be no confusion with the Brownian motion under the true probability measure.

3.1 Deterministic Volatility Models

If \( \sigma \) is nonstochastic, \( \varphi \) collapses to a deterministic function of time and the spot rate is Markovian by itself. Two homogeneous cases, with one-dimensional uncertainty, are of interest: \( \sigma(t, s) = \sigma \), a constant, and \( \sigma(t, s) = \sigma e^{-a(s-t)} \) with \( a \) and \( \sigma \) constants. The first case is the well known continuous-time version of the Ho-Lee model, which corresponds to the so called extended Vasicek model and is studied in example 1 of HJM. In this case all forward rates have the same constant volatility, and a shock shifts all rates uniformly. The second example has the volatility decreasing exponentially with maturity, thus making short rates more volatile than long rates.

Example 2 of HJM studies a mixture of the two examples above. The Brownian motion driving the uncertainty of forward rates is two-dimensional and the two components of the volatility vector of forward rates are the ones described above. Shocks to the first Brownian motion affect the ‘level’, while shocks to the second Brownian motion affect the ‘steepness’ of the forward curve.

Deterministic volatility models are the only Markovian arbitrage-free models known in the literature so far\(^7\) and, as such, have been extensively studied both for closed-form pricing.\(^8\) and in empirical studies\(^9\).

\(^7\)In fact, Hull and White (1993) have shown that deterministic volatility is a necessary and sufficient condition for the spot interest rate to be Markovian (by itself).

\(^8\)See, among many others, Jamshidian (1991), Flesaker (1993a) and Brace and Musiela (1994).

\(^9\)See Flesaker (1993b).
3.2 Constant Elasticity of Variance Models

Apart from the deterministic volatility case, which is in a sense degenerate, the most parsimonious examples of our class of models are such that, with one dimensional Brownian motion, \( \sigma(t, s; r(t)) = x(r(t), t) l(s) \).

A simple example that covers a great many models studied in the literature\(^{10}\) is \( x(r(t)) = \sigma r(t) \gamma \), for constant \( \sigma \) and \( \gamma \). When \( \gamma = 0 \) we have the Vasicek volatility, when \( \gamma = 1 \) we have versions of the Brennan and Schwartz (1977) volatility, when \( \gamma = 1/2 \) we have the CIR volatility and when \( \gamma = 3/2 \) we have Richard’s (1994) volatility. For the case of exponentially declining forward rate volatilities,

\[
dr(t) = [f_2(0, t) + a(f(0, t) - r(t)) + \varphi(t)] dt + \sigma r(t) \gamma dZ(t)
\]

where

\[
\varphi(t) = \int_0^t \sigma^2 e^{-2a(t-v)} r(v)^2 \gamma dv
\]

with dynamics given by

\[
d\varphi(t) = [\sigma^2 r(t)^2 \gamma - 2a \varphi(t)] dt
\]

Bond prices are then given by (9).

3.3 General Affine Models

A particularly tractable choice for the volatility function is an affine function of the spot rate \( x(r(t), t) = \sigma_0^2 + \sigma_1^2 r(t) \). Note that the extended Vasicek model is a special case of this model and is obtained if \( \sigma_1^2 = 0 \). Also, the CEV model with square root volatilities, \( \gamma = 1/2 \), is a special case of the affine model (\( \sigma_0^2 = 0 \)).\(^{11}\) These assumptions lead to the following dynamics for the state variables under the risk-neutral distribution

\[
dr(t) = [f_2(0, t) + a(f(0, t) - r(t)) + \varphi(t)] dt + \sqrt{\sigma_0^2 + \sigma_1^2 r(t)} dZ(t)
\]

\[
d\varphi(t) = [\sigma_0^2 + \sigma_1^2 r(t) - 2a \varphi(t)] dt
\]

Notice that both the drift and the instantaneous variance are linear in the state variables. Although there is only one source of noise, the process cannot be reduced to a one-dimensional Markov process for \( r \). The only exception is when there is no feedback from the short rate to the second state variable, i.e. when \( \sigma_1^2 = 0 \). In that case, the volatility is constant and the short rate is Gaussian, as in the Vasicek (1977) model. But in general the dynamics of this model are different from the one-factor affine models such as the CIR model, which assumes that the spot rate itself is a one-dimensional Markov process.

Again, the bond price formula is (9). Recall that this formula links the prices of bonds at time \( t \) to the current value of the state variables and the exogenous bond prices at some initial date \( t = 0 \). The dependence on the initial bond prices is a characteristic of all exogenous bond pricing

\(^{10}\)This example covers all cases studied by Chan, Karolyi, Longstaff and Sanders (1992)

\(^{11}\)In that model, pricing of some derivatives is possible in closed form since the joint distribution of \( e^{\sigma(s)} \) and \( e^{\nu(s)} = e^{\sigma_1^2 \int_0^t r(v)dv} \), for some future \( s \), is computable, see Beaglehole and Tenney (1991).
models. However, by studying the limiting process for the state variables, and given some suitable assumption on the forward curve we can make the Markovian HJM model directly comparable with traditional (CIR type) term structure model and solve for bond prices as functions of the state variables only.

In any stationary model, the yields on long bond converge to a constant. Therefore, if we consider the limiting behavior of the Markovian model we essentially can assume that the initial forward curve is flat, \( f(0, s) = \theta \). With this assumption the risk-neutral dynamics of the state variables become

\[
d \begin{pmatrix} r(t) \\ \varphi(t) \end{pmatrix} = \begin{pmatrix} a \theta \\ \sigma_0^2 \end{pmatrix} + \begin{pmatrix} -a & 1 \\ \sigma_1^2 & -2a \end{pmatrix} \begin{pmatrix} r(t) \\ \varphi(t) \end{pmatrix} dt + \begin{pmatrix} \sqrt{\sigma_0^2 + \sigma_2^2 r(t)} dZ(t) \\ 0 \end{pmatrix}
\] (10)

It follows that the model is a special case of a two-dimensional affine term structure model of the class introduced by Duffie and Kan (1996). It is a very special case in the sense that, although there are two relevant state variables, there is only one source of randomness in the model. Bond prices are given by

\[
P(t, s) = \exp \left\{ -A(s-t) - B_1(s-t)r(t) - B_2(s-t)\varphi(t) \right\}
\] (11)

where the coefficients \( A(s-t) \), \( B_1(s-t) \) and \( B_2(s-t) \) can be determined by solving the differential equations (adapted from Duffie and Kan (1996))

\[
\frac{dA(\tau)}{d\tau} = a\theta B_1(\tau) + \sigma_0^2 B_2(\tau) - \frac{1}{2} \sigma_0^2 B_1^2(\tau)
\]

\[
\frac{dB(\tau)}{d\tau} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} -a & 1 \\ \sigma_1^2 & -2a \end{pmatrix}' \begin{pmatrix} B_1(\tau) \\ B_2(\tau) \end{pmatrix} - \frac{1}{2} \begin{pmatrix} \sigma_1^2 B_1^2(\tau) \\ 0 \end{pmatrix}
\]

It can easily be verified that

\[
B_1(\tau) = \frac{1 - \exp(-a\tau)}{a}, \quad B_2(\tau) = \frac{1}{2} \left( \frac{1 - \exp(-a\tau)}{a} \right)^2
\]

and \( A(\tau) = \theta(\tau - B_1(\tau)) \) solve this system of equations. Substitution in the bond price equation (11) results in

\[
P(t, s) = \exp \left\{ -\theta(s-t) - \frac{1 - e^{-a(s-t)}}{a} (r(t) - \theta) - \frac{(1 - e^{-a(s-t)})^2}{2a^2} \varphi(t) \right\}
\] (12)

It is easily verified that this is consistent with the HJM bond price equation (9).

### 3.4 Stochastic Volatility Models

Our specification easily accommodates models with stochastic volatility. Let the short rate process be

\[
dr(t) = [f_2(0, t) + \varphi(t)] dt + \sqrt{v(t)} dZ_1(t)
\]

with stochastic volatility given by

\[
dv(t) = \theta dt + \sigma_1(r(t), v(t)) dZ_1(t) + \sigma_2(r(t), v(t)) dZ_2(t)
\]
with $Z_1$ and $Z_2$ orthogonal. We still have

$$d\varphi(t) = [v(t) - 2a\varphi(t)]dt$$

and bond pricing obtains, as before.

This example can of course be extended to any set of additional state variables, besides stochastic volatility.

### 3.5 Forward-Factor Models

In this example the additional state variables are themselves forward rates\(^{12}\). Note that we use forward rates with fixed time to maturity, instead of the fixed maturity case usually considered, in order to be able to separate the volatility function in both time arguments.

Consider now a vector of additional state variables which are forward rates with fixed time to maturity $F(t) = (f(t, t + \theta_1), f(t, t + \theta_2), \ldots, f(t, t + \theta_n))$ for constant $\theta_1, \theta_2, \ldots, \theta_n$. We can now consider forward rate volatilities of the form $\sigma(t, s, F(t)) = x(F(t), t)l(s)$. Then, the dynamics of the state variables are given, for $i = 1, 2, \ldots, n$, by

$$dr(t) = df(t, s)|_{s=t} + \left( \varphi'f(t, s)/\varphi|_{s=t} \right) dt$$

$$= \left[ f_2(0, t) + \frac{l'(t)}{l(t)}(r(t) - f(0, t)) + \varphi(t) \right] dt$$

$$+ l(t)x(F(t), t)'dZ(t)$$

and

$$df(t, t + \theta_i) = df(t, s)|_{s=t+\theta_i} + \left( \varphi'f(t, s)/\varphi|_{s=t+\theta_i} \right) dt$$

$$= \left[ f_2(0, t + \theta_i) + \frac{l'(t + \theta_i)}{l(t + \theta_i)}(f(t + \theta_i, t + \theta_i) - f(0, t + \theta_i)) \right] dt$$

$$+ l(t + \theta_i) \left( \int_t^{t+\theta_i} l(y)dy \right) x(F(t), t)'x(F(t), t) + \frac{l(t + \theta_i)^2}{l(t)^2} \varphi(t) \right] dt$$

$$+ l(t + \theta_i)x(F(t), t)'dZ(t)$$

where

$$\varphi(t) = l(t)^2 \int_0^t x(F(v), v)'x(F(v), v)dv$$

with dynamics

$$d\varphi(t) = \left[ l(t)^2 x(F(t), t)'x(F(t), t) + \frac{l'(t)}{l(t)} \varphi(t) \right] dt$$

Bond prices are still given by (7).

\(^{12}\)In the spirit of El Karoui and Lacoste (1992) and Duffie and Kan (1996).
4 Econometric Analysis

In this section we provide an empirical analysis of the Markovian HJM models. To keep matters simple we focus on one and two-factor Markovian HJM models with an affine volatility specification, as described in section 3.3. Recall that these models can be written as a special case of a multidimensional, traditional affine term structure model. Therefore, the estimation methodology that has been developed for this class of models is directly applicable to the estimation of the Markovian HJM models. In particular, we employ the Kalman filter methodology proposed in Duan and Simonato (1995) and Geyer and Pichler (1995).

Our empirical approach differs in several respects from the work of Bliss and Ritchken (1996), who analysed a similar two-state HJM model. Firstly, Bliss and Ritchken use (pooled) cross sections of bond prices to estimate the HJM model, whereas we use a panel dataset which exploits both cross sectional and time series information in a consistent way. This allows separate treatment of the empirical and the risk neutral dynamics, and identifies the market price of risk parameter. Secondly, Bliss and Ritchken estimate the state variables by ‘inverting’ the term structure at two points. The choice of these points is rather arbitrary and Bliss and Ritchken show that the parameter estimates are sensitive to this choice. Instead, we treat the state variables as unobserved time series, and use the Kalman filter to construct estimates for these series. The Kalman filter uses the full cross section of bond prices to estimate the state variables.

We also provide an explicit comparison of the HJM model with traditional models of the term structure, such as the Vasicek (1977) and Cox, Ingersoll and Ross (1985) models. Finally, we briefly study two-factor versions of the HJM and the traditional term structure models.

4.1 Econometric Method

The model we estimate is the Markovian HJM model with an affine volatility specification, described in equation (10). That equation specifies the dynamics under the risk-neutral distribution, but for the empirical model we need the dynamics of the state variables under the true probability distribution. For analytical tractability, we make the assumption that the market price of risk is proportional to the square root of the spot rate,

$$\phi(t) = \lambda \sqrt{\sigma_0^2 + \sigma_t^2 r(t)}$$

Under this assumption, the short rate process under the empirical probability distribution is

$$dr(t) = \left[ a \theta + \lambda \sigma_0^2 - (a - \lambda \sigma_t^2) r(t) + \varphi(t) \right] dt + \sqrt{\sigma_0^2 + \sigma_t^2 r(t)} dW(t)$$

where $W$ has dynamics given by

$$dW(t) = dZ(t) - \phi(t) dt$$

is a Brownian motion under the true probability measure. The bivariate process for the state variables under the empirical measure is

$$d\begin{pmatrix} r(t) \\ \varphi(t) \end{pmatrix} = \begin{pmatrix} a \theta + \lambda \sigma_0^2 \\ \sigma_0^2 \end{pmatrix} dt + \begin{pmatrix} a \theta + \lambda \sigma_0^2 \\ -2a \lambda \sigma_t^2 \end{pmatrix} \begin{pmatrix} r(t) \\ \varphi(t) \end{pmatrix} dt$$
\[ + \left( \frac{\sqrt{\sigma_0^2 + \sigma_1^2 \tau(t)} dW(t)}{0} \right) \] (13)

The differences with the risk-neutral dynamics, described in (10), are only in the intercept and the coefficient of \( r(t) \) in the first equation of this system.

For the empirical analysis it is convenient to write the model in state space form. The state space form consists of a measurement equation, which relates the bond prices to the state variables, and a transition equation which describes the discrete time dynamics of the state variables.

Due to the exponential-affine form of the bond pricing equation (12), the measurement equation is very simple. The bond yields are linear functions of the state variables

\[ y(t, t + \tau) \equiv -\ln P(t, t + \tau) / \tau = a(\tau) + b_1(\tau) r(t) + b_2(\tau) \varphi(t) + c(t, t + \tau) \] (14)

where \( a(\tau) = A(\tau) / \tau \) and \( b_1(\tau) = B_1(\tau) / \tau \). Notice that the intercept and the coefficients of the state variables depend only on the time to maturity \( \tau \). In our estimation we use multiple points on the yield curve and hence the model cannot fit all bond prices exactly. Therefore, a measurement error term \( e(t, t + \tau) \) was added to each yield. In vector notation, the measurement equation can be written

\[ y_t = A x_t + e_t \] (15)

where \( y_t \) is the vector that collects all observed interest rates, and \( x_t \) is the vector of state variables, \( x_t = (r_1, \varphi_1)' \). The coefficients \( A \) and \( B \) equal

\[ A = \begin{pmatrix} a(\tau_1) \\ \vdots \\ a(\tau_n) \end{pmatrix}, \quad B = \begin{pmatrix} b_1(\tau_1) & b_2(\tau_1) \\ \vdots & \vdots \\ b_1(\tau_n) & b_2(\tau_n) \end{pmatrix} \]

For simplicity, we assume that the error terms are both cross-sectionally and serially uncorrelated and have identical variance \( h^2 \) for yields at all maturities:

\[ \text{Var}(e_t) = h^2 I \]

The second part of the state space form is a transition equation that describes the discrete time dynamics of the state variables under the empirical probability measure. Rewriting equation (13) as

\[ dx(t) = [\psi + \Lambda x(t)] dt + S(t) dW(t) \] (16)

it can be demonstrated that the discrete values of the state variable follow a first order vector autoregression

\[ x_t = c + \Phi x_{t-1} + \eta_t \] (17)

with parameters

\[ \Phi = \exp \{ \Lambda \Delta t \}, \quad c = (\Phi - I) \Lambda^{-1} \psi \]

where \( \Delta t \) is the time interval between two consecutive observations. The variance of the innovations, \( q_t \equiv \text{Var}_t(\eta_t) \) is a complicated function of the model parameters and the state vector.\(^{13}\)

\(^{13}\) Exact expressions for the conditional variance can be found in the Appendix.
Given this transition equation, the Kalman filter can be used to make predictions (conditional on the past observations) and estimates of the unobserved state variables. In addition, the Kalman filter can be used to construct a likelihood function for the observed data. The Kalman filter starts from initial values for the state variables. In our application, we assume that the first state vector, \( \hat{x}_0 \), is equal to the unconditional mean, \( E(x_t) = -\Lambda^{-1} \psi \), and \( \hat{P}_0 \) is equal to the unconditional variance, \( \text{Var}(x_t) \). From these initial conditions, the filter iterates between the prediction equations

\[
x_t[t-1] = c + \Phi \hat{x}_{t-1}
\]

\[
P_t[t-1] = \Phi \hat{P}_{t-1} \Phi' + q_t
\]

and the updating equations

\[
\hat{x}_t = x_t[t-1] + P_t[t-1] B' F_t^{-1} \nu_t
\]

\[
\hat{P}_t = P_t[t-1] - P_t[t-1] B' F_t^{-1} B P_t[t-1]
\]

In these equations

\[
\nu_t = y_t - A - B x_t[t-1]
\]

\[
F_t = B P_t[t-1] B' + \kappa^2 I
\]

denote the prediction errors and the conditional variance of the prediction errors, respectively. If the prediction errors are normally distributed, the likelihood function immediately follows as

\[
\ln L = -\frac{nT}{2} \ln 2\pi - \frac{1}{2} \sum_{t=1}^T \left( \ln |F_t| + \nu_t' F_t^{-1} \nu_t \right)
\]

Under normality, the Maximum Likelihood estimator of the model parameters are asymptotically normal and efficient. If the conditional distribution of the prediction errors is not normal, the Kalman filter is still the best linear filter and the ML estimator has a pseudo maximum likelihood interpretation. Asymptotic standard errors are calculated by the method of White (1982).

### 4.2 Data

The data used for estimation are weekly US swap rates for 2, 3, 4, 5, 7, and 10 year maturities. The sample period is January 8, 1988 to November 16, 1996, which contains 459 weekly observations. Summary statistics of the data are provided in Table 1. Figure 1 graphs the 2-year and the 10-year rates. The interest rates decline over most of the sample period, with the exception of 1994 when there was a strong increase. The slope of the term structure is typically increasing, except for a short period in 1989, when the 2-year rate was higher than the 10-year rate. The data exhibit very strong serial correlation. The first order autocorrelation is very close to one, and even the autocorrelation at lag 30 is around 0.97. So, mean reversion in the data is extremely slow.

---

14 For a detailed exposition of the Kalman filter we refer to Harvey (1989, 1990).

15 The dimension of the vector of observations \( y_t \) is \( n \), and the number of time series observations is \( T \).

16 See Gourieroux, Monfort and Trognon (1984). This is only strictly true if the conditional variance does not depend on the unobserved state variables.
Table 1: Summary statistics of the data

<table>
<thead>
<tr>
<th>maturity</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>7</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>0.0672</td>
<td>0.0702</td>
<td>0.0726</td>
<td>0.0744</td>
<td>0.0771</td>
<td>0.0795</td>
</tr>
<tr>
<td>st.deviation</td>
<td>0.0171</td>
<td>0.0157</td>
<td>0.0147</td>
<td>0.0139</td>
<td>0.0130</td>
<td>0.0119</td>
</tr>
<tr>
<td>$\rho(1)$</td>
<td>0.9990</td>
<td>0.9990</td>
<td>0.9989</td>
<td>0.9990</td>
<td>0.9990</td>
<td>0.9990</td>
</tr>
<tr>
<td>$\rho(30)$</td>
<td>0.9666</td>
<td>0.9689</td>
<td>0.9701</td>
<td>0.9714</td>
<td>0.9720</td>
<td>0.9736</td>
</tr>
</tbody>
</table>

$\rho(k)$ is the $k^{th}$ order autocovariance

4.3 Estimation Results

Estimation results for the Markovian HJM models are reported in Table 2. In the HJM model with an affine volatility function, the estimated intercept, $\sigma_0$ was very close to zero and insignificant. Hence, the model with square root volatility is the preferred model, and we report only the results of this variance specification.

The mean reversion of the short rate is very strong. The estimate for the parameter $\alpha$ implies a ratio between the volatilities of the 10-year forward and the spot rate of about 50%, which is in accordance with casual observation. Compared with the results in Pearson and Sun (1994), who estimate a two-factor CIR model, the estimated mean reversion is somewhat stronger and the instantaneous variance is smaller. This may however reflect the difference in sample periods.

It’s instructive to consider the dynamics of the HJM model more carefully. Substituting the estimated values of the parameters, the bivariate system becomes

$$dr(t) = [0.014 + \varphi(t) - 0.16r(t)]dt + 0.055\sqrt{r(t)}dW(t)$$
$$d\varphi(t) = [-0.33\varphi(t) + 0.003r(t)]dt$$

This equation shows that the short rate is mean reverting toward a level determined partly by the current value of the second state variable, $\varphi(t)$. This second state variable is an exponentially weighted moving average of historical short rates. The decay of this moving average is very rapid, the half-life is around 2 years. The feedback from the short rate is small but important. We graph the smoothed values of the state variables in Figure 2. The gradual decline of the second state variable captures the general decrease of the interest rates over the sample period, whereas the short rate itself also tracks the short-term movements in the yield curve.

The estimated standard deviation of the measurement error is relatively low, around 21 basis points, but not negligible. Summary statistics of the residuals of the HJM model are reported in Table 3. The fit of the model is best in the middle range maturities, around 4 years, and worst at the extremes of the maturity spectrum. This is consistent with the results of Bliss and Ritchen (1996). There is no systematic bias in the fitted yield curves as the average of the residuals is close to zero for all maturities. We also calculated the first-order serial correlation coefficients of the residuals. It turned out that for the shorter rates and the long rates there is substantial residual serial correlation, with coefficients over 0.9. For the middle range maturities, the serial correlation is weaker. These numbers reflect the goodness-of-fit of the model.

---

17Pearson and Sun’s implicit estimate of the volatility parameter is 0.21.
Table 2: Estimation results one-factor models

<table>
<thead>
<tr>
<th></th>
<th>Duffie-Kan</th>
<th>Vasicek</th>
<th>CIR</th>
<th>HJM</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{a}$</td>
<td>0.1194</td>
<td>0.1191</td>
<td>0.1862</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0070)</td>
<td>(0.0070)</td>
<td>(0.0105)</td>
<td></td>
</tr>
<tr>
<td>$\hat{\mu}$</td>
<td>0.0699</td>
<td>0.0687</td>
<td>0.0654</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0037)</td>
<td>(0.0015)</td>
<td>(0.0025)</td>
<td></td>
</tr>
<tr>
<td>$a$</td>
<td>0.1193</td>
<td>0.1191</td>
<td>0.1120</td>
<td>0.1611</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.0091)</td>
</tr>
<tr>
<td>$\theta$</td>
<td>0.1015</td>
<td>0.1015</td>
<td>0.1002</td>
<td>0.0881</td>
</tr>
<tr>
<td></td>
<td>(0.0013)</td>
<td>(0.0011)</td>
<td></td>
<td>(0.0010)</td>
</tr>
<tr>
<td>$\sigma_0$</td>
<td>0.0117</td>
<td>0.0117</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0013)</td>
<td>(0.0011)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>0.0019</td>
<td></td>
<td>0.0481</td>
<td>0.0579</td>
</tr>
<tr>
<td></td>
<td>(0.0005)</td>
<td></td>
<td>(0.0037)</td>
<td>(0.0558)</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>-31.75</td>
<td>-32.85</td>
<td>-32.03</td>
<td>47.86</td>
</tr>
<tr>
<td></td>
<td>(6.15)</td>
<td>(5.32)</td>
<td>(2.55)</td>
<td>(58.28)</td>
</tr>
<tr>
<td>$100h$</td>
<td>0.2366</td>
<td>0.2366</td>
<td>0.2365</td>
<td>0.2122</td>
</tr>
<tr>
<td></td>
<td>(0.0094)</td>
<td>(0.0094)</td>
<td>(0.0095)</td>
<td>(0.0108)</td>
</tr>
<tr>
<td>$2\ln L$</td>
<td>30023.45</td>
<td>30023.48</td>
<td>30011.77</td>
<td>30510.36</td>
</tr>
</tbody>
</table>

For the Duffie-Kan, Vasicek and CIR models, $\hat{a}$ and $\hat{\mu}$ are free parameters in the estimation. The reported values of $\alpha$ and $\theta$ are the values implied by the estimates. For the HJM model, $\alpha$ and $\theta$ are free parameters in the estimation. Asymptotic standard errors, calculated by the method of White (1982), are reported in parentheses.
An informal specification test of the HJM model is performed by a regression of the observed yields on the fitted state variables. If the model is correctly specified, the regression coefficients should be close to the coefficients $B(\tau)$ generated by the model. Figure 3 shows the results of this test. The regression coefficients of the bond yields on the first state variable, the short rate, are larger than the sensitivities predicted by the model. This overestimation is of similar magnitude for all maturities. The regression coefficients of yields on the second state variable are pretty close to the sensitivities predicted by the model.

A final test of the model is performed by calculating an exponentially weighted moving average (EMWA) of the short rate, multiplied by $\sigma_1^2$, with decay parameter $2a$. This EMWA should be close to the estimated series of the second state variable, $\dot{\phi}_t$. Figure 4 shows that the two series do not quite look alike; the EMWA of the short rate increases over the first part of the sample, whereas the second factor steadily declines.

### 4.4 Comparison with Tradional Models

For benchmarking purposes, it is instructive to compare the Markovian HJM model with one-factor affine term structure models of the class proposed by Duffie and Kan (1996). In that class of models, the short rate is the only state variable, which satisfies the stochastic differential equation

$$dr(t) = a(\mu - r(t))dt + \sqrt{\sigma_0^2 + \sigma_1^2 r(t)}dZ(t)$$

where $Z(t)$ is a Brownian motion under the risk-neutral probability distribution. The solution for bond prices in this model can be found by solving a system of ordinary differential equations. The form of the pricing equation is again exponential-affine,

$$P(t, s) = \exp\{-A(s - t) - B(s - t)r(t)\}$$

It is worth mentioning that if the model is stationary ($a > 0$), the long-run forward curve implied by the model is flat, $f(t, s) = \theta$ for large $s$, where $\theta$ is a function of all other model parameters. This value is directly comparable to the long run forward rate in the Markovian HJM model.

For estimation purposes, we also need the short rate process under the true distribution. Like before, we assume that the market price of risk is proportional to the square root of the spot rate, $\phi(t) = \lambda\sqrt{\sigma_0^2 + \sigma_1^2 r(t)}$. Hence, the empirical distribution of the short rate in the affine model is

$$dr(t) = \dot{\alpha}(\mu - r(t))dt + \sqrt{\sigma_0^2 + \sigma_1^2 r(t)}dW(t)$$

where $W$ has dynamics given by

$$dW(t) = dZ(t) - \phi(t)dt$$

so that $\dot{\alpha} = a - \lambda\sigma_1^2$ and $\dot{\mu} = a\mu + \lambda\sigma_0^2$. The state space model and the Kalman filter described in the previous section can be applied immediately for the empirical analysis of this model.

The estimation results of the Vasicek, CIR and Duffie-Kan models are presented in Table 2. Among these traditional one-factor models, the constant volatility Vasicek model has the highest likelihood.\footnote{See Duffie and Kan (1996) for more details.}
Table 3: Summary statistics of residuals

<table>
<thead>
<tr>
<th></th>
<th>maturity</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>7</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIR</td>
<td>mean</td>
<td>-0.0010</td>
<td>0.0000</td>
<td>0.0006</td>
<td>0.0007</td>
<td>0.0004</td>
<td>-0.0007</td>
</tr>
<tr>
<td></td>
<td>st.deviation</td>
<td>0.0031</td>
<td>0.0015</td>
<td>0.0006</td>
<td>0.0011</td>
<td>0.0022</td>
<td>0.0031</td>
</tr>
<tr>
<td></td>
<td>$\rho(1)$</td>
<td>0.9642</td>
<td>0.8462</td>
<td>0.5393</td>
<td>0.8123</td>
<td>0.9333</td>
<td>0.9648</td>
</tr>
<tr>
<td></td>
<td>$\rho(30)$</td>
<td>0.6899</td>
<td>0.6132</td>
<td>0.5633</td>
<td>0.6875</td>
<td>0.6403</td>
<td>0.6797</td>
</tr>
<tr>
<td>HJM</td>
<td>mean</td>
<td>-0.0006</td>
<td>-0.0001</td>
<td>0.0001</td>
<td>0.0002</td>
<td>0.0001</td>
<td>-0.0002</td>
</tr>
<tr>
<td></td>
<td>st.deviation</td>
<td>0.0029</td>
<td>0.0012</td>
<td>0.0006</td>
<td>0.0012</td>
<td>0.0021</td>
<td>0.0028</td>
</tr>
<tr>
<td></td>
<td>$\rho(1)$</td>
<td>0.9823</td>
<td>0.9161</td>
<td>0.5948</td>
<td>0.8933</td>
<td>0.9588</td>
<td>0.9694</td>
</tr>
<tr>
<td></td>
<td>$\rho(30)$</td>
<td>0.7229</td>
<td>0.6654</td>
<td>0.3237</td>
<td>0.6304</td>
<td>0.6425</td>
<td>0.6083</td>
</tr>
</tbody>
</table>

$\rho(k)$ is the $k^{th}$ order autocovariance.

value and therefore is the preferred model. The CIR model is not very different in its properties, however. The mean reversion coefficient of the short rate (under the risk-neutral distribution) is around 0.11 for both models, which implies a half-life of about 6 years. For all the affine models, the implied long-run forward rate $\theta$ is around 0.10. The fit of the estimated yield curves is far from perfect, the estimated standard deviation of the measurement error is around 24 basis points. However, such values compare favorably with results of one-factor models on longer time series such as the McCullogh data.\(^9\)

Comparing the fit of the the traditional models with the fit of the Markovian HJM model, we find that the HJM model clearly dominates. The standard error of the HJM residuals is typically around 10% lower, and the residual serial correlation smaller than for the traditional models. Also, the value of the log-likelihood is highest for the HJM model. Although the models are not nested, so a formal likelihood ratio test is not possible, we still feel that the one-factor Markovian HJM model dominates the traditional one-factor models.

4.5 Two-Factor Models

The one-factor models all have fairly large and persistent pricing errors. It is well known from the empirical term structure literature that typically (at least) two factors are necessary to describe the term structure.\(^20\) In this section we extend the previous analysis to models with two stochastic factors. For convenience, the factors are assumed to be independent.

The HJM model is easily extended to a multi-factor framework by assuming that the instantaneous short rate is the sum of a number of underlying factors, each of which follows the dynamics described in equation (10). If these factors are independent, the solution for the bond prices is simply

$$P(t,s) = \exp \left\{ \sum_i A_i(s-t) - B_{i1}(s-t)r_i(t) - B_{i2}(t-s)\varphi_i(t) \right\}$$

where $A_i(s-t)$ and $B_i(s-t)$ satisfy differential equations similar to the ones discussed before. Assuming two stochastic factors, this specification implies a model with four state variables: two

\(^9\)See e.g. de Jong (1997).
\(^20\)See e.g. Pearson and Sun (1994), De Jong (1997), and Boudoukh, Whitelaw, Richardson, and Stanton (1997).
stochastic factors, whose sum is the short rate, and two locally riskless state variables which aggregate the stochastic factors. Extending the traditional one-factors model to a model with multiple, independent factors proceeds in exactly the same way.

Table 4 reports the estimates of the two-factor models. The main finding in all estimated models is the different speed of mean reversion of the factors. There is one factor which exhibits very slow mean reversion, whereas the mean reversion in the other factor is much stronger. Indeed, in the CIR model the first factor is not even stationary under the risk-neutral distribution.\textsuperscript{21} In the Vasicek and HJM model both factors are stationary under the risk neutral distribution.\textsuperscript{22} The volatility parameters for the HJM model tend to be somewhat higher than for the CIR model.

Compared with the one-factor models, the fit of the two-factor models improves dramatically. The standard error of the measurement error is down to 4 basis points, which is a remarkably good fit. This is confirmed by the descriptive statistics of the residuals in Table 4. The serial correlation of the residuals is much less than the serial correlation of the residuals of the one-factor models. For example, most autocorrelations at lag 30 are close to zero. Just like in the one-factor case, the Vasicek model has a slightly better fit than the CIR and HJM model. The likelihood for the Vasicek model is also higher. The differences between the two-factor CIR model and the two-factor HJM model are not very substantial. The estimated factors (or the first state variable for each factor in the HJM model) follow by and large the same pattern for both models, see Figure 5. This suggests that the two state variables that determine the short rate are sufficient to explain the term structure movements and the path volatility state variables are less important in the two-factor model.

5 Conclusion

We develop a class of term structure models under no arbitrage such that bonds and interest rate derivatives depend on two state variables that are jointly Markovian. One of these state variables is the instantaneous interest rate, while the other captures the information in the history of interest rates that is relevant for pricing. The condition for the Markovian representation of the term structure is that the diffusion function of forward rates be separable in its two time arguments, with only the current time part dependent on stochastic factors. Although restrictive, this class of models is shown to include several interesting parametric examples.

For all the models in this class, we obtain bond pricing formulas that depend only on the two state variables. Furthermore, the pricing formulas depend only on the parameters that fit the current term structures of interest rates and volatilities and not on the parameters of the dynamics of the state variables. The pricing formulas do not even depend on any additional state variables that form a joint diffusion with the original two variables.

The Markovian nature of the term structure in our class of models is particularly important for

\textsuperscript{21}Notice that the bond price equations in the CIR model remain well-defined in this case.

\textsuperscript{22}For models we imposed the restriction that the process was stationary under the empirical distribution. For the Vasicek model this requires $a > 0$ and for the CIR model $a - \lambda\sigma^2 > 0$ for both factors. In the HJM model, stationarity requires the eigenvalues of the mean reversion matrix in equation (13) to be negative. A sufficient condition for this is $2a(a - \lambda\sigma^2) > \sigma^2$. 

17
Table 4: Estimation results two-factor models

<table>
<thead>
<tr>
<th></th>
<th>Vasicek</th>
<th>CIR</th>
<th>HJM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0.4861</td>
<td>0.8103</td>
</tr>
<tr>
<td>$\hat{\sigma}$</td>
<td>0.0337</td>
<td>(0.0028)</td>
<td>(0.0069)</td>
</tr>
<tr>
<td>$\hat{\mu}$</td>
<td>0.0515</td>
<td>(0.0092)</td>
<td>0.0188</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0324</td>
<td>(0.0018)</td>
</tr>
<tr>
<td>$\theta$</td>
<td>0.0415</td>
<td>0.0230</td>
<td>0.0580</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0491</td>
<td>(0.0014)</td>
</tr>
<tr>
<td>$\sigma_0$</td>
<td>0.0103</td>
<td>0.0127</td>
<td>0.0788</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0009)</td>
<td>0.0492</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0011)</td>
<td>(0.0006)</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td></td>
<td>0.0775</td>
<td>0.0796</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0059)</td>
<td>(0.0090)</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>-11.72</td>
<td>-70.09</td>
<td>-27.53</td>
</tr>
<tr>
<td></td>
<td>(4.74)</td>
<td>(24.94)</td>
<td>(4.47)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(15.43)</td>
<td>(0.0060)</td>
</tr>
<tr>
<td>$100\delta h$</td>
<td>0.0415</td>
<td>0.0426</td>
<td>0.0426</td>
</tr>
<tr>
<td></td>
<td>(0.0021)</td>
<td>(0.0022)</td>
<td>(0.0024)</td>
</tr>
<tr>
<td>$2 \ln L$</td>
<td>37642.08</td>
<td>37319.39</td>
<td>37328.33</td>
</tr>
</tbody>
</table>

For the Duffie-Kan, Vasicek and CIR models, $\hat{\sigma}$ and $\hat{\mu}$ are free parameters in the estimation. The reported values of $\sigma$ and $\theta$ are the values implied by the estimates. For the HJM model, $\sigma$ and $\theta$ are free parameters in the estimation. In the Vasicek model, the mean of the second factor is not identified and normalized to zero. Asymptotic standard errors in parentheses.

Table 5: Summary statistics of residuals of two-factor models

<table>
<thead>
<tr>
<th></th>
<th>CIR</th>
<th>HJM</th>
</tr>
</thead>
<tbody>
<tr>
<td>maturity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean</td>
<td>-0.0001</td>
<td>-0.0000</td>
</tr>
<tr>
<td>st. deviation</td>
<td>0.0004</td>
<td>0.0003</td>
</tr>
<tr>
<td>$\rho(1)$</td>
<td>0.6771</td>
<td>0.5352</td>
</tr>
<tr>
<td>$\rho(30)$</td>
<td>0.1144</td>
<td>0.0910</td>
</tr>
<tr>
<td>maturity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean</td>
<td>-0.0000</td>
<td>-0.0000</td>
</tr>
<tr>
<td>st. deviation</td>
<td>0.0004</td>
<td>0.0003</td>
</tr>
<tr>
<td>$\rho(1)$</td>
<td>0.6917</td>
<td>0.5352</td>
</tr>
<tr>
<td>$\rho(30)$</td>
<td>0.0972</td>
<td>0.0440</td>
</tr>
</tbody>
</table>

$\rho(k)$ is the $k^{th}$ order autocovariance

18
the pricing of derivatives. In contrast to non-Markovian arbitrage-free models, there is a PDE that must be satisfied by all interest rate dependent contingent claims. Pricing derivatives by simulation or binomial approximation is also greatly enhanced by the formulation with state variables.

Finally, we estimate a parametric version of our model by pseudo maximum likelihood and extract the implied state variables using a Kalman filter. The model seems to explain very well the cross section of the term structure, with reasonably small yield errors. There is however considerable persistence remaining in the yield errors, showing that not all the dynamic characteristics of the term structure are captured by the one-factor model. The two-factor extensions of the CIR and HJM model fit the data much better. The fitted models have very small errors and the serial correlation in the residuals dies out quickly. Whereas in the one-factor case the HJM model clearly dominates the traditional term structure models, in the two-factor case the performance of the CIR and HJM models is comparable.
A Moments of the State Variables

In this appendix we derive the first two conditional and unconditional moments of the state variables for the one factor HJM model with square root volatility.\textsuperscript{23} To simplify notation we use $\sigma$ rather than $\sigma_1$ for the volatility parameter.

We start by the conditional expectation. Denote the vector that piles $E_t r(t)$ and $E_t \phi(t)$, for $t < s$, by $Y_1$. Then, this vector must satisfy the ODE

$$\frac{dY_1(s)}{ds} = A_1 Y_1(s) + Q_1(s)$$

with initial condition $Y_1(t) = (r(t), \phi(t))^t$, where

$$A_1 = \begin{pmatrix} -(a - \lambda \sigma^2) & 1 \\ \sigma^2 & -2a \end{pmatrix}$$

and $Q_1(s) = (a \theta, 0)^t$.

The solution of the ODE that is to be replaced in $M_s(y_{t-1})$, substituting $s$ by $t$ and $t$ by $t - 1$, is given by\textsuperscript{24}

$$Y_1(s) = S_1 X_1(s)$$

where

$$X_1(s) = e^{A_1(s-t)} X(t) + e^{A_1(s-t)} \int_t^s e^{-A_1(u-t)} P_1(u) du$$

with

$$X_1(t) = S_1^{-1} Y_1(t)$$

$$P_1(t) = S_1^{-1} Q_1(t)$$

Thus,

$$X_1(s) = e^{A_1(s-t)} X_1(t) + I_1(s)$$

where

$$I_1(1)(s) = \frac{\sigma^2}{\sqrt{\Delta}} \frac{a \theta}{\Lambda_1(1,1)} (e^{A_1(1,1)(s-t)} - 1)$$

and

$$I_1(2)(s) = -\frac{\sigma^2}{\sqrt{\Delta}} \frac{k_1}{\Lambda_1(2,2)} (e^{A_1(2,2)(s-t)} - 1)$$

The unconditional first moments can be obtained by making $s \to \infty$, under the condition that $\sqrt{\Delta} - \mu < 0$, for stationarity. Then

$$Y(\infty) = \begin{pmatrix} -\frac{a \theta}{\Lambda_1(1,1)} \\ 0 \end{pmatrix}$$

\textsuperscript{23}The conditional mean and variance for the short rate in the CIR model are well known and given, for example, in Chen and Scott (1993).

\textsuperscript{24}In order to make system computationally easier to solve, we first orthogonalized the differential equations and then solved the resulting system of decoupled ODE's.
Note that the unconditional mean for the spot rate is positive.

In all of the above, $S_1$ is a matrix with columns equal to the eigenvectors of $A_1$

$$S_1 = \left( \begin{array}{cc}
\frac{a+\lambda \sigma^2 + \sqrt{\Delta}}{2\sigma^2} & \frac{a+\lambda \sigma^2 - \sqrt{\Delta}}{2\sigma^2} \\
1 & 1
\end{array} \right)$$

and $\Lambda_1$ is the diagonal matrix of the eigenvalues of $A_1$

$$\Lambda_1 = \left( \begin{array}{cc}
\frac{\sqrt{\Delta}-\mu}{2} & 0 \\
0 & -\frac{\sqrt{\Delta}+\mu}{2}
\end{array} \right)$$

Finally, the following simplifications where made

$$\mu = 3a - \lambda \sigma^2$$

$$\Delta = (a + \lambda \sigma^2)^2 + 4\sigma^2$$

The second conditional moments involve the solution of a system of three ODE's. Denote the vector that piles $E_t r^2(s)$, $E_t r(s) \varphi(s)$ and $E_t \varphi^2(s)$, for $t < s$, by $Y_2$. Then, this vector must satisfy the ODE

$$\frac{dY_2(s)}{ds} = A_2 Y_2(s) + Q_2(s)$$

with initial condition $Y_2(t) = (r^2(t), r(t) \varphi(t), \varphi^2(t))'$, where

$$A_2 = \left( \begin{array}{ccc}
-2(a + \lambda \sigma^2) & 2 & 0 \\
\sigma^2 & -\mu & 1 \\
0 & 2\sigma^2 & -4a
\end{array} \right)$$

and

$$Q_2(s) = \left( \begin{array}{c}
(2f(s) + \sigma^2)E_t r(s) \\
f(s)E_t \varphi(s) \\
0
\end{array} \right)$$

The solution of the ODE is given by

$$Y_2(s) = S_2 X_2(s)$$

where

$$X_2(s) = e^{A_2(s-t)} X_2(t) + e^{A_2(s-t)} \int_t^s e^{-A_2(u-t)} P_2(u) du$$

with

$$X_2(t) = S_2^{-1} Y_2(t)$$

$$P_2(t) = S_2^{-1} Q_2(t)$$

Thus,

$$X_2(s) = e^{A_2(s-t)} X_2(t) + e^{A_2(s-t)} \begin{pmatrix}
I_2(1)(s) \\
I_2(2)(s) \\
I_2(3)(s)
\end{pmatrix}$$
where, for \( i = 1, 2, 3 \)

\[
I_2(i)(s) = \int_t^s e^{-\Lambda_2(i,i)(u-t)} P_2(i)(u) du
\]

The second conditional moments can be obtained from the above, after some simple but tedious algebra, making

\[
\begin{pmatrix}
\text{Var}_t[\tau(s)] \\
\text{Cov}_t[\tau(s), \varphi(s)] \\
\text{Var}_t[\varphi(s)]
\end{pmatrix}
= Y(s) - \begin{pmatrix}
(\text{E}_t \tau(s))^2 \\
\text{E}_t \tau(s) \text{E}_t \varphi(s) \\
(\text{E}_t \varphi(s))^2
\end{pmatrix}
\]

These are the elements to be replaced in the matrix \( V_y(y_{t-1}) \). The unconditional moments can be obtained by again making \( s \to \infty \). The condition for stationarity is the same as for the first moments.

In the above,

\[
\Lambda_2(s) = \begin{pmatrix}
-\mu & 0 & 0 \\
0 & (-\mu + \sqrt{\lambda}) & 0 \\
0 & 0 & (-\mu - \sqrt{\lambda})
\end{pmatrix}
\]

is the diagonal matrix of the eigenvalues of \( A_2 \), and

\[
S_2(s) = \begin{pmatrix}
1 & \frac{(a + \lambda \sigma^2)(a + \lambda \sigma^2 + \sqrt{\lambda} + 2\sigma^2)}{2q^4} & \frac{(a + \lambda \sigma^2)(a + \lambda \sigma^2 + \sqrt{\lambda} + 2\sigma^2)}{2q^4} \\
-a + \lambda \sigma^2 & \frac{a + \lambda \sigma^2 + \sqrt{\lambda}}{2\sigma^2} & \frac{a + \lambda \sigma^2 - \sqrt{\lambda}}{2\sigma^2} \\
-\frac{a + \lambda \sigma^2}{\sigma^2} & \frac{a + \lambda \sigma^2 + \sqrt{\lambda}}{2\sigma^2} & \frac{a + \lambda \sigma^2 - \sqrt{\lambda}}{2\sigma^2}
\end{pmatrix}
\]

is the matrix of the eigenvectors of \( A_2 \).
References


Figure 1: Evolution of the short and long end of the term structure of interest rates.
Figure 2: Filtered state variables in the one-factor affine HJM model.
Figure 3: Regressions of yields onto the two state variables in the one-factor affine HJM model. The solid lines are the sensitivities of the yields to the state variables predicted by the model. The dashed lines represent the coefficients of a regression of first differences of the yields on first differences of the fitted state variables.
Figure 4: Comparison of the filtered second state variable with an exponentially weighted moving average of the short rate according to the definition of the state variable, in the one-factor affine HJM model.
Figure 5: Filtered first state variables corresponding to each factor in the two-factor affine HJM model.