On Filtering in Markovian Term Structure Models
(an approximation approach)

Carl Chiarella
School of Finance and Economics, University of Technology Sydney
PO Box 123, Broadway NSW 2007, Australia
carl.chiarella@uts.edu.au

Sara Pasquali
Dipartimento di Matematica, Universitá degli Studi di Parma
85, Via M. D'Azeglio, I - 43100 Parma, Italy
sara.pasquali@unipr.it

Wolfgang J. Runggaldier
Dipartimento di Matematica Pura ed Applicata
Universitá di Padova, 7 Via Belzoni, I - 35131Padova, Italy
runggal@math.unipd.it

September 12, 2001

Abstract

We study a nonlinear filtering problem to estimate, on the basis of noisy observations of forward rates, the market price of interest rate risk as well as the parameters in a particular term structure model within the Heath-Jarrow-Morton family. An approximation approach is described for the actual computation of the filter.

Key words : Filter approximations, Heath-Jarrow-Morton model, market price of interest rate risk, Markovian representations, measure transformation, nonlinear filtering, term structure of interest rates.

MSC Classification : 90A09, 93E11, 60G35, 62F15, 62M05.

1 Introduction

The paper by Heath, Jarrow and Morton [14] (henceforth HJM) marked an important step in the development of models of the term structure of interest rates. The HJM model had been presaged by the simpler (and less general) Ho-Lee [15] model. The HJM model distinguished itself from previous term structure models, which were essentially based conceptually on the approach of Vasicek [24], by providing a pricing framework that is consistent with the currently observed yield curve and whose major input is a function specifying the volatility of forward interest rates. To this extent it can be viewed as the complete analogue, in the world of stochastic interest rates, to the Black-Scholes model of the deterministic interest rate world that prices derivatives consistently with respect to the price of the underlying asset (of which the currently observed yield curve is the analogue) and requires as its major input the volatility of returns of the underlying asset (to which the forward rate volatility function is the analogue).

The challenges posed in implementing the HJM model arise from the fact that in its most general form the stochastic dynamics are non-Markovian in nature. As a result most implementations of the HJM model revolve around some procedure, and/or assumptions, that allow the stochastic dynamics to be re-expressed in Markovian form - usually by employing the "trick" of expanding the state-space.

As we have stated above the major input into the HJM model is the forward rate volatility function and indeed its specification will determine the nature of the stochastic dynamics and whether and how it then can be reduced to Markovian form.

In view of finite dimensional realizations of HJM models (for a general study see [6]), Chiarella and Kwon [8], [9] have shown that a broad, and important for applications, class of interest rate derivative models whose dynamics can be "Markovianised" can be obtained by assuming forward rate volatility functions that depend on a finite set of forward rates with given maturities as well as time to maturity.

An important practical problem faced in implementing such term structure models is the estimation of the parameters entering into the specification of the forward rate volatility function. In fact, one of the major aims of this paper is to show how this estimation problem can be approached within a filtering framework.

In section 2 we introduce our basic model that is a particular version of the HJM model set-up within the Chiarella-Kwon [8], [9] framework in which the volatility function depends on the instantaneous spot rate of interest (maturity of zero), one forward rate of fixed maturity and, time to maturity. Under the risk-neutral probability measure the stochastic dynamics of the spot rate and of the fixed maturity forward rate are given by a two-dimensional Markovian stochastic differential equation system. However as our observations occur under the so-called historical probability measure, we need to introduce also the market price of interest rate risk (that connects the two probability measures). We assume that the market price of risk follows a mean reverting process and so, under the historical measure, we are left with a three-dimensional Markovian stochastic differential system. A truncation factor is furthermore added to the coefficients thereby guaranteeing existence and uniqueness of a strong solution that takes values in a compact set. Assuming that the information comes from noisy observations of the fixed-maturity forward rate, in this same section 2 we also formulate the filtering problem, whose solution leads to the estimation of the market price of risk and of the unobserved instantaneous rates of interest and as well as of the parameters in the model.

The resulting filtering problem is highly nonlinear so that approximation methods have to be used for its solution. We shall describe a method, based on time discretization that, together with further approximations (quantization), leads to a discrete time approximating problem for which a filter of fixed finite dimension can be derived. Provided the discretization is sufficiently fine, the optimal filter for the approximating problem can be shown to be an arbitrarily good approximation to the filter for the original problem. Time and spatial discretization methods for nonlinear filtering were pioneered by H.Kushner and his co-workers (for a general exposition see [18]). Our method here differs in various respects from those in [18] and extends previous work in [12], [17], [23] (see also [20], [22] and the references in those papers).

In section 3 we discuss the time discretization and show the convergence of the time discretized filter for each observed trajectory and not merely in the mean with respect to the observations. We also mention further discretizations (quantizations) that lead to finite-dimensional approximating filters. We point out that the time discretization does not even need to be looked at as an approximation per se, since the real observations take place in discrete time only and so the true filtering problem is actually one in discrete time. In this sense the convergence of the time discretized filter can be viewed as guaranteeing the consistency of the discrete time models with the original continuous-time setup.

2 Stochastic Dynamics and Filter Setup

Let $f(t,T)$ be the rate we contract at time t for instantaneous borrowing at time T $(> t)$. The Heath, Jarrow and Morton (HJM) [14] model for the term structure of interest rates is based on modelling the forward rates according to

$$f(t,T) = f(0,T) + \int_0^t \sigma^*(u,T)du + \int_0^t \sigma(u,T)d\tilde{w}_u \tag{1}$$

Here $f(0,T)$ is the observed forward rate curve at time 0 and $\tilde{w}_t$ is a scalar Wiener process on a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, Q)$ with Q the HJM "martingale measure". The quantity $\sigma(t,T)$ is the volatility function of the forward rate process which in general is an adapted process (in t), that we may view as being parametrized by T. From the HJM drift restriction we have that

$$\sigma^*(t,T) = \sigma(t,T)\int_t^T \sigma(t,u)du \tag{2}$$

The two major inputs into the HJM model are the initially observed forward curve $f(0,T)$ and the forward rate volatility function $\sigma(t,T)$. The $f(0,T)$ is imposed by the market, but $\sigma(t,T)$ remains at the discretion of the model builder. In fact, equation (1) specifies an entire family of models depending on how $\sigma(t,T)$ is specified and, as stated in the Introduction, in its most general form is non-Markovian.

Bhar, Chiarella, El-Hassan and Zheng [2] have modellled the randomness of the volatility function through dependence on the unobserved instantaneous spot rate of interest $r_t =$

$f(t,t)$ [1] and a forward rate $f_t = f(t,\tau)$ with fixed maturity τ. In particular, they take (with obvious abuse of notation)

$$\sigma(t,T) = \sigma(t,T;r_t,f_t) = g(r_t,f_t)\,e^{-\lambda(T-t)} \tag{3}$$

with $0 \le t < \tau < T$, where $\lambda > 0$ is a parameter and g a sufficiently well behaved function. The motivation for this particular specification is that it allows reduction of the forward rate dynamics to Markovian form. Furthermore, it generalizes in an obvious way the class of volatility functions introduced by Ritchken and Sankarasubramanian [21] in which g depends only on r_t. It turns out that, under the specification (3), the dynamics of a generic forward rate $f(t,T)$, of the fixed maturity forward rate f_t, and of the short rate r_t are then, according to [8], [9], driven by the Markovian system of stochastic differential equations

$$\begin{cases} df(t,T) = D_t(T)\,\sigma^2\,(t,T;r_t,f_t)\,dt + \sigma\,(t,T;r_t,f_t)\;d\tilde{w}_t \\ df_t = D_t\,\sigma^2\,(t,\tau;r_t,f_t)\,dt + \sigma\,(t,\tau;r_t,f_t)\;d\tilde{w}_t \\ dr_t = [A_t + B_t r_t + C_t f_t]\;dt + \sigma\,(t,t;r_t,f_t)\;d\tilde{w}_t \end{cases} \tag{4}$$

The function $g(r,f)$ in (3) is assumed to be of the form

$$g(r,f) = |\,a_0 + a_1 r + a_2 f\,|^{\delta} \tag{5}$$

for some positive parameters a_0, a_1, a_2, δ. Furthermore,

$$\begin{array}{l} D_t(T) = \lambda^{-1}\left(e^{\lambda(T-t)} - 1\right) \quad ; \quad D_t = D_t(\tau) \\ B_t = -\lambda\left[\left(e^{-\lambda(\tau-t)} - 1\right)^{-1} + 1\right] \quad ; \quad C_t = -\lambda e^{\lambda(\tau-t)}\left(e^{-\lambda(\tau-t)} - 1\right)^{-1} \\ A_t = f_T(0,t) - B_t f(0,t) - C_t f(0,\tau) \end{array} \tag{6}$$

where $f(0,t), f(0,\tau)$ are the initial forward rates for the maturities t and τ respectively, and $f_T(0,t)$ represents the partial derivative of $f(0,t)$ with respect to the second variable. We shall refer to f_t and r_t as *state variables* in the Markovian system (4). From model (4) we can derive by Ito's lemma the dynamics for the price $P(t,T)$ of a zero-coupon bond [2] with generic maturity T, namely

$$dP(t,T) = P(t,T)\,[r_t dt - D_t(T)\sigma(t,T;r_t,f_t)\,d\tilde{w}_t] \tag{7}$$

For later empirical implementations it is important to keep in mind how the stochastic dynamic system (4) should be interpreted. Suppose our observation period is 1^{st} June to 30^{th} June, and we have daily observations. On the first of June we have a zero coupon forward curve, $f(0,T)$ (T indicates maturity), reconstructed from a whole set of (noisily)

[1] The instantaneous spot rate of interest, r_t, is treated as unobserved since the shortest rate we observe in most markets is a 30-day rate. In many empirical studies in finance this latter rate is treated as a proxy for r_t. Part of our contribution is the development of a methodology that avoids such an approximation. We should however also point out that [7] discusses situations in which certain market observed short rates (such as 30-day and 90-day rates) are reasonable proxies for r_t.

[2] Recall that $P(t,T) = \exp\left(-\int_t^T f(t,u)du\right)$

observed forward rates. It is more likely that agents observe the prices of available zero-coupon bonds, however, since there is a one-to-one correspondence between these prices and forward rates, we may as well assume that the agents have access to the latter (forward rates can be reconstructed from observable data). Whether we take available bond prices or forward rates as the observed quantities, these have to be reconstructed from actually accessible data, and so such observations have to be considered as noisy. In spite of the fact that the forward rates are noisy, we take the reconstructed $f(0,T)$ as the "true" zero coupon yield curve on 1^{st} June. This viewpoint is consistent with the one we shall adopt in setting up the (Bayesian) filtering algorithm (see Remark 2.4).

The SDE system (4) that we are considering tells us how the zero coupon forward curve of 1 June will be projected over the month of June under the proposed forward rate volatility function. Recall that under the assumptions of the model, the evolution of the forward curve on any day is driven by that of the state variables.

This evolution of the initial forward curve can be depicted as shown in Fig. 1.

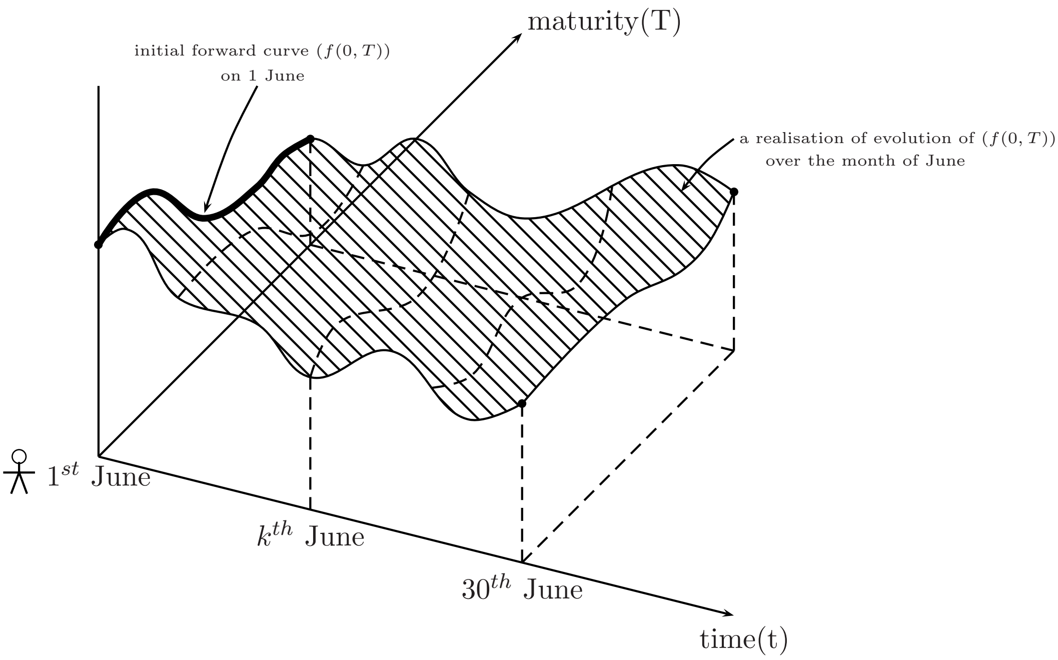

Figure 1: The Evolution of the Forward Curve

- Solid curve — initial (reconstructed from forward rates of many maturities) forward curve
- Dashed curves — realisations of the evolution of $f(0,T)$

In order to focus on perhaps the simplest filtering problem in the framework of the stochastic dynamical system (4), we shall assume (see (16) below) that the available observations are noisy observations of the forward rate with the fixed maturity τ (one may obviously add noisy observations of forward rates with other maturities as well as of any other economic quantity, whose dynamics can be derived from (4)).

Since $f_t = f(t,\tau)$ has to be treated here as an underlying quantity as opposed to a derivative quantity, we have to model its observations under the "historical" or "real world" probability measure P. [3] We shall therefore introduce the "market price of interest rate risk" process ψ_t, that corresponds to the translation of the Wiener process when passing from the measure Q to P, and assume that it satisfies, under the measure P, a mean reverting diffusion model. The market price of interest rate risk is essentially the additional compensation that a rational investor, operating under conditions of absence of arbitrage, would require for bearing an additional unit of interest rate risk as measured by a unitary increase in volatilty of the forward rate curve (see e.g. [4]).

Denote then by X_t the "state" process

$$X_t := [f_t, r_t, \psi_t]' \tag{8}$$

and, given a (large) $H > 0$ and a (small) $\epsilon > 0$, let

$$\chi(X) = \begin{cases} 1 \quad \text{if} \quad \max\{|f_t|,\, |r_t|\} \le H \\ 0 \quad \text{if} \quad \min\{|f_t|,\, |r_t|\} \ge H+\epsilon \\ \text{else a Lipschitz interpolation} \end{cases}; \quad \bar{\chi}(\psi) = \begin{cases} 1 & \text{if} \quad |\psi| \le H \\ 0 & \text{if} \quad |\psi| > H+\epsilon \\ \frac{H+\epsilon-|\psi|}{\epsilon} & \text{if} \quad H < |\psi| < H+\epsilon \end{cases} \tag{9}$$

Under the measure P with Wiener process $w_t = \tilde{w}_t - \int_0^t \psi_s ds$, we now let the processes f_t, r_t, ψ_t satisfy the dynamics

$$\begin{cases} df_t = (D_t\, \sigma(t,\tau; r_t, f_t) + \psi_t)\, \sigma(t,\tau; r_t, f_t)\; \chi(X_t)\, dt + \sigma(t,\tau; r_t, f_t)\; \chi(X_t)\, dw_t \\ dr_t = [A_t + B_t r_t + C_t f_t + \psi_t \sigma(t,t; r_t, f_t)]\; \chi(X_t)\, dt + \sigma(t,t; r_t, f_t)\; \chi(X_t)\, dw_t \\ d\psi_t = \kappa\, \left(\bar{\psi} - \psi_t\right)\; \bar{\chi}(\psi_t) dt + b\, |\psi_t|^\gamma \bar{\chi}(\psi_t)\, dw_t \end{cases} \tag{10}$$

where the totality of the parameters is given by the vector

$$\theta := (a_0, a_1, a_2, \delta, \kappa, \bar{\psi}, b, \gamma, \lambda) \tag{11}$$

and each of them is supposed to take values in a compact subset of the positive halfline. With the vector X_t as in (8), we shall write the dynamics in (10) in compact form as

$$dX_t = F_t(X_t)\, dt + G_t(X_t)\, dw_t \tag{12}$$

where $F_t(\cdot)$ and $G_t(\cdot)$ are implicitly defined in (10). In what follows, the generic $i-$th ($i = 1,2,3$) components of $F_t(\cdot)$ and $G_t(\cdot)$ will be denoted by $F_t^{(i)}(\cdot)$ and $G_t^{(i)}(\cdot)$ respectively. We have

Proposition 2.1 *The system (10) (equivalently (12)) has a unique strong and bounded solution.*

[3] On the other hand, if one takes as observations any of the derivative quantities, one would have the choice (depending on the intended application) of modelling their observations either under the martingale measure Q or under the real world measure P.

Proof : The boundedness of the solution follows from the truncation factors in the coefficients. It then suffices to show that, for a bounded solution, the drift and diffusion coefficients in the three equations in (10) are globally Lipschitz and for this purpose it is easily seen that it suffices to show the Lipschitzianity with respect to the spatial variable.

For the first drift coefficient we have

$$\begin{aligned}&\left| \, [D_t\sigma(t,\tau;r,f)+\psi] \; \sigma(t,\tau;r,f)\chi(X) - [D_t\sigma(t,\tau;r',f')+\psi'] \; \sigma(t,\tau;r',f')\chi(X') \, \right| \\ &\leq C \, \left|\sigma^2(t,\tau;r,f)\chi(X) - \sigma^2(t,\tau;r',f')\chi(X')\right| + \left| \, \sigma(t,\tau;r,f)\psi\chi(X) - \sigma(t,\tau;r',f')\psi'\chi(X')\right|\end{aligned} \tag{13}$$

with C a constant and from here the Lipschitzianity follows by the boundedness of $\sigma(\cdot)$, $\chi(\cdot)$ and ψ and the Lipschitzianity of $\sigma(\cdot)$ and $\chi(\cdot)$ (recall that r, f and ψ are solutions of (10) and therefore bounded).

Coming to the second drift term we have

$$\begin{aligned}&|[A_t + B_t r + C_t f + \psi_t\sigma(t,t;r,f)] \, \chi(X) - [A_t + B_t r' + C_t f' + \psi_t'\sigma(t,t;r',f')] \, \chi(X')| \\ &\leq |A_t| \, |\chi(X) - \chi(X')| + |B_t| \, |r\chi(X) - r'\chi(X')| \\ &\quad + |C_t| \, |f\chi(X) - f'\chi(X')| + |\sigma(t,t;r,f)\psi_t\chi(X) - \sigma(t,t;r',f')\psi_t'\chi(X')|\end{aligned} \tag{14}$$

The function A_t is an input and is bounded, uniformly in t, together with B_t and C_t. The Lipschitzianity then follows for the same reasons as before.

For the last drift term the Lipschitzianity follows again straightforwardly for the same reasons as before since

$$\left|\kappa \, \left(\bar{\psi} - \psi\right) \bar{\chi}(\psi) - \kappa \, \left(\bar{\psi} - \psi'\right) \bar{\chi}(\psi')\right| \leq C \, \left(|\bar{\chi}(\psi) - \bar{\chi}(\psi')| + |\psi\bar{\chi}(\psi) - \psi'\bar{\chi}(\psi')|\right) \tag{15}$$

Finally, the Lipschitzianity of the diffusion coefficients follows by complete analogy with the drift coefficients.

□

Remark 2.2 *In the literature one can find results on the existence of a strong solution to equations of the form (4) with volatilities according to (3) and (5) (see e.g. [10]). These results hold however for specific ranges of the parameter δ in (5). In our application δ may take any positive value and so we preferred to introduce the Lipschitz truncation factors (9) to ensure in any case the existence of a strong and bounded solution. From a practical point of view this truncation is hardly any restriction at all.*

Model (12), resulting from (10) is a minimal Markovian model for the term structure of interest rates : the dynamics of the various other forward rates $f(t,T)$ with generic maturity T (as well as the corresponding zero-coupon bond prices) can be derived from the first equation in (4) and from (7), whose dynamics depend only on the vector X_t. In what follows we shall denote by $\mathcal{X}$ the compact subset of $\mathbb{R}^3$ for which $X_t \in \mathcal{X}$.

In line with the foregoing, we shall assume that agents have access to noisy observations of $f_t = f(t,\tau)$. Denoting the observation process by y_t, we assume that it satisfies

$$dy_t = f_t dt + \hat{\epsilon} d\hat{w}_t \tag{16}$$

with $\hat{\epsilon} > 0$ small and $\hat{w}_t$ a $P-$Wiener, independent of w_t.

The goal here is a recursive Bayesian-type estimation of X_t and θ on the basis of the past and present observations of y_t, i.e. the combined filtering and parameter estimation of (X_t, θ), given $\mathcal{F}_t^y$, which is the filtration generated by the process y_t. The most complete solution to this problem is the recursive computation of the conditional joint distribution $p\,(X_t, \theta \mid \mathcal{F}_t^y)$. This is a highly nonlinear filtering problem and so in section 3 we shall compute a weak approximation to $p\,(X_t, \theta \mid \mathcal{F}_t^y)$ in the sense that we shall compute an approximation of the conditional expectation

$$E\left\{\bar{\Gamma}(X_t;\theta) \mid \mathcal{F}_t^y\right\} = \int \bar{\Gamma}(X;\theta)\, dp\,(X;\theta \mid \mathcal{F}_t^y) \tag{17}$$

where, for each θ, $\bar{\Gamma}(\cdot;\theta)$ is Lipschitz. The approximation is by discretization in time, which is motivated not only by the difficulty of computing (17) exactly, but also by the fact that, in reality, y_t is observed in discrete time. Additional possible approximations will also be mentioned in section 3

Remark 2.3 *Since the solution X_t of (12) takes values in the compact set $\mathcal{X}$, we may, without changing the value in (17), assume that $\bar{\Gamma}(X;\theta) = 0$ for $X \notin \mathcal{X}$. Notice also that from the econometric literature one has an indication of what could be possible values of the parameter vector θ. We shall thus assume that θ takes already from the outset only a finite number of possible values to which we may assign a uniform prior. This implies that the time discretization below concerns only the process X_t and, to emphasize this fact, we shall put $\Gamma_\theta(X) := \bar{\Gamma}(X;\theta)$ so that, instead of (17), we shall compute/approximate*

$$E\left\{\Gamma_\theta(X_t) \mid \mathcal{F}_t^y\right\} \tag{18}$$

Remark 2.4 *Stochastic filtering can be viewed as a dynamic generalization of Bayesian statistics. The "prior distribution" in this dynamic setup is given by the joint distribution of the (unobservable) state process X_t and of the parameter vector θ. This distribution is implied by the dynamic model for X_t (see (10) and (12)) and by the prior distribution on θ. This joint prior distribution is then successively updated on the basis of empirical data, namely of the noisy observations y_t of f_t. Analogously to classical Bayesian statistics, also in its dynamic generalization the "prior" is specified on the basis of extra-experimental information and/or on the basis of prior empirical information. As explained in the paragraph below equation (6), this is also the sense in which our double use of observations of forward rates is being interpreted : the one time initial observations of $f(0,t), f(0,\tau)$, $f_T(0,t)$ correspond to "prior" empirical information which is used, see (6), to determine the function A_t that is part of the dynamic model for X_t (see (10)), and thus of the "prior" for X_t. The successive noisy observations y_t of f_t on the other hand constitute the successively increasing empirical information, on the basis of which the prior of (X_t, θ) is being updated.*

We want to point out that, in Bayesian statistics, the current distributions turn out to be more informative, if one is able to assign a more informative prior. To this effect notice that, although the solution of (12) takes values in the compact set $\mathcal{X}$, there is no guarantee on the positivity of the instantaneous rates r_t and f_t. Since these rates are essentially positive, we should get more informative results if the "prior", i.e. our dynamic model for X_t guarantees positivity of these rates. For this purpose notice next that, if two quantities are in a one-to-one correspondence with each other, observing one of them or updating the distribution of one of them turns out to be equivalent to observing the other or updating its distribution respectively. We may therefore apply to the rates r_t and f_t an invertible transformation that transforms them into positive rates. For this purpose we use the $\mathcal{C}^2-$transformation

$$\bar{x} = T(x) := \begin{cases} x & \text{if} \quad x \geq \epsilon + \eta \\ (\epsilon + \eta) + \frac{2\eta}{\pi} \text{ arctan } \left[\frac{\pi}{2\eta}(x - \epsilon - \eta) \right] & \text{if} \quad x < \epsilon + \eta \end{cases} \tag{19}$$

where ϵ is, again, a small positive real and $0 < \eta < \epsilon$ (see Figure 2).

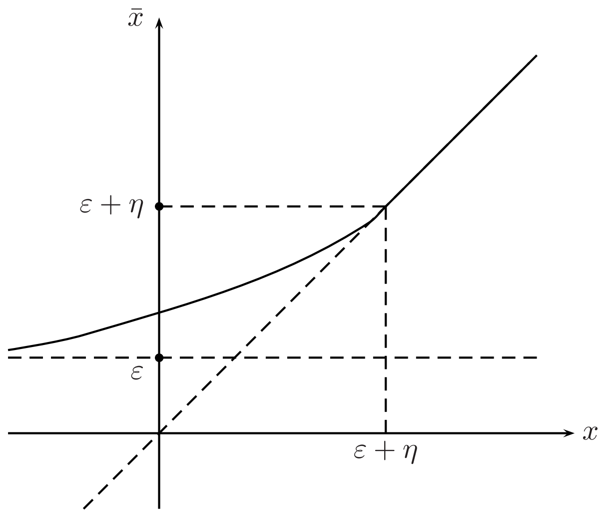

Figure 2: The Transformation $\bar{x} = T(x)$

Define $\rho_t := T(r_t)$, $\phi_t := T(f_t)$ and notice that, with the same H as in (9), $\rho_t, \phi_t \geq T(-H - \epsilon) > \epsilon$ and, on $[\epsilon + \eta, H]$, we have $\rho_t = r_t$, $\phi_t = f_t$. Putting $\bar{X}_t := [\phi_t, \rho_t, \psi_t]'$, we may, with some abuse of notation, also write $\bar{X}_t = T(X_t)$ and, applying Ito's rule, obtain from (12)

$$d\bar{X}_t = \bar{F}_t(X_t)dt + \bar{G}_t(X_t)\, dw_t \tag{20}$$

where the $i-$th $(i = 1, 2, 3)$ components of $\bar{F}_t(\cdot)$ and $\bar{G}_t(\cdot)$ are

$$\begin{aligned} \bar{F}_t^{(i)}(X_t) &= \begin{cases} F_t^{(i)}(X_t) & \text{if} \quad i = 3 \\ \dot{T}(X_t^{(i)})F_t^{(i)}(X_t) + \frac{1}{2}\ddot{T}(X_t^{(i)})(G_t^{(i)})^2(X_t) & \text{if} \quad i = 1, 2 \end{cases} \\ \bar{G}_t^{(i)}(X_t) &= \begin{cases} G_t^{(i)}(X_t) & \text{if} \quad i = 3 \\ \dot{T}(X_t^{(i)})G_t^{(i)}(X_t) & \text{if} \quad i = 1, 2 \end{cases} \end{aligned} \tag{21}$$

and they are bounded since all the individual factors on the right in (21) are. Since $T(\cdot)$ is invertible, the Ito process $\bar{X}_t$ in (20) can be represented as solution of

$$d\bar{X}_t = \bar{F}_t(T^{-1}(\bar{X}_t))dt + \bar{G}_t(T^{-1}(\bar{X}_t))\, dw_t \tag{22}$$

Proposition 2.5 *Equation (22) admits a unique strong solution.*

Proof : Notice first from (19) that the inverse transformation $T^{-1}(\bar{x})$ is given by

$$T^{-1}(\bar{x}) = \begin{cases} \bar{x} & \bar{x} \geq \epsilon + \eta \\ (\epsilon + \eta) + \frac{2\eta}{\pi} \tan\left[\frac{\pi}{2\eta}(\bar{x} - \epsilon - \eta)\right] & \epsilon < \bar{x} < \epsilon + \eta \end{cases}$$

and is Lipschitz so that $F_t^{(i)}\left(T^{-1}(\bar{X})\right)$ and $G_t^{(i)}\left(T^{-1}(\bar{X})\right)$ are also Lipschitz in addition to being bounded. To obtain the global Lipschitzianity of the coefficients in (22) and thus the existence of a strong solution, by (21) it suffices thus to show Lipschitzianity and boundedness of $\dot{T}\left(T^{-1}(\bar{x})\right)$ and $\ddot{T}\left(T^{-1}(\bar{x})\right)$. This follows immediately from their explicit expression, namely

$$\dot{T}\left(T^{-1}(\bar{x})\right) = \begin{cases} 1 & \bar{x} \geq \epsilon + \eta \\ \frac{1}{1+\left[\tan\left(\frac{\pi}{2\eta}(\bar{x}-\epsilon-\eta)\right)\right]^2} & \epsilon < \bar{x} < \epsilon + \eta \end{cases}$$

$$\ddot{T}\left(T^{-1}(\bar{x})\right) = \begin{cases} 0 & \bar{x} \geq \epsilon + \eta \\ \frac{\frac{\pi}{\eta}\tan\left[\frac{\pi}{2\eta}(\bar{x}-\epsilon-\eta)\right]}{\left[1+\left[\tan\left(\frac{\pi}{2\eta}(\bar{x}-\epsilon-\eta)\right)\right]^2\right]^2} & \epsilon < \bar{x} < \epsilon + \eta \end{cases}$$

□

In what follows we shall always refer to the same model (12) also in the case when we apply the transformation $T(\cdot)$. In this latter case X_t stands for $\bar{X}_t$, and the functions $F_t(X)$ and $G_t(X)$ then correspond to $\bar{F}_t\left(T^{-1}(\bar{X})\right)$ and $\bar{G}_t\left(T^{-1}(\bar{X})\right)$ respectively. Similarly, f_t in equation (16) stands for ϕ_t in case we apply the transformation $T(\cdot)$.

Notice that alternative approaches to obtain positive rates can be found in the recent literature (see e.g. [13]).

Notice finally that the filtering approach to HJM term structure models can also be seen as a possible way to overcome consistency problems in the calibration of HJM models (for the latter see e.g. the overview in [5]).

3 Time discretization and convergence results

In the following we implicitly assume that a generic value of θ has been fixed. Consider the partition of $[0, T]$ into subintervals of the same width $\Delta = \frac{T}{N}$ and perform an Euler discretization of (12), namely

$$X_{n+1}^N - X_n^N = F_n\left(X_n^N\right)\Delta + G_n\left(X_n^N\right)\Delta w_n \tag{23}$$

with $\Delta w_n = w_{(n+1)\Delta} - w_{n\Delta}$. Notice that, while the solution of the continuous-time model (12) is bounded, its discretized version (23) does not guarantee boudedness of (X_n^N). Denote by X_t^N the piecewise constant time interpolation of X_n^N, namely

$$X_t^N := \begin{cases} X_n^N & n\Delta \leq t < (n+1)\Delta \\ X_N^N & t = T \end{cases} \tag{24}$$

and simply write X_n for $X_{n\Delta}^N$ as well as $X^{(i)}$ for the $i-$th component of X.

Consider next a Girsanov-type change of measure which allows us to transform the original filtering problem into one with independent state and observations. Denote by P^0 the measure under which y_t is a Wiener process, independent of X_t and thus also of X_t^N. In fact, the change of measure affects only the distribution of y_t and not also of X_t . The corresponding Radon-Nikodym derivative is

$$\frac{dP}{dP^0} = \exp\left[\frac{1}{\hat{\epsilon}^2}\int_0^T f_s dy_s - \frac{1}{2\hat{\epsilon}^2}\int_0^T f_s^2 ds\right] \tag{25}$$

Analogously, denote by P^N the measure under which y_t satisfies the equation

$$dy_t = f_t^N dt + \hat{\epsilon} dw_t^N \tag{26}$$

with w_t^N a P^N-Wiener process and where, with some abuse of notation, we denote by f_t^N the first component of X_t^N, truncated upon exit from $[-(H+\epsilon),(H+\epsilon)]$ (H and ϵ are the same as in (9)); as a consequence, in what follows f_t^N will be treated as having the same bounds as f_t. We thus have that, under P^N, y_t has the same form as under P, but as a function of the discretized state.

Applying the so-called Kallianpur-Striebel formula (see [16]), the filter in (18) can be expressed as

$$E\left\{\Gamma_\theta(X_t)\,|\,\mathcal{F}_t^y\right\} = \frac{E^0\left\{\Gamma_\theta\left(X_t\right)\frac{dP}{dP^0}\,|\mathcal{F}_t^y\right\}}{E^0\left\{\frac{dP}{dP^0}\,|\mathcal{F}_t^y\right\}} \tag{27}$$

It follows that it suffices to approximate, for each value of θ,

$$V_t\left(\Gamma_\theta; y\right) := E^0\left\{\Gamma_\theta\left(X_t\right)\frac{dP}{dP^0}\,|\mathcal{F}_t^y\right\} \tag{28}$$

(the denominator in (27) is in fact simply $V_t(1;y)$).

Define

$$z_t := E^0\left\{\frac{dP}{dP^0}|\mathcal{F}_t\right\} = \exp\left[\int_0^t \frac{1}{\hat{\epsilon}^2} f_s dy_s - \frac{1}{2\hat{\epsilon}^2}\int_0^t f_s^2 ds\right] \tag{29}$$

$$z_t^N := E^0\left\{\frac{dP^N}{dP^0}|\mathcal{F}_t\right\} = \exp\left[\int_0^t \frac{1}{\hat{\epsilon}^2} f_s^N dy_s - \frac{1}{2\hat{\epsilon}^2}\int_0^t (f_s^N)^2 ds\right] \tag{30}$$

where $\mathcal{F}_t = \mathcal{F}_t^y \vee \mathcal{F}_t^X$. By analogy to (28) define, for $N \in \mathbb{N}$,

$$V_t^N\left(\Gamma_\theta; y\right) := E^0\left\{\Gamma_\theta\left(X_t^N\right) z_T^N\,|\mathcal{F}_t^y\right\} \tag{31}$$

By the "smoothing property" of conditional expectations we have

$$V_t\left(\Gamma_\theta; y\right) = E^0\left\{\Gamma_\theta\left(X_t\right) z_t\,|\mathcal{F}_t^y\right\}, \quad V_t^N\left(\Gamma_\theta; y\right) = E^0\left\{\Gamma_\theta\left(X_t^N\right) z_t^N\,|\mathcal{F}_t^y\right\} \tag{32}$$

We first have the following

Proposition 3.1 *The processes* $\{X_t\}$ *and* $\left\{X_t^N\right\}$ *satisfy, for* $t \in [0,T]$

$$E\left\|X_t - X_t^N\right\|^4 \le K\Delta^2 \quad \textit{and} \quad E^0\left\|X_t - X_t^N\right\|^4 \le K\Delta^2$$

where K *is a positive constant.*

Proof : The proof can easily be adapted from [12], where the components of X_t^N are not truncated, while here we have truncated the first component f_t^N. Notice however that, given the recursions (23) and our assumptions, the difference in fourth mean of the truncated and non-truncated values of f_t^N is, for all $t \in [0,T]$, bounded from above by $E\left\{ \left| F_n^{(1)}(\cdot)\Delta + G_n^{(1)}(\cdot)\Delta w_n \right|^4 \right\} \leq const \cdot \left(\|F^{(1)}\|^4 + \|G^{(1)}\|^4 \right) \Delta^2$ and the coefficient of Δ^2 in this latter quantity is bounded due to the fact that F and G are bounded by definition and this also in the case when we apply the transformation $T(\cdot)$ in (19) (see (21) and the proof of Proposition 2.5).

□

Notice that, according to Remark 2.3, the value of $V_{n\Delta}^N(\Gamma_\theta; y)$ in (31) does not change if we change the values of X_t^N outside of $\mathcal{X}$. Consequently, we shall truncate the process X_t^N as soon as it exits from $\mathcal{X}$ and denote by X_n the so truncated process ($X_n^{(i)}$ will denote the $i-$th ($i = 1,2,3$) component of X_n and notice that for $f_n^N = f_n = X_n^{(2)}$ we have already used this truncation after (26)). The process X_n is now bounded Markov with a well-defined transition kernel $P(X_{n+1}|\, X_n)$. The explicit expression of $P(X_{n+1}|\, X_n)$ is somewhat complicated but, for the actual calculations, we need its expression only in the interior of $\mathcal{X}$ and there it is given by

$$
\begin{aligned}
P(X_{n+1}|\, X_n) = & \left\{ \frac{1}{\sqrt{2\pi\, (G_n^{(1)})^2(X_n)\Delta}} \exp\left[-\frac{\left(X_{n+1}^{(1)} - X_n^{(1)} - F_n^{(1)}(X_n)\Delta \right)^2}{2(G_n^{(1)})^2(X_n)\Delta} \right] \right. \\
& \cdot \prod_{i=2}^{3} \delta \left(X_{n+1}^{(i)} - X_n^{(i)} - F_n^{(i)}(X_n)\Delta \right. \\
& \qquad \left. \left. - G_n^{(i)}(X_n)) \cdot \frac{X_{n+1}^{(1)} - X_n^{(1)} - F_n^{(1)}(X_n)\Delta}{G_n^{(1)}(X_n)} \right) \right\} \\
& \cdot \prod_{i=1}^{3} 1_{[-(H+\epsilon),(H+\epsilon)]} \left(X_{n+1}^{(i)} \right)
\end{aligned}
\tag{33}
$$

We also make the following assumption, which is in line with our observation model (16)
Assumption A.1 : The actually observed trajectory (y_t) satisfies, for $n = 0, \cdots, N-1$,

$$
\sup_{s,t\in[n\Delta,(n+1)\Delta]} |\, y_s - y_t \,| \leq K\Delta^{1/2}
$$

Lemma 3.2 *Given an observed trajectory y_s $(s \leq t)$ satisfying A.1, we have for $t = n\Delta$*

$$
E^0\left\{ z_t^2 | \mathcal{F}_t^y \right\} \leq K(y) \quad ; \quad E^0\left\{ (z_t^N)^2 | \mathcal{F}_t^y \right\} \leq K(y) \tag{34}
$$

$$
E^0\left\{ \left| z_t - z_t^N \right| | \mathcal{F}_t^y \right\} \leq \bar{K}(y) \cdot \Delta^{\frac{1}{2}} \tag{35}
$$

where $K(y)$, $\bar{K}(y)$ depend only on the observed trajectory y_s, $s \leq t$.

Proof. We start with the proof of the first inequality in (34). Using the stochastic integration by parts formula and the fact that $X_n = [f_n, r_n, \psi_n]$ is bounded together with the coefficients

$F_t^{(i)}(\cdot)$, $G_t^{(i)}(\cdot)$ $(i=1,2,3)$, we have

$$
\begin{aligned}
&E^0\left\{z_t^2 \mid \mathcal{F}_t^y\right\} = E^0\left\{\exp\left[\frac{2}{\hat{\epsilon}^2}\int_0^t f_s dy_s - \frac{1}{\hat{\epsilon}^2}\int_0^t f_s^2 ds\right] \mid \mathcal{F}_t^y\right\} \\
&\leq E^0\left\{\exp\left[\frac{2}{\hat{\epsilon}^2} f_t y_t - \frac{2}{\hat{\epsilon}^2}\int_0^t y_s df_s\right] \mid \mathcal{F}_t^y\right\} \\
&\leq K(y)\, E^0\left\{\exp\left[-\frac{2}{\hat{\epsilon}^2}\int_0^t y_s F_s^{(1)}(X_s) ds - \frac{2}{\hat{\epsilon}^2}\int_0^t y_s G_s^{(1)}(X_s) dw_s\right] \mid \mathcal{F}_t^y\right\} \\
&\leq \bar{K}(y)\, E^0\left\{\exp\left[\int_0^t H_s(y) dw_s - \frac{1}{2}\int_0^t H_s^2(y) ds\right] \exp\left[\tfrac{1}{2}\int_0^t H_s^2(y) ds\right]\right\} \\
&\leq \tilde{K}(y)
\end{aligned}
\tag{36}
$$

for appropriate constants $K(y), \bar{K}(y), \tilde{K}(y)$ and an adapted bounded process $H_s(y)$ that depends on the observed trajectory of y (recall that, under P^0, the processes X_t and y_t are independent).

Coming to the second inequality in (34) and recalling that the values of f_n^N are bounded, we have, for $t = n\Delta$ and with $\Delta y_{i+1} := y_{(i+1)\Delta} - y_{i\Delta}$,

$$
\begin{aligned}
E^0\left\{(z_t^N)^2 \mid \mathcal{F}_t^y\right\} &= E^0\left\{\exp\left[\frac{2}{\hat{\epsilon}^2}\sum_{i=0}^{n-1} f_i^N \Delta y_{i+1} - \frac{1}{\hat{\epsilon}^2}\sum_{i=0}^{n-1}(f_i^N)^2\Delta\right] \mid \mathcal{F}_t^y\right\} \\
&\leq E^0\left\{\exp\left[\frac{2}{\hat{\epsilon}^2}\sum_{i=0}^{n-1} f_i^N \Delta y_{i+1}\right] \mid \mathcal{F}_t^y\right\} \leq K(y)
\end{aligned}
\tag{37}
$$

Next we come to (35). Using $|e^x - e^y| \leq |x-y|\,|e^x + e^y|$ and (34) we obtain

$$
\begin{aligned}
E^0\left\{\mid z_t - z_t^N \mid \mid \mathcal{F}_t^y\right\} \leq\ & K(y)\left[E^0\left\{\left(\mid \int_0^t \left(f_s - f_s^N\right) dy_s \mid^2\right.\right.\right. \\
&+ \left.\left.\left.\tfrac{1}{4}\mid \int_0^t \left(f_s^2 - (f_s^N)^2\right) ds \mid^2\right) \mid \mathcal{F}_t^y\right\}\right]^{1/2}
\end{aligned}
\tag{38}
$$

By Proposition 3.1 , the fact that without loss of generality we may assume $\Delta < 1$, and the independence, under P^0, of the processes X_t and y_t, it suffices to show that

$$
E^0\left\{\mid \int_0^t \left(f_s - f_s^N\right) dy_s \mid^2 \mid \mathcal{F}_t^y\right\} \leq \bar{K}(y)\,\Delta \tag{39}
$$

For this purpose, putting $\Delta y_{i+1} = y_{(i+1)\Delta} - y_{i\Delta}$, we use the stochastic integration by parts formula as well as the fact that

$$
f_n y_{n\Delta} = \sum_{i=0}^{n-1} f_{i+1}\Delta y_{i+1} + \sum_{i=0}^{n-1} y_i \Delta f_{i+1} \tag{40}
$$

together with $y_0 = 0$ to obtain, for $t = n\Delta$, $(n = t/\Delta)$,

$$
\begin{aligned}
&| \int_0^{n\Delta} \left(f_s - f_s^N\right) \, dy_s \, |^2 = | \, y_{n\Delta} f_{n\Delta} - \int_0^{n\Delta} y_s df_s - \sum_{i=0}^{n-1} f_i \Delta y_{i+1} \, |^2 \\
&= | \, y_{n\Delta} f_{n\Delta} - \int_0^{n\Delta} y_s df_s - \sum_{i=0}^{n-1} f_{i+1} \Delta y_{i+1} + \sum_{i=0}^{n-1} \Delta f_{i+1} \Delta y_{i+1} \, |^2 \\
&\leq K \mid y_{n\Delta}(f_{n\Delta} - f_n) \, |^2 + K \, | \sum_{i=0}^{n-1} \left(y_{i+1} \Delta f_{i+1} - \int_{i\Delta}^{(i+1)\Delta} y_s df_s \right) |^2 \qquad (41) \\
&\leq K_1(y) \left(f_{n\Delta} - f_n\right)^2 + K_1 \, | \sum_{i=0}^{n-1} \left(y_{i+1} \Delta f_{i+1} - \int_{i\Delta}^{(i+1)\Delta} y_{i+1} df_s \right) |^2 \\
&\qquad + K_2 \, | \sum_{i=0}^{n-1} \int_{i\Delta}^{(i+1)\Delta} \left(y_{i+1} - y_s\right) df_s \, |^2 = I + II + III
\end{aligned}
$$

To obtain (39) it suffices now to show that a similar relation holds when replacing the $| \int_0^t \left(f_s - f_s^N\right) \, dy_s \, |^2$ there by the expressions corresponding to I, II, and III respectively.

By Proposition 3.1 the expression corresponding to I is immediately seen to be bounded by $\bar{K}_1(y)\Delta$ for a suitable $K_1(y)$.

For the expression corresponding to II we have

$$
\begin{aligned}
&E^0 \left\{ | \sum_{i=0}^{t/\Delta - 1} \left(y_{i+1} \Delta f_{i+1} - \int_{i\Delta}^{(i+1)\Delta} y_{i+1} df_s \right) |^2 \mid \mathcal{F}_t^y \right\} \\
&= E^0 \Bigg\{ \sum_{i=0}^{t/\Delta - 1} y_{i+1} \Big[F_i^{(1)}(X_i)\Delta + G_i^{(1)}(X_i)\Delta w_i \\
&\qquad - \int_{i\Delta}^{(i+1)\Delta} F_s^{(1)}(X_s) ds - \int_{i\Delta}^{(i+1)\Delta} G_s^{(1)}(X_s) dw_s \Big] \mid \mathcal{F}_t^y \Bigg\}^2 \\
&\leq 2 \, E^0 \left\{ \sum_{i=0}^{t/\Delta - 1} y_{i+1} \left(\int_{i\Delta}^{(i+1)\Delta} \left(F_i^{(1)}(X_i) - F_s^{(1)}(X_s) \right) ds \right) \mid \mathcal{F}_t^y \right\}^2 \\
&\qquad + 2 \, E^0 \left\{ \sum_{i=0}^{t/\Delta - 1} y_{i+1} \left(\int_{i\Delta}^{(i+1)\Delta} \left(G_i^{(1)}(X_i) - G_s^{(1)}(X_s) \right) dw_s \right) \mid \mathcal{F}_t^y \right\}^2 \\
&\leq 2 \, \max_{i,j} |y_i \cdot y_j| \, E^0 \Bigg\{ \sum_{i=0}^{t/\Delta - 1} \left(\int_{i\Delta}^{(i+1)\Delta} L_F \left(\Delta + \|X_i - X_s\|\right) ds \right)^2 \\
&\qquad + \sum_{i,j=0}^{t/\Delta - 1} \left(\int_{i\Delta}^{(i+1)\Delta} L_F \left(\Delta + \|X_i - X_s\|\right) ds \right) \cdot \left(\int_{j\Delta}^{(j+1)\Delta} L_F \left(\Delta + \|X_j - X_s\|\right) ds \right) \\
&\qquad + \sum_{i=0}^{t/\Delta - 1} \left(\int_{i\Delta}^{(i+1)\Delta} L_G^2 \left(\Delta + \|X_i - X_s\|\right)^2 ds \right) \Bigg\}
\end{aligned} \tag{42}
$$

where we have used the fact that, under P^0, the processes X_t and y_t are independent so that conditioning on $\mathcal{F}_t^y$ is equivalent to fixing a trajectory of y. Furthermore, we have used the global (also with respect to the time variable) Lipschitzianity of $F_t^{(1)}(\cdot)$ and $G_t^{(1)}(\cdot)$ (Lipschitz constants L_F and L_G respectively) and for the rightmost part we computed the expectation of the conditional expectation exploiting the property that

$$E^0 \left\{ \int_{i\Delta}^{(i+1)\Delta} \left(G_i^{(1)}(X_i) - G_s^{(1)}(X_s) \right) dw_s \mid \mathcal{F}_{i\Delta} \right\} = 0 \quad \text{and that}$$

$$E^0 \left\{ \left(\int_{i\Delta}^{(i+1)\Delta} \left(G_i^{(1)}(X_i) - G_s^{(1)}(X_s) \right) dw_s \right)^2 \right\}$$
$$= \int_{i\Delta}^{(i+1)\Delta} E^0 \left\{ \left(G_i^{(1)}(X_i) - G_s^{(1)}(X_s) \right)^2 \right\} ds$$

Notice next that, for $s \in [i\Delta, (i+1)\Delta)$ we have $X_s^N = X_i$ so that $\|X_i - X_s\| = \|X_s^N - X_s\|$ and therefore, by Proposition 3.1 , $E^0 \{\|X_i - X_s\|\} \leq K\sqrt{\Delta}$, $E^0 \{\|X_i - X_s\|^2\} \leq K\Delta$. Assuming without loss of generality that $\Delta < 1$, we can then continue the above relation (42) to become

$$\begin{aligned} \textit{expression}\, II \ &\leq K(y) \left[L_F^2 \left(\Delta^2 + 2\Delta^{5/2} + \Delta^3 \right) + L_F^2 \left(\Delta^2 + 2\Delta^{5/2} + \Delta^3 \right) \right. \\ &\left. + L_G^2 \left(\Delta + 2\Delta^{3/2} + \Delta^2 \right) \right] \leq \bar{K}(y) \cdot \Delta \end{aligned} \tag{43}$$

for suitable $K(y)$, $\bar{K}(y)$ depending on the observed trajectory of y.

Finally, for the expression corresponding to III we have

$$\begin{aligned} & E^0 \left\{ \mid \sum_{i=0}^{n-1} \int_{i\Delta}^{(i+1)\Delta} (y_{i+1} - y_s)\, df_s \mid^2 \mid \mathcal{F}_t^y \right\} \\ &= E^0 \left\{ \mid \sum_{i=0}^{n-1} \int_{i\Delta}^{(i+1)\Delta} (y_{i+1} - y_s) \left[F_s^{(1)}(X_s) ds - G_s^{(1)}(X_s) dw_s \right] \mid^2 \mid \mathcal{F}_t^y \right\} \\ &\leq 2\, E^0 \left\{ \left(\sum_{i=0}^{n-1} \int_{i\Delta}^{(i+1)\Delta} |y_{i+1} - y_s| \mid F_s^{(1)}(X_s) \mid \, ds \right)^2 \mid \mathcal{F}_t^y \right\} \\ &\qquad + 2\, E^0 \left\{ \left(\sum_{i=0}^{n-1} \int_{i\Delta}^{(i+1)\Delta} (y_{i+1} - y_s) G_s^{(1)}(X_s) dw_s \right)^2 \mid \mathcal{F}_t^y \right\} \\ &\leq \tilde{K}\Delta \left[T\, ||F^{(1)}||^2 + T\, ||G^{(1)}||^2 \right] \leq \bar{K} \cdot \Delta \end{aligned} \tag{44}$$

where we have used assumption A.1 and the boundedness of $F^{(1)}(\cdot)$ and $G^{(1)}(\cdot)$ (norms $||F^{(1)}||$ and $||G^{(1)}||$). □

Theorem 3.3 *For each $n = 0, 1, ..., N$, for $t = n\Delta$, for each observed trajectory y_s, $s \leq t$ satisfying A.1 and for each value of θ*

$$\left| V_t \left(\Gamma_\theta; y \right) - V_t^N \left(\Gamma_\theta; y \right) \right| \leq K_1(y) \Delta^{\frac{1}{2}}. \tag{45}$$

where $K_1(y)$ depends only on the observed trajectory y_s, $s \leq t$.

Proof. We have

$$\begin{aligned} \left| V_t \left(\Gamma_\theta; y \right) - V_t^N \left(\Gamma_\theta; y \right) \right| &\leq E^0 \left\{ \left| \Gamma_\theta(X_t) z_t - \Gamma_\theta(X_t^N) z_t^N \right| |\mathcal{F}_t^y \right\} \\ \leq E^0 \left\{ z_t \left| \Gamma_\theta(X_t) - \Gamma_\theta(X_t^N) \right| |\mathcal{F}_t^y \right\} &+ E^0 \left\{ \left| \Gamma_\theta(X_t^N) \right| \left| z_t - z_t^N \right| |\mathcal{F}_t^y \right\} \end{aligned} \tag{46}$$

Applying Hölder's inequality and the fact that $\Gamma_\theta(\cdot)$ is, uniformly in θ (recall that θ takes a finite number of values) Lipschitz (with $L-$constant $\hat{\Gamma}$) and bounded (by $\tilde{\Gamma}$), (46) is majorized by

$$\left(E^0\left\{z_t^2|\mathcal{F}_t^y\right\}\right)^{\frac{1}{2}}\hat{\Gamma}\left(E^0\left\|X_t-X_t^N\right\|^2\right)^{\frac{1}{2}}+\tilde{\Gamma}E^0\left\{\left|z_t-z_t^N\right||\mathcal{F}_t^y\right\}. \tag{47}$$

By Lemma 3.2 and Proposition 3.1 we then obtain the thesis. □

Remark 3.4 *Theorem 3.3 implies convergence of the filter for each observed trajectory. This is a stronger form of convergence than those in the traditional filtering literature (see e.g.[19]), where convergence is obtained in the mean with respect to y.*

Consider next the sequence of nonnegative measures $q_n(B;y^n)$, where B denotes the generic Borel subset of $\mathcal{X}$ and $y^n=(y_1^\Delta,\cdots,y_n^\Delta)$ with $y_n^\Delta:=y_{n\Delta}-y_{(n-1)\Delta}$, and that are recursively defined by

$$\begin{aligned}&q_0(B):=p_0(B)\\&q_{n+1}\left(B;y^{n+1}\right):=\int_B\int_{\mathcal{X}}\exp\left[\frac{1}{\hat{\epsilon}^2}f_ny_{n+1}^\Delta-\frac{1}{2\hat{\epsilon}^2}f_n^2\Delta\right]P\left(X_{n+1}\left|X_n\right.\right)\,dq_n\left(X_n;y^n\right)\,dX_{n+1}\end{aligned} \tag{48}$$

where p_0 is the initial distribution and f_n corresponds to $X_n^{(1)}$, which is also the same as f_t^N in (26) and (30).

Proposition 3.5 *For any bounded function Ψ we have*

$$E^0\left[\Psi\left(X_n\right)z_T^N\left|\mathcal{F}_{n\Delta}^y\right.\right]=\int_{\mathbf{X}}\Psi\left(X\right)\,dq_n\left(X;y^n\right). \tag{49}$$

For a proof see e.g. [1].

Applying this proposition we immediately obtain (writing V_n^N for the $V_{n\Delta}^N$ (31))

$$V_n^N\left(\Gamma_\theta;y\right)=\int_{\mathbf{X}}\Gamma_\theta\left(X\right)\,dq_n\left(X;y^n\right) \tag{50}$$

for $n=0,1,...,N$ and this also implies that, when computing $V_n^N(\Gamma_\theta;y)$, we do not lose information by considering only y^n instead of the entire filtration $\mathcal{F}_{n\Delta}^y$.

Using (50) and (27) it is easily seen that the measures $q_n(B;y^n)$ can be given the interpretation of unnormalized conditional distributions. To determine the time discretized filter it suffices thus to compute the recursions (48). This is still an infinite-dimensional problem and so further approximations are needed, specifically discretizations in the spatial variable (quantization). This can be done in a variety of ways, for which we refer e.g. to [1],[12], [17], [18], [20],[22], [23]. In particular, for problems that are already reduced to discrete time, in [20],[22], [23] a specific methodology is described to arrive at a finite-dimensional approximating filter. Alternatively, always for problems already in discrete time, one could also use the recent so-called "particle approach" to nonlinear filtering, that is based on a simulation methodology (see e.g.[11]).

4 Conclusion

We have considered a version of the Heath-Jarrow-Morton model with a volatility depending on time-to-maturity, the instantaneous spot rate and one fixed maturity forward rate. We have seen how estimation of this model may be set up as a non-linear filtering problem under the historical measure. We have proposed a framework in which a recursive (Bayesian-type) filtering algorithm may be developed.

We have provided convergence results that demonstrate the consistency of the discretized filtering model with the original continuous time counterpart.

Future research needs to focus on actual implementation of the filtering framework proposed here. Results of [3] using a recursive (Bayesian) filtering algorithm for estimation in a model of the instantaneous spot rate of interest indicate the feasibility of this general approach.

References

[1] A.Bensoussan, W.Runggaldier, *An approximation method for stochastic control problems with partial observation of the state - a method for constructing $\epsilon-$optimal controls*, Acta Appl. Math., 10 (1987), 145-170.

[2] R.Bhar, C.Chiarella, N.El-Hassan, X.Zheng, *The Reduction of Forward Rate Dependent Volatility HJM Models to Markovian Form: Pricing European Bond Options*, Journal of Computational Finance, 3(3), 47-72, 2000.

[3] R. Bhar, C. Chiarella, W. Runggaldier, *Estimation in Models of the Instantaneous Short Term Interest Rate by use of a Dynamic Bayesian Algorithm*, QFRG Working Paper, University of Technology, Sydney, 2001.

[4] T.Björk, *Arbitrage Theory in Continuous Time*, Oxford University Press, 1998.

[5] T.Björk, *A geometric view of interest rate theory*, SSE/EFI Working Paper Series in Economics and Finance No 419, Dec. 2000. To appear in *Handbook of Mathematical Finance*, Cambridge University Press.

[6] T.Björk, L.Svensson, *On the existence of finite dimensional realizations for nonlinear forward rate models*, Mathematical Finance, 11 (2), 205-243, 2001.

[7] D.A.Chapman, J.B.Long Jr., N.D.Pearson, *Using Proxies for the Short Rate : When Are Three Months Like an Instant ?*, Review of Financial Studies, 12, 763-806, 1999.

[8] C.Chiarella, O.K.Kwon, *Forward rate dependent Markovian transformations of the Heath-Jarrow-Morton term structure model*, Finance and Stochastics, 5, 237-257, 2001.

[9] C.Chiarella, O.K.Kwon, *Classes of Interest Rate Models under the HJM Framework*, Asia Pacific Financial Markets, 8, 1-22, 2001.

[10] G.Deelstra, F.Delbaen, *Convergence of Discretized Stochastic (Interest Rate) Processes with Stochastic Drift Term*, Appl. Stoch. Models Data Anal. 14, 77-84, 1998.

[11] P.Del Moral, *Measure valued processes and interacting particle systems. Applications to nonlinear filtering problems*, Ann.Appl.Probab., 8, 438-495, 1998.

[12] G.B. Di Masi, M.Pratelli, W.J.Runggaldier, *An Approximation for the Nonlinear Filtering Problem, with Error Bound*, Stochastics, 14, 247-271, 1985.

[13] B.Flesaker, L.P. Hughston, *Positive interest*, Risk 9, 46-49, 1996.

[14] D.Heath, R.Jarrow, A.Morton, *Bond Pricing and the Term Structure of Interest Rates: A New Methodology for Contingent Claim Valuation*, Econometrica, 60(1),77-105, 1992.

[15] T.S.Y.Ho, S.B.Lee,*Term Structure Movements and Pricing Interest Rate Contingent Claims*,Journal of Finance, 1011-1029, 1986 .

[16] G.Kallianpur, C.Striebel, *Estimation of stochastic processes. Arbitrary system processes with additive white noise observation error*, Ann. Math. Stat., 39 (1968), 785-801.

[17] H. Korezlioglu, W.J. Runggaldier, *Filtering for Nonlinear Systems Driven by Nonwhite Noises : an Approximation Scheme*, Stochastics and Stochastics Reports, 44 (1993), 65-102.

[18] H.J.Kushner, P.Dupuis, *Numerical Methods for Stochastic Control Problems in Continuous Time*, Springer-Verlag, New York 1992.

[19] R.S. Liptser, O. Zeitouni, *Robust Diffusion Approximation for Nonlinear Filtering*, J.Math.Systems, Estimation and Control, 8 (1998), 1-22.

[20] S. Pasquali, W.J. Runggaldier, *Approximations of a controlled diffusion model for renewable resource exploitation*, Kluwer Volume on Markov Processes and Controlled Markov Chains, in print.

[21] P.Ritchken, L.Sankarasubramanian, *Volatility structures of forward rates and the dynamics of the term structure*, Mathematical Finance, 5 (1), 55-72, 1995.

[22] W.J. Runggaldier, *On the construction of $\epsilon-$optimal strategies in partially observed MDP's*, Annals of Oper. Res., 28, 81-96, 1991.

[23] W.J. Runggaldier, O. Zane, *Approximations for Discrete-Time Adaptive Control : Construction of $\epsilon-$Optimal Controls*, Math. Control Signals Systems, 4, 269-291, 1991.

[24] O.Vasicek, *An Equilibrium Characterisation of the Term Structure*, Journal of Financial Economics, 5, 177-188, 1977.