Exile on Main Street –
Excursions in Fixed Income Modelling

Rolf Poulsen
Department of Operations Research
Departments of Mathematical Sciences, Building 530
University of Aarhus
DK-8000 Aarhus C, Denmark

July 2, 1999

PhD Thesis

Submitted to the Faculty of Science, University of Aarhus

Let no one else's work evade your eyes,
Remember why the good Lord made your eyes,
So don't shade your eyes,
But plagiarize, plagiarize, plagiarize ...
Only be sure always to call it please research.

-Tom Lehrer in *The Lobchevsky Song*

I have left out many of the things which could have been added so as to make the practice of the analysis more easy. I can assure you, nevertheless, that I have omitted all that quite deliberately since I felt sure that some people who boast that they know everything would not miss the chance of saying that they knew already what I had written, if I had made myself easily intelligible to them.

-Rene Descartes (as quoted by Hersh, *Math. Intell.*, 1997)

Contents

Paper II: A simple regime switching term structure model, 28 pages

Paper III: Stability of Derivative Prices in Market Models, 30 pages

Paper IV: A Comparison of Approximation Techniques for Transition Densities of Diffusion Processes, 57 pages

Paper V: Approximate Maximum Likelihood Estimation of Discretely Observed Diffusion Processes, 33 pages

Paper VI: Optimal Martingale and Likelihood Methods for Models of the Short Rate of Interest, With Monte Carlo Evidence for the CKLS Specification and Applications to Non-Linear Drift Models, 46 pages

Paper VII: Should He Stay or Should He Go? Estimating the Effect of Sacking the Manager in Association Football, 8 pages

1 Introduction

How to Read This Thesis

This thesis is the outcome of my 4-year PhD study at the Department of Operations Research at the University of Aarhus. The thesis has three parts. The first part, which you are reading now, is a brief non-technical overview (mindmap, summaries of included manuscripts) of the thesis. The second part is a survey of the relevant literature and how my work relates to, contributes to, and extends the various fields. Cynics would say that the second part exists so that you will not have to read the third part. The survey has two subparts. One is on issues from mathematical finance, fixed income models in particular, and one is on estimation of discretely observed diffusion processes. It is the intention that each subpart has a chronologically logic structure. This, however, does not mean that my work was done in that order since the lines of thought are not always straight. The third part of the thesis contains the "original manuscripts". These can – and preferably should – be read seperately. This means that there is some (but not tremendously much) overlap between the papers and that I have made no attempt to harmonize notation across manuscripts.

How to Avoid Reading This Thesis

Place the thesis on your desk/shelf, right next to the other n things you really must read. As time passes n will tend to infinity. (And do so frighteningly fast.)

Acknowledgements

Many people have provided me with valuable input during my PhD study. I thank Morten Bai Andersen, Jesper Andreasen, Ken Bechmann, Jochen Beisser, Tomas Björk, Claus Vorm Christensen, Asbjørn Hansen, Peter Honoré, Bjarke Jensen, Peter Løchte Jørgensen, Jesper Lund, Marek Musiela, Ken Nyholm, Jan Pedersen, Jesper Lund Pedersen, Michael Sørensen, Ole Østerby, and last but not least my advisors Bent Jesper Christensen and Jørgen Aase Nielsen.

Summaries of the Papers in the Thesis

A picture can say more than a thousand words. In this case I am content if Figure 1 can say roughly the same as the (according to the word-counting program `wc`) next 591 words.

Refereed Publications

Andreasen, Jesper, Bjarke Jensen, and Rolf Poulsen (1998), "Eight Valuation Methods in Financial Mathematics: The Black-Scholes Formula as an Example", Mathematical Scientist, Vol 23(1), pp. 18-40.
ABSTRACT: This paper describes a large number of valuation techniques used in modern financial mathematics. Though the approaches differ in generality and rigour, they are consistent in a very noteworthy sense: each model has the celebrated Black-Scholes formula for the price of a call-option as a special case.

Working Papers

"Stability of Derivative Prices in Market Models"
Describes the "market model" lognormal LIBOR specification that has recently enjoyed much success in term structure modelling. A simulation algorithm based on measure relations is developed and is applied in order to justify some of the approximations used when deriving closed-form expressions for advanced fixed income derivatives.

"Approximate Maximum Likelihood Estimation of Discretely Observed Diffusion Processes" (CAF working paper no. 29, submitted to *Econometric Theory*)
A new estimation technique for discretely observed diffusion processes based on numerical solution of a partial differential equation related to the likelihood function is developed, implemented and applied. The estimator is shown to be asymptotically equivalent to the maximum likelihood estimator, a computationally optimal discretization rule is found, and the CKLS short rate model is estimated.

"A simple regime switching term structure model" (joint with A. T. Hansen) (CAF working paper no. 30, conditional acceptance at *Finance and Stochastics*)

We extend the classical Vasicek Ornstein-Uhlenbeck short rate model by allowing the unconditional mean parameter to exhibit jumps. Decomposition formulae for bonds and bond options are found and an efficient numerical method for price calculations is implemented.

"A Comparison of Approximation Techniques for Transition Densities of Diffusion Processes" (joint with B. Jensen)
We give detailed description of many of the competing techniques for approximating the densities of a diffusion process. A numerical comparison of the techniques for some models with known transition densities gives a remarkably clear ordering based on speed/accuracy trade-off considerations.

"Optimal Martingale and Likelihood Methods for Models of the Short Rate of Interest, With Monte Carlo Evidence for the CKLS Specification and Applications to Non-Linear Drift Models " (joint with B. J. Christensen)
New improved estimation methods for discretely observed diffusion models for the short rate of interest are introduced. We consider both optimal martingale estimating equations and maximum likelihood methods based on second order convergent numerical solution of the forward partial differential equation for the transition density. The new methods are compared to well-known methods, namely GMM, Indirect Inference and GQML, both theoretically, in an application to U.S. data, and in Monte Carlo experiments. The benchmark model used for illustration and comparison is the CKLS short rate model. We find that the new martingale and likelihood methods reduce bias, true standard errors, and bias in estimated standard errors, relative to the established methods, in particular for the key parameter of interest, the elasticity of variance. In weekly data from 1982 to 1995, the new methods estimate this parameter to about 0.78. Finally, we use the approximate maximum likelihood method to estimate non-linear drift short rate models. We find the terms commonly suggested as drift augmentations insignificant.

"Should He Stay or Should He Go? Estimating the Effect of Sacking the Manager in Association Football" (submitted to *Chance*)
A not-so-serious paper where I use English data and a generalized linear model to empirically document that football clubs that are performing poorly should indeed sack their managers.

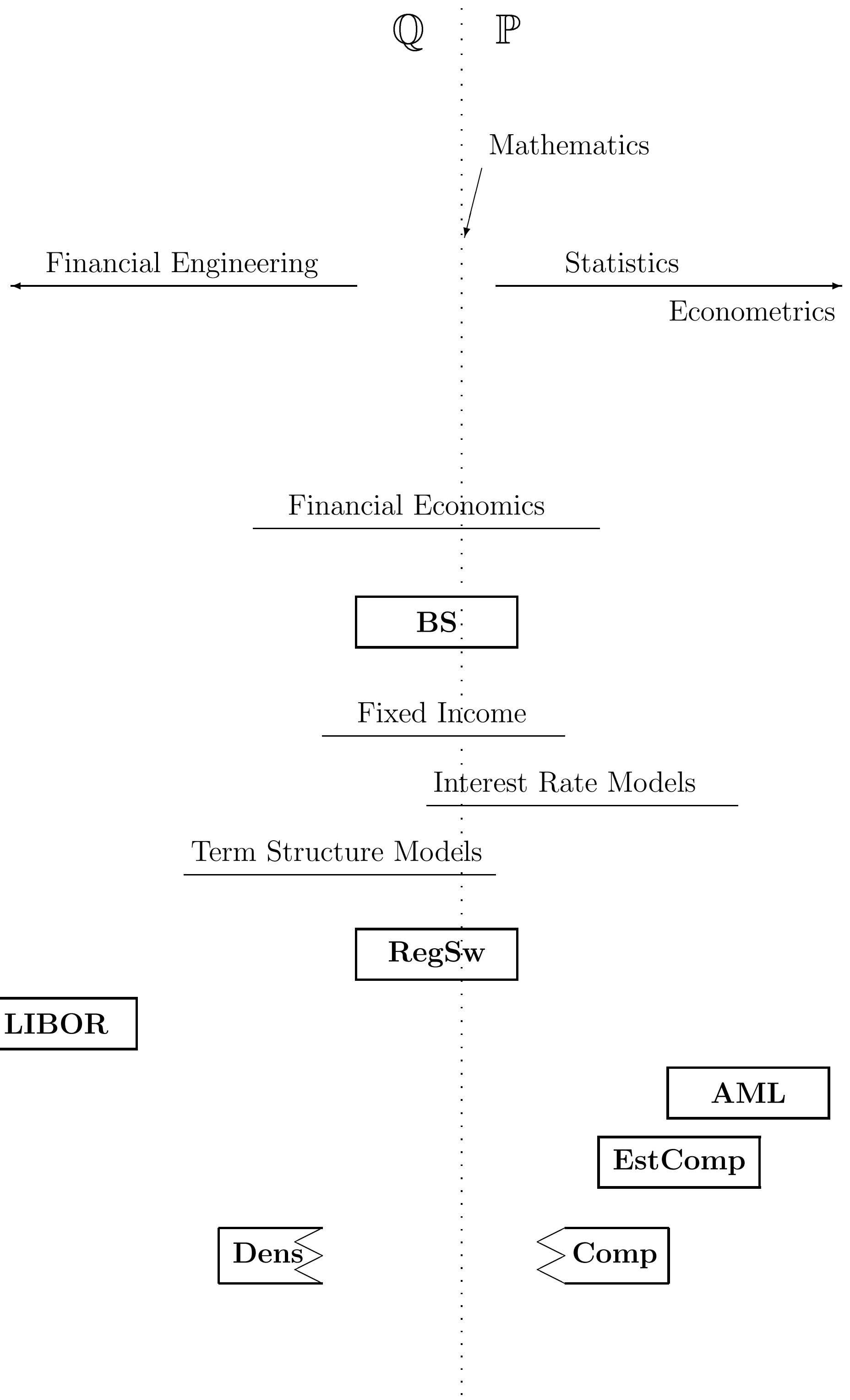

Figure 1: A mindmap of the thesis. Referring to the summaries, the abbreviations/acronyms should be obvious. It is not an error that **Sack** does not appear on the map.

2 Mathematical Finance

We take a guided tour of the main results of mathematical finance in general and fixed income modelling in particular (the "main street", in reference to the title of the thesis) with descriptions and appropriate placement of my own contributions/investigations. These are labelled "excursions", and have nothing to do with the probabilistic concept of that name (cf. Karatzas & Shreve (1992, Chapter 6)). An even more "pun intended"-term would be random field trips.

2.1 General Definitions and Results for Modelling Continuous Financial Markets

The results in this section are our fixed points when travelling on the high waters of mathematical finance. The "standard references" for this probabilistic/martingale approach are the seminal papers Harrison & Kreps (1979) and Harrison & Pliska (1981). These two references are a "must-have" for many papers; sometimes they are credited results that do not appear in the actual papers. But the blame for this lies solely on the "citer" not on the "citee". Therefore: The presentation in the rest of Section 2.1 is similar to that given in Jamshidian (1997) since this is the most self-contained description (but neither the most general one, nor the one most easy to understand).

Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0;T]}, \mathbb{P})$ be a filtered probability space satisfying the usual conditions. Let $\mathcal{S}$ denote the space of real continuous (so there goes the Poisson-process) semimartingales (so there goes fractional Brownian motion), $\mathcal{S}_+$ the space of strictly positive continuous semimartingales, and $\mathcal{S}^n$ and $\mathcal{S}^n_+$ their n-dimensional analogies. For $X, Y \in \mathcal{S}$, $< X, Y >$ denotes the quadratic covariation ($< X >=< X, X >$). For $X \in \mathcal{S}$, u^X denotes the additive compensator (i.e., $X - u^X$ is a local martingale), while for $X \in \mathcal{S}_+$, U^X is the multiplicative compensator (i.e., X/U^X is a local martingale). Finally, for $X \in \mathcal{S}$ we let $L(X)$ be the space of "processes that can reasonably be integrated w.r.t. X", i.e.,

$$L(X) = \left\{ \text{predictable process } Z : \int_0^T Z^2(s)d < X >_s + \int_0^T |Z(s)||du^X(s)| < \infty \;\; a.s. \right\}$$

We now give a financially motivated and carefully chosen definition.

Definition 2.1 *$S \in \mathcal{S}^n$ is called (locally) arbitrage-free if there exists $\xi \in \mathcal{S}_+$ (called a* state price deflator*) with $\xi(0) = 1$ such that*

$$\forall i \ : \ \xi S_i \textit{ is a (local) } \mathbb{P}\textit{-martingale.}$$

This implies that there are "no free lunches".

Theorem 2.1 *Let $C \in \mathcal{S}$ and $\xi \in \mathcal{S}_+$. Suppose ξC is a martingale and $C(T) > 0$ a.s. Then*

$$i) \ \ C(t) \geq 0 \ \ a.s. \textit{ for all } t,$$

and

$$ii) \ \ C(0) = 0 \ \ a.s. \ \Rightarrow C(t) = 0 \ a.s. \textit{ for all } t.$$

We are saving ourselves a lot of trouble by not trying to state the converse i.e., "something like Theorem 2.1" implies "something like Definition 2.1". The statement "Theorem 2.1 $\Leftrightarrow$ Definition 2.1" is called the *fundamental theorem of asset pricing* and does not hold in the general setting we use – but almost. See the discussion in Musiela & Rutkowski (1997, 10.1.5).

Locally arbitrage-free price systems and their state price deflators can be characterized from a linear constraint between the "volatilities" and the "expected rates of return".

Theorem 2.2 *Let $S \in \mathcal{S}^n_+$ and assume that there exist (pathwise bounded predictable) processes $V = [v_{ij}]$ and $\mu = [\mu_i]$ such that*

$$d < \ln S_i, \ln S_j >= v_{ij} dt,$$
$$d < U^{S_i} >= \mu_i dt.$$

Suppose there exist sufficiently regular processes $\lambda = [\lambda_j]$ and r such that

$$\mu = r\mathbf{1} + V\lambda \quad \textit{and} \quad \mathbf{1}^\top \lambda = 0.$$

Then S is locally arbitrage-free and the state price deflator is given by

$$\begin{aligned} \xi(t) &= \exp\left(-\sum_{j=1}^{n} \int_0^t \frac{\lambda_j}{S_j} dS_j + \int_0^t (\frac{1}{2}\lambda^\top V \lambda - r) ds \right) \\ &:= \exp\left(-\int_0^t r(s) ds \right) \eta(t). \end{aligned}$$

The main regularity condition *is that the above η process (which is always a non-negative local martingale, hence a supermartingale) is a martingale.*

Definition 2.2 *A pair (θ, S) where $S \in \mathcal{S}^n$ and $\theta \in L(S)$ is called a self-financing trading strategy if*

$$d(\theta^\top S) = \theta^\top dS.$$

Mathematically the above definition is general, and financially it makes good sense if we think of S as the price af a traded cash security with no intermediate cash flows. (As a rule there will be no dividends in this thesis. This is not completely without loss of generality, but dividends can often be treated with only minor adjustments, see Björk (1998, Chapter 11).)

The next theorem states when a claim can be hedged.

Theorem 2.3 *Let $S \in \mathcal{S}^n$ be arbitrage-free, $S_n > 0$ and ξS_n be a martingale. Suppose further that dU^ξ/dt and $d < S_i, S_j > /dt$ are bounded and that*

$$\mathit{rank}([d < S_i/S_n, S_j/S_n > /dt]) = d \ \ a.s.$$

Assume that there exists a d-dimensional Brownian motion $\widetilde{W}$, whose completed filtration we denote by $\widetilde{\mathcal{F}}$, such that S_i/S_n is $\widetilde{\mathcal{F}}$-adapted for all i. Finally, let $C(T)$ be an integrable and $\widetilde{\mathcal{F}}_T$-measurable random variable. Then there exists a self-financing strategy θ (called the replicating strategy*) such that $C(T) = \theta^\top(T)S(T)$ (and we say that $C(T)$ can be* hedged*) and $(\theta^\top S)/S_n$ is a martingale.*

Theorem 2.3 is basically a martingale representation theorem. The reason it may look a little more complicated than in most financial textbooks is that we have not assumed the existence of “a locally risk free asset”, an asset whose price process is of bounded variation.

Theorem 2.4 *Let $S \in \mathcal{S}^n$ be arbitrage free with state price deflator ξ. Let $A \in \mathcal{S}_+$ be such that ξA is a $\mathbb{P}$-martingale. Then $\xi A/A(0)$ is a non-negative martingale with mean 1 and we define the* numeraire measure *$\mathbb{P}^A \sim \mathbb{P}$ on $(\Omega, \mathcal{F}_T)$ by*

$$\frac{d\mathbb{P}^A}{d\mathbb{P}} = \frac{\xi(T)A(T)}{A(0)}.$$

Suppose that ξC is a $\mathbb{P}$-martingale. Then C/A is a $\mathbb{P}^A$-martingale,

$$\frac{C(t)}{A(t)} = \mathbb{E}_t^{\mathbb{P}^A}\left(\frac{C(T)}{A(T)}\right) \quad \textit{for all } t \leq T.$$

Theorem 2.4 is the well-known result that "no arbitrage implies existence of an equivalent martingale measure", but it also stresses that for any asset with positive price there is a measure such that prices discounted by this asset are martingales under that measure. This is commonly called "numeraire invariance", and when used in conjunction with Girsanov's Theorem it is a (surprisingly) powerful tool as we shall see. (For Girsanov's Theorem, as well as most of the other results from stochastic calculus used in finance, the reader is referred to Øksendal (1995); this is not the most prestigious reference (something French or Russian would be better) but a terribly effective one.)

2.2 The Black-Scholes Model

Combining Theorems 2.2 and 2.3 we get what might be termed "The Portmanteau Result of Contingent Claim Pricing".

Theorem 2.5 *Suppose we have an arbitrage-free economy with $n+1$ assets. One asset (the* savings account*) has a price process that is of bounded variation*

$$d\beta(t) = r(t)\beta(t)dt,$$

*where r is a sufficiently regular stochastic process. The other n assets (*stocks*) have a price process $S = (S_1, \dots, S_n)$ that follows*

$$dS(t) = I_S\mu(t)dt + I_S\Sigma^\top(t)dW(t),$$

where $I_S = diag(S_1, \dots, S_n)$, $\mu \in \mathbb{R}^n$, $\Sigma \in \mathbb{R}^{d\times n}$ (adapted processes) and W is a d-dimensional Brownian motion. Suppose that there exists a regular process ν such that

$$\mu - r\mathbf{1} = \Sigma^\top\nu \ \textit{ for all } t \textit{ and } \mathbb{P}-a.a.\ \omega.$$

Then there exists a measure $\mathbb{Q} \sim \mathbb{P}$ ($d\mathbb{Q}/d\mathbb{P}|_{\mathcal{F}_t} = \exp(\int_0^t \nu dW - 1/2\int_0^t \nu^\top \nu ds)$) and a $\mathbb{Q}$-Brownian motion $W^{\mathbb{Q}}$ ($dW^{\mathbb{Q}} = dW + \nu dt$) such that

$$dS = I_S r(t)dt + I_S\Sigma^\top(t)dW^{\mathbb{Q}}(t),$$

implying in particular that

$$S_i(t) = I_S\mathbb{E}_t^{\mathbb{Q}}\left(\exp\left(-\int_t^T r(s)ds\right)S_i(T)\right).$$

Suppose further that Σ is adapted to the Brownian filtration and has an invertible $d\times d$ submatrix. Then any claim in the filtration generated by β and S can be hedged.

And to avoid confusion we give the following definition.

Definition 2.3 *When we refer to* the martingale measure*, or write $\mathbb{Q}$, we mean (any of) the measure(s) $\mathbb{Q} \sim \mathbb{P}$ mentioned in Theorem 2.5, i.e., a measure such that all price processes discounted by the savings account are $\mathbb{Q}$-martingales.*

By the Feynman-Kac formula (cf. Øksendal (1995, Theorem 8.6)), Theorem 2.5 has a partial differential equation (PDE) version that applies to Markovian cases.

Theorem 2.6 *Suppose we have a $\mathbb{Q}$-Markovian setting,*

$$r = r(S,t), \qquad \Sigma = \Sigma(S,t).$$

Consider an asset whose terminal price is

$$C(T) = g(S(T)).$$

Then $C(t) = F(S(t),t)$ where $F : \mathbb{R}^{n+1} \mapsto \mathbb{R}$ solves the the PDE

$$rF = F_t + r\mathbf{1}^\top I_S F_S + \frac{1}{2} tr(I_S \Sigma^\top \Sigma I_S F_{SS}) \tag{1}$$

subject to the terminal condition $F(S(T),T) = g(S(T))$. Further, the replicating strategy is $\theta^\top = (\theta_\beta, \theta_{S\top})$ where

$$\theta_S = F_S$$

is the number of stocks held and

$$\theta_\beta = \frac{F - \theta_S^\top S}{\beta}$$

is the number of units of the locally risk-free asset held (so "we have $F - \theta_S^\top S$ \$ in the savings account").

We say that a PDE is *parabolic* if it involves a first derivative in "time", and the coefficient to the second derivative in "space" is positive definite - preferably strictly (actually, the statement is meaningless, but the reader should get the point). A PDE formulation as in Theorem 2.6 is important because parabolic PDEs lend themselves to numerical methods that can be much more rapidly converging than the simulation approach that Theorem 2.5 would inspire, see Wilmott, Dewynne & Howison (1998, Chapters 16-22), Duffie (1996, Chapter 11) or Strikwerda (1989).

Theorem 2.6 deals only with claims that have a simple payoff structure. However, the PDE approach can be applied to a much larger class. Typical ways of doing this are "Markovization by extra state variables" (works for example for Asian and Lookback options) and "reformulation of boundary conditions" (American and barrier options). For an excellent description of how to formulate exotic option pricing problems as PDEs, see Wilmott et al. (1998).

We now look at the classic Black-Scholes model (cf. Black & Scholes (1973) and Merton (1973)). It would be more accurate to call it the Black-Scholes-Merton (-Samuelson) model, but it is usual to compensate Merton by referring to more advanced versions of the model (with dividends and stochastic interest rates) with his name only. See Bernstein (1992, Chapter 11) for an interesting historic account of the Black-Scholes formula. The model consists of three assets.

i) A savings account with constant interest rate r, i.e.,

$$\beta(t) = \exp(rt).$$

ii) A non-dividend paying stock whose price follows a geometric Brownian motion

$$dS(t) = S(t)\mu dt + S(t)\sigma dW(t), S(0) = s(0),$$

where μ and $\sigma > 0$ are constants, so

$$S(t) = s(0)\exp((\mu - \sigma^2/2)t + W(t)).$$

iii) A call-option written on the stock with expiry date T and strike K, i.e., an asset whose price at time T is

$$C(T) = (S(T) - K)^+.$$

The question is then: What should the call-option cost? If C/β is a $\mathbb{Q}$-martingale then the economy is arbitrage-free, i.e., we should require that

$$C(t)/\beta(t) = \mathbb{E}_t^{\mathbb{Q}}(C(T)/\beta(T))$$

or in other words that

$$C(t) = e^{-r(T-t)}\mathbb{E}_t^{\mathbb{Q}}((S(T) - K)^+).$$

By the latter part of Theorem 2.5 there is a hedging strategy, so the above price is the only one that precludes arbitrage. Calculating the conditional mean leads to the celebrated Black-Scholes formula.

Theorem 2.7 *The unique arbitrage-free call-option price in the Black-Scholes model is $Call(t) = F(S(t), t)$ where*

$$\begin{aligned} F(x,t) &= x\Phi(z) - e^{-r(T-t)} K\Phi(z - \sigma\sqrt{T-t}) \\ z &= \frac{\ln(x/K) + (r + \frac{\sigma^2}{2})(T-t)}{\sigma\sqrt{T-t}} \end{aligned} \tag{2}$$

and Φ denotes the cumulative density of the standard normal distribution. In order to hedge a call-option we should hold

$$\theta_S(t) = F_x(S(t), t) = \Phi(z) \textit{ shares of stock}$$

and

$$C(S(t), t) - \theta_S(t)S(t) \textit{ \$ on the savings account.}$$

If you could pick only one result from mathematical finance, this would have to be it; the most astonishing thing at a first sight is that the $\mathbb{P}$-expected rate of return on the stock, μ, does not enter into the option price formula. The reason for this is that we are allowed to continuously readjust our portfolio and we are trying to determine the price of a derivative (the call-option) relative to a traded asset (the stock) that entails all the uncertainty in the economy ("no. assets = no. sources of risk = dim (W)").

Excursion; 8+ B-S proofs

The derivation of the Black-Scholes formula that was just outlined is "an 80'ies proof", but the formula is older, and financial mathematics has progressed since, so there must be other proofs. There are. In fact there are so many proofs that it is interesting to collect, harmonize, and compare them. This is done in Andreasen, Jensen & Poulsen (1998), where 8 methods of proof are given. The methods are all more or less sampled from the literature, and we refer the paper for complete references.

The Backward PDE Approach This is a direct derivation of the PDE (1) based on the following argument:

- Assume (reasonably) that the call price satisfies $Call(t) = F(S(t), t)$ for some smooth (but yet unknown) function F. Apply the Itô formula to find $dCall(t)$.

- Assume there exists a self-financing trading strategy, $\theta^\top = (\theta_0, \theta_1)$, in the savings account and the stock such that $\theta_0(t)\beta(t) + \theta_1(t)S(t) = Call(t)$ for all t. By the self-financing property and linearity of stochastic integration this gives another expression for $dCall(t)$.

- By a uniqueness theorem the drift and diffusion coefficients must be the same in the two $dCall$-expressions. Matching diffusion terms shows that $\theta_1 = F_x$ and matching drift terms gives the PDE (1), where we note that the call-option structure of the derivative is used "only" in the boundary condition $F(x,T) = (x-K)^+$.

- The PDE can be transformed into the heat equation and solved using a Fourier transform technique.

The Martingale Approach This is, more or less, the argument first outlined. The great advantage of this approach is that it does not hinge on the Markovian structure, neither in the stock price process nor in the derivative payoff.

The Numeraire Invariance Approach It is easy to see that

$$\frac{d\mathbb{Q}^S}{d\mathbb{Q}}|_{\mathcal{F}_t} = e^{-rt}\frac{S(t)}{S(0)}$$

defines a measure $\mathbb{Q}^S \sim \mathbb{Q}$. By Girsanov's Theorem $W^{\mathbb{Q}^S} = W^{\mathbb{Q}} - \sigma t$ is a Brownian motion under this measure, $dS = S(r+\sigma^2)dt + S\sigma dW^{\mathbb{Q}^S}$, and by the abstract Bayes rule (see Musiela & Rutkowski (1997, Lemma A.0.4), for instance) we have for any derivative with price process C that

$$\frac{C(t)}{S(t)} = \mathbb{E}_t^{\mathbb{Q}^S}\left(\frac{C(T)}{S(T)}\right).$$

The latter formula is useful for some options (e.g. Asian), but not here. Rather, consider WLOG $t = 0$ and use the general martingale approach to write

$$\begin{aligned} Call(0) &= \mathbb{E}^{\mathbb{Q}}(e^{-rT}S(T)\mathbf{1}_{\{S(T)>K\}}) - e^{-rT}K\mathbb{E}^{\mathbb{Q}}(\mathbf{1}_{\{S(T)>K\}}) \\ &= S(0)\mathbb{Q}^S(S(T)>K) - e^{-rT}K\mathbb{Q}(S(T)>K). \end{aligned}$$

Since S is also lognormally distributed under $\mathbb{Q}^S$ this yields the Black-Scholes formula without calculating a single integral.

A Proof Based on Local Time of Brownian Motion Consider for a moment the following trading strategy: If the present value of the strike price K is below the stock price hold one share of the stock. Finance this by using borrowed funds. If the stock price falls below the present value of the strike price liquidate the position. Formally, this means that $\theta_1(t) = \mathbf{1}_{\{S(t)>K\beta(T-t)\}}$ and $\theta_0(t) = -\mathbf{1}_{\{S(t)>K\beta(T-t)\}}K$. It is easy to see that the value of this strategy, say Y, satisfies

$$Y(t) = (S(t) - K\beta(T-t))^+ \text{ for all } t. \tag{3}$$

In particular, the terminal pay-off equals the call-option payoff and if $S(0) < K/\beta(T)$ the strategy cost nothing to initiate. Provided our general theory is correct, then this strategy cannot be self-financing. However, a careless application of the Itô formula to (3) could easily lead one to believe otherwise: The function is piecewise linear, so first derivatives are constants, second derivatives are 0, and self-financing follows immediately, right? Wrong! The function on the RHS of (3) is not even once differentiable. And even though this problem occurs only at a single value (0) it still renders an application of the Itô formula invalid. But all is not lost, the function is convex so we may apply a generalized Itô formula (as given in Karatzas & Shreve (1992, Theorem 6.22)) to find dY. When doing this, some rather lengthy calculations involving the local time of Brownian motion (basically the time a Brownian motion spends at 0) yields the correction term needed for self-financing to hold and eventually recovers the Black-Scholes formula.

A Proof Based on a Forward PDE If we let $\phi(x,T)$ denote the $\mathbb{Q}$-density of $S(T)$ in the point x given $S(0)$ we can write the martingale pricing relation for the call-option as

$$Call(0) = e^{-rT}\int_K^\infty (x-K)\phi(x,T)dx.$$

Combining this with the Fokker-Plank forward equation for ϕ (see Øksendal (1995, Chapter 8) or Section 3 in this paper) a number of partial integrations show that if we let $\widetilde{Call}(K,T)$ denote the initial price of a call-option with strike K and expiry T then $\widetilde{Call}$ solves the PDE

$$0 = -\frac{\partial \widetilde{Call}}{\partial T} - rK\frac{\partial \widetilde{Call}}{\partial K} + \frac{1}{2}\sigma^2 K^2 \frac{\partial^2 \widetilde{Call}}{\partial K^2} \tag{4}$$

subject to the initial boundary condition $\widetilde{Call}(K,0) = (S(0)-K)^+$. Solution of this PDE again yields to Black-Scholes formula. It should be stressed that the PDE (4), as opposed to (1) holds only for call-options. But by noting that (4) also holds for volatility of the form $\sigma = \sigma(S(t), t)$, we arrive at an important feature. Supposing that we can estimate the partial derivatives in (4) empirically from market data, then we can use this to get an estimate of the σ-function. This advanced implied modelling was first suggested in Dupire (1997) and has since been successfully extended and applied in for example Andersen & Brotherton-Ratcliffe (1998) and Andreasen (1996).

The B-S formula as the Limit of Binomial Models The binomial option pricing model suggested in Cox, Ross & Rubinstein (1979) is widely used both in trading and as an illustrative class-room tool to explain the fundamental concepts of arbitrage pricing. That call-option prices in the binomial model (can be made to) converge to the Black-Scholes price can by shown by using the put-call parity and a Lindeberg-Feller version of the Central Limit Theorem.

It is also possible to do
A Proof Based on The Consumption Based Capital Asset Pricing Model
and
A Proof Based on a Representative Investor
But since we have not developed the notation and terminology for a short and accurate description of these approaches in this paper we refer to Andreasen et al. (1998). It is, however, interesting to note that Black and Scholes first derivation of the fundamental pricing PDE was based on the standard capital asset pricing model.

More Proofs It has later been brought to our attention by John van der Hoek that the above list is not complete.

- In van der Hoek (1998) the formula is derived using a probabilistic approach, but without the mentioning of "market prices of risk", "measure changes" and "Girsanov's Theorem", in fact much like Section 2.1 (except Theorem 2.4) works. (And the derivation also gives the main idea in a proof of Girsanov's Theorem.)

- Mceneaney (1997) considers a robust control framework for option pricing. This framework in useful when dealing with market imperfections, e.g. stochastic volatility or (worse) transaction costs. But as a special case the Black-Scholes formula drops out.

2.3 Fixed Income Modelling

We now turn to the modelling of interest rates, bond prices and their derivatives and we shall use the term “fixed income modelling” to describe this. Without doubt, fixed income modelling is one of the most active fields of finance. There are many reasons for this. First, bond markets are (especially in Denmark) very large. Second, there is no empirical or theoretical agreement as to which model – or even which modelling framework – to use. And last but not least, the subject poses intriguing consistency problems, basically because there is (to a larger extent than in stock markets) a limit to how differently different bonds can behave; a fact that is reflected in Theorems 2.8 and 2.17. Berkeley professor William Keirstad expressed it thus “if one understands fixed income modelling, then everything else in finance is easy.”

For the rest of this thesis we shall uphold the following notation.

- $P(t,T)$ is the price at time t of a zero coupon bond (ZCB) maturing a time T (i.e., an asset whose sole payment is 1\$ at time T).

- The yield (to maturity) is $y(t,T) = -(\ln P(t,T))/(T-t)$

- The instantaneous forward rate with maturity date T contracted at time t is

$$f(t,T) = -\frac{\partial \ln P(t,T)}{\partial T},$$

  so $P(t,T) = P(t,s)\exp(-\int_s^T f(t,u)du)$ for all $t \leq s \leq T$. This rate can be interpreted as the interest rate one must pay on a loan between T and $T+dT$ when the loan in agreed upon at time t.

- The instantaneous short rate is

$$r(t) = f(t,t).$$

- The savings account is

$$\beta(t) = \exp\left(\int_0^t r(s)ds\right).$$

- By “the term structure (of interest rates)” we mean the mapping

$$T \mapsto P(t,T),$$

  or some deterministic translation hereof. So it may be ZCB prices (sometimes also called discount factors), yields or forward rates.

On an over-all scale there are two approaches to building a model for the fixed income market

- **The Indirect Approach**
  Here we specify the short rate and try to determine bond prices by arbitrage arguments.

- **The Direct Approach**
  Here we specify directly the all ZCB prices (or at least those of interest to us) and try to work from there.

The question "Which of the two approaches is the right one?" is roughly similar to "Which came first, the chicken or the egg?" They are not totally equivalent (see in the vicinity Theorem 2.20 for details), but on the other hand we cannot have one without the other. A further discussion of this topic will be given at the end of this section when "the smoke clears & we have done the math" (see also Rebonato (1998, 18.5)).

2.3.1 The Indirect Approach

The main result in the indirect approach is the following which may be derived form the results in Section 2.1 or found in any good text book, for instance Musiela & Rutkowski (1997, Proposition 12.2.1) or Björk (1998, Proposition 16.3).

Theorem 2.8 *Suppose we have an economy consisting only of a savings account,*

$$\beta(t) = r(t)\beta dt,$$

where r has dynamics,

$$dr(t) = \mu_r(t)dt + \sigma_r^\top(t)dW(t),$$

and W is a d-dimensional $\mathbb{P}$-Brownian motion. Now suppose ZCBs of all maturities are introduced in the economy. Then a necessary - and essentially sufficient - condition for the economy to remain arbitrage-free is the existence of a d-dimensional process ν such that with

$$\frac{d\mathbb{Q}}{d\mathbb{P}}|_{\mathcal{F}_t} = \exp\left(\int_0^t \nu^\top dW - \frac{1}{2}\int_0^t \nu^\top \nu ds\right) \tag{5}$$

we have

$$P(t,T) = \mathbb{E}_t^{\mathbb{Q}}\left(\exp\left(-\int_t^T r(s)ds\right)\right), \tag{6}$$

where the $\mathbb{Q}$*-dynamics of* r *is given by*

$$dr(t) = (\mu_r(t) - \sigma_r^\top(t)\nu(t))dt + \sigma_r^\top dW^{\mathbb{Q}}(t),$$

where $W^{\mathbb{Q}}$ *is a d-dimensional* $\mathbb{Q}$*-Brownian motion (*$dW^{\mathbb{Q}} = dW + \nu dt$*).*
Further, if μ_r *and* σ_r *are adapted to the Brownian filtration we have*

$$dP(t,T) = P(t,T)(\mu_P(t,T)dt + \sigma_P^\top(t,T)dW(t)),$$

where (for any vector $T = [T_1, \ldots, T_d]$ *(*$T_i > t$*) of maturity dates)* ν *solves*

$$\Sigma_P(t,T)\nu(t) = (\eta_P(t,T) - r(t)\mathbf{1}),$$

$\eta_P(t,T) = [\mu_P(t,T_1), \ldots, \mu_P(t,T_d)]^\top$*, and*

$$\Sigma_P(t,T) = \begin{bmatrix} \sigma_P^\top(t,T_1) \\ \vdots \\ \sigma_P^\top(t,T_d) \end{bmatrix}.$$

In other words, ν *is "universal".*

Theorem 2.8 tells us that ZCB prices are *not* uniquely determined by the $\mathbb{P}$-dynamics of the short rate, whereas by definition a specification of $\mathbb{Q}$-dynamics of the short rate enables us to calculate ZCB prices. A process of risk-premia, ν, is needed. These cannot be asset specific, have to satisfy certain integrability conditions, but can otherwise be arbitrary. But if we know the drift and volatility of d ZCBs then we can find ν. The determination/specification of ν is not "just an academic problem for pricing", the risk premia influence the $\mathbb{Q}$-dynamics of short rate, i.e., they effect the conditional mean we have to calculate to find bond prices. This is because we are trying to express bond prices by means of something that is not a traded quantity (the short rate), which means that we do not a priori know what its drift is under $\mathbb{Q}$.

Of course by the Feynman-Kac Formula, Theorem 2.8 has a PDE version in the case when r is a Markov process.

Theorem 2.9 *Let the general setting be as in Theorem 2.8, but assume further that* $r(t) = R(X(t))$*, where* X *follows*

$$dX(t) = \mu(t, X(t))dt + \Sigma^\top(t, X(t))dW(t).$$

Then $P(t,T) = F(X(t),t,T)$*, where* F *solves*

$$F_t + (\mu - \Sigma^\top \nu)^\top F_x + \frac{1}{2} tr(\Sigma^\top \Sigma F_{xx}) = RF, \quad F(\cdot, T, T) = 1. \tag{7}$$

The PDE version was initially established by Vasicek (1977) when he applied the hedge argument used by Black, Scholes and Merton to bonds. In fact Vasicek was very close to also giving the probabilistic version (cf. Eqn. (18) in Vasicek (1977)).

A recipe for a research paper involves one or more of the following ingredients:

a) Pick out some "factors", that you think explain bond prices.

b) Specify a description on the $\mathbb{P}$-dynamics of the factors, and how the factors influence the short rate.

c) Determine risk-premia, i.e., the ν-process. Ideally, this is a purely empirical question. In reality it usually amounts to either *i*) assuming them have a form that preserves the structure of the model (and then using them as "degrees of freedom" to fit observed bond prices once a ZCB pricing formula has been found) or, *ii*) making up a story as to why they should be 0.

d) Solve the PDE or alternatively calculate the conditional expectation, in order to find ZCB prices.

e) Find conditional expectations of more or less advanced functions of ZCB prices or the factors.

Referring to Figure 1 these points drift from right to left on the mindmap. The points a) and b) are empirical questions, c) is sort of in-between, and d) and e) gradually take us further and further into financial engineering.

We shall now give what we believe are the main results and findings in the literature.

Empirical Findings 2.1 (Factors in the Bond Market) *Approximately 3 factors give an adequate description of yield curve movements. The predominant of these factors is (closely related to) the short rate itself (in a Markovian sense).*

The analysis leading to the above findings are usually based on a regression approach or principal components analysis. The reader is referred to Steeley (1990), Litterman

& Scheinkman (1991) and Knez, Litterman & Scheinkman (1994). It may be tempting to use as "factors" some specific bond market related quantities, for example the yield of a particular bond. This may, however, introduce subtle consistency problems: Suppose that a factor is some (non-linear) function of ZCB prices. To calculate ZCB prices the factor dynamics (and risk premia) are needed. But by the Itô formula the factor dynamics can be derived from ZCB prices and in general there is no guarantee that the derived factor dynamics are the same as those initially postulated (and empirically verified). A problem of this nature, that is as of yet unresolved, arises in the Brennan & Schwartz (1979) two-factor model, where one of the state variables is taken as the yield of a consol bond, see Duffie, Ma & Young (1995).

But all in all, we may feel reasonably reassured that single-factor short rate models are a good place to start empirical investigations. A *very* frequently considered parametric model is the CKLS-equation (such acronym'ed after Chan, Karolyi, Longstaff & Sanders (1992))

$$dr_t = (\alpha + \beta r_t)dt + \sigma r_t^{\gamma} dW_t. \tag{8}$$

This specification has several virtues.

- It nests many of the commonly used short rate models.
- If $\beta < 0$ it has a (linear) mean reversion feature. This means that if the volatility is "not too nasty" (for precise, and somewhat surprising, statements see Conley, Hansen, Luttmer & Scheinkmann (1997)) then r_t has a limiting distribution (with mean $-\alpha/\beta$) as t tends to infinity.
- Linear mean reversion is analytically tractable; a standard trick is to remove r-dependency in the drift by using an "integrating factor", i.e., looking at $(\exp(-\beta t)r(t))$.

The conditional densities in the CKLS model are unknown (so likelihood inference is hard, but see Section 3); we do not even know conditional moments beyond the first (so consistent moment based inference is problematic). Despite this – or maybe because of this – a huge body of literature in which the CKLS model is estimated exists. The methods differ (some will be described in Section 3), the data differ, and results also differ. But we can summarize the "median findings".

Empirical Findings 2.2 (Estimation of the CKLS-model) *Many empirical studies have found the for the CKLS-model (8)*

i) β *is negative but not "significantly", i.e., there is weak evidence of linear mean reversion.*

ii) γ *is large, typically larger than 1, i.e., only weak solutions need exist and usual discretization schemes explode.*

Further, $-\alpha/\beta$ *is almost the sample mean and* σ *and* γ *are almost perfectly correlated.*

The reason we put " "-signs around significantly is that if β is 0 then standard asymptotic results used to derive confidence intervals cannot be applied (typically) because the model is non-stationary. An incomplete list of papers that discuss estimation of the CKLS-model is Chan et al. (1992), Broze, Scaillet & Zakoian (1995), Andersen & Lund (1996), Nowman (1997), Koedijk, Nissen, Schotman & Wolff (1997) and Honoré (1998*a*, Essay 1). (Quite possibly, every single author on this list will argue that Empirical Findings 2.2 do not reflect the finer points in his analysis.)

Point *ii*) was the first people tried to remedy, probably because many tools and models for this already existed. In the CKLS-model the short rate volatility depends only on the current level of the short rate. A high γ-value means that the volatility is very sensitive to the short rate level. It is easy to imagine that this could be induced if there was in fact "an extra stochastic term in from of 'dW"' that we were forced to explain solely by the short rate dependency. And that is how stochastic volatility (SV) models were born. The main problem with SV models is that the volatility is not directly observable (unless you can come up with a very good story).

SV models can be formulated both discretely (by some *GARCH* type process) as for example in Koedijk et al. (1997) or in a diffusion setting (by letting σ be governed by an SDE). Examples of this include Longstaff & Schwartz (1992), Andersen & Lund (1996). Despite these papers giving nice results, SV models have not nearly "caught on" as much in fixed income modelling as in the modelling of stock returns. A good reason for this is that for the purposes of ZCB price modelling the "stochastic component" of the short rate volatility is "only a third order effect" (the "first order effect" is the drift and the "second order effect" is the average volatility), see for example Schlögl & Sommer (1997).

Therefore, it should come as no surprise that point *i*) of Empirical Findings 2.2 has received much attention recently. At first it might appear to be "a good idea" that the short rate is a martingale. But at least as a property under $\mathbb{Q}$ it is far from desirable since it causes yields to tend to $-\infty$ as T tends to ∞. Ergo, something has to be done. And that is how non-linear drift models were born. The idea is that for "usual" interest rate levels the process behaves "as a martingale" but for very high or low levels the rate will have a very strong "pull" down- or upwards through basic economic forces ("the invisible hand" of Adam Smith 200+ years on?) Non-linear drift models are considered in Conley et al. (1997), Ait-Sahalia (1996), Stanton (1997), Pritsker (1998), Honoré (1998*a*, Essay 3) and Chapman & Pearson (1999) and discussed in further detail in the following excursion.

Excursion; Non-linear Drift

Consider a non-linear drift short rate model given by the SDE

$$dr(t) = \left(\frac{a_{-1}}{r} + \kappa(\theta - r) + a_2 r^2\right) dt + \sigma r^\gamma dW(t). \tag{9}$$

This model is the "parametric intersection" of the models proposed in Conley et al. (1997) (were the drift is allowed the more general form $\sum_{i=-l}^{k} a_i r^i$) and Ait-Sahalia (1996) (where the volatility is allowed the more general form $b_0 + b_1 r + b_2 r^{b_3}$).
Ait-Sahalia considers the following estimation procedure:

- Estimate the unconditional (or stationary or marginal) short rate density by a non-parametric kernel smoothing method; from this we get the estimate $\widehat{\phi}^{NP}$.

- For a diffusion process we have an explicit relation between the parametrized drift μ and volatility σ^2 and the stationary density (cf. Karlin & Taylor (1981, Section 15.6)),

$$\phi^\infty(x;\psi) = \frac{\xi(\psi)}{\sigma(x;\psi)} \exp\left(\int_{x^\#}^{x} \frac{2\mu(y;\psi)}{\sigma^2(y;\psi)} dy\right), \tag{10}$$

  where the lower limit of integration can be any interior point in the domain of the diffusion and ξ ensures that ϕ^∞ integrates to 1. (At this point one can see identification problems – but that is a minor point.)

- Given T observations, $r_1, \ldots, r_T$, estimate ψ by

$$\widehat{\psi} = \arg\min \frac{1}{T} \sum_i (\phi^\infty(r_i;\psi) - \widehat{\phi}^{NP}(r_i)).$$

  This estimator is consistent and $\sqrt{T}(\widehat{\psi} - \psi_0)$ is asymptotically normally distributed.

(This is in fact not a completely loyal description of what Ait-Sahalia does. He has some considerations about using the transition densities in estimation, but his major conclusions are based on the above method.) Using daily observations on U.S. short rate data covering the period 1973-95 (as well as the sub-sample 1982-95), Ait-Sahalia concludes that the non-linearities in the short rate drift specification offer significant improvement; he strongly rejects models with affine drift.

Later, the above approach has been criticized, primarily because its conclusions are based on asymptotic results for the estimators that - while not incorrect - do not give adequate descriptions of finite sample properties for relevant sample sizes. The asymptotic distribution of the non-parametric kernel estimator treats data as if they were independent and identically distributed, which means that in finite samples with persistent (positively autocorrelated) data standard deviations of estimates may be "too optimistic" (i.e., small). In fact Pritsker (1998) considers the Vasicek model, "reverses the question" and concludes that "to attain the accuracy of the kernel estimator implied by its asymptotic distribution with 22 years of data ... in fact requires 2255 years of data". Chapman & Pearson (1999) also address small sample behaviour. They give Monte Carlo evidence (for the CIR model) indicating that the Ait-Sahalia estimation technique tends to find non-linearities in the drift even where there are none. Further, they also find clear indications that Ait-Sahalia's standard errors are "too small to small samples" for realistic data, and stress co-linearity problems with the proposed drift specification.

In Christensen & Poulsen (1999) we also investigate non-linear drift specifications. However, we do it from a strictly parametric and likelihood based point of view. We use a numerical solution technique for the Fokker-Plank PDE to find the likelihood function numerically; this method will be described in detail in Section 3 (see also Poulsen (1999*a*)). We consider weekly observations of U.S. short rate data covering the period 1982-95. This is a different data-set than Ait-Sahalia, but a well-recognized one (it is used, among other places, in Andersen & Lund (1997)). Some results are reported in Table 1. We find that:

- The coefficients a_{-1} and a_2 are *not* significantly different from 0.

- Drift coefficient estimates are *highly* correlated.

- The discretization bias effect of using a simple Euler approximation to the transition density is small for weekly data (and thus even smaller in daily data.) (Cannot be seen from Table 1.)

- Using the uniform residuals suggested in Pedersen (1994) for model control, we find that the CKLS-model *is* indeed well-specified (the test-probability is about 13%). (Cannot be seen from Table 1.)

- The estimate of the γ-parameter is about 0.8 (with a standard error of about 0.07).

- But life is not a complete bed of roses; the estimate of the mean-reversion parameter is not "significantly" (see previous discussion about " "'s) different from 0.

All this indicates that (as it is put in Chapman & Pearson (1999)) "non-linearity of the short rate drift is not a robust stylized fact". But it is perhaps not so surprising that we are unable to detect the non-linearities with conventional statistical tools even if we use the best method we can think of (maximum likelihood). The very motivation behind the models is that the non-linearities should "kick in" for very large or very small interest rates. Or in other words for extremal events, so by definition we will have very few observations for which the non-linearities can be expected to have any effect. Notice also that the high persistence of high-frequency interest rate data means that data sets that "look big" really are not when it comes to drift estimation. So in short rate data sets of realistic sizes we would think it is very hard to find significant non-linearities. Therefore we find the approach taken in Honoré (1998*b*) where cross-sectional information (interest rates of 10 different maturities) is used the right way to proceed.

Estimator	Parameter estimate					
	a_{-1}	θ	κ	a_2	σ	γ
ψ^{AML}	0.0021	0.0530	2.315	-14.37	0.0955	0.7880
	(0.003)	(0.010)	(3.983)	(23.32)	(0.019)	(0.068)

Std. dev./correlation matrix of est. param.

0.003	0.200	-0.977	-0.950	-0.061	-0.067
0.200	0.010	0.010	0.115	-0.088	-0.097
-0.977	0.010	3.983	0.994	0.030	0.033
-0.950	0.115	0.994	23.322	0.032	0.034
-0.061	-0.088	0.030	0.032	0.019	0.994
-0.067	-0.097	0.033	0.034	0.994	0.068

Loglikelihood at non-linear ψ^{AML}	3347.8352 (mag. of num. err.: 0.1)
Loglikelihood at CKLS ψ^{AML}	3346.9673 (mag. of num. err.: 0.05)
Test probability of restricted model	42.0 % (in $\chi^2(2)$-dist.)

Table 1: Approximate Maximum Likelihood estimation of the non-linear drift short rate model (9) on U.S. weekly data covering 1982-95.

As pointed out earlier, indirect modelling (or more general any model with non-hedgeable claims, in the sense of Theorem 2.3) requires determination/specification of risk premia. The "high road" to achieving this would be empirical analysis ("the market decides which martingale measure is the right one", as Björk (1998) puts it) or introduction of *agent preferences* in the model – although we have never heard of anybody who knew their utility function despite the Von Neumann-Morgenstern axioms (cf. Kreps (1990)) being reasonable or Luenberger's humorous risk quiz (cf. Luenberger (1998, Figure 9.5)). In practice we just use a convenient specification. Therefore, considerable commotion ways stirred up when Cox, Ingersoll & Ross (1985, Section 5) gave an example of a seemingly harmless specification that lead to a model that was clearly not arbitrage-free. However, a careful inspection (cf. Morton (1988)) shows that this specification would cause the η-process in Theorem 2.2 not to be a martingale, hence $d\mathbb{Q}/d\mathbb{P}|_{\mathcal{F}_t}$ does not define a probability measure and no state price deflator exists. So the problem lies not with the "martingale measure pricing methodology". In fact, from Rogers (1995) we have the following reassuring result.

Theorem 2.10 *Suppose $S \in \mathcal{S}^n$ is locally arbitrage-free. Then there exists an economy for which S is an equilibrium price process.*

One might suspect the proof of Theorem 2.10 to be very abstract; for example to involve a martingale representation theorem. It is not; you basically just have to equip the (only) agent with logarithmic utility and an appropriate endowment process (which is also how a similar result is obtained from Exercise 10.3 in Duffie (1996).) Theorem 2.10 does not imply that there is no use for references to preferences (or: equilibrium based models). In models that are "grossly incomplete" it can be a simple and elegant way of finding/justifying pricing relations. For example, this is the case in "random jump size" models such as the Merton jump-diffusion model (cf. Merton (1976)) and many of the models used in insurance. Nonetheless, for the rest of Section 2.3 we shall unless clearly indicated focus on models formulated directly under measures relevant for pricing.

From Duffie & Kan (1996) we have the following major structural result.

Theorem 2.11 *Suppose that $r(t) = R(X(t))$, where X is a d-dimensional process satisfying*

$$dX(t) = \mu(X(t))dt + \Sigma^\top(X(t))dW^{\mathbb{Q}}(t),$$

where $W^{\mathbb{Q}}$ is a d-dimensional $\mathbb{Q}$-Brownian motion. Then under mild regularity and degeneracy conditions:

ZCB prices are exponentially affine, i.e., $\ln P(t,T) = A(T-t) + B(T-t)X(t)$

$\Updownarrow$

μ, $\Sigma^\top\Sigma$ *and R are affine functions.*

Further, A and B solve ODEs of the form

$$B'(t) = Q^B(B(t)), \quad B(0) = 0,$$
$$A'(t) = Q^A(B(t)), \quad A(0) = 0,$$

where $Q^A : \mathbb{R}^d \mapsto \mathbb{R}$ and $Q^B : \mathbb{R}^d \mapsto \mathbb{R}^d$ are quadratic functions.

The basic idea in the proof is quite simple, namely that if $\alpha + \beta^\top x = 0$ for all x in some open set, then $\alpha = 0$ and $\beta = 0$. This is then used in conjunction with the fundamental PDE (7).

Suppose in the notation of Theorem 2.11 we have $d = 1$ and $R(x) = x$. If we have

$$\mu(x) = \kappa(\theta - x), \quad \Sigma(x) = \sigma$$

say that we have a Vasicek model (after Vasicek (1977)).
If we have

$$\mu(x) = \kappa(\theta - x), \quad \Sigma(x) = \sigma\sqrt{x}$$

say that we have a Cox-Ingersoll-Ross (CIR) model (after Cox et al. (1985)).
These models have known closed form solutions for (among other things) ZCB prices (the ODEs in Theorem 2.11 are by far the easiest way of determining ZCB prices in the CIR model) and have been throughly analyzed, applied and extended.

Excursion; Jumping Mean

As the previous discussion indicates, recently there has been considerable focus on the modelling of the drift of the short rate. In Hansen & Poulsen (1999) we suggest a model that has some common features with the non-linear drift models (e.g. the possibility of several roots of the drift), but is quite different in other respects (two sources of risk, semi-closed ZCB price and ZCB option price formulae are available).

We extend the usual Brownian filtration setting in indirect models by letting the short rate dynamics depend also on a Poisson process. We are by no means the first to consider fixed income models with jumps; general treatments are given in Shirakawa (1991) and Björk, Kabanov & Runggaldier (1997), while specific models with a similar structure to the one we propose can be found in Naik & Lee (1994)

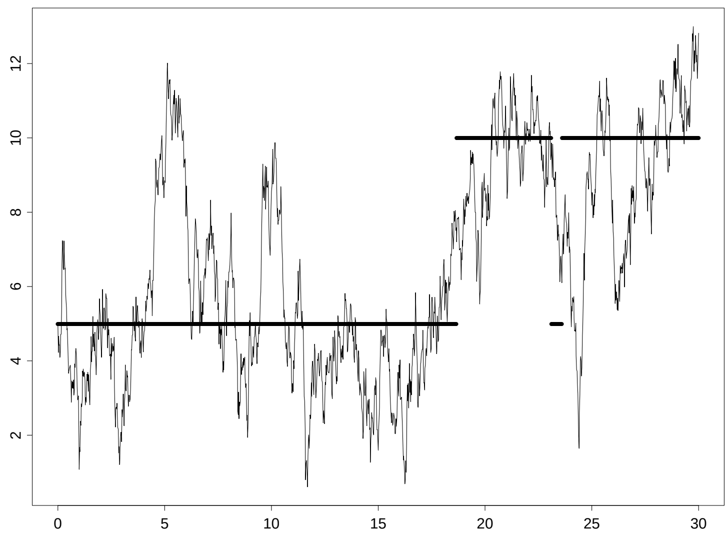

Figure 2: A sample path of the “jumping mean” process described by (11).

and Landén (1999). We aim to keep the model simple and intuitive while at the same time seeing how far we can carry the analysis. More formally, we look at a model where $\mathbb{Q}$-dynamics of the short rate are given by

$$dr_t = \kappa(\theta_t + \sigma\lambda(t)/\kappa - r_t)dt + \sigma dW_t^{\mathbb{Q}}, \tag{11}$$

with $W^{\mathbb{Q}}$ being a $\mathbb{Q}$-Brownian motion and

$$d\theta(t) = (\theta^H - \theta^L)\left(\mathbf{1}_{\{(\theta(t-)=\theta^L\}} - \mathbf{1}_{\{\theta(t-)=\theta^H\}}\right) dN^{\mathbb{Q}}(t), \tag{12}$$

where $\theta_0 \in \{\theta^L, \theta^H\}$ and $N^{\mathbb{Q}}$ is a $\mathbb{Q}$-Poisson process (independent of $W^{\mathbb{Q}}$) with intensity $\lambda^{\mathbb{Q}^N}$. Figure 2 shows a sample path of such a process. Supposing for a moment that $\lambda(t) \equiv 0$, then the short rate process is mean reverting towards a stochasticly shifting long term level $(\theta(t))$ with κ being the speed of mean reversion and σ being the volatility parameter. The local mean level $(\theta(t))$ can take on two values; a high level θ^H and a low level θ^L. Because the changes are controlled by arrivals from a Poisson-process, θ is “memoryless”; the (distribution of the) time we have to wait for the next change does not depend on how long we have been in the current state. The λ process is assumed to be deterministic and to represent the translation (of the Brownian motion) from $\mathbb{P}$ into $\mathbb{Q}$ (so putting $\lambda(t) \equiv 0$ gives - up to the intensity of the Poisson process - the $\mathbb{P}$-dynamics). By using the Itô formula on $(e^{\kappa t} r_t)$ we get a decomposition result for the short rate.

Theorem 2.12 *Let $r(0) = r_0$, $\theta(0) = \theta_0 \in \{\theta^L, \theta^H\}$, and A denote the set of odd numbers. The short rate can then be written as*

$$r(t) = a(t) + b(t) + c(t)$$

where

$$\begin{aligned} a(t) &= r_0 e^{-\kappa t} + \kappa \int_0^t e^{-\kappa(t-s)}\theta_0 ds + \sigma \int_0^t e^{-\kappa(t-s)} dW^Q(s), \\ b(t) &= \begin{cases} \kappa(\theta^H - \theta^L)\int_0^t e^{-\kappa(t-s)} \mathbf{1}_{\{N_s^Q \in A\}} ds & \text{if } \theta_0 = \theta^L \\ \kappa(\theta^H - \theta^L)\int_0^t e^{-\kappa(t-s)} \mathbf{1}_{\{N_s^Q \notin A\}} ds & \text{if } \theta_0 = \theta^H \end{cases} \end{aligned}$$

and

$$c(t) = \sigma \int_0^t e^{-\kappa(t-s)} \lambda(s) ds.$$

In particular $(a(t))$ is independent of $(b(t))$ and

$$a(t) \sim N\left(\theta^L + e^{-\kappa t}(r_0 - \theta^L), \frac{\sigma^2}{2\kappa}(1 - e^{-2\kappa t})\right). \tag{13}$$

This decomposition is important because of the general bond price formula (6). Before stating a decomposition result for zero coupon bond prices, let us now introduce some further notation. For arbitrary real numbers θ_0 and $\widehat{\theta}_0$ we define

$$h(\theta_0, T) = \mathbb{E}^{\mathbb{Q}}\left(\exp\left(-\kappa(\widehat{\theta}_0 - \theta_0)\int_0^T \int_0^t e^{-\kappa(t-s)} \mathbf{1}_{\{N_s^{\mathbb{Q}} \in A\}} ds dt\right)\right)$$

and

$$\widehat{h}(\theta_0, T) = \mathbb{E}^{\mathbb{Q}}\left(\exp\left(-\kappa(\theta_0 - \widehat{\theta}_0)\int_0^T \int_0^t e^{-\kappa(t-s)} \mathbf{1}_{\{N_s^{\mathbb{Q}} \notin A\}} ds dt\right)\right).$$

Since $\mathbf{1}_{\{N_s^{\mathbb{Q}} \in A\}} + \mathbf{1}_{\{N_s^{\mathbb{Q}} \notin A\}} = 1$, we get that

$$h(\theta_0, T) = \widehat{h}(\theta_0, T) e^{m(\widehat{\theta}_0; 0, T) - m(\theta_0; 0, T)},$$

where

$$n(t, T) = \frac{1 - \exp(-\kappa(T - t))}{\kappa}, \tag{14}$$

$$m(\theta; t, T) = \frac{(n(t,T) - (T - t))(\kappa^2\theta - \sigma^2/2)}{\kappa^2} - \frac{\sigma^2 n(t,T)^2}{4\kappa}. \tag{15}$$

We shall adapt the convention that the $\widehat{\ }$-notation is used to denote "complementary state" i.e.,

$$\begin{aligned} \widehat{\theta}_0 = \theta^L &\qquad \text{if } \theta_0 = \theta^H, \\ \widehat{\theta}_0 = \theta^H &\qquad \text{if } \theta_0 = \theta^L, \end{aligned}$$

and similarly $\widehat{\theta}(t) = \theta^L$ if $\theta(t) = \theta^H$, $\widehat{\theta}(t) = \theta^H$ if $\theta(t) = \theta^L$.

Theorem 2.13 *Let $r(0) = r_0$, $\theta(0) = \theta_0 \in \{\theta^L, \theta^H\}$, and A be the odd numbers. Then the price of the T-maturity ZCB equals*

$$P(t, T|r_t, \theta_t) = P_{\mathrm{V}}^{\theta(t)}(r(t), t, T) \ \exp\left(-\int_0^T c(s)ds\right) h(\theta(t), T - t), \tag{16}$$

and

$$P(t, T|r_t, \theta_t) = P_{\mathrm{V}}^{\widehat{\theta}(t)}(r(t), t, T) \ \exp\left(-\int_0^T c(s)ds\right) \widehat{h}(\theta(t), T - t), \tag{17}$$

where $P_{\mathrm{V}}^{\theta}(r_0, t, T)$ denotes the price of a ZCB in a Vasicek model with θ as the long term level of the short rate and is explicitly given by

$$P_{\mathrm{V}}^{\theta}(r_t, t, T) = \exp(m(\theta, t, T) - r_t n(t, T)),$$

where the functions m and n are given by (15) and (14), respectively.
Alternatively, we have

$$\begin{aligned} P(t, T|r_0, \theta_0) &= P_{\mathrm{V}}^{\theta^L}(r(t), t, T) \exp\left(-\int_0^T c(s)ds\right) \\ &\quad \times \left(h(\theta^L, T - t)\mathbf{1}_{\{\theta(t)=\theta^L\}} + \widehat{h}(\theta^H, T - t)\mathbf{1}_{\{\theta(t)=\theta^H\}}\right) \end{aligned} \tag{18}$$

and

$$\begin{aligned} P(t, T|r_0, \theta_0) &= P_{\mathrm{V}}^{\theta^H}(r(t), t, T) \exp\left(-\int_0^T c(s)ds\right) \\ &\quad \times \left(h(\theta^H, T - t)\mathbf{1}_{\{\theta(t)=\theta^H\}} + \widehat{h}(\theta^L, T - t)\mathbf{1}_{\{\theta(t)=\theta^L\}}\right). \end{aligned} \tag{19}$$

This result can be seen as a (slight) generalization of the Duffie-Kan result from Theorem 2.11 (for more general results about stochastic mean models see Balduzzi, Das, Foresi & Sundaram (1998)). It follows easily from Theorem 2.13 that ZCB prices in this model exhibit jumps (despite the short rate being a continuous process). Further, the theorem makes it easy to calibrate the model to market data using the ideas in Hull & White (1990). Specifically, let $P^{\mathrm{obs}}(0, T)$ denote the observed ZCB prices and put

$$f(T) = -\ln\left(\frac{P^{\mathrm{obs}}(0, T)}{P_{\mathrm{V}}^{\theta^L}(0, T)\mathbb{E}^{\mathbb{Q}}\left(\exp(-\int_0^T b(s)ds)\right)}\right).$$

If we assume κ, θ^L, θ^H, σ, and λ^{Q^N} have been estimated, for example from time series data, the model is calibrated by choosing λ such that

$$f(T) = \sigma \int_0^T \int_0^s e^{-\kappa(s-u)}\lambda(u)duds \quad \text{for all } T. \tag{20}$$

Differentiating (20) twice w.r.t. T gives the very neat result that

$$\lambda(T) = \frac{f'' - \kappa f'}{\sigma} \quad \text{for all } T.$$

To actually calculate ZCB prices we may WLOG assume that $\theta(0) = \theta^L$ and then have to calculate

$$g(\theta_t, t; T) = \mathbb{E}^{\mathbb{Q}}\left(\exp(-\int_t^T b(s,t)ds)|\mathcal{F}_t\right),$$

where

$$b(s,t) = \kappa(\theta^H - \theta^L)\int_t^s \exp(-\kappa(s-u))\mathbf{1}_{\{\theta_u = \theta^H\}}du.$$

It is possible to evaluate this in terms of certain special functions, but rather we give two numerical methods that are easy to comprehend, program and extend, yet also effective. One is simulation based and the other is an approximation (that converges at our will) based on backward induction. In particular the latter is very efficient; determination of the fully term structure up to maturities of 10 years with an accuracy of 1 basispoint (one hundredth of a percent) takes around 0.1 seconds.

Finally, the homogeneity of the payoff function for a call-option and the multiplicative ZCB decomposition enables us to derive (by a careful application of the "useful rule", see Hoffmann-Jørgensen (1994, 6.8.14)) a valuation formula for a call-option expiring at time T_E on a ZCB maturing at time T_M that involves only the θ-process and only the time-interval $[0; T_E]$.

Theorem 2.14 *Let $r(0) = r_0$, $\theta(0) = \theta_0$. Assume that $c \equiv 0$ and let $\mathrm{Call}_V^{\theta^L}(r_0, K)$ denote the call-option price in a θ^L-Vasicek model (an explicit formula can be found e.g. in Jamshidian (1989)). The price (at time 0) of a call-option with strike price K expiring at time T_E on a zero coupon bond maturing at time T_M satisfies*

$$\begin{aligned}\mathrm{Call}(r_0, \theta_0, K) \quad = \quad & \mathbb{E}^{\mathbb{Q}}\left(\exp\left(-\int_0^{T_E} b(s)ds\right) e^{-n(T_E,T_M)b(T_E)}\,\xi(\theta(T_E), T_M - T_E)\right. \\ & \left. \times\, \mathrm{Call}_V^{\theta^L}\left(r_0, \frac{K}{\xi(\theta(T_E); T_M - T_E)}\right)\right),\end{aligned}$$

where the random variable $\xi(\theta(T_E), T_M - T_E)$ *is given by*

$$\xi(\theta(T_E); T_M - T_E) \quad = \quad h(\theta^L, T_M - T_E)\mathbf{1}_{\{\theta(T_E)=\theta^L\}} + \widehat{h}(\theta^H, T_M - T_E)\mathbf{1}_{\{\theta(T_E)=\theta^H\}}$$

with

$$\theta(T_E) = \theta_0 + (\widehat{\theta}_0 - \theta_0)\mathbf{1}_{\{N^{\mathbb{Q}}_{T_E} \in A\}} = \widehat{\theta}_0 + (\theta_0 - \widehat{\theta}_0)\mathbf{1}_{\{N^{\mathbb{Q}}_{T_E} \notin A\}}.$$

Gaussian models, such as the Vasicek model, are analytically tractable (for an almost exhaustive account of what you can do in Gaussian term structure models, see Jamshidian (1991)), but have the drawback that negative interest rates are possible. If you have "something" that you want to make non-negative, then two ways immediately come to mind, *i*) squaring it or, *ii*) exponentiating it. It also works like that in interest rate modelling. (For a quite different way of achieving positive interest rates, see Flesaker & Hughston (1996).)

Let us first look at the "squaring" option. Modelling the interest rates as "something Gaussian squared" was proposed in Beaglehole & Tenney (1991), the idea is also used in Constantinides (1992), and has a tendency to be "rediscovered" from time to time. If $Y = \alpha + \beta^\top X$ for constants $\alpha \in \mathbb{R}$, $\beta \in \mathbb{R}^n$ and an n-dimensional random variable X, where X_i is non-central χ^2-distributed with ν_i degrees of freedom and non-centrality parameter λ_i (cf. Johnson, Kotz & Balakrishnan (1994, Chapter 29)), then we write $Y \in \chi^2_n(\alpha, \beta, \nu, \lambda)$. From Jamshidian (1996) we have the following interesting result.

Theorem 2.15 *i*) *Suppose that* $r(t) = \sum_{i=1}^n X_i(t)$*, where the* X_i*'s follow independent CIR-processes under* $\mathbb{Q}$*. Then* $\ln P(t,T) \in \chi^2_n$*. Further, ZCB option prices can be expressed in terms of a single (i.e., one-dimensional) integral.*

ii) *Suppose that* $r(t) = \sum_{i=1}^n X_i(t)$*, where the* X_i*'s follow independent Vasicek-processes under* $\mathbb{Q}$*. Then* $\ln P(t,T) \in \chi^2_n$*. Further, ZCB option prices can be expressed in terms of a single integral.*

iii) *Suppose that*

$$r(t) = Q^r(X(t)),$$

where $Q^r : \mathbb{R}^n \mapsto \mathbb{R}$ *is a quadratic function (with time homogeneous coefficients and symmetric second order coefficient matrix) and*

$$dX(t) = (\alpha + \beta X)dt + \Sigma^\top dW^{\mathbb{Q}}(t)$$

for constants α, β, and Σ. Then

$$\ln P(t,T) = Q^P(X(t)),$$

where $Q^P : \mathbb{R}^n \mapsto \mathbb{R}$ is a quadratic function with time dependent coefficients. These coefficients can by characterized as the solution to a nonlinear system of ODEs with no known closed-form solution.

The main point in the proofs (besides the calculation of truncated means of exponentiated χ_n^2-variables) of parts i) and ii) in Theorem 2.15 is that the χ_n^2 property is preserved under changes to forward measures (that will be formally described in Definition 2.5 and Theorem 2.21) which reduce problems to being one-dimensional. Part iii) is an "extended Duffie-Kan" result, but other than that primarily negative.

Next, we consider the "exponentiating" option. The first problem is that in lognormal short rate models no closed form solution for ZCB prices is known, basically because the calculation involves (the exponential of) the "stylized sum" $\int r(s)ds$ of lognormal variables, which is not lognormal and generally has an unknown distribution (however, for the case of exponentiated Brownian motion with drift, one should consult Geman & Yor (1993)). This has given rise to many suggestions regarding efficient numerical procedures for the lognormal short rate models (see for instance Black, Derman & Toy (1990), Hull & White (1994), Buttimer, Muller & Reeves (1995), Bjerksund & Stensland (1996)). But problems do not end there, in Hogan & Weintraub (1993) we find the following discomforting fact.

Theorem 2.16 *Suppose that either*

$$dr(t) = \alpha r dt + \sigma r dW^{\mathbb{Q}}(t),$$

or

$$d\ln r(t) = (\alpha + \beta \ln r)dt + \sigma dW^{\mathbb{Q}}(t).$$

Then for any $s < t < T$ we have

$$\mathbb{E}_s^{\mathbb{Q}}(P^{-1}(t,T)) = \infty, \qquad \mathbb{E}_s^{\mathbb{Q}}(\beta(t,T)) = \infty.$$

The result in Theorem 2.16 is surprisingly difficult to prove (Hoffman (1993) offers a different proof). You might think it could be done using the fact that lognormal variables do not have exponential moments along with elementary "Jensen Inequality"-estimates. This appears not to be the case.

Looking at Theorems 2.15 and 2.16 you would think that the former, which contains mainly positive results, had induced widespread investigation and application of χ_n^2-models, while the latter, a negative result, meant the end of the road for models with lognormal interest rates. But as we shall see towards the end of the next subsection, old habits die hard and there are plenty of people "qualified to give them life support".

2.3.2 The Direct Approach

The first steps in the "Direct Approach", i.e., in trying to model directly all ZCB prices simultaneously, were taken in a discrete model in Ho & Lee (1986). Direct modelling in continuous-time models was introduced by Heath, Jarrow & Morton (1992). When comparing those two articles we see the usual "discrete vs. continuous" effect; The ideas may be the same, the mathematics in the discrete models is not too deep, but the formulas get quite messy, whereas the continuous models use quite deep mathematical results, but the formulas can look much simpler (cf. (24)).

From Heath et al. (1992) we have the following.

Theorem 2.17 *Suppose we have a model where ZCBs of all maturities less that $\mathcal{T}$ are traded and that the forward rates f follow*

$$df(t,T) = \mu_f(t,T)dt + \sigma_f^\top(t,T)dW(t). \tag{21}$$

Then a necessary, and essentially sufficient, condition for the model to be arbitrage-free is the existence of a process ν such that

$$\mu_f(t,T) = -\sigma_f^\top(t,T)\left(\nu(t) - \int_t^T \sigma_f(t,u)du\right). \tag{22}$$

*The "*essentially*" part relates to integrability conditions on ν.*

The drift condition in Theorem 2.17 involves an unspecified risk-premium process and is not particularly intuitive at first sight. This, however, all changes when we formulate the result in terms of ZCB price related objects.

Theorem 2.18 *Consider a set-up as in Theorem 2.17, put $\eta(t) = \exp\left(\int_0^t \nu(s)dW(s) - \frac{1}{2}\int_0^t \nu^\top(s)\nu(s)ds\right)$, and define $\mathbb{Q} \sim \mathbb{P}$ by $d\mathbb{Q}/d\mathbb{P}|_{\mathcal{F}_t} = \eta(t)$. Then*

$$dP(t,T) = r(t)P(t,T)dt - \sigma_P(t,T)P(t,T)dW^{\mathbb{Q}}(t), \tag{23}$$

where $W^{\mathbb{Q}}$ *is a* $\mathbb{Q}$*-Brownian motion and* $-\sigma_P(t,T) = \int_t^T \sigma_f(t,u)du$*. In short:* $\mathbb{Q}$ *is an equivalent martingale measure. Further, with* $\mathbb{Q}$*-superscripts indicating* $\mathbb{Q}$*-drifts we have*

$$\mu_f^{\mathbb{Q}}(t,T) = \sigma_f^{\top}(t,T)\int_t^T \sigma_f(t,u)du. \tag{24}$$

The "4-line" way to see is to rely on Theorem 2.5 and say that under $\mathbb{Q}$ all prices (in particular those of ZCBs) have the short rate as expected rate of return, hence (23). Then use the Itô formula on $\ln P(t,T)$ and differentiate (formally) w.r.t. T to obtain (24). A HJM-economy consists of an uncountably infinite number of assets, but at the same time we maintain our definitions from Section 2.1. For the results regarding state price deflators it does not matter much, but note that trading strategies can by definition only involve a finite number of assets. This has the consequence that we cannot strictly construct the savings account from ZCBs since this would involve continuously rolling investments over in the instantly maturing ZCB. This can be remedied in two ways; we may explicitly assume there exists an asset with dynamics "$d\beta(t) = r(t)\beta dt$" or we may allow for so-called measure-valued portfolios (cf. Björk, di Masi, Kabanov & Runggaldier (1997)). We will not go into that, but the reader should rest assured that using the savings account or talking about $\mathbb{Q}$ can be rigorously justified.

For the forward rates $f(t,T)$, T is the time *of* maturity (a specific date, e.g. "September 24th"). Often it is more relevant to use a fixed time *to* maturity (e.g. "3 months"). In this case the drift restriction looks slightly different (when trying to determine the dynamics of $f(t,t+\tau)$ we have to remember t enters both arguments), explicitly we have the following result from Musiela (1993).

Theorem 2.19 *Consider the set-up as in Theorem 2.18. For any* $x \geq 0$ *define the forward rate* $\widetilde{f}$ *by* $\widetilde{f}(t,x) = f(t,t+x)$*. Put* $\widetilde{\sigma}(t,x) = \sigma_f(t,t+x)$ *and* $D(t,x) = \widetilde{\sigma}^{\top}(t,x)\int_0^t \widetilde{\sigma}(t,u)du$*. Then we have*

$$d\widetilde{f}(t,x) = \left(\frac{\partial}{\partial x}\widetilde{f}(t,x) + D(t,x)\right)dt + \widetilde{\sigma}(t,x)dW^{\mathbb{Q}}(t).$$

One thing is that this result is a useful reminder in empirical contexts when data are often given in time to maturity format. But more importantly the whole $\widetilde{\ }$-notation underlines the fact that (21) is really an infinite dimensional SDE or a stochastic PDE. For each T there is an equation, and it is an inherent part of the model that they are "connected" (by (22) or (24)). This leads to the realization that it might not

- and typically will not - be possible to give a Markovian representation of the whole term structure in terms of a finite number of observable quantities, even though the "number of sources of risk" $= \dim(W)$ is finite.

Fortunately, it turns out to be easy to characterize the models where such a representation is possible. This issue has been addressed by several authors and the conclusion is as follows.

Theorem 2.20 *Suppose the forward rate volatility can be written as*

$$\sigma_f(t,T) = \Big(\beta_1(t,r(t))\exp(-\int_t^T \kappa_1(u)du), \ldots, \beta_d(t,r(t))\exp(-\int_t^T \kappa_d(u)du)\Big)^\top,$$

for functions $\beta_1, \ldots, \beta_d$ and $\kappa_1, \ldots, \kappa_d$. Then the term structure has Markovian representation by means of $d(d+3)/2$ state variables. These variables can be chosen such that for the first d of them, say $X_1, \ldots, X_d$, we have

$$r(t) = f(0,t) + \sum_{i=1}^{d} X_i(t),$$

while the last $d(d+3)/2 - d = d(d+1)/2$, say $(V_{ij})_{j\le i}$, are of bounded variation. Further, if the β_i's do not depend on r then the V_{ij}'s are deterministic.
Conversely, if r is Markovian w.r.t. its own filtration then

$$\sigma_f(t,T) = \Big(\beta_1(t)\exp(-\int_t^T \kappa_1(u)du), \ldots, \beta_d(t)\exp(-\int_t^T \kappa_d(u)du)\Big)^\top$$

for functions $\beta_1, \ldots, \beta_d$ and $\kappa_1, \ldots, \kappa_d$.

Theorem 2.20 basically says that forward rate volatility must by multiplicative separable in T. The first part of the theorem (as well as explicit characterization of the state variables and their dynamics) follows from Ritchken & Sankarasubramanian (1995) (for $d = 1$) or Cheyette (1995) (for $d \ge 1$), while the best proof of the second part can be found in Musiela & Rutkowski (1997, Proposition 13.2.3). In general the V_{ij}'s in the representation are "integrated volatility".

One could contemplate using a "proportional volatility" model,

$$\sigma_f(t,T) = \sigma f(t,T),$$

but only for a brief moment, though. That model does not have a finite-dimensional Markovian representation so hopes of closed form solutions (for, say, options) seems futile since we would have to keep track of the whole term structure in our calculations. And even worse: Morton (1988) shows (in a quite cunning way) that with proportional volatility forward rates become infinite in finite time with positive probability. Note that this is a different and more severe problem than the "Hogan-Weintraub" problem outlined in Theorem 2.16.

A major breakthrough in models of this type came by going "back to basics". Among first things you learn in a finance course is "always quote interest rates on a continuously compounded basis to avoid confusion". (Actually, it is an elementary school exercise to inspire hostility against banks by letting pupils show that 3% compounded quarterly is not equivalent to 12% compounded yearly.) Among first things you learn in a finance course with dynamic modelling is "always to work with instantaneous interest rates". This, however, is not how practioners think of or quote interest rates. They typically use an "add-on" convention as given in the next definition.

Definition 2.4 *Let $\delta > 0$ be given. The simple forward rate, or the forward δ-LIBOR, $K_\delta(t,T)$, is defined by the relation*

$$P(t,T) = (1 + \delta K_\delta(t,T))P(t,T+\delta). \tag{25}$$

Originally, LIBOR was an acronym for London Interbank Offer Rate, an average of several interbank rates, i.e., rates that leading banks offered other leading banks on deposits (so there should be no default risk involved). But we shall use LIBOR as a generic term for any rate that has the "add-on" form

$$\text{return} = \text{principal} + \text{investment period} \times \text{interest rate} \times \text{principal},$$

as described by (25).

An invaluable tool in interest rate models is the *forward measure*. It is a special case of "numeraire invariance" which was described abstractly in Theorem 2.4 and revealed its computational powers in one of the Black-Scholes proofs (cf. Section 2.2). Now it becomes an invaluable tool in the model construction. There is some debate whether the construction and usefulness was first established by Hélyette Geman or by Farshid Jamshidian. We have no remarks on that, just the following definition.

Definition 2.5 *A probability measure $\mathbb{Q}^T \sim \mathbb{Q}$ on $(\Omega, \mathcal{F}_T)$ with Radon-Nikodym derivative given by*

$$\frac{d\mathbb{Q}^T}{d\mathbb{Q}} = \frac{\beta^{-1}(T)}{\mathbb{E}^{\mathbb{Q}}(\beta^{-1}(T))} = \frac{1}{\beta(T)P(0,T)}$$

is called the forward measure *for the settlement date T.*

When restricted to $\mathcal{F}_t$, the Radon-Nikodym derivative satisfies

$$\eta(t) := \frac{d\mathbb{Q}^T}{d\mathbb{Q}}|_{\mathcal{F}_t} = \mathbb{E}_t^{\mathbb{Q}}\left(\frac{1}{\beta(T)P(0,T)}\right) = \frac{P(t,T)}{\beta(t)P(0,T)}.$$

Suppose that ZCB prices fulfill (23), then we may write

$$\eta(t) = \exp\left(-\int_0^t \sigma_P(u,T)dW(u) - \frac{1}{2}\int_0^t \sigma_P^\top(u,T)\sigma_P(u,T)du\right),$$

and get by Girsanov's Theorem that

$$W^{\mathbb{Q}^T}(t) = W^{\mathbb{Q}}(t) + \int_0^t \sigma_P(u,T)du$$

defines a $\mathbb{Q}^T$-Brownian motion. Theorem 2.4 and/or an application of the Itô formula gives us the next result.

Theorem 2.21 *Suppose we have an arbitrage-free economy. Then the price of any asset, say C, discounted by the T-ZCB is a $\mathbb{Q}^T$-martingale, i.e.,*

$$\frac{C(s)}{P(s,T)} = \mathbb{E}_s^{\mathbb{Q}^T}\left(\frac{C(t)}{P(t,T)}\right) \quad \textit{for all } s \le t \le T.$$

Further, $K_\delta(t,T)$ is a $\mathbb{Q}^{T+\delta}$-martingale.

Of course, you could use the first (or the second, for that matter) part of Theorem 2.21 to define the forward measure and Definition 2.5 would then be a theorem. It is also useful to note that Brownian motions under different forward measures are related through

$$dW^{\mathbb{Q}^{T+T'}}(t) = dW^{\mathbb{Q}^T}(t) + \int_T^{T+T'} \sigma_f(t,u)dudt \tag{26}$$

With the first part of Theorem 2.21, it is now easy to see why the use of the forward measure may offer a simplification in interest rate modelling. Suppose we want to find the price (at time t) of a call-option (expiring at T_E) on a ZCB (maturing at T_M). We then have to calculate

$$\frac{Call(t)}{\beta(t)} = \mathbb{E}_t^{\mathbb{Q}}\left(\frac{(P(T_E,T_M)-K)^+}{\beta(T_E)}\right).$$

In a model with stochastic interest rates $\beta(T_E)$ and $P(T_E,T_M)$ can evidently not be independent, so calculation of the above expression would involve a 2-dimensional integral. But changing to the T_E forward measure and using that $P(T_E, T_E) = 1$ we get

$$\frac{Call(t)}{P(t,T_E)} = \mathbb{E}_t^{\mathbb{Q}}\left(\frac{(P(T_E,T_M)-K)^+}{P(T_E,T_E)}\right) = \mathbb{E}_t^{\mathbb{Q}^{T_E}}((P(T_E,T_M)-K)^+),$$

which involves only a 1-dimensional integral. Of course, we now have to determine the $\mathbb{Q}^{T_E}$-distribution of $P(T_E, T_M)$, but this is often easy (using Girsanov's Theorem which boils down to (26)).

We shall now illustrate how the second part of Theorem 2.21 is helpful in model building. To this end let us for given T_0 and δ define the δ-tenor structure as the set given by

$$\{T_0 + \delta n \mid n \in \mathbb{Z}\},$$

and in the rest of the paper $T_j = T_0 + j\delta$, for $j \geq 0$, unless clearly otherwise stated. We always think of the tenor-structure as capturing the dates where something "interesting" happens.

Recall that to completely specify HJM-type model we need (along with an initial term structure) to specify the bond price volatility $-\int_t^T \sigma_f(t,u)du$ for all t and T (or at least for $T \geq t$.) Two specifications have primarily been considered (see Andersen & Andreasen (1998) for more general volatility choices). Both have $\sigma_f(t,T) = 0$ for $T \leq t+\delta$ (this is a perhaps not totally realistic but quite convenient initial condition).

- The Gaussian δ-LIBOR model in which

  $$\int_T^{T+\delta} \sigma_f(t,u)du = \gamma_G(t,T) \quad \text{for } T \geq t+\delta, \tag{27}$$

  where $\gamma_G : \mathbb{R}^2 \mapsto \mathbb{R}^d$ is deterministic. This means that

  $$dK_\delta(t,T_{j-1}) = \gamma_G^\top(t,T_{j-1})dW^{\mathbb{Q}^{T_j}}(t), \tag{28}$$

  so $K_\delta(t,T_{j-1})$ is normally distributed under $\mathbb{Q}^{T_j}$. And we note that from (26) we have

  $$dW^{\mathbb{Q}^{T+\delta}}(t) = dW^{\mathbb{Q}^T}(t) + \gamma_G(t,T)dt,$$

  so normality is preserved under forward measure changes.

- The lognormal δ-LIBOR model in which

$$\int_T^{T+\delta} \sigma_f(t,u)du = \frac{\delta K_\delta(t,T)}{1+\delta K_\delta(t,T)}\gamma(t,T), \quad \text{for } T \geq t+\delta, \tag{29}$$

where $\gamma : \mathbb{R}^2 \mapsto \mathbb{R}^d$ is deterministic. This means that

$$dK_\delta(t,T_{j-1}) = K_\delta(t,T_{j-1})\gamma^\top(t,T_{j-1})dW^{\mathbb{Q}_j^T}(t), \tag{30}$$

implying that $K_\delta(t,T_{j-1})$ is lognormally distributed under $\mathbb{Q}^{T_j}$. So interest rates are positive (under any of the equivalent forward measures), but from (26) we have

$$dW^{\mathbb{Q}^{T+\delta}}(t) = dW^{\mathbb{Q}^T}(t) + \frac{\delta K_\delta(t,T)}{1+\delta K_\delta(t,T)}\gamma(t,T)dt, \tag{31}$$

which shows that the lognormality is *not* preserved under forward measure changes.

The above volatility definitions ((27) and (29)) are for arbitrary Ts; in other words we have models where formally ZCBs of all maturities exists. Really, this is “overkill”. We could (typically) make do with only ZCBs of maturities equal to a tenor-structure and still obtain the same pricing an hedging results. We shall not discuss these finer points (see Musiela & Rutkowski (1997, Chapter 16)), rather we are content when final pricing formulas indicate only dependence on a few ZCB prices.

As short rate specifications these models are quite hopeless. For example Brace, Gatarek & Musiela (1997) note that if r is a semimartingale then it is of bounded variation and can (thus) only be Markovian if it is deterministic. However, as we shall soon see, the models have other virtues.

In the following we shall focus on the lognormal specification. From (25) it is clear that $\mathbb{E}_t^{\mathbb{Q}^{T+\delta}}(P^{-1}(T,T+\delta)) < \infty$. In fact from Remark 2.3 in Brace et al. (1997) we get that $\mathbb{Q}$-expected roll-over gains are finite (the fact that lognormality of interest rates with strictly positive compounding periods removes “Hogan-Weintraub” problems was first shown in Sandmann & Sondermann (1997).)

Theorem 2.22 *In the lognormal δ-LIBOR model we have that for any $s < t$*

$$\mathbb{E}_s^{\mathbb{Q}}(P^{-1}(t,T)) < \infty, \qquad \mathbb{E}_s^{\mathbb{Q}}(\beta(t,T)) < \infty.$$

Theorem 2.22 is a nice result, but the lognormal δ-LIBOR model is of little interest unless we are able to find closed form expressions for prices of “plain vanilla” derivatives. As we shall soon see, such prices can be found. To this end we first consider a δ-caplet settled in arrears with strike level κ, which is a contract with pay-off

$$\delta(K_\delta(T_{j-1}, T_{j-1}) - \kappa)^+ \text{ at } T_j.$$

Note that the payment is made at T_j but known one δ-period in advance. The caplet price satisfies

$$\frac{C(t)}{P(t,T_j)} = \mathbb{E}_t^{\mathbb{Q}^{T_j}} \left(\delta \frac{(K_\delta(T_{j-1}, T_{j-1}) - \kappa)^+}{P(T_j, T_j)} \right),$$

or

$$C(t) = P(t,T_j)\mathbb{E}_t^{\mathbb{Q}^{T_j}} \left(\delta(K_\delta(T_{j-1}, T_{j-1}), -\kappa)^+ \right) \tag{32}$$

where $K_\delta(\cdot, T_{j-1})$ is a lognormal martingale under $\mathbb{Q}^{T_j}$. Hence we get the following result by the same calculation as used in the computation of the Black-Scholes formula.

Theorem 2.23 *In the lognormal δ-LIBOR model the price of a δ-caplet settled in arrears with strike level κ is given by*

$$Caplet(t) = \delta P(t,T_j) \left(K_\delta(t, T_{j-1})\Phi(\tilde{d}_1) - \kappa\Phi(\tilde{d}_2) \right)$$

where

$$\tilde{d}_{1,2} = \frac{\ln(K_\delta(t, T_{j-1})/\kappa) \pm \frac{1}{2}\tilde{v}^2}{\tilde{v}}$$

and

$$\tilde{v}^2 = \int_t^{T_{j-1}} \lambda^\top(u, T_{j-1})\lambda(u, T_{j-1})du.$$

One could ask why the apparently very simple result in Theorem 2.23 is a major step forward in interest rates modelling. The first answer is that it is interesting *because* it is simple, but besides that:

- Equation (32) and the line of text immediately following it reflects what has been "market practice" for many years when pricing caplets, see the discussion in Musiela & Rutkowski (1997, Section 17.3) or Miltersen, Sandmann & Sondermann (1997). This is why models of this type are often called "market models";

they mimic what the market (or rather: the market participants) does. It was believed that this was "only an approximation" (just looking at (32) it would appear that you are taking the discount factor and the underlying to be independent), or at least it was unclear whether the formula could be supported by a formal term structure model.

- It is not just caplets we can price. The price of a (forward) cap contract is just the sum of caplet prices and a floor contract is priced by parity arguments. Further, since easy calculations show that

$$\mathbb{E}_t^{\mathbb{Q}}\left(\delta\frac{(K_\delta(T_{j-1},T_{j-1})-\kappa)^+}{\beta_{T_j}}\right) = \mathbb{E}_t^{\mathbb{Q}}\left(\frac{\left(\frac{1}{1+\kappa\delta}-P(T_{j-1},T_j)\right)^+}{\beta_{T_{j-1}}}\right),$$

  we see that caplet pricing is similar to pricing put options on ZCBs.

- An academic "whodunit" debate always adds extra spice. It is probably safe to say that the idea was first developed by Miltersen et al. (1997) (its origin dating back to earlier papers by Sandmann), while Brace et al. (1997) were more rigorous in the model construction.

- Despite no finite dimensional representation of the whole term structure being available in the lognormal δ-LIBOR model (the volatility is evidently not multiplicatively T-separable) it is often still possible to express the derivative prices of interest by a the prices of a few bonds (alternatively by a few LIBORs), just as we see in Theorem 2.23. This also means that the derivatives can be hedged easily; in fact a hedge-PDE type argument using ZCBs was exactly how Theorem 2.23 was derived in Miltersen et al. (1997). In a subtle way this is also reflected in the formulation of Theorem 2.3, but rather than elaborating in that we refer the reader to Jamshidian (1997).

Another frequently considered contract in fixed income markets is the *swap* and its "natural" derivative the *swaption*.
A swap contract is a private agreement between two companies to exchange cash flows in the future after some prearranged formula. A payer swap settled in arrears exchanges LIBOR $K_{q\delta}(T_{j-q},T_{j-q})$ against a fixed rate κ at n intervals of length $q\delta$, i.e., payments are made at dates $T_0+q\delta, T_0+2q\delta,\ldots,T_0+nq\delta$ but known one $q\delta$-period in advance. The cash flow structure of payer swap is illustrated in Figure 3.

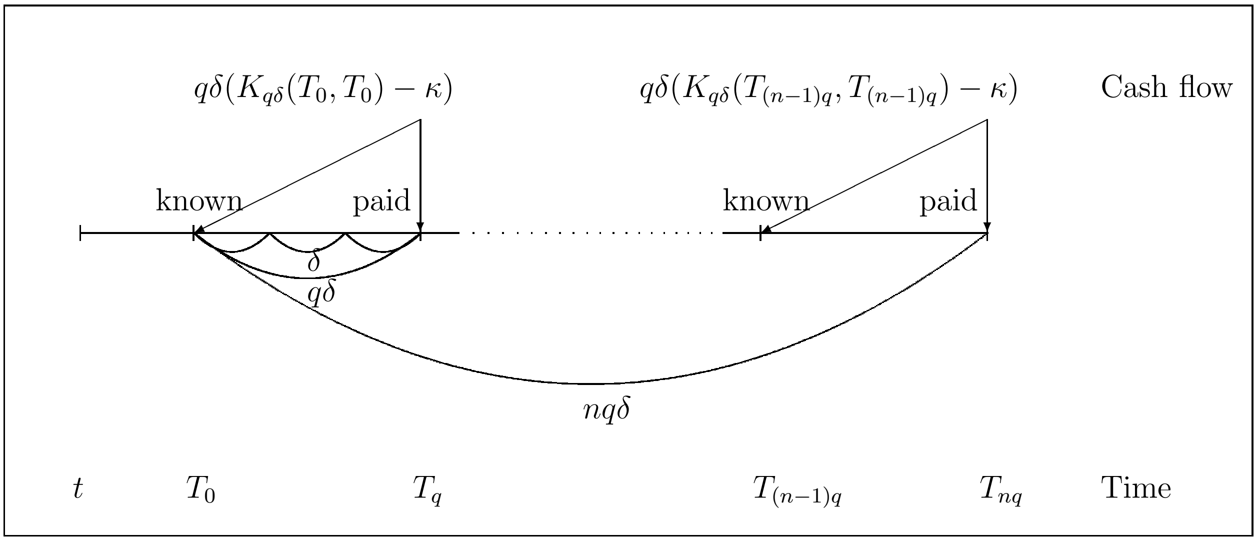

Figure 3: The "time & money"-structure of an interest rate payer swap settled in arrears.

For $t \leq T_0$ general theory gives us that the value of the swap is

$$Swap^{(q)}(t) = \mathbb{E}_t^{\mathbb{Q}}\left(\sum_{j=1}^{nq} qoz(j,q)\delta\frac{\beta(t)}{\beta(T_j)}[K_{q\delta}(T_{j-q},T_{j-q})-\kappa]\right),$$

where the "q or zero" function is defined by $qoz(j,q) = q\mathbf{1}_{\{(j \bmod q)=0\}}(j)$. By simple manipulations we can recast the swap price as

$$Swap^{(q)}(t) = \sum_{j=1}^{nq} \delta P(t,T_j)[K_\delta(t,T_{j-1}) - qoz(j,q)\kappa]. \tag{33}$$

The swap rate $\omega^{(q)}$ (corresponding to the specific swap described) is the value of κ that makes the value of the swap 0, i.e.,

$$\omega^{(q)}(t) = \frac{P(t,T_0)-P(t,T_{nq})}{\sum_{j=1}^{nq}\delta qoz(j,q)P(t,T_j)}. \tag{34}$$

A swaption is an option on the swap rate. Specifically, the basis swaption exchanges the time T_0 swap rate against a strike, κ, when $\omega^{(q)}(T_0) \geq \kappa$, thus it has a price determined by

$$Swaption^{(q)}(t) = \mathbb{E}_t^{\mathbb{Q}}\left(\frac{\beta_t}{\beta_{T_0}} Swap(T_0)^+\right).$$

Again, simple manipulations give the following equivalent forms for the swaption price

$$
\begin{aligned}
Swaption^{(q)}(t) &= \sum_{j=1}^{nq} \delta P(t,T_j) \mathbb{E}_t^{\mathbb{Q}_j^T} \left([K_\delta(T_0,T_{j-1}) - qoz(j,q)\kappa] \mathbf{1}_A \right) &\quad (35)\\
&= \mathbb{E}_t^{\mathbb{Q}} \left(\frac{\beta_t}{\beta_{T_0}} \left(1 - \sum_{j=1}^{nq} C_j P(T_0,T_j) \right)^+ \right), &\quad (36)
\end{aligned}
$$

where $A = \{\omega^{(q)}(T_0) \geq \kappa\} = \{Swap^{(q)}(T_0) \geq 0\}$, $C_j = qoz(j,q)\delta\kappa$ for $j = 1, \dots, nq - 1$
and $C_{nq} = 1 + qoz(j,q)\delta\kappa$. This shows two important things

i) A swaption is similar to a put-option on a coupon bearing bond. This gives us a strong idea as to how to price it: Use the techniques in Jamshidian (1989). These techniques are part of the "vocabulary of the financially articulate".

ii) The determination of the swaption price requires assessment of the distribution of a variety of LIBORs under a variety of measures. Unfortunately, given a δ-tenor structure and a lognormal δ-LIBOR model we have

$$K_{i\delta}(t,T_n) \overset{\mathbb{Q}^{T_n+j\delta}}{\sim} logN \quad \Leftrightarrow \quad i = j = 1.$$

So it does not seem possible to obtain a closed-form exact solution for the general swaption in the lognormal δ-LIBOR model.

In Brace et al. (1997), *ii)* is overcome by making "various lognormal approximations" and the ideas from *i)* are then used to derive the following approximate formula for the swaption price.

Theorem 2.24 *Consider a δ-tenor structure, the lognormal δ-LIBOR model, and the swaption described above. Define*

$$
\begin{aligned}
\mu(T_{l-1}) &= \frac{\delta K_\delta(0,T_{l-1})}{1+\delta K_\delta(0,T_{l-1})}, &\quad (37)\\
\Delta = (\triangle_{ij}) &:= \left(\int_0^{T_0} \gamma^\top(u,T_{i-1})\gamma(u,T_{j-1})du \right) \quad (\in \mathbb{R}^{nq\times nq}), &\quad (38)\\
D = (D_j) &:= \left(\sum_{l=1}^{j} \mu(T_{l-1})\triangle_{jl} \right) \quad (\in \mathbb{R}^{nq}). &\quad (39)
\end{aligned}
$$

Consider the rank-1 approximation

$$\Delta \approx \Gamma\Gamma^\top,$$

where Γ is the product of the square root of the largest of the eigenvalues of Δ (all of which are positive) and its corresponding eigenvector. Define the function J by

$$J(s) := \sum_{j=1}^{nq} \frac{\delta K_\delta(0, T_{j-1})(\exp(\Gamma_j(s+d_j) - \frac{1}{2}\Gamma_j^2) - qoz(j,q)\kappa)}{\prod_{i=1}^{j}(1 + \delta K_\delta(0, T_{j-1})\exp(\Gamma_i(s+d_i) - \frac{1}{2}\Gamma_i^2))}.$$

Then J has a unique root s^* and an approximation to the swaption price is given by

$$Swaption^{(q)}(0) = \sum_{j=1}^{nq} \delta P(0, T_j)[K_\delta(0, T_{j-1})\Phi(h_j) - qoz(j,q)\kappa\Phi(h_j - \Gamma_j)], \qquad (40)$$

where $h_j = \Gamma_j - s^* - d_j$ and Φ is the standard normal distribution function.

Excursion; Simulation and Approximation Quality in Lognormal LIBOR Models

The swaption price approximation in Theorem 2.24 leaves some open questions that are investigated in Poulsen (1997).

The formulation of Theorem 2.24 indicates that the two approximations are used.

i) The rank d (=the dimension of the driving Brownian motion) covariance matrix $\triangle$ is approximated by a rank 1 matrix. This means that all results can be expressed directly in terms of Φ; otherwise we would get expressions involving multidimensional normal integrals.

ii) The stochastic drift correction from (31) is substituted by a deterministic correction based on initial data (cf. (37)). In Brace et al. (1997) it is referred to as a "Wiener Chaos expansion of order 0", but still one should note that two the terms we interchange do not even have to same mean.

A further, somewhat more subtle question is

iii) How do prices depend on δ? Different values of δ (and corresponding tenor structure) give models with different dynamic properties and different derivative prices. To illustrate: Recall the definition of the swap contact. Suppose we were interested in pricing a swaption on a swap with $\delta = 1$, $q = 1$ and $n = 3$. This is readily done by Theorem 2.24. But exactly the same swaption is described by a model with $\delta = 1/2$, $q = 2$ and $n = 3$, or indeed any choice $\delta = 1/i$, $q = i$ and $n = 3$. Theorem 2.24 still works, but the prices are not the same. Note also that

$\lim_{\delta \to 0} K_\delta(t,T) = f(t,T)$, so by letting δ tend to 0 we get (at least formally) the "proportional volatility" model, which we saw earlier was an inherently bad model. The question is then when - or indeed if - things go awry.

Another logic question is:

iv) How *do* we determine exact prices numerically? More specificly, how do we simulate in lognormal LIBOR models?

In Poulsen (1997) issues *ii*)-*iv*) are addressed (issue *i*) is also relevant, but does not particularly hinge on the lognormal LIBOR model being used). The paper concludes the following.

- Simulation in lognormal LIBOR model is not hard, but it does require rewriting of pricing expressions, knowledge of the measure transforms used in defining the model, and some careful bookkeeping.

In the simple case of flat initial term structures and volatility we find

- The deterministic drift approximation appears not to be a problem. Even for swaptions with very long maturities (10 years) the approximation gives prices that are with 1 % of the true prices (and the error is increasing in the length of the swaption).

- The δ-discrepancy can be detected even for medium length swaptions (both for approximate and true prices). Pricing an at-the-money " $\delta = 1$, $q = 1$, $n = 3$" swaption in its "natural" " $\delta = 1$, $q = 1$, $n = 3$" model as in the $\lim_{i \to \infty}$ of "$\delta = 1/i$, $q = 1$, $n = 3$'-models gives a relative pricing difference of 5 %. This, however is the most extreme case w.r.t. sensitivity. First, the payments on the underlying are quite far apart (1 year), so there is "a lot of room for δ-effects". Second the swaption is at-the-money (by which we mean that the strike (κ) in equal to the current swap rate). And third, a decreasing volatility structure (by which we mean that $T \mapsto \gamma(t,T)$ is decreasing, there are sound empirical and theoretical arguments for this) also tends to lower the δ-effect.

Reflections on Fixed Income Modelling

The modelling methodologies described in the fixed income section show a clear tendency towards higher and higher inclusion of market information. We started with the indirect approach where a few interest rates were described and then claimed to be representative of the market. We then went to the direct HJM-approach where the whole yield curve was modelled and the current term structure – which by the way is not trivial to estimate – thus directly included. In practice, however, we want models with simple finite dimensional representations, so we often end up with something that looks like a yield factor model (see Brace & Musiela (1994) and Björk & Gombani (1999)). Finally, we came to market models which could be understood as a spin-off of the HJM-approach were volatility specifications were created directly to achieve simple formulae for liquid fixed income derivatives such as caps and swaptions which effectively means that these prices can be fitted be choosing appropriate volatility parameters (a.k.a. implied modelling). This pretty much takes us to "the state of the art" in fixed income modelling which we shall now briefly reflect upon. From both a theoretical and a practical point of view the methodology is sound enough. Arbitrage pricing is always a question of pricing something *relative* to something else (and with the widespread use of numeraire changes this is to be taken quite literally). For example in the Black-Scholes model we do not say what the stock "should" cost, but given the current stock price (and of course its dynamics) we say what the option has to cost. Likewise it makes good sense in fixed income markets to take a number of objects as given. The market models have structures that make any hope of closed form solutions for objects other than those they are "made" to price seem futile. Certainly this is true for the highly specialized fixed income products that are traded "over-the-counter" by many investment banks. Thus there is a wide playing field for numerical creativity or financial engineering. So the models have obvious advantages and possibilities of future research. But there are other directions or skeptic remarks. One was the mutual inconsistency of models addressed in the previous excursion (see also Rebonato (1999)). Another is the question "How much model is actually left?" Since the models take "almost everything as given" it may be argued that their explanatory power is low. There is no "good story" behind the models and quite unclear how to do critical model control/comparison. Regarding the "good story"-issue we fully agree that models should have some plausible motivation, but the reflection of traders beliefs is not a bad such. The "model control"-issue is highly relevant and leads us into hitherto unchartered land. The models fit today's prices but may give us a false sense of security. If the dynamics – the volatility specifications – are wrong, then our prices and dynamic hedging strategies will be wrong. In other words, we will be losing money or taking unnecessary risks. We do not see how to solve the problem with "conventional" statistical methods. First we must find a good criterion for comparison and then we need to take into account both time-series and cross-sectional information and use data on many different types of contracts. It is all a question of finding a balance between historical data and the current state of the market.

3 Estimation of Discretely Observed Diffusion Processes

This section has only one subsection and the main part of it consists of 3 "excursions". We feel that the ordering of the material reflects a "natural progression" when it comes to estimation of discretely observed diffusion processes (so of course the papers were written the other way round). First, we adapt results form the literature that tell us that *i*) diffusions are Markovian so to determine the likelihood function we only have to find the (1-step) transition densities (it is enough to condition only on the previous observation) and, *ii*) (under weak conditions) the maximum likelihood estimator has the usual good properties. Unfortunately, the transition densities and hence the likelihood function are not generally known in closed form.

It is then natural to look for approximations to the densities. Since many characterizing features (probabilistic and analytic) of diffusions are known, many approximations have been suggested. The **first excursion** (based on Jensen & Poulsen (1999)) compares some of these methods. We ask questions such as: How fast are the approximations to calculate? How accurate are they? How can we tell? Do they converge? Of which order and what can we do to help them? How easy are they to implement?

The **second excursion** (based on Poulsen (1999*a*)) then proposes a new estimator based on one of the methods that "scores highly" in the first excursion; namely solution of the PDE for the transition density by finite difference methods. At the risk of sounding arrogant or ignorant, for people with "basic training" in mathematical finance it seems a blatantly obvious way to construct an estimator, yet we have not seen it used before.

In the **third excursion** (based on a part of Christensen & Poulsen (1999), the other part of that paper we encountered in the excursion related to non-linear drift models in Section 2.3.1) we compare the estimator proposed in the second excursion to some of the (many) other estimators used in the literature. We use both simulated data and "real" data for a model with unknown likelihood function. Compared to the other estimators, we find that the proposed new estimator has properties that most closely resemble those we would expect the maximum likelihood estimator to have based on asymptotic results. But we also find differences in the behaviour of the various estimators are quite small for models with "financially realistic parameters".

3.1 General Definitions and Results

We consider a set-up with a n-dimensional diffusion process X with dynamics governed by the SDE

$$dX_t = \mu(X_t;\psi)dt + \sigma(X_t;\psi)dW_t, \tag{41}$$

where $\mathbb{W}$ is a d-dimensional Brownian motion on some filtered probability space, $\psi \in \Psi \subseteq \mathbb{R}^d$, $\mu :\ \mathbb{R}^n \times \Psi \mapsto \mathbb{R}^n$ $\sigma :\ \mathbb{R}^n \times \Psi \mapsto \mathbb{R}^{d\times n}$ are functions such that (41) (with some initial condition) is well-defined and has a (weak) solution and (versions of) transition densities that are absolutely continuous w.r.t. the Lebesque measure and sufficiently smooth. By

$$\phi(t,x,y)$$

we denote the density of X_t given $X_0 = x$, i.e., ϕ is a density 'in y for fixed t and x' and we put $l(x_{i-1}, x_i;\psi) = \ln\phi(\triangle, x_{i-1}, x_i)$.

The parameter ψ is unknown but we seek statistical inference about it from (discretely observed) data points $x_0, x_1, \ldots, x_T$. We make the following notational convention/assumption.

Assumption 3.1 *Assume that:*

i) The true parameter is ψ_0.

ii) The observations are equidistant, $\triangle$ apart.

iii) X is stationary and ergodic.

Part *iii*) of this assumption may be tricky to check for multidimensional diffusions, i.e., when $n > 1$. Even for one-dimensional models the condition may fail to hold (but this is relatively easy to check). The trick is then (with the Itô formula as a key tool) to find an appropriate transform of the model that is stationary and ergodic. Typical example: look at log-returns rather that stock prices. In fact this idea can be applied "backwards" to build diffusion models (see for example Bibby & Sørensen (1997), Jensen & Pedersen (1997), Rydberg (1999)).

Since X is a Markov process w.r.t. its own filtration(cf. Øksendal (1995, Theorem 7.2)), the loglikelihood function l_T is easy to write down,

$$l_T(\psi) = \sum_{i=1}^{T} l(x_{i-1}, x_i;\psi),$$

and the maximum likelihood estimator (MLE) is defined in the usual way,

$$\psi_T^{ML} \quad = \quad \arg\sup_{\psi \in \Psi} \frac{1}{T} l_T(\psi).$$

Theorem 3.1 *Under Assumption 3.1 and weak regularity conditions (see e.g. Barndorff-Nielsen & Sørensen (1994)), which we assume hold, the MLE is consistent and asymptotically normal,*

$$\begin{aligned} \psi_T^{ML} &\overset{Pr}{\to} \psi_0, \\ \sqrt{T}(\psi_T^{ML} - \psi_0) &\overset{\sim}{\to} N(0, i^{-1}(\psi_0)), \end{aligned}$$

where $i(\psi)$ is the Fisher information matrix,

$$i = \lim_{T \to \infty} -\mathbb{E}\left(\frac{1}{T} \frac{\partial^2 l_T(\psi)}{\partial \psi \partial \psi^\top} \right).$$

Also, for any other consistent and asymptotically normal estimator with asymptotic covariance matrix V, we have that

$$V - i \quad \textit{is positive semi-definite,}$$

i.e., asymptotically ψ_T^{ML} has the "smallest possible variance".

Unfortunately, we do not generally know what ϕ looks like.

Excursion; Density Approximation

Several times we have stated that we do not know the transition densities of a diffusion process in general. But many ways of approximating the densities have been suggested. In Jensen & Poulsen (1999) we investigate the quality of such approximations. We look at 5 different suggestions.

Simple Analytic Approximation If $\triangle$ is small we can think of dX_t as $X_\triangle - X_0 = X_\triangle - x_0$, approximate any X_t's occurring on the RHS of (41) with X_0 and perceive 'dW_t' as a $N(0, \triangle)$-variable. This leads to the approximation

$$\phi^A(\triangle, x, y) = n(y; x + \mu(x)\triangle, \sigma^2(x)\triangle).$$

This is called an Euler-approximation and is (a certain sense) a first order approximation w.r.t. time, i.e., it is "good" when $\triangle$ is "small". Alas, $\triangle$ is typically not a quantity we (the statisticians or the financial modellers) can control.
For one-dimensional diffusions it is easy to characterize the stationary density (see (10)). If $i)$ $\triangle$ is large or, $ii)$ ϕ does not 'depend very much' on x, then this can be used as an approximation.

Simulation Recall that we seek the distribution of $X_\triangle$ given some value of X_0. If we could simulate outcomes of $X_\triangle$, say $(x^i_\triangle)_{i=1}^n$ then we could use statistical/econometric methods to estimate the density of $X_\triangle$. But since we cannot solve the SDE (if we could we would effectively know the transistion densities), we cannot simulate outcomes of $X_\triangle$ exactly. This means that a simulation based determination of transition densities has to address 3 major questions.

- How do we simulate solutions of SDEs?
  To do this we use discretization schemes (such as the Euler, the Milstein, or some higher order scheme) described in Kloeden, Platen & Schurz (1991).

- How do we generate long sequences of normally distributed numbers?
  The standard way to do this is to use pseudo random numbers. Specifically, we use numbers that "behave like independent $U(0,1)$ draws" supplied from a linear congruential operator and transform these into "independent $N(0,1)$" by the Box-Muller transform. The independence means that we can use the Central Limit Theorem to asses quantitatively the accuracy of the results.
  Recently, the use of so-called quasi random numbers has been advocated. Quasi random numbers, or more tellingly low discrepancy sequences, are series of numbers where the marginal distributions are, say, $N(0,1)$, but the numbers have a non-zero correlation structure. It can be shown that by choosing "cunning sequences", we can obtain higher rates of convergence of terms like $\frac{1}{n}\sum_{i=1}^n x^i_\triangle$ than indicated in the Central Limit Theorem.

- How do we estimate the density from the $x^i_\triangle$s?
  We are now in the land of kernel smoothers and bandwidth choices. There is a trade-off between how smooth you want our density estimate and how accurate you want it. Since smoothness is not our main concern, we use a rectangular kernel.

Binomial Approximation With Cox et al. (1979), binomial models were an instant classic in finance. In a binomial model we construct a tree (with splitting index 2) with probabilities assigned to branches and each route through the tree representing a sample path. In Nelson & Ramaswamy (1989) it is shown how to construct a sequence of recombining trees – lattices – such that the binomial models converge weakly to a specified diffusion. A binomial approximation in a natural way gives rise to an approximation of the density we seek.

Numerical Solution of the Fokker-Planck PDE We already saw in the excursion about Black-Scholes proofs in Section 2.2 that ϕ solves the one-dimensional parabolic PDE (where the ψ-dependence has been suppressed)

$$\frac{\partial}{\partial t}\phi(t,x,y) = -\frac{\partial}{\partial y}\left(\mu(y)\phi(t,x,y)\right) + \frac{1}{2}\frac{\partial^2}{\partial y^2}\left(\sigma^2(y)\phi(t,x,y)\right), \tag{42}$$

with initial condition $\phi(0,x,y) = \delta(y-x)$, where $\delta(\cdot)$ is the Dirac-δ function. If we are careful, quite useful approximations to the solution of this PDE can be obtained using finite difference methods, such as the Crank-Nicolson (CN) method. Since the CN method is pivotal for the calculation of the approximate likelihood estimator we present in the next excursion, we shall now describe it in some detail.

Let subscripts denote differentiation, suppress "x" and other arguments where it causes no confusion, and rewrite (42) as

$$\phi_t(t,y) = a(y)\phi + b(y)\phi_y + c(y)\phi_{yy}, \tag{43}$$

where

$$\begin{aligned} a(y) &= (\sigma_y)^2 + \sigma\sigma_{yy} - \mu_y = \frac{1}{2}(\sigma^2)_{yy} - \mu_y, \\ b(y) &= 2\sigma\sigma_y - \mu = (\sigma^2)_y - \mu, \\ c(y) &= \frac{1}{2}\sigma^2. \end{aligned}$$

Now consider a time/space grid (as depicted in Figure 4) with step sizes k and h. For any grid point (m,n) away from the boundaries (say y_L and y_H) we consider the approximation $v_m^n \approx \phi(nk, y_L + mh)$ defined by the equation

$$\frac{v_m^{n+1} - v_m^n}{k} = a\delta_0(h) + b\delta_1(h) + c\delta_2(h), \tag{44}$$

where the δ's are difference operators and the arguments of a, b, and c have been notationally suppressed; the functions should all be evaluated at the space point

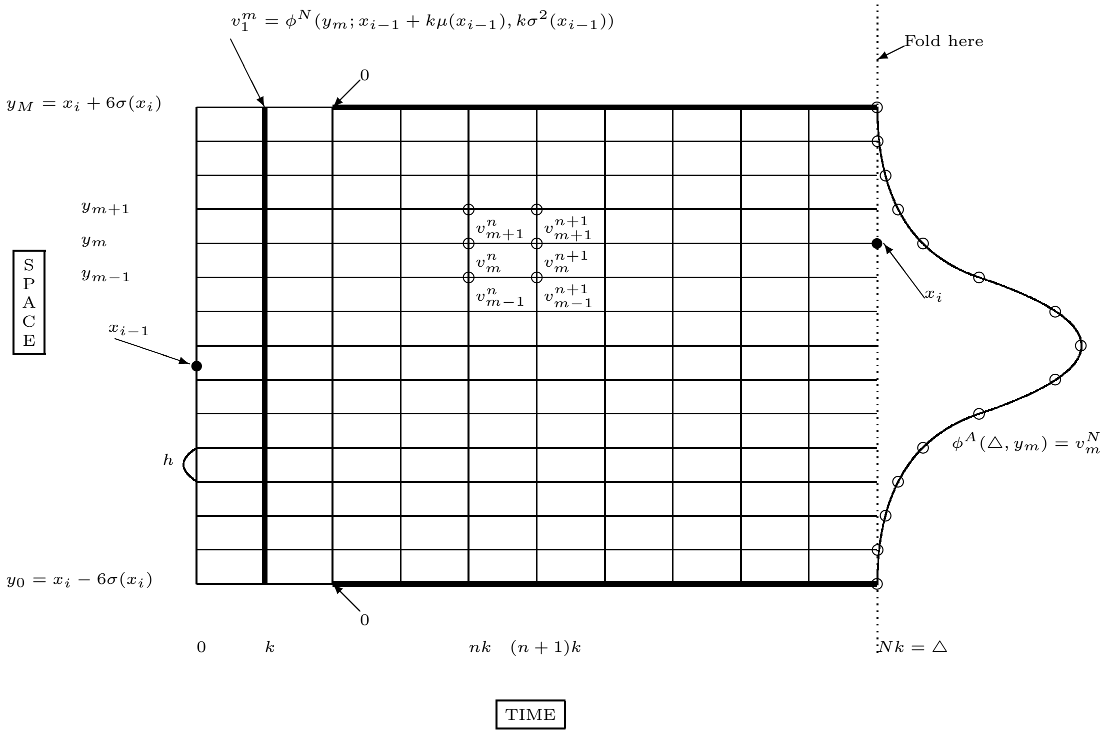

Figure 4: A finite difference grid for numerical solution of the forward PDE (42) for the transition density.

corresponding to the LHS, i.e., at $y_L + mh$. It is the particular choice of δ's that determines the accuracy and stability (and popular name) of the finite difference method. For the CN method the operators act in the following way:

$$\delta_0(h) = \frac{1}{2}v_m^{n+1} + \frac{1}{2}v_m^n, \tag{45}$$

$$\delta_1(h) = \frac{1}{2}\frac{v_{m+1}^{n+1} - v_{m-1}^{n+1}}{2h} + \frac{1}{2}\frac{v_{m+1}^n - v_{m-1}^n}{2h}, \tag{46}$$

$$\delta_2(h) = \frac{1}{2}\frac{v_{m+1}^{n+1} - 2v_m^{n+1} + v_{m-1}^{n+1}}{h^2} + \frac{1}{2}\frac{v_{m+1}^n - 2v_m^n + v_{m-1}^n}{h^2}. \tag{47}$$

Taylor expansions show that this method is locally second order accurate in both h and k. Also, the Crank-Nicolson method is unconditionally stable for parabolic PDEs (for a proof of this, or perhaps just to find out what it means, see Strikwerda (1989)). This allows us to use the Lax-Richtmeyer Equivalence Theorem Strikwerda (1989, Theorem 10.5.1) to conclude that the numerical solution converges. This theorem, however, does not give any statement about the order of convergence.

Defining $\lambda = kh^{-2}$ and inserting in (44) gives

$$\left(\frac{bh\lambda}{4} - \frac{c\lambda}{2}\right) v_{m-1}^{n+1} + \left(1 - \frac{ak}{2} + c\lambda\right) v_m^{n+1} + \left(-\frac{bh\lambda}{4} - \frac{c\lambda}{2}\right) v_{m+1}^{n+1} = \left(-\frac{bh\lambda}{4} + \frac{c\lambda}{2}\right) v_{m-1}^{n} + \left(1 + \frac{ak}{2} - c\lambda\right) v_m^{n} + \left(\frac{bh\lambda}{4} + \frac{c\lambda}{2}\right) v_{m+1}^{n}. \tag{48}$$

Considering v as known on the boundaries and using the initial condition, (48) defines a sequence of tridiagonal linear systems of equations. These systems can be solved recursively and each system requires only a number of operations that is proportional to the number of state space steps when the tridiagonal structure is exploited, e.g. by the routine `tridag` from Numerical Recipes. If we consider only diffusions with inaccessible boundaries (no reflection or absorption) then it is reasonable to put $v = 0$ on the grid boundaries. The grid boundary levels are chosen "sufficiently far apart" (the actual numbers are parameter dependent, but choosing them is not a problem). The initial condition is treated like this:

$$v_m^1 = n(y_l + mh; x + \mu(x)k, \sigma^2(x)k), \tag{49}$$

where n is the normal density. This is a first-order approximation to the density (in a certain sense). By definition of the Dirac-δ function this converges to the true initial condition as $k \to 0$. The CN method tends to behave badly for non-smooth initial data (as pointed out in Strikwerda (1989, page 121)), therefore the above choice is crucial.

Hermitian Expansion The idea of expanding an (unknown) density function by the use of Hermite polynomials goes back to Cramér (1925). Unfortunately, he showed that the class of densities, for which the Hermite expansion converges is rather limited. The density has to be "almost" normal to be in the class. Since most of the transition densities looked at in finance do not belong to that class, the idea of Hermite expansions has not been of much use in finance, until recently. A novel paper by Ait-Sahalia (1999) has changed this view completely. The main idea of his paper is to transform the diffusion (41) into another diffusion with a density that belongs to the class of converging Hermite expansions. A brief overview of the transformations, assumptions, main results and some 'trick of the trade' now follows. For a thorough understanding and in-depth analysis of the idea of Hermite expansions of a financial transition density, the reader is referred to Ait-Sahalia (1999). We first transform X into Y, where Y satisfies

$$dY_t = \mu_Y dt + dW(t),$$

(the transformation is basically $f : x \mapsto= \int_{x^{\#}}^{x} \frac{1}{\sigma(u)} du$), and then further into

$$Z = \sqrt{\triangle}(Y_{\triangle} - y_0).$$

Now define the Hermite polynomials as

$$H_j(z) := e^{\frac{z^2}{2}} \frac{d^j}{dz^j} \left[e^{-\frac{z^2}{2}} \right], \ j \geq 0.$$

(yes, that does define polynomials) and the J'th truncated density of Z as

$$\phi_Z^{(J)}(\Delta, y_0, z) := n(z) \sum_{j=0}^{J} \eta_j(\Delta, y_0) H_j(z), \tag{50}$$

where $n(z)$ is the standard normal density and

$$\eta_j(\Delta, y_0) := \frac{1}{j!} \int_{-\infty}^{+\infty} H_j(z) \phi_Z^{(J)}(\Delta, y_0, z) \, \mathrm{d}z.$$

The J'th truncated density of Y and X are then

$$\phi_Y^{(J)}(\Delta, y_0, y) = \Delta^{-\frac{1}{2}} \phi_Z^{(J)} \left(\Delta, y_0, \Delta^{-\frac{1}{2}} (y - y_0) \right), \tag{51}$$

$$\phi_X^{(J)}(\Delta, x_0, x) = \frac{\phi_Y^{(J)}(\Delta, f(x_0), f(x))}{\sigma(x)}. \tag{52}$$

In Ait-Sahalia (1999) it is shown that under regularity conditions

$$\phi_X^{(J)}(\Delta, x_0, y) \to \phi(\Delta, x_0, y), \quad \text{for } J \to \infty,$$

for all x_0, y in the domain of the diffusion. Suppose that the f-transformation and its inverse have been found and some J sufficiently large for our liking has been fixed. We then need the coefficients $\eta_j(\Delta, y_0)$, $j = 0, \ldots, J$, to compute $\phi_Z^{(J)}(\Delta, y_0, z)$. Explicit calculations yield

$$\eta_j(\Delta, y_0) = \frac{1}{j!} \mathbb{E} \left(H_j \left(\Delta^{-\frac{1}{2}} (Y_{t+\Delta} - y_0) \right) | Y_t = y_0 \right),$$

and this expectation can be evaluated using a Taylor-approximation around Δ.

We use the Vasicek, the CIR and the Black-Scholes models as benchmarks and compare the methods with a quantitative measure of the speed/accuracy trade-off,

run-time vs. error (maximal absolute and relative average),

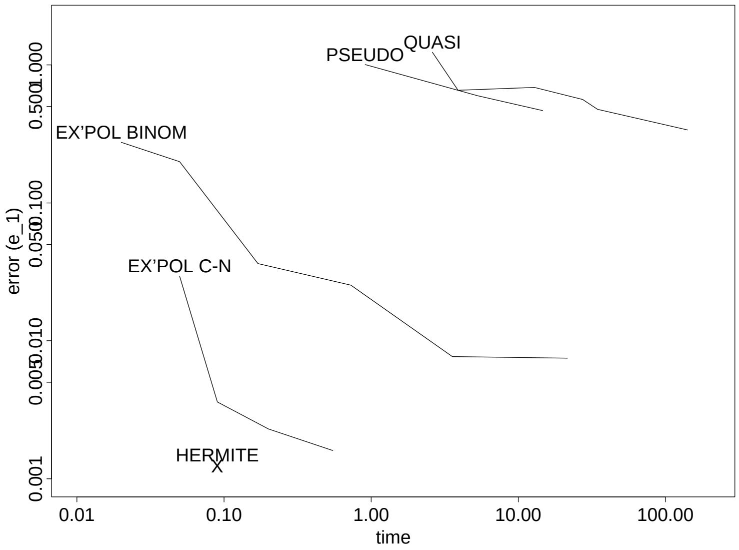

Figure 5: Speed (run-time in seconds) vs. accuracy (maximal absolute error) trade-off for density approximation techniques applied to the CIR model. The binomial and CN approximations have been Richardson extrapolated. There should also be a line for the Hermitian expansion, but it would be almost vertical.

as well as discuss the more qualitative features of the methods, such as flexibility and algorithm complexity. In short, the ordering of the descriptions of the methods also (and strongly) gives the speed/accuracy ordering. Figure 5 shows what a typical graphical representation of the speed/accuracy trade-off looks like. The run-time (x-axis with log-scale) is measured in seconds, the y-axis (also log-scale) measures the maximal absolute deviation from the true density function.

The simple analytic approximations are fast and easy to understand, but are inaccurate (and it is hard to tell *how* much) and their convergence is not within our control. The simulation methods are slow; it does not matter much whether we use pseudo random numbers or the more "fancy" quasi random numbers. The binomial approximation technique is reasonably fast, but can be a bit inflexible. Solution of the Fokker-Planck PDE by the Crank-Nicolson method may require some finesse to "fine tune", but is very fast and simple to program. However, on speed/accuracy considerations the Hermitian expansion developed by Ait-Sahalia stands head and shoulders above the other methods. The method does require some tedious model dependent analytic calculations, but (nowadays) these can easily be done with a symbolic calculator.

Excursion; AML Estimation

In Poulsen (1999*a*) we use the CN solution technique for the PDE to construct a new estimator. The idea for this approach comes from papers by Asger Pedersen (Pedersen (1995*a*) and Pedersen (1995*b*)), where the idea is to construct an approximation to the diffusion model in such a way that the quality of this approximation is controlled by the statistician and can be made arbitrarily good.

Formally, assume we have a second order approximation to the true likelihood.

Assumption 3.2 *Assume that the approximate loglikelihood can be written as*

$$\begin{aligned} l^A_{T,h}(\psi) &= \sum_{i=1}^{T} l^A_h(x_{i-1}, x_i; \psi) && (53)\\ &= \sum_{i=1}^{T} \Big(l(x_{i-1}, x_i; \psi) + h^2 a(x_{i-1}, x_i; \psi) + o(h^2) b(x_{i-1}, x_i; \psi) \Big) && (54)\\ &= l_T(\psi) + h^2 a_T(\psi) + o(h^2) b_T(\psi), && (55)\end{aligned}$$

where l_T is the true (but generally unknown) loglikelihood function and h is a parameter chosen by the statistician such that the time needed to compute the left hand side of (53) does not grow faster than T/h^2.

The reader should note that Assumption 3.2 could equally well be formulated with "just h's" (instead of h^2's), but we use h^2 deliberately as suggestive notation, since h will then correspond to the discretization step size in our numerical solution procedure. The reader may argue that if we use the CN method, then there are two discretization step sizes (one in time, one in space), but if a method is second order accurate in both directions then clearly we can combine these conceptually into one and obtain an approximation satisfying Assumption 3.2.

Define the approximate maximum likelihood estimator (AML) by

$$\psi^{AML}_{T,h} = \arg\sup_{\psi \in \Psi} \frac{1}{T} l^A_{T,h}(\psi).$$

In Poulsen (1999*a*) the following result is shown to hold under regularity conditions.

Theorem 3.2 *Suppose $h = h(T) = T^{-\delta}$. Then:*

i) $\psi_{T,h}^{AML}$ is consistent if (and only if) $\delta > 0$, i.e.,

$$\psi_{T,h}^{AML} \stackrel{Pr}{\rightarrow} \psi_0 \quad \textit{for } T \rightarrow \infty.$$

ii) If $\delta = 1/4$, then $\sqrt{T}(\psi_{T,h}^{AML} - \psi_0)$ converges in distribution to a normal variable or more precisely,

$$\sqrt{T}\left(\psi_{T,h}^{AML} - \psi_0\right) \stackrel{\sim}{\rightarrow} N\left(\frac{1}{2}\frac{\partial^2 \psi_h}{\partial h^2}|_{h=0}, i^{-1}(\psi_0)\right).$$

iii) If $\delta > 1/4$ then

$$\sqrt{T}(\psi_{T,h}^{AML} - \psi_0) \stackrel{\sim}{\rightarrow} N(0, i^{-1}(\psi_0)),$$

i.e., the AML estimator is asymptotically equivalent to the ML estimator.

This result shows that the AML estimator can be made to behave like the MLE and gives an asymptotically optimal trade-off between sample size and discretization choice in the finite difference solution method. In a way, the result should not be very surprising; for example in Cox & Hinkley (1974) it is noted that "estimates that differ from the ML estimate by o$(1/\sqrt{T})$ [which holds if $\delta > 1/4$] are also efficient". But note that the error term is controlled by us, the statisticians. The *regularity conditions* in Poulsen (1999*a*) can probably be weakened considerably. But that is not the aim of the paper. The main point is that by using the CN method described in the previous excursion it is possible - and in fact easy - to construct a likelihood approximation satisfying 3.2. It is also described how to verify numerically the second order accuracy; otherwise you should never trust "global" order statements based on "local" order considerations (derived from Taylor expansions).

It it now clear how to proceed: Use a "Black Box" to optimize, apply to data, and compare to other estimators. In practice we use a Quasi-Newton method for optimization. The optimization of a numerically determined function could be tricky, but is not since the function is concave, quite smooth and we have good starting points. The results of the application and comparison are given in the next excursion.

Excursion; Estimation of Short Rate Models

In Christensen & Poulsen (1999) we describe and compare some of the many different estimators that have been proposed for the estimation of discretely observed diffusion processes. In that paper we give a thorough description of the methods including both theoretical and implementational issues as well as Monte Carlo comparison and empirical results. Rather than repeating full method description here, we will just mention the methods considered in “buzz-word”/”key-reference”form.

(For obvious reasons:) **Approximate Maximum Likelihood** as described in the previous excursion.
Generalized Method of Moments Hansen (1982) is a standard 'theory' reference, which has little to do with SDEs, Chan et al. (1992) is a standard 'application' reference, the main objections to which have little to do with GMM.

Martingale methods The martingale property can also be used to great effect in this part of statistics as illustrated in Bibby & Sørensen (1995) in Sørensen (1997).

Indirect Inference Gouriéroux, Monfort & Renault (1993), Broze et al. (1995), Broze, Scaillet & Zakoian (1998).

Efficient Method of Moments A particularly cunning form of indirect inference, see Gallant & Long (1997), Gallant & Tauchen (1996*b*), and Gallant & Tauchen (1996*a*), Andersen & Lund (1997) is a successful application.

For Monte-Carlo comparison we use the CKLS-model with $\alpha = 0.2$, $\beta = -2$, $\sigma = 0.2$ and $\gamma = 0.8$ and a sampling frequency corresponding to weekly data and 500 observations in each of the (500) data sets. The β parameter may a bit large numerically, but otherwise the parameters are financially realistic. We find that

- All methods estimate the unconditional mean $(-\alpha/\beta)$ accurately; but this should be a minimal requirement.

- All methods that use correct first moments exhibit considerable bias in the estimation of β; β is estimated to be too large. This is bad news from a financial point of view since the β-estimates are already considerably closer to 0 than we like them. In fact, if we use the values estimated from real data in Monte Carlo, then we get quite horrible β-estimates.

- Methods that use (almost) correct second moments estimate γ quite accurately, and among these the AML performs best (albeit narrowly.)

- The standard error estimates produced by the various methods are quite reliable.

- The AML estimator tends to have the smallest variance among the estimators, as we should expect from its asymptotic equivalence to the MLE.

- It is interesting to note that undoubtedly the AML estimators have lower standard errors than the estimators that correspond to the starting points in the numerical optimization. The means that we are *not* just adding "white noise through a complicated numerical procedure".

- Quantile plots indicate normality of estimators.

Beside the conclusions already mentioned we arrive at the following conclusions applying the estimators to actual U.S. short rate data.

- Clear indication of a separate interest rate regime between October 1979 and October 1981 (this is well-known in the literature.)

- High γ-estimates (> 1.2 for all methods) when no 1979-1981 dummy is included, while γ is less than 1 for the period after 1981. This indicates that "misspecification is picked up in γ-estimates".

- Considerably larger variation between estimators than for simulated data.

4 Conclusion

4.1 What's New?

What have we done in this thesis? Or perhaps more accurately: What are the contributions of the manuscripts included?

Andreasen et al. (1998) is a pedagogical "general to specific" introduction to financial mathematics.

Hansen & Poulsen (1999) presents a toy model for the short term interest rate, where the unconditional mean in a Vasicek model is allowed to exhibit jumps. More general models nesting ours have been proposed, but we are able to do a great deal of analysis in an explicit and direct way.

Poulsen (1997) discusses the much-considered "lognormal LIBOR models". A number of the more or less subtle problems in derivative pricing in models of that type are investigated an generally found to be of little influence in practice.

Jensen & Poulsen (1999) gives a detailed description and comparison of techniques suggested for density approximation of diffusion processes. If Andreasen et al. (1998) might be called "Introduction to Financial Mathematics", then an alternative title to Jensen & Poulsen (1999) is "Introduction to Numerical Methods in Financial Mathematics". While the paper documents that the finite difference technique used in the AML construction is faster and more accurate than most other methods, it is "overtaken" by the Hermitian expansion technique suggested by Ait-Sahalia. This technique, however, does require some rather tedious "long-hand" calculations.

Poulsen (1999*a*) introduces a new estimator in the much-considered statistical/econometric field "estimation of discretely observed diffusion processes". The estimator (referred to as the AML estimator) is based on numerical solution of the PDE for the likelihood function and is (in an appropriate sense and under ditto conditions) equivalent to the MLE. Further, the method is easy to program, quite flexible and relatively fast. (Some might want to interchange the two latter adjectives.)

Christensen & Poulsen (1999) documents the advantages of the AML method and estimates the much-considered "non-linear drift short rate model". As opposed to previous studies we find virtually no evidence of non-linearities in the drift, and we prove that the CKLS-models passes goodness-of-fit tests.

Poulsen (1999*b*) is one of the first papers to quantitatively analyze the much-used option "sacking the manager". While a Poisson general linear model is used, and there are some delicate points regarding inference (related to "self-selection"), the paper is mainly included to "end on a happy note". (That is, unless you happen to manage an ill-performing football team.)

4.2 What's Next?

I am always quite skeptical of sections entitled "topics for future research". The ideas that people outline in such sections are by their very nature likely to be either extremely difficult to carry out, very vague, or plain "wild goose chases". Otherwise, they would already have pursued them themselves, or at the very least they wold not tell me.[1] This does not mean that research will not progress, but does mean that you should have little faith (and the younger the person, the smaller the faith) in people telling you what a "sure bet" is. In other words the "market" for research does not have many "arbitrage opportunities". Nonetheless, there are a number of "obvious" areas that call for further investigation.

- The numerical procedures in the lognormal LIBOR model may be used to solve "real" rather than "academic" problems which is what they are primarily used for as it stands.

- It may be possible to improve the numerical solution procedure used to obtain the approximate likelihood function used in the construction of the AML estimator.

- The conclusion based on sound likelihood analysis that the CKLS-model is well-specified and that no significant improvement is provided by non-linear drift terms is "nicely controversial". But still, all is not well short rate modelling. Firstly, when we perform Monte Carlo studies with parameter values equal to the estimates from real data, the estimates in simulated data behave very little like the original estimates. This suggests that the "large sample effect" has not "kicked in" yet, and may lead us on a quest to find estimators with better small sample properties.

[1] There are, of course, exceptions. Two famous examples are Hilbert's 23 open problems in mathematics presented at a conference in Paris 1900 (and they have been the focus of much the mathematics in this century) and Watson and Crick who proposed the double helix structure of DNA in the conclusion of an April 1953 article in *Nature* (and delivered the promised "details" the following month in the same journal).

Finally, on a more general note, the papers in this thesis largely take "partial views", it's either $\mathbb{P}$ or $\mathbb{Q}$; either time-series or cross-sectional. The future calls for *serious* attempts to merge the two approaches. The analysis can be both abstract/theoretical (as in Björk & Christensen (1999), Björk, Christensen & Gombani (1998) and Björk & Gombani (1999)) or empirical (as in Honoré (1998*b*)). By serious we mean that the $\mathbb{P} - \mathbb{Q}$ cross-over is taken as an integral part of the analysis. In other words we must seek to perform analysis that incorporates as much of the information in the market and its history as possible/reasonable. At the same time we must be critical about any particular parametric model specification. But because controlled experiments are not within our realm as economic researchers we should always require good or at least plausible stories behind the models, so we should not let the analysis be totally data driven and fall into the trap of data mining.

References

Ait-Sahalia, Y. (1996), 'Testing Continuous-Time Models of the Spot Interest Rate', *Review of Financial Studies* **9**(2), 385–426.

Ait-Sahalia, Y. (1999), 'Transition Densities for Iterest Rate and Other Nonlinear Diffusions', *Journal of Finance* **54**(4).

Andersen, L. & Andreasen, J. (1998), Volatility Skews and Extensions of the Libor Market Model. Working paper, General Re Financial Products, New York.

Andersen, L. & Brotherton-Ratcliffe, R. (1998), 'The equity option volatility smile: an implicit finite difference approach', *Journal of Computational Finance* **1**(2), 5–38.

Andersen, T. G. & Lund, J. (1996), Stochastic volatility and mean drift in the short rate diffusion: sources of steepness, level and curvature in the yield curve. Working Paper. Aarhus School of Business.

Andersen, T. G. & Lund, J. (1997), 'Estimating Continuous-Time Stochastic Volatility Models of the Short-Term Interest Rate', *Journal of Econometrics* **77**, 343–377.

Andreasen, J. (1996), Implied Modelling: Stable Implementation, Hedging and Duality. Working paper, Department of Operations Research, University of Aarhus.

Andreasen, J., Jensen, B. & Poulsen, R. (1998), 'Eight Valuation Methods in Financial Mathematics: The Black-Scholes Formula as an Example', *Mathematical Scientist* **23**(1), 18–40.

Balduzzi, P., Das, S., Foresi, S. & Sundaram, R. (1998), Stochastic mean models of the term structure. Forthcoming in Advanced Tools for the Fixed-Income Professional, edited by N. Jegadeesh and B. Tuckman, Wiley.

Barndorff-Nielsen, O. E. & Sørensen, M. (1994), 'A review of some aspects of asymtotic likelihood theory for stochastic processes', *International Statistical Review* **62**(1), 133–165.

Beaglehole, D. R. & Tenney, M. S. (1991), 'General solutions of some interest rate-contingent claim pricing equations', *Journal of Fixed Income* **1**, 69–83.

Bernstein, P. L. (1992), *Capital Ideas*, Free Press.

Bibby, B. M. & Sørensen, M. (1995), 'Martingale estimating functions for discretely observed diffusion processes', *Bernoulli* **1**, 17–39.

Bibby, B. M. & Sørensen, M. (1997), 'A hyperbolic diffusion model for stock prices', *Finance and Stochastics* **1**(1).

Bjerksund, P. & Stensland, G. (1996), 'Implementing the Black-Derman-Toy Interest Rate Model', *Journal of Fixed Income* **6**(2), 67–75.

Björk, T. (1998), *Arbitrage Theory in Continuous Time*, Oxford.

Björk, T. & Christensen, B. J. (1999), Interest rate dynamics and consistent forward rate curves. Working paper no. 27, Centre for Analytical Finance, University of Aarhus.

Björk, T., Christensen, B. J. & Gombani, A. (1998), 'Some System Theoretic Aspects of Interest Rate Theory', *Insurance, Mathematics and Economics* pp. 17–23.

Björk, T., di Masi, G., Kabanov, Y. & Runggaldier, W. (1997), 'Towards a General Thoery of Bond Markets', *Finance and Stochastics* **1**, 141–174.

Björk, T. & Gombani, A. (1999), 'Minimal Realizations of Interest Rate Models', forthcomming in *Finance and Stochastics* .

Björk, T., Kabanov, Y. & Runggaldier, W. (1997), 'Bond Market Structure in the Presence of Marked Point Processes', *Mathematical Finance* **7**, 211–239.

Black, F., Derman, E. & Toy, W. (1990), 'A one-factor model of interest rates and its application to treasury bond options', *Financial Analysts Journal* pp. 33–39.

Black, F. & Scholes, M. (1973), 'The pricing of options and corporate liabilities', *Journal of Political Economy* **81**, 637–654.

Brace, A., Gatarek, D. & Musiela, M. (1997), 'The market model of interest rate dynamics', *Mathematical Finance* **7**(2), 127–154.

Brace, A. & Musiela, M. (1994), 'A multifactor gauss markov implementation of heath, jarrow, and morton', *Mathematical Finance* **4**, 259–283.

Brennan, M. J. & Schwartz, E. S. (1979), 'A continuous time approach to the pricing of bonds', *Journal of Banking and Finance* pp. 133–155.

Broze, L., Scaillet, O. & Zakoian, J.-M. (1995), 'Testing for Continuous-Time Models of the Short-Term Interest Rate', *Journal of Empirical Finance* **2**, 199–223.

Broze, L., Scaillet, O. & Zakoian, J.-M. (1998), 'Quasi-Indirect Inference for Diffusion Processes', *Econometric Theory* **14**, 161–186.

Buttimer, R. J., Muller, W. J. & Reeves, R. D. (1995), 'An Alternative Trinomial Formulation for One-Factor Term Structure Models', *Journal of Financial Engineering* **4**(1).

Chan, K. C., Karolyi, G. A., Longstaff, F. A. & Sanders, A. B. (1992), 'An Empirical Comparison of Alternative Models of the Short-Term Interest rate', *Journal of Finance* **47**(3), 1209–1227.

Chapman, D. A. & Pearson, N. D. (1999), 'Is the Short Rate Drift Actually Nonlinear?', forthcomming in *Journal of Finance* .

Cheyette, O. (1995), Markov Representation of the Heath-Jarrow-Morton Model. Working paper. BARRA Inc. Berkeley, California.

Christensen, B. J. & Poulsen, R. (1999), Optimal Martingale and Likelihood Methods for Models of the Short Rate of Interest, With Monte Carlo Evidence for the CKLS Specification and Applications to Non-Linear Drift Models. Working paper. Department of Operations Research. University of Aarhus.

Conley, T. G., Hansen, L. P., Luttmer, E. G. & Scheinkmann, J. A. (1997), 'Short-Term Interest Rates as Subordinated Diffusions', *Review of Financial Studies* **10**(3), 525–577.

Constantinides, G. M. (1992), 'A theory of the nominal term structure of interest rates', *Review of Financial Studies* **5**, 531–552.

Cox, D. R. & Hinkley, D. V. (1974), *Theoretical Staistics*, Chapman-Hall.

Cox, J. C., Ingersoll, J. E. & Ross, S. A. (1985), 'A Theory of the Term Structure of Interest Rates', *Econometrica* **53**, 385–407.

Cox, J., Ross, S. & Rubinstein, M. (1979), 'Option pricing: A simplified approach', *Journal of Financial Economics* **7**, 229–263.

Cramér, H. (1925), On Some Classes of Series Used in Mathematical Statistics, *in* 'Proceedings of the Sixth Scandinavian Mathematical Congress', pp. 399–425.

Duffie, D. (1996), *Dynamic Asset Pricing Theory*, 2. edn, Princeton University Press.

Duffie, D. & Kan, R. (1996), 'A Yield-Factor Model of Interest Rates', *Mathematical Finance* **6**(4), 379–406.

Duffie, D., Ma, J. & Young, J. (1995), 'Black's consol rate conjecture', *The Annals of Applied Probability* **5**(2), 356–382.

Dupire, B. (1997), Pricing and Hedging with Smiles, *in* M. Dempster & S. Pliska, eds, 'Mathematics of Derivative Securities', Cambridge University Press, pp. 103–111.

Flesaker, B. & Hughston, L. (1996), 'Positive Interest', *Risk* **9**(1).

Gallant, A. R. & Long, J. R. (1997), 'Estimating Stochastic Differential Equations Efficiently by Minimum Chi-Square', *Biometrika* **84**, 125–142.

Gallant, A. R. & Tauchen, G. E. (1996*a*), Specification Analysis of Continuous Time Models in Finance, *in* P. E. Rossi, ed., 'Modelling Stock Market Volatility: Bridging the Gap to Continuous Time', Academic Press, pp. 357–383.

Gallant, A. R. & Tauchen, G. E. (1996*b*), 'Which Moments to Match', *Econometric Theory* **12**, 657–681.

Geman, H. & Yor, M. (1993), 'Bessel Processes, Asian Options, and Perpetuities', *Mathematical Finance* **3**(4), 349–375.

Gouriéroux, C., Monfort, A. & Renault, E. (1993), 'Indirect Inference', *Journal of Applied Econometrics* **8**, 85–118.

Hansen, A. T. & Poulsen, R. (1999), A simple regime switching model. Working paper no. 30, Centre for Analytical Finance, University of Aarhus.

Hansen, L. P. (1982), 'Large Sample Properties of Generalized Method of Moments Estimators', *Econometrica* **50**, 1029–1054.

Harrison, J. M. & Kreps, D. M. (1979), 'Martingales and arbitrage in multiperiod securities markets', *Journal of Economic Theory* **20**, 381–408.

Harrison, J. M. & Pliska, S. R. (1981), 'Martingales and stochastic integrals in the theory of continuous trading', *Stochastic Processes and their Applications* **11**, 215–260.

Heath, D., Jarrow, R. & Morton, A. (1992), 'Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation', *Econometrica* **60**(1), 77–105.

Ho, T. S. Y. & Lee, S. (1986), 'Term Structure Movements and Pricing Interest Rate Contingent Claims', *Journal of Finance* **41**(5), 1011–1029.

Hoffman, I. D. (1993), Lognormal Processes in Finance, PhD thesis, University of New South Wales.

Hoffmann-Jørgensen, J. (1994), *Probability with a View Toward Statistics*, Chapman & Hall.

Hogan, M. & Weintraub, K. (1993), The Log-normal Interest Rate Model and Eurodollar Futures. Working Paper. Citibank, New York.

Honoré, P. (1998*a*), Five Essays on Financial Econometrics in Continuous-Time Models, PhD thesis, Department of Finance. Aarhus School of Business.

Honoré, P. (1998*b*), Panel-Data Estimation of Non-Linear Term-Structure Models. Working paper no. 19, Centre for Analytical Finance, University of Aarhus.

Hull, J. & White, A. (1990), 'Pricing interest rate derivative securities', *Review of Financial Studies* **3**, 573–592.

Hull, J. & White, A. (1994), 'Numerical Procedures for Implementing Term Structure Models', *Journal of Derivatives* **2**.

Jamshidian, F. (1989), 'An exact bond option formula', *Journal of Finance* **44**, 205–209.

Jamshidian, F. (1991), 'Bond and option valuation in the Gaussian interest rate model', *Research in Finance* **9**, 131–170.

Jamshidian, F. (1996), 'Bond, futures and option evaluation in the quadratic interest rate model', *Applied Mathematical Finance* **3**, 93–115.

Jamshidian, F. (1997), 'Libor and swap market models and measures', *Finance and Stochastics* **1**, 293–330.

Jensen, B. & Poulsen, R. (1999), A Comparsion a Approximation Techniques for the Transistion Densities of Diffusion Processes. Working paper, Department of Operations Research, University of Aarhus.

Jensen, J. L. & Pedersen, J. (1997), A note on model for stock prices. Working paper no. 7, Centre for Analytical Finance, University of Aarhus.

Johnson, N., Kotz, S. & Balakrishnan, N. (1994), *Continuous Univariate Distributions*, Vol. 2, 2 edn, Wiley.

Karatzas, I. & Shreve, S. (1992), *Brownian Motion and Stochastic Calculus*, 2 edn, Springer-Verlag.

Karlin, S. & Taylor, S. (1981), *A Second Course in Stochastic Processes*, Academic Press.

Kloeden, P. E., Platen, E. & Schurz, H. (1991), *Numerical Solution of SDE Through Computer Experiments*, Springer.

Knez, P., Litterman, R. & Scheinkman, J. (1994), 'Explorations into factors explaining money market returns', *Journal of Finance* **49**, 1861–1682.

Koedijk, K., Nissen, F., Schotman, P. & Wolff, C. (1997), 'The Dynamics of the Short Rate Reconsidered', *European Finance Review* **1**(1), 105–130.

Kreps, D. M. (1990), *A Course in Microeconomic Theory*, Harvester Wheatsheaf.

Landén, C. (1999), Bond pricing in a hidden markov model of the short rate. forthcomming in Finance and Stochastics.

Litterman, R. & Scheinkman, J. (1991), 'Common Factors Affecting Bond Returns', *Journal of Fixed Income* pp. 54–61.

Longstaff, F. A. & Schwartz, E. (1992), 'Interest rate volatility and the term structure: a two-factor general equilibrium model', *Journal of Finance* **47**, 1259–1282.

Luenberger, D. G. (1998), *Investment Science*, Oxford.

Mceneaney, W. (1997), 'A Robust Control Framework for Option Pricing', *Mathematics of Operations Research* **22**(1), 202–221.

Merton, R. (1973), 'The theory of rational option pricing', *Bell Journal of Economics and Management Science* **4**, 141–183.

Merton, R. (1976), 'Option pricing when the underlying stock returns are discontinuous', *Journal of Financial Economics* **5**, 125–144.

Miltersen, K., Sandmann, K. & Sondermann, D. (1997), 'Closed form solutions for term structure derivatives with log-normal interest rates', *Journal of Finance* **52**, 409–430.

Morton, A. (1988), Arbitrage and Martingales, Technical Report 821, College of Engineering, Cornell University.

Musiela, M. (1993), Stochastic PDEs and term structure models. *Journées Internationales de Finance, IGR-AFFI.*

Musiela, M. & Rutkowski, M. (1997), *Martingale Methods in Financial Modelling*, 2. edn, Springer-Verlag.

Naik, V. & Lee, M. H. (1994), The yield curve and bond option prices with discrete shifts in economic regimes. Working Paper. University of British Columbia.

Nelson, D. & Ramaswamy, K. (1989), 'Simple Binomial Processes as Diffusion Approximations in Financial Models', *Review of Financial Studies* **3**, 393–430.

Nowman, K. B. (1997), 'Gaussian Estimation of Single-Factor Continuous Time Models of the Term Structure of Interest Rates', *Journal of Finance* **52**(4), 1695–1706.

Øksendal, B. (1995), *Stochastic Differential Equations*, 4. edn, Springer-Verlag.

Pedersen, A. R. (1994), Uniform Residuals for Discretely Observed Diffusion Processes. Research report no. 292. Department of Statistics. University of Aarhus.

Pedersen, A. R. (1995*a*), 'A New Approach to Maximum Likelihood Estimation for Stochastic Differential Equations Based on Discrete Observations', *Scandinavian Journal of Statistics* **22**(1), 55–71.

Pedersen, A. R. (1995*b*), 'Consistency and asymptotic normality of an approximate maximum likelihood estimator for discretely observed diffusion processes', *Bernoulli* **1**(3), 257–279.

Poulsen, R. (1997), Stability of Derivative Prices in Market Models. Working paper. Department of Operations Research. University of Aarhus.

Poulsen, R. (1999*a*), Approximate Maximum Likelihood Estimation of Discretely Observed Diffusion Processes. Working paper no. 29, Centre for Analytical Finance, University of Aarhus.

Poulsen, R. (1999*b*), Should He Stay or Should He Go? Estimating the Effect of Sacking the Manager in Association Football. Working paper. Department of Operations Research. University of Aarhus.

Pritsker, M. (1998), 'Nonparametric Density Estimation and Tests of Continuous Time Interest Rate Models', *Review of Financial Studies* **11**(3), 449–487.

Rebonato, R. (1998), *Interest-Rate Option Models*, 2 edn, Wiley.

Rebonato, R. (1999), 'On the pricing implications of the joint lognormal assumption for the swaption and cap markets', *Journal of Computational Finance* **2**, 57–76.

Ritchken, P. & Sankarasubramanian, L. (1995), 'Volatility Structure of Forward Rates and the Dynamics of the Term Structure', *Mathematical Finance* **5**, 66–72.

Rogers, L. C. G. (1995), Which Model for Term-Structure of Interest Rates Should One Use?, *in* M. Davis, ed., 'IMA Vol 65: Mathematical Finance', Springer.

Rydberg, T. H. (1999), 'Generalized hyperbolic diffusions with applications towards finance', *Mathematical Finance* **9**, 183–201.

Sandmann, K. & Sondermann, D. (1997), 'A Note on the Stability of Lognormal Interest Rate Models and the Pricing of Eurodollar futures', *Mathematical Finance* **7**(2), 119–125.

Schlögl, E. & Sommer, D. (1997), Factor models and the shape of the term structure. Discussion paper B-395. University of Bonn.

Shirakawa, H. (1991), 'Interest rate option pricing with Poisson-Gaussian forward rate curve processes', *Mathematical Finance* **1**(4), 77–94.

Sørensen, M. (1997), Estimating Functions for Discretely Observed Diffusions : A Reviev, *in* I. Basawa, V. P. Godambe & T. R. L., eds, 'Selected Procedings of the Symposium on Estimating Functions', Vol. 32 of *IMS Leture notes - Monograph series*, pp. 305–325.

Stanton, R. (1997), 'A Nonparametric Model of Term Structure Dynamics and the Market Price of Interest Rate Risk', *Journal of Finance* **52**, 1973–2002.

Steeley, J. M. (1990), 'Modelling the Dynamics of the Term Structure of Interest Rates', *Economic and Social Review* **21**(4), 337–361.

Strikwerda, J. (1989), *Finite Difference Schemes and Partial Differential Equations*, Wadsworth & Brooks/Cole.

van der Hoek, J. (1998), Derivative Pricing (A Note). Working paper. Department of Applied Mathematics, University of Adelaide.

Vasicek, O. (1977), 'An Equilibrium Characterization of the Term Structure', *Journal of Financial Economics* **5**, 177–188.

Wilmott, P., Dewynne, J. & Howison, S. (1998), *Option Pricing: Mathematical Models and Computation*, Oxford.

Manuscripts

Paper I: Eight Valuation Methods in Financial Mathematics: The Black-Scholes Formula as an Example, 30 pages

Paper II: A simple regime switching term structure model, 28 pages

Paper III: Stability of Derivative Prices in Market Models, 30 pages

Paper IV: A Comparison of Approximation Techniques for Transition Densities of Diffusion Processes, 58 pages

Paper V: Approximate Maximum Likelihood Estimation of Discretely Observed Diffusion Processes, 33 pages

Paper VI: Optimal Martingale and Likelihood Methods for Models of the Short Rate of Interest, With Monte Carlo Evidence for the CKLS Specification and Applications to Non-Linear Drift Models, 46 pages

Paper VII: Should He Stay or Should He Go? Estimating the Effect of Sacking the Manager in Association Football, 8 pages

Eight Valuation Methods in Financial Mathematics: The Black-Scholes Formula as an Example

Jesper Andreasen*

Bjarke Jensen†

Rolf Poulsen‡

June 24, 1999

Abstract

This paper describes a large number of valuation techniques used in modern financial mathematics. Though the approaches differ in generality and rigour, they are consistent in a very noteworthy sense: each model has the celebrated Black-Scholes formula for the price of a call-option as a special case.

Key Words: Financial mathematics, option pricing, forward and backward partial differential equations, martingale measures, change of numeraire, local time, binomial model, equilibrium.

JEL-code: G-12.

Running Head: Valuation techniques in finance.

*General Re Financial Securities Ltd, 199 Bishopsgate, London EC2M 3TY.

†Department of Management, School of Economics and Management, University of Aarhus, dk-8000 Denmark.

‡Department of Theoretical Statistics and Operations Research, Institute of Mathematical Sciences, University of Aarhus, dk-8000 Denmark.

1 Introduction

This paper serves as an account of the sophisticated and varied techniques that have been developed in financial mathematics during the last 20 years. We illustrate how these can be used to derive a result that most people with interest in finance are familiar with, the Black-Scholes formula. The real power of the techniques naturally lies in their ability to cope with different generalisations of the basic set-up of Black and Scholes. Some methods are easily described in a more general setting and the Black-Scholes formula then appears as a special case. Other methods are not possible to describe in generality within limited space. When this is the case, we consider the basic Black-Scholes set-up, derive the formula using the particular method, and then indicate the range and applicability of the method. In either case this should allow the reader not familiar with the subject to gain insight into the techniques used in financial mathematics. For people who are familiar with the field, 'we aim to please.' The hope is that any person in this group will see at least one proof and say 'Ah, yes. I hadn't thought of that one.'

The outline of the paper is as follows: In Section 2 we formulate the problem, introduce some central concepts, and present the Black-Scholes formula. In other words we ask the question *and* give the answer up front. The next 8 sections describe ways of getting from point A to point B(.S.) Ways that lead us past answers to more general questions than the one originally posed. By a hedge argument (which was the ingenious insight of Black, Scholes and Merton) the fundamental partial differential equation (PDE) for arbitrage free asset prices is derived in Section 3. Section 4 shows how martingale techniques can be used to solve the pricing problem and stresses the relationship between means of solutions of stochastic differential equations (SDEs) and PDEs. Section 5 shows how the seemingly neutral concept of using different numeraires can turn out to be a very powerful tool, in fact we derive the Black-Scholes formula without calculating a single integral. In Section 6 we initially try to 'mess with your head' by proposing a strategy that seemingly contradicts the previous results. But we show that a careful inspection and some advanced stochastic calculus not only resolves the paradox, but also provides an extra proof. It is shown in Section 7 that the price of the call-option also satisfies a PDE that runs in strike price and maturity date, a forward equation. Not only does this give another proof of the result but also has practical implications for numerical purposes. Section 8 derives the formula as a limiting case of a discrete binomial model. The proposed convergence proof is different from most other convergence proofs in the literature and highlights an interesting similarity between numeraire/measure changes in discrete and continuous cases. Section 9 shows that we can also derive the formula from the continuous time CAPM model, which links together two of the most celebrated results in financial economics. Utility maximisation of a representative agent with a utility function exhibiting constant relative risk aversion is shown also to do the trick in Section 10. Section 11 sums up the contributions of the paper and discusses the results.

2 The question and the answer

The basic Black-Scholes set-up consists of non-dividend paying stock, the price of which is assumed to be the solution to the SDE

$$\frac{dS_t}{S_t} = \mu dt + \sigma dW_t^{\mathcal{P}}, \tag{1}$$

where μ and σ are constants and $(W_t^{\mathcal{P}})$ is a Brownian motion on some filtered probability space $(\Omega, (\mathcal{F}_t), \mathcal{P})$, and a bond with price dynamics given by

$$\frac{dB_t}{B_t} = rdt, \quad B_T = 1,$$

where r is the (continuously compounded) interest rate which is assumed to be constant.[1]

Our aim is to price a European call-option on the stock with maturity date T and strike price K. This is a security that gives the bearer the right, but not the obligation, to buy one share of stock at time T (and only at that time) for a price of K \$. Hence the contract has a terminal pay-off of

$$\max(S_T - K, 0) \equiv (S_T - K)^+$$

and no intermediate payments.

We will assume that there are no transactions costs, no short-selling constraints and that all assets are perfectly divisible. Furthermore we will allow investors to continuously readjust their portfolios. Specifically, a *trading strategy* (a_t, b_t) is a *predictable* stochastic process satisfying certain technical conditions.[2] To us, a_t will represent the

[1]Notice that this is an ordinary differential equation with solution given by: $B_t = e^{-r(T-t)}$.

[2]The trading strategy will be stochastic because it depends on the stock and bond whose price evolution is stochastic. However, at any given time t we will know how much to hold. Among other reasons the technical conditions the strategy has to fulfill is to exclude *doubling strategies*. See Duffie (1992) for details.

number of stocks held at time t, while b_t is the bond holdings. $V_t = a_t S_t + b_t B_t$ is the *value process.* A trading strategy is called *self-financing* if

$$dV_t = a_t dS_t + b_t dB_t,$$

which means that we only make an investment today. The gains are reinvested, and we do not use extra funds to cover our losses (this does not necessarily mean constantly calling your stock-broker, 'buy-and-hold'-strategies are evidently self-financing). An *arbitrage opportunity* is a self-financing trading strategy such that either

$$V_0 \leq 0,\ \mathcal{P}(V_T \geq 0) = 1,\ \mathcal{P}(V_T > 0) > 0,$$

or[3]

$$V_0 < 0,\ \mathcal{P}(V_T \geq 0) = 1.$$

So, an arbitrage opportunity is 'something for nothing' or 'a free lunch'. Reasonably, though human intuition about stochastic phenomena is notoriously poor, we cannot have such strategies in the economic equilibrium. There would be an infinite demand for the 'arbitrage strategy', while no agent (without a serious financial death wish) would be willing to supply it.

From pure static arbitrage considerations the only bounds that can be put on the call-option price are:[4]

$$S_t \geq C_t \geq (S_t - B_t K)^+ .$$

The main contribution of Black & Scholes (1973) is that they close the gap and give an exact pricing formula by dynamic arbitrage arguments.

Result 1 (The Black-Scholes Formula) *If the setting is as described above then to prevent arbitrage opportunities we must have $C_t = C(S_t, t)$ where*

$$\begin{aligned} C(x,t) &= x\Phi(z) - e^{-r(T-t)} K \Phi(z - \sigma\sqrt{T-t}) \qquad (2) \\ z &= \frac{\ln(\frac{x}{K}) + (r + \frac{\sigma^2}{2})(T-t)}{\sigma\sqrt{T-t}} \end{aligned}$$

and Φ denotes the cumulative density of the standard normal distribution.

In order to replicate the pay-off of the call-option we should hold

$$a_t = \frac{\partial C}{\partial S}(S_t, t) = \Phi(z)$$

[3]In incomplete markets the two conditions are not equivalent.

[4]This is given that the call-option contract is the only existing derivative security written on the stock. If there exists several option contracts in the economy, say with different strikes, there would be static arbitrage bounds between these contracts.

shares of stock and

$$C(S_t, t) - a_t S_t$$

in bonds.

This is a remarkable result that has been twisted and turned in the literature for more than twenty years. The most noteworthy thing is that the instantaneous expected return of the stock, μ, does not enter the expression. In other words: Two investors need not agree on the expected return of the stock in order to agree about the option price.[5]

Note the, at least mathematical, simplicity of the replicating portfolio, and that the position in the stock is bounded above by 1.

This method of pricing is *relative*. We price the option in terms of the stock and bond, whose prices are taken as given. We do not need any general equilibrium constraints on the economy other than there being 'no free lunches'. We shall later see that we can arrive at the result from a general equilibrium model, but this is in a sense 'overkill': The above conditions are exactly what we need.

3 The hedge argument and the fundamental PDE

The technique presented in this section is the one originally used by Black & Scholes (1973) to derive the formula that now bears their names. The result was simultaneously and independently derived by Merton (1973*b*). Let Y_t denote the price of a call-option with strike K and maturity T. Now *assume* that Y_t can be written as a twice continuously differentiable function of S_t (hence, no dependence on past S_u's) and t. That is

$$Y_t = C(S_t, t).$$

The Ito formula applied to Y_t yields

$$dY_t = \left(\mu S_t \frac{\partial C}{\partial S} + \frac{\partial C}{\partial t} + \frac{1}{2}\sigma^2 S_t^2 \frac{\partial^2 C}{\partial S^2}\right) dt + \frac{\partial C}{\partial S}\sigma S_t dW_t^{\mathcal{P}}, \tag{3}$$

where some of the dependences have been notationally suppressed.

In the notation of Section 2 *assume* that a self-financing trading strategy (a_t, b_t) exists

[5] It is worth noting that increasingly frequent discrete sampling of the underlying stock gives an improved estimate of the volatility but high frequency sampling does not necessarily improve the estimate of the drift. So if the stock price is only observed at frequent but discrete times it is likely that investors will agree on the volatility but not necessarily on the drift. For a derivation of this see Ingersoll (1987).

such that

$$a_t S_t + b_t B_t = Y_t, \quad \forall t \in [0,T]. \tag{4}$$

By linearity of stochastic integrals and the self-financing condition we have

$$dY_t = (a_t \mu S_t + b_t r B_t)dt + a_t \sigma S_t dW_t^{\mathcal{P}}.$$

We now have two Ito-expressions for dY_t. This means (by the Unique Decomposition Theorem, see e.g. Duffie (1992)) that the drift and diffusion terms in these must be equal.
Matching diffusion terms yields (since $S_t > 0$ $\mathcal{P}$ -a.s.)

$$a_t = \frac{\partial C}{\partial S}(S_t, t),$$

which gives us the number of shares of stock to hold. On the other hand from (4)

$$S_t \frac{\partial C}{\partial S}(S_t, t) + b_t B_t = C(S_t, t),$$

so

$$b_t = \frac{1}{B_t}(C(S_t, t) - S_t \frac{\partial C}{\partial S}(S_t, t)).$$

From the drift terms we get

$$r S_t \frac{\partial C}{\partial S}(S_t, t) + \frac{1}{2}\sigma^2 S_t^2 \frac{\partial^2 C}{\partial S^2}(S_t, t) + \frac{\partial C}{\partial t}(S_t, t) = rC(S_t, t),$$

which holds if the function C satisfies the PDE

$$rx \frac{\partial C}{\partial x}(x, t) + \frac{1}{2}\sigma^2 x^2 \frac{\partial^2 C}{\partial x^2}(x, t) + \frac{\partial C}{\partial t}(x, t) = rC(x, t). \tag{5}$$

We immediately note that the exact same argument holds for all types of derivative assets whose prices depend only on time and some Markov process S_t . The only things that differ are the boundary conditions. For this reason (5) is referred to as the fundamental PDE for arbitrage free asset pricing.
As we are considering a European call-option, C should further satisfy the boundary condition[6]

$$C(x, T) = (x - K)^+. \tag{6}$$

To find the arbitrage free call-option price we have to solve (5)-(6). Given Section 2, the easiest thing is to verify the result by a direct calculation. Originally, Black and

[6]In fact we should also require $x \geq C(x,t) \geq (x - e^{-r(T-t)}K)^+$, but this turns out to be automatically satisfied in this case.

Scholes (probably) used a Fourier transform technique which is mainly of historical interest nowadays.

Is this price unique, one might ask. Yes it is. Given that (5)-(6) has been solved we have found a (hopefully) self-financing trading strategy that replicates the option. If the option had any other price than the initial investment in this replicating portfolio, we would just sell the option and buy the replicating portfolio (or the other way round, depending on which is the cheaper alternative). This would leave us with a risk-free profit. An arbitrage opportunity !

Now, the reader might argue: 'How do we know that the proposed trading strategy is indeed self-financing?' This is a fair point to make, we have just assumed so far, that this was the case, and made explicit use of it. However, using (5) (which in particular means that C is $\mathbf{C}^2$) and the Ito formula it is easy to show that (a_t, b_t) is indeed self-financing. We make this seemingly 'round-about' comment because later (in Section 6) we will meet a trading strategy that is obviously self-financing. Except: It isn't !

By a hedge argument this section derived the fundamental PDE for asset prices. By specifying the right boundary condition, the Black-Scholes formula emerged. This approach generalises easily to the case of dividend paying assets. Provided that covariance matrix for stock returns is of full rank ('a complete market'), the results also carry over to higher dimensions (see e.g. Duffie (1992)). The technique also works for American type assets (see Ingersoll (1987)), though the boundary conditions become more subtle, in fact determination of these becomes part of the solution. What this approach is NOT suitable for is non-Markovian settings, such as stochastic volatility in the stock or path dependent features in the derivative (e.g. Asian options). The next section develops a more general valuation technique, that coincides with the PDE approach when the latter is applicable.

4 The martingale approach

In this section we will derive the martingale pricing approach that does not depend on the Markov property of the stock and draw the analogy to the PDE obtained in the previous section. Using the martingale valuation technique we calculate the Black-Scholes formula.

The martingale pricing technique was pioneered by Cox & Ross (1976) and later on further developed and refined by Harrison & Kreps (1979), Harrison & Pliska (1981), and others.

We start by defining the quantity $\eta = (\mu - r)/\sigma$ and the Girsanov factor

$$\xi_t = \exp(-\frac{1}{2}\eta^2 t - \eta W_t^{\mathcal{P}}).$$

The Girsanov factor is a positive $\mathcal{P}$-martingale with mean 1, so we can define a new probability measure $\mathcal{Q}$, equivalent to $\mathcal{P}$, by setting

$$d\mathcal{Q} = \xi_t d\mathcal{P} \tag{7}$$

on $\mathcal{F}_t$. According to the Girsanov Theorem we have that under $\mathcal{Q}$

$$W_t^{\mathcal{Q}} = W_t^{\mathcal{P}} + \eta t$$

is a Brownian motion.[7] Plugging this into (1) yields that under $\mathcal{Q}$, the stock price evolves according to the SDE

$$\frac{dS_t}{S_t} = rdt + \sigma dW_t^{\mathcal{Q}}. \tag{8}$$

Let us now define

$$G_t = B_t \mathrm{E}_t^{\mathcal{Q}}\left[(S_T - K)^+\right] = B_t \mathrm{E}_t^{\mathcal{P}}\left[\frac{\xi_T}{\xi_t}(S_T - K)^+\right].$$

We observe that from the definition $(e^{-rt}G_t)$ must be a $\mathcal{Q}$-martingale with respect to the filtration $(\mathcal{F}_t)$. By the Martingale Representation Theorem we therefore have that

$$d\left[e^{-rt}G_t\right] = \gamma_t dW_t^{\mathcal{Q}}$$

for some process γ.[8] Introducing $\Gamma = e^{rt}\gamma/G$, using the Ito formula and reintroducing the $\mathcal{P}$ Brownian motion we get

$$dG_t = G_t(rdt + \Gamma_t dW_t^{\mathcal{Q}}) = G_t((r + \Gamma_t\eta)dt + \Gamma_t dW_t^{\mathcal{P}}).$$

Now consider a self-financing strategy with value V and no consumption flow before T consisting of a stocks and the remaining amount is put in $b = (V - aS)/B$ bonds. Such a strategy evolves according to

$$\begin{aligned} dV_t &= a_t dS_t + \frac{V_t - a_t S_t}{B_t} dB_t \\ &= (a_t S_t \sigma\eta + rV_t)dt + a_t\sigma S_t dW_t^{\mathcal{P}}. \end{aligned} \tag{9}$$

[7] For the Girsanov Theorem in the context of financial economics see for example Duffie (1992).

[8] A possible reference for the Martingale Representation Theorem is Duffie (1992).

Choosing $V_0 = G_0$ and

$$a = \frac{\Gamma V}{\sigma S}$$

we see that G and V have the same evolution and the same starting value, hence $V_t = G_t$ for all t. Since $C_T = G_T$ we conclude that $C_t = G_t$ for all t. Otherwise there exists an arbitrage because G_t can be replicated through the dynamic trading strategy outlined above. We have now obtained the martingale pricing equation

$$C_t = e^{-r(T-t)} \mathrm{E}_t^{\mathcal{Q}} \left[(S_T - K)^+ \right]. \tag{10}$$

Note that this relation does not depend on μ and/or σ being constants. In fact the measure $\mathcal{Q}$ is unique and the replication argument goes through as long as only the diffusion coefficient, but not necessarily the drift, of the stock is adapted to the filtration generated by the stock.[9] Investors need not have the same beliefs about the drift of the stock for the arbitrage pricing to be valid. All they have to agree about is the diffusion coefficient. In a continuous-time economy this means that all they have to agree about are the zero-sets for the stock price evolution.
If (S_t) happens to be a Markov process under under $\mathcal{Q}$, then (virtually by definition) we deduce from (10) that the option price is a smooth function of current time and stock price only, i.e. $C_t = C(S_t, t)$. But then the Ito formula and 'coefficient matching' recovers the fundamental PDE of Section 3. In other words, there is consistency between the methods. Note that it is sufficient that (S_t) is Markov under $\mathcal{Q}$, so nasty drifts do not prohibit us from using a PDE approach. This is of course what one would conjecture since the drift does not enter the fundamental PDE.
To compute the Black-Scholes formula we use that sitting at time 0, $\ln S_T$ is normal under $\mathcal{Q}$ with mean and variance

$$\begin{aligned} m &= \mathrm{E}^{\mathcal{Q}} [\ln S_T] = \ln S_0 + (r - \frac{1}{2}\sigma^2)T && (11) \\ v^2 &= \mathrm{Var}^{\mathcal{Q}} [\ln S_T] = \sigma^2 T, && (12) \end{aligned}$$

so

$$\begin{aligned} C_0 &= e^{-rT} \int_{\frac{\ln K - m}{v}}^{\infty} (e^{m+vx} - K) \frac{e^{-\frac{1}{2}x^2}}{\sqrt{2\pi}} dx \\ &= S_0 \Phi \left(\frac{\ln(S_0/K) + rT}{\sigma\sqrt{T}} + \frac{1}{2}\sigma\sqrt{T} \right) - e^{-rT} K \Phi \left(\frac{\ln(S_0/K) + rT}{\sigma\sqrt{T}} - \frac{1}{2}\sigma\sqrt{T} \right). \end{aligned}$$

[9] For obscure stock price processes ξ might not be a $\mathcal{P}$-martingale, which implies that $\mathcal{Q}$ as defined by (7) is not an equivalent probability measure. But if $\mathcal{Q}$ is well-defined it is unique.

5 Change of numeraire

This section reviews a popular technique for solving the valuation equation

$$C_0 = e^{-rT}\mathrm{E}^{\mathcal{Q}}\left[(S_T - K)^+\right].$$

The technique is often referred to as the change of numeraire technique because it involves a change of discounting factor from the bank-account to the underlying asset. The technique will prove to be extremely powerful: in this section we will derive the Black-Scholes formula without evaluating a single integral. But more than being a technical tool the change of numeraire approach also exposes an intriguing interpretation of the probabilities in the Black-Scholes formula as will be demonstrated in this section.

The change of numeraire technique showed up in several papers in the late eighties but it was probably known in the financial research community long before.

The idea is the following. We note that the martingale approach in the last section does not depend on the bank-account being used as numeraire for the pay-offs. In fact we could choose S as the numeraire of another martingale measure, $\mathcal{Q}'$, and under this measure C_t/S_t would be a martingale. To see this observe that

$$e^{-rt}\frac{S_t}{S_0} = \exp(-\frac{1}{2}\sigma^2 t + \sigma W_t^{\mathcal{Q}})$$

is a positive $\mathcal{Q}$-martingale with mean 1. Hence, we can define a new equivalent probability measure related to $\mathcal{Q}$ and $\mathcal{P}$ by:

$$d\mathcal{Q}' = e^{-rt}\frac{S_t}{S_0}d\mathcal{Q} = e^{-rt}\frac{S_t}{S_0}\xi_t d\mathcal{P}$$

on $\mathcal{F}_t$. The Brownian motion under $\mathcal{Q}'$ is then given by

$$W_t^{\mathcal{Q}'} = W_t^{\mathcal{Q}} - \sigma t = W_t^{\mathcal{P}} + \eta t - \sigma t.$$

Straightforward application of this yields the valuation equation

$$C_0 = S_0\mathrm{E}^{\mathcal{Q}'}\left[\frac{(S_T - K)^+}{S_T}\right], \tag{13}$$

where

$$\frac{dS_t}{S_t} = (r + \sigma^2)dt + \sigma dW_t^{\mathcal{Q}'}.$$

But this does not reduce the complexity of derivation of the Black-Scholes formula. We will still have to evaluate an integral like the one in the previous section. So instead we reconsider our initial valuation equation. We split up the pay-off to get:

$$\begin{aligned} C_0 &= \mathrm{E}^{\mathcal{Q}}\left[e^{-rT} S_T \mathbf{1}_{\{S_T>K\}}\right] - e^{-rT} K \mathrm{E}^{\mathcal{Q}}\left[\mathbf{1}_{\{S_T>K\}}\right] \\ &= S_0 \mathcal{Q}'(S_T > K) - e^{-rT} K \mathcal{Q}(S_T > K). \end{aligned} \tag{14}$$

We feel that this equation has a very nice interpretation: Given that the European option finishes in-the-money, the option pay-off can be decomposed into two components, the first component is the uncertain amount S_T, and the second component is the fixed amount $-K$. The present value of receiving S_T at time T is of course S_0. But this has to be weighted with some risk-adjusted probability of finishing in-the-money. $\mathcal{Q}'$ is the right measure to use because it exactly off-sets the 'S'-risk. In other words under $\mathcal{Q}'$ pay-offs are valued as if one were 'risk-neutral' with respect to the risk of the underlying stock. The second component $-K$ is a fixed dollar amount. The proper probability measure to apply is therefore the measure $\mathcal{Q}$ under which pay-offs are measured relative to the risk-less bond.

The formula is general, in the sense that it does not depend on the underlying stock following a geometric Brownian motion. In fact, if an equivalent martingale measure with the bank account as numeraire exists (and this measure need not be unique) then one can derive the above formula. In a subsequent section we will show that the European option price of the Cox, Ross & Rubinstein (1979) model has a similar interpretation.

To obtain the Black-Scholes formula we simply have to evaluate the two probabilities in the above equation. We observe that under $\mathcal{Q}'$, $\ln S_T$ is normal with mean and variance given by

$$\mathrm{E}^{\mathcal{Q}'}[\ln S_T] = \ln S_0 + (r + \frac{1}{2}\sigma^2)T \tag{15}$$

$$\mathrm{Var}^{\mathcal{Q}'}[\ln S_T] = \sigma^2 T. \tag{16}$$

Using this and the distribution of $S(T)$ under $\mathcal{Q}$ given in the previous section we immediately obtain

$$\begin{aligned} \mathcal{Q}'(S_T > K) &= \Phi\left(\frac{\ln(S_0/K) + rT}{\sigma\sqrt{T}} + \frac{1}{2}\sigma\sqrt{T}\right) \\ \mathcal{Q}(S_T > K) &= \Phi\left(\frac{\ln(S_0/K) + rT}{\sigma\sqrt{T}} - \frac{1}{2}\sigma\sqrt{T}\right), \end{aligned}$$

and thereby the Black-Scholes formula.

The change of numeraire technique is a very powerful tool that can be applied to other types of option contracts and to more general models. In the fixed income literature this technique has elegantly been applied to option pricing problems under the name of "forward-risk-adjustment", see for example Jamshidian (1989) and El Karoui & Rochet (1989). In the context of exotic options the technique has shown useful in the evaluation of Asian options, lookback options, barrier options, and various other exotica. See for example Ingersoll (1987), Babbs (1992), Dufresne, Kierstad & Ross (1996), and Graversen & Peškir (1995).

6 Shaking your foundation

In this section we will use the so-called *local-time* to derive the Black-Scholes formula. The idea of using local-times in finance is due to Carr & Jarrow (1990). Analysing what is known as the *stop-loss start-gain strategy* they get terms involving local-times. The stop-loss start-gain strategy has also been carefully analysed in the literature by Seidenverg (1988) and Dybvig (1988). The financial insight using this somewhat cumbersome method is the proposal of a trading strategy which is not self-financing and still gives the Black-Scholes formula by taking care of the extra external financing. Moreover one should notice that recently local-times have been used to price American options. See for instance Myneni (1992) and Carr, Jarrow & Myneni (1992). Consider the following trading strategy:
If the present value of the strike price K is below the stock price hold one share of the stock. Finance this by using borrowed funds. If the stock price falls below the present value of the strike price liquidate the position. As we shall see below this strategy will at terminal date T be worth exactly the same as the call-option. Now if the stock price initially is worth less than the present value of the strike price the strategy costs nothing initially. Therefore if the strategy is self-financing this would create arbitrage-opportunities in the Black Scholes-economy. To analyse this strategy we proceed more formally. Let:

$$\begin{aligned} a_t &= \mathbf{1}_{\{S_t > KB_t\}}, \\ b_t &= -\mathbf{1}_{\{S_t > KB_t\}} K, \quad \forall t \in [0,T]. \end{aligned}$$

Then the value of the portfolio at time t, Y_t, is equal to:

$$\begin{aligned} Y_t &= a_t S_t + b_t B_t \\ &= \mathbf{1}_{\{S_t > KB_t\}} S_t - \mathbf{1}_{\{S_t > KB_t\}} KB_t \\ &= (S_t - KB_t)^+ . \end{aligned} \tag{17}$$

Now we see that the value of the portfolio is always the lower bound for a call-option and furthermore since $B_T = 1$ we see that we have duplicated the call's payoff. Therefore if $S_0/B_0 < K$ then the portfolio initially costs nothing. To examine if this portfolio is self-financing we notice that the self-financing condition described with the bond as numeraire is:

$$\frac{Y_t}{B_t} = \frac{Y_0}{B_0} + \int_0^t a_u dF_u, \quad \forall t \in [0, T], \tag{18}$$

where $F_t = S_t/B_t$ is the stock price with the bond as numeraire (the forward price). Inserting a_t and Y_t in (18) we get that the stop-loss and start-gain strategy is self-financing if and only if

$$(F_t - K)^+ = (F_0 - K)^+ + \int_0^t \mathbf{1}_{\{F_u > K\}} dF_u, \quad \forall t \in [0, T]. \tag{19}$$

Fortunately this is not the case - otherwise there could be arbitrage in the economy as described above. To see that the strategy is not self-financing we use the Tanaka-Meyer-formula on Y_t/B_t.[10] It gives us:

$$(F_t - K)^+ = (F_0 - K)^+ + \int_0^t \mathbf{1}_{\{F_u > K\}} dF_u + \Lambda_t(K), \quad \forall t \in [0, T]. \tag{20}$$

We see that the difference between (19) and (20) is the term $\Lambda_t(K)$ which is called the *local time* at K by time t in the stochastic calculus literature. Now we will show that $\Lambda_t(K)$ is positive with positive probability for any t, which shows us that the stop-loss and start-gain strategy is not self-financing.
If we take the "risk-neutral" expectation of (20) we get:

$$\mathrm{E}_0^{\mathcal{Q}}\left[(F_t - K)^+\right] = (F_0 - K)^+ + \mathrm{E}_0^{\mathcal{Q}}\left[\Lambda_t(K)\right], \quad \forall t \in [0, T], \tag{21}$$

where the expectation of the "dF_u" integral is zero because F is a $\mathcal{Q}$-martingale and thereby we have that the integral-term is a $\mathcal{Q}$-martingale. It is obvious from the results in the previous sections that $\mathcal{Q}(F_t > K) > 0, \ \mathcal{Q}(F_t < K) > 0$.
Furthermore: Because $g(x) = (x - K)^+$ is strictly convex over an interval containing K we get that Jensen's inequality holds strictly for $g(x)$. I.e.:

$$\mathrm{E}_0^{\mathcal{Q}}\left[(F_t - K)^+\right] > \left(\mathrm{E}_0^{\mathcal{Q}}\left[F_t\right] - K\right)^+ = (F_0 - K)^+. \tag{22}$$

Combining (21) with (22) we get $\mathrm{E}_0^{\mathcal{Q}}\left[\Lambda_t(K)\right] > 0$. Since $\Lambda_t(K) \geq 0$ it follows that $\mathcal{Q}\left(\Lambda_t(K) > 0\right) > 0, \quad \forall t \in (0, T]$. Therefore $\mathcal{P}\left(\Lambda_t(K) > 0\right) > 0, \quad \forall t \in (0, T]$, which

[10]The Tanaka-Meyer formula can for instance be found in Karatzas & Shreve (1988) p. 220.

shows us that the strategy is not self-financing.

In our setting $\Lambda_t(K)$ has a very nice interpretation. Suppose that we change our strategy in the following way:

Buy one share of stock each time F rises from K to $K+\epsilon$, $\epsilon > 0$. In this case we should also go short in K bonds. We see that every time the transaction takes place it requires an additional ϵ bonds. Furthermore we liquidate the portfolio every time F goes back to K.

Now let $U_t(\epsilon)$ denote the number of times F has risen from K to $K+\epsilon$ until time t. Then we see that with the above mentioned strategy we would have to invest in $\epsilon U_t(\epsilon)$ bonds at time t to handle the external financing. Now it can be shown that:

$$\lim_{\epsilon \downarrow 0} \epsilon U_t(\epsilon) = \Lambda_t(K).$$

That is: The additional local time term from equation (20) can be interpreted as the external financing required to trade by the stop-loss start-gain strategy.

Now we will show that evaluating (21) for $t = T$ yields the Black-Scholes formula. From the previous section we notice that $C_0/B_0 = E_0^Q\left[(F_T - K)^+\right] = E_0^Q\left[(S_T - K)^+\right]$. I.e.:

$$C_0 = (S_0 - e^{-rT}K)^+ + e^{-rT} E_0^Q\left[\Lambda_T(K)\right]. \tag{23}$$

(23) has a nice interpretation: The first term on the right-hand side is the option's intrinsic value and is according to (17) equal to the initial investment required in the stop-loss start-gain strategy. The residual ($e^{-rT} E_0^Q\left[\Lambda_T(K)\right]$) is referred to as the option's time value which in this case is the present value of the expected external financial costs.

From the previous section we know that F is a Q-martingale. By Girsanov's theorem we therefore have that:

$$dF_t = \sigma F_t dW_t^Q.$$

That is:

$$F_t = F_0 \exp\{\sigma W_t^Q - \frac{1}{2}\sigma^2 t\}.$$

Therefore we get the following transition density for F:

$$\psi(F_t, t; F_0, 0) = \frac{1}{F_t \sigma \sqrt{t}} \phi\left(\frac{\ln\left(\frac{F_0}{F_t}\right) - \frac{1}{2}\sigma^2 t}{\sigma\sqrt{t}}\right), \tag{24}$$

where $\phi(z) \equiv \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right)$ is the standard normal density function.

Using a theorem about local times yields:[11]

$$\mathrm{E}_0^{\mathcal{Q}}\left[\int_0^T k(F_s)d < F >_s\right] = \mathrm{E}_0^{\mathcal{Q}}\left[2\int_{-\infty}^{\infty} k(x)\Lambda_T(x)dx\right], \tag{25}$$

where k is a Borel-measurable function. Using that $d < F >_t = \sigma^2 F^2(t)dt$ on the left side and employing Fubini‘s theorem on the right side of (25) gives us:

$$\int_{-\infty}^{\infty} k(x)\int_0^T \sigma^2 x^2 \psi(x,t;F_0,0)dt\ dx = \int_{-\infty}^{\infty} k(x)2\mathrm{E}_0^{\mathcal{Q}}\left[\Lambda_T(x)\right]dx. \tag{26}$$

Now choose $k(x) = \mathbf{1}_{\{x\in A\}}$, where $A \in \mathcal{F}$. (26) then becomes:

$$\int_A\int_0^T \sigma^2 x^2 \psi(x,t;F_0,0)dt\ dx = \int_A 2\mathrm{E}_0^{\mathcal{Q}}\left[\Lambda_T(x)\right]dx.$$

Realizing that the integrands are nonnegative and that both integrals are equal for any $A \in \mathcal{F}$ we get:

$$\int_0^T \sigma^2 x^2 \psi(x,t;F_0,0)dt = 2\mathrm{E}_0^{\mathcal{Q}}\left[\Lambda_T(x)\right]. \tag{27}$$

Combining (24) with (27) yields:

$$\mathrm{E}_0^{\mathcal{Q}}\left[\Lambda_T(K)\right] = \frac{\sigma K}{2}\int_0^T \frac{1}{\sqrt{t}}\phi\left(\frac{\ln\left(\frac{F_0}{K}\right) - \frac{1}{2}\sigma^2 t}{\sigma\sqrt{t}}\right)dt. \tag{28}$$

If we substitute (28) back into (23) we get:

$$C_0 = \left(S_0 - e^{-rT}K\right)^+ + e^{-rT}\frac{\sigma K}{2}\int_0^T \frac{1}{\sqrt{t}}\phi\left(\frac{\ln\left(\frac{S_0}{Ke^{-rT}}\right) - \frac{1}{2}\sigma^2 t}{\sigma\sqrt{t}}\right)dt.$$

Now changing variable by $\nu \equiv \frac{\sqrt{t}\sigma}{\sqrt{T}}$ gives us:

$$C_0 = \left(S_0 - e^{-rT}K\right)^+ + e^{-rT}K\sqrt{T}\int_0^\sigma \phi\left(\frac{\ln\left(\frac{S_0}{Ke^{-rT}}\right) - \frac{1}{2}\nu^2 T}{\nu\sqrt{T}}\right)d\nu.$$

Finally we have reached the Black-Scholes formula. This is seen by noticing that (2) differentiated with respect to σ is:

$$e^{-rT}K\sqrt{T}\phi\left(\frac{\ln\left(\frac{S_0}{Ke^{-rT}}\right) - \frac{1}{2}\sigma^2 T}{\sigma\sqrt{T}}\right), \tag{29}$$

with the boundary condition that $C_0 = (S_0 - e^{-rT}K)^+$ when $\sigma = 0$.

[11] The theorem can for instance be found in Karatzas & Shreve (1988) p.218.

7 The forward equation of European option prices

Compared to the technique applied in Section 5 the following derivation of the Black-Scholes formula might seem cumbersome. On the other hand the result that we will derive will shed further light on the European option pricing problem and will expose an interesting duality of the pricing problem that we consider. In the spirit of Dupire (1993) we derive a forward partial differential equation for the European option prices. In this equation the variables are the strike and the maturity whereas the current spot and time are fixed. This is opposed to the standard backward partial differential equation derived in Section 3, where the spot and time are the variables and strike and maturity are fixed. Examining this forward equation reveals that the option pricing problem can be solved in a dual economy where every parameter is turned upside down: the strike price is the underlying, the option is a put with strike equal to the current spot, time is reversed, etc.

Again we start from the valuation equation

$$C_0 = e^{-rT}\mathrm{E}^{\mathcal{Q}}\left[(S_T - K)^+\right] = e^{-rT}\int_K^\infty (x-K)\psi(x,T)dx, \tag{30}$$

where $\psi(x,T)$ is the $\mathcal{Q}$-density of S_T in the point x given S_0 at time 0.

Due to the Markov property of the spot price we have that ψ solves the forward Fokker-Planck equation[12]

$$0 = -\frac{\partial \psi}{\partial T} - \frac{\partial}{\partial x}\left[rx\psi\right] + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left[\sigma^2 x^2 \psi\right]$$

subject to the initial boundary condition $\psi(x,0) = \delta(x - S_0)$, where $\delta(\cdot)$ is the Dirac Delta function.[13] We will use this to derive a forward equation for the option prices. Assuming that

$$rx\psi(x,T) \to 0, \quad \sigma^2 x^2 \psi(x,T) \to 0, \quad \frac{\partial}{\partial x}\left[\sigma^2 x^2 \psi(x,T)\right] \to 0,$$

for $x \to \infty$, which is clearly satisfied in the Black-Scholes model, integration of the forward equation over the interval (y,∞) yields:

$$0 = -\frac{\partial}{\partial T}\int_y^\infty \psi(x,T)dx + ry\psi(y,T) - \frac{1}{2}\frac{\partial}{\partial y}\left[\sigma^2 y^2 \psi(y,T)\right].$$

Integrating once more, this time over (K,∞), yields

$$0 = -\frac{\partial}{\partial T}\int_K^\infty \int_y^\infty \psi(x,T)dx\ dy + r\int_K^\infty y\psi(y,T)dy + \frac{1}{2}\sigma^2 K^2 \psi(K,T)$$

[12]See Revuz & Yor (1991) p 269 for the Fokker-Planck equation.

[13]The Dirac Delta function is defined by $\delta(x) = 0$, for all $x \neq 0$, and $\int_{-\epsilon}^{\epsilon} \delta(x)dx = 1$, for all $\epsilon > 0$.

Now we go back to the pricing equation. Integrating by parts we get that

$$C_0 = e^{-rT} \int_K^\infty \int_y^\infty \psi(x, T) dx\ dy,$$

so

$$\frac{\partial}{\partial T} \int_K^\infty \int_y^\infty \psi(x, T) dx\ dy = re^{rT} C_0 + e^{rT} \frac{\partial C_0}{\partial T}.$$

Further we have that

$$\begin{aligned} \int_K^\infty y\psi(y, T) dy &= e^{rT} C_0 - K e^{rT} \frac{\partial C_0}{\partial K} \\ \psi(K, T) &= e^{rT} \frac{\partial^2 C_0}{\partial K^2}. \end{aligned} \tag{31}$$

If we let $C(K, T)$ denote the initial price of a European call-option with strike K expiring at time T, we obtain the following forward partial differential equation for the European call-option prices

$$0 = -\frac{\partial C}{\partial T} - rK \frac{\partial C}{\partial K} + \frac{1}{2} \sigma^2 K^2 \frac{\partial^2 C}{\partial K^2} \tag{32}$$

subject to the initial boundary condition $C(K, 0) = (S_0 - K)^+$.
The forward equation can now be solved to yield the Black-Scholes formula.
The advanced reader might observe that the forward equation could be derived from the valuation equation under the $\mathcal{Q}'$ measure, (13), combined with the time homogeneity of the stock price process in the Black-Scholes model. Under the assumption of a positive dividend yield of the underlying stock, Andreasen & Gruenewald (1996) apply this technique to obtain a forward equation for American call-options in the Black-Scholes model as well as in the jump-diffusion model of Merton (1976).[14] But the forward equation (32) is more general; under assumption of sufficient regularity it holds for all Ito processes of the type

$$\frac{dS_t}{S_t} = \mu(t; \omega) dt + \sigma(S_t, t) dW_t^{\mathcal{P}}$$

if additionally the interest rate is only a function of time and the stock price. For stock option pricing and short maturities it is in most cases reasonable to assume (at most) time-dependent interest rates. Given todays yield curve it is then possible to uniquely determine the function $\sigma(S, t)$ from a full double continuum of option prices, $C(K, T)$, by the forward equation (32). The trick is simply to estimate the

[14] The positive dividend yield implies that the American call-option might be exercised prematurely.

derivatives in (32) and isolate the function $\sigma(\cdot,\cdot)$. In other words from a full set of marketed options we can 'infer the option pricing model of the market'. This was first observed by Dupire (1993), but already Breeden & Litzenberger (1978) noted the relation (31). This relation tells us that we can infer the stock's risk-adjusted distribution at a given maturity date from a continuum of option prices of different strikes.[15]

Another interesting implication of the forward equation is that when the volatility coefficient (now possibly a function of time and spot) is given we can price all options on the market by only solving one partial differential equation numerically. Andreasen (1996) observes that this also goes for the hedge ratios of the European options. To see this define

$$\Delta(K,T) = \frac{\partial C(K,T)}{\partial S}|_{S=S_0}. \tag{33}$$

Differentiation of the forward equation (32) now yields

$$0 = -\frac{\partial \Delta}{\partial T} - rK\frac{\partial \Delta}{\partial K} + \frac{1}{2}\sigma^2 K^2 \frac{\partial^2 \Delta}{\partial K^2}$$

subject to the initial boundary condition $\Delta(K,0) = \mathbf{1}_{\{S_0 \geq K\}}$.

Similar forward equations might be derived for other 'Greeks', i.e partial derivatives of the option price w.r.t. other parameters.

The equation can also be used to show how to hedge exotic options statically. In Carr, Ellis & Gupta (1998) a model with zero interest and a symmetry condition on $\sigma(\cdot,\cdot)$, which is clearly satisfied in the Black-Scholes setting, is developed. It is then shown how to hedge for instance down-and-out calls statically by 'standard' options.

The last point to be stated is the duality of the option pricing problem implied by the forward equation. Suppose that the time axis is reversed, S_0 is a fixed quantity, and that we are sitting at time T evaluating

$$\mathrm{E}^{\mathcal{R}}\left[(S_0 - K_0)^+ | K_T = K\right] \tag{34}$$

for the process

$$\frac{dK_t}{K_t} = (-r)d(-t) + \sigma dW_t^{\mathcal{R}}$$

where $W^{\mathcal{R}}$ is some backward running Brownian motion under some probability measure $\mathcal{R}$. Then the forward equation (32) is the backward equation resulting for this problem. So we conclude that the option pricing problem might be solved in a dual economy where time is reversed, the strike is the underlying that pays a proportional

[15]For further exploration of this see for example Shimko (1991), Derman & Kani (1994), Rubinstein (1994), and Jackwerth & Rubinstein (1996).

dividend of r, the option is a put on the initial stock price, and finally the interest rate is equal to zero. Further we see that in this "space" the hedge ratio of the original economy will be a digital option. The option pricing can therefore be performed in a reversed binomial tree.[16]

Note that under $\mathcal{R}$ we have

$$K_0 = K \exp\left((-r - \frac{1}{2}\sigma^2)T + \sigma(W_0^{\mathcal{R}} - W_T^{\mathcal{R}})\right).$$

Using this in (34) gives the Black-Scholes formula by essentially the same calculations as those at the end of Section 4.

Using (33) we get the following expression for the hedge ratio

$$\mathrm{E}^{\mathcal{R}}\left[\mathbf{1}_{\{K_0 \leq S_0\}} | K_T = K\right] = \Phi\left(\frac{\ln(S_0/K) + rT}{\sigma\sqrt{T}} + \frac{1}{2}\sigma\sqrt{T}\right).$$

8 A convergence proof

Cox et al. (1979) were the first to publish a paper with a formal convergence proof along the lines of this section. A much less known paper with the same result (and from the same year) is by Rendleman & Bartter (1979). But the use of binomial models for economic reasoning is much older, dating (at least) back to Arrow and Debreu in the 50'ies.

Let us consider the following situation: A stock today has a price of S_0 and can in the next period either go up to uS_0 or down to dS_0. This happens with probabilities p and $1-p$, respectively. In the economy there further exists a risk-free zero coupon bond maturing in the next period with (discretely compounded) interest rate r_d ($u > 1 + r_d > d > 0$, to avoid dominance), and a call-option on the stock with exercise price K. The situation is illustrated in Figure 1.

We are interested in hedging the option by trading a shares of stock and b bonds. A perfect hedge, i.e. an exact replication of the option's pay-off in every possible future state, is achieved by letting

$$a = \frac{C_u - C_d}{(u-d)S_0} \qquad b = \frac{uC_d - dC_u}{(u-d)}$$

(notice that $a \approx \frac{\partial C}{\partial S}$, so the analogy to the continuous case is striking).

To prevent arbitrage opportunities the price of the hedge portfolio must be equal to

[16] For more on the duality see Dupire (1994) and Andreasen (1996).

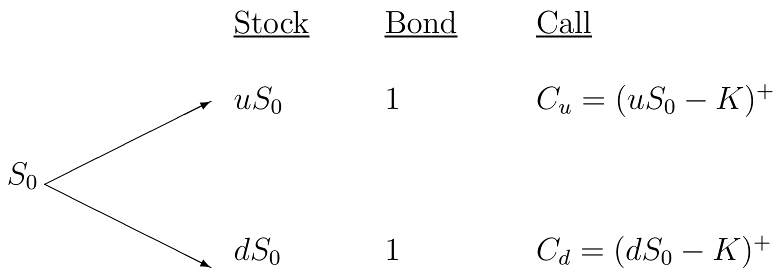

Figure 1: The One-Period Binomial Model

the present price of the call-option. Writing this out and reshuffling leads to

$$C_0 = R^{-1}(qC_u + (1-q)C_d) \tag{35}$$

where $q = (R-d)/(u-d)$ and $R = 1 + r_d$. From this we see that the price of the call-option is the discounted expected future value where the expectation is under a measure that gives probability q of an 'up-jump'. Notice that the original probabilities do not enter the expression. This hedge argument is the key in Cox et al. (1979). Notice that we can write (35) as

$$\frac{C_0}{B_0} = \mathrm{E}_0^{\mathcal{Q}}\left(\frac{C_T}{B_T}\right), \quad B_0 = R^{-T}$$

with obvious subscript notation and $\mathcal{Q}$ denoting the measure naturally induced by q. So: Using the bond as numeraire, the call price is a $\mathcal{Q}$-martingale. In other words the notation is consistent with that of Section 5. At this point let us make an observation. If we let $q' = (uq)/(1 + r_d)$ then a direct inspection reveals that

$$\frac{C_0}{S_0} = \mathrm{E}_0^{\mathcal{Q}'}\left(\frac{C_T}{S_T}\right)$$

with $\mathcal{Q}'$ being the measure induced by q'. Again the notation is consistent: Using the stock as numeraire, the call price is a $\mathcal{Q}'$-martingale.

The argument is easily extended to a setting with n independent multiplicative binomial movements per unit of time ensuring us that the martingale pricing techniques of Section 4 carry over in a discrete setting. Using the arguments from Section 5 we can thus still write out the call price as

$$C_0 = S_0\mathcal{Q}'(S_T^{(n)} > K) - KB_0\mathcal{Q}(S_T^{(n)} > K) \tag{36}$$

where

$$S_T^{(n)} = S_0 u^j d^{Tn-j} \;, \qquad B_0 = R^{-Tn} \;,$$
$$j \overset{\mathcal{Q}}{\approx} \mathrm{bi}(Tn, q) \;, \qquad j \overset{\mathcal{Q}'}{\approx} \mathrm{bi}(Tn, q'),$$

and 'bi' denotes the binomial distribution. Again we claim that the call price at any time can only be a function of current stock price and time. This claim is then justified by our ability to exactly replicate the final pay-off which only depends on S_T by a dynamic trading strategy in the stock and the bond.
For computational purposes (36) is often rewritten as

$$C_0 = S_0 \tilde{\Phi}(m; Tn, q') - K B_0 \tilde{\Phi}(m; Tn, q) \tag{37}$$

with m being the smallest non-negative integer greater than $\ln(K/(S_0 d^{Tn}))/\ln(u/d)$ and $\tilde{\Phi}(m; Tn, q')$ denoting the complementary binomial distribution function.
Now let anything with an 'n' on it refer to a binomial model with n moves per time unit. Our aim is to show that as n approaches infinity the call price in the n-model converges to that of the Black-Scholes model. Because of the decompositions and (14) and (36) and the distribution results (11)-(12) and (15)-(16) our main task is to choose the parameters of the binomial model such that

$$\ln S_T^{(n)} \overset{\mathcal{Q}}{\to} N(\ln S_0 + (r - \frac{\sigma^2}{2})T, \sigma^2 T) \tag{38}$$
$$\ln S_T^{(n)} \overset{\mathcal{Q}'}{\to} N(\ln S_0 + (r + \frac{\sigma^2}{2})T, \sigma^2 T). \tag{39}$$

Regarding interest rates we don't have much choice but to let $R_n = e^{r/n}$. This means that the key parameters we have to choose are the sizes of the up and down moves, u_n and d_n. A good choice is

$$\ln u_n = \frac{\sigma}{\sqrt{n}} \tag{40}$$
$$\ln d_n = -\frac{\sigma}{\sqrt{n}}. \tag{41}$$

With M_n and V_n denoting mean and variance of $\ln S_T^{(n)}$ we then have

$$\begin{aligned} M_n^{\mathcal{Q}} &= \ln S_0 + Tn(q_n \ln u_n + (1 - q_n) \ln d_n) \\ V_n^{\mathcal{Q}} &= Tnq_n(1 - q_n)(\ln u_n - \ln d_n)^2, \end{aligned}$$

and likewise for $\mathcal{Q}'$. Remembering that $q_n = \frac{e^{r/n} - d_n}{u_n - d_n}$ allows us to rewrite M_n and V_n by Taylor expanding the exponential function to the second order. This reveals that

$$\begin{aligned} M_n^{\mathcal{Q}} &\to \ln S_0 + (r - \frac{\sigma^2}{2})T \\ V_n^{\mathcal{Q}} &\to \sigma^2 T, \end{aligned}$$

and by similar calculations we get convergence of $\mathcal{Q}'$-moments. So the first and second moments converge, under the respective measures, and the jumps vanish in the limit. This allows us to invoke (basically) a Lindeberg-Feller version of the Central Limit Theorem (see e.g. Duffie (1992)) to confirm the validity of (38) and (39). Finally dominated convergence ensures that the elements of the binomial decomposition (36) converge to their continuous counterparts, which establishes the desired result.
We have not used the original probabilities for anything (except that we have implicitly assumed them to be non-zero). It is easy to see that we can add any term of order n or higher in (40) and (41) and still have the same $\mathcal{Q}$ (and $\mathcal{Q}'$) convergence results. This could, if we were such inclined, help us establish convergence of the underlying process.
The model described in this section illustrates the fundamentals of pricing by no arbitrage using only linear algebra. Therefore it is ideal for teaching purposes. It is not the most advanced model in the paper, but it is 'solid as a rock.' It is also very handy when we want initial price estimates for exotic derivatives in cases where it is unclear how more advanced methods work, if indeed they do.

9 The continuous-time CAPM

This section shows that one might also obtain the Black-Scholes formula in the continuous-time capital asset pricing model by Merton (1971). The derivation is basically taken from Ingersoll (1987) but a similar derivation appears in Cox & Rubinstein (1985). Suppose that the market in total contains N (non-dividend paying) risky assets that evolve according to the N-dimensional stochastic differential equation

$$dS_t = I_{S_t}(\mu dt + \Sigma dW_t^{\mathcal{P}}),$$

where I_S is the diagonal matrix with diagonal elements $(S_1, \ldots, S_N)$, μ is an N-dimensional constant vector, Σ is a constant $N \times N$ matrix, and $W^{\mathcal{P}}$ is an N-dimensional Brownian motion under $\mathcal{P}$. For simplicity we will suppose that Σ has full rank. Suppose that there additionally exists a risk free asset paying a constant continuously compounded interest rate r. Consider an investor that maximises expected additive utility on some time horizon $[0, \tau]$,

$$\mathrm{E}^{\mathcal{P}}\left[\int_0^\tau u(x_t, t)dt\right],$$

over consumption flow, x, and risky portfolio holdings vector, a, subject to the self-financing constraint or dynamic budget constraint (where $'$ denotes transposition)

$$\begin{aligned} dV_t &= a_t' dS_t + (V_t - a_t' S_t) r dt - x_t dt \\ &= (V_t \theta_t' (\mu - r\mathbf{1}) + rV_t - x_t) dt + V_t \theta_t' \Sigma dW_t^{\mathcal{P}} \end{aligned}$$

where V_t is the current wealth and θ is the N dimensional vector with elements $\theta_i = a_i S_i / V$. θ_i is the fraction of the investor's wealth invested in risky asset i. Defining the indirect utility as

$$J(V_t, t) = \max_{(\theta_s, x_s)_{s \geq t}} \mathrm{E}_t^{\mathcal{P}} \left[\int_t^{\mathcal{T}} u(x_s, s) ds \right],$$

we get the Bellman-Hamilton equation[17]

$$0 = \max_{\theta, x} u + \frac{\partial J}{\partial t} + (V\theta'(\mu - r\mathbf{1}) + rV - x)\frac{\partial J}{\partial V} + \frac{1}{2} V^2 \theta' \Sigma \Sigma' \theta \frac{\partial^2 J}{\partial V^2}.$$

The first order conditions imply that in optimum

$$\theta = -\frac{\partial J / \partial V}{V \partial^2 J / \partial V^2} (\Sigma \Sigma')^{-1} (\mu - r\mathbf{1}).$$

Observe that for all i, j the ratio θ_i / θ_j is independent of wealth and utility. So if all investors have additive separable utility they will all hold the same portfolio of risky assets. This means that the market portfolio of risky assets will be given by

$$\theta_M = k(\Sigma \Sigma')^{-1} (\mu - r\mathbf{1})$$

for some one-dimensional process k. The expected instantaneous excess return of the (risky) market portfolio is therefore

$$\mu_M - r = k(\mu - r\mathbf{1})'(\Sigma \Sigma')^{-1} (\mu - r\mathbf{1}),$$

and the local variance of the market return is

$$v_M^2 = k^2 (\mu - r\mathbf{1})'(\Sigma \Sigma')^{-1} (\mu - r\mathbf{1}).$$

The vector of local covariances between the market portfolio and instantaneous return of the assets is given by

$$c = k(\mu - r\mathbf{1}).$$

[17] For an intuitive proof of the Bellman-Hamilton equation see for example Ingersoll (1987).

Combining these equations we get

$$\mu_i = r + \frac{c_i}{v_M^2}(\mu_M - r).$$

Now suppose an option contract on S_i is introduced on the market in zero net supply. Since the market is dynamically complete the market equilibrium is not changed and the above expected return relation is still valid. If the option price is only a function of the underlying stock and time, the Ito formula implies that the local covariance of the return of the option with the market return can be written as

$$\frac{1}{C}\frac{\partial C}{\partial S}c_i$$

Therefore the expected instantaneous return of the option contract is given by

$$r + \frac{1}{C}\frac{\partial C}{\partial S}\frac{c_i}{v_M^2}(\mu_M - r).$$

Using the Ito formula on the option price, $C(S_i(t), t)$, yields that the instantaneous return of the option contract is given by

$$\frac{1}{C}\left[\frac{\partial C}{\partial t} + \mu_i S_i \frac{\partial C}{\partial S_i} + \frac{1}{2}\|\Sigma_i\|^2 S_i^2 \frac{\partial^2 C}{\partial S_i^2}\right].$$

Equating this to the return of the option and inserting the expected return of the underlying stock yields the partial differential equation

$$rC = \frac{\partial C}{\partial t} + rS\frac{\partial C}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 C}{\partial S^2},$$

where we have omitted the subscript on the stock and introduced the notation $\sigma^2 = \|\Sigma_i\|^2$. We have thereby derived the Black-Scholes partial differential equation in the context of the continuous-time CAPM. The formula for the European call can be calculated as in Section 3. This approach relies on the assumption that the option price is a function of time and current stock only. As shown in Section 3 this assumption can be justified by the Black-Scholes hedging argument that uniquely fixes the option price given no arbitrage possibilities. But here our argumentation is not based on a hedging argument but rather a risk-return relation. Therefore the above derivation shows the consistency of the Black-Scholes formula with the CAPM pricing relation.

From the last equations it is tempting to conclude that the Black-Scholes formula or a preference-free pricing formula can be derived in the context of any linear factor model of expected asset return like the continuous-time CAPM. This is only true though if the market additionally is dynamically complete or effectively complete. In general, if a new asset is introduced in an incomplete economy, a new equilibrium will be the outcome and the prices of the existing assets will change. As mentioned, an exception to this is when an incomplete market is effectively complete. An example of this situation is the consumption based capital asset pricing model described in Merton (1973*a*) where the state variables determining the investment opportunity set are spanned by the marketed assets.[18]

10 A representative investor approach

In this section we show that the Black-Scholes formula can be derived in a model where the continuous-trade assumption is replaced by the assumption of a representative investor with power utility. The approach was introduced by Rubinstein (1976).
First, let us consider a one-period model where trading can be performed at the times $0, T$. Suppose an agent maximises expected utility of terminal consumption

$$\mathrm{E}^{\mathcal{P}}\left[u(x_T)\right]$$

subject to the budget constraint

$$\begin{aligned} x_T &= a' S_T \\ V_0 &= a' S_0, \end{aligned}$$

where a is the vector of portfolio holdings, S is the vector of prices of the marketed assets, and V_0 is the initial wealth. Forming the Lagrangian yields the first order condition

$$S_0 = \lambda^{-1} \mathrm{E}^{\mathcal{P}}\left[u'(x_T) S_T\right],$$

where the prime denotes the first derivative and λ is the Lagrange multiplier of the budget constraint. Specifically we get for the risk-free asset

$$e^{-rT} = \lambda^{-1} \mathrm{E}^{\mathcal{P}}\left[u'(x_T)\right].$$

Combining these equations we get the valuation equation

$$S_0 = e^{-rT} \mathrm{E}^{\mathcal{P}}\left[\frac{u'(x_T)}{\mathrm{E}^{\mathcal{P}}\left[u'(x_T)\right]} S_T\right].$$

[18]See also Christensen, Graversen & Miltersen (1996).

With these preliminaries let us now assume that the market has a representative investor with power utility function

$$u(x) = \frac{x^{1+\gamma}}{1+\gamma},$$

with $\gamma < 0$, i.e. $-\gamma$ is the constant relative risk aversion of the representative investor. Now we redefine the notation; let S be the price of one particular stock, and V_0 be initial aggregate wealth.

Assume that aggregate consumption at time T and the time T stock price are jointly log-normally distributed, so that we can write

$$\begin{aligned} S_T &= S_0 e^{(\mu - \frac{1}{2}\sigma^2)T + \sigma W_T^{\mathcal{P}}} \\ x_T &= V_0 e^{(\mu_x - \frac{1}{2}\sigma_x^2)T + \sigma_x W_{x,T}^{\mathcal{P}}}, \end{aligned}$$

where $W^{\mathcal{P}}, W_x^{\mathcal{P}}$ are $\mathcal{P}$-Brownian motions with constant correlation ρ.

Note that

$$\frac{u'(x_T)}{\mathrm{E}^{\mathcal{P}}\left[u'(x_T)\right]} = e^{-\frac{1}{2}\gamma^2\sigma_x^2 T + \gamma\sigma_x W_{x,T}^{\mathcal{P}}}.$$

For the market to be in equilibrium we must have that the valuation equation holds for the stock. Inserting the above in the valuation equation yields

$$\begin{aligned} S_0 &= S_0 e^{-rT}\mathrm{E}^{\mathcal{P}}\left[e^{(\mu - \frac{1}{2}(\sigma^2 + \gamma^2\sigma_x^2))T + \sigma W_T^{\mathcal{P}} + \gamma\sigma W_{x,T}^{\mathcal{P}}}\right] \\ &= S_0 e^{(\mu + \sigma\gamma\sigma_x\rho - r)T}, \end{aligned}$$

so

$$\mu = r - \gamma\rho\sigma\sigma_x. \tag{42}$$

Now we want to evaluate a call-option on S_T with strike K. Using the valuation equation and the above derivations we get:

$$\begin{aligned} C_0 &= e^{-rT}\mathrm{E}^{\mathcal{P}}\left[e^{-\frac{1}{2}\gamma^2\sigma_x^2 T + \gamma\sigma_x W_{x,T}^{\mathcal{P}}}(S_T - K)^+\right] \\ &= S_0\mathrm{E}^{\mathcal{P}}\left[e^{-\frac{1}{2}(\sigma^2 + 2\gamma\rho\sigma\sigma_x + \gamma^2\sigma_x^2)T + \sigma W_T^{\mathcal{P}} + \gamma\sigma_x W_{x,T}^{\mathcal{P}}}\mathbf{1}_{\{S_T \geq K\}}\right] \\ &\quad - K e^{-rT}\mathrm{E}^{\mathcal{P}}\left[e^{-\frac{1}{2}\gamma^2\sigma_x^2 T + \gamma\sigma_x W_{x,T}^{\mathcal{P}}}\mathbf{1}_{\{S_T \geq K\}}\right] \end{aligned}$$

By introducing the joint density of $(W^{\mathcal{P}}, W_x^{\mathcal{P}})_T$ we could calculate the expectations to give us the Black-Scholes formula. But it is much easier to make use of the change of measure induced by the Girsanov factors under the expectations.

Define two new equivalent probability measure $\mathcal{Q}'$ and $\mathcal{Q}$ by the Radon-Nikodym derivatives

$$\begin{aligned} \frac{d\mathcal{Q}'}{d\mathcal{P}} &= e^{-\frac{1}{2}(\sigma^2 + 2\gamma\rho\sigma\sigma_x + \gamma^2\sigma_x^2)T + \sigma W_T^{\mathcal{P}} + \gamma\sigma_x W_{x,T}^{\mathcal{P}}} \\ \frac{d\mathcal{Q}}{d\mathcal{P}} &= e^{-\frac{1}{2}\gamma^2\sigma_x^2 T + \gamma\sigma_x W_{x,T}^{\mathcal{P}}}. \end{aligned}$$

Using these probability measures we can write

$$C_0 = S_0 \mathcal{Q}'(S_T > K) - e^{-rT} K \mathcal{Q}(S_T > K).$$

The Girsanov Theorem together with relation (42) imply that

$$\begin{aligned} S_T &= S_0 e^{rT + \frac{1}{2}\sigma^2 T + \sigma W_T^{\mathcal{Q}'}} \\ &= S_0 e^{rT - \frac{1}{2}\sigma^2 T + \sigma W_T^{\mathcal{Q}}}, \end{aligned}$$

where $W^{\mathcal{Q}'}, W^{\mathcal{Q}}$ are some standard normal Brownian motions under the two respective probability measures.[19] Using this we immediately obtain the Black-Scholes formula. In this section the assumption of continuous trade was replaced by the assumption of existence of a representative investor. Unless investors have identical or very similar preferences a representative investor is in general not guaranteed to exist in an incomplete market like the one analysed. Even if a representative investor exists, the market equilibrium, prices of existing assets, and the representative preferences might change when a new asset (in this case the option) is introduced on an incomplete market. Despite these drawbacks this approach is widely used in models of incomplete markets.

11 Discussion

Economics has been described as the only field where people can win Nobel Prizes for saying the exact opposite things. This paper has shown that in the subset of economics known as 'financial mathematics' there is a very high degree of consistency between models and approaches.

We did this by showing that as special cases they could produce the Black-Scholes formula which, despite its widespread recognition and use, certainly is no trivial result. Some of the methods seemed different (compare the PDEs of Section 3 to the SDEs of Section 4) - but were in fact very similar. Some approaches seemed at a first glance to offer little extra (comparing Section 4 to Sections 5 and 7 it is unclear what could possibly be the benefit of 'counting in units of the stock' or 'letting time run backwards' in a time-homogeneous model) - but they did. One model was very intuitive (Section 8) - one was very much the opposite (Section 6). Two models (Sections 9 and 10) build on the long-honoured economic concept of utility

[19] Notice: The correlation between the two coordinates is ρ no matter which of the two measures ($\mathcal{Q}$ or $\mathcal{Q}'$) we use.

maximisation, and produced the Black-Scholes formula as a special case when utility functions and/or distributions of returns were restricted.
Still, after eight proofs of the Black-Scholes formula skeptics could ask if the finance community has not progressed beyond that result. We believe that it has, and that the criticism is unreasonable. We have illustrated that by, in each of the Sections 3 through 10 outlining the applicability of the particular model or approach to more general cases than the one originally considered by Black, Scholes and Merton. This included both more advanced dynamics of fundamentals and contractually more sophisticated derivatives. Hence we hope to have convinced the reader that each method has validity beyond the basic setting, which in turn justifies the research done in the past, as well as the research that will continue for a long time. If the reader was already aware of this, we hope to have provided 'a couple of cheap thrills.'

References

Andreasen, J. (1996), Implied modelling: Stable implementation, hedging, and duality. Working Paper. Aarhus University.

Andreasen, J. & Gruenewald, B. (1996), American option pricing in the jump-diffusion model. Working Paper. Aarhus University.

Babbs, S. (1992), Binomial valuation of lookback options. Working Paper, Midland Mortagu Capital Markets, London.

Black, F. & Scholes, M. (1973), 'The pricing of options and corporate liabilities', *Journal of Political Economy* **81**, 637–654.

Breeden, D. & Litzenberger, R. (1978), 'Prices of state-contingent claims implicit in option prices', *Journal of Business* **51**, 621–651.

Carr, P., Ellis, K. & Gupta, V. (1998), 'Static hedging of exotic options', *Journal of Finance* **53**(3), 1165–1190.

Carr, P. & Jarrow, R. (1990), 'The stop-loss start-gain paradox and option valuation: A new decomposition into intrinsic and time value', *The Review of Financial Studies* **3**(3), 469–492.

Carr, P., Jarrow, R. & Myneni, R. (1992), 'Alternative characterizations of american put options', *Mathematical Finance* **2**(2), 87–106.

Christensen, P., Graversen, S. & Miltersen, K. (1996), Dynamic spanning in the consumption-based capital asset pricing model. Unpublished. Odense University.

Cox, J. & Ross, S. (1976), 'The valuation of options for alternative stochastic processes', *Journal of Financial Economics* **3**, 145–166.

Cox, J., Ross, S. & Rubinstein, M. (1979), 'Option pricing: A simplified approach', *Journal of Financial Economics* **7**, 229–263.

Cox, J. & Rubinstein, M. (1985), *Options Markets*, Prentice-Hall.

Derman, E. & Kani, I. (1994), 'Riding a smile', *RISK* **7**, 32–39.

Duffie, D. (1992), *Dynamic Asset Pricing Theory*, Princeton University Press.

Dufresne, P., Kierstad, W. & Ross, M. (1996), Pricing the martingale way. Working Paper. Berkeley.

Dupire, B. (1993), Pricing and hedging with smiles. Working Paper. Paribas Capital Markets. London.

Dupire, B. (1994), 'Pricing with a smile', *RISK* **7**, 18–20.

Dybvig, P. (1988), 'Inefficient dynamic portfolio strategies or how to throw away a million dollars in the stock market', *The Review of Financial Studies* pp. 67–88.

El Karoui, N. & Rochet, J.-C. (1989), A pricing formula for options on coupon bonds. Unpublished, University of Paris.

Graversen, S. E. & Peškir, G. (1995), On the russian option: An alternate derivation and expected waiting time. Unpublished. Aarhus University.

Harrison, J. M. & Kreps, D. M. (1979), 'Martingales and arbitrage in multiperiod securities markets', *Journal of Economic Theory* **20**, 381–408.

Harrison, J. M. & Pliska, S. R. (1981), 'Martingales and stochastic integrals in the theory of continuous trading', *Stochastic Processes and their Applications* **11**, 215–260.

Ingersoll, Jr., J. (1987), *Theory of Financial Decision Making*, Rowman & Littlefield.

Jackwerth, J. & Rubinstein, M. (1996), 'Recovering probability distributions from option prices', *Journal of Finance* **51**(5), 1611–1631.

Jamshidian, F. (1989), 'An exact bond option formula', *Journal of Finance* **44**, 205–209.

Karatzas, I. & Shreve, S. (1988), *Brownian Motion and Stochastic Calculus*, Springer-Verlag.

Merton, R. (1971), 'Optimum consumption and portfolio rules in a continuous time model', *Journal of Economic Theory* **3**, 373–413.

Merton, R. (1973*a*), 'An intertemporal capital asset pricing model', *Econometrica* **41**, 867–888.

Merton, R. (1973*b*), 'The theory of rational option pricing', *Bell Journal of Economics and Management Science* **4**, 141–183.

Merton, R. (1976), 'Option pricing when the underlying stock returns are discontinuous', *Journal of Financial Economics* **5**, 125–144.

Myneni, R. (1992), 'The pricing of the american option', *The Annals of Applied Probability* **2**(1), 1–23.

Rendleman, R. & Bartter, B. (1979), 'Two-state option pricing', *Journal of Finance* pp. 1093–1110.

Revuz, D. & Yor, M. (1991), *Continuous Martingales and Brownian Motion*, 2. edn, Springer-Verlag.

Rubinstein, M. (1976), 'The valuation of uncertain income streams and the pricing of options', *Bell Journal of Economics* **7**, 407–425.

Rubinstein, M. (1994), 'Implied binomial trees', *The Journal of Finance* **69**(3), 771–818.

Seidenverg, E. (1988), 'A case of confused identity', *Financial Analysts Journal* pp. 63–67.

Shimko, D. (1991), Beyond implied volatility: Probability distributions and hedge ratios implied by option prices. Mimeo. Department of Finance and Business Economics. University of Southern California.

A simple regime switching term structure model*

Asbjørn T. Hansen
3rd Floor, Swaps
Greenwich NatWest, 135 Bishopsgate
London EC2M 3UR England

Rolf Poulsen
Department of Operations Research
University of Aarhus, Building 530
DK-8000 Denmark
email: rolf@imf.au.dk

June 6, 1999

Abstract

We extend the short rate model of Vasicek (1977) to include jumps in the local mean. Conditions ensuring existence of a unique equivalent martingale measure are given, implying that the model is arbitrage-free and complete. We develop efficient numerical methods for computation of zero coupon bond prices, illustrate how the model is easily calibrated to market data and show how other interest rate derivatives can be priced.

*We are grateful for the detailed and insightful comments from two anonymous referees; these comments have lead to significant improvements of the paper. Further, we thank Tomas Björk, Svend Erik Graversen, Peter Honoré, Jesper Lund, Ken Nyholm, and Jan Pedersen for valuable suggestions.

1 Introduction

Many papers have extended the classical mean reversion interest rate models of Vasicek (1977) and Cox, Ingersoll & Ross (1985) as well as their parametric generalisation in Chan, Karolyi, Longstaff & Sanders (1992). Examples of such extensions include Longstaff & Schwartz (1992), Andersen & Lund (1996), and Conley, Hansen, Luttmer & Scheinkmann (1997). Typically one or more additional sources of Wiener-noise are introduced through the volatility parameter or through the level towards which the short rate reverts.

The former approach introduces stochastic volatility interest rate models. However, recent analysis suggests that stochastic short rate volatility has little effect on bond prices; especially for short term bonds. This conclusion is for instance suggested in the work of Schlögl & Sommer (1997). They examine what impact different parametrisations of the diffusion coefficient has on the class of yield curves in otherwise similar interest rate models. They find that any term structure from Cox et al. (1985) can be well approximated by some term structure from the Vasicek (1977) or indeed by some term structure from any member of the CKLS-family. Thus if term structure modelling is our primary aim it seems more interesting to introduce an additional noise factor through the drift specification of the short rate.

We choose the computationally tractable Vasicek (1977) model as our basis model and augment it by letting the level to which the process reverts change. These changes are governed by arrivals from a Poisson-process. This means that jumps will be present in the drift of the short rate process. The short rate process remains continuous but as we shall see, zero coupon bond prices jump. In other words we will see rapid shifts in the term structures. Moreover the term structures generated over time can easily cross each other. Both these features will make the traditional static duration measures perform badly.

The paper proceeds as follows: In Section 2 the model is presented and in Section 3 we demonstrate how an equivalent martingale measure can be constructed. This is important for derivative pricing and since our modelling framework is "indirect" all assets in the fixed income market, including bonds, are derivatives on the short rate and the mean factor. Taking an equivalent martingale measure as given in Section 4 we show that zero coupon bond prices are multiplicatively separable in a "Vasicek" and a "jump" component. This makes calibration to the current term structure easy. We derive computationally efficient algorithms for calculation of bond prices. In

Section 5 we investigate the pricing of more complex derivatives, zero coupon bond options in particular. Section 6 concludes the paper and outlines topics for future research.

2 The model

We consider a time interval $[0, \mathcal{T}]$ on with the stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,\mathcal{T}]}, P)$ satisfying the usual conditions. On the stochastic basis we assume that there exists a Wiener process (W_t) and a Poisson process (independent of (W_t)) (N_t) with P-intensity λ^P.

The short rate process $(r(t))$ (or sometimes: (r_t)) is defined by the stochastic differential equation

$$dr(t) = \kappa(\theta(t-) - r(t-))dt + \sigma dW(t) \tag{1}$$

where κ, $\sigma > 0$ and

$$d\theta(t) = (\theta^H - \theta^L)\left(\mathbf{1}_{\{(\theta(t-)=\theta^L\}} - \mathbf{1}_{\{\theta(t-)=\theta^H\}}\right) dN(t) \tag{2}$$

given r_0 and $\theta_0 \in \{\theta^L, \theta^H\}$. The interest rate process is mean reverting towards a stochasticly shifting long term level $(\theta(t))$ with κ being the speed of mean reversion and σ being the volatility parameter. The local mean level $(\theta(t))$ can take on two values; a high level θ^H and a low level θ^L. Because the changes are controlled by arrivals from a Poisson-process, θ is "memoryless"; the (distribution of the) time we have to wait for the next change does not depend on how long we have been in the current state. Note that if θ were constant, (1) would be the short rate specification from Vasicek (1977).

3 Martingale measures

In this section we address the existence of equivalent martingale measures. Recall that the existence of an equivalent martingale measure will imply that our model is arbitrage consistent. Martingale measure characterizations in jump-diffusion interest rate models have earlier been obtained by Shirakawa (1991). However, in the model of Shirakawa (1991) the jumps directly impact the level of the short rate whereas the jump specification in the present paper only indirectly affects the short rate through the drift of this process.

Assumption 1. *Zero coupon bond prices depend only and smoothly on r, θ and time. Specifically,*

$$P(t,T) = H^T(r(t), \theta(t), t),$$

where $H^T \in \mathcal{C}^{2,2,1}(\mathbf{R}^3)$ for all T. Further, there exists a bank account in the market with $B(0) = 1$ and price dynamics given by

$$dB(t) = r(t)B(t)dt.$$

With subscripts denoting partial derivatives we can now define

$$\begin{aligned} \alpha^T(t) &= \frac{1}{H^T}\left(H_r^T \kappa(\theta(t) - r(t)) + \frac{1}{2}H_{rr}^T \sigma^2 + H_t^T\right), \\ \delta^T(t) &= \sigma \frac{H_r^T}{H^T}, \end{aligned}$$

and

$$C^T(t) = (\mathbf{1}_{\{\theta(t)=\theta^L\}} - \mathbf{1}_{\{\theta(t)=\theta^H\}})(H^T(r(t), \theta^H, t) - H^T(r(t), \theta^L, t)).$$

Assumption 2. *The system of equations*

$$\lambda(t) = \frac{\alpha^T(t) + C^T(t)\lambda^{Q^N} - r(t)}{\delta^T(t)} \tag{3}$$

has a continuous solution (λ, λ^{Q^N}) that (i) does not depend on T, (ii) is unique (iii) is deterministic, and (iv) has constant second argument, λ^{Q^N}.

The following result is proven in Appendix A.

Proposition 1. Assumptions 1 *and* 2 *imply that a unique equivalent martingale measure (corresponding to the bank account as numeraire) Q exists; in particular there are no arbitrage opportunities.*
The Q-dynamics are given by

$$dr_t = \kappa(\theta_t + \sigma\lambda(t)/\kappa - r_t)dt + \sigma dW_t^Q, \tag{4}$$

with W^Q being a Q-Brownian motion and

$$d\theta(t) = (\theta^H - \theta^L)\left(\mathbf{1}_{\{(\theta(t-)=\theta^L\}} - \mathbf{1}_{\{\theta(t-)=\theta^H\}}\right) dN^Q(t), \tag{5}$$

where $\theta_0 \in \{\theta^L, \theta^H\}$ and N_t^Q is a Q-Poisson process (independent of W^Q) with intensity λ^{Q^N}.

Assumption 2 and Proposition 1 can be interpreted in the following way. A P-specification of the short rate dynamics is not enough to determine zero coupon bond prices; risk premia corresponding to the "sources of risk" are needed (see for instance Björk (1998, Chapter 16).) These risk premia must be "universal", i.e. not asset specific. This is reflected in Assumption 2(i) and, when taken in conjunction with parts (iii) and (iv), means that the uniqueness part will rarely be a problem. The risk premia can then be used to introduce the martingale measure Q, the existence of which we know precludes arbitrage. Proposition 1 shows why Assumption 2$(iii)-(iv)$ are needed; they ensure the model has same structure under P and Q. Put differently, if we exogenously specify the dynamics of two zero coupon bonds in such a way that Assumption 2 $(ii)-(iv)$ hold then we know exactly what Q looks like, what dynamics zero coupon bonds must have, and how to price and hedge interest rate derivatives. Alternatively, one could take the latter part of Proposition 1 as the denfinition of the model, but this does not fully reflect the fact that despite jumps be present, the model is not "grossly incomplete". Ideally, Equations (1)-(2) and Assumption 2 are things that could and should be verified empirically. This, however, is far beyond the scope of this paper.

4 Bond prices

In Section 3 we established the existence of an equivalent martingale measure for our model. In this section we develop manageable expressions for zero coupon bond prices.

We have the following decomposition result.

Proposition 2. *Let $r(0) = r_0$, $\theta(0) = \theta_0 \in \{\theta^L, \theta^H\}$, and A denote the set of odd numbers. The short rate can then be written as*

$$r(t) = a(t) + b(t) + c(t)$$

where

$$\begin{aligned} a(t) &= r_0 e^{-\kappa t} + \kappa \int_0^t e^{-\kappa(t-s)} \theta_0 ds + \sigma \int_0^t e^{-\kappa(t-s)} dW^Q(s), \\ b(t) &= \begin{cases} \kappa(\theta^H - \theta^L) \int_0^t e^{-\kappa(t-s)} \mathbf{1}_{\{N_s^Q \in A\}} ds & \text{if } \theta_0 = \theta^L \\ \kappa(\theta^H - \theta^L) \int_0^t e^{-\kappa(t-s)} \mathbf{1}_{\{N_s^Q \notin A\}} ds & \text{if } \theta_0 = \theta^H \end{cases} \end{aligned}$$

and

$$c(t) = \sigma \int_0^t e^{-\kappa(t-s)} \lambda(s) ds.$$

In particular $(a(t))$ *is independent of* $(b(t))$ *and*

$$a(t) \sim N\left(\theta^L + e^{-\kappa t}(r_0 - \theta^L), \frac{\sigma^2}{2\kappa}(1 - e^{-2\kappa t})\right). \tag{6}$$

Proof. Using Ito's formula on $(e^{\kappa t} r_t)$ gives us that

$$r_t = r_0 e^{-\kappa t} + \kappa \int_0^t e^{-\kappa(t-s)}(\theta_s + \sigma\lambda(s)/\kappa) ds + \sigma \int_0^t e^{-\kappa(t-s)} dW^Q(s).$$

By noting that we may write

$$\theta(t) = \begin{cases} \theta_0 + (\theta^H - \theta^L)\mathbf{1}_{\{N_t^Q \in A\}} & \text{if } \theta_0 = \theta^L \\ \theta_0 + (\theta^H - \theta^L)\mathbf{1}_{\{N_t^Q \notin A\}} & \text{if } \theta_0 = \theta^H, \end{cases}$$

the first part follows. The independence is evident, and since $\int_0^t h(s) dW^Q(s) \sim N(0, \int_0^t h^2(s) ds)$ for any deterministic, continuous function h, we have the last part. $\diamondsuit$

This decomposition is important because of the generic bond price formula (which follows form the very definition of Q)

$$P(t,T) = \mathbf{E}^Q \left(\exp(-\int_t^T r_s ds) | \mathcal{F}_t \right).$$

But before stating a decomposition result for zero coupon bond prices, let us now introduce some further notation. For arbitrary real numbers θ_0 and $\widehat{\theta}_0$ we define

$$h(\theta_0, T) = \mathbf{E}^Q \left(\exp\left(-\kappa(\widehat{\theta}_0 - \theta_0) \int_0^T \int_0^t e^{-\kappa(t-s)} \mathbf{1}_{\{N_s^Q \in A\}} ds dt \right) \right)$$

and

$$\widehat{h}(\theta_0, T) = \mathbf{E}^Q \left(\exp\left(-\kappa(\theta_0 - \widehat{\theta}_0) \int_0^T \int_0^t e^{-\kappa(t-s)} \mathbf{1}_{\{N_s^Q \notin A\}} ds dt \right) \right).$$

Since $\mathbf{1}_{\{N_s^Q \in A\}} + \mathbf{1}_{\{N_s^Q \notin A\}} = 1$, it is clear that

$$\begin{aligned} h(\theta_0, T) &= \widehat{h}(\theta_0, T) \mathbf{E}^Q \left(\exp\left(-\kappa(\widehat{\theta}_0 - \theta_0) \int_0^T \int_0^t e^{-\kappa(t-s)} ds dt \right) \right) \\ &= \widehat{h}(\theta_0, T) e^{m(\widehat{\theta}_0; 0, T) - m(\theta_0; 0, T)}, \end{aligned}$$

where

$$n(t,T) = \frac{1-\exp(-\kappa(T-t))}{\kappa}, \tag{7}$$

$$m(\theta;t,T) = \frac{(n(t,T)-(T-t))(\kappa^2\theta-\sigma^2/2)}{\kappa^2} - \frac{\sigma^2 n(t,T)^2}{4\kappa}. \tag{8}$$

(It may seem that the m-function is defined in a strange way; it will soon become apparent why we do this.) We shall adapt the convention that the $\widehat{\ }$-notation is used to denote "complementary state" i.e.,

$$\widehat{\theta}_0 = \theta^L \qquad \text{if } \theta_0 = \theta^H,$$
$$\widehat{\theta}_0 = \theta^H \qquad \text{if } \theta_0 = \theta^L,$$

and similarly $\widehat{\theta}(t) = \theta^L$ if $\theta(t) = \theta^H$, $\widehat{\theta}(t) = \theta^H$ if $\theta(t) = \theta^L$.

Proposition 3. *Let $r(0) = r_0$, $\theta(0) = \theta_0 \in \{\theta^L, \theta^H\}$, and A be the odd numbers. Then the price of the T-maturity zero coupon bond equals*

$$P(0,T|r_0,\theta_0) = P_{\mathrm{V}}^{\theta_0}(r_0,0,T)\ \exp\left(-\int_0^T c(s)ds\right) h(\theta_0,T), \tag{9}$$

and

$$P(0,T|r_0,\theta_0) = P_{\mathrm{V}}^{\widehat{\theta}_0}(r_0,0,T)\ \exp\left(-\int_0^T c(s)ds\right) \widehat{h}(\theta_0,T). \tag{10}$$

where $P_{\mathrm{V}}^{\theta}(r_0,t,T)$ denotes the price of a zero coupon bond in a Vasicek model with θ as the long term level of the short rate and is explicitly given by

$$P_{\mathrm{V}}^{\theta}(r_t,t,T) = \exp(m(\theta,t,T) - r_t n(t,T)),$$

where functions m and n are given by (8) and (7), respectively.

The ratio of bond prices immediately before and after a jump is

$$\begin{aligned}\frac{P(0,T|r_0,\theta^L)}{P(0,T|r_0,\theta^H)} &= \frac{h(\theta^L,T)}{\widehat{h}(\theta^H,T)} \\ &= \frac{\mathbf{E}^Q\left(\exp(-\kappa(\theta^H-\theta^L)\int_0^T\int_0^t \mathbf{1}_{\{N_s^Q\in A\}}e^{-\kappa(t-s)}dsdt)\right)}{\mathbf{E}^Q\left(\exp(-\kappa(\theta^H-\theta^L)\int_0^T\int_0^t \mathbf{1}_{\{N_s^Q\notin A\}}e^{-\kappa(t-s)}dsdt)\right)}.\end{aligned} \tag{11}$$

Since $(r(t),\theta(t))$ is jointly Markov, we also have

$$P(t,T|r_t,\theta_t) = P_{\mathrm{V}}^{\theta(t)}(r(t),t,T)\ \exp\left(-\int_0^T c(s)ds\right) h(\theta(t),T-t), \tag{12}$$

and

$$P(t,T|r_t,\theta_t) = P_{\mathrm{V}}^{\widehat{\theta}(t)}(r(t),t,T)\ \exp\left(-\int_0^T c(s)ds\right)\widehat{h}(\theta(t),T-t). \tag{13}$$

Finally,

$$\begin{aligned} P(t,T|r_0,\theta_0) &= P_{\mathrm{V}}^{\theta^L}(r(t),t,T)\exp\left(-\int_0^T c(s)ds\right) \\ &\quad \times\left(h(\theta^L,T-t)\mathbf{1}_{\{\theta(t)=\theta^L\}}+\widehat{h}(\theta^H,T-t)\mathbf{1}_{\{\theta(t)=\theta^H\}}\right) \end{aligned} \tag{14}$$

and

$$\begin{aligned} P(t,T|r_0,\theta_0) &= P_{\mathrm{V}}^{\theta^H}(r(t),t,T)\exp\left(-\int_0^T c(s)ds\right) \\ &\quad \times\left(h(\theta^H,T-t)\mathbf{1}_{\{\theta(t)=\theta^H\}}+\widehat{h}(\theta^L,T-t)\mathbf{1}_{\{\theta(t)=\theta^L\}}\right). \end{aligned} \tag{15}$$

Proof. Note that the distribution of $a(t)$ is exactly that of the short rate in a θ_0-standard Vasicek model. Using this for $\theta_0=\theta^L$, combining it with Proposition 2, the generic bond price formula and the bond price formula in e.g. Vasicek (1977) gives (9). Equation (10) is also clear since in the Vasicek model

$$P_V^{\widehat{\theta}_0}(r_0,0,T) = P_V^{\theta_0}(r_0,0,T)e^{m(\widehat{\theta}_0;0,T)-m(\theta_0;0,T)}.$$

Equation (11) follows if we use (9) with $\theta_0=\theta^L$ and (10) with $\theta_0=\theta^H$. Equations (12)-(13) are immediate generalizations. Finally, when combining (12)-(13) we get (14)-(15). $\diamondsuit$

Let us note that the above arguments work only for a model with deterministic short rate volatility. Specifically, the decomposition of bond prices does not hold in a CIR- or CKLS version of the model.

Figure 1 depicts a typical trajectory for the short rate process. Note that we can do exact simulation of the short rate process by first simulating jumps from a Poisson process and then using the distribution result for $a(t)$ given in Proposition 2.

`Figure 1 GOES ABOUT HERE`

We will refer to the last factor on the LHS of (9) as a *jump premium*. If $\theta_L<\theta_H$, it is less that 1 meaning that bond prices are lower than in a θ^L-Vasicek model, which is what we expect because the short rate sometimes reverts to a higher level. Note also that the ratio in (11) clearly is not equal to 1, so bond prices, and hence yields ($y(t,T)=-(\ln P(t,T))/(T-t)$) of strictly positive maturities exhibit jumps in our model even though the short rate itself is a continuous process. Further the ratio is independent of σ (which is intuitively reasonable) and r_0. Also, (22), and the multiplicative decomposition of bond prices shows us that the Wiener-part is the same as in a Vasicek model.

Consider for a moment the jump premium as a known function (the following section will show how to calculate it.) Proposition 3 makes it very easy to calibrate the model to market data using the ideas in Hull & White (1990). Specifically, let $P^{\text{obs}}(0,T)$ denote the observed zero coupon bond prices and put

$$f(T) = -\ln\left(\frac{P^{\text{obs}}(0,T)}{P_{\text{V}}^{\theta^L}(0,T)\mathbf{E}^Q\left(\exp(-\int_0^T b(s)ds)\right)}\right).$$

If we assume κ, θ^L, θ^H, σ, and λ^{Q^N} have been estimated, for example from time series data[1], the model is calibrated by choosing λ such that

$$f(T) = \sigma \int_0^T \int_0^s e^{-\kappa(s-u)}\lambda(u)duds \quad \text{for all } T. \tag{16}$$

Differentiating (16) twice w.r.t. T gives the very neat result that

$$\lambda(T) = \frac{f'' - \kappa f'}{\sigma} \quad \text{for all } T.$$

In practice, however, it is doubtful whether enough points on the yield curve are known to actually perform the double differentiation. In this case we would calibrate by a forward algorithm applied to (16).

4.1 Backward induction

We now give an efficient numerical evaluation method for pricing of zero coupon bonds. The technique generalizes to any model where the short rate can be decomposed as a sum of independent Markov-processes.

To simplify the exposition suppose (without loss of generality for the purposes of this section) $\theta(0) = \theta^L$ and let

$$g(\theta_t, t; T) = \mathbf{E}^Q\left(\exp(-\int_t^T b(s,t)ds)|\mathcal{F}_t\right)$$

[1]This is not without subtle points, none-less-so because we need the Q-intensity of the Poisson process. Notice also that the λ-process will depend on the estimates.

where

$$b(s,t) = \kappa(\theta^H - \theta^L) \int_t^s \exp(-\kappa(s-u))\mathbf{1}_{\{\theta_u = \theta^H\}} du$$

(so the $b(s)$-notation in the previous section is then shorthand for $b(s,0)$.) Note that from (14) we have

$$P(t,T|r_0,\theta_0) = P_V^{\theta^L}(r_0,\theta_0)\exp(-\int_t^T c(s)ds)g(\theta(t),t;T),$$

so if we can find g, then we can calculate bond prices. For the rest of this section we suppress the last argument of g.

Since $g(\theta(T),T) = 1$ are able to compute $g(\theta_0, 0)$ by backward induction. This backward induction technique is for instance well-known from lattice methods of pricing contingent claims (e.g. finite difference methods and binomial methods.) As it is often the case in such situations we shall be able to compute values for g at all time points used in the backward equation. This in turn allows us to characterize the entire term structure going backward through time just once. In the present set-up the backward induction technique will be extremely efficient because the underlying Markov state variable θ can take two values only. We will now denote the intensity λ^{Q^N} simply by λ; this should cause no confusion.

Proposition 4. *Consider a discretization of the time interval $[0,T]$ through the $n+1$ points $0 = t_0 \leq t_1 \cdots \leq t_{n-1} \leq t_n = T$ and set $\triangle_i = t_i - t_{i-1}$. For $i < n$, a first-order accurate (as $\sup\{\triangle_j\} \to 0$) approximation to g is given by the recursive formulae*

$$\widehat{g}(\theta^L, t_{i-1}) = e^{-\lambda\triangle_i}\Big(\widehat{g}(\theta^L, t_i)\widehat{f}(\theta^L, 0; t_i) + \lambda\triangle_i\widehat{g}(\theta^H, t_i)\widehat{f}(\theta^L, 1; t_i)\Big) \tag{17}$$

$$\widehat{g}(\theta^H, t_{i-1}) = e^{-\lambda\triangle_i}\Big(\widehat{g}(\theta^H, t_i)\widehat{f}(\theta^H, 0; t_i) + \lambda\triangle_i\widehat{g}(\theta^L, t_i)\widehat{f}(\theta^H, 1; t_i)\Big) \tag{18}$$

where $\widehat{g}(\cdot, t_n) = \widehat{f}(\cdot,\cdot; t_n) = 1$ and

$$\begin{aligned}
\widehat{f}(\theta^L, 0; t_i) &= 1, \\
\widehat{f}(\theta^H, 0; t_i) &= \exp\left((\theta^L - \theta^H)\left(\triangle_i - \frac{e^{-\kappa(T-t_i)} - e^{-\kappa(T-t_{i-1})}}{\kappa}\right)\right), \\
\widehat{f}(\theta^H, 1; t_i) &= \exp\Big((\theta^L - \theta^H)\Big(\frac{\triangle_i}{2} - \frac{1}{\kappa}\Big(\frac{2(1 - e^{-\kappa\triangle_i} - \kappa\triangle_i e^{-\kappa\triangle_i})}{\kappa\triangle_i} \\
&\qquad + e^{-\kappa T}\Big(e^{\kappa t_i} - \frac{e^{\kappa t_i} - e^{\kappa t_{i-1}}}{\kappa\triangle_i}\Big)\Big)\Big)\Big), \\
\widehat{f}(\theta^L, 1; t_i) &= \exp\Big((\theta^L - \theta^H)\Big(\frac{\triangle_i}{2} - \frac{e^{-\kappa T}}{\kappa}\Big(\frac{e^{\kappa t_i} - e^{\kappa t_{i-1}}}{\kappa\triangle_i} - e^{\kappa t_i}\Big)\Big)\Big).
\end{aligned}$$

A proof can be found in Appendix B.

Using Proposition 4 and time-shifting we can calculate the whole term structure. In practical applications we prefer to use $e^{-\lambda\triangle}/(e^{-\lambda\triangle}(1+\triangle\lambda)) = 1/(1+\triangle\lambda)$ rather than $e^{-\lambda\triangle}$ as multiplication factor in equations (17)-(18). This does not change the order and it gives better results for small n (large $\triangle$) (an inspection of the $O(\cdot)$-term in the proof of Proposition 4 reveals why.)

Different yield curves shapes are shown in Figure 2.

`Figure 2 GOES ABOUT HERE`

The output from the algorithm is so smooth that the accuracy/speed can by improved by some extrapolation technique. We used Richardson extrapolation (cf. Press, Teukolsky, Vetterling & Flannery (1992).) This does not increase speed very much, but the calculations can be used to check whether the algorithm has *global* first-order accuracy (it has) and to give an error-estimate.

Figure 3 shows what the yield curve looks like immediately before and after a regime switch.

`Figure 3 GOES ABOUT HERE`

4.2 Simulation

There are other ways of looking at the jump premium. We now give an evaluation approach based on simple simulation techniques. An advantage of this approach is that it easily extends to path-dependent parameter specifications or models with more than two θ-levels.

Define τ_i as the time of the ith jump, i.e. $\tau_i = \inf\{t | N_t^Q > i-1\}$ ($\tau_0 \equiv 0$), and look at

$$\begin{aligned} I(T) &:= \int_0^T b(t)dt \\ &= (\theta^H - \theta^L) \int_0^T \sum_{i=1}^{\infty} (e^{-\kappa(t-\tau_{2i}\wedge t)} - e^{-\kappa(t-\tau_{2i-1}\wedge t)})dt, \end{aligned} \tag{19}$$

where the second equality hinges on A being the odd numbers. If we let $m(T) := \sup\{i | \tau_{2i-1} \le T\}$ (so $m(T) < \infty$ for almost all ω) then the terms in the seemingly infinite sum in (19) vanish for $i > m(T)$ and we get

$$I(T) = (\theta^H - \theta^L) \sum_{i=1}^{m(T)} \left\{ \int_0^T (e^{-\kappa(t-\tau_{2i}\wedge t)} - e^{-\kappa(t-\tau_{2i-1}\wedge t)})dt \right\}$$

$$= (\theta^H - \theta^L) \sum_{i=1}^{2m(T)} (-1)^i \left\{ \tau_i \wedge T - \frac{e^{-\kappa(T-\tau_i)} - e^{-\kappa(\tau_i \wedge T - \tau_i)}}{\kappa} \right\}. \qquad (20)$$

We can now implement a Monte Carlo simulation approach in a way that is both easy (w.r.t. "time needed to program") and efficient (compared to simulating paths of the short rate process.) To simulate an outcome of $I(T)$ (call this $\tilde{I}_k(T)$) we only have to simulate independent $\exp(\lambda)$-distributed variables, because we know that the successive waiting times, $\tau_i - \tau_{i-1}$, are independent and $\exp(\lambda)$-distributed. Finally, with K simulations we approximate by

$$\mathbf{E}^Q(\exp(-I(T))) \approx \frac{1}{K} \sum_k \exp(-\tilde{I}_k(T)).$$

In a computer implementation we simulate a (long) series of τ_i's and generate $\tilde{I}_k(T)$ for different values T thus giving us several points on the yield curve. To speed up this process we rewrite (20) as

$$\frac{I(T)}{(\theta^H - \theta^L)} = \sum_{i=1}^{2m(T)} (-1)^i (\tau_i \wedge T) - \frac{e^{-\kappa T}}{\kappa} \sum_{i=1}^{2m(T)} (-1)^i e^{\kappa \tau_i} + \frac{e^{-\kappa T}}{\kappa} \sum_{i=1}^{2m(T)} (-1)^i e^{-\kappa(\tau_i \wedge T - \tau_i)}.$$

A small increase in T effects only the last few terms in the first two sums, and the last sum is "almost telescoping". Thus, we can update in a simple fashion without having to redo all the sums. Further, careful programming allows us to calculate $\tilde{I}_k(T)$ without exponentiating numerically large numbers.

Table 1 compares numerical methods:

- "Naive simulation" is just that. We first simulate jump times and then use the Euler scheme on (4) (treating θ as deterministic) to generate paths of the short rate. Sufficiently many paths give an estimate of the term structure.
- The simulation method described in this section (unmodestly referred to as "smart".)
- The backward induction technique.

`Table 1 GOES ABOUT HERE`

The experiment has been designed such that the 10-year yield is determined within 1 basispoint (the error/uncertainty is quite insensitive to maturity.) The methods agree (as they should, of course) on yield-estimates but differ markedly in the time needed to obtain these. Unsurprisingly the "naive" method is very slow, but we see that the backward induction technique is about 15 times faster than the "smart" simulation. Also, the absolute runtime for the backward induction is of a magnitude that justifies thinking of the bond price as a "known function" in the following.

5 Pricing bond options

In this section we look at pricing of more advanced interest rate derivatives, call-options on zero coupon bonds in particular.

This means that we will be looking at the generic call-option price formula for the time t price of a call-option expiring at time T_E on a zero coupon bond maturing at time T_M

$$\text{call}(t) = \mathbf{E}^Q \left(e^{-\int_0^{T_E} r(s)ds} \left(P(T_E, T_M) - K\right)^+ |\mathcal{F}_t \right).$$

The homogeneity of the pay-off function and the multiplicative bond price decomposition are key features in reducing the problem to one that involves only the θ-process. Specifically, we have the following proposition.

Proposition 5. *Let $r(0) = r_0$, $\theta(0) = \theta_0$. Assume that $c \equiv 0$ and let $\text{call}_V^{\theta^L}(r_0, K)$ denote the call option price in a θ^L-Vasicek model (an explicit formula can be found e.g. in Jamshidian (1989).) The price (at time 0) of a call-option with strike price K expiring at time T_E on a zero coupon bond maturing at time T_M satisfies*

$$\begin{aligned}\text{call}(r_0, \theta_0, K) &= \mathbf{E}^Q \left(\exp\left(-\int_0^{T_E} b(s)ds \right) e^{-n(T_E,T_M)b(T_E)} \xi(\theta(T_E), T_M - T_E) \right. \\ &\qquad \left. \times \, \text{call}_V^{\theta^L} \left(r_0, \frac{K}{\xi(\theta(T_E); T_M - T_E)} \right) \right)\end{aligned}$$

where the random variable $\xi(\theta(T_E), T_M - T_E)$ is given by

$$\xi(\theta(T_E); T_M - T_E) = h(\theta^L, T_M - T_E)\mathbf{1}_{\{\theta(T_E)=\theta^L\}} + \widehat{h}(\theta^H, T_M - T_E)\mathbf{1}_{\{\theta(T_E)=\theta^H\}}$$

with

$$\theta(T_E) = \theta_0 + (\widehat{\theta}_0 - \theta_0)\mathbf{1}_{\{N^Q_{T_E} \in A\}} = \widehat{\theta}_0 + (\theta_0 - \widehat{\theta}_0)\mathbf{1}_{\{N^Q_{T_E} \notin A\}}$$

and (as previously)

$$b(t) = \begin{cases} \kappa(\theta^H - \theta^L) \int_0^t e^{-\kappa(t-s)} \mathbf{1}_{\{N^Q_s \in A\}} ds & \text{if } \theta_0 = \theta^L \\ \kappa(\theta^H - \theta^L) \int_0^t e^{-\kappa(t-s)} \mathbf{1}_{\{N^Q_s \notin A\}} ds & \text{if } \theta_0 = \theta^H \end{cases}$$

A proof can be found in Appendix C.

The advantage of the result in Proposition 5 is that the expectation involves only the θ-process and only the time-interval $[0; T_E]$. Therefore it can be easily evaluated by the simulation technique described in Section 4.2. The computation time needed to obtain a certain degree of accuracy is of course platform- and parameter dependent but for "reasonable" choices (meaning, in particular, the one considered in the following) accuracy within 0.1 % of the option price for at-the-money options takes less than 0.2 second. To compare to option prices in the Vasicek model define the spread as

$$\delta := \theta^H - \theta^L$$

and consider this set-up:

Bond: $r_0 = (\theta^H + \theta^L)/2$, $\kappa = 2$, $\sigma = 0.02$, $\lambda = 1$.

Option: $T_E = 1$, $T_M = 10$, $K = 0.50933$.

So, $\delta = 0$ gives a Vasicek-model (with a very flat yield curve) and $\delta < 0$ means that the short rate is "on its way up". The strike price has been chosen such that the option in the Vasicek model is "forward-at-the-money" meaning that the strike is equal to the forward price of the underlying i.e. $K = P(0, 10)/P(0, 1)$. Figure 4 shows call prices for negative spreads (for positive spreads the call-option price is monotonely increasing.) We notice a peculiar effect: For numerically small, negative values of δ the call price is less than the Vasicek price, but for δ numerically large, the price is higher. The explanation for this can be found from the decomposition

$$r_{T_E} = a(T_E) + \kappa\delta \int_0^{T_E} \mathbf{1}_{\{N_s^Q \in A\}} e^{-\kappa(T_E - s)} ds.$$

A larger value of θ^L (i.e. $\delta < 0$) will (by a symmetry-argument) increase the expected value of r_{T_E}, which tends to decrease bond price at time T_E, hence lowering the call price. On the other hand, a numerically larger δ increases the variance of r_{T_E} (only the last part of the decomposition is affected, so std.dev.$(r_{T_E}) \propto |\delta|$) which increases the variance of the time T_E bond price; this will tend to increase the call price.

However, it could easily be argued that the comparison of call-option prices is like comparing apples and pears since varying δ creates different initial term structures. In market applications we would include risk premia and fit to the (partially) observed initial term structure as outlined in the previous section. But another, more simple, way of making reasonable comparisons is to correct the strike price of the option such the it is always forward-at-the-money (and otherwise doing as outlined above.) Call-option prices obtained in this way are also shown in Figure 4; we see that the "mean-effect" disappears.

The algorithm for computing zero coupon bond option prices extends (at little extra computational cost) to options on coupon bearing bonds when the results in Jamshidian (1989) are used.

Finally, let us note that Eurodollar futures prices can also easily be found. Following Musiela & Rutkowski (1997, Section 16.1.3) we have

$$\begin{aligned}
\mathrm{EDF} &= 2 - \mathbf{E}^Q\left(\frac{1}{P(T, T+\delta)}\right) \\
&= 2 - \mathbf{E}^Q\left(\frac{1}{\xi(\theta(T);\delta)}\right)(2 - \mathrm{EDF}_V^{\theta^L}),
\end{aligned}$$

where $\mathrm{EDF}_V^{\theta^L}$, the Eurodollar futures price in the Vasicek model, is easy to find.

6 Conclusion

In this paper we developed a regime switching fixed income model. This was done by augmenting the Vasicek (1977) model allowing the interest local mean to shift between a high and a low state.

We gave sufficient conditions for the model to and arbitrage free and found a (convenient) equivalent martingale measure.

A semi-analytic formula for zero coupon bond prices was found and we gave efficient numerical methods for calculating these prices. This made calibration to the current term structure easy. Finally, we looked at pricing of other interest rate derivatives and found that these could also easily, though numerically, be priced in the model.

In analysing the term structures we remarked the possibility of extensive and rapid shifts in the curves. In particular we noted that when the steady state level of the short rate switches the yield curve might change shape and that we might experience non-parallel shifts in the term structure over a small horizon of time. We remarked that the large impact on the interest rate model of introducing jumps in the drift of the short rate process is consistent with the work of Schlögl & Sommer (1997).

The model is of course very simple since θ only has two levels. An obvious generalization is letting it follow a finite-state Markov chain independent of W. We then still obtain a decomposition into a "Vasicek" and a "jump" part (as in Proposition 3.) Some results for models of this type can be found in Naik & Lee (1994) and Landén (1998). However, implementing these more general models in a computationally efficient way seems to be hard.

Others have considered altering the specification of the drift of the short rate. By specifying the θ-process as a Gaussian mean-reverting process (under an equivalent martingale measure) we can end up in the class of affine yield-curve models in which closed-form solutions for bond prices are known (see Balduzzi, Das, Foresi & Sundaram (1998).) Models of this type do not have the bi- (or multi-)modal short rate distribution that is a characterizing feature of our and the above mentioned regime switching models.

A way to obtain qualitatively the same short rate behaviour (two "local levels") is to use the one-factor framework of Conley et al. (1997) were the drift is a linear combination of powers of r and can have multiple roots. Models of this type have enjoyed some empirical success but are not very feasible for bond pricing and term structure calibration.

In conclusion, many criteria should be considered (simplicity, "time-series"-, and "cross-sectional"-behaviour, derivative pricing, computational aspects) when choosing an interest rate model. We have presented a model that we feel meets a number of the requirements and has proved worthy of investigation.

A Proof of Proposition 1

Pick an arbitrary T. The equations in the following are then understood to be valid for $t < T$. An extended version of Ito's formula (see Jacod & Shiryaev (1987)) implies that

$$\begin{aligned} dH^T(r(t),\theta(t),t) &= H_r^T(r(t-),\theta(t-),t-)dr + \frac{1}{2}H_{rr}^T(r(t-),\theta(t-),t-)(dr)^2 \\ &+H_t^T(r(t-),\theta(t-),t-)dt + H^T(r(t),\theta(t),t) - H^T(r(t-),\theta(t-),t-). \end{aligned}$$

Noting that

$$\begin{aligned} &H^T(r(t),\theta(t),t) - H^T(r(t-),\theta(t-),t-) = \\ &(\mathbf{1}_{\{\theta(t-)=\theta^L\}} - \mathbf{1}_{\{\theta(t-)=\theta^H\}})(H^T(r(t-),\theta^H,t-) - H^T(r(t-),\theta^L,t-))dN(t) \end{aligned}$$

and inserting from (1) we obtain

$$\begin{aligned} dH^T(r(t),\theta(t),t) &= \left\{ H_r^T\kappa(\theta(t-) - r(t-)) + \frac{1}{2}H_{rr}^T\sigma^2 + H_t^T \right\} dt \\ &+(\mathbf{1}_{\{\theta(t-)=\theta^L\}} - \mathbf{1}_{\{\theta(t-)=\theta^H\}})(H^T(r(t-),\theta^H,t-) - H^T(r(t-),\theta^L,t-))dN(t) \\ &+H_r^T\sigma dW^P(t). \end{aligned} \tag{21}$$

We look for the probability measures under which the discounted price process H^T/B is a martingale. Applying Ito's formula and making use of (21) we get

$$d\left(\frac{H^T(t)}{B(t)}\right) = \frac{H^T(t-)}{B(t)}\left((\alpha^T(t-) - r(t-))dt + \delta^T(t-)dW^P(t) + C^T(t-)dN(t)\right). \tag{22}$$

Task is now changing to a measure that allows us to rewrite (22) as integrals w.r.t. martingales. We can define a measure Q^N by the Radon-Nikodym density

$$\frac{dQ^N}{dP} = \exp\left((\lambda^P - \lambda^{Q^N})\left(\frac{\lambda^P}{\lambda^{Q^N}}\right)^{N_T}\right).$$

From standard theory (and Assumption 2) it follows that N_t is a standard Q^N-Poisson process with intensity λ^{Q^N}. Since $(N_t - \lambda^{Q^N}t)$ is a Q^N-martingale we rewrite (22) as

$$\begin{aligned} d\left(\frac{H^T(t)}{B(t)}\right) &= \frac{H^1(t-)}{B(t)}\delta^T(t-)\left\{\left(\frac{\alpha^T(t-) - r(t-) + C^T(t-)\lambda^{Q^N}}{\delta^T(t-)}\right)dt \right. \\ &\left. +dW(t) + \frac{C^T(t-)}{\delta^1(t-)}(dN(t) - \lambda^{Q^N}dt)\right\} \end{aligned}$$

Hence, H^T/B is a local martingale under if and only if $W_t + \int_0^t \lambda(s)ds$ is a Q^W-Brownian motion for some measure Q^W. But from Girsanov's Theorem this is the case if and only if

$$\frac{dQ^W}{dP} = \exp\left(-\frac{1}{2}\int_0^t \lambda^2(s-)ds + \int_0^t \lambda(s-)dW_t\right).$$

Since the $(\lambda(t), \lambda^{Q^N})$ is deterministic we can use $Q = Q^W \otimes Q^N$ as martingale measure for the price processes. The same change of measure works for all T (by Assumption 2 (i)) and we conclude that we have found an equivalent martingale measure. It is well-known that this ensures absence of arbitrage.

Any attempt to construct a martingale measure would involve finding a T-independent solution to system (3), but by assumption there is just such one solution. Hence there is only one martingale measure.

The latter statements of the proposition follow easily.

B Proof of Proposition 4

We start by observing $g(\theta_{t_n}, t_n) = 1$. Now look at $0 < i < n$ and let $\triangle_i = t_i - t_{i-1}$. Having obtained $g(\theta_{t_i}, t_i)$, the key is to find $g(\theta_{t_{i-1}}, t_{i-1})$. We do this the following way: From the definition

$$g(\theta_{t_{i-1}}, t_{i-1}) = \mathbf{E}^Q \left(\exp(-\int_{t_{i-1}}^T b(s, t_{i-1}) ds) | \mathcal{F}_{t_{i-1}} \right).$$

Now,

$$\int_{t_{i-1}}^T b(s, t_{i-1}) ds = \int_{t_i}^T b(t, t_i) dt + \int_{t_i}^T (b(t, t_{i-1}) - b(t, t_i)) dt + \int_{t_{i-1}}^{t_i} b(s, t_{i-1}) ds$$

where

$$\int_{t_i}^T (b(s, t_{i-1}) - b(s, t_i)) ds = \kappa(\theta^H - \theta^L) \int_{t_i}^T \exp(-ks) \int_{t_{i-1}}^{t_i} \exp(\kappa u) \mathbf{1}_{\{\theta_u = \theta^H\}} du ds$$

and

$$\int_{t_{i-1}}^{t_i} b(s, t_{i-1}) ds = \kappa(\theta^H - \theta^L) \int_{t_{i-1}}^{t_i} \exp(-\kappa s) \int_{t_{i-1}}^{s} \exp(\kappa u) \mathbf{1}_{\{\theta_u = \theta^H\}} du ds.$$

Therefore, using the law of iterated expectations we find

$$g(\theta_{t_{i-1}}, t_{i-1}) = \mathbf{E}^Q \left(g(\theta_{t_i}, t_i) f(t_i) | \mathcal{F}_{t_{i-1}} \right),$$

where $f(t_i)$ is the $\mathcal{F}_{t_i}$-measurable variable given by

$$\begin{aligned} f(t_i) \quad &= \quad \exp \left(\kappa(\theta^L - \theta^H) \int_{t_i}^T \exp(-\kappa s) \int_{t_{i-1}}^{t_i} \exp(\kappa u) \mathbf{1}_{\{\theta_u = \theta^H\}} du ds \right) \\ &\quad \times \exp \left(\kappa(\theta^L - \theta^H) \int_{t_{i-1}}^{t_i} \exp(-\kappa s) \int_{t_{i-1}}^{s} \exp(\kappa u) \mathbf{1}_{\{\theta_u = \theta^H\}} du ds \right). \end{aligned} \tag{23}$$

By the characteristics of the Poisson-process

$$g(\theta_{t_{i-1}}, t_{i-1}) = \mathbf{E}^Q \left(\mathbf{1}_{\{N^Q_{t_i} - N^Q_{t_{i-1}} \leq 1\}} g(\theta_{t_i}, t_i) f(t_i) | \mathcal{F}_{t_{i-1}} \right) + O(\triangle_i^2),$$

which gives rise to the approximate value $\widehat{g}$ satisfying

$$\widehat{g}(\theta_{t_{i-1}}, t_{i-1}) = \mathbf{E}^Q \left(\mathbf{1}_{\{N^Q_{t_i} - N^Q_{t_{i-1}} \leq 1\}} \widehat{g}(\theta_{t_i}, t_i) f(t_i) | \mathcal{F}_{t_{i-1}} \right)$$

(and of course $\widehat{g}(\cdot, t_n) = 1$.) We have to look at 4 cases corresponding to the elements in $\{\theta^L, \theta^H\} \times \{0, 1\}$.

$\underline{\theta_{t_{i-1}} = \theta^H \text{ and } N^Q_{t_i} - N^Q_{t_{i-1}} = 0}$ In this case $\widehat{g}(\theta_{t_i}, t_i) = \widehat{g}(\theta^H, t_i)$ and (23) can be calculated exactly resulting in

$$\widehat{f}(\theta^H, 0; t_i) = \exp\left((\theta^L - \theta^H) \left(\triangle_i - \frac{e^{-\kappa(T-t_i)} - e^{-\kappa(T-t_{i-1})}}{\kappa} \right) \right).$$

$\underline{\theta_{t_{i-1}} = \theta^L \text{ and } N^Q_{t_i} - N^Q_{t_{i-1}} = 0}$ Clearly, $\widehat{g}(\theta_{t_i}, t_i) = \widehat{g}(\theta^L, t_i)$ and $f(t_i) = 1 = \widehat{f}(\theta^L, 0; t_i)$.

$\underline{\theta_{t_{i-1}} = \theta^H \text{ and } N^Q_{t_i} - N^Q_{t_{i-1}} = 1}$ Evidently, $\widehat{g}(\theta_{t_i}, t_i) = \widehat{g}(\theta^H, t_i)$. Letting τ denote the jump time and calculating (23) yields

$$\begin{aligned} f(t_i) &= \exp\left((\theta^L - \theta^H) \left(t_i - \tau - \frac{2(e^{-\kappa(t_i-\tau)} - e^{-\kappa(\triangle_i)}) + e^{-\kappa(T-t_{i-1})} - e^{-\kappa(T-\tau)}}{\kappa} \right) \right) \\ &:= \exp(X). \end{aligned}$$

Since $\tau \sim U([t_{i-1}; t_i])$ we have $\mathbf{E}^Q(\tau) = (t_i + t_{i-1})/2$ and $\mathbf{E}^Q(\exp(\kappa\tau)) = (\exp(\kappa t_i) - \exp(\kappa t_{i-1}))/(\kappa\triangle_i)$, hence we can find $\mathbf{E}^Q(X)$ and use the first order approximation $\mathbf{E}^Q(\exp(X)) \approx \exp(\mathbf{E}^Q(X))$. This gives the approximation

$$\begin{aligned} \widehat{f}(\theta^H, 1; t_i) &= \exp\Big((\theta^L - \theta^H) \Big(\frac{\triangle_i}{2} - \frac{1}{\kappa} \Big(\frac{2(1 - e^{-\kappa\triangle_i} - \kappa\triangle_i e^{-\kappa\triangle_i})}{\kappa\triangle_i} \\ &\qquad + e^{-\kappa T} \Big(e^{\kappa t_i} - \frac{e^{\kappa t_i} - e^{\kappa t_{i-1}}}{\kappa\triangle_i} \Big) \Big) \Big) \Big). \end{aligned}$$

$\underline{\theta_{t_{i-1}} = \theta^L \text{ and } N^Q_{t_i} - N^Q_{t_{i-1}} = 1}$ Similar to the $(\theta^H, 1)$-case, but now the the approximation becomes

$$\widehat{f}(\theta^L, 0; t_i) = \exp\Big((\theta^L - \theta^H) \Big(\frac{\triangle_i}{2} + \frac{e^{-\kappa T}}{\kappa} \Big(\frac{e^{\kappa t_i} - e^{\kappa t_{i-1}}}{\kappa\triangle_i} - e^{\kappa t_i} \Big) \Big) \Big).$$

All in all, this leads us to the recursive formulae (accurately, it might be called $\widehat{\widehat{g}}$ since it is an approximation to $\widehat{g}$ - but it is still a first order approximation to g)

$$\begin{aligned} \widehat{g}(\theta^L, t_{i-1}) &= Q(N^Q_{t_i} - N^Q_{t_{i-1}} = 0)\widehat{g}(\theta^L, t_i)\widehat{f}(\theta^L, 0; t_i) + Q(N^Q_{t_i} - N^Q_{t_{i-1}} = 1)\widehat{g}(\theta^H, t_i)\widehat{f}(\theta^L, 1; t_i) \\ \widehat{g}(\theta^H, t_{i-1}) &= Q(N^Q_{t_i} - N^Q_{t_{i-1}} = 0)\widehat{g}(\theta^H, t_i)\widehat{f}(\theta^H, 0; t_i) + Q(N^Q_{t_i} - N^Q_{t_{i-1}} = 1)\widehat{g}(\theta^L, t_i)\widehat{f}(\theta^H, 1; t_i) \end{aligned}$$

where $Q(N^Q_{t_i} - N^Q_{t_{i-1}} = k) = \exp(-\lambda\triangle_i)(\lambda\triangle_i)^k/k!$ for $k = 0, 1$.

C Proof of Proposition 5

We note that $r_{T_E} = a(T_E) + b(T_E)$ and use Equation (14) from Proposition 3 to get

$$\begin{aligned} P(T_E, T_M) &= \exp(m(\theta^L; T_E, T_M) - n(T_E; T_M)a(T_E)) \\ &\qquad \times \exp(-n(T_E, T_M)b(T_E))\xi(\theta(T_E); T_M - T_E). \end{aligned}$$

At time 0 the call-option price is

$$\text{call} = \mathbf{E}^Q \left(e^{-\int_0^{T_E} a(s)ds} e^{-\int_0^{T_E} b(s)ds} \left(P(T_E, T_M) - K\right)^+ \right),$$

from which we get

$$\begin{aligned} \text{call} &= \mathbf{E}^Q \Bigg(e^{-\int_0^{T_E} a(s)ds} e^{-\int_0^{T_E} b(s)ds} e^{-n(T_E,T_M)b(T_E)} \xi(\theta(T_E); T_M - T_E) \\ &\qquad \times \left(P_V^{\theta^L}(a(T_E); T_E, T_M) - K/\xi(\theta(T_E); T_M - T_E) \right)^+ \Bigg). \end{aligned}$$

Let $(\mathcal{F}_t^\theta)$ denote the filtration generated by (θ_t). We then condition on $\mathcal{F}_{T_E}^\theta$ and arrive at

$$\begin{aligned} \text{call} &= \mathbf{E}^Q \Big(e^{-\int_0^{T_E} b(s)ds} e^{-n(T_E,T_M)b(T_E)} \xi(\theta(T_E); T_M - T_E) \\ &\qquad \times \mathbf{E}^Q \left(e^{-\int_0^{T_E} a(s)ds} (P_V^{\theta^L}(a(T_E); T_E, T_M) - K/\xi(\theta(T_E); T_M - T_E))^+ | \mathcal{F}_{T_E}^\theta \right) \Big) \end{aligned}$$

Because $a(T_E)$ and $\int_0^{T_E} a(s)ds$ are independent of $\mathcal{F}_{T_E}^\theta$ and $\xi(\theta(T_E); T_M - T_E)$ is $\mathcal{F}_{T_E}^\theta$-measurable we can apply the useful rule (cf. Hoffmann-Jørgensen (1994), (6.8.14)) to calculate the inner expectation; it reduces to an unconditional expectation where $\xi(\theta(T_E); T_M - T_E)$ is treated as known. But this yields that call-option price in the θ^L-Vasicek model and the proposition follows.

References

Andersen, T. G. & Lund, J. (1996), Stochastic volatility and mean drift in the short rate diffusion: sources of steepness, level and curvature in the yield curve, Working Paper. Aarhus School of Business.

Balduzzi, P., Das, S., Foresi, S. & Sundaram, R. (1998), Stochastic mean models of the term structure, Forthcoming in Advanced Tools for the Fixed-Income Professional, edited by N. Jegadeesh and B. Tuckman, Wiley.

Björk, T. (1998), *Arbitrage Theory in Continuous Time*, 1 edn, Oxford.

Chan, K. C., Karolyi, G. A., Longstaff, F. A. & Sanders, A. B. (1992), 'An empirical comparison of alternative models of the short-term interest rate', *The Journal of Finance* **47**(3), 1209–1227.

Conley, T. G., Hansen, L. P., Luttmer, E. G. J. & Scheinkmann, J. A. (1997), 'Short-term interest rates as subordinated diffusions', *Review of Finacial Studies* **10**, 525–578.

Cox, J. C., Ingersoll, J. E. & Ross, S. A. (1985), 'A theory of the term structure of interest rates', *Econometrica* **53**, 385–407.

Hoffmann-Jørgensen, J. (1994), *Probability with a View Toward Statistics*, Vol. 1, Chapman & Hall.

Hull, J. & White, A. (1990), 'Pricing interest rate derivative securities', *Review of Financial Studies* **3**, 573–592.

Jacod, J. & Shiryaev, A. N. (1987), *Limit Theorems for Stochastic Processes*, Springer-Verlag.

Jamshidian, F. (1989), 'An exact bond option formula', *Journal of Finance* **44**, 205–209.

Landén, C. (1998), Bond pricing in a regime-switching short rate model, Working paper. Royal Institute of Technology, Stockholm.

Longstaff, F. A. & Schwartz, E. (1992), 'Interest rate volatility and the term structure: a two-factor general equilibrium model', *Journal of Finance* **47**, 1259–1282.

Musiela, M. & Rutkowski, M. (1997), *Martingale Methods in Financial Modelling*, 1. edn, Springer-Verlag.

Naik, V. & Lee, M. H. (1994), The yield curve and bond option prices with dicrete shifts in economic regimes, Working Paper. University of British Columbia.

Press, W., Teukolsky, S., Vetterling, W. T. & Flannery, B. (1992), *Numerical recipes in C*, 2. edn, Cambridge University Press.

Schlögl, E. & Sommer, D. (1997), Factor models and the shape of the term structure, Discussion paper B-395. University of Bonn.

Shirakawa, H. (1991), 'Interest rate option pricing with poisson-gaussian forward rate curve processes', *Mathematical Finance* **1**(4), 77–94.

Vasicek, O. (1977), 'An equilibrium characterization of the term structure', *Journal of Financial Economics* **5**, 177–188.

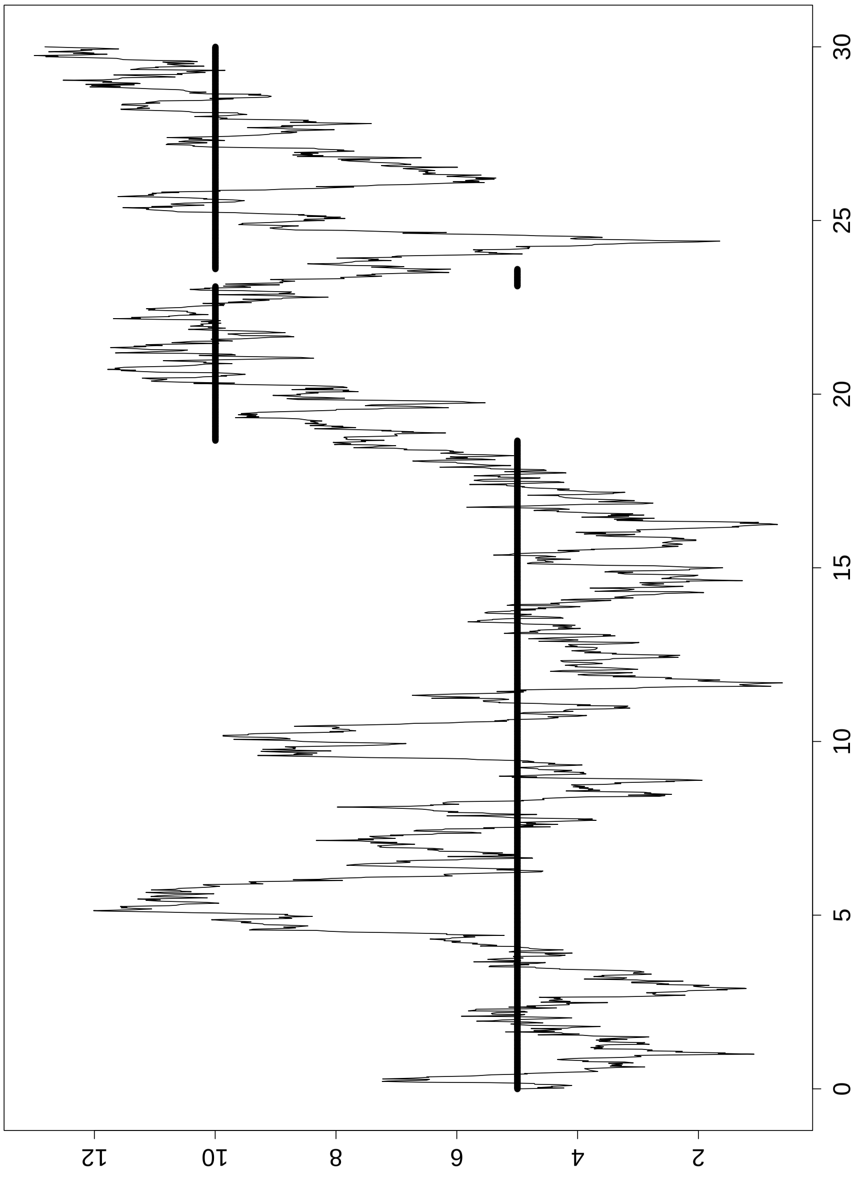

Figure 1: A simulated sample path of the short rate process. Time measured in years, $\theta^L = 5$, $\theta^H = 10$, $\kappa = 1$, $\sigma = \sqrt{10}$ and jumps expected every 7 years.

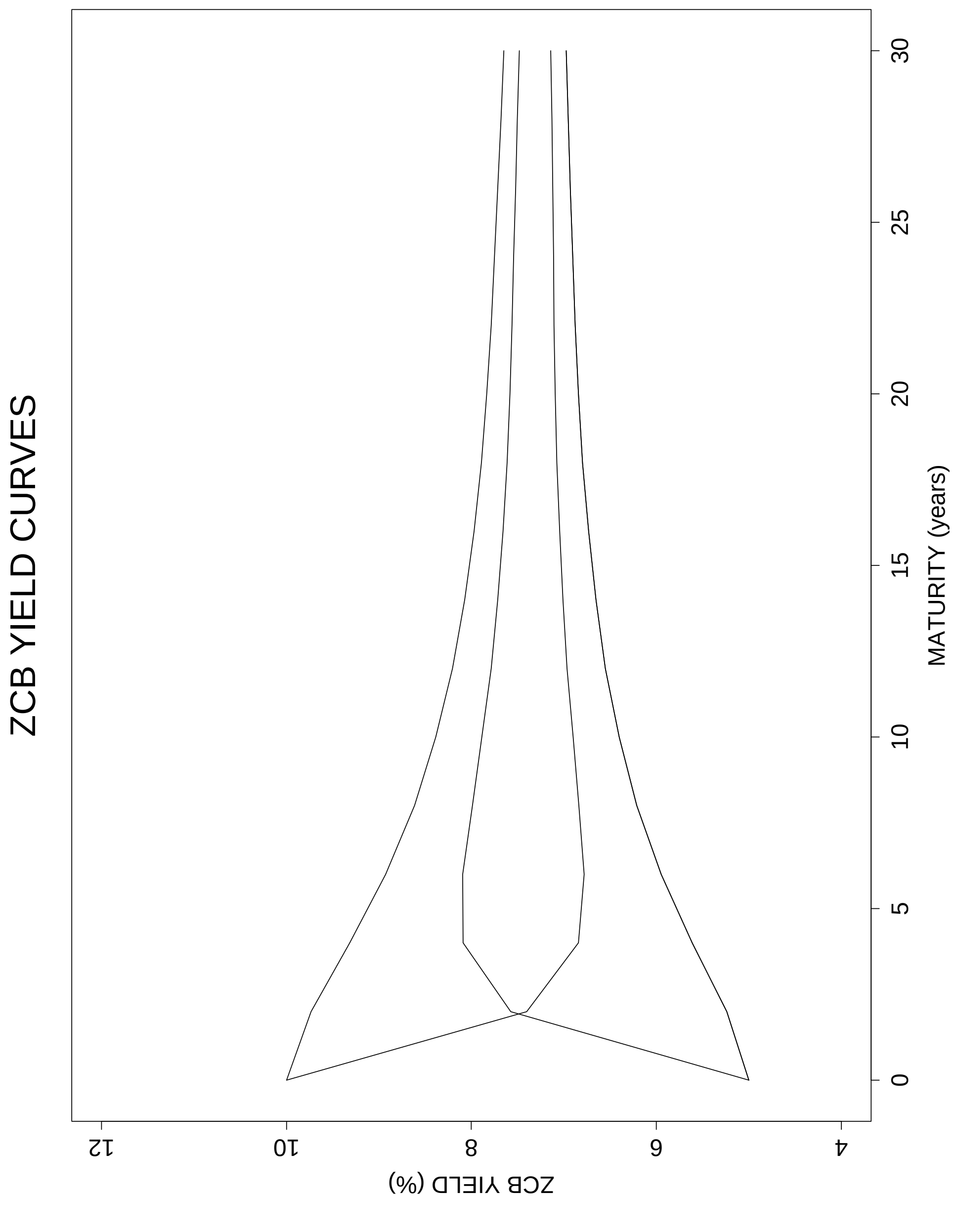

Figure 2: Yield curves with $c \equiv 0$ (i.e. zero Wiener risk premia), $\theta^L = 0.05$, $\theta^H = 0.10$, $\kappa = 1$, $\sigma = 0.02$, $r_0 = 0.05$ and $\lambda^{Q^N} = 7$ as parameters, and $r_0 = 0.05$, 0.10, $\theta_0 = 0.05$, 0.1 as values of the state variables.

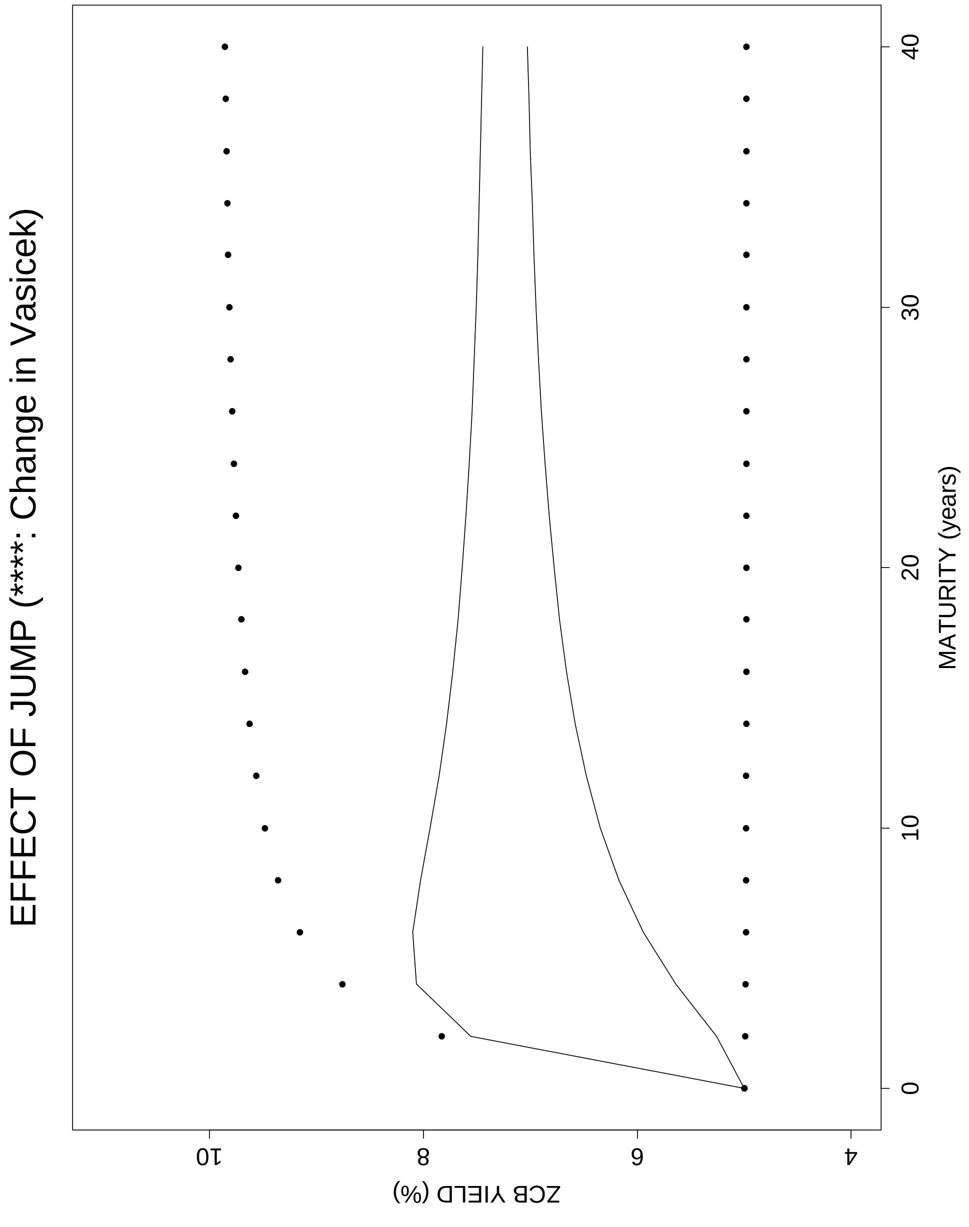

Figure 3: Fully drawn curves: Yield curves immediately before and after a jump. Again, $c \equiv 0$, $\theta^L = 0.05$, $\theta^H = 0.10$, $\kappa = 1$, $\sigma = 0.02$, $r_0 = 0.05$ and $\lambda^{Q^N} = 7$. Dotted curves: Corresponding effect in the "pure Vasicek" model.

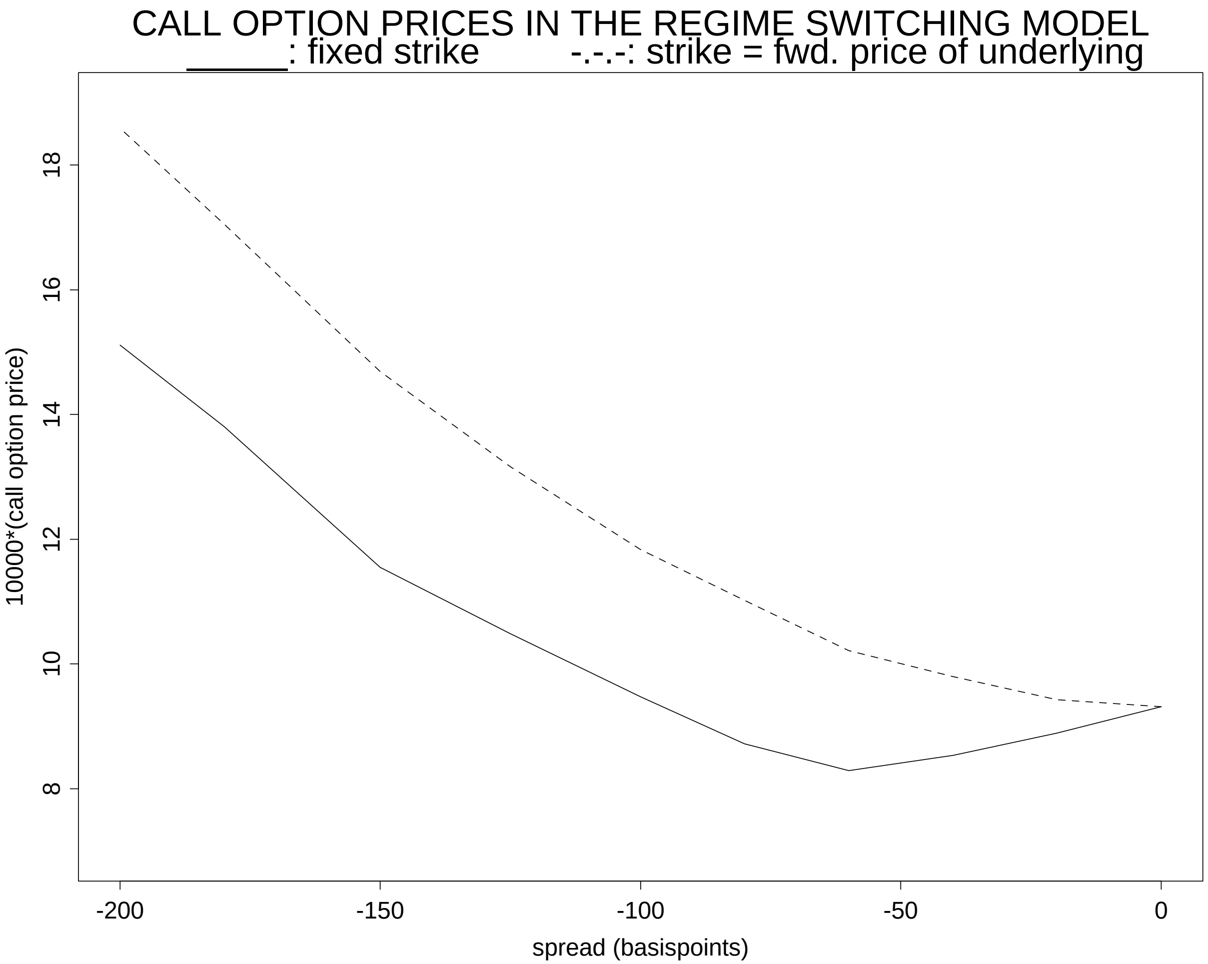

Figure 4: Call option prices with (dotted) and without (full) strike correction.

Numerical method	10-year yield (basispoints)	Uncertainty (± basispoints)	Average runtime (seconds)
Naive simulation	732.8	1.0	60
Smart simulation	732.9	1.0	1.7
Backward induction	733.1	1.0	0.1

Table 1: Comparison of numerical methods. Parameters: $\theta^L = 0.05$, $\theta^H = 0.10$, $\kappa = 1$, $\sigma = 0.02$, $r_0 = 0.05$, and $\lambda = 1$. Runtimes are given on a HP-9000 Unix computer. "Uncertainty" means (half) the width of 95%-confidence intervals for simulation methods and the error estimate produced by the Richardson extrapolation for the backward induction.

Stability of Derivative Prices in Market Models*

Rolf Poulsen
Department of Operations Research
University of Aarhus, Building 530
DK-8000 Aarhus C, Denmark
Phone: +45 8942 3549
email: rolf@imf.au.dk

June 7, 1999

Abstract

An arbitrage free multi-factor term structure model with nonnegative rates is constructed. It takes the current yield curve as input and uses directly observable LIBORs as fundamental building blocks. Closed form approximations to swaption prices are derived. The quality of the approximations is investigated and found to be high. Stability questions relating to the technical construction of the model are also investigated. We find that the problems with pricing of Eurodollar futures contracts that occur in a limiting case of the model do not have any practical effects.

*Written while visiting Marek Musiela at the University of New South Wales, Sydney. Without his ideas and support this paper would not exist. Further, detailed comments from Morten Bai Andersen are acknowledged. The usual disclaimers apply.

1 Introduction and Motivation

Traditional HJM-type term structure models (see Heath, Jarrow & Morton (1992)) focus on continuously compounded instantaneous (forward) rates. Theoretically this convenient, but in market practice such rates play a very little role (try calling your broker to get a quote.) Recently,[1] several authors (Brace, Gatarek & Musiela (1997) and Miltersen, Sandmann & Sondermann (1997) are key references) have shifted the focus (if not the actual framework, asomewhat subtle matter) to "market rates", such as LIBOR or swap rates. Firstly, this puts the emphasis on rates that are actually (and frequently) quoted in markets. Secondly, cunning model-building makes it possible create models that are arbitrage-free (both in a static and a dynamic sense) and in which price formulae of (certain) derivative assets look like the Black-Scholes formula, the *mother of all derivative prices*. Such closed form solutions are very useful since speed is imperative in financial markets. This practice of formulae that are "made to measure" is not without cost. The dynamic specifications that allow different derivatives to be priced in a "Black-Scholes-like" manner are not consistent with each other. Using models that are tailor-made to specific assets has recently become academic as well as market practice, see the discussion in Jamshidian (1997).

In this paper we present a family of LIBOR-models and give the generic derivation of a closed form derivative price. We investigate the consistency problems within the family and some of the approximations that are made to derive the closed form expressions. We also find a limiting case of the family of models and demonstrate that a problem with derivative prices in the limiting model does not effect the prices in any of the approximating models.

The outline of the rest of the paper is as follows. In Section 2 we describe the model and (re)derive some results concerning pricing of interest rate swaps and options on these. The model and some of the theoretical results are so new they are not part of the standard financial vocabulary which justifies the somewhat detailed exposition. In particular we have to set up a notation that facilitates the subsequent analysis. Section 3 looks at the dependence of the closed form expressions on the fundamental

[1] It can also been seen as a successful attempt of formalising what has been market practice for many years. According to Schmidt (1996) (and many others) standard market practice when pricing caps has been to assume lognormal distribution of LIBORs and then use a version of the Black-formula (see Black (1976)).

parameter δ, the compounding period, something that we have not seen discussed previously in the literature. In Section 4 we investigate a drift approximation that is crucial for the derivation of closed-form expression for prices in the model. And having set-up the simulation algorithm we show numerically (for those who do not think this is a contradiction in terms) that a problem with the limit model does not arise for any realistic discretisations of the model. Section 5 concludes, discusses, and outlines topics for future research of both practical and theoretic nature.

2 The Model and Derivation of the Swaption Price

The results in this section are not *original* (see e.g. Brace et al. (1997), Brace & Musiela (1997), Miltersen et al. (1997), and Musiela & Rutkowski (1997*b*)), but they are *new*.

2.1 General Framework and Results

We start by considering a HJM-framework where $P(t,T)$ denotes the zero coupon bond price and we have

$$dP(t,T) = P(t,T)\left(r_t dt - \int_t^{\top} \sigma(t,u)du \cdot d\mathbb{W}(t)\right), \tag{1}$$

where $\mathbb{W}$ is a d-dimensional Brownian motion under the spot martingale measure $\mathbb{P}$, and we work on a filtered probability space $(\Omega, \{\mathcal{F}_t\}, \mathcal{F}, \mathbb{P})$ satisfying "the usual conditions." Note that $\sigma(t,T)$ is the volatility of the instantaneous forward rates. Hence the form of the bond price volatility follows from the definition of instantaneous forward rates ($f(t,T) = -\partial_T \ln P(t,T)$, which also explains the sign convention) and the form of the drift is dictated by absence of arbitrage.

Let (β_t) be the bank account i.e., the value at time t of 1 \$ continuously rolled over in the shortest (instantly maturing, in this case) ZCB. Given T_0 and δ we define $\mathcal{T}(T_0,\delta)$, which we call the *on-beat* dates, to be the set given by

$$\mathcal{T}(T_0,\delta) = \{T_0 + \delta n \mid n \in \mathbb{Z}\},$$

and in the rest of the paper $T_j = T_0 + j\delta$, for $j \geq 0$, unless clearly otherwise stated. For $q \in \mathbb{N}$ we have by a simple arbitrage argument, that the forward price, $F_{T_{j-q}}(t,T_j)$, at time t for delivery at time T_{j-q} of a ZCB maturing at time T_j satisfies

$$F_{T_{j-q}}(t,T_j) = \frac{P(t,T_j)}{P(t,T_{j-q})} := \frac{1}{1 + q\delta K^{(q)}(t,T_{j-q})},$$

where the last equation defines $K^{(q)}(t, T_{j-q})$, called the forward rate over $[T_{j-q}; T_j]$. For an arbitrary T define the measure $\mathbb{P}_T$ on $(\Omega, \mathcal{F}_T)$, called the forward measure, by specifying its Radon-Nikodym derivative to be

$$\frac{d\mathbb{P}_T}{d\mathbb{P}} = \frac{\beta_T^{-1}}{\mathbb{E}(\beta_T^{-1})} = \frac{1}{\beta_T P(0,T)}$$

(evidently, this does define an equivalent probability measure.) When restricted to $\mathcal{F}_t$ ($t \in [0; T]$) the R-N derivative satisfies

$$\frac{d\mathbb{P}_T}{d\mathbb{P}}|_{\mathcal{F}_t} = \mathbb{E}\left(\frac{1}{\beta_T P(0,T)}|\mathcal{F}_t\right) = \frac{P(t,T)}{\beta_t P(0,T)} := \eta_t^T.$$

Using that $d\beta_t = r_t \beta_t dt$, (1), and the Ito formula gives us that (η_t^T) can also be represented as

$$\eta_t^\top = \exp\left(-\int_0^t \int_u^T \sigma(u,s)ds \cdot d\mathbb{W}(u) - \frac{1}{2}\int_0^t |\int_u^T \sigma(u,s)ds|^2 du\right),$$

which by Girsanov's theorem allows us to conclude that $(\mathbb{W}_T(t))$ defined by

$$d\mathbb{W}_T(t) = d\mathbb{W}(t) + \int_t^T \sigma(t,s)dsdt \tag{2}$$

is a BM under $\mathbb{P}_T$. It is now a straightforward exercise to show that $F_{T_{j-q}}(t, T_j)$ is a $\mathbb{P}_{T_{j-q}}$-martingale. Note that for any asset, say with price process S, with drift equal to the short rate under the spot martingale measure (which up to technicalities has to be true for every non-dividend paying asset to prevent arbitrage) the T-forward price, $S(t)/P(t,T)$, is a martingale under the T-forward measure. We note the T-forward price can also be seen as the price when the T-ZCB is used as numeraire. This is the economic intuition behind the different measures, as pointed out in Duffie (1992) (probably the first textbook recognising the computational power of forward measures.)

Moreover, we get that $1/F_{T_{j-q}}(t, T_j)$ is a $\mathbb{P}_{T_j}$-martingale, specifically

$$d\left(\frac{1}{F_{T_{j-q}}(t,T_j)}\right) = \frac{1}{F_{T_{j-q}}(t,T_j)} \int_{T_{j-q}}^{T_j} \sigma(t,u)du \cdot d\mathbb{W}_{T_j}(t), \tag{3}$$

this implying that $K^{(q)}(t, T_{j-q})$ is a $\mathbb{P}_{T_j}$-martingale.

We also introduce the "q or zero" function by

$$qoz(j,q) = q\mathbf{1}_{\{(j \bmod q)=0\}}(j).$$

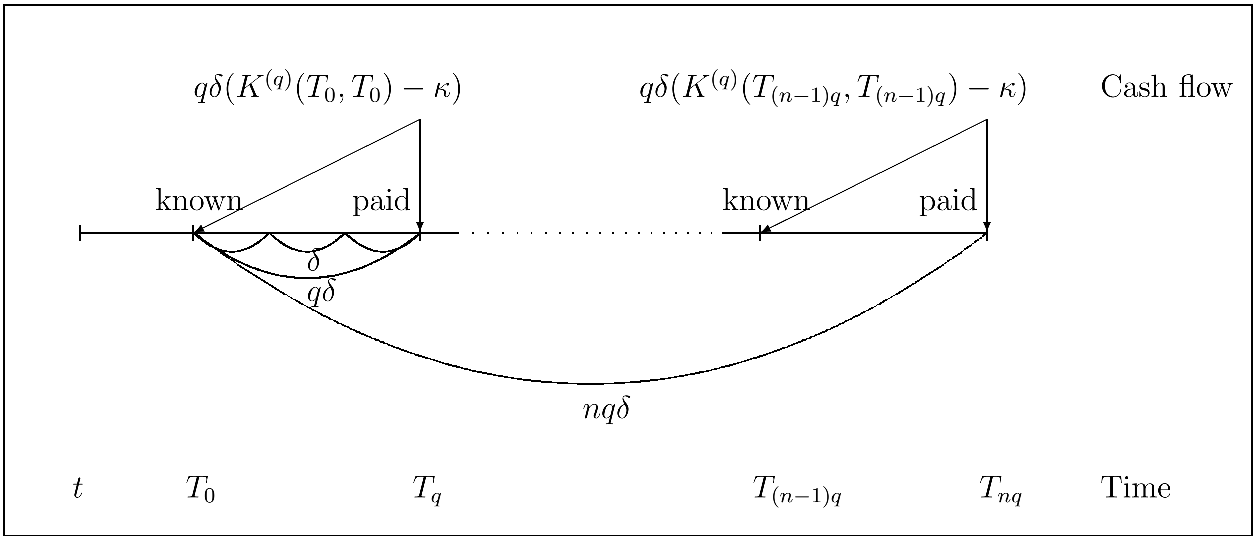

Figure 1: The "time & money"-structure of an interest rate payer swap settled in arrears.

2.2 Interest Rate Swaps

A swap contract is a private agreement between two companies to exchange cash flows in the future after some prearranged formula. This is a very broad definition, we now describe the simplest ("basis", "plain vanilla") interest rate swap. A payer swap settled in arrears exchanges the forward rate $K^{(q)}(T_{j-q}, T_{j-q})$ against a fixed rate κ at n intervals of of length $q\delta$ as shown in Figure 1 i.e., payments are made at dates $T_0 + q\delta, T_0 + 2q\delta, \ldots, T_0 + nq\delta$ (but known one $q\delta$-period in advance.) Thus, for $t \leq T_0$ general arbitrage pricing theory gives us that

$$\mathbf{Pswap}^{(q)}(t) = \mathbb{E}\left(\sum_{j=1}^{nq} qoz(j,q)\delta\frac{\beta_t}{\beta_{T_j}}[K^{(q)}(T_{j-q},T_{j-q})-\kappa] \mid \mathcal{F}_t\right)$$

is the value of the swap. Applying the abstract Bayes' rule, we get

$$\begin{aligned}
\mathbf{Pswap}^{(q)}(t) &= \sum_{j=1}^{nq} qoz(j,q)\delta\frac{\mathbb{E}_{T_j}\left((\eta_{T_j}^{T_j})^{-1}\frac{\beta_t}{\beta_{T_j}}[K^{(q)}(T_{j-q},T_{j-q})-\kappa]|\mathcal{F}_t\right)}{\mathbb{E}_{T_j}\left((\eta_t^{T_j})^{-1}|\mathcal{F}_t\right)} \\
&= \sum_{j=1}^{nq} qoz(j,q)\delta P(t,T_j)\mathbb{E}_{T_j}([K^{(q)}(T_{j-q},T_{j-q})-\kappa]|\mathcal{F}_t) \\
&= \sum_{j=1}^{nq} qoz(j,q)\delta P(t,T_j)[K^{(q)}(t,T_{j-q})-\kappa],
\end{aligned}$$

where the last equality follows by the martingale property.

Time	Action	Net cash flow
0	Sell short 1 T_j-ZCB Buy 1 T_{j-1}-ZCB	$P(0,T_j) - P(0,T_{j-1})$
T_{j-1}	Use the pricipal received from the T_{j-1}-ZCB to buy $1/P(T_{j-1},T_j)$ T_j-ZCBs	0
T_j	Close position	$1/P(T_{j-1},T_j) - 1 = K(T_{j-1},T_{j-1})$

Table 1: A simple trading strategy that pays out the LIBOR.

Introduce the simple forward rate, or the forward LIBOR (if $t = T_{j-1}$ we omit the term "forward") through

$$P(t,T_{j-1}) = (1 + \delta K(t,T_{j-1}))P(t,T_j) \tag{4}$$

Remark Originally, LIBOR was an acronym for London Interbank Offer Rate, an average of several interbank rates. But we shall use LIBOR as a generic term for any rate that has the "add-on" form

$$\text{return} = \text{principal} + \text{investment period} \times \text{interest rate} \times \text{principal},$$

as described by (4). It is convinient/important that this in accordance with market conventions. Note also that, contrary to the instantaneous rates we usually work with in mathematical finance, the LIBOR is in fact a traded asset, as can be seen from the simple trading strategy given i Table 1.

We can recast the swap price as

$$\mathbf{Pswap}^{(q)}(t) = \sum_{j=1}^{nq} \delta P(t,T_j)[K(t,T_{j-1}) - qoz(j,q)\kappa]. \tag{5}$$

The swap rate $\omega^{(q)}$ (corresponding to the specific swap described) is the value of κ that makes the value of the swap 0, i.e.

$$\omega^{(q)}(t) = \frac{P(t,T_0) - P(t,T_{nq})}{\sum_{j=1}^{nq} \delta qoz(j,q)P(t,T_j)}. \tag{6}$$

When dealers talk about "the swap price" they refer to the rate on the fixed leg and will quote it as a spread relative to the yield on a particular bond. This means that we have to be careful when comparing quotes for the fixed and the floating side since the money market (where the LIBOR "lives") and the bond market (where the spread stems from) use different have different conventions, e.g. regargding day count.

A swaption is an option on the swap rate. Specifically, the basis swaption exchanges the time T_0 swap rate against a strike, κ, when $\omega^{(q)}(T_0) \geq \kappa$, thus it has a price determined by

$$\begin{aligned}\mathbf{Pswptn}^{(q)}(t) &= \mathbb{E}\left(\frac{\beta_t}{\beta_{T_0}}\mathbf{Pswap}(T_0)^+|\mathcal{F}_t\right) \\ &= \mathbb{E}\left(\sum_{j=1}^{nq} qoz(j,q)\delta\frac{\beta_t}{\beta_{T_j}}[\omega^{(q)}(T_0)-\kappa]\mathbf{1}_A|\mathcal{F}_t\right),\end{aligned}$$

where $A = \{\omega^{(q)}(T_0) \geq \kappa\} = \{\mathbf{Pswap}^{(q)}(T_0) \geq 0\}$. We have the following equivalent formulations for the swaption price

$$\begin{aligned}\mathbf{Pswptn}^{(q)}(t) &= \mathbb{E}\left(\delta\frac{\beta_t}{\beta_{T_0}}\sum_{j=1}^{nq} qoz(j,q)P(T_0,T_j)[\omega^{(q)}(T_0)-\kappa]\mathbf{1}_A|\mathcal{F}_t\right) &(7)\\ &= P(t,T_0)\mathbb{E}_{T_0}\left(\mathbf{Pswap}(T_0)^+|\mathcal{F}_t\right) &(8)\\ &= \sum_{j=1}^{nq}\delta P(t,T_j)\mathbb{E}_{T_j}\left([K(T_0,T_{j-1})-qoz(j,q)\kappa]\mathbf{1}_A|\mathcal{F}_t\right) &(9)\\ &= \mathbb{E}\left(\frac{\beta_t}{\beta_{T_0}}\left(1-\sum_{j=1}^{nq}C_jP(T_0,T_j)\right)^+|\mathcal{F}_t\right), &(10)\end{aligned}$$

where $C_j = qoz(j,q)\delta\kappa$ for $j = 1, \ldots, nq-1$ and $C_{nq} = 1 + qoz(j,q)\delta\kappa$. (7), which follows by conditioning on $\mathcal{F}_{T_0}$ and using that $P(T_0,T_{j-1}) = \beta_{T_0}\mathbb{E}(\beta_{T_{j-1}}^{-1}|\mathcal{F}_{T_0})$, shows that we can also see the swaption as series of call options on a swap rate, where the options are not separately exercisable. In particular a one-period swaption is a caplet. From (10) we note the similarity with a put option on a coupon bearing bond. (8) is the most compact formulation, while (9) is the most handy expression for analytical computations.

The above formulae are general, in particular the value of the swap and the swap rate are independent of δ provided $q\delta$ is fixed. We have not used in any way assumptions about the functional form of $\sigma(t,T)$, just that the model was arbitrage free.[2] The formulas relating to basis swaps settled in arrears were derived using concepts of

[2] This does put some subtle constraints on $\sigma(t,T)$. We don't really know and we really don't care in this paper.

equivalent martingale measures and forward measures, but may derived using simple trading strategies. A number of market conventions may obscure the clean-cut analysis; different settlement conventions, different day-count conventions on the different legs, etc. Also a huge variety of exotic swaps contracts exist; the floating rate can be almost anything you want (the yield of a particular bond, an average of rates), option-like features, quantos,... So even though basis swaps can be priced by static hedge arguments, there are still good reasons to use a probabilistic framework.

2.3 A Specification and a Swaption Price Formula.

There is not yet consensus about the name the model we work with. A describtive name would be "the lognormal LIBOR model", but we choose to call it "the market model". This term covers other volatility specifications than the one we will give here, but seems to be the preferred one. Models have been developed (conditionally) independently by the authors of Brace et al. (1997) and Miltersen et al. (1997).
To completely specify the model we need to specify the bond price volatility $b(t,T) = -\int_t^T \sigma(t,u)du$ for all t and T (or at least for $T \geq t$.) We consider a specification where $\sigma(t,T) = 0$ for $T \leq t+\delta$ (this is a perhaps not totally realistic but quite convenient initial condition) and

$$\int_T^{T+\delta} \sigma(t,u)du = \frac{\delta K(t,T)}{1+\delta K(t,T)}\gamma(t,T), \quad \text{for } T \geq t+\delta,$$

where $\gamma : \mathbb{R}^2 \mapsto \mathbb{R}^d$ is deterministic. This completely specifies the model since

$$b(t,T) = -\sum_{k=1}^{\lfloor (T-t)/\delta \rfloor} \frac{\delta K(t,T-\delta k)}{1+\delta K(t,T-\delta k)}\gamma(t,T-\delta k).$$

Specifically, we get using the results in the previous section that our main objects of interest are ($\mathbb{P}_{T_j}$-)nicely behaved

$$dK(t,T_{j-1}) = K(t,T_{j-1})\gamma(t,T_{j-1}) \cdot d\mathbb{W}_{T_j}(t), \tag{11}$$

and thus (explaining why this called a lognormal model)

$$K(T_0,T_{j-1}) = K(t,T_{j-1}) \exp\left(\int_t^{T_0} \gamma(u,T_{j-1}) \cdot d\mathbb{W}_{T_j}(u) - \frac{1}{2}\int_t^{T_0} |\gamma(u,T_{j-1})|^2 du \right). \tag{12}$$

The astute reader may point out that we are cheating a bit here; can we be sure that the SDEs are well-defined? The bonds are specified by means of the LIBORs, who

are themselves specified in terms of the bonds. The answer is yes i.e., the SDEs are all well-defined, see Musiela & Rutkowski (1997*b*).
From (2) we get

$$d\mathbb{W}_{T_j}(t) = d\mathbb{W}_{T_0}(t) + \sum_{l=1}^{j} \frac{\delta K(t, T_{l-1})}{1 + \delta K(t, T_{l-1})} \gamma(t, T_{l-1}) dt,$$

which involves a nasty stochastic Girsanov drift correction. We therefore use a deterministic approximation based on the initial data,

$$\begin{aligned} \int_{T_0}^{T_j} \sigma(t,u) du &\approx \sum_{l=1}^{j} \frac{\delta K(0, T_{l-1})}{1 + \delta K(0, T_{l-1})} \gamma(t, T_{l-1}) && (13) \\ &:= \sum_{l=1}^{j} \mu(T_{l-1}) \gamma(t, T_{l-1}). && (14) \end{aligned}$$

This makes it easy for us to shift computations from one forward measure to another. We now derive a formula for the swaption price. Since we saw that the swaption was similar to a put option on a coupon bond, it is not surprising that the following argument is similar to that given in Jamshidian (1989). Assume WLOG $t = 0$ and recall that the exercise, or cut-off, region for the swaption is

$$\begin{aligned} \{\mathbf{Pswap}(T_0) \geq 0\} &= \left\{ \sum_{j=1}^{nq} \delta P(T_0, T_j) [K(T_0, T_{j-1}) - qoz(j,q)\kappa] \geq 0 \right\} \\ &= \left\{ \sum_{j=1}^{nq} \frac{\delta [K(T_0, T_{j-1}) - qoz(j,q)\kappa]}{\prod_{l=1}^{j} (1 + \delta K(T_0, T_{l-1}))} \geq 0 \right\}, && (15) \end{aligned}$$

so (12) tells us that considering

$$X = (X_j) := \left(\int_0^{T_0} \gamma(u, T_{j-1}) \cdot d\mathbb{W}_{T_j}(u) \right)$$

is the key to solving the swaption pricing problem. Using (14) we see that

$$X \sim N(D, \Delta) \quad \text{under} \quad \mathbb{P}_{T_0},$$

where

$$\Delta = (\triangle_{ij}) := \left(\int_0^{T_0} \gamma^\top(u, T_{i-1}) \gamma(u, T_{j-1}) du \right) \quad (\in \mathbb{R}^{nq \times nq})$$

$$D = (D_j) := \left(\int_0^{T_0} \sum_{l=1}^{j} \mu(T_{l-1}) \gamma^\top(u, T_{j-1}) \gamma(u, T_{l-1}) du \right) = \left(\sum_{l=1}^{j} \mu(T_{l-1}) \triangle_{jl} \right) \quad (\in \mathbb{R}^{nq})$$

Δ is a positive semi-definite $nq \times nq$ matrix of rank d. We consider the rank-1 approximation

$$\Delta \approx \Gamma\Gamma^\top,$$

where Γ is the product of the square root of the largest eigenvalue of Δ and its corresponding eigenvector. We then have $D = (\Gamma_j \sum_{l=1}^{j} \mu(T_{l-1})\Gamma_l) := (\Gamma_j d_j)$, and we can write X_j as

$$X_j = \Gamma_j(z + d_j),$$

where $z \sim N(0,1)$ under $\mathbb{P}_{T_0}$, and the simple forward as

$$K(T, T_{j-1}) = K(0, T_{j-1}) \exp(\Gamma_j(z + d_j) - \frac{1}{2}\Gamma_j^2).$$

This makes it possible to express the cut-off region in terms of a single stochastic variable. To evaluate (8) we have to integrate over the region where

$$J(s) := \sum_{j=1}^{nq} \frac{\delta K(0, T_{j-1})(\exp(\Gamma_j(s + d_j) - \frac{1}{2}\Gamma_j^2) - qoz(j,q)\kappa)}{\prod_{i=1}^{j}(1 + \delta K(0, T_{j-1}) \exp(\Gamma_i(s + d_i) - \frac{1}{2}\Gamma_i^2))} \geq 0.$$

Fortunately, one can show about that J has a unique root (it's easy to see that it has at least one root, the uniqueness follows by showing monotonicity.) This root i.e., the s such that $J(s) = 0$ is often called *Jamshidian's s* (although s is not the best letter in a world where integration dummies flow freely.) Using (9) (rather than (8)) we have

$$\mathbf{Pswptn}^{(q)}(0) = \sum_{j=1}^{nq} \delta P(0, T_j) \mathbb{E}_{T_j} \left([K(0, T_{j-1}) \exp(X_j - \frac{1}{2}\Gamma_j^2) - qoz(j,q)\kappa] \mathbf{1}_{\{X_j \geq \Gamma_j(s+d_j)\}} \right). \tag{16}$$

Because $X_j \sim N(0, \Gamma_j^2)$ under $\mathbb{P}_{T_j}$, this leads to the following formula for $\mathbf{Pswptn}^{(q)}(0)$ by essentially the same calculations as those leading to the Black-Scholes formula (which can be done very compactly, see Appendix A).

$$\mathbf{Pswptn}^{(q)}(0) = \sum_{j=1}^{nq} \delta P(0, T_j)[K(0, T_{j-1})\Phi(h_j) - qoz(j,q)\kappa\Phi(h_j - \Gamma_j)], \tag{17}$$

where $h_j = \Gamma_j - s - d_j$ and Φ is the standard normal distribution function. Though it is approximate such a closed form solution is important . Firstly, it is fast to evaluate, which is important if way are going to use the formula "backwards" to get out implied volatility. Secondly, we find hedge strategies much better than from a simulation approach (typically the latter would not give any hedging iformation.)

Note that this formula depends only on quantities corresponding to on-beat dates (and can thus be hedged using only these assets and one other ZCB.) The devil's advocate would say this in another way; that the formula can only be used for contracts with on-beat payments. Either way, to be on solid HJM-ground we had to assume the existence of a full double continuum of ZCB prices. But it does not seem surprising that this assumption can be relaxed (see Musiela & Rutkowski (1997a) and Jamshidian (1997).)

Remark If we model $\int_T^{T+\delta} \sigma(t,u)du$ as deterministic then we arrive at a model in which bond prices become lognormal, and forward rates are not precluded from being negative. In models with this specification, the above calculations virtually all carry over, some even as exact results not approximations, with $\delta K(t, T_{j-1})$ substituted by $1 + \delta K(t, T_{j-1})$. See Brace & Musiela (1997) and Rutkowski (1997). Another way of deriving closed form solutiopns for swaption prices is by noting (showing) that the swap rate is a martingale under the measure corresponding to an annuity bond as numeraire and the specifying dynamics under this measure.

Remark A computer implementation implementation of the model is pretty straightforward, at least if one makes use of the routines from Press, Teukolsky, Vetterling & Flannery (1992). One thing should be noted: Since we are going to work with small values of δ, the Δ-matrix, of which we have to find eigenvalues and eigenvectors, becomes rather large. Because standard routines for finding (and sorting) eigenvalues are cubic in the size of the matrix, this becomes a very time-consuming part of the process. However much faster algorithms can be applied if we remember that we only need the (few) largest eigenvalue(s) (and corresponding vector(s).)

3 Dependency on the Compounding Period

This section works with the previously described lognormal model, but many of the issues are just the same for the normal version mentioned in the **Remark**.

To apply the formulae to a particular swaption or other asset, we can simply choose a T and a δ that makes all payments *on-beat*, and then apply the relevant formula. But different δ's will give different models. In particular, a swaption paying at 6-months intervals will not have the same price in (the natural) "$\delta = 0.5$"-model as in a "$\delta = 0.25$"-model. In this section we investigate how the (approximate) derivative prices depend the choice on δ.

3.1 Swaption Prices

Consider the lognormal model with:

- Flat initial term structure at 10 %.
- One-dimensional flat volatility at 20 % i.e., $\gamma \equiv 0.2$.

Let

$$\mathbf{Swptn}(T, \delta, \kappa, q, n)$$

denote the time 0 swaption price in the lognormal δ-model of a payer swaption written on a swap settled in arrears with strike κ and cash flows at times $T + \delta q$, $T + 2\delta q$, ..., $T + n\delta q$. The "moneyness" is based on comparison between κ and the current swap rate, $\omega(0)$ (where the notational dependence on q has been dropped because we keep δq fixed).

Figure 2 depicts the functional relationship between δ and caplet prices, specifically

$$\left(\delta\ ,\ 100\mathbf{Swptn}(0.5, \frac{0.25}{i}, \kappa, i, i)\right)$$

for $i = 1, 2, \ldots, 64$ and $\kappa = 0.10126$, which means that that caplet is at-the-money, with the usage of the term given above. Figure 3 depicts

$$\left(\delta\ ,\ 100\mathbf{Swptn}(1, \frac{1}{i}, \kappa, i, 3i)\right)$$

for $i = 1, 2, \ldots, 64$ and $\kappa = 0.1051$, which makes this swaption at-the-money.

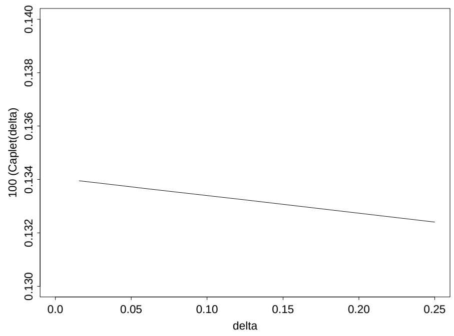

Figure 2: A-t-m-caplet prices as a function of δ in the lognormal model.

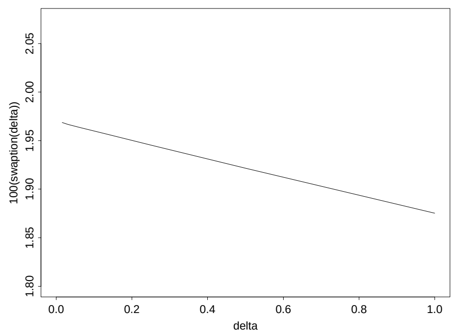

Figure 3: A-t-m-swaption prices as a function of δ in the lognormal model.

Once the computer programs have been made, we can vary the parameters to our hearts content and our readers dismay. The results can be summarised as follows.

1. For the at-the-money caplet as in Figure 2 the relative change in price when going from $\delta = 1/4$ to $\delta = 1/64$ is 1.2 %. Going from $\delta = 1$ to $\delta = 1/64$ for the at-the-money Figure 3-swaption changes the price by 5.0 %. The (perhaps more practically relevant) difference between "$\delta = 1/2$" and "$\delta = 1/4$" is around 1.2 %. Differences of this magnitude suggest that is might be advantageous to take into account the δ-effect when calibrating to, say, caplets in models with different δ's.

2. The prices are decreasing in δ. When the initial term structure is flat and γ is constant then it is not hard to show that the LHS of (13) increases as δ tends to 0 - but it does require cumbersome notation since $K(0, \cdot)$, j and T_{l-1} all implicitly depend on δ. We are basically approximating the integral of an increasing function by left sums. This means higher bond volatility, hence higher option prices.

3. The price is very close to being linear in δ. This is nice, we like linear functions ! In particular it means that the model is well-behaved (at least with respect to swaption prices) when $\delta \to 0$. Concerns about the stability of a model with lognormal continuously compounded rates was what originally lead to the

formulation of discrete-tenor models (that, and of course the fact that (semi-) closed form expressions for derivatives can be found).

4. The price is most δ-sensitive when the swaption is at-the-money, which again is what one would conjecture.

Remark The analysis was carried out under the assumption of flat yield curve and volatility structure. As long we do not look at too pathological examples the dependency on δ is smooth and not to large As an example of particular interest, the δ-dependency becomes less significant but still almost linear in a model with exponentially decaying volatility. Again tedious manipulation with the LHS of (13) can justify this. The "left sum integral approximation" is now of a product of an increasing $(K/(1+\delta K))$ and a decreasing (γ) function, so we don't know the sign of the error. Exponentially decaying volatility is often seen empirically, and as we will see in Subsection 3.2, it corresponds to mean reversion effect in (short) rates.

Remark The previous analysis was carried out under the assumption that payments were made in arrears at exactly equidistant dates. In real life market practice, weekends, holidays and the sheer construction of our calendar is bound to interfere with this. Step -1 towards remedying this is saying, as loud as you can, "I don't know & I don't care - and besides it's trivial". The former constitute legal points of view, the latter is not necessarily true, see Rutkowski (1997). Step 0 is finding out exactly when cash flows occur for a given contract, an equally tedious and important task. The next logical step is then to use the current term structure to construct an approximate standardised contract (e.g. adjust the κ slightly), and then use a version of our pricing formulae. Since the payments are random, this is only a first order approximation. This subject seems too tedious to be treated in the literature, with Schmidt (1996) as an exception. But it is important, especially because swap markets are very large and competitive, so "we need all we can get."

3.2 A Limiting Case of the Market Model

Let $K_\delta(t,T_i)$ denote the forward δ-LIBOR i.e., $K_\delta(t,T_i) = (P(t,T_i)/P(t,T_i+\delta) - 1)/\delta$. With $f(t,T_i)$ denoting the continuously compounded forward rate we have

$$K_\delta(t,T_i) \quad = \quad \frac{1}{\delta}(\underbrace{\exp(\int_{T_i}^{T_i+\delta} f(t,u)du)}_{:=l(\delta)} -1 \ )$$

$$
\begin{aligned}
&= \frac{l(\delta) - l(0)}{\delta} \\
&\overset{\delta \to 0}{\to} l'(0) \\
&= f(t, T_i),
\end{aligned}
$$

as, of course, it should be. Recall that $K_\delta(t, T_i)$ is modeled as being a lognormal martingale under the forward measure corresponding to the date $T_i + \delta$ i.e.,

$$
K_\delta(t, T_i) = K_\delta(0, T_i) + \int_0^t \gamma(u, T_i) K_\delta(u, T_i) d\mathbb{W}_{T_i+\delta}(u)\,, \quad \text{for } T_i \in \mathcal{T}(T_0, \delta), \tag{18}
$$

where $\gamma(\cdot, \cdot)$ is a deterministic function and $\mathbb{W}_{T_i+\delta}$, the BM under $\mathbb{P}_{T_i+\delta}$, is assumed to be one-dimensional. Because we want to consider the case of $\delta \to 0$, we are not immensely keen on having expressions involving a δ-dependent BMs. To fix this, recall that the BM under the spot martingale measure is linked to the BM under $\mathbb{P}_{T_i+\delta}$ through

$$
d\mathbb{W}(t) = d\mathbb{W}_{T_i+\delta}(t) - \int_t^{T_i+\delta} \sigma(t, u) du dt.
$$

Our specification of the market model links $\sigma(\cdot, \cdot)$ to $\gamma(\cdot, \cdot)$ through the equations

$$
\int_T^{T+\delta} \sigma(t, u) du = \frac{\delta K_\delta(t, T)}{1 + \delta K_\delta(t, T)} \gamma(t, T) \tag{19}
$$

Let τ_i's be the numbers of the form

$$
\tau_i = T_0 - (m_\delta - i)\delta\,, \quad \text{for } i = 1, 2, \ldots, (m+n)q,
$$

where $m_\delta = \lceil (T_0 - t)/\delta \rceil$. We can then write

$$
\int_t^{T_i+\delta} \sigma(t, u) du = \sum_{j=1}^{m_\delta+i+1} \frac{\delta K_\delta(t, \tau_j)}{1 + \delta K_\delta(t, \tau_j)} \gamma(t, \tau_j) + \mathrm{o}(\delta).
$$

Because $K_\delta(t, \tau_j)/(1+\delta K_\delta(t, \tau_j)) \overset{\delta \to 0}{\to} f(t, \tau_j)$, we have (pick your favourite convergence theorem)

$$
\sum_{j=1}^{m_\delta+i+1} \frac{\delta K_\delta(t, \tau_j)}{1 + \delta K_\delta(t, \tau_j)} \gamma(t, \tau_j) \overset{\delta \to 0}{\to} \int_t^{T_i} \gamma(t, u) f(t, u) du.
$$

Substituting into (18) we get

$$
\begin{aligned}
K_\delta(t, T_i) &= K_\delta(0, T_i) - \int_0^t \left(\sum_{j=1}^{m_\delta+i+1} \frac{\delta K_\delta(u, \tau_j)}{1 + \delta K_\delta(u, \tau_j)} \gamma(t, \tau_j) \right) \gamma(u, T_i) K_\delta(u, T_i) du - \int_0^t \mathrm{o}(\delta) du \\
&\quad + \int_0^t \gamma(u, T_i) K_\delta(u, T_i) d\mathbb{W}(u).
\end{aligned}
$$

Letting $\delta \to 0$, and merrily interchanging limits and integrals, we get the following as the δ-limiting case of the market model:

$$f(t,T_i) = f(0,T_i) - \int_0^t \gamma(s,T_i) \underbrace{\left(\int_s^{T_i} \gamma(s,u) f(s,u) du\right)}_{:=X_s^{T_i}} f(s,T_i) ds + \int_0^t \gamma(s,T_i) f(s,T_i) d\mathbb{W}(s),$$

or

$$df(t,T_i) = -\gamma(t,T_i) X_t^{T_i} f(t,T_i) dt + \gamma(t,T_i) f(t,T_i) d\mathbb{W}(t).$$

Note that this is, as it should be, exactly what we would get if we specified the forward rate volatility as $\gamma(t,T)f(t,T)$ in a classical HJM-formulation, and then applied the drift restriction.

By the Ito formula we get

$$d\ln f(t,T_i) = -(\gamma(t,T_i) X_t^{T_i} + \frac{1}{2}\gamma^2(t,T_i)) dt + \gamma(t,T_i) d\mathbb{W}(t).$$

The SDE for $\ln f(t,T_i)$ still contains $X_t^{T_i}$, which is a rather nasty integral over the yield curve. See Miltersen (1994) for a further discussion of of such SDEs.

Now, more specifically, assume that $\gamma(\cdot,\cdot)$ is multiplicatively separable,

$$\gamma(t,T) = g(t)h(T),$$

where g and h are positive functions. We want to look at the short rate r, which is defined by $r_t = \lim_{\delta \to 0} f(t, t+\delta)$. Letting $Y_t^T := \int_t^T h(u) f(t,u) du$, integrating, and setting $T = t$, we get

$$\ln r_t = \underbrace{\ln f(0,t) - \int_0^t g^2(s) h(t) (Y_s^t + \frac{1}{2} h(t)) ds}_{:=d(t)} + h(t) \int_0^t g(s) d\mathbb{W}(s).$$

Note that d is not a Markov process. This gives the following dynamics of $\ln r_t$

$$\begin{aligned} d\ln r_t &= d'(t)dt + h'(t) \underbrace{\int_0^t g(s) d\mathbb{W}(s)}_{=\frac{\ln r_t - d(t)}{h(t)}} dt + h(t) g(t) d\mathbb{W}(t) \\ &= (d'(t) - d(t)\frac{h'(t)}{h(t)} + \frac{h'(t)}{h(t)} \ln r_t) dt + g(t)h(t) d\mathbb{W}(t). \end{aligned} \tag{20}$$

So, if $h' < 0$ i.e., volatility is decreasing in maturity, then we have a linear mean reversion feature in the log of the short rate, albeit still to a nasty stochastic process, $h(t)d'(t)/h'(t) - d(t)$. If we approximate in similar fashion as previously,

$$X_t^{T_i} \approx \int_t^{T_i} \gamma(t,u) f(0,u) du,$$

then $d(t)$ becomes deterministic, $f(t, T_i)$ becomes lognormal, and we have that r follows an Ornstein-Uhlenbeck process in logs. The Black-Karasinski-model (see Black & Karasinski (1991)), in other words. If $\gamma(\cdot, \cdot)$ is exponentially decaying in time to maturity ,

$$\gamma(t, T) = \sigma \exp(-\kappa(T - t)),$$

where σ and κ are positive, then the volatility of the log of the short rate becomes time-homogeneous, $d(t) + d'(t)/\kappa$ determines the level that the log of the short rate locally reverts to, and κ is the speed of mean reversion.

Remark We speak quite uninhibitly about "convergence", without formally defining in which sense it is to be understood. For the sheer short rate considerations, the "correct" sense is "weak convergence of stochastic processes". We basically established convergence of the local first and second moment (the drift and volatility), which is about half-way there; we need a uniformity condition on the "discrete" process, something like "jumps vanishing uniformly on compacts". But one should be aware that "weak convergence" is just that, there is no guarantee that objects related to the processes (such as derivative prices and hedge ratios) also exhibit convergence He (1990) and Nelson & Ramaswamy (1989) among others treat the subject in a financial context.

4 Simulation Studies

The swaption price formula was derived using a deterministic approximation to the drift. We now perform a simulation study to check the accuracy of this approximation.

4.1 Swaption Prices

By general risk-neutral pricing and simple manipulations we have the following swaption price

$$\mathbf{Pswptn}(t) = \mathbb{E}\left(\frac{\beta_t}{\beta_T}\mathbf{Pswap}(T)^+ \mid \mathcal{F}_t\right), \tag{21}$$

where $\mathbf{Pswap}(T)$ is given by (5). The version of (21) that is most convenient for obtaining closed form expressions is not the most suitable one for simulation. We have by definition of the bank account, iterated expectations the abstract Bayes' rule that

$$
\begin{aligned}
\mathbf{Pswptn}(t) &= \mathbb{E}_{T_{nq}}\left(\frac{(\eta^{T_{nq}}_{T_{nq}})^{-1}}{(\eta^{T_{nq}}_{t})^{-1}}\frac{\beta_t}{\beta_{T_{nq}}}\frac{\mathbf{Pswap}(T_0)^+}{P(T,T_{nq})} \mid \mathcal{F}_t\right) \\
&= P(t,T_{nq})\mathbb{E}_{T_{nq}}\left(\frac{\mathbf{Pswap}(T_0)^+}{P(T_0,T_{nq})} \mid \mathcal{F}_t\right). \qquad (22)
\end{aligned}
$$

Now consider the market model, and note that we have

$$
d\mathbb{W}_{T_{nq-j}}(t) = d\mathbb{W}_{T_{nq}}(t) - \int_{T_{nq-j}}^{T_{nq}} \sigma(t,u)du dt\ , \quad \text{for } j = 1,\ldots,nq.
$$

We "simulate backwards" using the following algorithm. It works perfectly well also for the case of $d \geq 2$, but results are given only for the truly one-dimensional case.

1. Simulate a sample path of $(\mathbb{W}_{T_{nq}}(t))_0^{T_0}$ under $\mathbb{P}_{T_{nq}}$ i.e., simulate a standard BM. Assuming we want $\mathbb{W}_{T_{nq}}(\tau_i)$ at τ_i's that are ϵ (some small number compared to T_0) apart, then it is just a question of simulating independent $N(0,\epsilon)$-variables and adding these.

2. Using the simulated path of $\mathbb{W}_{T_{nq}}$ we simulate a path of $(K(t,T_{n-1}))_0^T$ using the relevant SDE. The smartest is using the well-known exact solution of in terms of the BM:

   $$
   K(\tau_i,T_{nq-1}) = K(t,T_{nq-1})\exp\left(\left(\int_t^{\tau_i}\gamma(u,T_{nq-1})d\mathbb{W}_{T_{nq}}(u)\right) - \frac{1}{2}\int_t^{\tau_i}\gamma^2(u,T_{nq-1})du\right),
   $$

   which is of course particularly simple in the case of time-homogeneous γ-function i.e., γ is constant in the first argument.

3. Then calculate a simulated path of the Girsanov correction by

   $$
   \int_{T_{nq-1}}^{T_{nq}} \sigma(\tau_i,u)du = \frac{\delta K(\tau_i,T_{nq-1})}{1+\delta K(\tau_i,T_{nq-1})}\gamma(\tau_i,T_{nq-1}),
   $$

   and a simulated path of $(\mathbb{W}_{T_{nq-1}}(t))_0^{T_0}$ through

   $$
   d\mathbb{W}_{T_{nq-1}}(\tau_i) = d\mathbb{W}_{T_{nq}}(\tau_i) - \epsilon\int_{T_{nq-1}}^{T_{nq}} \sigma(\tau_i,s)ds
   $$

   Note that this is $\mathbb{W}_{T_{nq-1}}$ simulated under $\mathbb{P}_{T_{nq}}$.

4. Repeat steps 1.-3., storing relevant results (thus making the algorithm linear in space) until $\mathbb{W}_{T_0}$ has been simulated (again, under $\mathbb{P}_{T_{nq}}$). A relevant result for (22) is $P(T, T_j)$, for $j = 1, \ldots, nq$, which is found by the recursive relation

$$P(T_0, T_{i+1}) = \frac{P(T_0, T_i)}{1 + \delta K(T_0, T_i)},$$

and the initial condition $P(T_0, T_0) = 1$.

5. Calculate

$$\mathbf{Pswptn}_m = P(t, T_{nq}) \frac{\mathbf{Pswap}(T_0)^+}{P(T_0, T_{nq})},$$

by using (5).

6. For anti-thetic variance reduction, repeat steps 1.-5. with $-\mathbb{W}_{T_j}(\tau_i)$ as simulated BM, hence calculating $\mathbf{Pswptn}_{m,A}$.

7. Repeat above steps a large number of times, say M. Calculate the approximate swaption price

$$\mathbf{Pswptn} \approx \mathbf{Pswptn}_{SIM} := \frac{1}{2M} \sum_m \left(\mathbf{Pswptn}_m + \mathbf{Pswptn}_{m,A} \right),$$

the hope being that if M is large and ϵ is small, then this will be an accurate approximation.

Note that for the swaption we only have to simulate the BMs between times 0 and T, not all the way out to T_{nq}.
In this particular case our knowledge of the structure of the model also allows us to do better than the above. To ease already cumbersome notation, we drop t-notational dependence and the "anti-thetic indications" in the following. By using the closed-form expression for the swaption price in the model with deterministic drift, we are able to use a (simple) control variate technique for further variance reduction. At little extra computational effort a simulated swaption price in the model deterministic drift, say $\mathbf{Pswptn}_m^D$, can be calculated for each simulation. With $\mathbf{Pswptn}^D$ denoting the known true swaption price in the deterministic drift model, we then consider the regression equation

$$\mathbf{Pswptn}_m - (\mathbf{Pswptn}_m^D - \mathbf{Pswptn}^D) = \mathbf{Pswptn} + \nu_m,$$

	Naive		Anti-thetic		Control variate	
Swptn (1,1,0.1,1,3)	$\mathbf{Swptn}_{SIM}$	$24.60 \cdot 10^{-3}$ $[35.6 \cdot 10^{-3}]$	$\mathbf{Swptn}_{SIM}$	$24.52 \cdot 10^{-3}$ $[35.5 \cdot 10^{-3}]$	$\mathbf{Swptn}_{SIM}$	$24.50 \cdot 10^{-3}$ $[0.5 \cdot 10^{-3}]$
$\mathbf{Swptn}^D$ (1,1,0.1,1,3)	$\mathbf{Swptn}^D_{SIM}$	$24.65 \cdot 10^{-3}$ $[35.7 \cdot 10^{-3}]$	$\mathbf{Swptn}^D_{SIM}$	$24.57 \cdot 10^{-3}$ $[35.6 \cdot 10^{-3}]$		
M =50000						
$\epsilon = 1/170$	rel. abs. diff.	0.20%	rel. abs. diff	0.20%	rel. abs. diff.*	0.20%

Table 2: Simulation techniques for determination of swaption prices and relative difference between the full model and the model with deterministicly approximated drift. $24.55 \cdot 10^{-3}$ is the true swaption price in the the latter model. [] indicate standard errors, $\tilde{\sigma}$.. *: relative absolute difference compared to the true "det. drift"-swaption price. Current term structure and volatility as in previous sections (flat at 10 and 20 %, respectively).

where the ν_m's are $iid(0, \sigma_\nu^2)$, and the hope is that σ_ν is "small" because the full model and the model with deterministic drift are "close". The regression equation provides us with a new estimate of **Pswptn** and a standard deviation of this. Firstly, this provides an estimate of **Pswptn** that we hope has a lower standard error than the first estimate. Secondly, looking at the change in standard errors of **Pswptn**-estimates, we get an indication of how good our control variate is i.e., how good the (or bad) the model with deterministic drift is. We can do this for any approximation that allows for closed-form swaption price expression (as long the approximate model can be simulated.)

We see from Table 2 that the anti-thetic method produces somewhat better estimates, but the control variate technique is is the way to go, it reduces the standard deviation with roughly a factor 70. Since the size the confidence intervals are approximately $3.92\tilde{\sigma}./\sqrt{M}$ determination of the price with three significant digits requires less than 1000 simulations with the control variate method (run-time: $2-3$ seconds, with $\epsilon = 1/170$), but about 2,000,000 simulations with the naive method (meaning that our potential clients have long left). If anything, this use gives an excellent reason for developing models in which we approximate at an analytic level in order to obtain closed-form expressions.

We also note that for the considered swaption, the underlying of which has payments 2, 3 and 4 years from today, the difference between the full model and the approximation is about 0.2 %, with the "deterministic price" being the higher. There is a

discretisation error in our simulations, so another question is "How should ϵ be chosen?"(note that the number of computations grows linearly in $(T_0 - t)/\epsilon$). Since the main difference between models with different ϵ's are the relative errors between the full and the deterministic drift model, we consider a fixed, large M and calculate the estimate of the aforementioned difference that various ϵ's lead to. In Figure 4 this has been done for our favourite 3-year swaption (same setting as always and $M = 50000$). We see that the model quickly stabilises "in ϵ", which is nice. 360 steps is clearly "overkill", computer-time and -power can be put to much better use.
Figure 5 shows that the δ-linearity also holds quite nicely for the full model. Note however, that the lines in Figure 5 are not parallel, the relative absolute difference in prices increases form 0.20 % to 0.34 % as δ goes from 1 down to 1/64. As Figure 6 shows, the discrepancy between the two models becomes larger as the volatility increases, as one might expect. Figure 7 shows that the longer the underlying swap is, the more imprecise is the deterministic approximation. Certainly, these observations will shake neither the market nor the academia. Or in less poetic fashion: "Just as we would expect."

Remark It would be possible to use other, better approximations to the drift. Sawa (1997) considers and approximation based on the idea that the swaption is exercised at T_0, so he uses an appropriate conditional mean of $K(T_0, \cdot)$. However, based on the results in this section, we doubt that there will be much gain in pursuing this line of research.

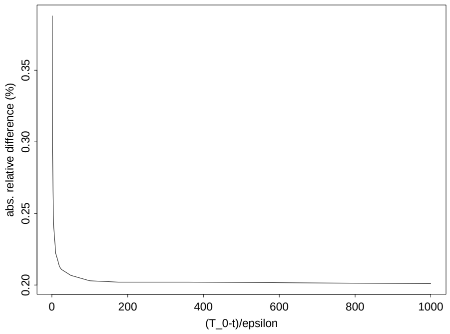

Figure 4: Absolute relative error vs. $(T_0 - t)/\epsilon$ for our favourite swaption.

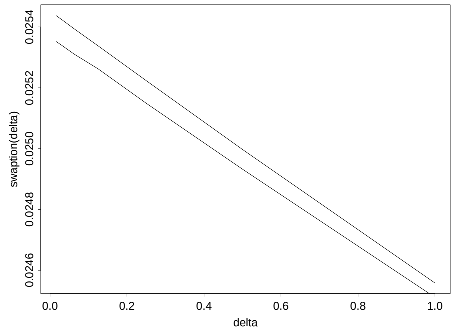

Figure 5: Lower curve: Difference between **Swptn** (1,1/i,0.1,i,3) and **Swptn** (1,1,0.1,1,3) in the full model. Upper curve: Same thing, but for the "det. drift" model.

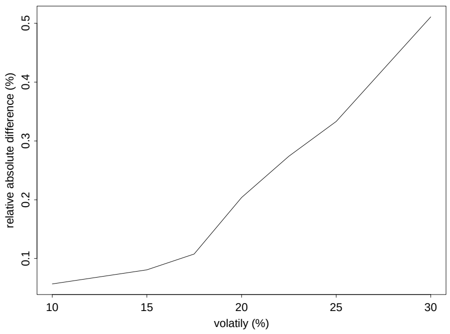

Figure 6: Absolute relative difference between **Swptn** (1,1,0.1,1,3) and $\mathbf{Swptn}^D$ (1,1,0.1,1,3) for different levels of volatility.

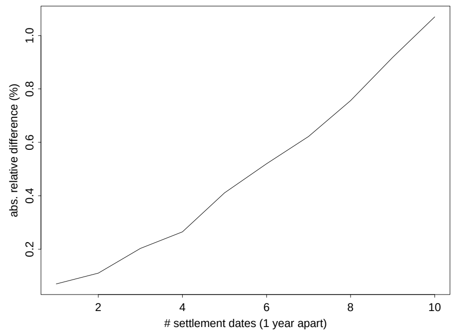

Figure 7: Absolute relative difference between **Swptn** (1,1,0.1,1,n) and $\mathbf{Swptn}^D$ (1,1,0.1,1,n) for different values of n.

4.2 Eurodollar Futures Prices

In Subsection 3.1 we showed that the swaption prices in the model with deterministic drift were well behaved as $\delta \to 0$. In Subsection 4.1 we concluded that the linearity also holds nicely for the swaption prices in the full model. But not all that glitters is gold, and not all derivative assets are (structurally similar to) swaptions. One type type of LIBOR-derivative for which the δ-stability of the market model is particularly interesting to investigate is the Eurodollar[3] futures contract. Why? Well, we showed that the market model converges to a model somewhat like the BK-model as $\delta \to 0$. In fact, for the deterministic approximation, it WAS the BK-model. On the other hand, in Hogan & Weintraub (1993) and Hoffman (1993) it is shown that Eurodollar futures prices are infinite in the BK-model. So something must go awry as $\delta \to 0$, or there is a discontinuity at "$\delta = 0$". In either case, we would like to know.

The conjecture is that for any reasonable δ (say, $>$ a day) the instability is not a problem. Two highly unscientific reasons for this are: (1) a discretised version of the BK-model is the Black-Derman-Toy-model (see Black, Derman & Toy (1990)), if this model produced unreasonable Eurodollar futures prices, Fisher Black and his co-workers would have noticed and not proposed and used the model, and (2) it is NOT easy to show the infinite prices, initial attempts, like using the Jensen inequality, fail (believe us!). But enough sweet-talk, let's get to work.

The Eurodollar futures contract is a futures contract with the LIBOR as underlying asset. By market convention, the Eurodollar futures "price" (as always, we should rather say "index") at maturity, T, is

$$\begin{aligned} U(T,T) &= 1 - \delta_U K_{\delta_U}(T,T) \\ &= 2 - P(T, T+\delta_U)^{-1}. \end{aligned}$$

It is well-known from general arbitrage pricing theory that the price of a continuously marked-to-market futures contract is a martingale under the spot martingale measure $\mathbb{P}$ (Musiela & Rutkowski (1997*b*) takes this as the definition, Duffie (1988) proves it from the specification of the futures contract as an index generating cash flows). This means that

$$U(t,T) = 2 - \mathbb{E}(P(T, T+\delta_U)^{-1} | \mathcal{F}_t),$$

[3] Why this peculiar name? See Hull (1993) or Hoffman (1993). Futures contracts of this type are among the most frequently traded interest rate derivatives.

which we as in the previous subsection would like to express as a $\mathbb{P}_{T+\delta_U}$-mean, because $\mathbb{P}_{T+\delta_U}$ is the natural measure to simulate under. Assuming WLOG $t = 0$, calculations as before gives us that

$$2 - U(0,T) = \mathbb{E}_{T+\delta_U}\left(\frac{P(0,T+\delta_U)\beta_{T+\delta_U}}{P(T,T+\delta_U)}\right).$$

Assume that $\delta_U/\delta \in \mathbb{N}$, and $T/\delta \in \mathbb{N}$, so the current date (0) is *on-beat*, which is NOT completely WLOG. In the δ-market model, the bank account at *on-beat* dates has a value equal to the value of 1 \$ repeatedly rolled over from time 0 in the shortest possible bond i.e.,

$$\beta_{T+\delta_U} = \prod_{i=0}^{\frac{T+\delta_U}{\delta}-1} P(\delta i, \delta(i+1))^{-1},$$

which can easily be recast in terms of δ-LIBORs, thus showing that the essential part is determining

$$\mathbb{E}_{T+\delta_U}\left(\prod_{i=0}^{\frac{T+\delta_U}{\delta}-1}(1+K_\delta(i\delta,i\delta))\prod_{i=\frac{T}{\delta}}^{\frac{T+\delta_U}{\delta}-1}(1+K_\delta(T,i\delta))\right), \tag{23}$$

which is illustrated in painstaking detail in Figure 8. We now have all the information we need to perform simulations along the lines of the previous subsection in order to determine the Eurodollar futures price. Let us stress, that when we say $\delta \to 0$, then this precisely what we mean i.e., δ_U is going nowhere fast !
In Figure 9 we see no indication of δ-instability for the full model. The only indication of some problem is that the confidence intervals are about 15 % broader in the left than in the right hand side of the figure. For the "deterministic drift"-model the picture is much the same, in fact for the given simulation parameters (M, ϵ) the test probability that the prices are equal is extremely high. The design of the experiment can be criticised. We should use a sample size, M, such that we are able to distinguish at some reasonable level of significance between prices that we know are cannot be the same. But this would take roughly forever, in particular for small δ's since the computation time grows quadratically in $1/\delta$.

Remark What the bank account is on *off-beat* dates does not concern us because we have been devious enough to make the current date *on-beat*. However, note that the process $G_t = P(t, T_{\lfloor t/\delta \rfloor+1})\prod_{i=0}^{t/\delta} 1/P(\delta i, \delta(i+1))$ which coincides with the bank account for *on-beat* t's, is not of bounded variation, hence does not qualify for a bank

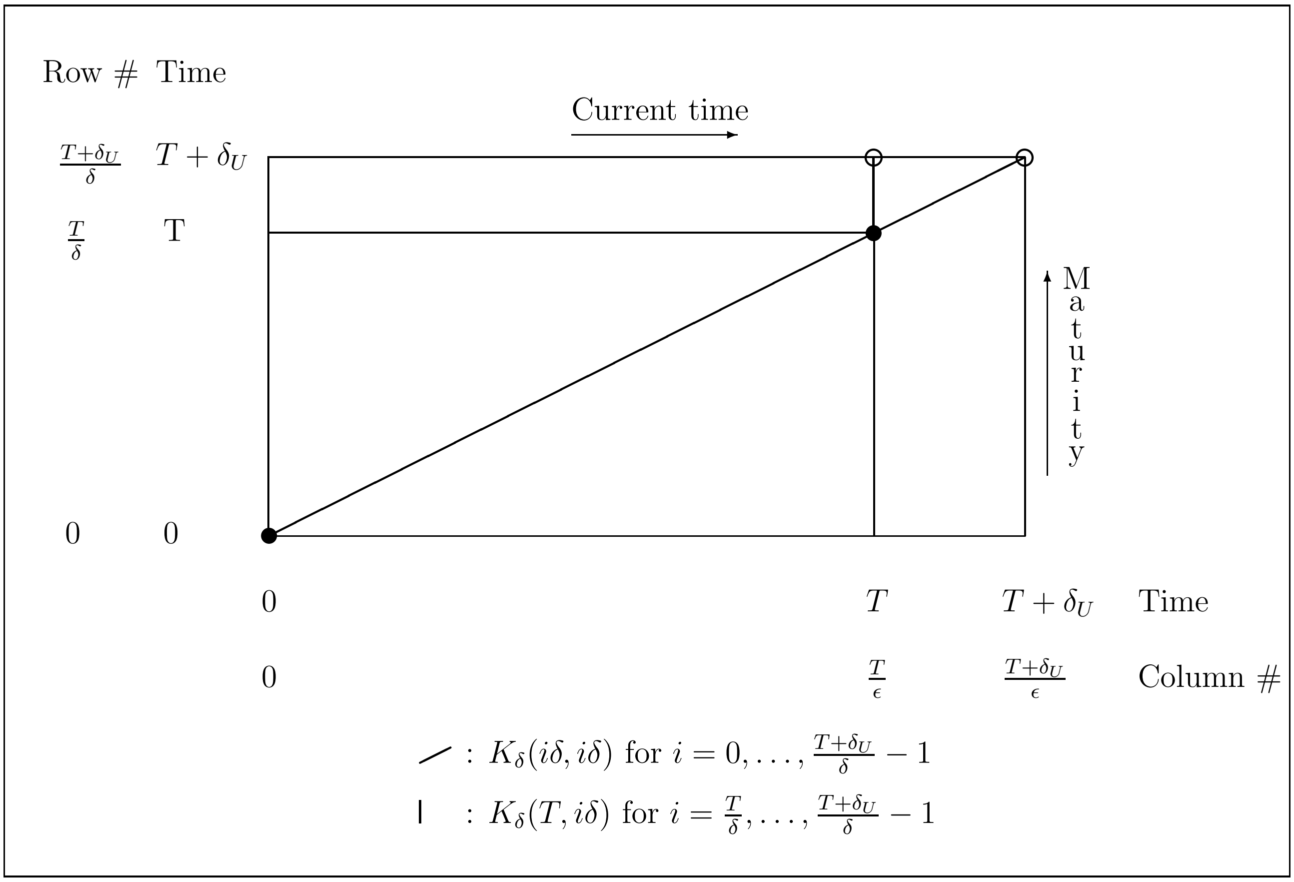

Figure 8: Schematic illustration of the simulation procedure for the Eurodollar future. Not a terribly important figure, but it took several hours to draw it, so now you are forced to look at it.

account. The martingale measure corresponding to G as numeraire is called the spot LIBOR measure and has been studied in Jamshidian (1997).

Remark Because lognormal distributions have moments of any order, it is easy to see by the Hölder inequality, that the mean we are trying to determine in the "deterministic drift" model is in fact finite for any $\delta > 0$. In the this model we know exactly the joint distribution of all the cells in Figure 8 (perceived as a matrix). One could get the impression that it is then "easy" to calculate (23) exactly and pass to the limit. This would tell us whether the model has a "discontinuity at 0", something that would be quite possible. However we have not been able to show a result like this analytically, the problem is basically that the lognormal distributions are not sum-(but product)stable, and we are dealing with quantities like "$\prod^n(1 + \log\mathrm{N}/n)$".

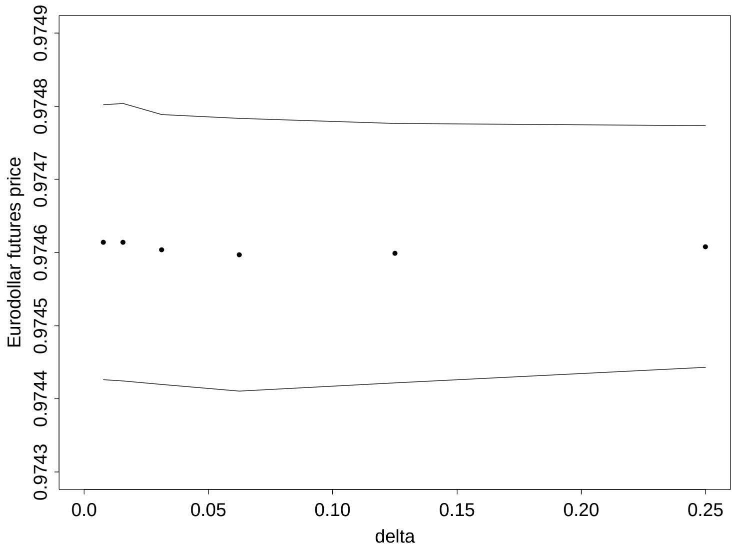

Figure 9: The simulated Eurodollar futures price in the full model. t=0, T=1, $\delta_U = 0.25$, otherwise the usual setting. $\delta/\epsilon = 5$, $M = 50000$. *'s are estimates, lines show 95 %-confidence interval.

5 Conclusion and Discussion

In this paper we have described a lognormal model for the LIBOR. The model was similar to the HJM-framework in that it took the current term structure as direct input, but the focus was shifted to rates with strictly positive compounding periods. This is surely this first, major step that is to be taken when we want a model the produce realistic prices i.e., prices that can be quoted in the market. We were able to find closed form solutions for swaptions if we were willing to make a number of approximations. These prices depended on the fundamental parameter δ in a non-obvious manner. We showed that this dependence was very well approximated linearly, in particular the prices were stable. We also showed that when a deterministic drift approximation was used, the model converged weakly to the well-known BK-model. This lead to concerns about the δ-stability of other derivatives, especially the Eurodollar futures contract, because the price connected with this contract has been shown to be infinite in the BK-model. Simulation studies revealed no indications of any instability for reasonable choices of δ. Without the deterministic drift approximation we have to determine swaption prices by simulation. We did this and showed that (a) there was not much difference between the two prices, and (b) using the closed form expression to generate control variates markedly reduced computer-time. The question of infinite Eurodollar futures prices was specific to a lognormal model, both otherwise a normal LIBOR-model can be treated with virtually the same, or even simpler, calculations. Structural similarities also allow for calculation of prices of other types of derivatives. We did not do this explicitly, but adapted a result from Brace & Musiela (1997) that allowed us to determine the risk neutral density of linear combinations of ZCBs.

Some of the approximations were not investigated. In Brace & Musiela (1997) it is shown how to get around the rank-1 approximation to covariance matrix in a normal model specification by using a "conditioning, linearising, and integrating out"-trick. Sawa (1997) considers approximations of this type in the lognormal model. There better, or at least more sophisticated, drift approximations are also considered.

Since the lognormal model, as opposed to the normal specification, forces us to make both analytical and numerical approximations as well as poses possible instability problems, the logical question is "Why use it?" A first answer is "Because it"s there!" A second is saying, in the same tone of voice that Buddhist monks use when reciting their mantras, that it prevents negative interest rates. A third reason is that might explain empirical behaviour better, which in turn leads to the issue of is model choice, one that there has not yet been much discussion about in the literature in the particular context described in this paper. In the framework of one-factor short term interest rates there seems to be no end to articles and working papers estimating and testing different specifications against each other. But believing that you can choose between models of this type by using conventional statistical methods seems futile. Firstly, the model is specified under numerous probability measures. The arbitrage free measure could be the objective probability measure ("the local expectations hypothesis"), but assuming that the forward measures coincides to the objective probability is easily seen to imply that interest rates cannot be stochastic. What could help us is that the volatility is the same under the different measures, so we could just do some factor analysis to validate models. But how do you consistently separate out drifts, whose functional form you do not know when you only have discrete observations ? In practical applications we would calibrate the model (meaning that we find the volatility) to the most liquid derivatives, typically caplets. This is very good on a day-to-day basis but can be dangerous in the long run since it is hard to determine whether the dynamics of your model are wrong. Indications of this would be wildly fluctuating implied estimates, poor hedges and, eventually, bankrupt! The widespread use of implied volatility also raises the question whether there is any relation between the implied volatility and the actual volatility (perhaps even the one the model claims there is?) Highly reliable data from both sides of the market can be obtained, which would greatly facilitate an analysis of this question.

Another question regarding model choice is where to start. Our model started in the LIBOR-market. This seems very reasonable; the market is liquid and LIBORs are easy to observe (as opposed to "the short rate", which is a highly theoretical concept). From this we then find the swap rates (see (6)), which have unknown, but numerically not too different from lognormal, distributions. Jamshidian (1997) starts by specifying swap rate dynamics (under a particular measure). A topic for research is how different, numerically, models of these types are. LIBOR and swap markets are probably the only ones liquid enough to justify they use a "dynamic building block".

References

Black, F. (1976), 'The pricing of commodity contracts', *Journal of Financial Economics* **3**, 167–179.

Black, F. & Karasinski, P. (1991), 'Bond and option pricing when short rates are lognormal', *Financial Analysts Journal* pp. 52–59.

Black, F., Derman, E. & Toy, W. (1990), 'A one-factor model of interest rates and its application to treasury bond options', *Financial Analysts Journal* pp. 33–39.

Brace, A. & Musiela, M. (1997), Dual formulae in the normal and lognormal models, Working Paper. University of New South Wales.

Brace, A., Gatarek, D. & Musiela, M. (1997), 'The market model of interest rate dynamics', *Mathematical Finance* **7**(2), 127–154.

Duffie, D. (1988), *Security Markets: Stochastic Models*, Academic Press, Boston.

Duffie, D. (1992), *Dynamic Asset Pricing Theory*, Princeton University Press.

He, H. (1990), 'Convergence from discrete- to continuous-time contingent claims prices', *Review of Financial Studies* **3**, 523–546.

Heath, D., Jarrow, R. & Morton, A. (1992), 'Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation', *Econometrica* **60**(1), 77–105.

Hoffman, I. D. (1993), Lognormal processes in finance, PhD thesis, University of New South Wales.

Hogan, M. & Weintraub, K. (1993), The log-normal interest rate model and eurodollar futures, Working Paper. Citibank, New York.

Hull, J. (1993), *Options, Futures, and Other Derivative Securities*, 2. edn, Prentice-Hall, Englewood Cliffs, N.J.

Jamshidian, F. (1989), 'An exact bond option formula', *Journal of Finance* **44**, 205–209.

Jamshidian, F. (1997), 'Libor and swap market models and measures', *Finance and Stochastics* **1**, 293–330.

Miltersen, K. (1994), 'An arbitrage theory of the term structure of interest rates', *The Annals of Applied Probability* **4**(4), 953–967.

Miltersen, K., Sandmann, K. & Sondermann, D. (1997), 'Closed form solutions for term structure derivatives with log-normal interest rates', *Journal of Finance* **52**, 409–430.

Musiela, M. & Rutkowski, M. (1997*a*), 'Continuous-time term structure models: forward measure approach', *Finance and Stochastics* **1**, 261–292.

Musiela, M. & Rutkowski, M. (1997*b*), *Martingale Methods in Financial Modelling*, 1. edn, Springer-Verlag.

Nelson, D. & Ramaswamy, K. (1989), 'Simple binomial processes as diffusion aprroximations in finacial models', *Review of Financial Studies* **3**, 393–430.

Press, W., Teukolsky, S., Vetterling, W. T. & Flannery, B. (1992), *Numerical recipes in C*, 2. edn, Cambridge University Press.

Rutkowski, M. (1997), Models of forward libor and swap rates, Working Paper. Politechnika Warszawska. Submitted to *Applied Mathematical Finance*.

Sawa, J. (1997), Approximations to swaption pricing formula in bgm model, Working Paper. University of New South Wales.

Schmidt, W. (1996), Pricing irregular interest cash flows, Working Paper. Deutcshe Morgan Grenfell, Frankfurt am Main.

A Lest we forget....

There may be a lot of things we can't do in finance. But we're good at calculating means of truncated normally and lognormally distributed variables. To get from (16) to (17) it suffices to look at

$$\mathbb{E}_{T_j}\left([K(0,T_{j-1})\exp(X_j-\frac{1}{2}\Gamma_j^2)-qoz(j,q)\kappa]\mathbf{1}_{\{X_j\geq\Gamma_j(s+d_j)\}}\right).$$

Under $\mathbb{P}_{T_j}$ we have $X_j\sim N(0,\Gamma^2)$, which we can write in the following fancy way; $X_j=X_j(1)$ where

$$dX_j(t)=\Gamma_j d\mathbb{W}_{T_j}(t).$$

By symmetry around 0 we have

$$\mathbb{E}_{T_j}\left(qoz(j,q)\kappa\mathbf{1}_{\{X_j\geq\Gamma_j(s+d_j)\}}\right)=qoz(j,q)\kappa\Phi(-s-d_j).$$

As a bolt from the blue introduce a probability measure $\mathbb{Q}$ through

$$\frac{d\mathbb{Q}}{d\mathbb{P}_{T_j}}=\exp(X_j-\frac{1}{2}\Gamma_j^2).$$

Girsanov's theorem gives us that $d\mathbb{W}_{\mathbb{Q}}(t)=d\mathbb{W}_{T_j}(t)-\Gamma_j dt$ defines a $\mathbb{Q}$-BM and that

$$dX_j(t)=\Gamma_j^2dt+\Gamma_j d\mathbb{W}_{\mathbb{Q}}(t),$$

so that $X_j\sim N(\Gamma_j^2,\Gamma_j^2)$ under $\mathbb{Q}$ and we get by simple manipulations that

$$\begin{aligned}\mathbb{E}_{T_j}\left(\exp(X_j-\frac{1}{2}\Gamma_j^2)\mathbf{1}_{\{X_j\geq\Gamma_j(s+d_j)\}}\right)&=\mathbb{E}_{\mathbb{Q}}\left(\mathbf{1}_{\{X_j\geq\Gamma_j(s+d_j)\}}\right)\\&=\Phi(\Gamma_j-s-d_j),\end{aligned}$$

which when collecting all the terms proves (17) without calculating a single integral.

A Comparison of Approximation Techniques for Transition Densities of Diffusion Processes

Bjarke Jensen* Rolf Poulsen†

June 10, 1999

Abstract

We give a detailed comparison of a large number of techniques used to approximate the transistion densities of a diffusion process. The techniques considered are simple normal approximations, simulation-based methods, binomial approximations, numerical solution of the Fokker-Plank partial differential equation, and Hermitian expansion. From speed/accuracy trade-off considerations this list also indicates the qualitative ordering of the methods.

Contents

*Department of Management, Building 350, University of Aarhus, DK-8000 Denmark.

†Department of Operations Research, Building 520, University of Aarhus, DK-8000, Denmark.

1 Introduction

Diffusion processes are widely used in financial and statistical modelling and are characterized by their transition densities. Unfortunately, we are in general unable to express these in a way that would commonly be accepted as "closed-form." We do, however, know many features of the densities; features that can be of both probabilistic and analytical nature and can lead in natural ways to approximations of the densities. In this paper we give a detailed comparison some of the methods most commonly and recently suggested in the literature.
The philosophy behind the comparisons can be expressed in two "bullet points":

- "General to specific"
  - *First* we describe the methods in a setting that allows for implementation for quite general parameter specifications, at least when we restrict ourselves, as we shall do, to one-dimensional process. *Then* we base error

comparisons on some of the few – but of course widely used – models where the transition densities are known.

- We also aim for generality by trying to find the transition densities. Teoretically, they characterize the process and with them at hand we can do many things easily. If we are "financially inclined" we can price any European-type contingent claim by (at most) numerical integration. Also, having implemented the algorithms a number of more exotic financial contracts can be priced with only minor revisions (but it may require quite some ingenuity to see *which* revisions). If we are "statistically inclined" we will use the transistion densities and the Markov property to write down the likelihood function. Maximizing this function (over "parameters", which may somewhat delicate numerically) gives us benefits in statistical inference ("under mild conditions": consistency, asymptotic normality, and efficiency).

• "Multi-criteria"

Contrary to what mathematical analysts will have you believe, approximation techniques should be compared not just by "the order of the remainder in a Taylor expansion". The following questions, with answers that are increasingly hard to quantify, should be asked:

- How accurate is the approximation?
- How long does it take to calculate the approximation?
- How long time does it take to program (and debug) the algorithm?
- How flexible is the method?

We shall try to address all these issues, even if the treatment of the latter points is by nature subjective.

The remainder of this section describes the general set-up. For the outline of the rest of the paper we refer to the table of contents.

We consider a set-up with a one-dimensional diffusion process X with dynamics governed by the stochastic differential equation

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t, \quad X_0 = x_0, \tag{1}$$

where W is a Brownian motion on some filtered probability space $(\Omega, (\mathcal{F})_t, \mathcal{F}, Q)$, and $\mu, \sigma : \mathbf{R} \mapsto \mathbf{R}$ are functions such that Equation (1) is well-defined and has a (weak) solution and (versions of) transition densities that are absolutely continuous wrt. the Lebesque measure and sufficiently smooth. We seek these transition densities. That is, we are looking for $\phi : \mathbf{R}_+ \times \mathbf{R} \times \mathbf{R} \mapsto \mathbf{R}_+$ (referred to as *the* transition density) such that

$$y \mapsto \phi(\triangle, x, y)$$

is the density of $X_\triangle$ given $X_0 = x$.

2 Models

In this section we describe the three most well-known models to the financial community. They are all characterized by a closed-form expression of the transition density.

2.1 The Vasicek Model

The term 'Vasicek model' is in tribute to Vasicek (1977); people outside the financial community would call it an Ornstein-Uhlenbeck process. The dynamics of the SDE are

$$dX_t = \kappa(\theta - X_t)dt + \sigma_V dW_t,$$

for $\kappa, \sigma_V \in \mathbf{R}_+$ and $\theta \in \mathbf{R}$. In this case $X_\triangle | X_0 = x$ is normally distributed, specifically the transition density is

$$\phi^V(\triangle, x, y) = n(y; m(\triangle, x), v(\triangle, x)),$$

where n is the normal density and the conditional moments are

$$\begin{aligned} m(t, x) &= E(X_t | X_0 = x) = e^{-\kappa t} x + \theta(1 - e^{-\kappa t}) \\ v(t, x) &= Var(X_t | X_0 = x) = \frac{\sigma_V^2 (1 - e^{-2\kappa t})}{2\kappa}. \end{aligned}$$

2.2 The Cox-Ingersoll-Ross Model

The term 'Cox-Ingersoll-Ross model' is in tribute to Cox, Ingersoll & Ross (1985), from which we have also adapted the density and moment formulae. The dynamics of the SDE are

$$dX_t = \kappa(\theta - X_t)dt + \sigma_{CIR}\sqrt{X_t}dW_t,$$

for $\kappa, \theta, \sigma_{CIR} \in R_+$ such that $2\kappa\theta > \sigma_{CIR}^2$. In this case $X_\triangle | X_0 = x$ is non-central χ^2-distributed, specifically the transition density is

$$\phi^{CIR}(\triangle, x, y) = c\exp(-u - v)\left(\frac{v}{u}\right)^{q/2} I_q(2\sqrt{uv}) \quad \text{for } x, y \in \mathbf{R}_+,$$

where

$$\begin{aligned} c = \frac{2\kappa}{\sigma_{CIR}^2(1 - \exp(-\kappa\triangle))}, &\qquad q = \frac{2\kappa\theta}{\sigma_{CIR}^2} - 1, \\ u = cx\exp(-\kappa\triangle), &\qquad v = cy, \end{aligned}$$

and I_q is the modified Bessel function of order q. Modified Bessel functions can be evaluated using routines from Numerical Recipes (Press, Teukolsky, Vetterling & Flannery (1992)). It should be noted that for large z (meaning $z > 2$) we have $I_q(z) \approx \exp(z)$. For the parameters we consider this is likely to be the case, so a slight modification in the routines is needed.
The conditional mean is given by the same expression as for the Vasicek model, while

$$v(t, x) = x\frac{\sigma_{CIR}^2(e^{-\kappa t} - e^{-2\kappa t})}{\kappa} + \theta\frac{\sigma_{CIR}^2(1 - e^{-\kappa t})^2}{\kappa}.$$

2.3 The Black-Scholes Model

Naturally, the term 'Black-Scholes model' is also a tribute to Robert Merton; others would call it a Geometric Brownian motion with drift. The dynamics of the SDE:

$$dX_t = rX_t dt + \sigma_{BS}X_t dW_t,$$

for $r, \sigma_{BS} \in \mathbf{R}_+$. In this case $X_\triangle | X_0 = x$ is lognormally distributed, specifically the transition density is

$$\phi^{BS}(\triangle, x, y) = \frac{1}{\sigma_{BS}y\sqrt{\triangle}\sqrt{2\pi}}\exp\left(-\frac{(\ln y - (\ln x + (r - \sigma_{BS}^2/2)\triangle))^2}{2\sigma_{BS}^2\triangle}\right), \quad \text{for } x, y \in \mathbf{R}_+,$$

and

$$\begin{aligned} m(t, x) &= xe^{rt} \\ v(t, x) &= x^2e^{2rt}(e^{\sigma_{BS}^2 t} - 1). \end{aligned}$$

3 Approximation Techniques

3.1 Simple Analytic Approximations

Two, quite different, approaches are used in the literature to obtain (effectively) analytic approximations to ϕ.

3.1.1 Small Time Steps

If $\triangle$ is small we can think of dX_t as $X_\triangle - X_0 = X_\triangle - x_0$, approximate any X_t's occurring on the RHS of (1) with X_0 and perceive 'dW_t' as a $N(0,\triangle)$-variable. This leads to the approximation

$$\phi^A(\triangle, x, y) = n(y; x + \mu(x)\triangle, \sigma^2(x)\triangle).$$

In a certain sense this (often called an Euler-approximation) is a first order approximation wrt. $\triangle$, i.e. it is 'good' when '$\triangle$ is small'. See Kloeden, Platen & Schurz (1991) for precise statements and examples of how to use such approximations (or: go to Section 3.2 of this paper).

Alas, $\triangle$ is typically not a quantity we (the statisticians or the financial modelers) can control; it is thrust upon us by God or 'The Market' or ... . Also, the error, which is highly dependent on the functional form of the drift and volatility functions, is hard to approximate in a simple and constructive manner.

Suppose the conditional first and second moments of X are known. Then we could use

$$\phi^A(\triangle, x, y) = n(y; m(\triangle, x), v(\triangle, x)).$$

This is also a first order approximation wrt. $\triangle$ but for many reasons a better one. For example, if it is used in statistical applications to discretely observed data, we obtain consistent estimates by maximizing the approximate loglikelihood function. Of course, there is a huge overlap between models where the transistion density is known and models where the conditional moments are known. But finding moments by numerical methods, e.g. simulation, seems an easier task than trying to determine the whole transistion density.

3.1.2 Large Time Steps

The material in this section is covered in 'second course'-books on stochastic processes (e.g. Karatzas & Shreve (1992) and Karlin & Taylor (1981)).

Let $D_X = (\underline{x}, \overline{x})$ denote the domain of X. In finance the only two cases relevant are the ones where $D_X = (-\infty, \infty)$ or $D_X = (0, \infty)$. Define

$$s(x) = \exp\left(-2\int_{x^{\#}}^{x} \frac{\mu(y)}{\sigma^2(y)} dy\right),$$

and

$$\nu(x) = \frac{1}{\sigma^2(x)s(x)},$$

where $x^{\#}$ is an arbitrary point in D_X. We introduce *the speed measure* by requiring its density to be ν. Assume also that the following conditions hold

$$\begin{aligned} \int_{\underline{x}}^{x^{\#}} s(x)dx &= \int_{x^{\#}}^{\overline{x}} s(x)dx = \infty \\ \int_{\underline{x}}^{\overline{x}} \nu(x)dx &\equiv \overline{\nu} < \infty. \end{aligned}$$

Then X is ergodic with invariant measure

$$\phi^{\infty}(y) = \lim_{t\to\infty} \phi(t, x, y) \tag{2}$$

given by

$$\phi^{\infty}(y) = \frac{\nu(y)}{\overline{\nu}}.$$

Finally, if $X_0 \sim \phi^{\infty}$ (which in statistical applications may largely be a matter of perception), then X is stationary, i.e. $X_t \sim \phi^{\infty}$ for all t. Note that it is the unconditional density that is the same for all t. Often ϕ^{∞} can be found in closed form (also for models with unknown conditional densities) or it can easily be determined by numerical integration. Therefore it is reasonable to think of the approximation

$$\phi^{A}(t, x, y) = \phi^{\infty}(y)$$

as analytic.

When is this a good approximation? Looking at (2) we note two things about ϕ^{∞}:

i) It is the limit as t tends to infinity.

ii) It does not depend on x.

This leads us to the conclusion that ϕ^∞ is a good approximation if $i)$ $\triangle$ is large or, $ii)$ ϕ does not 'depend very much' on x. The latter means that if X is used to model 'reasonably uncorrelated events' then ϕ^∞ is a good approximation. This is often the case when stock returns are modeled.
For the Vasicek model we see that

$$\phi^\infty(y) = n\left(y; \theta, \frac{\sigma_V^2}{2\kappa}\right).$$

For the Cox-Ingersoll-Ross model one can show that

$$\phi^\infty(y) = \frac{w^v}{\Gamma(v)} y^{v-1} e^{-wy}$$

where $v = 2\kappa\theta/\sigma_{CIR}^2$ and $w = 2\kappa/\sigma_{CIR}^2$ (so the invariant measure corresponds to a Γ-distribution).
The Black-Scholes model does not have an invariant measure (but $(\ln(e^{-rt}X_t))$ does).

3.2 Simulation

Monte Carlo simulation methods in finance are mainly used to price derivatives. The first paper published in finance using the idea of Monte Carlo simulation was Boyle (1977). In this paper the value of standard European call-options generated by Monte Carlo simulation were analyzed. The main idea of Monte Carlo simulations is to use the Law of Large Numbers known from basic probability theory. From basic arbitrage-theory it is known that the fair value of a European claim, c, maturing at time T with X as the underlying asset and pay-off structure given by $\widetilde{f}(\cdot)$ is:

$$c_0 = \mathrm{E}^Q\left[e^{-rT}\widetilde{f}(X_T)\right],$$

where r is the (constant) risk-free interest rate and the expectation is taken with respect to the risk-neutral measure Q. Therefore; simulating N independent realizations of X_T (under the measure Q) by the Monte Carlo method will result in the following unbiased estimate of the claim's price:

$$c_0^{est} = e^{-rT}\frac{1}{N}\sum_{j=1}^{N}\widetilde{f}\left(X_T^j\right),$$

where X_T^j denotes the j'th simulation of X_T. Before discussing how to simulate X_T it is worth to make some comments regarding the standard error of the estimate c_0^{est} :

1. Let $\sigma\left[\widetilde{f}\left(X_T\right)\right] := \sqrt{\text{VAR}\left(\widetilde{f}\left(X_T\right)\right)}$ denote the standard error coming from simulations of X_T. Then it is easily seen that the standard error for the estimate c_0^{est} is:

$$\text{STD}\left(c_0^{est}\right) = e^{-rT} \cdot \frac{\sigma\left[\widetilde{f}\left(X_T\right)\right]}{\sqrt{N}}.$$

   Therefore it is noticed that if a reduction in STD(c_0^{est}) of (for instance) a factor of 10 is required a number of $100 \cdot N$ simulations are required. From a computational point of view this can be a quite time-consuming task.

2. Instead of increasing the number of simulations in order to decrease the standard error of the estimate, STD(c_0^{est}) it may be more advantageous to simulate X_T such that the estimate c_0^{est} is still unbiased, and such that the standard error STD(c_0^{est}) is smaller than the standard error described above using the 'crude' Monte Carlo method, i.e. to reduce the standard error $\sigma\left[\widetilde{f}\left(X_T\right)\right]$. Several methods, some of which will be described in subsection 4.1.2, have been proposed. However, before implementing these methods, one has to be sure that the extra computer-time required for this 'sophistication' of the Monte Carlo method does not become larger than the extra computer-time required, increasing the number of simulations using the 'crude' method, in order to get the same decreased standard error, STD(c_0^{est}). An example:

   Suppose that the standard error estimated using N 'crude' simulations can (and should) be decreased by a factor of 10 using a 'sophisticated' method. For N simulations this method requires an addition of s seconds on the CPU. If an increase of simulations to $100 \cdot N$ results in an increase of less than s seconds on the CPU it is seen that it is more desirable to simply increase the number of simulations instead of implementing the 'sophisticated' method.

   For an excellent survey about this topic (and the general topic of Monte Carlo simulation in finance) see Boyle, Broadie & Glasserman (1997). However it should be noticed that 'sophisticating the simulation' techniques normally just increase the computer-time with a fraction compared to increasing the number of simulations - at least for a reasonable number of simulations.

After discussing how to estimate option-prices from simulations now comes a brief discussion regarding the simulation of X_T. Since (1) describes a continuous-time stochastic differential equation a discretization-scheme is needed to compute realizations of X_T. The book by Kloeden et al. (1991) gives a detailed analysis regarding this subject. In this paper such a detailed description will not be given. Two different discretization-schemes will now be described: The Euler scheme (EU) and The Milstein scheme (MI) which approximate (1) by the following discrete schemes:

$$\begin{aligned} X_{(i+1)\delta} - X_{i\delta} &= \mu\left(X_{i\delta}\right)\delta + \sigma\left(X_{i\delta}\right)\sqrt{\delta}\widetilde{N} \qquad \text{(EU)}, && (3)\\ X_{(i+1)\delta} - X_{i\delta} &= \mu\left(X_{i\delta}\right)\delta + \sigma\left(X_{i\delta}\right)\sqrt{\delta}\widetilde{N} \\ &\quad + \frac{1}{2}\sigma\left(X_{i\delta}\right)\sigma'\left(X_{i\delta}\right)\delta\left\{\widetilde{N}^2 - 1\right\} \quad \text{(MI)}, && (4)\\ i &= 0, 1, \ldots, M-1, \end{aligned}$$

where $X_0 = x_0$ is given, the volatility function $\sigma(\cdot)$ is differentiable, M is a fixed number of discretization-times, $\delta = \frac{T}{M}$ is the time-length (chosen by us) between two discretizations and $\widetilde{N}$ is a standard normal generator giving independent standard normal distributed realizations: $n_1, n_2, \ldots, n_M$.

It is observed from (3) and (4) that the Milstein-scheme is a refinement of the Euler-scheme leading to higher order accuracy (see Kloeden et al. (1991)).[1] However considering the Vasicek model it is seen that the two schemes coincide because the volatility in that model is independent of state-variable and therefore implies a zero extra term going from the Euler-scheme to the Milstein-scheme.

Generating M independent standard normal variables therefore lead to one realization of X_T using either The Euler-scheme or the Milstein-scheme. This leads us to conclude that $N \cdot M$ independent standard normal variables are needed to simulate a price of an option when the underlying process follows (1). However if the antithetic variates technique (as described above) is used, just half the size $\left(\frac{1}{2}N \cdot M\right)$ is needed.

Notice that although the explicit solutions to the differential equations analyzed in this paper could be written down and from that point on, one could start the simulation (requiring only $\left(\frac{1}{2}N\right)$ independent standard normal variables) it is not done.[2] This is not done because in the general case there exists no explicit solution to the

[1] It should be mentioned that Kloeden et al. (1991) introduce schemes of even higher orders. The implementation of one of these schemes might decrease the number of time approximation steps dramatically to get as good an approximation as, for instance, the Euler scheme.

[2] For instance: In the Black-Scholes model one could simulate X_T using the explicit solution:

$$X_T = X_0 \exp\left\{\left(r - \frac{1}{2}\sigma^2\right)T + \sigma\sqrt{T}\widetilde{N}\right\}.$$

stochastic differential equation and the only hope of simulating the density function is to use a discretization-scheme.

Another thing to consider is the optimal relationship between the numbers of simulations (N) and the number of time-discretizations (M) with a fixed amount of computer-time. Duffie & Glynn (1995) have done some investigations regarding that problem. Their main-result is an asymptotic result stating that it is optimal to quadruple the number of simulations with each doubling of time-discretizations for the Euler-scheme. For the Milstein-scheme one should make 16 times as many simulations when the number of time-discretizations are doubled.[3] We will return to this subject when the numerical results are discussed in a later section.

Generating random-numbers on a computer is a rather sophisticated thing to do. In finance mainly two types of generators are used: The pseudo-generator and the quasi-generator. A brief discussion of the two types and how they are implemented in the analysis of this paper follows:

1. The pseudo-generator: A linear congruential formula is used generating pseudo-random-numbers. The generator will return numbers in the interval $[0, 1)$ which can be seen as independent realizations of a uniform random variable taking its values in the interval $[0, 1)$.[4] The linear congruential formula is:

$$U_{n+1} = (aU_n + c) \text{ MOD } m, \ n \geq 1.$$

   To start the algorithm generating the numbers U_n one has to initialize the formula. This is done by stating a start-value for U_1. The n'th number displayed by this generator will then be $\frac{U_n}{m}$ which is located in the interval $[0, 1)$.

   To transform the uniform random variables to standard normal variables Box-Muller's method is used. This method is based on the theorem that if $\widetilde{U}$ and $\widetilde{V}$ are two independent $U(0, 1)$ distributed random variables, then:

$$\begin{aligned} \widetilde{N}_1 &= \sqrt{-2\ln\left(\widetilde{U}\right)}\cos\left(2\pi\widetilde{V}\right) \text{ and} \\ \widetilde{N}_2 &= \sqrt{-2\ln\left(\widetilde{U}\right)}\sin\left(2\pi\widetilde{V}\right) \end{aligned}$$

[3]It should be noticed that the Milstein-scheme used in the paper by Duffie & Glynn (1995) is a little different from the Milstein-scheme used in this paper. However the results from the theorem still applies.

[4]This hypothesis is accepted by various statistical tests for independence and distribution.

are two independent standard normal variables. Therefore transforming the realizations made by the above described generator using Box-Muller´s method generate independent standard normal distributed realizations.

The pseudo-generator has been criticized for generating too many 'clustered' numbers meaning that a histogram made on the interval $[0,1)$ of the realizations will result in too many 'peeks' and 'valleys' or if multi-dimensional simulation is used: plotting the numbers in the unit hypercube result in too many bare areas. To overcome this problem one has to simulate a lot of numbers (remember that the standard error of the simulated price is of order $N^{-\frac{1}{2}}$) or use what is known as quasi-random numbers.[5]

2. The quasi-random generator: Instead of using pseudo-random numbers one can use a quasi-random generator. This method generates sequences of numbers in the interval $[0,1)$ in such a way that the 'clustering' problem described above has decreased. Of course, one has to be very careful using these numbers because they can no longer be seen as independent draws. For a thorough understanding of these methods see Galanti & Jung (1997) and Joy, Boyle & Tan (1996).

   The quasi-random generator used in this study is the one generating Faure sequences. The base of the Faure sequence is going to be the smallest prime number (p) equal to or bigger than the number of time-approximations (in the above denoted M). Initializing the Faure sequence to start at the value p^4 should result in good numbers that do not cluster around zero.[6] For completeness here comes the algorithm generating Faure sequences. Furthermore an example with Faure numbers will be provided to indicate the way the numbers are constructed.

   Let n go through the list $\{p^4, p^4+1, \ldots, p^4-1+N\}$. For each n the following algorithm will provide M numbers in the interval $(0,1)$ that can be used to generate one pricepath for X_T according to one of the above described discretization-schemes.

Algorithm 1 • *Convert n into its representation in the base p number sys-*

[5]They are also known as low-discrepancy sequences.
[6]See Galanti & Jung (1997).

tem, i.e., find (the unique) numbers $a_0(n), a_1(n), \ldots, a_J(n)$ *such that:*

$$n = \sum_{j=0}^{J} a_j(n)\, p^j,$$

where J *equals the integer part of* $\frac{\ln(n)}{\ln(p)}$.

- *Reflect the found numbers* $a_j(n)$ *around the decimal point to get the first Faure number*

$$\alpha_1(n) \equiv \sum_{j=0}^{J} a_j(n)\, p^{-j-1}.$$

- *For each* k *in the list* $\{2, 3, \ldots, M\}$ *reorder the numbers* $a_j(n)$ *to* $b_0^k(n)$, $b_1^k(n), \ldots, b_J^k(n)$ *by the following rule:*

$$b_i^k(n) \equiv \left[\sum_{j\geq i}^{J} \binom{j}{i} (k-1)^{j-i} a_j(n)\right] \text{ MOD } p.$$

- *Reflect the found numbers* $b_i^k(n)$ *around the decimal point to get the k'th Faure number*

$$\alpha_k(n) \equiv \sum_{i=0}^{J} b_i^k(n)\, p^{-i-1}.$$

An example using the above algorithm (an actually used in the numerical study) now follows.

Example 2 *Let* $M = 4$*. Then it is obvious that* $p = 5$ *and that the sequence start at* $n = 5^4 = 625$*. For the three different numbers: 625,626 and 632 it is obvious that* J *equals 4. The corresponding values of* $a_j(n)$, $b_i^k(n)$ *and* $\alpha_k(n)$ *are therefore:*

$n = 625$						$n = 626$						$n = 632$					
a_0	a_1	a_2	a_3	a_4	α_1	a_0	a_1	a_2	a_3	a_4	α_1	a_0	a_1	a_2	a_3	a_4	α_1
0	0	0	0	1	$\frac{1}{3125}$	1	0	0	0	1	$\frac{626}{3125}$	2	1	0	0	1	$\frac{1376}{3125}$
b_0^2	b_1^2	b_2^2	b_3^2	b_4^2	α_2	b_0^2	b_1^2	b_2^2	b_3^2	b_4^2	α_2	b_0^2	b_1^2	b_2^2	b_3^2	b_4^2	α_2
1	4	1	4	1	$\frac{1171}{3125}$	2	4	1	4	1	$\frac{1796}{3125}$	4	0	1	4	1	$\frac{2546}{3125}$
b_0^3	b_1^3	b_2^3	b_3^3	b_4^3	α_3	b_0^3	b_1^3	b_2^3	b_3^3	b_4^3	α_3	b_0^3	b_1^3	b_2^3	b_3^3	b_4^3	α_3
1	2	4	3	1	$\frac{991}{3125}$	2	2	4	3	1	$\frac{1616}{3125}$	0	3	4	3	1	$\frac{491}{3125}$
b_0^4	b_1^4	b_2^4	b_3^4	b_4^4	α_4	b_0^4	b_1^4	b_2^4	b_3^4	b_4^4	α_4	b_0^4	b_1^4	b_2^4	b_3^4	b_4^4	α_4
1	3	4	2	1	$\frac{1111}{3125}$	2	3	4	2	1	$\frac{1736}{3125}$	1	4	4	2	1	$\frac{1236}{3125}$

Since the numbers in a Faure sequence can no longer be regarded as independent the Box-Muller method cannot be used. Instead Moro's inverse normal function approximation is used. Again, for completeness, here comes Moro's inverse normal function:[7]

Let $\Phi(x)$ denote the normal distribution function then Moro's inverse normal distribution is given by:

$$\Phi^{-1}(x) = \begin{cases} y\frac{\sum_{n=0}^{3} a_n y^{2n}}{\sum_{n=0}^{4} b_n y^{2n}} & \text{if } |y| \leq 0.42 \\ h(z) & \text{if } 0.42 < y < 0.5 \\ -h(z) & \text{if } -0.5 < y < -0.42 \end{cases},$$

where: $y = \Phi(x) - 0.5$, $z = k_1 [2\ln(-\ln[0.5 - |y|]) - k_2]$ and $h(z)$ is given recursively by:

$$\begin{aligned} d_{10} &= d_9 = 0, \\ d_j &= 2zd_{j+1} - d_{j+2} + c_j, \quad j = 8, 7, \ldots, 1, \\ h(z) &= d_0 = zd_1 - d_2 + \frac{c_0}{2}. \end{aligned}$$

The constants a_n, b_n, c_n and k_1 and k_2 are given by:

n	a_n	b_n	c_n
0	2.500662823884	1.00	7.7108870705487895
1	−18.61500062529	−8.47351093090	2.7772013533685169
2	41.39119773534	23.08336743743	0.3614964129261002
3	−25.44106049637	−21.06224101826	0.0373418233434554
4		3.13082909833	0.0028297143036967
5			0.0001625716917922
6			0.0000080173304740
7	k_1	k_2	0.0000003840919865
8	0.4179886424926431	4.2454686881376569	0.0000000129707170

The last thing to notice about this generator is that the difference

$$|c_0 - c_0^{est}|$$

[7]For technical details about this inverse distribution function the reader is referred to Joy et al. (1996).

asymptotically is of order $\frac{(\ln N)^M}{N}$, which for a suitable chosen N will be smaller than the error induced using pseudo numbers. However as will be seen in this numerical study the efficiency of the quasi numbers compared with the pseudo numbers is absolutely not overwhelming, especially not when computer-time is also included in the evaluation.

After describing the main features of Monte-Carlo simulation and the pricing of options it is time to focus on simulation of densities. Simulation of transition densities arising from diffusion processes have not been the topic of many research articles in finance. In fact we have not been able to find any papers considering Monte-Carlo simulation of transition-densities. All papers in finance using simulation are mainly focused on the techniques used; be it: The choice of generator, the choice of time-discretization-scheme and/or the choice of standard error reduction technique. Normally these choices are based on a specific option-pricing problem. However, as argued in the introduction: To measure the efficiency of a method, how well it approximates the (one-dimensional) density, is the right measure.
To simulate a transition density one has to make a histogram on the real axis $\mathbf{R}$ or the half-axis $[0, \infty)$, whichever is the supporting region for the underlying process $\{X_t\}$. Therefore suppose that the real axis $\mathbf{R}$ or the half-axis $[0, \infty)$ are divided into intervals, $[x_i, x_{i+1})$. The simulation of one realization of X_T is then recognized to be in one of the intervals $[x_i, x_{i+1})$. After N simulations one can count how many realizations that actually ended up in the interval $[x_i, x_{i+1})$. Dividing each of these numbers by N and we have the histogram that approximates the transition density. Actually the simulation of densities can be viewed as a simulation of some special digital options. If the real axis $\mathbf{R}$ or the half-axis $[0, \infty)$ are divided into intervals, $[x_i, x_{i+1})$ we see that simulation of the density is equivalent to the simulation of a basket of (non-discounted) binary options, C_i that pay off one dollar if X_T ends up in the interval $[x_i, x_{i+1})$ and nothing if X_T ends up somewhere outside this interval. Therefore the above described theory regarding option pricing using Monte Carlo simulation can also be applied to simulation of densities. Figure 1 shows the above described idea behind simulation of densities.

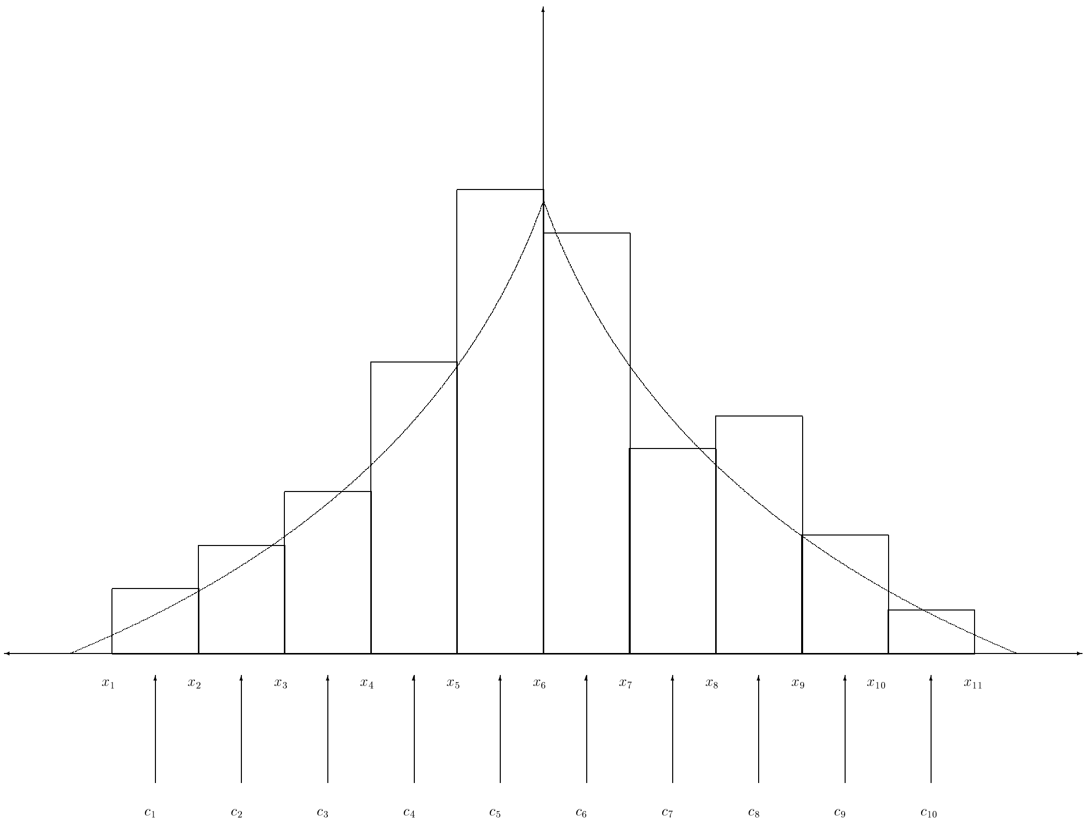

Figure 1: Example of a density approximation obtained by simulating a histogram.

A final comment regarding simulation of densities by histograms is worth making. One can consider the N draws of the random variable X_T as a dataset that can estimate the density using statistical devices other than the histogram. Especially, one could use kernel-smoothing known from non-parametric statistical analysis. But a histogram is actually kernel-smoothing, with the choice of kernel being the rectangular one, i.e.:

$$K(t) = \frac{1}{2} \text{ for } |t| < 1.$$

Other kernels could have been chosen, for instance the Gaussian density function. For our purposes their is not much to gain using other kernels. This statement is documented in Silverman (1986), p.43:

> ...there is very little to choose between the various kernels on the basis of mean integrated square error. It is perfectly legitimate, and indeed desirable, to base the choice of kernel on other considerations, for example the degree of differentiability required or the computational effort involved...

Since differentiability of the density is not an issue in this paper and since the rectangular kernel is the least computational time consuming kernel that exists, the choice is obvious! A histogram perfectly satisfies the demands set to our numerical analysis of transition densities.

3.3 Computationally Simple Binomial Models (CSBM)

This section largely follows Nelson & Ramaswamy (1989), which among other things is a brilliant 'translation into English' of general results from the probability literature (as found in e.g. Ethier & Kurtz (1986) and Stroock & Varadhan (1979)).
Think of a binomial model as a tree. More specificly as a tree with an orientation (given by 'time') where exactly two arcs go out from each node. With each node we associate a real number, an outcome of the binomial process at that point in time. With each arc we associate a number between 0 and 1, the probability of transition between the two nodes the arc connects. If the total number of nodes grows at most quadratically in the number of time intervals we say that the model is computationally simple. If the tree is recombining, it is a lattice.
In probabilistic terms: Suppose the time interval $[0; \triangle]$ has been divided into N pieces of equal length $k = \triangle/N$. For each value of k we consider a stochastic process $({}_kx_t)_{t\in[0;\triangle]}$ fulfilling

$$
\begin{aligned}
{}_0y_t &= x_0 \quad \forall k \\
{}_kx_t &= {}_kx_{hk} \quad hk \le t < (h+1)k \\
Q({}_kx_{(h+1)k} = c \,|\, {}_kx_{hk}) &= \begin{cases} q_k({}_kx_{hk}) & \text{for } c = X_k^+({}_kx_{hk}) \\ 1 - q_k({}_hy_{hk}) & \text{for } c = X_k^-({}_kx_{hk}) \end{cases} ,
\end{aligned}
$$

for functions $q_k : \mathbf{R} \mapsto [0; 1]$, $X_k^+, X_k^- : \mathbf{R} \mapsto \mathbf{R}$ such that $X_k^-(x) \le X_k^+(x)$ for all x.
Given a diffusion process, say by (1), we face three questions:

i) In what sense can we talk about (or better: prove) 'convergence' of a binomial process to a diffusion?

ii) How can we construct a binomial model that 'converges' to the diffusion?

iii) (How) Can we keep the binomial model computationally simple?

The type of converge most frequently considered in the literature is 'weak convergence on the space of RCLL functions'. This is a rather technical concept, but for our purposes it is important that it implies convergence in distribution of any vector,

$$({}_k x_{t_1}, \dots, {}_k x_{t_n}) \overset{\sim}{\to} (X_{t_1}, \dots, X_{t_n}) \quad \text{for} \quad k \to 0.$$

Weak convergence is just that; usually it is not enough to guarantee convergence of the objects we are interested in. This has to be done on a case-by-case basis (of course, some people can make a 'case' fairly general, see He (1990)). The best example is the convergence of the call-option price in the Cox-Ross-Rubinstein model (cf. Cox, Ross & Rubinstein (1979)) to the Black-Scholes price. To ensure weak convergence of a binomial model two conditions have to be fulfilled:

1) The local first and second moments of the binomial process converge to the drift and volatility of the diffusion process (and do so uniformly on compact sets).

2) The jump sizes in the binomial model tend to 0 (also in an appropriately uniform way).

This leaves us with considerable freedom in the choice of probabilities and up- and down-moves. One way to construct a binomial model is first to try with

$$\begin{aligned} X_k^+(x) &= x + \sqrt{k}\sigma(x), \\ X_k^-(x) &= x - \sqrt{k}\sigma(x), \\ q_k(y) &= \frac{1}{2} + \frac{1}{2}\sqrt{k}\frac{\mu(x)}{\sigma(x)}. \end{aligned}$$

This usually takes care of conditions 1) and 2); minor (but tedious) adjustments may be necessary to ensure uniformity (a problem that arises if $\sigma \not> 0$). This plan has serious flaw; unless σ is constant, it will not lead to a computationally simple model. But suppose we can find a function f such that

$$f'(x) = \frac{1}{\sigma(x)}$$

on the set where $\sigma(x) > 0$. Then the Ito formula tells us that the process $(f(X_t)) = (Y_t)$ has constant (unit) volatility. Our 'first try'-technique leads to a computationally simple binomial model for Y. Since (forgetting for a moment about the 0's of σ) f is strictly increasing it has an inverse function. Using this function we can transform the Y-lattice into an X-lattice. Convergence in distribution is preserved under continuous transforms, and thus we are done. That was the basis idea; some care need to be exercised to take care of the singularities of σ. This has been done in Nelson & Ramaswamy (1989) which contains both a general treatment and 'pret-a-porter'-versions for commonly used models. This idea of creating a diffusion with unit volatility is also used in the Hermitian expansion (see Section 3.5)
For the purposes of this paper there is a 'bonus question':

iv) How do we use the binomial model to get an estimate of the conditional density of the diffusion model?

Suppose we want an estimate of $\phi(\triangle, x_0, y)$ for some y corresponding to a terminal node. Keeping track of the q_k's through the lattice enables use to calculate

$$Q({}_kx_\triangle = y) = Q({}_kx_\triangle \in I_k)$$

where $I_k = [(y + X_k^-(y-k))/2; (y + X_k^+(y+k))/2[$. We then use the approximation

$$\phi^A(\triangle, x_0, y) = \frac{Q({}_kx_\triangle \in I_k)}{|I_k|}.$$

From a finite difference point of view (see Section 3.4) a binomial model can be seen as a so-called explicit method; this, however, is rarely fruitful and will not be investigated further.

3.3.1 A CSBM for the Vasicek Model

Since σ is constant the first try works if k is small enough, i.e.

$$f \; : \; x \mapsto x/\sigma_V, \tag{5}$$

$$g \; : \; y \mapsto \sigma_V y. \tag{6}$$

3.3.2 A CSBM for the Cox-Ingersoll-Ross Model

The f transformation is given by

$$x \mapsto 2\sqrt{x}/\sigma_{CIR}, \tag{7}$$

and the appropriate inverse transformation is given by

$$g : z \mapsto \begin{cases} \sigma_{CIR}^2 z^2/4 & \text{if } z > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

Following Nelson & Ramaswamy (1989) define

$$\begin{aligned} J_k^+(x) &= \begin{cases} \text{The smallest, odd, positive integer } j \text{ such that} \\ 4k\kappa\theta/\sigma_{CIR}^2 + x^2(1-\kappa k) < (x + j\sqrt{k})^2 \end{cases}, \\ J_k^-(x) &= \begin{cases} \text{The smallest, odd, positive integer } j \text{ such that} \\ 4k\kappa\theta/\sigma_{CIR}^2 + x^2(1-\kappa k) \geq (x - j\sqrt{k})^2 \text{ or } x - j\sqrt{k} \leq 0 \end{cases} \end{aligned}$$

A computationally simple binomial model is then given by

$$\begin{aligned} X_k^{\pm}(x) &= g(x \pm J_k^{\pm}\sqrt{k}), \\ q_k(x) &= \begin{cases} \frac{k\kappa(\theta - g(x)) + g(x) - X_k^-(x)}{X_k^+(x) - X_k^-(x)} & \text{if } X_k^+(x) > 0 \\ 0 & \text{otherwise} \end{cases}. \end{aligned}$$

3.3.3 Two CSBMs for the Black-Scholes Model

The f and g transformations here are

$$f \;:\; x \mapsto (\ln x)/\sigma_{BS}, \tag{9}$$

$$g \;:\; y \mapsto \exp(\sigma_{BS} y). \tag{10}$$

We can (for sufficiently small k) write the binomial approximation as

$$X_k^{\pm}(y) = y \exp(\pm \sigma_{BS}\sqrt{k}).$$

We could use the 'first approach' probabilities,

$$\tilde{q}_k(y) = \frac{1}{2} + \frac{1}{2}\sqrt{k}\frac{r}{\sigma_{BS}}.$$

Financial aesthetics, however, force us to do things slightly differently: The $\tilde{q}_k$'s do not represent risk-neutral probabilities in the binomial model. To illustrate:

$$\mathrm{E}^{\tilde{q}_k}\left[e^{-r\triangle}({}_k x_{\triangle} - K)^+\right]$$

would *not* be the (unique) arbitrage-free call-option price in the binomial model (augmented with a savings account). It would, though, converge to the Black-Scholes call-option price. The risk-neutral probabilities are

$$q_k = \frac{\exp(rk) - \exp(-\sigma_{BS}\sqrt{k})}{\exp(\sigma_{BS}\sqrt{k}) - \exp(-\sigma_{BS}\sqrt{k})}.$$

Fortunately, Taylor expansions show that the local moments also converge for this choice of q_k's, so all's well that ends well. We do not get problems of this kind for the Vasicek- and CIR-binomial approximations because these (thought of as short rate models) are not complete.

3.4 Partial Differential Equation Techniques

This section largely follows Poulsen (1999). Relying on diffusion theory (e.g. along the lines of Karatzas & Shreve (1992)), we have that under certain regularity conditions ϕ solves the parabolic PDE (called the *forward equation* or the *Fokker-Planck equation*)

$$\frac{\partial}{\partial t}\phi(t,x,y) = -\frac{\partial}{\partial y}\left(\mu(y)\phi(t,x,y)\right) + \frac{1}{2}\frac{\partial^2}{\partial y^2}\left(\sigma^2(y)\phi(t,x,y)\right), \tag{11}$$

with initial condition $\phi(0,x,y) = \delta(y-x)$, where $\delta(\cdot)$ is the Dirac-δ function. Strictly speaking this is not a PDE in the usual sense because the Dirac-δ function is not a function in the usual sense but a generalized function. It can be defined as a continuous, linear, real-valued mapping on the infinite dimensional space $\mathbf{C}^\infty(\mathbf{R})$ such that $\delta : f \mapsto f(0)$ for all $f \in \mathbf{C}^\infty(\mathbf{R})$. Therefore we cannot use standard theory to (define and) ensure existence and uniqueness of a solution. This has to be verified with other methods. Two possible ways are 'inspired guesses' and 'advanced probabilistic methods (Malliavan calculus) on the corresponding SDE'. Rigorous treatment of this is far beyond this (most) paper(s). Rather we will apply a numerical technique known from 'usual PDEs' and then see if it produces reasonable results.

3.4.1 The Crank-Nicolson Finite Difference Method

Let subscripts denote differentiation, suppress "x" and other arguments where it causes no confusion, and rewrite (11) as

$$\phi_t(t,y) = a(y)\phi + b(y)\phi_y + c(y)\phi_{yy} \tag{12}$$

where

$$\begin{aligned} a(y) &= (\sigma_y)^2 + \sigma\sigma_{yy} - \mu_y = \frac{1}{2}(\sigma^2)_{yy} - \mu_y, \\ b(y) &= 2\sigma\sigma_y - \mu = (\sigma^2)_y - \mu, \\ c(y) &= \frac{1}{2}\sigma^2. \end{aligned}$$

Now consider a time/space grid with step sizes k and h. For any grid point (m, n) away from the boundaries (say y_L and y_H) we consider the approximation $v_m^n \approx \phi(nk, y_L + mh)$ defined by the equation

$$\frac{v_m^{n+1} - v_m^n}{k} = a\delta_0(h) + b\delta_1(h) + c\delta_2(h), \tag{13}$$

where the δ's are difference operators and the arguments of a, b, and c have been notationally suppressed; the functions should all be evaluated at the space point corresponding to the LHS, i.e. at $y_L + mh$.

It is the particular choice of δ's that determine the accuracy and stability (and popular name) of the finite difference method. For the Crank-Nicolson method the operators act in the following way:

$$\delta_0(h) = \frac{1}{2}v_m^{n+1} + \frac{1}{2}v_m^n, \tag{14}$$

$$\delta_1(h) = \frac{1}{2}\frac{v_{m+1}^{n+1} - v_{m-1}^{n+1}}{2h} + \frac{1}{2}\frac{v_{m+1}^n - v_{m-1}^n}{2h}, \tag{15}$$

$$\delta_2(h) = \frac{1}{2}\frac{v_{m+1}^{n+1} - 2v_m^{n+1} + v_{m-1}^{n+1}}{h^2} + \frac{1}{2}\frac{v_{m+1}^n - 2v_m^n + v_{m-1}^n}{h^2}. \tag{16}$$

Taylor expansions show that this method is locally second order accurate in both h and k. Also, the Crank-Nicolson method is unconditionally stable for parabolic PDEs (for a proof of this, or perhaps just to find out what it means, see Strikwerda (1989)). Defining $\lambda = kh^{-2}$ and inserting in (13) gives

$$\begin{aligned}\left(\frac{bh\lambda}{4} - \frac{c\lambda}{2}\right) v_{m-1}^{n+1} + \left(1 - \frac{ak}{2} + c\lambda\right) v_m^{n+1} + \left(-\frac{bh\lambda}{4} - \frac{c\lambda}{2}\right) v_{m+1}^{n+1} &= \\ \left(-\frac{bh\lambda}{4} + \frac{c\lambda}{2}\right) v_{m-1}^n + \left(1 + \frac{ak}{2} - c\lambda\right) v_m^n + \left(\frac{bh\lambda}{4} + \frac{c\lambda}{2}\right) v_{m+1}^n&. \end{aligned}\tag{17}$$

Considering v as known on the boundaries and using the initial condition, (17) defines a sequence of tridiagonal linear systems of equations. These systems can be solved recursively and each system requires only a number of operations that is proportional to the number of state space steps when the tridiagonal structure is exploited, e.g. by the routine `tridag` from Numerical Recipes.

If we consider only diffusions with inaccessible boundaries (no reflection or absorption) then it is reasonable to put $v = 0$ on the grid boundaries. The grid boundary levels are chosen "sufficiently far apart" (the actual numbers are parameter dependent, but choosing them is not a problem).

The initial condition is treated like this:

$$v_m^1 = n(y_l + mh; x + \mu(x)k, \sigma^2(x)k), \tag{18}$$

This is a first-order approximation to the density (in a certain sense). By definition of the Dirac-δ function this converges to the true initial condition as $k \to 0$. The Crank-Nicolson method tends to behave badly for non-smooth initial data (as pointed out in Strikwerda (1989)[p.121]), therefore the above choice is crucial.

A very important point to make here is the following: The usual Taylor expansion analysis of finite difference schemes gives only local statements; when step sizes tend to 0 local errors become smaller, but there are a lot more of them (at any point the calculated solution depends on the calculated solution on earlier points in the grid because of the recursive structure), so potentially anything might happen. Problems of this nature can easily (and judging from part of the finance literature: undetectedly) occur when PDEs with non-smooth initial data or boundary conditions not of Dirichlet-type are solved numerically using the Crank-Nicolson scheme. Using a simple and effective technique developed in Østerby (1998), Poulsen (1999) demonstrate that second order accuracy holds globally in both the time- and space-dimension for the problems considered in this paper.

Clearly, if the order in the time- and space-dimension are the same, it is (asymptotically) computationally efficient the keep the step sizes proportional (i.e. let them tend to 0 at the same rate), where the above mentioned error/order analysis can be used to find the constant of proportionality.

3.5 Hermite Expansion

The idea of expanding an (unknown) density function by the use of Hermite polynomials go back to Cramér (1925). Unfortunately he showed that the class of densities, for which the Hermite expansion converges, is rather limited. The density has to be "almost" normal to be in the class. Since most of the transition densities looked at in finance do not belong to that class the idea of Hermite expansions has not been of much use in finance, until recently.[8] A novel paper by Ait-Sahalia (1998) has changed this view completely. The main idea of his paper is to transform the diffusion (1) into another diffusion with a density that belongs to the class of converging Hermite

[8]The, perhaps, most used density in finance, the log-normal density, does not belong to the class of densities that allow a Hermite expansion. See Ait-Sahalia (1998).

expansions. A brief overview of the transformations, assumptions, main results and some 'trick of the trade' now follows. For a thorough understanding and in-depth analysis of the idea of Hermite expansions of a financial transition density, the reader is referred to Ait-Sahalia (1998).

As before it is assumed that the underlying process follows a diffusion given by:

$$dX_t = \mu(X_t)\,dt + \sigma(X_t)\,dW_t.$$

Recall that $D_X = (\underline{x}, \overline{x})$ denotes the domain of X. Two assumption regarding the behaviour of the diffusion X are needed. Here they come:

Assumption 3 ***Smoothness***
The functions $\mu(x)$ and $\sigma(x)$ are infinitely differentiable in x on D_X.

Assumption 4 ***Non-Degeneracy of the Diffusion***

1. *If $D_X = (-\infty, \infty)$, then there exists a constant c such that $\sigma(x) > c > 0$ for all $x \in D_X$.*

2. *If $D_X = (0, \infty)$ and $\sigma(0) = 0$ then there exist constants $\varepsilon_0 > 0$, $a \geq 0$, $b \geq 0$ such that $\sigma(x) \geq ax^b$ for all $0 < x \leq \varepsilon_0$. Furthermore: for all $\varepsilon > 0$ there exists a constant c_ε such that $\sigma(x) \geq c_\varepsilon > 0$ for all $x > \varepsilon$.*

The first transformation of the diffusion X is given by the following function:

$$Y_t \equiv f(X_t) = \int_{x^\#}^{X_t} \frac{1}{\sigma(u)}\mathrm{d}u,$$

where $x^\#$ is an arbitrary point in D_X. Since $\sigma > 0$ the function f is increasing and invertible. The domain of Y denoted $D_Y = (\underline{y}, \overline{y})$ is defined by:

$$\begin{aligned} \underline{y} &\equiv \lim_{x \to \underline{x}} f(x), \text{ and} \\ \overline{y} &\equiv \lim_{x \to \overline{x}} f(x). \end{aligned}$$

Using Itô's Lemma yields:

$$\mathrm{d}Y_t = \mu_Y(Y_t)\,\mathrm{d}t + \mathrm{d}W_t,$$

where

$$\mu_Y(y) = \frac{\mu(f^{-1}(y))}{\sigma(f^{-1}(y))} - \frac{1}{2}\frac{\partial \sigma}{\partial x}\left(f^{-1}(y)\right). \tag{19}$$

We see that the first transformation has produced a diffusion with unit volatility, i.e. we are "getting nearer" the class of converging Hermite expansions. Before stating the main result an assumption regarding the behaviour of Y is needed. A detailed discussion of this assumption is given in Ait-Sahalia (1998).

Assumption 5 ***Boundary behaviour***
$\mu_Y(y)$, $\frac{\partial \mu_Y(y)}{\partial y}$ and $\frac{\partial^2 \mu_Y(y)}{\partial y^2}$ have at most exponential growth near the infinity boundaries. Furthermore:

$$\begin{aligned} \lim_{y \to \underline{y}} -\frac{1}{2}\left(\mu_Y^2(y) + \frac{\partial \mu_Y(y)}{\partial y}\right) &< +\infty, \text{ and} \\ \lim_{y \to \overline{y}} -\frac{1}{2}\left(\mu_Y^2(y) + \frac{\partial \mu_Y(y)}{\partial y}\right) &< +\infty. \end{aligned}$$

For the boundaries we have:

1. *Left boundary:*

   - *If $\underline{y} = 0^+$, there exists constants $\varepsilon_0 > 0$, a and b such that $\mu_Y(y) \geq a y^{-b}$ for all $0 < y \leq \varepsilon_0$, where either $b > 1$ and $a > 0$ or $b = 1$ and $a \geq \frac{1}{2}$.*
   - *If $\underline{y} = -\infty$, there exists constants $E_0 > 0$ and $K > 0$ such that $\mu_Y(y) \geq Ky$ for all $y \leq -E_0$.*

2. *Right boundary:*

   - *If $\overline{y} = +\infty$, there exists constants $E_0 > 0$ and $K > 0$ such that $\mu_Y(y) \leq Ky$ for all $y \geq E_0$.*
   - *If $\overline{y} = 0^-$, there exists constants $\varepsilon_0 > 0$, a and b such that $\mu_Y(y) \leq -a|y|^{-b}$ for all $-\varepsilon_0 < y \leq 0$, where either $b > 1$ and $a > 0$ or $b = 1$ and $a \geq \frac{1}{2}$.*

A last transformation is needed to reach the class of converging Hermite expansions. Define:

$$Z_t = \Delta^{-\frac{1}{2}}(Y_t - y_0),$$

where $y_0 = f(x_0)$. Before stating the main result from the paper by Ait-Sahalia (1998) here comes the relationships between the transition densities of the processes

X, Y and Z:

$$\begin{aligned}
\phi_Z(\Delta, y_0, z) &= \Delta^{\frac{1}{2}} \phi_Y \left(\Delta, y_0, \Delta^{\frac{1}{2}} z + y_0\right), \\
\phi_Y(\Delta, y_0, y) &= \Delta^{-\frac{1}{2}} \phi_Z \left(\Delta, y_0, \Delta^{-\frac{1}{2}} (y - y_0)\right), \qquad (20) \\
\phi_X(\Delta, x_0, x) &= \frac{\phi_Y(\Delta, f(x_0), f(x))}{\sigma(x)}, \qquad (21) \\
\phi_Y(\Delta, y_0, y) &= \sigma\left(f^{-1}(y)\right) \cdot \phi_X\left(\Delta, f^{-1}(y_0), f^{-1}(y)\right),
\end{aligned}$$

where the two first relations are merely due to a definition and the last two follows from the relationship between the processes X and Y. A few definitions are needed:

Definition 6

1. *The classical Hermite polynomials are :*

$$H_j(z) := e^{\frac{z^2}{2}} \frac{d^j}{dz^j} \left[e^{-\frac{z^2}{2}}\right], \ j \geq 0.$$

For instance we have: $H_0(z) \equiv 1$, $H_1(z) = -z$, $H_2(z) = z^2 - 1$, $H_3(z) = -z^3 + 3z$, $H_4(z) = z^4 - 6z^2 + 3$, $H_5(z) = -z^5 + 10z^3 - 15z$, $H_6(z) = z^6 - 15z^4 + 45z^2 - 15$.

2. *The J'th truncated density of Z is defined to be:*

$$\phi_Z^{(J)}(\Delta, y_0, z) \equiv n(z) \sum_{j=0}^{J} \eta_j(\Delta, y_0) H_j(z), \qquad (22)$$

where $n(z)$ *is the standard normal density and:*

$$\eta_j(\Delta, y_0) \equiv \frac{1}{j!} \int_{-\infty}^{+\infty} H_j(z) \phi_Z^{(J)}(\Delta, y_0, z) \, dz.$$

As in (20) and (21) we define the J'th truncated density of Y and X by:

$$\begin{aligned}
\phi_Y^{(J)}(\Delta, y_0, y) &= \Delta^{-\frac{1}{2}} \phi_Z^{(J)} \left(\Delta, y_0, \Delta^{-\frac{1}{2}} (y - y_0)\right), \qquad (23) \\
\phi_X^{(J)}(\Delta, x_0, x) &= \frac{\phi_Y^{(J)}(\Delta, f(x_0), f(x))}{\sigma(x)}. \qquad (24)
\end{aligned}$$

The main result now is:

Theorem 7 *Under assumptions 1-3, there exists* $\overline{\Delta} > 0$ *such that for every* $\Delta \in \left(0, \overline{\Delta}\right)$ *and* $(x_0, x) \in D_X^2$ *we have:*

$$\phi_X^{(J)}(\Delta, x_0, x) \longrightarrow_{J \to \infty} \phi_X(\Delta, x_0, x).$$

Notice that the three diffusions considered in this paper all fulfill assumptions 1-3. Having established the main theorem it is time to consider the practical implementation of it. First of all we fix a J. To compute $\phi_Z^{(J)}(\Delta, y_0, z)$ we need the coefficients $\eta_j(\Delta, y_0)$, $j = 0, \ldots, J$. Explicit calculations yield:

$$\eta_j(\Delta, y_0) = \frac{1}{j!}\mathrm{E}\left[H_j\left(\Delta^{-\frac{1}{2}}(Y_{t+\Delta} - y_0)\right)|Y_t = y_0\right].$$

This expectation can be evaluated using a Taylor-approximation around Δ. To use Taylor's formula the following lemma is relevant:

Lemma 8 *Under assumptions 1-3 let g be a function such that g and all its derivatives have at most exponential growth. Then for $\Delta \in \left(0, \overline{\Delta}\right)$, $y_0 \in \mathbf{R}$ there exists $\delta \in [0, \Delta]$ such that*

$$E[g(Y_{t+\Delta})|Y_t = y_0] = \sum_{j=0}^{J}\left(A^J \cdot g(y_0)\right)\frac{\Delta^j}{j!} + E\left[A^{J+1} \cdot g(Y_{t+\delta})|Y_t = y_0\right]\frac{\Delta^{J+1}}{(J+1)!},$$

where A is the infinitesimal operator of the diffusion Y defined by:

$$A : g \longmapsto \mu_Y(\cdot)\frac{\partial g}{\partial y}(\cdot) + \frac{1}{2}\frac{\partial^2 g}{\partial y^2}(\cdot),$$

and $A^j \cdot g(y_0)$ means A applied j times to the function $y \longmapsto g(y)$ and evaluated at $y = y_0$.[9]

Practical considerations then are: How many terms should be included in this Taylor-serie? As suggested in Ait-Sahalia (1998) one should first decide on the truncation point J and then Taylor-expand the mean such that the approximation of $\phi_Z^{(J)}(\Delta, x_0, x)$ has terms of at most order $\Delta^{\frac{J}{2}}$.

[9] For instance:

$$\begin{aligned}
A^0 \cdot g &= g, \\
A^1 \cdot g &= \mu_Y g' + \frac{1}{2}g'', \\
A^2 \cdot g &= \mu_Y\left(\mu_Y' g' + \mu_Y g'' + \frac{1}{2}g'''\right) + \frac{1}{2}\left(\mu_Y'' g' + 2\mu_Y' g'' + \mu_Y g''' + \frac{1}{2}g^{(4)}\right) \\
&= \mu_Y \mu_Y' g' + \mu_Y^2 g'' + \mu_Y g''' + \frac{1}{2}\mu_Y'' g' + \mu_Y' g'' + \frac{1}{4}g^{(4)},
\end{aligned}$$

where the functions $\mu_Y(\cdot)$ and $g(\cdot)$ are evaluated at $y = y_0$.

Following that suggestion, we fix $J = 6$, and therefore the approximations for η_j, $j = 0, \ldots, 6$ are:

$$
\begin{aligned}
\eta_0 &= \frac{1}{0!} \sum_{j=0}^{3} A^j \cdot H_0 \left(\Delta^{-\frac{1}{2}} \left(\cdot - y_0 \right) \right) \frac{\Delta^j}{j!} = 1, \\
\eta_1 &= -\mu_Y \sqrt{\Delta} - \left(\frac{1}{2} \mu_Y \mu_Y' + \frac{1}{4} \mu_Y'' \right) \left(\sqrt{\Delta} \right)^3 \\
&\quad - \left(\frac{1}{6} \mu_Y (\mu_Y')^2 + \frac{1}{6} \mu_Y^2 \mu_Y'' + \frac{1}{6} \mu_Y \mu_Y''' + \frac{1}{4} \mu_Y' \mu_Y'' + \frac{1}{24} \mu_Y'''' \right) \left(\sqrt{\Delta} \right)^5, \\
\eta_2 &= \frac{1}{96} \Delta \begin{pmatrix} 48\mu_Y^2 + 48\mu_Y' + 48\Delta\mu_Y^2\mu_Y' + 56\Delta\mu_Y\mu_Y'' + 32\Delta(\mu_Y')^2 + 16\Delta\mu_Y''' \\ +16\Delta^2\mu_Y^3\mu_Y'' + 3\Delta^2\mu_Y''''' + 21\Delta^2(\mu_Y'')^2 + 16\Delta^2(\mu_Y')^3 \\ +28\Delta^2\mu_Y^2\mu_Y''' + 16\Delta^2\mu_Y\mu_Y'''' \\ +32\Delta^2\mu_Y'\mu_Y''' + 88\Delta^2\mu_Y\mu_Y''\mu_Y' + 28\Delta^2\mu_Y^2(\mu_Y')^2 \end{pmatrix}, \\
\eta_3 &= \frac{1}{6} \begin{bmatrix} -3\left(\sqrt{\Delta}\right)^3 \mu_Y\mu_Y' - \left(\sqrt{\Delta}\right)^3 \mu_Y'' - \frac{7}{2}\left(\sqrt{\Delta}\right)^5 \mu_Y(\mu_Y')^2 \\ -\frac{11}{4}\left(\sqrt{\Delta}\right)^5 \mu_Y^2\mu_Y'' - \left(\sqrt{\Delta}\right)^3 \mu_Y^3 - \frac{7}{4}\left(\sqrt{\Delta}\right)^5 \mu_Y\mu_Y''' \\ -3\left(\sqrt{\Delta}\right)^5 \mu_Y'\mu_Y'' - \frac{3}{8}\left(\sqrt{\Delta}\right)^5 \mu_Y'''' - \frac{3}{2}\left(\sqrt{\Delta}\right)^5 \mu_Y^3\mu_Y' \end{bmatrix}, \\
\eta_4 &= \frac{1}{240} \Delta^2 \begin{pmatrix} 60\mu_Y^2\mu_Y' + 40\mu_Y\mu_Y'' + 30(\mu_Y')^2 + 10\mu_Y''' + 50\mu_Y^3\mu_Y''\Delta + 4\mu_Y'''''\Delta \\ +34(\mu_Y'')^2\Delta + 40(\mu_Y')^3\Delta + 50\mu_Y^2\mu_Y'''\Delta + 23\mu_Y\mu_Y''''\Delta + 52\mu_Y'\mu_Y'''\Delta \\ +180\mu_Y\mu_Y''\mu_Y'\Delta + 10\mu_Y^4 + 100\mu_Y^2(\mu_Y')^2\Delta + 20\Delta\mu_Y^4\mu_Y' \end{pmatrix}, \\
\eta_5 &= -\frac{1}{120} \left(\sqrt{\Delta} \right)^5 \left(5\mu_Y\mu_Y''' + 10\mu_Y^3\mu_Y' + 15\mu_Y(\mu_Y')^2 + 10\mu_Y'\mu_Y'' + 10\mu_Y^2\mu_Y'' + \mu_Y^5 + \mu_Y^{(4)} \right), \\
\eta_6 &= \frac{1}{720} \Delta^3 \begin{pmatrix} 10(\mu_Y'')^2 + \mu_Y^{(5)} + 15(\mu_Y')^3 + 45\mu_Y^2(\mu_Y')^2 + 6\mu_Y\mu_Y^{(4)} + 15\mu_Y'\mu_Y''' + 20\mu_Y^3\mu_Y'' \\ +15\mu_Y^4\mu_Y' + 15\mu_Y^2\mu_Y''' + 60\mu_Y\mu_Y''\mu_Y' + \mu_Y^6 \end{pmatrix}.
\end{aligned}
$$

Using (22) give us the approximations for $\phi_Z^{(j)}(\Delta, x_0, x)$, $j = 0, \ldots, 6$. Then using (23) and (24) yields the required approximations for $\phi_X^{(j)}(\Delta, x_0, x)$, $j = 0, \ldots, 6$.
To complete this section here comes the transformation from the process X to the process Y for the three investigated diffusions:

- The Vasicek model: The appropriate transform and its inverse are given in (5) and (6), so

$$
\mu_Y(y) = \frac{\kappa(\theta - \sigma_V y)}{\sigma_V} = \frac{\kappa\theta}{\sigma_V} - \kappa y.
$$

- The CIR model: The f transform and its inverse are given in (7) and (8) and therefore:

$$\begin{aligned}\mu_Y(y) &= \frac{\kappa\left(\theta - \frac{1}{4}\sigma_{CIR}^2 y^2\right)}{\sigma_{CIR}\sqrt{\frac{1}{4}\sigma_{CIR}^2 y^2}} - \frac{1}{2}\left[\frac{1}{2}\frac{\sigma_{CIR}}{\sqrt{\frac{1}{4}\sigma_{CIR}^2 y^2}}\right] \\ &= \frac{\frac{2\kappa\theta}{\sigma_{CIR}^2} - \frac{1}{2}}{y} - \frac{\kappa y}{2}.\end{aligned}$$

- The Black-Scholes model: The f transform and its inverse are given in (9) and (10), and from this we get

$$\begin{aligned}\mu_Y(y) &= \frac{r\exp\{\sigma_{BS}y\}}{\sigma_{BS}\exp\{\sigma_{BS}y\}} - \frac{1}{2}\sigma_{BS} \\ &= \frac{r}{\sigma_{BS}} - \frac{1}{2}\sigma_{BS}.\end{aligned}$$

4 Implementation of Algorithms and Results

4.1 Tricks of the Trade

4.1.1 Computers & Random Numbers

The random numbers used in this study are either generated by a pseudo-generator or a quasi-generator as described in the theoretical section about simulation.
The pseudo generator uses the following 48 bit integer arithmetic linear congruential formula:

$$U_{n+1} = (aU_n + c) \text{ MOD } m, \; n \geq 1,$$

where $m = 2^{48}$, $c = 013$ (base 8) and $a = 0273673163155$ (base 8). To initialize the generator a commando using the first 32 bit of the current (physical) clock is used thereby ensuring (almost) independent runs of the generator.
The quasi generator uses the described Faure sequences. The base of the Faure sequences is the smallest prime number (p) equal to or bigger than the number of time-approximations (in the above denoted M), i.e., $M = 2 \Rightarrow p = 2$, $M = 4 \Rightarrow p = 5$, $M = 8 \Rightarrow p = 11$, $M = 16 \Rightarrow p = 17$, $M = 32 \Rightarrow p = 37$ and $M = 64 \Rightarrow p = 67$. The algorithm is initialized as described in the theoretical section by starting at the value p^4.

4.1.2 Variance Reduction

There are a number of different ways to decrease the standard error $\sigma\left[\widetilde{f}\left(X_T\right)\right]$. Some are generic, some depend (highly) on the random variable X_T and the pay-off function $\widetilde{f}(\cdot)$. For a survey see Boyle et al. (1997). Next, the two most common standard error reduction methods will be described.

1. - The antithetic variates technique: This method of variance reduction is generic. The basic idea behind it is simply that a realization of a symmetric random variable Z is paired with the identical distributed random variable $-Z$. For instance, the processes considered in this paper are all of the form:

   $$dX_t = \mu\left(X_t\right)dt + \sigma\left(X_t\right)dW_t,$$

   where W is a Brownian Motion. Simulating one realization of this process (up to time T) requires M simulations of standard normal variables. Since the standard normal distribution is symmetric, the outcome of a simulation is equally likely with the negated outcome. This is used to generate two realizations of X_T, the first using the realizations $n_1, n_2, \ldots, n_M$ of the standard normal distribution, the second using the negated realizations $-n_1, -n_2, \ldots, -n_M$. Later on a brief discussion of how to generate realizations of X_T will be discussed.

   Before moving on to the next standard error reduction technique it should be noticed that the antithetic variates method is most efficiently used when options with a monotone pay-off (like vanilla call's and put's) are simulated.[10] In section 3.2 it was argued that simulating densities can be viewed as simulating some binary options, which are not monotone in their pay-off. Therefore it might be questioned how efficient antithetic variates are when used to simulate densities. Despite this question the method will be used in this numerical examination mainly because of the ease of inclusion in the computer-programming.

   - The control variates technique: This method depends on the relationship between a known (analytical) price for an option and the price of an option found by simulation. For example; let c_0^{est}, p_0^{est} and p_0^{an} denote the price of an option, c, found by simulation; the price of an option, p, found by

[10] A detailed discussion about this result can (again) be found in Boyle et al. (1997).

simulation and the price of the option, p, found analytically, respectively. The control variate estimate for the option, c, is then given by:

$$c_0^{cv} \equiv c_0^{est} + \left(p_0^{an} - p_0^{est}\right).$$

It is seen that this estimate is unbiased. Furthermore calculations yield the following estimate of the standard error of the estimated option price:

$$\mathrm{STD}\left(c_0^{cv}\right) = \sqrt{\mathrm{Var}\left(c_0^{est}\right) + \mathrm{Var}\left(p_0^{est}\right) - 2\mathrm{Cov}\left[c_0^{est}, p_0^{est}\right]}.$$

Therefore, the standard error of the estimate c_0^{cv} will be less than the standard error of the estimate c_0^{est} if the covariance between the options c and p is positive and large enough. If this covariance instead is negative and "large" one can take the control variate estimate to be:

$$c_0^{cv'} \equiv c_0^{est} + \left(p_0^{est} - p_0^{an}\right),$$

and the standard error of the estimate c_0^{cv} will again be less than the standard error of the estimate c_0^{est}. In fact one can find the best possible control variate by minimizing:

$$c_0^{cv''} \equiv c_0^{est} + \omega\left(p_0^{an} - p_0^{est}\right).$$

The optimal choice of ω is found to be:

$$\omega^* = \frac{\mathrm{Cov}\left[c_0^{est}, p_0^{est}\right]}{\mathrm{Var}\left(p_0^{est}\right)}.$$

This means that an efficient use of this method requires some knowledge about the structure of the options c and p, i.e., this is a highly problem-dependent method.

In this study we will always use the antithetic technique since it is totally independent of the considered problem. On the other hand: Since the control variates method is so problem-dependent it has been excluded from the numerical investigation done in this paper. However one should notice that for small time-steps, Δ, the density-function $\phi(\Delta, x, y)$ is approximately normal (see section 3.1). Therefore using a normal density function as a control variate might decrease the standard error of the estimated density function.

4.1.3 Extrapolation

Extrapolation is a classical numerical discipline. This section describes the technique commonly known as 'Richardson extrapolation'.
Suppose we know that v_h is a first order approximation of u,

$$v_h(y) = u(y) - hf(y) - h^2 g(y) + o(h^2) \tag{25}$$

Equation (25) should be understood in following way: We choose y and h and are then able to compute the LHS, but we do not know the functions u, f and g. Let $e_h(y) = u(y) - v_h(y)$ denote the (unknown) error. Performing the calculation for step sizes h and $h/2$ we get

$$\begin{aligned} v_h(y) &= u(y) - hf(y) - h^2 g(y) + o(h^2), \\ v_{h/2}(y) &= u(y) - \frac{h}{2} f(y) - \frac{h^2}{4} g(y) + o(h^2). \end{aligned}$$

Note that

$$e_{h/2} = \frac{h}{2} f + \frac{h^2}{4} g + o(h^2)$$

and that

$$v_{h/2} - v_h = \frac{h}{2} f + \frac{3h^2}{4} g + o(h^2).$$

Therefore, if h is small, the error in the '$h/2$'-computation is well approximated by $v_{h/2} - v_h$ and we may use

$$v^R = 2v_2 - v_1$$

as an approximation to u. The point here is that v^R is likely to be a considerably better approximation than $v_{h/2}$ and is obtained cheaply.
If we are dealing with a second order approximation, i.e. $f \equiv 0$, then similar analysis shows that

$$v^R = v_{h/2} + \frac{1}{3}(v_{h/2} - v_h)$$

is likely to be a better approximation to u than $v_{h/2}$.
The argument can be extended to functions with multiple arguments in a straightforward manner.

We have said nothing as to how v_h is calculated or how we know that (25) is valid. Of course these two problems are intimately related. Typically, a first attempt of justification of (25) (or: an investigation of which terms are likely to be in there) is based on Taylor-expansions. But, as discussed in Section 3.4.1 this not always enough if v_h is obtained through a series of calculations each involving an error. Formal analysis of error propagandation or global error analysis is difficult. But as pointed out in Østerby (1998) experimental verification can be done by considering the (computable) ratio

$$\frac{v_h - v_{h/2}}{v_{h/2} - v_{h/4}} = 2\frac{f + 6hg + \dots}{f + 3hg + \dots}.$$

for a large number of x's. If h is small and (25) is valid then we should observe many numbers close to 2.0 for a first order method while a second order method is likely to produce numbers in the vicinity of 4.0.

We shall be using Richardson extrapolation for the CSBMs and the finite difference method. For CSBMs we only have one quantity to vary, N, and the extrapolation is based on first order accuracy. (Technically, we are forced to use N and $4N$ to ensure overlapping terminal nodes.) For the Crank-Nicolson method we can vary time and space steps independently and the extrapolation based on second order accuracy in both directions. (This after both order assumptions have been justified by the 'ratio'-computation.)

Note that the extrapolated values are only available at points corresponding to the coarsest discretization. We may then use some interpolation scheme or we may be so fortunate that it is a specific y-value that is of interest to us, we then make sure that this y-value belongs the the coarsest discretization.

4.2 Numerical Results

Since the object we are trying to approximate is a whole function,[11] there are many ways of measuring the quality of the approximation. We give two error measures, one absolute, one relative. (The y_i's are the grid/lattice points.)

[11]Strictly speaking, we are trying to approximate a function of three variables, $t \sim$ time, $x \sim$ "current state", and $y \sim$ "future state". We give only error measurements related to one of these, namely y, while keeping the others fixed, since the "error in y" is often the most relavant one. Further our studies indicate that results are insensitive to the chosen values of t and x.

- Maximal absolute error,

$$e_1 = \sup_i \{|\phi^A(y_i) - \phi(y_i)|\}.$$

Evidently, this is a reasonable way to measure the error. We should be aware that our convergence results are mostly of a pointwise nature, so since e_1 is a uniform measure, we could be in for a few surprises. Further note that, generally, it does not make sense to compare e_1's for different models.

- Average relative error (quoted in parts per million in tables),

$$\begin{aligned} e_2 &= \sum_i |\phi^A(y_i) - \phi(y_i)| \triangle y \\ &\approx \int_{\mathbf{R}} |\phi^A(y) - \phi(y)| dy = \int_{\{y|\phi(y)>0\}} \frac{|\phi^A(y) - \phi(y)|}{\phi(y)} \phi(y) dy. \end{aligned}$$

The relative error tells us how many significant digits of the solution we can trust. One could wonder why do not measure, say, the maximal relative error. This would, however, not produce very informative numbers, since the functions we are working with are very close to 0 in large areas. But fortunately we have a reasonable "tool" for measuring the importance of the relative errors, namely the function itself. Errors of this type are supposed to make sense across models.

We use the following, financially realistic parameters.
The Vasicek model:

$\theta = 0.08$	$\kappa = 0.24$	$\sigma_V = 0.025$
$x_0 = 0.08$	$\triangle = 1/12$	

The Cox-Ingersoll-Ross model:

$\theta = 0.08$	$\kappa = 0.24$	$\sigma_{CIR}\sqrt{\theta} = 0.025 \Rightarrow \sigma_{CIR} = 0.08838$
$x_0 = 0.08$	$\triangle = 1/12$	

The Black-Scholes model:

$r = 0.08$	$\sigma_{BS} = 0.25$
$x_0 = 100$	$\triangle = 1$

Technique specific parameters are indicated in the table-captions.
'Time' is the CPU-time (in seconds) on a HP-9000 Unix machine.
Tabular evidence, which allows for easy comparison with results obtained "in the comfort in privacy of your own home", is given only for the CIR model. Similar tables for the Vasicek-model and the BS-model are available from the authors upon request.

As pointed out earlier skeptics would say that we are using too small a $\triangle$-value to make interesting/reasonable comparisons; for such small time steps all densities are quite close to normal. But Table 1 reveals that we have no problems detecting non-normality. This does not mean that quantities, typically estimators or asset prices, found using such approximations are unreasonable, but it is encouraging for the suggestion, implementation , comparison and application of refined density approximation techniques.

From Table 2 we see that in reasonable time we get better approximations with the binomial approximations than with the simulation approach (cf. Tables 5-13).

Table 3 concerns the Crank-Nicolson method. We see that when the number of time and space steps is doubled, the errors are roughly reduced by a factor 4, as we would expect them to be if we have a second order accurate approximation. Therefore the extrapolation is fruitful. Note that there is nothing wrong with errors being increasing in one discretization step size if the other is kept fixed; it just means that the "time" and "space errors" have opposite sign (cf. Section 4.1.3). There are some difficulties for very small step sizes; we attribute this to a "computer number representation effect", we are performing large number of operations with small numbers.

From Table 4 we see that very low errors are produced for all of the Hermitian approximations. Furthermore none of the approximations are very time-consuming. One might even be tempted to ask the question (for the fifth and the sixth approximation): "Which approximation of the modified Bessel-function is best - The one using a fifth or sixth Hermitian approximation or the one proposed in Press et al. (1992)?"

In Tables 5-13 we see the results of different simulations. It is evident that for a small number of time approximations (M) the Milstein-scheme is superior to the the Euler-scheme. However when M is big we see almost no difference. Looking at the tables produced using pseudo-numbers we see - as expected - that for a large number of simulations (N) the errors e_1 and e_2 decrease at the rate $\sqrt{N}$. Considering the tables produced using quasi-numbers we again see that the errors e_1 and e_2 decrease at the rate $\sqrt{N}$. From a theoretical point of view this is more surprising, at least if we think that the the number of simulations are large enough to cover some of the asymptotic properties of quasi-numbers. However, the same findings comparing quasi- and pseudo-numbers are reported in Berman (1998). For all the tables we find that the number of simulations is more important than the number of time approximations which supports the theorem regarding trade-off between the number of simulations and the number of time approximations found in Duffie & Glynn (1995). The last thing to notice from the tables is that the production of quasi-numbers give smaller errors e_1 and e_2 but the technique is significantly more time-consuming than the similar production of pseudo-numbers.

Figures 2-4 depict

$$e(y_i) = \phi^A - \phi$$

over (most of) the range of the space variable for different models and approximations. For each model the approximation parameters have been chosen such that the e_1's are roughly the same size. We note a strong similarity between the Crank-Nicolson and the Hermitian approximation. And as a further regularity we see that $\arg\max e \approx \arg\max \phi$.

The 6 (time, error) graphs (Figures 5-5) give a clear picture of the ranking of the methods considered in this paper. When the computational effort needed to produce the quasi-numbers is taken into account then the error behave roughly as when pseudo-random numbers are used - although the histograms analysed with quasi-numbers have a smaller number of intervals than the histograms analysed with pseudo-numbers (see also Tables 5-13). But these simulation methods are easily outperformed by the binomial method. Binomial approximation is then outperformed by the technique that uses the Crank-Nicolson method to solve the forward PDE, despite this method having "something of a reputation" for being "over-sensitive". But really neither of those methods can compete with the Hermitian expansion. The Hermite approximation does require some rather tedious algebraic calculations, but they are well worth the effort. Note also that the figures are very similar across error-types and model specifications. This is also encouraging.

5 Conclusion

A clear ranking of the different approximation techniques emerges from this study. Figures 5-5 give the result when two different errors e_1 (maximal absolute) and e_2 (average relative) are compared with the time consumed on a computer.

Among the methods we have considered to approximate the transition density of a diffusion process, the Hermite polynomial expansion described in Section 3.5 is inarguable the best one when the speed/accuracy trade-off is considered. The second best approximation-technique is to use the Crank-Nicolson method to solve the partial differential equation for the transistion density as described in Section 3.4. The third best approximation stems from the the binomial technique described in Section 3.3. The significantly worst way to approximate the density by is the simulation approach described in Section 3.2. However if one insists on simulating diffusions this study finds that pseudo-random numbers should be used - at least from a time-consumption point of view.

The issues flexibility (by which we for instance mean how easy it is to change the code if the volatility function is changed) and algorithm-complexity are very much in the mind of the beholder. Once the program has been made and is running correctly, almost any programmer would say that it is really flexible and not very complex! By giving very detailed descriptions we hope to help people make informed decisions for themselves. Nonetheless a brief discussion is in order.

- Hermite Expansion: This method is relatively inflexible. If μ or σ is changed a new calculation of (19) is needed. Furthermore derivatives of this function are required - something that might be calculated automatically depending one the program used. Regarding complexity of the algorithm we have just one thing to say: Quite easy! Just program the formulae for η_j, μ_Y and H_j and everything is running!

- Crank-Nicolson: Once the code for has been written it is flexible. The programming does require careful book-keeping and storage as well as an algorithm for solving tridiagonal linear systems. This is not too hard when a "real programming language" is used, but if you only have a spreadsheet is would be more difficult. The method also requires some "fingerspitzgefühl", for example when the ratio of time steps to space steps is chosen, but those that exercise this will be duly rewarded.

- Binomial: As with the Hermitian Expansion this method is made somewhat inflexible by the need for a transformation that can really only be done "by hand". But when this calculation has been performed the actual programming is quite easy because of the explicit structure (we do not have to solve large systems of equations).

- Simulation: This method is flexible. If μ or σ is changed the only line of coding that has to be changed is the one containing the approximation scheme ((3) or (4)). It is a relatively complex thing to program when starting "from scratch". First one has to produce a uniform generator (pseudo or random). Then a normal transform is needed (Box-Muller or Moro). Then the required loops and the discretization scheme (3) or (4) are programmed. Finally a loop determining the histogram is needed.

An obvious continuation of this study is to examine multi-dimensional diffusions. While not impossible, neither in theory nor in practice, it would be a quite heavy task. For instance if we consider a two-dimensional diffusion different techniques using the PDE-approach are needed. Furthermore, in principle one could use a two-dimensional version of the Hermite expansions proposed above. Until now these two-dimensional expansions have not been reported in the financial literature - perhaps because of the dramatically increased complexity of the problem. On the other hand, simulations of multi-dimensional diffusions will not be a problem. An almost identical problem as the one in the one-dimensional diffusion case arise. Therefore simulations might be the appropriate choice when multi-dimensional diffusions are considered.

References

Ait-Sahalia, Y. (1998), Maximum Likelihood Estimation of Discretely Observed Diffusion Processes: A Closed-Form Approach. Unpublished, University of Chicago.

Berman, L. (1998), 'Accelerating Monte Carlo: Quasirandom Sequences and Variance Reduction', *Journal of Computational Finance* **1**(2), 79–95.

Boyle, P. (1977), 'Options: A Monte Carlo Approach', *Journal of Financial Economics* **4**, 323–338.

Boyle, P., Broadie, M. & Glasserman, P. (1997), 'Monte Carlo Methods for Security Pricing', *Journal of Economic Dynamics and Cotrol* **21**, 1267–1321.

Cox, J. C., Ingersoll, J. E. & Ross, S. A. (1985), 'A Theory of the Term Structure of Interest Rates', *Econometrica* **53**, 385–407.

Cox, J., Ross, S. & Rubinstein, M. (1979), 'Option pricing: A simplified approach', *Journal of Financial Economics* **7**, 229–263.

Cramér, H. (1925), On Some Classes of Series Used in Mathematical Statistics, *in* 'Proceedings of the Sixth Scandinavian Mathematical Congress', pp. 399–425.

Duffie, D. & Glynn, P. (1995), 'Efficient monte carlo simulation of security prices', *The Annals of Applied Probability* **5**(4), 897–905.

Ethier, S. N. & Kurtz, T. G. (1986), *Markov Processes: Characterization and Convergence*, Wiley.

Galanti, S. & Jung, A. (1997), 'Low-Discrepancy Sequences: Monte Carlo Simulation of Option Prices', *Journal of Derivatives* pp. 63–83.

He, H. (1990), 'Convergence from discrete- to continuous-time contingent claims prices', *Review of Financial Studies* **3**, 523–546.

Joy, C., Boyle, P. & Tan, K. (1996), 'Quasi-Monte Carlo Methods in Numerical Finance', *Management Science* **42**(6), 926–938.

Karatzas, I. & Shreve, S. (1992), *Brownian Motion and Stochastic Calculus*, 2 edn, Springer-Verlag.

Karlin, S. & Taylor, M. H. (1981), *A Second Course in Stochastic Processes*, Academic Press.

Kloeden, P. E., Platen, E. & Schurz, H. (1991), *Numerical Solution of SDE Through Computer Experiments*, Springer.

Nelson, D. & Ramaswamy, K. (1989), 'Simple Binomial Processes as Diffusion Approximations in Financial Models', *Review of Financial Studies* **3**, 393–430.

Østerby, O. (1998), The Error of the Crank-Nicolson Method for Linear Parabolic Equations with a Derivative Boundary Condition. Working paper. Department of Computer Science. University of Aarhus.

Poulsen, R. (1999), Approximate Maximum Likelihood Estimation of Discretely Observed Diffusion Processes. Working paper no. 29, Centre for Analytical Finance, University of Aarhus.

Press, W., Teukolsky, S., Vetterling, W. T. & Flannery, B. (1992), *Numerical recipes in C*, 2. edn, Cambridge University Press.

Silverman, B. (1986), *Density estimation for Statistics and Data Analysis*, Chapman and Hall.

Strikwerda, J. (1989), *Finite Difference Schemes and Partial Differential Equations*, Wadsworth & Brooks/Cole.

Stroock, D. W. & Varadhan, S. R. S. (1979), *Multidimensional Diffusion Processes*, Springer.

Vasicek, O. (1977), 'An Equilibrium Characterization of the Term Structure', *Journal of Financial Economics* **5**, 177–188.

A Tables

Model	Euler		"Correct Moment"	
	e_1	e_2	e_1	e_2
Vasicek	0.5535	9643	0	0
CIR	2.001	35164	1.79	33996
BS	0.0027	188990	0.0031	191891

Table 1: Analytical normal approximations to densities.

M	Errors before ex'pol e_1	e_2	Errors after ex'pol e_1	e_2	Time incl. ex'pol
40	0.34778	5999	0.27426	4916	0.02
80	0.17436	2914	0.19916	2013	0.05
160	0.08720	1455	0.03634	736	0.17
320	0.04673	728	0.02534	249	0.73
640	0.02327	369	0.00769	130	3.57
1280	0.01316	190	0.00750	57	21.6

Table 2: Finding the CIR-density by binomial approximation.

M	N	Errors before ex'pol e_1	e_2	Errors after ex'pol e_1	e_2	Time incl. ex'pol
50	2	1.7637	19500	N/A	N/A	0.02
50	4	0.1884	3761	N/A	N/A	0.02
50	8	0.3748	6580	N/A	N/A	0.02
50	16	0.5087	8340	N/A	N/A	0.04
100	4	0.3748	5797	0.2037	3678	0.05
100	8	0.0526	1010	0.0294	524	0.05
100	16	0.1084	1747	0.0336	617	0.05
100	32	0.1267	2129	0.0611	800	0.06
200	8	0.0983	1636	0.0271	470	0.09
200	16	0.0150	270	0.0036	98	0.09
200	32	0.0257	448	0.0099	152	0.11
200	64	0.0328	538	0.0123	179	0.14
400	16	0.0262	435	0.0067	99	0.17
400	32	0.0051	80	0.0023	35	0.20
400	64	0.0085	120	0.0024	39	0.22
400	128	0.0093	136	0.0036	43	0.40
800	32	0.0074	113	0.0030	32	0.38
800	64	0.0022	24	0.0016	14	0.55
800	128	0.0032	31	0.0013	19	1.11
800	256	0.0032	35	0.0016	20	2.12

Table 3: Finding the CIR-density by the Crank-Nicolson method.

Approx.	e_1	e_2	Time
One	0.514300	8936	0.04
Two	0.028410	525	0.05
Three	0.006744	98	0.06
Four	0.001785	16	0.07
Five	0.001208	12	0.09
Six	0.001259	12	0.09

Table 4: Finding the CIR-density by Hermite expansions.

M	N	e_1	std(e_1)	e_2	std(e_2)	Time
2	20000	2.0988	0.7539	25884	3857	0.22
	40000	1.8398	0.5389	24213	3884	0.44
	80000	1.6914	0.4145	22471	2744	0.88
	160000	1.3864	0.2212	20320	2218	1.75
	320000	1.2909	0.1454	19592	1001	3.44
4	20000	2.3825	0.9842	23731	7010	0.32
	40000	1.5030	0.4442	18739	2624	0.64
	80000	1.1979	0.3150	14528	2102	1.25
	160000	1.0855	0.1992	13933	2399	2.52
	320000	0.9188	0.1495	11960	624	5.17
8	20000	1.6487	0.7120	24548	7176	0.51
	40000	1.5849	0.4712	19506	4201	1.02
	80000	1.1064	0.3109	14021	2520	2.13
	160000	0.9160	0.2388	10859	1642	4.23
	320000	0.6520	0.1231	8565	1173	8.20
16	20000	1.8769	0.5501	22065	3857	0.94
	40000	1.3627	0.3471	17952	3664	1.82
	80000	0.9359	0.2112	12541	2457	3.62
	160000	0.6250	0.1303	9049	1666	7.49
	320000	0.6398	0.1890	7902	1908	14.50

Table 5: Finding the CIR-density by pseudo-simulation. Time-discretization-scheme: Euler. Number of intervals in histogram: 50.

M	N	e_1	std(e_1)	e_2	std(e_2)	Time
4	40000	2.0325	0.3586	23785	3069	0.82
	80000	1.8807	0.3460	20357	3456	1.64
	160000	1.2601	0.2282	13950	998	3.28
	320000	1.0711	0.2412	12338	682	6.54
	640000	0.9111	0.1511	10212	648	13.08
8	40000	2.5740	0.4677	25885	4009	1.22
	80000	1.5408	0.2717	17168	3124	2.43
	160000	1.0906	0.2521	12582	1501	4.85
	320000	0.8766	0.2500	8973	1259	9.72
	640000	0.6028	0.1508	7227	624	19.46
16	40000	2.3595	0.5138	23585	4078	2.00
	80000	1.8323	0.3207	17757	2110	4.01
	160000	1.0275	0.2223	11550	1549	8.07
	320000	0.8669	0.2610	9040	1925	16.14
	640000	0.6294	0.1311	6538	1234	32.23
32	40000	2.1270	0.3913	22105	2652	3.67
	80000	1.4633	0.2929	16618	2158	7.19
	160000	1.0085	0.1999	11283	1853	14.41
	320000	0.7801	0.2305	8533	1671	28.79
	640000	0.5258	0.1434	5994	739	57.45

Table 6: Finding the CIR-density by pseudo-simulation. Time-discretization-scheme: Euler. Number of intervals in histogram: 100.

M	N	e_1	std(e_1)	e_2	std(e_2)	Time
8	80000	2.7755	0.6351	26301	2346	3.07
	160000	1.5090	0.3063	17343	2411	6.16
	320000	1.3833	0.3579	12078	1017	12.30
	640000	0.9092	0.1915	9410	963	24.63
	1280000	0.7761	0.1225	7026	575	49.04
16	80000	2.5304	0.3652	24513	2430	4.74
	160000	1.7214	0.3731	17214	1384	9.41
	320000	1.2577	0.1509	12807	1363	18.78
	640000	0.8281	0.1912	8158	655	37.46
	1280000	0.6985	0.1131	6371	890	74.84
32	80000	2.7103	0.8436	24980	3512	7.87
	160000	1.7185	0.4061	16259	1296	15.72
	320000	1.2523	0.2730	12266	952	31.50
	640000	0.8915	0.1438	8519	652	6303
	1280000	0.6339	0.1435	6014	1016	125.87
64	80000	2.5626	0.6120	24284	2090	14.26
	160000	1.7733	0.4065	17217	2555	28.50
	320000	1.1017	0.2386	11096	1265	56.87
	640000	0.8503	0.0657	8507	535	113.65
	1280000	0.6506	0.1319	6013	802	227.38

Table 7: Finding the CIR-density by pseudo-simulation. Time-discretization-scheme: Euler. Number of intervals in histogram: 200.

M	N	e_1	std(e_1)	e_2	std(e_2)	Time
2	20000	1.8889	0.3936	24531	3472	0.22
	40000	1.1091	0.2204	16788	2458	0.45
	80000	1.0052	0.3169	13316	3662	0.91
	160000	1.0351	0.3607	11896	2971	1.73
	320000	0.7129	0.1301	10527	1669	3.47
4	20000	1.6832	0.4606	22487	2756	0.33
	40000	1.3011	0.4013	16499	2950	0.67
	80000	0.8084	0.2395	11654	2690	1.27
	160000	0.7437	0.1570	9950	1796	2.53
	320000	0.5997	0.1870	7808	1395	5.27
8	20000	1.8496	0.3677	22648	6739	0.54
	40000	1.1011	0.2551	15238	2073	1.02
	80000	0.9281	0.3256	11428	1836	2.20
	160000	0.8128	0.3216	9782	2651	4.37
	320000	0.6686	0.2083	7455	1776	8.18
16	20000	1.5504	0.5748	20468	5865	0.95
	40000	1.4159	0.5231	18517	6397	1.92
	80000	0.8691	0.3174	11011	3615	3.66
	160000	0.6324	0.1019	8618	1714	7.78
	320000	0.4673	0.2386	5831	1931	14.68

Table 8: Finding the CIR-density by pseudo-simulation. Time-discretization-scheme: Milstein. Number of intervals in histogram: 50.

M	N	e_1	std(e_1)	e_2	std(e_2)	Time
4	40000	2.1934	0.4444	22682	2996	0.87
	80000	1.5928	0.4594	17606	3870	1.65
	160000	1.0336	0.1277	11958	1524	3.47
	320000	0.7964	0.1503	8589	723	6.55
	640000	0.6381	0.1415	6452	906	14.04
8	40000	1.9430	0.5227	22363	3602	1.23
	80000	1.5470	0.2058	16669	2213	2.45
	160000	1.1777	0.3553	12589	2296	5.10
	320000	0.6667	0.1362	7943	922	9.90
	640000	0.5941	0.0640	6490	933	19.52
16	40000	2.3730	0.9505	21624	3967	2.05
	80000	1.5663	0.4590	17240	2895	4.05
	160000	1.0773	0.3282	12160	2569	8.54
	320000	0.8108	0.2235	8660	1561	16.37
	640000	0.4955	0.0750	6219	1413	32.59
32	40000	2.1046	0.4760	22937	2601	3.72
	80000	1.4085	0.4706	16569	3488	7.30
	160000	0.9792	0.3697	10617	2844	14.60
	320000	0.6995	0.1666	8472	2089	29.09
	640000	0.5718	0.1828	6008	938	58.13

Table 9: Finding the CIR-density by pseudo-simulation. Time-discretization-scheme: Milstein. Number of intervals in histogram: 100.

M	N	e_1	std(e_1)	e_2	std(e_2)	Time
8	80000	2.5351	0.2743	24193	2157	3.10
	160000	1.7121	0.5819	16559	1778	6.22
	320000	1.2690	0.2288	12221	1372	12.37
	640000	0.8922	0.1638	8803	900	24.76
	1280000	0.6674	0.1435	5835	733	49.58
16	80000	2.6482	0.7353	23764	3482	4.68
	160000	1.4988	0.2907	14969	1592	9.39
	320000	1.2621	0.3085	11785	1642	18.60
	640000	0.8049	0.1525	8354	1057	37.60
	1280000	0.6287	0.1675	5702	560	75.06
32	80000	2.4608	0.4005	24432	2669	7.87
	160000	1.8992	0.4566	16938	1749	15.73
	320000	1.2440	0.2473	11424	1550	31.48
	640000	0.8002	0.1322	7916	740	62.93
	1280000	0.6573	0.1596	5953	856	127.12
64	80000	2.7533	0.5735	23944	2971	14.25
	160000	1.9112	0.4446	17474	1751	28.45
	320000	1.1826	0.2676	11882	1231	56.95
	640000	0.9140	0.1970	8913	898	115.04
	1280000	0.6264	0.1623	5907	662	226.67

Table 10: Finding the CIR-density by pseudo-simulation. Time-discretization-scheme: Milstein. Number of intervals in histogram: 200.

		Euler			Milstein		
M	N	e_1	e_2	Time	e_1	e_2	Time
2	20000	1.4283	19140	3.85	0.6537	10604	3.90
	40000	1.0445	19095	9.02	0.8037	9632	10.38
	80000	1.0390	18788	20.84	0.5921	9251	20.96
	160000	1.0796	18407	47.93	0.5100	8858	48.38
	320000	1.1437	18559	110.32	0.5171	8578	110.30
4	20000	1.5360	23529	2.53	1.2360	20405	2.61
	40000	0.8765	12525	5.66	0.6733	9986	5.73
	80000	1.1890	12605	12.77	0.6890	10924	12.92
	160000	0.8921	11556	26.84	0.5608	8952	27.43
	320000	0.7499	11559	62.66	0.5850	7016	63.92
8	20000	1.4890	17697	4.28	1.5390	17287	4.33
	40000	1.2537	16434	8.56	1.1537	16986	8.54
	80000	0.6320	11544	17.01	0.6453	9499	17.39
	160000	0.6943	10133	33.96	0.4764	7331	34.67
	320000	0.5945	9643	70.31	0.5827	7607	71.36
16	20000	1.1858	17144	8.71	1.2812	17322	8.73
	40000	1.6162	14934	17.53	1.7287	15209	17.53
	80000	1.2350	10267	35.39	1.2412	10069	35.11
	160000	0.7881	8047	70.10	0.8225	7066	72.05
	320000	0.3828	5685	140.92	0.3383	4704	142.01

Table 11: Finding the CIR-density by quasi-simulation. Number of intervals in histogram: 50.

		Euler			Milstein		
M	N	e_1	e_2	Time	e_1	e_2	Time
4	40000	1.5456	20714	5.81	1.5713	18862	6.03
	80000	1.1908	16829	12.96	1.0516	13072	13.39
	160000	0.8831	14132	31.66	0.6984	11574	28.24
	320000	0.7454	11507	63.27	0.5234	7542	65.82
	6400000	0.7164	9991	135.83	0.3950	5931	141.20
8	40000	1.9033	22039	8.72	2.1065	23022	8.80
	80000	1.3258	14644	17.23	0.9383	13017	18.16
	160000	1.0149	11374	34.47	0.6649	10165	35.46
	320000	1.0367	11467	71.18	0.7483	11189	73.47
	640000	0.6430	8392	166.89	0.5421	7558	169.60
16	40000	2.7046	21397	17.78	2.6796	21154	20.97
	80000	1.6921	17335	35.43	1.6671	16588	42.28
	160000	1.1234	11503	70.69	1.2921	12241	72.71
	320000	0.8110	7972	141.20	0.7794	8800	142.93
	640000	0.5969	5005	282.49	0.3536	5233	286.03
32	40000	1.7741	20776	35.50	1.7991	20074	36.08
	80000	1.6845	16249	71.29	1.6991	16321	72.15
	160000	0.7783	8180	142.29	0.7616	9408	146.50
	320000	0.7602	6533	285.50	0.7945	7256	287.60
	640000	0.5781	6738	571.92	0.5774	6547	581.77

Table 12: Finding the CIR-density by quasi-simulation. Number of intervals in histogram: 100.

		Euler			Milstein		
M	N	e_1	e_2	Time	e_1	e_2	Time
8	80000	2.2570	19810	17.79	1.9570	21344	18.04
	160000	1.4706	14883	35.82	1.4013	14783	36.28
	320000	1.2456	13238	73.67	1.2776	13586	73.99
	640000	0.8689	10028	169.34	0.6745	9389	170.56
	1280000	0.4938	6432	360.87	0.3921	4638	366.19
16	80000	2.4710	21716	35.66	2.4807	19496	36.96
	160000	1.6507	18210	71.02	1.7182	16710	72.37
	320000	1.3317	10174	142.62	1.1442	11564	147.78
	640000	0.7297	7087	285.67	0.6752	7707	293.45
	1280000	0.6750	6802	570.33	0.5265	6719	588.07
32	80000	2.7018	20034	71.64	2.3018	20690	74.23
	160000	1.1388	13640	143.63	1.1491	14408	149.03
	320000	1.1138	11362	287.59	1.0179	11717	295.77
	640000	0.9820	8470	574.84	1.1570	8726	589.77
	1280000	0.6945	6228	1148.19	0.7086	6103	1182.77
64	80000	2.8557	27458	143.01	3.0057	27181	149.19
	160000	1.7642	15456	286.52	1.5017	16065	298.64
	320000	1.1517	10522	574.45	1.1205	10933	600.07
	640000	0.7866	7814	1149.16	0.7732	7689	1186.73
	1280000	0.5914	5486	2293.88	0.6289	5104	2364.65

Table 13: Finding the CIR-density by quasi-simulation. Number of intervals in histogram: 200.

B Figures

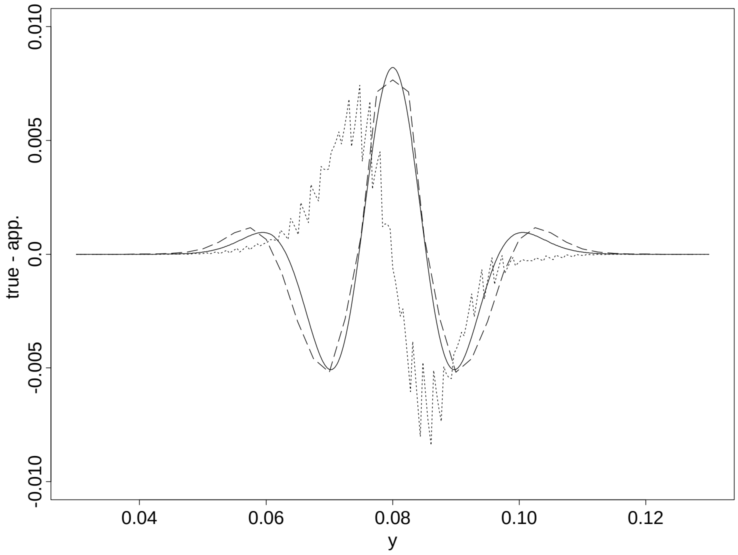

Figure 2: "True- app." density for the Vasicek model. —:Hermite ($t < 0.01$ s), - - -: Fin. Diff. ($t = 0.05$ s), ...:Binom ($t = 4.0$s)

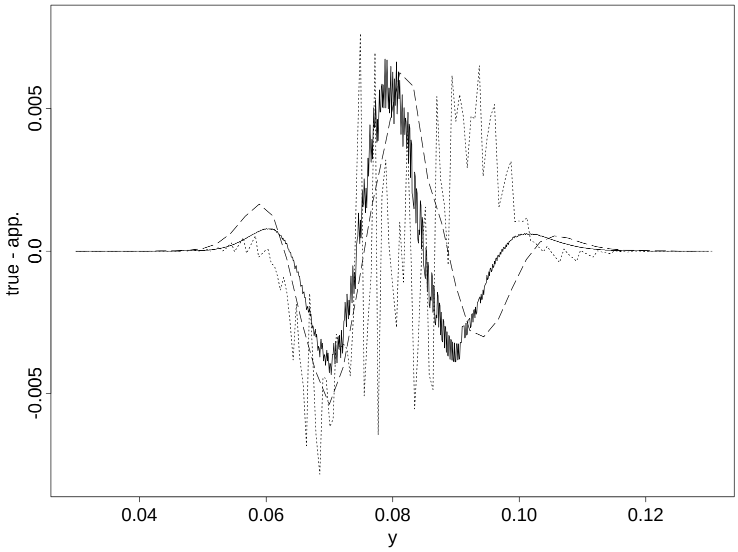

Figure 3: "True- app." density for the CIR model. —:Hermite ($t < 0.01$ s), - - -: Fin. Diff. ($t = 0.05$ s), ...:Binom ($t = 4.0$s)

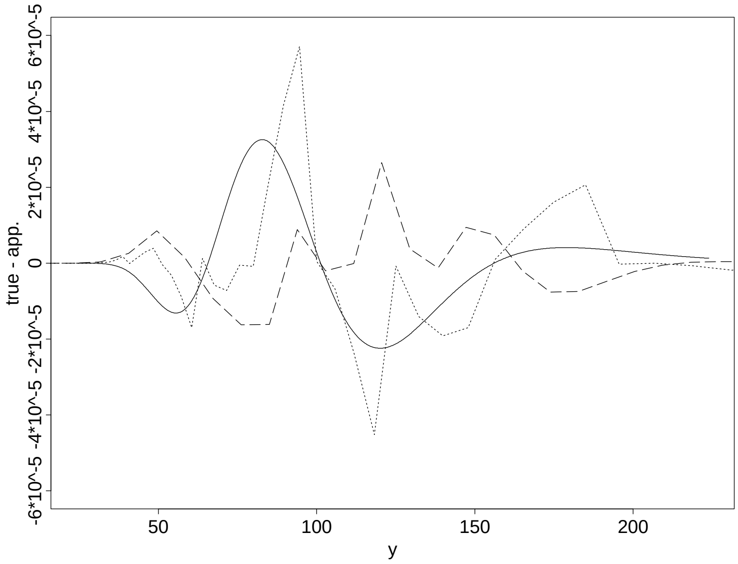

Figure 4: "True- app." density for the BS model. —:Hermite ($t < 0.01$ s), - - -: Fin. Diff. ($t = 0.05$ s), ...:Binom ($t = 0.05$s)

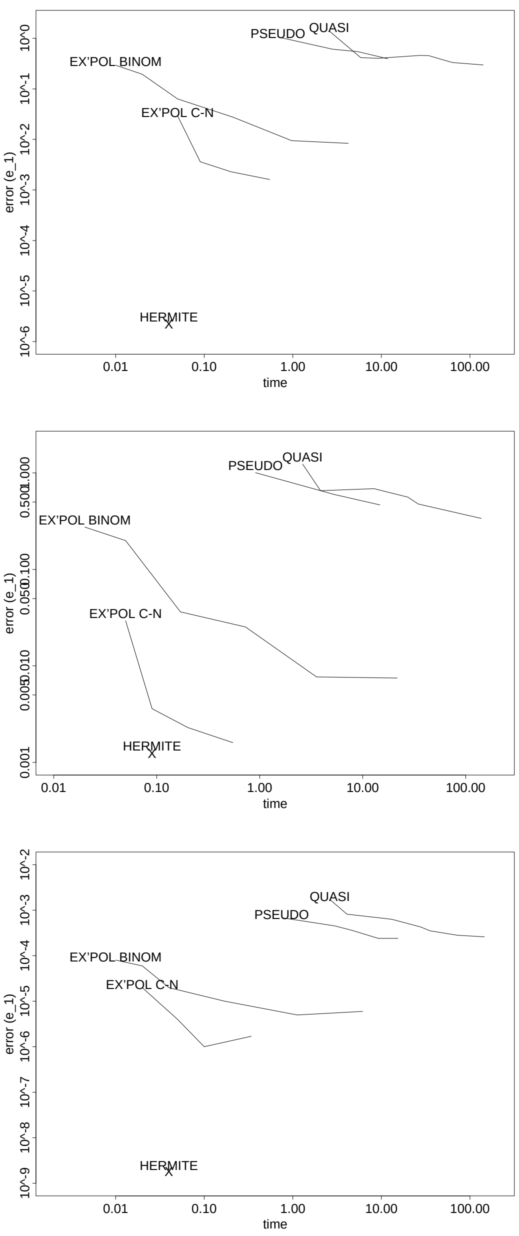

Figure 5: Speed vs. accuracy ($\sim e_1$) for different approximation techniques. Top: Vasicek, middle: CIR, bottom: BS.

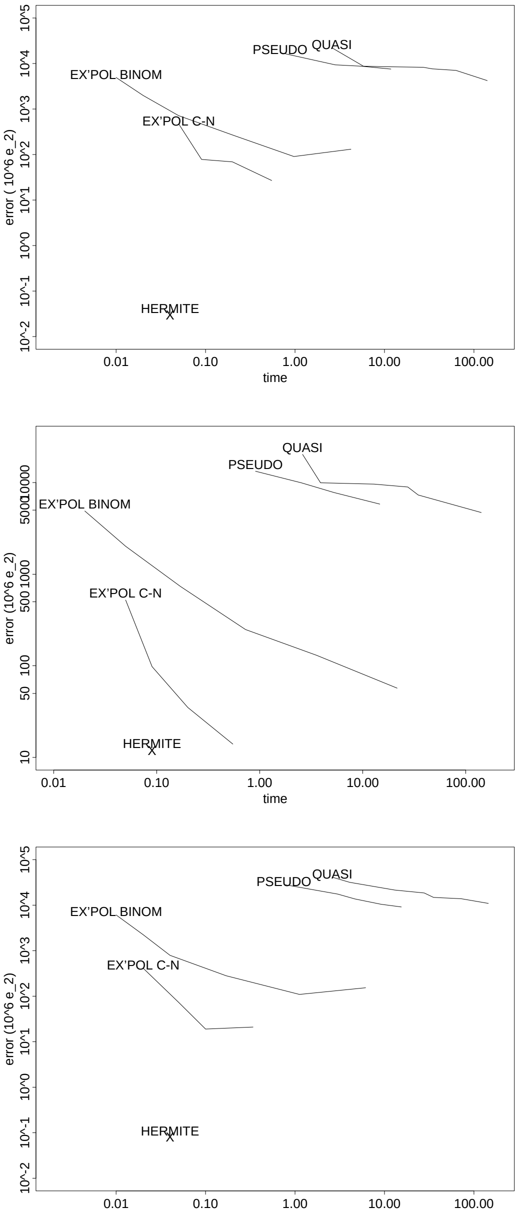

Figure 6: Speed vs. accuracy ($\sim e_2$) for different approximation techniques. Top: Vasicek, middle: CIR, bottom: BS.

Approximate Maximum Likelihood Estimation of Discretely Observed Diffusion Processes*

Rolf Poulsen
Department of Operations Research
University of Aarhus, Building 530
DK-8000 Aarhus C, Denmark
Phone: +45 8942 3549
email: rolf@imf.au.dk

Abstract

We introduce a new estimator using an approximation to the true, but generally unknown, loglikelihood function based on discrete observations of a diffusion process. The approximation is calculated using numerical solution techniques for partial differential equations and its quality is controlled – both theoretically and in numerical practice – by the statistician. The estimator is shown to be asymptotically equivalent to the maximum likelihood estimator. We also show how this equivalence can be achieved in a computationally optimal way. The method is easy to implement for one-dimensional diffusions and we demonstrate its computational feasibility for various diffusions. Finally, we estimate the CKLS short rate model on 1982-95 U.S. data and find the elasticity of variance parameter to be about 0.78.

*Helpful comments from Bent Jesper Christensen and Peter Honoré are gratefully acknowledged.

1 Introduction

Typically no closed-form solution is available for the transistion densities of a process governed by a stochastic differential equation (SDE). Therefore it is hard to base parametric inference from discrete observations on the likelihood function. Thus deprived of our ideal plan we can do two things *i*) find a good approximation to the likelihood, or *ii*) something else. A huge number of articles recently published in the statistical, the econometric and the financial literatures fall in the latter category. An incomplete list of keywords and key-references is:

- **Generalized Method of Moments**
  Hansen (1982) is a standard 'theory' reference, which has little to do with SDEs, Chan, Karolyi, Longstaff & Sanders (1992) is a standard 'application' reference, the main objections to which have little to do with GMM.

- **Martingale methods**
  Bibby & Sørensen (1995), Sørensen (1997)

- **Indirect Inference**
  Gouriéroux, Monfort & Renault (1993), Broze, Scaillet & Zakoian (1995), Broze, Scaillet & Zakoian (1998).

- **Efficient Method of Moments**
  A particularly cunning form of indirect inference, see Gallant & Long (1997), Gallant & Tauchen (1996*b*), and Gallant & Tauchen (1996*a*), Andersen & Lund (1997) is a successful application.

In the former category we find for example Pedersen (1995*b*), Pedersen (1995*a*) (with applications in Honoré (1998)), and Santa-Clara (1997). And now this paper. Several questions should be asked about a likelihood approximation: Is it reasonable? Is it accurate? How accurate? How can we tell? (Similar questions apply to the resulting estimators.) In this paper we propose an approximation to the likelihood and try to answer all those questions. The construction of the approximation is based on the combination of two well-known concepts:

- **The Fokker-Planck equation**
  This is a parabolic partial differential equation (PDE) for the transistion density (it is also called the *forward* or the *Kolmogorov* equation). It is known from the

probability literature; in textbooks (e.g. Karatzas & Shreve (1992), Øksendal (1995)) it is usually found close to the pages containing the very construction of diffusions. In the econometric literature this characterization (in a more general jump-diffusion setting) is also noted by Lo (1988).

- **The Crank-Nicolson method**
  This is a potentially second order accurate numerical solution technique of implicit finite difference type for PDEs, see e.g. Strikwerda (1989), Mitchell & Griffiths (1980). It is frequently used in the finance literature and industry see e.g. Wilmott, Dewynne & Howison (1993), Duffie (1996).

The main advantages of the method are: The approximation can be made arbitrarily accurate (and we can get a good idea *how* accurate a given approximation is), quite fast to calculate and easy to program. Further, the estimator can be made asymptotically equivalent the maximum likelihood (ML) estimator, there are not many 'loose ends' convergence-order-wise (which we think is a requirement that some of the methods relying heavily on simulation do not meet) and it is very flexible wrt. the statistical parametrization.

The outline of the paper is as follows. The rest of this section describes the basic set-up. Taking last things first, we show in Section 2 that the estimator we are later going to construct is consistent, asymptotically normal (CAN), and can be asymptotically equivalent to the ML estimator. Further we show how this desirable feature can be achieved at (asymptotically) minimal computational cost. Section 3 deals with the actual calculation of the approximation. It has two parts, construction and verification, of which the latter is often ignored, despite it being both important and potentially fruitful. In Section 4 we conduct numerical experiments on short rate models common in finance. Section 5 contains what has become a benchmark for estimation techniques for discretely observed diffusions: Estimation of the CKLS-short rate model. This necessitates consideration on optimization procedures. The very short summary: 'We find $\gamma = 0.78 \pm 0.133$ (95% confidence interval) for U.S. 1982-95 data'. In Section 6 we conclude and outline topics for future research.

1.1 The Basic Set-up

We consider a set-up with a one-dimensional diffusion process X with dynamics governed by the SDE

$$dX_t = \mu(X_t; \psi)dt + \sigma(X_t; \psi)d\mathbb{W}_t, \tag{1}$$

where $\mathbb{W}$ is a Brownian motion on some filtered probability space, $\psi \in \Psi \subseteq \mathbb{R}^d$, and $\mu, \sigma : \mathbb{R} \times \Psi \mapsto \mathbb{R}$ are functions such that (1) (with some initial condition) is well-defined and has a (weak) solution and (versions of) transition densities that are absolutely continuous wrt. the Lebesque measure and sufficiently smooth. By

$$\phi(t, x, y)$$

we denote the density of X_t given $X_0 = x$, i.e. ϕ is a density 'in y for fixed t and x' and we put $l(x_{i-1}, x_i; \psi) = \ln \phi(\triangle, x_{i-1}, x_i)$.

The parameter ψ is unknown but we seek statistical inference about it from (discretely observed) data points $x_0, x_1, \ldots, x_T$. The time between observations x_{i-1} and x_i , say $\triangle_i$, could vary with i, but we will make the following assumption.

Assumption 1 *Assume that:*

i) The true parameter is ψ_0.

ii) The observations are equidistant.

iii) X is stationary and ergodic.

Often a particular theoretical model violates part *iii*) of this assumption. The trick is then (with the Ito formula as a key tool) to find an appropriate transform of the model that is stationary and ergodic (typical example: look at log-returns rather that stock prices).

2 CAN of the AML Estimator

Typically, the transition densities are unknown, but our first assumption ensures that we can approximate them to our hearts' (if not our system administrator's) content.

Assumption 2 (Main) *Assume that the approximate loglikelihood can be written as*

$$
\begin{aligned}
l^A_{T,h}(\psi) &= \sum_{i=1}^{T} l^A_h(x_{i-1}, x_i; \psi) & (2)\\
&= \sum_{i=1}^{T} \Big(l(x_{i-1}, x_i; \psi) + h^2 a(x_{i-1}, x_i; \psi) + o(h^2) b(x_{i-1}, x_i; \psi) \Big) & (3)\\
&= l_T(\psi) + h^2 a_T(\psi) + o(h^2) b_T(\psi), & (4)
\end{aligned}
$$

where l_T is true true (but generally unknown) loglikelihood function and h is a parameter chosen by the statistician such that the time needed to compute the left hand side of (2) does not grow faster than T/h^2.

Assumption 2 gives a very specific structure on the approximation; (3) is the fundamental equation stating that each term in the loglikelihood function is well-approximated. (2) reflects that a diffusion is a Markov-process wrt. its own filtration and (4) is merely compact notation.

What makes the proofs in the following fairly easy is the fact that the approximation is 'strictly analytical' – no extra stochastic terms are introduced, as would be the case if quantities were to be determined by simulation. Another very crucial fact is that we shall be able to check the assumption numerically (which includes finding (numerical) estimates of a and b) when the Crank-Nicolson finite difference method is used to find the approximate loglikelihood. The Crank-Nicolson method is not the only approximation satisfying Assumption 2, but most of the approximations suggested in the literature violate Assumption 2 because either i) h is not controlled by the statistician, or ii) time-consumption is worse. Of course such approximations may have other virtues.

In this paper we focus on one-dimensional diffusions but (2)-(4) would look exactly the same in the case where data points are multi-dimensional. Likewise, the assumptions, results, and proofs in the following would be similar. But it would be harder to construct a finite difference second order approximation to the loglikelihood and the time-consumption would (at least) be proportional to $T/h^{\dim(x_i)+1}$. Also, for the multi-dimensional case it is a far more delicate matter whether a solution the SDE (looking like) (1) exists, and even more so whether it has a density. (In fact, the Malliavan calculus which has recently received considerable interest in the finance community was originally developed to shed light on this problem; see Pedersen (1995a).)

The order of time-consumption does not play any role in the proofs, but we have chosen to specify it for clarification.

Assumption 3 (Technical) *Assume that:*

i) $\psi \mapsto l_T(\psi)$ *is twice continuously differentiable and has a unique maximum point* $\psi_T^{ML} \in \Psi$.

ii) $\psi \mapsto l_{T,h}^A(\psi)$ *is twice continuously differentiable.*

iii) There exists finite constants K_a *and* K_b *(independent of* x *and* y*) such that* $\sup_{\psi\in\Psi} |a(x,y;\psi)| \leq K_a$ *and* $\sup_{\psi\in\Psi} |b(x,y;\psi)| \leq K_b$.

The 'least harmless' part of Assumption 3 is the 'boundedness in ψ'-part because it hard to verify, especially when the a and b functions are determined numerically. The assumption ensures that

$$\sup_{\psi\in\Psi} |l_{T,h}^A(\psi) - l_T(\psi)| \to 0 \text{ for } \ h \to 0. \tag{5}$$

and that the following estimators (the latter referred to as 'the AML estimator' for obvious reasons) are well-defined:

$$\begin{aligned} \psi_T^{ML} &= \arg\sup_{\psi\in\Psi} \frac{1}{T} l_T(\psi)\,, \\ \psi_{T,h}^{AML} &= \arg\sup_{\psi\in\Psi} \frac{1}{T} l_{T,h}^A(\psi)\,. \end{aligned}$$

It is only the AML estimator we are able to compute since the true loglikelihood will typically be unknown. As it is common in this part of the literature we shall (except for a brief discussion on Section 5.1) neglect all problems (convergence, stability, time-consumption) associated with the optimization involved in finding $\psi_{T,h}^{AML}$. We do this despite these problems being very 'real', in particular for high-dimensional parameters.

It follows from Corollary 2 in Pedersen (1995*b*) (this requires only Assumption 3 provided we substitute *iii*) by (5)) that if the ML estimator is CAN then there exists a sequence $h(T)$ (tending to 0 as T tends to infinity) such that the AML estimator has the same asymptotic properties. We shall use Assumption 2 to obtain more precise information on the h-sequence. In particular: How slowly can we allow it to tend to 0? But first we need two well-known results.

Theorem 1 (ML Has Usual Good Properties) *Under weak regularity conditions (see e.g. Barndorff-Nielsen & Sørensen (1994)), which we assume hold, the ML estimator is CAN with usual limits,*

$$\begin{aligned}\psi_T^{ML} &\overset{Pr}{\to} \psi_0 \\ \sqrt{T}(\psi_T^{ML} - \psi_0) &\overset{\sim}{\to} N(0, i^{-1}(\psi_0)),\end{aligned}$$

where $i(\psi)$ is the Fisher information matrix,

$$i = \lim_{T\to\infty} -\mathbf{E}\left(\frac{1}{T}\frac{\partial^2 l_T(\psi)}{\partial\psi\partial\psi^\top}\right).$$

Also, any other CAN estimator has an asymptotic covariance matrix no less (in the partial order of positive semi-definite matrices) than that of the ML estimator.

Consistency of the ML estimator, ergodicity and the assumption that the errors (by which we mean the a and b functions in (3)) are uniformly bounded in ψ ensures that the AML estimator has a limit as the number of observations tends to infinity. This is a result from the realm of 'theory of misspecified models'.

Theorem 2 (AML Has a Limit) *Under weak regularity conditions (see e.g. White (1994)), which we assume hold, we have that*

$$\exists\, h_0 > 0 \;\; \forall\, h \in]-h_0; h_0[\;\; \exists\, \psi_h \in \mathbb{R} : \psi_{T,h}^{AML} \overset{Pr}{\to} \psi_h \; for \; T \to \infty.$$

The following lemma shows that the AML estimator inherits the second order h-accuracy property of the approximate loglikelihood.

Lemma 1 (Inheritance) *There exists $\widetilde{\psi}_T$, $\widetilde{\psi}$, $\widetilde{\widetilde{\psi}}_T$ and $\widetilde{\widetilde{\psi}}$ such that:*

i) $\psi_{T,h}^{AML} = \psi_T^{ML} + h^2\widetilde{\psi}_T + o(h^2)\widetilde{\widetilde{\psi}}_T$.

ii) $\widetilde{\psi}_T \overset{Pr}{\to} \widetilde{\psi}$ *for* $T \to \infty$ *and* $\widetilde{\widetilde{\psi}}_T \overset{Pr}{\to} \widetilde{\widetilde{\psi}}$ *for* $T \to \infty$.

iii) $2\widetilde{\psi} = \partial^2\psi_h/\partial h^2|_{h=0}$.

Proof. Since $\psi_{T,h}^{AML}$ is an interior maximum point it satisfies usual first order conditions, which by Assumption 2 read:

$$\partial_\psi l_{T,h}^A(\psi_{T,h}^{AML}) = l_T(\psi_{T,h}^{AML}) + h^2\partial_\psi a_T(\psi_{T,h}^{AML}) + o(h^2)\partial_\psi b_T(\psi_{T,h}^{AML}) = 0.$$

Because $\partial_\psi l^A_{T,h}$ is a $\mathbf{C}^1$-function of h^2, the Implicit Function Theorem (see Rudin (1982), the technical 'non-degenerate Jacobian' condition is part of the implicit assumptions in Theorem 1) allows us to conclude that $\psi^{AML}_{T,h}$ is a $\mathbf{C}^1$-function of h^2. Since '$h = 0$' corresponds to the ML estimator, $i)$ is established. To prove $ii)$ choose a sufficiently small $\widetilde{h}$ and apply $i)$ for values $\widetilde{h}$ and $\widetilde{h}/2$:

$$\begin{aligned} \widetilde{h}^2\widetilde{\psi}_T + o(\widetilde{h}^2)\widetilde{\widetilde{\psi}}_T &= \psi^{AML}_{T,h} - \psi^{ML}_T, \\ \widetilde{h}^2\widetilde{\psi}_T/4 + o(\widetilde{h}^2)\widetilde{\widetilde{\psi}}_T &= \psi^{AML}_{T,h} - \psi^{ML}_T. \end{aligned}$$

Subtracting four times the latter equation from the former makes the $\widetilde{\psi}_T$ terms disappear and we may assume that the term in front of $\widetilde{\widetilde{\psi}}_T$ is non-zero,

$$o(\widetilde{h}^2)\widetilde{\widetilde{\psi}}_T = 3(\psi^{ML}_T - \psi^{AML}_{T,h}).$$

Letting $T \to \infty$, the RHS of this equation converges in probability (by Assumption 1 and Theorem 1), hence $\widetilde{\widetilde{\psi}}_T$ and (from $i)$) $\widetilde{\psi}_T$ converge in probability and $ii)$ is proved. Note that we must have $\psi_h = \psi_{-h}$. Rearranging and letting $T \to \infty$, we get from $i)$

$$\widetilde{\psi} = \frac{\psi_h - \psi_0 + o(h^2)\widetilde{\widetilde{\psi}}_T}{h^2} = \frac{1}{2}\left(\frac{\psi_h - 2\psi_0 + \psi_{-h}}{h^2}\right) + o(1)\widetilde{\widetilde{\psi}}_T.$$

Statement $iii)$ follows as we let $h \to 0$ ■

We can now finally prove the main result about CAN of the AML estimator.

Theorem 3 ('How Slow Can You Go?') *Suppose $h(T) = T^{-\delta}$. Then:*

i) $\psi^{AML}_{T,h}$ is consistent if (and only if) $\delta > 0$.

ii) If $\delta = 1/4$, then $\sqrt{T}(\psi^{AML}_{T,h} - \psi_0)$ converges in distribution to a normal variable or more precisely,

$$\sqrt{T}\left(\psi^{AML}_{T,h} - \psi_0\right) \overset{\sim}{\to} N\left(\frac{1}{2}\frac{\partial^2\psi_h}{\partial h^2}|_{h=0}, i^{-1}(\psi_0)\right).$$

iii) If $\delta > 1/4$ then

$$\sqrt{T}(\psi^{AML}_{T,h} - \psi_0) \overset{\sim}{\to} N(0, i^{-1}(\psi_0)),$$

i.e. the AML estimator is asymptotically equivalent to the ML estimator.

Proof. The consistency part (i)) is a consequence of continuity, Lemma 1 and CAN of the ML estimator and statements ii) and iii) follow if we write

$$\sqrt{T}(\psi_{T,h}^{AML} - \psi_0) \;\; = \;\; \sqrt{T}(\psi_T^{ML} - \psi_0) + T^{1/2-2\delta}\widetilde{\psi}_T + o(T^{1/2-2\delta})\widetilde{\widetilde{\psi}}_T.$$

The former term converges in distribution by Theorem 1 and the latter vanishes if we keep $1/2 - 2\delta \leq 0$. If $1/2 - 2\delta < 0$ the middle term also vanishes, and if we have '=' it converges in probability to $\widetilde{\psi}$ ■

In the light of Lemma 1 – or even Assumption 2 – this result should not be very surprising; for example in Cox & Hinkley (1974) it is noted that 'estimates that differ from the ML estimate by o($1/\sqrt{T}$) [which holds if $\delta > 1/4$] are also efficient'. But note that the error term is controlled by us, the statisticians. In that respect the estimation technique is related to the quasi-indirect inference suggested in Broze et al. (1998); their problem is choosing the step size in a discretization used for simulation of observations from a SDE. Theorem 3 is of the same type as the results on page 170 in that paper. But note that the asymptotic covariance matrix in our case is the inverse Fisher information.

Numerically it may be hard to obtain a reliable estimate of this correction term ('estimation' perceived as both a statistical and a numerical problem) but that is not a problem great of practical importance.

3 Construction of the AML estimator

In this section we show how to find an approximate loglikelihood function satisfying Assumption 2 using numerical methods. It consists of two independent parts; first an approximation is constructed from local considerations, then a simple but very effective technique is used to check the global behaviour of the approximation. The latter step can also be used to achieve considerably higher accuracy at little extra computational cost. In this section dependence on ψ will be notationally suppressed, while – for reasons that will soon become clear – time dependency is made explicit.

Relying on diffusion theory (e.g. along the lines of Karatzas & Shreve (1992)), we have that under certain regularity conditions ϕ solves the parabolic PDE (called the *forward* or the *Fokker-Planck* or the *Kolmogorov equation*)

$$\frac{\partial}{\partial t}\phi(t,x,y) = -\frac{\partial}{\partial y}\left(\mu(y)\phi(t,x,y)\right) + \frac{1}{2}\frac{\partial^2}{\partial y^2}\left(\sigma^2(y)\phi(t,x,y)\right), \tag{6}$$

with initial condition $\phi(0,x,y) = \delta(y-x)$, where $\delta(\cdot)$ is the Dirac-δ function. Strictly speaking this is not a PDE in the usual sense (it is often called a *functional* PDE) because the Dirac-δ function is not a function in the usual sense but a generalized function. It can be defined as a continuous, linear, real-valued mapping on the infinite dimensional space $\mathbf{C}^\infty(\mathbb{R})$ such that $\delta : f \mapsto f(0)$ for all $f \in \mathbf{C}^\infty(\mathbb{R})$. Therefore we cannot use standard theory to (define and) ensure existence and uniqueness of a solution. This has to be verified with other methods. Two possible ways are 'inspired guesses' and 'advanced probabilistic methods (Malliavan calculus) on the corresponding SDE'. Rigorous treatment of this is far beyond this (most) paper(s), we refer to Pedersen (1995*a*).

The PDE characterization of the likelihood is also noted in Lo (1988) where it is said that: 'when the existence of a density representation for a specific process has been assured by other means, (6) may often be solved by standard methods (Fourier transforms etc.) to yield the likelihood', but the issue is not further elaborated on. This is what we will do and to us 'standard methods' means 'numerical methods'. So, we will apply a numerical technique known from 'usual PDEs' and then see if it produces reasonable results.

3.1 The Crank-Nicolson Finite Difference Method

Let subscripts denote differentiation, suppress 'x' and other arguments where it causes no confusion, and rewrite (6) as

$$\phi_t(t,y) = a(y)\phi + b(y)\phi_y + c(y)\phi_{yy} \tag{7}$$

where

$$\begin{aligned} a(y) &= (\sigma_y)^2 + \sigma\sigma_{yy} - \mu_y = \frac{1}{2}(\sigma^2)_{yy} - \mu_y, \\ b(y) &= 2\sigma\sigma_y - \mu = (\sigma^2)_y - \mu, \\ c(y) &= \frac{1}{2}\sigma^2. \end{aligned}$$

Now consider a time/space grid, see Figure 1,

INSERT FIGURE 1 ABOUT HERE

with step sizes k and h. For any grid point away from the boundaries (say y_0 and y_M) we consider the approximation $v_m^n \approx \phi(nk, y_0 + mh)$ defined by the equation

$$\frac{v_m^{n+1} - v_m^n}{k} = a\delta_0(h) + b\delta_1(h) + a\delta_2(h), \tag{8}$$

where the δ's are difference operators. It is the particular choice of δ's that determine the accuracy and stability (and popular name) of the finite difference method. For the Crank-Nicolson method the operators act in the following way:

$$\begin{aligned}
\delta_0(h) &= \frac{1}{2}v_m^{n+1} + \frac{1}{2}v_m^n, &(9)\\
\delta_1(h) &= \frac{1}{2}\frac{v_{m+1}^{n+1} - v_{m-1}^{n+1}}{2h} + \frac{1}{2}\frac{v_{m+1}^n - v_{m-1}^n}{2h}, &(10)\\
\delta_2(h) &= \frac{1}{2}\frac{v_{m+1}^{n+1} - 2v_m^{n+1} + v_{m-1}^{n+1}}{h^2} + \frac{1}{2}\frac{v_{m+1}^n - 2v_m^n + v_{m-1}^n}{h^2}. &(11)
\end{aligned}$$

Taylor expansions show that this method is locally second order accurate in both h and k. Also, the Crank-Nicolson method is unconditionally stable for parabolic PDEs (see Strikwerda (1989)).

Defining $\lambda = kh^{-2}$ and inserting in (8) gives

$$\begin{aligned}
&\left(\frac{bh\lambda}{4} - \frac{c\lambda}{2}\right)v_{m-1}^{n+1} + \left(1 - \frac{ak}{2} + c\lambda\right)v_m^{n+1} + \left(-\frac{bh\lambda}{4} - \frac{c\lambda}{2}\right)v_{m+1}^{n+1} = \\
&\left(-\frac{bh\lambda}{4} + \frac{c\lambda}{2}\right)v_{m-1}^n + \left(1 + \frac{ak}{2} - c\lambda\right)v_m^n + \left(\frac{bh\lambda}{4} + \frac{c\lambda}{2}\right)v_{m+1}^n. \qquad (12)
\end{aligned}$$

Considering v as known on the boundaries and using the initial condition, (12) defines a sequence of tridiagonal linear systems of equations. These systems can be solved recursively and each system requires only a number of operations that is proportional to the number of state space steps (when the tridiagonal structure is exploited, e.g. by the routine `tridag` from Numerical Recipes (Press, Teukolsky, Vetterling & Flannery (1992)).)

If we consider only diffusions with inaccessible boundaries (no reflection or absorption) then it is reasonable to put $v = 0$ on the grid boundaries. The grid boundary levels are chosen 'sufficiently far apart' (the actual numbers are parameter dependent, but choosing them is not a problem).

The initial condition is treated like this:

$$v_m^1 = \phi^N(y_0 + mh; x + \mu(x)k, \sigma^2(x)k), \tag{13}$$

where $\phi^N(y; m, v)$ is the density function for a normally distributed variable with mean m and variance v evaluated at y. This is a first-order approximation to the density (in a certain sense). By (another) definition of the Dirac-δ function this converges to the true initial condition as $k \to 0$. The Crank-Nicolson method tends to behave badly for non-smooth initial data (as pointed out in Strikwerda (1989)[p.121]), therefore the above choice is crucial. In spirit this is similar to the first step in the approximation technique suggested in Pedersen (1995*a*).

Other choices would be possible, but for example

$$v_m^0 = \frac{1}{h}\mathbf{1}_{[y_0+mh;y_0+(m+1)h[}(x) \tag{14}$$

works very poorly in conjunction with the Crank-Nicolson method.

It may be argued that the Crank-Nicolson finite difference method is not best way of solving the Fokker-Planck equation numerically. Finite element methods or special polynomial approximations have been suggested, see Harrison (1988). In particular, these techniques perform better when the problem is 'convection dominated', i.e. when the drift is large compared to the volatility. For the interest rates models we shall be working with this turns out not to be a major problem, but it does indicate that it is very important to have methods that can provide numerical verification of the validity of the method.

3.2 Experimental Verification of Assumption 2

It is important to be able to provide error estimates for the numerical method when the analytical solution to the PDE is not known. This section describes a method (developed in Østerby (1998)) that does just that. We consider first the case where we have a function u of one variable and a numerical approximation v. We say that v is a globally first order accurate approximation if

$$v(x) = u(x) - hc(x) - h^2 d(x) - h^3 f(x) - \dots,$$

where c, d and f are smooth functions. If $c = 0$ then we say that the approximation is (at least) globally second order accurate. Notice that the usual Taylor expansion analysis of finite difference schemes gives only local statements; when step sizes tend to 0 local errors become smaller, but there are a lot more of them (at any point the calculated solution depends on the calculated solution in earlier points in the grid because of the recursive structure), so potentially anything might happen. Problems of this nature can easily (and judging from part of the finance literature: undetectedly) occur when PDEs with non-smooth initial data or boundary conditions not of Dirichlet-type are solved numerically using the Crank-Nicolson scheme.

Now calculate the numerical solution with step sizes h, $2h$ and $4h$ thus obtaining

$$\begin{aligned} v_1 &= u - hc - h^2 d - h^3 f, \\ v_2 &= u - 2hc - 4h^2 d - 8h^3 f, \\ v_3 &= u - 4hc - 16h^2 d - 64h^3 f. \end{aligned}$$

By subtracting and dividing we can calculate the following coefficient in all points corresponding to the '$4h$'-calculation

$$\frac{v_3 - v_2}{v_2 - v_1} = 2\frac{c + 6hd + 28h^2 f + \dots}{c + 3hd + 7h^2 f + \dots}. \tag{15}$$

If $c \neq 0$ and h is small then the ratio in (15) should be close to 2. This means that is we observe a lot of numbers close to 2.0 then we conclude that the approximation is globally first order accurate. Likewise, 4.0's all over the place indicate that the method is globally second order accurate. It is possible for c to have several roots so we should always calculate the ratio in (15) for a large number of points. The quantification of terms like 'a lot of' and 'all over the place' is in the mind of the beholder; but most times you are not in doubt.

If there is a nice pattern of 2.0's or 4.0's we can also obtain an estimate of the error on the 'h'-calculation,

$$\epsilon_1 = u - v_1 = hc + h^2 d + h^3 f + \dots .$$

If we have a first order method then $|c|$ is larger than $h|d|$ and we see that the error is fairly well represented be $v_1 - v_2 = hc + 3hd^2 + \dots$. For a second order method $|c|$ is 0, while $|d|$ is larger than $h|f|$ and $\epsilon_1 = h^2 d + h^3 f + \dots$ is well approximated by $(v_1 - v_2)/3 = h^2 d + 7h^3 f/3 + \dots$. By adding the approximated error term (a.k.a. Richardson extrapolating) we hope to obtain a more accurate estimate. Note also that the extrapolation schemes depend on the order of the approximation, therefore they should only be applied after justifying the order.

When u depends on several variables (time and space for our PDE-solution) we perform the order determination separately in each direction, 2's and 4's are still what makes us happy. Assuming that we have a second order method we may want to make an approximation that removes possible h^2k^2-terms. With v_4, v_5 and v_6

denoting '$2k$'-, '$4k$'- and '$(2h, 2k)$'-calculations respectively, a bit of algebra shows that

$$v_1 + \frac{4}{9}(v_1 - v_2) + \frac{4}{9}(v_1 - v_4) - \frac{1}{9}(v_1 - v_6) = u + O(h^3 + k^3). \tag{16}$$

When we have a method that is globally second order accurate it is clear that the computationally feasible thing is to keep the time step size proportional to the space step size with a proportionality factor chosen such that the leading error terms are roughly the same size. This means that effectively we only have to chose one step size and we get an approximation fulfilling Assumption 2. Except: Assumption 2 is concerned with the logarithm of the density, we are solving for the actual density. This, however, is not a problem. Simple manipulations and appropriate Taylor expansions on the (absolutely convergent) series

$$\ln x = \sum_{i=1}^{\infty} \frac{1}{2i-1} \left(\frac{1-x}{1+x} \right)^i$$

shows that we also have a second order approximation to the loglikelihood. Furthermore, in the regions where we need function values, the absolute error is often much smaller after taking logarithms.

In the numerical results and empirical applications that are to follow we shall be using the extrapolation given in (16). Note that extrapolated function values are only available for the points on the coarsest grid, but fortunately this turns out not to be a problem for the empirical applications. Despite using (16) we maintain Assumption 2 as stating second order accuracy in order not to be over-optimistic. (And time-consumption-wise we should be more than adequately compensated for the need to increase $y_M - y_0$ as h and k tend to 0.)

4 Numerical Experiments with Short Rate Models

We quote two error measures in the cases where the transition densities are known.

i) Maximal absolute error,

$$e_1 = \sup_m \{ |\phi^A(y_m) - \phi(y_m)| \}.$$

$\triangle$	x	θ	κ	$\sigma\theta^\gamma$
1/12	0.08	0.08	0.24	0.025

Table 1: Common parameters for the numerical experiments.

ii) Average relative error (quoted in part per million in tables),

$$\begin{aligned} e_2 &= \sum_m |\phi^A(y_m) - \phi(y_m)|h \\ &\approx \int_{\mathbb{R}} |\phi^A(y) - \phi(y)|dy = \int_{\{y|\phi(y)>0\}} \frac{|\phi^A(y) - \phi(y)|}{\phi(y)} \phi(y)dy. \end{aligned}$$

We do not report errors regarding $\ln\phi$. Time is the CPU-time (in seconds) on a HP-9000 Unix machine.

We consider the parametric specification suggested in Chan et al. (1992),

$$dX_t = \kappa(\theta - X_t)dt + \sigma X_t^\gamma d\mathbb{W}_t. \tag{17}$$

We use the financially realistic parameters given in Table 1.

4.1 The Vasicek Model

This is the '$\gamma = 0$'-case, i.e. we consider an Ornstein-Uhlenbeck process (the financial terminology is after (a section in) Vasicek (1977)),

$$dX_t = \kappa(\theta - X_t)dt + \sigma d\mathbb{W}_t.$$

The transition density is

$$\phi^V(\triangle, x, y) = \phi^N(y; m(\triangle, x), v(\triangle, x)),$$

where

$$\begin{aligned} m(t,x) &= \mathbf{E}(X_t|X_0 = x) = e^{-\kappa t}x + \theta(1 - e^{-\kappa t}) \\ v(t,x) &= \mathbf{V}(X_t|X_0 = x) = \frac{\sigma^2(1 - e^{-2\kappa t})}{2\kappa}. \end{aligned}$$

We use $y_0 = 0.03$ and $y_M = 0.13$ as upper and lower boundaries in the grid.

Figure 2 gives an indication of

INSERT FIGURE 2 ABOUT HERE

what the functions we will be working with look like. One could object to the relatively low $\triangle$-value, saying that 'for small time steps diffusions are close to Gaussian', so we are not investigating the numerical method for the whole range of densities that can occur for diffusions. However, we have several counter-objections: *i*) 'close to Gaussian' is a qualitative and parameter-dependent statement, *ii*) small time steps is an appropriate way of 'stress-testing' the method, *iii*) for the model and ψ in question it corresponds to monthly observations which is a common, perhaps even low, sampling frequency for financial data, and *iv*) numerical experiments indicate that error results are qualitatively similar for larger (and smaller) $\triangle$-values.

Detailed results for the Vasicek model are shown in Tables 2 and 3.

INSERT TABLE 2 ABOUT HERE

We see that it is easy to obtain (average) accuracy to 4-5 significant digits (which will give even more accurate values for the loglikelihood).

If we use a simple Euler approximation to the density, which corresponds to setting $N = 1$, the errors are 0.55 and 9640, respectively.

The Richardson extrapolations are (as all those that follow) based on the method being second order accurate in both time and space. If we insist, as we should, on performing the order determination then the extrapolation is virtually 'free of computational charge'. Even if we do not, then it only roughly doubles the number of calculations and we see that it is worth the effort, especially for large step sizes. Table 3

INSERT TABLE 3 ABOUT HERE

contains the h- and k-ratios for the experimental order determination. Overall, it is a pretty picture. In the time direction the numbers may fall a little short of the desired 4's, but much worse results have been encountered. For both directions the problems seem to arise at the '$y = 0.075$'- and '$y = 0.085$'-lines. These are the y-values closest to the inflexion points for the true solutions (at $m(\triangle, x) \pm v(\triangle, x)^{1/2}$ $= 0.0800 \pm 0.0071$) – more an observation than an explanation.

Figure 3

INSERT FIGURE 3 AND TABLE 4 ABOUT HERE

illustrates what happens if (14) is used as initial condition. For $x = 0.08$ things look nice, but looks can be deceiving. When is x changed slightly the errors can change dramatically; the average relative error can become more than 100 times larger. A calculation of h- and k-ratios (see Table 4, the numbers are equally wayward irrespective of what x is) also tells us that something is wrong and indicate that it is the space dimension that needs more careful treatment (with h-ratios like those in Table 4 one should check the estimated errors to see if it is due to a 'division by 0'-problem – it is not in this case).

4.2 The Cox-Ingersoll-Ross Model

This is the '$\gamma = 1/2$'-case of (17),

$$dX_t = \kappa(\theta - X_t)dt + \sigma\sqrt{X_t}d\mathbb{W}_t.$$

The transition density (cf. Cox, Ingersoll & Ross (1985)) is

$$\phi^{CIR}(\triangle, x, y) = c\exp(-u - v)\left(\frac{v}{u}\right)^{q/2} I_q(2\sqrt{uv}) \quad \text{for } y \in \mathbb{R}_+,$$

where

$$c = \frac{2\kappa}{\sigma^2(1 - \exp(-\kappa\triangle))}, \qquad q = \frac{2\kappa\theta}{\sigma^2} - 1,$$
$$u = cx\exp(-\kappa\triangle), \qquad v = cy,$$

and I_q is the modified Bessel function of order q. Again, we use $y_0 = 0.03$ and $y_M = 0.13$.

Results are shown in Tables 5 and 6.

INSERT TABLES 5 AND 6 ABOUT HERE

We draw the same conclusions as for the Vasicek model: 4-5 significant digits are easy to get, the method appears to be globally second order accurate in both time space, and extrapolation is fruitful. The Euler approximation has $e_1 = 2.0$ and $10000000e_2 = 35200$. Another fully analytical approximation, namely using a normal density with correct mean and variance (expressions for these can be found in Cox et al. (1985)) has $e_1 = 1.8$ and $10000000e_2 = 34000$.

4.3 The CKLS Model

For general γ-value the transition density to (17) is unknown but for $\gamma \in [1/2; 1]$ it is known to exist. (If no solution exists the numerical procedure is likely to indicate this.)

Results for different values of γ are shown in Table 7.

`INSERT TABLE 7 ABOUT HERE`

We use $y_0 = 0.03$ and $y_M = 0.17$. The method still works nicely, if anything it appears to work better.

5 Estimation of the CKLS Model

5.1 A Brief History of Optimization

In theory we could just apply the Crank-Nicolson method to the forward equations, calculate the approximate loglikelihood, feed this to some 'black-box' optimization routine and wait for the estimate to be delivered. In practice, however, several matters need to be considered:

i) A lot of time is spend determining the density in areas where it is flat and close to 0.

ii) To calculate the approximate loglikelihood value we have to run through T different finite difference grids since each 'x_{i-1}' gives a different initial condition.

iii) We only need the value of the density function in a single point, namely for '$y = x_i$'.

iv) Where do we start the optimization procedure?

v) Which optimization procedure should we use?

Using an unevenly spaced grid or transforming the SDE are the obvious ways of improving on *i*). But this is not without cost; unevenly spaced grid may make extrapolation difficult and a transform of the SDE could mean that we loose the 'natural' (Dirichlet-type) boundary conditions. Ergo, we stick with even spacing on the original SDE.

Basically, there is nothing we can do about *ii*). Solving only for relatively few initial conditions and then applying some interpolation technique can bring down computation time (a lot), but it induces a very large risk of 'throwing the baby out with the bath water'.

Since we are going to run through a grid for each pair of consecutive observations we may as well make the best of *iii*). We do this by constructing the $'4h'$-grid such that x_i lies exactly on a grid point. In this way we do not have to do interpolations and we get the full effect of the Richardson extrapolation. We use $y_{0/M} = x_i \mp 6\sigma(x_i; \psi_{T,h}^{AML,0})$ ($\psi_{T,h}^{AML,0}$ will be defined in the next paragraph.)

As starting point, $\psi_{T,h}^{AML,0}$, we use the estimates we get from '$N = 1$', that is by using the Euler discretization ('Euler' will be used to indicate estimators of this type).

Finally, we determine $\psi_{T,h}^{AML}$ by a quasi-Newton optimization (more precisely the Brent, Hall, Hall, Hausman technique). This means that we 'update until convergence' according to the rule

$$\psi_{T,h}^{AML,j} = \psi_{T,h}^{AML,j-1} - \lambda_{j-1} M^{-1} g,$$

where g is the gradient of the approximate likelihood function (at $\psi_{T,h}^{AML,j-1}$), M is a matrix that 'looks like the matrix of second derivatives' (ditto), $\lambda_0 = 1$ and the following λ_j's are chosen such that the objective function never decreases. The quantities g and M have to determined numerically. For g we use symmetric perturbations (in ψ) while M is determined from an outer product of first derivatives by invoking the information equality,

$$\mathbf{E}\left(\frac{\partial^2 l(x_{i-1}, x_i; \psi)}{\partial\psi\partial\psi^\top}\right) = -\mathbf{E}\left(\frac{\partial l(x_{i-1}, x_i; \psi)}{\partial\psi}\frac{\partial l(x_{i-1}, x_i; \psi)^\top}{\partial\psi}\right)$$

(which is of course only really relevant for large samples and does not strictly apply when $h > 0$). Our numerical analyses indicate that M is accurately determined even for fairly coarse grids provided we are careful about the parameter perturbation involved in the numerical ψ-differentiation.

5.2 Data and Estimation Results

We use (weekly observed) data on the U.S. three-month T-bill rates covering the period 1982-95. We use only 'post-October 1982'-data to avoid problems stemming from the well-known fact that the Federal Reserve Board used a different targeting policy between October 1979 and October 1982. We use weekly sampling frequency as a 'compromise'; on one hand we do not like to throw away information (=observations), on the other hand, say, daily observations would make us much more prone to market micro-structure effects. Results are fairly robust to using monthly data, the major difference is higher standard errors of the σ and γ estimates. The data are, after appropriate conversion, obtained from the Federal Reserve Board statistical release H.15 (see `http://www.bog.frb.fed.us/releases/`). This is (a subset of) the data set used in Andersen & Lund (1997). Summary statistics are given in Table 8.

INSERT TABLES 8 AND 9 ABOUT HERE

Table 9 reports estimation results. We give three sets of estimates:

i) AML estimates.

ii) Euler estimates, i.e. the starting point of the AML search. Since this corresponds to a fixed $h > 0$ these estimates are inconsistent.

iii) GMM estimates obtained using the technique described in Chan et al. (1992). This GMM *implementation* is based on an Euler-type approximation to the conditional moments and thus also yields inconsistent estimates (but estimates that are easy to calculate).

The methods agree on θ, but then one would not expect the unconditional mean to be that hard to estimate. There is bit more variation regarding κ, the speed of mean reversion, but none of the methods produce an estimate significantly different from 0.[1] From a financial point of view this is uncomfortable, and things might be even worse than they seem since it can be shown that the κ-estimates are positively biased for low values of κ. However, it is not all bad news. The γ-estimates, which have been a focal point in the literature, all lie between 1/2 and 1, which is where we (for many reasons) want them to lie. More often than not papers have reported estimates in excess of 1 (in this case a weak solution to the SDE does exist, but it is not a particularly nice one). The GMM estimates have higher standard deviations than the other estimates, which might not be so surprising because no attempt is made to mimic the likelihood function. AML and Euler estimates are quite similar which supports the conjecture that '$\triangle = 1$ week *is* small'. There is no definitive evidence of lower standard errors for the AML estimates. There may be a number of

[1] It should be noted that for $\kappa \leq 0$ (while $1/2 \leq \gamma \leq 1$) no invariant measure (limiting distribution) exists so we cannot test this using standard asymptotic results.

reasons for this: i) T is finite, anything may happen in small samples (the statistical equivalent of 'crying wolf'), ii) estimated standard errors are obtained with numerical differentiation, and iii) the Euler estimates can 'buy' lower standard deviations with their inconsistency.

6 Conclusion

In this paper we introduced an estimator based on a second order approximation to the true but unknown loglikelihood function for a discretely observed diffusion process. The quality of the approximation could be controlled – both in theory and numerical practice – by the statistician. We proved that for appropriate choice of the discretization step size h the estimator was asymptotically equivalent to the ML estimator; we even found the (computationally) optimal rate at which h should tend to 0. We demonstrated how (and how not) to calculate the approximation using the Crank-Nicolson method for PDEs, how to verify global properties, and how improve accuracy at low computational cost. The strength of the numerical technique was illustrated by numerical experiments on members of the CKLS-family. Finally, the CKLS-model was estimated using weekly observations on U.S. short rate data covering the period 1982 to 1995. We found the mean-reversion parameter insignificant, while γ was estimated around 0.78 with a standard deviation of 0.068. Standard deviations were markedly lower than those obtained when the GMM estimates suggested in Chan et al. (1992) was calculated, but on weekly data there appeared to be little difference between the AML estimates and the simple estimates based on an Euler approximation to transition density.

Topics of ongoing research include:

i) Simulation studies to investigate the small sample behaviour of the AML estimator.

ii) Detailed comparison of the AML estimator to other estimators proposed in the literature.

iii) Application of the AML estimator to short rate models with non-linear drift specification (as proposed for example in Ait-Sahalia (1996) and Conley, Hansen, Luttmer & Scheinkmann (1997)). This will remove a shortcoming of the methods frequently used to estimate these non-linear models: In some respects observations are treated as being *iid*, which they are clearly not (cf. Table 8). See Pritsker (1998) for further discussion.

Preliminary results can be found in Christensen, Poulsen & Sørensen (1999).

Also, a natural next task is to implement the numerical PDE solution technique for multi-dimensional diffusions. This is certainly not impossible, but numerically it is fairly difficult.

References

Ait-Sahalia, Y. (1996), 'Testing Continuous-Time Models of the Spot Interest Rate', *Review of Financial Studies* **9**(2), 385–426.

Andersen, T. G. & Lund, J. (1997), 'Estimating Continuous-Time Stochastic Volatility Models of the Short-Term Interest Rate', *Journal of Econometrics* **77**, 343–377.

Barndorff-Nielsen, O. E. & Sørensen, M. (1994), 'A review of some aspects of asymtotic likelihood theory for stochastic processes', *International Statistical Review* **62**(1), 133–165.

Bibby, B. & Sørensen, M. (1995), 'Martingale estimating functions for discretely observed diffusion processes', *Bernoulli* **1**, 17–39.

Broze, L., Scaillet, O. & Zakoian, J.-M. (1995), 'Testing for Continuous-Time Models of the Short-Term Interest Rate', *Journal of Empirical Finance* **2**, 199–223.

Broze, L., Scaillet, O. & Zakoian, J.-M. (1998), 'Quasi-Indirect Inference for Diffusion Processes', *Econometric Theory* **14**, 161–186.

Chan, K. C., Karolyi, G. A., Longstaff, F. A. & Sanders, A. B. (1992), 'An Empirical Comparison of Alternative Models of the Short-Term Interest rate', *Journal of Finance* **47**(3), 1209–1227.

Christensen, B. J., Poulsen, R. & Sørensen, M. (1999), Optimal Martingale and Likelihood Methods for Models of the Short Rate of Interest, With Monte Carlo Evidence for the CKLS Specification and Applications to Non-Linear Drift Models. Working paper. Department of Operations Research. University of Aarhus.

Conley, T. G., Hansen, L. P., Luttmer, E. G. & Scheinkmann, J. A. (1997), 'Short-Term Interest Rates as Subordinated Diffusions', *Review of Financial Studies* **10**(3), 525–577.

Cox, D. R. & Hinkley, D. V. (1974), *Theoretical Staistics*, Chapman-Hall.

Cox, J. C., Ingersoll, J. E. & Ross, S. A. (1985), 'A Theory of the Term Structure of Interest Rates', *Econometrica* **53**, 385–407.

Duffie, D. (1996), *Dynamic Asset Pricing Theory*, 2 edn, Princeton University Press.

Gallant, A. R. & Long, J. R. (1997), 'Estimating Stochastic Differential Equations Efficiently by Minimum Chi-Square', *Biometrika* **84**, 125–142.

Gallant, A. R. & Tauchen, G. E. (1996*a*), Specification Analysis of Continuous Time Models in Finance, *in* P. E. Rossi, ed., 'Modelling Stock Market Volatility: Bridging the Gap to Continuous Time', Academic Press, pp. 357–383.

Gallant, A. R. & Tauchen, G. E. (1996*b*), 'Which Moments to Match', *Econometric Theory* **12**, 657–681.

Gouriéroux, C., Monfort, A. & Renault, E. (1993), 'Indirect Inference', *Journal of Applied Econometrics* **8**, 85–118.

Hansen, L. P. (1982), 'Large Sample Properties of Generalized Method of Moments Estimators', *Econometrica* **50**, 1029–1054.

Harrison, G. W. (1988), 'Numerical Solution of the Fokker Planck Equation Using Moving Finite Elements', *Numrerical Methods for Partial Differential Equations* **4**, 219–232.

Honoré, P. (1998), Five Essays on Financial Econometrics in Continuous-Time Models, PhD thesis, Department of Finance. Aarhus School of Business.

Karatzas, I. & Shreve, S. (1992), *Brownian Motion and Stochastic Calculus*, 2 edn, Springer-Verlag.

Lo, A. W. (1988), 'Maximum Likelihood Estimation of Generalized Ito Processes with Discretely Sampled Data', *Econometric Theory* **4**(2), 231–247.

Mitchell, A. R. & Griffiths, D. F. (1980), *The Finite Difference Method in Partial Differential Equations*, Wiley.

Øksendal, B. (1995), *Stochastic Differential Equations*, 4. edn, Springer-Verlag.

Østerby, O. (1998), The Error of the Crank-Nicolson Method for Linear Parabolic Equations with a Derivative Boundary Condition. Working paper. Department of Computer Science. University of Aarhus.

Pedersen, A. R. (1995*a*), 'A New Approach to Maximum Likelihood Estimation for Stochastic Differential Equations Based on Discrete Observations', *Scandinavian Journal of Statistics* **22**(1), 55–71.

Pedersen, A. R. (1995*b*), 'Consistency and asymptotic normality of an approximate maximum likelihood estimator for discretely observed diffusion processes', *Bernoulli* **1**(3), 257–279.

Press, W., Teukolsky, S., Vetterling, W. T. & Flannery, B. (1992), *Numerical recipes in C*, 2. edn, Cambridge University Press.

Pritsker, M. (1998), 'Nonparametric Density Estimation and Tests of Continuous Time Interest Rate Models', *Review of Financial Studies* **11**(3), 449–487.

Rudin, W. (1982), *Principles of Mathematical Analysis*, 2 edn, McGraw-Hill.

Santa-Clara, P. (1997), Simulated Likelihood Estimation of Diffusion with an Application to the Short Term Interest Rate. Working paper no. 12-97 , UCLA.

Sørensen, M. (1997), Estimating Functions for Discretely Observed Diffusions : A Reviev, *in* I. Basawa, V. P. Godambe & T. R. L., eds, 'Selected Procedings of the Symposium on Estimating Functions', Vol. 32 of *IMS Leture notes - Monograph series*, pp. 305–325.

Strikwerda, J. (1989), *Finite Difference Schemes and Partial Differential Equations*, Wadsworth & Brooks/Cole.

Vasicek, O. (1977), 'An Equilibrium Characterization of the Term Structure', *Journal of Financial Economics* **5**, 177–188.

White, H. (1994), *Estimation, Inference, and Specification Analysis*, Cambridge University Press.

Wilmott, P., Dewynne, J. & Howison, S. (1993), *Option Pricing: Mathematical models and Computation*, Oxford Financial Press.

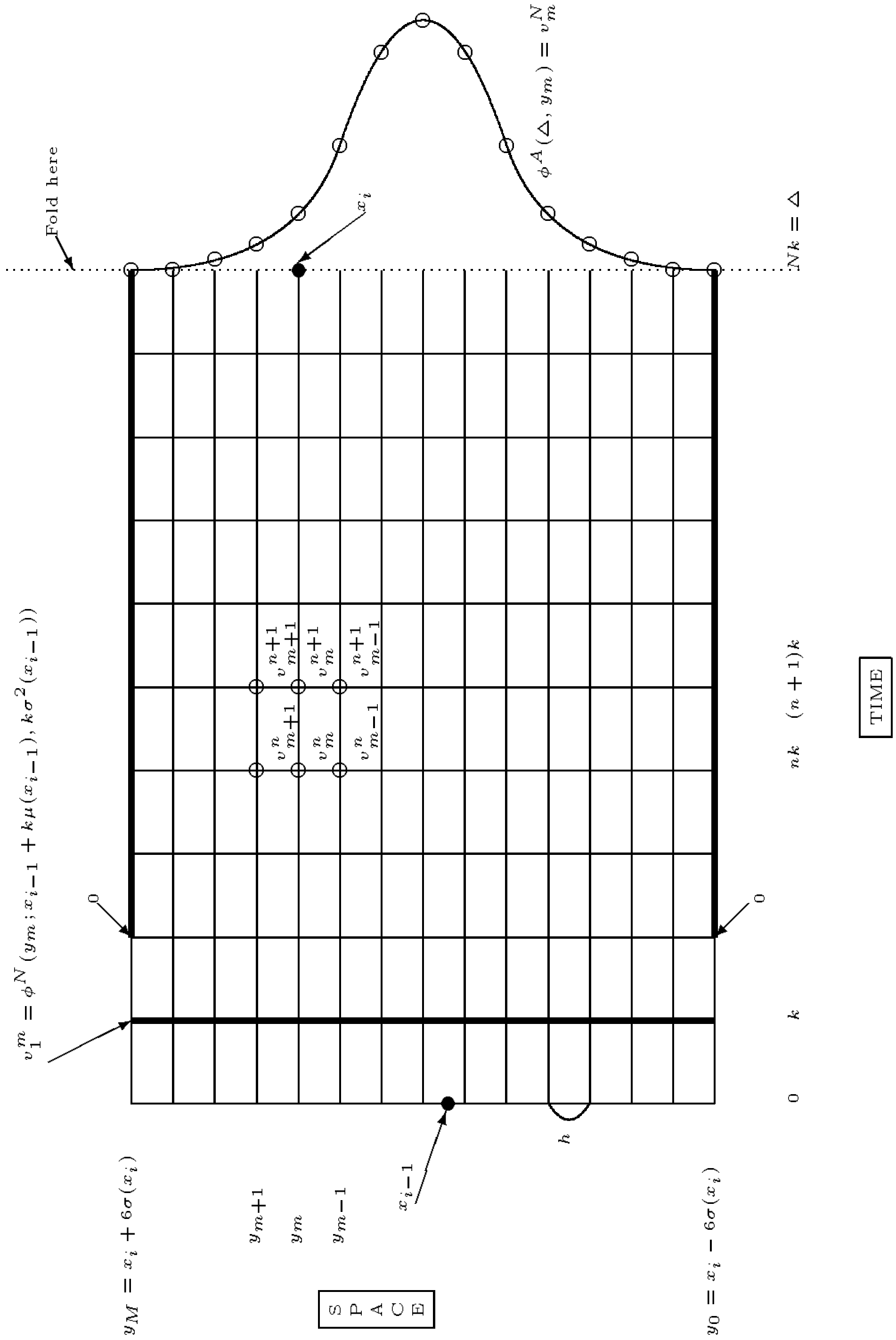

Figure 1: A finite difference grid for numerical solution of the forward PDE (6) for the transition density. How to specify the initial condition and making sure that x_i lies on a grid point are numerically important "tricks of the trade".

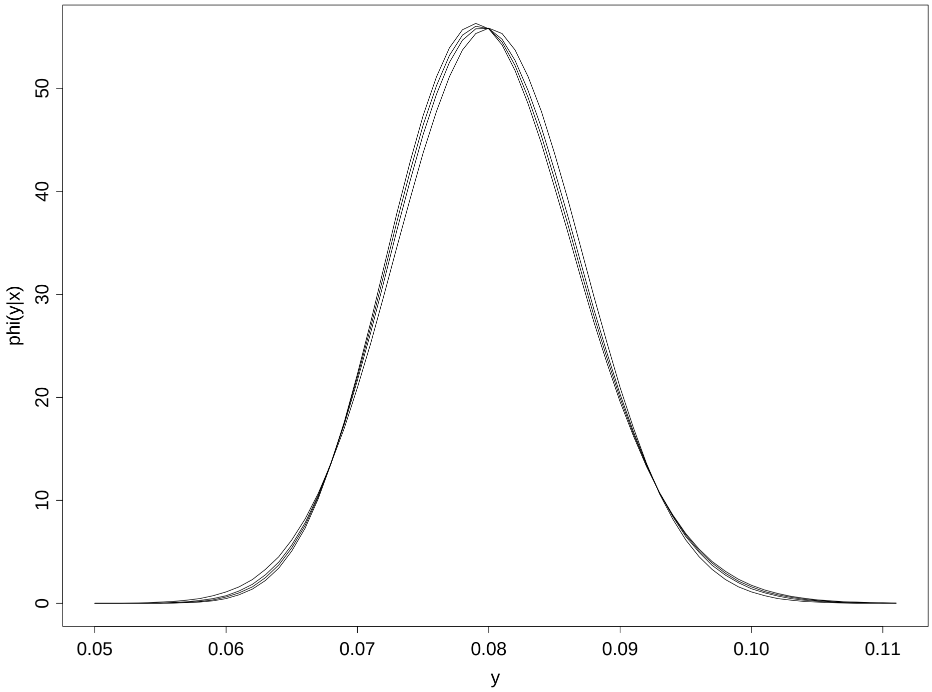

Figure 2: Densities for the CKLS-diffusion for different values of γ. The higher γ is, the further to the left the maximum of the density. The plots are for $\gamma = 0.0$ (Vasicek), $\gamma = 0.5$ (CIR), $\gamma = 0.75$, and $\gamma = 1.0$ and $\sigma\theta^\gamma$ is kept fixed. Other parameters: $x = \theta = 0.08$, $\kappa = 0.24$, and $\triangle = 1/12$.

M	N	Errors before ex'pol e_1	e_2	Errors after ex'pol e_1	e_2	Time incl. ex'pol
50	2	1.3926	18636	N/A	N/A	0.03
50	4	0.1545	3094	N/A	N/A	0.04
50	8	0.3762	6139	N/A	N/A	0.04
50	16	0.5096	7871	N/A	N/A	0.04
100	4	0.1619	5482	0.1797	3973	0.05
100	8	0.0352	913	0.0221	471	0.05
100	16	0.0980	1724	0.0322	615	0.05
100	32	0.1277	2111	0.0587	785	0.06
200	8	0.0970	1595	0.0234	414	0.09
200	16	0.0096	248	0.0037	78	0.09
200	32	0.0233	417	0.0077	140	0.11
200	64	0.0252	444	0.0087	148	0.14
400	16	0.0237	426	0.0081	96	0.17
400	32	0.0041	93	0.0034	69	0.20
400	64	0.0084	121	0.0027	61	0.22
400	128	0.0103	142	0.0031	64	0.40
800	32	0.0075	125	0.0081	113	0.38
800	64	0.0055	92	0.0072	115	0.55
800	128	0.0049	63	0.0025	62	1.11
800	256	0.0084	108	0.0097	173	2.12

Table 2: Finding the Vasicek-density by the Crank-Nicolson method applied to (7).

y / t	1	2	3	4	5	6	7	8
6.0	5.6	4.0	4.0	4.0	4.2	3.0	4.0	3.8
6.5	4.1	3.9	4.5	3.9	3.8	3.9	4.0	3.8
7.0	3.9	3.8	3.9	4.0	3.8	3.8	4.1	3.5
7.5	4.1	4.1	3.5	3.8	5.4	4.7	4.0	5.2
8.0	4.2	4.2	4.2	4.1	4.4	4.4	4.1	4.7
8.5	4.1	4.0	3.4	4.1	5.9	5.1	4.2	5.7
9.0	3.9	3.8	3.9	4.0	3.8	3.8	3.9	3.3
9.5	4.1	3.9	4.5	3.9	3.8	3.8	3.9	3.7
10.0	5.6	4.0	4.0	4.0	4.3	2.9	3.8	3.7

h-ratio

y / t	1	2	3	4	5	6	7	8
6.0	4.1	3.7	3.5	3.8	3.8	3.8	4.0	2.6
6.5	4.1	3.7	3.8	3.9	5.2	4.2	3.8	4.5
7.0	2.9	3.8	3.8	3.8	4.0	4.1	3.8	5.0
7.5	3.2	3.7	3.5	3.6	-0.6	1.7	3.7	1.9
8.0	3.7	3.7	3.7	3.8	3.4	3.1	3.6	2.6
8.5	3.8	3.7	3.9	4.0	-0.3	1.4	2.7	1.6
9.0	3.2	3.7	3.8	3.9	4.1	4.4	4.5	6.0
9.5	2.9	3.7	3.8	3.8	5.8	4.4	4.5	4.9
10.0	4.1	3.7	3.4	3.8	3.9	3.8	3.6	2.8

k-ratio

Table 3: Experimental order determination for the Crank-Nicolson method applied to (6) in the Vasicek model. Starting set-up has $h = 2.5 \cdot 10^{-4}$ and $k = 1/768$, y measured in %, t in 48'ths. If the method is globally $O(h^2+k^2)$, then the two columns to the right should contain a large amount of numbers fairly close to 4.

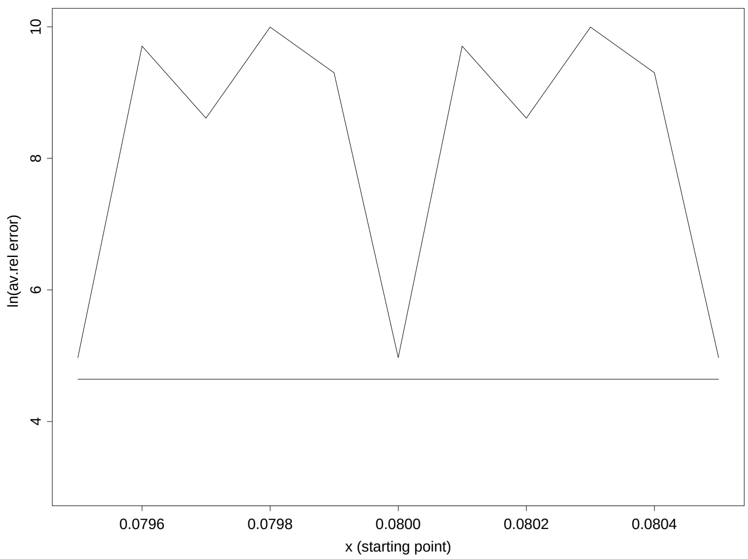

Figure 3: Sensitivity to the initial condition in the Vasicek model. The x-axis represents the initial value (x) and the y-axis is the logarithm of the average relative error (e_2). Jagged curve: Using (14). Straight line: Using (13). Step sizes: $h = 2.5 \cdot 10^{-4}$ and $k = 1/768$.

t / y	6.0	6.5	7.0	7.5	8.0	8.5	9.0	9.6	10.0
0.0833	277.9	-1309.5	-1158.4	17500.3	-4911.2	20605.5	-1702.5	-2316.2	588.2

h-ratio

t / y	6.0	6.5	7.0	7.5	8.0	8.5	9.0	9.6	10.0
0.0833	4.0	4.0	4.0	5.3	9.1	4.9	3.9	4.1	4.0

k-ratio

Table 4: Experimental order determination for the Crank-Nicolson method applied to (6) with (14) as initial condition in the Vasicek model. Starting set-up has $h = 2.5 \cdot 10^{-4}$ and $k = 1/768$ and $x = 0.080$.

M	N	Errors before ex'pol e_1	e_2	Errors after ex'pol e_1	e_2	Time incl. ex'pol
50	2	1.3637	19500	N/A	N/A	0.03
50	4	0.1884	3761	N/A	N/A	0.04
50	8	0.3748	6580	N/A	N/A	0.04
50	16	0.5087	8340	N/A	N/A	0.04
100	4	0.3748	5797	0.2037	3678	0.05
100	8	0.0526	1010	0.0294	524	0.05
100	16	0.1084	1747	0.0336	617	0.05
100	32	0.1267	2129	0.0611	800	0.06
200	8	0.0983	1636	0.0271	470	0.09
200	16	0.0150	270	0.0036	98	0.09
200	32	0.0257	448	0.0099	152	0.11
200	64	0.0328	538	0.0123	179	0.14
400	16	0.0262	435	0.0067	99	0.17
400	32	0.0051	80	0.0023	35	0.20
400	64	0.0085	120	0.0024	39	0.22
400	128	0.0093	136	0.0036	43	0.40
800	32	0.0074	113	0.0030	32	0.38
800	64	0.0022	24	0.0016	14	0.55
800	128	0.0032	31	0.0013	19	1.11
800	256	0.0032	35	0.0016	20	2.12

Table 5: Finding the CIR-density by the Crank-Nicolson method applied to (7)

y / t	1	2	3	4	5	6	7	8
6.0	7.7	4.0	3.9	4.0	4.1	5.0	3.8	3.6
6.5	4.3	3.9	4.2	3.7	3.9	3.8	3.8	3.9
7.0	5.7	3.6	3.9	3.9	3.9	3.7	3.6	3.7
7.5	-6.7	3.1	3.7	6.0	4.9	3.8	4.9	4.9
8.0	25.5	4.9	4.2	4.2	4.4	4.4	4.5	4.5
8.5	-8.1	2.9	3.2	5.6	5.6	5.2	5.2	5.4
9.0	7.7	3.6	3.9	3.9	3.7	3.6	3.3	3.3
9.5	4.2	4.0	3.3	3.9	3.8	3.8	3.7	3.7
10.0	5.2	3.9	4.1	4.1	2.3	3.7	3.6	3.7

h-ratio

y / t	1	2	3	4	5	6	7	8
6.0	4.2	3.9	4.2	3.6	3.7	3.8	3.6	3.8
6.5	5.7	3.5	3.8	3.8	4.6	3.8	3.8	3.8
7.0	3.6	4.8	3.9	3.8	3.8	3.6	3.7	3.7
7.5	3.6	3.4	3.8	10.8	3.6	3.8	4.0	4.0
8.0	3.7	4.0	3.7	3.8	3.7	3.8	3.9	3.9
8.5	3.8	3.7	4.0	3.9	5.5	5.4	5.1	5.1
9.0	3.8	16.7	3.8	3.8	4.0	3.9	3.8	3.8
9.5	4.0	3.8	3.9	4.2	4.1	3.9	3.9	3.9
10.0	3.9	4.0	3.9	3.9	3.9	4.0	3.8	3.9

k-ratio

Table 6: Experimental order determination for the Crank-Nicolson method applied to (6) in the CIR model. Starting set-up has $h = 2.5 \cdot 10^{-4}$ and $k = 1/768$, y measured in %, t in 48'ths.

h- and k-ratios

y	$\gamma = 0.6$		$\gamma = 0.7$		$\gamma = 0.8$		$\gamma = 0.9$		$\gamma = 1.0$	
0.060	4.0	3.3	4.0	2.5	4.0	10.4	3.9	4.9	3.9	4.5
0.065	3.8	3.8	3.8	3.7	3.8	3.7	3.7	3.7	3.7	3.6
0.070	4.0	3.7	4.0	3.8	4.1	3.8	4.0	3.7	4.0	3.8
0.075	3.6	5.7	3.5	4.0	3.8	4.5	3.6	5.0	3.5	4.0
0.080	4.0	3.8	4.0	3.8	4.1	3.8	4.0	4.0	4.0	3.8
0.085	3.7	3.8	3.9	4.1	4.0	4.1	3.7	3.7	4.2	4.1
0.090	4.0	3.8	4.0	4.0	4.0	3.9	4.0	3.8	3.9	3.9
0.095	4.0	4.2	3.9	3.7	3.9	4.2	4.0	4.3	3.8	4.4
0.100	4.0	3.9	4.0	3.9	4.1	3.9	4.1	3.9	4.7	3.9

Table 7: Experimental order determination (at time 1/12) for the Crank-Nicolson method in the CKLS model. Starting set-up has $h = 2.5 \cdot 10^{-4}$ and $k = 1/768$.

Weekly U.S. three-month T-bill rates, 1982-95. (FED)

Variables	T	Mean	Standard Deviation	ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	ρ_6
r_t	655	0.006379	0.02044	0.99	0.99	0.99	0.99	0.99	0.98
$r_t - r_{t-1}$	654	-0.00004	0.00166	-0.06	-0.05	0.07	0.06	0.10	-0.02

Table 8: Summary statistics. The ρ's are autocorrelation coefficients.

Estimation Technique	θ	κ	σ	γ
CKLS-GMM	0.0524	0.1724	0.0698	0.6729
	(0.0165)	(0.1465)	(0.0235)	(0.1217)
Euler	0.0510	0.1506	0.0953	0.7873
	(0.0184)	(0.1477)	(0.0186)	(0.0678)
AML	0.0511	0.1322	0.0931	0.7794
	(0.0173)	(0.1489)	(0.0182)	(0.0682)

Table 9: Estimates of the CKLS-model $dr_t = \kappa(\theta - r_t)dt + \sigma r_t^{\gamma} d\mathbb{W}_t$ on weekly FED data covering 1982 to 1995. The AML estimates are based on 8 time steps and space step sizes of 1 basis point (0.0001, that is). Parentheses indicate estimated standard errors. The run-time for the optimization is around 4 minutes.

Optimal Martingale and Likelihood Methods for Models of the Short Rate of Interest, With Monte Carlo Evidence for the CKLS Specification and Applications to Non-Linear Drift Models

Bent Jesper Christensen
School of Economics and Management
University of Aarhus, Building 350
DK-8000 Aarhus C, Denmark

Rolf Poulsen
Department of Operations Research
University of Aarhus, Building 530
DK-8000 Aarhus C, Denmark

June 30, 1999

Abstract

Improved estimation methods for discretely observed diffusion models for the short rate of interest are introduced. We consider both optimal martingale estimating equations and approximate maximum likelihood methods based on second order convergent numerical solution of the forward partial differential equation for the transition density. The methods are compared to well-known methods, namely the Generalized Method of Moments, Indirect Inference and Gaussian Quasi Maximum Likelihood, both theoretically, in an application to U.S. data, and in Monte Carlo experiments. The benchmark model used for illustration and comparison is the Chan, Karolyi, Longstaff & Sanders (1992, henceforth CKLS) short rate model. The new martingale and likelihood methods reduce bias, true standard errors, and bias in estimated standard errors, relative to the established methods, in particular for the key parameter of interest, the elasticity of variance. In weekly data from 1982 to 1995, the new methods estimate this parameter to about 0.78. Finally, we use the approximate maximum likelihood method to estimate non-linear drift short rate models. We find the terms commonly suggested as drift augmentations insignificant.

Contents

1 Introduction

Recently, a large number of articles in both the statistical, financial and econometric literatures have developed and applied techniques related to estimation of discretely observed diffusion processes. The bulk of the empirical applications concern interest rate data. There are good reasons for this. Firstly, diffusions (solutions of stochastic differential equations) are a key tool for modern financial model building. Secondly, as opposed to equity markets, where the Black & Scholes (1973) Model is typically employed as a benchmark (although numerous studies document its empirical inadequacies), a benchmark model is yet to emerge in fixed income markets. Thus our application focuses on models for the short rate of interest, which is only one aspect of term structure modelling, but a very important and relevant one that receives much attention. Though comparative studies had been performed earlier (e.g. Marsh & Rosenfeld (1983)), the popular breakthrough in this literature was Chan et al. (1992), who introduced a simple framework for comparison of many models from the literature. Numerous articles have followed up on this seminal work and suggested improvements in three conceptually separate directions:

1) Improvements of the **estimation technique**.

2) Reinvestigation of the **empirical results**.

3) Improvements of the **model**.

In our application we address all of these issues, albeit in the realm of one-factor models.

We consider five different estimation techniques. We first present the more familiar methods and then proceed to introduce martingale methods and maximum likelihood in the context of short rate models. Section 2 presents the different approaches: In Section 2.1 we look at the Generalized Method of Moments (GMM) used in Chan et al. (1992) and we include an improvement by calculating the moments more precisely. In Section 2.2 an Indirect Inference-type estimator is described and its relation to the so-called Efficient Method of Moments (EMM) is discussed. In Section 2.3, we investigate Gaussian Quasi Maximum Likelihood. Section 2.4 introduces martingale estimation equations and develops optimal martingale estimators. In Section 2.5 we propose a new approach to estimation of discretely observed diffusion processes, based on second order convergent numerical solution techniques for partial differential equations. In Section 3 we compare the estimators when applied to *i*) a U.S. data set with monthly sampling frequency used in Chan et al. (1992)), and *ii*) U.S. data with weekly sampling frequency covering the period from 1982 to 1995 (a subsample of the data set used in Andersen & Lund (1997).) Section 4 is a comprehensive simulation study. The proposed approximate maximum likelihood estimator is quite flexible w.r.t. the statistical parametrization of drift and volatility functions. We therefore conclude the paper by estimating a non-linear drift model (formally it is an intersection of the models proposed in Ait-Sahalia (1996) and Conley, Hansen, Luttmer & Scheinkmann (1997).)

2 Alternative Estimation Methods

In general an estimation method should work for d-dimensional diffusions

$$dX = \mu(X;\psi)dt + \sigma(X;\psi)d\mathbb{W}_t, \tag{1}$$

where $\mathbb{W}$ is a Brownian motion on a filtered probability space equipped with the usual filtration $\{\mathcal{F}_t\}$ and ψ is the parameter we seek statistical inference on.

Our recurrent example is the CKLS-equation (such acronym'ed after Chan et al. (1992)) describing the short term interest rate,

$$dr_t = (\alpha + \beta r_t)dt + \sigma r_t^{\gamma} d\mathbb{W}_t, \tag{2}$$

This family parametrized by $\psi^\top = (\alpha, \beta, \sigma, \gamma)$ nests many popular - and some rather esoteric - one-factor models, see Table 1.

We do not observe r continuously, rather we have $T+1$ observations

$$\{r_{t_0}, r_{t_1}, \ldots, r_{t_T}\}.$$

"The rules of the game" in economics/finance is that we cannot control the time between observations, at least we cannot let it tend to 0.

We shall let

$$y \mapsto p(\triangle, x, y; \psi)$$

denote the Lebesque density of $r_\triangle | r_0 = x$ and refer to p as the transition density. For application of standard distributional results it is also necessary that the invariant measure,

$$p^\infty(y;\psi) = \lim_{\triangle\to\infty} p(\triangle, x, y; \psi),$$

	Model	α	β	σ^2	γ
Merton	$dr_t = \alpha dt + \sigma d\mathbb{W}_t$		0		0
Vasicek	$dr_t = (\alpha + \beta r_t)dt + \sigma d\mathbb{W}_t$				0
Cox-Ingersoll-Ross	$dr_t = (\alpha + \beta r_t)dt + \sigma {r_t}^{1/2} d\mathbb{W}_t$				1/2
Dothan	$dr_t = \sigma r_t d\mathbb{W}_t$	0	0		1
Geometric Brownian motion	$dr_t = \beta r_t dt + \sigma r_t d\mathbb{W}_t$	0			1
Brennan-Schwartz	$dr_t = (\alpha + \beta r_t)dt + \sigma r_t d\mathbb{W}_t$				1
Constantinedes-Ingersoll	$dr_t = \sigma {r_t}^{3/2} d\mathbb{W}_t$				3/2
Constant Elasticity of Variance	$dr_t = \beta r_t dt + \sigma {r_t}^{\gamma} d\mathbb{W}_t$	0			

Table 1: Alternative short rate models. The names on the left are common, but some debatable.

exists. For one-dimensional diffusions it is often easy to determine if p^∞ exists (by speed and scale measure calculations, see Karlin & Taylor (1981, Section 15.6)) and characterize it explicitly. This measure can then be used to base statistical inference on, this is the idea in for instance Kessler (1996) and Conley et al. (1997). There are, however, several problems with this, *i*) if observations are highly correlated (which depends on the time between them and the functional form of drift and volatility) there may be loss of efficiency and, *ii*) often (as it is the case for the CKLS-model) ψ is not identified in p^∞.

If we knew p then we could use the Markov property of diffusions w.r.t. their own filtrations (see for instance Øksendal (1995, Theorem 7.2)) to write the likelihood function as the product of the $p(\triangle, r_{t_{i-1}}, r_{t_i}; \psi)$'s. (For multi-dimensional diffusions it is important that all the coordinates are observed, otherwise we would have to "condition & integrate out" (or do something else.)) The ML estimator would have the usual good properties (cf. Cambell, Lo & MacKinlay (1997, Appendix A) or for more technical aspects Barndorff-Nielsen & Sørensen (1994))

Result 1 *Under weak conditions, the ML estimator is consistent and asymptotically normal. Supposing observations are equidistant we have the usual limits:*

$$\begin{aligned} \psi_T^{ML} &\overset{Pr}{\rightarrow} \psi \quad \text{for } T \to \infty, \\ \sqrt{T}(\psi_T^{ML} - \psi) &\overset{\sim}{\rightarrow} N(0, i^{-1}(\psi)) \quad \text{for } T \to \infty. \end{aligned}$$

where $i(\psi)$ is the Fisher information matrix,

$$i = \lim_{T\to\infty} -\mathbf{E}\left(\frac{1}{T}\frac{\partial^2 l_T(\psi)}{\partial\psi\partial\psi^\top}\right).$$

Further the ML estimator is (statistically) efficient, i.e., for any other consistent and asymptotically normal estimator, say with asymptotic covariance matrix V, we have that

$$V - i \quad \textit{is positive semi-definite.}$$

But, alas, generally we do not have an analytical expression for p. So we have to something else.

2.1 Generalized Method of Moments

The standard reference for the Generalized Method of Moments is Hansen (1982), but the subject is treated in most econometric textbooks. The discrete time econometric specification of the model (2) that immediately springs to mind is

$$r_{t_i+\triangle_i} - r_{t_i} = (\alpha + \beta r_{t_i})\triangle_i + \epsilon_{t_i+\triangle_i}, \tag{3}$$

with $\triangle_i = t_{i+1} - t_i$ and the following conditional moment restrictions on the error terms:

$$\begin{aligned}
\mathbf{E}(\epsilon_{t_i+\triangle_i}|\mathcal{F}_{t_i}) &= 0 & (4)\\
\mathbf{E}(\epsilon^2_{t_i+\triangle_i}|\mathcal{F}_{t_i}) &= \triangle_i\sigma^2 r^{2\gamma}_{t_i}. & (5)\\
\mathbf{E}(\epsilon_{t_i+\triangle_i}\epsilon_{t_j+\triangle_j}|\mathcal{F}_{t_i}) &= 0 \quad \text{for } i < j & (6)
\end{aligned}$$

This is not an exact discrete-time representation of the model, rather it is a first order approximation[1] (w.r.t. time) a.k.a. an Euler-approximation. Later we will discuss how to make better approximations. For notational simplicity assume that $\triangle_i \equiv 1$, write the data as $\{r_0, r_1, \ldots, r_T\}$ and let $\rho^\top = (\alpha, \beta, \sigma, \gamma)$. We use this notation to distinguish in some way between the parameters in Equations (2) and (3) which are the same by name but not by nature. It will be the general notational philosophy in this paper to use ρ to denote a parameter in a more or less misspecified model

[1] There are different ways of considering the quality of such approximations. The Euler-approximation is of first order in the *weak* sense, viewed as a *strong* approximation its order is 1/2, see Kloeden, Platen & Schurz (1991). For statistical purposes the weak sense is often the relevant one.

whereas ψ's will indicate parameters that (can be made by us to) converge to the true parameter of the diffusion model.

The difference between (2) and (3) also means that the models are well-behaved/defined on different parameter sets and is probably one of the reasons there seems to be some confusion about stationarity properties. The following resolve some of these problems; the most striking thing about it models with γ-values larger than 1 can have stationary densities.

Result 2 *i) A model defined by*

$$r_{t_i+\triangle_i} - r_{t_i} = (\alpha + \beta r_{t_i})\triangle_i + \epsilon_{t_i+\triangle_i},$$

and

$$\begin{aligned} \epsilon_{t_i+\triangle_i}|\mathcal{F}_{t_i} &\sim N(0, \triangle_i\sigma^2|r_{t_i}|^{2\gamma}) \\ \mathbf{E}(\epsilon_{t_i+\triangle_i}\epsilon_{t_j+\triangle_j}|\mathcal{F}_{t_i}) &= 0 \quad \textit{for } i < j \end{aligned}$$

is stationary and ergodic if and only if

$$\alpha > 0, \quad \beta \in]-1;0[, \quad \sigma > 0, \quad \gamma \in [0;1].$$

ii) Weak solutions to (2) with an invariant measure exist if $\sigma > 0$, α satisfies one the following 3 conditions

$$\begin{aligned} \alpha > 0 &\quad \textit{if } \gamma = 0, \\ \alpha > \sigma^2/2 &\quad \textit{if } \gamma = 1/2 \quad (\textit{a.k.a. the Feller condition}) \\ \alpha > 0 &\quad \textit{if } \gamma > 1/2 \end{aligned}$$

and β satisfies one the following 3 conditions (of which the third is mainly stated for clarity)

$$\begin{aligned} \beta < 0 &\quad \textit{if } \gamma < 1, \\ \beta < \sigma^2/2 &\quad \textit{if } \gamma = 1, \\ \beta \textit{ free} &\quad \textit{if } \gamma > 1. \end{aligned}$$

Proof. The first part follows from Broze, Scaillet & Zakoian (1995). The second part follows from Conley et al. (1997) and can be checked using speed and scale measure calculations.

Now consider

$$f_{t+1} = f(r_{t+1}, r_t; \rho) = \begin{pmatrix} \epsilon_{t+1} \\ \epsilon_{t+1}^2 - \sigma^2 r_t^{2\gamma} \end{pmatrix} = \begin{pmatrix} r_{t+1} - r_t - \alpha - \beta r_t \\ (r_{t+1} - r_t - \alpha - \beta r_t)^2 - \sigma^2 r_t^{2\gamma} \end{pmatrix}.$$

Under the econometric null hypothesis (Equations (3)-(6)) we have that $\mathbf{E}(f_{t+1}|\mathcal{F}_t) = 0$. Introduce the instruments

$$I_t = I(r_t; \rho) = \begin{pmatrix} 1 \\ r_t \end{pmatrix},$$

and put $h_{t+1} = f_{t+1} \otimes I_t$. We then have that

$$\mathbf{E}(h_{t+1}|\mathcal{F}_t) = 0, \tag{7}$$

in fact we could have used *any* function of past values and parameters and still have this orthogonality condition. Define

$$g_T(\rho) = \frac{1}{T} \sum_{t=1}^{T} h_t.$$

The Generalized Method of Moments estimates ρ as

$$\rho^{GMM} = \arg\min \underbrace{g_T^\top(\rho) W_T(\rho) g_T(\rho)}_{:=J_T(\rho)},$$

where $W_T(\rho)$ is a positive-definite weighting matrix. It is this weighting matrix - possibly data- and parameter dependent - that constitutes the generalization in Hansen's approach. The GMM applies not only under the conditional orthogonality condition (7), which ensures that g_T is a martingale, but also under the much weaker unconditional version $\mathbf{E}(h_{t+1}) = 0$.

If the model is stationary then under the null hypothesis this estimator is consistent, i. e.,

$$\rho^{GMM} \overset{\Pr}{\to} \rho \quad \text{for } T \to \infty.$$

This is "fixed-$\triangle$" asymptotics, some sources in the literature consider asymptotics as $\triangle \to 0$. This might be reasonable if we were modeling the outcome of some controlled (say physical) experiment, but in economics it is characteristic that we as modelers do not control the experiments.

Let us also stress that the consistency is under the assumption that (3)-(6) (and not (2)) defines the true model. If (2) is the true model then the convergence of ρ^{GMM}

to something can often be justified using results form what is known in econometrics as "theory of misspecified models", but we would hesitate to call it consistency.

For the drift parameters the CKLS-discretization is a reparametrization of the conditional mean of the process in (2) (cf. Equation (17)), for the volatility this is not the case.

Note that in the case of a just identified system ($\dim(h_{t+1}) = \dim(\rho)$) we attain 0 as the minimum by solving $g_T = 0$ (unless we are very silly or extremely unlucky), irrespective of the choice of weighting. In the over identified case this is not true. But note also that different instruments will give rise to different estimators even in the just identified case. One can show that the optimal choice of weighting matrix (i.e., the one that has the smallest asymptotic covariance matrix in the partial order positive definite symmetric matrices) is $W_T(\rho) = S^{-1}(\rho)$ where

$$S(\rho) = \mathbf{E}(g_T(\rho) g_T^\top(\rho)).$$

In this case the estimator is asymptotically normal,

$$\sqrt{T}\left(\rho^{GMM} - \rho\right) \stackrel{\sim}{\to} N(0, \Omega) \quad \text{for} \quad T \to \infty,$$

where

$$\Omega = (\, D^\top(\rho) S^{-1}(\rho) D(\rho))^{-1} \tag{8}$$

and $D(\rho) = T^{-1} \sum(\partial h^i / \partial \rho^j)$ is the Jacobian of g_T (asymptotic normality also holds with non-optimal weighting matrices, but the covariance matrix becomes more complicated, see Hansen (1982).) This can be used to create confidence intervals and perform t-tests on individual parameters in standard fashion. Composite hypotheses can by tested by evaluating the test statistic

$$T(J_T(\rho^{GMM,R}) - J_T(\rho^{GMM})),$$

where $\rho^{GMM,R}$ is the restricted estimate, in a $\chi^2(k)$-distribution, where k is the number of restrictions. For this asymptotic distribution result to apply we have to use the efficient weighting matrix from in unrestricted model in the optimizations.

Another important question is whether we can produce a better discrete time approximation than the one given by (3)-(6). The first conditional moment of (2) is well-known, but higher order moments are not. Only if we use exact/true moments (but they could be unconditional) of the diffusion model will we have $\rho^{GMM} \to \psi$. We could try to determine the moments by intensive simulation or by numerical integration of solutions to partial differential equations. But, rather we develop (in Appendix A) what we consider a very accurate analytical approximation that has the virtue of being exact for the Vasicek, Cox-Ingersoll-Ross and Brennan-Schwartz models. In Tables GMM estimates with these moment approximations will be called $\rho^{GMM(2)}$ and we shall in the following refer to GMM estimates based on Euler moment approximation as $\rho^{GMM(1)}$ – this "(1)/(2)"-notational philosophy will be maintained throughout the paper.

We can perform the exact same estimation procedure with the new approximation substituted in appropriate places - the main difference being that the last two rows of the Jacobian D are more complicated.

2.2 Indirect Inference

The basic idea behind indirect inference (see Gouriéroux, Monfort & Renault (1993)) is to invert the relation – often called the binding function – between parameters in a misspecified model (such as in the previous subsection) and parameters in data generating process (the diffusion); in the case where the misspecified model has more parameters than the diffusion the inversion is more of a projection.

We consider an auxiliary model where an estimator, ρ^{AUX}, is found by solving

$$\frac{1}{T}\sum_{t=1}^{T} s_g(r_{t-1}, r_t; \rho) = 0.$$

The function s_g (mapping into $\mathbb{R}^{d_{AUX}}$, where $d_{AUX} = \dim(\rho)$) is called the score generator and should be a known function. It could be the logarithmic derivative of certain a class of densities or it could be a moment-inspired estimation function. The dimension of ρ could be (much) larger the dimension of the unknown parameter in the data generating process (the underlying diffusion.)

Define

$$m(\psi, \rho) = \int\int s_g(x, y; \rho) p(\triangle, x, y; \psi) p^{\infty}(x; \psi) dy dx.$$

The Indirect Inference estimator is then defined as

$$\psi^{II} = \arg\min_{\psi} m^{\top}(\psi, \rho^{AUX})\, W\, m(\psi, \rho^{AUX}), \tag{9}$$

where W is a positive define weighting matrix (like the one in the GMM estimation.) Note that ψ^{II} depends on the data only through ρ^{AUX}. Thus we can argue qualitatively that if some "effect in the data" is not captured by the auxiliary estimation procedure, then it will not be captured by the II-estimator, even if the diffusion has "the effect".

There are then 3 increasingly realistic cases to consider (where $\boxed{\textbf{TURC}}$ means "Then/That/True under regularity conditions".)

Case 0 is where m is known exactly or "with neglible error". $\boxed{\textbf{TURC}}$ as T tends to infinity we have

$$\psi^{II} \xrightarrow{Pr} \psi$$

and

$$\sqrt{T}(\psi^{II} - \psi) \xrightarrow{\sim} N(0, (D_\psi^\top W D_\psi)^{-1}),$$

where $D_\psi = \partial m/\partial \psi^\top$. This case is not very realistic in practice. It is, however, useful to neglect any error/uncertainty related to m at first when we want to prove consistency and asymptotic normality. Also, it allows easier analysis of the dependency of the asymptotic covariance matrix on s_g. For example the idea behind the efficient method of moments is to show that if the s_g-functions (indexed by ρ with $\dim(\rho) \to \infty$) are logarithmic derivatives of densities in a "dense subset" of "all densities" then the asymptotic covariance matrix becomes the same as if we had performed ordinary maximum likelihood in the diffusion model.
In **Case 0** the time-consumption needed to compute the criterion function grows linearly in T. (The time needed to solve the optimization part is usually ignored in the part of the literature.)

Case 1 is where m is determined by exact simulation. Typically this would be as

$$\widetilde{m} = \widetilde{m}_{N(T)}(\psi, \rho^{AUX}) = \frac{1}{N(T)} \sum_{n=1}^{N(T)} s_g(\tilde{r}_{n-1}(\psi), \tilde{r}_n(\psi); \rho^{AUX}), \tag{10}$$

where $N(T)$ is a large integer chosen by the statistician and $\tilde{r} = \{\tilde{r}_n(\psi)\}_{n=0}^{N(T)}$ is a simulated sequence of observations ($\triangle$ apart) from the diffusion (ideally started with a draw from the invariant distribution, in practice this is often achieved by using a "burn-in" period.) The indirect inference estimator is now computed with m substituted by $\widetilde{m}$. $\boxed{\textbf{TURC}}$ ψ^{II} is consistent provided $N(T) \to \infty$. Suppose further that

$$\frac{T}{N(T)} \to \tau \quad \text{for } T \to \infty \tag{11}$$

for some $\tau \in [0; \infty]$. Then (Gouriéroux et al. (1993) or Duffie & Singleton (1993))

$$\sqrt{T}(\psi^{II} - \psi) \xrightarrow{\sim} N(0, (1+\tau)(D_\psi^\top W D_\psi)^{-1}),$$

i.e., $N(T)$ must grow at least as fast as T for a "standard" asymptotic normality result to hold. We also get a "natural scale" for large (but finite) T's; for example if we use 100 times as many simulated observations as "real" observations we would (asymptotically) get standard errors 0.5% higher than in **Case 0**; an increase that is likely to be neglible when considering that m (meaning $\widetilde{m}$ in practice) has to differentiated numerically to determine the standard errors, anyway. In this case the time-consumption needed to compute the criterion function grows at least quadratically in T.

Case 2 is the most realistic one, namely where the diffusion process cannot be simulated exactly, but we have to use some discretization scheme (cf. Kloeden et al. (1991)), say characterized by step size $\delta_S(T)$. This case is investigated in Broze, Scaillet & Zakoian (1998). For any $\delta_S(T) > 0$ an approximation to m is defined analogously to Equation (10). Suppose that the discretization scheme has weak order $\nu > 0$, i.e., in obvious notation

$$|\mathbf{E}(r_1^{(\delta_S)}) - \mathbf{E}(r_1)| \le C\delta_S^\nu,$$

where C is a constant that does not depend on δ_S. Typically ν is 1 or 2, but "pret-a-porter" higher order schemes can be found in Kloeden et al. (1991).[2] $\boxed{\textbf{TURC}}$ ψ^{II} is consistent provided $N(T) \to \infty$ and $\delta_S(T) \to 0$. Suppose that $\delta_S(T) \propto T^c$. To ensure asymptotic behaviour as in **Case 1** we must require that $\sqrt{T}\delta_S(T)^\nu \to 0$ or in other words that $c > 1/(2\nu)$. At the expense of including "non-standard correction terms" on the LHS of the "asymptotic normality relation", it may be possible to use lower c-values, see Broze et al. (1998).
In this case the time-consumption needed to compute the criterion function grows must grow at least as fast as $T^{2+1/(2\nu)}$.

We shall be using a very simple type of indirect inference, namely the one where the score generator is given by the h function of the previous subsection, so $\rho^{AUX} = \rho^{GMM(1)}$, the W matrix is also chosen similarly. Adhering to the rules in **Case 2** we

[2]On a practical note: High order schemes often dramatically increase the number of `flops` per step in the iteration.

should expect this give to us an estimator, ψ^{II}, that behaves asymptotically as what might be termed ψ^{GMM}, the GMM-estimator of Section 2.1 if correct conditional moments were used.

In practice we use $N/T > 30$ and a weak second order Taylor scheme in the simulations. The Jacobian D_ψ is calculated using symmetric perturbation of the parameter. The starting point in the optimization is naturally chosen as $\rho^{GMM(1)}$ and we search by a Quasi-Newton method, i.e., by repeatedly updating our estimate by the gradient pre-multiplied by "something that looks like" the inverse of the matrix of second partial derivatives of the object function. Specifically we are inspired by the form of the (**Case 0**) asymptotic covariance and use (since D_ψ is quadratic in this simple case)

$$\psi_j^{II} = \psi_{j-1}^{II} - \lambda_{j-1} D_\psi^{-1} m,$$

where $\lambda_0 = 1$ and the following λ_j's are chosen such that the objective function never increases. This is done until some convergence criteria are met. (In short: We are a BHHH optimization technique.) Since numerical differentiation is involved and the second derivative is only "asymptotically correct" it is somewhat optimistic to hope for second order convergence, but the method works fairly well (and fairly quickly.)

2.2.1 EMM as (or rather: is) Indirect Inference

In this subsection a very brief description of the ideas in the Efficient Method of Moments (EMM) is given, see for instance Gallant & Long (1997), Gallant & Tauchen (1996*b*), Gallant & Tauchen (1996*a*), or Andersen & Lund (1996) for more details. In the EMM, which has received considerable interest recently, the idea is to be able to pick the s_g's used when finding an II estimator from a very flexible class. It is first shown (cf. references just given) that if the s_g's can approximate the unknown loglikelihood arbitrarily well, then an estimator that is as efficient as the ML estimator can be constructed despite the need for $\dim(\rho)$ to tend to infinity. The next step (which was chronologically the first) is then to actually write down such a class of s_g's. The catch-22 for practical applications is that this has been done and implemented; the Semi-Non-Parametric class of densities introduced by Gallant and Nychka is dense in a Sobolev sense (see Gallant & Nychka (1987)). The corresponding SNP estimation program developed by Gallant and Tauchen is free available by ftp (direct your browser to `http://www.econ.duke.edu/~get/snp.html`.) It should be noted that many of the regularity conditions have still not formally been verified (or are not strictly true, cf. Andersen & Lund (1996))

A strong side of the EMM is its usefulness in multi-dimensional models with unobserved components (such as stochastic volatility models.) Also long as i) we can make an econometrically satisfactory model of the observed components and, ii) we are able to simulate from the data generating process, then we can apply the EMM.

2.3 Gaussian Quasi Maximum Likelihood

Letting $F(x;\psi) = \mathbf{E}(r_\triangle | r_0 = x)$ and $\phi(x;\psi) = \mathbf{V}(r_\triangle | r_0 = x)$, we may be inspired to consider (minus twice) the Gaussian quasi loglikelihood function (quasi because it is not the exact likelihood "but close", no other meaning in the word than that)

$$l_T^G(\psi) = \sum_{t=1}^{T} l_t = \sum_{t=1}^{T} \left[\ln \phi(r_{t-1};\psi) + \frac{(r_t - F(r_{t-1};\psi))^2}{\phi(r_{t-1};\psi)} \right]$$

We find estimators by minimizing the Gaussian quasi loglikelihood, which can be seen as solving $\partial_\psi l_T^G = 0$, or as minimizing $(\partial_\psi l_T^G)^\top \partial_\psi l_T^G$. In the literature the variance matrix is usually estimated as $2(\sum_t l_t l_t^*)^{-1}$. This estimator relies on the information equality from likelihood theory

$$-\mathbf{E}\left(\frac{\partial^2 \ln L}{\partial\psi \partial\psi^\top} \right) = \mathbf{E}\left(\partial_\psi \ln L \partial_\psi^\top \ln L \right),$$

i.e., that the expected information equals the expected outer product of the score vector. This easy to calculate but not entirely correct since we are not using the true likelihood. We could use (8), but that's like rearranging the deck chairs on the Titanic.

Note that

$$\begin{aligned} \mathbf{E}(\partial_\psi l_T^G | \mathcal{F}_{T-1}) &= \partial_\psi l_{T-1}^G - \frac{2\partial_\psi F}{\phi} \mathbf{E}((r_t - F)|\mathcal{F}_{T-1}) - \frac{\partial_\psi \phi}{\phi^2}(\mathbf{E}((r_t - F)^2 | \mathcal{F}_{T-1}) - \phi) \\ &= \partial_\psi l_{T-1}^G, \end{aligned}$$

so looking backwards we could say that it is a GMM technique and looking forward we could say that we use an unbiased martingale estimating function.

Of course we still do not know the conditional second moments so in the implementation we can use i) the Euler approximation used in the first part of the section on GMM (thus obtaining $\rho^{GQML(1)}$) or, ii) the improved approximations from Appendix A (giving $\rho^{GQML(2)}$.) Following the **Case 2** rules of the Indirect Inference section we could determine the conditional moments by simulation and get an estimator, ψ^{GQML}, that would behave asymptotically as if the moments were known analytically.

2.4 Martingale Estimating Equations

The GMM and II/EMM described in the two previous subsections stem from the econometrics literature. Estimation techniques for discretely observed diffusions have also been treated in the statistics literature (the intersection between statistics and econometrics is smaller than one would think.) An estimating function, $G_T(\psi)$, is a function of ψ (and the observations, notationally suppressed) that we use for finding an estimator as a solution to the estimating equations

$$G_T(\psi) = 0.$$

It turns out to be more convenient/general to work with the estimating functions rather than the estimators themselves.

The unbiased martingale estimating functions constitute a particularly convenient class. An estimation function is in this class if G_T is a (0-mean)-martingale, by which we mean that

$$\mathbf{E}(G_i(\psi)|\mathcal{F}_{i-1}) = G_{i-1}(\psi) \text{ for all } 1 \leq i \leq T.$$

In Bibby & Sørensen (1995) it is shown (in a diffusion setting) $\boxed{\textbf{TURC}}$ estimators found by using unbiased martingale estimating functions *i*) exist (eventually), *ii*) are consistent (but not unbiased), and *iii*) are asymptotically normal. Largely, this should be no surprise, since the martingale property induces a (strong) orthogonality condition and qualitatively we would think of using the arguments from the GMM technique. An advantage, however, of the work of Bibby & Sørensen (1995) is that the technical conditions needed are easier to verify and that the proofs are more direct.

Another advantage is that we are able to give stronger quantitative statements because of our knowledge of the structure of the space of estimating functions. To illustrate this, we will focus on the class of quadratic martingale estimating functions. We see that the function

$$\begin{aligned} G_T^{\text{QM}} &= \sum_{t=1}^{T} \big(g_1(r_{t-1};\psi)[r_t - F(r_{t-1};\psi)] \\ &\quad + g_2(r_{t-1};\psi)[(r_t - F(r_{t-1};\psi))^2 - \phi(r_{t-1};\psi)]\big) \end{aligned} \tag{12}$$

is a martingale for any choice of deterministic functions g_1, g_2. These functions (similar to the instruments in the GMM estimation) can depend on ψ in essentially any way we like. The asymptotic covariance matrix of the estimator will depend on the choice of g-functions. Note that the Gaussian loglikelihood derivative in the previous subsection, $\partial_\psi l_T^G$, is an example of a member of this class. From Sørensen (1997) we have the following result.

Result 3 *The optimal estimators in the class of quadratic martingale estimating functions corresponds to the following choice of weight functions*

$$g_1^\star(x;\psi) = \frac{\nu(\triangle,x;\psi)\partial_\psi\phi(\triangle,x;\psi) - \Psi(\triangle,x;\psi)\partial_\psi F(\triangle,x;\psi)}{\phi(\triangle,x;\psi)\Psi(\triangle,x;\psi) - \nu(\triangle,x;\psi)^2} \tag{13}$$

$$g_2^\star(x;\psi) = \frac{\nu(\triangle,x;\psi)\partial_\psi F(\triangle,x;\psi) - \phi(\triangle,x;\psi)\partial_\psi\phi(\triangle,x;\psi)}{\phi(\triangle,x;\psi)\Psi(\triangle,x;\psi) - \nu(\triangle,x;\psi)^2} \tag{14}$$

with the notation

$$\begin{aligned}\nu(\triangle,x;\psi) &= \mathbf{E}((r_\triangle - F(\triangle,x;\psi))^3 | r_0 = x)\\ \Psi(\triangle,x;\psi) &= \mathbf{E}((r_\triangle - F(\triangle,x;\psi))^4 | r_0 = x) - \phi(\triangle,x;\psi)^2\ .\end{aligned}$$

Notice that if observations are truly Gaussian then $\nu = 0$ and $\Psi = 2\phi^2$ and we get the same weights as for Gaussian quasi likelihood analysis. The optimality is both asymptotically and in finite samples (in the sense defined in Godambe & Heyde (1987)).

Because the derivatives of the 1st and 2nd centered conditional moments enter into expressions (13) and (14), it is all the more important from a time consumption point of view to have analytical expressions for (approximations to) the moments. We could calculate the estimators by finding moments using simulation (in Pedersen (1994*a*) it is even shown how to simulate differentiated moments directly); but the fun of simulation is starting to flatten so we calculate *i*) $\rho^{OQM(1)}$ based on Euler moment approximations (so in fact we do not calculate it because it is identically equal to $\rho^{GQML(1)}$ by the remark above) and, *ii*) $\rho^{OQM(2)}$ based on the improved moment approximations of Appendix A. The quality of the approximations in the appendix decreases with order, but using only approximations in the weight functions does not "add to the misery of misspecification", it only effects the asymptotic covariance matrix. And our experience tells us that in practice it does not matter much. To calculate an estimate of the asymptotic covariance matrix of $\rho^{GQML(2)}$ we think of the problem in GMM terms, i.e., as solving $\min G_T^\top G_T$ rather than $G_T = 0$ and then use (8) (this ensures that we get the scaling right.)

Note by the way that if we put $G_T(\psi) = \sum_{i=1}^{T} \partial_\psi \ln p(\triangle, X_{i-1}, X_i)$, and assume regularity conditions that allow us to interchange integration and differentiation, we get

$$\begin{aligned}
\mathbf{E}\left(G_i(\psi)|\mathcal{F}_{i-1}\right) &= G_{i-1}(\psi) + \mathbf{E}\left(\partial_\psi \ln p(\triangle, X_{i-1}, X_i)|\mathcal{F}_{i-1}\right) \\
&= G_{i-1}(\psi) + \int_{\mathbb{R}} \frac{\partial_\psi p(\triangle, X_{i-1}, y)}{p(\triangle, X_{i-1}, y)} p(\triangle, X_{i-1}, y) dy \\
&= G_{i-1}(\psi) + \partial_\psi \int_{\mathbb{R}} p(\triangle, X_{i-1}, y) dy = G_{i-1}(\psi) + \partial_\psi 1 \\
&= G_{i-1}(\psi),
\end{aligned}$$

so maximization of the loglikelihood function results in an unbiased martingale estimation function.

2.5 Approximate Maximum Likelihood

In general we do not know an analytical expression for the transition density p. But we do know something. Relying on diffusion theory (e.g. along the lines of Karatzas & Shreve (1992)), we have $\boxed{\textbf{TURC}}$ the transition density p of the solution to a one-dimension version of the SDE (1) solves the parabolic partial differential equation (PDE) (called the *forward* or the *Fokker-Planck* or the *Kolmogorov equation*)

$$\frac{\partial}{\partial t} p(t,x,y) = -\frac{\partial}{\partial y}\left(\mu(y)p(t,x,y)\right) + \frac{1}{2}\frac{\partial^2}{\partial y^2}\left(\sigma^2(y)p(t,x,y)\right), \tag{15}$$

with initial condition $p(0,x,y) = \delta(y-x)$, where $\delta(\cdot)$ is the Dirac-δ function. (In the multi-dimensional case the transition density solves a similar PDE but it would be much harder (though not impossible) to solve it numerically since the space variable would be multi-dimensional and the "cross-derivatives" do not vanish.)

Strictly speaking this is not a PDE in the usual sense (it is often called a *functional* PDE) because the Dirac-δ function is not a function in the usual sense but a generalized function. It can be defined as a continuous, linear, real-valued mapping on the infinite dimensional space $\mathbf{C}^\infty(\mathbb{R})$ such that $\delta : f \mapsto f(0)$ for all $f \in \mathbf{C}^\infty(\mathbb{R})$. Therefore we cannot use standard theory to (define and) ensure existence and uniqueness of a solution. This has to be verified with other methods. Two possible ways are "inspired guesses" and "advanced probabilistic methods (Malliavan calculus) on the corresponding SDE". Rigorous treatment of this is far beyond this (most) paper(s), we refer to Pedersen (1995). But after (or before, for that matter) well-definedness

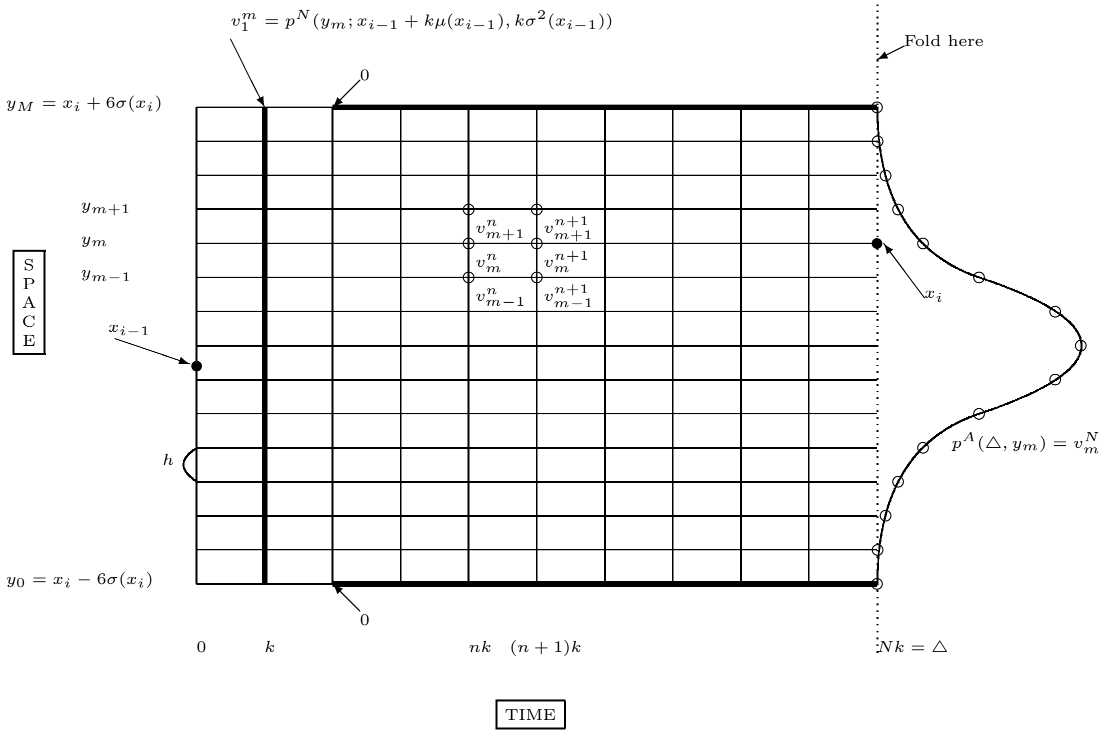

Figure 1: A finite difference grid for numerical solution of the forward PDE (15) for the transition density. How to specify the initial condition and making sure that x_i lies on a grid point are numerically important "tricks of the trade".

has been established we can apply numerical techniques known from "usual PDEs" to our hearts' content.

This is investigated in Poulsen (1999), where it is shown that a *careful* implementation of the Crank-Nicolson finite difference method does lead to a globally second order accurate approximation, p^A, to the true density and in turn to the true loglikelihood (and it also demonstrated that a *careless* does not.) The solution technique is illustrated in Figure 1. The idea basic to use a number of intermediate points (a grid) and then approximate the differential operators in (15) by difference operators, i.e., by linear combinations of the v's in Figure 1 (these then constitute the approximate solution.) For the Crank-Nicolson method the approximation involves six v's (the o's in the figure.) Since (or rather: if) the v's are known on the boundaries (the thick lines in the figure) this leads to a sequence of tridiagonal linear equations which can be solved with time-consumption $O(MN)$. Because the approximation is second order accurate in both time and space, it is always computationally optimal to have $h \propto k$, so we can pool them in the following and are able to construct (in a "numerically self-checking" manner) an approximate loglikelihood function that satisfies

$$l^A_{T,h}(\psi) \quad = \quad l_T(\psi) + h^2 a_T(\psi) + o(h^2) b_T(\psi), \tag{16}$$

where a_T and b_T are smooth (and deterministic) functions and h is a parameter chosen by the statistician such that the time needed to compute the left hand side of Equation (16) does not grow faster than T/h^2 (the "h^2" here is only place we can see the "h,k"-pooling.)

Let $\psi^{AML}_{T,h}$ denote the estimator obtained by maximizing $l^A_{T,h}$ (the maximization is performed using the same technique as in the Indirect Inference section.) The theory of misspecified models (cf. White (1994)) can be used to establish the convergence

$$\psi^{AML}_{T,h} \overset{Pr}{\to} \rho_h \text{ for any fixed } h$$

for some function (which turns out to be smooth) ρ of h. Since $h = 0$ corresponds to the ML estimator we have $\rho_0 = \psi$. The question is then: How slowly can we allow h to tend to 0 as T tends to infinity? (And precisely what happens then?) The answer is given in Poulsen (1999):

Result 4 *Suppose* $h(T) = T^{-c}$. $\boxed{\textbf{TURC}}$*:*

i) $\psi^{AML}_{T,h}$ *is consistent if (and only if)* $c > 0$.

ii) If $c = 1/4$, *then* $\sqrt{T}(\psi^{AML}_{T,h} - \psi)$ *converges in distribution to a normal variable or more precisely,*

$$\sqrt{T}\left(\psi^{AML}_{T,h} - \psi\right) \overset{\sim}{\to} N\left(\frac{1}{2}\frac{\partial^2 \rho_h}{\partial h^2}|_{h=0}, i^{-1}(\psi)\right).$$

iii) If $c > 1/4$ *then*

$$\sqrt{T}(\psi^{AML}_{T,h} - \psi) \overset{\sim}{\to} N(0, i^{-1}(\psi)),$$

i.e., the AML estimator is asymptotically equivalent to the ML estimator.

This means that for the AML method consistency, asymptotic normality and statistical efficiency can be achieved at a time-consumption that does not grow faster than $T^{3/2}$ (T (no. obs.) $\times\, T^{1/4}$ (time) $\times\, T^{1/4}$ (space).) This should be compared to $T^{2+1/(2\nu)}$ (where ν is the weak order of the discretization scheme) for indirect inference, and there we did not achieve statistical efficiency.

2.5.1 Model Control Using AML

By applying the ideas in Pedersen (1994*b*) it is easy to do model control with the AML method. Let $\Phi(\cdot, x)$ denote the distribution function of $r_\triangle | r_0 = x$ and introduce the *uniform residuals* through

$$U_i = \Phi(r_i, r_{i-1}) \quad \text{for } i = 1, \ldots, T.$$

Then (no matter how dependent the r_i's are)

$$\{U_i\} \text{ is a series of independent } U(0,1)\text{-distributed random variables.}$$

So we have to calculate

$$u_i = \int_{-\infty}^{x_i} p(\triangle, x_{i-1}, y) dy.$$

Now drop $\triangle$ and x_{i-1} from the notation and recall that for a "nice" function f we have

$$\int_a^b f(y)dy = \sum_{i=1}^{N-1} \left(\frac{1}{2} f(a+ih) + \frac{1}{2} f(a+(i+1)h) \right) h + O(h^2),$$

where $h = (b-a)/N$. Since (16) stems from a second order, appropriately uniform approximation to p, we get by not worrying too much about having "$a = -\infty$" that

$$\sum_{m: y_m \le x_i} \left(p^A(y_{m-1}) + p^A(y_m) \right) \frac{h}{2} = u_i + O(h^2),$$

so we retain second order accuracy in the calculation of the uniform residuals using only the points (the y_m's) where we already have an approximation to p.

We can then apply a batch of graphical and formal tests to the calculated uniform residuals to see if they "look like *iid* $U(0,1)$-variables." One of the most popular such is the Kolmogorov-Smirnov test statistic (cf. Press, Teukolsky, Vetterling & Flannery (1992) or any semi-advanced statistics book.) We should, however, be a bit careful with this (and other formal tests) because its distributional characteristics are developed under under the assumption that we use the true parameter in the residual calculation. But of course we cannot do that, we have to use an estimate (and what is more natural than the AML estimate we have at hand.) This changes the small sample and probably also the asymptotic distribution of the statistic. We do not know *how*, but conjecture that when the AML estimate is used it increases the variability of the statistic, so if the K-S value is not extreme in the "standard" asymptotic distribution, it certainly is not extreme in the "estimator dependent".

3 Data and Estimation Results

In theory the short rate is unambiguous - in practice it is not. Many rates may serve as proxy of the yield on an instantly maturing zero coupon bond. A full discussion of this matter is beyond the scope of this paper, but we refer to Duffee (1996) and Honoré (1998, Essay 2).

Estimation results in the literature depend on - besides the estimation technique which is the focus of this paper - the type of rate, sampling frequency, time period, and country considered. To separate effects we consider two data sets.

A) Monthly observations of U.S. one-month T-bill yields from June 1964 to December 1989 obtained from the CRSP bond data file. This is the exact the data set used in Chan et al. (1992). These data can be found on the CRSP bond data file.

B) Weekly observations of U.S. three-month T-bill yields from October 1982 to December 1995 obtained from H.15 release of the Federal Reserve System.

Both sets of rates have to be converted into continuously compounded yields. Summary statistics are given in Table 9.

All empirical results are given after the reparametrization

$$\theta = -\alpha/\beta, \qquad \kappa = -\beta.$$

Both for numerical optimization and for interpretation of parameters this is convenient (is does rule out the Merton model from Table 1, but that is a minor problem.)

3.1 The CKLS data (CRSP, monthly, 1964-89)

Results for the monthly data are given in Tables 3 and 4. We make to following observations:

- The GMM-estimates agree with those reported in Chan et al. (1992) - as of course they should when the same moment approximations are employed (but it is a good way of eliminating blatant errors.). When the new moment approximation is used the estimates change, but by no means "significantly" when the magnitude of the standard errors is considered. We note a fairly large reduction in the (estimated) standard error of the σ-parameter. But we should be very careful with the estimates since all methods estimate γ to be larger that 1, meaning that the econometric model is non-stationary (cf. Result 2.)

- In our implementation the II-estimates highly resemble the GMM-estimates.

- The moment approximations the GQML(2)-method differ only slightly from those used in Nowman (1997), and the estimates are also very similar, except for a change of the time-scale. (Nowman defines the time between observations is to be 1, so for weekly data his κ estimate must be multiplied by 12 and his σ estimate multiplied by $\sqrt{12}$ to yield our results. We think one should always measure time in years in interest rate models (and then make the r-units time^{-1}).) The γ-estimate (1.384, unaffected by using a different time-scale) is somewhat smaller than for the GMM-estimation, but the most noteworthy change is in the standard errors; these are reduced by a factor 2-3.

- The OQM method gives estimates that are very close to the GQML(2)-estimates indicating that the GQML(2)-estimates are "close to optimal". One might wonder why the estimated standard errors of the OQM estimates are in fact slightly larger than the GQML(2) estimated standard errors. We attribute this to small sample and numerical effects.

- The AML estimates are fairly close to the GQML(2) estimates but there appears to be a slight movement towards the GMM-estimates.

Between October 1979 and October 1982 the U.S. Federal Bank employed a monetary rather than an interest rate targeting policy (see Sanders & Unal (1988) and the references therein; we regard the change points as known dates rather than parts of the estimation problem). Often, this has lead people to allow different parameters in the pre-, during-, and post-monetary targeting periods. Table 4 reports the results of applying the AML technique to the three sub-periods (similar conclusions are reached with the other estimation procedures.) We make these observations:

- We strongly reject that parameters are equal in sub-periods. Twice the difference in optimal loglikelihood values equals 29.44, which when evaluated in a $\chi^2(8)$-distribution gives a test probability less than 0.01%. The same conclusion is reached in several places in the literature, but not in Chan et al. (1992). This is because they only allow one structural change (occurring in October 1979.)

- The speed of mean-reversion, κ, a very important parameter for term structure considerations differs a great deal between periods. Note that

$$\kappa_{\text{whole period}} < \inf\{\kappa_{\text{sub-period}}\}.$$

- The "long term levels" or unconditional means, θ, appear to be equal before and after the monetary targeting policy period and markedly higher during the period. Indeed, a(n approximate) likelihood-ratio test confirms this.

- The "average" short rate standard deviation per time unit, $\sigma\theta^{\gamma}$, is not significantly (again, by an approximate likelihood-ratio test) different "before" and "after", but is 3 times higher "during".

- The γ-parameter changes over time and much to our delight it is going down – a bit too much, perhaps, for those of us who like to think of the short rate as a diffusion (recall that Equation (2) is not well-defined for $\gamma \in]0; 1/2[$.) Analogously to the κ-estimates we have that

$$\gamma_{\text{whole period}} > \sup\{\gamma_{\text{sub-period}}\}.$$

  This indicates that one reason for "unsatisfactory" (i.e., very high) γ-estimates is failure to incorporate structural changes. In general it is our experience that the γ-estimates are not very robust and misspecifications or outliers tend to raise γ-estimates.

3.2 The Andersen-Lund data (FED, weekly, 1982-95)

The conclusions for the CKLS data set make it worthwhile to take a closer look at the period after 1982; this is where we use data set B. This data set has a weekly sampling frequency. Estimation results are shown in Table 5.[3] We remark that:

- Again, the GMM and II estimators are similar to each other, while the GQML, OQM and AML are quite alike and the three latter "claim" to have lower standard errors.

- All estimates indicate that $\gamma \in [1/2; 1]$, the OQM, GQML and AML give estimates around 0.78.

[3] This is after removal of the observation from Oct. 14 1987, the week of the stock market crash.

- The γ and σ estimates have much lower standard errors than for the whole CKLS-period.
- The evidence of mean reversion is still weak, none of the methods give a κ estimate significantly different from 0. But it should be noted that no invariant measure exists for $\kappa = 0$, so we cannot apply standard results to test this hypothesis.
- Graphical analysis of the uniform residuals (see Figure 8) gives no indication of a misspecified model. The sample mean is 0.5088 (1/2 for $U(0,1)$) and the sample variance is 0.07834 ($1/12 = 0.08333$ for $U(0,1)$.) There is no significant autocorrelation. A standard Kolmogorov-Smirnov test (which should be taken with the reservation outlined in Section 2.5.1) gives a test probability of 12.9 % of accepting the model.

4 A Sampling Experiment

We perform a Monte Carlo study of the behaviour of the various estimators. We consider the parameter settings given in Table 2.

Tables 6 and 7 report these results (where in fact some the ψ's should be ρ's, and probably vice versa):

<u>Estimates</u>
sample mean of $\{\hat{\psi}_i^j - \psi_i\}_{j=1}^M$
(sample standard deviation of $\{\hat{\psi}_i^j - \psi_i\}_{j=1}^M$) =: ($\widehat{se}_i$)

<u>Bias in standard errors</u>
For each $\hat{\psi}_i^j$ the algorithm provides an estimate of the standard error of this parameter estimate, say $\widehat{se}_i^j$.

bias(se_i) = sample mean of $\{\widehat{se}_i^j - \widehat{se}_i\}_{j=1}^M$
(sample standard deviation of $\{\widehat{se}_i^j - \widehat{se}_i\}_{j=1}^M$)

We note the following:

- We know that GMM(1) and GQML(1) are both biased and inconsistent, but for κ the effects are opposite. These estimates tend to "trade standard error for

Description	Symbol	Value
True parameter	ψ	$(0.1, 2, 0.2, 0.8)^\top$
Time between observations	$\triangle$	$1/12 = 0.08333$
No. observations in "real" data set	T	300
No. repetitions	M	500
No. observations in "sim." data set	N	10000
Max. no. searches in optimizations		10
Seed in `gasdev`		541
Perturbation for num. ψ-diff.		0.1%
FD grid width		6 app. std. dev.
FD space step size		10 basis points
FD no. time steps		16

Table 2: Parameter settings for the Monte Carlo experiments in Tables 6 and 7.

inconsistency". Once we use correct mean and a better variance approximation the standard errors increase.

- The κ estimates are in general quite poor (this is also the conclusion in, for instance, Jiang & Knight (1999).)
- Among the consistent or "almost" consistent estimates AML performs best.
- The estimated standard errors can be trusted.

The methods agree (unsurprisingly) almost perfectly on θ and unless we want (which we do not) to consider estimators (e.g. based on some kind of order statistics) of more robust but less efficient types, there is not much we can do about the small sample bias of κ-estimates. This means that the interesting thing is to take a closer look at the γ-estimates produced by the various methods (because γ- and σ-estimates are very highly correlated, much the same story would result from looking at the latter.) This is done in Figures 5 and 6, in particular we are interested in performance of AML estimates. We note the following:

- From Figure 5 we see that there is "some movement" of AML away from other estimates. And it is not just "very advanced numerical white noise" because then the sample std.dev. of estimates in Table 6 would increase for AML and the est.std.dev. in Table 7 would exhibit negative bias.

- Figure 6 shows the densities of the γ-estimates produced by the different techniques.

5 Estimating Non-Linear Drift Models Using AML

Models with non-linear drift effects have recently been suggested in the finance literature. One reason for this is the unsatisfactory κ-estimates, i.e., the very weak evidence of mean reversion. Specifications have been suggested in for example Ait-Sahalia (1996), Conley et al. (1997) and Stanton (1997). We look at the "parametric intersection" of the models suggested in two former articles,

$$dr_t = \left(\frac{a_{-1}}{r_t} + \kappa(\theta - r_t) + a_2 r_t^2 \right) dt + \sigma r_t^\gamma d\mathbb{W}_t.$$

The AML estimation method can readily be applied to this specification, even though the numerical optimization becomes somewhat more delicate. (Almost similar results can be obtained using a Gaussian approximation of Euler type to the likelihood function.) We focus on the FED data set, i.e., on the weekly data from the period 1982 to 1995. Estimation results are given in Table 8. We note the following:

- There appears to be no extra explanatory power in the non-linear drift terms. The uncertainty on the drift estimate, see Figure 9, is very high and the parameters are highly correlated. This agrees with the results in Chapman & Pearson (1999) where (towards to end) a non-linear least squares estimator based on an Euler approximation is used. (That paper also suggests a "parameter orthogonalization" that is easy to do in a regression framework.)

- Ait-Sahalia (1996) reports the non-linearity parameters in the drift to be significant. However, his way of estimating the standard errors of estimates may be somewhat optimistic in finite samples of highly persistent data; for a further discussion of this see Pritsker (1998). From the previous section we know that the finite sample behaviour of AML and other "almost likelihood" estimates of drift parameters may leave a lot to be desired, but still it is hard to see how one can do better.

- We think that the very motivation for these non-linear drift models makes it hard to verify them with conventional statistical methods. The non-linearities are "supposed" to have an effect for very large and very small values of the short rate. By definition such incidents will be rare, so it will be hard to obtain reliable inference. And a higher sampling frequency (= more observations) will not help when it comes to drift estimation.

- One approach (taken in Honoré (1998, Essay 3)) to obtaining significant and financially useful drift estimates is to use information from the whole term structure, not just from the short rate.

6 Conclusion

In this paper we investigated estimation methods for discretely observed diffusion models for the short rate of interest. In particular, we considered novel methods such as optimal martingale estimating equation methods and approximate maximum likelihood (AML) methods based on second order convergent numerical solution of the forward partial differential equation for the transition density. Using the CKLS short rate model as benchmark these methods were compared to well-known methods, namely GMM, II and GQML, both theoretically, in an application to U.S. data, and in Monte Carlo experiments. We find that the new martingale and likelihood methods reduce bias, true standard errors, and bias in estimated standard errors, relative to the established methods, in particular for the key parameter of interest, the elasticity of variance. In weekly data from 1982 to 1995, the new methods estimate this parameter to about 0.78. The quality of this result is strongly enhanced by the fact that an analysis of the uniform residuals from the AML estimation gave no indication of model misspecification. A related, but also interesting, result was that there is no (significant) extra explanatory power in the inclusion of non-linear terms in the drift function. This is in stark contrast to the conclusions reached in Ait-Sahalia (1996), but in agreement with the analyses and results in Pritsker (1998) and Chapman & Pearson (1999).

References

Ait-Sahalia, Y. (1996), 'Testing Continuous-Time Models of the Spot Interest Rate', *Review of Financial Studies* **9**(2), 385–426.

Andersen, T. G. & Lund, J. (1996), Stochastic volatility and mean drift in the short rate diffusion: sources of steepness, level and curvature in the yield curve. Working Paper. Aarhus School of Business.

Andersen, T. G. & Lund, J. (1997), 'Estimating Continuous-Time Stochastic Volatility Models of the Short-Term Interest Rate', *Journal of Econometrics* **77**, 343–377.

Barndorff-Nielsen, O. E. & Sørensen, M. (1994), 'A review of some aspects of asymtotic likelihood theory for stochastic processes', *International Statistical Review* **62**(1), 133–165.

Bibby, B. & Sørensen, M. (1995), 'Martingale estimating functions for discretely observed diffusion processes', *Bernoulli* **1**, 17–39.

Black, F. & Scholes, M. (1973), 'The pricing of options and corporate liabilities', *Journal of Political Economy* **81**, 637–654.

Broze, L., Scaillet, O. & Zakoian, J.-M. (1995), 'Testing for Continuous-Time Models of the Short-Term Interest Rate', *Journal of Empirical Finance* **2**, 199–223.

Broze, L., Scaillet, O. & Zakoian, J.-M. (1998), 'Quasi-Indirect Inference for Diffusion Processes', *Econometric Theory* **14**, 161–186.

Cambell, J., Lo, A. & MacKinlay, A. C. (1997), *The Econometrics of Financial Markets*, Princeton University Press.

Chan, K. C., Karolyi, G. A., Longstaff, F. A. & Sanders, A. B. (1992), 'An Empirical Comparison of Alternative Models of the Short-Term Interest rate', *Journal of Finance* **47**(3), 1209–1227.

Chapman, D. A. & Pearson, N. D. (1999), 'Is the Short Rate Drift Actually Nonlinear', Forthcomming in *Journal of Finance* .

Conley, T. G., Hansen, L. P., Luttmer, E. G. & Scheinkmann, J. A. (1997), 'Short-Term Interest Rates as Subordinated Diffusions', *Review of Financial Studies* **10**(3), 525–577.

Duffee, G. R. (1996), 'Idiosyncratic Variation of Treasury Bill Yields', *Journal of Finance* **51**, 527–551.

Duffie, D. & Singleton, K. (1993), 'Simulated Moments Estimation of Markov Models of Asset Prices', *Econometrica* **61**(4), 929–952.

Gallant, A. R. & Long, J. R. (1997), 'Estimating Stochastic Differential Equations Efficiently by Minimum Chi-Square', *Biometrika* **84**, 125–142.

Gallant, A. R. & Nychka, W. (1987), 'Semi-nonparametric maximum likelihood estimation', *Econometrica* **55**(2), 363–390.

Gallant, A. R. & Tauchen, G. E. (1996*a*), Specification Analysis of Continuous Time Models in Finance, *in* P. E. Rossi, ed., 'Modelling Stock Market Volatility: Bridging the Gap to Continuous Time', Academic Press, pp. 357–383.

Gallant, A. R. & Tauchen, G. E. (1996*b*), 'Which Moments to Match', *Econometric Theory* **12**, 657–681.

Godambe, V. & Heyde, C. (1987), 'Quasi Likelihood and Optimal Estimation', *International Statistical Review* **55**, 231–244.

Gouriéroux, C., Monfort, A. & Renault, E. (1993), 'Indirect Inference', *Journal of Applied Econometrics* **8**, 85–118.

Hansen, L. P. (1982), 'Large Sample Properties of Generalized Method of Moments Estimators', *Econometrica* **50**, 1029–1054.

Honoré, P. (1998), Five Essays on Financial Econometrics in Continuous-Time Models, PhD thesis, Department of Finance. Aarhus School of Business.

Jiang, G. & Knight, J. (1999), 'Finite sample comparsion of alternative estimators of ito diffusion processes: a Monte Carlo study', *Journal of Computational Finance* **2**, 5–38.

Karatzas, I. & Shreve, S. (1992), *Brownian Motion and Stochastic Calculus*, 2 edn, Springer-Verlag.

Karlin, S. & Taylor, S. (1981), *A Second Course in Stochastic Processes*, Academic Press.

Kessler, M. (1996), Simple estimating functions for a discretely observed diffusion process. Research report no. 336, Dept. of Theoretical Statistics. University of Aarhus.

Kloeden, P. E., Platen, E. & Schurz, H. (1991), *Numerical Solution of SDE Through Computer Experiments*, Springer.

Marsh, T. A. & Rosenfeld, E. R. (1983), 'Stochastic Processes for Interest Rates and Equilibrium Bond Prices', *Journal of Finance* **38**(2), 635–645.

Nowman, K. B. (1997), 'Gaussian Estimation of Single-Factor Continuous Time Models of the Term Structure of Interest Rates', *Journal of Finance* **52**(4), 1695–1706.

Øksendal, B. (1995), *Stochastic Differential Equations*, 4. edn, Springer-Verlag.

Pedersen, A. R. (1994*a*), Quasi-likelihood for Discretely Observed Diffusion Processes. Research report no. 295. Department of Theoretical Statistics. University of Aarhus.

Pedersen, A. R. (1994*b*), Uniform Residuals for Discretely Observed Diffusion Processes. Research report no. 292. Department of Theoretical Statistics. University of Aarhus.

Pedersen, A. R. (1995), 'A New Approach to Maximum Likelihood Estimation for Stochastic Differential Equations Based on Discrete Observations', *Scandinavian Journal of Statistics* **22**(1), 55–71.

Poulsen, R. (1999), Approximate Maximum Likelihood Estimation of Discretely Observed Diffusion Processes. Working paper no. 29, Centre for Analytical Finance, University of Aarhus.

Press, W., Teukolsky, S., Vetterling, W. T. & Flannery, B. (1992), *Numerical recipes in C*, 2. edn, Cambridge University Press.

Pritsker, M. (1998), 'Nonparametric Density Estimation and Tests of Continuous Time Interest Rate Models', *Review of Financial Studies* **11**(3), 449–487.

Sanders, A. & Unal, H. (1988), 'On the intertemporal stability of the short term rate of interest', *Journal of Financial and Quantitative Analysis* **23**, 417–423.

Sørensen, M. (1997), Estimating Functions for Discretely Observed Diffusions : A Reviev, *in* I. Basawa, V. P. Godambe & T. R. L., eds, 'Selected Procedings of the Symposium on Estimating Functions', Vol. 32 of *IMS Leture notes - Monograph series*, pp. 305–325.

Stanton, R. (1997), 'A Nonparametric Model of Term Structure Dynamics and the Market Price of Interest Rate Risk', *Journal of Finance* **52**, 1973–2002.

White, H. (1994), *Estimation, Inference, and Specification Analysis*, Cambridge University Press.

A Approximate Moments for the CKLS Model

The following approximate moments are used for the CKLS model for calculation of the GMM(2), GQML(2) and OQM(2) estimators. Applying the Ito formula to $(\exp(\kappa t)X_t)$ and using the martingale property of stochastic integrals w.r.t. Brownian motion we get the well-known result

$$F(t,x;\psi) = \mathbf{E}(X_t|X_0 = x) = x\exp(-\kappa t) + \theta(1 - \exp(-\kappa t)). \tag{17}$$

This means that we can find consistent estimators of κ and θ without any knowledge of the volatility parameters. To obtain inference about σ and γ (and to find more efficient estimates of κ and θ) we need higher order moments. The following lemma is easy to verify.

Lemma 1 *The solution to the ODE*

$$g' = a_0 + a_1 t + a_2 e^{a_3 t} + a_4 g$$

with initial condition $g(0) = 0$ is given by

$$\begin{aligned} g(t) \quad = \quad & \Big(a_0a_3a_4 + a_1a_3 - a_0a_4^2 - a_1a_4 \\ & +(a_1a_3a_4 - a_1a_4^2)t \\ & +(-a_0a_3a_4 - a_1a_3 + a_4a_1 + a_4^2a_0 + a_2a_4^2)\exp(a_4t) \\ & + \ a_2a_4^2\exp(a_3t)\Big) / a_4^2(a_4 - a_3). \end{aligned}$$

For $n \geq 2$ the Ito formula gives us that

$$X_t^n - x_0^n = \int_0^t \left(n\kappa(\theta - X_s)X_s^{n-1} + \frac{\sigma^2 n(n-1)}{2} X_s^{n-2(1-\gamma)} \right) ds + \int_0^t n\sigma X_s^{n-1+\gamma} dW_s.$$

Taking conditional mean, using the Fubini theorem and the martingale property leads to

$$\begin{aligned} \mathbf{E}(X_t^n|X_0 = x) - x_0^n \quad = \quad & n\kappa\theta \int_0^t \mathbf{E}(X_s^{n-1}|X_0 = x)ds \\ & -n\kappa \int_0^t \mathbf{E}(X_s^n|X_0 = x)ds \\ & +\frac{\sigma^2 n(n-1)}{2} \int_0^t \mathbf{E}(X_s^{n-2(1-\gamma)}|X_0 = x)ds. \end{aligned}$$

Define $g_n(t) = \int_0^t \mathbf{E}(X_s^n|X_0 = x)ds$. For $\gamma \in \{0, 1/2, 1\}$ the above expression reduces to a recursive system of ODEs for $\{g_n|n \geq 2\}$ - where in fact g_n' is our object of interest.

Note that all g_n's are linear combinations of affine and exponential functions. Thus we can find g_n for $n = 2, 3, \ldots$ by (first using Equation (17) and then) repeatedly applying Lemma 1. By the binomial formula there is no problem as to whether we consider centered or raw moments.

For $\gamma \notin \{0, 1/2, 1\}$ we cannot determine moments by the above ODE method. However, we suggest using the log-quadratic approximation

$$\begin{aligned}\ln \tilde{\phi} &= (2 \ln \phi_0 - 4 \ln \phi_{1/2} + 2 \ln \phi_0)\gamma^2 \\ &\quad +(-3 \ln \phi_0 + 4 \ln \phi_{1/2} - \ln \phi_0)\gamma \\ &\quad + \ln \phi_0,\end{aligned}$$

where $\phi_j = \phi(\ldots; (\ldots, j))$ for $j = 0, 1/2, 1$. This approximation is exact for the Vasicek, the Cox-Ingersoll-Ross and the Brennan-Schwarz models. Also, it can be seen as a generalization of the approximation used in Nowman (1997). In Figure 7 the difference between this log-quadratic approximation, the CKLS-approximation and the mean determined by simulation is shown. It takes many simulations to determine errors in the log-quadratic approximation.

We use similar approximations for 3rd and 4th moments (these are needed for calculation of the optimal quadratic martingale estimators.)

B Tables of Results

Estimator	θ	κ	σ	γ
$\rho^{GMM(1)}$	0.0687	0.5971	1.2800	1.4940
	(0.010)	(0.381)	(0.817)	(0.248)
$\rho^{GMM(2)}$	0.0686	0.6124	1.2768	1.523
	(0.011)	(0.401)	(0.656)	(0.255)
ψ^{II}	0.0656	0.5988	1.2795	1.4935
	(0.012)	(0.394)	(0.642)	(0.252)
$\rho^{GQML(1)}$	0.0734	0.3264	0.9191	1.3658
	(0.019)	(0.204)	(0.204)	(0.078)
$\rho^{GQML(2)}$	0.0734	0.3307	0.9341	1.384
	(0.017)	(0.210)	(0.205)	(0.082)
$\rho^{OQM(2)}$	0.0734	0.3370	0.9531	1.388
	(0.017)	(0.220)	(0.207)	(0.086)
ψ^{AML}	0.0734	0.3265	1.023	1.396
	(0.020)	(0.228)	(0.238)	(0.083)

Table 3: Estimates of the unrestricted model $dr_t = \kappa(\theta - r_t)dt + \sigma r_t^{\gamma} d\mathbb{W}_t$ on data set A ("the CKLS U.S. data set"). The II estimation uses 50000 simulated observations. The AML uses 16 time steps and space steps of 1 basispoint.

Period	κ	θ	σ	γ	Approximate Loglikelihood
Jan. '64 - Dec. '89	0.0734	0.3265	1.023	1.396	1408.75
	(0.020)	(0.228)	(0.238)	(0.083)	
Jan. '64 - Sep. '79	0.0675	0.3368	0.6367	1.270	897.31
	(0.024)	(0.318)	(0.266)	(0.140)	
Oct. '79 - Sep. '82	0.1126	2.378	0.1361	0.3733	127.24
	(0.014)	(1.668)	(0.258)	(0.889)	
Oct. '82 - Dec. '89	0.07696	1.014	0.0393	0.2283	398.92
	(0.007)	(0.861)	(0.020)	(0.339)	

$2 \times$ diff. in loglikelihood values = 29.44
Test probability of restricted model = 0.0265 % (in $\chi^2(8)$-dist.)

Table 4: What is the effect of monetary targeting policy in the CKLS data set? The estimates are based on AML with the same settings as in Table 3.

Estimator	θ	κ	σ	γ
$\rho^{GMM(1)}$	0.0524	0.1724	0.0698	0.6729
	(0.017)	(0.147)	(0.024)	(0.122)
$\rho^{GMM(2)}$	0.0522	0.1720	0.0714	0.6740
	(0.015)	(0.140)	(0.024)	(0.122)
ψ^{II}	0.0522	0.1720	0.0702	0.6736
	(0.015)	(0.140)	(0.024)	(0.122)
$\rho^{GQML(1)}$	0.0510	0.1506	0.0953	0.7873
	(0.0184)	(0.1477)	(0.0186)	(0.0678)
$\rho^{GQML(2)}$	0.0512	0.1420	0.0892	0.7853
	(0.019)	(0.125)	(0.018)	(0.070)
$\rho^{OQM(2)}$	0.0522	0.1413	0.0897	0.7770
	(0.019)	(0.130)	(0.018)	(0.072)
ψ^{AML}	0.0511	0.1322	0.0931	0.7794
	(0.0173)	(0.1489)	(0.0182)	(0.0682)

Table 5: Estimates of the unrestricted model $dr_t = \kappa(\theta - r_t)dt + \sigma r_t^{\gamma} d\mathbb{W}_t$ on data set B. The II estimation uses 50000 simulated observations. The AML uses 8 time steps and space steps of 1 basispoint.

Estimator	Estimates			
	θ	κ	σ	γ
$\rho^{GMM(1)}$ *GMM2B*	0.1000477 (0.0031666)	1.992767 (0.393090)	0.1778859 (0.1387007)	0.683943 (0.294565)
$\rho^{GMM(2)}$ *GMM2B*	0.1000475 (0.0031663)	2.188506 0.474942	0.2331604 (0.1903302)	0.7498757 (0.321452)
ψ^{II} *EMM6*	0.1000042 (0.0031852)	2.190512 (0.480614)	0.2433365 (0.1665152)	0.7953451 (0.2733099)
$\rho^{GQML(1)} = \rho^{OQM(1)}$ *GQML1*	0.100054 (0.003162)	1.969823 (0.383635)	0.1858347 (0.1286432)	0.7124268 (0.2866288)
$\rho^{GQML(2)}$ *GQML2*	0.1000481 (0.0031630)	2.15927 (0.46402)	0.246081 (0.185447)	0.7822098 (0.3157391)
$\rho^{OQM(2)}$ *OQM*	0.100053 (0.003184)	2.155632 (0.46303)	0.242798 (0.181953)	0.7761021 (0.314758)
ψ^{AML} *AML*	0.100031 (0.003156)	2.158259 (0.462048)	0.2448245 (0.165242)	0.7998988 (0.2774589)

Table 6: Simulation based comparison of estimators, I: Estimates. All files in `optimize/`*filename*`.C`

Estimator	Bias in standard errors			
	θ	κ	σ	γ
$\rho^{GMM(1)}$ *GMM2B*	-9.4997e-05 (0.000686)	0.0059732 (0.044136)	-0.022599 (0.1062)	-0.022594 (0.047050)
$\rho^{GMM(2)}$ *GMM2B*	-9.5022e-05 (0.000686)	0.0056671 (0.067228)	-0.03007 (0.13590)	-0.025396 (0.055944)
ψ^{II} *EMM6*	-9.4123e-05 (0.000754)	-0.0066188 (0.082539)	-0.014507 (0.077618)	0.019707 (0.056468)
$\rho^{GQML(1)} = \rho^{OQM(1)}$ *GQML1*	4.9383e-06 (0.000685)	0.019623 (0.041800)	-0.003732 (0.090359)	0.0039019 (0.046275)
$\rho^{GQML(2)}$ *GQML2*	2.1283e-06 (0.000686)	0.020465 (0.066817)	-0.0043287 (0.13330)	0.0064468 (0.055439)
$\rho^{OQM(2)}$ *OQM*	-5.669672e-05 (0.0006612096)	-0.005071177 (0.06743576)	-0.01772764 (0.124072)	-0.01068873 (0.059056)
ψ^{AML} *AML*	-4.6828e-05 (0.000691)	0.012461 (0.062307)	-0.0060109 (0.12219)	0.00080375 (0.044669)

Table 7: Simulation based comparison of estimators, II: Bias in standard errors.

Estimator	Parameter estimate					
	a_{-1}	θ	κ	a_2	σ	γ
ψ^{AML}	0.0021	0.0530	2.315	-14.37	0.0955	0.7880
	(0.003)	(0.010)	(3.983)	(23.32)	(0.019)	(0.068)

Std. dev./correlation matrix of est. param.

0.003	0.200	-0.977	-0.950	-0.061	-0.067
0.200	0.010	0.010	0.115	-0.088	-0.097
-0.977	0.010	3.983	0.994	0.030	0.033
-0.950	0.115	0.994	23.322	0.032	0.034
-0.061	-0.088	0.030	0.032	0.019	0.994
-0.067	-0.097	0.033	0.034	0.994	0.068

Loglikelihood at non-linear ψ^{AML}	3347.8352 (mag. of num. err.: 0.1)
Loglikelihood at CKLS ψ^{AML}	3346.9673 (mag. of num. err.: 0.05)
Test probability of restricted model	42.0 % (in $\chi^2(2)$-dist.)

Table 8: Estimation of the non-linear drift model on FED data. AML uses 0.5 basis point as state space step size, grid width of 6 app.std.dev. and 16 steps in the time dimension.

C Figures

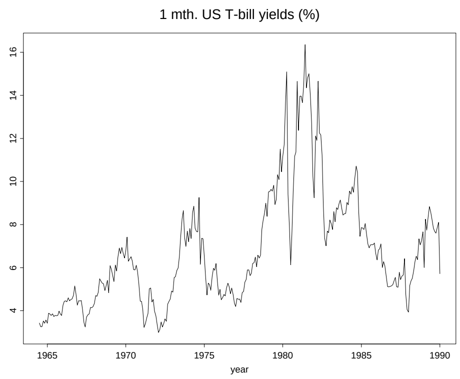

Figure 2: The CRSP data.

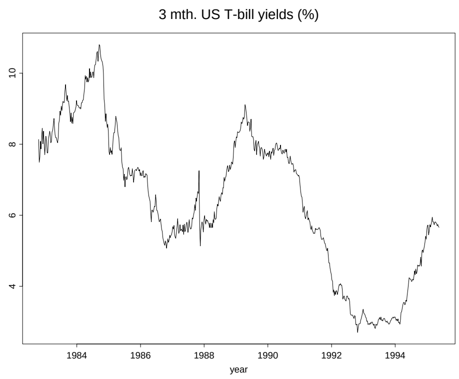

Figure 3: The FED data.

Data set A: Monthly yields of U.S. one-month Treasury Bills, 1964-89. (CRSP)

Variables	T	Mean	Standard Deviation	ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	ρ_6
r_t	307	0.06716	0.02671	0.95	0.91	0.86	0.82	0.80	0.78
$r_t - r_{t-1}$	306	0.00007	0.00824	-0.09	0.07	-0.12	-0.13	-0.03	-0.02

Data set B: Weekly U.S. three-month T-bill rates, 1982-95. (FED)

Variables	T	Mean	Standard Deviation	ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	ρ_6
r_t	655	0.006379	0.02044	0.99	0.99	0.99	0.99	0.99	0.98
$r_t - r_{t-1}$	654	-0.00004	0.00166	-0.06	-0.05	0.07	0.06	0.10	-0.02

Table 9: Summary statistics. The ρ's are autocorrelation coefficients.

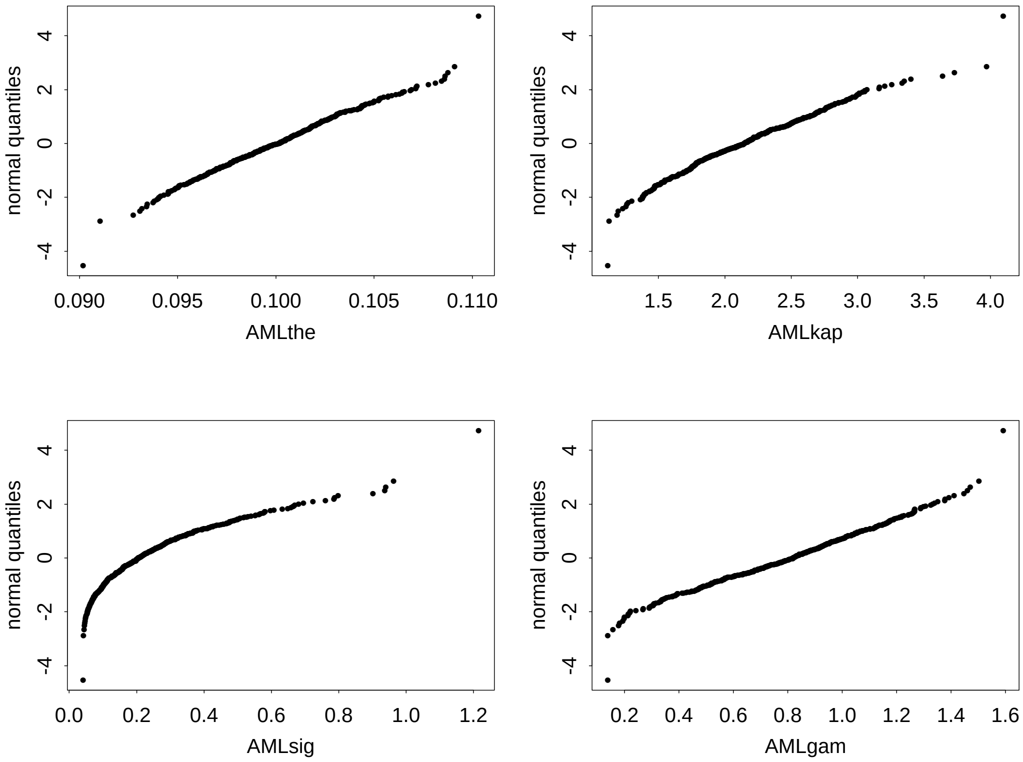

Figure 4: Normal quantile plots of AML-estimators. To the naked eye QQ-plots for the other proposed estimators are the same.

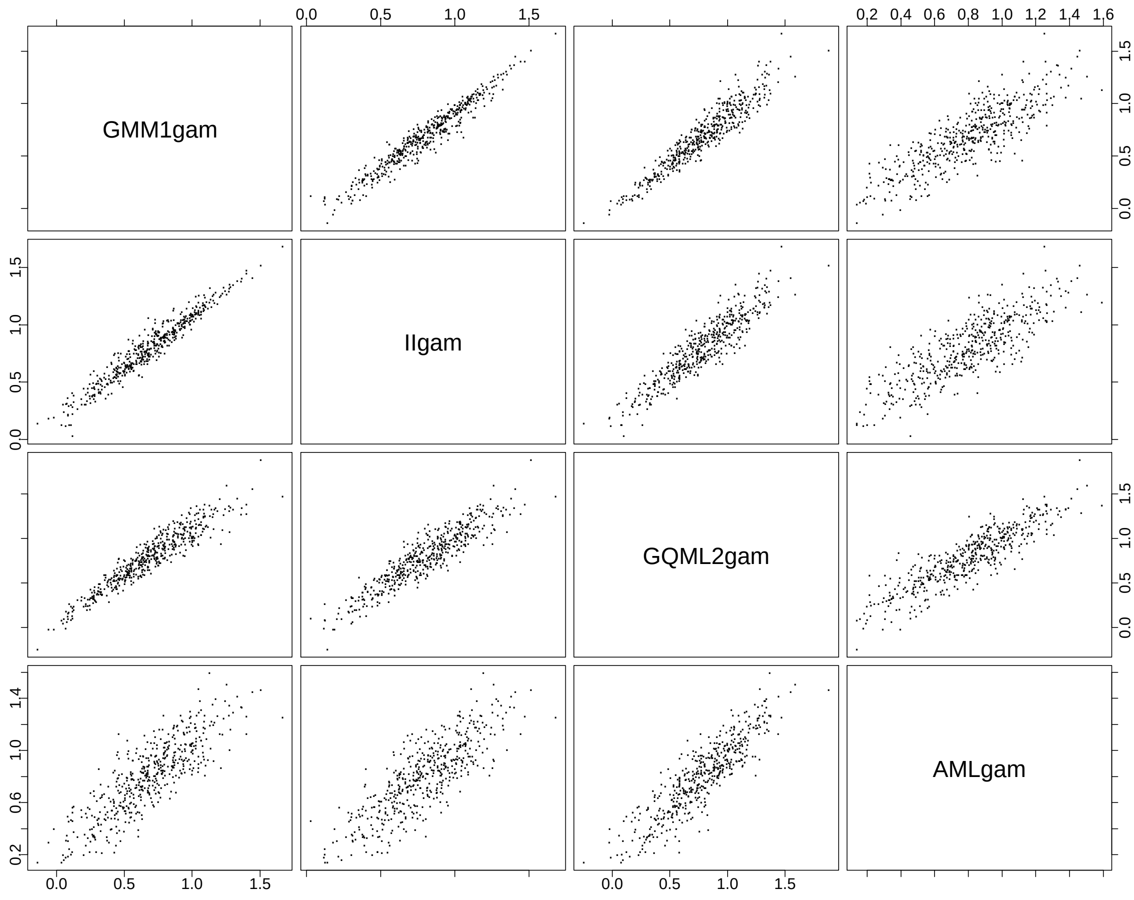

Figure 5: Scatter plots of γ estimates from the of GMM(1), II, GQML(2), and AML estimators.

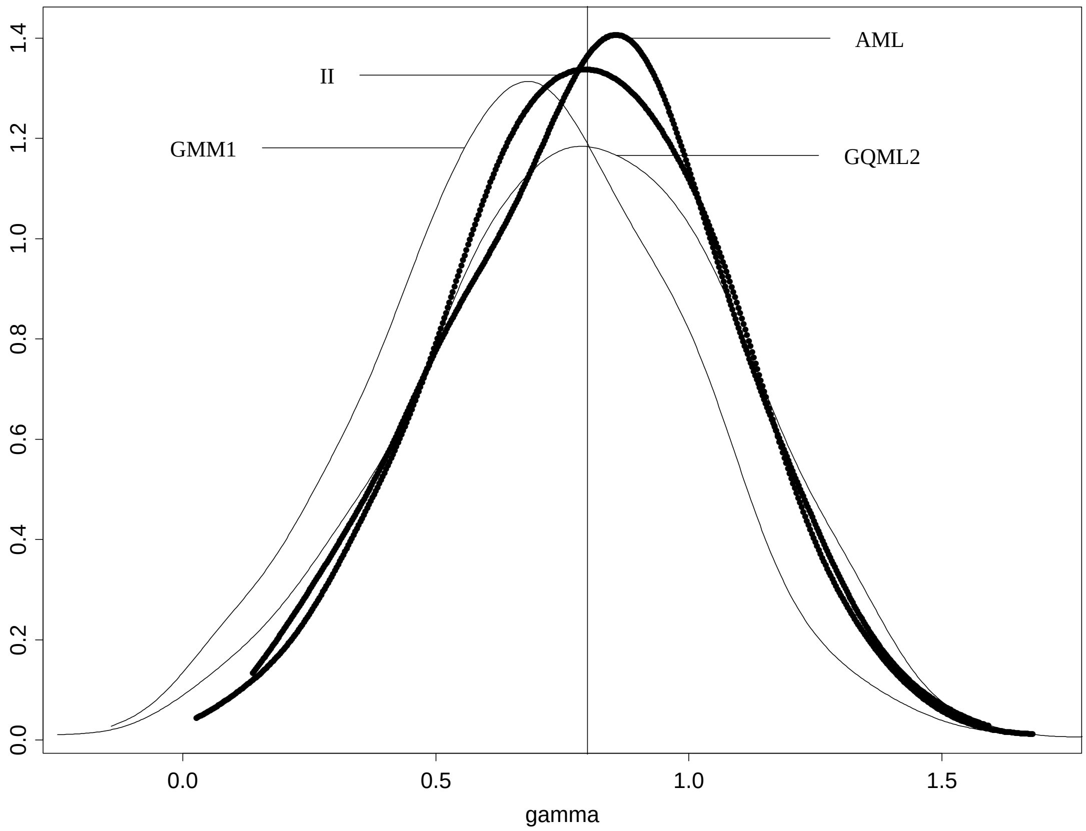

Figure 6: Empirical density functions of γ estimates (after appropriate smoothing for graphical purposes) for GMM(1), II, GQML(2), and AML estimators.

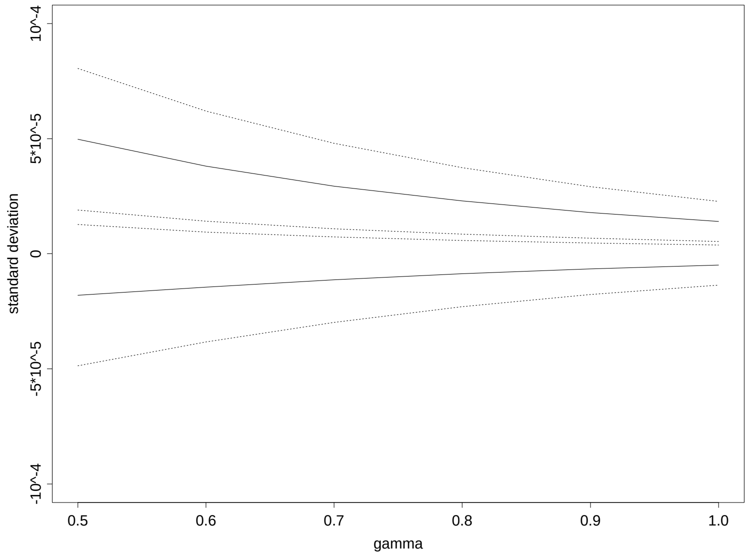

Figure 7: Approximations to second moments in the CKLS-model with $\kappa = 0.2339$, $\theta = x_0 = 0.0808$, $\sigma^2 = 0.0073$, and γ-value on the x-axis. Upper full curve: Difference between the CKLS approximation to the standard deviation of $r_\triangle$ and the standard deviation estimated by simulation of 1000000 realizations of $r_\triangle$ (dotted lines are 95% percent confidence intervals.) Low full curve: As upper full curve, except now the log-quadratic approximation is used.

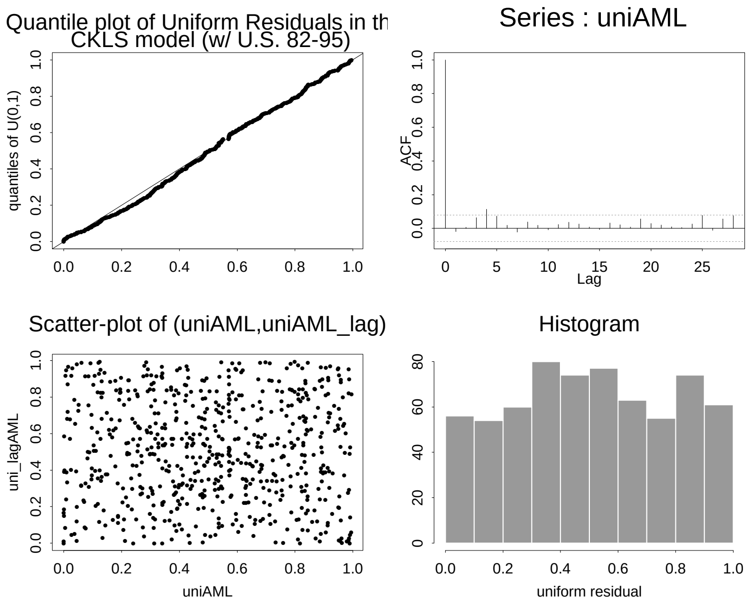

Figure 8: Graphical analysis of the uniform residuals from AML estimation of the CKLS-model on U.S. rates 1982-95 (FED data.)

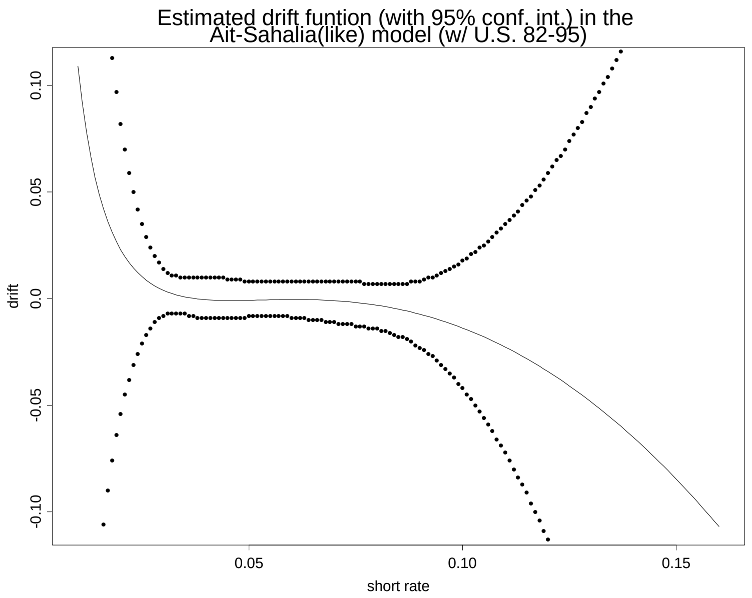

Figure 9: Estimated drift function and 95% confidence intervals for the non-linear drift model.

Should He Stay or Should He Go? Estimating the Effect of Sacking the Manager in Association Football

Rolf Poulsen
Department of Operations Research
University of Aarhus, Building 530
DK-8000 Aarhus C Denmark
rolf@imf.au.dk

June 3, 1999

1 Introduction

Professional football(=soccer) clubs often sack their managers. But does it help? Let's find out.

We use data from the top two English divisions. From memory, the English league is not the "most dangerous" for managers, there are other leagues where sackings are more frequent. It is, however, the only league where the amount of data accessible to us makes it possible to do "statistical analysis" rather than "case studies".

2 Collecting Data

We need two sets of data.

<u>Results of football matches.</u> These are easy to quantify. We need the results (and dates) of all matches from the season(s). These can be obtained either from Rothmans Football Yearbook (RFY), bought in computer readable form, or found (for recent seasons) on the Internet. Directing your browser to the home page of The

Association of Football Statisticians (`http://www.innotts.co.uk/~soccerstats/`) is a good way to start. We only consider league games.

Sacking of managers. This is somewhat more tangible information and - what is worse - not registered in structured form. This matter can be solved by cross-referencing the "Milestones diary" and the "English League Managers" sections in a volume of RFY (yielding information for one season). In this way we get exact dates and are sure that we have not missed any sackings. (The data used in the analysis can be found on the web-page `http://www.imf.au.dk/~rolf/`) The board of directors or the chairman of the club decides if the manager is sacked. The source may say that the manager "has been sacked", "has resigned", "leaves after mutual consent", or use some more poetic term. It is all the same to us and we will refer to it as a "sacking". In our analysis we use the time of sacking, not the time at which a new manager is appointed. We could make up a story that justified this, but the primary reason is that it is the easiest thing to observe. Many changes on the managerial front take place during the summer break (mid-May to mid-August). Several of these are in effect sackings but we are only able to analyse regular season sackings. The number of regular season manager sackings in the top two English divisions for the 1993-94 to 1997-98 seasons are given in Table 1. Judging from this Premier League managers have 25% risk of being sacked, while Division One managers have a sacking risk of over 40%. But as one would probably expect, this risk is not "evenly spread", clubs from top halves of tables rarely sack their managers.
Of course we would like more observations of sackings, but for now we will have to make do with the data in Table 1.

3 Estimating the "Sacking Effect"

Sackings rarely come as bolt from the blue; typically managers are sacked after a spell of poor results. In other words the "mechanism" that produces our observations (the sackings) is not independent of "the underlying stochastic process" (the results). This means that we have to be careful when we want to draw inference otherwise we might get out spurious results due to selection anomalies ("self-selection").

Interesting examples of this are given in Wainer, Palmer & Bristow (1998). To illustrate in this context, suppose a team has, purely by chance, performed very poorly and that its manager is sacked. Then we are likely to see an increase in performance

Season(s)	PL	D1
1993-94	6 (22)	9(24)
1994-95	8 (22)	13 (24)
1995-96	3*(20)	8*(24)
1996-97	7 (20)	9 (24)
1997-98	4 (20)	13 (24)
1993-98	28	52

Table 1: Number of regular season sackings in the Premier League and Division One. (*: Crude estimates, exact dates not available) Numbers parentheses indicate the total number of teams in the division. Note that it is possible for a club to sack more than one manager in a season.

after the sacking when compared to the "before" results. An obvious way to remedy this is to compare the results following the sacking to the "average" or "expected performance". This may remove spurious effects to some degree but not completely. To tell what "expected performance" is, we need an estimate. This estimate could be negatively affected by the string of poor, but unlucky, results that lead to the sacking in the first place.

We try to circumvent the problems by comparing the "after sacking" performance (the "treatment group") to that of teams which had an equally poor run of form/luck but did not sack the manager (the "control group"), see Figure 1.

We use two estimation methods:

i) A fully robust and very simple method that requires little more than counting and adding. However, this method is not very efficient. The strength of the opposition is not taken into account. In the long run this "noise" will "average out", but try telling that to a newly appointed manager whose first fixture is "Manchester United, away" rather than "Barnsley, home".

ii) The generalized linear (independent Poisson) model suggested in Lee (1997) that can be used to estimate home-advantage and relative strength of teams. This can be used to make a new measure of team performance. We then compare the "before" and "after" performance of "sacking" and "non-sacking" teams based on this measure. There is a slight possibility that this is a "Trojan Horse", i.e. it

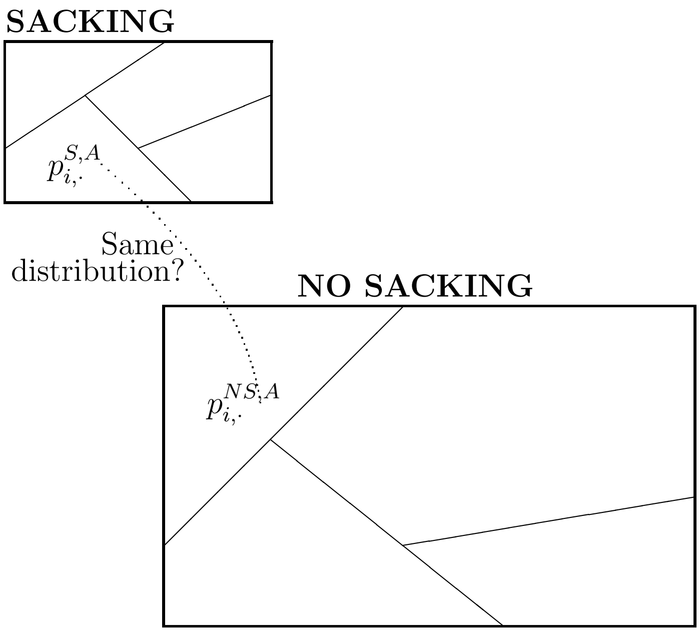

Figure 1: The sampling design.

brings back in the spurious effects, but since we still do the "treatment/control" comparison, this effect should be small compared to the "noise" that is removed.

3.1 Robust Estimation

The "treatment"and "control" samples are:

$\{p_{i,j}^{S,A}\}_j$	observed # points obtained in n_A matches after sacking following i points in the n_B-run prior to the sacking.
$\{p_{i,j}^{NS,A}\}_j$	observed # points obtained in n_A matches following a n_B-run that yielded i points with no sacking in the $n_A + n_B$-run.

We require that the "after"-observations in the "control" sample are from non-overlapping intervals for the individuals teams.

We choose $n_B = 6$ and $n_A = 3$ in the reported analysis, but, within reason, the results are insensitive to this choice. It is common to look at a period of 6 matches when discussing the "form" of team (as people reading the gambling sections of newspapers will know.) The length of the "after" period is a compromise; on one hand we want a short period so that "all other things are equal", but on the other hand we want a longer period because that gives us more data. It would be possible

# points “before”	No sacking #obs.	mean	std. dev.	Sacking #obs.	mean	std. dev.	Same mean? test pr.
≤ 3	243	3.653	2.09	11	4.091	2.09	0.494
4	208	3.735	2.14	11	4.818	2.40	0.109
5	257	3.836	2.03	7	4.000	2.82	0.834
6	298	3.721	2.13	9	4.333	2.00	0.394
7	335	4.065	2.17	7	4.142	3.71	0.927
8	357	3.767	2.25	6	3.833	2.99	0.897
≥ 9	722	4.297	2.23	8	2.875	2.30	0.073

Table 2: “High resolution” robust analysis with $n_B = 6$ and $n_A = 3$. The test probability is calculated using a likelihood ratio test with a independence, normality, and “row-wise equal variances” assumptions. The normality assumption is crude but not terribly important and the variances do in fact not differ significantly.

to use a more “continuous” weighting scheme, but we will not do that.

Table 2 gives descriptive statistics for the five seasons. The observed number of sackings in this table is smaller than the number of sackings in Table 1 since we have had to discard observations when sackings took place less than 6 games into the season or less than 3 before the end of the season.

We see that for all “# points before”-classes (the rows in Table 2) except the one corresponding to the best performing sacking teams, the teams that sack their manager have higher estimated expected points gain than teams that have an equally poor run of form but do not sack their manager. But, alas, none of these differences are (even vaguely) significant. (A pooled test over all rows accepts equal row-means with a probability of 44%.)

3.2 Refined Estimation

Suppose that the score X of a particular team in a particular match is Poisson(λ)-distributed,

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!} \text{ for } x = 0, 1, 2 \ldots.$$

Suppose further that *i*) matches are independent (at least with “sufficiently little information”), *ii*) the number of home and away goals are independent (this may be questionable, Dixon & Coles (1997) suggest an improvement, but it is not totally

Excess points "before"	No Sacking			Sacking			Same mean?
	#obs.	mean	std. dev.	#obs.	mean	std. dev.	test pr.
$]-6;-4]$	85	0.1909	2.26	9	0.7981	2.32	0.441
$]-4;-3]$	69	0.0508	2.39	10	1.8281	2.61	0.030
$]-3;-2]$	107	0.1913	2.30	10	0.2648	1.60	0.921
$]-2;-1]$	128	0.0292	2.35	10	-0.9798	1.90	0.184
$]-1;0]$	132	-0.1981	2.52	7	0.6007	2.51	0.410
$]0;1]$	178	0.0191	2.42	7	0.0759	2.21	0.950
$]1;6]$	376	-0.1787	2.39	6	-1.5978	1.18	0.147

Table 3: "High resolution" refined analysis with $n_B = 6$ and $n_A = 3$. Again, test probabilities calculated from the likelihood ratio under the assumptions of independence, normality and row-wise equal variances.

Excess points "before"	No Sacking			Sacking			Same mean?
	#obs.	mean	std. dev.	#obs.	mean	std. dev.	test pr.
$]-6;-3]$	154	0.1281	2.32	19	1.3402	2.48	0.033
$]-3;-1]$	235	0.1030	2.33	20	-0.3575	1.76	0.386
$]-1;6]$	686	-0.1311	2.42	20	-0.2425	2.08	0.838

Table 4: "Low resolution" refined analysis with $n_B = 6$ and $n_A = 3$. Test probabilities calculated as in Tables 2 and 3.

unreasonable), and *iii*) home and away scoring intensities in a particular match satisfy

$$\begin{aligned} \ln \lambda_{HOME} &= \beta + \beta_{HOME} + \beta_{OFF}(HOME\_TEAM) + \beta_{DEF}(AWAY\_TEAM), \\ \ln \lambda_{AWAY} &= \beta + \beta_{OFF}(AWAY\_TEAM) + \beta_{DEF}(HOME\_TEAM). \end{aligned}$$

To identify and interpret parameters we impose the restrictions

$$\sum \beta_{OFF} = 0, \qquad \sum \beta_{DEF} = 0.$$

Strictly speaking this specification is inconsistent with any "sacking effects", but we use it as a reasonable overall approximation to the season that can be used to account for some information that is definitely relevant.

This generalized linear model can easily be estimated using maximum likelihood, for example by the procedure `glim` in `Splus`. Once this has been done the entire probability distribution of the outcome of any match can be determined. In particular we can find the number of points a team can expect from a match. Now imagine that all points are substituted by "excess points", that is by "actual points" - "expected points". We can then perform largely the same analysis as before – except now the "before" points are real numbers so we have to group based on intervals. In Tables 3 and Tables 4 this has been done for a "high" and a "low resolution" case.

We see that for the sacking teams with the worst runs of form there is in fact a significant increase in performance compared to teams that do not sack their manager. But if teams are doing well, or just "not too bad" then there is no significant gain from sacking the manger, if anything there are some indications of a negative effect. But we have to do the individual row-wise analysis to find the effects, a pooled tests of equal row-means are accepted with test probabilities between 15% and 20%.

4 Conclusion

Overall, we can give no definitive answer to the question posed in the title of the paper. If we could, we would have used a different title. But if you are the chairman of a club that is performing really poorly, then sacking the manager is probably not a bad idea.

We find it quite encouraging for further research that we obtain markedly higher levels of significance with the refined analysis. Further research primarily means gathering more data. With more data we could also try to make a "real dynamic model", that is a model where among other things the cause of the sacking is included (as more or less stochastic effect dependent on performance) as well as a measure of "after performance".

References

Dixon, M. J. & Coles, S. G. (1997), 'Modelling Association Football Scores and Inefficiencies in the Football Betting Market', *Applied Statistics* **46**(2), 265–280.

Lee, A. J. (1997), 'Modeling Scores in the Premier League: Is Manchester United Really the Best?', *Chance* **10**(1), 15–19.

Wainer, H., Palmer, S. & Bristow, E. T. (1998), 'A Selction of Selection Anomalities', *Chance* **11**(2), 3–7.