Term structure of interest models: concept and estimation problem in a continuous-time setting

Orazio Di Miscia*

20 November 2004

*This research was carried out when the author was visiting the School of Economics and Management at the University of Aarhus. This research has been supported by a Marie Curie Fellowship of the European Community Programme Improving the Human Research Potential and the Socio-Economic Knowledge Base under contract number HPMT-CT-2000-00139. I would like to thank Bent Jesper Christensen and participants at the workshop "Topics in Applied Economics and Finance", University of Aarhus. Only the author is responsible for any omission and mistakes. E-mail *orazio.dimiscia1@bancaintesa.it*

Abstract

Continuous-time models have a large range of applications. They have been used for a long time to model phenomena evolving randomly and continuously in time. However, data are essentially always recorded at discrete points in time only and this is one of the main source of difficulties when the researcher is interested about their estimation. This paper review some of the estimation problems and focus the attention about the link between continuous-time stochastic process and the estimation of term structure interest rate.

1 Introduction

Continuous-time stochastic processes arise in many applications in economics, but perhaps nowhere do they play as large a role as in finance. Following the pathbreaking work of Merton (1969, 1973), and Black and Scholes (1973) the use of continuous–time stochastic processes has become a common feature of many applications, especially is asset pricing models.

This survey will look at the specification and estimation of continuous-time stochastic processes with a glance at the theory of term structure of interest rate models. Theoretical research have extensively used continuous-time models but few empirical research have pointed their attention to the estimation methods and relative problems. However there is at least a reason for the lack of empirical research in this area. Except for a small number of specifications, estimation of continuous-time stochastic processes can be a daunting challenge. In large part, the difficulties stem from "measurement" problems. Economic data do not come in the form of a continuous-time record. Available data are a sequence of observations. The data are often irregularly or randomly spaced. Moreover, the observations on the continuous-time process can take on a variety of forms. They may be measurements on the level of a process ('stocks'), such as an asset price, or an integral of a process between two points in time ('flows'), such as consumption, or a mixture of both types. The map that relates observed data to the parameters of the underlying process is generally very untidy. The specifications for which this mapping is fairly tractable have therefore received most of the attention of the empirical researchers. But this inordinate attention to "tractable" functional forms has been costly. Although these popular functional forms may make estimation easier, they are often inappropriate for providing useful answers. For example, the popular geometric Brownian motion specification does a notoriously poor job of mimicking high frequency asset price data.

The absence of estimation strategies for more "realistic" stochastic specifications has led to a variety of *ad hoc* empirical strategies that are less than satisfactory.

Section 2 briefly consider the term structure problem and state some basic definitions. Section 3 is devoted to define the continuous-time model, discussing the characterization of the likelihood of an Itô processes with discretely sampled data. Section 4 summarize , shortly, *some* of the estimation methods recently used. The appendix review definitions and some results about stochastic processes. A recommendation for notation: capital letters will be used for generic random variable and lower case letters for sample random variables. This could be an abuse of notation but in my opinion it will help to understand different set-up.

2 The term structure of interest rate

The quest for understanding what moves bond yields has produced an enormous literature. The difficulties are different and from different point of view.

For example, bond yield movements over time could be captured by simple vector autoregressions in yield and maybe other macroeconomic variables. Several aspects of bond yields, however, set them apart from other variables typically used in VAR studies. One aspect is that bonds are assets, and that bonds with many different maturities are traded at the same time. Bonds with long maturities are risky when held over short horizons, and risk-averse investors demand compensation for bearing such risk. Arbitrage opportunities in these market exist unless long yield are risk-adjusted expectations of average future shorts rate. Movements in the cross section of yield are therefore closely tied together. These ties show up as *cross-equation restrictions* in a yield-VAR, that are not so easy to impose. Another aspect of yields is that they are not normally distributed. This makes it difficult to compute risk-adjusted expected value of future short rates.

In order to build models that capture exactly these aspects of bond yields, the literature has evolved mostly in continuous time with a massive use of stochastic calculus and partial differential equation. The reason is the possibility to use the well established and elegance results of stochastic process to the mechanism of generating interest rate. However this has exacerbated some mathematical aspects increasing the gap between theoretical and empirical research in this field.

2.1 Purposes of term structure models

Understanding what moves bond yield is important for at least four reasons. One of these reasons is forecasting. Yields on long-maturity bonds are expected valued of average future short yields, at least after an adjustment for risk. This means that the current yields curve contains information about the future path of the economy. Yield spreads have indeed been useful for forecasting not only future short yields (see Campbell and Shiller (1991)) but also real activity and inflation (Ang, Piazzesi, and Wei (2002), Fama (1990)). These forecasts provide basis for investment decisions of firms, savings decisions of consumers, and policy decisions.

Monetary policy is a second reason for studying the yield curve.In most industrialized countries, the central bank seem to be able to move the short end of the yield curve. What matters for "aggregate demand", however, are long-term yields. For example householder base their decision on whether

to buy or rent a house on long-term mortgage rates not on the rate in the funds market which seems to be controlled by the central banks. For a given state of the economy, a model of the yield curve helps to understand how movements at the short end translate into longer-term yields.

Debt policy constitutes a third reason. When issuing new debt, governments need to decide about the maturity of the new bonds. For example, in the U.S.A. the Kennedy administration actively managed the maturity structure of public debt in the early 1960s in what is known as "operation twist". The treasury at the time was trying to flatten or invert the yield curve by selling short maturity debt and buying long maturity notes. The outcome of such operations depends crucially on how bond yields depend on the supply of bonds with different maturities.

Derivative pricing and hedging provide a fourth reason. For example, coupon bonds are priced as baskets of coupon payments weighted by the price of a zero-coupon bond that matures on the coupon date. Even the price of more complicated securities, such as swaps, caps and floors, futures and options on interest rates are computed from a given model of the yield-curve. Banks need to manage the risk of paying short-term interest rates on deposits while receiving long-term interest rates on loans. Hedging strategies involve contracts that are contingent on future short rates, such as swap contracts. To compute these strategies, banks need to know how the price of these derivative securities depends on the state of the economy.

2.2 Bond pricing in continuous time: some basic definitions

An inexpert researcher should affirm that term structure models are particular simple, since bond prices are just the expected value of the discount factor.

In equation, the price at time t of a zero-coupon bond that comes due at time $t + j$ is[1]:

$$P_t^j = \mathrm{E}_t(m_{t,t+j})$$

Thus, once you specify a time-series process for one-period discount factor $m_{t,t+j}$, you can in principle find the price of any bond by chaining together the discount factors and finding

$$P_t^j = \mathrm{E}_t(m_{t,t+1} m_{t+1,t+2} \dots m_{t+j-1,t+j})$$

[1]Large part of this paragraph is "randomly" drawn from Cochrane (2001) Duffie (2001) and Piazzesi (2003)

This "chain" can be hard to do. If you buy an N-period bond and then sell it (it has now become an $N-1$ period bond) you achieve a Hold Period Return at $t+1$ for an N-period bond equal to:

$$\begin{aligned} HPR_{t+1}^{N} &= \frac{P_{t+1}^{(N-1)}}{P_t^N} \\ \text{or} & \\ hpr_{t+1}^{N} &= \log(HPR_{t+1}^{N}) = \log P_{t+1}^{(N-1)} - \log P_t^N = p_{t+1}^{(N-1)} - p_t^N \end{aligned}$$

The holding period return at $t+n$,with $n < N$ is usually random, because it depends on the resale value of the bond P_{t+n}^{N-n} which is generally not known at time t.

But, the resale value is equal to its payoff (in our case the normalized price 1) when the bond matures, so that holding the bond until matures generates a return which is known at time t (in this case $P_{t+N}^{(N-N)} = 1$ and $hpr_{t+N}^{N} = -\log P_t^{(N)}$). In this case we can define the *yield to maturity* as:

$$y_t^{(N)} = \frac{hpr_{t+N}^{N}}{N} = \frac{-\log P_t^N}{N} \tag{1}$$

The *short rate* is the limit of yields as maturity approaches

$$r_t = \lim_{N \downarrow 0} y_t^{(N)}$$

Bonds are usually priced with the help of a so-called "risk-neutral probability measure" Q^*. Just like the name, risk-neutral pricing applies under Q^*. In other words, asset prices are the expected values of their future payoffs discounted at the riskless rate, where the expectation is computed using the probability measure Q^*. When agents are risk-neutral, the pricing result applies under the data-generating measure Q, but in general Q^* will be different from Q. Within this framework the price of a zero-coupon bonds is:

$$P_t^{(N)} = \mathrm{E}_t^* \left[exp\left(- \int_t^{t+N} r_u du \right) \right] \tag{2}$$

where E^* denote expectation under Q^*- Standard results show that if there is a risk-neutral probability measure Q^*, a system of asset prices is arbitrage free. Under the risk-neutral measure, expected excess returns on bonds are zero.

The pricing relation (2) shows that any *yield-curve model* consists of two ingredients:

1. the change of measure from Q to Q^* and

2. the dynamics of the short rate r under Q^*

In so-called *factor models* of the yield curve, point 2 is replaced by the following assumption:

2'. the short rate r is a function $g(X)$ of X and $X \in \mathbb{R}^n$ is a time-homogeneous Markov process under Q^*

This means that X is the relevant state vector, a vector of *factors*[2]. This modified (2') assumption implies that the conditional expectation in eq. (2) is some function f or time-to-maturity τ and the state X_t at time t, or

$$P_t^{(N)} = f(X_t, N)$$

This big advantage of pricing bonds (or any other assets) in continuous time is Ito's Lemma. The lemma says that smooth function f of some Ito process X at time t is again Ito processes. The lemma thus preserves the Ito property even if f is nonlinear. Ito's lemma allow to turn the problem of solving conditional expectation in eq. (2) into the problem of solving partial differential equation (PDE) for the bond price $f(X, N)$. The trick of computing eq. (2) by solving a PDE is called the Feyman-Kac approach.

Before give some details about this approach in what follow i consider that the agents are risk-neutral and so the risk-neutral probability measure Q^* and the true probability measure, Q, are equal. Under this condition it is possible to "skip" point 1 above and obviously the framework is again arbitrage-free under the true measure.

The idea of the Feynman-Kac approach is the following: the conditional expected value in eq. (2) can be see as the solution of the PDE for the bond price $f(X_t, N)$. The PDE can be obtained in four steps. First, the pricing equation (2) implies that the price of the bond at maturity is equal to its payoffs (here the bond price is taken to be cum-dividend). This means that $f(X_t, 0) = 1$ for all states of the world. Second, the pricing equation also shows that the bond price is the expected value of an exponential function, so $f(X_t, N)$ is strictly positive (which makes it possible to divide by f). Third, Ito's Lemma implies that $f(X_t, N)$ itself is an Ito process[3]:

[2]The notation could be cumbersome since that we define the price of the bond in terms of the short rate r. In a single-factor model $g(X) = X$ and the factor is imposed equal to r the short rate. In a multivariate framework X is a vector and one component is, usually, the short rate r

[3]For an unifying treatment of the continuous-time model from a mathematical point of view the Ito's process is defined in section 3.2. Here it is used in order to derive the PDE without "skip the origin"

$$\frac{df(X_t, N))}{f(X_t, N)} = a_f(X_t, N)dt + b_f(X_t, N)dz_t \tag{3}$$

with instantaneous expected bond return:

$$a_f = -\frac{f_N(X_t, N)}{f(X_t, N)} + \frac{f_X(X_t, N)}{f(X_t, N)} a_X(X_t) + \frac{1}{2} b_X^2 \frac{f_{XX}(X_t, N)}{f(X_t, N)}$$

where f_N,f_X,f_{XX} are partial derivatives of f. In an arbitrage-free world under the risk-neutral probability measure Q^* (that could be equal to the true measure Q for risk-averse agents) the expected return $a_f(\cdot)$ is equal to the short rate $r = g(X)$. The following *Cauchy problem* summarizes these steps:

$$\begin{aligned} a_f(X_t, N) &= R(X) \\ f(X, 0) &= 1 \end{aligned} \tag{4}$$

for all states of the world. Bond prices can now be computed in different ways. The conditional expected value in eq. (2) can be computed using Monte-Carlo methods. PDE in (4) can be solved numerically. For small dimensional systems ($N \leq 3$), solving the PDE is precise and relatively fast. For larger dimensional systems ($N > 3$), Monte-Carlo methods tend to be more attractive. The alternative is to make strong functional form assumptions on the coefficients a and b and the short-rate function $R(X)$ so that the PDE has a closed form solution.

3 General discussion and statement of the estimation problem

3.1 On the continuity of paths

Although path continuity is compelling in many physical applications, it is hard to make the same argument in finance. Research dealing with the microstructure of market often examines data on a transaction by transaction basic. Data at this frequency are becoming increasingly available and force the researcher to take seriously the path properties of asset prices. For example, on organized exchanges, asset prices are discrete and there are often institutional regularities that govern the set of possible transitions. In fact, since the bid-ask spread is large relative to the changes in observed transactions

prices, one must recognise when working at such high frequencies that there are really two prices - one if you are buying and another if you are selling. It may even be necessary to think of these prices as points on schedules with values depending upon the size and timing of the order flow. It is unsettling that diffusion models of asset prices should look so inappropriate for describing extremely high frequency data, since this is exactly the setting where they should be most relevant. Diffusion models can describe these data, but only if they are combined with some fairly complicated measurement scheme.

In many cases, the data are available at some fixed frequency such as daily, weekly, monthly, etc. Ironically, as we go to these lower frequencies, the "micro" details about path properties seem much less important,a and a diffusion model for asset prices (with point sampling) appears less questionable. Indeed, given even daily data one might well ask how can continuity of sample paths have any empirical content?

Path continuity by itself is not restrictive is we simply wish to describe the distribution of an asset's price measured at some fixed frequency , such as daily. However, it is important if we impose restrictions suggested by economic theroy on the joint distribution of asset prices and other variable. Economic theory gives empirical content to sample path properties, because the links across asset prices and other variables are sometimes sensitive to path properties. Consider, for example, the case of an option written on an asset. If the asset's price follows a diffusion process, then the options payoff can be replicated by a dynamically weighted portfolio of the asset and "cash". This is the insight behind the Black-Scholes formula linking the price of the asset to options written on it. If the asset's price follows a diffusion process with jumps, then this replication is no longer possible. The price of the option relative to the underlying asset will be sensitive to the decomposition of the asset's price into its continuous and jump components.

This is why the problem of the estimation of continuous-time model using discrete data arise: find consistent estimators of the underlying continuous-time model using discrete (and so "jumping") data.

3.2 The continuous-time model

The following model arises often in finance. We are given a complete filtered probability space $(\Omega, \mathcal{F}, \mathbf{F}, P)$. On this space is defined an adapted process X of the form (to avoid cumbersome notation we consider only an univariate Itô process even if the appendix is developed in a multivariate framework):

$$X_t(\omega) = X_{t_0}(\omega) + \int_{t_0}^{t} a(X, \tau; \alpha) d\tau + \int_{t_0}^{t} b(X, \tau; \beta) dW(\tau) \tag{5}$$

or the equivalent and perhaps more familiar representation given by the following stochastic differential equation:

$$dX(t) = a(X, \tau; \alpha)dt + b(X, \tau; \beta)dW(t) \tag{6}$$

where W_t denotes a standard wiener process and $a(X, \tau; \alpha)$ is called the drift coefficient and $b(X, \tau; \beta)$ is called the diffusion coefficient. The different specifications of the model derive if the researcher take into account some feature as the dependency on the time of $a(\cdot)$ and $b(\cdot)$ (if not the model is time homogeneous otherwise time-inhomogeneous). Moreover the need to work with fat tails for the distribution of X can be considered specifying an appropriate state-dependence for the diffusion $b(X)$.

However, in a general framework these coefficients are known functions which depend upon (X, t) and an unknown parameter vector $\theta = [\alpha^{'} \beta^{'}]^{'}$. The integrals must be interpreted in the sense of Itô if the functions $a(\cdot)$, $b(\cdot)$ satisfy the following restrictions:

A1. Let $\mathbf{F}$ the right-continuous filtration σ-field defined on $(\Omega, \mathcal{F}, P)$ and let the pure Wiener process $\{W(t) : t \in T\}$ be adapted to this filtration.

A2. If $\mathcal{B}(\mathbb{R}_+))$ is the σ-field of borel sets on $\mathbb{R}_+$, then for all θ in the parameter space $\Theta \equiv A \times B$ the functions a and b are measurable in the product σ-field $\mathcal{B} \times \mathbf{F}$

A3. For all $\theta \in \Theta$ the functions a and b satisfy the following inequalities almost surely

$$\int_{t_0}^{t} |a| d\tau < \infty \quad \int_{t_0}^{t} |b|^2 d\tau < \infty \tag{7}$$

In order to ensure the existence and uniqueness of a solution to the stochastic integral [or differential] equation given by eq.(5) [(6)] we require:

A5. There exists some constant $K > 0$ such that the functions a,b satisfy the following conditions for all X, $X^{'} \in \mathbb{R}$ and t, $t^{'}$:

$$\begin{aligned}
&|a(X,t) - a(X^{'}, t^{'})| + |b(X,t) - b(X^{'}, t^{'})| \leq K|X - X^{'}| \\
&|a(X,t) - a(X^{'}, t^{'})| + |b(X,t) - b(X^{'}, t^{'})| \leq K|t - t^{'}| \\
&a^2(X,t) + b^2(X,t) \leq K^2(1 + X^2)
\end{aligned}$$

3.3 A characterization of the likelihood function

Suppose that the process $X(t)$ is sampled at $n+1$ discrete points in time $t_0, t_1, \ldots, t_n$ *not* necessarily equally spaced apart and let $X \equiv (X_0, X_1, \ldots, X_n)$ denote this random sample where $X_k \equiv X(t_k)$. Given the discretely sampled data X and the stochastic specification of the process $X(t)$, denote by the $P(X_0, X_1, \ldots, X_n; \theta)$ the finite-dimensional distribution of X and let $\rho(X; \theta)$ denote the density representation of P. When considered a function of θ, this joint density is obviously the desired exact likelihood function of X. Since $X(t)$ is a markov process (see Arnold (1971) chapter 9) the joint density ρ may be rewritten as the following product of conditional densities:

$$\rho(X) = \rho_0(X_0) \prod_{k=1}^{n} \rho_k(X_k, t_k | X_{k-1}, t_{k-1}) \tag{8}$$

Deriving the likelihood function then reduced to calculating density function ρ_k. Following the result of Lo (1988) the density ρ_k is the solution of the corresponding Fokker-Plank equation:

$$\frac{\partial}{\partial t}\rho_k = -\frac{\partial}{\partial X} a \rho_k + \frac{1}{2}\frac{\partial^2}{\partial X^2}(b^2 \rho_k) \tag{9}$$

The result is particular relevant because highlight the necessity to impose some restriction to the coefficient function a, b in order to solve to functional partial differential equation. However when the existence of a density representation for a specific process has been assured by other means the solution of the p.d.e is standard even if cumbersome. In fact sometimes the solution is often obtained, following Lo (1988) terminology, by "educated guess".

3.4 Maximum likelihood Estimation Via Discretized Itô Processes

Having characterized the likelihood function as the solution of the functional partial differential equation cited above and assuming its existence, define the maximum likelihood estimator in the usual manner

$$\theta_{ML} \equiv \ \text{argMax}_\theta G(\theta; X)$$

where

$$G(\theta; X) \equiv \ln \rho_0(X_0, t_0) + \sum_{k=1}^{n} \ln \rho_k(X_k, t_k | X_{k-1}, t_{k-1}; \theta) \tag{10}$$

Since θ_{ML} is the true maximum likelihood estimator, it possesses the standard properties of consistency and asymptotic normality under appropriate regularity conditions.

In the event that the functional partial differential equation cannot be solved explicitly to obtain the likelihood function of the sample X, several authors have estimated θ by applying maximum likelihood to a suitably discretized standard differential equation.

Specifically, equally spaced discretely sampled data are assumed to be generated by the following difference equation.

$$X_{k+1} = X_k + a(X_k, t_K; \alpha) + b(X_k, t_k; \beta)\Delta W(t_{k+1} \tag{11}$$

where

$$\Delta W(t_{k+1}) \equiv W(t_{k+1} - W(t_k)) \quad t_k \equiv kh$$

for $k = 0, 1, \ldots, n$ and $h \equiv T/n$. The parameters $\theta \equiv [\alpha'\beta']$ are then estimated via maximum likelihood using eq. (11). Because it is well-known that the sample paths of the discretization (11) converge to those of the continuous-time Itô $X(t)$ as h approaches 0 (see, for example Kloeden and Platen (1992)), such an estimation procedure may seem useful. However as reported by Lo (1988) the "discretized maximum likelihood" estimator need not be consistent.

For example if consider $X(t)$ a lognormal diffusion on the interval [0,T]:

$$dX(t) = \alpha X(t)dt + \beta X(t)dW(t) \tag{12}$$

and consider its discretization form:

$$X_{k+1} = \alpha X_k h + \beta X_k \Delta W_{k+1} \equiv \alpha X_k h + X_k \epsilon_{k+1} \tag{13}$$

where ϵ_{k+1} is an i.i.d $N(0, \beta^2 h)$ random variable. From (13), it is apparent that the discretized maximum likelihood estimators of α and β^2 are given by:

$$\hat{\alpha_D} = \frac{1}{T}\sum_{k=1}^{n}\left[\frac{X_k}{X_{k-1}} - 1\right] \quad \hat{\beta_D^2} = \frac{1}{T}\sum_{k=1}^{n}\left[\frac{X_k}{X_{k-1}} - 1 - \hat{\alpha_D}h\right]^2 \tag{14}$$

But for fixed observation intervals h, as the number of observation increases, the discretized estimators do not converge to the population parameters of interest:

$$\hat{\alpha}_D \xrightarrow{p} \frac{1}{h}[e^{\alpha h} - 1] \neq \alpha \qquad \hat{\beta}_D^2 \xrightarrow{p} \frac{1}{h} e^{2\alpha h}[e^{\beta^2 h} - 1] \neq \beta^2 \tag{15}$$

It is evident that for small h the asymptotic bias may be negligible. Of course, whether or not the bias is economically meaningful is an empirical question which is process-specific and must be resolved for each application individually. However, it should be clear that for arbitrary coefficient functions a, b and c the discretized ML estimator is generally inconsistent. Since eq. (15) indicates that the asymptotic is decreasing in the observation interval h, it might be conjectured that consistency may be restored if we draw observations more frequently within the fixed time span [0,T], thereby letting h approach zero as n increases without bound so as to keep $T \equiv nh$ fixed. In fact, it may be shown that such a limiting operation, (the so-called "continuous data recording"), does guarantee that $\hat{\beta}_D$ converges to β in probability. However, the same cannot be said for the drift estimator $\hat{\alpha}_D$. In particular, using functional central limit theory techniques, we conclude that:

$$\hat{\alpha} \xrightarrow{p} \alpha + \frac{\beta}{\sqrt{T}}[W(1) - W(0)]$$

Thus, with more frequent sampling within a fixed time span, $\hat{\alpha}$ converges weakly to a Gaussian random variable with mean α and variance β^2/T.

Of course, the inconsistency of the discretized maximum likelihood estimators does not imply that there exists no consistent estimator of the parameters (see Lo (1988)) but it means that the maximum likelihood of the discretized process is inappropriate when it is the parameters of the continuous-time process that are of interest.

That this is so should not be surprising since the discretization may be viewed as a misspecification of the true maximum likelihood function.

4 Estimation methods of continuous-time stochastic process for term structure interest rate

As discussed above, most of the literature about term structure estimation is evolved in a *parametric* framework in order to obtain a "tractable" functional form of the stochastic differential equation. In this setting various choices have to be made regarding measurement errors and estimation methods.

From an econometric point of view we can see the model as a *state space systems*:

$$
\begin{aligned}
y_t^{(N)} &= g(X_t) + \epsilon_t^{(N)} \\
dX(t) &= a(X)dt + b(X)dz_t
\end{aligned} \tag{16}
$$

with an observation equation which links observable yields $y_t^{(N)}$ to the state vector and a state equation $dX(t)$ which describes the dynamics of the state.

The reason to add a measurement error depends on the so-called stochastic singularity (see Piazzesi (2003)).

Consider for simplicity that in the market we can observe zero-coupon bond price for each maturity or analogous the corresponding yields [4]. Using eq. (1) and substituting P_t^N with the equation of the state variable we obtain the equation of the yield as a function of the state variable.

Parametric models rely on low-dimensional state vector to describe what drives the yield curve. Data on K different yields can therefore be used to back out K state variables. The K yields $y^{N_1}, y^{N_2}, \ldots, y^{N_K}$ can be used to invert eq. (1) for $N_1, N_2, \ldots, N_K$ to obtain the model-implied state vector X. This means that any additional yield is predicted by the model with an R^2 of 1. The model can therefore be rejected with a single observation on $y^{(N+1)}$. Put differently, variance-covariance matrix of $N+1$ yields in the model is singular, (stochastic singularity).

Stochastic singularity is a problem, because we have lots of cross-sectional yield data (many different N's) and want to use models with few state variables. Adding measurement error $\epsilon^{(N)}$ to the yield equation, as done in eq. (16), breaks this singularity. Now different assumptions can be made on the properties of these measurement errors. Either all of the yields are observed with error or only a subset of yields are observed with error.

The assumption that all yields are observed with error seems plausible. Data-entry mistakes and interpolation methods for constructing zero-coupon yields are among the obvious sources for such errors. When all yields have errors, we cannot invert the yield coefficients in eq.(16) to compute the state vector. Kalman filtering is useful here, especially when the state vector is normally distributed.

This field of financial econometrics is one of the more active and different solution have been proposed in order to estimate eq. (16). In what follow i give a brief explanation of some estimation techniques that ideally can be grouped in two different classes:

- likelihood based methods

[4] To be precise, usually,for long maturity we observe prices of coupon bonds. Nevertheless, we can reconstruct the zero-coupon price at each maturity, using the "bootstrapping" technique, under the assumption that they are basket of zero-coupon bonds

- GMM methods (*Generalized Method of Moments*)
- nonparametric methods

4.1 Likelihood based methods

Maximizing the likelihood function relies on being able to compute $f(x_{t+1}|x_t)$ of the state vector x_{t+1} given x_t. The conditional density of an N-dimensional vector of observed yields Y can be obtained by a change of variable. The density of Y is the product of the conditional density of x and the determinant of the Jacobian:

$$f(y_{t+1}|y_t) = f(x_{t+1}|x_t) \left| \frac{\partial x_{t+1}}{\partial y_{t+1}} \right|$$

The log-likelihood function of observed yields is then constructed as the usual sum of log density $logf(y_{t+1}|y_t)$ over the sample. To maximize the log-likelihood, the state x_{t+1} is backed out from y_{t+1} for any given parameter vector.

For Gaussian processes, f is multivariate normal. Zero-coupon yields that are "affine[5]" in x are also Gaussian. Their likelihood function is therefore particularly easy to compute.

For independent square-root processes, f is the product of noncentral chi-square densities. The formula for the densities is based on the modified Bessel function of the first kind of order q (see Cox, Ingersoll, and Ross (1985) p. 391-392).

In this approach the main problem is that for general specification is not possible to know the closed form of $f(x_{t+1}|x_t)$. As said before, the strategy has been the discretization of SDE. Roughly speaking, the estimation problem of a continuous time stochastic process is transformed into the estimation of a difference equation:

$$\Delta x_{t+h} = \mu(x_t)h + \sigma(x_t)\varepsilon_{t+h}\sqrt{h} \tag{17}$$

where ε_{t+h} has a standard Gaussian distribution (if the discretization follows the Euler scheme Kloeden and Platen (1992)) and h is the length of the time interval. The conditional density of the discretized process respect to x_t, is Gaussian with mean $\mu(x_t)h$ and conditional variance $\sigma(x_t)\sigma(x_t)^\top$. It approximates the conditional density of SDE when h shrink to 0. However, as

[5]A function $F : \mathbb{R}^n \to \mathbb{R}$ is *affine* if there exists some coefficients $a \in \mathbb{R}$ and $b \in \mathbb{R}^n$ such that $F(x) = a + b^\top x \quad \forall x \in \mathbb{R}^n$

we have seen before Lo (1988) shows that the maximum likelihood estimator of discretized process, is not consistent for each h (*discretization bias*).

Other solutions based on the likelihood are the Fourier inversion of the characteristic function. Duffie, Pan, and Singleton (2000) show that the characteristic function can be computed in close form for affine diffusions. Singleton (2001) estimates an affine process with one state variable maximizing the log-likelihood function obtained by Fourier inversion. Problem arise at higher dimensional state space. The computation of the conditional density becomes costly.

Pedersen (1995) and Santa-Clara (1995) propose to simulate the likelihood function (SML methods) which is unknown. The approach is based on simulating (17) whose conditional density is Gaussian.

The idea is to write the density of x_{t+1} conditional on the last observation x_t using Bayes' Rule and the Markov property of x:

$$f(x_{t+1}|x_t) = \int_D f(x_{t+1}|x_{t+1-h}) f(x_{t+1-h}|x_t) dx_{t+1-h}$$

for any time interval h. The density $f(x_{t+1}|x_t)$ is now approximated with the density $\hat{f}$ of the discretized process (17).

This density is normal with mean $x_{t+1-h} + \mu(x_{t+1-h})h$ and standard deviation $\sigma(x_{t+1-h})\sqrt{(}h)$. The integral above can then be computed using Monte Carlo:

$$f(x_{t+1}|x_t) \approx \frac{1}{S} \sum_{s=1}^{S} \hat{f}(x_{t+1}|\hat{x}^{x_t}_{t+1-h}[s]))$$

where the summation is over a total of S simulated paths of the state that start at the last observation x_t at time t. The computer only needs to store the terminal simulated value $\hat{x}^{x_t}_{t+1-h}[s]$ for each simulation s, not the entire simulated path. Standard variance reduction techniques, such as antithetic sampling, can be used to improve the efficiency of Monte Carlo integration. Brandt and Santa-Clara (2002) use this simulated maximum likelihood method to estimate a multi-factor diffusion model.

4.2 GMM methods

Within GMM methods, the *Efficient Method of Moments* (EMM) developed in Gallant and Tauchen (1996) is particular useful when the continuous time process is readily simulated but the likelihood function of the structural model is intractable. Roughly speaking, Gallant and Tauchen (1996) have

developed a systematic strategy for choosing the moments for GMM estimation of a structural model. The idea is relatively straightforward: it is based on the use of the expectation with respect to the structural model of the score function of an auxiliary model as the vector of moment conditions for GMM estimation. The score function is the derivative of the logarithm of the density of the auxiliary model with respect to the parameters of the auxiliary model. The moment conditions obtained by taking the expectation of the score depend directly upon the parameters of the auxiliary model and indirectly upon the parameters of the structural model through the dependence of expectation operator on the parameters of the structural model. The parameters of the auxiliary model are eliminated from the moment conditions by replacing them with their quasi-maximum likelihood estimates, which are obtained by maximizing the likelihood of the auxiliary model. This leaves a random vector of moment conditions that depends only on the parameters of the structural model; the randomness is due to the random fluctuations of the quasi-maximum likelihood estimates of the parameters of the auxiliary model. When this vector of moment conditions is evaluated at the true values of the structural parameters, it tends to zero as sample size increases, presuming that the structural model is correctly specified. The parameters of the structural model can be estimated by minimizing the magnitude of the vector of moment conditions as measured by appropriate GMM metric. It is obvious that the problems arise from the choice of a adequate auxiliary model and from a good selection of the parameters of structural models when it is simulated for expected value calculation.

4.3 Nonparametric methods

One potentially serious problem with any parametric model, particularly when there is no economic reason why we should prefer one functional form over another, is misspecification. The reason is that even if a model fits interest rate movements well in-sample, this does not necessarily imply that it will price securities well. This is because the price today of an interest rate dependent security depends not on the past interest rates, but on the entire distribution of possible future interest rates between today and the maturity of the security. Fitting historical data well is no guarantee of matching this entire distribution, leading to the possibility of large pricing and hedging errors (see Backus, Foresi, and Zin (1995)). To avoid misspecification, resent research has used nonparametric estimation techniques in order to avoid having to specify (arbitrary) functional forms for a and/or b. Within the nonparametric framework the approaches are different.

4.3.1 The Aït-Sahalia's approach

Aït-Sahalia (1996) use the equivalence between (a, b) and densities but works in the other direction, from the densities back to (a, b). The idea is that, the drift and diffusion are not observed and cannot be estimated directly. However, the densities of the process can be straightforwardly estimated from the data on the short-term rate (that is the state variable in an univariate framework). Then it is possible to reconstruct the drift and diffusion by matching these density.

In a univariate model, Aït-Sahalia start with the following identification: the drift $a(\cdot;\theta)$ depends on an unknown parameter vector θ, while the diffusion $b^2(\cdot)$ is an unknown function.

Let $\pi(\cdot)$ be the marginal density of the state variable, and $p(\Delta, X_{t+\Delta}|X_t)$ the transition density function between two successive observations. The latter density use only the information at time t because the assumption is that the process is Markovian.

Consider the Kolmogorov forward equation:

$$\begin{aligned} \frac{\partial p(\Delta, X_{t+\Delta}|X_t)}{\partial \Delta} &= -\frac{\partial}{\partial X_{t+\Delta}}(a(X_{t+\Delta,}\theta)p(\Delta, X_{t+\Delta}|X_t)) \\ &+ \frac{1}{2}\frac{\partial^2}{\partial X_{t+\Delta}^2}(b^2(X_{t+\Delta})p(\Delta, X_{t+\Delta}|X_t)) \end{aligned} \tag{18}$$

It depends, in some way, on the drift, diffusion's functions and the transition probability density.

To construct an estimator of b^2 from the densities of the process we use eq. (18) to characterize the diffusion function but we need to rule out the dependency on the transition probability that depend also by Δ[6].

In order to do this, we note that by stationarity, $\int_0^{+\infty} p(\Delta, X_{t+\Delta}|X_t)\pi(X_t)dX_t = \pi(X_{t+\Delta})$ has a partial derivatives with respect to time equal to zero. Therefore, we can use this result multiplying eq. (18) by $\pi(X_t)$ and integrating the resulting equation with respect to the conditioning variable X_t:

$$\begin{aligned} \frac{\partial p(\Delta, X_{t+\delta}|X_t)}{\partial \Delta}\pi(X_t) &= -\frac{\partial}{\partial X_{t+\Delta}}(a(X_{t+\Delta,}\theta)p(\Delta, X_{t+\delta}|X_t))\pi(X_t) \\ &+ \frac{1}{2}\frac{\partial^2}{\partial X_{t+\Delta}^2}(b^2(X_{t+\Delta})p(\Delta, X_{t+\delta}|X_t))\pi(X_t) \end{aligned} \tag{19}$$

[6]This is the step that permit to achieve asymptotic results without imposing that the sampling interval shrink to zero.

$$\frac{\partial}{\partial \Delta}\int_0^{+\infty} p(\Delta, X_{t+\Delta}|X_t)\pi(X_t)dX_t =$$
$$= -\frac{\partial}{\partial X_{t+\Delta}}\left[\int_0^{+\infty} a(X_{t+\Delta;}\theta)p(\Delta, X_{t+\Delta}|X_t)\pi(X_t)dX_t\right] +$$
$$+\frac{1}{2}\frac{\partial^2}{\partial X_{t+\Delta}^2}\left[\int_0^{+\infty} b^2(X_{t+\Delta})p(\Delta, X_{t+\Delta}|X_t)\pi(X_t)dX_t\right]$$

Simple algebra give the final result:

$$\frac{\partial^2}{\partial X_{t+\Delta}^2}\big(b^2(X_{t+\Delta})\pi(X_{t+\Delta})\big) = 2\frac{\partial}{\partial X_{t+\Delta}}a(X_{t+\Delta};\theta)\pi(X_{t+\Delta}) \tag{20}$$

This equation must be satisfied at any point in the domain and at the true parameter value θ.

To find and explicit expression for b^2 we have to integrate eq. (20) two times with the boundary condition $\pi(0) = 0$ now yields:

$$b^2(X) = \frac{2}{\pi(X)}\int_0^r a(u;\theta)\pi(u)du \tag{21}$$

This equation shows that once the drift parameter vector θ has been identified, the diffusion function can be identified from the marginal distribution $\pi(\cdot)$.

In a parametric framework, for identification of θ is necessary to impose a functional form to the drift[7]. A possible specification is the linear mean–reverting form: $a(X_t;\theta) \equiv \beta(\alpha - X_t)$, $\theta \equiv (\alpha, \beta)'$, that is consistent with the parameterization of the drift used in most spot rate models in the literature. Heuristically, the interest rate is elastically attracted to its equilibrium value α at a speed βdt. Once indentified, the drift can be plugged into eq.(21), that is valid for a generic drift function: the equation requires to estimate parameter θ and marginal density $\pi(u)$.

In order to estimate θ, the first step is to derive the expression of the first conditional moment of the process X_t. The reason is related to the property of diffusion processes: their transition probability is, under certain regularity assumptions, uniquely determined by the drift and the diffusion

[7]This is the reason why Aït-Sahalia (1996)'s approach is a semiparametric approach: the drift is identified parametrically while the diffusion is estimated with nonparametric methods

coefficients and so, by definition of diffusion process, by the first two moments (see Arnold (1971))[8]

Now, using standard argument of stochastic calculus like, the infinitesimal operator of the process X_t and Dynkin formula (see Oksendal (1995)), to obtain:

$$\mathrm{E}[X_{t+\Delta}|X_t] = \alpha + e^{-\beta\Delta}(X_t - \alpha) \tag{22}$$

The drift's parameters can be estimate using eq. (22) and OLS technique with $\mathrm{E}[r_{t+\Delta}|r_t] = \lambda + \delta r_t$ and the following one–to–one transformation $\alpha = -\lambda/\delta$ and $\beta = -\ln 1 + \delta/\Delta$. The OLS estimator can be plugged in the eq. (21) and together with a nonparametric estimation of the marginal density is possible to estimate the diffusion function.

5 Conclusions

Continuous-time models have a large range of applications. They have been used for a long time to model phenomena evolving randomly and continuously in time, e.g. in physics and biology. During the last thirty years or so the models have also been applied intensively in mathematical finance for describing stock prices, exchange rates, interest rates, etc. (although it is wellknown that such quantities do not really change continuously in time). Data are essentially always recorded at discrete points in time only (e.g. weekly, daily or each minute) and can thus be interpreted as time series data. Still, continuoustime models are often preferred to classical time series models. There are (at least) two reasons for this. First, if data are sampled at irregularly spaced timepoints, then an appropriate discretetime model should incorporate this explicitly. As opposed to this, continuous-time models implicitly define transitions over time intervals of any length in a consistent way. For example, missing data in a sample where timepoints for observations are otherwise regularly spaced, do not give rise to serious problems in the continuoustime setting as they are treated just like the values not observed due to discretetime sampling. Second, all the machinery from stochastic calculus is at our disposal when we use continuous-time models. This has proved important in finance theory where derivation of various price formulas usually relies heavily on this theory. Thus convinced that continuous models are important and useful alternatives to classical time series models I turn to the statistical analysis. For a few models, estimation

[8]This means that the first two moments are sufficient to know the entire transition probability and this is not true in general

is straightforward because the corresponding stochastic differential equation can be solved explicitly. This is the case for the geometric Brownian motion, the OrnsteinUhlenbeck process and the squareroot process which have lognormal, normal and noncentral chisquare transition probabilities respectively. However, "nature" (or "the market") most often generates data not adequately described by such simple models. For example, empirical studies clearly reveal that increments of logarithmic stock prices are not independent and Gaussian as implied by the geometric Brownian motion classically used for stock price modelling. Rather, they exhibit temporal dependence and leptokurtosis. Consequently, more complex models are needed in order to obtain reasonable agreement with data. This complicates the statistical analysis considerably because the discretetime transitions (implicitly defined by the model) are no longer known analytically. Specifically, the likelihood function is usually not tractable. In other words, one has to use models for which likelihood analysis is not possible, and there is consequently a need for alternative.

Appendix

The purpose of this appendix is to collect some of the background concepts and definitions. There exist a variety of textbook treatments of stochastic processes and this appendix is drawn from Oksendal (1995), Karlin and Taylor (1981) and Karatzas and Shreve (1991).

Let $(\Omega, \mathcal{F}, P)$ denote a probability space. For our purposes, a stochastic process is a family $X = (X_t; t \in \mathbb{R}_+)$ of $\mathcal{F}/\mathcal{B}(\mathbb{R}^n)$-measurable mappings from Ω into $\mathbb{R}^n$ where $\mathcal{B}(\mathbb{R}^n)$ denotes the Borel σ-field on $\mathbb{R}^n$. The collection of the mappings $(X_t(\omega); t \in \mathbb{R}_+)$ for fixed $\omega \in \Omega$ is called a path (sample path, trajectory, or realization) of the process.

Sometimes it is natural to view the a process as a mapping from $\Omega \times \mathbb{R}_+ \to \mathbb{R}^m$. This can be done by strengthening slightly the measurability requirements. Let $\mathcal{B}([0,\infty))$ denote the Borel σ-field on $\mathbb{R}_+$ and $\mathcal{F} \otimes \mathcal{B}([0,\infty))$ the product σ-fiels. Sets in $\mathcal{F} \otimes \mathcal{B}([0,\infty))$ are called random sets. A process X is said measurable (jointly measurable) if it is $\mathcal{F} \otimes \mathcal{B}([0,\infty))/\mathcal{B}(\mathbb{R}^m)$-measurable. If X is measurable, then its sample paths are a Borel-measurable function of $t \in [0,\infty)$

The notion that the information is revealed by the passage of time is captured by the concept of a filtration. A filtration $\mathbf{F} = (\mathcal{F}_t; t \in \mathbb{R}_+)$ is an increasing family of σ-algebra (i.e $\mathcal{F}_s \subseteq \mathcal{F}_t$ for $s \leq t$). $\mathcal{F}_\infty \equiv \mathcal{F}$ or $\mathcal{F}_\infty \equiv \sigma(\mathcal{F}_t; t \in \mathbb{R}_+)$.

A process is said to be adapted (or non-anticipative) (to the filtration $\mathbf{F}$) if X_t is $\mathcal{F}_t$-measurable, for every t.

Let $\mathcal{F}_t^X \equiv \sigma[X_s; 0 \leq s \leq t]$, i.e. the σ-algebra generated by the process up to time t. Heuristically, $\mathcal{F}_t^X$ contains all the events based upon observing the history of the process X up to and including time t. $\mathbf{F}^X \equiv (\mathcal{F}_t^X; t \in \mathbb{R}_+)$ provides an example of a filtration (in this case the so-called *natural filtration*) and X is adapted to $\mathbf{F}^X$ by construction.

A probability space endowed with a filtration, denoted $(\Omega, \mathcal{F}, \mathbf{F}, P)$ is called a filtered probability space.

For technical reasons (essentially concerned with measurability properties), further restrictions on $\mathbf{F}$ are often imposed. The filtration $\mathbf{F}$ is said to be right-continuous if $\mathcal{F}_t = \cap_{s>t}\mathcal{F}_s$. Associated with each filtration $\mathbf{F}$ is the P-augmented filtration $\mathbf{F}^P = (\mathcal{F}_t^P; t \in \mathbb{R}_+)$ where $\mathcal{F}^P$ is the P-completion of $\mathcal{F}$ and $\mathcal{F}_t^P$ is the σ-algebra generated by $\mathcal{F}_t$ and the null sets of $\mathcal{F}^P$. A filtration that is right-continuous and augmented is said to satisfy the *usual conditions*. The filtered probability space $(\Omega, \mathcal{F}^P, \mathbf{F}^P, P)$ is said to be complete.

References

Aït-Sahalia, Y. (1996): "Nonparametric Pricing of Interest Rate Derivatives Securities," *Econometrica*, 64, 527–560.

Ang, A., M. Piazzesi, and M. Wei (2002): "What does the yield curve tell us about GDP growth," Working paper, UCLA.

Arnold, L. (1971): *Stochastic Differential Equations: Theory and Applications*. Wiley, New York.

Backus, D. K., S. Foresi, and S. E. Zin (1995): "Arbitrage opportunities in arbitrage free models of bond pricing," Working paper, New York University.

Black, F., and M. Scholes (1973): "The Pricing of Options and Corporate Liabilities," *Journal Political Economics*, 81, 384–404.

Brandt, M. W., and P. Santa-Clara (2002): "Simulated likelihood estimation of diffusions with an application to exchange rates dynamics in incomplete markets," *Journal of Financial Economics*, 63, 161–210.

Campbell, J., and R. Shiller (1991): "Yield spreads and interest rates: A bird's eye view," *Review of Economic Studies*, 58, 495–514.

Cochrane, J. (2001): *Asset Pricing*. Princeton University Press.

Cox, J., J. Ingersoll, and S. Ross (1985): "A Theory of the Term Structure of Interest Rates," *Econometrica*, 53, 385–408.

Duffie, D. (2001): *Dynamic Asset Pricing Theory*. Princeton University Press, Princeton.

Duffie, D., J. Pan, and K. Singleton (2000): "Transform analysis and asset pricing for affine-jump diffusion," *Econometrica*, 68, 1343–76.

Fama, E. (1990): "Term-structure forecasts of interest rates, inflation, and real returns," *Journal of Monetary Economics*, 25, 59–76.

Gallant, A., and G. Tauchen (1996): "Which moments to match?," *Econometric Theory*, 12, 657–681.

Karatzas, I., and S. Shreve (1991): *Brownian Motion and Stochastic Calculus*. Springer Verlag, New York.

Karlin, S., and H. Taylor (1981): *A Second Course in Stochastic Processes*. Academic Press, New York.

Kloeden, P., and E. Platen (1992): *Numerical Solutions to Stochastic Differential Equations*. Springer, New York.

Lo, A. W. (1988): "Maximum likelihood estimation of Generalized Ito Processes with discretely-sampled data," *Econometric Theory*, 4, 231–247.

Merton, R. (1969): "Lifetime Portfolio Selection under Ucertainty: The Continuous Time Case," *Review of Economics and Statistics*, 51, 247–257.

——— (1973): "The Theory of Rational Option Pricing," *Bell Journal of Economics and Management Science*, 4, 141–183.

Oksendal, B. (1995): *Stochastic Differential Equations*. Springer, Berlin.

Pedersen, A. (1995): "A New Approach to Maximum Likelihood Estimation for Stochastic Differential Equations based on Discrete Observations," *Scandinavian Journal of Statistics*, 22, 55–71.

Piazzesi, M. (2003): "Affine term structure models," Working paper, UCLA and NBER.

Santa-Clara, P. (1995): "Simulated Likelihood Estimation of Diffusions with an application to the Short Term Interest Rate," Ph.D. thesis, Insead, France.

Singleton, K. (2001): "Estimation of Affine Asset Pricing Models using the Empirical Characteristic Function," *Journal of Econometrics*, 102, 111–41.