Likelihood Methods for Continuous-Time Models in Finance

Yacine Aït-Sahalia

Princeton University and NBER

July 7, 2011

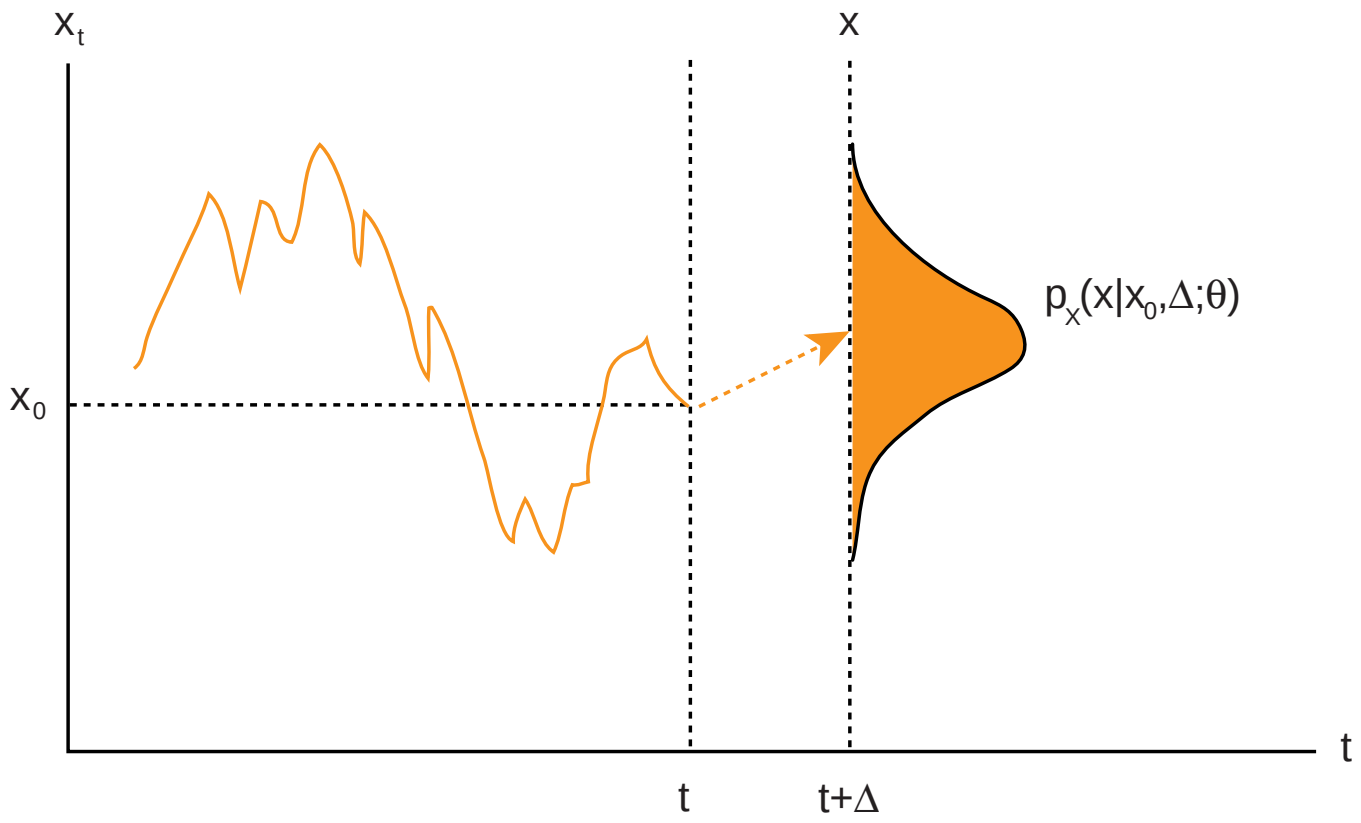
1. Introduction

- Thanks to Merton's seminal work in the 1970s, continuous-time modelling has become one of the main tools of asset pricing.

- Diffusions and more generally continuous-time Markov processes are generally specified in economics and finance by their evolution over infinitesimal instants

$$dX_t = \mu\left(X_t; \theta\right) dt + \sigma\left(X_t; \theta\right) dW_t$$

- The transition function of the Markov process X is the conditional density $p_X\left(\Delta, x | x_0; \theta\right)$ for the values of the state variable x at a fixed future time Δ, given the current level x_0 of the state vector.


x_t
x
$p_X(x|x_0,\Delta;\theta)$
x_0
t
t
$t+\Delta$

1.1. Maximum-Likelihood Estimation

- Suppose that we observe the process at discrete dates $t = i\Delta$ for $i = 0, \ldots, n$.

- Random times can be accommodated as long as they do not cause the X process.

- The form of the likelihood function is particularly simple due to the Markovian nature of the model.

- Using Bayes' Rule and the Markov property, the probability of observing the data $\{X_0, X_\Delta, \ldots, X_{n\Delta}\}$ given that the parameter vector is θ is

$$\begin{aligned}\mathbb{P}\left(X_{n\Delta}, X_{(n-1)\Delta}, \ldots, X_0; \theta\right) &= \mathbb{P}\left(X_{n\Delta} | X_{(n-1)\Delta}, \ldots, X_0; \theta\right) \\ &\times \mathbb{P}\left(X_{(n-1)\Delta}, \ldots, X_0; \theta\right) \\ &= \mathbb{P}\left(X_{n\Delta} | X_{(n-1)\Delta}; \theta\right) \times \mathbb{P}\left(X_{(n-1)\Delta}, \ldots, X_0; \theta\right) \\ &= \mathbb{P}\left(X_{n\Delta} | X_{(n-1)\Delta}; \theta\right) \\ &\times \ldots \times \mathbb{P}\left(X_\Delta | X_0; \theta\right) \times \mathbb{P}\left(X_0; \theta\right).\end{aligned}$$

- Ignoring the initial observation, one would like to maximize over values of θ the log of the corresponding density

$$\ell_n(\theta) \equiv \sum_{i=1}^{n} \ln p_X\left(\Delta, X_{i\Delta} | X_{(i-1)\Delta}; \theta\right).$$

- So this would be rather easy, except that for most models of interest, the function p_X is not known in closed-form.

- Rare exceptions:
  - Geometric Brownian motion, Black and Scholes (1973): $dX_t = \beta X_t dt + \sigma X_t dW_t$
  - Ornstein-Uhlenbeck process, Vasicek (1977): $dX_t = \beta(\alpha - X_t)\,dt + \sigma dW_t$
  - Feller's square root process, Cox et al. (1985): $dX_t = \beta(\alpha - X_t)\,dt + \sigma X_t^{1/2} dW_t$.

- In many cases that are relevant in finance, however, the transition function p_X is unknown:
  - Courtadon (1982): $dX_t = \beta(\alpha - X_t)\,dt + \sigma X_t dW_t$
  - Marsh and Rosenfeld (1982): $dX_t = (\alpha X_t^{-(1-\delta)} + \beta)dt + \sigma X_t^{\delta/2} dW_t$
  - Cox (1975), Chan et al. (1992): $dX_t = \beta(\alpha - X_t)\,dt + \sigma X_t^{\gamma} dW_t$
  - Constantinides (1992): $dX_t = \left(\alpha_0 + \alpha_1 X_t + \alpha_2 X_t^2\right) dt + (\sigma_0 + \sigma_1 X_t)\,dW_t$
  - Duffie and Kan (1996), Dai and Singleton (2000), affine models: $dX_t = \beta(\alpha - X_t)\,dt + (\sigma_0 + \sigma_1 X_t)^{1/2}\,dW_t$
  - Aït-Sahalia (1996), nonlinear mean reversion: $dX_t = (\alpha_0 + \alpha_1 X_t + \alpha_2 X_t^2 + \alpha_{-1}/X_t)dt + (\beta_0 + \beta_1 X_t + \beta_2 X_t^{\beta_3})dW_t$.

- I will survey the development of closed form likelihood expansions for models such as these, following in Aït-Sahalia (1999) (examples and application to interest rate data), Aït-Sahalia (2002) (univariate theory) and Aït-Sahalia (2008) (multivariate theory).

- *Matlab* code is available upon request to compute these expansions for arbitrary models.

- For a long time, continuous-time models in finance have been predominently univariate

- Most models nowadays are multivariate
  - Asset pricing models with multiple explanatory factors
  - Term structure models with multiple yields or factors
  - Stochastic volatility or stochastic mean reversion models

- I will describe the construction of approximations to the transition density of arbitrary multivariate diffusions.

- Extensions:
  - To time-inhomogenous univariate diffusions: Egorov et al. (2003) and Choi (2008)
  - To univariate models driven by Lévy processes other than Brownian motion, including jumps: Schaumburg (2001) and Yu (2007)
  - To a Bayesian setting: DiPietro (2001), Stramer et al. (2009)
  - To damped diffusions: Li (2010)

- This method is applicable also to other estimation strategies which require an expression for p_X, such as:
  - Bayesian methods where one needs a posterior distribution for θ;
  - To generate simulated paths at the desired frequency from the continuous-time model;
  - Or to serve as the instrumental or auxiliary model in indirect inference and simulated/efficient methods of moments.

- Other methods require an expression for p_X : for instance, derivative pricing relies on expectations taken with respect to the risk-neutral density which is the p_X corresponding to the risk-neutral dynamics from the model.

- Computations using a closed-form p_X are straightforward and fast because everything is in closed form, compared to other methods, such as:
  - Solving numerically the FPK partial differential equation: Lo (1988)
  - Simulating the process to determine p_X by Monte Carlo integration: Pedersen (1995) and Brandt and Santa-Clara (2002)
  - Using binomial or other trees: Jensen and Poulsen (2002)
- The computation of the closed-form expression for p_X corresponding to a given model is only done once and for all.
- A *Matlab* library for existing models is now available, with models added as they are requested.

2. The Univariate Case

- The intuition is to transform the data into something that is amenable to an explicit correction around a leading term.

- As in the Central Limit Theorem, if we could get "close enough" to a Normal variable, then corrections are possible for the fact that the sample size is never quite infinity.

- Still not an Edgeworth expansion: we want convergence as the number of terms increase, not the sample size

- In general, p_X cannot be approximated for a fixed sampling interval Δ around a Normal density by standard series
  - because the distribution of X is too far from that of a Normal
  - for instance, if X follows a geometric Brownian motion, the right tail is too thick, and the Edgeworth expansion diverges

2.1. Main Idea

1. Make two successive transformations $X \mapsto Y \mapsto Z$ such that Z that is sufficiently close to a Normal

2. Construct a sequence for p_Z around a Normal

3. Revert the transformation $Z \mapsto Y \mapsto X$

$\Rightarrow$ Expansion for p_X around a deformed Normal.

2.2. First Transformation: $X \mapsto Y$

- Define

$$Y_t \equiv \gamma\left(X_t;\theta\right) = \int^{X_t} du/\sigma\left(u;\theta\right)$$

- By Itô's Lemma:

$$dY_t = \mu_Y\left(Y_t;\theta\right)dt + dW_t$$

where

$$\mu_Y\left(y;\theta\right) = \frac{\mu\left(\gamma^{-1}\left(y;\theta\right);\theta\right)}{\sigma\left(\gamma^{-1}\left(y;\theta\right);\theta\right)} - \frac{1}{2}\frac{\partial\sigma}{\partial x}\left(\gamma^{-1}\left(y;\theta\right);\theta\right)$$

- Despite its local unit variance, Y is still not close enough to a Normal...

2.3. Second Transformation: $Y \mapsto Z$

- Z is a "standardized" version of Y:

$$Z_t \equiv \Delta^{-1/2} (Y_t - y_0)$$

- Since $\Delta \to 0$ is not required:
  - No claim regarding the degree of accuracy of this standardization device
  - Rather, it will turn out that this Z_t is "close enough" to a Normal for this to work.

2.4. The Approximation Sequence for p_Z

- Hermite polynomials:

$$H_j(z) \equiv e^{z^2/2} \frac{d^j}{dz^j} \left[e^{-z^2/2} \right]$$

- $N(0,1)$ density: $\phi(z) \equiv e^{-z^2/2}/(2\pi)^{1/2}$

- Hermite expansion for the density function of Z, p_Z, at order J :

$$p_Z^{(J)}(\Delta, z|y_0;\theta) \equiv \phi(z) \sum_{j=0}^{J} \eta_j(\Delta, y_0;\theta)\ H_j(z)$$

- This is an expansion in z, for fixed Δ, y_0 and θ.

- The unknowns are the coefficients η_j.

- By orthogonality of the Hermite polynomials, we have

$$\eta_j(\Delta, y_0; \theta) \equiv \frac{1}{j!} \int_{-\infty}^{+\infty} H_j(z)\, p_Z(\Delta, z|y_0; \theta)\, dz$$

- This means that η_j is in the form of an expected value over the density p_Z.

- This makes it possible to compute these coefficients explicitly, which will be done below.

2.5. Revert the Transformations $Z \mapsto Y \mapsto X$

- Assume for now that $p_Z^{(J)}$ has been obtained.

- From $p_Z^{(J)}$, we get a sequence of approximations to p_Y:

$$p_Y^{(J)}(\Delta, y|y_0;\theta) \equiv \Delta^{-1/2} p_Z^{(J)}\left(\Delta,\ \Delta^{-1/2}(y-y_0)\ |y_0;\theta\right)$$

- And then to p_X:

$$p_X^{(J)}(\Delta, x|x_0;\theta) \equiv \sigma(x;\theta)^{-1}\, p_Y^{(J)}(\Delta,\ \gamma(x;\theta)\,|\gamma(x_0;\theta)\ ;\theta)$$

- Recall that $Y_t \equiv \gamma\left(X_t; \theta\right) = \int^{X_t} du/\sigma\left(u; \theta\right)$

- So Y is a nonlinear transformation of X unless σ is a constant parameter, in which case $Y_t = X_t/\sigma$.

- As a result, the leading term of $p_X^{(J)}$ will in general not be Gaussian since it is a Gaussian term evaluated at a nonlinear function of X.

2.6. Convergence of the Density Sequence

- Theorem : There exists $\bar{\Delta} > 0$ such that for every $\Delta \in \left(0, \bar{\Delta}\right)$, $\theta \in \Theta$ and $(x, x_0) \in D_X^2$:

$$p_X^{(J)}(\Delta, x|x_0; \theta) \;\rightarrow\; p_X(\Delta, x|x_0; \theta) \quad \text{as } J \rightarrow \infty$$

- In addition:
  - the convergence is uniform in θ over Θ
  - in x over D_X
  - and in x_0 over compact subsets of D_X.

2.7. A Sequence of Approximations to the MLE

- Theorem: Maximizing

$$\ell_n^{(J)}(\theta) \equiv \sum_{i=1}^{n} \ln p_X^{(J)}\left(\Delta, X_{i\Delta} | X_{(i-1)\Delta};\, \theta\right)$$

  results in an estimator $\hat{\theta}_n^{(J)}$:

  - which converges to the exact (but uncomputable) MLE $\hat{\theta}_n$ as $J \longrightarrow \infty$
  - and inherits all its asymptotic properties.

- Methods to do small sample bias corrections for these maximum-likelihood estimates are provided by Phillips and Yu (2009) and Tang and Chen (2009).

- They are relevant in particular for the speed of mean reversion parameter when near a unit root.

- This is the case for US interest rate data for instance.

- Latest results on the asymptotics of the approximate MLE
  - Chang and Chen (2011): Consistency and rate of convergence of the approximate MLE as a function of the number of terms used in the density approximation (J) and Δ
  - Jeong and Park (2011): Optimal rates of convergence achieved by the approximate MLE differentiated by type of parameter to be estimated as a function of both the horizon of estimation (T) and the sampling interval (Δ)

2.8. Explicit Expressions for the Expansion

- Limit as the number of Hermite polynomials $J \to \infty$:

$$p_Z^{(\infty)}(\Delta, z|y_0;\theta) = \phi(z) \sum_{j=0}^{\infty} \eta_j(\Delta, y_0;\theta)\ H_j(z)$$

- To compute the coefficients $\eta_j(\Delta, y_0;\theta)$:
  - Taylor expand η_j in Δ up to order Δ^K
  - This gives rise to the expansion denoted $\tilde{p}_Z^{(K)}$

- Theorem:

$$\tilde{p}_Y^{(K)}(\Delta, y|y_0;\theta) = \Delta^{-1/2}\phi\left(\frac{y-y_0}{\Delta^{1/2}}\right)\exp\left(\int_{y_0}^{y}\mu_Y(w;\theta)\,dw\right) \times \sum_{k=0}^{K} c_k(y|y_0;\theta)\frac{\Delta^k}{k!}$$

- The coefficients are obtained in closed form:

$$c_0(y|y_0;\theta) = 1$$

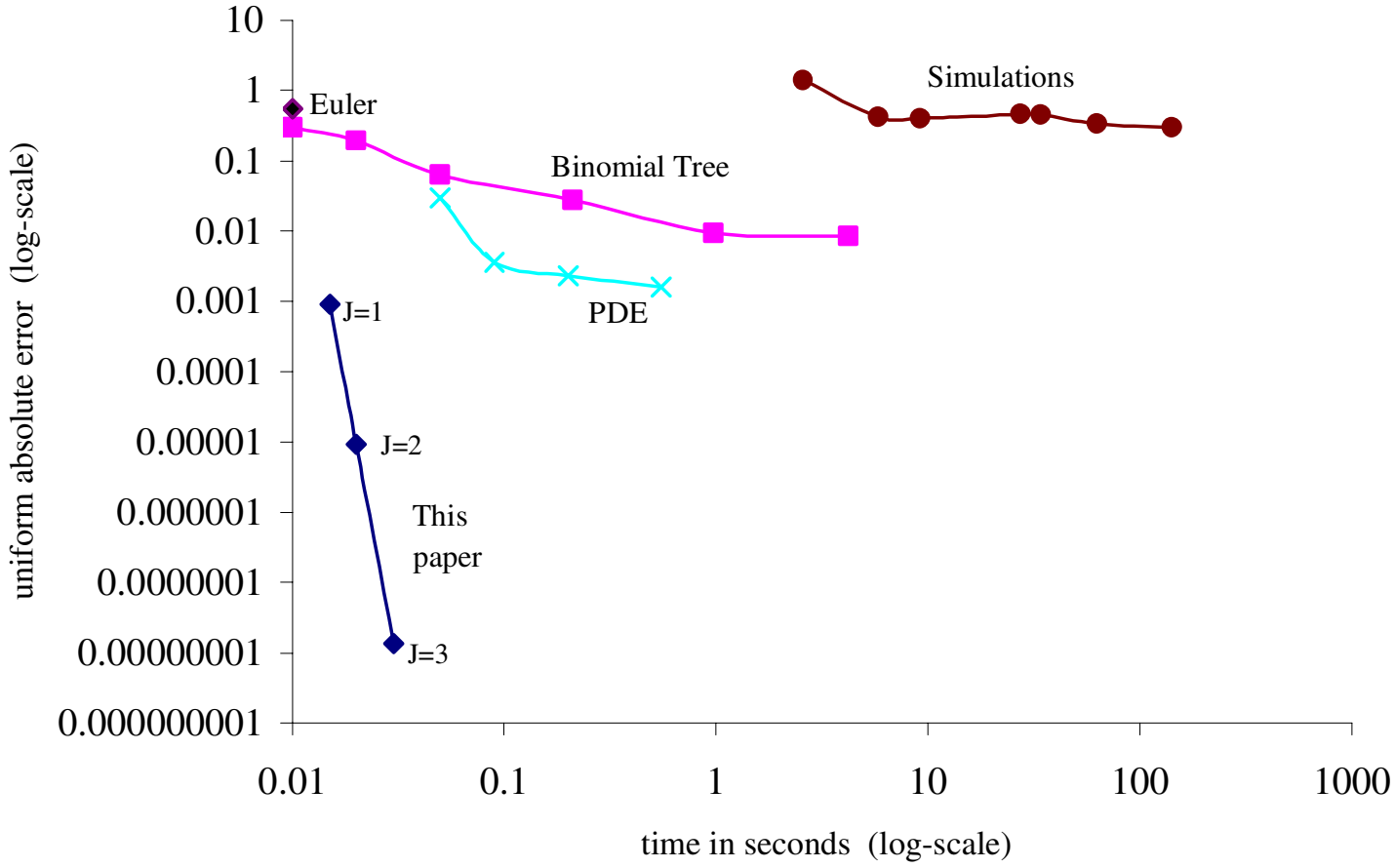
$$c_j(y|y_0;\theta) = j\,(y-y_0)^{-j}\int_{y_0}^{y}(w-y_0)^{j-1}\left\{\lambda_Y(w)\,c_{j-1}(w|y_0;\theta) + \frac{1}{2}\frac{\partial^2 c_{j-1}(w|y_0;\theta)}{\partial w^2}\right\}dw$$

$$\lambda_Y(y;\theta) \equiv -\frac{1}{2}\left(\mu_Y^2(y;\theta) + \frac{\partial\mu_Y(y;\theta)}{\partial y}\right)$$

2.9. Comparisons with Other Methods

Comparison studies include:

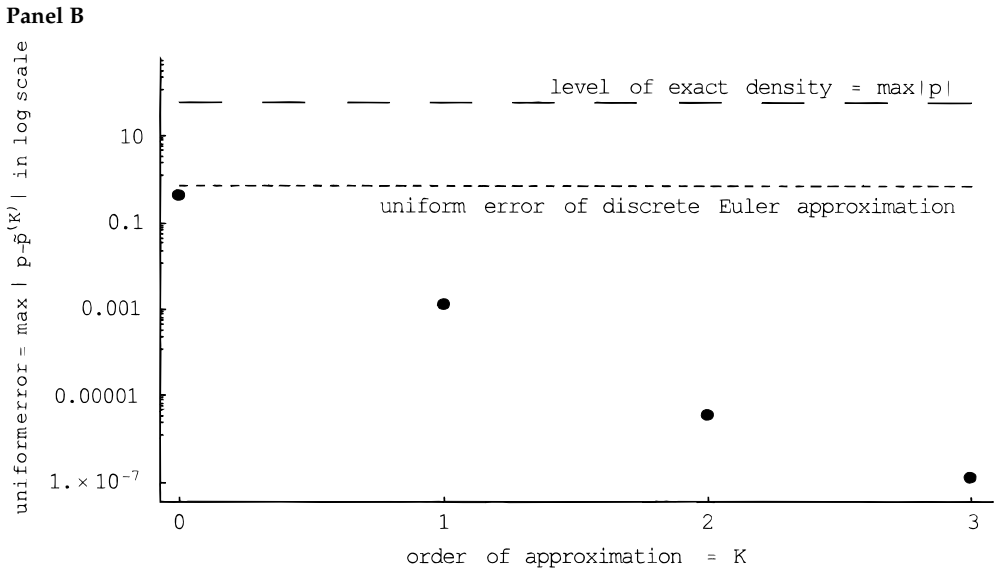
- Jensen and Poulsen (2002)
- Hurn et al. (2007)
- Stramer and Yan (2007)


10
1
0.1
0.01
0.001
0.0001
0.00001
0.000001
0.0000001
0.00000001
0.000000001
uniform absolute error (log-scale)
0.01
0.1
1
10
100
1000
time in seconds (log-scale)
Euler
Binomial Tree
Simulations
PDE
J=1
J=2
J=3
This
paper

2.10. Examples

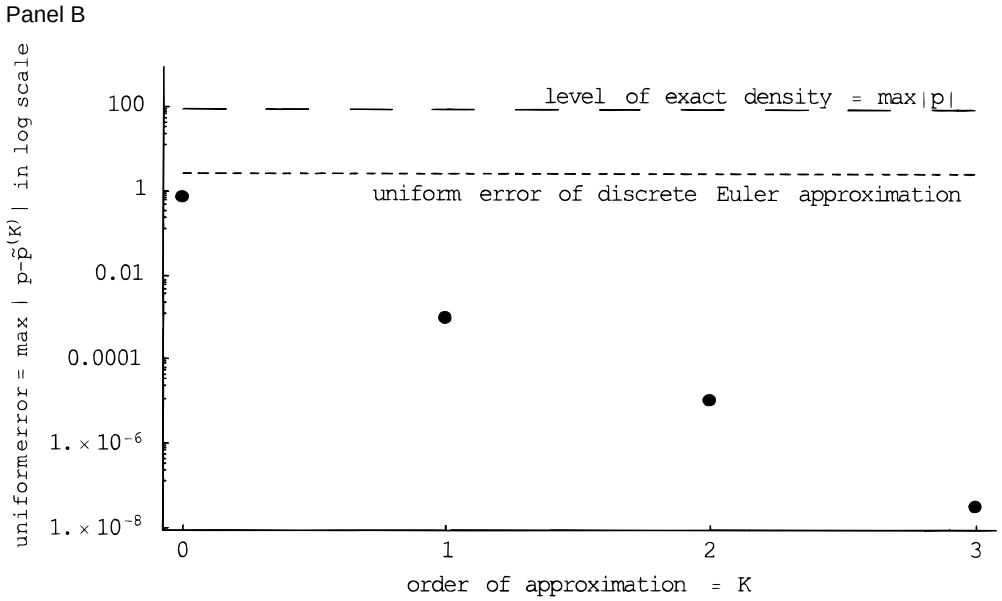
- Vasicek Model: $dX_t = \kappa(\alpha - X_t)\,dt + \sigma dW_t$

- Uniform approximation errors in log-scale



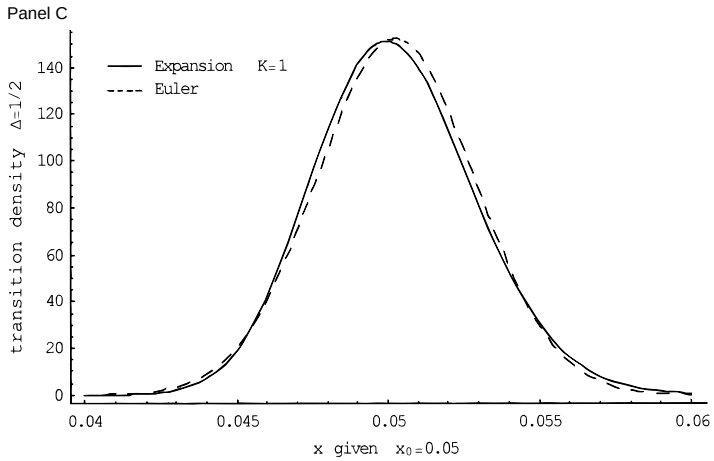
- CIR Model: $dX_t = \kappa(\alpha - X_t)\,dt + \sigma X_t^{1/2} dW_t$

- Uniform approximation errors in log-scale



- CEV Model: $dX_t = \kappa(\alpha - X_t)\,dt + \sigma X_t^{\gamma} dW_t$

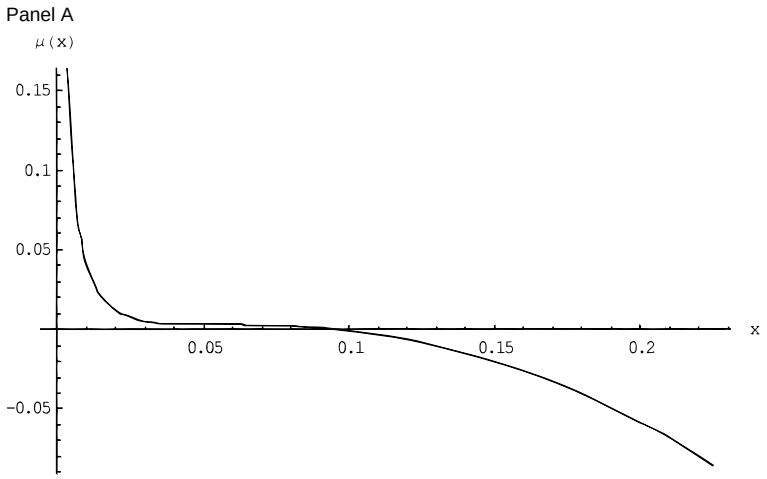
- Density from this method compared to Euler approximation



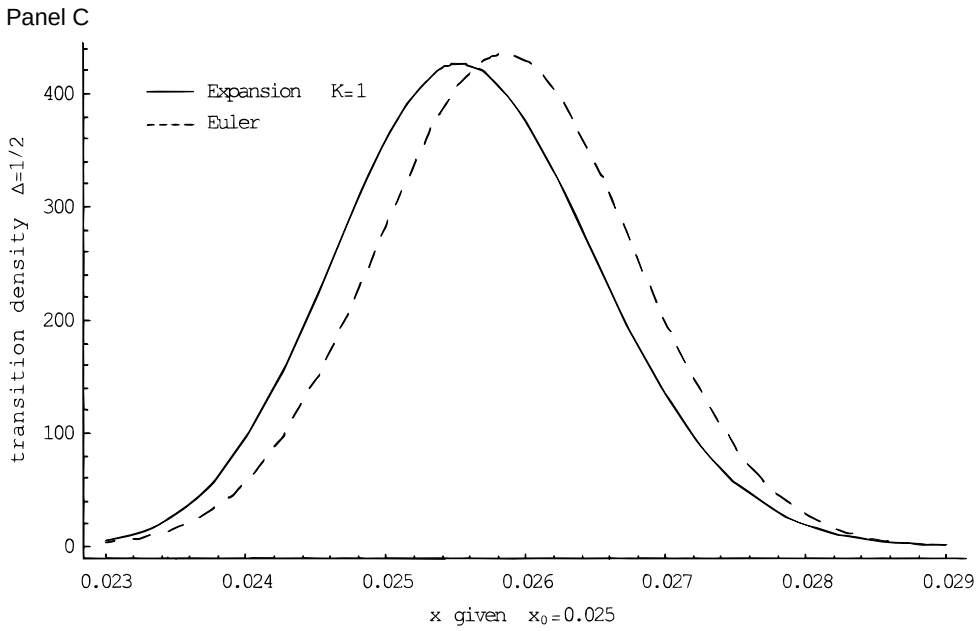
- Nonlinear Drift Model in Aït-Sahalia (1996b):

$$dX_t = \left(\alpha_{-1}X_t^{-1} + \alpha_0 + \alpha_1 X_t + \alpha_2 X_t^2\right) dt + \sigma X_t^{\gamma} dW_t$$

- Drift

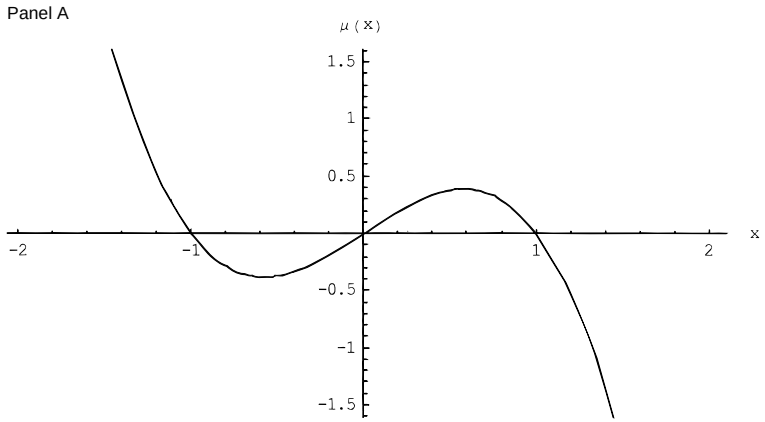


- Nonlinear Drift Model: Density from this method compared to Euler approximation

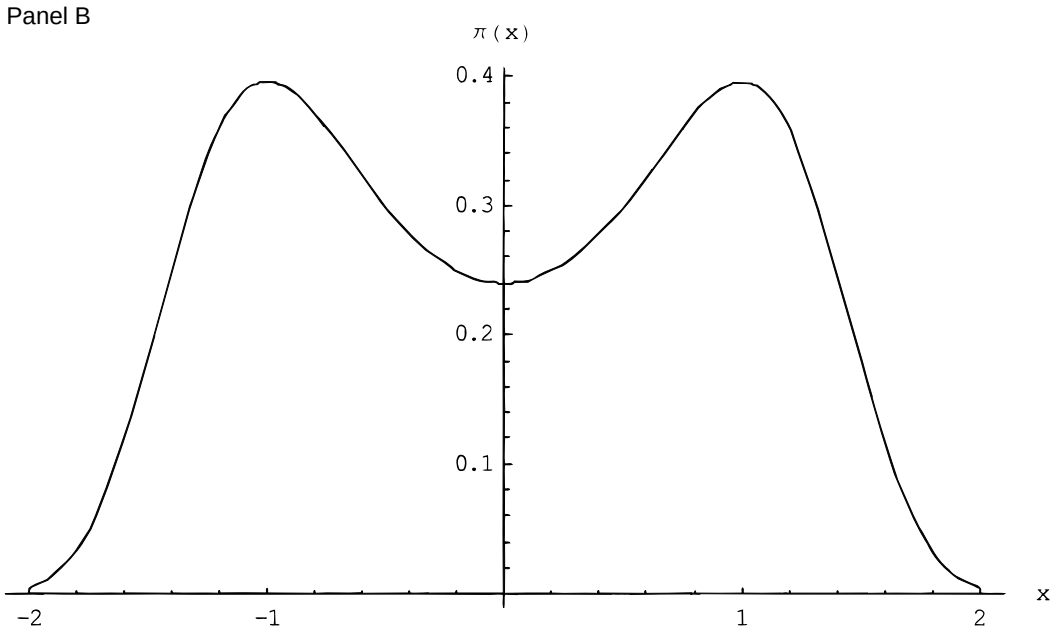


- Double-Well Model: $dX_t = \left(\alpha_1 X_t - \alpha_3 X_t^3\right) dt + dW_t$

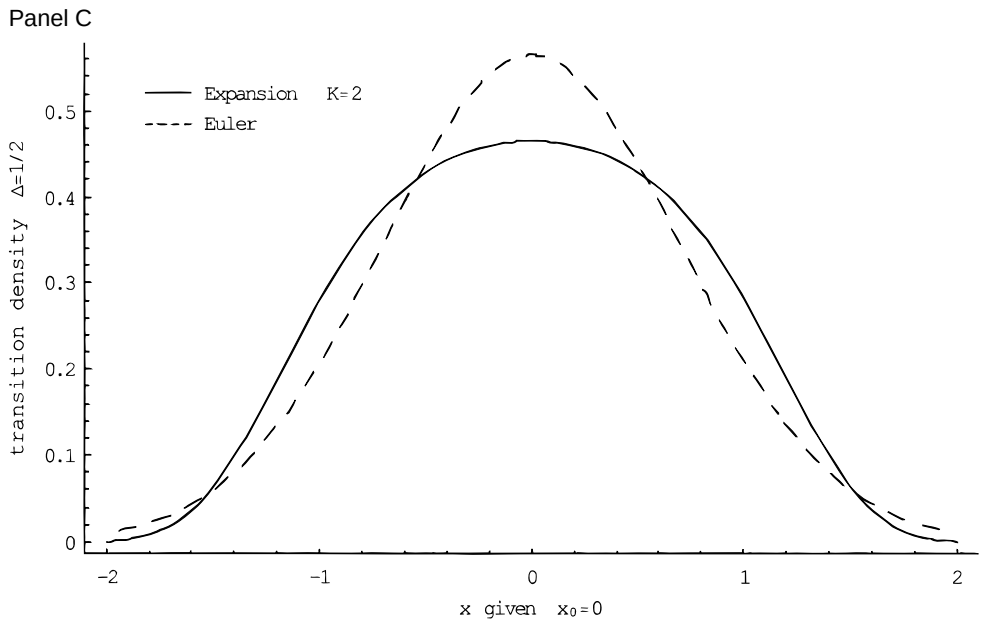
- Drift



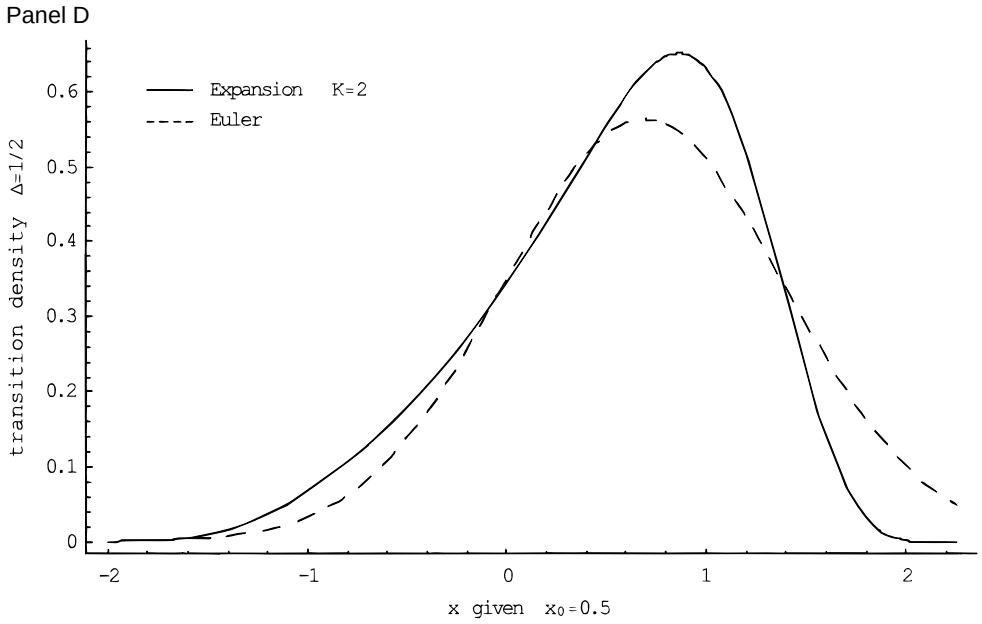
- Double-Well Model: Marginal Density



- Double-Well Model: Density from this method compared to Euler approximation, conditioned on center value ($x_0 = 0.0$)



- Double-Well Model: Density from this method compared to Euler approximation, conditioned on high value ($x_0 = 0.5$)



3. Multivariate Expansions

- Every univariate diffusion can be transformed into one with unit diffusion, whose density can then be approximated around a standard Normal.

- This is no longer the case for multivariate diffusions.

- I therefore introduce the concept of reducibility for multivariate diffusions, which characterizes diffusions for which such a transformation exists.

- For reducible multivariate diffusions, I will construct an expansion for the log-likelihood function in the form of a single Taylor series in the time variable Δ.

- For irreducible diffusions, however, one must proceed differently, in the form of a double Taylor expansion in the time variable Δ and the state vector change $x - x_0$.

3.1. Reducible Diffusions

- Definition: *The diffusion X is reducible to unit diffusion if and if only if there exists a one-to-one transformation of the diffusion X into a diffusion Y whose diffusion matrix σ_Y is the identity matrix:*

$$dY_t = \mu_Y\left(Y_t;\theta\right)dt + dW_t$$

- Whether or not a given multivariate diffusion is reducible depends on the specification of its σ matrix.

- Proposition: (Necessary and Sufficient Condition for Reducibility) The diffusion X is reducible if and only if the inverse diffusion matrix $\sigma^{-1} = \left[\sigma_{i,j}^{-1}\right]_{i,j=1,...,m}$ satisfies the condition that

$$\frac{\partial \sigma_{ij}^{-1}(x;\theta)}{\partial x_k} = \frac{\partial \sigma_{ik}^{-1}(x;\theta)}{\partial x_j}$$

for each triplet $(i,j,k) = 1,...,m$ such that $k > j$.

- Each model corresponds to a particular $\sigma(x;\theta)$ and so this can be checked for the model of interest.

Example 1: Diagonal Systems

- Since $\sigma_{ii}^{-1} = 1/\sigma_{ii}$ in the diagonal case, reducibility is equivalent to the fact that σ_{ii} depends only on x_i (and θ) for each i.

- Note that this is not the case if off-diagonal elements are present.

Example 2: Stochastic Volatility

- If

$$\sigma(x;\theta) = \begin{pmatrix} \sigma_{11}(x_2;\theta) & 0 \\ 0 & \sigma_{22}(x_2;\theta) \end{pmatrix}$$

then the process is not reducible, as this is a diagonal system where σ_{11} depends on x_2.

- However, if

$$\sigma(x;\theta) = \begin{pmatrix} a(x_1;\theta) & a(x_1;\theta)b(x_2;\theta) \\ 0 & c(x_2;\theta) \end{pmatrix}$$

then the process is reducible.

3.2. Reducible Expansion

- One particularly convenient way of gathering the terms of the expansion consists in grouping them in powers of Δ :

$$l_Y^{(K)}(\Delta, y|y_0;\theta) = -\frac{m}{2}\ln(2\pi\Delta) + \frac{C_Y^{(-1)}(y|y_0;\theta)}{\Delta} + \sum_{k=0}^{K} C_Y^{(k)}(y|y_0;\theta)\frac{\Delta^k}{k!}$$

- Leaving us with the computation of the coefficients $C_Y^{(k)}$, $k = -1, 0, 1, ...$

Theorem (Reducible Diffusions): The coefficients of the log-density Taylor expansion $l_Y^{(K)}(\Delta, y|y_0;\theta)$ *are given explicitly by:*

$$C_Y^{(-1)}(y|y_0;\theta) = -\frac{1}{2}\sum_{i=1}^{m}(y_i - y_{0i})^2$$
$$C_Y^{(0)}(y|y_0;\theta) = \sum_{i=1}^{m}(y_i - y_{0i})\int_0^1 \mu_{Yi}(y_0 + u(y - y_0);\theta)\,du$$

and, for $k \geq 1$,

$$C_Y^{(k)}(y|y_0;\theta) = k\int_0^1 G_Y^{(k)}(y_0 + u(y - y_0)|y_0;\theta)\,u^{k-1}du$$

where

$$G_Y^{(k)}(y|y_0;\theta) = -\sum_{i=1}^{m}\mu_{Yi}(y;\theta)\frac{\partial C_Y^{(k-1)}(y|y_0;\theta)}{\partial y_i} + \frac{1}{2}\sum_{i=1}^{m}\frac{\partial^2 C_Y^{(k-1)}(y|y_0;\theta)}{\partial y_i^2}$$
$$+\frac{1}{2}\sum_{i=1}^{m}\sum_{h=0}^{k-1}\binom{k-1}{h}\frac{\partial C_Y^{(h)}(y|y_0;\theta)}{\partial y_i}\frac{\partial C_Y^{(k-1-h)}(y|y_0;\theta)}{\partial y_i}.$$

3.3. Irreducible Expansion

- Mimicking the form of the Taylor expansion in Δ obtained in the reducible case:

$$l_X^{(K)}(\Delta, x|x_0;\theta) = -\frac{m}{2}\ln(2\pi\Delta) - D_v(x;\theta) + \frac{C_X^{(-1)}(x|x_0;\theta)}{\Delta} + \sum_{k=0}^{K} C_X^{(k)}(x|x_0;\theta)\frac{\Delta^k}{k!}.$$

- The idea now is to derive an explicit Taylor expansion in $(x - x_0)$ of the coefficients $C_X^{(k)}(x|x_0;\theta)$, at order j_k.

- For a balanced expansion:

$$j_k = 2(K - k)$$

for $k = -1, 0, ..., K$, will provide an aproximation error due to the Taylor expansion in $(x - x_0)$ of the same order Δ^K for each one of the terms in the series.

- This means in particular that the highest order term $(k = -1)$ is Taylor-expanded to a higher degree of precision than the successive terms.

3.3.1. Coefficients in the Irreducible Case

- The coefficients are determined one by one, starting with the leading term $C_X^{(j_{-1},-1)}$.

- Given $C_X^{(j_{-1},-1)}$, the next term $C_X^{(j_0,0)}$ is calculated explicitly, and so on.

3.4. Examples

3.4.1. A univariate example where $X \mapsto Y$ is not available in closed form

- Model proposed for the short term interest rate in Ait-Sahalia (1996)

$$dX_t = \left(\alpha_{-1}X_t^{-1} + \alpha_0 + \alpha_1 X_t + \alpha_2 X_t^2\right) dt + (\beta_0 + \beta_1 X_t + \beta_2 X_t^{\beta_3}) dW_t$$

- For this model, $1/\sigma$ cannot be integrated in closed-form in the general case.

- Bakshi and Ju (2005) proposed a method to circumvent the integrability of $1/\sigma$.

- An alternative is to use the irreducible method in that case, thereby bypassing the need for the $X \mapsto Y$ transformation.

- For instance, at order $K = 1$ in Δ, the irreducible expansion for the generic model $dX_t = \mu(X_t)dt + \sigma(X_t)dW_t$ is

$$\tilde{l}_X^{(1)}(\Delta, x|x_0;\theta) = -\frac{1}{2}\ln(2\pi\Delta) - \ln(\sigma(x;\theta)) + \frac{C_X^{(4,-1)}(x|x_0;\theta)}{\Delta} + C_X^{(2,0)}(x|x_0;\theta) + C_X^{(0,1)}(x|x_0;\theta)\Delta.$$

- The coefficients $C_X^{(j_k,k)}$, $k = -1, 0, 1$ are given by

$$C_X^{(4,-1)}(x|x_0;\theta) = \frac{\left(4\sigma(x_0;\theta)\sigma''(x_0;\theta) - 11\sigma'(x_0;\theta)^2\right)}{24\sigma(x_0;\theta)^4}(x-x_0)^4 + \frac{\sigma'(x_0;\theta)}{2\sigma(x_0;\theta)^3}(x-x_0)^3 - \frac{1}{2\sigma(x_0;\theta)^2}(x-x_0)^2$$

$$C_X^{(2,0)}(x|x_0;\theta) = \frac{\left(\left(\sigma'(x_0;\theta)^2 + 2\mu'(x_0;\theta)\right)\sigma(x_0;\theta) - 4\mu(x_0;\theta)\sigma'(x_0;\theta) - \sigma''(x_0;\theta)\sigma(x_0;\theta)^2\right)}{4\sigma(x_0;\theta)^3} \times (x-x_0)^2 + \frac{(2\mu(x_0;\theta) - \sigma(x_0;\theta)\sigma'(x_0;\theta))}{2\sigma(x_0;\theta)^2}(x-x_0)$$

$$C_X^{(0,1)}(x|x_0;\theta) = \frac{1}{8}\left(2\sigma(x_0;\theta)\sigma''(x_0;\theta) - \frac{4\mu(x_0;\theta)^2}{\sigma(x_0;\theta)^2} + \frac{8\sigma'(x_0;\theta)\mu(x_0;\theta)}{\sigma(x_0;\theta)} - \sigma'(x_0;\theta)^2 - 4\mu'(x_0;\theta)\right)$$

- Bakshi et al. (2006) use these expansions to estimate models for volatility.

3.4.2. The Bivariate Ornstein-Uhlenbeck Model

$$\begin{pmatrix} dX_{1t} \\ dX_{2t} \end{pmatrix} = \begin{pmatrix} \beta_{11}(\alpha_1 - X_{1t}) + \beta_{12}(\alpha_2 - X_{2t}) \\ \beta_{21}(\alpha_1 - X_{1t}) + \beta_{22}(\alpha_2 - X_{2t}) \end{pmatrix} dt + \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} \begin{pmatrix} dW_{1t} \\ dW_{2t} \end{pmatrix}$$

- This model has Gaussian transitions.

- X is reducible, and

$$dY_t = (\gamma - \kappa Y_t)\, dt + dW_t$$

- The coefficients of the expansion are:

$$C_Y^{(-1)}(y|y_0;\theta) = -\tfrac{1}{2}(y_1 - y_{01})^2 - \tfrac{1}{2}(y_2 - y_{02})^2$$

$$\begin{aligned} C_Y^{(0)}(y|y_0;\theta) = &-\tfrac{1}{2}(y_1 - y_{01})((y_1 + y_{01})\kappa_{11} + (y_2 + y_{02})\kappa_{12} - 2\gamma_1) \\ &- \tfrac{1}{2}(y_2 - y_{02})((y_1 + y_{01})\kappa_{21} + (y_2 + y_{02})\kappa_{22} - 2\gamma_2) \end{aligned}$$

$$
\begin{aligned}
C_Y^{(1)}(y|y_0;\theta) = & -\tfrac{1}{2}\left(\left(\gamma_1 - y_{02}\kappa_{12} - y_{01}\kappa_{11}\right)^2 + \left(\gamma_2 - y_{01}\kappa_{21} - y_{02}\kappa_{22}\right)^2 - \kappa_{11} - \kappa_{22}\right) \\
& + \tfrac{1}{2}\left(y_1 - y_{01}\right)\left(\gamma_1\kappa_{11} + \gamma_2\kappa_{21} - y_{01}\left(\kappa_{11}^2 + \kappa_{21}^2\right) - y_{02}\left(\kappa_{11}\kappa_{12} + \kappa_{21}\kappa_{22}\right)\right) \\
& + \tfrac{1}{24}\left(y_1 - y_{01}\right)^2\left(-4{\kappa_{11}}^2 + {\kappa_{12}}^2 - 2\kappa_{12}\kappa_{21} - 3\kappa_{21}^2\right) \\
& + \tfrac{1}{2}\left(y_2 - y_{02}\right)\left(\gamma_1\kappa_{12} + \gamma_2\kappa_{22} - y_{01}\left(\kappa_{11}\kappa_{12} + \kappa_{21}\kappa_{22}\right) - y_{02}\left(\kappa_{12}^2 + \kappa_{22}^2\right)\right) \\
& + \tfrac{1}{24}\left(y_2 - y_{02}\right)^2\left(-3\kappa_{12}^2 - 2\kappa_{12}\kappa_{21} + \kappa_{21}^2 - 4\kappa_{22}^2\right) \\
& - \tfrac{1}{3}\left(y_1 - y_{01}\right)\left(y_2 - y_{02}\right)\left(\kappa_{11}\kappa_{12} + \kappa_{21}\kappa_{22}\right)
\end{aligned}
$$

and

$$
\begin{aligned}
C_Y^{(2)}(y|y_0;\theta) = & -\tfrac{1}{12}\left(2\kappa_{11}^2 + 2\kappa_{22}^2 + \left(\kappa_{12} + \kappa_{21}\right)^2\right) \\
& + \tfrac{1}{6}\left(y_1 - y_{01}\right)\left(\kappa_{12} - \kappa_{21}\right)\left(-\gamma_1\kappa_{12} - \gamma_2\kappa_{22} + y_{01}\left(\kappa_{11}\kappa_{12} + \kappa_{21}\kappa_{22}\right) + y_{02}\left(\kappa_{12}^2 + \kappa_{22}^2\right)\right) \\
& + \tfrac{1}{12}\left(y_1 - y_{01}\right)^2\left(\kappa_{12} - \kappa_{21}\right)\left(\kappa_{11}\kappa_{12} + \kappa_{21}\kappa_{22}\right) \\
& + \tfrac{1}{6}\left(y_2 - y_{02}\right)\left(\kappa_{21} - \kappa_{12}\right)\left(-\gamma_1\kappa_{11} - \gamma_2\kappa_{21} + y_{02}\left(\kappa_{11}\kappa_{12} + \kappa_{21}\kappa_{22}\right) + y_{01}\left(\kappa_{11}^2 + \kappa_{21}^2\right)\right) \\
& + \tfrac{1}{12}\left(y_2 - y_{02}\right)^2\left(\kappa_{21} - \kappa_{12}\right)\left(\kappa_{11}\kappa_{12} + \kappa_{21}\kappa_{22}\right) \\
& + \tfrac{1}{12}\left(y_1 - y_{01}\right)\left(y_2 - y_{02}\right)\left(\kappa_{12} - \kappa_{21}\right)\left(\kappa_{22}^2 + \kappa_{12}^2 - \kappa_{11}^2 + \kappa_{21}^2\right)
\end{aligned}
$$

3.4.3. A Stochastic Volatility Model

- Consider as a second example a typical stochastic volatility model

$$\begin{pmatrix} dX_{1t} \\ dX_{2t} \end{pmatrix} = \begin{pmatrix} \mu \\ \kappa(\alpha - X_{2t}) \end{pmatrix} dt + \begin{pmatrix} \gamma_{11} \exp(X_{2t}) & 0 \\ 0 & \gamma_{22} \end{pmatrix} \begin{pmatrix} dW_{1t} \\ dW_{2t} \end{pmatrix}$$

  where X_{1t} plays the role of the log of an asset price and $\exp(X_{2t})$ is the stochastic volatility variable.

- This model has no closed-form density.

- And is not reducible.

4. Expansions for Models with Jumps

- Consider jumps of finite activity with intensity $\lambda(x;\theta)$ and jump measure $\nu(y-x;\theta)$.

- By Bayes' Rule, we have

$$p(y|x;\theta) = \sum_{n=0}^{+\infty} p(y|x, N_\Delta = n;\theta)\,\mathsf{Pr}\,(N_\Delta = n|x;\theta)$$

- With $\mathsf{Pr}\,(N_\Delta = 0|x;\theta) = O(1)$, $\mathsf{Pr}\,(N_\Delta = 1|x;\theta) = O(\Delta)$ and $\mathsf{Pr}\,(N_\Delta > 1|x;\theta$ $o(\Delta)$, and the fact that when at least one jump occurs the dominant effect is due to the jump (vs. the increment due to the Brownian increment), an expansion at order K in Δ of p obtained by extending

the pure diffusive result to jump-diffusions is shown by Yu (2007) to be

$$\tilde{p}^{(K)}(y|x;\theta) = \exp\left(-\frac{1}{2}\ln\left(2\pi\Delta\sigma^2(y;\theta)\right) + \frac{c_{-1}(y|x;\theta)}{\Delta}\right) \times \sum_{k=0}^{K} c_k(y|x;\theta)\frac{\Delta^k}{k!} + \sum_{k=1}^{K} d_k(y|x;\theta)\frac{\Delta^k}{k!}.$$

- The unknowns are the coefficients c_k and d_k of the series.

- Relative to the pure diffusive case, the coefficients d_k are the new terms needed to capture the presence of the jumps in the transition function and will capture the different behavior of the tails of the transition density when jumps are present.

- These tails are not exponential in y, hence the absence of a the factor $\exp(c_{-1}\Delta^{-1})$ in front of the summation of d_k coefficients.

- The coefficients c_k and d_k can be computed analogously to the pure diffusive case, resulting in a system of equations that can be solved in closed form, starting with c_{-1} and c_0.

- Coefficients of higher order of the diffusive part of the expansion (i.e., c_k, $k \geq 1$) are no longer functions of the diffusive characteristics of the process only; instead, they also involve the characteristics of the jump part.

- In particular, for $k = 1$,

$$c_1(y|x;\theta) = -\left(\int_x^{y_t} \frac{du}{\sigma(u;\theta)}\right)^{-1} \times \int_x^y \left\{ \frac{du}{\sigma(u;\theta)} \exp\left(\int_x^s \frac{\mu(u;\theta)}{\sigma^2(u;\theta)} du - \int_x^{y_t} \frac{\partial\sigma(u;\theta)/\partial u}{2\sigma(u;\theta)} du\right) \times \left(\int_s^y \frac{du}{\sigma(u;\theta)}\right) (\lambda(s;\theta) - \mathcal{A} \cdot c_0(y|s;\theta)) \right\} \frac{ds}{\sigma(s;\theta)}$$

where the operator $\mathcal{A}$ is the generator of the diffusive part of the process only, defined by its action

$$\mathcal{A} \bullet f = \mu \frac{\partial f}{\partial x} + \frac{1}{2}\sigma^2 \frac{\partial^2 f}{\partial x^2}$$

on functions in its domain.

- As for the leading term in the jump part of the expansion (i.e., d_1), it is given by

$$d_1(y|x;\theta) = \lambda(x;\theta)\,\nu(y-x;\theta).$$

- As in the pure diffusive case, higher order terms d_k, $k \geq 2$, are obtained recursively from the preceding ones.

5. Connection to Saddlepoint Approximations

- Aït-Sahalia and Yu (2006) developed an alternative strategy for constructing closed form approximations.

- Instead of expanding the transition function in orthogonal polynomials around a leading term, they rely on the saddlepoint method.

  – They replace the characteristic function by an expansion in Δ.

  – The expansion is then computed from the infinitesimal generator $\mathcal{A}$ of the Markov process.

6. Inference When the State is Partially Observed

- In many cases, the state vector is of the form $X_t = [S_t; V_t]'$, where the $(m-q)-$dimensional vector S_t is observed but the $q-$dimensional V_t is not.

- Two typical examples:

  - Stochastic Volatility Models: S_t is an asset price, V_t is its volatility;

  - Term Structure Models: V_t is a vector of latent state variables (say, macro factors) but observations are yields of bonds of different maturities.

- Inference Strategy:
  - Write down in closed form an expansion for the log-likelihood of the state vector X, including its unobservable components.
  - Then enlarge the observation state by adding variables that are observed and functions of X.
    - For example, in the stochastic volatility case, an option price or an option-implied volatility.
    - In term structure models, as many bonds as there are factors.

– Then, using the Jacobian formula, write down the likelihood function of the pair consisting of the observed components of X and the additional observed variables:

  * Let the observed state be $G_t = [S_t; C_t]' = f(X_t; \theta)$.

  * The log-likelihood of G is

$$\ln p_G(g|g_0, \Delta; \theta) = -\ln J_t(g|g_0, \Delta; \theta) + \ln p_X(f^{-1}(g; \theta)\,|f^{-1}(g_0; \theta); \Delta, \theta)$$

– And conduct inference using the augmented data $G_t = [S_t; C_t]'$ and the above likelihood function.

6.1. Stochastic Volatility Models

- We can estimate the model using observations on asset price S_t and an option C_t, or an implied volatility: see Aït-Sahalia and Kimmel (2007).

- Let $V(t,T) = \int_t^T V_u du$.

- If V_t is instantaneously uncorrelated with S_t, then we can calculate option prices by taking the expected value of the Black-Scholes option price over the probability distribution of $V(t,T)$.

- If not, then the price of the option is a weighted average of Black-Scholes prices evaluated at different stock prices and volatilities.

6.2. Term Structure Models

- Another example of practical interest in finance consists of term structure models.

- A multivariate term structure model specifies that the instantaneous riskless rate r_t is a deterministic function of an $m-$dimensional vector of state variables, X_t:

$$r_t = r\left(X_t; \theta\right).$$

- An affine yield model is any model where the short rate is an affine function of the state vector and the risk-neutral dynamics are affine:

$$dX_t = \left(\tilde{A} + \widetilde{B} X_t\right) dt + \mathbf{\Sigma}\ S\left(X_t; \alpha, \beta\right)^{1/2} dW_t$$

where $\tilde{A}$ is an m–element column vector, $\widetilde{B}$ and Σ are $m \times m$ matrices, and $S(X_t; \alpha, \beta)$ is the diagonal matrix with elements $S_{ii} = \alpha_i + X_t' \beta_i$, with each α_i a scalar and each β_i an $m \times 1$ vector, $1 \leq i \leq m$ (see Dai and Singleton (2000)).

- It can then be shown that, in affine models, bond prices have the exponential affine form $\exp\left(-\gamma_0(\tau;\theta) - \gamma(\tau;\theta)' x\right)$ where $\tau = T - t$ is the bond's time to maturity.

- Aït-Sahalia and Kimmel (2010) derive the likelihood expansions for the nine canonical models of Dai and Singleton (2000).

- The method is implemented in various contexts and with various datasets by Bakshi et al. (2006), Cheridito et al. (2007), Mosburger and Schneider (2005), Thompson (2008) and Egorov et al. (2011).

7. Application to Specification Testing

- When attempting to select a model among many possible choices, a natural idea is to try to pick the model that produces the functions μ and σ that are closest to those that are inferred from the data.

- Without high frequency observations, however, it is not possible to infer those infinitesimal functions from the data.

- Aït-Sahalia (1996) used the mapping between the drift and diffusion on the one hand, and the marginal density π_X and transition density p_X on the other, to test the model's specification using densities at the observed discrete frequency (π_X, p_X) instead of the infinitesimal characteristics of the process (μ, σ^2):

  - Any parametrization of μ and σ^2 corresponds to a parametrization of the marginal and transitional densities.

  - For example, the Ornstein-Uhlenbeck process $dX_t = \beta(\alpha - X_t)\,dt + \gamma dW_t$ generates Gaussian marginal and transitional densities. The square-root process $dX_t = \beta(\alpha - X_t)\,dt + \gamma X_t^{1/2} dW_t$ yields a Gamma marginal and non-central chi-squared transitional densities.

- Chen and Gao (2004) and Thompson (2008) proposed tests based on the empirical likelihood method.

- Hong and Li (2005) use the fact that under the null hypothesis, the random variables $\{P_X(X_{i\Delta}|X_{(i-1)\Delta}, \Delta, \theta)\}$ are a sequence of i.i.d.

uniform random variables; see also Chen et al. (2008) and Corradi and Swanson (2005).

- Using for P_X the closed form approximations described above, they detect the departure from the null hypothesis by comparing the kernel-estimated bivariate density of $\{(Z_i, Z_{i+\Delta})\}$ with that of the uniform distribution on the unit square, where $Z_i = P_X(X_{i\Delta}|X_{(i-1)\Delta}, \Delta, \theta)$.

- Other test statistics can be constructed, including some that are more powerful for detecting local departures from the null model, or more powerful for detecting global departures.

- An alternative specification test for the transition density of the process is proposed by Aït-Sahalia et al. (2009), based on a direct comparison

of a nonparametric estimate of the transition function to the assumed parametric transition function

$$\begin{aligned} & H_0 : p_X(y|x, \Delta) = p_X(y|x, \Delta, \theta) \\ \text{vs.} \quad & H_1 : p_X(y|x, \Delta) \neq p_X(y|x, \Delta, \theta). \end{aligned}$$

- To estimate the transition density nonparametrically, they use the method of Fan et al. (1996), yielding an estimator $\hat{p}_X(y|x, \Delta)$ which is then compared to $p_X(y|x, \Delta, \hat{\theta})$ using the likelihood ratio under the null and the alternative hypotheses.

- This leads to the test statistic, where w is a weight function,

$$\begin{aligned} T = \sum_{i=1}^{n} & \ln \left(\hat{p}_X(X_{(i+1)\Delta}|X_{i\Delta}, \Delta) / p_X(X_{(i+1)\Delta}|X_i, \Delta, \hat{\theta}) \right) \\ & \times w(X_{i\Delta}, X_{(i+1)\Delta}). \end{aligned}$$

8. Closed-Form Derivative Pricing

- As long as Δ is not too large, one can use the expression for the transition density p_X^* corresponding to the SDE:
  - Say a call option with payoff function $g(\cdot, K)$ for a strike K and maturity Δ when the underlying asset is worth S_t
  - Get a closed form approximation of the derivative price $Call$ in the form

$$Call(X_t, \Delta, K) = e^{-r\Delta} \int_0^{+\infty} g(y, K)\, p_X^*(y|x, \Delta, \theta) dy$$

- The formula obtained with is a closed-form expansion of p_X^* of a different nature than the corrections to the Black-Scholes formula as in for example Jarrow and Rudd (1982).

- These corrections are based on assuming an arbitrary parametric density family that nests the Black-Scholes models.

- As a result, these corrections break the link between the derivative price and the dynamic model for the underlying asset price by assuming directly a functional form for p_X.

- By contrast, if we use p^*_X obtained using the method above, then this is the option pricing formula (of finite order in Δ) that matches the assumed risk-neutral dynamics of the underlying asset.

- Being in closed form, comparative statics, etc. are possible.

- Being an expansion in small time, accuracy will be limited to relatively small values of Δ, of the order of up to 3 months in practical applications.

- The equation can also be implemented with payoff functions other than those of a call or put options.

9. Conclusions

- These methods make it possible to estimate and test continuous-time models using financial data, consisting of either observations on the underlying asset or factor, or on derivatives' prices.

- In either case, the transition density of the process is the key object to move from the infinitesimal specification of the model in the form of a stochastic differential equation to its implications at the discrete frequency of observation.

- Closed-form expansions for p_X that are now available for arbitrary models remove a basic constraint that had previously limited the range of applications for which this was possible.

References

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

Aït-Sahalia, Y., 1999. Transition densities for interest rate and other nonlinear diffusions. The Journal of Finance 54, 1361–1395.

Aït-Sahalia, Y., 2002. Maximum-likelihood estimation of discretely-sampled diffusions: A closed-form approximation approach. Econometrica 70, 223–262.

Aït-Sahalia, Y., 2008. Closed-form likelihood expansions for multivariate diffusions. Annals of Statistics 36, 906–937.

Aït-Sahalia, Y., Fan, J., Peng, H., 2009. Nonparametric transition-based tests for jump-diffusions. Journal of the American Statistical Association 104, 1102–1116.

Aït-Sahalia, Y., Kimmel, R., 2007. Maximum likelihood estimation of stochastic volatility models. Journal of Financial Economics 83, 413–452.

Aït-Sahalia, Y., Kimmel, R., 2010. Estimating affine multifactor term structure models using closed-form likelihood expansions. Journal of Financial Economics 98, 113–144.

Aït-Sahalia, Y., Yu, J., 2006. Saddlepoint approximations for continuous-time Markov processes. Journal of Econometrics 134, 507–551.

Bakshi, G., Ju, N., Ou-Yang, H., 2006. Estimation of continuous-time models with an application to equity volatility dynamics. Journal of Financial Economics 82, 227–249.

Bakshi, G. S., Ju, N., 2005. A refinement to Aït-Sahalia's (2002) "Maximum likelihood estimation of discretely sampled diffusions: A closed-form approximation approach". Journal of Business 78, 2037–2052.

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

Brandt, M., Santa-Clara, P., 2002. Simulated likelihood estimation of diffusions with an application to exchange rate dynamics in incomplete markets. Journal of Financial Economics 63, 161–210.

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 48, 1209–1227.

Chang, J., Chen, S. X., 2011. On the approximate maximum likelihood estimation for diffusion processes. Tech. rep., Peking University and Iowa State University.

Chen, S. X., Gao, J., 2004. An adaptive empirical likelihood test for time series models. Tech. rep., Iowa State University.

Chen, S. X., Gao, J., Tang, C., 2008. A test for model specification of diffusion processes. Annals of Statistics 36, 167–198.

Cheridito, P., Filipović, D., Kimmel, R. L., 2007. Market price of risk specifications for affine models: Theory and evidence. Journal of Financial Economics 83, 123–170.

Choi, S., 2008. Closed-from likelihood expansions for multivariate time-inhomogeneous diffusions. Tech. rep., University of Adelaide.

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

Corradi, V., Swanson, N. R., 2005. A bootstrap specification test for diffusion processes. Journal of Econometrics 124, 117–148.

Courtadon, G., 1982. The pricing of options on default free bonds. Journal of Financial and Quantitative Analysis 17, 75–100.

Cox, J. C., 1975. The constant elasticity of variance option pricing model. The Journal of Portfolio Management (1996 special Issue) 22, 15–17.

Cox, J. C., Ingersoll, J. E., Ross, S. A., 1985. A theory of the term structure of interest rates. Econometrica 53, 385–408.

Dai, Q., Singleton, K. J., 2000. Specification analysis of affine term structure models. The Journal of Finance 55, 1943–1978.

DiPietro, M., 2001. Bayesian inference for discretely sampled diffusion processes with financial applications. Ph.D. thesis, Department of Statistics, Carnegie-Mellon University.

Duffie, D., Kan, R., 1996. A yield-factor model of interest rates. Mathematical Finance 6, 379–406.

Egorov, A. V., Li, H., Ng, D., 2011. A tale of two yield curves: Modeling the joint term structure of dollar and euro interest rates. Journal of Econometrics 162, 55–70.

Egorov, A. V., Li, H., Xu, Y., 2003. Maximum likelihood estimation of time inhomogeneous diffusions. Journal of Econometrics 114, 107–139.

Fan, J., Yao, Q., Tong, H., 1996. Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems. Biometrika 83, 189–206.

Hong, Y., Li, H., 2005. Nonparametric specification testing for continuous-time models with applications to term structure of interest rates. Review of Financial Studies 18, 37–84.

Hurn, A., Jeisman, J., Lindsay, K., 2007. Seeing the wood for the trees: A critical evaluation of methods to estimate the parameters of stochastic differential equations. Journal of Financial Econometrics 5, 390–455.

Jarrow, R., Rudd, A., 1982. Approximate option valuation for arbitrary stochastic processes. Journal of Financial Economics 10, 347–369.

Jensen, B., Poulsen, R., 2002. Transition densities of diffusion processes: Numerical comparison of approximation techniques. Journal of Derivatives 9, 1–15.

Jeong, M., Park, J. Y., 2011. Asymptotic theory of maximum likelihood estimator for diffusion model. Tech. rep., Indiana University.

Li, M., 2010. A damped diffusion framework for financial modeling and closed-form maximum likelihood estimation. Journal of Economic Dynamics and Control 34, 132–157.

Lo, A. W., 1988. Maximum likelihood estimation of generalized Itô processes with discretely sampled data. Econometric Theory 4, 231–247.

Marsh, T., Rosenfeld, E., 1982. Stochastic processes for interest rates and equilibrium bond prices. The Journal of Finance 38, 635–646.

Mosburger, G., Schneider, P., 2005. Modelling international bond markets with affine term structure models. Tech. rep., Vienna University of Economics and Business Administration.

Pedersen, A. R., 1995. A new approach to maximum-likelihood estimation for stochastic differential equations based on discrete observations. Scandinavian Journal of Statistics 22, 55–71.

Phillips, P. C., Yu, J., 2009. Maximum likelihood and Gaussian estimation of continuous time models in finance. In: Andersen, T., Davis, R., Kreib, J.-P., Mikosch, T. (Eds.), Handbook of Financial Time Series. Springer, New York, NY.

Schaumburg, E., 2001. Maximum likelihood estimation of jump processes with applications to finance. Ph.D. thesis, Princeton University.

Stramer, O., Bognar, M., Schneider, P., 2009. Bayesian inference for discretely sampled markov processes with closedform likelihood expansions. Journal of Financial Econometrics 8, 450–480.

Stramer, O., Yan, J., 2007. On simulated likelihood of discretely observed diffusion processes and comparison to closed-form approximation. Journal of Computational and Graphical Statistics 16, 672–691.

Tang, C. Y., Chen, S. X., 2009. Parameter estimation and bias correction for diffusion processes. Journal of Econometrics 149, 65–81.

Thompson, S., 2008. Identifying term structure volatility from the LIBOR-swap curve. Review of Financial Studies 21, 819–854.

Vasicek, O., 1977. An equilibrium characterization of the term structure. Journal of Financial Economics 5, 177–188.

Yu, J., 2007. Closed-form likelihood approximation and estimation of jump-diffusions with an application to the realignment risk of the chinese yuan. Journal of Econometrics 141, 1245–1280.