

Closed-Form Likelihood Approximation and Estimation of Jump-Diffusions with an Application to the Realignment Risk of the Chinese Yuan

Jialin Yu*
Department of Finance & Economics
Columbia University[†]

First draft: January 30, 2003
This version: November 12, 2006

Abstract

This paper provides closed-form likelihood approximations for multivariate jump-diffusion processes widely used in finance. For a fixed order of approximation, the maximum-likelihood estimator (MLE) computed from this approximate likelihood achieves the asymptotic efficiency of the true yet *uncomputable* MLE estimator as the sampling interval shrinks. This method is then used to uncover the realignment probability of the Chinese Yuan. Since February 2002, the realignment intensity implicit in the financial market has increased fivefold. The term structure of the *forward realignment rate*, which completely characterizes realignment probabilities in the future, is hump-shaped and peaks at six months from the end of 2003. The realignment probability responds quickly to news releases on the Sino-US trade surplus, state-owned enterprise reform, Chinese government tax revenue and, most importantly, government interventions.

*I thank Yacine Aït-Sahalia for helpful discussions throughout this project. I am especially grateful to Peter Robinson (the Editor), an associate editor and two anonymous referees for comments and suggestions that greatly improved the paper. I also thank Laurent Calvet, Gregory Chow, Martin Evans, Rodrigo Guimaraes, Harrison Hong, Bo Honoré, Robert Kimmel, Adam Purzitsky, Hélène Rey, Jacob Sagi, Ernst Schaumburg, José Scheinkman, Wei Xiong and participants of the Princeton Gregory C. Chow Econometrics Research seminar and Princeton Microeconomic Workshop for helpful discussions. All errors are mine.

[†]421 Uris Hall, 3022 Broadway, New York, NY 10027. Email: jy2167@columbia.edu

1 Introduction

Jump-diffusions are very useful for modeling various economic phenomena such as currency crises, financial market crashes, defaults etc. There are now substantial evidence of jumps in financial markets. See for example Andersen, Bollerslev and Diebold (2006), Barndorff-Nielsen and Shephard (2004) and Huang and Tauchen (2005) on jumps in modeling and forecasting return volatility, Andersen, Benzoni and Lund (2002) and Eraker, Johannes and N. Polson (2003) on jumps in stock market, Piazzesi (2005) and Johannes (2004) on bond market, Bates (1996) on currency market, Duffie and Singleton (2003) on credit risk. From a statistical point of view, jump-diffusion nests diffusion as a special case. Being a more general model, jump-diffusion can approximate the true data-generating process better as measured by, for example, Kullback-Leibler Information Criterion (KLIC) (see White (1982)). As illustrated by Merton (1992), jumps model “Rare Events” that can have a big impact over a short period of time. These rare events can have very different implications from diffusive events. For example, it is known in the derivative pricing theory that, with jumps, the arbitrage pricing argument leading to the Black-Scholes option pricing formula breaks down. Liu, Longstaff and Pan (2003) have shown that the risks brought by jumps and stochastic volatility dramatically change an investor’s optimal portfolio choice.

To estimate jump-diffusions, likelihood-based methods such as maximum-likelihood estimation are preferred. This optimality is well documented in the statistics literature. However, maximum-likelihood estimation is difficult to implement because the likelihood function is available in closed-form for only a handful of processes. This difficulty can be seen from Sundaresan (2000): “The challenge to the econometricians is to present a framework for estimating such multivariate diffusion processes, which are becoming more and more common in financial economics in recent times. ... The development of estimation procedures for multivariate AJD processes is certainly a very important step toward realizing this hope.” Note the method proposed in this paper will apply to both affine jump-diffusion (AJD in the quote) and non-affine jump-diffusions. The difficulty of obtaining closed-form likelihood function has led to likelihood approximation using simulation (Pedersen (1995), Brandt and Santa-Clara (2002)) or Fourier inversion of the characteristic function (Singleton (2001), Aït-Sahalia and Yu (2002)). Generalized method of moments estimation (Hansen and Scheinkman (1995), Kessler and Sørensen (1999), Duffie and Glynn (2004), and Carrasco, Chernov, Florens and Ghysels (2006)) has also been proposed to sidestep likelihood evaluation. Nonparametric and semiparametric estimations have also been proposed for diffusions (Aït-Sahalia (1996), Bandi and Phillips (2003), Kristensen(2004)) and for jump-diffusions (Bandi and Nguyen (2003)).

The first step towards a *closed-form* likelihood approximation is taken by Aït-Sahalia (2002) who provides a likelihood expansion for univariate diffusions. Such closed-form approximations are shown to be extremely accurate and are fast to compute by Monte-Carlo studies (see Jensen and Poulsen (2002)). The method was subsequently refined by Bakshi and Ju (2005) and Bakshi, Ju and Ou-Yang (2005) under the same setup of univariate diffusions and was extended to multivariate diffusions by Aït-Sahalia (2001) and univariate Levy-driven processes by Schaumburg (2001).

Building on this closed-form approximation approach, this paper provides a closed-form approximation of the likelihood function of multivariate jump-diffusion processes. It extends Aït-Sahalia (2002) and Aït-Sahalia (2001) by constructing an alternative form of leading term that captures the jump behavior. The approximate likelihood function is then solved from Kolmogorov equations. It extends Schaumburg (2001) by relaxing the i.i.d. property inherent in Levy-driven randomness and by addressing multivariate processes. The maximum-likelihood estimators using the approximate likelihood function provided in this paper are shown to achieve the asymptotic efficiency of the true yet *uncomputable* maximum-likelihood estimators as the sampling interval shrinks for a fixed order of

expansion. This differs from Aït-Sahalia (2001) where the asymptotic results are obtained when the order of expansion increases holding fixed the sampling interval. Such infill (small sampling interval) and long span asymptotics have been employed in nonparametric estimations (see for example Bandi and Phillips (2003) and Bandi and Nguyen (2003), among others). This sampling scheme allows the extension of likelihood expansion beyond univariate diffusions and reducible multivariate diffusions to the more general jump diffusions (see section 3). It also permits the use of lower order approximations which is further helped by the numerical evidence that even a low order approximation can be very accurate for typically sampling intervals, e.g. daily or weekly data (see section 2.4.4). The accuracy of the approximation improves rapidly as higher order correction terms are added or as the sampling interval shrinks. One should caution however and not take the asymptotic results literally by estimating a jump-diffusion model aimed at capturing longer-run dynamics using ultra high frequency data. With ultra high frequency data, one encounters market microstructure effects and the parametric model becomes misspecified (see for example Aït-Sahalia, Mykland and Zhang (2005) and Bandi and Russell (2006)). Estimation at such frequencies as daily or weekly can strike a sensible balance between asymptotic requirements and market microstructure effects.

The theory on closed-form likelihood approximation maintains the assumption of observable state variables. Nonetheless, these methods can still be applied to models with unobservable states if additional data series are available such as the derivative price series in the case of stochastic volatility models. In these cases, one may be able to create a mapping between the unobservable volatility variable and the derivative price and conduct inference using the approximate likelihood of the observable states and the unobservable states mapped from the derivative prices. Please see Aït-Sahalia and Kimmel (2005), Pan (2002) and Ledoit, Santa-Clara and Yan (2002) for more details on this approach. Alternatively, one can use methods such as Efficient Method of Moments (EMM), Indirect Inference, and Markov Chain Monte Carlo (MCMC) to handle processes with unobservable states, see for example Gallant and Tauchen (1996), Gouriéroux, Monfort and Renault (1993), and Elerian, Chib and Shephard (2001). Andersen, Benzoni and Lund (2002) provides more details on alternative estimation methods for jump diffusions with stochastic volatilities.

Applying the proposed estimation method, the second part of the paper studies the realignment probability of the Chinese Yuan. The Yuan has been essentially pegged to the US Dollar for the past seven years. A recent export boom has led to diplomatic pressure on China to allow the Yuan to appreciate. The realignment risk is important to both China and other parts of the world. Foreign trade volume accounts for roughly 40 percent of China's GDP in 2002.¹ Any shift in the terms of trade and any changes in the competitive advantage of exporters brought by currency fluctuation can have a significant impact on Chinese economy. Between 1992 and 1999, Foreign Direct Investment (FDI) into China accounted for 8.2 percent of worldwide FDI and 26.3 percent of FDI going into developing countries (see Huang (2003)), all of which can be subject to currency risks. Beginning in August 2003, China started to open its domestic financial markets to foreign investors through Qualified Foreign Institutional Investors and is planning to allow Chinese citizens to invest in foreign financial markets. Currency risk will be important for these investors of financial markets, too.

This paper uncovers the term structure of the *forward realignment rate* which completely characterizes future realignment probabilities.² The term structure is hump-shaped and peaks at six months from the end of 2003. This implies the financial market is anticipating an appreciation of the Yuan in the next year and, conditioning on no realignments in that period, the chance of a realignment is perceived to be small in the further future. Since February 2002, the upward realignment intensity for the Yuan implicit in the financial market has increased fivefold. The realignment probability

¹Data provided by Datastream International.

²The forward realignment rate is defined in section 4.2.1.

responds quickly to news releases on Sino-US trade surplus, state-owned enterprise reform, Chinese government tax revenue and, most importantly, both domestic and foreign government interventions.

The paper is organized as follows. Section 2 provides the likelihood approximation. Likelihood estimation using the approximate likelihood is discussed in Section 3. Section 4 studies the realignment probability of the Chinese Yuan and section 5 concludes. Appendix 1 collects all the technical assumptions. Appendix 2 contains the proofs. Appendix 3 extends the approximation method. Appendix 4 provides the history of the Yuan's exchange rate regime. Important news releases that influence the realignment probability are documented in Appendix 5.

2 Likelihood Approximation

2.1 Multivariate jump-diffusion process

We consider a multivariate jump-diffusion process X defined on a probability space $(\Omega, \mathfrak{F}, P)$ with filtration $\{\mathfrak{F}_t\}$ satisfying the usual conditions (see, for example, Protter (1990)).

$$dX_t = \mu(X_t, \theta) dt + \sigma(X_t, \theta) dW_t + J_t dN_t \quad (2.1)$$

where X_t is an n -dimensional state vector and W_t is a standard d -dimensional Brownian motion. $\theta \in \mathbb{R}^p$ is a finite dimensional parameter. $\mu(\cdot, \theta) : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is the drift and $\sigma(\cdot, \theta) : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times d}$ is the diffusion. The pure jump process N has stochastic intensity $\lambda(X_t, \theta)$ and jump size 1.³ The jump size J_t is independent of \mathfrak{F}_{t-} and has probability density $\nu(\cdot, \theta) : \mathbb{R}^n \rightarrow \mathbb{R}$ with support $C \in \mathbb{R}^n$. In this section, we consider the case where C has a non-empty interior in \mathbb{R}^n .⁴ $V(x, \theta) \equiv \sigma(x, \theta) \sigma(x, \theta)^T$ is the variance matrix, $\sigma(x, \theta)^T$ denotes the matrix transposition of $\sigma(x, \theta)$.

Equivalently, X is a Markov process with infinitesimal generator A^B , defined on bounded function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ with bounded and continuous first and second derivatives, given by

$$A^B f(x) = \sum_{i=1}^n \mu_i(x, \theta) \frac{\partial}{\partial x_i} f(x) + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n v_{ij}(x, \theta) \frac{\partial^2}{\partial x_i \partial x_j} f(x) + \lambda(x, \theta) \int_C [f(x+c) - f(x)] \nu(c, \theta) dc \quad (2.2)$$

where $\mu_i(x, \theta)$, $v_{ij}(x, \theta)$ and x_i are, respectively, elements of $\mu(x, \theta)$, $V(x, \theta)$ and x .⁵

Definition. *The transition probability density $p(\Delta, y|x, \theta)$, when it exists, is the conditional density of $X_{t+\Delta} = y \in \mathbb{R}^n$ given $X_t = x \in \mathbb{R}^n$*

To save notation, the dependence on θ of the functions $\mu(\cdot, \theta)$, $\sigma(\cdot, \theta)$, $V(\cdot, \theta)$, $\lambda(\cdot, \theta)$, $\nu(\cdot, \theta)$ and $p(\Delta, y|x, \theta)$ will not be made explicit when there is no confusion.

Proposition 1. *Under assumptions 1–2, the transition density satisfies the backward and forward Kolmogorov equations given by*

$$\frac{\partial}{\partial \Delta} p(\Delta, y|x) = A^B p(\Delta, y|x) \quad (2.3)$$

$$\frac{\partial}{\partial \Delta} p(\Delta, y|x) = A^F p(\Delta, y|x). \quad (2.4)$$

³See page 27-28 of Brémaud (1981) for definition of stochastic intensity.

⁴This implies, when N jumps, all the state variables can jump. Cases where some state variables do not jump, together with some other extensions, are considered in Appendix 3.

⁵See, for example, Revuz and Yor (1999) for details on infinitesimal generators.

(2.3) and (2.4) are, respectively, the backward and forward equations. The infinitesimal generators A^B and A^F are defined as

$$\begin{aligned} A^B p(\Delta, y|x) &= L^B p(\Delta, y|x) + \lambda(x) \int_C [p(\Delta, y|x+c) - p(\Delta, y|x)] \nu(c) dc \\ A^F p(\Delta, y|x) &= L^F p(\Delta, y|x) + \int_C [\lambda(y-c) p(\Delta, y-c|x) - \lambda(y) p(\Delta, y|x)] \nu(c) dc. \end{aligned}$$

The operators L^B and L^F are given by

$$\begin{aligned} L^B p(\Delta, y|x) &= \sum_{i=1}^n \mu_i(x) \frac{\partial}{\partial x_i} p(\Delta, y|x) + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n v_{ij}(x) \frac{\partial^2}{\partial x_i \partial x_j} p(\Delta, y|x) \\ L^F p(\Delta, y|x) &= - \sum_{i=1}^n \frac{\partial}{\partial y_i} [\mu_i(y) p(\Delta, y|x)] + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2}{\partial y_i \partial y_j} [v_{ij}(y) p(\Delta, y|x)]. \end{aligned} \quad (2.5)$$

2.2 Method of transition density approximation

In this paper, we will find a closed-form approximation to $p(\Delta, y|x)$ using the backward and forward equations. Specifically, we conjecture

$$p(\Delta, y|x) = \Delta^{-n/2} \exp \left[-\frac{C^{(-1)}(x, y)}{\Delta} \right] \sum_{k=0}^{\infty} C^{(k)}(x, y) \Delta^k + \sum_{k=1}^{\infty} D^{(k)}(x, y) \Delta^k \quad (2.6)$$

for some function $C^{(k)}(x, y)$ and $D^{(k)}(x, y)$ to be determined. We then plug (2.6) into the backward and forward equations, match the terms with the same orders of Δ or $\Delta \exp \left[-\frac{C^{(-1)}(x, y)}{\Delta} \right]$, set their coefficients to 0 and solve for $C^{(k)}(x, y)$ and $D^{(k)}(x, y)$. An approximation of order $m > 0$ is obtained by ignoring terms of higher orders.

$$p^{(m)}(\Delta, y|x) = \Delta^{-n/2} \exp \left[-\frac{C^{(-1)}(x, y)}{\Delta} \right] \sum_{k=0}^m C^{(k)}(x, y) \Delta^k + \sum_{k=1}^m D^{(k)}(x, y) \Delta^k. \quad (2.7)$$

The transition density further satisfies the following two conditions.

Condition 1. $C^{(-1)}(x, y) = 0$ if and only if $x = y$.

Condition 2. $C^{(0)}(x, x) = (2\pi)^{-n/2} [\det V(x)]^{-1/2}$.

When $\Delta \rightarrow 0$, condition 1 guarantees the transition density peaks at x and condition 2 comes from the requirement that the density integrates to one with respect to y as Δ becomes small (see Appendix 2 for proof).

The approximation is a small- Δ approximation in that it does not need more correction terms to deliver better approximation for fixed Δ . Instead, for a fixed number of correction terms, the approximation gets better as Δ shrinks, much like Taylor expansion around $\Delta = 0$.

Before calculating $C^{(k)}(x, y)$ and $D^{(k)}(x, y)$, we discuss why (2.6) is the right form of approximation to consider. Intuitively, the first term in (2.6) captures the behavior of $p(\Delta, y|x)$ at y near x where the diffusion term dominates and the second term captures the tail behavior corresponding to jumps. Let $A_{t, \Delta}$ denote the event of no jumps between time t and $t + \Delta$. $A_{t, \Delta}^c$ denotes set

complementation.

$$p(\Delta, y|x) = \Pr(A_{t,\Delta}|X_t = x) \text{pdf}(X_{t+\Delta} = y|X_t = x, A_{t,\Delta}) + \Pr(A_{t,\Delta}^c|X_t = x) \text{pdf}(X_{t+\Delta} = y|X_t = x, A_{t,\Delta}^c) \quad (2.8)$$

The Poisson arrival rate implies the second term is $O(\Delta)$ as $\Delta \rightarrow 0$. This gives the second term in (2.6).

Conditioning on no jumps, $\text{pdf}(X_{t+\Delta} = y|X_t = x, A_{t,\Delta})$ is the transition density for a diffusion. As shown in Varadhan (1967 Theorem 2.2) (see also Section 5.1 in Ait-Sahalia 2002),

$$\lim_{\Delta \rightarrow 0} [-2\Delta \log \text{pdf}(X_{t+\Delta} = y|X_t = x, A_{t,\Delta})] = d^2(x, y)$$

for some function $d^2(x, y)$.⁶ Let $C^{(-1)}(x, y) = \frac{1}{2}d^2(x, y)$, we obtain the leading term $\exp\left[-\frac{C^{(-1)}(x, y)}{\Delta}\right]$ in (2.6). It is pre-multiplied by $\Delta^{-n/2}$ because, as $\Delta \rightarrow 0$, the density at $y = x$ goes to infinity at the same speed as a standard n -dimensional normal density with variance of order $O(\Delta)$ in light of the driving Brownian motion. $\sum_{k=0}^m C^{(k)}(x, y) \Delta^k$ corrects for the fact that Δ is not 0. When there is no jump, the expansion coincides with that in Ait-Sahalia (2001).

2.3 Closed-form expression of the approximate transition density

Theorems 1 and 2 in this section give a set of restrictions on $C^{(k)}(x, y)$ and $D^{(k)}(x, y)$ imposed by the forward and backward equations which, together with conditions 1 and 2, can be used to solve for the approximate transition density (2.7). Corollaries 1 and 2 give explicit expressions of $C^{(k)}(x, y)$ and $D^{(k)}(x, y)$ in the univariate case.

Theorem 1. *The backward equation imposes the following restrictions,*

$$\begin{aligned} 0 &= C^{(-1)}(x, y) - \frac{1}{2} \left[\frac{\partial}{\partial x} C^{(-1)}(x, y) \right]^T V(x) \left[\frac{\partial}{\partial x} C^{(-1)}(x, y) \right] \\ 0 &= C^{(0)} \left[L^B C^{(-1)} - \frac{n}{2} \right] + \left[\frac{\partial}{\partial x} C^{(-1)}(x, y) \right]^T V(x) \left[\frac{\partial}{\partial x} C^{(0)}(x, y) \right] \\ 0 &= C^{(k+1)} \left[L^B C^{(-1)} + (k+1) - \frac{n}{2} \right] + \left[\frac{\partial}{\partial x} C^{(-1)}(x, y) \right]^T V(x) \left[\frac{\partial}{\partial x} C^{(k+1)}(x, y) \right] \\ &\quad + [\lambda(x) - L^B] C^{(k)} \quad \text{for nonnegative } k \\ 0 &= D^{(1)} - \lambda(x) \nu(y - x) \\ 0 &= D^{(k+1)} - \frac{1}{1+k} \left[A^B D^{(k)} + (2\pi)^{n/2} \lambda(x) \sum_{r=0}^k \frac{1}{(2r)!} \sum_{s \in S_{2r}^n} M_s^n \frac{\partial^s}{\partial w^s} g_{k-r}(x, y, w) \Big|_{w=0} \right] \quad \text{for } k > 0 \end{aligned}$$

where $g_k(x, y, w) \equiv \frac{C^{(k)}(w_B^{-1}(w), y) \nu(w_B^{-1}(w) - x)}{\left| \det \frac{\partial}{\partial u^T} w_B(u, y) \Big|_{u=w_B^{-1}(w)} \right|}$, $w_B(x, y) \equiv [V^{1/2}(x)]^T \frac{\partial}{\partial x} C^{(-1)}(x, y)$. Fixing y , $w_B(\cdot, y)$

is invertible in a neighborhood of $x = y$ and $w_B^{-1}(\cdot)$ is its inverse function in this neighborhood. (For the ease of notation, the dependence of $w_B^{-1}(\cdot)$ on y is not made explicit henceforward.)

⁶ $d^2(x, y)$ is the square of the shortest distance from x to y measured by the Riemannian metric defined locally as $ds^2 = dx^T \cdot V^{-1}(x) \cdot dx$ where $V^{-1}(x)$ is the matrix inverse to $V(x)$, dx is the vector $(dx_1, dx_2, \dots, dx_n)^T$.

S_r^n in the theorem denote the set of n -tuple non-negative even integers (s_1, s_2, \dots, s_n) with the property that $|(s_1, s_2, \dots, s_n)| \equiv s_1 + s_2 + \dots + s_n = r$. For $s \in S_r^n$ and $x \in \mathbb{R}^n$, $x^s \equiv x_1^{s_1} x_2^{s_2} \dots x_n^{s_n}$ and differentiation of a function $q(\cdot) : \mathbb{R}^n \rightarrow \mathbb{R}$ is denoted $\frac{\partial^s}{\partial x^s} q(x) \equiv \frac{\partial^{|s|}}{\partial x_1^{s_1} \dots \partial x_n^{s_n}} q(x)$. M_s^n denotes the s -th moment of n -dimensional standard normal distribution given by $M_s^n \equiv (2\pi)^{-n/2} \int_{\mathbb{R}^n} \exp\left[-\frac{w^T w}{2}\right] w^s dw$. $\det(\cdot)$ is the determinant of a square matrix. For ease of notation, we sometimes use $C^{(k)}$ and $D^{(k)}$ for $C^{(k)}(x, y)$ and $D^{(k)}(x, y)$. $L^B C^{(-1)}(x, y)$ is defined the same way as in (2.5) with $C^{(-1)}(x, y)$ replacing $p(\Delta, y|x)$. Similar use extends to L^F , A^B and A^F .

Theorem 2. *The forward equation imposes the following restrictions,*

$$\begin{aligned}
0 &= C^{(-1)}(x, y) - \frac{1}{2} \left[\frac{\partial}{\partial y} C^{(-1)}(x, y) \right]^T V(y) \left[\frac{\partial}{\partial y} C^{(-1)}(x, y) \right] \\
0 &= C^{(0)} \left[-\sum_{i=1}^n \mu_i(y) \frac{\partial}{\partial y_i} C^{(-1)} + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n H_{ij}(x, y) - \frac{n}{2} \right] + \left[\frac{\partial}{\partial y} C^{(-1)}(x, y) \right]^T V(y) \left[\frac{\partial}{\partial y} C^{(0)}(x, y) \right] \\
0 &= C^{(k+1)} \left[-\sum_{i=1}^n \mu_i(y) \frac{\partial}{\partial y_i} C^{(-1)} + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n H_{ij}(x, y) - \frac{n}{2} + (k+1) \right] \\
&\quad + \left[\frac{\partial}{\partial y} C^{(-1)}(x, y) \right]^T V(y) \left[\frac{\partial}{\partial y} C^{(k+1)}(x, y) \right] + [\lambda(y) - L^F] C^{(k)} \quad \text{for nonnegative } k \\
0 &= D^{(1)} - \lambda(x) \nu(y - x) \\
0 &= D^{(k+1)} - \frac{1}{1+k} \left[A^F D^{(k)} + (2\pi)^{n/2} \sum_{r=0}^k \frac{1}{(2r)!} \sum_{s \in S_{2r}^n} M_s^n \frac{\partial^s}{\partial w^s} [\lambda(w_F^{-1}(w)) h_{k-r}(x, y, w)] \Big|_{w=0} \right] \\
&\quad \text{for } k \text{ positive}
\end{aligned}$$

where $H_{ij}(x, y) \equiv \left[\frac{\partial}{\partial y_i} v_{ij}(y) \right] \left[\frac{\partial}{\partial y_j} C^{(-1)}(x, y) \right] + \left[\frac{\partial}{\partial y_j} v_{ij}(y) \right] \left[\frac{\partial}{\partial y_i} C^{(-1)}(x, y) \right] + v_{ij}(y) \frac{\partial^2}{\partial y_i \partial y_j} C^{(-1)}(x, y)$, $h_k(x, y, w) \equiv \frac{C^{(k)}(x, w_F^{-1}(w)) \nu(y - w_F^{-1}(w))}{\left| \det \frac{\partial}{\partial u^T} w_F(x, u) \Big|_{u=w_F^{-1}(w)} \right|}$, $w_F(x, y) \equiv [V^{1/2}(y)]^T \frac{\partial}{\partial y} C^{(-1)}(x, y)$. Fixing x , $w_F(x, \cdot)$ is invertible in a neighborhood of $y = x$ and $w_F^{-1}(\cdot)$ is its inverse function in this neighborhood. (For the ease of notation, the dependence of $w_F^{-1}(\cdot)$ on x is not made explicit henceforward.)

The first equation in either theorem characterizes $C^{(-1)}(x, y)$. Knowing $C^{(-1)}(x, y)$, the second equation can be solved for $C^{(0)}(x, y)$. $C^{(k+1)}(x, y)$ is then solved recursively through the third equation. The last two equations give $D^{(k)}(x, y)$ which do not require solving differential equations.

Implementation. *To compute the approximate transition density, we need to use Conditions 1 and 2, together with either Theorem 1 or Theorem 2. Regarding the inverse functions $w_B^{-1}(w)$ and $w_F^{-1}(w)$ in Theorems 1 and 2, only the values of these inverse functions at $w = 0$ are required to evaluate the transition density. At $w = 0$, $w_B^{-1}(w = 0) = y$ and $w_F^{-1}(w = 0) = x$. The derivatives involving the inverse function can be calculated using the implicit function theorem (see Rudin(1976)), $\frac{\partial}{\partial w^T} w_B^{-1}(w) = \left(\frac{\partial}{\partial x^T} w_B(x, y) \Big|_{x=w_B^{-1}(w)} \right)^{-1}$ and $\frac{\partial}{\partial w^T} w_F^{-1}(w) = \left(\frac{\partial}{\partial y^T} w_F(x, y) \Big|_{y=w_F^{-1}(w)} \right)^{-1}$.*

In the univariate case, $n = 1$, the functions $C^{(k)}(x, y)$ and $D^{(k)}(x, y)$ can be solved explicitly.

Corollary 1. *Univariate case. From the backward equation,*

$$\begin{aligned}
C^{(-1)}(x, y) &= \frac{1}{2} \left[\int_x^y \sigma(s)^{-1} ds \right]^2 \\
C^{(0)}(x, y) &= \frac{1}{\sqrt{2\pi}\sigma(y)} \exp \left[\int_x^y \frac{\mu(s)}{\sigma^2(s)} - \frac{\sigma'(s)}{2\sigma(s)} ds \right] \\
C^{(k+1)}(x, y) &= - \left[\int_x^y \sigma(s)^{-1} ds \right]^{-(k+1)} \int_x^y \left\{ \frac{\exp \left[\int_s^x \frac{\sigma'(u)}{2\sigma(u)} - \frac{\mu(u)}{\sigma^2(u)} du \right] \sigma(s)^{-1}}{\left[\int_s^y \sigma(u)^{-1} du \right]^k} [\lambda(s) - L^B] C^{(k)}(s, y) \right\} ds \text{ for } k \geq 0 \\
D^{(1)}(x, y) &= \lambda(x) \nu(y-x) \\
D^{(k+1)}(x, y) &= \frac{1}{1+k} \left[A^B D^{(k)}(x, y) + \sqrt{2\pi} \lambda(x) \sum_{r=0}^k \frac{M_{2r}^1}{(2r)!} \frac{\partial^{2r}}{\partial w^{2r}} g_{k-r}(x, y, w) \Big|_{w=0} \right] \text{ for } k > 0
\end{aligned}$$

where $g_k(x, y, w) \equiv C^{(k)}(w_B^{-1}(w), y) \nu(w_B^{-1}(w) - x) \sigma(w_B^{-1}(w))$, $M_{2r}^1 \equiv \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp\left(-\frac{s^2}{2}\right) s^{2r} ds$ and $w_B(x, y) = \int_x^y \sigma(s)^{-1} ds$.

Corollary 2. *Univariate case. From forward equation,*

$$\begin{aligned}
C^{(-1)}(x, y) &= \frac{1}{2} \left[\int_x^y \sigma(s)^{-1} ds \right]^2 \\
C^{(0)}(x, y) &= \frac{1}{\sqrt{2\pi}\sigma(x)} \exp \left[\int_x^y \frac{\mu(s)}{\sigma^2(s)} - \frac{3\sigma'(s)}{2\sigma(s)} ds \right] \\
C^{(k+1)}(x, y) &= - \left[\int_x^y \sigma(s)^{-1} ds \right]^{-(k+1)} \int_x^y \left\{ \frac{\exp \left[\int_y^s \frac{3\sigma'(u)}{2\sigma(u)} - \frac{\mu(u)}{\sigma^2(u)} du \right] \sigma(s)^{-1}}{\left[\int_x^s \sigma(u)^{-1} du \right]^k} [\lambda(s) - L^F] C^{(k)}(x, s) \right\} ds \text{ for } k \geq 0 \\
D^{(1)}(x, y) &= \lambda(x) \nu(y-x) \\
D^{(k+1)}(x, y) &= \frac{1}{1+k} \left[A^F D^{(k)}(x, y) + \sqrt{2\pi} \sum_{r=0}^k \frac{M_{2r}^1}{(2r)!} \frac{\partial^{2r}}{\partial w^{2r}} [\lambda(w_F^{-1}(w)) h_{k-r}(x, y, w)] \Big|_{w=0} \right] \text{ for } k > 0
\end{aligned}$$

where $h_k(x, y, w) \equiv C^{(k)}(x, w_F^{-1}(w)) \nu(y - w_F^{-1}(w)) \sigma(w_F^{-1}(w))$, $M_{2r}^1 \equiv \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp\left(-\frac{s^2}{2}\right) s^{2r} ds$ and $w_F(x, y) = \int_x^y \sigma(s)^{-1} ds$.

The choice of Theorem 1 or Theorem 2 in practice largely depends on computational ease. For example, it is easier to use the results from the forward equation to verify that, in univariate case without jumps, the approximate transition density obtained here coincides with that in Ait-Sahalia (2002). Ait-Sahalia (2001) gives approximation for the log-likelihood of multivariate diffusion. Without jumps, the approximate log-likelihood obtained in this paper takes the form

$$\log p^{(m)}(\Delta, y|x) = -\frac{n}{2} \log \Delta - \frac{C^{(-1)}(x, y)}{\Delta} + \log \sum_{k=0}^m C^{(k)}(x, y) \Delta^k$$

which coincides with the approximation in Ait-Sahalia (2001) when the last term is Taylor expanded around $\Delta = 0$.

2.4 Examples

2.4.1 Brownian motion

Let $dX_t = \sigma dW_t$. The true transition density is $N(0, \sigma^2 \Delta)$. In this case, $C^{(-1)}(x, y) = \frac{(y-x)^2}{2\sigma^2}$, $C^{(0)}(x, y) = \frac{1}{\sqrt{2\pi\sigma^2}}$, $C^{(k)}(x, y) = D^{(k)}(x, y) = 0$ for $k \geq 1$. The approximate density is exact.

$$p^{(m)}(\Delta, y|x) = \frac{1}{\sqrt{2\pi\sigma^2\Delta}} \exp\left[-\frac{(y-x)^2}{2\sigma^2\Delta}\right].$$

2.4.2 Brownian motion with drift

For the process $dX_t = \mu dt + \sigma dW_t$, the true transition density is $N(\mu\Delta, \sigma^2\Delta)$, or equivalently

$$p(\Delta, y|x) = \frac{1}{\sqrt{2\pi\sigma^2\Delta}} \exp\left[-\frac{(y-x)^2}{2\sigma^2\Delta} + \frac{\mu}{\sigma^2}(y-x)\right] \exp\left(-\frac{\mu^2}{2\sigma^2}\Delta\right)$$

We can calculate that $C^{(-1)}(x, y) = \frac{(y-x)^2}{2\sigma^2}$, $C^{(0)}(x, y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{\mu}{\sigma^2}(y-x)\right]$, $C^{(1)}(x, y) = -C^{(0)}(x, y) \frac{\mu^2}{2\sigma^2}$, $D^{(k)}(x, y) = 0$ for all k . The approximate density $p^{(1)}(\Delta, y|x)$ is

$$p^{(1)}(\Delta, y|x) = \frac{1}{\sqrt{2\pi\sigma^2\Delta}} \exp\left[-\frac{(y-x)^2}{2\sigma^2\Delta} + \frac{\mu}{\sigma^2}(y-x)\right] \left(1 - \frac{\mu^2}{2\sigma^2}\Delta\right)$$

which approximates $p(\Delta, y|x)$ by replacing $\exp\left(-\frac{\mu^2}{2\sigma^2}\Delta\right)$ with its first-order Taylor expansion.

2.4.3 Jump-diffusion

Consider the univariate jump-diffusion $dX_t = \mu dt + \sigma dW_t + S_t dN_t$, where N_t is a Poisson process with arrival rate λ . The jump size S_t is i.i.d. $N(\mu_S, \sigma_S^2)$. Conditioning on j jumps, the increment of X is $N(x + \mu\Delta + j\mu_S, \sigma^2\Delta + j\sigma_S^2)$. The true and approximate transition densities are

$$\begin{aligned} p(\Delta, y|x) &= \sum_{j=0}^{\infty} \frac{e^{-\lambda\Delta} (\lambda\Delta)^j}{j!} \frac{1}{\sqrt{2\pi}\sqrt{\sigma^2\Delta + j\sigma_S^2}} \exp\left[-\frac{(y-x-\mu\Delta-j\mu_S)^2}{2(\sigma^2\Delta + j\sigma_S^2)}\right] \\ &= \frac{1}{\sqrt{2\pi\sigma^2\Delta}} \exp\left[-\frac{(y-x)^2}{2\sigma^2\Delta} + \frac{\mu}{\sigma^2}(y-x)\right] \exp\left[-\left(\frac{\mu^2}{2\sigma^2} + \lambda\right)\Delta\right] \\ &\quad + \frac{e^{-\lambda\Delta}\lambda}{\sqrt{2\pi}\sqrt{\sigma^2\Delta + \sigma_S^2}} \exp\left[-\frac{(y-x-\mu\Delta-\mu_S)^2}{2(\sigma^2\Delta + \sigma_S^2)}\right] \Delta + O(\Delta^2). \end{aligned}$$

$$\begin{aligned} p^{(1)}(\Delta, y|x) &= \frac{1}{\sqrt{2\pi\sigma^2\Delta}} \exp\left[-\frac{(y-x)^2}{2\sigma^2\Delta} + \frac{\mu}{\sigma^2}(y-x)\right] \left[1 - \left(\frac{\mu^2}{2\sigma^2} + \lambda\right)\Delta\right] \\ &\quad + \frac{\lambda}{\sqrt{2\pi\sigma_S^2}} \exp\left[-\frac{(y-x-\mu_S)^2}{2\sigma_S^2}\right] \Delta \end{aligned}$$

Here $C^{(-1)}(x, y) = \frac{(y-x)^2}{2\sigma^2}$, $C^{(0)}(x, y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{\mu}{\sigma^2}(y-x)\right]$, $C^{(1)}(x, y) = -C^{(0)}(x, y) \left(\frac{\mu^2}{2\sigma^2} + \lambda\right)$, $D^{(1)}(x, y) = \frac{\lambda}{\sqrt{2\pi\sigma_S^2}} \exp\left[-\frac{(y-x-\mu_S)^2}{2\sigma_S^2}\right]$. $p(\Delta, y|x)$ is approximated by expanding $\exp\left[-\left(\frac{\mu^2}{2\sigma^2} + \lambda\right)\Delta\right]$ to its first-order around $\Delta = 0$, by approximating $\frac{e^{-\lambda\Delta}}{\sqrt{2\pi}\sqrt{\sigma^2\Delta + \sigma_S^2}} \exp\left[-\frac{(y-x-\mu\Delta-\mu_S)^2}{2(\sigma^2\Delta + \sigma_S^2)}\right]$ with its limit at $\Delta = 0$ and by ignoring the terms corresponding to at least two jumps which are of order Δ^2 .

2.4.4 A Numerical Example

Consider the Ornstein-Uhlenbeck process with jump $dX_t = -kX_t dt + \sigma dW_t + S_t dN_t$ where $k, \sigma > 0$, N is a standard Poisson process with constant intensity λ . S_t is i.i.d. and has double exponential distribution with mean 0 and standard deviation σ_S , which has a fatter tail than normal distribution.

X is an affine process whose characteristic function is known according to Duffie et al (2000).⁷ An approximate transition density, $p^{FFT}(\Delta, X_\Delta|X_0)$, can be obtained via fast Fourier transform. Treating $\log p^{FFT}(\Delta, X_\Delta|X_0)$ as the ‘‘true’’ log-likelihood, we now investigate numerically the accuracy of the closed-form log-likelihood approximation and how the accuracy varies with the sampling interval Δ and the order of the approximation.⁸

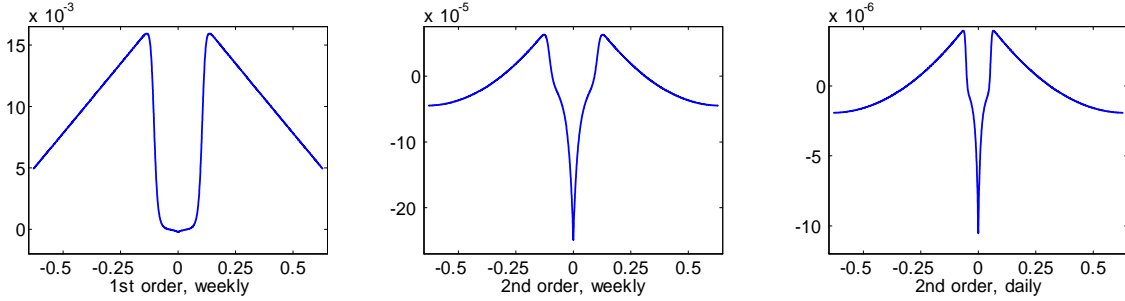


Figure 1: Approximation Error of Log-Likelihood

The first and the second graphs in Figure 1 show the effect of adding correction terms. The first graph uses first order approximation while the second uses second order approximation. Both graphs correspond to weekly sampling. The accuracy of the approximation increases rapidly with additional terms. For weekly sampling, the standard deviation of $X_\Delta|X_0 = 0$ is 0.03. The first two graphs in fact show the log-likelihood approximation is good from negative twenty to positive twenty standard deviations. That the approximation is good in the large deviation area is useful in case a rare event occurs in the observations.

The second and the third graphs in Figure 1 show the effect of shrinking sampling interval. The third graph uses daily instead of weekly sampling in the second graph. Both graphs use second order approximation. The approximation improves quickly as sampling interval shrinks.

2.5 Calculate $C^{(-1)}(x, y)$

The equations characterizing $C^{(-1)}(x, y)$ in Section 2.3 are, in multivariate cases, PDEs which do not always have explicit solutions. In this section, we will discuss the conditions under which they can be

⁷ $E\left[e^{i u X_\Delta} | X_0 = x\right] = \exp\left[\mathbf{i} u x e^{-k\Delta} + \frac{(e^{-2k\Delta} - 1)u^2 \sigma^2}{4k}\right] \left(\frac{1 + e^{-2k\Delta} u^2 \sigma_S^2}{1 + u^2 \sigma_S^2}\right)^{\frac{\lambda}{2k}}$

⁸ The parameters are set to $k = 0.5$, $\sigma = 0.2$, $\lambda = 1/3$, $\mu_S = 0$, $\sigma_S = 0.2$, $X_0 = 0$.

solved explicitly. When they cannot be explicitly solved, we will discuss how to find an approximate solution with a second expansion in the state space.

2.5.1 The reducible case

Consider the equation characterizing $C^{(-1)}(x, y)$ in Theorem 1.

$$C^{(-1)}(x, y) = \frac{1}{2} \left[\frac{\partial}{\partial x} C^{(-1)}(x, y) \right]^T V(x) \left[\frac{\partial}{\partial x} C^{(-1)}(x, y) \right].$$

Remember $w_B(x, y)$ is defined as $w_B(x, y) \equiv [V^{1/2}(x)]^T \frac{\partial}{\partial x} C^{(-1)}(x, y)$ in Theorem 1, we have

$$\begin{aligned} C^{(-1)}(x, y) &= \frac{1}{2} w_B^T(x, y) w_B(x, y) \\ \frac{\partial}{\partial x} C^{(-1)}(x, y) &= \left[\frac{\partial}{\partial x} w_B^T(x, y) \right] w_B(x, y). \end{aligned}$$

Therefore,

$$w_B(x, y) = [V^{1/2}(x)]^T \left[\frac{\partial}{\partial x} w_B^T(x, y) \right] w_B(x, y). \quad (2.9)$$

Let I_n be the n -dimensional identity matrix. It is easy to see $[V^{1/2}(x)]^T \left[\frac{\partial}{\partial x} w_B^T(x, y) \right] = I_n$ characterizes a solution for $w_B(x, y)$ and hence a solution for $C^{(-1)}(x, y)$. Let $V^{-1/2}(x)$, whose element in the i -th row and j -th column is denoted $v_{ij}^{-1/2}(x)$, be the inverse matrix of $V^{1/2}(x)$. The condition for the existence of a vector function $w_B(x, y)$ whose first derivative $\frac{\partial}{\partial x} w_B(x, y) = V^{-1/2}(x)$ is, intuitively, that the second derivative matrix of each element of $w_B(x, y)$ is symmetric,

$$\frac{\partial}{\partial x_k} v_{ij}^{-1/2}(x) = \frac{\partial}{\partial x_j} v_{ik}^{-1/2}(x) \quad \text{for all } i, j, k = 1, \dots, n. \quad (2.10)$$

This is the reducibility condition given by Proposition 1 in Ait-Sahalia (2001). All the univariate cases are reducible. Therefore, it is not surprising that $C^{(-1)}(x, y)$ can be solved explicitly for the univariate case in Corollaries 1 and 2.

2.5.2 The irreducible case

When the reducibility condition (2.10) does not hold, we can no longer solve (2.9) for $C^{(-1)}(x, y)$ explicitly in general. In this case, we will approximate $C^{(-1)}(x, y)$ instead. To approximate $C^{(-1)}(x, y)$, we notice that the derivatives of $C^{(-1)}(x, y)$ are needed to compute $C^{(0)}(x, y)$ whose derivatives will in turn be used to compute $C^{(1)}(x, y)$ and so on. So we have to approximate not only $C^{(-1)}(x, y)$ but also its derivatives of various orders. A natural method of approximation is therefore Taylor expansion of $C^{(-1)}(x, y)$ in the state space proposed by Ait-Sahalia (2001). We will Taylor expand $C^{(-1)}(x, y)$ around $y = x$ with coefficients to be determined, plug the expansion into the differential equation in Theorem 1 or Theorem 2, match and set to 0 the coefficients of terms with the same orders of $y - x$, and solve for the Taylor expansion of $C^{(-1)}(x, y)$. This expansion is discussed in detail in Ait-Sahalia (2001) (see especially Theorem 2 and the remarks thereafter).

We will first give an example to illustrate this second expansion in the state space and then discuss how to determine the order of the Taylor expansion. We will use $C^{(-1, J)}(x, y)$ to denote

the expansion of $C^{(-1)}(x, y)$ to order J and use $C^{(k, J)}(x, y)$ for $k \geq 0$ to denote other coefficients computed from using $C^{(-1, J)}(x, y)$. To make the error of $p^{(m)}(\Delta, y|x)$ have the correct order, J will depend on m . But for notational convenience, this dependence is not written out explicitly.

As an example, we take the equation characterizing $C^{(-1)}(x, y)$ in Theorem 2 and seek a second order expansion $C^{(-1, 2)}(x, y)$. Taylor expand this equation around $y = x$ and notice Condition 1,

$$\begin{aligned} & \frac{\partial}{\partial y} C^{(-1)}(x, y) \Big|_{y=x} \cdot (y-x) + \frac{1}{2} (y-x)^T \cdot \frac{\partial^2}{\partial y^2} C^{(-1)}(x, y) \Big|_{y=x} \cdot (y-x) + R_1 \\ = & \frac{1}{2} \left[\frac{\partial}{\partial y} C^{(-1)}(x, y) \Big|_{y=x} + \frac{\partial^2}{\partial y^2} C^{(-1)}(x, y) \Big|_{y=x} \cdot (y-x) + R_2 \right]^T \cdot [V(x) + R_3] \\ & \cdot \left[\frac{\partial}{\partial y} C^{(-1)}(x, y) \Big|_{y=x} + \frac{\partial^2}{\partial y^2} C^{(-1)}(x, y) \Big|_{y=x} \cdot (y-x) + R_2 \right] \end{aligned}$$

where R_1 , R_2 and R_3 are remainder terms that will not affect the result.

Match the terms with the same orders of $(y-x)$ and set their coefficients to 0. We get

$$\frac{\partial}{\partial y} C^{(-1)}(x, y) \Big|_{y=x} = 0, \quad \frac{\partial^2}{\partial y^2} C^{(-1)}(x, y) \Big|_{y=x} = V^{-1}(x)$$

The second order expansion of $C^{(-1)}(x, y)$ around $y = x$ is therefore

$$C^{(-1, 2)}(x, y) = \frac{1}{2} (y-x)^T \cdot V^{-1}(x) \cdot (y-x) \quad (2.11)$$

Now we discuss the choice of J . We can achieve an error of $o_p(\Delta^m)$ if each term $C^{(k, J)}(x, y) \Delta^k$ differs from $C^{(k)}(x, y) \Delta^k$ by $o_p(\Delta^m)$. For jump-diffusions, $y-x \sim O_p(\sqrt{\Delta})$. It is clear then to make $C^{(k, J)}(x, y) - C^{(k)}(x, y) = o_p(\Delta^{m-k})$, we need $J \geq 2(m-k)$ for all $k \geq -1$ as shown in equation (5.3) in Ait-Sahalia (2001). Therefore,

$$J = 2(m+1).$$

This choice of J also ensures the relative error incurred by $\sum_{k=1}^m D^{(k)}(x, y) \Delta^k$ is $o_p(\Delta^m)$, which can be verified from Theorems 1 and 2.

For $k \geq 0$, $C^{(k, J)}(x, y)$ are exact solutions to the linear PDEs characterizing them.⁹ However, they only approximate $C^{(k)}(x, y)$ because these PDEs involve $C^{(-1)}(x, y)$ for which only an approximation $C^{(-1, J)}(x, y)$ is available. In the case that these linear PDEs are too cumbersome to solve, we can find an approximate solution $C^{(k, J(k))}(x, y)$ in the same way as the expansion for $C^{(-1)}(x, y)$ with the order of expansion being $J(k) = 2(m-k)$. See also Ait-Sahalia (2001).

3 Likelihood Estimation

To estimate the unknown parameter $\theta \in \mathbb{R}^p$ in the jump diffusion model, we collect observations in a time span T with sampling interval Δ . The sample size is $n_{T, \Delta} = T/\Delta$. Let $\{X_0, X_\Delta, X_{2\Delta}, \dots, X_T\}$

⁹Linear PDEs can be solved using, for example, the method in chapter 6 of Carrier et al (1988).

denote the observations. Let

$$l_{T,\Delta}(\theta) = \sum_{i=1}^{n_{T,\Delta}} \log p(\Delta, X_{i\Delta} | X_{(i-1)\Delta}, \theta)$$

be the log-likelihood function conditioning on X_0 . The loss of information contained in the first observation X_0 is asymptotically negligible. Let $\hat{\theta}_{T,\Delta} \equiv \arg \max_{\theta \in \Theta} l_{T,\Delta}(\theta)$ denote the (uncomputable) estimator from maximum-likelihood estimation (MLE). Let

$$l_{T,\Delta}^{(m)}(\theta) = \sum_{i=1}^{n_{T,\Delta}} \log p^{(m)}(\Delta, X_{i\Delta} | X_{(i-1)\Delta}, \theta)$$

be the m -th order approximate likelihood function. Denote $\hat{\theta}_{T,\Delta}^{(m)} \equiv \arg \max_{\theta \in \Theta} l_{T,\Delta}^{(m)}(\theta)$. Let $I_{T,\Delta}(\theta) \equiv E_{\theta} \left[\dot{l}_{T,\Delta}(\theta) \dot{l}_{T,\Delta}(\theta)^T \right]$ be the information matrix. To simplify notations, $\dot{l}_{T,\Delta}(\theta)$ indicates derivative with respect to θ and the dependence of $\hat{\theta}_{T,\Delta}$ and $\hat{\theta}_{T,\Delta}^{(m)}$ on T and Δ will not be made explicit when there is no confusion. Let $i_{\Delta}(\theta) \equiv E_{\theta} \left[\frac{\partial}{\partial \theta} \log p(\Delta, X_{\Delta} | X_0, \theta) \cdot \frac{\partial}{\partial \theta^T} \log p(\Delta, X_{\Delta} | X_0, \theta) \right]$.

The asymptotics of $\hat{\theta}_{T,\Delta}^{(m)}$ is obtained in two steps. We first establish that the difference between $\hat{\theta}_{T,\Delta}^{(m)}$ and the true MLE estimator $\hat{\theta}$ is negligible compared to the asymptotic distribution of $\hat{\theta}$. The asymptotic distribution of $\hat{\theta}_{T,\Delta}^{(m)}$ can then be deduced from that of $\hat{\theta}$ for the purpose of statistical inference.

Under assumption 6, it can be shown¹⁰ that the true MLE estimator $\hat{\theta}_{T,\Delta}$ satisfies

$$I_{T,\Delta}^{1/2}(\theta_0) \left(\hat{\theta}_{T,\Delta} - \theta_0 \right) = G_{T,\Delta}(\theta_0)^{-1} S_{T,\Delta}(\theta_0) + o_p(1) \quad (3.1)$$

as $T \rightarrow \infty$, uniformly for all $\Delta \leq \bar{\Delta}$, where $G_{T,\Delta}(\theta) \equiv -I_{T,\Delta}^{-1/2}(\theta) \ddot{l}_{T,\Delta}(\theta) I_{T,\Delta}^{-1/2}(\theta)$ and $S_{T,\Delta}(\theta) \equiv I_{T,\Delta}^{-1/2}(\theta) \dot{l}_{T,\Delta}(\theta)$. $G_{T,\Delta}(\theta_0)^{-1} S_{T,\Delta}(\theta_0) = O_p(1)$ as $T \rightarrow \infty$ uniformly for $\Delta \leq \bar{\Delta}$ under very general conditions including but not restricted to stationarity, see for example, Ait-Sahalia (2002) and Jeganathan (1995). $\|I_{T,\Delta}(\theta_0)\|^{-1/2}$ captures the magnitude of the statistical noise in the true MLE estimator. The next theorem shows that the difference between the approximate and the true MLE estimator is negligible compared to $\|I_{T,\Delta}(\theta_0)\|^{-1/2}$ when the sampling interval is small.

Theorem 3. *Under Assumptions 1–7, there exists a sequence $\Delta_T^* \xrightarrow{T \rightarrow \infty} 0$ so that for any $\{\Delta_T\}$ satisfying $\Delta_T \leq \Delta_T^*$,*

$$\left\| I_{T,\Delta_T}^{1/2}(\theta_0) \left(\hat{\theta}_{T,\Delta_T}^{(m)} - \hat{\theta}_{T,\Delta_T} \right) \right\| = o_p(\Delta_T^m)$$

as $T \rightarrow \infty$.

Therefore, the asymptotic distribution of $\hat{\theta}_{T,\Delta_T}^{(m)}$ inherits that of the true MLE estimator. For statistical inference, we provide a primitive condition for stationarity in Proposition 2 below and derive under stationarity the asymptotic distribution of the true MLE estimator (hence the asymptotic distribution of the approximate MLE estimator). Under non-stationarity, primitive condition of sufficient generality is unavailable for the asymptotic distribution of the true MLE estimator to

¹⁰See the proof of Theorem 4.

the author's knowledge, with the exception of Ornstein-Uhlenbeck model (see Aït-Sahalia (2002)). Therefore under non-stationarity, the asymptotic distribution of the true MLE estimator needs to be derived case by case in practice using methods in for example Jeganathan (1995). Note that the main result of the paper – that the approximate MLE estimator inherits the asymptotic distribution of the true MLE estimator as the sampling interval shrinks (Theorem 3) – applies irrespective of stationarity or the asymptotic distribution of the true MLE estimator.

Proposition 2. *The jump-diffusion process X in (2.1) is stationary if there exists a non-negative function $f(\cdot) : \mathbb{R}^n \rightarrow \mathbb{R}$ such that for x in the domain of X ,*

$$\lim_{|x| \rightarrow \infty} A^B f(x) = -\infty \text{ and } \sup_x A^B f(x) < \infty$$

where A^B is the infinitesimal generator defined in (2.2).

This condition is easy to verify. As an example, consider the process X with $\mu(x) = -kx$ for some $k > 0$, $\sigma(x) = s\sqrt{x}$, constant jump intensity, and constant jump distribution with finite variance. We can see this process is stationary by letting $f(x) = x^2$ in Proposition 2 because $A^B f(x) = -2kx^2 + o(x^2)$ when $|x| \rightarrow \infty$.

Theorem 4. *Under Assumptions 1–7, there exists $\bar{\Delta} > 0$ such that $\hat{\theta}_{T,\Delta}$ is consistent as $T \rightarrow \infty$ for all $\Delta \leq \bar{\Delta}$. Further, if X process is stationary, $\hat{\theta}_{T,\Delta}$ has the following asymptotic distribution:*

$$(n_{T,\Delta} i_{\Delta}(\theta_0))^{1/2} (\hat{\theta}_{T,\Delta} - \theta_0) = N(0, I_{p \times p}) + o_p(1)$$

as $T \rightarrow \infty$, uniformly for $\Delta \leq \bar{\Delta}$.

Corollary 3. *Under Assumptions 1–7, there exists a sequence $\Delta_T^* \xrightarrow{T \rightarrow \infty} 0$ so that for any $\{\Delta_T\}$ satisfying $\Delta_T \leq \Delta_T^*$,*

$$(n_{T,\Delta_T} i_{\Delta_T}(\theta_0))^{1/2} (\hat{\theta}_{T,\Delta_T}^{(m)} - \theta_0) = N(0, I_{p \times p}) + o_p(1) + o_p(\Delta_T^m)$$

as $T \rightarrow \infty$.

Therefore, $\hat{\theta}_{T,\Delta}^{(m)}$ inherits the asymptotic property of $\hat{\theta}_{T,\Delta}$ if the sampling interval becomes small. The remainder term in Corollary 3 is of order $o_p(1)$. It is explicitly decomposed into two parts. The first part $o_p(1)$ is the usual approximation error of large sample asymptotics ($T \rightarrow \infty$). The second part $o_p(\Delta_T^m)$ results from using an approximate likelihood estimator. Even a low order (m) of approximation yields the same asymptotic distribution as the true MLE estimator. This benefits from the use of in-fill asymptotics (small Δ) in addition to large sample asymptotics (large T). A small Δ asymptotics is required because the correction term of the approximate density is analytic at $\Delta = 0$ only for univariate diffusions or reducible multivariate diffusions. For these processes, asymptotics can be derived by letting $T \rightarrow \infty$ and letting the number of correction terms $m \rightarrow \infty$ holding Δ fixed as in Aït-Sahalia (2001). However, for irreducible multivariate diffusions (see Aït-Sahalia (2001)) and general jump diffusions, analyticity at $\Delta = 0$ may not hold. Therefore, the likelihood expansion (2.7) is to be interpreted strictly as a Taylor expansion and the asymptotics hold for small Δ holding fixed the order of expansion m . In practice, Δ is not zero and T and m are both finite. The numerical examples in Figure 1 show that even the first order likelihood approximation can be very accurate for weekly data. The accuracy improves rapidly when the order of approximation increases.

4 Realignment Risk of the Yuan

The currency of China is the Yuan. Since 1994, daily movement of the exchange rate between the Yuan and the US Dollar (USD) is limited to 0.3% on either side of the exchange rate published by People’s Bank of China, China’s central bank. Since 2000, the Yuan/USD rate has been in a narrow range of 8.2763 to 8.2799 and is essentially pegged to USD. A recent export boom has led to diplomatic pressures on China to allow the Yuan to appreciate. What is the realignment probability of the Yuan implicit in the financial market? What factors does this probability respond to? These are the questions to be addressed in the rest of the paper.

4.1 Data

“Ironically, what began as a protection against currency devaluation has now become the chief tool for betting on currency appreciation, particularly in China – where an export boom has led to diplomatic pressure to allow the yuan to rise in value.”

—— “Feature – NDFs, the secretive side of currency trading”
Forbes, August 28, 2003

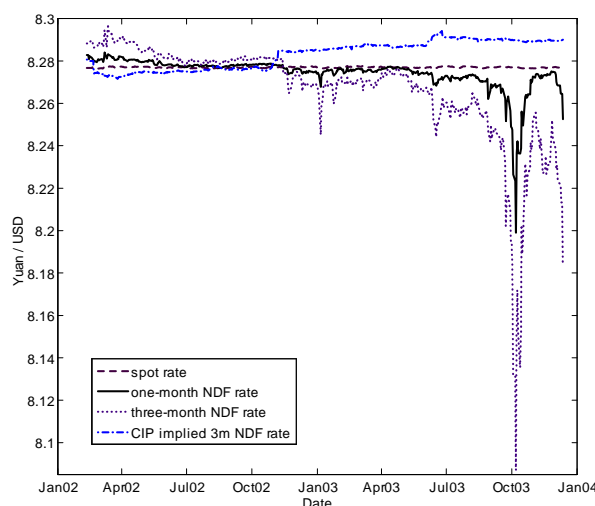


Figure 2: Spot and NDF rates

We obtained daily Non-Deliverable Forward (NDF) rates traded in major off-shore banks. The data, covering February 15, 2002 to December 12, 2003, are sampled by WM/Reuters and are obtained from Datastream International.¹¹

A NDF contract traded in an off-shore bank is the same as a forward contract except that, on the expiration day, no physical delivery of the Yuan takes place. Instead, profit (loss) based on the difference between the NDF rate and the spot exchange rate at maturity is settled in USD. The daily trading volume in the offshore Chinese Yuan NDF market is around 200 million USD.¹²

¹¹Datastream daily series CHYUA\$, USCNY1F and USCNY3F are obtained for the spot Yuan/USD rate, 1 month NDF rate and 3 month NDF rate. All the data are sampled by WM/Reuters at 16:00 hrs London time.

¹²From “Feature – NDFs, the secretive side of currency trading”, Forbes, August 28, 2003.

Figure 2 plots the spot Yuan/USD exchange rate, the one-month NDF rate and the three-month NDF rate. A feature of the plot is the downward crossing over spot rate of both forward rates near the end of year 2002. In early 2002, both forward rates are above the spot rate; however, since late 2002, both forward rates have been lower than the spot rates. This reflects the increasing market expectation of the Yuan’s appreciation, which is confirmed by the estimation results in Section 4.5.

For a currency that can be freely exchanged, the forward price is pinned down by the spot exchange rate and domestic and foreign interest rates (the covered interest rate parity) and does not contain additional information. Due to capital control, the arbitrage argument leading to the covered interest rate parity does not hold for the Yuan. The “CIP implied rate” in Figure 2 is the three-month forward rate implied by the covered interest rate parity if the Yuan were freely traded.¹³ Both the level and the trend of the CIP implied rate differs from that of the NDF rate. Therefore, the NDF rate does provide additional information which will be used in the following study.

4.2 Setup

The spot exchange rate S in USD/Yuan is assumed to follow a pure jump process

$$\frac{dS_t}{S_{t-}} = J_t^u dU_t + J_t^d dD_t$$

where U and D are Poisson processes with arrival rate $\lambda_{u,t}$ and $\lambda_{d,t}$. The jump size J_t^u (J_t^d) is a positive (negative) i.i.d. random variable. Therefore, U and D are associated with upward (the Yuan appreciates) and downward (the Yuan depreciates) realignments, respectively. The time-varying realignment intensities are parametrized as $\lambda_{u,t} = \lambda_u e^{\Gamma_{t-}}$ and $\lambda_{d,t} = \lambda_d e^{-\Gamma_{t-}}$ where Γ_t follows

$$d\Gamma_t = -k\Gamma_t dt + \sigma dW_t + K_t dL_t$$

for some $k, \sigma > 0$. W is a Brownian motion, $K_t \sim N(0, \sigma_K^2)$, L is a Poisson process with arrival rate η . W , L and K are assumed to be independent of other uncertainties in the model.

This specification parsimoniously characterizes how the realignment intensity varies over time. When $\Gamma > 0$, $\lambda_{u,t} > \lambda_u$ and appreciation is more likely than long-run average. When $\Gamma < 0$, $\lambda_{d,t} > \lambda_d$ and depreciation is more likely. The process Γ mean reverts to 0 at which point appreciation and depreciation intensities equate their long-run averages.

Under regularity conditions, the rate F_d of a NDF with maturity d satisfies

$$0 = E^Q \left[e^{-\int_0^d r_s ds} (S_d - F_d) \middle| S_0, \Gamma_0 \right]$$

where r is the short-term USD interest rate and Q is a risk-neutral equivalent martingale measure.

We make a simplifying assumption that the realignment risk premium is zero which allows us to identify the risk-neutral probability Q with the actual probability. Risk-neutral probability consists of actual probability and risk premium. We will mostly be interested in changes in realignment probability. This assumption is therefore an assumption of constant risk premium and setting it to zero just makes the notation simpler. Incorporating time-varying risk-premium requires writing down a formal model which depends on the different parametrization of its time variation. To avoid the results being driven by potential misspecification of the time variation of risk premium, we take the first order approximation by using a constant risk premium. To compute the forward price,

¹³Three-month eurodollar deposit rate and three-month time deposit rate for the Yuan are used. They are obtained from series FREDD3M and CHSRW3M in Datastream International.

we also assume that $e^{-\int_0^d r_s ds}$ and S_d are uncorrelated. The main advantage of this assumption is that it avoids potential misspecification of interest rate dynamics in light of the recent finding that available term structure models do not fit time series variations of interest rates very well (See Duffee (2002)). We will use one-month forward rate in the estimation. When the time-to-maturity d is small, $\int_0^d r_s ds = r_0 d + o(d)$, i.e. changes in interest rates is of smaller order relative to $\int_0^d r_s ds$. The above two assumptions also make the empirical analysis easy to replicate.

Under these two assumptions, the forward pricing formula simplifies to

$$F_d = E[S_d | S_0, \Gamma_0] \quad (4.1)$$

We have performed a robustness check of the model specification. Let $sp_d = F_d - S$ be the spread of a d -maturity forward rate over the spot rate. Let \hat{sp}_d be the spread implied by (4.1) using the estimated parameters. The following regression

$$sp_d = \alpha_d + \beta_d \cdot \hat{sp}_d + \varepsilon_d$$

produced estimates $\hat{\alpha}_d$ close to 0 and estimates $\hat{\beta}_d$ close to 1 for all maturities d longer than one month, which would be the case if the model is correctly specified. The joint hypothesis that $\hat{\beta}_d = 1$ for all d is not rejected. The detailed robustness check results are available upon request.

4.2.1 Term Structure of the Forward Realignment Rate

Let $q(t)$ denote the probability of *no* appreciation before time t .¹⁴ When the Poisson intensity is a constant λ , the probability of no jumps before t is $e^{-\lambda t}$ which converges to one as $t \rightarrow 0$. Under the current setup, $q(t) = E_0 \left[e^{-\int_0^t \lambda_{u,\tau} d\tau} \right]$ which is a stochastic version for the probability of no jumps.¹⁵ Because the jump intensity is positive, $q(t)$ decreases with t and we can find a function $f(t) \geq 0$ such that $q(t) = e^{-\int_0^t f(\tau) d\tau}$. It can be verified that $f(t) = -\frac{q'(t)}{q(t)}$ and that, for $s \geq t$, the probability of no realignment before s conditioning on no realignment before time t (denoted by $q(s|t)$) satisfies

$$q(s|t) = e^{-\int_t^s f(\tau) d\tau}.$$

We call the function f the *forward realignment rate*. Knowing the term structure of $f(t)$ (i.e., its variation with respect to t) gives complete information regarding future realignment probabilities.¹⁶

This paper uses one factor Γ to model the evolution of the realignment intensity. This simple setup already affords a rich set of possibilities for the term structure of f . Figure 3 plots three such possibilities including decreasing, increasing and hump-shaped term structure.

With a decreasing term structure, the market perceives an imminent realignment. Conditioning on no immediate realignment, the chance of a realignment decreases over time. An increasing term structure has the opposite implication: realignment probability is considered small and reverts back to the long-run mean. When the term structure is hump-shaped, the market could be expecting a realignment at a certain time in the future and, if nothing happens by then, the chance of a

¹⁴The term structure of downward realignment rate can be defined in direct parallel. However, it is suppressed since the current interest is in the appreciation of the Yuan.

¹⁵This can be shown using a doubly stochastic argument in Brémaud (1981) by first drawing $\{\lambda_{u,t}\}$ and then determining the jump probability conditioning on $\{\lambda_{u,t}\}$.

¹⁶Readers familiar with duration models will recognize this is the hazard rate of the Yuan's realignment. This concept is also parallel to the forward interest rate in term structure interest rate models and to the forward default rate in credit risk models (see Duffee and Singleton (2003)).

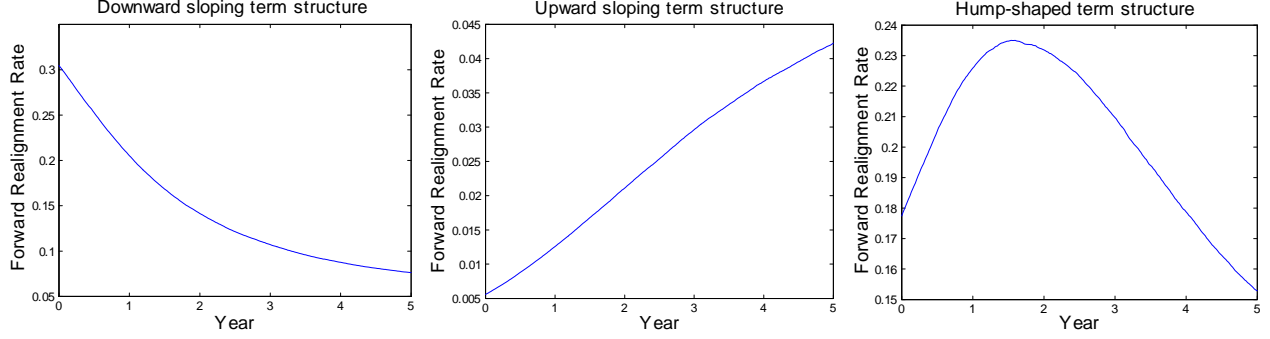


Figure 3: Possible Shapes of the Term Structure

realignment in the further future is perceived to be smaller.

4.2.2 Approximation of the Forward Price

The forward price in (4.1) does not have a closed form. We will approximate the forward price with its third order expansion around time-to-maturity $d = 0$ using the infinitesimal generator \mathcal{A} of the process $\{S, \Gamma\}$, i.e.,

$$F_d \approx S_0 + \sum_{i=1}^3 \frac{d^i}{i!} \mathcal{A}^i f(S_0, \Gamma_0) \quad (4.2)$$

where $f(S_d, \Gamma_d) \equiv S_d$. Since a NDF with short maturity (one month) will be used in the estimation, the approximation is expected to be good. We have verified using simulation that such approximation has an error of no more than 1×10^{-4} . The accuracy is sufficient for the application since the forward price data reported by Datastream International have just four digits after the decimal point.

4.2.3 Identification

Let $g(S_d) = S_d$. The forward price (4.1) can be rewritten as

$$F_d = S_0 + E \left[\int_0^d \mathcal{L}g(S_s) ds \middle| S_0, \Gamma_0 \right]$$

where \mathcal{L} is the infinitesimal generator of the process S . It can be calculated that

$$\mathcal{L}g(S_t) = \left[e^{\Gamma t} \lambda_u E(J_t^u) + e^{-\Gamma t} \lambda_d E(J_t^d) \right] S_t. \quad (4.3)$$

Therefore, λ_u , $E(J_t^u)$, λ_d and $E(J_t^d)$ are not separately identified from the forward rate alone. Intuitively, a higher intensity and a larger magnitude of realignment have the same implication for the forward rate.

In principle, the parameters λ_u , $E(J_t^u)$, λ_d and $E(J_t^d)$ are identified from the time series of exchange rates. However, in the sample period from February 2002 to September 2003, no realignments took place. To overcome this problem, we collected all the realignment events since the creation of the Yuan in 1955 (see Appendix 4). In the past forty-eight and a half years, there have been two appreciations for the Yuan: 9 percent in 1971 and 11 percent in 1973 with an average of 10 percent.

There have been four depreciations: a roughly 28 percent depreciation in 1986, a 21 percent and a 10 percent depreciation in 1989 and 1990 respectively; and finally a 33 percent depreciation in 1994. The average of these depreciations is 23 percent.

Therefore, we will set $\lambda_u = \frac{2}{48.5}$, $E(J_t^u) = 0.1$, $\lambda_d = \frac{4}{48.5}$, $E(J_t^d) = -0.23$. I.e., we will consider realignment risk with an average magnitude of 10 percent appreciation and 23 percent depreciation. This is a normalization that will not affect the results later. To see this, given the prevailing anticipation of Yuan's appreciation, the term $e^{\Gamma t - \lambda_u E(J_t^u)}$ corresponding to appreciation will dominate in (4.3). Suppose the true value of $\lambda_u E(J_t^u) = v$ and we instead set $\lambda_u E(J_t^u) = cv$, then $e^{\Gamma t - \lambda_u E(J_t^u)} = e^{\Gamma t - v} = e^{\Gamma t - \log c cv}$. This amounts to a level effect of magnitude $\log c$ on estimates of $\{\Gamma_t\}$ which will not affect conclusions on time-variations of Γ_t which are our focus.

4.3 Iterative Maximum-Likelihood Estimation

Let $\theta \equiv \{k, \sigma, \eta, \sigma_K\}$ with θ_0 being the true parameter value. Starting from an initial estimate $\theta^{(m)}$, we can invert the forward pricing formula (4.2) to get $\widehat{\Gamma}^{(m)} = \Gamma(\theta^{(m)})$. The vector $\widehat{\Gamma}^{(m)} = \{\widehat{\Gamma}_0^{(m)}, \widehat{\Gamma}_\Delta^{(m)}, \widehat{\Gamma}_{2\Delta}^{(m)}, \dots, \widehat{\Gamma}_T^{(m)}\}$ estimates the process Γ at different points in time. (To save notation, Γ is used for both the process Γ and the function inverting the forward pricing equation. The dependence of $\Gamma(\theta^{(m)})$ on observed spot and forward rates is not made explicit.) Knowing $\widehat{\Gamma}^{(m)}$, we can apply maximum-likelihood estimation to get another estimate of the parameter $\theta^{(m+1)} = H(\widehat{\Gamma})$. The process Γ does not admit a closed-form likelihood. Fortunately, the approximate likelihood introduced in the first part of the paper can be applied to estimate $\theta^{(m+1)}$. In particular, a third order likelihood approximation will be used. This defines an iterative estimation procedure.

$$\theta^{(m+1)} = H\left(\Gamma\left(\theta^{(m)}\right)\right) \quad (4.4)$$

Let $\theta_{T,\Delta}^*$ denote the *uncomputable* maximum-likelihood estimator if we can observe the process Γ , i.e., $\theta_{T,\Delta}^* = H(\Gamma(\theta_0))$. $\theta_{T,\Delta}^*$ is uncomputable because the process Γ is not directly observed by the econometrician. The following proposition shows that the iterative maximum-likelihood estimation procedure converges and the resulting estimator approaches the efficiency of $\theta_{T,\Delta}^*$. As a reminder, d is the maturity of the forward contract (one month here), Δ is the sampling interval (daily here) and T is the sample period (February 2002 to December 2003).

Proposition 3. *Given Δ and T , there exists a positive function $M(d) \xrightarrow{d \rightarrow 0} 0$ so that as $d \rightarrow 0$, with probability approaching one, the mapping $H(\Gamma(\cdot))$ is a contraction with modulus $M(d)$ and the iterative maximum-likelihood estimation procedure converges to an estimate $\widehat{\theta}_{T,\Delta}$. Further,*

$$\left\| \widehat{\theta}_{T,\Delta} - \theta_0 \right\| \leq \frac{1}{1 - M(d)} \left\| \theta_{T,\Delta}^* - \theta_0 \right\|$$

This proposition ensures we can approach efficiency feasible when the process Γ is observable. Theorem 3 ensures $\theta_{T,\Delta}^*$ approaches the efficiency of the maximum-likelihood estimator using the exact likelihood function. Therefore, the iterative MLE estimator asymptotically achieves the efficiency of the true MLE estimator which is uncomputable both because the process Γ is unobservable and because the likelihood function is unavailable in closed form. When $\widehat{\theta}_{T,\Delta}$ is estimated, $\widehat{\Gamma} = \Gamma(\widehat{\theta}_{T,\Delta})$ estimates the time series of the process Γ .

4.4 Estimation Result

To invert Γ from the forward pricing formula, forward rate of one maturity suffices. The data contain forward rates of nine different maturities: one day, two days, one week, one month, two months, three months, six months, nine months and one year. Proposition 3 demands the use of the shortest maturity. However, we did not take Proposition 3 literally and use the forward price with the shortest maturity (one day) because, when the time-to-maturity is too small, the realignment risk becomes negligible and the price can be prone to noises not modelled here. In the sample, the forwards with extremely short maturities behave very differently from other forwards. The forward contracts with maturities one month or longer are highly correlated (correlation coefficient above 0.95) and any one of them captures most of the variations of other forwards. Therefore, the one-month forward rate will be used in the estimation, striking a balance between Proposition 3 and the short-term noises. Given that we are mostly concerned with the realignment risk in the next several years, this choice is more sensible than the one-day forward rate. We applied the iterative MLE procedure on the spot exchange rate and the one-month NDF rate. The estimates are in Table 1. The standard errors in the parentheses are computed from Theorem 4. A statistically significant η indicates the presence of jumps. The jump standard deviation σ_K is harder to estimate due to the infrequent nature of jumps.

Table 1: Parameter Estimates

k	σ	η	σ_K
0.2545 (0.042)	0.5159 (0.012)	31.414 (10.38)	0.2393 (0.240)

The future realignment probability can be assessed from the term structure of the forward realignment rate. Using the parameter estimates and the filtered Γ , this term structure on December 12, 2003, the last day in the sample, is recovered and plotted in Figure 4.¹⁷

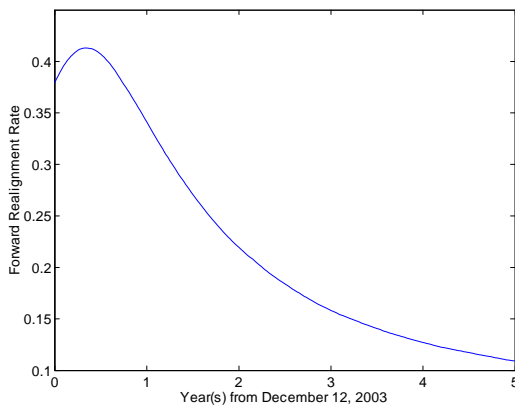


Figure 4: Term Structure of Forward Realignment Rate

Interestingly, the term structure is hump-shaped and peaks at about six months from December 12, 2003. The market considers the Yuan likely to appreciate in 2004 and, conditioning on no

¹⁷Simulation with 100,000 sample paths is used to compute the term structure. Each week is discretized with 50 points in between using the Euler Scheme.

realignment in that period, realignment in the further future is perceived less likely.

4.5 What does the Realignment Probability Respond to?

Figure 5 plots the daily estimates of process $\hat{\Gamma}$. The realignment intensity can vary dramatically over a short period of time. In this section, we infer from the estimates what information the realignment intensity, especially its jumps, responds to. Such information helps identify the factors influencing exchange rate realignment.

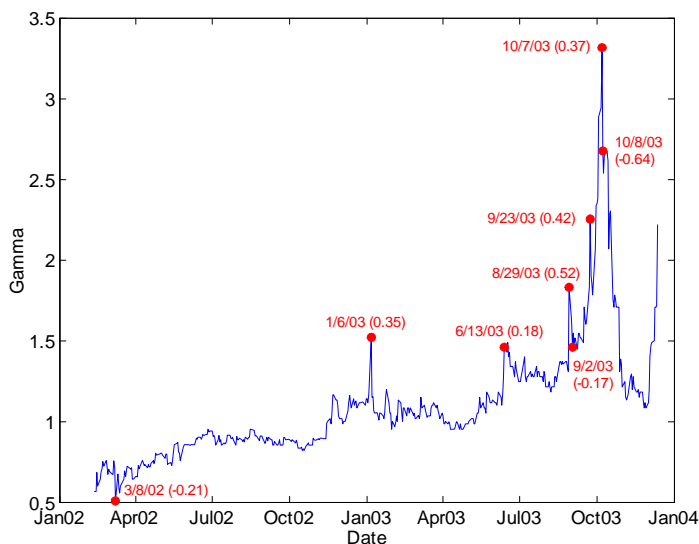


Figure 5: Filtered Γ and Eight of its Jumps

To identify when a jump takes place, we computed the standard deviation over one day of the process Γ should there be no jumps. Days over which Γ changed by more than five such standard deviations are singled out. They contain at least one jump with probability close to one by Bayes rule. There are twenty four jumps identified in the sample period using this method.

To find out what these jumps respond to, we collected new releases on these jump days. Eight of these jumps are found to coincide with economic news releases. These eight days are marked in Figure 5 by dots and the jump magnitudes are in the parentheses. The realignment intensity is proportional to the exponential of Γ , so a 0.01 increase in Γ translates into about a one percentage increase of the intensity. On some of these jump days, the realignment intensity swings by as much as 50 percent. News releases on the eight jump days are listed in Table 3 in Appendix 5. The jumps can be seen to coincide with news releases on the Sino-US trade surplus, state-owned enterprise reform, Chinese government tax revenue and, most importantly, both domestic and foreign government interventions. Given the tightly managed nature of the Chinese Yuan, it is not surprising that the market's second-guess of the Chinese government's intention plays a dominant role. Diplomatic pressures from foreign governments are also important.

To illustrate how the term structure of the forward realignment rate responds to the news release, Figure 6 plots the term structures on August 28, August 29 and September 2 in 2003. On August 29, 2003, before the visit of the US Treasury Secretary, the Japanese Finance Minister publicly urged China to let the Yuan float freely and the state-owned enterprises in China showed improved profitability. On the same day, the level of the forward realignment rate almost doubled relative

to the previous day. More interestingly, the peak of the term structure moved from ten months in the future to six months in the future, indicating the market is anticipating the Yuan to appreciate sooner. Four days later, after the diplomatic pressure receded, the peak of the term structure moved back to roughly eight months in the future and the level of the term structure dropped.

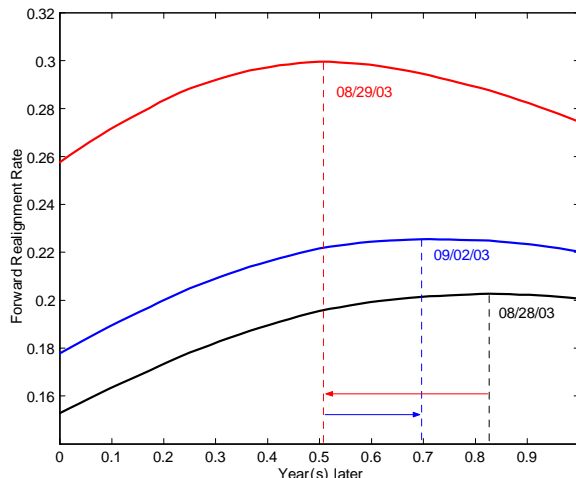


Figure 6: Changes of the Term Structure on Jump Days

Economic theory is needed to gain further insights on the Yuan’s realignment. Most of the existing theory on exchange rate is about depreciation, unlike the case of the Yuan where the market is expecting an appreciation. The results here lend empirical evidence to modelling realignment where a currency is under pressure to appreciate.

5 Conclusion

This paper provides closed-form likelihood approximations for multivariate jump-diffusion processes. For a fixed order of approximation, the maximum-likelihood estimator computed from this approximate likelihood achieves the asymptotic efficiency of the true yet uncomputable maximum-likelihood estimator as the sampling interval shrinks. It facilitates the estimation of jump-diffusions which are very useful in analyzing many economic phenomena such as currency crises, financial market crashes, defaults, etc. The approximation method is based on expansions from Kolmogorov equations which are common to Markov processes and therefore can potentially be generalized to other classes of Markov processes such as the time-changed Levy processes in Carr et. al. (2003) and the non-Gaussian Ornstein-Uhlenbeck-based processes in Barndorff-Nielsen and Shephard (2001). The question then is to find an appropriate leading term for these processes to replace (2.6).

The approximation technique is applied to the Chinese Yuan to study its realignment probability. The term structure of the forward realignment rate is hump-shaped and peaks at six months from the end of 2003. The implication is that the financial market is anticipating an appreciation in the next year and, conditioning on no realignment before then, the chance of realignment is perceived to be smaller in the further future. Since February 2002, the realignment intensity of the Yuan has increased fivefold. The realignment probability responds quickly to news releases on the Sino-US

trade surplus, state-owned enterprise reform, Chinese government tax revenue and, most importantly, both domestic and foreign government interventions.

References

- [1] Aït-Sahalia, Yacine (2002): “Maximum Likelihood Estimation of Discretely Sampled Diffusions: A Closed-Form Approximation Approach,” *Econometrica*, Vol 70, No. 1.
- [2] Aït-Sahalia, Yacine (2001): “Closed-Form Likelihood Expansions for Multivariate Diffusions,” working paper.
- [3] Aït-Sahalia, Yacine (1996): “Nonparametric Pricing of Interest Rate Derivative Securities,” *Econometrica* 64, 527-560.
- [4] Aït-Sahalia, Yacine and Robert Kimmel (2005): “Maximum Likelihood Estimation of Stochastic Volatility Models,” *Journal of Financial Economics*, forthcoming.
- [5] Aït-Sahalia, Yacine, Per Mykland, and Lan Zhang (2005): “How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise,” *Review of Financial Studies*, Vol 18, 351-416.
- [6] Aït-Sahalia, Yacine and Jialin Yu (2006): “Saddlepoint Approximations for Continuous-Time Markov Processes,” *Journal of Econometrics*, forthcoming.
- [7] Andersen, Torben, Luca Benzoni and Jesper Lund (2002): “An Empirical Investigation of Continuous-Time Equity Return Models,” *Journal of Finance*, Vol 57, 1239-1284.
- [8] Andersen, T.G., T. Bollerslev and F.X. Diebold (2006): “Roughing It Up: Including Jump Components in the Measurement, Modeling and Forecasting of Return Volatility,” *Review of Economics and Statistics*, forthcoming.
- [9] Bakshi, G. and N. Ju (2005): “A Refinement to Aït-Sahalia’s (2002) “Maximum Likelihood Estimation of Discretely Sampled Diffusions: A Closed-Form Approximation Approach”,” *Journal of Business*, 2037-2052.
- [10] Bakshi, G., N. Ju, and H. Ou-Yang (2005): “Estimation of Continuous-Time Models with an Application to Equity Volatility Dynamics,” *Journal of Financial Economics*, forthcoming.
- [11] Bandi, Federico M. and Thong H. Nguyen (2003): “On the Functional Estimation of Jump-Diffusion Models”, *Journal of Econometrics*, Vol 116, 293-328.
- [12] Bandi, Federico M. and Peter C. B. Phillips (2003): “Fully Nonparametric Estimation of Scalar Diffusion Models,” *Econometrica* Vol 71, No. 1, 241-284.
- [13] Bandi, Federico M. and Jeffrey R. Russell (2006): “Volatility”, *Handbook of Financial Engineering*, Elsevier, forthcoming.
- [14] Barndorff-Nielsen, O.E. and N. Shephard (2001): “Non-Gaussian Ornstein-Uhlenbeck-Based Models and Some of Their Uses in Financial Economics,” *Journal of the Royal Statistical Society, Series B*, Vol 63, 167-241.
- [15] Barndorff-Nielsen, O.E. and N. Shephard (2004): “Power and Bipower Variation with Stochastic Volatility and Jumps,” *Journal of Financial Econometrics*, Vol 2, 1-48.
- [16] Bates, D. (1996): “Jumps and Stochastic Volatility: Exchange Rate Processes Implicit in Deutsche Mark Options,” *Review of Financial Studies*, Vol 9, 69-107.

- [17] Bichteler, Klaus, Jean-Bernard Gravereaux and Jean Jacod (1987): *Malliavin Calculus for Processes with Jumps*, Gordon and Breach Science Publishers.
- [18] Brandt, Michael and Pedro Santa-Clara (2002): "Simulated Likelihood Estimation of Diffusions with an Application to Exchange Rate Dynamics in Incomplete Markets," *Journal of Financial Economics*, Vol 63, 161-210.
- [19] Brémaud, Pierre (1981): *Point Processes and Queues, Martingale Dynamics*, Springer-Verlag.
- [20] Carr, P., H. Geman, D. Madan and M. Yor (2003): "Stochastic Volatility for Lévy Processes", *Mathematical Finance*, Vol 13, 345-382.
- [21] Carrasco, Marine, Mikhail Chernov, Jean-Pierre Florens, and Eric Ghysels (2006): "Efficient Estimation of General Dynamic Models with a Continuum of Moment Conditions", *Journal of Econometrics*, forthcoming.
- [22] Duffee, Gregory (2002): "Term Premia and Interest Rate Forecasts in Affine Models," *Journal of Finance*, Vol 57, 405-443.
- [23] Duffie, Darrell and Peter Glynn (2004): "Estimation of Continuous-Time Markov Processes Sampled at Random Time Intervals," *Econometrica*, Vol 72, 1773-1808.
- [24] Duffie, Darrell, Jun Pan and Kenneth Singleton (2000): "Transform Analysis and Asset Pricing for Affine Jump Diffusions," *Econometrica*, Vol 68, 1343-1376.
- [25] Duffie, Darrell and Kenneth Singleton (2003): *Credit Risk*, Princeton University Press.
- [26] Elerian, Ola, Siddhartha Chib and Neil Shephard (2001): "Likelihood Inference for Discretely Observed Nonlinear Diffusions," *Econometrica*, Vol 69, No. 4, 959-993.
- [27] Eraker, B., M. Johannes and N. Polson (2003): "The Impact of Jumps in Volatility and Returns," *Journal of Finance*, 1269-1300.
- [28] Gallant, A. R. and G. Tauchen (1996): "Which Moments to Match?" *Econometric Theory* 12, 657-681.
- [29] Gouriéroux, C., A. Monfort and E. Renault (1993): "Indirect Inference," *Journal of Applied Econometrics*, S85-S118.
- [30] Hansen, L.P. and J.A. Scheinkman (1995): "Back to the Future: Generating Moment Implications for Continuous-Time Markov Processes," *Econometrica* 63, 767-804.
- [31] Has'minskii, R. Z. (1980): *Stochastic Stability of Differential Equations*, Sijthoff & Noordhoff.
- [32] Hoffman, Kenneth (1975): *Analysis in Euclidean Space*, Prentice Hall.
- [33] Huang, X. and G. Tauchen (2005): "The Relative Contribution of Jumps to Total Price Variance," *Journal of Financial Econometrics*, Vol 3, 456-499.
- [34] Huang, Yasheng (2003): *Selling China*, Cambridge University Press.
- [35] Jeganathan, P. (1995): "Some Aspects of Asymptotic Theory with Applications to Time Series Models," *Econometric Theory*, Vol 11, 818-887.
- [36] Jensen, Bjarke and Rolf Poulsen (2002): "Transition Densities of Diffusion Processes: Numerical Comparison of Approximation Techniques," *Journal of Derivatives*, Vol 9, No. 4, 18-32.
- [37] Johannes, M. (2004): "The Statistical and Economic Role of Jumps in Continuous-Time Interest Rate Models," *Journal of Finance*, 227-260.
- [38] Kessler, M. and M. Sørensen (1999): "Estimating Equations Based on Eigenfunctions for a Discretely Observed Diffusion Process," *Bernoulli*, 5, 299-314.

- [39] Kristensen, Dennis (2004): “Estimation in Two Classes of Semiparametric Diffusion Models”, working paper.
- [40] Ledoit, Olivier, Pedro Santa-Clara and Shu Yan (2002): “Relative Pricing of Options with Stochastic Volatility,” working paper.
- [41] Liu, Jun, Francis A. Longstaff and Jun Pan (2003): “Dynamic Asset Allocation with Event Risk,” *Journal of Finance*, Vol 58, No. 1, 231-259.
- [42] Merton, Robert (1992): *Continuous-Time Finance*, Blackwell Publishers.
- [43] Pan, Jun (2002): “The jump-risk premia implicit in options: evidence from an integrated time-series study,” *Journal of Financial Economics*, Vol 63, 3-50.
- [44] Pedersen, A.R. (1995): “A New Approach to Maximum-Likelihood Estimation for Stochastic Differential Equations Based on Discrete Observations,” *Scandinavian Journal of Statistics*, Vol 22, 55-71.
- [45] Piazzesi, M. (2005): “Bond Yields and the Federal Reserve,” *Journal of Political Economy*, Vol 113, 311-344.
- [46] Protter, Philip (1990): *Stochastic Integration and Differential Equations*, Springer-Verlag.
- [47] Revuz, D. and Marc Yor (1999): *Continuous Martingales and Brownian Motion*, Springer Verlag, Third Edition.
- [48] Robinson, P. M. (1988): “The Stochastic Difference between Econometric Statistics,” *Econometrica*, Vol. 56, No. 3, 531-548.
- [49] Rudin, W. (1976): *Principles of Mathematical Analysis*, McGraw-Hill, Third Edition.
- [50] Schaumburg Ernst (2001): “Maximum Likelihood Estimation of Levy Type SDEs,” Ph.D. dissertation, Princeton University.
- [51] Singleton, Kenneth (2001), ”Estimation of Affine Asset Pricing Models Using the Empirical Characteristic Function,” *Journal of Econometrics*, Vol. 102, 111-141.
- [52] Stokey, Nancy, Robert Lucas with Edward Prescott (1996): *Recursive Methods in Economic Dynamics*, Harvard University Press.
- [53] Sundaresan, Suresh (2000): “Continuous-Time Methods in Finance: A Review and an Assessment,” *Journal of Finance*, Vol. 55, No. 4, 1477-1900.
- [54] Varadhan, S.R.S. (1967): “On the Behavior of the Fundamental Solution of the Heat Equation with Variable Coefficients,” *Communications on Pure and Applied Mathematics*, Vol XX, 431-455.
- [55] White, Halbert (1982): “Maximum Likelihood Estimation of Misspecified Models,” *Econometrica*, Vol 50, No. 1.

Appendix 1: Assumptions

To save notation, the dependence on θ of the functions $\mu(\cdot, \theta)$, $\sigma(\cdot, \theta)$, $V(\cdot, \theta)$, $\lambda(\cdot, \theta)$, $\nu(\cdot, \theta)$ and $p(\Delta, y|x, \theta)$ will not be made explicit when there is no confusion.

Assumption 1. *The variance matrix $V(x)$ is positive definite for all x in the domain of the process X .*

This is a nondegeneracy condition. Given this assumption, we can find, by Cholesky decomposition, an $n \times n$ positive definite matrix $V^{1/2}(x)$ satisfying $V(x) = V^{1/2}(x) [V^{1/2}(x)]^T$.

Assumption 2. *The stochastic differential equation (2.1) has a unique solution. The transition density $p(\Delta, y|x)$ exists for all x, y in the domain of X and all $\Delta > 0$. $p(\Delta, y|x)$ is continuously differentiable with respect to Δ , twice continuously differentiable with respect to x and y .*

Theorem 2-29 in Bichteler et al (1987) provides the following sufficient conditions for assumption 2 to hold: (1) jump intensity $\lambda(\cdot)$ is constant, (2) $\mu(\cdot)$ and $\sigma(\cdot)$ are infinitely continuously differentiable with bounded derivatives, (3) the eigenvalues of $V(\cdot)$ is bounded below by a positive constant, and (4) The jump distribution of J_t has moments of all orders.

Assumption 3. *The boundary of the process X is unattainable.*

This assumption implies that the transition density $p(\Delta, y|x)$ is uniquely determined by the backward and forward Kolmogorov equations.

Assumption 4. *$\nu(\cdot)$, $\mu(\cdot)$, $\sigma(\cdot)$ and $\lambda(\cdot)$ are infinitely differentiable almost everywhere in the domain of X .*

Assumption 5. *The model (2.1) describes the true data-generating process and the true parameter, denoted θ_0 , is in a compact set Θ .*

Let $I_{T,\Delta}(\theta) \equiv E_\theta \left[\dot{l}_{T,\Delta}(\theta) \dot{l}_{T,\Delta}(\theta)^T \right]$ be the information matrix. To simplify the notation, $\dot{l}_{T,\Delta}(\theta)$ is used for derivative with respect to the vector θ . Let $G_{T,\Delta}(\theta) \equiv -I_{T,\Delta}^{-1/2}(\theta) \ddot{l}_{T,\Delta}(\theta) I_{T,\Delta}^{-1/2}(\theta)$ and $S_{T,\Delta}(\theta) \equiv I_{T,\Delta}^{-1/2}(\theta) \dot{l}_{T,\Delta}(\theta)$.

Assumption 6. *$I_{T,\Delta}(\theta)$ is invertible and $l_{T,\Delta}(\theta)$ is thrice continuously differentiable with respect to $\theta \in \Theta$. There exists $\bar{\Delta} > 0$ such that*

1. $\|I_{T,\Delta}(\theta)\|^{-1} \xrightarrow{P} 0$ as $T \rightarrow \infty$, uniformly for $\theta \in \Theta$ and $\Delta \leq \bar{\Delta}$;
2. Uniformly for $\theta \in \Theta$ and $\Delta \leq \bar{\Delta}$, $G_{T,\Delta}^{-1}(\theta)$ and $S_{T,\Delta}(\theta)$ converge to $G_\Delta(\theta)$ and $S_\Delta(\theta)$ that are bounded in probability as $T \rightarrow \infty$;
3. $\left\| I_{T,\Delta}^{-1/2}(\theta) \ddot{l}_{T,\Delta}(\tilde{\theta}) I_{T,\Delta}^{-1/2}(\theta) \right\|$ is uniformly bounded in probability for $\theta, \tilde{\theta} \in \Theta$ and $\Delta \leq \bar{\Delta}$.

The last condition in this assumption is stronger than the usual one which only requires uniform boundedness in probability for $\tilde{\theta}$ in a shrinking neighborhood of θ . This stronger version is used to prove the asymptotic property of $\hat{\theta}^{(m)}$. If the process is stationary, this condition will automatically hold because $\left\| I_{T,\Delta}^{-1/2}(\theta) \ddot{l}_{T,\Delta}(\tilde{\theta}) I_{T,\Delta}^{-1/2}(\theta) \right\|$ will converge in probability to a constant $\eta(\Delta, \theta, \tilde{\theta})$ which is bounded for parameters in the compact set Θ and $\Delta \leq \bar{\Delta}$ (see equation (A.49) in Ait-Sahalia (2002) for the stationary diffusion case), so no cost is incurred from strengthening this condition.

Assumption 7. *For any $t < \infty$, $M_t \equiv \sup_{0 \leq s \leq t} \|X_s\|$ is bounded in probability.*

Appendix 2: Proofs

Proof of Proposition 1: The result in this proposition is standard and the proof is therefore omitted and is available upon request.

Derivation of Condition 2: It is clear that this initial condition does not affect the result in Theorem 1 which is satisfied by any solution to the backward equation.

Knowing $C^{(-1)}(x, y) = \frac{1}{2} \left[\frac{\partial}{\partial x} C^{(-1)}(x, y) \right]^T V(x) \left[\frac{\partial}{\partial x} C^{(-1)}(x, y) \right]$,

$$\begin{aligned} \int p(\Delta, y|x) dy &= \int \Delta^{-n/2} \exp \left[-\frac{C^{(-1)}(x, y)}{\Delta} \right] C^{(0)}(x, y) dy + o(\Delta) \\ &= \int \Delta^{-n/2} \exp \left\{ -\frac{\left[\frac{\partial}{\partial x} C^{(-1)}(x, y) \right]^T V(x) \left[\frac{\partial}{\partial x} C^{(-1)}(x, y) \right]}{2\Delta} \right\} C^{(0)}(x, y) dy + o(\Delta) \\ &= (2\pi)^{n/2} \int \frac{(2\pi\Delta)^{-n/2} \exp \left(-\frac{w^T w}{2\Delta} \right) C^{(0)}(x, w_x^{-1}(w))}{\left| \det V(x)^{1/2} \frac{\partial^2}{\partial x \partial y^T} C^{(-1)}(x, w_x^{-1}(w)) \right|} dw + o(\Delta) \end{aligned}$$

by a change of variable formula. $w_x^{-1}(\cdot)$ is defined as the inverse function of $V(x)^{1/2} \frac{\partial}{\partial x} C^{(-1)}(x, \cdot)$ and satisfies $w_x^{-1}(0) = x$. It can be verified as in the proof of Theorem 1 that this change of variable is legitimate.

$(2\pi\Delta)^{-n/2} \exp \left(-\frac{w^T w}{2\Delta} \right)$ converges to a dirac-delta function at 0 implies, as $\Delta \rightarrow 0$,

$$\int p(\Delta, y|x) dy \rightarrow (2\pi)^{n/2} C^{(0)}(x, x) \left| \det V(x)^{1/2} \frac{\partial^2}{\partial x \partial y^T} C^{(-1)}(x, y) \right|_{y=x}^{-1}.$$

Requiring the right-hand side equals 1 gives condition 2, noticing from equation (2.11) that

$$\frac{\partial^2}{\partial x \partial y^T} C^{(-1)}(x, y) \Big|_{y=x} = -V(x)^{-1}.$$

The proofs to Theorems 1 and 2 are very similar. We, therefore, prove Theorem 1. Corollaries 1 and 2 are obtained by explicitly solving the equations in Theorems 1 and 2.

Proof of Theorem 1: To get $C^{(k)}(x, y)$ and $D^{(k)}(x, y)$, we replace $p(\Delta, y|x)$ with (2.6) in the backward Kolmogorov equation (2.3), match the terms with the same orders of Δ^k and the terms with the same orders of $\Delta^k \exp \left[-\frac{C^{(-1)}(x, y)}{\Delta} \right]$ and set their coefficients to 0. Most of the steps are careful bookkeeping, only the following two points need to be addressed.

- How to expand $\int_C p(\Delta, y|x+c) \nu(c) dc$ in increasing orders of Δ ?
- Prove that, for fixed y , $w_B(\cdot, y)$ is invertible in a neighborhood of $x = y$ and its inverse function $w_B^{-1}(\cdot)$ is continuously differentiable.

Notice

$$\int_C p(\Delta, y|x+c) \nu(c) dc \rightarrow \nu(y-x) \quad \text{as } \Delta \rightarrow 0$$

implies the expansion of $\int_C p(\Delta, y|x+c) \nu(c) dc$ does not involve terms in the form of $\Delta^k \exp\left[-\frac{C^{(-1)}(x,y)}{\Delta}\right]$. Therefore, we first group terms with the same orders of $\Delta^k \exp\left[-\frac{C^{(-1)}(x,y)}{\Delta}\right]$, set their coefficient to 0 and get

$$C^{(-1)}(x, y) = \frac{1}{2} \left[\frac{\partial}{\partial x} C^{(-1)}(x, y) \right]^T V(x) \left[\frac{\partial}{\partial x} C^{(-1)}(x, y) \right] \quad (5.1)$$

which is the equation for $C^{(-1)}(x, y)$ in Theorem 1. Take the derivative of $w_B(x, y)$ and notice $\frac{\partial}{\partial x} C^{(-1)}(x, y)|_{x=y} = 0$,

$$\begin{aligned} \frac{\partial}{\partial x^T} w_B(x, y) \Big|_{x=y} &= \left[V^{1/2}(y) \right]^T \frac{\partial^2}{\partial x \partial x^T} C^{(-1)}(x, y) \Big|_{x=y} \\ &= V(y)^{-1/2} \end{aligned}$$

by equation (2.11). By the Inverse Mapping Theorem (Section 8.5 in Hoffman (1975)), there is an open neighborhood of $x = y$, so that $w_B(x, y)$, for fixed y , maps this neighborhood 1-1 and onto an open neighborhood of 0 in \mathbb{R}^n and its inverse function is continuously differentiable.

Now, knowing $w_B(x, y)$ is invertible, we will expand $\int_C p(\Delta, y|x+c) \nu(c) dc$ in increasing orders of Δ . First, for a given $k \geq 0$, by (5.1)

$$\begin{aligned} &\int_C \Delta^{-n/2} \exp\left[-\frac{C^{(-1)}(x+c, y)}{\Delta}\right] C^{(k)}(x+c, y) \nu(c) dc \\ &= \int_C \Delta^{-n/2} \exp\left[-\frac{w_B^T(x+c, y) w_B(x+c, y)}{2\Delta}\right] C^{(k)}(x+c, y) \nu(c) dc. \end{aligned} \quad (5.2)$$

Let $C_1 \subset C$ be an open set containing $y-x$ on which $w_B(x+., y)$ is invertible for the given x and y . Denote by $W \in \mathbb{R}^n$ the image of C_1 under $w_B(x+., y)$. As shown previously, W is open and contains 0. As $\Delta \rightarrow 0$, the last integral differs exponentially small from

$$\begin{aligned} &\int_{C_1} \Delta^{-n/2} \exp\left[-\frac{w_B^T(x+c, y) w_B(x+c, y)}{2\Delta}\right] C^{(k)}(x+c, y) \nu(c) dc \\ &= \int_W \Delta^{-n/2} \exp\left[-\frac{\omega^T \omega}{2\Delta}\right] g_k(x, y, \omega) d\omega \end{aligned}$$

by a change of variable. The function $g_k(x, y, \omega)$ is defined in the theorem. Using the multivariate differentiation notation in Section 2.3, we Taylor expand $g_k(x, y, \omega)$ around $\omega = 0$ to get

$$\begin{aligned} &= \int_W \Delta^{-n/2} \exp\left[-\frac{\omega^T \omega}{2\Delta}\right] \sum_{r=0}^{\infty} \frac{1}{r!} \sum_{|s|=r} \omega^s \frac{\partial^s}{\partial u^s} g_k(x, y, u) \Big|_{u=0} d\omega \\ &= \sum_{r=0}^{\infty} \frac{1}{r!} \sum_{|s|=r} \frac{\partial^s}{\partial \omega^s} g_k(x, y, \omega) \Big|_{\omega=0} \int_W \Delta^{-n/2} \exp\left[-\frac{\omega^T \omega}{2\Delta}\right] \omega^s d\omega \\ &\approx \sum_{r=0}^{\infty} \frac{1}{r!} \sum_{|s|=r} \frac{\partial^s}{\partial \omega^s} g_k(x, y, \omega) \Big|_{\omega=0} \Delta^{r/2} \int_{\mathbb{R}^n} \exp\left[-\frac{\omega^T \omega}{2}\right] \omega^s d\omega \end{aligned}$$

where the last term follows by first changing the variable $\frac{\omega}{\sqrt{\Delta}}$ and then ignore terms that are exponentially small as $\Delta \rightarrow 0$. We note that, to decide which terms are exponentially small, we need not worry about the change of limit operators of the form: $\lim_{\Delta \rightarrow 0} \sum_{r=0}^{\infty}$ to $\sum_{r=0}^{\infty} \lim_{\Delta \rightarrow 0}$ simply because only finite number of terms in the expansion $\sum_{r=0}^{\infty}$ will be used to compute the approximate transition density $p^{(m)}(\Delta, y|x)$ for fixed m . I.e., the notation $\sum_{r=0}^{\infty}$ here does not mean convergence of the infinite sum, it simply represents a Taylor expansion from which only a finite number of terms will be used.

Let $M_s^n \equiv (2\pi)^{-n/2} \int_{\mathbb{R}^n} \exp\left[-\frac{w^T w}{2}\right] w^s dw$ denote the s -th moment of n -variate standard normal distribution. $M_s^n = 0$ if any element of s is odd. The last expression therefore equals

$$= (2\pi)^{n/2} \sum_{r=0}^{\infty} \frac{\Delta^r}{(2r)!} \sum_{s \in S_{2r}^n} \frac{\partial^s}{\partial w^s} g_k(x, y, w) \Big|_{w=0} M_s^n \quad (5.3)$$

with S_{2r}^n defined as in Section 2.3.

Note, we have implicitly assumed $y - x \in C$. In the case $y - x \notin C$, assumption 4 implies $\nu(\cdot)$ and its derivatives of all orders vanish at $y - x$ which implies the right hand side of (5.3) is 0. It is also clear that (5.2) is exponentially small in Δ if $y - x \notin C$. Therefore, (5.3) is still valid.

Now, knowing the above expansion,

$$\begin{aligned} & \int_C \Delta^{-n/2} \exp\left[-\frac{C^{(-1)}(x+c, y)}{\Delta}\right] \sum_{k=0}^{\infty} C^{(k)}(x+c, y) \Delta^k \nu(c) dc \\ &= (2\pi)^{n/2} \sum_{k=0}^{\infty} \Delta^k \sum_{r=0}^{\infty} \frac{\Delta^r}{(2r)!} \sum_{s \in S_{2r}^n} \frac{\partial^s}{\partial w^s} g_k(x, y, w) \Big|_{w=0} M_s^n \\ &= (2\pi)^{n/2} \sum_{k=0}^{\infty} \Delta^k \sum_{r=0}^k \frac{1}{(2r)!} \sum_{s \in S_{2r}^n} \frac{\partial^s}{\partial w^s} g_{k-r}(x, y, w) \Big|_{w=0} M_s^n. \end{aligned}$$

This last expression, expanded in increasing orders of Δ^k , can be used to match other terms in orders of Δ . This completes the proof.

Proof of Theorem 3 and 4: The true maximum-likelihood estimator sets the score to 0. Therefore,

$$\dot{l}_{T,\Delta}(\theta_0) = -\ddot{l}_{T,\Delta}(\tilde{\theta}) \left(\hat{\theta}_{T,\Delta} - \theta_0\right)$$

for some $\tilde{\theta}$ in between θ_0 and $\hat{\theta}_{T,\Delta}$.

$$\begin{aligned} I_{T,\Delta}^{1/2}(\theta_0) \left(\hat{\theta}_{T,\Delta} - \theta_0\right) &= I_{T,\Delta}^{1/2}(\theta_0) \left[-\ddot{l}_{T,\Delta}(\tilde{\theta})\right]^{-1} \dot{l}_{T,\Delta}(\theta_0) \\ &= -\left[I_{T,\Delta}^{-1/2}(\theta_0) \ddot{l}_{T,\Delta}(\tilde{\theta}) I_{T,\Delta}^{-1/2}(\theta_0)\right]^{-1} I_{T,\Delta}^{-1/2}(\theta_0) \dot{l}_{T,\Delta}(\theta_0) \\ &= -\left[I_{T,\Delta}^{-1/2}(\theta_0) \ddot{l}_{T,\Delta}(\theta_0) I_{T,\Delta}^{-1/2}(\theta_0)\right]^{-1} \left[I_{T,\Delta}^{-1/2}(\theta_0) \dot{l}_{T,\Delta}(\theta_0)\right] + O_p(1) \\ &= G_{\Delta}^{-1}(\theta_0) S_{\Delta}(\theta_0) + O_p(1) \end{aligned}$$

as $T \rightarrow \infty$ uniformly for $\Delta \leq \bar{\Delta}$ and $\theta_0 \in \Theta$ by assumption 6. We have now proved the consistency of $\hat{\theta}_{T,\Delta}$.

Knowing that $\hat{\theta}_{T,\Delta}$ will asymptotically be in an $I_{T,\Delta}^{-1/2}(\theta_0)$ -neighborhood of θ_0 , by repeating the previous argument, we have

$$\begin{aligned} I_{T,\Delta}^{1/2}(\theta_0) \left(\hat{\theta}_{T,\Delta} - \theta_0\right) &= -\left[I_{T,\Delta}^{-1/2}(\theta_0) \ddot{l}_{T,\Delta}(\tilde{\theta}) I_{T,\Delta}^{-1/2}(\theta_0)\right]^{-1} I_{T,\Delta}^{-1/2}(\theta_0) \dot{l}_{T,\Delta}(\theta_0) \\ &= -\left[I_{T,\Delta}^{-1/2}(\theta_0) \ddot{l}_{T,\Delta}(\theta_0) I_{T,\Delta}^{-1/2}(\theta_0)\right]^{-1} \left[I_{T,\Delta}^{-1/2}(\theta_0) \dot{l}_{T,\Delta}(\theta_0)\right] + o_p(1) \\ &= G_{\Delta}^{-1}(\theta_0) S_{\Delta}(\theta_0) + o_p(1). \end{aligned}$$

The asymptotic distribution of $\hat{\theta}_{T,\Delta}$ now follows from the limiting distributions of G_{Δ} and S_{Δ} . In the stationary case, $G_{\Delta}(\theta_0)$ is non-random and it follows as in Theorem 4 that

$$(n_{T,\Delta} \dot{l}_{\Delta}(\theta_0))^{1/2} \left(\hat{\theta}_{T,\Delta} - \theta_0\right) = N(0, I_{p \times p}) + o_p(1).$$

We next investigate the stochastic difference (see Robinson (1988)) between $\widehat{\theta}_{T,\Delta_T}^{(m)}$ and $\widehat{\theta}_{T,\Delta_T}$.

$$\begin{aligned}
& \dot{l}_{T,\Delta_T} \left(\widehat{\theta}_{T,\Delta_T}^{(m)} \right) - \dot{l}_{T,\Delta_T}^{(m)} \left(\widehat{\theta}_{T,\Delta_T}^{(m)} \right) \\
&= \dot{l}_{T,\Delta_T} \left(\widehat{\theta}_{T,\Delta_T}^{(m)} \right) \\
&= \dot{l}_{T,\Delta_T} \left(\widehat{\theta}_{T,\Delta_T} \right) + \ddot{l}_{T,\Delta_T}(\bar{\theta}) \left(\widehat{\theta}_{T,\Delta_T}^{(m)} - \widehat{\theta}_{T,\Delta_T} \right) \\
&= \ddot{l}_{T,\Delta_T}(\bar{\theta}) \left(\widehat{\theta}_{T,\Delta_T}^{(m)} - \widehat{\theta}_{T,\Delta_T} \right)
\end{aligned}$$

for some $\bar{\theta}$ between $\widehat{\theta}_{T,\Delta_T}$ and $\widehat{\theta}_{T,\Delta_T}^{(m)}$. Therefore,

$$\begin{aligned}
I_{T,\Delta_T}^{1/2}(\theta_0) \left(\widehat{\theta}_{T,\Delta_T}^{(m)} - \widehat{\theta}_{T,\Delta_T} \right) &= \left[I_{T,\Delta_T}^{-1/2}(\theta_0) \ddot{l}_{T,\Delta_T}(\bar{\theta}) I_{T,\Delta_T}^{-1/2}(\theta_0) \right]^{-1} I_{T,\Delta_T}^{-1/2}(\theta_0) \left[\dot{l}_{T,\Delta_T} \left(\widehat{\theta}_{T,\Delta_T}^{(m)} \right) - \dot{l}_{T,\Delta_T}^{(m)} \left(\widehat{\theta}_{T,\Delta_T}^{(m)} \right) \right] \\
&= (G_{\Delta}^{-1}(\theta_0) + O_p(1)) I_{T,\Delta_T}^{-1/2}(\theta_0) \left[\dot{l}_{T,\Delta_T} \left(\widehat{\theta}_{T,\Delta_T}^{(m)} \right) - \dot{l}_{T,\Delta_T}^{(m)} \left(\widehat{\theta}_{T,\Delta_T}^{(m)} \right) \right].
\end{aligned}$$

We will show later that there exists a sequence $\{\Delta_T^*\}$ so that for any sequence $\{\Delta_T\}$ satisfying $\Delta_T \leq \Delta_T^*$, $\dot{l}_{T,\Delta_T}^{(m)}(\cdot)$ incurs a relative error of order $o_p(\Delta_T^m)$ in approximating $\dot{l}_{T,\Delta_T}(\cdot)$, uniformly for the parameters in the compact set Θ . Assuming this holds now, we have, for some $\tilde{\theta}$ between $\widehat{\theta}_{T,\Delta_T}^{(m)}$ and θ_0 ,

$$\begin{aligned}
I_{T,\Delta_T}^{1/2}(\theta_0) \left(\widehat{\theta}_{T,\Delta_T}^{(m)} - \widehat{\theta}_{T,\Delta_T} \right) &= o_p(\Delta_T^m) I_{T,\Delta_T}^{-1/2}(\theta_0) \dot{l}_{T,\Delta_T} \left(\widehat{\theta}_{T,\Delta_T}^{(m)} \right) \\
&= o_p(\Delta_T^m) I_{T,\Delta_T}^{-1/2}(\theta_0) \left[\dot{l}_{T,\Delta_T}(\theta_0) + \ddot{l}_{T,\Delta_T}(\tilde{\theta}) \left(\widehat{\theta}_{T,\Delta_T}^{(m)} - \theta_0 \right) \right] \\
&= o_p(\Delta_T^m) \left[I_{T,\Delta_T}^{-1/2}(\theta_0) \dot{l}_{T,\Delta_T}(\theta_0) + O_p(1) I_{T,\Delta_T}^{1/2}(\theta_0) \left(\widehat{\theta}_{T,\Delta_T}^{(m)} - \theta_0 \right) \right] \\
&= o_p(\Delta_T^m) \left[S + O_p(1) I_{T,\Delta_T}^{1/2}(\theta_0) \left(\widehat{\theta}_{T,\Delta_T}^{(m)} - \widehat{\theta}_{T,\Delta_T} + \widehat{\theta}_{T,\Delta_T} - \theta_0 \right) \right]
\end{aligned}$$

where S is the probability limit of $I_{T,\Delta_T}^{-1/2}(\theta_0) \dot{l}_{T,\Delta_T}(\theta_0)$ under P_{θ_0} as $T \rightarrow \infty$.

Therefore,

$$\begin{aligned}
(I - o_p(\Delta_T^m)) I_{T,\Delta_T}^{1/2}(\theta_0) \left(\widehat{\theta}_{T,\Delta_T}^{(m)} - \widehat{\theta}_{T,\Delta_T} \right) &= o_p(\Delta_T^m) \left[O_p(1) + O_p(1) I_{T,\Delta_T}^{1/2}(\theta_0) \left(\widehat{\theta}_{T,\Delta_T} - \theta_0 \right) \right] \\
&= o_p(\Delta_T^m)
\end{aligned}$$

which proves Theorem 3.

Finally, we show the existence of $\{\Delta_T^*\}$, which was used in proving the theorem.

Uniformly for θ in the compact set Θ , $p^{(m)}(\Delta, X_{i\Delta}|X_{(i-1)\Delta}, \theta)$ is an expansion of $p(\Delta, X_{i\Delta}|X_{(i-1)\Delta}, \theta)$ with relative error $\Delta^m \varepsilon_1(\tilde{\Delta}_i, X_{i\Delta}, X_{(i-1)\Delta}, \theta)$ for some $\tilde{\Delta}_i \leq \Delta$, i.e., $p^{(m)}(\Delta, X_{i\Delta}|X_{(i-1)\Delta}, \theta)$ is equal to $p(\Delta, X_{i\Delta}|X_{(i-1)\Delta}, \theta) \left(1 + \Delta^m \varepsilon_1(\tilde{\Delta}_i, X_{i\Delta}, X_{(i-1)\Delta}, \theta) \right)$. Similarly, $\frac{\partial}{\partial \theta} p^{(m)}(\Delta, X_{i\Delta}|X_{(i-1)\Delta}, \theta)$ is an expansion of $\frac{\partial}{\partial \theta} p(\Delta, X_{i\Delta}|X_{(i-1)\Delta}, \theta)$ with relative error $\Delta^m \varepsilon_2(\widehat{\Delta}_i, X_{i\Delta}, X_{(i-1)\Delta}, \theta)$ for some $\widehat{\Delta}_i \leq \Delta$.

Now we bound the functions ε_1 and ε_2 . Let $\{q_t\}$ and $\{\xi_t\}$ be two sequences of positive numbers converging to 0. Under the assumption that M_T is bounded in probability for any $T < \infty$, X_t for all $t < T$ is in a compact set with probability $1 - q_T$ and we can always choose the sequence $\{\Delta_T^*\}$ converging to 0 fast enough so that

for any $\{\Delta_T\}$ satisfying $\Delta_T \leq \Delta_T^*$, $\|\varepsilon_1\|$ and $\|\varepsilon_2\|$ for observations inside this compact set are bounded by ξ_T , uniformly for $\theta \in \Theta$. With this choice of $\{\Delta_T^*\}$, $l_{T,\Delta_T}(\widehat{\theta}_{T,\Delta_T}^{(m)}) - l_{T,\Delta_T}^{(m)}(\widehat{\theta}_{T,\Delta_T}^{(m)})$ equals $o_p(\Delta_T^m) l_{T,\Delta_T}(\widehat{\theta}_{T,\Delta_T}^{(m)})$.

Proof of Corollary 3: This follows from Theorem 3 and Theorem 4 and that

$$I_{T,\Delta_T}^{1/2}(\theta_0) \left(\widehat{\theta}_{T,\Delta_T}^{(m)} - \theta_0 \right) = I_{T,\Delta_T}^{1/2}(\theta_0) \left(\widehat{\theta}_{T,\Delta_T}^{(m)} - \widehat{\theta}_{T,\Delta_T} \right) + I_{T,\Delta_T}^{1/2}(\theta_0) \left(\widehat{\theta}_{T,\Delta_T} - \theta_0 \right).$$

Proof of Proposition 2: By the non-negativity of $f(\cdot)$,

$$\begin{aligned} -f(X_0) &\leq E[f(X_t)] - f(X_0) \\ &= \int_0^t E[A^B f(X_s)] ds. \end{aligned}$$

$\lim_{|x| \rightarrow \infty} A^B f(x) = -\infty$ implies for $|x| > r$, $A^B f(x)$ is bounded above by a constant $-C_r$ where $C_r \rightarrow \infty$ as $r \rightarrow \infty$. Let $\sup_x A^B f(x) = k < \infty$. Therefore,

$$\begin{aligned} -f(X_0) &\leq -\int_0^t C_r P(|X_s| > r) ds + kt. \\ \frac{1}{t} \int_0^t P(|X_s| > r) ds &\leq -\frac{f(X_0)}{tC_r} - \frac{k}{C_r} \leq -\frac{k}{C_r} \rightarrow 0 \text{ as } r \rightarrow \infty. \end{aligned}$$

Stationarity of X now follows from Theorem III.2.1 in Has'minskiĭ (1980).

Proof of Proposition 3: Let $G(\cdot) = H(\Gamma(\cdot))$. First, we show, with probability approaching one as $d \rightarrow 0$, $\|G'\| < M(d)$ for some $M(d) \xrightarrow{d \rightarrow 0} 0$. By equation (4.1), the forward pricing satisfies

$$F_d(S, \Gamma) = S + df_1(\Gamma) + \frac{1}{2}d^2 f_2(\Gamma, \theta) + o(d^2)$$

for some function f_1 and f_2 . In particular, $f_1 = \lambda_u e^{\Gamma} E(J^u) + \lambda_d e^{-\Gamma} E(J^d)$. Notice, the parameter θ does not affect f_1 . Therefore, the function Γ from inverting this pricing formula satisfies

$$\frac{\partial \Gamma}{\partial \theta} = -\frac{\frac{\partial f_2(\Gamma, \theta)}{\partial \theta}}{\frac{\partial f_1(\Gamma, \theta)}{\partial \Gamma}} d + o(d).$$

Given Γ , the next-step iterative estimator satisfies $\frac{\partial}{\partial \theta} L(\Gamma, \theta) = 0$. Therefore,

$$\frac{\partial}{\partial \Gamma} H = -\frac{\frac{\partial^2}{\partial \theta \partial \Gamma} L}{\frac{\partial^2}{\partial \theta^2} L}.$$

θ is in a compact set and with probability approaching one, Γ stays in a compact set for given Δ and T . Therefore, $\frac{\partial}{\partial \Gamma} H$ is bounded, $\frac{\partial \Gamma}{\partial \theta}$ is $O(d)$. It is clear then we can find $M(d) \xrightarrow{d \rightarrow 0} 0$ so that $\|G'\| < M(d) < 1$ with probability approaching one as $d \rightarrow 0$. Therefore, $H(\Gamma(\cdot))$ is a contraction. The convergence of the iterative MLE procedure now follows from the contraction mapping theorem (see Theorem 3.2 in Stokey et al (1996)). Now, with probability approaching one as $d \rightarrow 0$,

$$\begin{aligned} \left\| \widehat{\theta}_{T,\Delta} - \theta_0 \right\| &\leq \left\| \widehat{\theta}_{T,\Delta} - \theta_{T,\Delta}^* \right\| + \left\| \theta_{T,\Delta}^* - \theta_0 \right\| \\ &= \left\| H\left(\Gamma\left(\widehat{\theta}_{T,\Delta}\right)\right) - H\left(\Gamma\left(\theta_0\right)\right) \right\| + \left\| \theta_{T,\Delta}^* - \theta_0 \right\| \\ &\leq M(d) \left\| \widehat{\theta}_{T,\Delta} - \theta_0 \right\| + \left\| \theta_{T,\Delta}^* - \theta_0 \right\| \end{aligned}$$

and

$$\left\| \widehat{\theta}_{T,\Delta} - \theta_0 \right\| \leq \frac{1}{1 - M(d)} \left\| \theta_{T,\Delta}^* - \theta_0 \right\|.$$

Appendix 3: Extensions of the Approximation Method

1. When some state variables do not jump

The state variables in the main part of the paper jump simultaneously. However, it is sometimes of interest to consider cases where a subset of the state variables jump while others do not¹⁸. In this case, the solution to the Kolmogorov equations no longer has an expansion in the form of (2.6). However, the methodology still applies and all we need is another form of leading term in the expansion. This new leading term, together with the closed-form approximation, will be provided below.

Let the model setup be the same as Section 2.1 except that the state variable X can be partitioned into two subsets.

$$X_{n \times 1} = \begin{pmatrix} X_{n_1 \times 1}^C \\ X_{n_2 \times 1}^D \end{pmatrix}$$

where X^D can jump and X^C cannot.

To simplify the algebra, only the following case is considered where the variance matrix can be partitioned as

$$V(X) = \begin{pmatrix} V_{11}(X^C) & 0 \\ 0 & V_{22}(X^D) \end{pmatrix}.$$

This condition simplifies the algebra below without losing any intuition on the new leading term. It can be relaxed, though details are suppressed.

The transition density in this case has an expansion in the form of

$$p(\Delta, y|x) = \Delta^{-n/2} \exp \left[-\frac{C^{(-1)}(x, y)}{\Delta} \right] \sum_{k=0}^{\infty} C^{(k)}(x, y) \Delta^k + \Delta^{-n_1/2} \exp \left[-\frac{D^{(-1)}(x^C, y^C)}{\Delta} \right] \sum_{k=1}^{\infty} D^{(k)}(x, y) \Delta^k.$$

Intuitively, when jump happens, X^C has to diffuse to the new location while X^D can jump to the new location thus the new form of leading term. (See also Section 2.2). This conjecture can be substituted into the Kolmogorov equations and solve for the unknown coefficient functions.

It turns out the functions $C^{(k)}(x, y)$ are exactly the same as those in Theorem 1 and 2. This is intuitive since a change in jump behavior will not affect terms for the diffusion part. The other coefficient functions are characterized below.

$$\begin{aligned} D^{(-1)}(x^C, y^C) &= \frac{1}{2} \left[\frac{\partial}{\partial x^C} D^{(-1)}(x^C, y^C) \right]^T V_{11}(x^C) \left[\frac{\partial}{\partial x^C} D^{(-1)}(x^C, y^C) \right] \\ D^{(1)}(x, y) &= \lambda(x) \nu(y-x) - D^{(1)} \left[L^B D^{(-1)} - \frac{n_1}{2} \right] - \left[\frac{\partial}{\partial x} D^{(-1)}(x, y) \right]^T V(x) \left[\frac{\partial}{\partial x} D^{(1)}(x, y) \right] \\ D^{(k+1)}(x, y) &= \frac{1}{1+k} \left\{ \begin{aligned} &A^B D^{(k)} + (2\pi)^{n_2/2} \lambda(x) \sum_{r=0}^k \frac{1}{(2r)!} \sum_{s \in S_{2r}^{n_2}} M_s^{n_2} \frac{\partial^s}{\partial w^s} g_{k-r}(x, y, w) \Big|_{w=0} \\ &- D^{(k+1)} \left[L^B D^{(-1)} - \frac{n_1}{2} \right] - \left[\frac{\partial}{\partial x} D^{(-1)} \right]^T V(x) \left[\frac{\partial}{\partial x} D^{(k)}(x, y) \right] \end{aligned} \right\} \\ &\text{for } k > 0 \end{aligned}$$

where $g_k(x, y, w) \equiv \frac{C^{(k)} \left(\left(\begin{pmatrix} x^C \\ w_B^{-1}(w) \end{pmatrix}, y \right) \nu(w_B^{-1}(w) - x^D) \right)}{\left| \det \frac{\partial}{\partial (x^D)^T} w_B(x^D, y^D) \Big|_{x^D = w_B^{-1}(w)} \right|}$, $w_B(x^D, y^D) \equiv \left[V_{22}^{1/2}(x^D) \right]^T \frac{\partial}{\partial x^D} H(x^D, y^D)$. The function H satisfies $H(x^D, y^D) = \frac{1}{2} \left[\frac{\partial}{\partial x^D} H(x^D, y^D) \right]^T V_{22}(x^D) \left[\frac{\partial}{\partial x^D} H(x^D, y^D) \right]$. Fixing y^D , $w_B(\cdot, y^D)$ is

¹⁸As an example, one could model an asset price as a jump diffusion with stochastic volatility where the volatility process follows a diffusion.

invertible in a neighborhood of $x^D = y^D$ and $w_B^{-1}(\cdot)$ is its inverse function in this neighborhood. (For the ease of notation, the dependence of $w_B^{-1}(\cdot)$ on y^D is not made explicit.)

2. State-dependent jump distribution

Now let us consider a process characterized by its infinitesimal generator A^B on a bounded function f with bounded and continuous first and second derivatives

$$A^B f(x) = \sum_{i=1}^n \mu_i(x) \frac{\partial}{\partial x_i} f(x) + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n v_{ij}(x) \frac{\partial^2}{\partial x_i \partial x_j} f(x) + \lambda(x) \int_C [f(x+c) - f(x)] \nu(x, c) dc.$$

The jump distribution $\nu(x, \cdot)$ is now state-dependent. The solution of the Kolmogorov equations admits an expansion in the form of (2.6). It turns out that the coefficient functions $C^{(k)}(x, y)$ and $D^{(k)}(x, y)$ in this case are characterized by Theorem 1 and Corollary 1, with the exception that $\nu(\cdot)$ is replaced everywhere by $\nu(x, \cdot)$.

3. Multiple jump types

Consider a process whose infinitesimal generator A^B , on a bounded function f with bounded and continuous first and second derivatives, is given by

$$A^B f(x) = \sum_{i=1}^n \mu_i(x) \frac{\partial}{\partial x_i} f(x) + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n v_{ij}(x) \frac{\partial^2}{\partial x_i \partial x_j} f(x) + \sum_u \lambda_u(x) \int_C [f(x+c) - f(x)] \nu_u(x, c) dc.$$

This is a process with multiple jump types with different jump intensity $\lambda_u(\cdot)$ and jump distribution $\nu_u(x, \cdot)$. The solution of the Kolmogorov equations also admits an expansion in the form of (2.6) where the coefficient functions $C^{(k)}(x, y)$ and $D^{(k)}(x, y)$ are given below.

$$\begin{aligned} 0 &= C^{(-1)}(x, y) - \frac{1}{2} \left[\frac{\partial}{\partial x} C^{(-1)}(x, y) \right]^T V(x) \left[\frac{\partial}{\partial x} C^{(-1)}(x, y) \right] \quad \text{and } C^{(-1)}(x, y) = 0 \text{ when } y = x \\ 0 &= C^{(0)} \left[L^B C^{(-1)} - \frac{n}{2} \right] + \left[\frac{\partial}{\partial x} C^{(-1)}(x, y) \right]^T V(x) \left[\frac{\partial}{\partial x} C^{(0)}(x, y) \right] \\ 0 &= C^{(k+1)} \left[L^B C^{(-1)} + (k+1) - \frac{n}{2} \right] + \left[\frac{\partial}{\partial x} C^{(-1)}(x, y) \right]^T V(x) \left[\frac{\partial}{\partial x} C^{(k+1)}(x, y) \right] \\ &\quad + \left[\sum_u \lambda_u(x) - L^B \right] C^{(k)} \quad \text{for nonnegative } k \\ 0 &= D^{(1)} - \sum_u \lambda_u(x) \nu_u(x, y-x) \\ 0 &= D^{(k+1)} - \frac{1}{1+k} \left[A^B D^{(k)} + (2\pi)^{n/2} \sum_{r=0}^k \frac{1}{(2r)!} \sum_u \lambda_u(x) \sum_{s \in S_{2r}^n} M_s^n \frac{\partial^s}{\partial w^s} g_{u, k-r}(x, y, w) \Big|_{w=0} \right] \quad \text{for } k > 0 \end{aligned}$$

where $g_{u, k}(x, y, w) \equiv \frac{C^{(k)}(w_B^{-1}(w), y) \nu_u(w_B^{-1}(w) - x)}{\left| \det \frac{\partial}{\partial x^T} w_B(x, y) \Big|_{x=w_B^{-1}(w)} \right|}$, $w_B(x, y) \equiv [V^{1/2}(x)]^T \frac{\partial}{\partial x} C^{(-1)}(x, y)$. Fixing y , $w_B(\cdot, y)$ is invertible in a neighborhood of $x = y$ and $w_B^{-1}(\cdot)$ is its inverse function in this neighborhood. (For the ease of notation, the dependence of $w_B^{-1}(\cdot)$ on y is not made explicit.)

Appendix 4: History of the Yuan's Exchange Rate Regime¹⁹

Table 2 documents the history of the Yuan's exchange rate regime since its creation²⁰.

Table 2: History of the Yuan's Exchange Rate Regime

Date	Exchange Rate Regime	Official rate
March 1, 1955	The Yuan was created and pegged to USD	2.46
August 15, 1971	A new official rate against the USD was announced.	2.267
February 20, 1973	The official rate (the "Effective Rate") was realigned.	2.04
August 19, 1974 — December 31, 1985	The Effective Rate was pegged to a trade-weighted basket of fifteen currencies with undisclosed compositions. The rate was fixed almost daily against that basket.	1.5 – 2.8
January 1, 1986	The Effective Rate was placed on a controlled float.	
July 5, 1986	The Effective Rate was fixed against the USD until December 15, 1989	3.72
November 1986	A second rate, Foreign Exchange Swap Rate determined by market force, was created. The Foreign Exchange Swap Rate quickly moved to 5.2 in one month.	3.72
December 15, 1989	The Effective Rate was realigned.	4.72
November 17, 1990	The Effective Rate was realigned.	5.22
April 9, 1991	The Effective Rate start to adjust frequently depending on certain economic indicators.	
December 31, 1993	The Effective Rate was 5.8. The Foreign Exchange Swap Rate was 8.7.	5.8
January 1, 1994	The Effective Exchange Rate and the swap market rate were unified at the prevailing swap market rate. Daily movement of the exchange rate of the Yuan against the USD is limited ot 0.3% on either side of the reference rate as announced by the PBC.	

¹⁹ Courtesy of the Economics Department at the Chinese University of Hong Kong.

²⁰ Official rate is quoted in Yuan/USD.

Appendix 5: News Release on Jump Days

Table 3 documents the news releases that coincide with the jumps in Γ .

Table 3: News Release on the Jump Days

Date	Events	News Source
3-8-2002	1. State-owned enterprises, which contribute 60 percent of China's tax revenue, see a profit drop of 1.4% last year. 2. China plans to incur 80 billion yuan, a quarter of its planned 2003 deficit, on restructuring SOEs.	ChinaOnline InfoProd
1-6-2003	1. China's foreign trade increased 21 percent over 2002 with surplus expanding to USD 27 billion for the first 11 month of 2002. 2. China's tax revenue rose 12.1% year-over-year with GDP estimated to grow by 8%.	Business Daily China Daily
6-13-2003	Goldman Sachs forecasts an appreciation of the yuan to 8.07 by the year-end based on the view that China was preparing for a change in currency policy, as SARS and deflationary pressures recede and the dollar weakens.	AFX News
8-29-2003	1. Japanese Finance Minister urges China to let the yuan float freely before meeting US treasury secretary. 2. China's state-owned enterprises posted a profit of 212 billion yuan from Jan to July, up 47% year-over-year.	Japan Economic Newswire AFX News
9-2-2003	US Treasury Secretary wraps up first day of talks in China, no comments on the yuan.	AFX news
9-23-2003	G-7 requests China to move rapidly towards more flexible exchange rates.	South China Morning Post
10-7-2003	The People's Bank of China is considering revaluing the yuan by about 30 percent over the next five years, subject to a final decision from the State Council led by Premier Wen Jiabao.	Jiji Press Japan
10-8-2003	China's Premier Wen Jiabao suggested he would resist foreign pressure to revalue the yuan.	Xinhua News