Jump-Diffusion Model of Exchange Rate Dynamics — Estimation via Indirect Inference

George J. Jiang *

May, 1998

Abstract

Jump-diffusion processes have been widely used to model financial time series to reflect discontinuity of asset returns. However, difficulty involved in the estimation of general jump-diffusion processes has prevented their implementation in empirical applications. This paper proposes the simulation-based indirect inference approach to the estimation of general parametric continuous-time jump-diffusion processes from discretely observed data. Applications to currency exchange rate models are undertaken to illustrate the estimation procedure and to present interesting empirical results. The estimation results suggest that jumps are important components of the currency exchange rate dynamics even when conditional heteroscedasticity and mean-reversion are taken into account. However, models assuming conditional homoscedasticity tend to overestimate the jump frequency and underestimate the jump size.

Keywords: Jump-Diffusion, Indirect Inference, Exchange Rate Dynamics

* I wish to thank John Knight, Pieter van der Sluis and Tom Wansbeek for helpful suggestions and comments. I would also like to thank the seminar participants at the Tinbergen Institute in Amsterdam and the CentER at Tilburg University for helpful discussion and comments. The usual disclaimer applies. Department of Econometrics, University of Groningen, P.O. Box 800, 9700 AV Groningen, The Netherlands, e-mail: G.Jiang@eco.rug.nl
1. Introduction

Jump-diffusion processes have been widely used to model financial time series to reflect discontinuity of asset returns. However, difficulties involved in the estimation of general jump-diffusion processes have prevented their implementation in empirical applications. This paper proposes the indirect inference approach, developed by Gouriéroux, Monfort and Renault (1993), and Gallant and Tauchen (1996), to the estimation of the general parametric continuous-time jump-diffusion processes from discretely observed data. The associated instrumental or auxiliary model is an approximate discrete model which can be easily estimated via maximum likelihood (ML) estimation method and the discretization biases of the estimators are then corrected based on simulations. Applications of the indirect estimation method to currency exchange rate models are undertaken to illustrate its implementation on the one hand and to present interesting empirical results on the other. The estimation results suggest that jumps are important components of the currency exchange rate dynamics even when conditional heteroscedasticity and mean-reversion are taken into account. However, models assuming conditional homoscedasticity tend to overestimate the jump frequency and underestimate the jump size.

The paper is organized as follows: Section 2 details the basic assumptions and properties of the continuous-time jump-diffusion processes with presentation of two examples, and briefly reviews the available estimation methods; Section 3 proposes the indirect inference approach to the estimation of general parametric jump-diffusion processes via an instrumental or auxiliary model from discretely observed data; in Section 4, applications of the indirect estimation method to various currency exchange rate models are undertaken to illustrate its implementation on the one hand and to present interesting empirical results on the other.

2. Jump-Diffusion Processes and Available Estimation Methods

2.1 The Model: Jump-Diffusion Processes

The general parametric jump-diffusion processes, as a mixture of both continuous diffusion path and discontinuous jump path, can be written as:

\[
dS_t / S_t = (\alpha_t(\beta) - \lambda \mu_0)dt + \sigma_t(\beta) dW_t + (Y_t(\beta) - 1)d\xi_t(\lambda)
\]

(1)

where

- \( S_t \) — asset price at time \( t \);
- \( \alpha_t \) — the instantaneous expected return;

...
\( \sigma_t^2 \) — the instantaneous volatility of the asset’s return conditional on that the Poisson jump event does not occur;

\( W_t \) — a standard Gauss-Wiener process or the Brownian motion process;

\( q_t(\lambda) \) — a Poisson process which is iid over time;

\( \lambda \) — the intensity parameter of the Poisson distribution;

\( Y_t \) — the random jump size with \( Y_t > 0 \);

\( \mu_0 \) — expectation of the relative jump size, i.e. \( \mu_0 = E[Y_t - 1] \);

\( dq_t(\lambda), dW_t \) — statistically independent;

\( \theta = (\beta, \mu_0, \lambda) \in \Theta \) — the parameter space which parameterizes the coefficient functions, the jump sizes, as well as the intensity of Poisson process.

Alternatively, (1) can be rewritten in terms of the logarithmic asset prices, i.e., \( s_t = \ln S_t \), as:

\[
ds_t = \mu_t(\beta)dt + \sigma_t(\beta)dW_t + \ln Y_t(\beta) dq_t(\lambda)
\]

where \( \mu_t = \alpha_t - \lambda \mu_0 - \frac{1}{2} \sigma_t^2 \).

The jump-diffusion process defined in (2) is a Markov process with discrete parameter space and continuous state space. Such a jump-diffusion process can be used to approximate a wide range of Markovian or non-Markovian processes. In particular, the weak convergence of a driving noise process to a Lévy process implies the weak convergence of the driving noise process to a jump-diffusion process (see e.g. Jacob and Shiryaev, 1987, Fujiwara, 1990, and Blasikiewicz and Brown, 1996). Both the Wiener process \( W_t \) and the compound Poisson process \( \int_0^t \ln Y_t dq_t(\lambda) \) are infinitely-divisible in time, appropriately scaled, and have independent increments. It is noted, however, that even though \( W_t \) is a martingale process, the compound Poisson process is in general not a martingale. Define the compensated Poisson process

\[
d\tilde{q}_t(\lambda) = dq_t(\lambda) - \lambda dt
\]

since \( E[\int_{t}^{t+\Delta t} \ln Y_t d\tilde{q}_t(\lambda)] = E[\int_{t}^{t+\Delta t} \ln Y_t (dq_t(\lambda) - \lambda d\tau)] = 0 \), i.e. \( \int_{t}^{t+\Delta t} \ln Y_t d\tilde{q}_t(\lambda) \) is a martingale process. Therefore, without loss of generality, we only need to consider the case that the compound Poisson process is a martingale process.

Under certain regularity conditions, such as the Lipschitz condition and linear growth condition for the coefficient functions (see Jiang, 1996), solution of (2) is a Markov process with killing time equal to infinity, satisfying

\[
s_t = s_0 + \int_0^t \mu_t d\tau + \int_0^t \sigma_t dW_t + \int_0^t \ln Y_t dq_t(\lambda)
\]

The solution is unique in probability, i.e. let solutions \( s_t^1 \) and \( s_t^2 \) of (2) corresponding to initial values \( s_0^1 \) and \( s_0^2 \), then \( P_{s_0^1,t}^1, s_0^2 \in [s^2_{t+\Delta}, s^2_{t+\Delta}] \geq \epsilon \rightarrow 0 \) as \( |s^1_0 - s^2_0| \rightarrow 0 \) for any \( \epsilon > 0, \Delta > 0 \). In addition, the process is uniformly bounded over fi-
nite time interval, essentially discontinuous but continuous from right a.s. Let \( 0 \leq t \leq T < \infty \), and initial condition \( s_0 \), we have \( E_{s_0} \left[ \max_{0 \leq \Delta t \leq \Delta} |s_{t+\Delta t} - s| > \epsilon \right] = O(\Delta) \), where \( O(\cdot) \) is uniform in \( t \) and \( s \), but depends on \( \epsilon \) (see, e.g., Kushner, 1967; p.19). Using the above results, it can be shown that \( s_t \) is a Feller process and, since the paths are continuous from right, also a strong Markov process.

Special cases of the model defined in (2) include those by Press (1967) with \( t = 0 \); \( \mu = 0 \), Merton (1976) with \( t = \sigma = \mu = \mu_0 = 0 \), Beckers (1981) and Ball and Torous (1985) with \( t = \sigma = \mu = \mu_0 = 0 \), and Lo (1988) with \( \ln Y_t = \kappa(s_t) \), i.e. the jump size is determined by the process itself.

2.2 Examples of Jump-Diffusion Processes

Example 1: Geometric Brownian Motion with i.i.d. Lognormal Jump (Merton (1976a)). The model assumes that in (2) \( \mu_1(\cdot) = \mu \), \( \sigma_1(\cdot) = \sigma \), and \( \ln Y_t \sim i.i.d.N(\mu_0, \nu^2) \), where \( \mu \), \( \mu_0 \), \( \sigma \), \( \nu \) are all constants. This model results in the asset return, \( x = \ln S_t/S_{t-1} = s_t - s_{t-1} \), with density given by

\[
p(x) = \sum_{n=0}^{\infty} \frac{e^{-\lambda_n} \phi(x; \mu + n\mu_0, \sigma^2 + n\nu^2)}{n!}
\]

which has an infinite series representation, where \( \phi(x; \mu, \delta^2) = (2\pi \delta^2)^{-1/2} \exp(-(x - \mu)^2/2\delta^2) \). Let \( \varphi_x(u) \) denote the characteristic function of the asset return \( x \), then

\[
\ln \varphi_x(u) = \mu u - \frac{1}{2} \sigma^2 u^2 + \lambda(\exp(\mu_0 u) - \frac{1}{2} \nu^2 u^2) - 1
\]

It is easy to derive that \( E[x] = \mu + \lambda \mu_0 \), \( Var[x] = \sigma^2 + \lambda (\mu_0^2 + \nu^2) \) and the distribution of \( x \) is leptokurtic, more peaked in the vicinity of its mean than the distribution of a comparable normal random variable, asymmetric if \( \mu_0 \neq 0 \), and the skewness has the same sign as that of \( \mu_0 \). Furthermore, \( s_t \) is a process with stationary and independent increments and \( E[s_t] = C + (\mu + \lambda \mu_0)t \), \( Var[s_t] = (\sigma^2 + \lambda (\mu_0^2 + \nu^2))t \), \( Cov[s_t, s_{t'}] = (\sigma^2 + \lambda (\mu_0^2 + \nu^2))\min(t, t') \).

Example 2: Ornstein-Uhlenbeck Process with Exponentially Decaying Jump. This model assumes that in (2) \( \mu(s, t) = -\eta s \), where \( \beta > 0 \), \( \sigma(s, t) = \sigma \), and \( \ln Y_t \sim i.i.d.N(0, e^{-2\beta(t-t_0)}\nu^2) \), where \( t_0 \) is the initial time. The process is governed by the following transition density function

\[
f(s(t+1)|s_t) = \sum_{n=0}^{\infty} \frac{e^{-\lambda_n} \phi(s(t+1); e^{-\beta}s, \frac{\sigma^2}{2\beta}(1 - e^{-2\beta})} \phi(s_t; e^{-\beta}s_t, \frac{\sigma^2}{2\beta}(1 - e^{-2\beta}))}{n!}
\]

where \( \lambda_n \) is a sequence of positive numbers such that \( \sum_{n=0}^{\infty} \frac{e^{-\lambda_n} \phi(s(t+1); e^{-\beta}s, \frac{\sigma^2}{2\beta}(1 - e^{-2\beta})\phi(s_t; e^{-\beta}s_t, \frac{\sigma^2}{2\beta}(1 - e^{-2\beta}))}{n!} \).
The conditional mean and variance are given by \( E[s(t + 1)|s_t] = e^{-\beta} s_t, \) \( Var[s(t + 1)|s_t] = \frac{\sigma^2}{\theta^2} (1 - e^{-2\beta}) + \lambda e^{-2\beta(t-t_0)} v^2. \) This process converges to the Ornstein-Uhlenbeck process since the jump component dies out as time goes to infinity.

Suppose \( s(t_0) \sim N(0, \frac{\sigma^2}{\theta^2}), \) then \( E[s_t] = 0, \) \( Var[s_t] = \frac{\sigma^2}{\theta^2} + \lambda e^{-2\beta(t-t_0)} v^2(t - t_0), \) \( Cov[s_t, s(\tau)] = e^{-\beta(t-t_0)} (\frac{\sigma^2}{\theta^2} + \lambda e^{-2\beta(t-t_0)} v^2(t - t_0)). \) As \( t_0 \to -\infty, \) or \( t - t_0 \to +\infty, s_t \) converges to a stationary process, i.e. the Ornstein-Uhlenbeck process, with marginal density following a normal distribution \( N(0, \frac{\sigma^2}{\theta^2}). \) Further difference from the geometric Brownian motion with i.i.d. lognormal jump is that the Ornstein-Uhlenbeck process with exponentially decaying jump no longer has independent increments even thought its driving processes, the Brownian motion and compound Poisson, both do. Let \( \hat{s}(t) = s_t - S_{t-1}, \) we have \( E[\hat{s}(t)] = 0, \) \( Var[\hat{s}(t)] = (1 - e^{-\beta})^2 (\frac{\sigma^2}{\theta^2} + \lambda e^{-2\beta(t-t_0)} v^2(t - t_0 - 1)) + \frac{\sigma^2}{\theta^2} (1 - e^{-2\beta}) + \lambda e^{-2\beta(t-t_0)} v^2, \) \( Cov[\hat{s}(t), \hat{s}(\tau)] = e^{-\beta(t-t_0)} ((1 - e^{-\beta})(\frac{\sigma^2}{\theta^2} + \lambda e^{-2\beta(t-t_0)} v^2(t - t_0)) + (1 - e^{-\beta})(\frac{\sigma^2}{\theta^2} + \lambda e^{-2\beta(t-t_0)} v^2(t - t_0 - 1)) \neq 0 \) for \( \tau \leq t - 1. \) Again as \( t_0 \to -\infty, \) or \( t - t_0 \to +\infty, \) we have \( E[\hat{s}(t)] = 0, \) \( Var[\hat{s}(t)] \to (1 - e^{-\beta}) \frac{\sigma^2}{\theta^2}, \) \( Cov[\hat{s}(t), \hat{s}(\tau)] \to -e^{-\beta(t-t_0)} (1 - e^{-\beta})^2 \frac{\sigma^2}{\theta^2} < 0 \) for \( \tau \leq t - 1, \) and \( Corr[\hat{s}(t), \hat{s}(\tau)] \to -\frac{1}{2} e^{-\beta(t-t_0)} (1 - e^{-\beta}) < 0 \) for \( \tau \leq t - 1. \)

### 2.3 Available Estimation Methods of Jump-Diffusion Processes

Identification of the jump-diffusion process can be complicated if its coefficient functions are not well behaved enough for the Markov transition density functions to be well defined. Appendix A.1 imposes regularity conditions to ensure the process in (2) is a well-defined strong Markov process. Various estimation methods of jump-diffusion models have been developed in the literature based on restrictive specification of coefficient functions. The problem with empirically estimating the parameters based on arbitrarily chosen moments of jump-diffusion models, such as the method of cumulants matching, the method of moments or GMM, originates from the difficulty in distinguishing between whether movements in the underlying process are part of the continuous path dynamics, or whether they are part of the jump path dynamics. It might be possible to estimate the parameters through minimizing a penalty function of the prediction error, e.g. the implied parameter estimation method used in finance literature based on fitting the option pricing formula to market option prices, but it is unlikely that the identification problem will be resolved. In addition, to apply the implied parameter estimation method, researchers have to assume a simple structure for the underlying model in order to obtain closed form or tractable solutions of option prices. From an econometrics point of view, these identification problems are important. If there is a way of disentangling the jumps from the rest of the process, then the parameters could be more accurately estimated. Identifying the jumps without any
restrictions, however, is an ad-hoc task by the researcher, and prone to introducing sample selection error to the model. In the rest of this section, we will briefly review the available estimation methods proposed in the literature for jump-diffusion processes, namely the method of cumulant matching, the method of moment or GMM, and the ML method.

Press (1967) proposed a simplified jump-diffusion model by, a priori, constraining that $\mu(\cdot) = 0$, $\sigma(\cdot) = \sigma$, and $lnY \sim i.i.d. N(\mu_0, \nu^2)$. These restrictions result in the asset return, $x = ln(s_t/S_{t-1}) = s_t - S_{t-1}$, with density given by (5) with $\mu = 0$. Since the likelihood function has an infinite series representation, the ML method does not yield explicit estimators. Press proposes the method of “cumulant matching”, a variant of the method of moments, which sets the sample cumulants equal to the population cumulants, and solves the resultant equations for the parameter estimators (see Appendix A.2). Parameter estimation by cumulant matching is known to yield consistent estimators, but it is known that they are not always efficient. Employing the method of cumulants to a sample of monthly returns of ten NYSE listed common stocks over the period 1926 through 1960, Press frequently obtains negative estimates of the variance parameters $\sigma^2$ and $\nu^2$. Although Press (1967) believes that a larger sample size will remedy the inherent estimation problems, Beckers (1981) contends that it is highly unlikely to be true. Beckers (1981) modified the Press procedure and set the mean logarithmic jump size equal to zero, i.e. $\mu_0 = 0$, and $\mu(\cdot) = \mu$. His assumptions yield a symmetric density of the asset return, $x = ln(s_t/S_{t-1}) = s_t - S_{t-1}$, given by (5) with $\mu_0 = 0$. Due to symmetry of the distribution, the odd cumulants except the first one all vanish. Applying these estimators to 47 NYSE listed common stocks, each with 500 daily return observations over September 15, 1975 to September 7, 1977, Beckers also often obtains negative estimates of the variance parameters $\nu^2$ and $\sigma^2$. Only for stocks with high sample kurtosis, this estimation generates sensible parameter values.

Technically speaking, estimation of the jump-diffusion process can be conveniently performed via ML method. This is because, due to the Markov nature of the process, the calculation of the likelihood function from discretely sampled data which are not necessarily equally spaced in time is considerably simplified. That is, the joint density $f(s(t_0), t_0; s(t_1), t_1; \ldots; s(t_n), t_n)$ can be written as the product of conditional densities:

$$f(s(t_0); s(t_1); \ldots; s(t_n)) = f_0(s(t_0), t_0) \prod_{k=1}^{n} f_k(s(t_k), t_k|s(t_{k-1}), t_{k-1})$$

Ball and Torous (1983) considered the estimation of the jump-diffusion with Bernoulli jump process as approximation of Poisson jump process. They also impose the same restrictions as in Beckers (1981) guaranteeing a symmetric return distribution. Due to
the simplified asset return generating structure, Ball and Torous are able to implement
the ML procedure for the parameter estimation. They apply it to Beckers’ sample of
NYSE listed common stocks. Significantly, all variance estimates are positive and
standard errors of the estimates are readily available. Ball and Torous (1985) further
extend the ML method to the Poisson jump-diffusion process as considered in Beck-
ners (1981), with the density of the asset returns given by (5) with $\mu_0 = 0$, through
numerically maximizing a truncated likelihood function. Using the data consisting
of 30 NYSE listed stocks, each with 500 daily return observations over the period
from January 1, 1981 to December 31, 1982. The ML estimation based on truncated
likelihood function provides consistently positive estimates of variance. And based
on the likelihood ratio test, the majority of stocks has significant jump feature in their
returns.

Lo (1988) considers a Poisson jump-diffusion process, with deterministic exogenous
jump-size of the compound Poisson event, i.e. $\ln Y_t = \kappa(s, \tau; \theta)$ in (2). Lo (1988)
showed that under certain regularity conditions and the condition that the function
$\tilde{\kappa}(s, \tau; \theta) = s + \kappa(s, \tau; \theta)$ is bijective and $|\frac{\partial \tilde{\kappa}}{\partial s} + 1| \neq 0$ for all $s, \tau$, and $\theta$, the transition
density function $f_k(s(t_k), t_k|s(t_{k-1}), t_{k-1})$ solves the following functional PDE:

$$\frac{\partial f_k}{\partial t} = -\frac{\partial \mu f_k}{\partial s} + \frac{1}{2} \frac{\partial^2 \sigma^2 f_k}{\partial s^2} - \lambda f_k + \lambda f_k \frac{\partial \tilde{\kappa}^{-1}}{\partial s}$$  (8)

subject to $f_k(s, t_{k-1}) = \delta(s - s(t_{k-1}))$ and any other relevant boundary conditions,
where $\tilde{f}_k = f_k(\tilde{\kappa}^{-1}, \tau)$, and $\delta(s - s(t_{k-1}))$ is the Dirac-delta generalized function
centered at $s(t_{k-1})$. However, existence and uniqueness of the solution for the func-
tional PDE (8) is not guaranteed for general coefficient functions. Moreover, even
though there exists a unique solution, an analytical form of the solution is usually un-
available. Then the solution has to rely on numerically solving the PDE and the ML
estimation of $(\theta, \lambda)$ on numerical maximization of the likelihood function, which can
be very computation intensive.

Fehr and Rosenfeld (1979) consider ML estimation of the Poisson jump-diffusion
process under the simplifying assumption that the jump size is a fixed known con-
stant. In their empirical applications, unfortunately, the results are not statistically
powerful in that the corresponding parameter estimates have standard errors which
are generally quite large. Sørensen (1991) considers ML estimation and inference
of parameters that determine the drift and the jump mechanism of a jump-diffusion
process. The non-jump conditional volatility function is assumed not containing any
unknown parameters and the time series data of the stochastic process is assumed to
be a continuously observed sample path. The likelihood function and the asymptotic
distribution of ML estimators are derived. It is also shown that the ML estimation
and inference can be largely simplified when the likelihood function or a part of it
belongs to an exponential family. In addition to the cumulant matching method and ML method, Powell (1989), Longstaff (1989) and Ho, Perraudin and Sørensen (1996) also considered GMM to estimate the parameters based on arbitrarily chosen moment conditions.

3. Indirect Estimation of Jump-Diffusion Processes

3.1 Estimation of Jump-Diffusion Processes based on Indirect Inference

This paper proposes a simulation-based indirect inference approach to the estimation of the general parametric jump-diffusion models as specified in (2) where \( s_t \) is assumed to be a stationary regular Markov process and \( Y_t \) follows a lognormal distribution. It is noted that for more general distributions of \( Y_t \), the estimation described below can also be followed except with different random generator of \( Y_t \).

The basic idea of indirect inference method proposed by Gourieroux, Monfort and Renault (1993), and Gallant and Tauchen (1996) is that when a model leads to a complicated structural or reduced form and therefore to intractable likelihood functions, estimation of the original model \( (M_o) \) can be indirectly achieved by estimating an instrumental or auxiliary model \( (M_i) \) which is constructed as an approximation of the original one. The estimation of jump-diffusion process based on indirect inference can be summarized as following four steps.

Step 1: Choice of the instrumental or auxiliary model: For the continuous-time jump-diffusion process (2), a natural choice of the instrumental or auxiliary model is its simple discretization. However, when \( \lambda \Delta_i \) is small \((< 1)\) as in our application, a much simpler instrumental or auxiliary model is given by

\[
  s_{i+1} = s_i + \mu(t_i; \beta_t) \Delta_i + \sigma(t; \beta_t) \Delta_i^{1/2} \epsilon_{i+1}^0 + \eta(\mu_{0t} + v_t \epsilon_{i+1}^1)
\]  

with instrumental parameter \( \theta_I = (\beta_I, \mu_{0t}, v_I, \lambda_I) \in \Theta_I \) - the parameter space, where \( s_i = s_{ti}, \forall i, \Delta_i = t_{i+1} - t_i, \epsilon_{i+1}^j \sim iid N(0, 1), j = 0, 1; i = 0, 1, \ldots, M - 1, \) and \( \eta \sim \) Bernoulli distribution with \( P(\eta = 1) = \lambda_I \Delta_i \) and \( P(\eta = 0) = 1 - \lambda_I \Delta_i \).

The model defined in (9) has a one-to-one relationship with (2), and the parameter space \( \Theta_I \) has the same dimension as \( \Theta \), i.e. \( dim(\Theta_I) = dim(\Theta) \). Both models are Markov processes in continuous state space, with (2) being in continuous-time while (9) in discrete-time. The conditional density function of \( s_{i+1} \) given \( s_i \) for the instrumental model can be written as

\[
  f_I(s_{i+1} | s_i; \theta_I) = (1 - \lambda_I \Delta_i) \phi_1(s_{i+1}) + \lambda_I \Delta_i \phi_2(s_{i+1})
\]  

where \( \phi_1(\cdot) \) is the pdf of the normal distribution with mean \( s_i + \mu(t_i; \beta_t) \Delta_i \) and variance \( \sigma(t; \beta_t)^2 \Delta_i \) and \( \phi_2(\cdot) \) is the pdf of the normal distribution with mean \( s_i + \mu(t_i; \beta_t) \Delta_i + \sigma(t; \beta_t) \lambda_I \Delta_i \Delta_i^{1/2} \epsilon_{i+1}^0 \).
\( \mu(t; \beta) \Delta t + \mu_{0 \text{i}} \) and variance \( \sigma(t; \beta)^2 \Delta t + \nu_i^2 \).

**Step 2: Estimation of the instrumental or auxiliary model:** The ML estimator of \( \theta_{\text{i}} \) is given by

\[
\hat{\theta}_{\text{i}} = \arg \max_{\theta_{\text{i}}} \sum_{i=0}^{M-1} \ln f_t(s_{i+1} | s_i; \theta_{\text{i}}) \tag{11}
\]

which is much easier to implement than the Poisson driven jump-diffusion process.

**Step 3: Path simulation of the original jump-diffusion process and estimation of the instrumental model based on simulated sampling path:** Since the instrumental or auxiliary model is the discretized approximation of the original one, i.e. \( \mathbf{M}_I \) and \( f_I(\cdot) \) are misspecified, the pseudo maximum likelihood (PML) estimator \( \hat{\theta}_{\text{i}} \) is generally biased and inconsistent estimator of the true parameter \( \theta \). The indirect inference estimation method uses simulations performed under the original model to correct for the asymptotic biases of \( \hat{\theta}_{\text{i}} \). Given values of \( \theta \) and initial values of \( s_t \) at \( t = t_0 \), we can simulate the sampling path \( \tilde{s}_t \) of \( s_t \), observed at \( (t_0, t_1, \ldots, t_M) \), for the original jump-diffusion model (2). As in most simulation exercises, we can redraw such simulations \( H \) times. Then we estimate the parameter \( \theta_{\text{i}} \) of the instrumental model, denoted by \( \hat{\theta}_{\text{i}}^{\text{HM}}(\theta) \), from the observations of the simulated sampling path via ML method

\[
\hat{\theta}_{\text{i}}^{\text{HM}}(\theta) = \arg \max_{\theta_{\text{i}}} \sum_{h=1}^{H} \sum_{i=0}^{M-1} \ln f_t(\tilde{s}_{i+1}^h(\theta) | \tilde{s}_i^h(\theta); \theta_{\text{i}}) \tag{12}
\]

Simulation of the exact dynamic sampling path of the jump-diffusion model defined in (2) is in general impossible unless its transition density functions are explicitly known whereas in this case the model could be estimated directly. Hence in general we can only simulate approximated sampling path of the jump-diffusion model. Discrete approximations of the continuous-time jump-diffusion process can be obtained through the Euler scheme. Alternatively, the Milstein scheme can be used for the continuous part of the process as it has better convergence rate than the Euler scheme for the convergence in \( L^p(\Omega) \) and the almost sure convergence, see (Talay, 1996).

**Step 4: Indirect estimator of jump-diffusion parameter:** An indirect estimator of \( \theta \) based on \( M \) observations of the simulated sampling path with \( H \) drawings, denoted by \( \hat{\theta}_{\text{i}}^{\text{HM}} \), is defined by choosing values of \( \theta \) from which \( \hat{\theta}_{\text{i}} \) and \( \hat{\theta}_{\text{i}}^{\text{HM}}(\theta) \) are as close as possible, i.e.

\[
\hat{\theta}_{\text{i}}^{\text{HM}}(\Omega) = \arg \min_{\hat{\theta}_{\text{i}}} (\hat{\theta}_{\text{i}} - \hat{\theta}_{\text{i}}^{\text{HM}}(\theta)) \Omega (\hat{\theta}_{\text{i}} - \hat{\theta}_{\text{i}}^{\text{HM}}(\theta))^t \tag{13}
\]

where \( \Omega \) is a symmetric nonnegative matrix, defining the metric or the weighting scale. The functions \( \hat{\theta}_{\text{i}}^{\text{HM}}(\theta) \) relate the parameters of the original model to the esti-
Computation of the standard derivation of the estimator is based on the so-called binding function. Consider the criterion in the asymptotic optimization problem, i.e. given a set of observations of $s_t$ at $\{t_0, t_1, \ldots, t_M\}$, assume that $\lim_{M \to \infty} \frac{1}{M} \sum_{i=0}^{M-1} \ln f_I(s_{i+1}|s_i; \theta_I) = E_\theta[\ln f_I(s_{i+1}|s_i; \theta_I)]$, which, as a function of $\theta$ and $\theta_I$, has a deterministic limit. The solution of this asymptotic problem defined by

$$b(\theta) = \arg \max_{\theta_I} E_\theta[\ln f_I(s_{i+1}|s_i; \theta_I)]$$

(14)

is called the binding function. Since $M_I$ and $M_O$ have a one-to-one relationship given well-defined coefficient functions and $M_O$ is the limiting case of $M_I$, $b(\theta)$ is thus in general injective. Thus, given pseudo-true values of $\theta_0$ to which $\hat{\theta}_I$ converges in probability, there are unique true values of $\theta$, say $\theta_0$, given by $b(\theta_0) = \theta_0$. Instead of estimating $\theta$ through the PML estimators of $\theta_I$, one can also consider using directly the score functions of the instrumental or auxiliary model. An equivalent efficient method of moments (EMM) approach proposed by Gallant and Tauchen (1996) selects values of $\theta$ such that $\sum_{h=1}^{H} \sum_{l=0}^{M-1} \frac{\partial \ln f_I}{\partial \theta}(s_{i+1}^h(\theta); \hat{\theta}_I)$ is as close as possible to zero, which defines another set of indirect estimator of $\theta$. If the gradient of the score functions also have a closed form, the Gallant and Tauchen (1996) method has computational advantage because it necessitates only one optimization in $\theta$. However, in our case the gradient of the score functions does not have a closed form. Moreover, the EMM technique involves the construction of the semi-nonparametric (SNP) score generator which embeds the original model. This can be much more computationally intensive in the estimation of jump-diffusion processes. Gouriéroux, Monfort and Renault (1993) show that, under certain regularity conditions, the indirect inference estimator $\hat{\theta}^{HM}(\Omega)$ is consistent, asymptotically normal, when $H$ is fixed and $M$ goes to infinity:

$$\sqrt{M}(\hat{\theta}^{HM}(\Omega^*) - \theta) \xrightarrow{d} N(0, V(H, \Omega^*))$$

(15)

where $V(H, \Omega^*) = (1 + \theta_H)^{-1} \frac{\partial b(\theta)}{\partial \theta} \Omega^* \frac{\partial b(\theta)}{\partial \theta}^{-1}$, and $\Omega^*$ is the optimal weighting matrix. The choice of $\Omega$ in the above application is the identity matrix as the dimensions of the parameter space are equal for the original model and the instrumental model, i.e. $\text{dim}(\Theta) = \text{dim}(\Theta_I)$.

### 3.2 Path Simulation of Jump-Diffusion Processes

Simulation of the exact dynamic sampling path of the jump-diffusion model defined in (2) is in general impossible unless its transition density functions are explicitly known whereas in this case the model could be estimated directly. Hence in general we can only simulate approximated sampling path of the jump-diffusion model.
crete approximations of the continuous-time jump-diffusion process can be obtained through two equivalent ways: one is to divide the time interval \([t_i, t_{i+1}]\) further into very small subintervals, i.e \([t_i + k \Delta_i/n, t_i + (k + 1) \Delta_i/n]\), where \(k = 0, 1, ..., n - 1; i = 0, 1, ..., M - 1\) and \(n\) is a large number. Or equivalently construct step functions of the coefficient functions as \(\mu^a(s_i, t) = \mu(s(t_i + k \Delta_i/n), t_i + k \Delta_i/n), \sigma^a(s_i, t) = \sigma(s(t_i + k \Delta_i/n), t_i + k \Delta_i/n), \) for \(t_i + k \Delta_i/n \leq t < t_i + (k + 1) \Delta_i/n\). Both ways lead to the following simulation model:

\[
d\tilde{s}(t) = \mu^a(\tilde{s}(t), t; \theta)dt + \sigma^a(\tilde{s}(t), t; \theta)dB(t) + \ln Y d\xi(t) (16)
\]

or

\[
\tilde{s}(t_i + (k + 1) \Delta_i/n) = \tilde{s}(t_i + k \Delta_i/n) + \mu(\tilde{s}(t_i + k \Delta_i/n), t_i + k \Delta_i/n; \theta)\Delta_i/n + \sigma(\tilde{s}(t_i + k \Delta_i/n), t_i + k \Delta_i/n; \theta)\Delta_i/n^{1/2} \epsilon_{ik}^0 + \sum_{j=1}^{N_{\Delta_i/n}} (\mu_0 + v \epsilon_{ik}^j) (17)
\]

where \(\Delta_i = t_{i+1} - t_i, \epsilon_{ik}^j \sim i.i.d. N(0, 1), j \geq 0, k = 0, 1, ..., n - 1, i = 0, 1, ..., M - 1, \) and \(N_{\Delta_i/n} \sim \text{Poisson distribution with intensity } \lambda \Delta_i/n\). Discretization of the process described in the above is called the Euler scheme. As \(n \to \infty\), it can be shown that the simulated path \(\tilde{s}(t)\) converges to the paths of \(s_t\) uniformly in probability on compact sets, i.e., \(\sup_{0 \leq t \leq T} |\tilde{s}(t) - s_t|\) converges to zero in probability and the expected error can be expanded in powers of \(1/n\) under certain regularity conditions (see Appendix A.4). Alternatively, the following Milstein scheme can be used for the continuous part of the process as it has better convergence rate than the Euler scheme for the convergence in \(L^p(\Omega)\) and the almost sure convergence (see Talay, 1996):

\[
\tilde{s}(t_i + (k + 1) \Delta_i/n) = \tilde{s}(t_i + k \Delta_i/n) + \mu(\tilde{s}(t_i + k \Delta_i/n), t_i + k \Delta_i/n; \theta)\Delta_i/n + \sigma(\tilde{s}(t_i + k \Delta_i/n), t_i + k \Delta_i/n; \theta)\Delta_i/n^{1/2} \epsilon_{ik}^0 + \frac{1}{2} \sigma^2(\tilde{s}(t_i + k \Delta_i/n), t_i + k \Delta_i/n; \theta)\Delta_i/n + \sum_{j=1}^{N_{\Delta_i/n}} (\mu_0 + v \epsilon_{ik}^j) (18)
\]

Two points regarding the dynamic path simulation are noted here. First, to reduce the computational burden and at the same time to achieve a high level of accuracy, a variety of variance reduction techniques have been proposed for random number simulation, such as control variates, antithetic variates, stratified sampling, importance sampling, and Quasi-random number generator (see Boyle, Broadie, and Glasserman,
Second, the optimization procedure of indirect parameter estimation involves re-simulating and re-sampling the dynamic path based on updated parameter values. It is required that within the optimization procedure the random numbers or the seeds used to generate the underlying random variables are kept fixed. For Poisson random number generation, it is therefore more convenient to first generate a set of i.i.d. uniform random numbers with given seeds, then convert them to i.i.d. Poisson random numbers through inverse cumulative distribution function (CDF) with updated intensity parameter values.

4. Applications of Indirect Estimation to Exchange Rate Models

4.1 Modeling Exchange Rate Dynamics Using Jump Diffusion Processes

Since the introduction of jump-diffusion process by Press (1967) to model the dynamics of stock returns, and by Merton (1976) to derive option prices based on more general processes of underlying asset returns, jump-diffusion processes have also been widely used to model the dynamics of currency exchange rates. It is believed that shifts in interest rate differentials between two countries, or changes in monetary and fiscal policies usually result in a revaluation of foreign currency. As Jorion (1988) points out, the foreign exchange market is characterized by active exchange rate management policies which do not exist in other securities’ markets, e.g. the stock market. Therefore, stochastic processes which incorporate jumps might reflect the rate of change of foreign currency prices better than the pure continuous Wiener process. Empirical evidence based on simple jump-diffusion models (see e.g. Jorion, 1988, Akgiray and Booth, 1988, and Johnson and Schneeweis, 1994) suggest that jumps are important components of foreign exchange rate processes and hence both the theoretical and empirical studies of exchange rates under uncertainty should explicitly allow for the presence of discontinuity.

Unfortunately, up to current writing, only Merton’s (1976) model has been estimated using time series data of the underlying asset returns and empirically implemented and tested for pricing options. This is primarily due to the difficulty involved in estimating general jump-diffusion processes. This paper extends Merton’s (1976) model through more general specifications of the coefficient functions in order to incorporate other possible important features of the currency exchange rate time series together with jumps.

Model I: Black-Scholes Non-Jump Model

\[ ds_t = \mu dt + \sigma dW_t \] (19)
Model II: Merton's Jump Model
\[ ds_t = (\mu - \lambda_0)dt + \sigma dW_t + \ln Y_t dq_t(\lambda) \] (20)

Model III: Conditional Heteroscedasticity and Jump
\[ ds_t = (\mu - \lambda_0)dt + (\sigma + \sigma_1 s_t) dW_t + \ln Y_t dq_t(\lambda) \] (21)

Model IV: Mean-Reversion, Conditional Heteroscedasticity and Jump
\[ ds_t = (\mu - \beta s_t - \mu_0 \lambda) dt + (\sigma + \sigma_1 s_t) dW_t + \ln Y_t dq_t(\lambda) \] (22)

where \( \ln Y_t \sim \text{iid } N(\mu_0, \nu^2) \).

Model I, as a benchmark model which assumes the logarithmic exchange rates follow a Brownian motion with drift process. Model II is Merton’s (1976) model which explicitly allows for the presence of asymmetric (i.e. if \( \mu_0 \neq 0 \)) iid lognormal jumps to exchange rates. Model III extends Merton’s model to allow for conditional heteroscedasticity (i.e. if \( \sigma_1 \neq 0 \)) in addition to jumps. Model IV specifies a linear drift term allowing the model to capture the mean-reversion feature of the underlying process (i.e. if \( \beta > 0 \)) in addition to conditional heteroscedasticity and asymmetric jumps. Models III & IV are important in that they allow for the conditional heteroscedasticity and jumps at the same time. A number of authors have addressed the issue of testing for jumps within models that also allow for heteroscedasticity (see, e.g. Jorion, 1988, and Vlaar and Palm, 1993). These models are all constructed at arbitrarily chosen sampling frequency (determined by the available data). Instead, the jump-diffusion processes considered in this paper (Model III & IV) define the conditional heteroscedasticity through non-jump instantaneous volatility, i.e.,

\[ \lim_{h \to 0} E \left[ \frac{(s_{t+h} - s_t)^2}{h} \right]_{s_t, \lambda = 0} = \sigma_t^2 \] (23)

where in our case \( \sigma_t \) is assumed to be of the simplest form, a reflected line, i.e. the non-jump conditional volatility can be varying with the level of the stochastic process. When \( \sigma_t \neq 0 \), the conditional volatility, as a convex function of \( s_t \), either decreases in a decreasing rate or increases in a increasing rate.

4.2 The Data

The time series data are weekly observations on the exchange rates of UK pound, German mark, Japanese yen, and French franc against US dollar. The data are sampled on Wednesdays. If Wednesday is a holiday, Thursday data is used. The data
Table 4.1: Summary Statistics of Currency Exchange Rates

(a) Statistics of Logarithmic Exchange Rates $s_t$:

<table>
<thead>
<tr>
<th></th>
<th>Pound</th>
<th>Mark</th>
<th>Yen</th>
<th>Franc</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.527</td>
<td>0.685</td>
<td>5.159</td>
<td>1.749</td>
</tr>
<tr>
<td>Std.Dev.</td>
<td>0.152</td>
<td>0.211</td>
<td>0.347</td>
<td>0.213</td>
</tr>
</tbody>
</table>

(b) Statistics of Differenced Logarithmic Exchange Rates $\Delta s_t$:

<table>
<thead>
<tr>
<th></th>
<th>Pound</th>
<th>Mark</th>
<th>Yen</th>
<th>Franc</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean(10E$^{-4}$)</td>
<td>$-2.755$</td>
<td>$-5.473$</td>
<td>$-10.18$</td>
<td>1.150</td>
</tr>
<tr>
<td>Std. Dev.(10E$^{-2}$)</td>
<td>1.521</td>
<td>1.515</td>
<td>1.444</td>
<td>1.473</td>
</tr>
<tr>
<td>Skewness</td>
<td>$-0.263$</td>
<td>$-0.092$</td>
<td>$-0.344$</td>
<td>0.034</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>2.167</td>
<td>0.857</td>
<td>0.739</td>
<td>1.460</td>
</tr>
<tr>
<td>Min (10E$^{-2}$)</td>
<td>$-8.669$</td>
<td>$-8.113$</td>
<td>$-6.546$</td>
<td>$-7.741$</td>
</tr>
<tr>
<td>Max (10E$^{-2}$)</td>
<td>7.397</td>
<td>7.274</td>
<td>6.582</td>
<td>6.858</td>
</tr>
</tbody>
</table>

Note: Pound is UK pound, Mark is German mark, Yen is Japanese yen, Franc is French franc. The kurtosis is measured as excess kurtosis.

covers the period from January 7, 1976 to January 17, 1996 with 1,046 observations. Let $s_t$ be the exchange rate and define $s_t = \ln s_t$. Table 1 reports summary statistics of both $s_t$ and $\Delta s_t = s_t - s_{t-1}$. The distribution of $\Delta s_t$ is generally asymmetric, skewed either to the right or the left, with positive excess kurtosis. The minimum and maximum observations of $\Delta s_t$ are all further away from the mean points considering the magnitudes of standard derivations, suggesting jumps are possible and are needed to model the dynamics of $s_t$. Observations of the plots of $\Delta s_t$ as time series also suggest time-varying volatility or heteroscedasticity of the exchange rate processes.

The weekly sampling frequency is chosen here as a compromise between using as many available data as possible and avoiding additional modeling issues. First, the weekly sampling frequency would much reduce spurious market microstructure distortions in the data which are quite difficult to deal with satisfactorily in our continuous-time model framework; Second, using the weekly data instead of daily (or even higher frequency data) can avoid the treatment of weekends and other seasonal day-of-the-week effects; Third, since our model aims to identify the mean jump frequency and mean jump size over the sampling time interval, choice of too high sampling frequency would make such identification much more difficult and the estimation results much more subject to noises in the data.
4.3 The Estimation Results and Implications

Models I & II are estimated directly via ML method, while models III & IV are estimated using indirect inference method described in section 3. In performing the estimation, the minimization problem of the optimal indirect inference estimator is solved numerically with given analytically gradient functions using the procedure “Optimum” of GAUSS 3.2. Simulation of the dynamic sampling path of the jump-diffusion model is based on the Milstein scheme (with number of drawings $H = 20$), and the antithetic variate approach is used to reduce the variability of estimation results. Parameter estimation of the instrumental or auxiliary models is performed using ML estimation. Tables 2(a), 2(b), 2(c) and 2(d) report the estimation results of for the exchange rates of UK pound, German mark, Japanese yen, and French franc against US dollar. To reduce the table width, the estimates for parameters $\mu$ and $\sigma$ are omitted as they are relatively less important and of less interest to us. Tests of various hypotheses are also performed for each model. Hypothesis for the presence of jumps in Model II is tested based on likelihood ratio test statistic. Hypotheses for the presence of jumps, conditional heteroscedasticity, or mean-reversion in Model III & IV are tested based on the test statistic derived in Gouriéroux, Monfort and Renault (1993) which measures the difference between the optimal values of the objective function. This test statistic is asymptotically distributed as $\chi^2(p)$, where $p$ is the difference between the dimensions of the unrestricted parameter space and restricted parameter space, i.e $p = \text{dim}(\theta^U) - \text{dim}(\theta^K)$. All one-sided tests are based on standard normal test statistics using the property of asymptotic normality of the indirect inference estimators. The estimation results and hypothesis tests indicate that all exchange rate processes exhibit significant jumps. For model II, mean jump size is significant only for the exchange rate process of Japanese yen/US dollar, while for models III & IV, mean jump size is significant for UK pound/US dollar and Japanese yen/US dollar at 99% confidence level and also for French franc/US dollar at 95% confidence level. The sign of mean jump size for each process is consistent with the direction of skewness reported in Table 1. Significant conditional heteroscedasticity is present in all exchange rates, except for German mark/US dollar in model IV. Mean-reversion is not significantly present in any of the exchange rate processes.

The major conclusions from the estimation results and hypothesis tests are: First, jumps are significant components of all exchange rate processes even when conditional heteroscedasticity and other features are considered. Second, jump frequencies are significantly lower for models which consider conditional heteroscedasticity, especially for those with strongly significant presence of conditional heteroscedasticity, namely the exchange rates of UK pound/US dollar, Japanese yen/US dollar, and French franc/US dollar. However, the jump size tends to be higher for models with
Table 4.2: Parameter Estimates for Alternative Models

(a) UK pound/US dollar

<table>
<thead>
<tr>
<th></th>
<th>Diffusion</th>
<th>Jump</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\beta$ ($10^{-2}$)</td>
<td>$\sigma_1$ ($10^{-3}$)</td>
</tr>
<tr>
<td>II</td>
<td>2.70</td>
<td>(0.38)</td>
</tr>
<tr>
<td>III</td>
<td>$-13.2$</td>
<td>(2.70)</td>
</tr>
<tr>
<td>IV</td>
<td>4.02</td>
<td>(11.9)</td>
</tr>
</tbody>
</table>

(b) German mark/US dollar

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>II</td>
<td>2.94</td>
<td>(0.34)</td>
<td>$-0.45$</td>
<td>8.59</td>
</tr>
<tr>
<td>III</td>
<td>$-1.93$</td>
<td>(1.01)</td>
<td>2.41</td>
<td>$-0.73$</td>
</tr>
<tr>
<td>IV</td>
<td>3.71</td>
<td>(8.74)</td>
<td>1.64</td>
<td>$-0.32$</td>
</tr>
</tbody>
</table>

(c) Japanese yen/US dollar

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>II</td>
<td>1.47</td>
<td>(0.27)</td>
<td>$-16.8$</td>
<td>11.0</td>
</tr>
<tr>
<td>III</td>
<td>$-0.50$</td>
<td>(1.40)</td>
<td>0.96</td>
<td>$-16.7$</td>
</tr>
<tr>
<td>IV</td>
<td>6.41</td>
<td>(5.44)</td>
<td>$-0.51$</td>
<td>0.98</td>
</tr>
</tbody>
</table>

(d) French franc/US dollar

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>II</td>
<td>2.81</td>
<td>(0.38)</td>
<td>4.39</td>
<td>8.58</td>
</tr>
<tr>
<td>III</td>
<td>$-21.3$</td>
<td>(2.47)</td>
<td>1.73</td>
<td>4.57</td>
</tr>
<tr>
<td>IV</td>
<td>8.72</td>
<td>(10.9)</td>
<td>$-21.2$</td>
<td>1.74</td>
</tr>
</tbody>
</table>

Note: The numbers in the brackets are standard derivations of the above estimates.
conditional heteroscedasticity. This confirms the fact that conditional heteroscedasticity can help to remove the spurious volatility that leads to mis-identification of jump size and frequency. Third, even though mean-reversion is believed to be an important feature for many financial time series, e.g. short-term interest rates, our estimation and hypothesis test results suggest that it is not an important feature for exchange rate processes. A simple explanation of this finding is that, considering the time span of the sampling period (over 20 years in our sampling observations), exchange rate processes are essentially not stationary and therefore exhibit no unique stable long-run mean or equilibrium level.

5. CONCLUSION

This paper proposes the indirect inference approach to estimating the jump-diffusion processes with general parametric specifications of the coefficient functions. Applications of the indirect estimation method are undertaken to various currency exchange rate models. The estimation results suggest strong evidence for the misspecification of both the non-jump Black-Scholes model and Merton’s constant conditional volatility jump model as representation of currency exchange rate processes. In particular, models assuming conditional homoscedasticity tends to overestimate the jump frequency but underestimate the jump size. It would be interesting to investigate the exact impact of various model specifications on exchange rate option prices and we leave it for future research.
Appendix

A.1. Regularity Conditions on Jump-Diffusion Processes

To ensure that the generalized mixing SDE defined in (2) is well behaved and can be applied with Itô’s stochastic calculus, we impose the following conditions:

C1. Let \( \{F_t : t \in [0, T]\} \) denote a sequence of right-continuous filtrations of sub-\(\sigma\)-algebras of the \(\sigma\)-field \(F\) such that (i) \(F_t \subset F_s\) for \(t \leq s\); (ii) \(F_t = \cap_{s > t} F_s\), the Brownian motion or Wiener process \(B(t)\) be \(F_t\) measurable for all \(t \in [0, T]\);

C2. For all \(\theta \in \Theta\), the functions \(\mu(\cdot), \sigma(\cdot), \kappa(\cdot),\) and \(\gamma(\cdot)\) are measurable in the product \(\sigma\)-algebra \(B \times F\), where \(B\) is the \(\sigma\)-field of the Borel sets on \(\Omega\), and

\[
\int_0^t |\mu|^2 d\tau < \infty, \int_0^t |\sigma|^2 d\tau < \infty, \int_0^t |\kappa|^2 d\tau < \infty, \int_0^t |\gamma|^2 d\tau < \infty;
\]

C3. \(q_\theta(t)\) represents an independent Poisson step process with the probability \(\lambda \Delta + o(\Delta)\) that \(s(t)\) will experience a jump;

C4. Let \(P(\gamma(dY, t))\) be the probability measure on the jump amplitude, then \(P(\gamma(dY, t))\) has a compact support.

To ensure the existence and uniqueness of a solution to the SDE in (2), we further impose:

C5. The coefficient functions satisfy both Lipschitz condition and linear growth condition, i.e., there exists a positive constant \(K\) for which

\[
|\mu(s, t) - \mu(s', t)| + |\sigma(s, t) - \sigma(s', t)| + |\kappa(s, t) - \kappa(s', t)| \leq K |s - s'|
\]

\[
|\mu(s, t) - \mu(s, t')| + |\sigma(s, t) - \sigma(s, t')| + |\kappa(s, t) - \kappa(s, t')| \leq K |t - t'|
\]

and

\[
\mu^2(s, t) + \sigma^2(s, t) + \kappa^2(s, t) \leq K(1 + s^2), a.s.
\]

With above conditions, the differential generator of the process (2) is defined by the operator

\[
\mathcal{L} = \frac{\partial}{\partial t} + \mu(s, t) \frac{\partial}{\partial s} + \frac{1}{2} \sigma^2(s, t) \frac{\partial^2}{\partial s^2} + \lambda \int [\Delta_\kappa(s,t)\gamma(Y,t)] P(dY)
\]

where, for any function \(g(s, t), \Delta_\kappa(s,t)\gamma(Y,t)g(s, t) = g(s + \kappa(s, t)\gamma(Y, t), t) - g(s, t)\). Suppose that the function \(g(s, t)\) is bounded and has bounded and continuous first and second partial derivatives with respect to \(s\) and bounded and continuous derivative with respect to \(t\), it can be verified that for each \(t\) and \(s\), \(\lim_{\Delta \to 0} \frac{\mathcal{L}g(s, t)}{\Delta} = 0\).

Finally, for the application of generalized Itô’s lemma and for the purpose of indirect estimation, we further impose following conditions:
C6. The functions $\mu(t), \sigma(t), \kappa(t)$ and $\gamma(t)$ are twice continuously differentiable w.r.t. both $s(t)$ and $t$ or $Y$ and $t$, and three times continuously differentiable w.r.t. $\theta$, $\kappa^{-1}(\cdot, t)$ as an inverse function of $\kappa(s(t), t)$ w.r.t. $s(t)$ exists and is bijective. The true but unknown parameters $(\theta, \lambda)$ lie in the interior of a finite dimensional closed and compact parameter space $\Theta \times \Lambda$.


Let $K_i^{(p)}$ denote the $i^{th}$ cumulant associated with the density in (5), with $\mu = 0$, for any integer $i$. Press (1967) derives (with some corrections later on by Beckers (1981)):

\[
\begin{align*}
K_1^{(p)} &= \lambda \mu_0, \\
K_2^{(p)} &= \sigma^2 + \lambda (\mu_0^2 + v^2), \\
K_3^{(p)} &= \lambda \mu_0 (\mu_0^2 + 3v^2), \\
K_4^{(p)} &= \lambda (\mu_0^4 + 6\mu_0^2v^2 + 3v^4).
\end{align*}
\]

Set the sample cumulants ($\hat{K}_i^{(p)}$) equal to the population cumulants, and solves the resultant equations for the parameter estimators $\hat{\lambda}, \hat{\sigma}^2, \hat{v}^2$, and $\hat{\mu}_0$:

\[
\begin{align*}
\hat{\mu}_0^4 - \frac{2\hat{K}_3^{(p)}}{K_1^{(p)}} \hat{\mu}_0^2 + \frac{3\hat{K}_4^{(p)}}{2K_1^{(p)}} \hat{\mu}_0 - \frac{\hat{K}_2^{(p)}}{2K_1^{(p)}} &= 0, \\
\hat{\lambda} &= \frac{\hat{K}_1^{(p)}}{\hat{\mu}_0}, \\
\hat{v}^2 &= \frac{\hat{K}_3^{(p)} - \mu_0^2 \hat{K}_1^{(p)}}{3\hat{K}_1^{(p)}}, \\
\hat{\sigma}^2 &= \frac{\hat{K}_2^{(p)}}{K_1^{(p)}} - \frac{\hat{K}_3^{(p)}}{\hat{\mu}_0^2} (\hat{\mu}_0^2 + \frac{\hat{K}_3^{(p)} - \mu_0^2 \hat{K}_1^{(p)}}{3\hat{K}_1^{(p)}}).
\end{align*}
\]

The polynomial in $\hat{\mu}_0$ has four roots, two complex and two real. Of these, the real root which yields a positive $\hat{\lambda}$ is chosen.

Let $K_i^{(b)}$ denote the $i^{th}$ cumulant associated with the density in (5), with $\mu_0 = 0$, for any integer $i$, Beckers (1981) shows that

\[
\begin{align*}
K_1^{(b)} &= \mu, \\
K_2^{(b)} &= \sigma^2 + \lambda v^2, \\
K_3^{(b)} &= K_5^{(b)} = 0, \\
K_4^{(b)} &= 3v^4\lambda, \\
K_6^{(b)} &= 15v^6\lambda.
\end{align*}
\]

Solving this system of equations gives

\[
\begin{align*}
\hat{\mu} &= \hat{K}_1^{(b)}, \\
\hat{\lambda} &= \frac{25\hat{K}_4^{(b)}}{3\hat{K}_6^{(b)}}, \\
\hat{v}^2 &= \frac{\hat{K}_5^{(b)}}{5\hat{K}_4^{(b)}}, \\
\hat{\sigma}^2 &= \hat{K}_2^{(b)} = \frac{5\hat{K}_4^{(b)} \hat{K}_4^{(b)}}{3\hat{K}_6^{(b)}}.
\end{align*}
\]

A.3. Optimal Choice of Weighting Matrices $\Omega$ and $\Sigma$ and Computation of Variance-Covariance Matrix

Let $Q_M((s(t_i), t_i)_{i=1:M}; \theta, \lambda)$ be the criterion function (e.g., the likelihood function, the sum of squares of mean absolute differences) based on the instrumental or
auxiliary model, define

\[ I_0 = \lim_{M \to \infty} V_0 \sqrt{M} \frac{\partial Q_M}{\partial \theta_{1,1} \partial \lambda_{1,1}}(\{ \hat{s}^h(t_i; \theta^0, \lambda^0), t_i \}_{i=1:M; \theta^0, \lambda^0}) \]

\[ J_0 = \lim_{M \to \infty} - \frac{\partial^2 Q_M}{\partial \theta_{1,1} \partial \lambda_{1,1}}(\{ \hat{s}^h(t_i; \theta^0, \lambda^0), t_i \}_{i=1:M; \theta^0, \lambda^0}) \]

where \( V_0 \) indicates variance w.r.t. the true distribution of the process \( s(t) \). Since the model does not contain pre-determined exogenous variables, the optimal choice of \( \Omega \) and \( \Sigma \) are (see Gouriéroux, Monfort and Renault, 1993):

\[ \Omega^* = J_0 I_0^{-1} J_0 \]
\[ \Sigma^* = I_0^{-1} \]

Since \( (\hat{\theta}_1, \hat{\lambda}_1) \overset{p}{\longrightarrow} (\theta^0_1, \lambda^0_1) \), \( J_0 \) can be consistently estimated by

\[ - \frac{\partial^2 Q_M}{\partial \theta_{1,1} \partial \lambda_{1,1}}(\{ s(t_i), t_i \}_{i=1:M; \hat{\theta}_1, \hat{\lambda}_1}) \]

where \( \{ s(t_i) \}_{i=1:M} \) are actual observations of \( s(t) \) at \( \{ t_i \}_{i=1:M} \).

In the case of PML, the criterion function \( Q_M(\{ s(t_i), t_i \}_{i=1:M; \theta_1, \lambda_1}) \) is defined as

\[ Q_M = \frac{1}{M} \sum_{i=0}^{M-1} \log f_l(s(t_{i+1}), t_{i+1} | s(t_i), t_i; \theta_1, \lambda_1) \]
\[ = \frac{1}{M} \sum_{i=1}^{M-1} \psi_l(\theta_1, \lambda_1) \]

Since there are no exogenous variables,

\[ I_0 = \lim_{M \to \infty} V_0 \left[ \frac{1}{\sqrt{M}} \sum_{i=0}^{M-1} \frac{\partial \log f_l(s(t_{i+1}), t_{i+1} | s(t_i), t_i; \theta_1, \lambda_1)}{\partial \theta_{1,1} \partial \lambda_{1,1}} \right] \]

and \( I_0 \) can be approximated by (see Newey and West, 1987)

\[ \hat{\Gamma}_0 + \sum_{k=1}^{K} (1 - \frac{k}{K + 1})(\hat{\Gamma}_k + \hat{\Gamma}_k') \]

with

\[ \hat{\Gamma}_k = \frac{1}{M} \sum_{i=k+1}^{M} \frac{\partial \psi_{i-k}}{\partial \theta_{1,1}}(\hat{\theta}_1, \hat{\lambda}_1) \frac{\partial \psi_i}{\partial \lambda_{1,1}}(\hat{\theta}_1, \hat{\lambda}_1) \]

where, for the choice of the bound \( K \), see Newey and West (1987). The expression of the asymptotic variance-covariance matrix contains the derivative of the binding
function at the true value. It is possible to estimate this quantity consistently without determining the binding function and its derivative. An equivalent expression of the asymptotic variance-covariance matrix of the optimal indirect inference estimator which may be directly computed from the criterion function is (see Gouriéroux and Monfort, 1996, p.71)

\[
V(H, \Omega^*) = (1 + \frac{1}{H})(\frac{\partial^2 Q_\infty}{\partial(\theta, \lambda)\partial(\theta_1, \lambda_1)'})^{-1} \frac{\partial^2 Q_\infty}{\partial(\theta_1, \lambda_1)\partial(\theta, \lambda)'}
\]

A consistent estimator of this matrix can be obtained by replacing \( Q_\infty \) by \( Q_M \), \( b(\theta^0, \lambda^0) \) by \( (\hat{\theta}_1, \hat{\lambda}_1) \) and \( I_0 \) by the approximation derived above.

### A.4. Uniform Convergence of Simulated Path \( \tilde{s}(t) \) to \( s(t) \) in Probability and Expansion of Expected Error

Under regularity conditions in A.1, we have \( \lim_{n \to \infty} \mu^n(s(t), t) = \mu(s(t), t) \), \( \lim_{n \to \infty} \sigma^n(s(t), t) = \sigma(s(t), t) \), and \( \lim_{n \to \infty} \kappa^n(s(t), t) = \kappa(s(t), t) \). With \( \tilde{s}(t_0) = s(t_0) \), using the results in Protter (1990) (p.209, Theorem 15), we have \( \tilde{s}(t) \) based on the Euler scheme converges to \( s(t) \) uniformly in probability on compact sets, i.e. \( \sup_{0 \leq t \leq T} |\tilde{s}(t) - s(t)| \) converges to zero in probability. Further, apply Theorem 2.2 for general Lévy driven process in Protter and Talay (1995) to model (2), we can show that the expected error can be expanded in powers of \( 1/n \) under certain conditions. That is, there exists a function \( C(\cdot) \) and an increasing function \( K(\cdot) \) such that for any discretization step of type \( T/n \),

\[
E[s(t) - \tilde{s}(t)] = \frac{C(T)}{n} + R^n_T
\]

and \( \sup_n n^2 |R^n_T| \leq exp(K(T)\rho_{16}(\infty)) \), where \( \rho_{16}(\cdot) \) is defined in Protter and Talay (1995).
References:


Asse, K.K. (1988) ‘Contingent claims valuation when the security price is a combination of an Itô process and a random point process,’ Stochastic processes and their applications 28, 185-220.


