UNIVERSITY OF SOUTHAMPTON

ECONOMETRICS OF JUMP-DIFFUSION PROCESSES: APPROXIMATION, ESTIMATION AND FORECASTING

Sanghoon Lee

Doctor of Philosophy

Department of Economics

June 2001

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF SOCIAL SCIENCE

ECONOMICS

Doctor of Philosophy

ECONOMETRICS OF JUMP-DIFFUSION PROCESSES: APPROXIMATION, ESTIMATION AND FORECASTING

By SANGHOON LEE

In this thesis we consider the relationship between jump-diffusion processes and ARCH models with jump components. In the theoretical financial economics literature, jump-diffusion processes in continuous time have been used to model financial markets. Most empirical works use either directly discretised jump-diffusion processes or ARCH models with jump components to estimate the underlying processes. There is, however, no guarantee that those models used in empirical works are discrete counterparts of the continuous time jump-diffusion process.

In Chapter 2, Survey on Jump-Diffusion Processes in Financial Economics, we survey the existing literature of jump-diffusion processes. During the 80's and 90's, it started to draw more attention as an alternative tool to the ARCH type models. The most significant theoretical developments and empirical findings are reviewed.

In Chapter 3, Approximation of Jump-Diffusion Processes, we show that a discrete time stochastic difference equation (e.g. ARCH with jumps) converged weakly to the continuous time stochastic differential equation (e.g. jump-diffusion limit) as the length of sampling interval goes to zero. It is shown that, as examples, GARCH(1,1)-M with jumps and EGARCH with jumps converge to their jump-diffusion limits.

In Chapter 4, Filtering with Jump-Diffusion Processes, we study the properties of the conditional covariance estimates generated by a misspecified model with jumps. We show that a misspecified model can consistently estimate the conditional covariance of the true data generating process. I.e., the difference between a conditional covariance estimate and the true conditional covariance converges to zero in probability as the sampling interval of length h goes to zero.

In Chapter 5, Forecasting with Jump-Diffusion Processes, we investigate the forecasting ability of jump-diffusion processes. It is shown that forecast generated by a misspecified model with jumps converges weakly to forecast generated by the true data generating process. That is, the difference between the forecasts generated by a misspecified model and those generated by the true underlying process becomes zero as the length of sampling interval approaches to zero.

Finally, in Chapter 6, we conclude the thesis by summing up the important features of the results. Some further directions for future research are proposed.

Contents

Acknowledgments

Thanks to my wife, Ree, who has encouraged and supported me all the way through, my daughter, Chaemin, who has given me all the happiness and my parents for their spiritual and financial support for many years.

Thanks to Jan Podivinsky for his invaluable supervision and help.

Thanks to John Aldrich, Maozu Lu, In-Ho Lee, Giovanni Forchini, Jong-Hee Hahn, Pat Rice, Raymond O'Brien, Sally Srinivasan, Chong-Cheul Cheong, Sung-Kyu Lee, and many other people for their encouragements, help, comments, ideas, and friendship.

A faculty scholarship from the Faculty of Social Sciences and partial financial support from the Department of Economics and the University of Southampton are gratefully acknowledged.

Chapter 1

Introduction

In the financial markets we can easily observe shocks causing market volatilities. When important information arrives at the markets, the underlying processes, for example, stock prices or foreign exchange rates, are disturbed and jump discretely to some other levels. Therefore, it has been an important issue in the financial economics literature to find adequate tools to model such behaviour of financial markets. Since Press (1967) , there has been growing interest in jump-diffusion processes during the last three decades or so. Merton (1976a, b) divides the total changes in stock prices into two parts. The first part is normal changes in stock prices caused by normal economic activities, such as temporal imbalance between supply and demand, changes in capitalisation rates, changes in economic outlook, or other new information causing marginal changes in the values of stock. The second part is abnormal changes in stock prices caused by unanticipated shocks to markets, such as some important new information causing more than a marginal changes in the values of stock. He suggests to model the normal vibration by a standard Brownian motion and the abnormal vibration by a Poisson jump process.

Since then, as the solution to a stochastic asset optimisation problem, jump-diffusion processes are popularly used in the dynamic asset pricing literature

[Amin (1993) , Ball and Torous (1983, 1985) , Chang (1995) , Jarrow and Rosenfeld (1984) , Kim, Oh and Brooks (1994) , and Oldfield, Rogalski and Jarrow (1977)], in foreign exchange rates [Ball and Roma (1993) , Jorion (1988) , Park, Ahn and Fujihara (1993) , and Vlaar and Palm (1993)], and in the term structure of interest rates [Ahn and Thompson (1988) and Das (1997)].

In the theoretical financial economics literature, jump-diffusion processes in continuous time have been used to model financial markets. The jump-diffusion processes in continuous time can be represented by a continuous time stochastic differential (integral) equation which is the linear combination of a diffusion process and a Poisson jump process. Most empirical works use either directly discretised jump-diffusion processes or $ARCH$ type models with jump components to estimate the underlying processes. However, there is no guarantee that the discretised jump-diffusion processes or $ARCH$ (AutoRegressive Conditional Heteroskedasticity) type models with jump components are the discrete counterparts of the continuous time jump-diffusion processes. So, in this thesis we would like to develop the relationship between the continuous time stochastic differential equations and the discrete time stochastic difference equations.

First of all, it is necessary to review the existing literature of the jump-diffusion processes. Although the literature has initiated earlier than the $ARCH$ type models[1], the literature is relatively small compared to that of $ARCH$ type models. Starting from Press (1967) , Cox and Ross (1975, 1976) , and Merton (1976*a*, *b*) established the cornerstone of the literature. Then, during the 1980's and 90's, it started to draw more attention as an alternative modeling tool to the $ARCH$ type models. In Chapter 2, the most significant theoretical developments and empirical findings are reviewed.

There, we try to answer the following three questions: 1) Can we use the

[1]Since Engle(1982) , the $ARCH$ type models have been extensively used in the financial economics literature and the literature has grown massively during the last two decades or so. There are a couple of excellent surveys about $ARCH$ type models. Interested readers are refered to Bollerslev, Chou and Kroner(1992) and Bollerslev, Engle and Nelson(1994) .

discrete time stochastic difference equation as a discrete time counterpart of the continuous time stochastic differential equation? 2) Can the misspecified models correctly identify the conditional covariance structure of the true data generating process? And 3) how close can a misspecified model produce forecasts to the forecasts generated by the true data generating process?

In Chapter 3, to answer the first question, we show that a discrete time stochastic difference equation converges weakly to the continuous time stochastic differential equation as the length of sampling intervals goes to zero. Then, as examples, we show that $GARCH\,(1,1)$-M with jumps and $EGARCH$ with jumps are jump-diffusion approximations.

Next, as economic models are rough approximations of the real economy, it is inevitable that those models are misspecified. We, however, require that these misspecified models should enable reasonable understanding of the real economy. So, in Chapter 4, we answer the second question by showing that the misspecified models can consistently estimate the conditional covariance structure of the true data generating process.

Then, as most market participants would like to predict market behaviour in the future so that they can minimise the risk existing in the future, it is necessary to raise the third question. In Chapter 5, we show that the misspecified models can produce forecasts close to those generated by the true data generating process. That is, we derive the forecast functions for the misspecified and true data generating processes and show that the misspecified models can produce at least consistent forecasts of the true data generating process.

Finally, we conclude the thesis by summing up the important features of the results. Some future directions for further research are proposed.

Chapter 2

Survey on Jump-Diffusion Processes in Financial Econometrics

2.1 Introduction

In financial economics, it has been an important issue to find a distribution accurately describing the behaviour of financial time series. Many earlier researchers' attempts [such as the Stable Paretian by Mandelbrot (1963), Poisson Mixture of lognormal distribution by Press (1967), Scaled t distribution by Praetz (1972), and Subordinate Stochastic process by Clark (1973)] have not been accepted in general. However, the work of Black and Scholes (1973) has been accepted as one of the most significant developments in the history of financial economics and has been extensively used to model financial time series. Yet, their model cannot fully describe the behaviour of financial time series, such as fatter tails than normal, high concentration of mass near zero, and high volatilities.

Later, Merton (1976a, b) introduced jump-diffusion processes to model financial time series. His model allows jumps caused by arrival of important news at the market. He decomposed the total changes in stock prices into two parts. First, the normal vibration in prices due to a temporary imbalance between supply and demand, changes in capitalisation rates, changes in economic outlook, or other new information causing marginal changes in the values of stock. This kind of change can be modelled by a standard geometric Brownian motion. Second, the abnormal vibration in prices due to the arrival of important new information causing more than a marginal changes in the values of stock. This part of the changes can be modelled by a jump process. With a jump-diffusion process, we expect to take the empirically observed properties of financial time series into account more successfully than the Black-Scholes model.

With the introduction of jump-diffusion processes, there have been quite a number of researches on the issues of jump-diffusion processes in financial economics. It has been used in the analysis of financial markets, such as stock markets, foreign exchange markets, and term structure of interest rates. Much empirical work found that the jumps in those financial time series are significant, especially when the sampling interval gets smaller. As the $ARCH$ type models cannot explain alone the stochastic nature of financial markets, there have been several works to incorporate jump processes into $ARCH$ framework [Feng and Smith (1997) , Jorion (1989) , and Vlaar and Palm (1993)]. When $ARCH$ models are used along with jump-diffusion processes, it becomes more powerful to describe the volatilistic nature of financial time series as well.

In this survey, we are looking for the extensions and developments of the Poisson Mixture of lognormal distribution by Press (1967) . In the next section, we present the jump-diffusion process developed during the last three decades or so. Section 2.3 considers several estimation methods popularly used in the literature. In section 2.4, we survey the empirical evidence of jump-diffusion process used in the dynamic asset pricing literature, foreign exchange markets, etc. Then, we conclude the survey in section 2.5.

2.2 Jump-Diffusion Processes

In this section, we will explore the developments of the jump-diffusion processes in the financial economics literature. The development of jump-diffusion processes is based on the Poisson mixture of lognormal distribution by Press (1967). Then, Cox and Ross (1975, 1976) and Merton (1976 a, b) established the cornerstone of the literature.

2.2.1 Development of Jump-Diffusion Processes

Black and Scholes (1975) assume that changes in the price of the stock, S, are governed by a diffusion process of the return

$$\frac{dS}{S} = \mu dt + \sigma dW \tag{2.1}$$

where W is a Weiner process. This equation can be interpreted as the percentage change in the stock price over a small interval of time will be given by a deterministic drift component, μdt, and a random increment which is normally distributed with mean zero and variance $\sigma^2 dt$. As dt gets smaller, S_{t+dt} will not differ much from S_t. That is, in this diffusion process only local changes in the stock price are permitted.[1] The arrival of information at a market, however, will be in discrete lumps rather than in a smooth flow, and assets in such markets are likely to have discontinuous jumps in value. So this behavior violates the basic assumption of a diffusion process. Price movements of this sort can be captured by assuming that the asset follows a jump process rather than a diffusion process. Unlike a diffusion process, a jump process is characterized by the property that with high probability, approaching 1 as $dt \to 0$, its movement within the interval $[t, t + dt)$ will be certain, but with a low and continuing probability it will jump

[1] Within a small interval of time $[t, t + dt)$, S_t will move in a random fashion, but with high probability, approaching 1 as $dt \to 0$. So S_{t+dt} will be in an arbitrary small neighbourhood of S_t.

to a new value.

Press (1967) assumed that the logged price changes are following a Poisson mixture of normal distribution. He had shown that the distribution agrees with the characteristic of the logged price changes found in empirical work. A cumulant matching method is suggested for estimating the first four moments of the distribution. He assumed that $S(t)$, the natural logarithm of the price of a security at time t, is stationary and has independent increments whose basic mechanism is composed of a compound Poisson process augmented by a Wiener process,

$$S(t) = C + \sum_{i=1}^{N(t)} Y_i + X(t) \tag{2.2}$$

where $S(0) = C$, Y_i, $i = 1, \ldots k, \ldots$, are the size of jumps mutually independent normally distributed with mean θ and δ^2, $N(t)$ is a Poisson counting process with parameter λt, which represents the number of random jumps occurring in time t, and $\{N(t), t \geq 0\}$ is independent of Y_i, and $\{X(t), t \geq 0\}$ is a Wiener process with mean 0 and variance $\sigma^2 t$. It is assumed that the process has uncorrelated increments.

If we difference $S(t)$ in (2.2),[2] we have

$$\begin{aligned} \Delta S(t) &= S(t) - S(t-1) \\ &= \sum_{i=N(t-1)+1}^{N(t)} Y_i + \varepsilon(t) \end{aligned} \tag{2.3}$$

where $\varepsilon(t) = X(t) - X(t-1)$ is the stationary independent normal with mean 0 and variance σ^2. He showed that the first four moments of $\Delta S(t)$ agree with what is found empirically, such as, skewness, leptokurtosis, and thick tailed.

In the study of the problem of option valuation when the stock follows a jump

[2] Here, by differencing $S(t)$, he implicitly assumed that the diffusion component has zero drift.

process, Cox and Ross(1975) proposed a simple jump process that can be written as

$$\frac{dS}{S} = \mu dt + (k-1)\, d\widetilde{\Pi} \tag{2.4}$$

$$= \begin{cases} \mu dt + k - 1 \text{ with probability } \lambda dt \\ \mu dt \text{ with probability } (1 - \lambda dt) \end{cases} \tag{2.5}$$

where $\widetilde{\Pi}$ is a Poisson process and $d\widetilde{\Pi}$ takes value 0 with probability $1 - \lambda dt$ and 1 with probability λdt. The parameter λ is called the intensity of the process. Therefore, if no jump occurs $S(t)$ moves at the exponential rate μ, but if a jump occurs $S(t)$ changes by $(k-1)\,S(t)$ to $S(t) + (k-1)\,S(t) = kS(t)$. They also showed that the Black and Scholes' diffusion process is simply a limiting case of the jump process, and can be approximated by a jump process [Section 3, Cox and Ross (1975)].

Cox and Ross (1976) developed several alternative processes to (2.5) to model the jump processes. First, they suppose that changes in stock price follow a pure birth and death process with $k^{+} > 1$ and $k^{-} < 1$ with respective probability π^{+} and π^{-}, and that the intensity is proportional to value, S :

$$dS = \begin{cases} k^{+} - 1 & \text{with probability } \pi^{+}\lambda S dt, \\ k^{-} - 1 & \text{with probability } \pi^{-}\lambda S dt, \\ 0 & \text{with probability } 1 - \lambda S dt. \end{cases} \tag{2.6}$$

The limit of (2.6) as $k^{+} \to 1$, $k^{-} \to 1$, and $\lambda \to \infty$, is a diffusion with instantaneous mean μS and variance $\sigma^2 S$ where μ and σ^2 are

$$E\{dS\} = [\mu + \lambda E\{k-1\}]\, S dt$$

$$Var\{dS\} = \lambda E\left\{(k-1)^2\right\} S dt.$$

So, the stochastic differential equation can be written as

$$dS = \mu S dt + \sigma \sqrt{S} dz. \tag{2.7}$$

Another interesting process is one where the intensity, λ, is constant and the value of the increment is constant.

$$dS = \mu S dt + \begin{cases} k^+ - 1 \text{ with probability } \pi^+ \lambda S dt \\ k^- - 1 \text{ with probability } \pi^- \lambda S dt \\ 0 \qquad \text{with probability } 1 - \lambda S dt \end{cases} \tag{2.8}$$

The local mean and variance of the process are given by

$$\begin{aligned} E\{dS\} &= \left\{\mu S + \lambda \left[\pi^+ \left(k^+ - 1\right) + \pi^- \left(k^- - 1\right)\right]\right\} dt \\ Var\{dS\} &= \lambda \left\{\pi^+ \left(k^+ - 1\right)^2 + \pi^- \left(k^- - 1\right)^2\right\} dt \end{aligned}$$

The diffusion limit of (2.8) can be written as

$$dS = \mu S dt + \sigma dz \tag{2.9}$$

Merton (1976 a, b) divided the total change in a stock price into two components. The first is the normal vibrations due to a temporary imbalance between supply and demand, changes in capitalization rates, changes in the economic outlook, or other information that causes marginal changes in stock's value. The second is the abnormal vibrations in price due to the arrival of important new information about the stock that has more than a marginal effect on price. The normal vibration can be modelled by a standard geometric Brownian motion, and the abnormal vibration can be modelled by a jump process. The Poisson driven process can be defined as follows

$$\begin{aligned}\Pr\left[N(t,t+h)=0\right] &= 1-\lambda h+o(h),\\ \Pr\left[N(t,t+h)=1\right] &= \lambda h+o(h),\\ \Pr\left[N(t,t+h)\geq 2\right] &= o(h),\end{aligned}$$

where $N(t,t+h)$ is a number of Poisson distributed events per unit time and λ is the mean number of arrivals of important news per unit time. If $S(t)$ is the stock price at time t and Y is the random variable of the drawing from a distribution to determine the impact of the information on the stock price, then neglecting the continuous part, the stock price at time $t+h$, $S(t+h)$, will be the random variable $S(t+h)=S(t)Y$, given that one such arrival occurs between t and $t+h$. The $\{Y\}$ from successive drawings are assumed independent and identically distributed. Then the stochastic differential equation for the stock price return is

$$\frac{dS}{S}=\alpha dt+\sigma dW+zdq \tag{2.10}$$

where α is an instantaneous expected return on the stock, σ^2 is the instantaneous variance of the return, conditional on no arrivals of important new information, dW is a standard Wiener process, z is the percentage change in share price resulting from a jump,[3] and $q(t)$ is a jump process.

The σdW term describes the instantaneous part of unanticipated return due to the normal price change and dq describes the part due to the abnormal price changes. If $\lambda=0$, then the return dynamics would be the same as given in the Black and Scholes (1973).

[3] Then, $Z=(z+1)$ will be the jump amplitude, and $\ln Z\sim N(\mu,\delta^2)$.

A solution for the differential equation in (2.10) is

$$S(t+s) = S(t) Z(0) Z(1) \ldots Z(N) \exp\left\{\left(\alpha - \sigma^2/2\right) s + \sigma\sqrt{s}W\right\}, \qquad (2.11)$$

where s is the time between observed prices $S(t+s)$ and $S(t)$. The number of jumps during the interval s is N, and the $Z(i)$ are the jump size where $Z(0) = 1$ and $Z(i) \geq 0$ for $i = 1, \ldots, N$. Dividing (2.11) by $S(t)$ and take natural logarithms gives

$$\ln\left[\frac{S(t+s)}{S(t)}\right] = \left(\alpha - \sigma^2/2\right) s + \sigma\sqrt{s}W + \sum_{i=1}^{N} \ln Z(i). \qquad (2.12)$$

The first two terms are due to the continuous diffusion process, and the rest is due to the jump process. If $N = 0$, then $\ln[S(t+s)/S(t)]$ is normally distributed with mean $(\alpha - \sigma^2/2) s$ and variance $\sigma^2 s$. If the $\ln Z(i)$ are assumed to be identically distributed with mean μ and finite variance δ^2, a general form of the joint density for $\ln Z(i)$ can be represented by

$$\phi(\ln Z(1), \ldots, \ln Z(N)) = \int_{-\infty}^{\infty} f(\ln Z(1), \ldots, \ln Z(N), W)\, dW, \qquad (2.13)$$

with

$$\begin{aligned} &E[\ln Z(i)] = \mu && \text{for } i = 1, \ldots, N \\ &Var[\ln Z(i)] = \delta^2, && \text{for } i = 1, \ldots, N \\ &Cov[\ln Z(i), \ln Z(|i-j|)] = \rho_j \delta^2 && \text{for } j \geq 0 \end{aligned}$$

where ρ_j is the correlation between $\ln Z(i)$ and $\ln Z(|i-j|)$. The index i represents the jump number while the index j denotes the number of lags between jumps.

2.2.2 The Conditional Density

For $N \geq 1$, the conditional density of $\ln\left[S(t+s)/S(t)\right]$ can be derived with the transformation technique using (2.12) and (2.13):

$$c\left(\ln\left[\frac{S(t+s)}{S(t)}\right]\middle| N\right)$$
$$= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f\left(\left[\ln\left[\frac{S(t+s)}{S(t)} - \left(\alpha - \sigma^2/2\right)s\right]\right.\right.$$
$$\left.\left. -\sigma\sqrt{s}W - \sum_{i=2}^{N} y_i\right], y_2, \ldots, y_n, W\right) dy_2 \cdots dy_N dW, \qquad (2.14)$$

where $-\infty < \ln\left[S(t+s)/S(t)\right] < \infty$, $y_i = \ln Z(i)$, and $N \geq 1$. The moments of (2.14) can be derived by taking the expectation of (2.12) given N jumps and using $E\{\ln Z(i)\} = \mu$ and $E\{W\} = 0$.

$$E\left\{\ln\left[\frac{S(t+s)}{S(t)}\right]\middle| N\right\} = \left(\alpha - \sigma^2/2\right)s + N\mu. \qquad (2.15)$$

The first term on the right-hand side of (2.15) represents the contribution to the mean due to the diffusion process. The second term is due to the jump process. The presence of autocorrelation does not affect the conditional mean.

The variance of (2.14) can be derived from (2.12) in a similar way:

$$var\left\{\ln\left[\frac{S(t+s)}{S(t)}\right]\middle| N\right\} = \sigma^2 s + var\left\{\sum_{i=1}^{N}\ln Z(i)\middle| N\right\}. \qquad (2.16)$$

The variance of the sum of $\ln Z(i)$ given N can be shown by using mathematical induction:

$$var\left\{\sum_{i=1}^{N}\ln Z(i)\middle| N\right\} = N\delta^2 + 2\delta^2\sum_{j=1}^{N-1}(N-j)\rho_j, \quad \text{for } N \geq 1. \qquad (2.17)$$

Substituting (2.17) into (2.16) gives the conditional variance

$$var\left\{\ln\left[\frac{S(t+s)}{S(t)}\right]\bigg| N\right\} = \sigma^2 s + N\delta^2 + 2\delta^2\sum_{j=1}^{N-1}(N-j)\,\rho_j, \quad \text{for } N \geq 1. \quad (2.18)$$

There are several important features of the conditional variance in (2.18) that should be pointed out. First, the conditional variance of returns is influenced by autocorrelation between the jump sizes ρ_j.[4] Second, the portion of the variance due to the diffusion process and the portion due to the jump process enter separately. Thus, if no jump process present, the conditional variance is a linear function of the time between transactions. When no diffusion process operates, (2.18) reduces to

$$var\left\{\ln\left[\frac{S(t+s)}{S(t)}\right]\bigg| N\right\} = N\delta^2 + 2\delta^2\sum_{j=1}^{N-1}(N-j)\,\rho_j$$

which is a function of the number of jumps during s, but not the length of s. Finally, the variance in (2.18) is derived without requiring a specific form for the distribution of $\ln Z(i)$.

2.2.3 The Unconditional Density

The unconditional density corresponds to the probability function that can be used to examine returns over fixed time intervals s where the number of transactions N is variable. This is important because it provides a means of modelling daily returns where N varies from day to day but s remains fixed at one day.

The unconditional density for $\ln[S(t+s)/S(t)]$ is found to be

$$h\left(\ln\left[\frac{S(t+s)}{S(t)}\right]\right) = \sum_{N=0}^{\infty} c\left(\ln\left[\frac{S(t+s)}{S(t)}\right]\bigg| N\right) q(N), \quad (2.19)$$

[4] Merton(1976) assumes that $\rho_j = 0$ for all j and the time between jumps Exponentially distributed. So, Merton's model is not autocorrelated and N has Poisson density.

where $q(N)$ is the probability that N jumps occur during $[t, t+s]$. The unconditional mean is the expectation of conditional mean in (2.16) which gives

$$E\left\{\ln\left[\frac{S(t+s)}{S(t)}\right]\right\} = \left(\alpha - \sigma^2/2\right)s + \mu E\{N\}. \tag{2.20}$$

The unconditional variance is given by

$$\begin{aligned} & var\left\{\ln\left[\frac{S(t+s)}{S(t)}\right]\right\} \\ = & \sum_{N=0}^{\infty} var\left\{\ln\left[\frac{S(t+s)}{S(t)}\right]\middle| N\right\} q(N) + \mu^2 var\{N\}. \end{aligned} \tag{2.21}$$

Now, to specify a specific form for the mean and the variance, we need to define a particular density for N.

Assume that the time intervals between the transactions are independent and identically distributed random variables with a gamma density function:

$$g(\Delta t) = [\exp(-\lambda\Delta t)]\,\lambda\,(\lambda\Delta t)^{r-1}/\Gamma(r) \tag{2.22}$$

with

$$\begin{aligned} E\{\Delta t\} &= \frac{r}{\lambda} \\ var\{\Delta t\} &= \frac{r}{\lambda^2}, \quad \text{for } 0 \leq \Delta t < \infty,\ r > 0, \text{ and } \lambda > 0. \end{aligned}$$

The gamma random variable Δt is the continuous amount of time required to observe the first jump starting at an arbitrary point in the process. r and λ represent parameters which determine the shape of a gamma distribution.[5] For a standardised gamma (mean zero and variance one), as r increases the shape of the gamma becomes less skewed until in the limit the gamma distribution tends

[5]While λ is just a scale factor, r indicates the height of the density

to a normal density.

With a density in (2.22) the density for the number of jumps between $(t, t+s)$ can be derived. The probability that N jumps occur in $[t, t+s]$ is

$$\begin{aligned} \Pr(N) &= \Pr(\Delta t_1 + \Delta t_2 + \ldots + \Delta t_N + s) \\ &\quad - \Pr(\Delta t_1 + \Delta t_2 + \ldots + \Delta t_{N+1} \leq s), \quad \text{for } N \geq 1, \end{aligned} \tag{2.23}$$

where Δt_i is time between ith and $(i-1)$st jump. So, the probability that N jumps occur in $[t, t+s]$ is the probability that at least N jumps occur minus the probability that at least $(N+1)$ jumps occur. This requires that the time intervals are independent. Using moment generating functions, it can be shown that

$$g(a) = [\exp(-\lambda a)] \frac{\lambda (\lambda a)^{rN-1}}{\Gamma(rN)}, \tag{2.24}$$

with

$$\begin{aligned} E\{a\} &= \frac{Nr}{\lambda} \\ var\{a\} &= \frac{Nr}{\lambda^2}, \quad \text{for } a = \sum_{i=1}^{N} \Delta t_i \text{ and } 0 \leq a < \infty. \end{aligned}$$

Substituting (2.24) into (2.23) and integrating gives the density for N steps in a time period s :

$$q(N) = \frac{\gamma(rN, \lambda s)}{\Gamma(rN)} - \frac{\gamma(rN + r, \lambda s)}{\Gamma(rN + r)}, \quad N \geq 1 \tag{2.25}$$

where $\gamma(c, d)$ is the incomplete gamma function with arbitrary parameters c and d,

$$\gamma(c, d) = \int_0^d \exp[(-u)] u^{c-1} du, \quad c > 0. \tag{2.26}$$

Specific forms of the unconditional density function and moments can now be obtained by substituting the assumed density (2.25) into (2.19), (2.20), and (2.21). This gives the unconditional density $h\left(\ln\left[S(t+s)/S(t)\right]\right)$:

$$
\begin{aligned}
& h\left(\ln\left[\frac{S(t+s)}{S(t)}\right]\right) \\
= & \sum_{N=0}^{\infty} c\left(\ln\left[\frac{S(t+s)}{S(t)}\right]\middle| N\right) q(N) \\
= & \left[1-\frac{\gamma(r,\lambda s)}{\Gamma(r)}\right] c\left(\ln\left[\frac{S(t+s)}{S(t)}\right]\middle| N=0\right) \\
& +\sum_{N=1}^{\infty}\left[\frac{\gamma(rN,\lambda s)}{\Gamma(rN)}-\frac{\gamma(rN+r,\lambda s)}{\Gamma(rN+R)}\right] c\left(\ln\left[\frac{S(t+s)}{S(t)}\right]\middle| N\right)
\end{aligned} \tag{2.27}
$$

and the unconditional mean and the variance are

$$
\begin{aligned}
& E\left\{\ln\left[\frac{S(t+s)}{S(t)}\right]\right\} \\
= & \left(\alpha-\sigma^2/2\right)s+\mu E\{N\} \\
= & \left(\alpha-\sigma^2/2\right)s+\mu\sum_{N=1}^{\infty} N\left[\frac{\gamma(rN,\lambda s)}{\Gamma(rN)}-\frac{\gamma(rN+r,\lambda s)}{\Gamma(rN+r)}\right],
\end{aligned} \tag{2.28}
$$

$$
\begin{aligned}
& var\left\{\ln\left[\frac{S(t+s)}{S(t)}\right]\right\} \\
= & \sum_{N=0}^{\infty} var\left\{\ln\left[\frac{S(t+s)}{S(t)}\right]\middle| N\right\} q(N)+\mu^2 var\{N\} \\
= & \beta^2 s+\delta^2 E\{N\}+\mu^2 var\{N\} \\
& +2\delta^2\sum_{N=2}^{\infty}\sum_{i=1}^{N-1}(N-i)\rho_i\left[\frac{\gamma(rN,\lambda s)}{\Gamma(rN)}-\frac{\gamma(rN+r,\lambda s)}{\Gamma(rN+r)}\right].
\end{aligned} \tag{2.29}
$$

Both the unconditional mean and variance are functions of time s, not N.[6] As s increases, both the mean and variance increase. For example, if $r=1$, then the

[6] Here, note that N is an integer.

two unconditional moments are reduced to

$$E\left\{\ln\left[\frac{S(t+s)}{S(t)}\right]\right\} = \left(\alpha - \sigma^2/2\right)s + \mu\lambda s,$$

and

$$\begin{aligned} var\left\{\ln\left[\frac{S(t+s)}{S(t)}\right]\right\} &= \beta^2 s + \left[\delta^2 + \mu^2\right]\lambda s \\ &\quad + 2\delta^2 \sum_{N=2}^{\infty} e^{-\lambda s}\frac{(\lambda s)^N}{N!}\sum_{i=1}^{N-1}(N-i)\,\rho_i. \end{aligned}$$

This is an important property because it implies that the mean and variance of returns over longer period of time intervals than transactions are scaled by time s.

2.2.4 Other Issues in Jump-Diffusion Processes

Bardhan and Chao (1996) studied the nature of equilibrium in a multi-agent production exchange economy with jump-diffusion processes. In this study, they assume that each agent maximises his/her total utility from consumption over a finite time horizon, the market is complete, the agents take prices of securities in the market as given, and they solve their optimisation problem. They prove the existence and uniqueness of equilibrium in the presence of jumps, while several earlier works studied with only diffusion information.

2.3 Estimation Methods

2.3.1 Cumulant Matching Method

Press (1967), Beckers (1981), and Ball and Torous (1983) used the cumulant matching method to estimate the stock prices processes. The method relies upon the theoretical relationship between the population cumulants and the parameters of the distribution. Since the population cumulants are unknown, the corresponding sample cumulants are substituted into the system of equations which is then solved for the parameter estimates. Although the method of moments does not always yield efficient estimators, it is the prime alternative procedure in cases in which maximum likelihood estimation is mathematically cumbersome. Since the resulting estimators are consistent, the method is usually acceptable when large numbers of observations are available.[7]

Press (1967) derived the first four cumulants of the distribution of $\Delta S(t)$ in (2.3) with the a priori assumption that the diffusion component has zero drift $(\alpha = 0)$. They are given as:

$$\begin{aligned} K_1 &= \lambda\theta \\ K_2 &= \sigma^2 + \lambda\left(\theta^2 + \delta^2\right) \\ K_3 &= \lambda\theta\left(\theta^2 + 3\delta^2\right) \\ K_4 &= \lambda\left(\theta^4 + 6\theta^2\delta^2 + 3\delta^4\right). \end{aligned}$$

The kurtosis is

$$\gamma_1 = \frac{K_4}{K_2^2} = \frac{\lambda\left(\theta^4 + 6\theta^2\delta^2 + 3\delta^4\right)}{\left[\sigma_1^2 + \lambda\left(\theta^2 + \delta^2\right)\right]^2} \tag{2.30}$$

and the skewness is

$$\gamma_2 = \frac{K_3}{K_2^{3/2}} = \frac{\lambda\theta\left(\theta^2 + 3\delta^2\right)}{\left[\sigma_1^2 + \lambda\left(\theta^2 + \delta^2\right)\right]^{3/2}} \tag{2.31}$$

As the leptokurtosis is defined, γ_1 is positive. γ_2, the skewness, has the same sign as that of θ. Therefore, $\gamma_2 = 0$ if and only if $\theta = 0$, the mean of $\Delta S(t)$.

[7] The relationship between moments and cumulants are dealt in great detail in Kendall and Stuart (1969).

Now let $\bar{K}_i$ denote the ith sample cumulant[8] for $i = 1,2,3,4$. By setting $K_i = \bar{K}_i$ for $i = 1,2,3,4$, we have four equations for the unknown parameters $(\lambda, \sigma^2, \delta^2, \theta)$, which are

$$\begin{aligned}
\lambda &= \frac{\bar{K}_1}{\hat{\theta}}, \\
\hat{\sigma}_1^2 &= \bar{K}_2 - \frac{\bar{K}_1}{\hat{\theta}}\left(\hat{\theta}^2 + \frac{\bar{K}_3 - \bar{K}_1\hat{\theta}^2}{3\bar{K}_1}\right) \\
\hat{\sigma}_2^2 &= \frac{\bar{K}_3 - \bar{K}_1\hat{\theta}^2}{3\bar{K}_1} \\
0 &= \hat{\theta}^4 - \frac{2\bar{K}_3}{\bar{K}_1}\hat{\theta}^2 + \frac{3}{2}\frac{\bar{K}_4}{\bar{K}_1}\hat{\theta} - \frac{\bar{K}_3^2}{2\bar{K}_1^2}
\end{aligned}$$

Press (1967) pointed that if the sample size is not large enough, then the estimates for those two variances can be negative valued. So, they should be zeroed. However, Beckers (1981) paid attention on the fact that Press used unconstrained $\hat{\delta}^2$ in computing $\hat{\sigma}^2$, including negative values. The $\hat{\sigma}^2$ and $\hat{\delta}^2$ parameters are incompatible in those cases where $\hat{\delta}^2$ is put equal to zero.

Beckers (1981) imposed restriction on the mean jump being equal to zero rather than on the diffusion drift. It means that he needed up to the sixth moments to solve for the four parameter without restriction By solving the

[8] If we define the rth sample moments as

$$m_r = \frac{1}{T}\sum_{k=1}^{T}[\Delta Z(k)]^r, \; r = 1, \ldots, 4,$$

then the required relationships between the sample moments and the sample cumulants are

$$\begin{aligned}
\bar{K}_1 &= m_1, \\
\bar{K}_2 &= m_2 - m_1^2, \\
\bar{K}_3 &= m_3 - 3m_1m_2 + 2m_1^3, \\
\bar{K}_4 &= m_4 - 3m_2^2 - 4m_1m_3 + 12m_1^2m_2 - 6m_1^4.
\end{aligned}$$

following system of equations

$$\begin{aligned} K_1 &= \alpha\tau \\ K_2 &= \sigma^2\tau + \lambda\tau\delta^2 \\ K_3 &= K_5 = 0 \\ K_4 &= 3\delta^4\lambda\tau \\ K_6 &= 15\delta^6\lambda\tau, \end{aligned}$$

the estimators are obtained

$$\hat{\alpha} = \bar{K}_1, \tag{2.32}$$

$$\hat{\lambda} = \frac{25\bar{K}_4^3}{3\bar{K}_6^2}, \quad \hat{\delta}^2 = \frac{\bar{K}_6}{5\bar{K}_4}, \quad \hat{\sigma}^2 = \bar{K}_2 - \frac{5\bar{K}_4^2}{3\bar{K}_6}. \tag{2.33}$$

2.3.2 Maximum Likelihood Estimation

The maximum likelihood estimation method for the jump-diffusion processes is considered by Lo(1988). He characterised the exact likelihood function of a discrete sample as a solution to a particular functional partial differential equation.

Suppose that $S(t)$ is a stochastic process defined on a complete probability space, and satisfies a stochastic differential equation given by

$$dS(t) = a(S,t;\alpha)\,dt + b(S,t;\beta)\,dW_t + c(S,t;\gamma)\,dq_\lambda(t),$$

where W_t is the pure Wiener process, $q_\lambda(t)$ is a Poisson counter a, b, and c are known functions which depend upon (S,t) and an unknown parameter vector $\overline{\theta} = [\alpha'\ \beta'\ \gamma']$. With some regularity conditions, the stochastic differential equation has a unique solution.

Now, suppose $S(t)$ is sampled at $n+1$ discrete points in time $t_0,\ t_1,\ \ldots,\ t_n$, not necessary equally spaced and $S \equiv (S_0, S_1, \ldots, S_n)$ denote the random sample where $S_k = S(t_k)$. With the discretely sample data S and the stochastic specification of $S(t)$, let $P(S_0, \ldots, S_n \,;\, \theta)$ denote the finite-dimensional distribution of S and let $\rho(S \,;\, \theta)$ denote the density representation of P. As $X(t)$ is a Markov process, ρ can be written as the product of conditional densities

$$\rho(S) = \rho_0(S_0) \prod_{k=1}^{n} \rho_k(S_k, t_k \mid S_{k-1}, t_{k-1}).$$

It is needed to find the transition density function ρ_k as a solution to a functional partial differential equation

$$\frac{\partial}{\partial t}[\rho_k] = -\frac{\partial}{\partial S}[a\rho_k] + \frac{1}{2}\frac{\partial^2}{\partial^2 S}\left(b^2\rho_k\right) - \lambda\rho_k + \lambda\widetilde{\rho}_k \left|\frac{\partial}{\partial S}\left[\widetilde{c}^{-1}\right]\right| \tag{2.34}$$

subject to

$$\rho_k(S, t_{k-1} | \, S_{k-1}, \, t_{k-1}) = \delta(S - S_{k-1}) \tag{2.35}$$

where $\widetilde{c} = S + c$, $\widetilde{\rho}_k \equiv \rho_k(\widetilde{c}^{-1}, t)$ and δ is the Dirac-delta generalised function centered at S_{k-1}.

Having characterised the likelihood function as the solution to (2.34) and (2.35), we assume its existence and define the maximum likelihood estimator as

$$\hat{\theta}_{ML} \equiv \arg\max_{\theta} G(\theta; S)$$

where

$$G(\theta\,;\, S) \equiv \ln \rho_0(S_0, t_0) + \sum_{k=1}^{n} \ln \rho_k(S_k, t_k \,|\, S_{k-1}, t_{k-1}\,;\, \theta)$$

Since $\hat{\theta}_{ML}$ is a true maximum likelihood estimator, it posseses the standard properties of consistency and asymptotic normality under appropriate regularity conditions.

He showed that when (2.34) and (2.35) do not have explicit solutions, equally spaced discrete sample data are assumed to be generated by a difference equation given by

$$S_{k+1} = S_k + a\left(S_k, t_k\,;\,\alpha\right)h + b\left(S_k, t_k\,;\,\beta\right)\Delta W_{t_{k+1}} + c\left(S_k, t_k\,;\,\gamma\right)\Delta q_\lambda\left(t_{k+1}\right) \quad (2.36)$$

where

$$\begin{aligned} \Delta W_{t_{k+1}} &= W_{t_{k+1}} - W_{t_k}, \quad \Delta q_\lambda\left(t_{k+1}\right) = q_\lambda\left(t_{k+1}\right) - q_\lambda\left(t_k\right), \\ t_k &\equiv kh, \quad \text{for } k = 0, 1, \ldots, n, \quad h \equiv T/n \end{aligned}$$

As discretised sample paths are well-known to converge to those of the continuous-time Itô process $S(t)$ as h converges to 0, this procedure seems sensible. However, the discretized maximum likelihood estimator $\hat{\theta}_D$ need not be consistent.

For example, let $S(t)$ denote the lognormal diffusion process on the time interval $[0, T]$,

$$dS(t) = \alpha S(t)\,dt + \beta S(t)\,dW_t,$$

and consider its discretization according to (2.36) is

$$S_{k+1} = \alpha S_k h + \beta S_k \Delta W_{k+1} \equiv \alpha S_k h + S_k \varepsilon_{k+1},$$

where ε_{k+1} is an i.i.d. $N\left(0, \beta^2 h\right)$ random variate. Then, the discretized maximum likelihood estimator of α and β^2 are

$$\widehat{\alpha}_D = \frac{1}{T}\sum_{k=1}^{n}\left[\frac{S_k}{S_{k-1}} - 1\right], \quad \widehat{\beta}_D^2 = \frac{1}{T}\sum_{k=1}^{n}\left[\frac{S_k}{S_{k-1}} - 1 - \widehat{\alpha}_D h\right]^2, \quad T \equiv nh.$$

However, with fixed h,

$$p\lim_{n\to\infty}\widehat{\alpha}_D = \frac{1}{h}\left(e^{\alpha h} - 1\right) \neq \alpha, \; p\lim\widehat{\beta}_D^2 = \frac{1}{h}e^{2\alpha h}\left(e^{\beta^2 h} - 1\right) \neq \beta.$$

Although, for samll h, the asymptotic bias may be negligoble, it should be clear that for arbitrary coefficient functions a, b, and c the discretized ML estimator is generally inconsistent.

To restore the consistency, we may draw observations more frequently within the fixed time span $[0,T]$. That is, let $h \to 0$, as $n \to \infty$ so as to keep $T \equiv nh$ fixed. Then, $\hat{\beta}_D$ converges to β in probability by continuous data recording whereas $\hat{\alpha}_D$ converges to a Gaussian variate with mean α and variance β^2/T using functional central limit theory.

2.3.3 Indirect Estimation

Jiang (1998) proposed a simulation based indirect estimation method for jump-diffusion processes. Here is the outline of the proposed estimation method following. In his model

$$dS_t = \mu_t(\beta)\,dt + \sigma_t(\beta)\,dW_t + \ln Y_t dq_t,$$

S_t is assumed to be a stationary Markov process and Y_t follows a lognormal distribution, and q_t denotes a Poisson process which is i.i.d over time. The basic idea of indirect inference method is that when a model leads to a complicated structural or reduced from and therefore to intractable likelihood functions, estimation of 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. There are four steps.

First, choose an instrumental or auxiliary model. The simple discretisation of the original model can be a natural choice. So the M_I is given as

$$S_{i+1} = S_i + \mu(t_i;\beta_I)\,\Delta_i + \sigma(t_i;\beta_I)\,\Delta_i^{1/2}\varepsilon_{i+1}^o + \eta\left(\mu_{0I} + \upsilon_I\varepsilon_{i+1}^1\right)$$

with instrumental parameter $\theta_I = (\beta_I, \mu_{0I}, \upsilon_I, \lambda_I) \in \Theta_I$, where $S_i = S_{t_i}$, $\forall i$,

$\Delta_i = t_{i+1} - t_i$, $\varepsilon_{i+1}^j \sim iid\ N(0,1)$, $j = 0,1$; $i = 0,1,\ldots M-1$, and η ~Bernoulli distribution with $P(\eta = 1) = \lambda_I \Delta_i$ and $P(\eta = 0) = 1 - \lambda_I \Delta_i$. M_I and M_O have a one-to-one relationship and the parameter space of M_I has the same dimension as that of M_O. The conditional density function of S_{i+1} given S_i for M_I is

$$f_I(S_{i+1}|\ S_i; \theta_I) = (1 - \lambda_I \Delta_i)\,\phi_1(S_{i+1}) + \lambda_I \Delta_i \phi_2(S_{i+1})$$

where $\phi_1(.)$ is the pdf of the normal distribution with mean $S_i + \mu(t_i; \beta_I)\,\Delta_i$ and variance $\sigma^2(t_i; \beta_I)\,\Delta_i$ and $\phi_2(.)$ is the pdf of the normal distribution with mean $S_i + \mu(t_i; \beta_I)\,\Delta_i + \mu_{0I}$ and the variance $\sigma^2(t_i; \beta_I)\,\Delta_i + \upsilon_I^2$.

Second, the ML estimator of θ_I is given by

$$\hat{\theta}_I = \arg\max_{\theta_I} \sum_{i=0}^{M-1} \ln f_I(S_{i+1}|\ S_i; \theta_I).$$

Then, thirdly, perform path simulation of the original jump-diffusion process and estimation of the instrumental model based on simulated sampling path. Since M_I and f_I are misspecified, the pseudo maximum likelihood (PML) estimator $\hat{\theta}_I$ is generally biased and inconsistent estimator of the true parameter θ. Given values of θ and initial values of S_t at $t = t_0$, simulate the sample path $\widetilde{S}_t$ of S_t, observed at t_0, t_1, t_M for the original model M_O. Redraw the simulations H times. Then, the estimate of the parameter θ_I of M_I, $\hat{\theta}_I^{HM}$ can be obtained from the observations of the simulated sampling path via ML method

$$\hat{\theta}_I^{HM}(\theta) = \arg\max_{\theta_I} \sum_{h=1}^{H} \sum_{i=0}^{M-1} \ln f_I\left(\widetilde{S}_{i+1}^h(\theta) \mid \widetilde{S}_i^h(\theta)\,;\, \theta_I\right)$$

Finally, an indirect estimator θ based on M observations of the simulated sampling path with H drawings, denoted by $\hat{\theta}^{HM}$, is defined by choosing values of θ from which $\hat{\theta}_I$ and $\hat{\theta}_I^{HM}$ are as close as possible. That is,

$$\hat{\theta}^{HM} = \arg\min_{\theta} \left(\hat{\theta}_I(\theta) - \hat{\theta}_I^{HM}(\theta)\right) \Omega \left(\hat{\theta}_I(\theta) - \hat{\theta}_I^{HM}(\theta)\right)'$$

were Ω is a symmetric nonnegative matrix, defining the matrix or the weighting scale.

2.4 Empirical Evidence

2.4.1 Foreign Exchange Rates

For the fixed exchange-rate regimes, discontinuities obviously occur when the parity values are realigned. But with flexible exchange rates, realignment in cross exchange rates, for example, within the European Monetary System (EMS) could be reflected in the exchange rate against the dollar. The arrival of important "news" in the market can generate jumps in exchange rates as well.

Jorion (1989) analysed and compared the empirical distribution of returns in the stock market and in the foreign exchange market. He compares two classes of models and tests whether one is more appropriate than the other. The first model considered is the jump-diffusion process which could explain the skewness in exchange rate distributions and the second one is a diffusion process with time varying parameters, $ARCH$ process. The discontinuities are to be identified even after allowing for diffusion process with time varying parameters. He used daily observations for exchange rates obtained for the period June 1973 to December 1985. Daily stock market return were taken from the Centre for Research in Security Prices (CRSP) database and AMEX stocks. He analysed the weekly and monthly data which is usually chosen for tests of asset pricing models and of models of exchange rate determination. The weekly \$/DM exchange rate exhibits more skewness and more excess kurtosis than the monthly observations.

The high asymptotic t-statistics reveal fat-tailed distributions. When the jump-diffusion process was estimated for the fix-rate period, January 1959 to May 1971, and the jump component is factored in, the volatility of the remaining diffusion process drops dramatically and the drift term becomes much smaller. So, fixed-exchange-rate regimes are characterised by discontinuities that can be modelled by jump processes.

He found the simple diffusion process provide an adequate description for monthly stock returns. For monthly exchange rates, the hypothesis of pure diffusion process is rejected against both the jump-diffusion and $ARCH$ models. The Schwarz Criterion (SC) suggests that the $ARCH$ model is a posteriori most probable by a small margin over the diffusion process. Overall, these results do not present overwhelming evidence against the diffusion model for monthly exchange rate movements.

For weekly data, χ^2 tests indicate that the jump-diffusion model is a significant improvement over the simple diffusion model in both the foreign exchange and stock markets. The estimates of the $ARCH$ process suggest economically important movement in exchange rate volatility. The jump-diffusion process is a posteriori more probable than either the diffusion or the $ARCH$ model for weekly data. The discontinuities are present in the distribution of weekly exchange rates even after explicitly accounting for heteroskedasticity.

Ball and Roma (1993) examined the EMS mechanism and put forward a stochastic model for the affected exchange rates. They model the price movement by means of a bivariate structure which incorporates the particular exchange rate and the institutional restrictions which governs it. An important feature of the model is the explicit inclusion of the central parity into the information evolution process. Their model is

$$S_t = T_t + I_t$$

$$dI_t = -\alpha I_t dt + \beta dW_t$$

where

S_t : the logarithm of the exchange rate,

T_t : the logarithm of the target value of the exchange rate enforced by the central banks,

I_t : the percentage change in exchange rate from the central parity,

W_t : standard Wiener process.

Here I_t follows an Ornstein-Uhlenbeck (OU) process. This OU process is a diffusion with a central restoring tendency. This models the tendency of EMS exchange rates, constrained by central banks' intervention, to fluctuate around a central parity. T_t is assumed to follow a Jump process. If a jump in T occurs, I is restarted randomly and independently around the new value of T. The jump in the logarithm of central parity is defined as J_T and the simultaneous jump in the logarithm of the exchange rate J_X. They set the jump size J_T to be a linear function of displacement from the central parity:

$$J_T(I) = cI + h$$

where $c = (\alpha + \lambda)/\lambda$, $h = \xi/\lambda$, λ is intensity of jump occurring over unit time interval, and ξ is constant.

They used six exchange rates in terms of the Deutsche Mark and time series of weekly observations from March 1979 when the EMS began to March 1991. The dates of realignments are given and the corresponding new value of the bilateral central parity for each exchange rate measured in Deutsche Marks per unit currency. For empirical comparison, they compared three processes for the diffusion component: i) the OU process, ii) Brownian motion and iii) Brownian motion with reflecting barriers. The first and the third processes are mean reverting, which they have the central parity as their long-term mean. In case of the first model, mean reversion is present even when the exchange rate is away from its bilateral limit. In case iii), .the mean reversion is induced by the barrier behaviour and will not be significantly evident inside the band.

They compared the fit of OU and Brownian motion with reflecting barriers by means of an empirical maximum likelihood criterion. A simulation analysis was performed to assess the p-value of the log likelihood ratio $L_{BM} - L_{OU}$. However, as the OU process does not respect the barriers, it should be viewed as a statistic for indicating generic mean reversion within the EMS bands. For each starting point I_0, estimated volatility σ_{BM}, and time series length, 10000 time series were generated with these same characteristics. Reflecting barriers are set at ± 2.25 percent limits (± 6 per cent for the Italian Lira). They generated unrestricted Brownian motion from its increments which are independent and normally distributed. The increment for reflected Brownian motion is generated as in the unrestricted case except when the new value of the process X exceeds the barrier b say. In this case, the value of the reflected process becomes $b - (X - b)$, its reflection about b.

They found that the estimated autocorrelation coefficient, $\widehat{\rho}$, is always below 1 for regimes lasting more than 40 weeks (77 weeks for the Italian Lira), which shows that unrestricted Brownian motion is not an appropriate model. They noted that there is broad evidence of fewer realignments as the EMS system becomes more established. At the same time for the first eight or nine years, the Brownian motion with reflecting barriers provides quite a reasonable model fit.

Vlaar and Palm (1993) modeled EMS exchange rates. First, to model the mean-reversion nature of Exchange Rate Mechanism (ERM) exchange rates, they used MA parameters. For the skewness and leptokurtosis, the combination of normal and stochastic jump process is introduced. The presence of conditional heteroskedasticity is taken into account by using a $GARCH$ specification.

For the jump process, they compared Bernoulli and Poisson process, and for most currencies, they obtained similar results from the two processes. Without considering the stochastic jumps, the MA parameter is not always negative, although not significant. However, after jumps are taken into account, the MA parameter becomes negative and significant for all ERM currencies and the GARCH specification is weakly stationary for all currencies. They found that the expected jumps size is positive, which is in accordance with the positive skewness.

Jiang (1998) estimated weekly observations on the exchange rates of the UK pound, German mark, Japanese yen, and French franc against US dollar by using indirect inference method. He compared four models, which are Black-Scholes diffusion model, Merton's Jump model, $ARCH$ with Jump and Mean-Reversion $ARCH$ with Jump model. The presence of a jump in the second model is tested using likelihood ratio test statistics. The hypothesis for the presence of jump, conditional heteroskedasticity, or mean-reversion in the third and fourth models are tested based on the test statistics derived in Gouriéroux, Monfort, and Renault (1993). He found that jump components are significant in all exchange rates even when conditional heteroskedasticity is considered. Jump frequencies are significantly low for models with strongly significant presence of conditional heteroskedasticity although the jump size tends to be higher in those models confirming that conditional heteroskedasticity can help to remove the volatility that leads to mis-identified jump size and frequency. The mean-reversion is not significant. While the mean-reversion is an important feature for many financial time series, exchange rate processes are essentially not stationary and therefore exhibit no unique stable long-run mean or equilibrium level.

De Jong, Drost, and Werker (1999) develope a relatively simple target zone model. The stylised facts of EMS exchange rates in a target zone, which are mean reversion within the band, strong heteroskedasticity due to a time-varying volatility of the exchange rates and jumps due to realignment also within the band, are captured. They introduce the model, specified by a stochastic differential equation, that takes into account most stylized facts. The models are with realignment and without realignment. In the former model, they fixed the central parity and in the latter, they allowed the central parity to follow a Poisson process with fixed intensity of jumps. With the reason that maximum likelihood estimation can be complicated and cannot be easily extended to other target models, they used the generalised method of moments (GMM) to estimate the models. As the result of GMM estimation of the model with realignments, they found the large estimates for the jump intensity, which says that number of jumps per year is around 2. This may mean that some large changes within the band could be taken for realignment. So they calculated a simple frequency estimator which is the actual number of realignment divided by the total number of observations, and remaining parameters are estimated by GMM considering the intensity of jump as given. There is another interesting point. They found that a smaller estimated variance values from a model with realignment if the jump probability is high than from a model without realignment. This means that the jump process accounts for a substantial part of the variance in exchange rates.

2.4.2 Asset Pricing

As we have mentioned in the earlier section, there have been several estimation methods used to estimate jump-diffusion processes. In most estimation, they have found that the jump components are significant. However, some of the empirical work with data collected in longer sampling period found that jumps are significant but less frequent.

Press (1967) estimated ten of the stocks listed in Dow Jones Industrial average by using the cumulant matching method. The data are collected monthly and range between 1926 and 1960. The parameters are estimated for three different time periods, (a) 1926-1950, (b) 1926-1955, and (c) 1926-1960, so that he can see how much variation in estimates could be attributed to the time span of observation. In fact the resulting estimates were not clearly close to the population value. For example, some of variance estimates turned out to be negative and they had to be zeroed. He considered that is caused by not sufficient data set. However, according to Beckers(1981) , there have been some model specification errors. As Beckers pointed out that Press used unconstrained $\hat{\delta}^2$, the variance of the jump size, when he calculates $\hat{\sigma}^2$, the variance of diffusion part. So, the reported values of those parameters in Press are incompatible in those cases where $\hat{\delta}^2$ is put equal to zero. By restricting, $\hat{\lambda}$, intensity of jump, to be positive, $\hat{\mu}$, mean of jump size, has been restricted implicitly to be positive. This in turn forces $\hat{\delta}^2$ to be negative whenever $\bar{K}_3$ is negative.

Ball and Torous (1983) estimated diffusion with Bernoulli jump processes using the cumulant matching method. They used 47 NYSE listed stocks each with 500 daily returns. They compared Beckers' cumulant method and Bernoulli cumulant estimates, then reported maximum likelihood estimates. Although Beckers' cumulant method produced negative variances, $\hat{\delta}^2$ and $\hat{\sigma}^2$, in 60% of the sampled stocks, it is reduced to 20% by Bernoulli cumulant method. As it is expected, the maximum likelihood estimation does not produce any negative variances. When the cumulant method produces positive variances, the parameter estimates are similar to maximum likelihood estimates. The result confirms the presence of jump in the majority of the sampled common stock returns as well. In their work (1985) , they used a sample of daily return to 30 NYSE common stocks. The estimation confirms the presence of statistically significant jumps in a majority of these returns. But they found there were no significant difference between the Black-Scholes and Merton model prices of the call options written on this sample of common stocks. The only significant differences occur when the underlying common stock return process is predominated by large jumps which occur infrequently.

Kremer and Roenfeldt (1992) investigated jump-diffusion and Black-Scholes models to determine which model provides theoretical values closest to market determined warrant prices. As the warrant has longer maturity period, it will be more probable that there are arrivals of the important new information causing a stock price jump than for the option. Beckers' specification of cumulant matching method used with data of 1,549 observations on 75 warrants from 71 companies during the 56-month period. Considering the whole sample, the jump-diffusion model does not provide expected improvement over the Black-Scholes model. When Merton's dividend adjustment is employed, Black-Scholes model provides the more efficient estimates, while jump-diffusion model is the least biased. They concluded that large reduction in bias accompanied by minor losses in efficiency indicate that the jump-diffusion model probably should be considered when valuing out-of-money, noncallable nonsenior security warrant with maturities in excess of one year, or warrant with underlying stock exhibiting an historically large jump impact.

Oldfield, Rogalski and Jarrow (1977) investigated a common stock returns. The data set they considered is individual transaction information for the 22 trading days during September 1976 for 20 stocks listed on NYSE. The time interval between transactions appears to be highly peaked around the mean then tails off slowly as the number of minutes increases. The distributions for transaction time are strictly positive and highly skewed. The important findings of their work are: i) the variance of the diffusion part is zero. If so, the geometric Brownian motion process may not be correct over the period sampled. ii) common stock returns follow an autoregressive jump process. iii) the gamma distribution shows better fit for the time interval between transactions than the exponential distribution. And iv) with their data set they could not draw definite conclusion on the normality of the jump amplitude.

In the theoretical finance literature, it is assumed that the option prices follow diffusion processes with continuous sample paths. Jarrow and Rosenfeld (1984) generalized the assumption to discrete sample paths by extending Merton's (1973) intertemporal asset pricing model, in the special case of a constant investment opportunity set, to include discontinuous sample paths for asset prices. The constant investment opportunity set assumption is imposed because of obtaining sufficient condition under which an instantaneous capital asset pricing model (CAPM) results. Jarrow and Rosenfeld (1984) tested the hypothesis whether jump risk is diversifiable or not by using two market indices. For the daily market index, 2 out of 4,133 observations exhibit a daily return of greater than 5%, which makes these two observations prime candidate for jumps and for the monthly observations, 14 out of 633 observations show larger than 15% movement in a given month. This does not indicate a significant jump component for this market index. For some daily sample periods, a statistically significant jump components were found. This is confirmed by likelihood ratio tests. All the daily sample periods reject the null of a continuous sample path process at a 99% significant level. There could be measurement errors that can be induced as jumps. But after the correction procedure, the jump components are still significant although the size of jumps for some of the stocks is very low.

2.4.3 Others

There are several works investigating jump-diffusion processes in the areas of financial economics other than asset pricing or foreign exchange rates. Some of the research is in the literature of term structure of interest rates. The first recognised attempt in the area is the work by Ahn and Thompson (1988). They investigated the effect of jump components of the underlying processes on the term structure of interest rates. They found that the Merton's multi-beta CAPM does not hold in general. The discontinuous movements of the investment opportunities cannot be completely captured by a single consumption beta. The equilibrium interest rate under jump-diffusion process is strictly lower than one under diffusion process, ceteris paribus. The traditional expectation theory is not consistent with the equilibrium models under jump-diffusion processes. This is due to the term premium is additionally affected by the jump risk. Finally, the covariability with technological jump changes is priced even in the case of the logarithmic utility function under the jump diffusion processes. The consumers with logarithmic utility functions would appreciate the hedging service of an asset against the uncertain jump changes of the underlying technologies.

In the literature of commodity future prices, the jump-diffusion processes are applied, as the log-return on commodity future prices are not normally distributed according to the empirical findings. Hilliard and Reis (1999) modelled the stochastic process underlying commodity option prices with Bates' (1991) European option pricing formula. They collected data for call and put options on soybean futures and for the underlying future contracts traded on the Chicago Board of Trade. The data consist of the time and price for every transaction in which the price changed from the previous transaction for the period July 1990 until June 1992. As the log-return on commodity future prices not normally distributed, they compared the out-of-sample performance of diffusion and jump-diffusion models. They concluded that the mean-jump size and frequency of jumps are consistently positive for all estimation days. By using the pure diffusion model, the put option can be overpriced and the call option underpriced.

2.5 Conclusion

In this chapter, we have surveyed most of the literature related to the jump-diffusion processes in financial economics. As Black-Scholes' diffusion process can not fully describe the behaviour of financial time series, Merton (1976) introduced the Poisson jump process to model the abnormal changes in the stock prices. Since then, the jump-diffusion process have draw attentions from many researchers. Many empirical findings show that the financial time series indeed includes jump components.

At the moment, the research is most focused on the stock market analysis. But it is believed that the underlying processes of foreign exchange rates have similar characteristic as those of common stocks. Some empirical work on the ERM foreign exchange rates had confirmed this fact. This jump-diffusion model can be very useful to apply to the commodity future markets as well.

So far most of the works imposed restrictions on the intensity of the jump, λ, as a fixed constant. For future research, it might be interesting to find out the statistical properties of the intensity of jumps.

Chapter 3

Approximation of Jump-Diffusion Processes

3.1 Introduction

During the last couple of decades or so, many researchers have found that the value of option prices is not continuous with probability one. Cox and Ross (1975) assumed that the new information arriving at a market is a lump sum causing a discrete jump in the values of options and derived the option pricing formula by using Poisson jump process. Merton, in his works ($1976a, b$), decomposed the total change in the stock prices into two components: 1) systematic risk which is typically modelled by a Brownian motion, 2) nonsystematic risk which represents the arrivals of new information, in other word, shock, to the market, which can be modelled by the Poisson jump process, and derived the option pricing formula with jump-diffusion process. After these researches, as the solution to a stochastic asset optimization problem, the jump-diffusion process is popularly used in the dynamic asset pricing literature [e.g., Oldfield, Rogalski, and Jarrow (1977), Ball and Torous (1983, 1985), Jarrow and Rosenfeld (1984), Amin (1993), Kim, Oh, and Brooks (1994), Chang (1995)] as well as other financial economic literature such as in the term structure of interest rates [e.g., Ahn and Thompson (1988), Das (1997)], in foreign exchange rates [Jorion (1988), Ball and Roma (1993) , Park, Ahn and Fujihara (1993) , Vlaar and Palm (1993)], etc.

Ball and Torous (1983) considered the Bernoulli process to model the arrivals of information in a market and estimate the model with 47 NYSE listed stocks each with 500 daily return observations. In their other work (1985) , they estimated the Poisson jump-diffusion process with 30 daily common stock returns. They found the evidence that jump components are present in a majority of the stocks examined.

Another point we need to consider is that the financial time series is found to be highly heteroskedastic over time. There are massive amount of literature documenting the heteroskedastic nature of the financial time series data. With $ARCH$ models introduced by Engle (1983) , the heteroskedastic nature of the data is well explained by the $ARCH$ type models. There are good survey papers about the $ARCH$ type models such as, Bollerslev, Chou, and Kroner (1992) and Bollerslev, Engle and Nelson (1994) .

In this chapter, we are trying to develop the relationship between the continuous time stochastic differential equation used in the theoretical literature of financial economics and the discrete time difference equation used in the lots of empirical works. In section 3.2, we will sketch the main results in weak convergence of a sequence of stochastic difference equation to a jump-diffusion process. Section 3.3 will, then, present the $ARCH$ jump-diffusion approximation. We show that it can approximate a wide variety of Generalised Ito processes which are jump-diffusion processes. We show examples based on the $AR(1)$ Exponential $ARCH$ model of Nelson (1991) . Finally, we will conclude this chapter in section 3.4.

In Appendix A, we show the higher moments up to fourth order for a process considered in section 3.3. The higher moments greater than second order for continuous diffusion part of the process converge to zero as the sampling intervals go to zero, while those of Poisson jump part of the process converge to a finite limit. Some regularity conditions for a martingale problem are stated in Appendix B. These conditions will be useful to prove the results in weak convergence. All the proofs of theorems will be presented in Appendix C.1.

3.2 Weak Convergence of the Processes

3.2.1 The Main Result in Weak Convergence

Here we want to show the weak convergence of a discrete time process to a jump-diffusion process. That is, we present general conditions for a sequence of finite-dimensional discrete time Markov process $\{{}_hX_t\}_{h\downarrow 0}$ to converge weakly to a jump-diffusion process. The basic theoretical setup is following.

Let $D([0,\infty), R^n)$ be the space of mappings from $[0,\infty)$ into R^n that are right continuous having finite left limits and let $B(R^n)$ denote the Borel sets on R^n. With introducing an appropriate Skorohod metric, $D([0,\infty), R^n)$ becomes a complete metric space.[1] For each $h>0$, let $\mathfrak{M}_{kh}$ be the σ-algebra generated by ${}_hX_0,\ {}_hX_h,\ {}_hX_{2h},\ \ldots,\ {}_hX_{kh}$, and let ν_h be a probability measure on $(R^n, B(R^n))$. For each $h>0$, and each $k = 0,\ 1,\ 2,\ \ldots$, let $\Pi_{h,kh}(x,\cdot)$ be a transition function on R^n. That is,

i) $\Pi_{h,kh}(x,\cdot)$ is a probability measure on $(R^n, B(R^n))$ for all $x \in R^n$,

ii) $\Pi_{h,kh}(\cdot,\Gamma)$ is $B(R^n)$ measurable for all $\Gamma \in B(R^n)$.

For each $h>0$, let P_h be the probability measure on $D([0,\infty), R^n)$ such that

$$P_h[{}_hX_0 \in \Gamma] = \nu_h(\Gamma) \text{ for any } \Gamma \in B(R^n), \tag{3.1}$$

$$P_h[{}_hX_t = {}_hX_{kh},\ kh \le t < (k+1)h] = 1, \tag{3.2}$$

[1] See Kushner (1984) Section 4.3 in Chapter 2.

$$P_h\left[{}_hX_{(k+1)h}\in\Gamma\mid\mathfrak{M}_{kh}\right]=\Pi_{h,kh}\left({}_hX_{kh},\Gamma\right)\text{ almost surely under }P_h \quad \text{for all } k\geq 0 \text{ and } \Gamma\in B\left(R^n\right). \tag{3.3}$$

Here, for each $h>0$, we specify the distribution of the random starting point by (3.1) and form a continuous time process ${}_hX_t$ from the discrete time process ${}_hX_{kh}$ by (3.2) making ${}_hX_t$ a step function with jumps at time h, $2h$, $3h$, and so on. (3.3) specifies the transition probabilities of n-dimensional discrete time Markov process ${}_hX_{kh}$.

Now, define, for each $h>0$,

$$a_h^{ij}(x,t) \equiv h^{-1}\int_{\|y-x\|\leq 1}(y_i-x_i)(y_j-x_j)\,\Pi_{h,h[t/h]}(x,dy), \tag{3.4}$$

$$b_h^{i}(x,t) \equiv h^{-1}\int_{\|y-x\|\leq 1}(y_i-x_i)\,\Pi_{h,h[t/h]}(x,dy), \tag{3.5}$$

$$\Delta_h^{\varepsilon}(x,t) \equiv h^{-1}\int_{\|y-x\|>\varepsilon}\Pi_{h,h[t/h]}(x,dy). \tag{3.6}$$

$$g_h(x,t) = x_t-x_{t-},\quad x_{t-}=\lim_{s\to t}x_s \quad \text{for } s<t \tag{3.7}$$

where $[t/h]$ is the integer part of t/h.

$a_h(x,t)$ is a measure of the truncated second moment per unit of time, $b_h(x,t)$ is a measure of truncated drift per unit of time, and $\Delta_h^{\varepsilon}(x,t)$ is a probability that the process has magnitude of a jump greater than ε per unit of time. We define the truncated first and second moment for the process x, since the usual conditional moments for the process may not be finite. For example, if $X_t=\exp\left[\exp W_t\right]$, where W_t is a Wiener process, X_t is a diffusion process, but there exist no moments of any order. And $g_h(x,u)$ measures the size of jump per unit of time. We suppose that the jumps occur with probability $\lambda h+o(h)$ in the time interval $(t,t+h]$. As we assume the process is right continuous with finite left limit, there will be only discontinuity of the first kind (i.e., discrete jumps) and the jump size will be finite.

Now, we state the assumptions which are required to obtain the weak convergence result. Let S^n denote the space of $n \times n$ matrices and let S_+^n denote the space of $n \times n$ symmetric nonnegative definite matrices.

Assumption 1. Let $a(x,t) : R^n \times [0,\infty) \to S_+^n$, $b(x,t) : R^n \times [0,\infty) \to R^n$ and $g(x,t) : R^n \times [0,\infty) \to R^n$ be continuous measurable mappings which are continuous in x for each $t \geq 0$. We assume that for all $R > 0$, $T > 0$ and $\varepsilon > 0$

$$\lim_{h\downarrow 0} \sup_{\|x\|\leq R,\, 0\leq t\leq T} \|a_h(x,t) - a(x,t)\| = 0 \tag{3.8}$$

$$\lim_{h\downarrow 0} \sup_{\|x\|\leq R,\, 0\leq t\leq T} \|b_h(x,t) - b(x,t)\| = 0 \tag{3.9}$$

$$\lim_{h\downarrow 0} \sup_{\|x\|\leq R,\, 0\leq t\leq T} \|g_h(x,t) - g(x,t)\| = 0 \tag{3.10}$$

$$\lim_{h\downarrow 0} \sup_{\|x\|\leq R,\, 0\leq t\leq T} \Delta_h^{\varepsilon}(x,t) = \lambda \tag{3.11}$$

This assumption requires that the second moment, drift, and jumps per unit of time converge uniformly on compact sets to well-behaved functions of time and the state variables x. The probability of jump of size greater then ε converges to a constant λ. So, the sample paths of the limit process will have only discontinuity of the first kind with probability one.

Assumption 2. Let $\sigma(x,t) : [0,\infty) \times R^n \to S^n$ be a continuous measurable mapping such that for all $x \in R^n$ and all $t \geq 0$,

$$a(x,t) = \sigma(x,t)\,\sigma(x,t)'. \tag{3.12}$$

This assumption requires that the function $a(x,t)$, the second moment per unit of time of the limit process, has a well-behaved matrix square root $\sigma(x,t)$.

Assumption 3. As $h \to 0$, ${}_hX_0$ converges in distribution to a random variable X_0 with probability measure ν_0 on $(R^n, B(R^n))$.

This assumption requires that the probability measure ν_h of the random starting points ${}_hX_0$ to converge to a limit measure ν_0 as $h \to 0$.

With all the assumptions we made above, we specified an initial probability measure ν_0 of the limit process, an instantaneous drift function $b(x,t)$, an instantaneous covariance matrix $a(x,t)$, and a jump amplitude $g(x,t)$. We have supposed that the sample path of the process is discontinuous with probability one. However, there is no guarantee that a limit process is finite or is uniquely defined. There are a number of works considering the conditions under which ν_0, $a(x,t)$, and $b(x,t)$ uniquely define a diffusion limit process. Especially Strook and Varadhan (1979) studied extensively about the diffusion limit process. Either and Kurtz (1986) considers the martingale problems with Levy measure. The conditions of unique existence of a solution to a jump-diffusion limit process can be found in Gihman and Skorohod (1972)[2]. The non-exploding condition for jump-diffusion limit will be stated in Appendix B.

Assumption 4. ν_0, $a(x,t)$, $b(x,t)$, and $g(x,t)$ uniquely specify the distribution of a jump diffusion process X_t with initial distribution ν_0, diffusion matrix $a(x,t)$, drift vector $b(x,t)$, and jump amplitude $g(x,t)$.

Theorem 1 *Under Assumptions 1 through 4, the sequence of ${}_hX_t$ processes defined by* (3.1) *to* (3.3) *converges weakly as $h \to 0$ to the X_t process defined by the stochastic integral equation*

$$X_t = X_0 + \int_0^t b(X_s,s)\,ds + \int_0^t \sigma(X_s,s)\,dW_s + \int_0^t \int g(X_{s^-},s)\,\widetilde{N}_\lambda(ds) \quad (3.13)$$

where W_t is an n-dimensional standard Brownian motion, independent of X_0, $\widetilde{N}_\lambda(ds)$ is the compensated Poisson process defined as $\widetilde{N}_\lambda(ds) = N(ds) - \lambda ds$[3] *and where for any $\Gamma \in B(R^n)$, $P(X_0 \in \Gamma) = \nu_0(\Gamma)$. Such an X_t process exists*

[2] See Chapter 2, Part II in Gihman and Skorohod (1972) for more detail.

[3] *$N(ds)$ is a Poisson process, and λ is a constant probability of jump in the Poisson process, and $\lambda > 0$.*

and is distributionally unique. This distribution does not depend on the choice of $\sigma(\cdot,\cdot)$ *made in Assumption 2. Finally,* X_t *remains finite in finite time intervals almost surely, i.e. for all* $T > 0$,

$$P\left[\sup_{0\leq t\leq T}\|X_t\| < \infty\right] = 1 \tag{3.14}$$

Proof. See Appendix C.1. ■

Now, we want to make the above result a bit more general. For each i, $i = 1, 2, \ldots, n$, each $\delta > 0$, and each $h > 0$, define

$$\gamma_{h,i,\delta}(x,t) \equiv h^{-1}\int_{R^n} |(y-x)_i|^{2+\delta}\,\Pi_{h,h[t/h]}(x,dy), \tag{3.15}$$

where $(y-x)_i$ is the ith element of the vector $(y-x)$. If for some $\delta > 0$ and all i, $i = 1, 2, \ldots, n$, $\gamma_{h,i,\delta}(x,t)$ is finite, then the following integral will be well-defined and finite

$$\begin{aligned} a_h^*(x,t) &\equiv h^{-1}\int_{R^n}(y-x)(y-x)'\,\Pi_{h,h[t/h]}(x,dy), \\ b_h^*(x,t) &\equiv h^{-1}\int_{R^n}(y-x)\,\Pi_{h,h[t/h]}(x,dy), \end{aligned}$$

They are the same measures as $a_h(\cdot,\cdot)$, $b_h(\cdot,\cdot)$ and $g_h(\cdot,\cdot)$, but integration is taken over R^n rather than $|y-x| \leq 1$.

Assumption 1′. There exist $\delta > 0$ such that for each $R > 0$, each $T > 0$, and each i, $i = 1, 2, \ldots, n$,

$$\lim_{h\to 0}\ \sup_{\|x\|\leq R,\ 0\leq t\leq T}\gamma_{h,i,\delta}(x,t) = 0. \tag{3.16}$$

Further, let $a(x,t) : R^n \times [0,\infty) \to S_+^n$, $b(x,t) : R^n \times [0,\infty) \to R^n$ and $g(x,t) : R^n \times [0,\infty) \to R^n$ be continuous measurable mappings which are continuous in x for each $t \geq 0$. We assume that for all $R > 0$, $T > 0$

$$\lim_{h \downarrow 0} \sup_{\|x\| \leq R,\, 0 \leq t \leq T} \|a_h^*(x,t) - a(x,t)\| = 0, \tag{3.17}$$

$$\lim_{h \downarrow 0} \sup_{\|x\| \leq R,\, 0 \leq t \leq T} \|b_h^*(x,t) - b(x,t)\| = 0, \tag{3.18}$$

$$\lim_{h \downarrow 0} \sup_{\|x\| \leq R,\, 0 \leq t \leq T} \|g_h(x,t) - g(x,t)\| = 0, \tag{3.19}$$

$$\lim_{h \downarrow 0} \sup_{\|x\| \leq R,\, 0 \leq t \leq T} \Delta_h^{\varepsilon}(x,t) = \lambda \tag{3.20}$$

Theorem 2 *Under Assumptions 1′, and 2 through 4, the conclusions of Theorem 1 hold.*

Proof. See Appendix C.1 ■

As stated in Merton(1990, Ch.3), Assumption 1′ implies that the moments higher than two vanish to zero at an appropriate rate as $h \longrightarrow 0$.

3.2.2 Example: $GARCH\,(1,1)$-M Model

In Engle and Bollerslev (1986), they presented the $GARCH\,(1,1)$-M process for the cumulative excess returns y_t on a portfolio. The excess return process is

$$\begin{aligned} y_t &= y_{t-1} + \mu\sigma_t^2 + \sigma_t Z_t, \\ \sigma_{t+1}^2 &= \omega + \sigma_t^2\left[\beta + \alpha Z_t^2\right]. \end{aligned}$$

where $Z_t \sim$iid $N(0,1)$.

Let us suppose that a stochastic process in discrete time is including the jump process as follows;

$$y_t = y_{t-1} + \mu\sigma_t^2 + \sigma_t Z_t + c\eta_t, \tag{3.21}$$

$$\sigma_{t+1}^2 = \omega + \sigma_t^2\left(\beta + \alpha Z_t^2\right), \tag{3.22}$$

where $Z_t \sim iid\ N(0,1)$ and η_t ~Bernoulli distributed with $\Pr(\eta_t = 0) = 1 - \lambda dt + o(dt)$[4] and $\Pr(\eta_t = 1) = \lambda dt + o(dt)$. Here c denotes the jump size of the process when a jump occurs.

Now, we partition the time interval more and more finely and examine the properties of the stochastic difference equation system. We allow the parameters α, β, and ω to depend on h and make the drift term in (3.21) proportional to h. Then we may rewrite the system of stochastic processes (3.21) and (3.22) as

$$\begin{aligned} {}_h y_{kh} &= {}_h y_{(k-1)h} + h \cdot \mu_h\, {}_h\sigma^2_{kh} + h^{1/2}\, {}_h\sigma_{kh} \cdot {}_h Z_{kh} \\ &\quad + {}_h\eta_{kh}\left(c_h + \phi_h\, {}_h Z_{kh}\right), \qquad (3.23) \\ {}_h\sigma^2_{(k+1)h} &= \omega_h + {}_h\sigma^2_{kh}\left(\beta_h + \alpha_h \cdot {}_h Z^2_{kh}\right), \qquad (3.24) \end{aligned}$$

and

$$\Pr\left[\left({}_h y_0,\ {}_h\sigma^2_0\right) \in \Gamma\right] = \upsilon_h(\Gamma) \qquad \text{for any } \Gamma \in B\left(R^2\right), \qquad (3.25)$$

where ${}_h Z_{kh}$ ~iid $N(0,1)$ and ${}_h\eta_{kh}$ ~Bernoulli distributed with $\Pr[{}_h\eta_{kh} = 0] = 1 - \lambda h + o(h)$ and $\Pr[{}_h\eta_{kh} = 1] = \lambda h + o(h)$. υ_h satisfies Assumption 3 as $h \to 0$, and for each $h \geq 0$, $\upsilon_h\left((y_0, \sigma^2_0) : \sigma^2_0 > 0\right) = 1$. And we create the continuous time processes ${}_h y_t$ and ${}_h\sigma^2_t$ by

$${}_h y_t \equiv {}_h y_{kh} \text{ and } {}_h\sigma^2_t \equiv {}_h\sigma^2_{kh} \text{ for } kh \leq t < (k+1)h. \qquad (3.26)$$

We want to find out which sequences $\{\omega_h, \alpha_h, \beta_h\}$ make the $\{{}_h\sigma^2_t, {}_h y_t\}$ process converge in distribution a jump-diffusion mixed process as $h \to 0$.

Let $\mathfrak{M}_{kh}$ be the σ-algebra generated by kh, ${}_h y_0$, ${}_h y_h$, $\ldots$, ${}_h y_{(k-1)h}$, and ${}_h\sigma^2_0$,

[4]Let $f(h)$ and $g(h)$ be functions of h. $f(h) = o(g(h))$, if $\lim_{h\to 0}[f(h)/g(h)] = 0$ and $f(h) = O(g(h))$, if $\lim_{h\to 0}[f(h)/g(h)]$ is bounded.

${}_h\sigma^2_h$, ..., ${}_h\sigma^2_{kh}$. Then the first moment of the process is

$$E\left[h^{-1}\left({}_hy_{kh} - {}_hy_{(k-1)h}\right)|\mathfrak{M}_{kh}\right] = \mu_h\, {}_h\sigma^2_{kh} + \lambda c_h \tag{3.27}$$

$$E\left[h^{-1}\left({}_h\sigma^2_{(k+1)h} - {}_h\sigma^2_{kh}\right)|\mathfrak{M}_{kh}\right] = h^{-1}\omega_h + h^{-1}\sigma^2_{kh}\left(\beta_h + \alpha_h - 1\right) \tag{3.28}$$

To satisfy Assumption 1′, we require the following limits exist and be finite;

$$\lim_{h\to 0} h^{-1}\omega_h = \omega \tag{3.29}$$

$$\lim_{h\to 0} h^{-1}\left(1 - \beta_h - \alpha_h\right) = \theta \tag{3.30}$$

As it is stated in Bollerslev(1986) , it is necessary to require that ω_h, α_h, and β_h be nonnegative because σ^2_t should be remain positive with probability one. Therefore, $\omega \geq 0$ while θ could be of either sign.

Then,

$$\lim_{h\to 0} E\left[h^{-1}\left({}_hy_{kh} - {}_hy_{(k-1)h}\right)|\mathfrak{M}_{kh}\right] = \mu\sigma^2 + \lambda c \tag{3.31}$$

$$\lim_{h\to 0} E\left[h^{-1}\left({}_h\sigma^2_{(k+1)h} - {}_h\sigma^2_{kh}\right)|\mathfrak{M}_{kh}\right] = \omega - \theta\sigma^2 \tag{3.32}$$

The second moment per unit of time is

$$E\left[h^{-1}\left({}_hy_{kh} - {}_hy_{(k-1)h}\right)^2|\mathfrak{M}_{kh}\right]$$
$$= h\mu^2_h\, {}_h\sigma^4_{kh} + {}_h\sigma^2_{kh} + \lambda\left(c^2_h + \phi^2_h\right) + 2\lambda h c_h\mu_h\, {}_h\sigma^2_{kh}, \tag{3.33}$$

$$E\left[h^{-1}\left({}_h\sigma^2_{(k+1)h} - {}_h\sigma^2_{kh}\right)^2|\mathfrak{M}_{kh}\right]$$
$$= h^{-1}\omega^2_h + h^{-1}\, {}_h\sigma^4_{kh}\left(\alpha_h + \beta_h - 1\right)^2 + 2h^{-1}\alpha^2_h\, {}_h\sigma^4_{kh}$$

$$+h^{-1}2\omega_{hh}^2\sigma_{kh}^4\left(\alpha_h+\beta_h-1\right), \tag{3.34}$$

$$E\left[h^{-1}\left({}_hy_{kh}-{}_hy_{(k-1)h}\right)\left({}_h\sigma^2_{(k+1)h}-{}_h\sigma^2_{kh}\right)|\mathfrak{M}_{kh}\right]$$
$$=\mu_h\,{}_h\sigma^2_{kh}\,\omega_h+\mu_h\,{}_h\sigma^4_{kh}\left(\alpha_h+\beta_h-1\right)$$
$$+\lambda_h c_h\omega_h+\lambda_h c_h\,{}_h\sigma^2_{kh}\left(\alpha_h+\beta_h-1\right) \tag{3.35}$$

With (3.29) and (3.30) and assuming that

$$\lim_{h\to 0}2h^{-1}\alpha_h^2=\alpha^2, \tag{3.36}$$

exist and is finite.. Also α^2 is always greater than 0. Then we have

$$E\left[h^{-1}\left({}_hy_{kh}-{}_hy_{(k-1)h}\right)^2|\mathfrak{M}_{kh}\right]={}_h\sigma^2_{kh}+\lambda\left(c_h^2+\phi_h^2\right)+o\left(1\right), \tag{3.37}$$

$$E\left[h^{-1}\left({}_h\sigma^2_{(k+1)h}-{}_h\sigma^2_{kh}\right)^2|\mathfrak{M}_{kh}\right]=\alpha_h^2\,{}_h\sigma^4_{kh}+o\left(1\right), \tag{3.38}$$

$$E\left[h^{-1}\left({}_hy_{kh}-{}_hy_{(k-1)h}\right)\left({}_h\sigma^2_{(k+1)h}-{}_h\sigma^2_{kh}\right)|\mathfrak{M}_{kh}\right]=o\left(1\right) \tag{3.39}$$

where if $\psi\left(h\right)=o\left(1\right)$, then $\lim_{h\to 0}\psi\left(h\right)=0$. We can show that the third and fourth moments of the process ${}_h\sigma_t^2$ exist and converge to zero as $h\to 0$, and those of the process ${}_hy_t$ exist and $O\left(h\right)$. The detailed calculation is provided in Appendix A.

Then, we can define the coefficients in the jump-diffusion mixed process as

$$b\left(y,\sigma^2\right)\equiv\begin{bmatrix}\mu\sigma^2+\lambda c\\ \omega-\theta\sigma^2\end{bmatrix} \tag{3.40}$$

$$a\left(y,\sigma^2\right) \equiv \begin{bmatrix} \sigma^2+\lambda\left(c^2+\phi^2\right) & 0 \\ 0 & \alpha^2\sigma^4 \end{bmatrix} \tag{3.41}$$

$$g\left(y,\sigma^2\right) \equiv \begin{bmatrix} c \\ 0 \end{bmatrix} \tag{3.42}$$

and, if α_h, β_h and ω_h satisfy the conditions in (3.29), (3.30) and (3.36), then Assumption 1′ holds. If we suppose that $\sigma(\cdot,\cdot)$ in Assumption 2 is the element-by-element square root of $a\left(y,\sigma^2\right)$, then Assumption 2 holds as well. From (3.40) - (3.42), we can write the jump-diffusion limit as

$$dy_t = \left(\mu\sigma^2+\lambda c\right)dt + \left[\sigma^2+\lambda\left(c^2+\phi^2\right)\right]^{1/2} dW_{1,t} + c\eta_t \tag{3.43}$$

$$d\sigma^2_{(t+1)} = \left(\omega-\theta\sigma^2\right)dt + \alpha\sigma^2 dW_{2,t} \tag{3.44}$$

$$P\left[\left(y_0,\sigma_0^2\right)\in\Gamma\right] = \nu_0(\Gamma) \quad \text{for any } \Gamma\in B\left(R^2\right) \tag{3.45}$$

where $W_{i,t}$, $i=1,2$, are independent standard Wiener processes and are also independent of the Bernoulli process, η_t. All those three independent processes are independent of the initial values $\left(y_0,\sigma_0^2\right)$.

3.3 Jump-Diffusion Approximation

In this section, we will present that *ARCH* models can be used as approximation of generalized Itô process (i.e., jump-diffusion process).

3.3.1 *ARCH* Jump-Diffusion Approximation

Define the stochastic differential equation system

$$\begin{aligned} dy_t &= f\left(y_t,s_t,t\right)dt + g\left(y_t,s_t,t\right)dW_{1,t} \\ &\quad + \left\{k\left(y_t,s_t,t\right) + dt^{-1/2}\phi\left(y_t,s_t,t\right)dW_{1,t}\right\}d\eta_t, \end{aligned} \tag{3.46}$$

$$ds_t = F(y_t, s_t, t)\, dt + G(y_t, s_t, t)\, dW_{2,t}, \tag{3.47}$$

$$\begin{bmatrix} dW_{1,t} \\ dW_{2,t} \end{bmatrix} \begin{bmatrix} dW_{1t} & dW'_{2,t} \end{bmatrix} = \begin{bmatrix} 1 & \Omega_{1,2} \\ \Omega_{2,1} & \Omega_{2,2} \end{bmatrix} dt = \Omega dt, \tag{3.48}$$

$$\begin{bmatrix} d\eta_t \\ \mathbf{0} \end{bmatrix} \begin{bmatrix} d\eta_t & \mathbf{0}' \end{bmatrix} = \begin{bmatrix} \lambda & \mathbf{0}_{1,2} \\ \mathbf{0}_{2,1} & \mathbf{0}_{2,2} \end{bmatrix} dt = \Lambda dt, \text{ and} \tag{3.49}$$

$$\Sigma dt = \Omega dt + \Lambda dt \tag{3.50}$$

where Ω and Λ are $(n+1) \times (n+1)$ positive semi-definite matrices of rank two or less, $\mathbf{0}_{1,2}$, $\mathbf{0}_{2,1}$, $\mathbf{0}_{2,2}$ are $(n \times 1)$ column vector, $(1 \times n)$ row vector and, $(n \times n)$ matrix of zeros, respectively, s_t is an n-dimensional vector of unobservable state variables, y is an observable scalar process, $W_{1,t}$ is one dimensional standard Wiener process, $W_{2,t}$ is an n-dimensional standard Wiener process, η_t is a Poisson process with intensity λ, $f(s_t, y_t, t)$, $g(s_t, y_t, t)$, and $k(s_t, y_t, t)$ are real-valued continuous scalar functions, and $F(s_t, y_t, t)$ and $G(s_t, y_t, t)$ are real, continuous $n \times 1$ and $n \times n$ valued functions, respectively. The initial values of the process (y_0, s_0) is assumed to be random and independent of $W_{1,t}$, $W_{2,t}$, and η_t, and $W_{1,t}$, $W_{2,t}$, and η_t, are independent of each other.

Define the vector and matrix functions a, b, and c by

$$a(y, s, t) = \begin{bmatrix} g^2 + \lambda\left(k^2 + \phi^2\right) & g\Omega_{1,2}G' \\ G\Omega_{2,1}g & G\Omega_{2,2}G' \end{bmatrix}, \tag{3.51}$$

$$b(y, s, t) = \begin{bmatrix} f + \lambda k & F' \end{bmatrix}', \tag{3.52}$$

$$c(y, s, t) = \begin{bmatrix} k & \mathbf{0}' \end{bmatrix}' \tag{3.53}$$

where $\mathbf{0}$ is an $n \times 1$ vector of zeros. Then, $b(y, s, t)$ and $c(y, s, t)$ are $(n+1) \times 1$ vectors and $a(y, s, t)$ is $(n+1) \times (n+1)$ matrix.

Now, we define a sequence of approximating processes that converge to (3.46)-

(3.48) in distribution as $h \to 0$.

$$
\begin{aligned}
{}_h y_{kh} &= {}_h y_{(k-1)h} + f\left({}_h y_{kh},\ {}_h s_{kh}, kh\right) h + g\left({}_h y_{kh},\ {}_h s_{kh}, kh\right)\ {}_h Z_{kh} \\
&\quad + \eta_{kh}\left(k\left({}_h y_{kh},\ {}_h s_{kh}, kh\right) + h^{-1/2}\phi\left({}_h y_{kh},\ {}_h s_{kh}, kh\right)\ {}_h Z_{kh}\right), \quad (3.54) \\
{}_h s_{(k+1)h} &= {}_h s_{kh} + F\left({}_h y_{kh},\ {}_h s_{kh}, kh\right) h + G\left({}_h y_{kh},\ {}_h s_{kh}, kh\right)\ {}_h Z^*_{kh}, \quad (3.55)
\end{aligned}
$$

where

$$
\begin{aligned}
{}_h Z_{kh} &\sim \text{i.i.d. } N(0, h), \quad (3.56) \\
{}_h Z^*_{kh} &= \begin{bmatrix} \theta_1\ {}_h Z_{kh} + \gamma_1\left[\left|{}_h Z_{kh}\right| - \left(\frac{2h}{\pi}\right)^{1/2}\right] \\ \theta_n\ {}_h Z_{kh} + \gamma_n\left[\left|{}_h Z_{kh}\right| - \left(\frac{2h}{\pi}\right)^{1/2}\right] \end{bmatrix}, \quad (3.57)
\end{aligned}
$$

and the coefficients $\{\theta_1, \gamma_1, \ldots, \theta_n, \gamma_n\}$ are selected so that

$$
E\begin{bmatrix} {}_h Z_{kh} + \eta_{kh} \\ {}_h Z^*_{kh} \end{bmatrix}\begin{bmatrix} {}_h Z_{kh} + \eta_{kh} & {}_h Z^*_{kh} \end{bmatrix} = \Sigma dt. \quad (3.58)
$$

Now we can convert the discrete time process $[{}_h y_{kh},\ {}_h s'_{kh}]$ into a continuous time process by defining

$$
{}_h y_t \equiv {}_h y_{kh},\ {}_h s_t \equiv {}_h s_{kh} \quad \text{for } kh \leq t < (k+1)h. \quad (3.59)
$$

Theorem 3 *If $a(y,s,t)$, $b(y,s,t)$ and $c(y,s,t)$ satisfy Assumption 4, with $x \equiv [y, s']$, and if the joint probability measures v_h of the starting values $({}_h y_0, {}_h s'_0)$ converges to the measure v_0 as $h \to 0$, then $({}_h y_t, {}_h s'_t) \Rightarrow (y_t, s'_t)$ as $h \to 0$.*

Proof. See Appendix C.1. ■

The proof of this theorem is a direct application of Theorem 2. Here $\left({}_h y_{kh} - {}_h y_{(k-1)h}\right)$, $\left({}_h s_{(k+1)h} - {}_h s_{kh}\right)$ and h are discrete correspondences of dy, ds, and dt, respectively. The Theorem 3.2 in Nelson (1990) shows that ${}_h Z_{kh}$ and

${}_hZ^*_{kh}$ are the discrete time counterpart of dW_1 and dW_2.

3.3.2 AR(1) Exponential *ARCH*

Here we consider a jump-diffusion process with conditional variance following an Ornstein-Uhlenbeck process. Those models in Wiggins (1987) and Nelson (1990) are variations of the process. Let's define a system of stochastic differential equations

$$d(\ln S_t) = \theta\sigma_t^2 dt + \sigma_t dW_{1,t} + k_t d\eta_t \tag{3.60}$$

$$d(\ln \sigma_t^2) = -\beta\left[(\ln\sigma_t^2) - \alpha\right]dt + dW_{2,t} \tag{3.61}$$

$$P\left[(\ln S_0, (\ln\sigma_0^2)) \in \Gamma\right] = \upsilon_0(\Gamma) \quad \text{for any } \Gamma \in B(R^2) \tag{3.62}$$

where S_t is the value of a stock at time t, $W_{1,t}$ and $W_{2,t}$ are Wiener processes with

$$\begin{bmatrix} dW_{1,t} \\ dW_{2,t} \end{bmatrix} \begin{bmatrix} dW_{1,t} & dW_{2,t} \end{bmatrix} = \begin{bmatrix} 1 & \Omega_{1,2} \\ \Omega_{1,2} & \Omega_{2,2} \end{bmatrix} dt \equiv \Omega dt \tag{3.63}$$

and $\Omega_{2,2} \geq \Omega_{1,2}^2$, and η_t is a Poisson process with parameter λ and

$$\begin{bmatrix} d\eta_t \\ 0 \end{bmatrix} \begin{bmatrix} d\eta_t & 0 \end{bmatrix} = \begin{bmatrix} \lambda & 0 \\ 0 & 0 \end{bmatrix} dt \equiv \Lambda dt. \tag{3.64}$$

Then the variance matrix of the system of equation is

$$\Sigma dt = \Omega dt + \Lambda dt. \tag{3.65}$$

We want to find a sequence of $ARCH$ models converging weakly to (3.60)-(3.65) by using Theorem 1. As we assume that $(\ln\sigma_t^2)$ follows a continuous time $AR(1)$ process in (3.61), $(\ln {}_h\sigma_{kh}^2)$ in the discrete counterpart of (3.61) will also follow an $AR(1)$ process. For each $h > 0$, we have

$$\begin{aligned}(\ln_h S_{kh}) &= \left(\ln_h S_{(k-1)h}\right) + h\theta\, {}_h\sigma^2_{kh} + {}_h\sigma_{kh}\, {}_hZ_{kh} \\ &\quad +\eta_{kh}\left(k_h + h^{-1/2}\phi_h\, {}_hZ_{kh}\right), \qquad (3.66)\\ \left(\ln_h \sigma^2_{(k+1)h}\right) &= (\ln_h \sigma^2_{kh}) - \beta\left[(\ln\, {}_h\sigma^2_{kh}) - \alpha\right] h + \Omega_{1,2}\, {}_hZ_{kh} \\ &\quad +\gamma\left[\left|{}_hZ_{kh}\right| - \left(\tfrac{2h}{\pi}\right)^{1/2}\right], \qquad (3.67)\\ P\left[(\ln S_0, (\ln \sigma_0^2)) \in \Gamma\right] &= v_0(\Gamma) \quad \text{for any } \Gamma \in B(R^2), \qquad (3.68)\end{aligned}$$

where $\gamma = \left[(\Omega_{2,2} - \Omega_{1,2}) / (1 - 2/\pi)\right]^{1/2}$ and ${}_hZ_{kh} \sim$iid $N(0,h)$. Then, we have

$$\begin{aligned} & E\begin{bmatrix} {}_hZ_{kh} + \eta_{kh} \\ \Omega_{1,2}\, {}_hZ_{kh} + \gamma\left[\left|{}_hZ_{kh}\right| - \left(\tfrac{2h}{\pi}\right)^{1/2}\right]\end{bmatrix} \\ & \times \begin{bmatrix} {}_hZ_{kh} + \eta_{kh} & \Omega_{1,2}\, {}_hZ_{kh} + \gamma\left[\left|{}_hZ_{kh}\right| - \left(\tfrac{2h}{\pi}\right)^{1/2}\right]\end{bmatrix} \\ = & \begin{bmatrix} 1+\lambda & \Omega_{1,2} \\ \Omega_{1,2} & \Omega_{2,2}\end{bmatrix} h \qquad (3.69) \\ \equiv & \Sigma h \end{aligned}$$

which is the discrete counterpart of (3.65). As before, we define a continuous time step function

$${}_hS_t \equiv {}_hS_{kh}, \quad {}_h\sigma^2_t \equiv {}_h\sigma^2_{kh} \quad \text{for } kh \le t < (k+1)h.$$

The discrete time counterparts of $d[\ln S]$, $d[\ln \sigma^2]$, $dW_{1,t}$, $dW_{2,t}$, $d\eta_t$ and dt are $(\ln_h S_{kh}) - \left(\ln_h S_{(k-1)h}\right)$, $\left(\ln_h \sigma^2_{(k+1)h}\right) - (\ln_h \sigma^2_{kh})$, ${}_hZ_{kh}$, η_{kh}, and h, respectively.

Theorem 4 *If the distribution of random starting point, v_h, converges to v_0 as $h \to 0$, then $\{{}_hS_t, {}_h\sigma^2_t\} \Rightarrow \{S_t, \sigma^2_t\}$ as $h \to 0$.*

Proof. See Appendix C.1. ■

3.4 Conclusions

In this chapter, we have shown that a stochastic difference equation converges weakly to a stochastic differential equation with jump component as length of sampling interval, h, goes to zero. We presented that, as an example, $GARCH\,(1,1)$-M process converges weakly to a jump-diffusion limit as h goes to zero. That is, a $ARCH$ type process can be approximated by stochastic jump-diffusion process. It is shown that $ARCH$ process is a discrete approximation of jump-diffusion process with using Exponential $ARCH$ process with Poisson jump component.

Therefore, we may use a discrete time $ARCH$ process as an approximation of a jump-diffusion process in estimation and forecasting. And we may use the jump-diffusion process as an approximation of $ARCH$ process when there is distributional results available for the jump-diffusion limit of the sequence of $ARCH$ processes.

While we show the weak convergence of those processes, we fixed the jump intensity, λ, as a constant. This may be extended to a more general case where the jump intensity varies over time with some probability structure.

Chapter 4

Filtering with Jump-Diffusion Processes

4.1 Introduction

In modern financial economics, it has been an important issue to model financial time series accurately. Since Press (1967), jump-diffusion processes which are the solution to the stochastic asset optimisation problem have been paid attention and have been used to model financial time series more widely [Cox and Ross (1975, 1976), Merton (1976a, b), and etc.].[1] During the last two decades or so, jump-diffusion processes play significant role in the financial economics literature, such as in the dynamic asset pricing literature [e.g., Oldfield, Rogalski, and Jarrow (1977), Ball and Torous (1983, 1985), Jarrow and Rosenfeld (1984), Amin (1993), Kim, Oh, and Brook (1994), Chang (1995)], in foreign exchange rates [e.g., Jorion (1988), Ball and Roma (1993), Park, Ahn and Fujihara (1993), Vlaar and Palm (1993), Jiang (1998)], in the term structure of interest rates [e.g., Ahn and Thompson (1988) and Das (1997)]. Merton (1976a, b), for example,

[1] See the Chaper 2 for more detail in usage of jump-diffusion processes in the financial economics literature.

decomposed the total changes in the stock prices into two components which were systematic risk and nonsystematic risk. The systematic risk can be modelled by a Brownian motion while the nonsystematic risk can be modelled by a Poisson jump process. This nonsystematic risk occurs mainly due to the arrivals of new important information to the market.

Since Engle (1982), the $ARCH$ models have been used widely to model the changes in conditional variances in the financial economic literature. There is no doubt that $ARCH$ models are the most useful method to model heteroskedastic financial time series data. As far as the reality is concerned, $ARCH$ models with jump components become a model describing the real economy more closely than $ARCH$ models alone. There have been few empirical researches on this issue [Feng and Smith (1997), Jorion (1989), and Vlaar and Palm (1993)]. These works found that there have been some improvement in explaining leptokurtic behaviour of financial time series with using $ARCH$ models with jump components.

In asset pricing theory, the expected return on the asset is estimated by using its variance and covariance. For example, Black and Scholes (1973), and Ross (1976) and many other empirical works have investigated the relationship between the risk and return on assets. As the literature has made clear, the variability of returns and the degree of co-movement between assets change stochastically over time. Then, many researchers have developed the asset pricing theory within the context of the conditional variances and the covariances of returns. We can find these works in, for example, Engle, Ng, and Rothschild (1990), Harvey (1989), and Merton (1973).

In this chapter, we study the properties of the conditional covariance estimates generated by misspecified $ARCH$ models with a jump process. According to Nelson (1992), if a process is well approximated by a diffusion, broad classes of $ARCH$ models provide consistent estimates of the conditional covariances. We show that a misspecified $ARCH$ with a jump process can still provide a consistent estimates of the conditional covariances. That is, the difference between a conditional covariance estimate and the true conditional covariance converges to zero in probability as the sampling interval of length h goes to zero.

We will briefly state the main frame of the chapter by defining the jump-diffusion data generating process and the consistent filter in the next section. In section 4.3, several assumptions are presented. Some of them are required to show the weak convergence of a discrete time process to a jump-diffusion process and others will be needed to show the convergence the conditional covariance estimate to the true conditional covariance in probability. Section 4.4 contains examples. We used $GARCH(1,1)$-M model and $AR(1)$ Exponential $ARCH$ model. In the last section, we briefly summarise and conclude the chapter.

4.2 Preliminary

We start from defining the $(n \times 1)$ jump-diffusion process $\{X_t\}$ by a stochastic integral equation

$$X_t = X_0 + \int_0^t b(X_s, s)\, ds + \int_0^t \sigma(X_s, s)\, dW_s + \int_0^t \int g(X_{s-}, s)\, \widetilde{N}_\lambda(ds), \quad (4.1)$$

where $\{W_t\}$ is an n-dimensional standard Brownian motion, independent of X_0, and $\widetilde{N}_\lambda(ds)$ is the compensated Poisson process defined as $\widetilde{N}_\lambda(ds) = N(ds) - \lambda ds$. And $b(X_t, t)$ is a continuous function from R^n to R^n, $\sigma(X_t, t)$ is continuous function from R^n into the space of $n \times n$ matrices and $g(X_{t-}, t)$ is a continuous function from R^n to R^n measuring the size of jump per unit of time.. Under some regularity conditions, $a(X_t, t) = \sigma(X_t, t)\, \sigma(X_t, t)'$ and $b(X_t, t)$ are, respectively, instantaneous conditional covariance matrix and instantaneous conditional mean per unit of time of increments in the $\{X_t\}$ process. X_0 is assumed random with probability measure v_0 and independent of $\{W_t\}_{0 \leq t < \infty}$.

Now, we assume that the econometricians are never able to observe some elements of $\{X_t\}$, but are able to observe others directly at discrete time intervals of length h only. We partition X_t as $\left[X'_{1:q,t}\ X'_{q+1:n,t}\right]'$, where $X'_{1:q,t}$ consists of the first q elements of X_t which are observable at time interval of length h, and $X_{q+1:n,t}$ consists of the last $n-q$ elements which are never observable. We partition b and σ accordingly.

For each t, we want to estimate $a_{1:q,1:q}\left(X_t, t\right)$, which is the instantaneous covariance of the increments in the observable variables $X_{1:q,t}$. The information given is very restricted such as the time index t and the history up to time t of the $\{X_t\}$ process and the smaller information set $\left\{t,\ X_{1:q,0},\ X_{1:q,h},\ \ldots,\ X_{1:q,h[t/h]}\right\}$, the past observed values of $X_{1:q,t}$. Under these assumptions, $a_{1:q,1:q}\left(X_t, t\right)$ is unobservable since $X_{q+1:n,t}$ is not observable and unless t is integer multiple of h, $X_{1:q,t}$ is as well. Therefore, $a_{1:q,1:q}\left(X_t, t\right)$ is a conditional covariance matrix, but is conditional on a larger information set than is possessed by the econometricians.

Now, we generate conditional covariance estimates $\{{}_h\widehat{a}_{1:q,1:q,t}\}$ with a sequence of $ARCH$ models with jump components, whose coefficient may depend on h. We say that a given sequence of conditional covariance estimates $\{{}_h\widehat{a}_{1:q,1:q,t}\}$ provides a consistent filter as $h \to 0$ if for each $t > 0$,

$$\text{p-}\lim_{h\to 0}\left[{}_h\widehat{a}_{1:q,1:q,t} - a_{1:q,1:q,t}\right] \to \mathbf{0}_{q\times q}, \quad \text{for all } t, \tag{4.2}$$

where $\mathbf{0}_{q\times q}$ is a $q \times q$ matrix of zeros. Thus the $ARCH$ models consistently estimates the true underlying conditional covariance as the length of sampling interval, h, goes to zero. Here the term 'estimate' is used the one in the filtering literature rather than the statistics literature.

4.3 Weak Convergence of Markov Processes to Jump-Diffusion Processes

Now, we will show the weak convergence of Markov processes to jump-diffusion processes in this section. We begin by defining the sequence of processes which we divide into two pieces: $\{_hX_t\}$ and $\{_hY_t\}$, $n \times 1$ and $m \times 1$ vector, respectively. $\{_hX_t\}$ and $\{_hY_t\}$ will be random step functions taking jumps at times h, $2h$, $3h$, and so on. With some sufficient conditions given below, we will show that $\{_hX_t\}$ converges weakly to $\{X_t\}$, where $\{X_t\}$ is a solution to (4.1). Here, $\{_hX_t\}$ represents the underlying stochastic system generating data, while $\{_hY_t\}$ represents the difference between the true conditional covariance matrix of $\{_hX_t\}$ and the estimate generated by an $ARCH$ model with jump component. Now, we will present some conditions that ensure the weak convergence of a Markov process to a jump-diffusion process and, for every $t > 0$, the convergence of $_hY_t$ to $m \times 1$ vector of zeros as $h \to 0$. That is, we will show that for $h > 0$, a Markov process converges to a jump-diffusion and the conditional covariance estimator is consistent as $h \to 0$.

The formal setup is following:

Let $D([0,\infty), R^n \times R^m)$ be the space of mappings from $[0,\infty)$ into $R^n \times R^m$ that are right continuous having finite left limits and let $B(R^n \times R^m)$ denote the Borel sets on $R^n \times R^m$. With introducing an appropriate Skorohod metric, $D([0,\infty), R^n \times R^m)$ becomes a complete metric space.[2] For each $h > 0$, let $\mathfrak{M}_{kh}$ be the σ-algebra generated by $_hX_0$, $_hX_h$, $_hX_{2h}$, $\ldots$, $_hX_{kh}$, and $_hY_0$, $_hY_h$, $_hY_{2h}$, $\ldots$, $_hY_{kh}$, and let ν_h be a probability measure on $(R^n, B(R^n \times R^m))$. For each $h > 0$, and each $k = 0, 1, 2, \ldots$, let $\Pi_{h,kh}(x, y, \cdot)$ be a transition function on $R^n \times R^m$. That is,

i) $\Pi_{h,kh}(x, y, \cdot)$ is a probability measure on $(R^n \times R^m, B(R^n \times R^m))$ for all $x \in$

[2]See [2] Kunhner (1984) Section 4.3 in Chapter 2.

R^n,

ii) $\Pi_{h,kh}(\cdot,\cdot,\Gamma)$ is $B(R^n \times R^m)$ measurable for all $\Gamma \in B(R^n \times R^m)$.

For each $h > 0$, let P_h be the probability measure on $D([0,\infty), R^n \times R^m)$ such that

$$P_h[({}_hX_0, {}_hY_0) \in \Gamma] = \nu_h(\Gamma) \text{ for any } \Gamma \in B(R^n \times R^m), \tag{4.3}$$

$$P_h[({}_hX_t, {}_hY_t) = ({}_hX_{kh}, {}_hY_{kh}),\ kh \le t < (k+1)h] = 1, \tag{4.4}$$

$$\begin{array}{l} P_h\left[\left({}_hX_{(k+1)h}, {}_hY_{(k+1)h}\right) \in \Gamma \mid \mathfrak{M}_{kh}\right] = \Pi_{h,kh}({}_hX_{kh}, {}_hY_{kh}, \Gamma) \\ \text{almost surely under } P_h \text{ for all } k \ge 0 \text{ and } \Gamma \in B(R^n \times R^m). \end{array} \tag{4.5}$$

Here, for each $h > 0$, we specify the distribution of the random starting point $({}_hX_0,\ {}_hY_0)$ by (4.3) and form a continuous time process $({}_hX_t,\ {}_hY_t)$ from the discrete time process $({}_hX_{kh},\ {}_hY_{kh})$ by (4.4) making $\{{}_hX_t,\ {}_hY_t\}$ a step function with jumps at time h, $2h$, $3h$, and so on. (4.5) specifies the transition probabilities of $(n+m)$-dimensional discrete time Markov process $\{{}_hX_{kh},\ {}_hY_{kh}\}$.

To ensure that there is no feed back from $\{{}_hY_{kh}\}$ to $\{{}_hX_{kh}\}$, we may need the following condition. For every Borel set Γ_x of R^n and for all $h > 0$,

$$\Pi_h(x, y, \Gamma_x \times R^m) \text{ is independent of } y. \tag{4.6}$$

Now, define, for each $h > 0$,

$$\begin{aligned} a_h(x,t) &\equiv h^{-1}E\left[\left({}_hX_{(k+1)h} - {}_hX_{kh}\right)\right. \\ &\qquad \times \left.\left({}_hX_{(k+1)h-h} - {}_hX_{kh}\right)'\right|\, {}_hX_{kh} = x, {}_hY_{kh} = y\Big], \end{aligned} \tag{4.7}$$

$$b_h(x,t) \equiv h^{-1}E\left[{}_hX_{(k+1)h} - {}_hX_{kh}\right|\, {}_hX_{kh} = x, {}_hY_{kh} = y\Big], \tag{4.8}$$

$$\Delta_h^{\varepsilon}(x,t) \equiv h^{-1}P_h\left[\left|\, {}_hX_{(k+1)h} - {}_hX_{kh}\right| > \varepsilon \,\middle|\, {}_hX_{kh} = x, {}_hY_{kh} = y\right], \tag{4.9}$$

$$g_h(x,t) \quad = \quad x_t - x_{t-}, \quad x_{t-} = \lim_{s \to t} x_s \quad \text{for } s < t \tag{4.10}$$

where the expectations in (4.7) and (4.8) are taken under P_h. $a_h(x,t)$ is a measure of the second moment per unit time, and $b_h(x,t)$ is a measure of drift per unit time, $\Delta_h^{\varepsilon}(x,t)$ is a probability that the process has a jump of size greater than ε, and $g_h(x,t)$ is a magnitude of jump, if jump occurs. $a_h(x,t)$, $b_h(x,t)$, and $\Delta_h^{\varepsilon}(x,t)$ are independent of y by (4.6).

Now, we need several assumption to achieve the weak convergence results.

Assumption 1. We assume that for all $R > 0$, $T > 0$ and $\varepsilon > 0$,

$$\lim_{h \downarrow 0} \sup_{\|x\| \leq R,\ 0 \leq t \leq T} \|a_h(x,t) - a(x,t)\| \quad = \quad 0 \tag{4.11}$$

$$\lim_{h \downarrow 0} \sup_{\|x\| \leq R,\ 0 \leq t \leq T} \|b_h(x,t) - b(x,t)\| \quad = \quad 0 \tag{4.12}$$

$$\lim_{h \downarrow 0} \sup_{\|x\| \leq R,\ 0 \leq t \leq T} \|g_h(x,t) - g(x,t)\| \quad = \quad 0 \tag{4.13}$$

$$\lim_{h \downarrow 0} \sup_{\|x\| \leq R,\ 0 \leq t \leq T} \Delta_h^{\varepsilon}(x,t) \quad = \quad \lambda \tag{4.14}$$

This assumption requires that the second moment, drift, and jumps per unit of time converge uniformly on compact sets to well-behaved functions of time and the state variables x. And the probability of jump of size greater then ε converges to a constant λ. So, the sample paths of the limit process will have only discontinuity of the first kind with probability one.

Assumption 2. For every $R \geq 0$ and all $i = 1, 2, \ldots, n$,

$$\lim_{h \to 0} \sup_{\|x\| \leq R} h^{-1} E\left[\left({}_hX_{i,(k+1)h} - {}_hX_{i,kh}\right)^4 |{}_hX_{kh} = x\right] = 0 \tag{4.15}$$

Assumption 3. As $h \to 0$, $({}_hX_0, {}_hY_0)$ converges in distribution to a random variable (X_0, Y_0) with probability measure ν_0 on $(R^n, B(R^n \times R^m))$.

This assumption requires that the probability measure ν_h of the random starting points ${}_hX_0$ to converge to a limit measure ν_0 as $h \to 0$.

With all the assumptions we made above, we specified a initial probability measure ν_0 of the limit process, an instantaneous drift function $b(x,t)$, an instantaneous covariance matrix $a(x,t)$, and a jump amplitude $g(x,t)$. We have supposed that the sample path of the process is discontinuous with probability one. However, there is no guarantee that a limit process is finite or is uniquely defined. There are a number of works considering the conditions under which ν_0, $a(x,t)$, and $b(x,t)$ uniquely define a diffusion limit process. Especially Strook and Varadhan (1979) studied extensively about the diffusion limit process. Either and Kurtz (1986) considers the martingale problems with Levy measure. Gihman and Skorohod (1972) gives the conditions of the unique existence of a solution to a jump-diffusion limit. The non-explosion condition for jump-diffusion limit will be stated in Appendix B.

Assumption 4. ν_0, $a(x,t)$, $b(x,t)$, and $g(x,t)$ uniquely specify the distribution of a jump diffusion process X_t with initial distribution ν_0, diffusion matrix $a(x,t)$, drift vector $b(x,t)$, and jump amplitude $g(x,t)$.

Theorem 5 *Under Assumptions 1-4,* $\{{}_hX_t\} \Rightarrow \{X_t\}$ *as* $h \to 0$.

Proof. See the proof of Theorem 1 in Appendix C.1. ■

Now, we need to some assumptions to guarantee the convergence of $\{{}_hY_t\}$ to zero as $h \to 0$. Let the following conditional expectations be well-defined:

$$c_{h,\delta}(x,y) \equiv h^{-\delta} E\left[\left({}_hY_{(k+1)h} - {}_hY_{kh}\right) \mid {}_hX_{kh} = x, {}_hY_{kh} = y\right], \quad (4.16)$$

$$d_{h,\delta}(x,y) \equiv h^{-\delta}E\left[\left({}_hY_{(k+1)h} - {}_hY_{kh}\right) \times \left({}_hY_{(k+1)h} - {}_hY_{kh}\right)' \mid {}_hX_{kh} = x, {}_hY_{kh} = y\right]. \quad (4.17)$$

Assumption 5. For some δ, $0 < \delta < 1$, and for every $R > 0$,

$$\lim_{h \to 0} \sup_{\|(x,y)\| \leq R} \|c_{h,\delta}(x,y) - c(x,y)\| = 0 \quad (4.18)$$

where for all $x \in R^n$, $c(x,0) = 0$ and

$$\lim_{h \to 0} \sup_{\|(x,y)\| \leq R} \|d_{h,\delta}(x,y)\| = 0 \quad (4.19)$$

We defined the first and second moments of the increments in $\{{}_hX_{kh}\}$ as $O(h)$. In Assumption 5, we defined the drift and second moment of the increments in $\{{}_hY_{kh}\}$ are $O(h^\delta)$ and $o(h^\delta)$, respectively. This implies that $\{{}_hY_t\}$ operates on a faster time scale than $\{{}_hX_t\}$, since the drift per unit of time of $\{{}_hY_t\}$ grow at a faster rate as $h \to 0$ than the drift and variance per unit of time of $\{{}_hX_t\}$. Therefore, if $\{{}_hY_t\}$ is mean-reverting to a vector of zeros, the drift converges to zero as well with the increasing speed as $h \to 0$. Another implication is that, as $h \to 0$, the drift of $\{{}_hY_t\}$ dominates the variance of $\{{}_hY_t\}$. This allows us to approximate the behaviour of $\{{}_hY_t\}$ by a deterministic differential equation.

Assumption 6. For each $x \in R^n$, $y \in R^m$, define the differential equation

$$dY(t,x,y)/dt = c(x, Y(t,x,y)), \quad (4.20)$$

with initial condition

$$Y(0,x,y) = y \quad (4.21)$$

Then $0_{m\times 1}$ is a globally asymptotically stable solution of (4.20) and (4.21) for bounded values of x, y. That is, for every $R > 0$,

$$\lim_{h\to 0} \sup_{\|(x,y)\|\leq R} \|Y(t,x,y)\| = 0 \tag{4.22}$$

This Assumption 6 ensure that the differential equation approximating the behaviour of $\{_hY_t\}$ is well behaved, pulling $\{_hY_t\}$ back to a vector of zeros. Now, we need a condition that $\{_hY_t\}$ does not diverge to infinity in finite time.

Assumption 7. There exist a nonnegative function $\rho(x,y,h)$, twice differentiable in x and y, and a positive function $\lambda(x,h)$ such that

$$\lim_{R\to\infty} \liminf_{h\to 0} \inf_{\|(x,y)\|\geq R} \rho(x,y,h) = \infty, \tag{4.23}$$

$$\limsup_{R\to\infty} \limsup_{h\to 0} M(R,h) < \infty, \tag{4.24}$$

$$\limsup_{h\to 0} E\left[\rho\left({}_hX_0, {}_hY_0, h\right)\right] < \infty, \tag{4.25}$$

and for every $R > 0$,

$$\limsup_{h\to 0} \sup_{\|(x,y)\|\leq R} h^{-1}E\left[\rho\left({}_hX_{(k+1)h}, {}_hY_{(k+1)h}, h\right)\right.$$

$$\left.-\rho(x,y,h) \,|\, {}_hX_{kh} = x, {}_hY_{kh} = y\right] - M(R,h)\,\rho(x,y,h) \leq 0 \tag{4.26}$$

Theorem 6 *(Ethier-Nagylaki) Let Assumptions 1-7 hold. Then*

$${}_hY_t \to \mathbf{0}_{m\times 1} \text{ in probability as } h\to 0 \text{ for every } t > 0 \tag{4.27}$$

Proof. See Ethier and Nagylaki (1988) Theorem 2.1. ■

4.4 Examples of Consistent $ARCH$ Filters

In this section, we will show $ARCH$ models can produce a consistent filter when (4.1) generates $\{_hX_t\}$. We will use $GARCH\,(1,1)$-M, and $AR\,(1)$ $EGARCH$ models as examples. For simplicity, we will deal with the univariate cases. The extension to a multivariate case wouldn't be a difficult task.

4.4.1 $GARCH\,(1,1)$-M Model

In Engle and Bollerslev (1986), they presented the $GARCH\,(1,1)$-M process for the cumulative excess returns X_t on a portfolio. If we suppose that the process includes jump components, then the model can be rewritten as follows:

$$X_t = X_{t-1} + \mu\sigma_t^2 + \sigma_t Z_t + c\eta_t, \tag{4.28}$$

$$\sigma_{t+1}^2 = w + \sigma_t^2\left(\beta + \alpha Z_t^2\right) \tag{4.29}$$

where $Z_t \sim i.i.d.\ N(0,1)$ and η_t ~Bernoulli distributed with $\Pr(\eta_t = 0) = 1 - \lambda dt + o(dt)$ and $\Pr(\eta_t = 1) = \lambda dt + o(dt)$. Here c denotes the jumps size of the process when a jump occurs. Now, if we partition the time interval more finely, then we may rewrite the system of difference equations as follows:

$$\begin{aligned} {}_hX_{kh} &= {}_hX_{(k-1)h} + h\cdot\mu_h\left({}_hX_{kh},\, {}_h\sigma_{kh}^2\right) + {}_h\sigma_{kh}\, {}_hZ_{kh} \\ &\quad + {}_h\eta_{kh}\left(c_h + h^{-1/2}v_h\, {}_hZ_{kh}\right), \end{aligned} \tag{4.30}$$

$${}_h\sigma_{(k+1)h}^2 = w_h + {}_h\sigma_{kh}^2\left(\beta_h + h^{-1}\alpha_h\, {}_hZ_{kh}^2\right), \tag{4.31}$$

and

$$\Pr\left[\left({}_hX_0,\, {}_h\sigma_0^2\right) \in \Gamma\right] = \upsilon_h(\Gamma) \quad \text{for any } \Gamma \in B\left(R^2\right) \tag{4.32}$$

where ${}_hZ_{kh} \sim i.i.d.\ N(0,h)$ and η_t ~Bernoulli distributed with $\Pr\left[{}_h\eta_{kh}=0\right] = 1-\lambda h+o(h)$, and $\Pr\left[{}_h\eta_{kh}=1\right] = \lambda h+o(h)$. υ_h satisfies Assumption 3 as $h \to 0$, and for each $h \geq 0$, $\upsilon_h\left((X_0,\sigma_0^2) : \sigma_0^2 > 0\right) = 1$.

Then the conditional covariance estimate can be obtained by

$$ {}_h\widehat{\sigma}^2_{(k+1)h} = w_h + {}_h\widehat{\sigma}^2_{kh}\left(\beta_h + h^{-1}\alpha_h\, {}_h\widehat{Z}^2_{kh}\right). \tag{4.33}$$

The $\left\{{}_h\widehat{Z}_{kh}\right\}$ in (4.33) are fitted residuals obtained by using the drift $\widehat{\mu}_h\left({}_hX_{kh},\, {}_h\sigma^2_{kh}\right)$:

$$ {}_h\widehat{Z}_{kh} = {}_hX_{(k+1)h} - {}_hX_{kh} - h\,\widehat{\mu}_h\left({}_hX_{kh},\, {}_h\widehat{\sigma}^2_{kh}\right). \tag{4.34}$$

For some $\delta > 0$, $0 < \delta < 1$, let

$$ \widehat{\mu}_h = o\left(h^{-1/2}\right), \tag{4.35}$$

$$ w_h = o\left(h^{\delta}\right), \tag{4.36}$$

$$ 1 - \beta_h - \alpha_h = o\left(h^{\delta}\right), \tag{4.37}$$

$$ \alpha_h = h^{\delta}\alpha + o\left(h^{\delta}\right), \tag{4.38}$$

where α is independent of h. Now, we define

$$ {}_hY_t \equiv {}_h\widehat{\sigma}^2_t - {}_h\sigma^2_t \tag{4.39}$$

This ${}_hY_t$ is the estimate error at time t. And

$$ h^{-1}E\left[\, {}_hY_{(k+1)h} - {}_hY_{kh}\middle|\, {}_hX_{kh} = x,\ {}_hY_{kh} = y\right] = -h^{\delta-1}\alpha y + O(1) \tag{4.40}$$

In this equation, we can observe that $\{{}_hY_t\}$ is mean-reverting process. As $h \to 0$, the difference between ${}_hY_{(k+1)h}$ and ${}_hY_{kh}$ converges to zero, and the speed of the convergence goes to infinity. The consistency in the filter can be achieved in this way.

Now we need several conditions on the jump-diffusion (4.1), on the other parameters, and the initial value of ${}_h\widehat{\sigma}_t^2$ guaranteeing Assumptions 1-7 hold. Then we will be able to apply the Theorem 2 in Chapter 3 to prove the measurement error, ${}_hY_t$, converges to zero in probability for each $t > 0$ as $h \to 0$.

Condition 1. For each $h > 0$, (4.1) generates $\{{}_hX_t\}$, and satisfies Assumptions 2-4 in section 3.

Condition 2. For some $\varepsilon > 0$, $\limsup_{h\to 0} E\left\|{}_hY_0\right\|^{2+\varepsilon} < \infty$.

Condition 3. There is a twice differentiable, nonnegative $\omega(x)$ and a $\theta > 0$ such that for every $R > 0$,

$$\liminf_{\|x\|\to\infty} \omega(x) = \infty \tag{4.41}$$

$$\lim_{h\to 0} E\left[\omega\left({}_hX_0\right)\right] < \infty \tag{4.42}$$

$$\lim_{h\to 0}\inf_{\|x\|\le R} h^{-1}E\left[\left|\omega\left({}_hX_{(k+1)h}\right) - \omega(x)\right|^{1+\theta} \mid {}_hX_{kh} = x\right] < \infty \tag{4.43}$$

and there is a $M > 0$ for all $x \in R^n$,

$$\begin{aligned} &\sum_{i=1}^{n} b_i(x)\frac{\partial\omega(x)}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^{n} a_{ij}(x)\frac{\partial^2\omega(x)}{\partial x_i\partial x_j} \\ &+\int\left[f(x+g(x,t)) - f(x) - \sum_{i=1}^{n} g_i(x,t)\frac{\partial f}{\partial x_i}\right]N(h) \\ \le\ & M\omega(x). \end{aligned} \tag{4.44}$$

Condition 4. For every $R > 0$, there is an $\varepsilon > 0$ such that

$$\lim_{h\to 0}\sup_{\|x\|\le R} h^{-1}E\left[\left|{}_h\sigma^2_{(k+1)h} - {}_h\sigma^2_{kh}\right|^{2+\varepsilon} \mid {}_hX_{kh} = x\right] = 0, \tag{4.45}$$

$$\lim_{h\to 0} \sup_{\|x\| \quad R} h^{-1} E\left[\left|{}_hX_{(k+1)h} - x\right|^{4+\varepsilon} \mid {}_hX_{kh} = x\right] = 0. \tag{4.46}$$

Condition 5. w_h, α_h, and β_h satisfy (4.35)-(4.37). And

$$\lim_{h\to 0} h^{-1}\alpha_h^2 > 0.$$

Theorem 7 *Let Conditions 1-5 hold. Then for each $t > 0$, $\|{}_hY_t\| \to 0$ in probability as $h \to 0$.*

Proof. See Appendix C.2. ∎

4.4.2 $AR(1)$ Exponential $ARCH$ Model

In Nelson (1991), the Exponential $ARCH$ process is introduced. We can write the $AR(1)$ Exponential $ARCH$ process as follows:

$$\begin{aligned} {}_hX_{kh} &= {}_hX_{(k-1)h} + h\,\mu_h\left({}_hX_{kh},\, {}_h\sigma_{kh}^2\right) + {}_h\sigma_{kh}\, {}_hZ_{kh} \\ &\quad + \eta_{kh}\left(k_h + h^{-1/2} v_h\, {}_hZ_{kh}\right), \qquad (4.47) \\ \ln\left({}_h\sigma_{(k+1)h}^2\right) &= \ln\left({}_h\sigma_{kh}^2\right) - \beta_h\, h\left[\ln\left({}_h\sigma_{kh}^2\right) - \alpha_h\right] + \theta_h\, {}_hZ_{kh} \\ &\quad + \gamma_h\left[\left|{}_hZ_{kh}\right| - \left(\frac{2h}{\pi}\right)^{1/2}\right], \qquad (4.48) \\ {}_hZ_{kh} &\sim i.i.d.\ N(0,h). \qquad (4.49) \end{aligned}$$

According to Theorem 4 in Chapter 3, if $\mu_h \to \mu$ uniformly on compact sets,

$$\begin{aligned} \lim\nolimits_{h\to 0} \beta_h &= \beta, \\ \lim\nolimits_{h\to 0} \alpha_h &= \alpha, \\ \lim\nolimits_{h\to 0} \theta_h &= \theta, \\ \lim\nolimits_{h\to 0} \gamma_h &= \gamma, \end{aligned} \tag{4.50}$$

if the initial values of the processes $({}_hX_0,\ \ln({}_h\sigma_0^2))$ converges in probability, and if μ_h satisfies some mild regularity condition, then $\{{}_hX_t,\ {}_h\sigma_t^2\}$ generated by (4.47)-(4.50) and (4.4) converges weakly to the jump-diffusion as $h \to 0$:

$$dX_t = \mu\left(X_t, \sigma_t^2\right) dt + \sigma_t dW_{1,t} + k_t d\eta_t, \tag{4.51}$$

$$d\ln\left(\sigma_t^2\right) = -\beta\left[\ln\left(\sigma_t^2\right) - \alpha\right] dt + dW_{2,t}, \tag{4.52}$$

where $W_{1,t}$ and $W_{2,t}$ are Wiener processes, η_t is a Poisson process with intensity λ, and $W_{i,t}$ and η_t are independent for $i = 1, 2$. Then

$$\begin{bmatrix} dW_{1,t} + d\eta_t \\ dW_{2t} \end{bmatrix} \begin{bmatrix} dW_{1,t} + \eta_t & dW_{2,t} \end{bmatrix}$$

$$= \begin{bmatrix} 1+\lambda & \theta \\ \theta & \theta^2 + \gamma^2\left[1 - \frac{2}{\pi}\right] \end{bmatrix} dt \tag{4.53}$$

We suppose that the data are generated by (4.1) and that a misspecified $EGARCH$ model is used to produce an estimate of the true underlying conditional variance process $\{\sigma_t^2\}$. We generate fitted conditional variance $\{\widehat{\sigma}_{kh}^2\}$ recursively by (4.48) with some initial value ${}_h\widehat{\sigma}_0^2$, and ${}_h\widehat{Z}_{kh}$ is generated by

$${}_h\widehat{Z}_{kh} \equiv \frac{\left[{}_kX_{kh} - {}_hX_{(k-1)h} - h \cdot \widehat{\mu}_h\left({}_hX_{kh;h}\,\widehat{\sigma}_{kh}^2\right)\right]}{{}_h\widehat{\sigma}_{kh}}. \tag{4.54}$$

We require that for some δ, $0 < \delta < 1$,

$$\widehat{\mu}_h = o\left(h^{-1/2}\right), \tag{4.55}$$

$$\beta_h = o\left(h^{\delta-1}\right), \tag{4.56}$$

$$\alpha_h\beta_h = o\left(h^{\delta-1}\right), \tag{4.57}$$

$$\theta_h = o\left(h^{(\delta-1)/2}\right), \tag{4.58}$$

$$\gamma_h = \gamma h^{\delta-1/2} + o\left(h^{\delta-1/2}\right), \quad \text{where } \gamma > 0. \tag{4.59}$$

Define a measurement error

$${}_hY_t \equiv \left[\ln\left({}_h\widehat{\sigma}_t^2\right) - \ln\left({}_h\sigma_t^2\right)\right], \tag{4.60}$$

then

$$\begin{aligned} c(x,y) &\equiv \lim_{h\to 0} h^{-\delta} E\left[{}_hY_{(k+1)h} - {}_hY_{kh}\,|{}_hX_{kh} = x,\ {}_hY_{kh} = y\right] \\ &= \gamma\left(\frac{2}{\pi}\right)\left[\exp\left(-y/2\right) - 1\right]. \end{aligned} \tag{4.61}$$

The differential equation of Assumption 6 is

$$\frac{dY}{dt} = \gamma\left(\frac{2}{\pi}\right)\left[\exp\left(-Y/2\right) - 1\right] \tag{4.62}$$

If we replace Condition 4 with the following condition, then we can achieve the convergence of ${}_hY_t$ to zero as $h \to 0$.

Condition 6. For every $R > 0$, there is an ε such that

$$\lim_{h\to 0} \sup_{\|x\|\le R} h^{-1} E\left[\left|\ln\left({}_h\sigma_{(k+1)h}^2\right) - \ln\left({}_h\sigma_{kh}^2\right)\right|^{2+\varepsilon} |{}_hX_{kh} = x\right] = 0. \tag{4.63}$$

Theorem 8 *Let* (4.55)-(4.63) *and Conditions 1-3 and 6 hold. Then* ${}_hY_t \to 0$ *in probability for every* $t > 0$ *as* $h \to 0$.

Proof. See Appendix C.2. ∎

4.5 Conclusion

It has become most important to identify the structure of conditional covariance correctly in modern financial economics. With *ARCH* type models, it has been greatly advantageous to model the heteroskedastic conditional covariance in a natural way. This may be the reason why so many econometricians have used $ARCH$ type models in modeling financial time series.

Nelson(1992) stated that if $\{X_t\}$ is a jump-diffusion, the consistency of the conditional covariance estimates generated by misspecified $ARCH$ models breaks down. However, by introducing the jump process into the $ARCH$ framework, we have shown that a misspecified $ARCH$ model with jumps can still produce consistent conditional covariance estimates. That is, we have shown that the conditional covariance estimates from misspecified $ARCH$ models with jump components converge to the conditional covariance of true data generating process as the length of sampling interval, h, goes to zero.

In the next chapter, we will investigate the forecasting abilities of the $ARCH$ models with jump components when the data are generated by the jump-diffusion process.

Chapter 5

Forecasting with Jump-Diffusion Processes

5.1 Introduction

In the two previous chapters, we have shown that an $ARCH$ type model with jumps is an approximation of a jump-diffusion limit and that an $ARCH$ type model with jumps estimates consistently conditional covariances of the true data generating process. As most market participants want to minimise their risk existing in the future, the forecasting ability of the model is also an important aspect. Now, it seems to be a natural stage to investigate the forecasting ability of the $ARCH$ with jumps.

Many researchers have found that financial time series are non-linear and volatile in nature. Since their introduction by Engle (1982), $ARCH$ type models have most popularly used in modeling financial time series. Among many other characteristics, the non-linearity of $ARCH$ type models makes themselves very successful candidates in modeling non-linear financial time series. These models also assist researchers to expand the ability of modeling heteroskedastic financial time series.

Recently, there have been a few works to test the predictability of $ARCH$ type models with financial time series [for example, Brooks and Burke (1998), Campa and Chang (1998), Chong, Ahmad and Abdullah (1999), Franses and Dijk (1996), Lamoureux and Lastrapes (1993), Mills (1996), and West and Cho (1995)]. Nelson (1992) suggested that the success of $ARCH$ type models is their ability to estimate consistent conditional covariances of the true data generating process. Later, Nelson and Foster (1995) developed the conditions under which a sequence of misspecified $ARCH$ type models generate consistent forecasts as well. I.e., as the length of sampling interval approaches to zero, the sequence of $ARCH$ type models not only produces consistent conditional covariance estimates, but also generates consistent forecasts of the true data generating process.

In this chapter, we investigate the forecasting ability of $ARCH$ type models with jumps as West and Cho (1995) suggested that including discrete jumps may improve the predictability of $ARCH$ type models. There have been a number of researches conducted on jump-diffusion processes during the last two decades or so. Since Cox and Ross (1976) and Merton (1976 a, b), many researchers applied jump-diffusion processes to model financial time series. Most empirical works found that financial time series contain discrete jumps caused by arrival of shocks at financial markets. They also found that $ARCH$ type models along with jump components may explain the real economy more successfully than $ARCH$ alone. However, there seem to be few researches investigating the issue of predictability of jump-diffusion processes yet.

We develop the conditions under which the forecasts generated by a sequence of misspecified $ARCH$ type models with jumps converge to the forecasts generated by the true underlying process as the length of sampling intervals approaches to zero. In the next section, we show that forecasts generated by a misspecified $ARCH$ model with jumps converge weakly to forecasts generated by the true data-generating process. In the earlier two chapters, we show that a jump-diffusion process is a continuous approximation of a discrete $ARCH$ model with jump component and that an $ARCH$ with jump component can perform filtering well. In section 5.3, with an $EGARCH$ model, we will show that a sequence of misspecified models can consistently estimate the conditional covariance matrix and produce reasonable forecasts for the true data-generating process as the length of sampling intervals become zero. Then, a summary and concluding remarks will be given in section 5.4.

5.2 Main Setup

In this section we develop the conditions under which forecasts generated by a sequence of misspecified $ARCH$ type models with jumps converge weakly to a forecast generated by the true data-generating process as the length of sampling intervals goes to zero.

For each $h > 0$, consider a pair of stochastic processes $\{{}_hX_t,\ {}_hU_t\}$. We suppose that $\{{}_hX_t\}$ is an $n \times 1$ directly observable process, and $\{{}_hU_t\}$ is an $m \times 1$ unobservable process. These processes are step functions with jumps only at times h, $2h$, $3h$, and so on. Here $\{{}_hX_t\}$, which is generated by, for example, a stochastic volatility model, is not Markovian. With the introduction of $\{{}_hU_t\}$, which will control the conditional covariance of the increments in the $\{{}_hX_t\}$ process, the pair $\{{}_hX_t,\ {}_hU_t\}$ becomes Markovian. We will, therefore, assume that the pair $\{{}_hX_t,\ {}_hU_t\}$ is Markovian with probability measure P_h. If an $ARCH$ model recursively defines $\left\{{}_h\hat{U}_t\right\}$ in a way that it is a function of ${}_hX_t$, ${}_hX_{t-h}$, and ${}_h\hat{U}_{t-h}$, then the $ARCH$ model assigns a probability measure $\hat{P}_h$ to $\left\{{}_hX_t,\ {}_h\hat{U}_t\right\}$ so that the pair $\left\{{}_hX_t,\ {}_h\hat{U}_t\right\}$ becomes Markovian. $\left\{{}_h\hat{U}_t\right\}$ is also assumed to have jumps at times h, $2h$, $3h$, ... The variable ${}_h\hat{U}_t$ can be computed from the previous values of ${}_hX_t$ and the initial value ${}_h\hat{U}_0$. So, $\{{}_hX_t,\ {}_hU_t\}$ can be considered as the true Markov process, and $\left\{{}_hX_t,\ {}_h\hat{U}_t\right\}$ is an approximation.

The main issue of this chapter is how closely the $ARCH$ approximation can forecast the true process. That is, how close the forecasts generated under the probability measure, $\hat{P}_h$, are to the forecasts generated under the probability measure, P_h. To compare the resulting forecasts, we need to bring those two processes into the same probability space. If we allow that ${}_hU_t = {}_h\hat{U}_t$ a.s. under $\hat{P}_h$, then $\left\{{}_hX_t,\ {}_hU_t,\ {}_h\hat{U}_t\right\}$ is Markovian under $\hat{P}_h$. As ${}_h\hat{U}_t$ is a function of ${}_hX_t$, ${}_hX_{t-h}$, and ${}_h\hat{U}_{t-h}$, $\left\{{}_hX_t,\ {}_hU_t,\ {}_h\hat{U}_t\right\}$ is already Markovian under P_h.[1]

Now we will present some conditions that ensure the weak convergence of a Markov process to a jump-diffusion process, and, for every τ, $0 < \tau < \infty$, the convergence of ${}_h\hat{U}_\tau - {}_hU_\tau$ to a $m \times 1$ vector of zeros as the length of sampling interval h goes to zero. In addition to that, we will investigate the conditions under which, for every τ, $0 < \tau < \infty$, $\left\{{}_hX_t,\ {}_h\hat{U}_t\right\}$ consistently estimates the forecast distribution of $\{{}_hX_t,\ {}_hU_t\}$ as $h \to 0$.

Let $D\left([0,\infty), R^n \times R^{2m}\right)$ be a space of mappings from $[0,\infty)$ into $R^n \times R^{2m}$ that are continuous from right with finite left limits. With the introduction of Skorohod metric,[2] the space $D\left([0,\infty), R^n \times R^{2m}\right)$ becomes a complete metric space. Let $\mathfrak{B}(E)$ be a Borel set of a metric space E. Let $\mathfrak{M}_{kh}^{X,U}$ be the σ-field generated by ${}_hX_0,\ {}_hX_h,\ {}_hX_{2h},\ \ldots$, and ${}_hU_0,\ {}_hU_h,\ {}_hU_{2h},\ \ldots$, and let $\mathfrak{M}_{kh}^{X,\hat{U}}$ be the σ-field generated by ${}_hX_0,\ {}_hX_h,\ {}_hX_{2h},\ \ldots$, and ${}_h\hat{U}_0$. Then we denote $\mathfrak{M}_{kh} \equiv \mathfrak{M}_{kh}^{X,U} \cup \mathfrak{M}_{kh}^{X,\hat{U}}$, so that $\mathfrak{M}_{kh}$ is the natural σ-field for $\left\{{}_hX_t,\ {}_hU_t,\ {}_h\hat{U}_0\right\}$.[3] Let v_h and $\hat{v}_h$ be probability measures on $\left(R^{n+2m}, \mathfrak{B}\left(R^n \times R^{2m}\right)\right)$.[4] For each $h > 0$, let $\Pi_h(x, u, \hat{u}, .)$ and $\hat{\Pi}_h(x, u, \hat{u}, .)$ be transition functions on R^{n+2m} such that for

[1]However, we need to note that those probability measures, P_h and $\hat{P}_h$, are mutually singular, since $\hat{P}_h\left[{}_hU_t = {}_h\hat{U}_t, \forall t\right] = 1$ and, in general, $P_h\left[{}_hU_t = {}_h\hat{U}_t, \forall t\right] = 0$. See Billingsley (1986) for the mutually singular measures.

[2]For formal definitions, see Ethier and Kurtz(1986), Ch. 3.

[3]The variable ${}_h\hat{U}_t$ can be computed from the previous values of ${}_hX_t$ and a starting value ${}_h\hat{U}_0$, so ${}_h\hat{U}_t$ is measurable with respect to the natural σ-field of the ${}_hX_t$ and ${}_h\hat{U}_0$. Thus the natural σ-field for all three processes is the union of the natural σ-field for$\{{}_hX_t,\ {}_hU_t\}$ and the natural σ-field generated by ${}_h\hat{U}_0$.

[4]These measures, v_h and $\hat{v}_h$, will be, respectively, the probability measures for the starting values $\left({}_hX_0,\ {}_hU_0,\ {}_h\hat{U}_0\right)$ under, respectively, the true model and the $ARCH$ approximation.

each $h > 0$ and for all $(x, u, \hat{u}) \in R^{n+2m}$

$$\int_{\{\hat{u}^*=U_h(x^*,x,\hat{u})\}} \hat{\Pi}_h (x, u, \hat{u}, d (x^*, u^*, \hat{u}^*))$$
$$= \int_{\{\hat{u}^*=U_h(x^*,x,\hat{u})\}} \Pi_h (x, u, \hat{u}, d (x^*, u^*, \hat{u}^*)) = 1 \quad (5.1)$$

Under $\hat{v}_h$ and $\hat{\Pi}_h$, u and $\hat{u}$ are treated as being equal almost surely. So for all $h > 0$ and all $(x, u, \hat{u}) \in R^{n+2m}$,

$$\int_{\{u^*=\hat{u}^*\}} \Pi_h (x, u, \hat{u}, d (x^*, u^*, \hat{u}^*)) = 1, \quad (5.2)$$
$$\int_{\{u^*=\hat{u}^*\}} \hat{v}_h (d (x^*, u^*, \hat{u}^*)) = 1 \quad (5.3)$$

We need to note that no feedback from $\left\{{}_h\hat{U}_{kh}\right\}_{k=0,1,2,...}$ into $\{{}_hX_{kh}, {}_hU_{kh}\}_{k=0,1,2,...}$ is allowed. That is, given ${}_hX_{kh}$ and ${}_hU_{kh}$, ${}_hX_{(k+1)h}$ and ${}_hU_{(k+1)h}$ are independent of ${}_h\hat{U}_{kh}$ under P_h, so for any $\Gamma \in \mathfrak{B} (R^{n+m})$ and all h,

$$\Pi_h (x, u, \hat{u}, \Gamma \times R^m) = \Pi_h (x, u, 0_{m\times 1}, \Gamma \times R^m), \quad (5.4)$$

where $0_{m\times 1}$ is an $m \times 1$ vector of zeros.

For each $h > 0$, let P_h and $\hat{P}_h$ be the probability measures on $D ([0, \infty), R^n \times R^{2m})$ such that

$$P_h \left[\left({}_hX_0, {}_hU_0, {}_h\hat{U}_0\right) \in \Gamma\right] = v_h (\Gamma) \text{ for any } \Gamma \in \mathfrak{B} \left(R^{n+2m}\right) \quad (5.5)$$
$$\hat{P}_h \left[\left({}_hX_0, {}_hU_0, {}_h\hat{U}_0\right) \in \Gamma\right] = \hat{v}_h (\Gamma) \text{ for any } \Gamma \in \mathfrak{B} \left(R^{n+2m}\right) \quad (5.5')$$

$$P_h \left[\left({}_hX_t, {}_hU_t, {}_h\hat{U}_t\right) = \left({}_hX_{kh}, {}_hU_{kh}, {}_h\hat{U}_{kh}\right), \ kh \le t < (k+1) h\right] = 1 (5.6)$$

$$\hat{P}_h\left[\left({}_hX_t,\ {}_hU_t,\ {}_h\hat{U}_t\right) = \left({}_hX_{kh},\ {}_hU_{kh},\ {}_h\hat{U}_{kh}\right),\ kh \le t < (k+1)\,h\right] \quad = \quad 1 \qquad (5.6')$$

and for all $k \ge 0$ and $\Gamma \in \mathfrak{B}\left(R^{n+2m}\right)$,

$$P_h\left[\left({}_hX_{(k+1)h},\ {}_hU_{(k+1)h},\ {}_h\hat{U}_{(k+1)h}\right) \in \Gamma \mid \mathfrak{M}_{kh}\right] = \Pi_h\left({}_hX_{kh},\ {}_hU_{kh},\ {}_h\hat{U}_{kh},\ \Gamma\right) \qquad (5.7)$$

almost surely under P_h, and

$$\hat{P}_h\left[\left({}_hX_{(k+1)h},\ {}_hU_{(k+1)h},\ {}_h\hat{U}_{(k+1)h}\right) \in \Gamma \mid \mathfrak{M}_{kh}\right] = \hat{\Pi}_h\left({}_hX_{kh},\ {}_hU_{kh},\ {}_h\hat{U}_{kh},\ \Gamma\right) \qquad (5.7')$$

almost surely under $\hat{P}_h$.

For each $h > 0$, (5.5) and (5.5′) specify the distribution of the starting point $\left({}_hX_0,\ {}_hU_0,\ {}_h\hat{U}_0\right)$ under P_h and $\hat{P}_h$, respectively. (5.6) and (5.6′) form the continuous time process $\left\{{}_hX_t,\ {}_hU_t,\ {}_h\hat{U}_t\right\}$ as a step function with jumps at times h, $2h, \ldots$. (5.7) and (5.7′) specify the transition probabilities for the jumps in $\left\{{}_hX_t,\ {}_hU_t,\ {}_h\hat{U}_t\right\}$.

Next we define the forecast functions for $\{{}_hX_t,\ {}_hU_t\}$. Let $A \in \mathfrak{B}\left(D\left[0,\infty\right), R^{n+m}\right)$.[5] The conditional probability under P_h that $\{{}_hX_t,\ {}_hU_t\}_{\tau\le t\le\infty} \in A$ is

$$\begin{aligned} & P_h\left[\{{}_hX_t, {}_hU_t\}_{\tau\le t\le\infty} \in A \mid \mathfrak{M}_\tau\right] \\ = \ & P_h\left[\{{}_hX_t, {}_hU_t\}_{\tau\le t\le\infty} \in A \mid {}_hX_\tau,\ {}_hU_\tau,\ {}_h\hat{U}_\tau\right] \quad \text{a.s. under } P_h \\ = \ & P_h\left[\{{}_hX_t, {}_hU_t\}_{\tau\le t\le\infty} \in A \mid {}_hX_\tau,\ {}_hU_\tau\right] \quad \text{a.s. under } P_h \qquad (5.8) \end{aligned}$$

The first equality holds since $\left\{{}_hX_{kh},\ {}_hU_{kh},\ {}_h\hat{U}_{kh}\right\}$ is a Markov chain under P_h. The second equality follows from (5.4), so $\{{}_hX_{kh},\ {}_hU_{kh}\}$ is also a Markov chain

[5] Note that R^{n+m}, not R^{n+2m}. Since we are interested in forecasting $\{{}_hX_t, {}_hU_t\}$ rather than $\left\{{}_hX_t, {}_hU_t, {}_h\hat{U}_t\right\}$, we use the Markov structure of $\{{}_hX_t, {}_hU_t\}$ under both P_h and $\hat{P}_h$ to drop $\left\{{}_h\hat{U}_t\right\}$.

under P_h. Now for every $(x, u, \hat{u}, \tau) \in R^{n+2m+1}$, define $P_{(h,x,u,\tau)}$ and $\hat{P}_{(h,x,u,\tau)}$ on $(D([0,\infty), R^{n+m}), \mathfrak{B}([0,\infty), \ R^{n+m}))$ by replacing P_h, $\hat{P}_h$, and $k \geq 0$ in (5.6)-(5.7) and (5.6′)-(5.7′) with $P_{(h,x,u,\tau)}$, $\hat{P}_{(h,x,u,\tau)}$ and $k \geq [\tau/h]$, and replacing (5.5) and (5.5′) with

$$P_{(h,x,u,\tau)}\left[\left({}_hX_\tau,\ {}_hU_\tau,\ {}_h\hat{U}_\tau\right)\right] = (x, u, u) = 1 \tag{5.5''}$$

$$\hat{P}_{(h,x,u,\tau)}\left[\left({}_hX_\tau,\ {}_hU_\tau,\ {}_h\hat{U}_\tau\right)\right] = (x, u, u) = 1 \tag{5.5'''}$$

Since the forecasting generated using these probabilities will regard only the future paths of $\{{}_hX_t,\ {}_hU_t\}$ and once ${}_hU_\tau$ is fixed, the value of ${}_h\hat{U}_t$ is irrelevant. So, there is no loss of generality to set ${}_hU_\tau = {}_h\hat{U}_\tau = u$ in (5.5″) and (5.5‴). Now we define the forecast functions $F_h(A, x, u, \tau)$ and $\hat{F}_h(A, x, u, \tau)$ by

$$F_h(A, x, u, \tau) \equiv P_{(h,x,u,\tau)}\left[\{{}_hX_t,\ {}_hU_t\}_{\tau \leq t \leq \infty} \in A\right] \tag{5.9}$$

$$\hat{F}_h(A, x, u, \tau) \equiv \hat{P}_{(h,x,u,\tau)}\left[\{{}_hX_t,\ {}_hU_t\}_{\tau \leq t \leq \infty} \in A\right] \tag{5.10}$$

$F_h(A, x, u, \tau)$ and $\hat{F}_h(A, x, u, \tau)$, here, are functions of $A, x, u,$ and τ and are not random since they are defined in terms of unconditional probabilities rather than conditional probabilities.

Since the aim of this chapter is to make the forecast function $F_h(\cdot)$ close to $\hat{F}_h(\cdot)$, we want to show that, for every $\zeta > 0$,

$$\lim_{h \to 0} P_h\left(\left|F_h(A,\ {}_hX_\tau,\ {}_hU_\tau,\ \tau) - \hat{F}_h\left(A,\ {}_hX_\tau,\ {}_h\hat{U}_\tau,\ \tau\right)\right| > \zeta\right) = 0 \tag{5.11}$$

In other words, we first generate, using P_h, the underlying data, namely the sample path of $\{{}_hX_t,\ {}_hU_t\}$. Next we use the $ARCH$ recursive updating formula, which is identical in P_h and $\hat{P}_h$, to generate the $\left\{{}_h\hat{U}_t\right\}$. Finally, at some time τ, we generate forecast for the future path of $\{{}_hX_t,\ {}_hU_t\}$ first, using the true state variables $({}_hX_\tau,\ {}_hU_\tau)$ and the forecast function generated by the true probability P_h, and second, using the $ARCH$ estimate ${}_h\hat{U}_\tau$ in place of ${}_hU_\tau$ and using the forecast function generated by the $ARCH$ probability measure $\hat{P}_h$. Then we compare the forecasts: if the difference between them converges to zero in probability under P_h as $h \to 0$ for all well-behaved events, then we say that the forecasts generated by the $ARCH$ model $\hat{P}_h$ are asymptotically correct.

There are three steps involved to prove that $\hat{P}_h$ consistently estimates the forecast distribution for P_h over the interval $[\tau, \infty)$.

5.2.1 Step 1

In this step we show that given ${}_h\hat{U}_\tau = {}_hU_\tau$, the forecasts generated at time τ by P_h and $\hat{P}_h$ become very close as $h \to 0$. That is, for $\tau > 0$ and for every $(x, u) \in R^{n+m}$, $F_h(A, x, u, \tau) - \hat{F}_h(A, x, u, \tau) \to 0$ as $h \to 0$.

The assumptions following will assure that $\{{}_hX_t,\ {}_hU_t\}$ and $\left\{{}_hX_t,\ {}_h\hat{U}_t\right\}$ converge weakly to limit processes $\{X_t, U_t\}$ and $\left\{X_t, \hat{U}_t\right\}$ under P_h and $\hat{P}_h$, respectively as $h \to 0$, where the limit processes $\{X_t, U_t\}$ and $\left\{X_t, \hat{U}_t\right\}$ are generated by the stochastic integral equations

$$\begin{aligned} (X_t', U_t')' &= (X_0', U_0')' + \int_0^t b(X_s, U_s)\, ds + \int_0^t \sigma(X_s, U_s)\, dW_s \\ &\quad + \int_0^t \int g(X_s)\, d\tilde{N}_\lambda(ds), \qquad (5.12) \\ \left(X_t', \hat{U}_t'\right)' &= \left(X_0', \hat{U}_0'\right)' + \int_0^t \hat{b}\left(X_s, \hat{U}_s\right) ds + \int_0^t \hat{\sigma}\left(X_s, \hat{U}_s\right) dW_s \\ &\quad + \int_0^t \int \widehat{g}(X_s)\, d\tilde{N}_{\hat{\lambda}}(ds), \qquad (5.12') \end{aligned}$$

where $\{W_t\}$ is an $(n+m) \times 1$ standard Brownian motion, independent of $\left(X_0, U_0, \hat{U}_0\right)$ and $\tilde{N}_\lambda(ds)$ is a compensated Poisson process defined as $\tilde{N}_\lambda(ds) = N_\lambda(ds) - \lambda ds$. We assume that $\{W_t\}$ and $\left\{\tilde{N}_\lambda\right\}$ are independent. Here $b(X_t, U_t)$ is the

$(n+m) \times 1$ instantaneous drift per unit time in $\{X_t, U_t\}$ and $\sigma^2(X_t, U_t)$ is the $(n+m) \times (n+m)$ instantaneous conditional covariance matrix per unit of time of the increments in $\{X_t, U_t\}$. (X_0, U_0) is assumed to be random with a distribution π.[6] In (5.12′), $\hat{b}$, $\hat{\sigma}^2$, $\hat{U}_t$ and $\hat{\pi}$ replace b, σ^2, U_t and π, respectively. P_0 and $\hat{P}_0$ are the probability measures on $D([0,\infty), R^{n+m})$ generated by (5.12) and (5.12′), respectively.

Now, define the first and second conditional moments, for each $h > 0$,

$$\begin{aligned}
b_h(x,u) &= h^{-1} E_{(h,x,u,kh)} \left[\left[\begin{array}{c} {}_hX_{(k+1)h} - {}_hX_{kh} \\ {}_hU_{(k+1)h} - {}_hU_{kh} \end{array} \right] \right], \\
\sigma_h^2(x,u) &= h^{-1} E_{(h,x,u,kh)} \left[\left[\begin{array}{c} {}_hX_{(k+1)h} - {}_hX_{kh} \\ {}_hU_{(k+1)h} - {}_hU_{kh} \end{array} \right] \left[\begin{array}{c} {}_hX_{(k+1)h} - {}_hX_{kh} \\ {}_hU_{(k+1)h} - {}_hU_{kh} \end{array} \right]' \right], \\
\Delta_h^{\varepsilon}(x,u) &= h^{-1} P_h \left[\left| {}_hX_{(k+1)h} - {}_hX_{kh} \right| > \varepsilon \,|\, {}_hX_{kh} = x,\ {}_hU_{kh} = u \right] \\
g_h(x) &= x_t - x_{t-}, \quad x_{t-} = \lim_{s \to t} x_s \ \text{ for } s < t
\end{aligned} \tag{5.13}$$

$$\begin{aligned}
\hat{b}_h(x,u) &= h^{-1} \hat{E}_{(h,x,\hat{u},kh)} \left[\left[\begin{array}{c} {}_hX_{(k+1)h} - {}_hX_{kh} \\ {}_h\hat{U}_{(k+1)h} - {}_h\hat{U}_{kh} \end{array} \right] \right], \\
\hat{\sigma}_h^2(x,u) &= h^{-1} \hat{E}_{(h,x,\hat{u},kh)} \left[\left[\begin{array}{c} {}_hX_{(k+1)h} - {}_hX_{kh} \\ {}_h\hat{U}_{(k+1)h} - {}_h\hat{U}_{kh} \end{array} \right] \left[\begin{array}{c} {}_hX_{(k+1)h} - {}_hX_{kh} \\ {}_h\hat{U}_{(k+1)h} - {}_h\hat{U}_{kh} \end{array} \right]' \right], \\
\hat{\Delta}_h^{\varepsilon}(x,u) &= h^{-1} \hat{P}_h \left[\left| {}_hX_{(k+1)h} - {}_hX_{kh} \right| > \varepsilon \,\middle|\, {}_hX_{kh} = x,\ {}_h\hat{U}_{kh} = \hat{u} \right]. \\
\hat{g}_h(x) &= x_t - x_{t-}, \quad x_{t-} = \lim_{s \to t} x_s \ \text{ for } s < t
\end{aligned} \tag{5.13'}$$

[6] Here, the initial distribution π is defined as

$$\pi(\Gamma) = \nu_0(\Gamma \times R^m)$$

for every $\Gamma \in B(R^{n+m})$. For $\hat{\pi}$, ν_0 is replace with $\hat{\nu}_0$.

$b_h(x,u)$ and $\hat{b}_h(x,u)$ are the conditional drift under probability measures P_h and $\hat{P}_h$, respectively, and $\sigma_h^2(x,u)$ and $\hat{\sigma}_h^2(x,u)$ are the conditional second moments under P_h and $\hat{P}_h$, respectively. And $g_h(x)$ and $\widehat{g}_h(x)$ measure the magnitude of jumps with intensity of λ in the process under P_h and $\hat{P}_h$, respectively. Note that each moment is normalised by the length of sampling interval, h.

Now, we state the assumptions which are required to obtain the weak convergence result.

Assumption 1. Under P_h, $\left({}_hX_0,\ {}_hU_0,\ {}_h\hat{U}_0\right) \Rightarrow \left(X_0,\ U_0,\ \hat{U}_0\right)$ as $h \to 0$ with probability measure ν_0. Under $\hat{P}_h$, $\left({}_hX_0,\ {}_hU_{0},{}_h\hat{U}_0\right) \Rightarrow \left(X_0,\ U_0,\ \hat{U}_0\right)$ as $h \to 0$ with probability measure $\hat{\nu}_0$.

This assumption requires that the random starting points $\left({}_hX_0,\ {}_hU_{0},{}_h\hat{U}_0\right)$ converge to those of the limit process under P_h and $\hat{P}_h$ as $h \to 0$.

Assumption 2. There exist an $\delta > 0$ such that for every $R > 0$ and every $k > 0$,

$$\lim_{h \to 0} \sup_{\|x\| \le R, \|u\| \le R} h^{-1} E_{(h,x,u,kh)} \left[\left| {}_hX_{i,(k+1)h} - {}_hX_{i,kh} \right|^{2+\delta} \right] = 0 \qquad (5.14)$$

for $i = 1, 2, \ldots, n$, and

$$\lim_{h \to 0} \sup_{\|x\| \le R, \|u\| \le R} h^{-1} E_{(h,x,u,kh)} \left[\left| {}_hU_{i,(k+1)h} - {}_hU_{i,kh} \right|^{2+\delta} \right] = 0 \qquad (5.15)$$

a.s. under P_h for all $i = 1, 2, \ldots, m$. Further (5.14) and (5.15) hold when E_h, P_h, u are replaced by $\hat{E}_h$, $\hat{P}_h$ and $\hat{u}$, respectively.

This assumption puts conditional moment restrictions to guarantee that the sample path of the jump-diffusion limit has only first kind discontinuities. That is, the sample paths for the diffusion part of the process are continuous with probability one, but for jump part, it allows only first kind (discrete) jumps.

Assumption 3. There exist continuous $(n+m)\times 1$ functions $b(x,u)$ and $\hat{b}(x,u)$ and $(n+m)\times(n+m)$ continuous positive semi-definite functions $\sigma^2(x,u)$ and $\hat{\sigma}^2(x,u)$ such that for every $R>0$,

$$\begin{aligned}
\lim_{h\to 0}\sup_{\|x\|\le R\|u\|\le R}\|b_h(x,u)-b(x,u)\| &= 0 \\
\lim_{h\to 0}\sup_{\|x\|\le R\|u\|\le R}\left\|\sigma_h^2(x,u)-\sigma^2(x,u)\right\| &= 0 \\
\lim_{h\to 0}\sup_{\|x\|\le R\|u\|\le R}\|g_h(x)-g(x)\| &= 0 \\
\lim_{h\to 0}\sup_{\|x\|\le R\|u\|\le R}\Delta_h^{\varepsilon}(x,u) &= \lambda \qquad (5.16)
\end{aligned}$$

$$\begin{aligned}
\lim_{h\to 0}\sup_{\|x\|\le R\|u\|\le R}\left\|\hat{b}_h(x,u)-\hat{b}(x,u)\right\| &= 0 \\
\lim_{h\to 0}\sup_{\|x\|\le R\|u\|\le R}\left\|\hat{\sigma}_h^2(x,u)-\hat{\sigma}^2(x,u)\right\| &= 0 \\
\lim_{h\to 0}\sup_{\|x\|\le R\|u\|\le R}\|\widehat{g}_h(x)-\widehat{g}(x)\| &= 0 \\
\lim_{h\to 0}\sup_{\|x\|\le R\|u\|\le R}\hat{\Delta}_h^{\varepsilon}(x,u) &= \hat{\lambda} \qquad (5.16')
\end{aligned}$$

Here, we assume that the second moments, drifts and jumps per unit time converge uniformly on compact sets to well-behaved functions of time and state variables x. The probability jumps of size greater than ε is assumed to converge to a constant λ. It is also required that $b(x,u)$, $\widehat{b}(x,u)$, $\sigma^2(x,u)$, $\hat{\sigma}^2(x,u)$, $g(x)$, $\widehat{g}(x)$, v_0, and $\hat{v}_0$ completely characterise the distributions of the jump-diffusion limit $\{X_t,U_t\}$ and $\left\{X_t,\hat{U}_t\right\}$.

Assumption 4. For any choice of π_0 and $\hat{\pi}_0$, distributionally unique solutions exists to the stochastic integral equations (5.12) and (5.12′).

Theorem 9 *Under Assumptions 1-4,* $\{{}_hX_t,\ {}_hU_t\} \Rightarrow \{X_t, U_t\}$ *under* P_h *as* $h \to 0$, *where the initial distribution* π *is given by*

$$\pi(\Gamma) \equiv \nu_0(\Gamma \times R^m), \tag{5.17}$$

for every $\Gamma \in B(R^{n+m})$. *If* π *and* ν_0 *in* (5.17) *are replaced by* $\hat{\pi}$ *and* $\hat{\nu}_0$, $\left\{{}_hX_t,\ {}_h\hat{U}_t\right\} \Rightarrow \left\{X_t, \hat{U}_t\right\}$ *under* $\hat{P}_h$ *as* $h \to 0$.

Proof. See the proof of Theorem 1 in Appendix C.1. ■

We have shown that the weak convergence of the pairs of $\{{}_hX_t,\ {}_hU_t\}$ and $\left\{{}_hX_t,\ {}_h\hat{U}_t\right\}$ to its jump-diffusion limit as $h \to 0$. Next, we define forecast functions informally for the Markov processes (5.12) and (5.12′) :

$$F_0(A, x, u, \tau) \equiv P_0(A \mid X_t = x, U_t = u) \tag{5.18}$$

$$\hat{F}_0(A, x, u, \tau) \equiv \hat{P}_0\left(A \mid X_t = x, \hat{U}_t = \hat{u}\right) \tag{5.18′}$$

where $A \in \mathfrak{B}(D[\tau, \infty), R^{n+m})$, and $P_0(\cdot)$ and $\hat{P}_0(\cdot)$ are the probability measures corresponding to (5.12) and (5.12′), respectively.

Assumption 5. For all $(x, u) \in R^{n+m}$, $\hat{b}(x, u) = b(x, u)$, $\hat{\sigma}^2(x, u) = \sigma^2(x, u)$, $\widehat{g}(x, u) = g(x, u)$, and $\hat{\lambda} = \lambda$.

This assumption says that the misspecified $ARCH$ model generating $\hat{P}_h$ correctly specifies the functional form of the first and second conditional moments of ${}_hX_t$ and ${}_hU_t$, and the structure of jumps. This assumption is the most important one to move from consistent filtering to consistent estimation of the forecast distribution.

Then with Assumption 5, the conditions of Theorem 9 accomplish the first step. I.e., if ${}_hU_\tau = {}_h\hat{U}_\tau$ and if $\hat{b}(x,u) = b(x,u)$, $\hat{\sigma}^2(x,u) = \sigma^2(x,u)$, $\widehat{g}(x,u) = g(x,u)$, and $\hat{\lambda} = \lambda$, the forecast distributions generated by $\hat{P}_h$ and P_h at time τ become close and both become close to the forecast distribution generated by the limit diffusion P_0 as $h \to 0$.

5.2.2 Step 2

In this section, we show that $\hat{P}_h$ is a consistent filter for P_h at time τ. That is, ${}_h\hat{U}_\tau - {}_hU_\tau \to 0$ in probability under P_h as $h \to 0$. This step is proving the properties of a misspecified $ARCH$ models as consistent filter.

Before we state some additional assumptions, we will define the measurement error process $\{{}_hY_t\}$ as, for all $h > 0$ and all $t > 0$, ${}_hY_t = {}_h\hat{U}_t - {}_hU_t$.

Assumption 6. For every $h > 0$ and $\delta > 0$ and every $(x, u, \hat{u}) \in R^{n+2m}$, the following are well-defined and finite

$$\begin{aligned} c_{h,\delta}(x,u,\hat{u}) &= h^{-\delta} E_h \left[{}_hY_{(k+1)h} - {}_hY_{kh} \mid {}_hX_{kh} = x, \right. \\ &\qquad \left. {}_hU_{kh} = u,\ {}_h\hat{U}_{kh} = \hat{u} \right] \qquad (5.19) \\ d_{h,\delta}(x,u,\hat{u}) &= h^{-\delta} E_h \left[\left({}_hY_{(k+1)h} - {}_hY_{kh}\right)\left({}_hY_{(k+1)h} - {}_hY_{kh}\right)' \mid \right. \\ &\qquad \left. {}_hX_{kh} = x,\ {}_hU_{kh} = u,\ {}_h\hat{U}_{kh} = \hat{u} \right]. \qquad (5.20) \end{aligned}$$

Further, there exists a function $c(c, u, \hat{u})$ with $c(x, u, \hat{u}) = 0$ whenever $u = \hat{u}$ such that for some δ, $0 < \delta < 1$, and for every $R > 0$,

$$\lim_{h \to 0} \sup_{\|x\| \le R, \|u\| \le R, \|\hat{u}\| \le R} \left\| c_{h,\delta}(x,u,\hat{u}) - c(x,u,\hat{u}) \right\| = 0 \qquad (5.21)$$

and

$$\lim_{h \to 0} \sup_{\|x\| \le R, \|u\| \le R, \|\hat{u}\| \le R} \left\| d_{h,\delta}(x,u,\hat{u}) \right\| = 0 \qquad (5.22)$$

By (5.21) and (5.22), the drift and second moment of the increments in $\{_hY_{kh}\}$ are $O\left(h^{\delta}\right)$ and $o\left(h^{\delta}\right)$, respectively. This implies that $\{_hY_t\}$ operates on a faster scale than $\{_hX_t\}$, since the drift per unit of time of $\{_hY_t\}$ grow at faster rate as $h \longrightarrow 0$ than the drift and variance per unit of time of $\{_hX_t\}$. Therefore, if $\{_hY_t\}$ is mean-revert to a vector of zeros, the drift converges to zero as well with increasing rate as $h \longrightarrow 0$. Another implication is that, as $h \longrightarrow 0$, the drift of $\{_hY_t\}$ dominates the variance of $\{_hY_t\}$. Thus, the behaviour of $\{_hY_t\}$ can be approximated by a deterministic differential equation.

Assumption 7. For each $(x, u, \hat{u}) \in R^{n+2m}$, define the ordinary differential equation

$$\frac{dY\left(t, x, u, \hat{u}\right)}{dt} = c\left(x, u, \left[Y\left(t, x, u, \hat{u}\right) + u\right]\right), \tag{5.23}$$

with initial condition $Y\left(0, x, u, \hat{u}\right) = \hat{u} - u$. Then $0_{m\times 1}$, an $m \times 1$ vector of zeros, is a globally asymptotically stable solution for bounded values of $(x, u, \hat{u})$. That is, for every $R \geq 0$,

$$\lim_{t\to\infty} \sup_{\|x,u,\hat{u}\|\leq R} \left\|Y\left(t, x, u, \hat{u}\right)\right\| = 0_{m\times 1}. \tag{5.24}$$

This assumption will ensure that the differential equation approximating the behaviour of $\{_hY_t\}$ is well behaved, pulling $\{_hY_t\}$ back to a vector of zeros. The next assumption is to guarantee that $\{_hY_t\}$ does not diverge to infinity in finite time.

Assumption 8. There exists a nonnegative, twice differentiable function $\rho\left(x, y, h\right)$ and a positive function $M\left(R, h\right)$ such that

$$\lim_{R\to\infty} \liminf_{h\to 0} \inf_{\|(x,y)\|\geq R} \rho\left(x, y, h\right) = \infty, \tag{5.25}$$

$$\limsup_{R\to\infty} \limsup_{h\to 0} M\left(R, h\right) < \infty, \tag{5.26}$$

$$\limsup_{h\to 0} E_h\left[\rho\left({}_hX_0, {}_hY_0, h\right)\right] \; < \; \infty, \tag{5.27}$$

and for every $R > 0$ and $h > 0$,

$$\sup_{\|(x,u,\hat{u})\|\leq R} h^{-1}E_h\left[\rho\left({}_hX_{(k+1)h}, {}_hY_{(k+1)h}, h\right) - \rho\left(x, \hat{u} - u, h\right) \mid \right.$$

$$\left. {}_hX_{kh} = x,\; {}_hU_{kh} = u,\; {}_h\hat{U}_{kh} = \hat{u}\right] - M\left(R, h\right)\rho\left(x, \hat{u} - u, h\right)$$

$$\leq \; 0 \tag{5.28}$$

Theorem 10 *(Ethier-Nagylaki, 1988). Let Assumptions 1-5 and 6-8 hold. Then for every* $\tau >$, $0 < \tau < \infty$, $\left\{{}_h\hat{U}_t\right\}$ *is a consistent filter for* $\{{}_hU_t\}$ *at time* τ *under* $\{P_h\}$ *as* $h \to 0$, *where we say that* $\left\{{}_h\hat{U}_t\right\}$ *is a consistent filter for* $\{{}_hU_t\}$ *at time* τ *under* $\{P_h\}$ *as* $h \to 0$, *if for all* $\varepsilon > 0$,

$$\lim_{h\to 0} P_h\left[\left\|{}_h\hat{U}_\tau - {}_hU_\tau\right\| > \varepsilon\right] = 0. \tag{5.29}$$

Proof. See Ethier and Nagylaki (1988) . ■

5.2.3 Step 3

This step is to show that the forecasts generated by the $ARCH$ model with jumps are smooth in the underlying state variables, so that as ${}_h\hat{U}_\tau - {}_hU_\tau$ approaches to zero, the forecasts generated by the $ARCH$ models with jumps converge to the forecast generated by the correct model. That is, $F_h\left(A, x, u, \tau\right) - \hat{F}_h\left(A, x, u, \tau\right) \to 0$ as $u \to \hat{u}$ and $h \to 0$.

We need couple of definitions about the consistent estimation of the forecast distribution. They are adapted from Nelson(1995) .

Definition 11 Let ∂A be the boundary of the set A.[7] Let M_τ be the collect of sets A such that $A \in \mathfrak{B}(D([\tau,\infty), R^{n+m}))$ and $P_0\left[\{X_t, U_t\}_{\tau\leq t<\infty} \in \partial A \mid (X_t, U_t) = (x,u)\right] = 0$ for all starting points (x,u). Here P_0 can be defined analogously as in (5.18), and we treat $P_0[\cdot \mid (X_t, U_t) = (x,u)]$ as a function of (x,u) rather than a random variable as we did with the forecast functions.

Definition 12 We say that $\left\{{}_hX_t, {}_h\hat{U}_t\right\}_{\tau\leq t<\infty}$ consistently estimates the forecast distribution of $\{{}_hX_t,\ {}_hU_t\}_{\tau\leq t<\infty}$ at time τ if for every $A \in M_\tau$ and every $\varepsilon > 0$,

$$P_h\left[\left|F_h(A,\ {}_hX_\tau,\ {}_hU_\tau,\tau) - \hat{F}_h\left(A,\ {}_hX_\tau,\ {}_h\hat{U}_\tau,\tau\right)\right| > \varepsilon\right] \to 0$$

as $h \to 0$.

Theorem 11 *If Assumptions 1-8 are satisfied, then for every $A \in M_\tau$, every $\delta > 0$, and every $\varepsilon > 0$, there exists an $h^* > 0$ such that for every h, $0 < h \leq h^*$*

$$P_h\left[\left|F_h(A,\ {}_hX_\tau,\ {}_hU_\tau,\tau) - \hat{F}_h\left(A,\ {}_hX_\tau,\ {}_h\hat{U}_\tau,\tau\right)\right| > \varepsilon\right] \leq \delta \quad (5.30)$$

$$P_h\left[\left|F_h(A,\ {}_hX_\tau,\ {}_hU_\tau,\tau) - F_0(A,\ {}_hX_\tau,\ {}_hU_\tau,\tau)\right| > \varepsilon\right] \leq \delta \quad (5.31)$$

$$P_h\left[\left|\hat{F}_h\left(A,\ {}_hX_\tau,\ {}_h\hat{U}_\tau,\tau\right) - \hat{F}_0(A,\ {}_hX_\tau,\ {}_hU_\tau,\tau)\right| > \varepsilon\right] \leq \delta \quad (5.32)$$

Proof. See Appendix C.3. ∎

Theorem 12 *Let Assumptions 1-4 and 6-8 be satisfied. Then for every $\varepsilon > 0$, every $\tau > 0$, $0 < \tau < \infty$, and every $A \in M_\tau$,*

$$P_h\left[\left|F_h(A,\ {}_hX_\tau,\ {}_hU_\tau,\tau) - F_0(A,\ {}_hX_\tau,\ {}_hU_\tau,\tau)\right| > \varepsilon\right] \to 0, \quad (5.33)$$

[7] The set of all points in $D([\tau,\infty), R^{n+m})$ which are limit points both of A and its complement.

$$P_h\left[\left|\hat{F}_h\left(A,\, {}_hX_\tau,\, {}_h\hat{U}_\tau,\tau\right)-\hat{F}_0\left(A,\, {}_hX_\tau,\, {}_hU_\tau,\tau\right)\right|>\varepsilon\right] \;\rightarrow\; 0. \tag{5.34}$$

Proof. See the Appendix C.3. ■

5.3 Example

In this section we will show an example that a misspecified $ARCH$ model can estimate the consistent forecast generated by the true data-generating process. We will assume that the true data-generating process is a stochastic volatility model and a $ARCH$ with jump components can approximate the stochastic volatility data-generating process. We propose an $ARCH$ approximation to a stochastic volatility model and show that it satisfies Theorem 9, 10, and 13.

Let S_t be the price of a non-dividend paying stock at time t. σ_t is its instantaneous returns volatility. We assume that

$$d\left[\ln\left(S_t\right)\right] = \left(\mu-\sigma^2/2\right)dt+\sigma_t dW_{1,t}+c\,d\eta_t \tag{5.35}$$

$$d\left[\ln\left(\sigma_t^2\right)\right] = -\beta\left[\ln\left(\sigma_t^2\right)-\alpha\right]dt+\Lambda dW_{2,t}. \tag{5.36}$$

$W_{1,t}$ and $W_{2,t}$ are standard Brownian motion with correlation ρ and η_t is a Poisson process with intensity of λ. c is the magnitude of a jump when a jump occurs. We assume that $W_{i,t}$ and η_t are independent for $i=1,2$. Then

$$\begin{bmatrix} dW_{1,t}+d\eta_t \\ dW_{2,t} \end{bmatrix}\begin{bmatrix} dW_{1,t}+d\eta_t & dW_{2,t} \end{bmatrix}=\begin{bmatrix} 1+\lambda & \rho \\ \rho & 1 \end{bmatrix}dt. \tag{5.37}$$

μ, Λ, β and α are constants. We allow $\{\sigma_t\}$ to vary randomly with $\{\ln\left(\sigma_t^2\right)\}$ following an Ornstein-Uhlenbeck process.

We assume that we observe $\{S_t\}$ at discrete intervals of length h, so for every t, ${}_hS_t \equiv {}_hS_{h[t/h]}$. We suppose that $S_0>0$ and $\sigma_0>0$ to be nonrandom. Because of the continuous time Markov structure, the discrete time process $\{_hX_{t,h}U_t\}$ is also Markov.

Now consider using a conditional normal $AR(1)$ $EGARCH$ model by Nelson (1991) to forecast when the data are generated by (5.35)-(5.37). The $AR(1)$ $EGARCH$ generated the fitted conditional variances $_h\hat{\sigma}^2_{kh}$ by the recursive formulae

$$\ln\left({}_h\hat{\sigma}^2_{(k+1)h}\right) = \ln\left({}_h\hat{\sigma}^2_{kh}\right) - \hat{\beta}\left[\ln\left({}_h\hat{\sigma}^2_{kh}\right) - \hat{\alpha}\right] + h^{1/2}\zeta\left({}_hZ_{(k+1)h}\right), \tag{5.38}$$

where $_h\hat{\sigma}^2_0 > 0$ is fixed for all h and

$$\zeta(z) \equiv \theta z + \gamma\left[|z| - (2/\pi)^{1/2}\right], \tag{5.39}$$

$$_hZ_{(k+1)h} \equiv \frac{\left[\ln\left({}_hS_{(k+1)h}\right) - \ln\left({}_hS_{kh}\right) - h\left(\hat{\mu} - {}_h\hat{\sigma}^2_{kh}/2\right) - \eta_{(k+1)h}\,\widehat{c}_h\right]}{\left[h^{1/2}\,{}_h\hat{\sigma}_{kh} + \eta_{kh}\,v_h\right]} \tag{5.40}$$

(5.35)-(5.40) completely specify the $\left\{{}_hS_{kh},\ {}_h\sigma^2_{kh},\ {}_h\hat{\sigma}^2_{kh}\right\}$ process under P_h. Next, we construct the $\hat{P}_h$ measure. Under $\hat{P}_h$ the recursive updating formulae (5.38)-(5.40) continue to hold, but (5.35)-(5.37) are replaced with

$$\begin{aligned}
\ln\left({}_hS_{(k+1)h}\right) &= \ln\left({}_hS_{kh}\right) + h\left(\hat{\mu} - {}_h\hat{\sigma}^2_{kh}/2\right) + h^{1/2}\,{}_h\hat{\sigma}_{kh}\,{}_hZ_{(k+1)h} \\
&\quad + \eta_{(k+1)h}\left(\widehat{c}_h + \hat{v}_h\,{}_hZ_{(k+1)h}\right) && (5.41) \\
{}_hZ_{kh} &\sim \text{iid } N(0,1), && (5.42) \\
{}_h\hat{\sigma}^2_{kh} &= \sigma^2_{kh} \quad \text{a.s. under } \hat{P}_h && (5.43)
\end{aligned}$$

The continuous time process $\left\{{}_hS_t,\ {}_h\sigma^2_t,\ {}_h\hat{\sigma}^2_t\right\}$ will be created by making them step functions having jumps at h, $2h$, $3h$, $\ldots$, as in (5.6) and (5.6′). Here we can obtain (5.41) by simply rearranging (5.40). Under P_h, (5.40) is a definition of

${}_hZ_{(k+1)h}$.[8] Under $\hat{P}_h$, however, (5.38)-(5.43) define the transition probabilities for the $\{{}_hS_{kh},\ {}_h\sigma^2_{kh},\ {}_h\hat{\sigma}^2_{kh}\}$ process. Note that the condition in (5.43) requires that $\hat{\sigma}^2_0 = \sigma^2_0$ under $\hat{P}_h$ which need not be true under P_h. Consistent filtering is not achieved at time 0 if $\hat{\sigma}^2_0 \neq \sigma^2_0$. Therefore, we required $\tau > 0$ in earlier section.

Nelson (1992) shows that the main requirement for consistent filtering for this model is $\gamma > 0$. For consistent estimation of forecast distributions, however, we must match the first two conditional moments of the $ARCH$ model considered as a data-generating process to the corresponding moments of the true data-generating process (5.35)-(5.37).

Under $\hat{P}_h$, the innovation in $\ln\left({}_h\hat{\sigma}^2_{(k+1)h}\right)$ is $h^{1/2}\zeta\left({}_hZ_{kh}\right)$, which has variance $h\left[\theta^2 + \gamma^2\ (1 - 2/\pi)\right]^{1/2}$. Under $\hat{P}_h$, the instantaneous correlation of the increments in $\ln\left({}_hS_{kh}\right)$ and $\ln\left({}_h\sigma^2_{kh}\right)$ is $\theta / \left[\theta^2 + \gamma^2 (1 - 2/\pi)\right]^{1/2}$. We need to match the conditional second moments under $\hat{P}_h$ and P_h, which requires that

$$\theta^2 + \gamma^2 [1 - 2/\pi] = \Lambda^2 \tag{5.44}$$

$$\rho = \theta / \left[\theta^2 + \gamma^2 (1 - 2/\pi)\right]^{1/2} \tag{5.45}$$

which is easily accomplished by setting $\theta = \rho{\cdot}\Lambda$ and $\gamma = |\Lambda| (1 - \rho^2)^{1/2} / (1 - 2/\pi)^{1/2}$. Also the drifts of (5.35) and (5.36) are $[(\mu - \sigma^2_t/2) + \lambda c]$ and $-\beta [\ln(\sigma^2_t) - \alpha]$ respectively. If $\hat{\alpha} = \alpha$, $\hat{\beta} = \beta$, $\hat{\mu} = \mu$ and $\widehat{c} = c$, then the drifts in (5.35) and (5.36) are equal to the drifts in (5.38) and (5.41).

5.4 Conclusion

In this chapter, we have derived that the conditions under which a misspecified $ARCH$ process with jumps performs well in forecasting as well as filtering. The

[8]Note that ${}_hZ_{kh}$ are used to generate the probability measure $\hat{P}_h$ over the stochastic process $\{{}_hS_{kh},\ {}_h\sigma^2_{kh},\ {}_h\hat{\sigma}^2_{kh}\}$. ${}_hZ_{kh}$ are not $iidN\,(0, 1)$ under P_h when they are obtained recursively.

conditions for the consistent forecasting are much stricter than those for consistent filtering. For example, we require the conditional second moments generated under P_h and $\hat{P}_h$, to match, although this condition is not required for the consistent filtering. Without this moment matching condition, the forecasts generated by $\hat{P}_h$ will not approach to the forecasts generated by P_h. Nelson (1995) gives an example with diffusion limit of $ARCH$ processes.

Although the jump-diffusion processes have been used to model financial time series to a degree, there is not much literature about the forecasting ability of jump-diffusion processes. Although we need more empirical evidence about the forecasting ability of jump-diffusion process, we may expect that $ARCH$ with jump approximation will produce reasonable forecasts with the results in this chapter when the data are generated by jump-diffusion processes.

To derive the result, we only considered the consistency of the forecasts, not the efficiency. This will be left for future research.

Chapter 6

Conclusion

For the last two decades or so, $ARCH$ type models introduced by Engle(1982) have been used rather successfully as a major tool to analyse financial markets. The success seems due to the ability of $ARCH$ models to characterise heteroskedasticity and non-linearity of financial time series data. Yet, there has been another model for analysing the financial markets in theoretical financial economics literature, which is the jump-diffusion process. In some empirical works, models including jump components are found to be advantageous in terms of fitting the financial time series data. For example, $ARCH$ with jumps seems to fit the data better than $ARCH$ alone [Jorion(1988)]. If it is the case, it seems to be natural to investigate the relationship between $ARCH$ models with jumps and their jump-diffusion limits.

In this thesis, we have dealt with the three econometric issues with jump-diffusion processes: i) the relationship between jump-diffusion processes in continuous time and their discrete time counterparts, ii) misspecified model can still produce consistent estimates of conditional covariance of the true data generating process, and iii) misspecified model can generated consistent forecasts of the true data generating process.

For the first issue, we have shown that a sequence of discrete time processes converges weakly to a jump-diffusion limit as the length of sampling interval goes to zero. That is, an $ARCH$ model with jump component can be an approximation of a jump-diffusion process. So, by including jump components into an $ARCH$ framework, it is expected to explain the real economy more closely than other modeling tools.

The next issue has arisen from the fact that economic or econometric models are rough approximations of the real economy. However closely they explain the real economy, they are inevitably misspecified. It has been shown that misspecified $ARCH$ models with jumps can estimate the conditional covariance of true data generating processes consistently. That is, we have shown that the measurement error between the conditional covariance estimates from a misspecified $ARCH$ model with jump and the true conditional covariance converges zero in probability as the length of sampling interval approaches to zero.

Then, the forecasting ability of misspecified $ARCH$ with jumps has been dealt with. Since most market participants try to minimise their risk existing in the future, it would be necessary to investigate the forecasting ability of those economic or econometric models. Under some regularity conditions, we have shown that the difference between forecasts generated by a misspecified model and those generated by true data generating process converges to zero as the length of sampling interval goes to zero.

Some empirical studies found that when the daily data are aggregated to either weekly or monthly data, the jump components tend to disappear. It may be that measurement error associated with daily data inducing jumps in the process. In Jarrow and Rosenfeld(1984), they tested the null hypothesis of continuous sample paths for stock prices. After adjusting for weekends and holidays, the null is rejected for daily data with small magnitudes of jumps, but not for the weekly and monthly data. This would be the reason that with the weekly and monthly data, the inclusion of weekends and holidays tends to cover up the small jump components. To avoid a false determination of a jump process, it might be important how we choose time intervals between observations. With the high technology employed in the modern financial markets, the data are collected even in second from the market these days. It would be interesting to research further to see how frequently we need to observe the markets to approximate models more accurately.

In the previous chapters, we assumed the convergence of the parameters of the processes to prove the weak convergence between a discrete time process and a continuous time process, when some regularity conditions are satisfied. Those regularity conditions are required to achieve the desired convergence. For example, in chapter 3, we imposed the conditions for the parameters given in Bollerslev(1986) , which are mainly for the variance to remain positive. However, we did not considered how these regularity conditions should be altered for different parameter estimators for the approximation or convergence. In some cases, we require more restrictive regularity conditions for the parameters to achieve a desired approximation or convergence. In other cases, we don't. We may extend our results to consideration of these regularity conditions in the convergence or the approximation of the processes.

While proving the weak convergence of $ARCH$ with jumps to its jump-diffusion limit, we restricted the jump intensity to a fixed constant λ for the simplicity of the discussion. In reality, this intensity is likely to have a certain probability structure. It is expected to improve the performance of jump-diffusion process in empirical researches, if the constant restriction on the jump intensity is relaxed. For the consistent filtering and forecasting of $ARCH$ models with jump, we only considered the consistency of those estimates and forecasts, but the efficiency, in the current research. These issues will require future research.

Appendix A

Higher Moments for a $GARCH\,(1,1)$-M Process with Jumps

The higher order moments up to fourth order of the process ${}_h y_t$ in (3.23) are obtained as follow;

$$
\begin{aligned}
& E\left[h^{-1}\left({}_h y_{kh} - {}_h y_{(k-1)h}\right)^3 |\mathfrak{M}_{kh}\right] \\
= & \left[h^2 \mu_h^3\, {}_h\sigma_{kh}^6 + 3h\mu_h\, {}_h\sigma_{kh}^4\right] \\
& +\lambda\left[c_h^3 + 3h^2\mu_h^2\, {}_h\sigma_{kh}^4 c_h + 6h^{3/2}\mu_h\, {}_h\sigma_{kh}^3\, \phi_h + 3h\mu_h\, {}_h\sigma_{kh}^2 c_h^2\right. \\
& \left.+3h\mu_h\, {}_h\sigma_{kh}^2\, \phi_h^2 + 3\, {}_h\sigma_{kh}^2 h c_h + 6_h\sigma_{kh} h^{1/2} c_h \phi_h + 3c_h^2\phi_h\right] \\
= & \lambda\left(c_h^3 + 3c_h\phi_h^2\right) + o\,(1) \qquad \text{(A.1)}
\end{aligned}
$$

$$
\begin{aligned}
& E\left[h^{-1}\left({}_h y_{kh} - {}_h y_{(k-1)h}\right)^4 |\mathfrak{M}_{kh}\right] \\
= & \left[h^3\mu_h^4\, {}_h\sigma_{kh}^8 + 6h^2\mu_h^2\, {}_h\sigma_{kh}^6 + 3\, {}_h\sigma_{kh}^4 h\right]
\end{aligned}
$$

$$
\begin{aligned}
& +\lambda\left[c_h^4+6c_h^2\phi_h^2+12\,{}_h\sigma_{kh}^3h^{3/2}\phi_h+6\,{}_h\sigma_{kh}^2hc_h^2+18\,{}_h\sigma_{kh}^2h\phi_h^2\right. \\
& +12\,{}_h\sigma_{kh}h^{1/2}c_h^2\phi_h+12\,{}_h\sigma_{kh}h^{1/2}\phi_h^3+12h^{5/2}\mu_h^2\,{}_h\sigma_{kh}^5\phi_h \\
& +4h\mu_h\,{}_h\sigma_{kh}^2c_h^3+6h^2\mu_h^2\,{}_h\sigma_{kh}^4\phi_h^2+12h^2\mu_h\,{}_h\sigma_{kh}^3c_h \\
& +24h^{3/2}\mu_h\,{}_h\sigma_{kh}^3c_h\phi_h+12h\mu_h\,{}_h\sigma_{kh}^2c_h\phi_h^2+4h^3\mu_h^3\,{}_h\sigma_{kh}^6c_h \\
& \left.+6h^2\mu_h^2\,{}_h\sigma_{kh}^4c_h^2+3\phi_h^4\right] \\
= & \ \lambda\left(c_h^4+6c_h^2\phi_h^2+3\phi_h^4\right)+o(1) \qquad (A.2)
\end{aligned}
$$

The limits of $(A.1)$ and $(A.2)$ as h goes to zero are

$$
\lim_{h\to 0} E\left[h^{-1}\left({}_hy_{kh}-\ {}_hy_{(k-1)h}\right)^3|\mathfrak{M}_{kh}\right]=\lambda\left(c^3+3c\phi^2\right),
$$

and

$$
\lim_{h\to 0} E\left[h^{-1}\left({}_hy_{kh}-\ {}_hy_{(k-1)h}\right)^4|\mathfrak{M}_{kh}\right]=\lambda\left(c^4+6c^2\phi^2+3\phi^4\right).
$$

Also the higher moments of the process ${}_h\sigma_t^2$ exist and converge to zero.

$$
\begin{aligned}
& E\left[h^{-1}\left(\sigma_{(k+1)h}^2-\sigma_{kh}^2\right)^3|\mathfrak{M}_{kh}\right] \\
= & \ E\left[h^{-1}\left(\omega_h+\sigma_{kh}^2\left(\beta_h+\alpha_h\,{}_hZ_{kh}^2-1\right)\right)^3|\mathfrak{M}_{kh}\right] \\
= & \ h^{-1}\omega_h^3+3h^{-1}\omega_h^2\,{}_h\sigma_{kh}^2\left(\alpha_h+\beta_h-1\right) \\
& +3h_h^{-1}\omega\,{}_h\sigma_{kh}^4\left(\alpha_h+\beta_h-1\right)^2 \\
& +6h^{-1}\beta_h^2\omega_h\,{}_h\sigma_{kh}^4 \\
& +h^{-1}\,{}_h\sigma_{kh}^6\left(\alpha_h+\beta_h-1\right)^3 \\
& +h^{-1}\,{}_h\sigma_{kh}^6\left(14\alpha_h^3-6\alpha_h^2+6\alpha_h\beta_h^2\right) \\
= & \ o(1) \qquad (A.3)
\end{aligned}
$$

$$
\begin{aligned}
& \lim_{h \to 0} E\left[h^{-1}\left(\sigma^2_{(k+1)h} - \sigma^2_{kh}\right)^3 |\mathfrak{M}_{kh}\right] \\
= & \lim_{h \to 0}\left[h^{-1}{}_h\sigma^6_{kh}\left(14\beta^3_h - 6\beta^2_h + 6\alpha_h\beta^2_h\right)\right] \\
= & \ 0 \qquad \text{(A.4)}
\end{aligned}
$$

$$
\begin{aligned}
& E\left[h^{-1}\left(\sigma^2_{(k+1)h} - \sigma^2_{kh}\right)^4 |\mathfrak{M}_{kh}\right] \\
= & \ E\left[h^{-1}\left(\omega_h + \sigma^2_{kh}\left(\alpha_h + \beta_h\,{}_hZ^2_{kh} - 1\right)\right)^4 |\mathfrak{M}_{kh}\right] \\
= & \ h_h^{-1}\omega^4_h + h^{-1}4\omega^3_h\,{}_h\sigma^2_{kh}\left(\alpha_h + \beta_h - 1\right) \\
& +h^{-1}6\omega^2_h\,{}_h\sigma^4_{kh}\left(\alpha_h + \beta_h - 1\right)^2 \\
& +h^{-1}12\omega^2_h\,{}_h\sigma^4_{kh}\alpha^2_h \\
& +h^{-1}4\omega_h\,{}_h\sigma^6_{kh}\left(\alpha_h + \beta_h - 1\right)^3 \\
& -h^{-1}4\omega_h\,{}_h\sigma^6_{kh}\alpha_h\beta_h\left(\alpha_h + \beta_h - 1\right) \\
& -h^{-1}4\omega_h\,{}_h\sigma^6_{kh}\left(-3\alpha_h\beta_h - 14\alpha^3_h + 6\alpha^2_h\right) \\
& +h_h^{-1}\omega\,{}_h\sigma^6_{kh}12\alpha_h\beta_h\left(\alpha_h + \beta_h - 1\right) \\
& +h_h^{-1}\omega\sigma^6 12\alpha_h\beta_h\left(2\alpha_h - 1\right) \\
= & \ o(1) \qquad \text{(A.5)}
\end{aligned}
$$

$$
\lim_{h \to 0} E\left[h^{-1}\left(\sigma^2_{(k+1)h} - \sigma^2_{kh}\right)^4 |\mathfrak{M}_{kh}\right] = 0 \qquad \text{(A.6)}
$$

Therefore, the higher moments of the $ARCH$ with jump components can be decided by the distribution of jump part of the process, not by the diffusion part of the process.

Appendix B

Conditions for Non-Explosion

Theorem 10.2.1 in Strook and Varadhan(1975) provides a non-explosion condition for the limit process. This condition ensures that the limit process does not explode in finite time. In the theorem, the condition is given for the case of diffusion process. Here, we adopt this condition for the jump-diffusion process with replacing the infinitesimal operator for a jump-diffusion process with that of a diffusion process.

Suppose that there exists a nonnegative function $\varphi(x,t)$ which is twice differentiable with respect to x and differentiable with respect to t such that for each $T > 0$,

$$\lim_{|x|\to\infty} \inf_{0\le t\le T} \varphi(x,t) = \infty \tag{B.1}$$

and there exist a positive locally bounded function $M(T)$ such that for each $T > 0$, all $x \in R^n$, and all t, $0 \le t \le T$,

$$\left(\frac{\partial}{\partial t} + \mathfrak{L}_t\right)\varphi(x,t) \le M(T)\,\varphi(x,t), \tag{B.2}$$

where

$$\mathfrak{L}\varphi(x,t) = \sum_{i=1}^{n} b_i(x,t)\frac{\partial\varphi(x,t)}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^{n} a_{ij}(x,t)\frac{\partial^2\varphi(x,t)}{\partial x_i\partial x_j}$$

$$+\int\left[\varphi\left(x+g\left(x,t\right)\right)-\varphi\left(x\right)-\sum_{i=1}^{n}g_{i}\left(x,t\right)\frac{\partial\varphi}{\partial x_{i}}\right]N\left(h\right),$$

If we assume that $X_t = x$, $(B.2)$ ensures the instantaneous drift of $\varphi(X_t, t)$ grows linearly with $\varphi(X_t, t)$. Therefore, it guarantees that $\varphi(X_t, t)$ does not explode in finite time. Also $(B.1)$ will guarantee that if $\varphi(X_t, t)$ does not explode, neither will X_t.

Appendix C

Proofs of The Theorems

C.1 Proofs of Theorems in Chapter 3

Proof of Theorem 1

Proof. Let the infinitesimal operator for the jump-diffusion process be

$$\begin{aligned} \mathfrak{L}f(x) &= \sum_{i=1}^{n} b_i(x,t) \frac{\partial f}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^{n} a_{ij}(x,t) \frac{\partial^2 f}{\partial x_i \partial x_j} \\ &\quad + \int \left[f(x+g(x,t)) - f(x) - \sum_{i=1}^{n} g_i(x,t) \frac{\partial f}{\partial x_i} \right] N(h), \quad \text{(C.1)} \end{aligned}$$

then, by the Lemma 11.2.1 in Strook and Varadhan(1979) , Assumption 1 is equivalent to the condition that for each $f \in C_0^\infty(R^n)$

$$\frac{1}{h} A_h f \to \mathfrak{L} f,$$

where the infinitesimal operator for a discrete Markov process A_h is defined as

$$Af(x) = \int [f(y) - f(x)] \Pi_{h,h[t/h]}(x, dy).$$

Let's define a random process M_f as

$$M_f(t) = f(x_t) - f(x_0) - \int_0^t \mathfrak{L} f(x_s)\, ds.$$

According to Kushner(1984, Sec. 1.6), if $M_f(t)$ is a martingale, then there exist a Wiener process W_t, and a Poisson process $N_\lambda(t)$ with independent increments and identically distributed jumps which solve a jump diffusion model

$$X_t = X_0 + \int_0^t b(x,s)\, ds + \int_0^t \sigma(x,s)\, dW_s + \int_0^t \int g(x,s)\, \widetilde{N}_\lambda(ds) \qquad \text{(C.2)}$$

where $g(x,t)$ is a bounded continuous function and $\widetilde{N}_\lambda(dt)$ is a compensated Poisson process defined as $\widetilde{N}_\lambda(ds) = N(ds) - \lambda ds$ with jump probability of λ.

In Assumption 4, the distribution of X_t is specified by $v_{0,}$ $a(x,t)$, $b(x,t)$ and $g(x,t)$. As $\sigma(x,t)$ only enters the equation through $a(x,t)$ function, the distribution of X_t does not depend on the choice of $\sigma(x,t)$ as long as $\sigma(x,t)\sigma'(x,t) = a(x,t)$. ∎

Proof of Theorem 2

Proof. By showing that Assumption 1′ implies Assumption 1, we can prove the theorem, since Theorem 2 follows immediately by Theorem 1. To do so, we only need to prove that

$$\lim_{h\to 0} \sup_{|x|\le R,\, 0\le t\le T} \frac{1}{h} \int_{\|y-x\|>1} (y-x)_i^2\, \Pi_{h,h[t/h]}(x,dy) = 0, \qquad \text{(C.3)}$$

$$\lim_{h\to 0} \sup_{|x|\le R,\, 0\le t\le T} \frac{1}{h} \int_{\|y-x\|>1} |y-x|_i\, \Pi_{h,h[t/h]}(x,dy) = 0, \qquad \text{(C.4)}$$

since the conditions for $c(x,u)$ and $\Delta_h^\varepsilon(x,t)$ remain same as in Assumption 1.

By Hőlder's Integral Inequality,

$$\frac{1}{h}\int_{\|y-x\|>1} |y-x|_i \, \Pi_{h,h[t/h]}(x,dy)$$
$$\leq \left[\gamma_{h,i,\delta}(x,t)\right]^{1/(2+\delta)} \left[\Delta_h^{\varepsilon}(x,t)\right]^{(1+\delta)/(2+\delta)}. \tag{C.5}$$

By (3.16), there is some $\delta > 0$ such that for all R, $T > 0$, the right hand side of the inequality vanishes to zero for every ε as $h \to 0$ uniformly on $\|x\| \leq R$, $0 \leq t \leq T$, proving $(C.4)$. Again, by Hőlder's Integral Inequality

$$\frac{1}{h}\int_{\|y-x\|>1} (y-x)_i^2 \, \Pi_{h,h[t/h]}(x,dy)$$
$$\leq \left[\gamma_{h,i,\delta}(x,t)\right]^{2/(2+\delta)} \left[\Delta_h^{\varepsilon}(x,t)\right]^{\delta/(2+\delta)} \tag{C.6}$$

which vanishes in the same manner as $(C.3)$. ■

Proof of Theorem 3

Proof. To prove this theorem, we need to show that Assumption 1 and 2 are satisfied. First we can factor $a(y,s,t)$ into $\sigma(y,s,t)\,\sigma'(y,s,t)$ which satisfies Assumption 2. To show this

$$\begin{aligned}
a(y,s,t) &= \begin{bmatrix} g^2+\lambda(k^2+v^2) & g\Omega_{1,2}G' \\ G\Omega_{2,1}g & G\Omega_{2,2}G' \end{bmatrix} \\
&= \begin{bmatrix} g^2 & g\Omega_{1,2}G' \\ G\Omega_{2,1}g & G\Omega_{2,2}G' \end{bmatrix} + \begin{bmatrix} \lambda(k^2+v^2) & \mathbf{0}_{1,2} \\ \mathbf{0}_{2,1} & \mathbf{0}_{2,2} \end{bmatrix} \\
&= \left\{ \begin{bmatrix} gdW_{1,t} \\ GdW_{2,t} \end{bmatrix} + \begin{bmatrix} \left(k+dt^{-1/2}vdW_{1,t}\right)d\eta_t \\ \mathbf{0} \end{bmatrix} \right\} \\
&\quad \times \left\{ \begin{bmatrix} gdW_{1t} & dW_{2,t}'G' \end{bmatrix} + \begin{bmatrix} \left(k+dt^{-1/2}vdW_{1,t}\right)d\eta_t & \mathbf{0}' \end{bmatrix} \right\}
\end{aligned}$$

$$
\begin{aligned}
&= \begin{bmatrix} gdW_{1,t} + \left(k + dt^{-1/2} v dW_{1,t}\right) d\eta_t \\ GdW_{2,t} \end{bmatrix} \\
&\quad \times \begin{bmatrix} gdW_{1t} + \left(k + dt^{-1/2} v dW_{1,t}\right) d\eta_t & dW'_{2,t} G' \end{bmatrix} \\
&= \sigma\sigma' \qquad \text{(C.7)}
\end{aligned}
$$

Now we must show that Assumption 1′ is satisfied. That is, we need to show that $a_h^*(y,s,t)$, $b_h^*(y,s,t)$, and $c_h(y,s,t)$ converges to $a(y,s,t)$, $b(y,s,t)$ and $c(y,s,t)$ respectively, and $\gamma_{h,i,\delta}(y,s,t)$ converges to zero uniformly on compacts as $h \to 0$. Since

$$
b_h^*(y,s,t) = \begin{bmatrix} f(y,s,t) + \lambda k(y,s,t) \\ F(y,s,t) \end{bmatrix}, \qquad \text{(C.8)}
$$

$b_h^*(s,y,t) = b(s,y,t)$.

$$
a_h^*(y,s,t) = \begin{bmatrix} f^2 h + g^2 + \lambda\left(k^2 + v^2\right) & fhF' + G\Omega_{1,2}G' \\ Ffh + G\Omega_{2,1}g & G\Omega_{2,2}G' \end{bmatrix}, \qquad \text{(C.9)}
$$

which will converge to $a(y,s,t)$ as $h \to 0$, since f, F, g and G are locally bounded.

Finally, if we choose $\delta = 1$, then, by stacking elements of $\gamma_{h,i,1}(y,s,t)$ in a vector, we have

$$
\begin{aligned}
\gamma_{h,1}(y,s,t) &= h^{-1} E \begin{bmatrix} \left|hf + g\,{}_hZ_{kh} + \eta_{kh}\left(k + h^{-1/2} v\,{}_hZ_{kh}\right)\right|^3 \\ \left|hF + G\,{}_hZ_{kh}^*\right|^3 \end{bmatrix} \\
&= \begin{bmatrix} \sqrt{h}\left(g^2 + \lambda\left(k^2 + v^2\right)\right)\sqrt{fh + g^2 + \lambda\left(k^2 + v^2\right)} \\ \sqrt{h} G\Omega_{22}G' \sqrt{FF'h + G\Omega_{22}G'} \end{bmatrix} \\
&= O\left(h^{1/2}\right)
\end{aligned}
$$

uniformly on compacts.

Therefore, we showed that Assumptions 2 and 1′ are satisfied. ■

Proof of Theorem 4

Proof. To prove the theorem, we need to show that Assumption 4 is satisfied, then the result follows immediately by Theorem 2 and Theorem 3. To do so, we need to show that the system of stochastic differential equations has a unique solution.

i)show that the martingale problem for $a(\cdot,\cdot)$, $b(\cdot,\cdot)$, and $c(\cdot,\cdot)$ is well posed,

ii) show that the limit process does not explode.

By Chapter 8 Theorem 3.3 in Ethier and Kurtz (1986) and Theorem 11.2.3 in Strook and Varadhan (1979) we can prove the statement i). To prove the statement ii), define for $K > 0$,

$$\varphi \equiv K + f(S)|S| + f(V)\exp(|V|),$$

where

$$\begin{aligned} f(x) &\equiv \exp\left(-\frac{1}{|x|}\right), && \text{if } x = 0, \\ &\equiv 0, && \text{otherwise.} \end{aligned}$$

$\varphi(V,S)$ is nonnegative, arbitrarily differentiable and satisfies $(B.1)$. Its derivatives are locally bounded, so that positive K and M can be chose to satisfy $(B.2)$ on any compact set. For large values of S and V[1]

$$\begin{aligned} \varphi_V(V,S) &\approx sign(V)\exp(|V|), \\ \varphi_{VV}(V,S) &\approx \exp(|V|), \\ \varphi_S(V,S) &\approx sign(S), \end{aligned}$$

[1]Here, $\varphi_V = \frac{\partial\varphi}{\partial V}$, $\varphi_{VV} = \frac{\partial^2\varphi}{\partial V^2}$, $\varphi_S = \frac{\partial\varphi}{\partial S}$, and $\varphi_{SS} = \frac{\partial^2\varphi}{\partial S^2}$

$$\varphi_{SS}(V,S) \approx 0,$$

so that with $M > 1 + \alpha\beta + \Omega_{2,2}/2 + |\theta| + \lambda k$, there exist a finite K satisfying $(B.2)$. Then, then result follows by Theorem 2. ∎

C.2 Proofs of Theorems in Chapter 4

To prove Theorem 7, we need the following lemma:

Lemma C1 Let $\{\zeta_t\}_{[0,T]}$ be generated by the stochastic integral equation

$$\zeta_t = \zeta_0 + \int_0^t m(\zeta_s, s)\, ds + \int_0^t \Lambda^{1/2}(\zeta_s, s)\, dW_s + \int_0^t \int g(\zeta_s, t)\, dN_\lambda(ds), \tag{C.10}$$

where ζ_0 is fixed, $\{W_t\}$ is $q \times 1$ standard Brownian motion, $\tilde{N}_\lambda(ds)$ is a compensated Poisson process with jump intensity, λ, and $m(\zeta_t, t)$ and $\Lambda(\zeta_t, t)$ are $q \times 1$- and $q \times q$-valued functions, respectively, and where $(C.10)$ has a unique weak-sense solution. Next, let $\gamma(t)$ be a $q \times 1$ function satisfying

$$\gamma(t) = o\left(t^{1/2}\right) \quad \text{as} \quad t \to 0 \tag{C.11}$$

And assume that $\gamma(t,u) \sim iid\ N(c,v)$. Then, for some $M \geq 0$

$$t^{-1/2}(\zeta_t - \gamma(t) - \zeta_0) \xrightarrow{d} N(0, \Lambda(\zeta_0, 0) + \Psi_N) \times M \quad \text{as} \quad t \to 0, \tag{C.12}$$

where Ψ_N represents the second moment from a Poisson Process. Further, let $f(\zeta, t)$ be a continuous function from R^{q+1} into R^1. Let there exists an $\varepsilon > 0$ such that

$$\limsup_{t \to 0} E \left| f\left(t^{-1/2}(\zeta_t - \gamma(t) - \zeta_0), t\right)\right|^{1+\varepsilon} < \infty \tag{C.13}$$

Then, if $\xi \sim N(0, \Lambda(\zeta_0, 0))$,

$$\lim_{t \to 0} E\left[f\left(t^{-1/2}(\zeta_t - \gamma(t) - \zeta_0), t\right)\right] = E\left[f(\xi, 0)\right] < \infty \tag{C.14}$$

Proof. If we prove $(C.12)$, then $(C.14)$ follows directly by Billingsley(1986, p286-291). To prove $(C.12)$, first note that $t^{-1/2}\gamma(t) \to 0$ as $t \to 0$ by $(C.11)$. Next rewrite $(C.10)$ as

$$\begin{aligned} & t^{-1/2}(\zeta_t - \zeta_0) \\ = \; & t^{-1/2}\Lambda^{1/2}(\zeta_0, 0) W_t + t^{-1/2}\int_0^t m(\zeta_s, s)\, ds \\ & + t^{-1/2}\int_0^t \left(\Lambda^{1/2}(\zeta_s, s) - \Lambda^{1/2}(\zeta_0, 0)\right) dW_s \\ & + t^{-1/2}\int_0^t \int g(\zeta_s, s)\, dN_\lambda(ds) \end{aligned} \tag{C.15}$$

Since $t^{-1/2}W_t \sim N(0, I)$ and $t^{-1/2}\gamma(t) \to 0$ as $t \to 0$, $(C.12)$ follows if the second and the third terms on the right hand side of $(C.15)$ converge in probability to zero as $t \to 0$. The second term is

$$\left| t^{-1/2}\int_0^t m(\zeta_s, s)\, ds \right| \le t^{-1/2} \max_{0 \le s \le t} |m(\zeta_s, s)| \tag{C.16}$$

which converges to a vector of zeros in probability as $t \to 0$. Next consider the last term. Define

$$\xi_t \equiv \int_0^t \left(\Lambda^{1/2}(\zeta_s, s) - \Lambda^{1/2}(\zeta_0, 0)\right) dW_s \tag{C.17}$$

If $t^{-1/2}\xi_t$ vanishes in probability to a vector of zeros, the proof of the lemma is completed. With $I(\cdot)$, the indicator function, we have

$$t_t^{-1}\xi\xi_t' = I(\|\xi_t\| < 1)\, t_t^{-1}\xi\xi_t' + I(\|\xi_t\| \ge 1)\, t_t^{-1}\xi\xi_t'. \tag{C.18}$$

The second term in $(C.18)$ vanishes in probability as $t \to 0$ since

$$\lim_{t\to 0} E\left[t^{-1} I\left(\|\xi_t\| \geq 1\right) | \zeta_0\right] = 0 \tag{C.19}$$

Since Λ is continuous, the first term also vanishes.

Similarly, define

$$\psi = t^{-1/2} \int_0^t \int g\left(\zeta_s, s\right) dN_\lambda\left(ds\right) \tag{C.20}$$

then, we have

$$t^{-1}\psi\psi' = I\left(\|\psi\| < 1\right) t^{-1}\psi\psi' + I\left(\|\psi\| \geq 1\right) t^{-1}\psi\psi'. \tag{C.21}$$

The first term on the right-hand side will vanish as $g\left(\zeta_s, s\right)$ is continuous, and the second term on the right-hand side will converge in probability as

$$\lim_{t\to 0} E\left[t^{-1} I\left(\|\psi\| \geq 1\right) | \zeta_0\right] = \lambda. \tag{C.22}$$

The proof is completed. ∎

Proof of Theorem 7

Proof. With Lemma C1 proved, the proof of this theorem is almost identical to the proof of Theorem 3.1 in Nelson (1992). ∎

Proof of Theorem 8

Proof. With Conditions 1-3, Assumptions 1-4 are satisfied. Proof of Theorem 7 is implies that Assumption 5 and 7 hold with replace Condition 4 with Condition 6.

From (4.62),

$$\frac{dY}{dt} = \gamma\left(\frac{2}{\pi}\right)\left[\exp\left(-\frac{Y}{2}\right) - 1\right],$$

which has a solution

$$Y_t = 2\ln\left\{1 - \exp\left[\frac{-\gamma\left(\frac{2}{\pi}\right)t}{2}\right] + \exp\left[\frac{1}{2}\left(Y_0 - \gamma\left(\frac{2}{\pi}\right)t\right)\right]\right\}. \tag{C.23}$$

Then, $(C.23)$ satisfies Assumption 6 as long as $\gamma > 0$. ∎

C.3 Proofs of Theorems in Chapter 5

To prove Theorem 13, we need couple of Lemmas.

Lemma C2 For any $\tau \geq 0$, define $\mu_{h,\tau}$ to be the probability measure for $({}_hX_\tau, {}_hU_\tau)$ generated by P_h. For any $\tau < \infty$ and for any $\delta > 0$, there exists a compact $\Lambda(\delta) \subset R^{n+m}$ such that for all h, $0 < h \leq h'$, $P_h\left[({}_hX_\tau, {}_hU_\tau) \in \Lambda(\delta)\right] > 1-\delta$.

Proof. The proof will be completed by Theorem 1 in Chapter 3, and Proposition 9.3.4 in Dudley (1989). ∎

Lemma C3 Let Assumptions 1-4 be satisfied. Let $(x_h, u_h) \to (x, u)$ as $h \to 0$. Then $P_{(0,x_h,u_h,\tau)} \Rightarrow P_{(0,x,u,\tau)}$, $P_{(h,x_h,u_h,\tau)} \Rightarrow P_{(0,x,u,\tau)}$, $\hat{P}_{(0,x_h,u_h,\tau)} \Rightarrow \hat{P}_{(0,x,u,\tau)}$, $\hat{P}_{(h,x_h,u_h,\tau)} \Rightarrow \hat{P}_{(0,x,u,\tau)}$ as $h \to 0$. If Assumption 5 is also satisfied, then $\hat{P}_{(h,x_h,u_h,\tau)} \Rightarrow \hat{P}_{(0,x,u,\tau)}$ as $h \to 0$. In each case the convergence is uniform on bounded subsets of R^{n+m}.

Proof. Theorem 11.2.3 in Strook and Varadhan (1979) together with Chapter 8 Theorem 3.3 in Ethier and Kurtz (1986) prove the theorem. ∎

Proof of Theorem 13

Proof. By Lemma C2, there exist a compact $\Lambda(\delta)$ and $h''(\delta) > 0$ such that $P_h\left[({}_hX_\tau,\ {}_hU_\tau) \in \Lambda(\delta)\right] > 1 - \delta/2$ where $h \leq h''(\delta)$. ${}_hU_\tau \to {}_h\hat{U}_\tau$ in probability under P_h as $h \to 0$ by Theorem 10. Therefore, for every $\varsigma > 0$ and $\delta > 0$, there exist an $h'''(\varsigma, \delta) > 0$ such that $P_h\left[\left\|{}_hU_\tau - {}_h\hat{U}_\tau\right\| > \varsigma\right] < \delta/2$ for $h < h'''(\varsigma, \delta)$. If we choose h as $h = \min\{h''(\delta), h'''(\varsigma, \delta)\}$, then

$$\begin{aligned} & P_h\left[\left|F_h(A, {}_hX_\tau, {}_hU_\tau, \tau) - \hat{F}_h\left(A, {}_hX_\tau, {}_h\hat{U}_\tau, \tau\right)\right| > \varepsilon\right] \\ < & \frac{\delta}{2} + \frac{\delta}{2} \\ & + \sup_{\substack{(x,u)\in\Lambda(\delta), \\ \|u-\hat{u}\|\leq\varsigma}} I\left[\left|F_h(A, {}_hX_\tau, {}_hU_\tau, \tau) - \hat{F}_h\left(A, {}_hX_\tau, {}_h\hat{U}_\tau, \tau\right)\right| > \varepsilon\right] \quad (C.24) \end{aligned}$$

where $I(\cdot)$ is the indicator function. By Lemma C3, there exist an $\varsigma(\delta) > 0$ and an $h''''(\delta, \varepsilon) > 0$ such that $\left|F_h(A, {}_hX_\tau, {}_hU_\tau, \tau) - \hat{F}_h\left(A, {}_hX_\tau, {}_h\hat{U}_\tau, \tau\right)\right| \leq \varepsilon$ whenever $(x, u) \in \Lambda(\delta)$ and $\|u - \hat{u}\| \leq \varsigma(\delta)$ and $h \leq h''''(\delta, \varepsilon)$. Thus, if $h \leq h'(\delta, \varepsilon) \equiv \min\{h''(\delta), h'''(\varsigma, \delta), h''''(\delta, \varepsilon)\}$, (5.30) is proved.

Next, for $h \leq h''(\delta)$,

$$\begin{aligned} & P_h\left[\left|F_h(A, {}_hX_\tau, {}_hU_\tau, \tau) - F_0(A, {}_hX_\tau, {}_hU_\tau, \tau)\right| > \varepsilon\right] \\ < & \frac{\delta}{2} + \sup_{(x,u)\in\Lambda(\delta)} I\left[\left|F_h(A, {}_hX_\tau, {}_hU_\tau, \tau) - F_0(A, {}_hX_\tau, {}_hU_\tau, \tau)\right| > \varepsilon\right] \quad (C.25) \end{aligned}$$

By Lemma C3, $P_{(h,x,u,\tau)} \Rightarrow P_{(0,x,u,\tau)}$ is uniform on compacts. So, the second term in the right hand side of $(C.25)$ will vanish uniformly in h. This will prove (5.31). Similarly, since $\hat{P}_{(h,x,u,\tau)} \Rightarrow \hat{P}_{(0,x,u,\tau)}$ is also uniform on compacts by Lemma C3, (5.32) is proved. ■

Proof of Theorem 14

Proof. By Lemma C3, the convergence of $P_{(h,x,u,\tau)} \Rightarrow P_{(0,x,u,\tau)}$ and $\hat{P}_{(h,x,u,\tau)} \Rightarrow \hat{P}_{(0,x,u,\tau)}$ are uniform in h. δ can be chosen very small in the fashion of h chosen in the proof of Theorem 13. Therefore, the theorem is proved. ■

Bibliography

[1] Ahn, C. M., and H. E. Thompson (1988), "Jump-Diffusion Processes and the Term Structure of Interest Rates", *Journal of Finance,* 43, 155-174.

[2] Amin, K. I. (1993), "Jump Diffusion Option Valuation in Discrete Time", *Journal of Finance,* 48, 1833-1863.

[3] Anderson, B. D. O., and J. B. Moore (1979), *Optimal Filtering*, Prentice-Hall, New York, NY.

[4] Arnold, L. (1973), *Stochastic Differential Equations: Theory and Applications*, John Wiley & Son, New York, NY.

[5] Ball, C. A., and A. Roma (1993), "A Jump Diffusion Model for the European Monetary System", *Journal of International Money and Finance,* 12, 475-492.

[6] Ball, C. A., and W. N. Torous (1983), "A Simplified Jump Process for Common Stock Returns", *Journal of Financial and Quantitative Analysis,* 18, 53-65.

[7] ________________________ (1985), "On Jumps in Common Stock Prices and Their Impact on Call Option Pricing", *Journal of Finance,* 40, 155-173.

[8] Bardhan, I., and X. Chao (1996), "Stochastic Multi-Agent Equilibria in Economies with Jump-Diffusion Uncertainty," *Journal of Economic Dynamics and Control,* 20, 361-384.

[9] Bates, D. (1991), "The Crash of '87: Was It Expected? The Evidence from Option Markets," *Journal of Finance*, 46, 1009-1044.

[10] ______ (1996*a*), "Dollar Jump Fears, 1984-1992: Distributional Abnormalities implicit in Currency Futures Options", *Journal of International Money and Finance,* 15, 65-93.

[11] ______ (1996*b*), "Jumps and Stochastic Volatility: Exchange Rate Processes Implicit in Deutsche Mark Options", *Review of Financial Studies,* 9, 69-107.

[12] Becker, S. (1981), "A Note on Estimating the Parameters of the Diffusion-Jump Model of Stock Returns", *Journal of Financial and Quantitative Analysis,* 16, 127-139.

[13] Bekaert, G., and S. F. Gray (1998), "Target Zones and Exchange Rates : An Empirical Investigation", *Journal of International Economics*, 45, 1-35.

[14] Bhattacharya, R. N. and E. C. Waymire (1990), *Stochastic Processes with Applications,* John Wiley & Son, New York, NY.

[15] Billingsley, P. (1986), *Probability and Measure,* 2nd ed., Wiley, New York, NY.

[16] Black, F., and M. Scholes (1973), "The Pricing of Options and Corporate Liabilities", *Journal of Political Economy,* 81, 637-654.

[17] Bollerslev, T. (1986), "Generalized Autoregressive Conditional heteroskedasticity," *Journal of Econometrics,* 31, 307-327.

[18] Bollerslev, T, R. Y. Chou, and K. F. Kroner (1992), "*ARCH* modelling in Finance: A Review of Theory and Empirical Evidence," *Journal of Econometrics,* 52, 5-59.

[19] Bollerslev, T., R. F. Engle, and D. B. Nelson(1994), "*ARCH* Models," in *Handbook of Econometrics,* Vol IV, eds. R. F. Engle and D. McFadden, Amsterdam, North-Holland, 2959-3038.

[20] Brooks, C. and S. P. Burke(1998), "Forecasting Exchange Rate Volatility Using Conditional Variance Models Selected by Information Criteria," *Economics Letters*, 61, 273-278.

[21] Campa, J. M. and P. H. K. Chang(1998), "The Forecasting Ability of Correlations Implied in Foreign Exchange Options," *Journal of International Money and Finance*, 17, 855-880.

[22] Chang, C. W. (1995), "A No-Arbitrage Martingale Analysis for Jump-Diffusion Valuation", *Journal of Financial Research,* 18, 351-381.

[23] Chib, S. (1998), "Estimation and Comparison of Multiple Change-Point Models," *Journal of Econometrics*, 86, 221-241.

[24] Chong, C. W., M. I. Ahmad and M. Y. Abdullah(1999), "Performance of *GARCH* Models in Forecasting Stock Market Volatility," *Journal of Forecasting*, 18, 333-343.

[25] Clark, P. K.(1973), "A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices", *Econometrica,* 41, 135-155.

[26] Cox, J. C. and S. A. Ross (1975), "The Pricing of Option for Jump Processes," Rodney L. White Centre Working Paper Series, No. 2-75, University of Pennsylvania, Philadelphia, Penn.

[27] ___________________ (1976), "The Valuation of Options for Alternative Stochastic Processes", *Journal of Financial Economics,* 3, 145-166.

[28] Das, S. R. (1997), "An Efficient Generalized Discrete Time Approach to Poisson-Gaussian Bond Option Pricing in the Heath-Jarrow-Morton Model," NBER Technical Working Paper, No. 212, Cambridge, MA.

[29] DeGrauwe, P., H. Dewachter, and D. Veestraeten (1999), "Price Dynamics under Stochastic Process Switching : Some Extensions and an Application to EMU," *Journal of International Money and Finance*, 18, 195-224.

[30] De Jong, F., F. C. Drost, and B J. M. Werker (1999), "A Jump-Diffusion Model for Exchange Rates in a Target Zone", Tilberg University, CentER Discussion Paper 9704.

[31] Dewachter, H. and D. Veestraeten (1998), "Expectation Revisions and Jumps in Asset Prices", *Economics Letters*, 59, 367-372.

[32] Dudley, R. M. (1989),*Real Analysis and Probability,* Wadsworth & Brooks/Cole, Pacific Grove, CA.

[33] Engle, R. F. (1982), "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, 987-1008.

[34] Engle, F. and T. Bollerslev, (1986), "Modelling the Persistence of Conditional Variances," *Econometrics Reviews,* 5, 1-50.

[35] Engle, F., V. Ng, and M. Rothschild (1990), "Asset Pricing with a Factor *ARCH* Covariance Structure : Empirical Estimates for Treasury Bills", *Journal of Econometrics*, 45, 213-238.

[36] Ethier, S. N. and T. G. Kurtz (1986), *Markov Process : Characterization and Convergence,* John Wiley & Son, New York, NY.

[37] Ethier, S. N. and T. Nagylaki (1988), "Diffusion Approximation of Markov Chains with Two Time Scales and Applications to Population Genetics II", *Advances in Applied Probability*, 20, 525-545.

[38] Feng, C. and S. D. Smith(1997), "Jump Risk, Time-Varying Risk Premia and Technical Trading Profits", Federal Reserve Bank of Atlanta, Working Paper 97-17.

[39] Franses, P. H. and D. V. Dijk (1996), "Forecasting Stock Market Volatility Using (Non-Linear) *GARCH* Models," *Journal of Forecasting*, 15, 229-235.

[40] Gihman, I. I. and A. V. Skorohod (1972), *Stochastic Differential Equation*, Springer-Verlag, Berlin, Heidelberg.

[41] Gouriéroux, C., A. Monfort, and E. Renault(1993), "Indirect Inference", *Journal of Applied Econometrics,* 4, S85-S118.

[42] Harvey, C. R. (1989), "Time-Varying Conditional Covariances in Tests of Asset Pricing Models", *Journal of Financial Economics*, 24, 289-318.

[43] Herzel, S. (1998), "A Simple Model for Option Pricing with Jumping Stochastic Volatility", *International Journal of Theoretical and Applied Finance,* 1, 487-505.

[44] Hilliard, J. E., and J. A. Reis (1999), "Jump Processes in Commodity Futures Prices and Option Pricing", *American Journal of Agricultural Economy,* 81, 273-286.

[45] Ho, M. S., W. R. M. Perraudin, and B. E. Sorensen (1996), "A Continuous-Time Arbitrage-Pricing Model with Stochastic Volatility and Jumps," *Journal of Business and Economic Statistics*, 14, 31-43.

[46] Jarrow, R. A. and E. R. Rosenfeld (1984), "Jump Risks and the Intertemporal Capital Asset Pricing Model," *Journal of Business*, 57, 337-351.

[47] Jiang, G. J. (1998), "Jump-Diffusion Model of Exchange Rate Dynamics-Estimation via Indirect Inference", University of Groningen, Research Institute SOM, Research Report No. 98A40.

[48] Jorion, P. (1988), "On Jump Processes in the Foreign Exchange and Stock Markets", *Review of Financial Studies*, 1, 427-445.

[49] Kendall, M. G., and A. Stuart (1969), *The Advanced Theory of Statistics,* London, Griffin and Co.

[50] Kim, M-J, Y-H Oh, and R. Brooks (1994), "Are Jumps in Stock Returns Diversifiable? Evidence and Implication for Option Pricing", *Journal of Financial and Quantitative Analysis,* 29, 609-631.

[51] Kremer, J. W., and R. L. Roenfeldt (1992), "Warrant Pricing: Jump-Diffusion vs. Black-Scholes", *Journal of Financial and Quantitative Analysis,* 28, 255-272.

[52] Kushner, H. J. (1984), *Approximation and Weak Convergence Methods for Random Processes, with Applications to Stochastic Systems Theory,* MIT Press, Cambridge, MA.

[53] Lamoureux, C. G. and W. D. Lastrapes (1993), "Forecasting Stock-Return Variance: Toward an Understanding of Stochastic Implied Volatilities," *Review of Financial Studies*, 6, 293-326.

[54] Lo, A. W. (1988), 'Maximum Likelihood Estimation of Generalised Ito Processes with Discretely Sampled Data", *Econometric Theory,* 4, 231-247.

[55] Mandelbrot, B. (1963), "The Variation of Certain Speculative Prices", *Journal of Business*, 36, 394-419.

[56] Merton, R. C. (1973), "An Intertemporal Capital Asset Pricing Model," *Econometrica*, 41, 867-887.

[57] ___________ (1976*a*), "The Impact on Option Pricing of Specification Error in the Underlying Stock Price Returns", *Journal of Finance,* 31, 333-350.

[58] ___________ (1976*b*), "Option Pricing When Underlying Stock Returns are Discontinuous", *Journal of Financial Economics,* 4, 125-144.

[59] ___________ (1990), *Continuous Time Finance,* Cambridge, MA: Basil Blackwell.

[60] Mills, T. C. (1996), "Non-Linear Forecasting of Financial Time Series: An Overview and Some New Models," *Journal of Forecasting*, 15, 127-135.

[61] Mulinacci, S. (1996), "An Approximation of American Option Prices in a Jump-Diffusion Model", *Stochastic Processes and Their Applications,* 62, 1-17.

[62] Nelson, D. B. (1990), "*ARCH* Models as Diffusion Approximations", *Journal of Econometrics,* 45, 7-38.

[63] ___________ (1991), "Conditional Heteroskedasticity in Asset Returns: New Approach," *Econometrica,* 59, 347-370.

[64] ___________ (1992), "Filtering and Forecasting with Misspecified ARCH Models I : Getting the right variance with the wrong model", *Journal of Econometrics*, 52, 61-90.

[65] Nelson, D. B., and D. P. Foster (1995), "Filtering and Forecasting with Misspecified ARCH Models II : Making the right forecast with the wrong model", *Journal of Econometrics*, 67, 303-335.

[66] Nimalendran, M. (1994), "Estimating the Effects of Information Surprises and Trading on Stock Returns Using a Mixed Jump-Diffusion Model", *Review of Financial Studies*, 7, 451-473.

[67] Oldfield, G. S., R. J. Rogalski, and R. A. Jarrow (1977), "An Autoregressive Jump Process for Common Stock Returns", *Journal of Financial Economics,* 5, 389-418.

[68] Park, K-H, C. M. Ahn, and R. Fujihara (1993), "Optimal Hedged Portfolios: the case of jump-diffusion risks", *Journal of International Money and Finance,* 12, 493-510.

[69] Praetz, P. D. (1972), "The Distribution of Share Price Changes", *Journal of Business,* 45, 49-55.

[70] Press, S. J. (1967), "A Compound Events Model for Security Prices", *Journal of Business,* 40, 317-335.

[71] Protter, P. (1990), *Stochastic Integration and Differential Equations: A New Approach*, Springer-Verlag, Berlin, Heidelberg.

[72] Ross, S. (1976), "The Arbitrage Theory of Capital Asset Pricing Model", *Journal of Economics Theory*, 13, 341-360.

[73] Schwert, G. W. (1989), "Why Does Stock Market Volatility Change Over Time?", *Journal of Finance,* 44, 1115-1153.

[74] Strook, D. W. and S. R. S. Varadhan (1979), *Multidimensional Diffusion Processes*, Springer-Verlag, New York.

[75] Vlaar, P. J. G., and F. C. Palm (1993), "The Message in Weekly Exchange Rates in the European Monetary System: Mean Reversion, Conditional Heteroskedasticity, and Jumps", *Journal of Business and Economic Statistics,* 11, 351-360.

[76] West, K. D. and D. Cho (1995), "The Predictive Ability of Several Models of Exchange Rate Volatility," *Journal of Econometrics*, 69, 367-391.

[77] Wiggins, J. B. (1987) , "Option Values Under Stochastic Volatility : Theory and Empirical Estimates", *Journal of Financial Economics",* 19, 351-372.

[78] Zhang, X. L. (1997) , "Numerical Analysis of American Option Pricing in a Jump-Diffusion Model," *Mathematics of Operation Research*, 22, 668-690.