Geometric Process Models for Financial Time Series

Connie Pui Yu LAM

A thesis submitted in fulfillment of
the requirements for the degree of
Master of Philosophy

School of Mathematics and Statistics
The University of Sydney

July 2011

Abstract

Time series is one of the main subjects in Statistics. It can apply to any kind of fields, such as finance, engineering, medical science, environmental science and even astronomy. A time series which exhibits a monotonic increasing or decreasing trend pattern is very common. Lam (1988) first proposed modelling directly a monotonic trend by a monotone process called the Geometric Process (GP). The definition is: Let $X_1,\ X_2, \dots$ be a set of positive random variables. If there exists a positive real number a, such that $\{Y_t = a^{t-1}X_t,\ t = 1, 2, \dots\}$ forms a renewal process (RP) (Feller, 1949) with mean $E(Y_t) = \mu$ and variance $Var(Y_t) = \sigma^2$, then $\{X_t,\ t = 1, 2, \dots\}$ is called a Geometric Process (GP), and the real number a is called the ratio of the GP. If $a > 1$, a GP is stochastically decreasing; if $0 < a < 1$, it is stochastically increasing; and if $a = 1$, it is stationary.

A financial time series often includes trend, and exhibits jump diffusion and volatility change. While most existing models can capture these features, they are complicated in the modelling approach and implementation. We propose GP model as a simple modelling alternative in modelling these features of a financial time series and in forecasting. To address these features, we first propose an autoregressive geometric process (ARGP) model in which an autoregressive (AR) term is added to the mean of the latent RP, in order to account for the autocorrelation of the series. Moreover, the ratio a of the GP model describes trend

movement, and allows volatility changes in the time series. Then the model is further extended to the conditional autoregressive geometric process (CARGP) range model by adopting a conditional autoregressive range (CARR)-type (Chou, 2005) mean function. This extended model captures the dynamics of the serial correlation better, and is easily applied in forecasting. Besides, by assigning the log-t distribution, expressed in scale mixtures form to the RP, the model is more robust for heavy-tailed time series data. Lastly, to capture the jumps that occur in many financial time series, we add the threshold and jump components to the mean function, and the advanced models are called CARGP threshold (CARGPT) and CARGP jump (CARGPJ) models, respectively. We show that the advanced models are the generalisation of the mean reverting jump diffusion (MRJD) model of Rambharat *et al.* (2005), and can implement easily and accurately in price forecasts.

The statistical inference of the ARGP model can be carried out easily by the least square error (LSE) method and maximum likelihood (ML) method. Due to the complicated likelihood functions when the mean of the data distribution adopts a CARR-type mean function, we implement the CARGP, CARGPT and CARGPJ models using Bayesian inference.

Declaration

I declare that this thesis represents my own work, except where due acknowledgement is made, and that it has not been previously included in a thesis, dissertation or report submitted to this University or to any other institution for a degree, diploma or other qualifications.

Signed _____

Connie Pui Yu LAM

Acknowledgements

I would like to express my gratitude to my supervisor, Dr. Jennifer S. K. Chan for her inspiration, guidance, support and encouragement during my master by research course. With her ideas and teaching, I can start my research work smoothly. With her support and understanding, I can finish this thesis satisfactorily. Moreover, I should show my thanks to Prof. Neville Weber, my associate supervisor, for his support on my study in the School. Besides, I would also like to thank all the staff in the School for their advice and help. Last but not least, I need to say thank you to my husband, Dr. Boris Choy and my son, Borison Choy for their encouragement and support on my continuing study. Also, without Boris's encouragement and belief, I will not have this achievement.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1. Background

We all live in a world of information. A time series presents the information of a certain event over time. Depending on the nature of the event, such time series may exhibit a wide range of characteristics from the monotonic trend movement to the heteroskedasticity. In the past decades, much research attention has been directed to investigate the many unique features in a financial time series. The remaining section is devoted to give an account of these features in a time series.

Data with increasing or decreasing *trend* patterns are common. In engineering, data of the successive operating times of a repairable system after repairs are monotonically decreasing because of the accumulated wears. On the other hand, due to the ageing effects, the failure rate of a medical treatment for a patient exhibits a monotonic increasing trend. In finance, the stock price may follow a monotonic increasing or decreasing trend over a certain period of time. The price of a certain commodity, such as electricity, is another example in our daily life, which exhibits monotonic increasing trend. A time series model which

fails to allow for the dynamics of the trend movement will not provide accurate forecast.

Another common feature of a time series is the cyclical behaviour or seasonal variation. *Seasonality* is presented in a time series when the outcome follows a cyclical pattern usually induced by some seasonal characteristics, and repeats itself again over a period of time, such as a year. If the seasonality is of secondary importance, it is usually removed from the data by a procedure called seasonal adjustment (Wallis, 1974). On the other hand, if forecasting is the major objective, a separate component is usually added to the mean function to allow for seasonality (Zekkberm, 1979).

Many linear models are governed by the intrinsic dynamics that small causes lead to small effect. However, such linearity is often violated in the course of time. A regime switching model is one of the models that allows for *nonlinearity*. In the model, a structural break or a model shift may occur when a time threshold or a threshold level for a certain risk variable is exceeded. See for example the threshold time model of Chan *et al.* (2006) and the threshold heteroskedastic models of Chen *et al.* (2008).

For some financial time series, Black (1976) revealed that a drop in price or return was associated with a rise in volatility. This phenomenon is known as the *leverage effect*, and is modelled by either introducing a correlation between the distribution of the log-return and

log-volatility in the stochastic volatility modelling framework, or simply adding a term of the lag-1 price in the volatility function. These two approaches are called parameter driven and observation driven models, respectively.

Apart from the seasonality, nonlinearity and leverage effects, another prominent feature for a time series is the *serial/auto/lagged correlation* (Box and Jenkins, 1976). Autocorrelation describes the similarity between observations, as a function of the time separation between them. The dynamics of autocorrelation may be induced by seasonality, leverage effects and many others. Many researches have been conducted to model the dynamics of serial correlation. See for example, the conditional autoregressive range model of Chou (2005) and the autoregressive moving average (ARMA) model of Box, Jenkins and Reinsel (1994).

Volatility clustering is visible in many financial time series. Volatility clustering refers to the phenomenon, as noted by Mandelbrot (1963), that large changes in observations tend to be followed by large changes, of either sign in volatility, and small changes tend to be followed by

small changes. An implication is that, while the observations themselves are uncorrelated, their absolute values or squares display a positive and significant autocorrelation function (ACF) that suggest the observations are dependent. The autoregressive conditional heteroskedastic (ARCH) (Engle, 1982) and generalised ARCH (GARCH) (Bollerslev, 1986) models are derived to describe the phenomenon of volatility clustering and related effects such as kurtosis.

Lastly, other financial time series such as the electricity market prices are highly volatile and often have spikes. They are caused by the low elasticity of supply and strong seasonality. To provide a robust inference that downweights the effect of spikes, some *heavy-tailed* or leptokurtic error distribution should be adopted. A common choice of these distributions is the log-t distribution for its lognormal counterpart. Alternatively, the *jump-diffusion* model of Merton (1973) models the spikes by two stochastic processes separately, for the jump and the diffusion. Moreover, the price for the non-storable commodity such as electricity may exhibit strong *mean-reversion* over a short time period, that would be inconsistent with the price of a storable good in an efficient market. In light of this, some mean reversion jump diffusion (MRJD) models are proposed. See for example the MRJD model of Rambharat *et al.* (2005) in electricity application.

It is challenging to derive models to address all these features in a time series, particularly a financial time series, which is the focus of

this thesis. The next section lists some important models that cater for the different features in a time series, to facilitate accurate modelling and forecasting.

1.2. Overview of finanical time series model

This section summarise briefly a few benchmark financial time series models that are developed and modified to address the various distinct characteristics of a financial time series. We begin with the autoregressive moving average (ARMA) model (Box, Jenkins and Reinsel, 1994) which allows for serial correlation in a stationary time series. It is a model combining two basic time series models, autoregressive (AR) model and moving average (MA) model, in order to describe the dynamic structure of time series adequately. Then, a more appropriate model for modelling non-stationary time series is its generalisation, the autoregressive integrated moving average (ARIMA) model. However, many time series exhibit nonlinear characteristics: the bilinear model (Granger and Anderson, 1978), the autoregressive model with random coefficients (Robinson, 1978) and the threshold models (Tong and Lim, 1980) are three important classes of nonlinear models. There are also the mean reverting jump diffusion (MRJD) models of Barz and Johnson (1998), Cartea and Figueroa (2005) and Rambharat *et al.* (2005), especially designed to model the spikes and mean reversions for high

frequency data. The jump and mean reversion components in the mean account partially the heteroskedasticity in a financial time series.

However, while the ARIMA and other nonlinear models study the dynamic of the mean of a time series, they do not allow for the time-varying nature of the volatility. Some models are constructed to include the modelling of the unobserved volatility of the series. They include the ARCH model of Engle (1982), the GARCH model of Bollerslev (1986) and stochastic volatility (SV) models of Hull and White (1987), Melino and Turnbull (1990) and Taylor (1994). The ARCH and GARCH models aim to describe the phenomenon of volatility clustering and related effects such as kurtosis more accurately. In the model, volatility is dependent upon past realisations of the price and volatility process.

The GARCH models determines the conditional volatility by a function of lagged squared residuals and lagged conditional variance while the SV models allow the conditional volatility to be modelled by an unobservable stochastic process. These approaches are sophisticated, adaptive and flexible but the model implementations are quite computationally demanding. This research proposes the Geometric Process (GP) model as a simple modelling alternative to model simultaneously the dynamics of the mean and variance.

1.3. Overview of GP model

To model the monotonic trend in a time series, the nonhomogeneous Poisson process and the Cox-Lewis model (Cox and Lewis, 1966) have been used for a while. Lam (1988) first proposed modelling directly a monotonic trend by a monotone process called the Geometric Process (GP).

Definition : Let $X_1,\ X_2, \dots$ be a set of positive random variables. If there exists a positive real number a, such that $\{Y_t = a^{t-1}X_t,\ t = 1, 2, \dots\}$ forms a renewal process (RP) (Feller, 1949), then $\{X_t,\ t = 1, 2, \dots\}$ is called a Geometric Process (GP), and the real number a is called the ratio of the GP.

If we denote the mean and variance for the latent RP $\{Y_t\}$ as

$$E(Y_t) = \mu \quad \text{and} \quad Var(Y_t) = \sigma^2,$$

the mean and variance for the observed data $\{X_t\}$ which form a GP are given by

$$E(X_t) = \mu/a^{t-1} \quad \text{and} \quad Var(X_t) = \sigma^2/a^{2(t-1)}, \tag{1.1}$$

respectively. If $a > 1$, a GP is stochastically decreasing; if $0 < a < 1$, it is stochastically increasing; and if $a = 1$, it is stationary.

The original GP model mainly focused on modelling X_t as the inter-arrival times of a series of events in reliability and maintenance problem in system engineering (Lam, 2007). In health science, Chan *et al.* (2006) first extended the GP model to allow for multiple trends at

different stages of development for the Severe Acute Respiratory Syndrome (SARS) epidemic in Hong Kong in 2003. More studies of the GP model on Poisson count times series with the applications to clinical trials can also be found in Wan and Chan (2009, 2011) and Chan *et al.* (2011). The GP model was also extended to binary data by Chan and Leung (2010) with an application to methadone clinic data in clinical trial.

In statistical inference, Lam (1992) first proposed the non-parametric least square error (LSE) method. Besides, by adopting some lifetime distributions to the RP $\{Y_t\}$, the model can also be implemented by a parametric approach. Lam and Chan (1998) investigated the statistical inference and properties of the maximum likelihood (ML) estimators for the GP models with the lognormal distribution and Chan *et al.* (2004) considered gamma distribution.

The GP model provides a new modelling approach for trend data. Lam *et al.* (2004) showed that the GP model out-performed the Cox-Lewis model (Cox and Lewis, 1966), the Weibull process model (Ascher, 1981) and the homogeneous Poisson process model. The original GP model assumes a constant mean μ and a constant ratio a over time. These assumptions reduce the flexibility in modelling time series. Chan *et al.* (2006) extended the GP model to include threshold effects for multiple trend data to model, in particular, the growing, stabilising and declining stages of the SARS epidemic. Wan and Chan (2009, 2011)

and Chan and Leung (2010) considered covariate effects in both the mean μ_t and ratio a_t functions which change across time t. Under this assumption, the unobserved $\{Y_t\}$ is no longer a RP but a stochastic process (SP) in general.

Despite these researches have extended the GP model considerably in terms of modelling techniques, methods of inference and fields of applications, few have focused on financial time series. The merit of the GP model in financial application is two folds: its geometric structure and a ratio component to model non-stationarity. Firstly, the model assumes that the observed data at time t form a latent 'detrended' SP after discounted $t-1$ times by a ratio parameter and the GP model focuses on modelling the latent SP. In fact, the latent SP forms the state parameter of a state space model and hence it can achieve model robustness by suitably adopting some parametric distributions such as lognormal, gamma and Weibull distributions. Chan *et al.* (2011) has shown the overdispersion property of a Poisson GP model with a Poisson data distribution.

Secondly, the ratio a_t measures the direction and strength of a dynamic trend movement whereas the mean μ_t of the SP indicates the level of the underlying stationary process. By separately modelling the ratio and the mean of the SP, the model separates the effects on trend movement from the effects on the underlying SP that generates the observed GP. This approach is natural and appealing. Moreover,

equation (1.1) shows that the ratio affects both the mean and variance of a GP and allows them to change coherently over time.

In summary, the GP model has shown many advantages over some common time series models, like the GARCH and SV models, as it provides a simple modelling approach to model simultaneously the trend movement and heteroskedasticity in a financial time series. It is definitely worthy of investigation in the developments of GP model for financial time series.

1.4. Objective and structure of thesis

This thesis extends the GP model to capture the additional features including autocorrelation, jump diffusion and heavy-tailed data distribution, etc, in a financial time series. The extended models can facilitate accurate forecasting. The methods of inference including the LSE, ML and Bayesian are also exploited to implement the proposed models.

To address the autoregressive feature of a financial time series, the GP model is firstly extended to the autoregressive geometric process (ARGP) model in which an autoregressive (AR) term is added to the mean of the latent SP. To allow for nonlinearity and covariant effects, the time-variant mean function adopts covariate effects and components which incorporate threshold effects on time and on some

lagged outcomes. Moreover, the ratio function a_t of the GP model describes the trend movement and permits the volatility to change over time. The model called the conditional autoregressive geometric process (CARGP) range model is further extended from the ARGP model by adopting a conditional autoregressive range (CARR)-type (Chou, 2005) mean function in order to capture the dynamics of the serial correlation better. By adopting the log-t distribution expressed in scale mixtures representation for the SP, the model is more robust than its normal counterpart in modelling leptokurtic time series. With the fact that jumps occur in many financial time series, we propose further extensions of the CARGP model by adding threshold and jump effects in the mean function of the SP. These two components capture mean reversion, together with the log-t distribution for the SP and the ratio function, and allow for the volatility clustering and heteroskedasticity. We show that the extended CARGP Threshold (CARGPT) and Jump (CARGPJ) models are the generalisation of the MRJD model of Rambharat *et al.* (2005).

The LSE and ML methods using the non-parametric and parametric approaches, respectively are used for the inference of the ARGP models. In the past decade, the Bayesian method becomes very popular. It has an advantage over the ML method as it avoids the evaluation of complicated likelihood functions, particularly when the mean of the GP model adopts a CARR-type mean function. Moreover, it has had

increasing popularity in recent years due to the advancement of computational power and the development of efficient sampling techniques. Furthermore, the Bayesian software, Bayesian analysis Using Gibbs Sampling for Window version; (WinBUGS; Spiegelhalter *et al.*, 2004) allows non-experts to implement the model and performs forecast fairly easily. Hence, we will then focus on the implementation of the CARGP, CARGPT and CARGPJ models using the Bayesian method.

For the remaining parts of this thesis, they are structured as below. The ARGP model using the LSE and ML methods of inference is presented in Chapter 2. The CARGP model with a CARR-type mean function using the Bayesian approach is introduced in Chapter 3. Further extensions of the CARGP model to adopt threshold and jump effects are described in Chapter 4. We will apply the proposed models to real data sets in each of the above chapters. Finally, the thesis is concluded in Chapter 5 by a discussion on the advantages and limitations of the proposed GP models in modelling financial time series. Some future research areas are proposed to advance the GP model and enrich its financial applications.

Chapter 2

Autoregressive Geometric Process (ARGP) Model

2.1. Introduction

Most of the financial time series have several noticeable features. They include trend movement, autocorrelation, heteroskedasticity, leverage effect and jump diffusion, etc. Among these features, heteroskedasticity is perhaps the most prominent feature that attracts more research attention. In financial markets, volatility has become a standard risk measure. Accurate forecasting of volatility is important but difficult, because financial time series often exhibit time-varying volatility and volatility clustering. They are periods of elevated volatility interspersed among more tranquil periods. As discussed in Section 1.2, two main classes of models are derived to capture the dynamics of the volatility precisely, and they are the generalised autoregressive conditional heteroskedastic (GARCH) models (Bollerslev, 1986) and the stochastic volatility (SV) models (Hull and White, 1987).

Essentially, the GARCH and SV models are return-based models as they are constructed using the data of closing prices, neglecting all intra-day price movement. Parkinson (1980) stated that the range

which is the difference between the intra-day high-low prices is an efficient proxy of volatility. This volatility estimator is based on the assumption that the asset price follows a driftless geometric Brownian motion, and is theoretically shown by Parkinson (1980) to be around 5 times more efficient than the classical estimator based on closing prices. Garman and Klass (1980) who included more historical data in the volatility estimator shown that it is 8 times more efficient. Beckers (1983) also concluded that the daily price range contained important information concerning the stock price variability. Although Garman and Klass (1980), and Marsh and Rosenfeld (1986) remarked that the extreme-value estimators are downward biased due to the discreteness in prices and even worst with relatively low trading volumes as shown in the estimator of Wiggins (1991). Wiggins (1992) demonstrated that it is still significantly more efficient than the classical estimator for some time series, like the S&P 500 futures price series which has a high trading volume and takes values over a more continuous range. This is because higher trading volume usually associates with higher continuity in trading price. Kunitomo (1992) improved the Parkinson's method by allowing a drift in the geometric Brownian motion process. Andersen and Bollerslev (1998) reported the favourable explanatory power of range data. The more recent studies, including Alizadeh *et al.* (2002) who used the ranged-based estimator for the SV model, Brandt and

Jones (2006) who obtained a better volatility forecast using the range-based exponential GARCH (EGARCH) model, Chou (2005, 2006) and Chen *et al.* (2008) also found strong support on the models using a range-based estimation for volatility. Hence, range, instead of closing prices, is adopted as the outcome variable in this chapter for better modelling and forecasting of volatility.

As mentioned, a financial time series often includes trend movement, and exhibits volatility changes. The Geometric Process (GP) model has shown distinct features to capture both dynamics simultaneously via the inherent geometric structure and the ratio parameter. To further allow for the rather high serial dependency, we first extend the GP model to the autoregressive GP (ARGP) model in which a first order autoregressive (AR) term and possibly some other covariates are added to the mean μ_t of the latent stochastic process (SP). Secondly, nonlinearity arises when there is a change in the trend movement and/or in the dynamics of the latent SP. As a result, threshold model can be considered when the model differs in some important way, for example, a shift in the mean after the occurrence of a certain event, such as, the global financial crisis in September 2008. This threshold model is called the threshold time ARGP (TARGP) model. In addition, another threshold model called the threshold level ARGP (LARGP) model assumes that the exposure to a risk (threshold) variable, such as the lagged outcome, causes a model shift only when the

risk variable exceeds a certain threshold level. Both types of threshold models will surely describe the dynamics of the range data better. For model implementation, both the non-parametric (NP) least square error (LSE) method and the parametric maximum likelihood (ML) method are used.

The layout of the sections in this chapter is given below. Section 2.2 introduces the model development of ARGP model and the extensions for capturing nonlinearity in a financial time series. Section 2.3 describes the LSE and the parametric ML methods for statistical inference. In Section 2.4, an empirical study is presented using the intra-day range data of the All Ordinaries (AORD) index of the Australian stock market. Finally, a conclusion is given in Section 2.5.

2.2. Model development

2.2.1. The parametric GP model. Assume that $\{X_t\}$ follows a GP such that $\{Y_t = a_t^{t-1} X_t\}$ forms a latent SP with the ratio a_t. Referring to equation (1.1), the mean and variance of the observed data $\{X_t\}$ are given by,

$$E(X_t) = \mu_t / a_t^{t-1} \quad \text{and} \quad Var(X_t) = \sigma_t^2 / a_t^{2(t-1)}, \tag{2.1}$$

respectively, where the mean μ_t and the ratio a_t may change over time in general.

By adopting some lifetime distributions including the lognormal and gamma distributions to the SP $\{Y_t\}$, the model can be implemented using the parametric method. Chan *et al.* (2004) investigated the statistical inference for the GP model with gamma distribution. Assuming that $Y_t \sim Ga(\alpha, \gamma_t)$, the probability density function (pdf) for Y_t is

$$f_{G,t}(y_t) = \frac{1}{\Gamma(\alpha)}\,\gamma_t^{\alpha} y_t^{\alpha-1}\exp(-\gamma_t y_t)\,I(y_t \geq 0),$$

where $\alpha(>0)$ is the shape parameter, $\gamma_t(>0)$ is the scale parameter and $\Gamma(\alpha)$ is the gamma function. The mean and variance of Y_t are

$$E(Y_t) = \mu_t = \frac{\alpha}{\gamma_t} \quad \text{and} \quad Var(Y_t) = \sigma_t^2 = \frac{\alpha}{\gamma_t^2},$$

respectively. Using $f_{G,t}(x_t) = \dfrac{dy_t}{dx_t} f_{G,t}(y_t) = a_t^{t-1} f_{G,t}(y_t)$, the pdf for the observed data X_t is given by

$$f_{G,t}(x_t) = \frac{1}{\Gamma(\alpha)}\,(a_t^{t-1}\gamma_t)^{\alpha} x_t^{\alpha-1}\exp(-a_t^{t-1}\gamma_t x_t)\,I(x_t \geq 0),$$

or equivalently,

$$X_t \sim Ga(\alpha, a^{t-1}\gamma_t).$$

The likelihood and log-likelihood functions for the observed data $\{X_t\}$ are

$$\begin{aligned} L_G(\boldsymbol{\theta}; \boldsymbol{x}) &= \prod_{t=1}^{n} \frac{1}{\Gamma(\alpha)}\,(a_t^{t-1}\gamma_t)^{\alpha} {x_t}^{\alpha-1}\exp(-a_t^{t-1}\gamma_t x_t) \quad \text{and} \\ \ell_G(\boldsymbol{\theta}; \boldsymbol{x}) &= \alpha\sum_{t=1}^{n}\ln\left(a_t^{t-1}\gamma_t\right) + (\alpha-1)\sum_{t=1}^{n}\ln x_t - \sum_{t=1}^{n} a_t^{t-1}\gamma_t x_t - n\ln\Gamma(\alpha), \end{aligned}$$

respectively, where $\boldsymbol{x}$ is a vector of x_t, $t = 1, ..., n$ and $\boldsymbol{\theta}$ is a vector of model parameters including parameters in the mean and ratio functions, and the shape parameter.

Alternatively, Lam and Chan (1998) considered a lognormal distribution, that is,

$$Y_t \sim LN(\upsilon_t, \tau^2). \tag{2.2}$$

Then the pdf for Y_t is

$$f_{LN,t}(y_t) = \frac{1}{\tau\sqrt{2\pi}y_t} \exp\left[-\frac{1}{2\tau^2}(\ln y_t - \upsilon_t)^2\right] I(y_t \geq 0),$$

and the mean and variance of Y_t are

$$E(Y_t) = \mu_t = \exp\left(\upsilon_t + \frac{1}{2}\tau^2\right) \quad \text{and} \quad Var(Y_t) = \sigma_t^2 = \exp\left(2\upsilon_t + \tau^2\right)\left[\exp(\tau^2) - 1\right],$$

respectively. Equivalently, $X_t \sim LN(\upsilon_t - \ln(a_t^{t-1}), \tau^2)$ with the pdf

$$f_{LN,t}(x_t) = \frac{1}{\tau\sqrt{2\pi}x_t} \exp\left\{-\frac{1}{2\tau^2}\left[\ln(a_t^{t-1}\, x_t) - \upsilon_t\right]^2\right\},$$

the mean

$$E(X_t) = \frac{\mu_t}{a_t^{t-1}} = \exp\left[\upsilon_t - \ln(a_t^{t-1}) + \frac{1}{2}\tau^2\right], \tag{2.3}$$

and the variance

$$Var(X_t) = \frac{\sigma_t^2}{a_t^{2(t-1)}} = \exp\left\{2\left[\upsilon_t - \ln(a_t^{t-1})\right] + \tau^2\right\}\left[\exp(\tau^2) - 1\right]. \tag{2.4}$$

The likelihood and log-likelihood functions are given by

$$L_{LN}(\boldsymbol{\theta};\boldsymbol{x}) = \prod_{t=1}^{n} \frac{1}{\tau\sqrt{2\pi}x_t} \exp\left\{-\frac{1}{2\tau^2}[\ln(a_t^{t-1}x_t) - \upsilon_t]^2\right\} \quad \text{and}$$

$$\ell_{LN}(\boldsymbol{\theta};\boldsymbol{x}) = -\frac{n}{2}\ln\tau^2 - \frac{n}{2}\ln(2\pi) - \sum_{t=1}^{n}\ln x_t - \frac{1}{2\tau^2}\sum_{t=1}^{n}[\ln(a_t^{t-1}x_t) - \upsilon_t]^2, \tag{2.5}$$

respectively.

2.2.2. Extension to autoregressive and covariate effects. Original GP model assumes a constant mean μ and a constant ratio a over time such that the latent $\{Y_t\}$ forms a renewal process (RP). The adoption of a homogeneous mean μ over time is over-simplified in many cases. On one hand, it fails to capture the high serial dependency, and on the other, it does not account for other covariate effects which may be highly correlated to the outcome variable. Adopting the framework of the generalised linear models (GLM), the constant mean is replaced by a mean function log linked to a linear function of r covariates $z_{tk},\ k = 1, \ldots, r$ as well as an first order AR term. Assuming that $r = 1$ for simplicity, the mean function for $\ln Y_t$ is written as

$$\upsilon_t = \beta_{\mu 0} + \beta_{\mu 1}\ln(y_{t-1}) + \beta_{\mu 2}\, z_t, \tag{2.6}$$

where $y_{t-1} = a_{t-1}^{t-2}\, x_{t-1}$ since from (2.2), $\ln(Y_t) \sim N(\upsilon_t, \tau^2)$. To allow a flexible trend movement, the ratio a which models the direction and strength of the trend movement can change gradually over time. The time effect in the ratio function can be modelled by different time functions, for example, t or $\ln t$. For a long time series, $\ln t$ is preferred

as it allows the time effect to level off when t becomes large. Other covariates can also be included to allow for their effects on the trend movement. In subsequent analyses, we consider

$$a_t = \exp(\beta_{a0} + \beta_{a1} \ln t). \tag{2.7}$$

The extended model is called the autoregressive GP (ARGP) model, and the vector of model parameters for the ARGP model is $\boldsymbol{\theta} = (\boldsymbol{\beta}, \tau^2)$ where $\boldsymbol{\beta} = (\beta_{\mu 0}, \beta_{\mu 1}, \beta_{\mu 2}, \beta_{a0}, \beta_{a1})$. Dropping the lognormal distribution assumption in non-parametric inference, the general mean function for X_t is given by,

$$E(X_t) \;=\; \exp\left[\upsilon_t - \ln(a_t^{t-1})\right].$$

2.2.3. Extension to threshold time. Particularly when the time series is long, some structural changes may occur over the course of time, so that model shift should be allowed for at some time points $\boldsymbol{\mathcal{T}}$ called the turning points. Chan *et al.* (2006) extended the GP model to the threshold GP (TGP) model by fitting a separate GP to each stage of development, growing, stabilising and declining, of an epidemic as identified by the turning points using the LSE method of inference.

Let $\mathcal{T}_g,\ g = 1, \ldots, G$ be the turning point or threshold time for the g-th GP, n_g be the number of observations for the g-th GP, and $\sum_{g=1}^{G} n_g = n$ such that $\mathcal{T}_1 = 1$ and $\mathcal{T}_g = 1 + \sum_{j=1}^{g-1} n_j,\ g = 2, \ldots, G$. For $\mathcal{T}_g \le t < \mathcal{T}_{g+1}$, $X_t \sim LN(\upsilon_{tg} - \ln(a_{tg}^{t-\mathcal{T}_g}), \tau_g^2)$ with the mean and variance

given by

$$E(X_t) = \frac{\mu_{tg}}{a_{tg}^{t-\mathcal{T}_g}} = \exp\left[\upsilon_{tg} - \ln(a_{tg}^{t-\mathcal{T}_g}) + \frac{1}{2}\tau_g^2\right], \quad \text{and} \tag{2.8}$$

$$Var(X_t) = \frac{\sigma_{tg}^2}{a_{tg}^{2(t-\mathcal{T}_g)}} = \exp\left\{2[\upsilon_{tg} - \ln(a_{tg}^{t-\mathcal{T}_g})] + \tau_g^2\right\}\left[\exp(\tau_g^2) - 1\right], \tag{2.9}$$

respectively, where

$$\upsilon_{tg} \;=\; \beta_{\mu 0g} + \beta_{\mu 1g}\,\ln(y_{t-1}) + \beta_{\mu 2g}\,z_t \tag{2.10}$$

and

$$a_{tg} \;=\; \exp[\beta_{a0g} + \beta_{a1g}\ln(t - \mathcal{T}_g)], \tag{2.11}$$

where $y_{t-1} = a_{t-1,g}^{t-\mathcal{T}_g-1}\,x_{t-1}$.

For non-parametric inference, the mean function for X_t is given by,

$$E(X_t) = \exp\left[\upsilon_{tg} - \ln(a_{tg}^{t-\mathcal{T}_g})\right]. \tag{2.12}$$

This extended model is called the threshold time ARGP (TARGP) model. The vector of model parameters is $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\tau}^2, \boldsymbol{\mathcal{T}})$ where $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_G)$, $\boldsymbol{\beta}_g = (\beta_{\mu 0g}, \beta_{\mu 1g}, \beta_{\mu 2g}, \beta_{a0g}, \beta_{a1g})$, $g = 1, ..., G$, $\boldsymbol{\tau}^2 = (\tau_1^2, \ldots, \tau_G^2)$ and $\boldsymbol{\mathcal{T}} = (\mathcal{T}_1, \ldots, \mathcal{T}_G)$.

2.2.4. Extension to threshold level. Nonlinearity occurs in a financial time series when a certain risk variable, other than time, has an effect on the outcome only, when the risk variable exceeds a certain threshold. Assume that a model shift occurs when an observable lag-d threshold variable, $L_{t-d},\ t = d+1, \ldots, n$, exceeds a certain threshold

level. Let $\mathcal{L}_g$, $g = 1, \ldots, G$ be the latent threshold levels for L_{t-d} with a lower bound $\mathcal{L}_1 = 0$ or $-\infty$ depending on whether L_t is positive continuous or continuous, and a upper bound $\mathcal{L}_{G+1} = \infty$.

When $\mathcal{L}_g \leq \mathfrak{l}_{t-d} < \mathcal{L}_{g+1}$, the mean and variance for X_t are given by (2.8) and (2.9), respectively with $\mathcal{T}_g = 1$.

The extended model is called the threshold level ARGP (LARGP) model and the vector of model parameters is $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\tau}^2, \boldsymbol{\mathcal{L}})$ where $\boldsymbol{\mathcal{L}} = (\mathcal{L}_1, \ldots, \mathcal{L}_G)$. The threshold variable L_t is preferable to be positively correlated to X_t and it should contain market information. Some choices of L_t include certain lagged value of X_t, as well as certain lagged value of an exogenous factor such as the international market movement indices and the interest rates.

2.3. Methodology of inference

2.3.1. Non-parametric inference. LSE method estimates the model parameters by minimising the sum of squared error (SSE) defined as

$$SSE \;=\; \sum_{t=1}^{n} [X_t - E(X_t)]^2,$$

where $E(X_t) = \mu_t / a_t^{t-1}$ for the GP model. Obviously, a small SSE indicates a better fit of the model to the data. To minimise the SSE, we solve the first order derivative equation $SSE' = 0$ for the parameters $\boldsymbol{\beta}$ using the standard Newton Raphson (NR) method. We start the

process with some arbitrary initial value $\boldsymbol{\beta}_0$. The estimates in the iterative procedures will converge and a local solution is found, provided that this initial guess is close enough to the true root, $\hat{\boldsymbol{\beta}}_{LSE}$, such that $SSE'(\hat{\boldsymbol{\beta}}_{LSE}) = 0$. Denote the current parameter estimates in the i^{th} iteration by $\boldsymbol{\beta}^{(i)}$ and the next updated estimates by $\boldsymbol{\beta}^{(i+1)}$, the updating procedure in each NR iteration is:

$$\boldsymbol{\beta}^{(i+1)} = \boldsymbol{\beta}^{(i)} - [SSE''(\boldsymbol{\beta}^{(i)})]^{-1} SSE'(\boldsymbol{\beta}^{(i)}),$$

where SSE' and SSE'' denote the first and second order of derivatives of SSE. The iteration continues until $\parallel \boldsymbol{\beta}^{(i+1)} - \boldsymbol{\beta}^{(i)} \parallel$ becomes sufficiently small. The estimates are now converged and the LSE estimates are given by $\hat{\boldsymbol{\beta}}_{LSE} = \boldsymbol{\beta}^{(i+1)}$.

The first and second order derivatives of SSE, SSE' and SSE'' respectively, as required in the NR procedure are given by

$$\begin{aligned} \frac{\partial SSE}{\partial \beta_{jkg}} &= -2 \sum_{t=1}^{n} z'_{jkt}\, \hat{x}_t\, (x_t - \hat{x}_t)\, I_t, \quad \text{and} \\ \frac{\partial^2 SSE}{\partial \beta_{j_1 k_1 g_1} \partial \beta_{j_2 k_2 g_2}} &= -2 \sum_{t=1}^{n} z'_{j_1 k_1 t}\, z'_{j_2 k_2 t}\, \hat{x}_t\, (x_t - 2\hat{x}_t)\, I_t, \end{aligned}$$

respectively, where $j, j_1, j_2 = \mu, a$; $k, k_1, k_2 = 0, 1, 2$; $g, g_1, g_2 = 1, \ldots, G$; $z'_{\mu 0t} = 1$, $z'_{\mu 1t} = \ln y_{t-1}$, $z'_{\mu 2t} = z_t$, $z'_{a0t} = -(t-1)$, $z'_{a1t} = -(t-1)\ln t$, $I_t = 1$ for the ARGP model, $I_t = I(\mathcal{T}_g \le t < \mathcal{T}_{g+1})$ for the TARGP model and $I_t = I(\mathcal{L}_g \le \mathfrak{l}_{t-d} < \mathcal{L}_{g+1})$ for the LARGP model, and $\hat{x}_t = E(X_t)$ is given by (2.3) with $\tau^2 = 0$ for the ARGP model, and by (2.8) with $\tau_g^2 = 0$ for the TARGP and LARGP models. The NR

iterative procedures are implemented in R. Denote the element in the u^{th} row and v^{th} column of $[SSE''(\hat{\boldsymbol{\beta}}_{LSE})]^{-1}$ by s_{uv}, the standard error of each estimate, $\hat{\beta}_{LSE,u}$ is given by the square root of s_{uu}.

The mean squared error (MSE) which is defined as the average of SSE is one of the measures for the goodness-of-fit (GOF) of a model and is used in this chapter for model assessment.

2.3.2. Parametric inference. ML method is a classical methodology in parametric inference. Although the ML method requires more distribution assumptions than the NP inference, it has more statistical power. Lam and Chan (1998) and Chan *et al.* (2004) showed that the ML method using the lognormal and gamma distributions, respectively performed better than the NP method in parameter estimation.

Taking the lognormal distribution as an example, the ML estimates $\hat{\boldsymbol{\theta}}_{ML}$ which maximise the log-likelihood function in (2.5) can be obtained again using the NR iterative method by solving $\ell'_{LN}(\boldsymbol{\theta};\boldsymbol{x}) = 0$. Writing $\ell(\boldsymbol{\theta}) = \ell_{LN}(\boldsymbol{\theta};\boldsymbol{x})$ for simplicity, the NR iterative procedure is given by

$$\boldsymbol{\theta}^{(i+1)} = \boldsymbol{\theta}^{(i)} - [\ell''(\boldsymbol{\theta}^{(i)})]^{-1}\ell\,'(\boldsymbol{\theta}^{(i)}), \tag{2.13}$$

and is updated iteratively until $\boldsymbol{\theta}^{(i+1)}$ converges for sufficiently large i. The ML estimates are given by $\hat{\boldsymbol{\theta}}_{ML} = \boldsymbol{\theta}^{(i+1)}$, and the first and second

order derivatives, $\ell'(\boldsymbol{\theta})$ and $\ell''(\boldsymbol{\theta})$, as required in (2.13) are given by

$$\begin{aligned}
\frac{\partial \ell(\boldsymbol{\theta})}{\partial \beta_{jkg}} &= \frac{1}{\tau_g^2}\sum_{t=1}^{n} z'_{jkt}\left[\ln(a_t^{t-1}x_t) - \upsilon_t\right] I_t, \\
\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_{j_1k_1g_1}\partial \beta_{j_2k_2g_2}} &= -\frac{1}{\tau_g^2}\sum_{t=1}^{n} z'_{j_1k_1t}\, z'_{j_2k_2t}\, I_t, \\
\frac{\partial \ell(\boldsymbol{\theta})}{\partial \tau_g^2} &= -\frac{n_g}{2\tau_g^2} + \frac{1}{2\tau_g^4}\sum_{t=1}^{n}[\ln(a_t^{t-1}x_t) - \upsilon_t]^2\, I_t, \\
\frac{\partial \ell(\boldsymbol{\theta})}{\partial \beta_{jkg}\partial \tau_g^2} &= \frac{1}{\tau_g^4}\sum_{t=1}^{n} z'_{jkt}\left[\ln(a_t^{t-1}x_t) - \upsilon_t\right] I_t, \\
\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \tau_g^2\partial \tau_g^2} &= \frac{n_g}{2\tau_g^4} - \frac{1}{\tau_g^6}\sum_{t=1}^{n}[\ln(a_t^{t-1}x_t) - \upsilon_t]^2\, I_t,
\end{aligned}$$

respectively, where υ_t and a_t are given by (2.6) and (2.7) for the ARGP model, and (2.10) and (2.11) for the TARGP and LARGP models.

To conduct hypothesis testing for a certain parameter, the large sample properties for the ML estimates $\hat{\boldsymbol{\theta}}_{ML}$ are derived using the Lindeberg-Levy Central Limit Theorem.

Lindeberg-Levy Central Limit Theorem :

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{ML} - \boldsymbol{\theta}) \xrightarrow{D} N(\mathbf{0}, n^{-1}\boldsymbol{\Sigma}),$$

where $\xrightarrow{D}$ indicates the convergence in distribution when n is large and $\boldsymbol{\Sigma}$ is the covariance matrix given by $-E[\ell''(\boldsymbol{\theta})]^{-1}$. The standard error of each estimate $\hat{\theta}_{ML,u}$ is given by the square root of s_{uu} where s_{uv} is the element in the u^{th} row and v^{th} column of the matrix $-[\ell''(\hat{\boldsymbol{\theta}}_{ML})]^{-1}$.

To assess the performance of each model in parametric inference, we adopt the Akaike's Information Criterion (AIC) (Akaike, 1973) given

by

$$AIC = 2 \times p - 2 \times \ell(\hat{\boldsymbol{\theta}}_{ML}),$$

where $\ell(\hat{\boldsymbol{\theta}}_{ML})$ is the log-likelihood function evaluated at the maximum likelihood estimates, and p is the number of independent parameters in the model.

To estimate other model parameters such as the threshold time $\mathcal{T}_g$ and threshold level $\mathcal{L}_g$, a direct way is to conduct a grid search condition on G. For example, for the TARGP model, we first set $G = 2$ and search $\mathcal{T}_2$ over a certain interval not too close to the end points 1 and n. Condition on each threshold time $\mathcal{T}_2$, two GP models are fitted to data with time $t < \mathcal{T}_2$ and $t \geq \mathcal{T}_2$, respectively. The optimal $\mathcal{T}_2$ is chosen to minimise either the MSE for the NP inference, or the AIC for parametric method. Then G is set to 3 and the search for $\mathcal{T}_3$ given $\mathcal{T}_2$ is similarly repeated from the remaining time points not too close to $\mathcal{T}_2$, 1 and n. The number of threshold times G can be chosen by the MSE or AIC.

To estimate the threshold levels $\mathcal{L}_g$ in the LARGP model, we first set $G = 2$. The search for the threshold level $\mathcal{L}_2$ is similar to that of the threshold time $\mathcal{T}_2$, but we make further condition on d. Setting $d = 1$, we search $\mathcal{L}_2$ over certain interval which is approximately the median to the 95 percentile of the risk variable L_t because the threshold level should be large in general. For each threshold value $\mathcal{L}_2$ in the interval, two GP models are fitted to data x_t with $\mathfrak{l}_{t-1} < \mathcal{L}_2$ and $\mathfrak{l}_{t-1} \geq \mathcal{L}_2$

respectively. Then we set $d = 2$ and fit two GP models for each $\mathcal{L}_2$ similarly. The optimal $\mathcal{L}_2$ is chosen to minimise the MSE or AIC over the median to 95 percentile of $\mathfrak{l}_{t-d}$ for a range of d. Then we may set $G = 3$ and repeat the search for $\mathcal{L}_3$ again.

Essentially the LSE and ML methods are partial LSE and ML methods because the shape parameter and parameters in the mean and ratio functions are estimated using the LSE and ML methods whereas the remaining parameters are estimated by a grid search.

2.4. Empirical Study

2.4.1. The data. We analyse the intra-day high-low prices from the All Ordinaries (AORD) Index of Australia stock market. The data is collected from 1 May 2006 to 30 April 2009 (n=763) in the website `finance.yahoo.com`. The variable of interest is the daily range X_t which is the differences between the log of the daily maximum price, $P_{max,t}$ and minimum price, $P_{min,t}$ indices defined as

$$X_t = [\ln(P_{max,t}) - \ln(P_{min,t})] \times 100. \tag{2.14}$$

As suggested in Chou (2005), the lag-one daily log return $Z_{t-1} = [\ln(P_{c,t-1}) - \ln(P_{c,t-2})] \times 100$, where $P_{c,t}$ is the closing price on day t, is taken as a covariate to allow for the leverage effect. Summary statistics and three test statistics for X_t, $\ln X_t$, Z_t and $|Z_t|$ are reported in Table 2.1. The first statistic, Ljung-Box Q_{12}, tests the overall randomness

of a time series based on 12 lagged autocorrelations in this application. The second and third statistics, the Cramér-von Mises W and Jarque-Bera JB, test for the normality in the data. W compares the empirical distribution with the hypothesized distribution while the JB measures the departure from normality based on the sample kurtosis and skewness. From Table 2.1, all tests are significant, showing non-randomness and non-normality, except that Z_t is random and $\ln X_t$ is normal, confirming the lognormal assumption for X_t. Furthermore, the time series plots of X_t and Z_{t-1} and the histogram of X_t are presented in Figure 2.1. The summary statistics in Table 2.1 and the histogram in Figure 2.1(b) show that X_t is leptokurtic and skewed. Moreover, the correlation between X_t and Z_t ($\hat{\rho}_{X,Z}$=-0.241) and their plots in Figure 2.1(a) show that the leverage effect is present in the data, because the daily range is high or the price is volatile when the return is low, particularly during the period of the global financial tsunami which started in October 2008.

The autocorrelation function (ACF) plot in Figure 2.2 shows that the daily range X_t of AORD index has strong correlation and hence the ARGP model with the mean function given by (2.6) is applied to allow for the strong autocorrelation. For the ratio function, we adopt a time-varying ratio in (2.7) for the ARGP model, and in (2.11) for the TARGP and LARGP models.

2.4.2. Numerical results. The ARGP model is first fitted to the data using both the LSE and ML methods. As Figure 2.1 shows that there is a general increasing trend before October 2008 that marks the beginning of the global financial tsunami, and a sharp decreasing trend thereafter. The TARGP model is then fitted to the data. In the TARGP model, the number of turning points, G is estimated by a grid search by first setting $G = 2$ with $\mathcal{T}_1 = 1$ and $\mathcal{T}_2$ being searched over a period of time, $t = 200$ to 700. The range is set to be wide enough to avoid subjective bias. From Table 2.2, the threshold time in the TARGP model is estimated to be 10 October 2008 ($\hat{\mathcal{T}}_2$=625) by the LSE method and 6 October 2008 ($\hat{\mathcal{T}}_2$=621) by the ML method. They marks the beginning of the nonlinear effects due to the global financial tsunami at early October 2008. Then another TARGP model is attempted with $G = 3$ in which $\mathcal{T}_2$ and $\mathcal{T}_3$ are sampled from 19 September 2006 to 10 September 2007 and 1 February 2008 to 12 February 2009, respectively. The two turning points are estimated to be 4 July 2007 ($\hat{\mathcal{T}}_2 = 302$) and 24 December 2008 ($\hat{\mathcal{T}}_3 = 678$), and they mark three trends which are stationary ($\hat{a}_1 = 1.000$), increasing ($\hat{a}_2 = 0.998$) and decreasing ($\hat{a}_3 = 1.003$), respectively. However, the AIC shows no improvement. These explain why $G = 2$ is suitable for this data.

From Figure 2.1(a), the jump dynamics seem to be very different across the time threshold $\mathcal{T}_2$ in October 2008. Before the time threshold, there are some obvious jumps in February 2007, August 2007,

January 2008 and July 2008, but not such kind of jumps after the time threshold. Hence, a separate LARGP model is fitted to each period before and after October 2008. We call this model the threshold LARGP (TLARGP) model. Again, we set $G = 2$ and search $\mathcal{L}_2$ over a range from 1.0 to 3.5 which correspond to the median (1.16) and 95 percentile (3.39) approximately, for lag $d = 1$, 2 and 3. Since the autocorrelation often drops with increasing lags, the optimal lag will not be very large. Experience shows that the optimal lag is often related to the lag with high autocorrelation. The lag-1 to 3 autocorrelations r_j for X_t are (0.57, 0.56, 0.55) which drop slowly but they are higher than the correlation between X_t and Z_t. From Table 2.2, $\mathcal{L}_2$ are estimated to be relatively large. Hence increasing G to 3 seems unnecessary. Moreover, especially in the period of volatile market, the daily range will cross the threshold levels $\mathcal{L}_g$ too often across time. To enable consistency for some effects including the autoregressive effect and trend pattern across time, we set $\beta_{\mu 1g}$, β_{a0g} and β_{a1g} to be the same across g.

Table 2.2 reports the parameter estimates, standard errors (in parenthesis), MSE and AIC for applying the daily range data to the models. Most parameter estimates are significant, except two cases: (1) $\beta_{\mu 0}$ before the threshold time in both TARGP and TLARGP models, and (2) β_{a0} and β_{a1} in the TLARGP model using the LSE method. In both TARGP and TLARGP models, $\beta_{\mu 1}$ are positive for the period before October 2008 but negative after that period, showing that the

autocorrelation is different across these two periods. A negative autocorrelation arises when the data are more volatile after October 2008. Since the sign of $\beta_{\mu 2}$ is consistently negative across models and methods of inference, the leverage effect is present. The general positive sign of β_{a0} and negative sign of β_{a1} show that the trend is decreasing at a decreasing rate, that is, the series show a decreasing trend, then a stationary trend and finally an increasing trend from December 2006, because a_{tg} in (2.11) changes from greater than 1 to less than 1. The only exception is the trend which is increasing at a decreasing rate for the ARGP model using the LSE method. For the TARGP and TLARGP models which consider threshold time in the models, the series follow a straightly decreasing trend after this threshold time, as a_{tg} is greater than 1. However, while this estimated downward trend captures the real downward trend closely, it fails to describe the enlarged volatility during this period while compared to the ARGP model.

Figure 2.3 plots the fitted lines, trends and threshold levels for the ARGP, TARGP and TLARGP models using both LSE and ML methods. It demonstrates graphically the trends described in the above paragraph. From Figure 2.3(a), the trend line μ_1/a_t^{t-1} of the ARGP model using the LSE method is higher than that using the ML method, and increasing all the way, because the jumps affect the parameters using the LSE method more, resulting in a more obvious upward trend. In Figure 2.3(b), the trend lines drop sharply and then slowly after the

threshold time. Moreover, the trend by the LSE method is much flatter than in the ARGP model after allowing model shift in time. Figure 2.3(c) shows that the fitted values in the TLARGP model are similar to that in the TARGP model by the ML method. However, the fitted values of the TLARGP model by the LSE method seem better than in the TARGP model at some points, especially during start of the downward trend period after the threshold time.

According to the MSE, the TARGP models improve over the ARGP models by 19.6% and 18.4% using the LSE method and ML method, respectively. Moreover, the TLARGP models improve over the TARGP models further by 8.2% and 3.2% for the LSE and ML methods, respectively. Further to the AIC in Table 2.2, it shows that the TLARGP model is the best out of the three models due to the lowest AIC.

Apart from the model fit, the model performance is also evaluated by comparing the fitted distribution with the hypothesized distribution. Figure 2.4 displays the histograms of the standardised residuals,

$$S_t = \frac{\ln(X_t) - [\upsilon_t - \ln(a_t^{t-1})]}{\tau} \sim N(0,1)$$

for the three models using the ML method, superimposed on the hypothesized density functions. In general, the distribution of S_t from all the three models is close to the hypothesized standard normal density, except perhaps the middle position where S_t shows a slightly higher density, and the right tail. In summary, the model performance is satisfactory for all the three models.

2.4.3. Forecasting. The TLARGP model using both LSE and ML methods are used to forecast 50 days of daily ranges from 1 May 2009 to 10 July 2009. The forecasts of X_t, $t = 764, \ldots, 813$, are based on the expected value $E(X_t)$ in (2.8) using the same set of parameter estimates as given by Table 2.2 and x_{t-1} as required in (2.10) is substituted by observed value in the forecasting period. Figure 2.3(c) includes these forecast values of $E(X_t)$. The forecast values $E(X_t)$ and the corresponding 95% prediction interval (PI)

$$\Big(\exp(\upsilon_t - \ln a_t^{t-1} + \Phi^{-1}(0.025)\,\tau^2),\, \exp(\upsilon_t - \ln a_t^{t-1} + \Phi^{-1}(0.975)\,\tau^2) \Big)$$

by the ML method are also displayed in the enlarged figure, Figure 2.5. From the result, 86% of PIs cover the true values. In general, the forecast values using the LSE method are smaller than those using the ML method. Although the forecast period lies on the downward trend of the TLARGP model, the data and so are the forecast values, using both methods, do not show an obvious downward trend. The MSE of the forecast values using the LSE method (0.2248) is as well smaller than that using the ML method (0.2495).

2.5. Conclusion

In this chapter, we demonstrate the ability of the GP model in modelling the trend movement and heteroskedasticity of a financial time series. In particular, we show that the extended ARGP models

can further allow for serial correlation in a financial time series. The leverage effect is accounted for by adding the lag-1 return as a covariate in the mean function. Besides, the TARGP and TLARGP models can capture well the nonlinear effects in the mean across a threshold in time or a threshold level of the outcome variable at a certain lag respectively. Result from the TARGP models shows that the threshold time $\hat{\mathcal{T}}_2$ marks the beginning of the financial crisis in October 2008 accurately. The models are implemented using both the LSE and the classical ML method, and R is used to perform the optimisation using the NR procedures in the LSE and ML methods. In summary, the proposed ARGP model and its extensions are simple, easy to implement and give reliable estimates of the mean and volatility dynamics. However, there is still room for further improvement to capture other features in a financial time series.

Although the TLARGP models allow for nonlinearity via a model shift as detected by a risk variable, the model is incapable of capturing the jumps when the market is highly volatile, particularly on 5 August 2008 when the jump is sudden. Further study is required to extend the model to capture the characteristics of these outliers. One approach is to consider heavy-tailed distribution, like log-t distribution for model robustness. Another approach is to consider a jump model which will be presented in Chapter 4. To enable more flexible

modelling of the serial dependency in a time series, the conditional autoregressive range (CARR) model can be incorporated into the mean function of the ARGP models. These extensions are considered in the next chapter.

Table 2.1. Summary statistics for the AORD stock market daily range data.

	Range X_t	Ln range $\ln(X_t)$	Return Z_t	Absolute return $\lvert Z_t \rvert$
Mean	1.4311	0.0723	-0.0471	1.0509
Median	1.1572	0.0634	0.0081	0.7943
SD	0.9908	0.2649	1.4823	1.0457
Kurtosis	7.0649	-0.2866	3.8305	8.3178
Skewness	2.1410	0.2131	-0.5032	2.3368
Range	7.8271	1.4980	13.914	8.5536
Minimum	0.2568	-0.5904	-8.5536	0.0000
Maximum	8.0839	0.9076	5.3601	8.5536
Ljung-Box, Q_{12}	2416	2838	20.22 †	721.7
Cramér-von Mises, W	5.647	0.110 †	1.337	6.374
Jarque-Beta, JB	2170	8	499	2894

† p-value > 0.05. All other p-values are less than 0.02.

Table 2.2. Parameter estimates, standard errors in *italics* and model assessment measures for the AORD data.

	Model	T	$\beta_{\mu 0}$	$\beta_{\mu 1}$	$\beta_{\mu 2}$	β_{a0}	β_{a1}	τ^2	GOF^
LSE	ARGP	-	-0.3014	0.4131	-0.0461	-0.0070	0.0008	-	0.5725
			0.1202	*0.0341*	*0.0077*	*0.0022*	*0.0003*	-	
	TARGP	$t < 625$	-0.0899†	0.3841	-0.0631	0.0038†	-0.0008	-	0.4604
			0.1178	*0.0456*	*0.0128*	*0.0028*	*0.0004*	-	
		$t \geq 625$	1.8764	-0.2891	-0.0377	0.0325	-0.0050	-	
			0.0792	*0.0563*	*0.0096*	*0.0084*	*0.0017*	-	
	TLARGP	$x_{t-2} < 1.9$	-0.1329†	0.2954	-0.0605	0.0013†	-0.0004†	-	0.4228
		85.26%*	*0.1189*	*0.0475*	*0.0224*	*0.0028*	*0.0004*	-	
		$x_{t-2} \geq 1.9$	0.1989†	-	-0.0727	-	-	-	
		14.74%*	*0.1274*	-	*0.0129*	-	-	-	
	$t \geq 625$	$x_{t-3} < 3.6$	1.6222	-0.2897	-0.0267	0.0196	-0.0027†	-	
		86.71%*	*0.1044*	*0.0546*	*0.0186*	*0.0088*	*0.0018*	-	
		$x_{t-3} \geq 3.6$	1.8802	-	-0.0472	-	-	-	
		13.29%*	*0.0767*	-	*0.0102*	-	-	-	
ML	ARGP	-	-0.1486	0.3434	-0.0544	0.0047	-0.0009	0.1942	0.5869
			0.0638	*0.0335*	*0.0109*	*0.0015*	*0.0002*	*0.0099*	*1181*
	TARGP	$t < 621$	-0.0463†	0.2878	-0.0671	0.0109	-0.0019	0.1887	0.4792
			0.0692	*0.0378*	*0.0142*	*0.0019*	*0.0003*	*0.0107*	*1095*
		$t \geq 621$	1.7042	-0.1423†	-0.0251	0.0352	-0.0055	0.1113	
			0.1517	*0.0782*	*0.0127*	*0.0104*	*0.0020*	*0.0132*	
	TLARGP	$x_{t-3} < 2.5$	-0.0465†	0.2126	-0.1008	0.0107	-0.0018	0.1732	0.4638
		93.39%*	*0.0505*	*0.0291*	*0.0120*	*0.0014*	*0.0002*	*0.0102*	*1060*
		$x_{t-3} \geq 2.5$	0.3383	-	0.0000†	-	-	0.2397	
		6.61%*	*0.0750*	-	*0.0225*	-	-	*0.0536*	
	$t \geq 621$	$x_{t-2} < 1$	1.9472	-0.1458	-0.0184†	0.0347	-0.0054	0.0243†	
		4.20%*	*0.0923*	*0.0337*	*0.0405*	*0.0044*	*0.0009*	*0.0140*	
		$x_{t-2} \geq 1$	1.7068	-	-0.0269	-	-	0.1127	
		95.80%*	*0.0648*	-	*0.0055*	-	-	*0.0136*	

† p-value > 0.05.

* proportion of lagged data below or above the threshold level.

^ The first GOF measure is MSE and the second GOF in italics is AIC.

FIGURE 2.1. (a) Observed daily range X_t (black) and log-return Z_t (grey) of the AORD stock market price from 1 May, 2006 to 30 April, 2009 and (b) histogram of X_t.

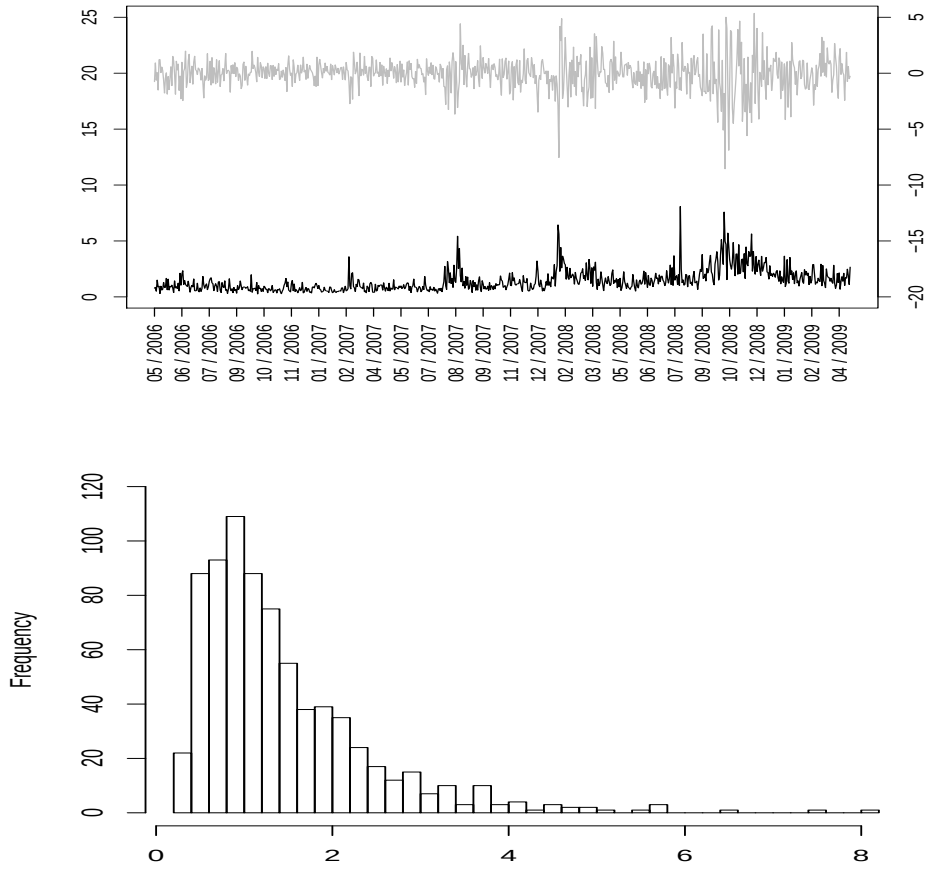


Figure 2.2. Autocorrelation function (ACF) of observed daily range X_t of AORD stock market price.

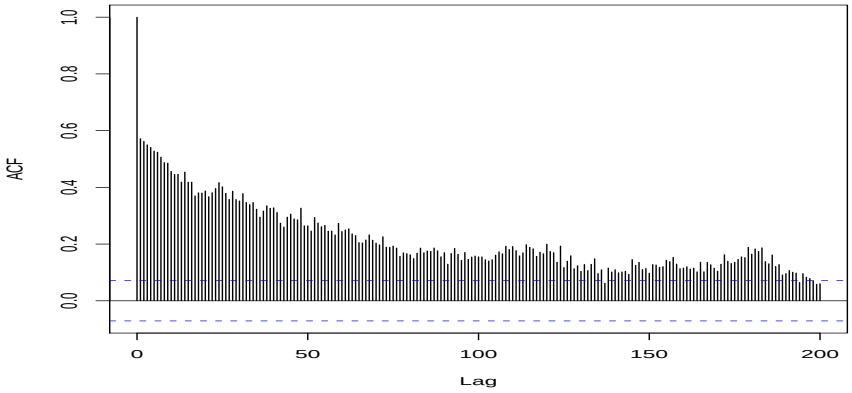


Figure 2.3. Observed (black) and fitted values of the (a) ARGP, (b) TARGP and (c) TLARGP models using the LSE (red) and ML (green) methods with forecast (orange for LSE and blue for ML methods), threshold levels (horizontal dotted) and trends (inclined dotted).

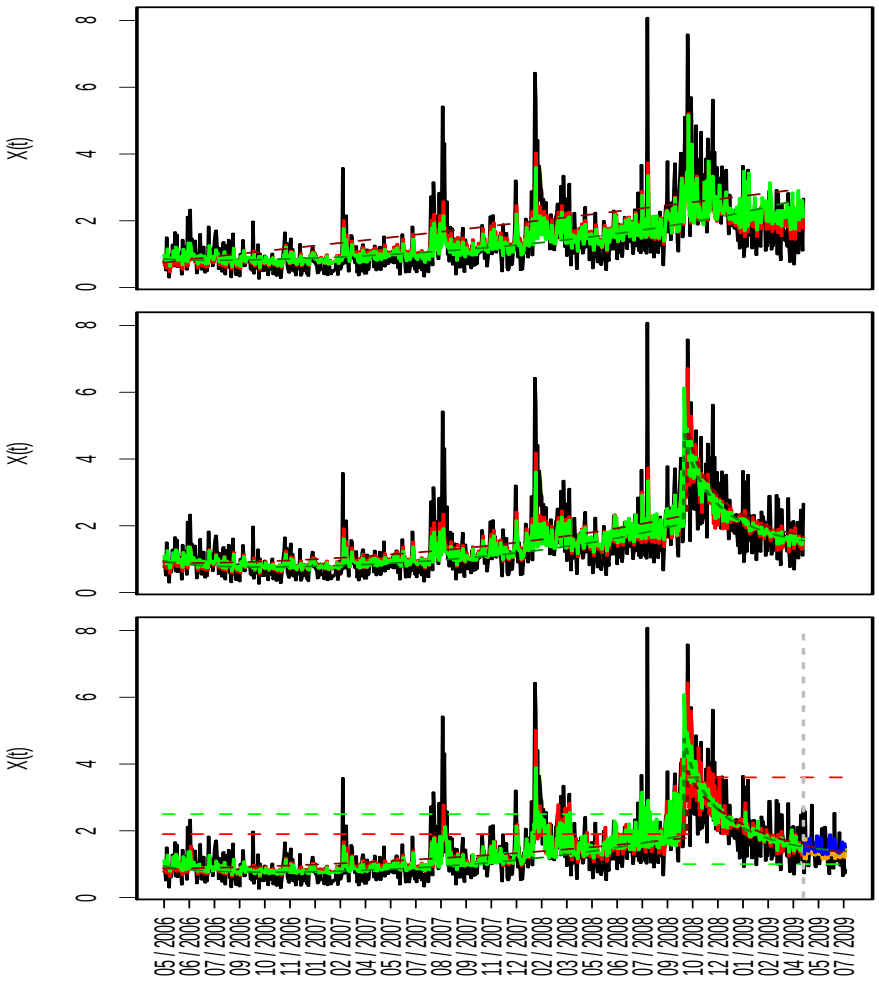


FIGURE 2.4. Comparison of observed and hypothesized distributions for standardised X_t between ARGP, TARGP and TLARGP models using ML method.

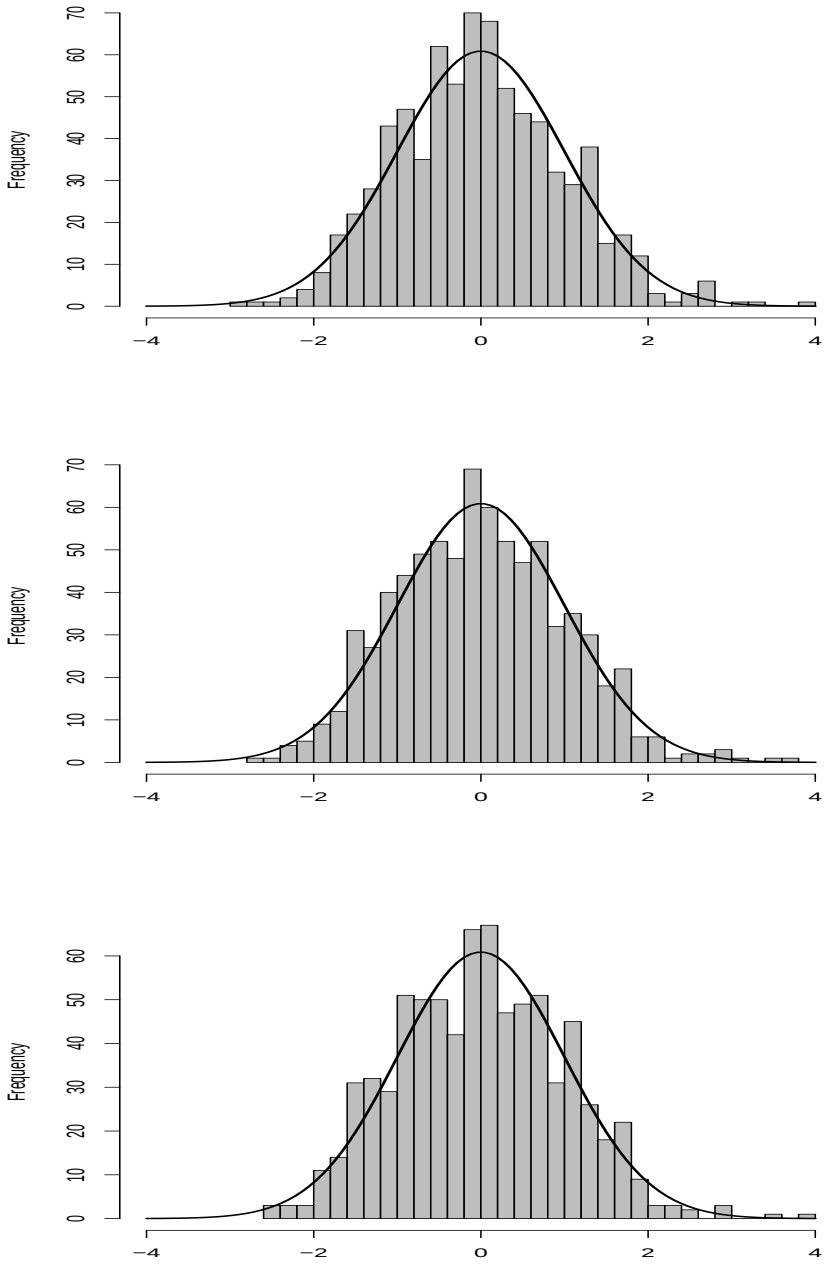
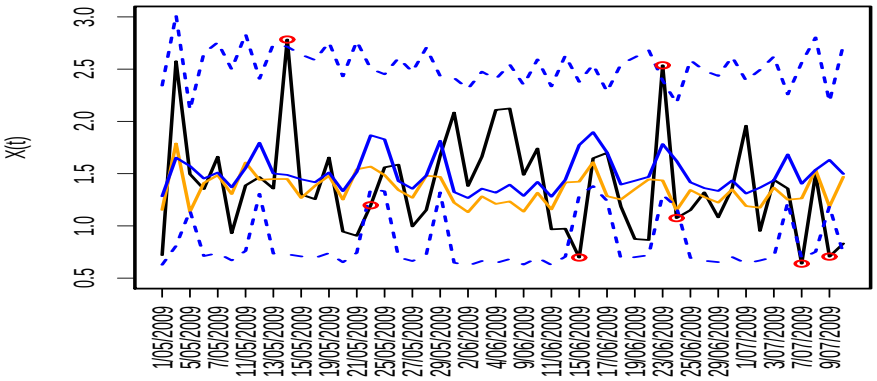


Figure 2.5. Forecast values using the TLARGP models by LSE method (orange) and ML method (blue) with 95% PI of the forecast values (blue dotted), and observed values outside PI of the forecast values (red circle).



Chapter 3

Conditional Autoregressive Geometric Process (CARGP) Model

3.1. Introduction

Chapter 2 has demonstrated the ability of Geometric Process (GP) model in capturing some common features of a financial time series, that includes trend movement, autocorrelation and heteroskedasticity. On top of these, we will show in this chapter, its strength in modelling a financial time series which has leptokurtic distribution. As mentioned in the last chapter, recent research has proposed using daily ranges to construct estimates of daily return volatility since daily ranges are known to be a more efficient measure of return volatility (see Parkinson, 1980; Andersen and Bollerslev, 1998; Alizadeh *et al.*, 2002) than daily returns.

Following the idea of Chou (2005), this chapter extends the modelling strategy of the GP model to the dynamic conditional autoregressive range (CARR) model for range data, to obtain a simple yet highly efficient model for capturing the dynamics of the volatility. In particular, the mean of the stochastic process (SP) is assigned a CARR-type mean function, and the extended model is called the conditional

autoregressive geometric process (CARGP) model. By incorporating lagged returns in the mean function, the model can capture leverage effects or volatility asymmetry, which refers to the negative return sequences associated with an increase in volatility of the stock returns. The strength of the CARGP model lies in its flexibility to adapt the dynamics of the volatility using the CARR-type mean function and the trend movement specified by the ratio parameter in the GP model. The proposed model is further extended to accommodate a model shift after some time points called thresholds, and is distinguished from the regime switching model, where the changes occur when the outcomes exceed certain threshold levels. Parameter estimation in threshold autoregressive (TAR) models is usually performed in two approaches: classical likelihood approach (Tong and Lim, 1980 and Tong, 1990) and Bayesian approach (Geweke and Terui, 1993, and Chen and Lee, 1995).

In this chapter, we adopt the Bayesian approach using Markov chain Monte Carlo (MCMC) algorithms, and we apply the Metropolis-Hastings algorithm to estimate the threshold time jointly with other model parameters. A variety of model structures and error distributions can be considered to provide a tailor-made analysis (Chiu and Wang, 2006). For robustness considerations, a heavy-tailed distribution such as Student's t-distribution is considered, and it is expressed in the scale mixture representation to allow a simpler Gibbs sampler

for model implementation and to enable outlier diagnosis (Choy and Chan, 2008).

For readers who are less familiar with Bayesian computation techniques, the WinBUGS (Bayesian analysis Using Gibbs Sampling; Spiegelhalter *et al.*, 2004) package provides a user-friendly means to perform Bayesian inference in a relatively simple program. We adopt the Bayesian approach using MCMC algorithms and implement the CARGP model using WinBUGS package.

This chapter is structured as follows. Section 3.2 introduces the extension of ARGP model to the CARGP model. Section 3.3 describes the Bayesian computational methods for statistical inference. Section 3.4 presents a simulation result to illustrate the performance of the CARGP model. In Section 3.5, a CARGP model is again fitted to the intra-day range data of the Australian All Ordinary stock market index (AORD). Models are comparied using both in-sample and out-of-sample model assessment. Finally, the conclusion is made in Section 3.6.

The shortened version of this chapter is in press as : J.S.K. Chan, C.P.L. Lam, P.L.H. Yu, S.T.B. Choy, C.W.S. Chen, A Bayesian conditional autoregressive geometric process model for range data. *Computational Statistics and Data Analysis (2011)*, doi: 10.1016/j.csda.2011.01.006.

3.2. Model development

3.2.1. The CARGP model. To specify a dynamic structure in the mean function that describes the persistence of market shocks to the range of prices, Chou (2005) proposed the following CARR(p, q) model for some positive continuous measurements, X_t, for example, the range data, as below:

$$\begin{aligned} x_t &= \mu_t \, \epsilon_t, \\ \mu_t &= \beta_0 + \sum_{j=1}^{p} \beta_{1j} \, \mu_{t-j} + \sum_{j=1}^{q} \beta_{2j} \, x_{t-j}, \\ \epsilon_t | \Im_{t-1} &\sim f(.|\Im_{t-1}), \end{aligned} \tag{3.1}$$

where $\Im_{t-1}$ is the set of information up to time $t-1$ and $f(.|\Im_{t-1})$ is the conditional density for the errors ϵ_t with unit mean. The stationary condition for the process is

$$C = \sum_{j=1}^{p} \beta_{1j} + \sum_{j=1}^{q} \beta_{2j} < 1, \tag{3.2}$$

where C determines the persistence of range shocks and the unconditional (long-term) mean of X_t is $\beta_0/(1-C)$. Chou (2005) adopted the Weibull distribution with exponential distribution as a special case. If the range X_t follows a Weibull distribution, that is, $X_t \sim W(\psi_t, \alpha)$, where ψ_t and α are the scale and shape parameters, respectively, $\psi_t = \mu_t / \Gamma(1 + \frac{1}{\alpha})$, and $\Gamma(\cdot)$ is a gamma function, the pdf for X_t is,

$$f_{w,t}(x_t) = \frac{\alpha}{\psi_t} \left(\frac{x_t}{\psi_t} \right)^{\alpha - 1} \exp\left(- \frac{x_t}{\psi_t} \right)^{\alpha}.$$

The mean and variance are given by μ_t and

$$\sigma_t^2 = \mu_t^2 \left[\frac{\Gamma(1+\frac{2}{\alpha})}{\Gamma^2(1+\frac{1}{\alpha})} - 1 \right],$$

respectively.

However, this CARR model does not allow for trend movement explicitly. To remedy this, we introduce the GP model and equate the mean function (3.1) to υ_t in (2.3) as

$$\upsilon_t = \beta_{\mu 0} + \sum_{j=1}^{p} \beta_{\mu 1 j}\, \upsilon_{t-j} + \sum_{j=1}^{q} \beta_{\mu 2 j}\, \ln(y_{t-j}),$$

where $y_{t-j} = a_{t-j}^{t-j-1}\, x_{t-j}$. The extended model combining the modelling approaches of the GP and CARR techniques is called the conditional autoregressive geometric process (CARGP(p,q)) model.

3.2.2. Extension to heavy-tailed distributions. Many financial data are heavy tailed, and hence the lognormal distribution is replaced by the log-t distribution to achieve a robust analysis. To facilitate efficient Bayesian MCMC computation and outlier diagnostics, the t-distribution is expressed as a scale mixtures of normal (SMN) distributions. Andrews and Mallows (1974) studied the class of SMN distributions, and Choy and Chan (2008) investigated different scale mixture distributions. Amongst the SMN distributions, the Student's t-distribution is the most popular member. Student's t-distribution with location μ, scale σ, and number of degrees of freedom ν has the

following SMN representation:

$$t_\nu(y|\,\mu,\sigma) \;=\; \int_0^\infty N\left(y\,\middle|\,\mu,\frac{\sigma^2}{\lambda}\right) G\left(\lambda\,\middle|\,\frac{\nu}{2},\frac{\nu}{2}\right) d\lambda$$

which can be expressed hierarchically as

$$Y|\mu,\sigma,\lambda \sim N\left(\mu,\frac{\sigma^2}{\lambda}\right), \qquad \text{and} \qquad \lambda \sim G\left(\frac{\nu}{2},\frac{\nu}{2}\right),$$

where $G(\alpha,\gamma)$ denotes the gamma distribution with mean α/γ. This hierarchical form is particularly useful in Bayesian modelling where the Gibbs sampler can be substantially simplified without a heavy computational cost. In addition, the posterior mean or median of the scale mixture variable λ can be acted as a proxy to identify potential outliers. See Choy and Chan (2008) for details.

In the CARGP model, we assume that X_t follows a GP such that the latent $Y_t = a_t^{t-1}X_t$ forms a detrended SP and a_t given by (2.7) in general is the ratio parameter of the GP. In particular, if $a_t = a = 1$, Y_t forms a renewal process (RP). In Chapter 2, we assume that $\ln Y_t \sim N(\upsilon_t,\tau^2)$. For model robustness, we assume, in this chapter, that $\ln Y_t \sim t_\nu(\upsilon_t,\tau^2)$ or $\ln Y_t|\lambda_t \sim N(\upsilon_t,\frac{\tau^2}{\lambda_t})$ by conditioning on λ_t. Hence $X_t|\lambda_t \sim LN(\upsilon_t - \ln(a_t^{t-1}),\frac{\tau^2}{\lambda_t})$ and the mean and variance of X_t are given by

$$E(X_t) = \frac{\mu_t}{a_t^{t-1}} = \exp\left[\upsilon_t - \ln(a_t^{t-1}) + \frac{\tau^2}{2\lambda_t}\right], \tag{3.3}$$

and

(3.4)
$$Var(X_t) = \frac{\sigma_t^2}{a_t^{2(t-1)}} = \exp\left\{2[\upsilon_t - \ln(a_t^{t-1})] + \frac{\tau^2}{\lambda_t}\right\}\left[\exp\left(\frac{\tau^2}{\lambda_t}\right) - 1\right],$$

respectively.

3.2.3. Incorporation of covariate and threshold time effects. As the daily range may evolve over time subject to certain external effects, exogenous variables Z_{tj} should be incorporated into the conditional mean function μ_t of the CARGP model via υ_t as

$$\upsilon_t = \beta_{\mu 0} + \sum_{j=1}^{p} \beta_{\mu 1 j}\, \upsilon_{t-j} + \sum_{j=1}^{q} \beta_{\mu 2 j}\, \ln(y_{t-j}) + \sum_{j=1}^{r} \beta_{\mu 3 j}\, z_{tj}. \tag{3.5}$$

Chou (2005) suggested the use of lagged return, trading volume, and seasonal factors as the exogenous variable, z_{tj}. Because a negative relationship is often found between the range and the lagged return, suggesting a leverage effect that a decrease in return leads to higher volatility, the leverage effect can be allowed for by including lagged returns as a covariate. As a positive relationship is often present between the range and the trading volume, trading volume can be another potential exogenous variable.

If we set $p = q = r = 1$ and drop the redundant subscript j in β in (3.5), it becomes

$$\upsilon_t = \beta_{\mu 0} + \beta_{\mu 1}\, \upsilon_{t-1} + \beta_{\mu 2}\, \ln(y_{t-1}) + \beta_{\mu 3}\, z_t, \tag{3.6}$$

for $t = 2, \ldots, n$, and $\upsilon_1 = \beta_{\mu 0} + \beta_{\mu 3} z_1$ for $t = 1$. This function can be rewritten as

$$\upsilon_t = \beta_{\mu 0} \sum_{i=1}^{t} \beta_{\mu 1}^{i-1} + \beta_{\mu 2} \sum_{i=2}^{t} \beta_{\mu 1}^{i-2} \ln(y_{t-i+1}) + \beta_{\mu 3} \sum_{i=1}^{t} \beta_{\mu 1}^{i-1} z_{t-i+1}, \tag{3.7}$$

showing the complexity of parameter $\beta_{\mu 1}$ in υ_t. Note that the stationary condition $C < 1$ in (3.2) does not apply to (3.6) with a log link function, as shown in (3.3). However, we find that the sum of parameters in υ_t is less than 1 for most of the models reported in Table 3.2 in the empirical study. When $a_t = a = 1$, X_t, which is just Y_t, is neither increasing nor decreasing, and the stationary constraint in (3.2) does not apply too.

On the other hand, the CARGP model can be extended to allow for multiple trends to describe different stages of development, the growing stage ($a < 1$), stabilising stage ($a = 1$) and declining stage ($a > 1$), for a certain event. In this case, the constant ratio a in (1.1) is replaced by a time-dependent ratio function log linked to a function of covariates as in equation (2.7) in section 2.2.2, to allow a flexible trend movement.

To incorporate threshold times effect in the CARGP model, we assume that $X_t|\lambda_t \sim LN(\upsilon_{tg} - \ln(a_g^{t-\mathcal{T}_g}), \frac{\tau_g^2}{\lambda_t})$ for $\mathcal{T}_g \leq t < \mathcal{T}_{g+1}$, where

$$\upsilon_{tg} = \beta_{\mu 0 g} + \sum_{j=1}^{p} \beta_{\mu 1 j g}\, \upsilon_{t-j,g} + \sum_{j=1}^{q} \beta_{\mu 2 j g} \ln(y_{t-j}) + \sum_{j=1}^{r} \beta_{\mu 3 j g}\, z_{tj}, \tag{3.8}$$

where $y_{t-j} = a_g^{t-\mathcal{T}_g-j}\, x_{t-j}$. The mean and variance for X_t become

$$E(X_t) = \frac{\mu_{tg}}{a_g^{t-\mathcal{T}_g}} = \exp\left[\upsilon_{tg} - \ln(a_g^{t-\mathcal{T}_g}) + \frac{\tau_g^2}{2\lambda_t}\right], \tag{3.9}$$

and

(3.10)
$$Var(X_t) = \frac{\sigma_g^2}{a_g^{2(t-\mathcal{T}_g)}} = \exp\left\{2[\upsilon_{tg} - \ln(a_g^{t-\mathcal{T}_g})] + \frac{\tau_g^2}{\lambda_t}\right\}\left[\exp\left(\frac{\tau_g^2}{\lambda_t}\right) - 1\right],$$

respectively.

Tiwari *et al.* (2005) estimated the number of turning points using different model selection criteria. In applications, the number of turning points G and the range from which each turning point $\mathcal{T}_g$ is sampled are determined by examining the empirical time series. In general, the best model among models with $G = 1, 2, 3\ldots$ thresholds can be selected based on some model selection criterion such as the Bayesian Information Criterion (BIC) and the Deviance Information Criterion (DIC) (Section 3.5.3).

The vector of all model parameters is $\boldsymbol{\theta} = (\boldsymbol{\beta}_{\mu 1}, \ldots, \boldsymbol{\beta}_{\mu G}, \boldsymbol{\beta}_{a1}, \ldots, \boldsymbol{\beta}_{aG}, \tau_1^2, \ldots, \tau_G^2)$ where $\boldsymbol{\beta}_{\mu g} = (\beta_{\mu 0g}, \beta_{\mu 11g}, \ldots, \beta_{\mu ijg}, \ldots, \beta_{\mu 3rg})$, $i = 1, 2, 3$, $j = 1, \ldots, k_i$, $k_1 = p$, $k_2 = q$, $k_3 = r$, and $\boldsymbol{\beta}_{ag} = (\beta_{a0g}, \beta_{a1g})$, $g = 1, \ldots, G$. For models without threshold effect, we write $G = 1$, and the subscript g is dropped, and for model with a constant ratio a_g, $a_g = \exp(\beta_{ag})$.

3.3. Bayesian Inference

The first and second order derivatives of the log-likelihood function as required in the classical likelihood approach are difficult to evaluate because υ_t in (3.7) is a complicated function of $\beta_{\mu 1}$. On the other hand,

the Bayesian method differs from the classical method as the prior probability or belief is used. Apart from the data, Bayesian method makes use of both prior and observed data information to calculate the posterior distribution $f(\boldsymbol{\theta}|\,\boldsymbol{x})$ of the parameters $\boldsymbol{\theta}$ given the observed data $\boldsymbol{x}$. Based on the Bayes theorem, the posterior distribution is equal to

$$\begin{aligned} f(\boldsymbol{\theta}|\,\boldsymbol{x}) &= \frac{f(\boldsymbol{x}|\,\boldsymbol{\theta})f(\boldsymbol{\theta})}{\int f(\boldsymbol{x}|\,\boldsymbol{\theta})f(\boldsymbol{\theta})d\boldsymbol{\theta}} \\ &\propto f(\boldsymbol{x}|\,\boldsymbol{\theta})f(\boldsymbol{\theta}), \end{aligned}$$

where $f(\boldsymbol{\theta})$ is the prior distribution and $f(\boldsymbol{x}|\,\boldsymbol{\theta})$ is the likelihood.

With insufficient prior information, non-informative prior with large variance is often used, just letting the data to speak for the posterior distribution.

The Bayesian approach using MCMC techniques converts an optimisation problem into a sampling problem, by simulation of a single or block of model parameters iteratively, conditional on other parameters and the data. The Gibbs sampling algorithm (Smith and Roberts, 1993; Gilks *et al.*, 1996) and Metropolis Hastings algorithm (Hastings, 1970; Metropolis *et al.*, 1953) are the most popular MCMC techniques that produce samples from the intractable posterior distributions. For readers who are less familiar with Bayesian computation techniques, we recommend using the WinBUGS (Bayesian analysis Using Gibbs Sampling) package. See Spiegelhalter *et al.* (2004).

In the simulation and empirical studies, different CARGP models are compared, and vague and non-informative priors are assigned to the model parameters. The Bayesian hierarchy for the CARGP models (Models 1-4) is

$$\text{Data:} \qquad X_t \sim LN(\upsilon_t - \ln(a^{t-1}), \tfrac{\tau^2}{\lambda_t}).$$

$$\text{Priors:} \quad a \sim U(0.95, 1.05), \quad \beta_{\mu ij} \sim N(0, \sigma_\beta^2), \quad \tau^2 \sim IG(\alpha_\tau, \gamma_\tau),$$

$$\lambda_t \sim G(\tfrac{\nu}{2}, \tfrac{\nu}{2}), \quad \nu \sim G(\alpha_\nu, \gamma_\nu)\, I(1, 30),$$

where $I(a, b)$ indicates a truncated distribution with support (a, b) and $\lambda_t = 1$ for Model 1. With a ratio function a_t (Model 4), the priors are $\beta_{ai} \sim N(0, \sigma_\beta^2),\ i = 0, 1$. The Bayesian hierarchy for the threshold CARGP model (Model 5) is

$$\text{Data:} \qquad X_t \sim LN(\upsilon_t - \ln(a_g^{t-\mathcal{T}_g}), \tfrac{\tau_g^2}{\lambda_t})\, I(\mathcal{T}_g \le t < \mathcal{T}_{g+1}).$$

$$\text{Priors:} \quad a_g \sim U(0.95, 1.05), \quad \beta_{\mu ijg} \sim N(0, \sigma_\beta^2), \quad \tau_g^2 \sim IG(\alpha_\tau, \gamma_\tau),$$

$$\lambda_t \sim G(\tfrac{\nu_g}{2}, \tfrac{\nu_g}{2}), \quad \nu_g \sim G(\alpha_\nu, \gamma_\nu)\, I(1, 30), \quad \mathcal{T}_g \sim U(c_g, d_g),$$

where $\mathcal{T}_g$ is assigned a discrete uniform prior on the range $[c_g, d_g]$. Lastly, the Bayesian hierarchy for the CARR model with covariate using the Weibull distribution (Models 6 ($\alpha = 1$) and 7) is

$$\text{Data:} \qquad X_t \sim W(\mu_t/\Gamma(1 + \tfrac{1}{\alpha}), \alpha).$$

$$\text{Priors:} \quad \alpha \sim G(c, d),\ \beta_{\mu 0} \sim N(0, \sigma_\beta^2),\ \beta_{\mu 1} \sim U(0, 1),\ \beta_{\mu 2} \sim U(0, 1 - \beta_{\mu 1}),$$

$$\beta_{\mu 3} \sim N(0, \sigma_\beta^2).$$

The hyperparameter σ_β^2 is set to be very large whereas α_τ, γ_τ, α_ν, γ_ν, c, and d are set to zero for non-informative priors. The full conditional distributions for the parameters in Model 5 are derived and reported in the Appendix to facilitate the MCMC sampler. In the Gibbs sampling scheme, a single Markov chain is run for 7000 iterations, discarding the initial 5000 iterations as the burn-in period to ensure convergence of parameter estimates. Convergence is also carefully checked by the history and autocorrelation function (ACF) plots. Simulated values from the Gibbs sampler after the burn-in period are taken to mimic a random sample of size 2000 from the joint posterior distribution for posterior inference. Parameter estimates are given by the posterior means or medians. To check if the posterior samples of 2000 iterations are sufficient, longer chains of 5000 iterations after burn-in are run for Models 1 and 2, and they give estimates similar to those from 2000 iterations. Moreover, the ACFs and history plots show that the posterior samples are quite uncorrelated. The computation time depends on the complexity of the model and the power of computer, and it is around 4 hours using a Core 2 Duo 2 GHz PC for fitting the CARGP models in the empirical study.

3.4. Simulation study

In this simulation study, we compare the model performance for models fitted to data of different sizes (small or medium) and adopted

different data distributions (lognormal or log-t) and trend patterns (increasing or decreasing). We simulate $N = 100$ data sets; each contains $n = 200$ or $n = 700$ observations. Two models using lognormal (LN) and log-t (LT) distributions are considered, and each model adopts two sets of parameters with decreasing (set 1) and increasing (set 2) trends. Table 3.1 reports the mean and standard deviation (SD) of the parameter estimates over $N = 100$ replications as given by

$$\hat{\theta} = \frac{1}{N}\sum_{i=1}^{N}\hat{\theta}_i, \quad \text{and} \quad SD = \left[\frac{1}{N-1}\sum_{i=1}^{N}(\hat{\theta}_i - \hat{\theta})^2\right]^{1/2},$$

respectively, where $\hat{\theta}_i$ is the posterior mean of θ in the i-th replication. The performance of the proposed models is further evaluated via three criteria: the absolute percentage bias (APB), root mean square error (RMS) and coverage percentage (CP), defined as

$$\begin{aligned} APB &= \left|\frac{\hat{\theta} - \theta}{\theta}\right|, \\ RMS &= \left[\frac{1}{N}\sum_{i=1}^{N}(\hat{\theta}_i - \theta)^2\right]^{1/2}, \qquad \text{and} \\ CP &= \frac{100}{N}\sum_{i=1}^{N} I[\theta \in (\hat{\theta}_{i,0.025}, \hat{\theta}_{i,0.975})], \end{aligned}$$

respectively, where $(\hat{\theta}_{i,0.025}, \hat{\theta}_{i,0.975})$ is the 95% credible interval of θ in the i-th replication and $I(E)$ is an indicator function for the event E. Models with smaller SD, APB and RMS and with CP closer to 95 are preferred.

From Table 3.1, the parameter estimates are close to their true values except $\hat{\nu}$ (both sets) and $\hat{\beta}_{\mu 1}$ (set 2) when $n = 200$. The complex function of $\beta_{\mu 1}$ in the mean function (3.7) explains the difficulty of estimating $\beta_{\mu 1}$ precisely. As for the number of degrees of freedom ν, it is well known that the shape of the t-distribution is rather insensitive to moderate to large numbers of degrees of freedom. However, both estimates improve substantially when the sample size increases to $n = 700$. The CP ranges from 80% to 100% for all CARGP models, showing satisfactory coverage. There is no obvious difference in model performance between models showing a decreasing (set 1) or increasing (set 2) trend, nor between models adopting a lognormal or log-t distribution. Generally speaking, the results in the simulation study are satisfactory when n=200 and are excellent when n=700.

3.5. Empirical Study

3.5.1. The data. We apply the daily range of intra-day high-low prices, X_t which defined in (2.14) from the All Ordinaries (AORD) index of the Australian stock market in Chapter 2, to the CARGP models.

3.5.2. Model selection. The basic CARGP(1,1) model with lognormal (Model 1) and log-t distributions (Model 2) are first utilised, and Model 2 is preferred according to the DIC because the heavier

tails of the log-t distribution can accommodate outliers. The BIC is slightly larger due to the rather heavy penalty for an additional parameter. Hence, the log-t distribution is adopted in all subsequent CARGP models. By setting $a = 1$, the trend movement is not modelled, similar to the CARR model (Model 2.1), but it adopts a log-t instead of Weibull distribution with a log link function and uses the Bayesian approach in parameter estimation.

To describe different levels of persistence, the CARGP(1,2) and CARGP(2,1) models are considered. However, the CARGP(2,1) model has a technical problem in the implementation and β_{22} in the CARGP(1,2) model (Model 3) is insignificant, showing that the basic CARGP(1,1) model describes the market persistence effect well. Moreover, both the BIC and DIC of Model 3 show no improvement either. Hence the basic CARGP(1,1) model is adopted hereafter.

To allow for the leverage effect, Z_t is added to υ_t as an exogenous variable. Moreover, Models 1 to 3 are restricted to monotone trend data. Figure 2.1(a) of Chapter 2 shows that the monotonic increasing trend applies only till the global financial tsunami in October 2008, and decreases thereafter. To allow a flexible trend movement, a ratio function a_t in (2.7) of Chapter 2 is adopted in Model 4 and a threshold time effect in Model 5. Moreover, we set $G = 2$ and the range for sampling $\mathcal{T}_2$ to be [610, 630], which covers the period from 19 September 2008 to 17 October 2008 for Model 5.

Lastly, the CARR models in (3.1) with the covariate Z_t in the mean μ_t using exponential (Model 6) and Weibull (Model 7) distributions and the Bayesian approach are also fitted for model comparison. Table 3.2 reports the posterior mean and the posterior standard error (in italics) of the model parameters, together with two model assessment criteria for Models 1 to 7.

3.5.3. Model assessment. To compare Models 1 to 7, the Bayes factor, BIC, and DIC (Spiegelhalter *et al.*, 2002) are often used in Bayesian analysis. However, the former is often commented on as being too difficult to calculate, especially for models that involve many random effects, large numbers of unknowns, or improper priors (Ntzoufras, 2009). Alternatively, the BIC and DIC are defined as

$$\begin{aligned} BIC &= -2\ln f(\boldsymbol{y}|\boldsymbol{\theta}) + p\,\ln n, \qquad \text{and} \\ DIC &= \overline{D(\boldsymbol{\theta})} + p_D, \end{aligned} \tag{3.11}$$

respectively, are adopted to approximate the Bayes factor. Both criteria contain two components: a measure of model fit and a penalty for model complexity, where $f(\boldsymbol{y}|\boldsymbol{\theta})$ is the likelihood function, $\overline{D(\boldsymbol{\theta})} = E_{\theta|y}[D(\boldsymbol{\theta})]$ is the posterior expectation of the deviance, and p_D is the effective number of parameters defined as the difference between the posterior mean of deviance and the deviance evaluated at the posterior mean of parameters; that is,

$$p_D = E_{\theta|y}(D(\boldsymbol{\theta})) - D(E_{\theta|y}(\boldsymbol{\theta})) = \overline{D(\boldsymbol{\theta})} - D(\bar{\boldsymbol{\theta}}).$$

Clearly, the model with the smallest BIC and/or DIC values is preferred. The BIC and DIC values for Models 1 to 7 are presented in Table 3.2.

Moreover, five more measures, namely the root mean squared error (RMS), mean absolute error (MAE), coverage percentage (CP), width of the 95% confidence interval (CI) for $E(X_t)$ (CI(EX)) and width of the 95% CI for X_t (CI(X)), are defined as

$$\begin{aligned}
\text{RMS}_{ih} &= \left[\frac{1}{n_h}\sum_{t=1}^{n_h}(\text{MR}_{t+s_h,i} - \hat{X}_{t+s_h})^2\right]^{1/2}, \\
\text{MAE}_{ih} &= \frac{1}{n_h}\sum_{t=1}^{n_h}|\text{MR}_{t+s_h,i} - \hat{X}_{t+s_h}|, \\
\text{CP}_h &= \frac{1}{n_h}\sum_{t=1}^{n_h} I(X_{t+s_h} \in (\text{CI}_{X_{t+s_h},low},\ \text{CI}_{X_{t+s_h},up})), \\
\text{CI}(EX)_h &= \frac{1}{n_h}\sum_{t=1}^{n_h}(\text{CI}_{E(X_{t+s_h}),up} - \text{CI}_{E(X_{t+s_h}),low}), \\
\text{CI}(X)_h &= \frac{1}{n_h}\sum_{t=1}^{n_h}(\text{CI}_{X_{t+s_h},up} - \text{CI}_{X_{t+s_h},low}),
\end{aligned}$$

where $h = 0$ indicates the in-sample estimation with start $s_0 = 0$, $h = 1$ indicates the out-of-sample forecast with start $s_1 = n_1 = 763$, $i = 1$ indicates the measure of range $\text{MR}_{t,1} = X_t$, and $i = 2$ indicates $\text{MR}_{t,2} = |Z_t|$ as a proxy of X_t (Chou, 2005). The standardised variables when $X_t \sim LN(\omega_t, \varsigma_t)$, where $\omega_t = \upsilon_t - \ln a_t^{t-1}$ and $\varsigma_t^2 = \frac{\tau^2}{\lambda_t}$, and when $X_t \sim W(\psi_t, \alpha)$, where $\psi_t = \mu_t/\Gamma(1 + \frac{1}{\alpha})$, are

(3.12)

$$S_{LT,t} = \frac{\ln(X_t) - \omega_t}{\varsigma_t} \sim N(0,1) \quad \text{and} \quad S_{W,t} = (X_t/\psi_t)^\alpha \sim \text{Exp}(1),$$

respectively. Hence the corresponding 95% CIs ($\mathrm{CI}_{X_t,low}$, $\mathrm{CI}_{X_t,up}$) for X_t are

$$\Big(\exp(\omega_t + \Phi^{-1}(0.025)\,\varsigma_t),\, \exp(\omega_t + \Phi^{-1}(0.975)\,\varsigma_t)\Big) \tag{3.13}$$

and

$$\Big([-\ln(0.975)]^{1/\alpha}\psi_t,\, [-\ln(0.025)]^{1/\alpha}\psi_t\Big), \tag{3.14}$$

respectively, where $\Phi(\cdot)$ is the standard normal distribution function. On the other hand, the CIs for $E(X_t)$ are obtained from the 2.5 and 97.5 percentiles of the posterior sample for $E(X_t)$, where $E(X_t)$ is given by (3.3) and (3.1), respectively, for the log-t and Weibull distributions. The first three criteria measure the accuracy of the model while the last two measure the precision of the CIs. Models with smaller values of these criteria except CP_h are preferred. For CP_h, it should be close to 95%. Table 3.3 reports these measures together with the three test statistics Q_{12}, W and JB. The standardised variables $S_{LT,t}$ and $S_{E,t}$ are used to test the log-t and Weibull data distributions, respectively.

3.5.4. Numerical results. The results in Table 3.2 show that the parameter estimates are qualitatively consistent across the models. In particular, the ratio a for Models 1 to 3 is less than 1 and significant, showing a general monotonic increasing trend. Since both $\beta_{\mu 11}, \beta_{\mu 21} > 0$ and are significant in all models, a persistence effect is present in the data. Moreover, τ^2 decreases across Models 1 to 5, showing an increase in model robustness while the number of degrees of freedom is around

20, indicating a moderate tail effect. Since Model 2 shows a better model fit than Model 2.1 ($a = 1$) according to both the BIC and DIC (1125 versus 1128 for the BIC and 1095 versus 1102 for the DIC), after allowing for model complexity, the superiority of the CARGP model in allowing trend movement is clear.

For Models 4 and 5, as $\beta_{\mu 31}$ is significant and negative, a leverage effect is present in the data. Parameters β_{a0} and β_{a1} in Model 4 show that a_t changes from greater than 1 to less than 1, indicating a mild and short decreasing trend followed by an increasing trend thereafter. Moreover, the substantially larger DIC for Models 1 and 4 with a gamma distribution (1223 and 1167 for the BIC and 1200 and 1130 for the DIC) supports the assertion in Section 2.4 that lognormal and log-t distributions give better fits than a gamma distribution. For Model 5, the ratios a_1 and a_2 show an increasing trend before 7 October 2008 ($t = 622$) and a decreasing trend thereafter. The market volatility increases sharply from September 2008 to the maximum on 7 October 2008, during which the market price dropped continuously. The trends of the mean, $E(X_t)$ in (3.3) and (3.9), variance, $Var(X_t)$ in (3.4) and (3.10) and the ratio $1/a_t^{t-1}$ for Models 4 and 5, as plotted in Figures 3.1(a) and 3.2(a), respectively, show that the mean and variance capture the volatility clustering well.

The 95% CIs for X_t in (3.13) are displayed in Figures 3.1(b) and 3.2(b). The coverage percentages (CPs), 96.5% and 96.3%, are reasonably close to 95%. Because there is no obvious volatility clustering after 7 October 2008, no outliers are detected, and $\beta_{\mu 112}$ and $\beta_{\mu 212}$ are both insignificant in Model 5, leading to a rather smooth trend after 7 October 2008. Obviously the significance of $\beta_{\mu 11}$ and $\beta_{\mu 21}$ for other models with a monotonic increasing trend is due to the volatility clustering before 7 October 2008. Moreover, even though there is no stationary constraint for Y_t, the sum of parameters in the mean function υ_t in (3.5) is less than 1 for all the models reported in Table 3.2. This stationary condition is only violated by the second-stage model in Model 5.

The results for Models 6 and 7 are qualitatively the same as for Models 1 to 5. The shape parameter α is estimated to be 2.186 for the Weibull distribution. Using both the BIC and DIC, the CARR model using an exponential distribution (Model 6) is far from satisfactory but the model using the Weibull distribution (Model 7) is still no better than any of the CARGP models (Models 1-5), because the CARGP models accommodate the trend effect and adopt the more robust log-t distribution. The trends of the mean and variance for Model 7 are plotted in Figure 3.3(a) and the 95% CI for X_t in (3.14) is plotted in Figure 3.3(b). Again, the mean and variance capture the volatility clustering well, but the lower bound of the CI is very close to zero, revealing the

characteristic of the Weibull distribution with higher density around zero when α is small.

Tests using Q_{12}, W and JB show that all the standardised residuals $S_{it}, i = LT, W$ in (3.12) are non-random and do not follow the hypothesized distribution except Model 5. Figure 3.4 displays the histograms of $S_{LT,t}$ for Models 4 and 5 and $S_{W,t}$ for Model 7 superimposed on their hypothesized density functions. Again, the distribution of $S_{LT,t}$ from Model 5 is closest to the hypothesized standard normal density. Hence Model 5 is preferable to the other models.

Tables 3.2 and 3.3 show that Model 5 outperforms Models 4 and 7 across all in-sample model assessment criteria, BIC and DIC. The only two exceptions are the shorter CI(EX) for Model 7 and the slightly lower BIC for Model 4. The latter is due to the heavy penalty term in the BIC ($-2\ln f(\boldsymbol{y}|\boldsymbol{\theta})$ in (3.11) is 1021 and 1001, respectively, for Models 4 and 5). Figure 3.5 compares the mean and 95% CI of X_t between Models 5 and 7. The CI estimate is clearly shorter for Model 5, giving more precise fitted values.

3.5.5. Forecasting. As in Chapter 2, we perform the forecast of $n_1 = 50$ daily ranges using Models 4, 5 and 7 with the forecasting period from 1 May 2009 to 10 July 2009, and the forecasting values are labelled as $x_{764}, \ldots, x_{813}$. There are two possible ways of computing the forecasts in a Bayesian approach.

Firstly, the forecasts can be obtained using the posterior predictive means of X_t which are sampled based on the predictive distribution. The joint predictive distribution is given by

$$f(x_{764},\ldots,x_{813}|\boldsymbol{x}) = \int \prod_{t=764}^{813} f_{LT}\left(x|\upsilon_t - \ln(a_2^{t-\mathcal{T}_2}),\tau^2,\alpha\right) f(\boldsymbol{\theta}|\boldsymbol{x})\, d\boldsymbol{\theta}, \tag{3.15}$$

where $\boldsymbol{x}$ denotes the vector of 763 observed daily ranges, $\boldsymbol{\theta}$ is the vector of model parameters, and $f_{LT}(x|b,c,d)$ is the density function of the log-t distribution with location b, scale c, and number of degrees of freedom d. Given a set of parameter values, $\boldsymbol{\theta}^{(i)}$, at the i-th iteration of the Gibbs sampling output, a set of predicted values can be simulated successively from

$$x_t|x_{t-1},\boldsymbol{\theta}^{(i)} \sim LT\left(\upsilon_t^{(i)} - \ln[(a_2^{(i)})^{t-\mathcal{T}_2^{(i)}}],(\tau_2^{(i)})^2,\alpha_2^{(i)}\right), \tag{3.16}$$

where $t = 764,\ldots,813$. The random variate generation from the log-t distribution can be done via its scale mixture of normal representation.

This forecast can be easily performed in WinBUGS by assigning a missing value 'NA' for the forecast X_t, $(t > 763)$ such that the vector of 813 observations to be uploaded is $(X_1,\ldots X_{763},\mathrm{NA},\ldots,\mathrm{NA})$. Posterior samples of model parameters together with the missing forecast X_t are then generated using MCMC techniques.

Secondly, the forecast can be based on the expected value $E(X_t)$ in (3.9) which are sampled in the MCMC similar to the forecast in Chapter 2 where $E(X_t)$ is analytically evaluated. Forecast based on

the expected value $E(X_t)$ is more robust to outliers. To enable comparison of forecasting performance with models in Chapter 2, forecasts using $E(X_t)$ are reported. The posterior means of $E(X_t)$ and the corresponding 95% Bayesian prediction intervals of $E(X_t)$ are displayed in Figures 3.1(b), 3.2(b), and 3.3(b) for Models 4, 5, and 7, respectively, and are also plotted in Figure 3.6 for clarity.

In general, the forecasting error increases across the forecast period due to the accumulated uncertainty. However, Model 5 has a substantially lower forecasting error, and hence a much shorter prediction interval, because of the fitted decreasing trend and the insignificance of $\beta_{\mu 112}$ and $\beta_{\mu 212}$, leading to a relatively constant υ_t in the expression of $E(X_t)$. Since the observed X_t shows a gentle decline during the forecast period, the forecast using Model 4 with a fitted increasing trend is less satisfactory. Model 5 still outperforms Model 7 across all out-of-sample forecasting criteria in Table 3.3.

3.5.6. Outlier diagnostic. It is well known that Student's t-distribution provides a robust inference by downweighting the distorting effects of outliers. Expressing the t-distribution as an SMN distribution, Choy and Smith (1997) was the first to propose performing outlier diagnostics using the scale mixture variable λ in the SMN representation. An outlier is associated with a large value of $1/\lambda$ which

inflates the variance of the corresponding normal distribution to accommodate the outlier. Therefore, the extremeness of observations is closely associated with the magnitude of λ.

Figure 3.7 plots the reciprocal $1/\lambda_t$ in Model 5 across time. From the figure, two outliers on 28 February 2007 and 5 August 2008 (t=214 and 577) are detected, because their variances are inflated nearly twice as much as the variances at other time points. Table 3.4 reports the reciprocal $1/\lambda_t$, the observed value X_t, the mean $E(X_t)$, and the 95% CI of X_t for the two outliers. As the CIs do not contain X_t, the daily ranges on 28 February 2007 and 5 August 2008 are indeed outlying.

3.6. Conclusion

This chapter extends the GP model to the CARGP model for range data to describe the persistence dynamics in the mean function μ_t. The CARR-type range model is simpler than the GARCH and SV models, but yet it was shown to provide a superior volatility forecast. The performance of the proposed CARGP model was shown to exceed that of the CARR-type models in four aspects: the accommodation of trend movement using an explicit ratio parameter or function, the adoption of heavy-tailed distributions such as the log-t distribution to describe different tail behavior, the use of the Bayesian approach via the Bayesian software WinBUGS to simplify the model implementation for non-experts, and, lastly, the representation of the t-distribution in an

SMN representation to facilitate the MCMC algorithm in the Bayesian simulation and enable outlier detection. The simulation study shows that the CARGP model provides highly accurate parameter estimates, particularly when the sample size is large. In the empirical study using the AORD daily range data, the CARGP model acheives a better model fit and provides a sharper volatility forecast, confirming the superiority of the CARGP model.

Range data is sensitive to outliers. In the daily range of AORD data, the daily ranges on 28 February 2007 and 5 August 2008 are detected to be outliers by the variable $1/\lambda_t$. However, Figure 3.2(a) and Table 3.4 show that even the best model, Model 5, fails to capture these instantaneous jumps and possibly mean reversions well. To capture the spikes in a highly volatile financial time series, a product of the random jump indicator and the random jump size in a jump diffusion model may be added to υ_t. We believe the extension of the CARGP model to incorporate the mean reverting jump diffusion modelling strategies is promising and will be considered in the next chapter.

Appendix

The full conditional distributions for the parameters in Model 5 are derived to facilitate the Gibbs sampling algorithm. Define $\boldsymbol{x} = (x_1, \ldots, x_n)$; $\boldsymbol{\beta}_g = (\beta_{\mu 0g}, \beta_{\mu 1g}, \beta_{\mu 2g}, \beta_{\mu 3g})$ (drop the redundant subscript j), $g = 1, 2$ indicate the GP model before and after the threshold $\mathcal{T}_2$ ($\mathcal{T}_1 = 1$) respectively; $(s_1, e_1) = (1, \mathcal{T}_2 - 1)$ and $(s_2, e_2) = (\mathcal{T}_2, n)$ indicate the start and end times of the two GPs respectively; $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n)$ and $\boldsymbol{\lambda}_{-t} = (\lambda_1, \ldots, \lambda_{t-1}, \lambda_{t+1}, \ldots, \lambda_n)$. The Gibbs sampler draws realisations iteratively from the following conditional distributions:

$$
\begin{aligned}
f\left(\ln(a_g)|\boldsymbol{\beta}_g, \tau_g^2, \boldsymbol{\lambda}, \nu_g, \mathcal{T}_2, \boldsymbol{x}\right) &= N\left(-\frac{\sum_{t=s_g+1}^{e_g}(t-\mathcal{T}_g)\left[\ln(x_t)-v_t\right]\tau_g^2}{\sum_{t=s_g+1}^{e_g}\lambda_t(t-\mathcal{T}_g)^2}, \frac{\tau_g^2}{\sum_{t=s_g+1}^{e_g}\lambda_t(t-\mathcal{T}_g)^2}\right) \\
&\quad \times I(\ln(0.95), \ln(1.05)) \\
f\left(\boldsymbol{\beta}_g|a_g, \tau_g^2, \boldsymbol{\lambda}, \nu_g, \mathcal{T}_2, \boldsymbol{x}\right) &\propto \prod_{t=s_g}^{e_g} N\left(\ln(x_t)\middle| v_t - (t-\mathcal{T}_g)\ln(a_g), \frac{\tau_g^2}{\lambda_t}\right) \\
f\left(\tau_g^2|a_g, \boldsymbol{\beta}_g, \boldsymbol{\lambda}, \nu_g, \mathcal{T}_2, \boldsymbol{x}\right) &= IG\left(\frac{n}{2}, \frac{1}{2}\sum_{t=1}^{n}\lambda_t\left[\ln(x_t)-v_t+(t-\mathcal{T}_g)\ln(a_g)\right]^2\right) \\
f\left(\lambda_t|a_{\mathtt{g}}, \boldsymbol{\beta}_{\mathtt{g}}, \boldsymbol{\lambda}_{-t}, \nu_{\mathtt{g}}, \mathcal{T}_2, \boldsymbol{x}\right) &= IG\left(\frac{\nu_{\mathtt{g}}+1}{2}, \frac{\nu_{\mathtt{g}}}{2}+\frac{1}{2\tau_{\mathtt{g}}^2}\left[\ln(x_t)-v_t+(t-\mathcal{T}_{\mathtt{g}})\ln(a_{\mathtt{g}})\right]^2\right) \qquad (3.17) \\
f\left(\nu_g|a_g, \boldsymbol{\beta}_g, \boldsymbol{\lambda}, \mathcal{T}_2, \boldsymbol{x}\right) &\propto \prod_{t=s_g}^{e_g}\frac{1}{\nu_g}G\left(\lambda_t\middle|\frac{\nu_g}{2}, \frac{\nu_g}{2}\right) \\
f\left(\mathcal{T}_2|a_g, \boldsymbol{\beta}_g, \boldsymbol{\lambda}_g, \nu_g, \boldsymbol{x}\right) &\propto \text{Multinomial}(\pi_{610}, \ldots, \pi_{630}),
\end{aligned}
$$

where $\pi_k = \dfrac{\prod_{t=1}^{k-1} f_{LN}\left(x_t|\upsilon_{t1} - \ln(a_1^{t-1}), \frac{\tau_1^2}{\lambda_t}\right)\prod_{t=k}^{n} f_{LN}\left(x_t|\upsilon_{t2} - \ln(a_2^{t-k}), \frac{\tau_2^2}{\lambda_t}\right)}{\sum_{k'=610}^{630}\left[\prod_{t=1}^{k'-1} f_{LN}\left(x_t|\upsilon_{t1} - \ln(a_1^{t-1}), \frac{\tau_1^2}{\lambda_t}\right)\prod_{t=k'}^{n} f_{LN}\left(x_t|\upsilon_{t2} - \ln(a_2^{t-k'}), \frac{\tau_2^2}{\lambda_t}\right)\right]}$, $k = 610, \ldots, 630$, $g = 1, 2$, $\mathtt{g} = 1\, I(t < \mathcal{T}_2) + 2\, I(t \geq \mathcal{T}_2)$ in (3.17), and υ_t is given by (3.8). The algorithm of Robert (1995) can be used to simulate the random variate $\ln(a_g)$ from a truncated normal distribution.

The conditional distributions of $\boldsymbol{\beta}_g$ and ν_g are non-standard, and random variate generations from these full conditional distributions can be performed using Metropolis-Hastings algorithms.

TABLE 3.1. Parameter estimates, standard deviation, absolute percentage bias, root mean square error and coverage percentage in the simulation study.

Dist.		Set	a	$\beta_{\mu 0}$	$\beta_{\mu 1}$	$\beta_{\mu 2}$	ν	σ^2	Set	a	$\beta_{\mu 0}$	$\beta_{\mu 1}$	$\beta_{\mu 2}$	ν	σ^2
							$n = 200$								
LT	θ	1	1.001	1.000	-0.200	0.030	5.000	0.500	2	0.99800	-0.020	0.700	0.200	5.000	1.000
	$\hat{\theta}$		1.001	0.995	-0.207	0.021	8.043	0.545		0.99790	-0.063	0.484	0.222	9.036	1.095
	SD		0.001	0.216	0.241	0.065	3.877	0.098		0.00415	0.185	0.214	0.055	3.968	0.168
	APB		0.000	0.005	0.035	0.311	0.609	0.089		0.00010	2.165	0.309	0.109	0.807	0.095
	RMS		0.001	0.215	0.240	0.066	4.91	0.107		0.00413	0.189	0.304	0.059	5.65	0.192
	CP		85	98	98	95	92	93		91	89	80	95	91	95
							$n = 700$								
LN	θ	1	1.001	1.000	-0.200	0.030	-	0.500	2	0.99800	-0.020	0.700	0.200	-	1.000
	$\hat{\theta}$		1.001	1.005	-0.212	0.031	-	0.501		0.99802	-0.023	0.659	0.212	-	0.999
	SD		0.000	0.255	0.280	0.038	-	0.028		0.00056	0.032	0.060	0.026	-	0.055
	APB		0.000	0.030	0.138	0.202	-	0.005		0.00002	0.128	0.059	0.062	-	0.001
	RMS		0.000	0.411	0.327	0.034	-	0.031		0.00056	0.031	0.072	0.029	-	0.055
	CP		100	89	89	92	-	94		94	92	88	96	-	95
							$n = 700$								
LT	θ	1	1.001	1.000	-0.200	0.030	5.000	0.500	2	0.99800	-0.020	0.700	0.200	5.000	1.000
	$\hat{\theta}$		1.001	1.011	-0.206	0.022	5.769	0.517		0.99800	-0.025	0.669	0.208	5.600	1.034
	SD		0.000	0.269	0.301	0.030	1.359	0.045		0.00065	0.031	0.046	0.027	1.224	0.088
	APB		0.000	0.202	0.729	0.230	0.114	0.011		0.00000	0.271	0.044	0.041	0.120	0.034
	RMS		0.000	0.462	0.351	0.035	1.32	0.043		0.00065	0.031	0.055	0.028	1.358	0.094
	CP		100	85	88	95	94	96		94	97	93	93	94	94

TABLE 3.2. Parameter estimates, standard errors in *italics*, BIC and DIC for the AORD daily range data.

Model	Dist.	Type	T	$\beta_{\mu 0}$	$\beta_{\mu 11}$	$\beta_{\mu 21}$	$\beta_{\mu 22}$	$\beta_{\mu 31}$	a or β_{a0}	β_{a1}	τ^2	ν or α	GOF^
CARGP													
M1	LN	(1,1)	-	-0.0234	0.7644	0.1879	-	-	0.9983	-	0.1762	-	1121
			-	*0.0118*	*0.0356*	*0.0253*	-	-	*0.0005*	-	*0.0090*	-	*1098*
M2	LT	(1,1)	-	-0.0192	0.7878	0.1735	-	-	0.9984	-	0.1568	19.26	1125
			-	*0.0095*	*0.0263*	*0.0202*	-	-	*0.0005*	-	*0.0103*	*5.93*	*1095*
M2.1	LT	(1,1)	-	0.0027†	0.8080	0.1776	-	-	-	-	0.1570	17.96	1128
($a = 1$)			-	*0.0031*	*0.0199*	*0.0182*	-	-	-	-	*0.0104*	*5.97*	*1102*
M3	LT	(1,2)	-	-0.0222	0.7806	0.1661	0.0110†	-	0.9983	-	0.1566	20.56	1132
			-	*0.0103*	*0.0328*	*0.0415*	*0.0521*	-	*0.0006*	-	*0.0104*	*5.40*	*1097*
M4	LT	(1,1)	-	-0.0129	0.8269	0.1124	-	-0.0536	0.0038	-0.0008	0.1441	18.63	1074
			-	*0.0061*	*0.0259*	*0.0197*	-	*0.0067*	*0.0005*	*0.0001*	*0.0094*	*5.91*	*1047*
M5	LT	(1,1)	622	-0.0225	0.8437	0.0972	-	-0.0765	0.9988	-	0.1490	21.56	1101
			5.27	*0.0084*	*0.0311*	*0.0233*	-	*0.0089*	*0.0003*	-	*0.0096*	*4.86*	*1037*
	LT	(1,1)	-	1.0530	0.1960†	-0.0734†	-	-0.0300	1.0070	-	0.1144	18.44	
			-	*0.3328*	*0.2279*	*0.0875*	-	*0.0135*	*0.0007*	-	*0.0153*	*6.29*	
CARR													
M6	Exp	(1,1)	-	0.1571	0.6075	0.2820	-	-0.1261	-	-	-	-	1949
			-	*0.0410*	*0.0733*	*0.0615*	-	*0.0365*	-	-	-	-	*1930*
M7	Wei	(1,1)	-	0.1883	0.5858	0.2766	-	-0.1378	-	-	-	2.1860	1339
			-	*0.0156*	*0.0286*	*0.0262*	-	*0.0181*	-	-	-	*0.0521*	*1316*

† p-value > 0.05.

^ The first GOF measure is BIC and the second GOF in italics is DIC.

Table 3.3. In-sample and out-of-sample model assessment for Models 4 to 7.

	In-sample model-fit criteria ($n = 763$)									
Model	RMS_1	MAE_1	RMS_2	MAE_2	CP	CI(EX)	CI(X)	Q_{12}	W	JB
M4	0.684	0.455	1.001	0.750	0.965	0.189	2.161	12.1	0.32	14.1
M5	*0.649*	*0.442*	*0.953*	*0.736*	*0.963*	0.284	*2.069*	*5.91*	*0.11* †	*6.40*
M6	0.700	0.485	0.994	0.763	0.996	0.362	5.265	21.2	23.7	-
M7	0.707	0.490	0.988	0.759	0.972	*0.165*	2.616	19.4	1.73	-

	Out-of-sample forecasting criteria ($n_1 = 50$)						
Model	RMS_1	MAE_1	RMS_2	MAE_2	CP	CI(EX)	CI(X)
M4	0.859	0.712	1.277	1.095	0.820	1.194	2.961
M5	*0.521*	*0.383*	*0.800*	*0.659*	*0.940*	*0.419*	*1.764*
M6	0.577	0.664	0.878	0.851	1.000	2.359	4.618
M7	0.604	0.483	0.922	0.770	0.980	1.382	2.472

† p-value > 0.05 while the p-values for other test statistics are all < 0.05.

Table 3.4. Summary information for the outliers in Model 5.

t	Date	$1/\lambda_t$	X_t	$E(X_t)$	95% CI of X_t
214	28 Feb 2007	1.8021	3.5846	0.7378	(0.2455, 1.8676)
577	5 Aug 2008	1.7176	8.0839	1.8200	(0.6091, 4.4166)

FIGURE 3.1. (a) Trends of the mean, variance, and ratio. (b) Observed, expected, 95% CI of X_t and 95% PI of the forecast $E(X_t)$ using Model 4.

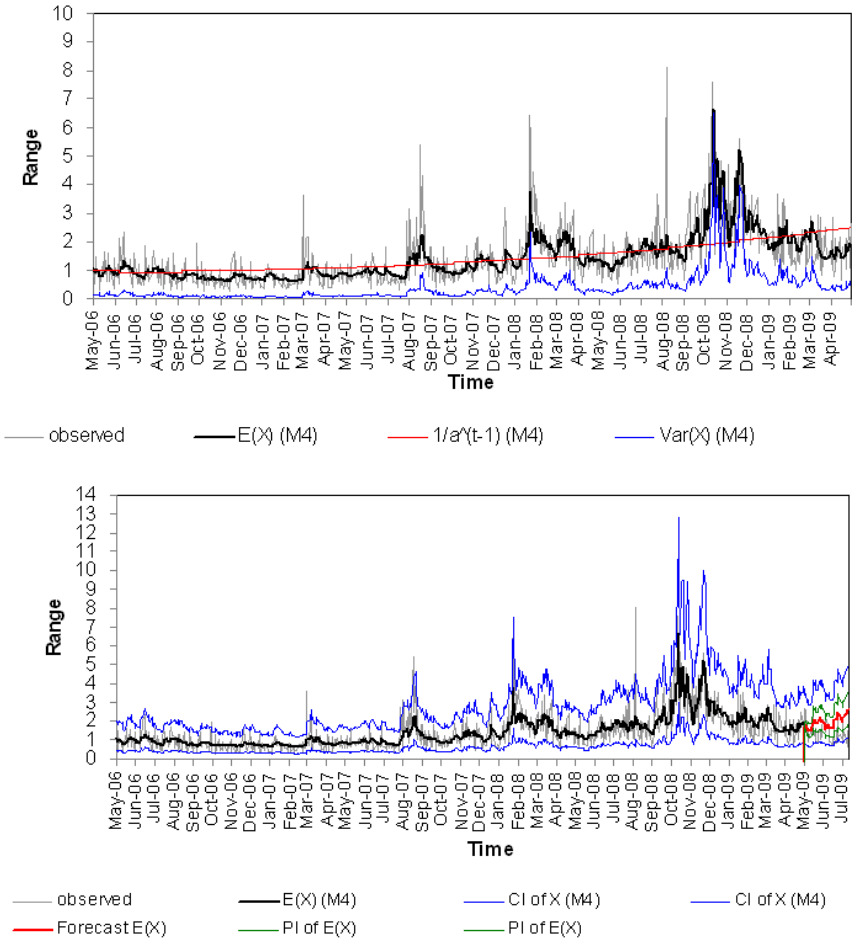


FIGURE 3.2. (a) Trends of the mean, variance, and ratio. (b) Observed, expected, 95% CI of X_t and 95% PI of the forecast $E(X_t)$ using Model 5.

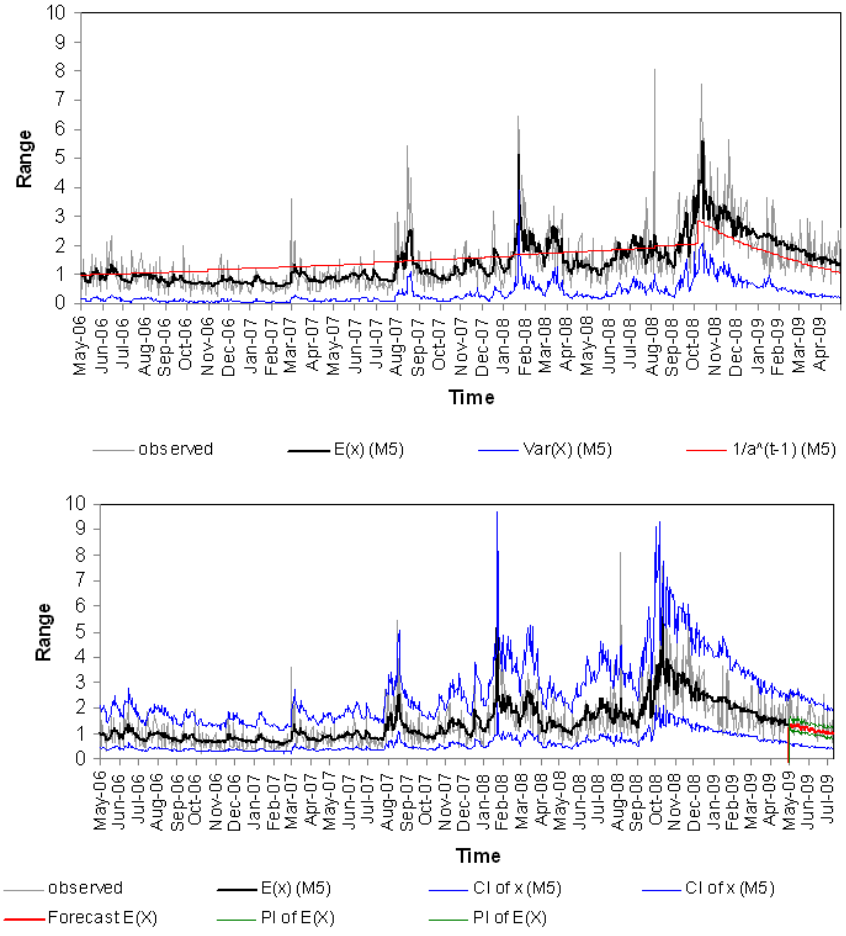


FIGURE 3.3. (a) Trends of the mean and variance. (b) Observed, expected, 95% CI of X_t and 95% PI of the forecast $E(X_t)$ using Model 7.

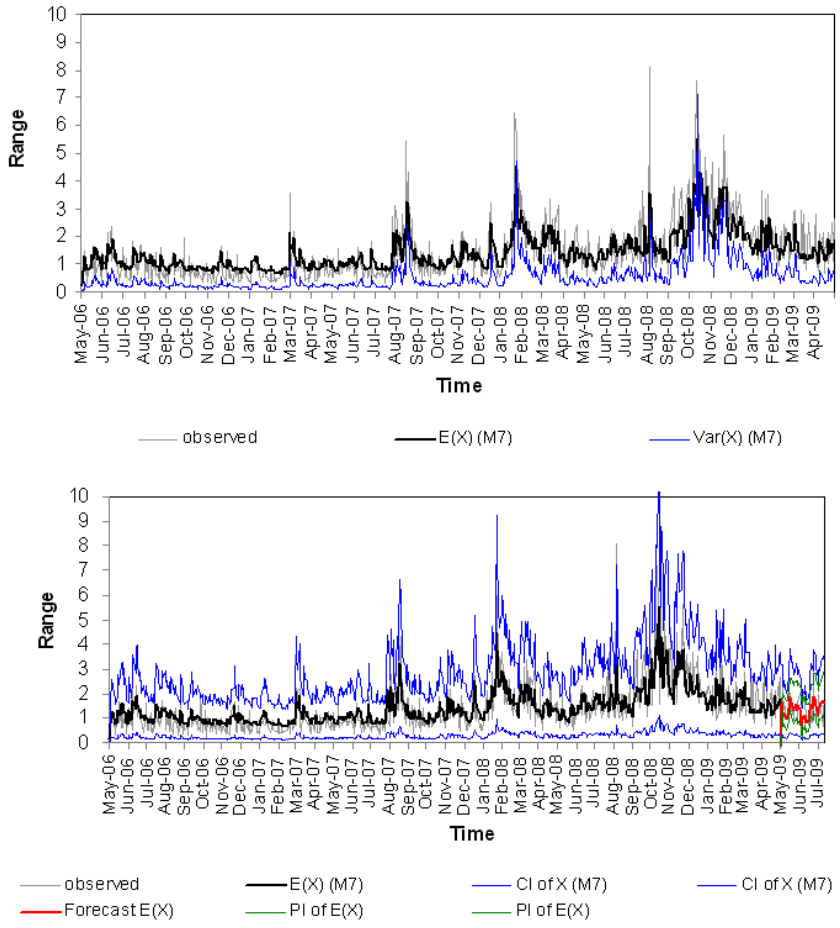


FIGURE 3.4. Comparison of observed and hypothesized distributions for standardised X_t between Models 4, 5, and 7.

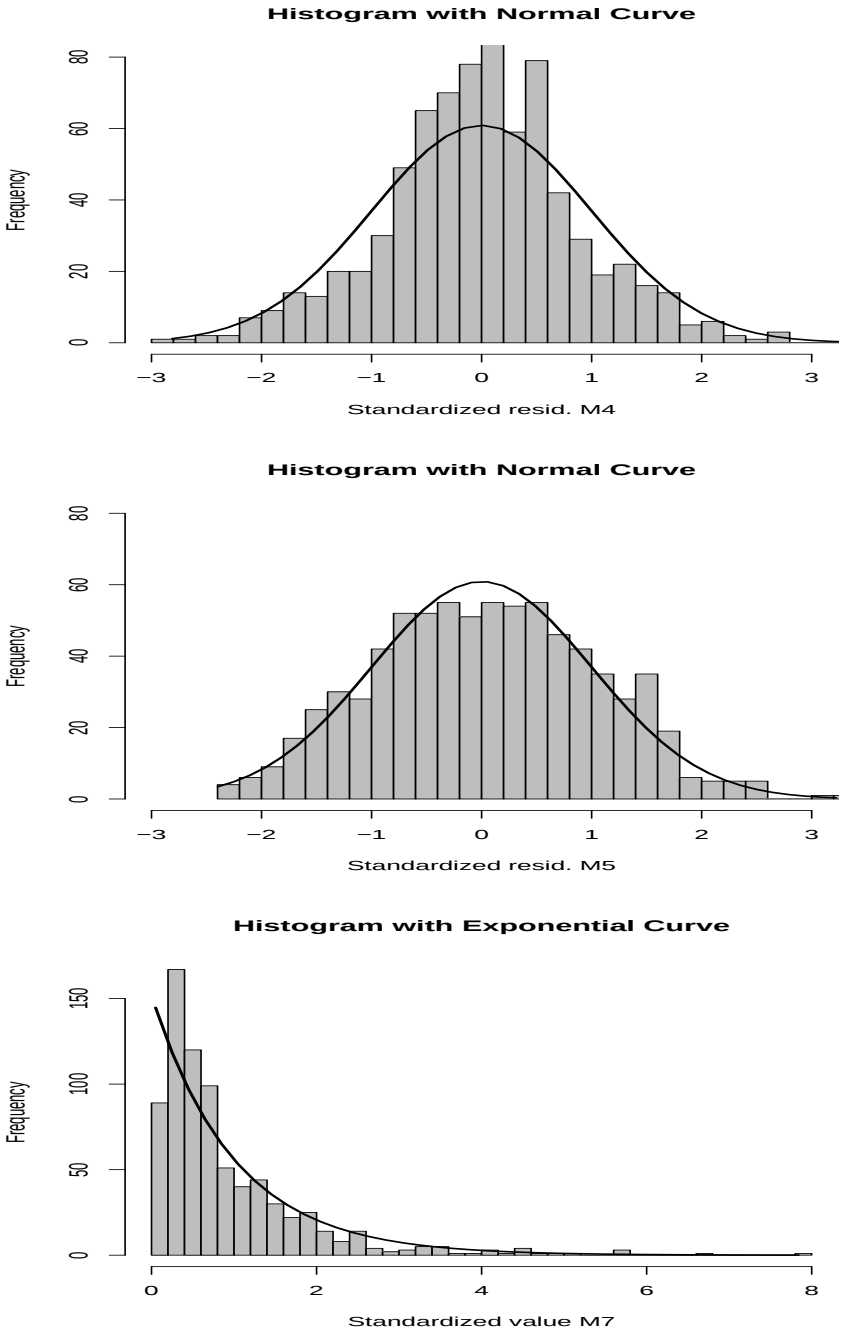


FIGURE 3.5. Comparison of the expected and 95% CI of X between Models 5 and 7.

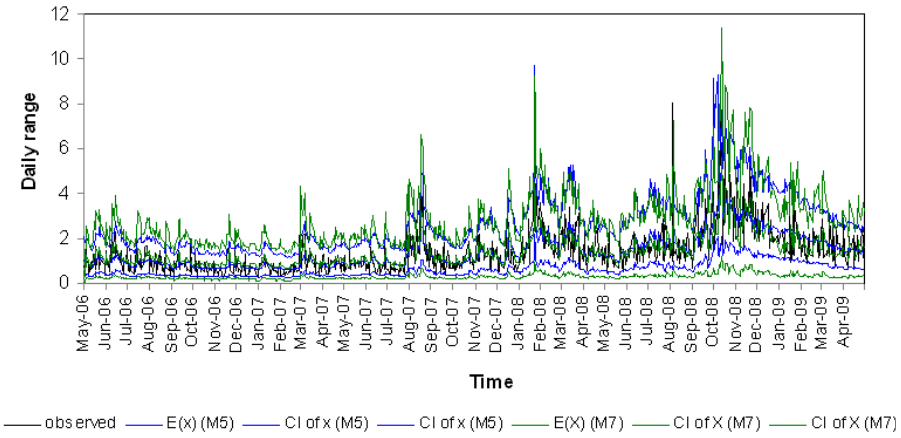


FIGURE 3.6. Forecast range and 95% PI of the forecast range using Models 4, 5 and 7 respectively.

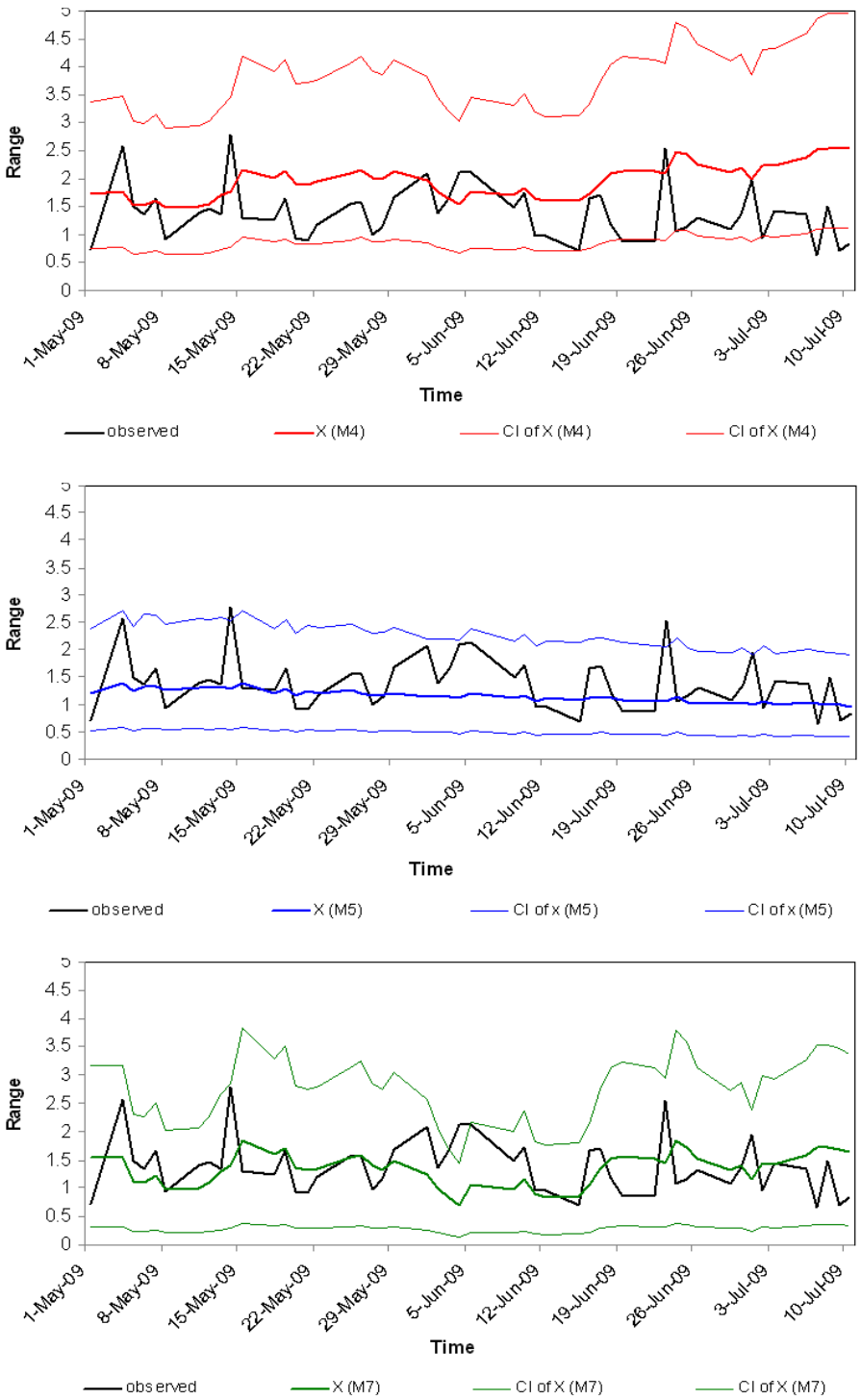
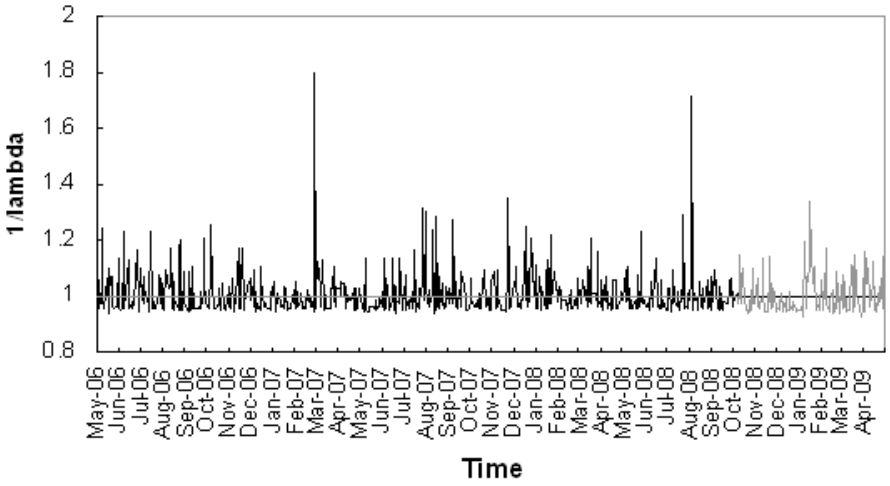


FIGURE 3.7. Reciprocal of lambda $1/\lambda_t$ in outlier diagnostic using Model 5.



Chapter 4

Extension to CARGP Threshold and CARGP Jump models

4.1. Introduction

In the last two chapters, we show that the extended geometric process (GP) models enjoy several nice properties. The models address trend movements, autocorrelation, leverage effect, leptokurtosis and heteroskedasticity which are the common features in a financial time series. Moreover, the ARGP and CARGP models are easy to implement using both the LSE and ML methods of inferences for the former model and Bayesian method for the latter model. However, the models are still incapable of detecting and describing the frequent and rather pronounced jumps and mean reversions in the daily range data. This chapter focuses on devising different modelling strategies to capture these features. To do this, we consider two stages of development. In the first stage, we adopt a data set which displays more prominent jump diffusion and mean reversion features, and hence, facilitates the training of more efficient mean reverting jump diffusion (MRJD) models. One favourable choice of such data is the electricity price data because

it is well known that electricity market is highly volatile due to its seasonality in demand and absence of effective storage. As we focus on the modelling of electricity price, daily maximum electricity price of New South Wales (NSW), Australia, is used in the first empirical study to test different models that describe mean reversions and jumps in the search of better models. Then, in the second stage, the best model(s) will be applied to the daily range of All Ordinaries (AORD) Index of Australia stock market used in Chapter 2 and 3.

Apart from providing insight in model development, research on electricity price has a practical context. Because of the increasing concerns over climate change and global warming, efficient energy use is a contemporary prominent and urgent issue. Governments and authorities set rules and regulations to ensure efficient power generation and security of supply. However, as efficiency is often achieved through economics of scale, many countries started to restructure and deregulate power markets to introduce competition. Consequently, it brings enormous uncertainty and market risk to electricity market participants, including generators and service providers.

Comparing to financial stock markets, electricity markets are more competitive and hence spot electricity prices are more volatile. This is attributed to several unique features of electricity prices. Firstly, the demand of electricity is highly seasonal. Peaks of demand during summer, winter, week-day, and intra-day high load hours are driven

by the effects of temperature, number of day light hours and human activities. Secondly, the prices contain frequent and short term extreme jumps caused by the low elasticity of supply. Electricity is continuously generated and consumed with no effective storage. If the instantaneous demand cannot be met because of a system breakdown say, the price can be pushed up substantially but it reverts back to equilibrium price shortly after the outage is over. Lastly, the prices often display a common feature called mean reversion which refers to the tendency to fluctuate around a long term equilibrium.

To reduce the financial exposure and maintain spot market stability, some market participants use financial contracts, known as derivatives, which include swaps or hedges, options and futures contracts, to lock electricity at a fix price for a certain time in the future. Alternatively, they monitor closely the price movement using models that describe the dynamics of the price volatility, and capture the major features of electricity prices for pricing, hedging and forecasting in the electricity markets (Schwartz 1997).

There are two main branches of electricity price models: fundamental models and econometric models. The former are based on the equilibrium of demand and supply (Barlow, 2002) whereas the latter on stochastic models in finance which include the MRJD models of Barz and Johnson (1998) and Cartea and Figueroa (2005). A general

MRJD model for the price X_t is defined as

$$dX_t = \mu_t\, dt + J dq_t + \sigma_t\, dW_t, \tag{4.1}$$

where the first term $\mu_t\, dt$ contains a deterministic function of covariates such as seasonality and a mean reversion term $-mX_{t-1}$, the second term $J\, dq_t$ describes a pure jump process which is a product of a jump size J with a lognormal or gamma distribution and a Poisson process dq_t with intensity $\pi\, dt$ (a compound Poisson process), and the last term $\sigma_t\, dW_t$ is the error or diffusion process with volatility σ_t and Wiener process W_t. The term $-mX_{t-1}$ with a mean reversion rate m causes the price process to be pulled towards a equilibrium value. Rambharat *el al.* (2005) showed that the threshold autoregressive (TAR) model using a Bayesian approach provided a better fit than the simple MRJD model for electricity spot prices, as it allows both the mean reversion rate m and jump intensity J to depend on the previous price through a threshold level $\mathcal{L}$, that is, $m = m_1 I(X_{t-1} \leq \mathcal{L}) + m_2 I(X_{t-1} > \mathcal{L})$ and $\pi = \pi_1 I(X_{t-1} \leq \mathcal{L}) + \pi_2 I(X_{t-1} > \mathcal{L})$, where $I(E)$ is an indicator function for event E. Other electricity price models include the neural networks model of Zhang *et al.* (2003), the autoregressive integrated moving average (ARIMA) model of Contreras *et al.* (2003), the general autoregressive conditional heteroskedastic (GARCH) models of Thomas and Mitchell (2005) and Higgs and Worthington (2005),

the data mining approach of Lu *el al.* (2005), the dynamic econometric model of Nogales and Conejo (2006) and the Gaussian stochastic volatility (SV) model of Smith (2010).

While many of these models capture the short-term jumps and mean reversions, they do not allow for a long-term trend on equilibrium price, particularly when the observation period is long so that the assumption of constant equilibrium price becomes unrealistic. In Chapter 3, we develop the Conditional Autoregressive Geometric Process (CARGP) model to allow for heteroskedasticity and flexible trend movement. To model the dynamics of electricity prices, this chapter extends the CARGP model and proposes two flexible types of models, namely the CARGP Threshold (CARGPT) model and CARGP Jump (CARGPJ) model, by including threshold and jump effects into the CARR-type mean function. The CARR-type mean function fosters flexible correlation, whereas the threshold model permits the mean reversion rate to be price dependent. Moreover, the extreme jumps are captured by the heavy-tailed t-distribution as well as an explicit jump component. Lastly, as compared to the very sophisticated GARCH and SV models, the GP model offers a simple modelling alternative to capture the long-term price movement and volatility dynamics, using a ratio parameter which describes the trend pattern. See equation (2.1) in Section 2.2.1 for details. Essentially, the model classifies the price dynamics into short-term jumps and long-term trend movement.

The jump and mean reversion components in the short-term jumps are based on the untrended prices, after discounting the observed prices geometrically by the ratio parameter, thereby eliminating the distorting effect of long-term price movement on modelling the jumps and mean reversions. In numerical analysis using electricity prices, the CARGPT and CARGPJ models are shown to be a generalisation of the MRJD model in allowing volatility clustering, long-term trend movement and leptokurtic data distribution.

For model implementation, Barlow (2002) and Huisman and Mahieu (2003) adopted the maximum likelihood approach. We use a Bayesian approach using Markov chain Monte Carlo (MCMC) algorithms via the user friendly Bayesian software WinBUGS.

This chapter is organised as follows. Section 4.2 introduces the development of CARGPT and CARGPJ models. In Section 4.3, the Bayesian inference for the proposed models is described. In Section 4.4, the proposed models are fitted to the daily maximum electricity price of NSW, Australia and AORD daily range data. Finally, the study is concluded in Section 4.5.

4.2. Model development

4.2.1. The CARGP model. To model the jump diffusion dynamics in a time series with trends, we begin with the CARGP model

in Chapter 3. In the model, we assume that $\ln Y_t \sim t_\nu(\upsilon_t, \tau^2)$ or equivalently, $\ln Y_t|\lambda_t \sim N(\upsilon_t, \frac{\tau^2}{\lambda_t})$ by conditioning on λ_t, where the CARR-type mean function is defined as

$$\upsilon_t = \beta_{\mu 0} + \beta_{\mu 1}\upsilon_{t-1} + \beta_{\mu 2}\ln(y_{t-1}) + \beta_{\mu 3}z_{1t} + \beta_{\mu 4}z_{2t}, \tag{4.2}$$

where $y_{t-1} = a_{t-1}^{t-2}x_{t-1}$ is the latent SP and z_{1t} and z_{2t} denote some time-varied covariates which measure, for example, the temperature and weekday effects. Hence, $\ln X_t|\lambda_t \sim N(\upsilon_{xt}, \frac{\tau^2}{\lambda_t})$, where the mean and ratio functions are given by

Model 0:

$$\upsilon_{xt} = \upsilon_t - \ln(a_t^{t-1}), \tag{4.3}$$

$$a_t = \exp(\beta_{a0} + \beta_{a1}\ \ln t), \tag{4.4}$$

respectively. Then the mean and variance of X_t in Model 0 and the forthcoming extended models (Models 1-5) are given by (3.3) and (3.4) in Section 3.2.2 respectively.

4.2.2. Extension to the CARGP model. A time series is sometimes subject to abrupt and unanticipated effects and behaves differently across a possibly unknown level of a certain risk variable. They are the jumps which occur prominently and frequently in some financial time series such as the daily maximum electricity price. These jumps usually do not stay in the new level but rather quickly revert to their previous levels. Time series model not cater for these jumps and mean reversions will underestimate the price at the time of jump, leading to inflated variance estimate, τ^2. To accurately model these jump

dynamics, we consider two important factors: the time and magnitude for the jumps. Two types of models, the threshold and jump models, are proposed to address these two factors on the jump dynamics.

4.2.2.1. *Threshold intercept model.* A threshold model assumes that the exposure to a risk variable has no influence on the model except when the risk variable exceeds a certain threshold $\mathcal{L}$. Because of the autoregressive feature in a time series, the risk variable is often taken to be certain lagged response, say the lag-1 response, X_{t-1}. To allow for the model shift across a certain threshold $\mathcal{L}$, a single component $Q_t = J \cdot I(x_{t-1} > \mathcal{L})$, where $I(E)$ is an indicator function of the event E, is added to υ_{xt} in (4.3), and the extended model is given by

$$\text{Model 1:} \quad (4.5) \qquad \upsilon_{xt} \;=\; \upsilon_t - \ln(a_t^{t-1}) + J \cdot I(x_{t-1} > \mathcal{L}),$$

where υ_t and a_t are given by (4.2) and (4.4), respectively. The component Q_t consists of a constant jump size J and a risk variable, say X_{t-1} which determines a possible jump time when X_{t-1} exceeds $\mathcal{L}$. This threshold model assumes that the mean reversion rate $m = 1 - \beta_{\mu 2}$ is a constant.

4.2.2.2. *Threshold intercept model with time-varied jump size.* A constant jump size J may be too restrictive. To allow a flexible jump mechanism, J is allowed to change over time and is written as J_t. The function υ_{xt} is similar to (4.5) but with J_t replacing J, and the resultant

model, called Model 2, is defined as

$$\text{Model 2:} \quad (4.6) \qquad \upsilon_{xt} \;=\; \upsilon_t - \ln(a_t^{t-1}) + J_t \cdot I(x_{t-1} > \mathcal{L}),$$

where $J_t \sim G(\alpha, \gamma)$ and $E(J_t) = \dfrac{\alpha}{\gamma}$.

4.2.2.3. *Threshold coefficient model.* Instead of assigning a single component to describe the model shift, a new set of parameters in the mean and ratio functions is applied when the threshold is exceeded. The new mean and ratio functions are defined as:

$$\text{Model 3:}\ \upsilon_{xt} \;=\; \begin{cases} \beta_{\mu 01} + \beta_{\mu 11}\upsilon_{t-1} + \beta_{\mu 21}\ln(y_{t-1}) + \beta_{\mu 31} z_{1t} + \beta_{\mu 41} z_{2t} - \ln(a_t^{t-1}) & \text{if } x_{t-1} \le \mathcal{L}, \\ \beta_{\mu 02} + \beta_{\mu 12}\upsilon_{t-1} + \beta_{\mu 22}\ln(y_{t-1}) + \beta_{\mu 32} z_{1t} + \beta_{\mu 42} z_{2t} - \ln(a_t^{t-1}) & \text{if } x_{t-1} > \mathcal{L}, \end{cases}$$

(4.7)

$$a_t \;=\; \begin{cases} \exp(\beta_{a01} + \beta_{a11}\ \ln t) & \text{if } x_{t-1} \le \mathcal{L}, \\ \exp(\beta_{a02} + \beta_{a12}\ \ln t) & \text{if } x_{t-1} > \mathcal{L}. \end{cases}$$

Note that Model 3 can be written as (4.3) and (4.4), where

$$\beta_{jk_j} = \beta_{j,k_j,1}\, I(x_{t-1} \le \mathcal{L}) + \beta_{j,k_j,2}\, I(x_t > \mathcal{L}), \quad j = \mu, a;\ k_\mu = 0, 1, \ldots, 4;\ k_a = 0, 1,$$

and hence, it is a generalisation of Model 1 when $\beta_{\mu 02} = \beta_{\mu 01} + J$ and $\beta_{j,k,2} = \beta_{j,k,1},\ k = 1, 2, 3, 4$ otherwise. The parameters $\beta_{\mu 2g}, g = 1, 2$ in (4.8) provide different mean reversions with rates $m_g = 1 - \beta_{\mu 2g}$ before $(g = 1)$ and during $(g = 2)$ the jumps when $X_{t-1} > \mathcal{L}$ signals a possible jump. They are similar to m_g in the TAR model of Rambharat *et al.* (2005). This threshold coefficient model can be interpreted as a regime switching model with two basic CARGP models for two subsets of data, according to whether $X_{t-1} \le \mathcal{L}$. The three threshold models,

Models 1 to 3, are called the CARGP Threshold (CARGPT) model.

4.2.2.4. *Jump model with time-varied jump size.* Although we can detect the jump time using say a lag-1 response X_{t-1} in the threshold models, the model falls short of identifying an instantaneous jump which is not related to the level of X_{t-1}. Similar to the jump component $\{G_t\}$ in Rambharat *et al.* (2005), we consider a jump component Q_t, where the indicator of jump is a stochastic variable q_t which follows a Bernoulli distribution, that is, $q_t \sim Bern(\pi)$ and π is the probability of jumps. Moreover, similar to Model 2, we allow the jump magnitude J_t to vary across time t. This extended model, called the CARGP Jump (CARGPJ) model (Model 4), is equivalent to adding the component $Q_t = J_t\, q_t$ to υ_{xt}, that is,

Model 4: $$\upsilon_{xt} = \upsilon_t - \ln(a_t^{t-1}) + J_t\, q_t, \tag{4.8}$$

where $J_t \sim G(\alpha, \gamma)$ and $q_t \sim Bern(\pi)$.

4.2.2.5. *Jump model with threshold effects.* While the threshold coefficient model (Model 3) allows the mean reversion to have a price dependent rate, and the jump model (Model 4) captures instantaneous jumps, a model which incorporates both features will model the characteristics of electricity prices better. The resultant model, called Model

5, is given by

$$\text{Model 5:} \qquad (4.9) \qquad \upsilon_{xt} = \upsilon_t - \ln(a_t^{t-1}) + J_t\, q_t,$$

where υ_t is given by (4.2),

$$\begin{aligned} (4.10) \qquad \beta_{\mu k} &= \beta_{\mu k1}\, I(x_{t-1} \le \mathcal{L}) + \beta_{\mu k2}\, I(x_{t-1} > \mathcal{L}), \quad k = 2,3,4, \\ J_t &\sim G(\alpha, \gamma), \\ q_t &\sim Bern(\pi), \\ h &= h_1\, I(x_{t-1} \le \mathcal{L}) + h_2\, I(x_{t-1} > \mathcal{L}), \quad h = \alpha, \gamma, \pi, \\ E(J_t) &= \frac{\alpha_1}{\gamma_1}\, I(x_{t-1} \le \mathcal{L}) + \frac{\alpha_2}{\gamma_2}\, I(x_{t-1} > \mathcal{L}). \end{aligned}$$

Writing $\mathsf{X}_t = \ln X_t$, $d\mathsf{X}_t = \mathsf{X}_t - \mathsf{X}_{t-1}$, $dt = 1$, $\mu_t = \beta_{\mu 0} - m\mathsf{X}_{t-1} + \beta_{\mu 3} z_{1t} + \beta_{\mu 4} z_{2t}$ and $\sigma_t^2 = \tau^2/\lambda_t$, the MRJD model in (4.1) can be expressed as

$$\mathsf{X}_t = \beta_{\mu 0} + (1-m)\mathsf{X}_{t-1} + \beta_{\mu 3} z_{1t} + \beta_{\mu 4} z_{2t} + J_t\, q_t + \frac{\tau}{\sqrt{\lambda_t}}\, Z_t,$$

where $Z_t \sim N(0,1)$ is a standard normal random variable, J_t follows a lognormal or gamma distribution, and q_t is a Poisson process with $q_t = 1$ at probability π and zero otherwise. Equivalently, Model 5 can be written as $\ln X_t \sim N(\upsilon_{xt}, \frac{\tau^2}{\lambda_t})$, where $\upsilon_{xt} = \beta_{\mu 0} + (1-m)\mathsf{X}_{t-1} + \beta_{\mu 3} z_{1t} + \beta_{\mu 4} z_{2t} + J_t\, q_t$, showing that the MRJD model, including the TAR model of Rambharat *et al.* (2005) (equation (2) and (3) in that paper), is closely related to a special case of Model 5 when $a_t = \lambda_t = 1$, $\beta_{\mu 1} = 0$, $\beta_{\mu 2g} = 1 - m_g$, $\alpha_1 = \alpha_2$ and $\gamma_1 = \gamma_2$. Releasing these

constraints, Model 5 captures the long-term trend movement for equilibrium price, heteroskedasticity, leptokurtic data distribution, flexible jump mechanism and mean reversion with price dependent rate and the price dependent jump probability, and covariates such as temperature and weekday effects.

4.3. Bayesian Inference

4.3.1. Model hierarchy and priors. The log-likelihood function and its derivatives as required in the classical likelihood approach are difficult to evaluate because υ_t in (4.2) is a complicated function of $\beta_{\mu 1}$ in (3.7). Bayesian approach is used to avoid the numerical difficulties associated with the maximisation of complicated high-dimensional likelihood functions. The Bayesian hierarchy of the proposed models can be represented by

$$\begin{aligned} X_t &\sim LN(\upsilon_{xt}, \tau^2/\lambda_t), \\ \lambda_t &\sim G(\nu/2, \nu/2), \end{aligned}$$

where υ_{xt} is given by (4.3), (4.5), (4.6), (4.8), (4.8) and (4.9) for Models 0-5, respectively,

$$\begin{aligned} J_t &\sim G(\alpha, \gamma) \text{ (Models 2, 4 \& 5)} \qquad \text{and} \\ q_t &\sim Bern(\pi) \text{ (Models 4 \& 5).} \end{aligned}$$

We assign prior distributions to the model parameters as

$$\beta_{ai}, \beta_{aij} \sim N(0, \sigma_\beta^2), \quad \beta_{\mu i}, \beta_{\mu ij} \sim N(0, \sigma_\beta^2), \quad \tau^2 \sim IG(c_\tau, d_\tau), \quad 1/\nu \sim U(c_\nu, 1),$$
$$\mathcal{L} \sim U(c_L, d_L) \text{ (Models 1 to 3, 5)}, \quad J \sim G(c_J, d_J) \text{ (Model 1)},$$
$$\pi, \pi_j \sim U(0,1) \text{ (Models 4 \& 5)}, \quad \alpha, \alpha_j, \gamma, \gamma_j \sim G(c_\alpha, d_\alpha) \text{ (Models 2, 4 \& 5)}. \tag{4.11}$$

To obtain non-informative priors, the hyperparameter σ_β^2 is set to be very large whereas c_τ, d_τ, c_α and d_α are set to zero. Other hyperparameters c_ν, c_L, d_L, c_J and d_J are set to some specific values depending on real data. In the Gibbs sampling scheme, a single Markov chain is run for 10000 iterations, discarding the initial 5000 iterations as the burn-in period to ensure convergence of parameter estimates. Convergence is also carefully checked by the history and auto-correlation function (ACF) plots. Simulated values from the Gibbs sampler after the burn-in period are taken to mimic a random sample of size 5000 from the joint posterior distribution for posterior inference. Parameter estimates are given by the posterior means or medians. The full conditional posterior distributions which are derived in the Appendix of Chapter 3 for Model 5 can be similarly modified for Models 1 to 5 in this chapter.

4.3.2. Model assessment. Model comparison is based on three measures. Firstly, the deviance information criterion (DIC) which was introduced in Section 3.5.3 is adopted. This criterion measures the

model fit while allowing for model complexity. The other two measures, namely the mean square error (MSE) and mean absolute percentage error (MAPE) are defined as

$$\text{MSE} = \frac{1}{n}\sum_{t=1}^{n}(X_t - \widehat{X}_t)^2 \quad \text{and} \tag{4.12}$$

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\frac{|X_t - \widehat{X}_t|}{X_t}, \tag{4.13}$$

respectively and they assess model fit without a penalty. For each criterion, a smaller value indicates a better model fit.

4.4. Empirical Study

We illustrate the ability of the models in capturing jumps through two sets of data. First of all, since the electricity prices exhibit jumps frequently, it is a good choice of data for deriving efficient jump models. In the second empirical study, we apply the model to the daily range data used in Chapter 2 and 3 to capture the pronounced jump and mean reversion effects in the data.

4.4.1. Electricity price data, NSW of Australia.

4.4.1.1. *The data.* In Australia, the National Electricity Market (NEM) started to work as a pooled exchange market, for supplying electricity to retailers and end-users in Queensland, NSW, ACT, Victoria, and South Australia in December 1998. Tasmania joined the

NEM in 2005. Due to the growing needs for multi-faceted energy services, the Australian Energy Management Operator (AEMO) was established, and replaced the National Electricity Market Management Company (NEMMCO) to manage the NEM and gas markets, according to the provisions of National Electricity Law and Statutory Rules (The Rules) from July 2009. The NEM operates the world's longest interconnected power system which is around 5000 kilometres in distance and supplies more than 10 billion dollars of electricity annually to over 8 million consumers.

NSW is the state that has highest average demand in electricity and also highest production of total energy in 2007/08. A study of the electricity prices of NSW gives participants valuable insight of the price dynamics, and provides them accurate forecast to reduce the risk in this volatile market. In wholesale electricity market, a dispatch price is determined every five minutes, and six of dispatched price are averaged every half-hour to determine the spot price for each trading interval. We analyse the logarithm of daily maximum electricity regional reference price (DMRRP) in mega watt per hour (mwh), $X_t = \ln(\text{DMRRP})$, of NSW, Australia over a 3-year period from 1 January 2007 to 31 December 2009. There are 1096 observations. The data are obtained from the website of AEMO, `www.aemo.com.au`.

Daily maximum prices are analysed to facilitate the modelling of spikes which is the main focus of this empirical study. Moreover, daily

prices are also free from the strong intraday on- and off-peak hours cycle. Since the DMRRP is highly skewed, we take logarithmic transformation of DMRRP. Figure 4.1 displays the time series plot of X_t. The histogram of X_t is presented in Figure 4.2, and summary statistics are reported in Table 4.1. Note that the maximum value of 9.2103 are the log of maximum price which is set to be \$10000 according to The Rules. Figure 4.2 and Table 4.1 show that X_t and $\ln(X_t)$ are right-skewed and heavy-tailed. The autocorrelation (ACF) plot in Figure 4.3 indicates the high correlation feature of this time series. To capture the prominent jump and mean reversion effects, the CARGPT and CARGPJ models with threshold and jump components are applied to fit the data.

Temperature and weekday effects are included as covariates in the analyses to account for the seasonality at annual and weekly levels (Huisman and Mahieu, 2003, and Rambharat *et al.*, 2005). Owing to the parabolic relationship between temperature $\mathtt{T}_t$ and response X_t, the temperature effect is measured by the absolute deviation from the mean $z_{1t} = |\mathtt{T}_t - \bar{\mathtt{T}}|$, whereas the weekday effect is given by the weekday indicator $z_{2t} = I(t \text{ is a weekday})$. We set some of the hyperparameters in the prior distributions in (4.11) as follow:

$c_\nu = 0.033$, $c_L = 5$, $d_L = 9$, $c_J = 1$, and $d_J = 1$ to ensure that $\nu \in (1, 30)$, $\mathcal{L} \in (5, 9)$ and $E(J) = 1$.

4.4.1.2. *Numerical results.* Table 4.2 reports parameter estimates for Models 0 to 5. The two sets of parameters for Model 3 and 5 are labelled as 'lower' and 'upper', according to $X_{t-d} \leq \mathcal{L}$ and $X_{t-d} > \mathcal{L}$, respectively, where d=1 for simplicity. To enable consistency of some effects across a time period, we set $\beta_{\mu 1g}, \beta_{a0g}$ and β_{a1g} to be consistent across g in Model 5 as in Section 2.4.2 for the LARGP and TLARGP models. Moreover, $\beta_{\mu 0g}, \nu_g$ and τ_g^2 are set to be the same across g because the jump size J_t, allows a different intercept when $X_{t-1} > \mathcal{L}$, and a unit set of parameters for the variance and the degrees of freedom simplifies the error distribution.

Parameters in υ_t and a_t are generally consistent in direction across models. The insignificance of $\beta_{\mu 1}$ except for Models 1 and 2 which have a negative effect, shows that the time series do not demonstrate a strong persistence feature in the model. On the other hand, the positive and significant $m = 1 - \beta_{\mu 2}$ gives the mean reversion rates. For Models 3 and 5, the mean reversion rates are higher when $X_{t-1} > \mathcal{L}$. Moreover the positive and significant $\beta_{\mu 3}$ and $\beta_{\mu 4}$ show that the price increases when the temperature is more different from the overall mean and during weekdays. These covariate effects are more pronounced when $X_{t-1} > \mathcal{L}$. Furthermore, the positive and significant β_{a0} except M0 and M2, indicates a decreasing trend of X_t but the rate of decline drops across time as β_{a1} is negative. The number of degrees of freedom estimate ν shows that the distribution is heavy-tailed and it is more

heavy-tailed when $X_{t-1} \leq \mathcal{L}$ relative to $X_{t-1} > \mathcal{L}$ in Model 3 with no jump component.

The threshold values L for Models 2, 3 and 5 are around 5.5, whereas that for Model 1, it is much higher (8.05) because the only jump size J in Model 1 may be pushed to a higher level to reveal the prominent jump effects. The corresponding threshold levels for electricity price (TLEP) after transformation are \$3137, \$287, \$261 and \$248 per mwh whereas the percentages of days exceeding L are 2.37, 6.66, 7.66, and 7.85, respectively for Models 1 to 3 and 5. Except for Model 1, the estimated TLEPs are close to \$300/mwh which is the current Administered Price Cap (APC) in all regions of NEM in Australia, set in May 2008, and is also the cap of the futures traded in Sydney Futures Exchange (SFE). According to The Rules, if the cumulative spot price over the previous 7 days or 336 half-hour trading intervals exceeds the current Cumulative Price Threshold (CPT) of \$187,500/mwh in the wholesale electricity market, the APC is applied to mitigate financial risk exposure. On the other hand, hedge contracts are used by the generators and wholesalers as agreements, to lock the electricity price for a certain time in the future to minimise their risks. To set up hedge contracts, it is necessary to determine the lock-up price levels. The estimated TLEPs can be used as references in determining the APC for the authority as well as the strike price in a hedge contract.

Table 4.2 shows that model fit improves across Models 0 to 5 in general and Model 5 is the best model, according to the three criteria of DIC, MSE and MAPE as given in (**??**) to (4.13). In Model 5, jumps occur at much higher probability ($\pi = 0.81$) with higher expected jump size $E(J_t)$ (0.39) when $X_{t-1} > \mathcal{L}$. Figure 4.4 illustrates graphically the improvements in model fit. Without both threshold and jump components, Model 0 fails to model the jumps adequately, but although Models 1 to 3 can capture the spikes better than Model 0, they still underestimate the magnitude of jumps. Figure 4.5 displays an enlarged portion of Figure 4.3 for Models 0, 2 and 5 during 31 October to 10 December 2009 when more spikes are observed. On 3, 19, 26 November and 7 December 2009 when the spikes occurred, Model 0 can only allow for part of the jump at lag-1 time whereas Models 2 (threshold model) and 5 model the jump sizes more closely. However, only Model 5 (jump model) can capture the instantaneous jump. Figure 4.6 compares the observed and hypothesized distributions of the standardised residuals

$$S_t = \frac{\ln(X_t) - \upsilon_{xt}}{\tau/\sqrt{\lambda_t}} \sim N(0,1),$$

in Models 0 and 5. The S_t in Model 0 are far more uniform than $N(0,1)$, whereas those for Model 5 approximate closely a truncated $N(0,1)$ because the DMRRPs are capped at \$10000.

4.4.1.3. *Forecasting.* A 1-step ahead forecast is performed using Model 5 for 10 selected days which include 4 days of jumps (12, 21

to 23 January 2010). As mentioned in Section 3.5.5, two types of forecasting values can be evaluated in the Bayesian approach. In this study, the forecasts are based on the first type where the predictive distribution for $\boldsymbol{x}_t$ is given by,

$$f(\boldsymbol{x}_t|\boldsymbol{x}) = \int f_{LT}\left(\boldsymbol{x}_t|\upsilon_t - \ln(a_t^{t-1}) + J_t\, q_t, \tau^2, \alpha\right) f(\boldsymbol{\theta}|\boldsymbol{x})\, d\boldsymbol{\theta}.$$

The predicted values x_t can be sampled successively from

$$x_t|x_{t-1}, \boldsymbol{\theta}^{(i)} \sim LT\left(\upsilon_t^{(i)} - \ln(a_t^{(i)\,t-1}), (\tau^{(i)})^2, \alpha^{(i)}\right).$$

As mentioned in Section 3.5.5, the forecast of X_t can be easily performed in WinBUGS by assigning a missing value 'NA' for the forecast X_t=ln (DMRRP) ($t > 1096$) such that the vector of 1097 observations to be uploaded is $(X_{t-1096}, \ldots X_{t-1}, \text{NA})$. Posterior samples of model parameters together with the missing forecast X_t are then generated using MCMC techniques. Forecast and its prediction interval are given by the posterior mean $\widehat{X}_t$ and the 2.5 ($\widehat{X}_{t,0.025}$) and 97.5 ($\widehat{X}_{t,0.975}$) percentiles of the posterior sample of X_t. The forecasts of the 10 selected days are reported in Table 4.4 together with the observed X_t, temperature effect z_{1t}, weekday effect z_{2t} and forecast error which is defined as $|X_t - \widehat{X}_t|/X_t$. The forecast errors range from 2.93% to 46.94%. Larger forecast errors occur at the days with spikes and among them, 23 January has an unexpected high temperature which pushes the forecast price to a much higher level with an inflated prediction interval. The only exception is on 13 January where a large forecast error is also

observed on a day with no spike. This is because the daily maximum price is mean-reverted to a rather low level after the spike on 12 January 2010. Figure 4.7 plots the observed (circle), forecast (dot) and prediction intervals (lines) for the ten selected days. Table 4.3 and Figure 4.7 show that there are three observations (30%) which lie just outside the prediction intervals. In summary, the forecast performance is satisfactory.

4.4.2. AORD daily range data.

4.4.2.1. *Numerical results.* As Model 5, the jump model with threshold effect, provides the best fit for the electricity price data which exhibit typical jump and mean reversion effects, we apply Model 5 to the AORD daily range data which are adopted in Chapter 2 and 3. We demonstrate that Model 5 is capable of allowing for the jump and mean reversion effects in the data. The data contain one covariate which is the lag-1 daily log return Z_{t-1} to account for the leverage effect as shown in Chapter 2 and 3. Hence, the last term, $\beta_{\mu 4}\ z_{2t}$ in (4.2) for Models 0 to 4 is omitted in this study and $k = 2, 3$ in (4.10) for Model 5. We set some of the hyperparameters in the prior distributions (4.11) as follow:

$c_\nu = 0.033$, $c_L = 1$, and $d_L = 3.5$ to ensure that $\nu \in (1, 30)$, and $\mathcal{L} \in (1, 3.5)$ which is also used as the searching range for threshold levels in Chapter 2. Again, the analyses begin with Model 0 which is actually Model 4 in Chapter 3 (Model 4.3). We apply Model 5 of

this chapter without and with threshold time effect (Model 5T) to the daily range to facilitate the comparison to Model 5 in Chapter 3 (Model 5.3). Table 4.4 reports the parameter estimates and three goodness-of-fit (GOF) measures for Models 4.3, 5.3, 5, and 5T.

The positive and significant $\beta_{\mu 1}$ for Models 5 and 5T supports the persistence effect in this time series, except for Model 5T after $\mathcal{T}_2$=622. Model 5T has more insignificant parameter estimates after $\mathcal{T}_2$=622 because of the reduced number of observations. The positive and significant $\beta_{\mu 2}$ and $m = 1-\beta_{\mu 2}$ gives the mean reversion rates, also except for Model 5T after $\mathcal{T}_2$=622. As $\beta_{\mu 3}$ is negative and significant, a leverage effect is present in the data, except for downward trend AORD daily range data. The positive and significant β_{a0} for Model 5 in this chapter without threshold time, indicates a decreasing trend of the range data but the rate of decline drops across time as β_{a1} is negative, and as a result followed by an increasing trend thereafter. On the other hand, for the model with threshold time, negative and significant β_{a0} indicates an upward trend with increasing rate as β_{a1} is positive for the period before $\mathcal{T}_2$, and vice versa for the period after $\mathcal{T}_2$. The degrees of freedom estimates show that the tail behaviour is moderate for all models whereas the decrease in τ^2s from Models 4.3 and 5.3 to Models 5 and 5T shows that Model 5 can capture jump effects.

For the jump components, the estimated threshold levels for Model 5T are similar to those of the TLARGP models using ML method,

probably because the sampling or search methods for these two models are conducted over the same range of (1,3.5). The proportion of the data before $\mathcal{L}$ that is higher than $\mathcal{L}$=2.595 is 5.48% while the proportion after $\mathcal{L}$ that is higher than $\mathcal{L}$=1.234 is 85.92%. The expected jump size is significant only in Models 5 and 5T for $x_{t-1} < \mathcal{L}$.

Lastly, all three GOF measures indicate a gradual improvement in model fit across Models 4.3, 5.3, 5 and 5T. As a result, Model 5T provides the best model fit. Figure 4.8 compares the observed and fitted using $E(X_t)$ across the 4 models. The fitted values $E(X)$ for all 4 models describe the trend movement well, although it is well understood that the variability is underestimated using $E(X)$. The last plot in Figure 4.8, as well as Figure 4.9 which compares Models 5.3 and 5T, shows that Model 5T capture the jumps better than other models particularly for the instantaneous jumps which occurs at 28 February 2007 and 5 August 2008 which are both detected as outliers in Section 3.5.6. Furthermore, the variability in the fitted values of Model 5T is closer to the observed than that of Model 5.3. Figure 4.10(a) shows the percentage of difference of the fitted values between Model 4 of Chapter 3 and Model 5 of Chapter 4 and (b) for the models with threshold time effect. Two outliers (28 Feb 2007 and 5 Aug 2008) found in Chapter 3 are modelled well in this chapter. It is also convinced by the plot of reciprocal, $1/\lambda_t$ in Figure 4.11(a) and (b). No obvious outlier is detected

for Model 5 in this chapter.

4.4.2.2. *Forecasting.* As in Chapter 2 and 3, forecast of 50 daily ranges is performed using Model 5T. To facilitate comparison, we use the second type of forecast based on $E(X_t)$ as given by

$$E(X_t) = \exp\left[\upsilon_{tg} - \ln(a_g^{t-\mathcal{T}_g}) + J_t\, q_t + \frac{\tau_g^2}{2\lambda_t}\right].$$

which is also adopted in Section 3.5.5.

Figure 4.9 includes the forecast values of $E(X_t)$ (blue line). It shows the forecast values are more volatile than those in Model 5.3 (orange line). The forecast values of TLARGP model by LSE (MSE=0.2248) and ML (MSE=0.2495) method in Chapter 2, Model 5.3 (MSE=0.2714) in Chapter 3 and Model 5T (MSE=0.2459) in this chapter are also displayed in the enlarged figure, Figure 4.12 for comparison. As Model TLARGP using LSE method is based on minimizing the sum of squared errors, it has the smallest MSE among these 4 models. On the other hand, Model 5T is slightly out-performed Model TLARCP using ML method as the former model has a jump diffusion component but they are similar in having both threshold time and threshold level effects. In general, all the forecast performance are satisfactory.

4.5. Conclusion

In this chapter, different models that cater for mean reversion and jump diffusion are considered. In wholesales electricity market, operational planning for the delivery of electricity requires short-term and accurate forecast to facilitate optimal decision making. Statistical models that do not properly capture the essential features of electricity prices, including the strong seasonality, frequent spikes with mean reversions, volatility clustering and long-term trend movement can substantially misprice electricity. In finance, the range models not cater for the infrequent but pronounced jumps and mean reversions will not provide accurate forecast of volatility for risk management. This study proposes three CARGPT models and two CARGPJ models which progressively improve the fit of log DMRRP of NSW, Australia in the first empirical study. The best model (Model 5) also enhanced the fit of the log daily range of AORD Index of Australia in the second empirical study.

In the study, the CARGPJ model with threshold effects (Model 5) is chosen to be the best model, according to three model selection criteria. The model is shown to be a generalisation of the MRJD models, including the TAR model of Rambharat *et al.* (2005), and hence, it allows the jump sizes, jump likelihoods and mean reversion rates to be all price-varied. With a ratio component, the CARR-type mean function,

and heavy-tailed data distribution, it further captures heteroskedasticity, long-term trend movement and leptokurtic data behavior. For model implementation, instead of adopting the two-step approach of Rambharat *et al.* (2005), we propose a full Bayesian approach via the user friendly Bayesian software WinBUGS.

Model performance is evaluated by the in-sample model-fit, out-of-sample forecast accuracy and residual plot. Figures 4.4 and 4.5 for the electricity price data and figures 4.8 to 4.11 for the AORD range data show that Model 5 can capture the spikes well in the in-sample estimation, and figure 4.7, 4.9 and 4.12 further shows that the out-of-sample forecast for both data is satisfactory. Furthermore, for the electricity price data, residual plot in figure 4.6 shows that the standardised residuals agree closely with the hypothesized standard normal distribution, and the estimated threshold levels help to set the control level price for the authority as well as the strike price for a hedge contract.

Table 4.1. Summary statistics for log daily maximum electricity price X_t and $\ln X_t$ for NSW, Australia (1 Jan 2007 - 31 Dec 2009)

	X_t	$\ln(X_t)$
Mean	4.3002	1.4377
Standard Error	0.0302	0.0059
Median	4.1190	1.4156
Mode	3.5553	1.2685
Standard Deviation	0.9993	0.1943
Kurtosis	8.9248	3.7834
Skewness	2.6154	1.5569
Range	6.2813	1.1456
Minimum	2.9291	1.0747
Maximum	9.2103	2.2203
% of Jump (DMRRP>$300)	6.02	
% of Jump (DMRRP>$1000)	3.38	
% of Jump (DMRRP>$3000)	2.55	

Table 4.2. Parameter estimates, standard errors in *italics* and model fit criteria for Models 0 to 5 using the electricity price data.

	M0	M1	M2	M3		M4	M5	
				lower	upper		lower	upper
$\beta_{\mu 0}$	0.1882	0.4226	0.4476	0.2893	1.0820	0.4699	0.4178	-
	0.0309	*0.0311*	*0.0398*	*0.0761*	*0.3255*	*0.0444*	*0.0361*	-
$\beta_{\mu 1}$	0.0903†	-0.0670	-0.0679	0.1646†	-0.1835†	-0.0310†	-0.0297†	-
	0.0785	*0.0146*	*0.0154*	*0.1408*	*0.1407*	*0.0194*	*0.0339*	-
$\beta_{\mu 2}$	0.7181	0.7453	0.7196	0.6090	0.3745	0.6754	0.7047	0.3719
	0.0660	*0.0265*	*0.0322*	*0.0941*	*0.1406*	*0.0308*	*0.0211*	*0.0319*
$\beta_{\mu 3}$	0.0088	0.0114	0.0110	0.0088	0.0534	0.0111	0.0096	0.0478
	0.0014	*0.0012*	*0.0012*	*0.0018*	*0.0103*	*0.0013*	*0.0012*	*0.0079*
$\beta_{\mu 4}$	0.0322	0.0388	0.0403	0.0334	0.1007†	0.0411	0.0478	0.1345
	0.0064	*0.0063*	*0.0064*	*0.0063*	*0.0641*	*0.0063*	*0.0079*	*0.0537*
$1000\beta_{a0}$	-4.3300	0.5021	-0.1120†	0.5128	0.9956†	0.5310	-0.1491†	-
	0.5547	*0.1357*	*0.1186*	*0.2050*	*0.7544*	*0.0957*	*0.1619*	-
$1000\beta_{a1}$	0.6282	-0.0452	0.0383	-0.0465†	-0.0926†	-0.0481	0.0469	-
	0.0799	*0.0190*	*0.0167*	*0.0294*	*0.1053*	*0.0136*	*0.0224*	-
ν	2.14	2.08	2.23	2.65	15.94	2.87	11.04	-
	0.1727	*0.1707*	*0.1857*	*0.2502*	*8.0810*	*0.3665*	*4.2560*	-
τ^2	0.0049	0.0044	0.0045	0.0045	0.0495	0.0045	0.0048	-
	0.0005	*0.0004*	*0.0004*	*0.0005*	*0.0100*	*0.0004*	*0.0004*	-
$\mathcal{L}$	-	8.0510	5.6590	5.5630	-	-	5.5150	-
	-	*0.5753*	*0.3253*	*0.0573*	-	-	*0.0825*	-
J or $E(J)$	-	0.1128	0.0296†*	-	-	0.1725*	0.1054*	0.3630*
	-	*0.0355*	*0.0174*	-	-	*0.0852*	*0.0511*	*0.0350*
π	-	-	-	-	-	0.1046	0.1328	0.8053
	-	-	-	-	-	*0.0498*	*0.0384*	*0.0977*
DIC	1398	1308	1264	1273	-	1015	841	-
MSE	0.5945	0.6334	0.5271	0.1525	-	0.2892	0.0776	-
MAPE	0.0896	0.0890	0.0853	0.0421	-	0.0750	0.0529	-

† p-value > 0.05.

* The estimates are $E(J_t)$.

Table 4.3. Observed, forecast and prediction interval using Model 5 for 10 selected days in Jan 2010 using the electricity price data.

Date	Observed	Temp	Weekday	Forecast	Prediction interval	Forecast error(%)
t	X_t	z_{1t}	z_{2t}	$\widehat{X}_t$	$(\widehat{X}_{t,0.025},\ \widehat{X}_{t,0.975})$	$\frac{\lvert X_t-\widehat{X}_t\rvert}{X_t}$
1/01/2010	3.173	6.745	1	3.997	(3.338, 5.366)	25.98
2/01/2010	3.212	8.138	0	3.430	(2.872, 4.490)	6.80
11/01/2010	4.532	3.226	1	4.258	(3.503, 5.614)	6.05
12/01/2010	7.194	7.723	1	4.634	(3.825, 6.422)	35.59
13/01/2010	3.712	5.122	1	5.455	(3.342, 9.060)	46.94
21/01/2010	5.930	4.821	1	3.904	(3.296, 5.147)	34.17
22/01/2010	8.415	10.026	1	6.767	(4.210, 11.53)	19.58
23/01/2010	7.849	18.716	0	10.400	(6.162, 17.29)	32.51
24/01/2010	3.373	0.418	0	3.713	(2.106, 6.758)	10.06
30/01/2010	3.692	4.714	0	3.800	(3.157, 4.896)	2.93

TABLE 4.4. Parameter estimates, standard errors in *italics* and model fit criteria for Models 4.3, 5.3, 5, and 5T with AORD range data.

	Model 4.3	Model 5.3		Model 5		Model 5T			
		$t < 622$	$t \geq 622$			$t < 622$		$t \geq 622$	
				lower	upper	lower	upper	lower	upper
$\beta_{\mu 0}$	-0.0129	-0.0225	1.0530	-0.0334	-	-0.0306	-	1.6710	-
	0.0061	*0.0084*	*0.3328*	*0.0108*	-	*0.0101*	-	*0.1977*	-
$\beta_{\mu 1}$	0.8269	0.8437	0.1960†	0.7652	-	0.8478	-	-0.1160†	-
	0.0259	*0.0311*	*0.2279*	*0.0408*	-	*0.0304*	-	*0.1633*	-
$\beta_{\mu 2}$	0.1124	0.0972	-0.0734†	0.1197	0.1953	0.0965	0.1177	0.0462†	-0.1206†
	0.0197	*0.0233*	*0.0875*	*0.0246*	*0.0394*	*0.0204*	*0.0591*	*0.3004*	*0.1036*
$\beta_{\mu 3}$	-0.0536	-0.0765	-0.0300	-0.0952	-0.0445	-0.0766	-0.0818	-0.3115†	-0.0192†
	0.0067	*0.0089*	*0.0135*	*0.0221*	*0.0083*	*0.0139*	*0.0189*	*0.2190*	*0.0150*
β_{a0}	0.0038	0.9988	1.0070	0.0041	-	-0.0051	-	0.0219	-
	0.0005	*0.0003*	*0.0007*	*0.0017*	-	*0.0018*	-	*0.0065*	-
β_{a1}	-0.0008	-	-	-0.0008	-	0.0006	-	-0.0029	-
	0.0001	-	-	*0.0003*	-	*0.0003*	-	*0.0013*	-
ν	18.63	21.56	18.44	20.13	-	21.11	-	15.27	-
	5.91	*4.86*	*6.29*	*5.69*	-	*5.54*	-	*6.70*	-
τ^2	0.1441	0.1490	0.1144	0.1359	-	0.1405	-	0.1019	-
	0.0094	*0.0096*	*0.0153*	*0.0094*	-	*0.0112*	-	*0.0163*	-
$\mathcal{L}$	-	-	-	1.4460	-	2.5950	-	1.2340	-
	-	-	-	*0.2209*	-	*0.7234*	-	*0.3518*	-
$E(J)$	-	-	-	0.3007	0.0499†	0.6354	0.0306†	2.0810†	0.0350†
	-	-	-	*0.0940*	*0.0424*	*0.1965*	*0.0325*	*90.600*	*0.0631*
π	-	-	-	0.1162	0.5635	0.0401†	0.4701†	0.0774†	0.4682†
	-	-	-	*0.0531*	*0.2688*	*0.0424*	*0.2650*	*0.0559*	*0.2673*
proportion	100%	81.39%	18.61%	63.96%	36.04%	94.52%	5.48%	14.08%	85.92%
DIC	1047	1037	-	1028	-	1010	-	-	-
MSE	0.4682	0.4216	-	0.4107	-	0.3531	-	-	-
MAPE	0.3623	0.3574	-	0.3400	-	0.3383	-	-	-

† p-value > 0.05.

FIGURE 4.1. Observed log daily maximum of electricity price, for NSW, Australia.

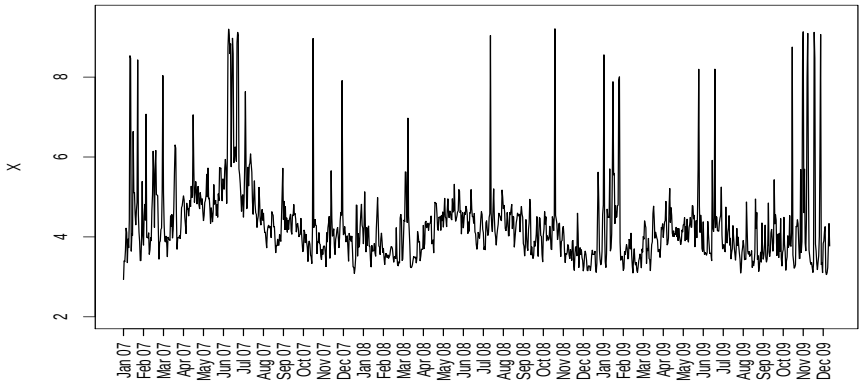


FIGURE 4.2. Histogram of log daily maximum electricity price for NSW, Australia.

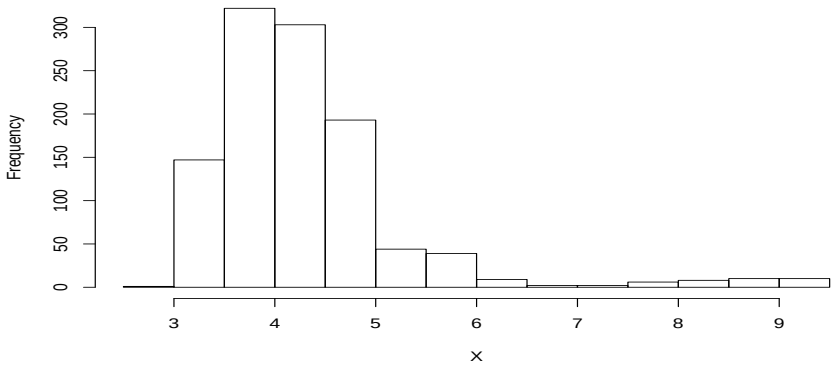


Figure 4.3. Autocorrelation function (ACF) of observed daily maximum of electricity regional reference price.

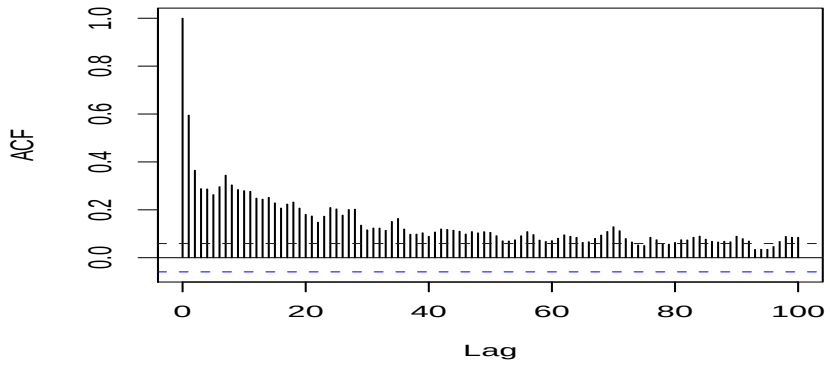


FIGURE 4.4. Observed and fitted values using Models 0 to 5 for the electricity price data.

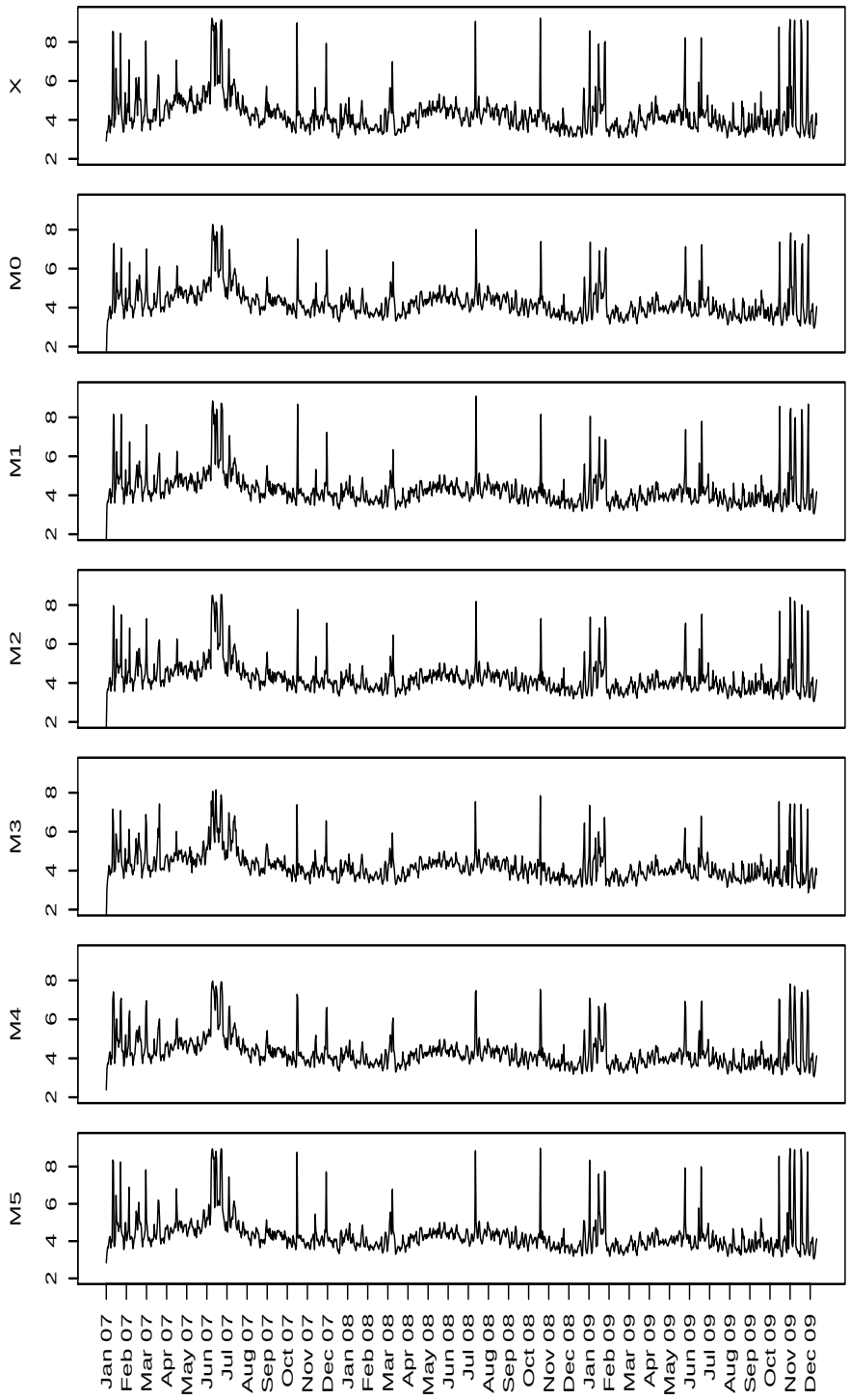


FIGURE 4.5. Observed (black line) and fitted values using Models 0 (green line), 2 (red line) and 5 (blue line) for the electricity price data from 31 Oct 2009 to 10 Dec 2009.

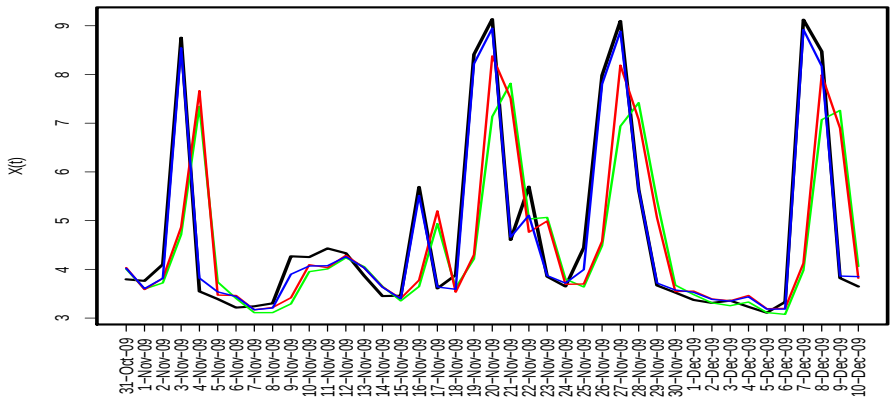


FIGURE 4.6. Observed and hypothesized distribution for standardised residuals in Models 0 (left) and 5 (right) for the electricity price data.

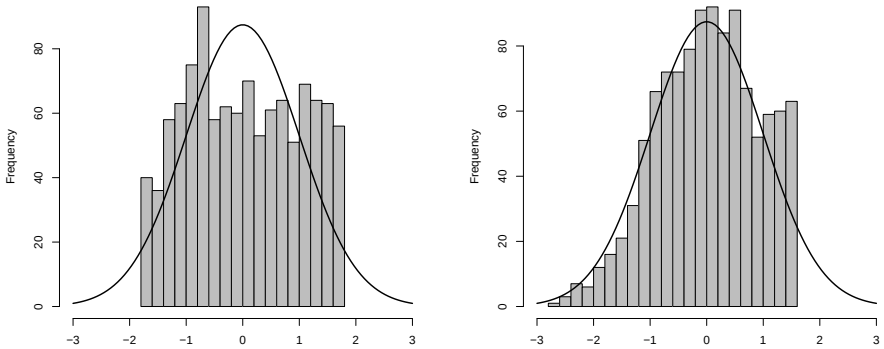


FIGURE 4.7. Observed, forecast and 95% PI using Model 5 for 10 selected days in Jan 2010 for the electricity price data.

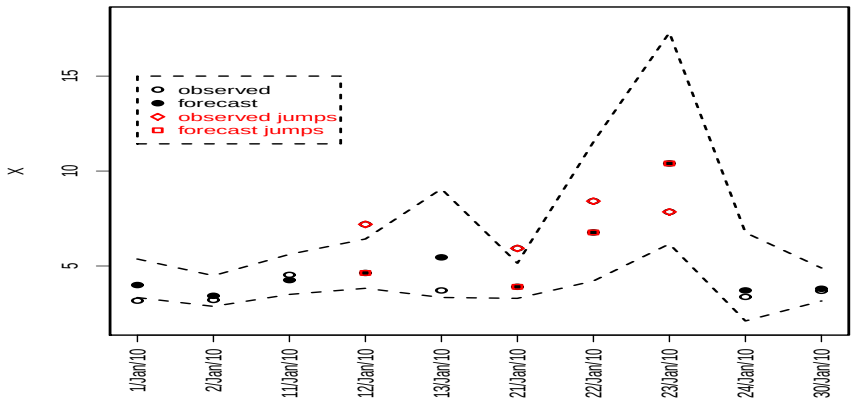


FIGURE 4.8. Observed and fitted values using Models 4.3, 5.3, 5, and 5T for the AORD range data.

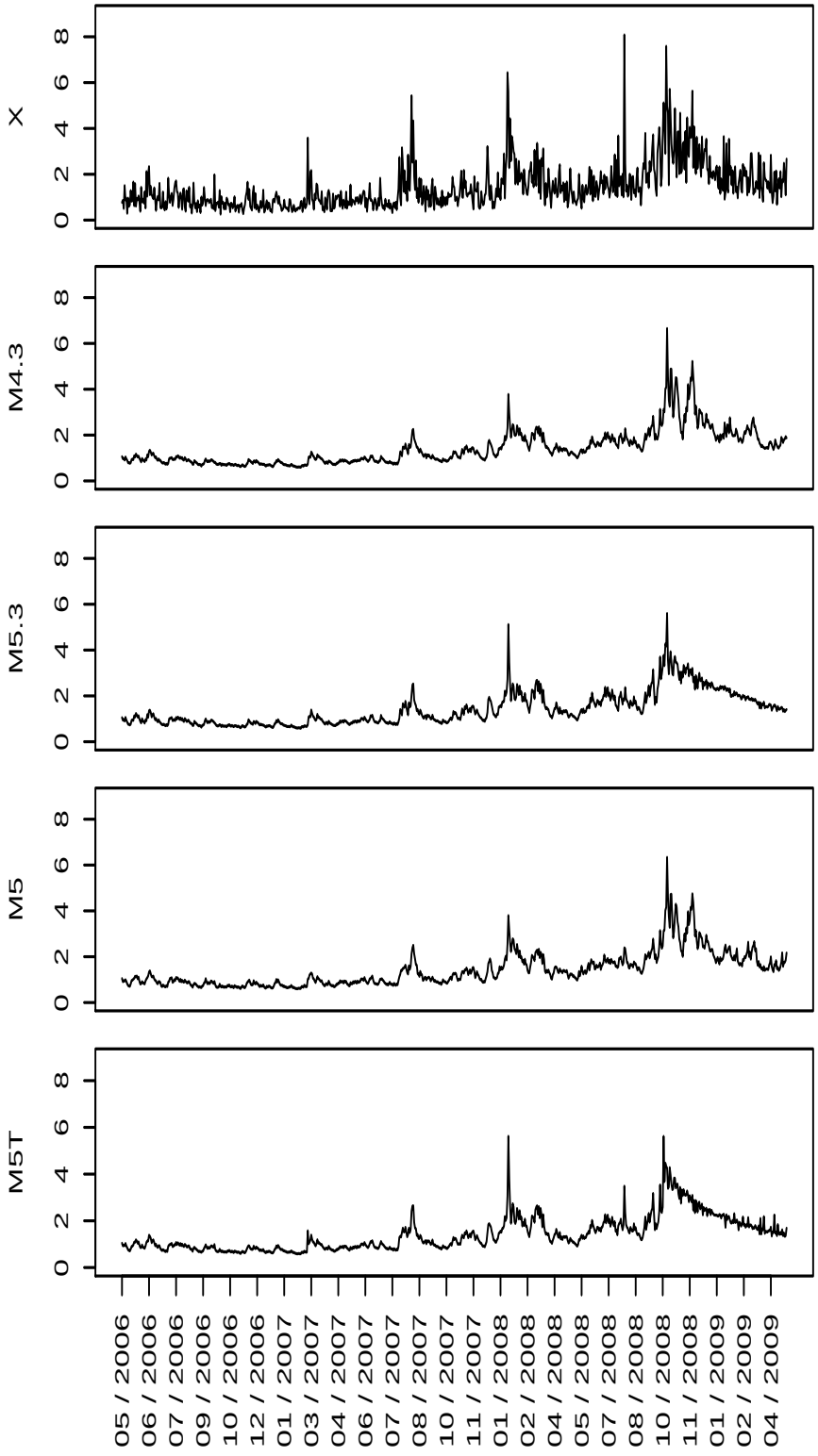


FIGURE 4.9. Observed (black line) and fitted values using Model 5.3 (red line) and forecast (orange line), Model 5T (green line) with threshold levels (green dotted lines) and forecast (blue line) for AORD range data.

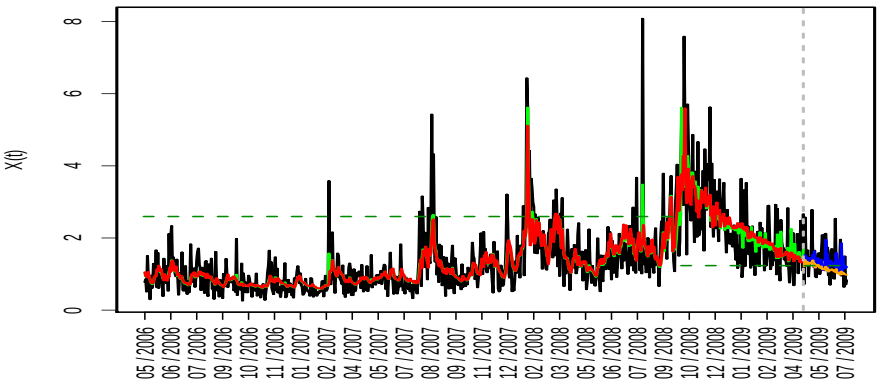


FIGURE 4.10. Percentage of difference of fitted values between (a) Models 4.3 and 5 (b) Models 5.3 and 5T for AORD range data.

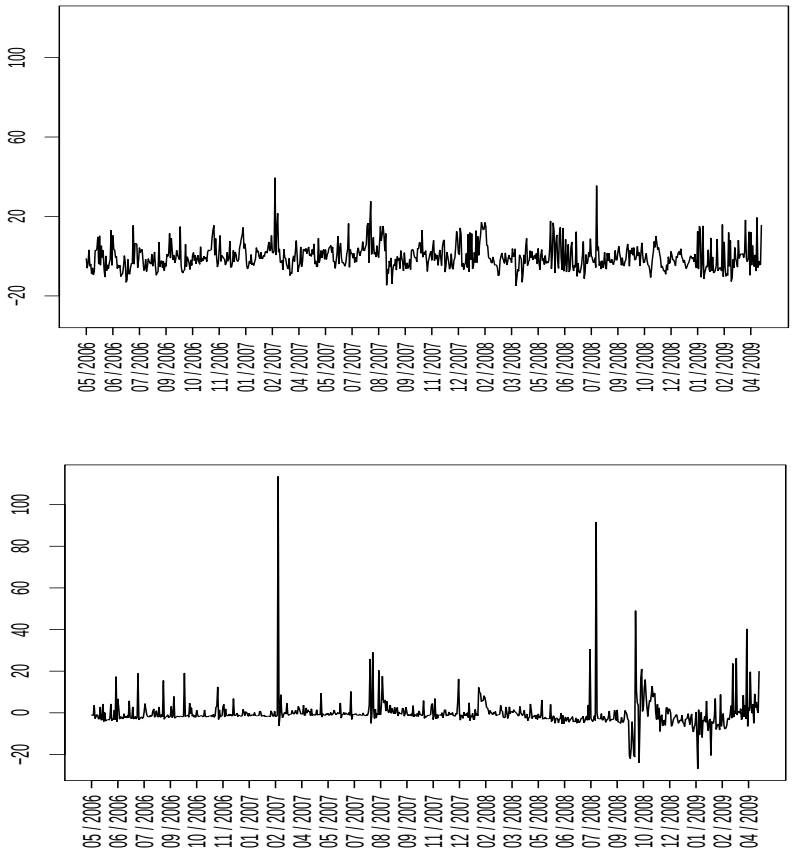


FIGURE 4.11. Inverse lambda values of (a) Model 4.3 (black line) and Model 5 (red line) (b) Model 5.3 (black line) and Model 5T (red line) for AORD range data.

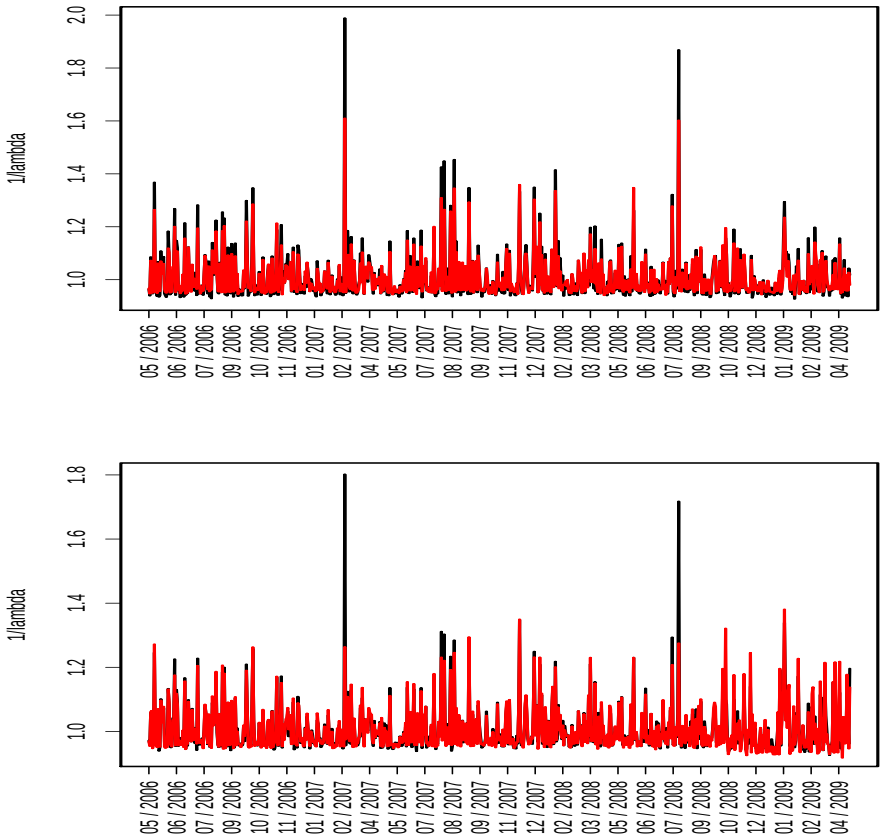
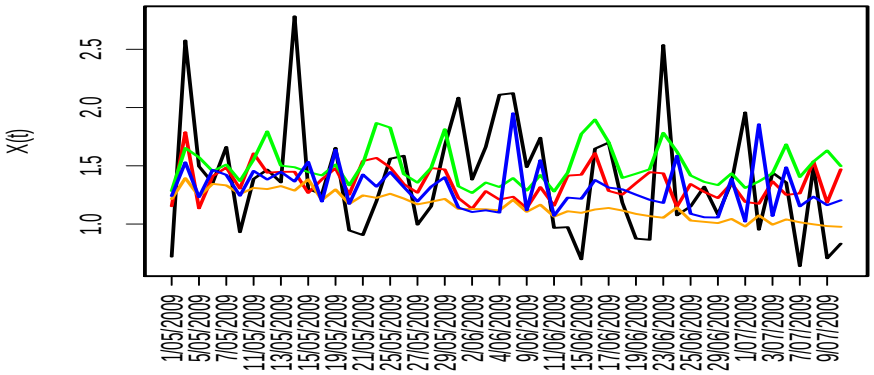


FIGURE 4.12. Observed (black line) and forecast values using TLARGP model by LSE (red line), TLARGP model by ML (green line), CARGP model (orange line), and CARGPJ model (blue line) for AORD range data.



Chapter 5

Overview and further studies

The analysis of a financial time series is of great interest to many researchers. Its contribution to the economy worldwide is highly recognised, as two econometricians Robert F. Engle who proposed an autoregressive conditional heteroskedastic (ARCH) model in 1982, and Clive Granger who developed the concept of cointegration and published a joint paper with Robert Engle in 1987 (Engle and Granger, 1987), were awarded the Nobel Prize in Economics in 2003. To develop some models that can help in understanding and predicting the markets or the economies is very important for the economists, the analysts and even the policy makers in their decision-making processes.

This thesis proposes, develops and improves the GP models to address different features of a financial time series including trend movement, seasonality, non-linearity such as threshold effects, autocorrelation, jumps, mean-reversion, volatility clustering and leptokurtic errors, etc. We begin with the autoregressive GP (ARGP) model in Chapter 2 by adding an autoregressive (AR) term to the mean of the latent stochastic process (SP) which is detrended from the observed data geometrically using a ratio parameter. The time-variant ratio function allow for the *volatility clustering* and *heteroskedasticity*. By taking the

lag-1 log return as a covariate in the mean function, the model allows for *leverage* effect. To further allow for *nonlinearity*, components with threshold effects across time and across levels of some lagged outcome are also included in the mean function.

In order to capture the dynamics of the autocorrelation better, the model is further extended to the conditional ARGP (CARGP) model by adopting a conditional autoregressive range (CARR)-type (Chou, 2005) mean function, in Chapter 3. To allow for *leptokurtic* data distribution for model robustness, we also adopt the heavy-tailed log-t distribution expressed in scale mixtures of normal (SMN) representation for the SP. Simulation and empirical studies show that the models provide accurate parameter estimates and forecasts respectively. In the empirical study, the model outperforms the original CARR models in both in-sample estimation and out-of-sample forecast. However, although the SMN representation of Student's t-distribution in the model can help to identify and downweight the effect of outliers which are the jumps in the data using the mixture variable λ_i, the models still fail to capture adequately the magnitude of the jumps.

To capture the *jumps* and possibly their *mean reversions* effectively, we propose further extensions of the CARGP model in Chapter 4, by adding components that describe the threshold and jump effects in the mean function of the SP. These two components capture the jump and

mean reversion features in a financial time series. The model can further capture *seasonality* by including a covariate in the mean function. We show that the extended CARGP Jump (CARGPJ) models are the generalisation of the mean reverting jump diffusion (MRJD) model of Rambharat *et al.* (2005). The results of two empirical studies show that the CARGPJ model can capture the jumps and mean reversions well, and hence, facilitate accurate forecast.

For model implementation, the non-parametric least square error (LSE) method and parametric maximum likelihood (ML) method are used for the inference of the ARGP models. The first and second order derivatives of the least square and log-likelihood functions are derived to facilitate the implementation in R. However, the likelihood functions become particularly complicated when the GP model adopts a CARR-type mean function. Hence, it is more appealing to use the Bayesian approach for the inference of the CARGP model and its extensions. The Bayesian software, Bayesian analysis Using Gibbs Sampling (WinBUGS; Spiegelhalter *et al.*, 2004) allows us to implement the model and performs forecast fairly easily. However, in order to facilitate the Markov chain Monte Carlo (MCMC) sampler, the full conditional distributions for the parameters in threshold CARGP model are also derived and reported in the Appendix of Chapter 3 for reference.

In summary, the proposed GP models are capable of modelling a financial time series which exhibits trend movement, heteroskedasticity,

leptokurtic data distribution, jumps and mean reversions. However, there are a few limitations of the GP models in modelling financial time series. One of them is the model failure in describing the enlarged volatility for the downward trend data. This is a challenging and favourable matter in further development of GP models on the financial time series. There are still many research areas that we can go on in modelling a financial time series with GP models. We state some as below.

For modelling methodologies, the CARGPJ model with a jump diffusion component is shown to be the most capable model for capturing the spikes. However, it still fails to model the spikes adequately in some cases. Perhaps the likelihood of jumps should be modelled using some dynamic modelling techniques that involve covariates. A model with dynamic jump probability provides more information to the likelihood of jumps, and hence improves the forecast accuracy for data with spikes.

Studies should also be directed to the choice of distributions which serves as an alternative modelling technique for the jumps. Heavy-tailed distributions, including the generalised error (exponential power) distribution with leptokurtic and platykurtic shapes, the generalised-t distribution which encompasses uniform, t and generalised error distributions, etc and their asymmetric extensions, have different skewness

and kurtosis to provide more flexible and adaptive models, and hence to foster better understanding of market activity.

The models considered in this thesis focus on modelling the conditional mean of a distribution. However, attention should also be drawn to the modelling of different quantiles of the data, particularly the upper quantiles which are important for risk management. Nonparametric quantile regression is first introduced by Koenker and Bassett (1978) to model the conditional quantile functions. Extension of the CARGP model to different quantile functions is feasible and promising as the model is robust to outliers, and can provide a more comprehensive view on the covariates effects across quantiles.

Univariate model is considered in this thesis. However, the price dynamics in most financial time series are often more complicated than a univariate model can describe, because a univariate model ignores the cross correlation among a number of interrelated time series from the same industry and/or region. For example, the electricity prices of New South Wales and Victoria are highly correlated due to their geographical proximity and the number of the interconnectors. Extension to multivariate CARGP model allows the cross correlation to be modelled in the correlation matrix of the multivariate distribution. This simultaneous modelling of a set of asset prices is practical and useful in portfolio setting.

Although Bayesian inference is adopted in Chapters 3 and 4 for its simplicity in model implementation particularly with the use of WinBUGS, the ML method is feasible for the CARGP model and its extension. The log-likelihood function and its derivatives can be derived without the need of integration based on (3.7). From our experience, estimates using the ML method of inference are often more efficient and, more importantly, the prolonged MCMC computation for complicated models can be avoided. Hence, the ML method should be considered as an alternative method of inference for the CARGP model and its extensions.

Lastly, in application, daily range and electricity price have been fitted to the proposed models. As range data is sensitive to outliers, Chou (2005) suggested using other frequency range intervals, say every hour or every quarter. With the rapid development of market automation and computational power, high frequency data (HFD) emerge and fuel the development and application of CARGP model to high frequency range data that may exhibit different trend movements and volatility dynamics.

References

[1] Akaike, H. (1973) Information theory and an extension of the maximum likelihood principle, *Proceedings of 2nd International Symposium on Information Theory*, 267-281.

[2] Alizadeh, S., Brandt, M.W., and Diebold, F.X. (2002) Range-based estimation of stochastic volatility models or exchange rate dynamics are more interesting than you think, *Journal of Finance*, **57**, 1047-1092.

[3] Andersen, T. and Bollerslev, T. (1998) Answering the skeptics: Yes, standard volatility models do provide accurate forecasts, *International Economic Review*, **39**, 885-905.

[4] Andrews, D.F. and Mallows, C.L. (1974) Scale mixtures of normal distributions, *Journal of The Royal Statistical Series, Series B*, **36**, 99-102.

[5] Ascher, H. (1981) Weibull distribution vs weibull process, *Proceedings Annual Reliability and Maintainability Symposium*, 426-431.

[6] Australia energy market commission (2008) *Determination of Schedule for the Administered Price Cap, May 2008*, http://www.aemc.gov.au/Market-Reviews/Completed.html, ref EMO0003.

[7] Australia energy market operator (2009) *An Introduction to Australia's National Electricity Market, Decemeber 2009*, http://www.aemo.com.au/corporate/publications.html, 0000-0262.pdf.

[8] Barlow, M. T. (2002) A diffusion model for electricity prices, *Mathematical Finance*, **12**, 287-298.

[9] Barz, G. and Johnson, B. (1998) Modeling the prices of commodities that are costly to store: the case of electricity, *Proceeding, Risk Management Conference, Chicago.*

[10] Beckers, S. (1983) Variances of security price returns based on high, low, and closing prices, *Journal of Business*, **56**, 97-112.

[11] Black F. (1976) Studies of stock price volatility changes, *Proceedings of the 1976 Meetings of the American Statistical Association, Business and Economical Statistics Section*, 177-181.

[12] Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics*, **31**, 307-328.

[13] Box, G.E.P. and Jenkins, G.M. (1976) *Time series analysis: forecasting and control*, Holden-Day, San Francisco.

[14] Box, G.E.P., Jenkins, G.M., and Reinsel G. (1994) *Time series analysis: forecasting and control, 3rd edition*, Prentice Hall, Englewood Cliffs, NJ.

[15] Brandt, M. and Jones, C. (2006) Volatility forecasting with range-based EGARCH models. *Journal of Business & Economic Statistics*, **24**, 470-486.

[16] Cartea A. and Figueroa M. (2005) Pricing in electricity markets: a mean reverting jump diffusion model with seasonality, *Applied Mathematical Finance*, **12**, 313-335.

[17] Chan, J.S.K., Lam, Y., and Leung, D.Y.P. (2004) Statistical inference for geometric processes with gamma distribution, *Computational Statistics and Data Analysis*, **47**, 565-581.

[18] Chan, J.S.K., Yu, P.L.H., Lam, Y., and Ho, A.P.K. (2006) Modelling SARS data using threshold geometric process, *Statistics in Medicine*, **25**, 1826-1839.

[19] Chan, J.S.K. and Leung, D.Y.P. (2010) A new approach to the modelling longitudinal binary data with trend: the binary geometric process model, *Computational Statistics*, **25**, 505-536.

[20] Chan, J.S.K., Wan, W.Y., Lee, C.K., Lin, C.K., and Yu, P.L.H. (2011) Predicting drop-out and committed first time blood donors: a Poisson geometric process approach. Submitted for publication.

[21] Chen, C.W.S. and Lee, J.C. (1995) Bayesian inference of threshold autoregressive models, *Journal of Time Series Analysis*, **16**, 483-492.

[22] Chen, C.W.S., Gerlach, R., and Lin, E. M. H. (2008) Volatility forecast using threshold heteroskedastic models of the intra-day

range, *Computational Statistics & Data Analysis*, on Statistical & Computational Methods in Finance, **52**, 2990-3010.

[23] Chen C.W.S., Gerlach, R., Choy, S.T.B., and Lin, C. (2010) Estimation and inference for exponential smooth transition nonlinear volatility models, *Journal of Statistical Planning and Inference*, **140**, 719-733.

[24] Chiu, H.C. and Wang, D. (2006) Using conditional autoregressive range model to forecast volatility of the stock indices. *Proceedings of Joint Conference on Information Science 2006.* Atlantis Press.

[25] Chou, R.Y. (2005) Forecasting financial volatilities with extreme values : the conditional autoregressive range (CARR) model, *Journal of Money Credit and Banking*, **37**, 561-582.

[26] Chou, R.Y. (2006) Modeling the asymmetry of stock movements using price ranges, *Advances in Econometrics*, **20**, 231-257.

[27] Choy, S.T.B. and Smith, A.F.M. (1997) Hierarchical models with scale mixtures of normal distribution, *TEST*, **6**, 205-221.

[28] Choy, S.T.B. and Chan, J.S.K. (2008) Scale mixtures distributions in statistical modelling, *Australian and New Zealand Journal of Statistics*, **50**, 135-146.

[29] Contreras, J., Espinola, R., Nogales, F.J., and Conejo, A.J. (2003) ARIMA models to predict next-day electricity prices, *IEEE Transactions on Power Systems*, **18**, 1014-1020.

[30] Cox, D.R. and Lewis, P.A.W. (1966) *The statistical analysis of series of events*, Chapman and Hall, UK.

[31] Engle, R.F. (1982) Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation, *Econometrica*, **50**, 987-1007.

[32] Engle, R.F. and Granger, C.W.J. (1987) Co-integration and error correction: representation, estimation and testing, *Econometrica*, **55**, 251-276.

[33] Feller, W. (1949) Fluctuation theroy of recurrent events, *Transactions of the American Mathematical Society*, **67**, 98-119.

[34] Fleming, J., Kirby, C., and Ostdiek, B. (2003) The economic value of volatility timing using "realized volatility", *Journal of Financial Economics*, **67**, 473-509.

[35] Garman, M.B. and Klass, M.J. (1980) On the estimation of security price volatilities from historical data, *Journal of Business*, **53**, 67-78.

[36] Geweke, J. and Terui, N. (1993) Bayesian threshold autoregressive models for nonlinear time series, *Journal of Time Series Analysis*, **14**, 441-454.

[37] Gilks, W.R., Richardson, S., and Spiegelhalter, D.J. (1996) *Markov Chain Monte Carlo in Practice*, Chapman and Hall, UK.

[38] Granger, C.W.J. and Anderson A.P. (1978) *An introduction to bilinear time series models*, Vandenhoeck and Ruprecht, GÃűttingen,

94.

[39] Hastings, W.K. (1970) Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, **57**, 97-109.

[40] Higgs, H., Worthington, A. (2005) Systematic features of high-frequency volatility in the Australian electricity market: intraday patterns, information arrival and calender effects, *The Energy Journal*, **26**, 1-20.

[41] Huisman, R. and Mahieu, R. (2003) Regime jumps in electricity prices, *Energy Economics*, **25**, 425-434.

[42] Hull, J. and White, A. (1987) The pricing of options on assets with stochastic volatility, *Journal of Finance*, **42**, 281-300.

[43] Koenker, R. and Bassett, G. (1978) Regression Quantiles, *Econometrica*, **46**, 33-50.

[44] Kunitomo, N. (1992) Improving the Parkinson method of estimating security price volatilities, *Journal of Business*, **65**, 295-302.

[45] Lam, Y. (1988) Geometric process and replacement problem, *Acta Mathematicae Applicatae Sinica*, **4**, 366-377.

[46] Lam, Y. (1992) Nonparametric inference for geometric processes. *Communications in Statistics: Theory and methods*, **21**, 2083-2105.

[47] Lam, Y. (2007) *The geometric process and its applications*. Singapore: World Scientific Publishing Co. Pte. Ltd.

[48] Lam, Y. and Chan, J.S.K. (1998) Statistical inference for geometric processes with lognormal distribution, *Computational Statistics and*

Data Analysis, **27**, 99-112.

[49] Lam, Y., Zhu, L.X., Chan, J.S.K., and Liu, Q. (2004) Analysis of data from a series of events by a geometric process model. *Acta Mathematica Applicatae Sinica*, **20**, 263-282.

[50] Lu, X., Dong, Z.Y., and Li, X. (2005) Electricity market price spike forecast with data mining techniques, *Electric Power Systems Research*, **73**, 19-29.

[51] Mandelbrot, B. (1963) The variation of certain speculative prices, *Journal of Business*, **36**, 394-419.

[52] Marsh, T.A. and Rosenfeld, E.R. (1986) Non-trading, market making, and estimates of stock price volatility, *Journal of Financial Economics*, **15**, 359-372.

[53] Melino, A. and Turnbull, S.M. (1990) Pricing foreign currency options with stochastic volatility, *Journal of Econometrics*, **45**, 239-265.

[54] Merton, R.C. (1973) Theory of rational option pricing, *Bell Journal of Economics and Management Science*, **4**, 141-183.

[55] Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N. and Teller, A.H. (1953) Equations of state calculations by fast computing machines, *Journal of Chemical Physics*, **21**, 1087-1091.

[56] Nogales, F.J. and Conejo, A.J. (2006) Electricity price forecasting through transfer function models, *Journal of the Operational Research Society*, **57**, 350-356.

[57] Ntzoufras, I. (2009) *Bayesian Modeling Using WinBUGS*, New Jersey: Wiley, 389-390.

[58] Parkinson, M. (1980) The extreme value method for estimating the variance of the rate of return, *Journal of Business*, **53**, 61-65.

[59] Rambharat, B. R., Brockwell, A.E., and Seppi, D.J. (2005) A threshold autoregressive model for wholesale electricity prices, *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, **54**, 287-299.

[60] Robert, C.P. (1995) Simulation of truncated normal variables, *Statistics and Computing*, **5**, 121-125.

[61] Robinson, P.M. (1978) Statistical inference for a random coefficient autoregressive model, *Scandinavian Journal of Statistics*, **5**, 163-168.

[62] Schwartz, E.S. (1997) The stochastic behaviour of commodity prices: implications for valuation and hedging, *Journal of Finance*, **52**, 923-973.

[63] Spiegelhalter, D., Best, N.G., Carlin, B.P., and van der Linde, A. (2002) Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B*, **64**, 583-639.

[64] Spiegelhalter, D., Thomas, A., Best, N.G., and Lunn, D. (2004) Bayesian inference using Gibbs sampling for Window version (WinBUGS) Version 1.4.1. The University of Cambridge. `www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml`.

[65] Smith, M.S. (2010) Bayesian inference for a periodic stochastic volatility model of intraday electricity prices. *In Statistical Modelling and Regression Structures: Festschrift in Honour of Ludwig Fahrmeir, Kneib T, Tutz G (eds)*, Springer: Berlin, 353-376.

[66] Smith, A.F.M. and Roberts, G.O. (1993) Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods, *Journal of the Royal Statistical Society, Series B*, **55**, 3-23.

[67] Taylor, S.J. (1994) Modelling stochastic volatility: A review and comparative study, *Mathematical Finance*, **4**, 183-204.

[68] Thomas, S. and Mitchell, H. (2005) *GARCH Modelling of High-Frequency Volatility in Australia's National Electricity Market.* Discussion paper. Melbourne Centre for Financial Studies.

[69] Tiwari, R. C., Cronin, K.A., Davis, W., and Feuer, E.J. (2005) Bayesian model selection for joint point regression with application to age-adjusted cancer rates. *Appl. Statist.*, **54**, 919-939.

[70] Tong, H. and Lim, K. S. (1980), Threshold autoregression, limit cycles and cyclical data (with discussion), *J. Roy. Statist. Soc. Ser. B*, **42**, 245-292.

[71] Wallis, K.F. (1974) Seasonal adjustment and relations between variables, *Journal of the American Statistical Association*, **69**, 18-31.

[72] Wan, W.Y. and Chan, J.S.K. (2009) A new approach for handling longitudinal count data with zero inflation and overdispersion: Poisson Geometric Process model. *Biometrical Journal*, **51**, 556-570.

[73] Wan, W.Y. and Chan, J.S.K. (2011) Bayesian analysis of robust Poisson geometric process model using heavy-tailed distributions, *Computational Statistics and Data Analysis*, **55**, 687-702.

[74] Wiggins, J.B. (1991) Empirical tests of the bias and efficiency of the extreme-value variance estimator for common stocks. *Journal of Business*, **64**, 417-432.

[75] Wiggins, J.B. (1992) Estimating the Volatility of S&P 500 Futures Prices Using the Extreme-Value Method, *Journal of Futures Markets*, **12**, 265-273.

[76] Zhang, L., Luh, P.B., and Kasiviswanathan, K. (2003) Energy Clearing Price Prediction and Confidence Interval Estimation with Cascaded Neural Networks, *IEEE Transactions on Power Systems*, **18**, 99-105.

[77] Zekkberm, Arnold (1979) Seasonal Analysis of Economic Time Series, NBER, 281-308.