Further Developments of Two Point Process Models for Fine-Scale Time Series

A thesis presented in partial fulfilment of the
requirements for the degree of

Doctor of Philosophy
in Statistics

at Massey University,
Albany (Auckland), New Zealand.

Gang Xie

2011

Abstract

Two point processes, the Autoregressive Conditional Duration (ACD) model and the Bartlett-Lewis Pulse (BLP) model, are further developed and used to model fine scale time series.

Six different ACD models are specified and fitted to two sets of real stock transaction data. The Akaike Information Criterion (AIC), Takeuchi Information Criterion (TIC) and Generalized Information Criterion (GIC) are the information-theoretic criteria for model evaluation. This is the first time that ACD models have been evaluated and ranked based on the information-theoretic criteria, and makes the comparison of ACD models with different autoregressive and error structures straightforward. A newly proposed ACD model with a mixed lognormal-gamma error term distribution is identified as the best model with minimum predictive error.

The original BLP model was developed and fitted to 60 years of 5-minute rainfall series from Kelburn, a place near Wellington, New Zealand, by Cowpertwait et al. (2007). The BLP model can be used to produce realistic simulation samples of fine scale rainfall series that are required in many applications, e.g. urban drainage system design. Following the continuous distributions of storm types approach as first proposed by Cowpertwait (2010), a more general BLP process characterization framework is formulated under which the original BLP model can be recovered as a special case. Statistical properties up to third order are derived for the BLP model characterized by continuous distributions of storm types. Without an increase in model parameters, a modified BLP model is specified, in which a conditional mean exponential distribution is used to represent the pulse depths distribution and a continuum of storm types within a process is assumed. Simulation studies show that the modified model improves the fit to the observed proportion of dry periods significantly, whilst retaining a good fit to moment properties. Also a better fit is obtained to the annual extreme rainfall values at the 5 minutes and 12 hours aggregation levels whilst retaining a good fit at other aggregation levels, and significant improvement is achieved in the goodness-of-fit to extremes for the individual months. The improvements of the original BLP model's performance are mainly due to the successful implementation of the within cell pulse depths dependence structure using a conditional mean exponential distribution.

Acknowledgements

Firstly, I would like to thank my supervisor, Associate Professor Paul Cowpertwait. I thank him for his careful planning, academically rigorous supervision, patient and to-the-point guidance in my studies. I thank my second supervisor, Dr Daniel Walsh, for his responsive professional input and sincerity. I would also like to thank Associate Professor Xiaoming Li, for being the co-supervisor of my first year study. I am grateful to Professor Valerie Isham and Dr Christian Onof for their valuable suggestions and comments on my research results of the empirical study of BLP models. I would like to express my appreciation to Dr Winston Sweatman for his assistance in vector calculus and help with some computation issues.

I am very grateful to Freda Mickisch for her effective support, in particular her kindness, flexibility, patience, and tolerance in coordinating my PhD studies. I would also like to thank Michelle Campbell, Annette Warbrooke, Colleen Van Es for their help with office matters. I owe thanks to the IT support team. Particular thanks go to Rushad Irani, Mike Yap, and Yan Ou, for their prompt and quality support that has made me 'worry free' with my computer. My sincere thanks also go to many of those people who have helped me in one way or the other but not been named here.

My fellow students Amanda Jane Elvin, Sena Galkadowite, Sharleen Harper, Joanne Mann, Qing Zhang, Haydn Cooper, Maarten McKubre-Jordens, Sam Dillon, Guy Kloss, Cho Hong Ling (Joe), Jingning Li, Norazlina Ismail, Adam Smith, Insha Ullah, Katharina Parry, and Oliver Hannaford have provided a great sense of community. I would like to give special thanks to Guy, Maarten, and Amanda for their kind help in using LaTeX.

I have really benefited from attending the weekly statistics group meetings and learning what expertise other statisticians have. It has given me a sense of belonging to a small but vital statistics family. I appreciate the opportunities I have been given in participation of the various of statistics teaching activities, e.g. assignment marking,

computer lab tutoring, summer school tutoring, supervision of exams, etc. In particular, I have had a great experience in tutoring the summer school workshops for the past three years in cooperation with Marie Fitch.

I am grateful to Massey University for the Doctoral Scholarship funding, and to the Institute of Information and Mathematical Sciences for hosting me for the last three plus years by providing quality facilities and services, and support for conferences.

Finally, I would like to thank my family. I dedicate this thesis to my wife, Min Zhou. I thank my son, David, for being an outstanding school graduate a father can be proud of.

Contents

List of Figures

List of Tables

Main abbreviations and symbols

$\exp(\cdot)$	exponential function such that $\exp(1) = e$
$\log(\cdot)$	the natural logarithm function such that $\log(e) = 1$
iid	independent and identically distributed
pdf	probability density function
pmf	probability mass function
Pr[ ], or P[ ]	probability
E[ ]	expected value, mean value
Var[ ]	variance
Cov[ ]	covariance
$L(\underline{\theta})$	likelihood function for the parameter $\underline{\theta}$
MLE	maximum likelihood estimator or maximum likelihood estimate
QMLE, quasi MLE	quasi-maximum likelihood estimator or estimate
M-estimator	maximum likelihood-type estimator
EDA	exploratory data analysis
qq-plot	quantile-quantile plot
$G(x)$, $g(x)$	the distribution and density functions representing the true distribution from which the observed sample data are generated
$\hat{G}$, $\hat{g}$	the empirical distribution and density functions of the true distribution
$F(x\|\underline{\theta})$ or $f(x\|\underline{\theta})$	a fitted distribution or density function, which is specified to approximate the true distribution
$\underline{x} = \{x_1, x_2, ..., x_n\}$	a random sample data set of n observations generated from the true distribution $G(x)$
$\{y_t\} = \{y_1, y_2, ..., y_n\}$	an observed time series which consists of n values sampled at discrete times $1, 2, ..., n$
AR model	(standard time series) Autoregressive model
MA model	(standard time series) Moving-average model
ARMA model	(standard time series) Autoregressive Moving-average model
L-B statistic	Ljung-Box statistic (see Section A.3)
Q-statistic	Pearson's Q-statistic (see Section 2.5.2)

K-L information	Kullback-Leibler information
$\mathbf{I}(g; f)$	symbol for K-L information (see Section 2.3.2)
AIC	Akaike Information Criterion
AICc	the small sample bias-corrected AIC
BIC	Bayesian Information Criterion
DIC	Deviance Information Criterion
TIC	Takeuchi Information Criterion
GIC	Generalized Information Criterion
EMD	effective model dimension
W_i	Akaike weights (see Section 2.4.1)
ACD model	Autoregressive Conditional Duration model (finance model)
EACD(1,1), EACD(2,2)	the first, second order ACD models that have an exponential distribution error term structures (see Section 3.3.1)
WACD(1,1), WACD(2,2)	the first, second order ACD models that have a Weibull distribution error term structures (see Section 3.3.1)
mixedE-ACD(1,1)	a first order ACD model that has a mixed exponential distribution error term (see Section 3.3.2)
mixedLG-ACD(1,1)	a first order ACD model that has a mixed lognormal-gamma distribution error term (see Section 3.3.2)
NSRP model	Neyman-Scott rectangular pulse model (rainfall model)
BLRP model	Bartlett-Lewis rectangular pulse model (rainfall model)
BLP model	Bartlett-Lewis pulse model (rainfall model)
CV	coefficient of variation
AC	autocorrelation
PD	proportion dry
FIT-O1	the original BLP model specification, single process of storms (six parameters in total) (see Section 6.3)
FIT-O	the original BLP model specification, superposition of two independent FIT-O1 processes (11 parameters in total) (see Section 6.3)
FIT-C1	based on FIT-O1, but with a within cell pulse depths dependence structure characterization (see Section 6.3)
FIT-C	superposition of two independent FIT-C1 processes (11 parameters in total) (see Section 6.3)
FIT-C-PD	based on FIT-C, but with the sample PD properties further included in the fitting procedure (see Section 6.3)

FIT-C1-βzθz — based on FIT-C1, specification with continuous distributions of storm types in terms of cell origins generation rate $\beta(z) = \beta_0(1 - z) + \beta_1 z$ and mean pulse depth $\theta(z) = \theta z$ (see Section 7.2.3)

FIT-C1-ξzθz — based on FIT-C1, specification with continuous distributions of storm types in terms of pulse generation rate $\xi(z) = \xi_0(1-z)+\xi_1 z$ and mean pulse depth $\theta(z) = \theta z$ (see Section 7.2.3)

FIT-C1-βzξz — based on FIT-C1, specification with continuous distributions of storm types in terms of cell origins generation rate $\beta(z) = \beta_0(1 - z) + \beta_1 z$ and pulse arrival rate $\xi(z) = \xi_0(1 - z) + \xi_1 z$ (see Section 7.2.3)

FIT-C1-βzξzθz — based on FIT-C1, specification with continuous distributions of storm types in terms of cell origins generation rate $\beta(z) = \beta_0(1 - z) + \beta_1 z$, pulse arrival rate $\xi(z) = \xi_0(1 - z) + \xi_1 z$, and mean pulse depth $\theta(z) = \theta z$ (see Section 7.2.3)

FIT-C-8 — based on FIT-C, but with the assumption of a common mean storm lifetime, a common rain cell origins generation rate and common mean cell duration for the two distinct processes of storms (a reduced FIT-C, eight parameters in total)

FIT-C-9 — based on FIT-C, but with the assumption of a common mean storm lifetime, a common cell origins generation rate for the two distinct processes of storms (a reduced FIT-C, nine parameters in total)

FIT-C-10 — based on FIT-C, but with the assumption of a common mean storm lifetime for the two distinct processes of storms (a reduced FIT-C, 10 parameters in total)

FIT-C1-βzθz(6) — based on FIT-C1, specification with continuous distributions of storm types in terms of cell origins generation rate $\beta(z) = \beta(1-z)$ and mean pulse depth $\theta(z) = \theta z$ (six parameters in total, see Section 8.1.1)

FIT-C1-βzθzE — based on FIT-C1, specification with continuous distributions of storm types in terms of cell origins generation rate $\beta(z) = \beta e^{-z}$ and mean pulse depth $\theta(z) = \theta z$ (six parameters in total, see Section 8.1.1)

FIT-C-βzθzE — superposition of two independent FIT-C1-βzθzE processes

Chapter 1

Introduction

1.1 Motivation and research objectives

1.1.1 Point processes and fine-scale time series

A stochastic process is a sequence of random variables generated by probabilistic laws. *Point processes* are a particular kind of stochastic process that, in general, describe the random scattering of point events in a certain space. Two point processes, which model random point events occurring along a time axis, are studied in this thesis.

Examples of point processes can be found in many fields. For example, an emission process of radioactive particles, where each emission is recorded in time, form an irregular point process sequence. The locations of a species of plant in a field is an example of spatial point process. The historical records of the location and time of earthquake occurrences, say, in New Zealand, represent a spatio-temporal point process series. Arrival times of customers at a fast-food counter, arrival times and length of stay of patients in an emergency ward, times and locations of residential house sales in Auckland region, and times and durations of telephone calls in a call center, etc, are the situations in everyday life which can all be considered as point processes. Point processes are the stochastic models formulated to analyse point occurrence phenomena. The above examples illustrate the breadth of potential applications of point processes.

Historical series are treated as realizations of sequences of random variables. The so-called high frequency financial data, which record transaction-by-transaction observations in (near) continuous time, may be treated as a fine-scale point process series (the transaction times are measured up to one second accuracy). The model per-

formances of Autoregressive Conditional Duration (ACD) models (Engle and Russell, 1998) in fitting to two sets of real stock transaction data are evaluated in the first part of this thesis. On the other hand, standard time series record observations in time over or at a fixed interval; such time series are referred to as discrete stochastic processes. A particular type of discrete series arises when continuous variables are measured by aggregating (or accumulating) the values over equal intervals of time. Rainfall records, traditionally of daily and hourly rainfall records, are typical examples of this type of time series. In this thesis, fine-scale rainfall data are defined as rainfall series measured in sub-hourly intervals. In the second part of this thesis, the Bartlett-Lewis Pulse (BLP) model (Cowpertwait et al., 2007), a continuous time stochastic rainfall model, is further developed and fitted to 60 years of 5-minute rainfall series.

Unlike the analysis of a standard time series, when modelling a point process series it is necessary to describe the random occurrences of the points. The Poisson process[1] is the simplest and most essential point process when modelling random point occurrences, because points in a Poisson process are independent. The intervals between successive points in a Poisson process are independent exponential random variables with a constant rate. An immediate extension of the Poisson process is obtained by defining successive intervals to be independent and identically distributed with any positively valued distribution. This produces an important class of point processes called *renewal processes*.

Using the Poisson process as a basis for qualitative comparison, we may distinguish two different types of point processes. The first type of stochastic process, which behaves more like a deterministic process, has the intervals between the points more regularly spaced (i.e. underdispersed) than the Poisson processes. A second type of process, that is overdispersed relative to the Poisson process, has a higher variance for the intervals. These second type of processes include *cluster processes* that are of interest here because stock transactions times and the rainfall events occur as clusters in time. The Poisson process can be generalized by replacing each point of the process by a cluster of points, where the original point is regarded as the cluster centre. If we further define that the cluster sizes are independent and identically distributed (iid) and the cluster points have some distribution of location relative to their cluster centre, a *Poisson cluster process* can be specified. The Bartlett-Lewis process, in which the

[1]Formal definition and properties of the Poisson process is given in Section B.1.

cluster points follow the cluster centre in a finite renewal process, is a special case of Poisson cluster process (Cox and Isham, 1980).

A point process that only models random time occurrences of point events is called a *simple point process*. A point process that not only describes the random point occurrences but also models a vector of random variables associated with the point occurrence times is called a *marked point process*. We may consider the stock transaction times as a simple point process and the transaction prices and volumes as the associated marks. Similarly, rainfall events occurrence times may be considered as a simple point process and corresponding rainfall intensities as the marks. Therefore, both the stock transaction-by-transaction data and the continuous rainfall data can be treated as marked point processes.

1.1.2 ACD model and research objective

According to the financial time series literature, the dynamic behaviour of inter-transaction durations contains useful information about intraday market activities (O'Hara, 1995; Tsay, 2005). The irregular occurrences of transaction times (in continuous time) require a point process analysis. ACD models were first used to describe the inter-transaction durations of the IBM stock transactions on the New York Stock Exchange (NYSE) by Engle and Russell (1998). Since then, ACD models have been widely used to study the market micro-structure of stock transaction processes by fitting to various real transaction data sets. However, probably due to the inherent limitation of the hypothesis testing approach in model evaluation, the literature so far has devoted little attention to model comparison despite the availability of many competing ACD model specifications. By following the hypothesis testing approach, nothing can generally be said about ranking models. Even if a best model can be identified, a fundamental concern is that no known theory can justify what kind of 'good' property this best model can have, in terms of the overall model performance. In contrast, Akaike Information Criterion (AIC; Akaike, 1973) and its generalized variants provide a valid solution for ranking ACD models based on the overall model performances and the result can be interpreted. However, this needs to be justified in theory and tested in application.

Takeuchi Information Criterion (TIC; Takeuchi, 1976) is a generalized version of AIC which enables misspecified models to be evaluated under the maximum likeli-

hood estimator (MLE) framework. The most general version, Generalized Information Criterion (GIC; Konishi and Kitagawa, 1996), extends the application of AIC model selection procedure into the framework of *M-estimator* (Maximum likelihood-type estimator; Huber and Ronchetti, 2009). The theoretical justification of an M-estimator version AIC by Konishi and Kitagawa (2008) may have removed the need to calculate TIC or GIC in model evaluation practice. Since AIC and its generalized variants are Kullback-Leibler information-based model evaluation criteria[2] which have a deep root in information theory and maximum likelihood theory, they are specifically referred to as *information-theoretic criteria* in the literature (Bozdogan, 1987; Burnham and Anderson, 2002). Hence, the model evaluation procedure based on information-theoretic criteria is referred to as the *information-theoretic approach*.

The first research topic of this thesis focuses on the evaluation of ACD models using information-theoretic criteria. The main objectives are: (1) to justify that information-theoretic criteria provide a valid and effective general approach for model evaluation; (2) to show that ACD models can be compared based on the information-theoretic criteria and the results can be interpreted.

1.1.3 BLP model and research objective

Stochastic rainfall models are used to produce the realistic simulation of observed rainfall series as may be required by many applications like urban drainage system designs, civil defence projects, ecosystem studies, or even radio transmission studies , etc. A Bartlett-Lewis pulse (BLP) model is formulated for generating realistic simulation samples of fine-scale rainfall series.

A Poisson rectangular pulse model[3] is a simple continuous time point process rainfall model in which we assume the arrival times of rectangular pulses follow a Poisson process. A Poisson rectangular pulse model is too simple to provide an adequate fit to observed rainfall time series on different time scales. The Bartlett-Lewis rectangular pulse (BLRP) model, first introduced by Rodriguez-Iturbe et al. (1987a), extends the Poisson rectangular pulse model by assuming the arrival times of rectangular pulses follow a Bartlett-Lewis process. The BLRP model has been widely studied over the last two decades. The basic BLRP model and its variants perform well in relation to

[2]Definition and more about the Kullback-Leibler information are given in Section 2.3.2

[3]More details about Poisson rectangular pulse model are given in Section 4.4.

observed statistical properties of rainfall series over a range of time scales from 1-hour upwards.

The BLP model is essentially an extension of the BLRP model formed by replacing the rectangular pulse profiles with a Poisson process of instantaneous pulse depths to give more realistic rainfall profiles for modelling fine-scale rainfall series. While the results in fitting the original BLP model to data, given by Cowpertwait et al. (2007), were excellent in terms of moment properties and extreme values, a tendency for over-estimating the historical proportion of dry periods is identified and more empirical work needed. This is the second research topic of this thesis and it focuses on the improvements to the BLP models. The objectives are (1) to improve the model structure in terms of the characterization of the physical process of precipitation; (2) to improve the model performance in fitting the proportion dry property and the extreme rainfall values for individual months; (3) to optimize the parametrisation of BLP models; (4) to carry out a comprehensive model performance assessment between different BLP model specifications by simulation.

1.2 Thesis outline

This thesis consists of two research topics, both related to the modelling of fine-scale time series using point processes. Chapters 2 and 3 present the research results about model evaluation for ACD models using the information-theoretic approach. Chapters 4, 5, 6, 7, and 8 present the research results of the further developments of the original BLP model through modelling a fine-scale series of 60 years of 5-minute rainfall.

The second chapter gives a comprehensive review of the methodology of model evaluation using information-theoretic criteria. AIC and its generalized versions, TIC and GIC, are formally defined and the theoretical basis is outlined for the justification of the information-theoretic approach as a valid and effective alternative to the traditional hypothesis testing approach for model evaluation. The historical development of the information-theoretic criteria is reviewed. The utilities of applications of AIC, TIC, and GIC for model evaluation are discussed to answer the question of 'which one to use in practice'. Examples of how to use AIC, TIC, and GIC for model evaluation are presented. In particular, AIC and GIC are tentatively applied to evaluate the basic ACD models. Chapter 2 provides the theoretical justification and practical examples for applying the information-theoretic criteria to evaluate ACD models in Chapter 3.

In the third chapter, four basic ACD models and two mixed-distribution ACD models are fitted to two real stock transaction data sets, the IBM data from a mature financial market (NYSE) and the Darby data from an emerging Asian market (KLSE). For the first time in the literature, ACD models are ranked based on AIC scores and a newly proposed mixed lognormal-gamma ACD model is identified as the best model in a clear-cut way. This result matches the model assessment results given by the Pearson's Q-statistics and the bootstrap quantile-quantile plots[4] in all cases examined. The best model selected by AIC should have the minimum expected predictive error as justified by its theoretical property specified in Chapter 2. The results presented in this chapter provide one more empirical analysis case to confirm that the information-theoretic approach is a valid and effective alternative for model evaluation in general.

The fourth chapter is the first of five consecutive chapters to present the research results of the study of the point process rainfall models for fine-scale time series. Among different types of rainfall models presented in the literature, the empirical statistical models and the point process models represent two main groups of stochastic rainfall models. These two groups of stochastic rainfall models are reviewed in the fourth chapter with a focus on the evolution of the Poisson-based point process models: from the simplest Poisson white noise model to the recent developments of temporal Poisson cluster process models, following the line of work developed in Rodriguez-Iturbe et al. (1984, 1987a,b). The review of the empirical statistical models emphasizes the connection and the comparison with the point process models.

In the fifth chapter, a detailed description and exploratory analysis are carried out on a 60-year 5-minute rainfall series, recorded at Kelburn (a place near Wellington, New Zealand). This is the data set which has been analysed by Cowpertwait et al. (2007) and is selected for the point process rainfall model study in this thesis. Chapters 4 and 5 provide the basis for presenting the research results of the improvements on the original BLP model.

The sixth chapter begins with a brief review of the Bartlett-Lewis process and the BLP model specification. The model structure characterization and fitting procedure of the original BLP model are then investigated in detail. Discussion of the original BLP model leads to the proposal of a different model specification, aimed at improving the fit in terms of extreme values at the 5-minute level and the proportion dry properties. The

[4]Details about the Pearson's Q-statistics and the bootstrap quantile-quantile plots are given in Section 2.5.2.

model performances of different BLP model specifications are assessed by an extensive simulation study. Detailed model performance comparison results are presented which include moment properties (up to 4th order), extreme values (annual maxima) patterns, and proportion dry patterns, at different aggregation levels. This chapter presents the results of an extensive empirical analysis of BLP models.

Following the continuous distributions of storm types approach as first proposed by Cowpertwait (2010), a more general BLP process characterization framework, BLP models with continuous distributions of storm types, is formulated in the seventh chapter. By expressing model parameters as functions of a continuous random variable, different continuous-storm-types BLP models can be specified and the original BLP model is recovered by setting the model parameters to be constants. The statistical properties up to third order are derived and four typical continuous-storm-types BLP models are specified for model comparison. The model performances are compared between the original BLP process and the four continuous-storm-types BLP processes based on the derived model properties, the objective function minimisation values, and the number of parameters.

Based on the research findings obtained in Chapter 7, a superposed continuous-storm-types BLP model is proposed in the eighth chapter. The model performances of this modified BLP model are compared, through a simulation study, with performances of the BLP model developed in Chapter 6, which has a conditional mean pulse depths distribution.

After an overall summary of the major research results, the ninth chapter proposes a few interesting and promising topics for future research. For example, a Bayesian approach seems to be a possible solution for a new model estimation procedure with BLP models so that an information-theoretic approach can be applied for model evaluation; the research findings with both the ACD models and the BLP models indicate that the application of BLP models in analysing real time stock transaction data is a research direction worth investigation.

Some standard theoretical results of mathematical statistics, which are required for the definition and justification of AIC and its generalized variants, are reproduced in Appendix A. Two more related information-theoretic criteria topics are also given in Appendix A. The formal definition and some important properties of Poisson processes are specified in Appendix B.1 as the background theory because it is the theoretical basis on which various point process rainfall models have been developed. Definitions of

the Markov property and Markov chains are given in Appendix B.2 as will be required for the description of empirical statistical rainfall models. The formulas for calculating statistical properties of the original BLP model were derived by Cowpertwait et al. (2007) and many of these are cited in Chapter 6. The third order moment formula of the original BLP model is reproduced in Appendix B.3. Parameter functions of five different BLP models with continuous distributions of storm types are given in Appendix B.4. A section for verification of the numerical solution using Simpson's rule to calculate the fitted proportion dry values is contained in Appendix B.5.

Throughout this study, the open source statistical package R (R Development Core Team, 2007) is used in most of the statistical analyses. The open source C language package (The GCC team, 2007) is used for raw data treatment and generation of simulation samples using the fitted BLP models. The mathematical programming language package Maple (e.g. Burkhardt, 1994, Heck, 1996) is used for symbolic operation in simplification of the algebraic expressions of the second and the third order properties of the BLP model.

This thesis is prepared using the free software typesetting program TEX which is based on the document preparation system LATEX originally written in 1984 by Leslie Lamport.

Chapter 2

Model evaluation using information-theoretic criteria – methodology

2.1 Introduction

In his seminal paper (Akaike, 1973), Japanese statistician Hirotugu Akaike proposed that the expected Kullback-Leibler information (Kullback and Leibler, 1951) should be used as a means of discriminating between competing statistical models. This idea leads to the introduction of Akaike Information Criterion (AIC) which is defined as

$$\mathrm{AIC}(k) \;=\; -2\,\log L(\hat{\underline{\theta}}) + 2k,$$

where $\log L(\hat{\underline{\theta}})$ is the sample log-likelihood evaluated at the MLE of $\underline{\theta}$, $\hat{\underline{\theta}}$, which is a $k \times 1$ parameter vector. If a model is estimated under the MLE framework, this model can be evaluated based on AIC. Among a group of candidate models, the model with a minimum AIC score is the best model selected by the AIC model evaluation procedure.

One area in which we will apply AIC is Autoregressive Conditional Duration (ACD) models (Engle and Russell, 1998). ACD models deal with the high-frequency financial time series data, the inter-transaction time intervals of a traded share in a stock market. ACD model has played an important role in study of the dynamic structure of a stock market over the last decade (Bauwens and Hautsch, 2006). Although many competing ACD model specifications are available nowadays, the literature shows that no effective method is available in evaluation of the overall performance of an ACD model using

the traditional hypothesis testing approach. Hence, the model comparison issue has been largely ignored (Bauwens et al., 2004; Meitz and Terasvirta, 2006).

Since AIC and its generalized variants have a deep root in information theory and maximum likelihood theory, they are referred to as information-theoretic criteria. Correspondingly, the model evaluation procedure based on information-theoretic criteria is referred to as the information-theoretic approach. The information-theoretic approach provides a solution for ranking models based on the overall model performance in fitting to the observed data. However, this needs to be justified in theory and testified for each application.

AIC has been used as a model evaluation criterion, most commonly for Gaussian linear regression models and the standard time series ARMA(p,q) models in the literature, for more than 30 years. (see, e.g. McQuarrie and Tsai, 1998; Chatfield, 2004; Tsay, 2005; Shumway and Stoffer, 2006; Crawley, 2007). However, it is less well-known that AIC has a deep root in the information theory and maximum likelihood theory so that it provides a general, theoretically sound, and unified model evaluation procedure. It is even less well-known that Takeuchi Information Criterion (Takeuchi, 1976) and Generalized Information Criterion (Konishi and Kitagawa, 1996), both of which are generalized versions of AIC, have extended AIC's application domain to include misspecified models and robust model estimation. The theoretical justification of an M-estimator version AIC by Konishi and Kitagawa (2008) may have removed the need to calculate TIC or GIC in model evaluation practice and this needs to be tested. The objective of this chapter is to give a theoretical justification on the information-theoretic approach as a valid and effective alternative for model evaluation in general. Based on the research results from this chapter, different ACD models will be evaluated and compared in the next chapter to verify the application of information-theoretic approach.

The current chapter is organized as follows. Section 2.2 specifies the general setting and notation which will be used in this and the next chapters. There are three subsections in Section 2.3. AIC, TIC, and GIC are formally defined and the model evaluation procedure using information-theoretic criteria is specified in Section 2.3.1. The theoretical basis of the information-theoretic criteria is outlined and some important properties are discussed in Section 2.3.2. A literature review on the historical development from AIC to GIC and some critical remarks are given in Section 2.3.3; an in-depth discussion on the connection between 'quasi-maximum likelihood' and 'quasi likelihood' in terms

of model evaluation using AIC is also included in this subsection. Section 2.4 presents a number of simple examples of applying the information-theoretic criteria to evaluate various statistical models. Section 2.5 is a discussion section in which we want to justify two conclusions: (1) AIC is a practical and parsimonious implementation of TIC and GIC for model evaluation; (2) Pearson's Q-statistic and bootstrap quantile-quantile plot can be used to provide valid additional evidence to confirm the model evaluation results obtained based on information-theoretic criteria. Finally, a concluding remarks section concludes this chapter.

2.2 General setting and notation

The general setting and notation specified in this section are mainly adapted from Konishi and Kitagawa (2008).

If a random variable X is defined on the sample space Ω, for any real value $x \in R$, $G(x) = \Pr(X \leq x)$ is referred to as the cumulative distribution function of X. We have

$$G(x) = \int_{-\infty}^{x} g(u)\mathrm{d}u, \tag{2.1}$$

(assuming X is continuous) where the function $g(x)$ is called a *probability density function* (pdf). If X is a discrete random variable such that $g(x) = \Pr(X = x)$, the function $g(x)$ is then called a *probability mass function* (pmf) and the cumulative distribution function $G(x)$ should be expressed in terms of summation of $g(x)$ rather than integration in Equation (2.1). For simplicity in notation, hereinafter $g(x)$ will be used to denote a probability density function or a probability mass function as required by context.

In mathematical statistics, a random sample of size n is defined by a multivariate random vector $\underline{X} = (X_1, X_2, ..., X_n) \in R^n$ where $X_1, X_2, ..., X_n$ are the independent and identically distributed (iid) random variables with some distribution $g(x)$. The observed sample vector, denoted by $\underline{x} = (x_1, x_2, ..., x_n)$, is a realization of the random sample $\underline{X}$. From a data analysis point of view, we will denote $\underline{x} = \{x_1, x_2, ..., x_n\}$ in this thesis to emphasize that $\underline{x}$ is a data set generated from the distribution $G(x)$ or $g(x)$. For convenience, we will also refer to $\underline{x} = \{x_1, x_2, ..., x_n\}$ as a *random sample*. The *empirical probability mass distribution function*, $\hat{g}$, is the probability mass distribution function $\hat{g}(x_r) = 1/n$ $(r = 1, 2, \ldots, n)$ that has equal probability $1/n$ for each of n

observations in a random sample $\{x_1, x_2, ..., x_n\}$.

We may fit a probability distribution model, $F(x)$ say, to the observed data $\underline{x}$ generated from $G(x)$. The function $G(x)$ or $g(x)$ is then referred to as the *true distribution*, or the *true model*. On the other hand, the distribution $F(x)$, used to approximate the true distribution, is referred to as *a model* and is assumed to have either a pdf or a pmf $f(x)$. If a model is specified by k parameters such that $\underline{\theta} = (\theta_1, \theta_2, ..., \theta_k)^T$, then the model can be written as $f(x|\underline{\theta})$. If the parameters are represented as a point in some parameter space $\Theta \subset R^k$, then $\{f(x|\underline{\theta}); \underline{\theta} \in \Theta\}$ is called a parametric family of probability distributions or models. We assume that the goodness of the model $f(x)$ is assessed in terms of the closeness as a probability distribution to the true distribution $g(x)$.

Another dual usage of notation is between $f(x|\underline{\theta})$ and the likelihood function $L(\underline{\theta})$. For a random sample $\underline{x} = \{x_1, x_2, ..., x_n\}$, the likelihood function is defined as

$$L(\underline{\theta}) = f(\underline{x}, \underline{\theta}) = \prod_{r=1}^{n} f(x_r|\underline{\theta}), \tag{2.2}$$

where we write $f(\underline{x}, \underline{\theta})$ to emphasize that the density is a function of both data and parameters. Then, the log-likelihood function is $\log L(\underline{\theta}) = \sum_{r=1}^{n} \log f(x_r|\underline{\theta})$. We now define the likelihood equation as

$$\frac{\partial \log L(\underline{\theta})}{\partial \underline{\theta}} = \sum_{r=1}^{n} \frac{\partial \log f(x_r|\underline{\theta})}{\partial \underline{\theta}} = \underline{0}, \tag{2.3}$$

where $\partial \log L(\underline{\theta})/\partial \underline{\theta}$ is a $k \times 1$ vector, the ith component of which is given by $\partial \log L(\underline{\theta})/\partial \theta_i$, and $\underline{0}$ is the $k \times 1$ zero vector, i.e. all the components of which are 0. Let $\hat{\underline{\theta}}$ denote a solution of the likelihood equation (2.3). Then $\hat{\underline{\theta}}$ is the *maximum likelihood estimate* based on n observations. For brevity, we will use the acronym 'MLE' to represent *maximum likelihood estimate* or *maximum likelihood estimator*. A formal definition of the *maximum likelihood estimator* can be found in standard mathematical statistics textbooks, e.g. DeGroot (1986) or Casella and Berger (2002).

To avoid being distracted by mathematical rigor, unless specified otherwise, we always assume that the suitable regularity conditions for MLE (e.g. see Casella and Berger, 2002, p516) are satisfied.

A time series of length n is represented by $\{y_t\} = \{y_1, y_2, ..., y_n\}$. It consists of n values sampled at discrete times $t = 1, 2, ..., n$. The observed time series is generated by a sequence of random variables sequentially through time. In general, the time series are

mutually correlated and the log-likelihood of the time series model cannot be expressed as the sum of the logarithms of the density function of each observation. However, the likelihood can generally be expressed by using the conditional distributions as follows:

$$L(\underline{\theta}) = f(y_1, ..., y_n|\underline{\theta}) = \left[\prod_{t=2}^{n} f(y_t|y_1, ..., y_{t-1}; \underline{\theta})\right] \times f(y_1|\underline{\theta}). \tag{2.4}$$

If a data set can be adequately described by a probability distribution, this takes the most fundamental form of statistical modelling, the *probability distribution model*. Real life data can seldom be modelled adequately by a simple probability distribution model. More sophisticated data analysis is usually dealt with by *conditional distribution models*. In general, if the distribution of a response random variable Y is determined in a manner that depends on some explanatory variables $\underline{X}$ and follows some probability distribution such that Y can be expressed as $F(y|\underline{x}, \underline{\theta})$ or $f(y|\underline{x}, \underline{\theta})$, this is called a conditional distribution model. Therefore, conditional distribution models are also constructed in terms of probability distribution model.

Some standard theoretical results of mathematical statistics are given in Appendix A.1 and A.2. These results will be required for the definition and theoretical justification of AIC and its generalized variants.

2.3 Definition and theoretical justification

2.3.1 Definition and model evaluation procedure

Let us assume that a random sample of n observations $\underline{x} = \{x_1, x_2, ..., x_n\}$ is generated from the true distribution $g(x)$ and a statistical model $f(x|\underline{\theta})$ is fitted to the sample data. Akaike Information Criterion (AIC) is then defined by

$$\text{AIC}(\mathbf{k}) \;=\; -\mathbf{2}\,\log \mathbf{L}(\hat{\underline{\theta}}) + \mathbf{2}\,\mathbf{k}, \tag{2.5}$$

where the sample MLE $L(\hat{\underline{\theta}})$ is defined by Equations (2.2) and (2.3).

The theoretical derivation of AIC is based on the following assumptions:

1. The sample size n is large enough so that the *large sample* theory applies;
2. Model parameters are estimated using the *maximum likelihood* method;
3. The true model $g(x)$ is contained in the class of parametric models $\{f(x|\underline{\theta}); \underline{\theta} \in \Theta \subset R^k\}$.

By convention, we refer to k, the number of estimated parameters or the number of free parameters in a fitted model, as the *model dimension*. Assumption 3 basically says that AIC is valid in application to determine the model dimension given the true model has been correctly specified. This is a very strong condition.

Takeuchi Information Criterion (TIC), first proposed by Takeuchi (1976), generalizes AIC by relaxing the third assumption condition above. TIC is defined by

$$\mathrm{TIC}(\mathbf{k}) \;=\; -\mathbf{2}\,\log \mathbf{L}(\hat{\underline{\theta}}) + \mathbf{2}\,\mathrm{tr}\big(\mathbf{K}(\hat{\underline{\theta}})\mathbf{J}(\hat{\underline{\theta}})^{-\mathbf{1}}\big), \tag{2.6}$$

where $K(\hat{\underline{\theta}})$ and $J(\hat{\underline{\theta}})$ matrices are defined by Equations (A.8) and (A.9), and tr denotes the matrix trace function. The matrices $K(\underline{\theta})$ and $J(\underline{\theta})$ involve first and second mixed partial derivative of the log-likelihood function.

If there exists a parameter vector $\underline{\theta}_o \in \Theta$ such that $g(x) = f(x|\underline{\theta}_o)$, we will have $K(\underline{\theta}_o) = J(\underline{\theta}_o)$. Therefore, the trace term in Equation (2.6) becomes $2\,\mathrm{tr}\big(K(\underline{\theta}_o)J(\underline{\theta}_o)^{-1}\big) = 2\,\mathrm{tr}(\mathrm{I}_k) = 2k$ as shown in Equation (A.7), where I_k stands for a $k \times k$ identity matrix. In general, $K(\underline{\theta}) \neq J(\underline{\theta})$, hence $\mathrm{AIC}(k) \neq \mathrm{TIC}(k)$. TIC is derived without any assumption that truth g is the same as the model f.

So long as we are under a maximum likelihood estimation framework, AIC and TIC have provided a very useful tool for evaluating the competitive models. The development of modern statistical data analysis has made some model estimation procedures other than MLE available. Konishi and Kitagawa (1996) proposed a Generalized Information Criterion (GIC) which can be applied to evaluate statistical models constructed by various types of estimation procedures including the M-estimator (the maximum likelihood-type estimator). AIC and TIC can be derived as special cases of GIC.

GIC is defined by

$$\begin{aligned}\mathrm{GIC}(\mathbf{k}) \;&=\; -\mathbf{2}\,\log \mathbf{L}(\hat{\underline{\theta}}_{\mathbf{M}}) + \frac{\mathbf{2}}{\mathbf{n}}\sum_{\mathbf{r=1}}^{\mathbf{n}} \mathrm{tr}\Big\{\mathbf{T}_{\mathbf{M}}^{(\mathbf{1})}(\mathbf{x_r};\hat{\mathbf{G}})\,\frac{\partial \log \mathbf{f}(\mathbf{x_r}|\underline{\theta})}{\partial \underline{\theta}^{\mathbf{T}}}\Big|_{\underline{\theta}=\hat{\underline{\theta}}_{\mathbf{M}}}\Big\} \\ &=\; -\mathbf{2}\,\log \mathbf{L}(\hat{\underline{\theta}}_{\mathbf{M}}) + \mathbf{2}\,\mathrm{tr}\Big\{\mathbf{R}(\psi,\hat{\mathbf{G}})^{-\mathbf{1}}\mathbf{Q}(\psi,\hat{\mathbf{G}})\Big\},\end{aligned} \tag{2.7}$$

where $\log L(\hat{\underline{\theta}}_M) = \sum_{r=1}^{n} \log f(x_r|\hat{\underline{\theta}}_M)$ and the empirical M-estimator $\hat{\underline{\theta}}_M$ is defined by Equation (A.12). Matrices $R(\psi,\hat{G})$, $Q(\psi,\hat{G})$, and $T_M^{(1)}(x_r;\hat{G})$ are defined by Equations (A.14), (A.15), and (A.18) in Appendix A.

The only assumption condition required for deriving GIC is the large sample property condition. In the case that model parameters are estimated by maximum likelihood, it is trivial to show that $\mathrm{GIC}(\mathbf{k}) = \mathrm{TIC}(\mathbf{k})$ based on Equations (A.16) and (A.17).

Given a certain regularity condition[1], Konishi and Kitagawa (2008), p131, shown that, if we would like to assume that the parametric family of probability distribution $\{f(x|\underline{\theta}); \underline{\theta} \in \Theta \subset R^k\}$ contains the true distribution $g(x)$, the AIC can be applied directly to evaluate statistical models within the framework of M-estimation. Therefore, an M-estimator version AIC is defined as

$$\text{AIC}(\mathbf{k}) \; = \; -\mathbf{2} \log \mathbf{L}(\hat{\underline{\theta}}_{\mathbf{M}}) + \mathbf{2k}, \tag{2.8}$$

where $\hat{\underline{\theta}}_M$ is defined by Equation (A.12). The discovery of this M-estimator version AIC may have removed the need to calculate TIC or GIC for model evaluation in practice. We will justify this conclusion in Section 2.5.1.

Akaike (1974) summarized the model evaluation procedure based on AIC as follows:

(1) the model which has a minimum AIC estimate should be selected as the best model among all candidate models;

(2) when there is only one unrestricted family of $f(x|\underline{\theta})$, the minimum AIC estimate is defined by the $f(x|\hat{\underline{\theta}})$ with $\hat{\underline{\theta}}$ identical to the classical MLE;

(3) when there are several specifications of $f(x|\underline{\theta})$ corresponding to several models, the minimum AIC estimate is defined by the $f(x|\hat{\underline{\theta}})$ which gives the minimum of AIC(k);

(4) when the maximum likelihood is identical for two models, the minimum AIC estimate is the one defined with the smaller number of parameters.

This procedure applies in the same way when models are evaluated based on TIC and GIC. Note that, under the large sample conditions, AIC, TIC, and GIC can be applied without being limited to any specific probability distribution. As long as the model likelihood function is properly defined, this model evaluation procedure is valid.

2.3.2 Theoretical basis and properties

Outline of theoretical basis

Model selection in statistical modelling implies a group of candidate models are evaluated against some model performance measurement criterion so that a 'good' model

[1]See Konishi and Kitagawa, 2008, p117, for detail.

may be determined. But what defines a good model? There is no absolute 'correct' answer for this question because it all depends on what general assumptions we like to make in construction of the required model selection criterion. A *general assumption* that underlies AIC, TIC, and GIC is that **the true model which generates the observed data is of infinite dimension and it is unknown** to investigators.

Akaike held the view that the purpose of statistical modelling is to predict future data as accurately as possible (Konishi and Kitagawa, 2008, p2), i.e. a model that fits well on one of the data sets representing the true generation process should fit well on any other data set. A model with a too detailed structure level (i.e. with too many parameters) but fits the current data set well may fit subsequent data sets poorly. A model that is too simple may fit none of the data sets well. In general, bias decreases and variance increases as the model dimension (k) increases. Therefore, a good model selection criterion should select a model best to balance the increase in fit against the increase in model complexity. That is, the model structure should be fitted to a sufficient detailed level in order to fit the observed data well (indicated by a higher likelihood so that bias is reasonably small) but the model dimension should also be constrained in a way that only a reasonable level of variance is allowed. Such property is also described as a *trade-off between bias and variance* which is the very essence of the Principle of Parsimony (Burnham and Anderson, 2002). All model selection methods are based to some extent on the principle of parsimony. Information criteria, derived based on minimizing the expected Kullback-Leibler information, are the model selection criteria which give an optimal balance between bias (model accuracy) and variance (model complexity) for the reasons given below.

Ludwig Boltzmann, a nineteenth-century physicist derived the fundamental theorem in 1877 that

$$\text{Entropy is proportional to } -\log(\text{probability}),$$

where *entropy* is considered as a measure of uncertainty. Boltzmann's concept of entropy built up the rigorous relationship between the information theory and statistics through probability: entropy is expressed in terms of information which is measured by $-\log(\text{probability})$ and of course probability is the fundamental concept in statistics. Boltzmann's concept of entropy is considered as one of the greatest achievements in the nineteenth-century science (Burnham and Anderson, 2002).

Following our general setting in Section 2.2, the truth is denoted by $g(x)$ and a

fitted approximation of the truth is denoted by $f(x)$. The Boltzmann's entropy is conceptually defined by

$$\text{Boltzmann's entropy} = -\log\Big[\frac{g(x)}{f(x)}\Big]. \tag{2.9}$$

The *Kullback-Leibler Information* (K-L information) is then defined by

$$\begin{aligned}\mathbf{I}(g;f) &= \mathrm{E}_g\big[-\text{Boltzmann's entropy}\big] \\ &= \mathrm{E}_g\Big\{\log\Big[\frac{g(x)}{f(x)}\Big]\Big\} \\ &= \mathrm{E}_g\big[\log g(x)\big] - \mathrm{E}_g\big[\log f(x)\big], \end{aligned} \tag{2.10}$$

where E_g stands for *taking expectation with respect to* $g(x)$ (Burnham and Anderson, 2002). This is how K-L information relates to the Boltzmann's entropy.

K-L information measures the expected information loss when a model $f(x)$ is used to approximate the full reality or truth $g(x)$. The third line equality in Equation (2.10) implies that $\mathbf{I}(g;f)$ is a measure of the difference between the expected log-likelihood of the true model $g(x)$ and the expected log-likelihood of a fitted model $f(x)$. Therefore, $\mathbf{I}(g;f)$ may be considered as a measure of the distance between $f(x)$ and $g(x)$, too. That is why K-L information is also referred to as *K-L distance* or *K-L discrepancy* in literature. From Equation (2.10), two K-L information properties can be drawn:

(i) $\mathbf{I}(g;f) \geq 0$;

(ii) $\mathbf{I}(g;f) = 0 \iff g(x) = f(x)$.

Konishi and Kitagawa (2008) interpreted *Boltzmann's entropy* as a quantity that varies proportionally with the logarithm of the probability with which the relative frequency of the samples obtained from a fitted model agrees with the true distribution. Therefore, K-L information " is thought of as the probability of obtaining the observed data from the model" (Konishi and Kitagawa, 2008, p34). Minimization of the K-L information implies the maximization of the probability of obtaining the observed data from a fitted model.

It is clear that, by Equation (2.10), K-L information is defined in terms of the expected log-likelihood with respect to the true distribution $g(x)$. Because the expected log-likelihood of the true model, $\mathrm{E}_g\big[\log g(x)\big]$, is a constant, it suffices to measure the discrepancy between a fitted model and the true model based on the relative K-L

information in terms of $\mathrm{E}_g\big[\log f(x)\big]$, the expected log-likelihood of a fitted model. However, the calculation of the relative K-L information also involves the true model which is unknown to us. Therefore, for defining an information criterion, the crucial issue is to obtain a good estimator for $\mathrm{E}_g\big[\log f(x)\big]$. One such estimator is the mean sample log-likelihood, $\frac{1}{n}\log L(\hat{\underline{\theta}})$.

Akaike (1973) suggested that (the relative) K-L information can be used as a *loss function* so that the expected K-L information is considered as a *risk function* in the process of optimization. The expected K-L information measures the unconditional mean information loss in fitting a candidate model to approximate the true distribution. In observational data analysis, the best information available about the true distribution is from the observed sample data. However, the K-L information estimated based on any single realization of the true distribution, $\mathbf{I}(\hat{g}, f)$, is bound to be biased due to the sampling variation. It can be shown that the mean sample log-likelihood as an estimator of the expected log-likelihood has a bias of order n^{-1} (e.g. Chapter 7, Konishi and Kitagawa, 2008). Hence, the bias correction for the estimation of the expected log-likelihood using the log-likelihood of a fitted model is essential for constructing an information criterion for model evaluation. The optimal strategy is, therefore, to estimate $\mathbf{I}(g, f)$ by minimizing the risk function, i.e. the expected mean information loss or the unconditional mean information loss.

Let us now define a concept called *Effective Model Dimension* (EMD). Then all information criteria may be considered taking a general form of

$$\mathrm{IC} = -(\text{log-likelihood}_{max}) + \mathrm{EMD}. \tag{2.11}$$

Theoretically, EMD is the bias correction term which is essential for constructing an information criterion (Konishi and Kitagawa, 2008). Intuitively, an EMD term may also be interpreted as a measure of the complexity of a fitted model. An information criterion defined by $\mathrm{aic}(k) = -(\text{log-likelihood}_{max}) + k$ will behave exactly as $\mathrm{AIC}(k) = -2(\text{log-likelihood}_{max}) + 2k$ in model selection. Just because Akaike multiplied through by -2 when he first proposed AIC, hence TIC and GIC also have been defined by following the same fashion purely for the historical reason. The terms k, $\mathrm{tr}\big(K(\hat{\underline{\theta}})J(\hat{\underline{\theta}})^{-1}\big)$, and $\mathrm{tr}\big(R(\psi, \hat{G})^{-1}Q(\psi, \hat{G})\big)$ are defined as the EMD for AIC, TIC, and GIC, respectively. Konishi and Kitagawa (2008) have shown that AIC, TIC, and GIC are defined based on the bias-corrected log-likelihood which are second-order correct or accurate for the estimation of relative K-L information. Second-order correct

means that the expectation of two quantities are in agreement up to and including the term of order n^{-1} and that the order of the remainder is n^{-2}. This is the theoretical framework under which AIC, TIC, and GIC are derived. For detailed derivations of these three information-theoretic criteria, probably the three recently available monographs, Burnham and Anderson (2002), Claeskens and Hjort (2008), and Konishi and Kitagawa (2008), are the most comprehensive and convenient reference sources.

The information-theoretic criteria have been formulated to measure the relative expected loss of information in fitting a statistical model to sample data. Therefore, the selected best-fit model based on a minimum IC score should have the minimum expected information loss about the true model among all candidate models. In statistical terms, since a lower IC score implies a higher likelihood, an information-theoretic criterion selected model has the highest probability of obtaining the observed data. The trace term EMD penalizes the model complexity to ensure a parsimonious model specification in terms of the total number of parameters. This penalty effect is a natural consequence of minimizing the expected Kullback-Leibler information in the process of deriving AIC, TIC, and GIC. Note that the penalty term of AIC does not depend on sample size n.

Some important properties

From the point of view of statistical performance of a model selection criterion, and intended context of its use, there are only two distinct classes of criteria: one kind is labeled as *efficient* and the other as *consistent* (Burnham and Anderson, 2002). These two different groups of criteria are formulated from different *general assumptions*. Both kinds have been widely used in statistical modelling practice. But there is little agreement on which kind performs better.

According to Shibata (1980), for large samples, a model selection criterion is said to be *asymptotically efficient* if it chooses the model with minimum mean squared error distribution. AIC is asymptotically efficient.

According to Claeskens and Hjort (2008), p109, a model selection criterion is called 'efficient' if it selects a model such that the ratio of the expected loss function at the selected model and the expected loss function at its theoretical minimizer converges to one in probability. AIC has also been shown to be efficient under this definition.

On the other hand, under the general assumptions that a *true model* is of fi-

nite dimension and it is included in the class of candidate models, a criterion that identifies the true model asymptotically with probability one is said to be *consistent*. The general assumptions for defining a consistent criterion imply that the investigator has enough knowledge about the true data generation mechanism so that a list of important variables can be determined and all these variables can be measured. These are strong conditions in data analysis practice. Obviously, AIC is not consistent under this definition. However, Claeskens and Hjort (2008), p101, pointed out that AIC is *weakly consistent* in a sense that the ratio of its penalty term over sample size (i.e. $2k/n$) tends to zero as n approaches $+\infty$. Heuristically, AIC's weak consistency property holds when sample size is getting infinitely large since the sample maximum log-likelihood term will then become infinitely large and the penalty term is fixed (not depends on sample size) so is negligible.

Based on different general assumptions, an *efficient* model selection criterion and a *consistent* criterion are designed to answer different questions even though they are often employed in the same situation. For an efficient selection criterion, like AIC, its goal is to select one model that best approximates the true model (which is assumed to be of infinite dimension) from a set of finite-dimensional candidate models. Therefore, the objective of an efficient criterion in model selection is to obtain a 'good' model rather than a 'true' model. The candidate model with the minimum K-L discrepancy, which implies a minimum mean squared prediction error, is considered as the best choice. For a consistent criterion, the goal is to identify the true model from the list of candidate models asymptotically with probability one. Claeskens and Hjort (2008) shown that consistency and efficiency cannot occur together, in particular, a consistent criterion can never be efficient. Bayesian Information Criterion (BIC) (Schwarz, 1978) is a well-known consistent model selection criterion. We will mention BIC again in Section 2.3.3.

The performance of a model selection method is often assessed by whether or not it overfits or underfits data. However, the notion '*overfitting*' or '*underfitting*' can be misleading without specifying what general assumptions one has made in making the assessment. McQuarrie and Tsai (1998), p8, have made a nice brief summary to clarify this possible confusion as we quote below: "The terms *overfitting* and *underfitting* can be defined in two ways. Under consistency, when a true model is itself a candidate model, overfitting is defined as choosing a model with extra variables, and underfitting is defined as choosing a model that either has too few variables or is incomplete.

We have no term to describe choosing a model with the correct order but the wrong variables. Using efficiency (observed or expected), overfitting can be defined as choosing a model that has more variables than the model identified as closest to the true model, thereby reducing efficiency. Underfitting is defined as choosing a model with too few variables compared to the closest model, also reducing efficiency."

As we have presented earlier that AIC is an efficient model selection criterion so that the conceptual framework underlying valid use of AIC assumes the true model has infinitely many parameters. However, it is quite common in the literature to see AIC's model selection performance studied under consistency general assumptions (i.e. finite dimensional model and the true model known) (see, for example, Bozdogan, 1987, and Davison, 2003, section 4.7). Therefore, AIC has been typically criticized for tending to overfit the order of the true model. We would like to make following points as our understanding of the issue of 'overfitting' with AIC.

(1) The consistency concept of overfitting is not applicable under the K-L information based model selection framework. If 'overfitting' is defined as one has selected a model with more estimated parameters than are in the true model, under the infinite dimensional true model assumption overfitting is not even possible. A proper model selection performance assessment with AIC should be based on predictive mean squared error as did in Burnham and Anderson (2002), p298, because information-theoretic criteria have been designed to select a 'good' model rather than the 'true' model.

(2) "Even if a true finite order exists, the order of a good model is not necessarily equal to the true order" (Konishi and Kitagawa, 2008, p74). We will show some examples of this in Sections 2.4.1 and 2.4.2.

(3) Overfitting does occur if AIC is applied improperly for model selection when the large sample assumption is not met. In this case, it is the AICc actually should be used so that the term $2kn/(n-k-1)$ imposes a heavier penalty on model complexity. (See Equation (2.12) and Appendix A.4 for more about AICc, the small sample version AIC.)

(4) Information-theoretic criteria are estimators of the relative expected K-L information which implies that there is no guarantee that the selected best model is a good approximation of the truth in an absolute sense. A model selection criterion

can do nothing if a reasonable good model is not even included in the specified group of candidate models. Therefore, it is an investigator's ultimate responsibility to be able to define the *a priori* set of candidate models which are based on the research objectives and what is known about the problem in concern.

Burnham and Anderson (2002) recommended that a small sample version AIC needs to be used if the sample size versus the number of free parameters ratio is less than 40 (i.e. $\frac{n}{k} < 40$). The most widely used small sample version AIC is derived as

$$\mathbf{AIC_c(k)} \;=\; \mathbf{-2 \log L(\hat{\theta}) + 2\,k\,\frac{n}{n-k-1}} \tag{2.12}$$

by Hurvich and Tsai (1989) adapted from Sugiura (1978)'s results. Note that, this formula is only valid for normally distributed models and there is no general form of the small sample version AIC.

Since, in all the data analyses of this study, the n/k ratio is very large (at least greater than 500 for real data or greater than 200 for simulated data), small sample issue will be largely ignored in the main body of this thesis. But some further investigations on AICc can be found in Appendix A.4.

2.3.3 From AIC to GIC – literature review and more

The oldest reference material mentioned in the information criterion literature, is about *Boltzmann's entropy* published in 1877 by L. Boltzmann (see the corresponding German reference entry in Bozdogan, 1987, or in Burnham and Anderson, 2002). Literature always cited Kullback and Leibler (1951) as the original reference for *Kullback-Leibler information*. Kullback-Leibler information is also referred to as negative *entropy*. Entropy is a well-known concept as a measure of the degree of disorder in thermodynamics. However, it is entropy's definition as *a measure of uncertainty* in information theory which is expressed by *logs of probabilities* that creates the term *information-theoretic* approach in statistics. For more about the relationship between entropy and probability, Applebaum (2008) is a good introductory book to look at.

The word of 'information' in *Fisher Information Matrix* has no relationship to *information theory* in the sense that it was in the 1920s that Fisher chose to name Equation (A.3) (the expected matrix of second mixed partial derivatives of a log pdf) 'information', whereas the information theory is a subject developed mostly since the

late 1940s. On the other hand, however, because all of the information in the observations about a given model is contained in the likelihood function for that assumed model family and the Fisher information is a second-order approximation part of it (Lindsey, 1996, p101), *Fisher Information Matrix* does deserve its name. As it is shown in Section A.1, the Fisher information matrix fundamentally relates to the precision of the MLE.

Akaike (1969) introduced an AIC type of the fitting procedure (Final Prediction Error procedure) for a univariate AR model. This is probably the very first one in a sequence of many articles contributed by Akaike over the following few decades for establishing the information-theoretic approach in statistical modelling. Most literature nowadays recognize Akaike (1973) as the paper in which AIC was originally developed. Some literature also cite Akaike (1974) as the original AIC reference for it proposed MAICE (minimum AIC estimators) as an AIC application procedure.

In an effort to provide a general derivation from K-L information to AIC, Takeuchi (1976) proposed TIC (as an intermediate result in his derivation) for model selection that allows for "misspecification" of the approximating models, i.e. the set of candidate models does not include the true model $g(x)$ or any model very similar to $g(x)$. TIC represents an important conceptual advance and further justifies AIC. However, TIC has received little attention in practice probably mainly because of the following three reasons. (1) Takeuchi (1976) was published in Japanese so TIC is much less well-known; (2) most statistics practitioners did not realize that there is a sound theoretical basis behind AIC so would not mind how AIC is derived; (3) more importantly, the estimation of the trace term in Equation (2.6) involves estimation of the elements of two $k \times k$ matrices which can cause instability of the results of model selection (see Burnham and Anderson, 2002, Chapter 7 and references therein). The form of the trace penalty term in TIC has also been reported by Stone (1977).

A third theoretical development comes from the second-order improvement or the small sample bias-corrected AIC. The original idea of AICc was from Sugiura (1978), but Equation (2.12) was formally introduced by Hurvich and Tsai (1989). Although AICc is derived specifically for normally distributed models, Burnham and Anderson (2002), p316, recommended that AICc is useful in general.

The most routine applications of AIC have been in linear regression and the standard time series ARMA models, often as for selection of explanatory variables or determination of the order of an ARMA process. McQuarrie and Tsai (1998) is a typical

reference of this kind worth mentioning which is entitled "Regression and time series model selection". McQuarrie and Tsai (1998) also gave a historical review on the development of information-theoretic approach up to then and a good summary of different kinds of model selection criteria.

The Generalized Information Criterion (GIC) proposed by Konishi and Kitagawa (1996) extends the application of information-theoretic criteria beyond the MLE framework, which marked another milestone in the development of the information-theoretic approach. The theoretical justification of an M-estimator version AIC may have a significant implication for model evaluation practice. Konishi and Kitagawa (2008) provided a systematic and comprehensive account on information-theoretic criteria from a statistical modelling point of view.

There are many other information criteria proposed for model selection among which Bayesian Information Criterion (BIC) is probably the most well-known. Motivated by finding a Bayesian solution for model selection problems, Schwarz (1978) proposed BIC as

$$\mathrm{BIC}(k) = -2 \log L(\hat{\underline{\theta}}) + k \log(n). \tag{2.13}$$

BIC has also been derived independently, without Bayesian considerations, by Rissanen (1978).

BIC is structurally similar to AIC but includes a penalty term dependent on the sample size. Because BIC selected best-fit models do not necessarily minimize K-L information, BIC is therefore fundamentally different from AIC. Some more discussions on different information criteria, e.g. AIC, BIC, and DIC (Deviance Information Criterion), and the applications under a Bayesian statistics framework is given in Appendix A.5.

Burnham and Anderson (2002), Claeskens and Hjort (2008), and Konishi and Kitagawa (2008) have provided a full treatment on various aspects of the information-theoretic criteria: e.g. the motivation of information-theoretic approach, the derivation of information-theoretic criteria, the justification of their theoretical properties and their application results, the comparison between various model selection criteria, etc are all attended in intensive depths.

In particular, Burnham and Anderson (2002) had a more applied emphasis and focused on a full exploitation of application of AIC and AICc in analysing real data. Burnham and Anderson (2002) also stated that they made a further extension of the

information-theoretic approach which was a simple modification to AIC and AICc for overdispersed count data. They denoted the modified criteria by QAIC and QAICc as they are derived from quasi-likelihood theory,

$$\text{QAIC}(k) \;=\; -[2\,\log L(\underline{\hat{\theta}})/\hat{c}] + 2k, \tag{2.14}$$

and

$$\text{QAICc}(k) \;=\; \text{QAIC}(k) + \frac{2k(k+1)}{n-k-1}, \tag{2.15}$$

where $\hat{c}$ is a variance inflation factor which is computed from the goodness-of-fit statistic, divided by its degrees of freedom, $\hat{c} = \chi^2/\text{df}$. Burnham and Anderson (2002) believed that in the case of overdispersed count data, the value of the maximum log-likelihood function is divided by the estimate of overdispersion to provide a proper estimate of the log-likelihood. When no overdispersion exists, $c = 1$, and the formulas for QAIC and QAICc reduce to AIC and AICc, respectively.

As we will see in Section 2.4.3, the model estimation for ACD models fits within the quasi-MLE framework. The concept of *quasi-maximum likelihood* with ACD models is equivalent to *misspecification* of a model. It seems that this *quasi-maximum likelihood* is very different from the concept of *quasi-likelihood* used by by Burnham and Anderson (2002) in defining QAIC and QAICc. From all the literature reviewed so far, we have found no one mentioning anything about the relationship between the so-called "quasi-maximum likelihood" and the "quasi-likelihood". We give our findings and understanding on these two concepts and how they relate to information-theoretic criteria as follows.

Hamilton (1994), p126, stated that "An estimate that maximizes a misspecified likelihood function (for example, an MLE calculated under the assumption of a Gaussian process when the true data are non-Gaussian) is known as a *quasi-maximum likelihood estimate*." The landmark result of QMLE is the well-known information sandwich variance matrix as it is defined in Section A.1. White (1982) is referred by literature as the original paper which introduced the *information sandwich variance matrix* (Theorem 3.2). Claeskens and Hjort (2008), p64, described the sandwich matrix as the generalization 50 years later than the standard likelihood theory. The information sandwich variance matrix provides us a powerful and useful tool in real life data analysis: we misspecify models almost surely. The derivation of the information sandwich variance matrix is based on the concepts of K-L information, M-estimator and large sample theory as shown in White (1982). As shown in Konishi and Kitagawa (2008),

the derivation of TIC and GIC is closely related to the information sandwich variance matrix and those concepts that underlie it. It is interesting to observe that White (1982) is actually a paper of investigating the hypothesis testing approach for model evaluation.

On the other hand, the concept of *quasi-likelihood* is particularly related to the fitting and inference of generalized linear models (GLMs) (e.g. see Wood (2006), p74). Most often Wedderburn (1974), sometimes Nelder and Wedderburn (1972), is cited as the paper responsible for the original introduction of quasi-likelihood for GLMs. Quasi-likelihood is considered as a robust model estimation method where the true model is only partially correctly specified, namely the variance-mean relationship is correctly specified (hence, this is a misspecified model case, too). Therefore, the so-called quasi-likelihood can be actually considered as a special case of the QMLE framework as defined in White (1982). Since we have shown that GIC is suitable for model evaluation under a QMLE framework in this chapter, a logical conclusion is that GIC should have covered QAIC and QAICc in utility of applications. The examples given in Konishi and Kitagawa (2008), Chapter 6, seem to support this conclusion.

As pointed out by Burnham and Anderson (2002), little work has been done in verification of the utilities of TIC and GIC in real data analysis. It is the theoretical definitions and empirical formulas on TIC and GIC given in Konishi and Kitagawa (2008) that enable us to use TIC and GIC in real data analysis, in which we will present more results in the next chapter.

Both Burnham and Anderson (2002) and Claeskens and Hjort (2008) have a focus on multimodel inference (e.g. model averaging). This is a topic outside the scope of this research.

2.4 Application of AIC, TIC, and GIC – simple examples

Even though the derivation of information-theoretic criteria depends heavily on mathematical statistics, it turns out that AIC, TIC, and GIC are relatively easy to use and the results can be interpreted. The crucial factor in applying information-theoretic criteria for model evaluation is to be able to define the likelihood function $L(\underline{\theta})$ properly.

2.4.1 Basic applications of AIC

In this section, we will fit several ARMA models to a simulated sample of data, which are generated from a predetermined AR(2) model, to show some basic applications of AIC in model selection. One focus of this section is on the usage of the *Akaike weights* in model evaluation which has been ignored by most routine AIC applications in practice so far. A brief summary on the usage of AIC and how it relates to the hypothesis test in general is given at the end of this section.

Given a univariate time series $\{y_t : \ t = 1, 2, ..., n\}$, a general ARMA model can be expressed as

$$y_t \ = \ \sum_{i=1}^{p} a_i y_{t-i} + \sum_{j=1}^{q} b_j \epsilon_{t-j} + \epsilon_t, \tag{2.16}$$

where (p,q) is called the order of an ARMA model (Chatfield, 2004). For an ARMA(p,q) model, we assume that $a_p \neq 0$ and $b_q \neq 0$. The error terms $\epsilon_r, \quad r = t-q, t-q+1, ..., t$ are assumed to be iid normal random variables with mean 0 and variance σ^2. Here we have defined a standard time series ARMA model with Gaussian white noise.

An ARMA(p,q) model is therefore expressed by the conditional density of y_t, given $y_{t-1}, y_{t-2}, ..., y_{t-m}$ where $m = \max(p, q)$, as

$$f(y_t|y_{t-1}, y_{t-2}, ..., y_{t-m}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big\{ -\frac{1}{2\sigma^2}\Big(y_t - \sum_{i=1}^{p} a_i y_{t-i} - \sum_{j=1}^{q} b_j \epsilon_{t-j}\Big)^2\Big\}. \tag{2.17}$$

Thus, assuming that $y_{1-m}, y_{2-m}, ..., y_0$ are known, the log-likelihood of an ARMA(p,q) model is therefore defined by

$$\log L(\underline{\theta}) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=1}^{n}\Big(y_t - \sum_{i=1}^{p} a_i y_{t-i} - \sum_{j=1}^{q} b_j \epsilon_{t-j}\Big)^2, \tag{2.18}$$

where $\underline{\theta} = (a_1, ..., a_p, b_1, ..., b_q, \sigma^2)^T$. Once the MLE $\hat{\underline{\theta}}$ is obtained, the AIC for an ARMA(p,q) model is given by

$$\begin{aligned} \text{AIC}(k) &= -2 \log L(\hat{\underline{\theta}}) + 2k \\ &= n\log(2\pi) + n\log(\hat{\sigma}^2) + n + 2(p+q+1), \end{aligned} \tag{2.19}$$

where $k = p + q + 1$ is the number of estimated parameters in an ARMA(p,q) model. Obviously, an AR(p) or a MA(q) model is a special case of an ARMA(p,q) model.

Now a simulated time series data set $y(t)$ of sample size $n = 1000$ is generated from an AR(2) model

$$y_t = 0.65y_{t-1} + 0.25y_{t-2} + \epsilon_t, \quad t = 1, 2, ..., 1000.$$

using the following R code.

```
set.seed(101)
a=c(0.65,0.25)
yt=e=rnorm(1002)
for(t in 3:1002) {yt[t]=a[1]*yt[t-1]+a[2]*yt[t-2]+e[t]}
yt <- yt[3:1002]
```

The R function `arima` is then used to fit various ARMA models. For example, we want to fit an ARMA(2,1) model to the data. According to Equation (2.16), an ARMA(2,1) can be expressed as

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + b_1 \epsilon_{t-1} + \epsilon_t.$$

The following R command line is used to fit an ARMA(2,1) model.

```
yt.arma21 = arima(yt,order=c(2,0,1),method="ML", include.mean=F)
```

Then AIC can be obtained by using either one of the following R commands,

```
AIC(yt.arma21)
yt.arma21$aic
-2*yt.arma21$loglik + 2*(3+1)
```

they all give us the same answer, 2768.378. Two little details need our attention in using R here. First, the argument `method="ML"` is necessary for the output AIC value to become valid according to the users help manual; Secondly, if you do not specify the `include.mean=F` argument, AIC score calculated by `AIC(yt.arma21)` (which is same as `yt.arma21$aic`) would be different from using `-2*yt.arma21$loglik + 2*(3+1)`

because the number of parameters will be one more (i.e. k=4+1 rather than k=3+1) by needing to estimate the mean level of the residuals.

Most statistical computer packages routinely return the value of the log-likelihood at its maximum or even AIC score itself as R does here. In the same way, eight models are fitted to the simulated data and AIC scores are calculated. The model fitting results are summarized in Table 2.1. We have that sample size $n = 1000$ and k is the number of estimated parameters in a model. An AR(1) model is denoted by 'yt.ar1' which is fitted to the data y_t. Similarly 'yt.ar2' denotes AR(2), 'yt.ar3' for AR(3), 'yt.ar4' for AR(4) and 'yt.ma1' for MA(1), 'yt.ma2' for MA(2), 'yt.arma11' for ARMA(1,1) and 'yt.arma21' for ARMA(2,1). Parameter estimate $\hat{\sigma}^2$ is obtained by extracting `sigma2` from a fitted model object in R.

Table 2.1: Basic application of AIC: ARMA models, AIC selected model is the true model

Models	k	AIC	ΔAIC$_i$	W_i	$\hat{\sigma}^2$
yt.ar1	2	2830.30	63.87	0.000	0.9871
yt.ar2	**3**	**2766.43**	**0**	**0.509**	**0.9241**
yt.ar3	4	2768.38	1.95	0.192	0.9241
yt.ar4	5	2770.32	3.89	0.073	0.9240
yt.ma1	2	3565.69	799.26	0.000	2.061
yt.ma2	3	3232.94	466.49	0.000	1.474
yt.arma11	3	2771.92	5.49	0.033	0.9292
yt.arma21	4	2768.38	1.95	0.193	0.9241

In Table 2.1, $\Delta\,\text{AIC}_i$ is the AIC difference of the fitted model f_i as defined in the formula below.

$$\Delta\,\text{AIC}_i \;=\; \text{AIC}_i - \min(\text{AIC}_i) \quad i = 1, ..., m, \tag{2.20}$$

where m is the number of candidate models, in our case here $m = 8$. The model with $\min(\text{AIC}_i)$ is considered as the K-L best model. In this particular example, AIC has identified the AR(2) model (the bold line model in Table 2.1) as the best-fit model which happens to be the true model. It is most common that AIC application practice stops here. What we want to emphasize here is the use of W_i for model selection.

Because the truth, g, is unknown, an individual AIC value is not interpretable by itself. AIC is only comparative, relative to other AIC values in the candidate model set. Thus AIC differences $\Delta\,\text{AIC}_i$ are very important and useful. From the calculated

$\Delta\,\mathrm{AIC}_i$, we can work out the so-called 'Akaike weights' W_i as named by Burnham and Anderson (2002) which is defined as

$$\mathrm{W}_i \;=\; \frac{\exp(-0.5\,\Delta\mathrm{AIC}_i)}{\sum_{i=1}^{m}\exp(-0.5\,\Delta\mathrm{AIC}_i)}. \tag{2.21}$$

W_i may be interpreted as the probability that model f_i is the K-L best model among all the specified models. The idea of W_i has been suggested for many years by Akaike (Akaike, 1978, also see Bozdogan, 1987). However, W_i has not been commonly used in most AIC application cases for model selection.

Note that W_i have been scaled to summing up to unity with respect to the entire model set. Therefore, if a model is added or dropped in the process of analysis, the W_i must be recomputed for all the models in the newly defined set.

If we look at W_i column in Table 2.1, we find that model AR(3) and ARMA(2,1) also have quite strong support from the data (0.192 and 0.195 respectively) which can not be ignored while AR(2) gets the highest support of 0.509. Actually, both AR(3) and ARMA(2,1) are models in which the true model AR(2) is nested. It is only due to the penalty term $2k$ that makes them not be picked as the best model. In real data analysis (in which a finite true model scenario is very unlikely), such results indicate that we may need to consider the model averaging strategy rather than picking up a single best model for the subsequent statistical inference. Therefore, always calculating $\Delta\,\mathrm{AIC}_i$ and W_i is a good practice in making full use of AIC (or TIC, or GIC) properties. The computation of $\Delta\,\mathrm{AIC}_i$ and W_i is nearly trivial once AIC scores are obtained. Note that the use of $\Delta\,\mathrm{AIC}_i$ and W_i in model selection has a sound theoretical basis. However, We could not find any of the traditional hypothesis testing methods can give us such a capacity as Akaike weights W_i have in model selection. As a rule of thumb, Burnham and Anderson (2002) suggested that models within two units of $\Delta\,\mathrm{AIC}_i$ should be considered as strongly supported by data so that a single best model selection may not be appropriate. In the case where the AIC selected best model is at least 10 units in $\Delta\,\mathrm{AIC}_i$ apart from the nearest competitive model, we should feel confident to choose a single best-fit model for subsequent statistical inference.

Finally, we notice that AR(2), AR(3), AR(4), and ARMA(2,1) have the same estimated variance $\hat{\sigma}^2$ (up to 3 decimal places) among which the AIC selected model AR(2) is the simplest model. This exhibits AIC's unique property in selecting a 'good' model based on the Principle of Parsimony. In contrast, we can not find a theoretical

basis to guarantee a hypothesis testing approach selected model having such a property (Burnham and Anderson, 2002).

The model fitting results in Table 2.1 presented a case in which the AIC selected best model happens to be the true model. In order to further display the property that AIC selects a 'good' model which is not necessarily the true model, we present a second example with exactly the same settings used in the previous example except for the specification of the random seed. Therefore, a different random sample is generated from the same AR(2) true model specified above using the same R code by a different random seed: `set.seed(17)`. The same group of fitted models in Table 2.1 are specified and the fitting results are presented in Table 2.2.

Table 2.2: Basic application of AIC: ARMA models, AIC selected model is not the true model

Models	k	AIC	ΔAIC_i	W_i	$\hat{\sigma}^2$
yt.ar1	2	2920.44	45.22	0.000	1.080
yt.ar2	3	2878.58	3.36	0.076	1.034
yt.ar3	4	2875.55	0.33	0.347	1.028
yt.ar4	**5**	**2875.22**	**0**	**0.409**	**1.026**
yt.ma1	2	3743.82	868.60	0.000	2.463
yt.ma2	3	3376.08	500.86	0.000	1.701
yt.arma11	3	2886.69	11.47	0.001	1.042
yt.arma21	4	2877.02	1.80	0.166	1.030

The AIC selected model, which is highlighted with bold font, is now AR(4) which is not the true model AR(2) as shown in Table 2.2. Model AR(4) has the most total number of parameters among all candidate models, but with the minimum estimated variance $\hat{\sigma}^2$. The resulting minimum AIC indicates that AR(4) has the optimal trade-off between model bias and variance so that is selected as the 'best' model in approximation of the true model. If we further take the W_i results into account, we realize that AR(4) is not a dominant 'good' model, similar to what presented in Table 2.1. AIC model selection procedure exhibits a good capability to cope with the sampling variation.

A few useful summary points on the usage of AIC and how it relates to the hypothesis test in general are given below. These points are essentially adapted from Burnham and Anderson (2002) and in agreement with our experience.

(1) AIC cannot be used to compare models estimated from fitting to different data sets. It is very obvious, but is easily ignored that data must be fixed, because

the inference is conditional on the data behind.

(2) If we want to compare models from fitting different distributions (to the same data set, of course), we need to keep all parts of the model likelihood expression with each of the candidate models (Section 2.4.1). Any dropout of constant terms or scaling will make the calculated AIC values become incomparable.

(3) Order is not important in computing AIC values. It is the relative differences (ΔAIC_i) between each pair of the candidate models rank models hence we can make a comparison (Section 2.4.1 and 2.4.2).

(4) If it is a small sample case, i.e. $n/k < 40$, use AICc instead of AIC. The AICc may be applied in general, i.e. not limited to normal distribution (Section A.4).

(5) If the calculated Akaike weights W_i show that there are more than one model strongly supported by the data, a single best model selection strategy would be inappropriate (Section 2.4.1).

(6) Information-theoretic approaches and tests of null hypotheses should not be used together. The theories underlying the information-theoretic approach and null hypothesis testing are fundamentally different. A criterion is not a test and the two approaches are very different analysis paradigms. It is meaningless to associate concepts such as test power or p-values or α levels with an information-theoretic criterion.

(7) The results of model selection under the information-theoretic approach and hypothesis testing might happen to be similar with simple problems. However, the real difference between these two approaches is that there is a theoretical basis to the former approach, while the latter approach for model selection must be considered as *ad hoc*.

(8) The hypothesis testing approach is a useful tool for statistical inference in a dichotomous decision making situation, e.g. testing normality or independence of a sample. The hypothesis testing approach will still play an important role in experimental data analysis where randomization and replication of observations are possible .

In application of AIC, we need to explicitly identify the number of estimated parameters as the model dimension k. "The true model should be included in the specified candidate models" is also a strong condition for a valid use of AIC. In applying TIC for model selection, the effective model dimension is derived from the trace term $\mathrm{tr}\big(K(\hat{\underline{\theta}})J(\hat{\underline{\theta}})^{-1}\big)$ calculation and model misspecification is allowed. Some simple examples using TIC follow.

2.4.2 Examples of using TIC

In this subsection, we will fit some simple *probability distribution models* to some simulated sample data to investigate how different AIC and TIC are in evaluating a fitted model.

It is more difficult to calculate TIC than AIC due to the computation of the trace term in Equation (2.6). In order to evaluate this trace term, two $k \times k$ matrices $K(\hat{\underline{\theta}})$ and $J(\hat{\underline{\theta}})^{-1}$ need to be evaluated based on Equations (A.8) and (A.9). It seems that, so far no built-in R function is available to calculate TIC. Therefore, some special R programs are written to do the job.

The major built-in R functions used in this study are: `optimize` (for univariate case) and `optim` as the minimizer functions, and function `D` for evaluating derivatives. `optim` is a general-purpose optimization function in R and we have kept using the default method in `optim` for optimization calculation in this research. The default method in `optim` is an implementation of the Nelder and Mead (1965) algorithm, that uses only function values and is robust but relatively slow. It will work reasonably well for non-differentiable functions according to the R manual. After the 2.7.1 version, the R function `D` can be used to evaluate the second derivative of a *gamma function*. This makes programming in R much easier for calculating TIC and GIC because several important probability distributions, e.g. *gamma distribution* and *Weibull distribution*, do involve the gamma function.

We have designed two scenarios to compare AIC and TIC in evaluation of different models.

Scenario 1: Data (of size $n = 1000$) are generated from a unit mean exponential distribution using the R commands: `set.seed(101); rexp(1000)`.
Three distribution models, denoted by 'fitexp' for exponential, 'fitwei' for Weibull, and 'fitgam' for gamma, are fitted to the data. Since the exponential distribution is a special case for both the Weibull and the gamma distributions, this is a situation that

we should expect to see AIC and TIC behave almost the same (except for differences due to sampling variation). The model evaluation results are summarized in Table 2.3.

Scenario 2: Data (of size $n = 1000$) are generated from a lognormal distribution using the R commands: `set.seed(101); rlnorm(1000, meanlog=0, sdlog=2)`. Four distribution models are fitted to the data: 'fitexp' for exponential; 'fitwei' for Weibull; 'fitgam' for gamma; and 'fitlnorm' for lognormal distribution. This is a different situation from Scenario 1, although both scenarios satisfy the three AIC assumption conditions (beginning of Section 2.3.1). We would like to see how different are AIC and TIC as summarized in Table 2.4.

In Tables 2.3 and 2.4, notation W_{aic} and W_{tic} stand for 'Akaike weights' calculated from ΔAIC and ΔTIC based on formula (2.21), respectively. The notation 'EMD' stands for 'Effective Model Dimension' as defined in Equation (2.11) and explained in the paragraph after the definition. Since, for AIC, EMD $= k$ is a constant for a fitted model, it makes more sense using the Effective Model Dimension notion for TIC or GIC only. Here we have EMD $= \text{tr}\big(K(\hat{\underline{\theta}})J(\hat{\underline{\theta}})^{-1}\big)$ for TIC.

Table 2.3: Comparison of AIC versus TIC for model selection (1)

Fitted models	AIC	ΔAIC_i	W_{aic}	TIC	ΔTIC_i	W_{tic}	EMD
fitexp (k=1)	1935.8	0.51	0.290	1936.0	0.70	0.271	1.07
fitwei (k=2)	1935.6	0.22	0.335	1935.5	0.23	0.344	1.99
fitgam (k=2)	**1935.3**	**0**	**0.375**	**1935.3**	**0**	**0.385**	**1.98**

Table 2.4: Comparison of AIC versus TIC for model selection (2)

Fitted models	AIC	ΔAIC_i	W_{aic}	TIC	ΔTIC_i	W_{tic}	EMD
fitexp (k=1)	5364.3	1360.1	0.000	5362.3	1358.2	0.000	0.020
fitwei (k=2)	4152.0	147.8	0.000	4154.3	150.2	0.000	3.14
fitgam (k=2)	4389.4	385.2	0.000	4403.1	399.0	0.000	8.86
fitlnorm (k=2)	**4004.2**	**0**	**1.000**	**4004.1**	**0**	**1.000**	**1.94**

In Table 2.3, we observed that EMD $\approx k$ as we expected. The reason is explained in Equation (A.7) and in the paragraph following Equation (2.6). The model evaluation results for AIC and TIC in Scenario 1 have no real difference. Note that both AIC and TIC have selected gamma distribution as the 'best' model ($\Delta\text{AIC} = \Delta\text{TIC} = 0$) which is NOT the 'true' model. From the W_{aic} and W_{tic} results, however, we notice that all three candidate models are strongly supported by the data in terms of bias-variance

trade-off. Again, we should not take the 'single one best-fit model' strategy in this situation as explained in Section 2.4.1.

Table 2.4 is different from Table 2.3 in several aspects. The EMD is no longer close to k for models fitexp, fitwei, and fitgam. EMD $\approx k$ (i.e. $1.94 \approx 2$) with model fitlnorm because the lognormal distribution is the true model which generates the data as we have designed. Both W_{aic} and W_{tic} results indicate that fitlnorm is dominantly supported by the data and this is a single best-fit model situation. We notice that although EMD is substantially different from k if a model is 'misspecified' (i.e. AIC score is different from TIC beyond the two-unit warning range), ΔAIC and ΔTIC end up picking the same 'best' model and this is not by coincidence. We will give more discussions on this issue in Section 2.5.1.

Table 2.4 shows us a substantial advantage in using information-theoretic criteria that they are valid for non-nested models. The likelihood ratio method of hypothesis testing is as widely applicable as maximum likelihood estimation (Casella and Berger, 2002). However, the traditional *likelihood ratio tests* are defined only for nested models. This represents another substantial limitation in the use of hypothesis testing in model selection.

2.4.3 Example of using GIC

Application of GIC is more general in that the parameter estimation is no longer restricted to MLE. We will fit two basic ACD models to a real data set and use GIC for model evaluation.

IBM transaction data for the period 01/11/1990 to 31/01/1991 are used for model fitting. The details about this data set and the raw data treatment are given in Section 3.2.1. After some basic cleaning and adjustments to the raw transaction data, we obtained 51304 observations at 46060 distinct transaction times. To avoid any possible concerns about the process stationarity, only data from the first 50 trading days are used. So the final data set used for this analysis is of sample size $n = 32965$. Because intraday transactions tend to exhibit diurnal patterns, which can not be ignored (Engle and Russell, 1998), ACD models are usually fitted to adjusted inter-transaction durations.

Let $\{y_t\} = \{y_1, y_2, ..., y_n\}$ be the adjusted inter-transaction durations series, a basic

ACD(1,1) model is defined by

$$\begin{aligned} y_t &= \psi_t\,\epsilon_t \\ \psi_t &= \omega + \alpha y_{t-1} + \beta\psi_{t-1} \qquad t = 1, 2, ..., n, \end{aligned} \tag{2.22}$$

where ϵ_t is the multiplicative error term assumed to be iid with unit mean.
Symbol $\psi_t = \omega + \alpha y_{t-1} + \beta\psi_{t-1} = \mathrm{E}(y_t|\mathcal{F}_{t-1})$ denotes the conditional expectation of the adjusted duration between the $(t-1)$st and tth trades, where $\mathcal{F}_{t-1}$ is the information set available at the $(t-1)$st trade.

Let EACD(1,1) denote an ACD(1,1) model where ϵ_t follows a unit mean exponential distribution. The conditional distribution model that describes an EACD(1,1) is

$$\begin{aligned} f(\epsilon_t|\mathcal{F}_{t-1}, \underline{\theta}) &= \exp\Big\{-\frac{y_t}{\psi_t}\Big\} \\ &= \exp\Big\{-\frac{y_t}{\omega + \alpha\ y_{t-1} + \beta\psi_{t-1}}\Big\}, \end{aligned} \tag{2.23}$$

where $\underline{\theta} = (\omega, \alpha, \beta)^T$ and $\omega > 0,\ \ \alpha, \beta \geq 0$. Hence, the conditional likelihood function used for model fitting evaluation is given by

$$\begin{aligned} L(\underline{y}|\hat{\underline{\theta}}, y_1) &= \prod_{t=2}^{n} f(\epsilon_t|\mathcal{F}_{t-1}, \hat{\underline{\theta}}) \\ &= \prod_{t=2}^{n} \exp\Big\{-\frac{y_t}{\hat{\omega} + \hat{\alpha}\ y_{t-1} + \hat{\beta}\psi_{t-1}}\Big\}. \end{aligned} \tag{2.24}$$

Whereas, $\hat{\underline{\theta}}$ is obtained from the quasi MLE based on the following conditional likelihood function

$$\begin{aligned} L^*(\underline{y}|\underline{\theta}, y_1) &= \prod_{t=2}^{n} f(y_t|\mathcal{F}_{t-1}, \underline{\theta}) \\ &= \prod_{t=2}^{n} \Big\{\frac{1}{\omega + \alpha\ y_{t-1} + \beta\psi_{t-1}} \exp\Big(-\frac{y_t}{\omega + \alpha\ y_{t-1} + \beta\psi_{t-1}}\Big)\Big\}, \end{aligned} \tag{2.25}$$

where the conditional distribution pdf of y_t is derived from the pdf of ϵ_t based the standard result of mathematical statistics

$$f(y_t) = f\Big(\epsilon_t = \frac{y_t}{\psi_t}\Big)\Big|\frac{d\epsilon_t}{dy_t}\Big|. \tag{2.26}$$

Quasi MLE $\hat{\underline{\theta}}$ in Equation (2.24) is a solution of the conditional likelihood equation

$$\frac{\partial \log L^*(\underline{y}|\underline{\theta}, y_1)}{\partial\underline{\theta}} = \underline{0}, \tag{2.27}$$

where $L^*(\underline{y}|\underline{\theta}, y_1)$ is given by Equation (2.25).

Similarly, let WACD(1,1) denote an ACD(1,1) model where ϵ_t follows a unit mean Weibull distribution. The conditional distribution model that describes a WACD(1,1) is

$$\begin{aligned} f(\epsilon_t|\mathcal{F}_{t-1}, \underline{\theta}) &= \frac{a\epsilon_t^{a-1}}{b^a} \exp\Big\{ -(\frac{\epsilon_t}{b})^a \Big\} \\ &= \frac{a\psi_t}{y_t}\Big(\frac{y_t\,\Gamma(1+1/a)}{\psi_t}\Big)^a \exp\Big\{ -\Big(\frac{y_t\,\Gamma(1+1/a)}{\psi_t}\Big)^a \Big\}, \end{aligned} \tag{2.28}$$

where $\underline{\theta} = (\omega, \alpha, \beta, a)^T$ and $\omega > 0,\ \alpha, \beta \geq 0$. Parameters a and b are the shape and scale parameters of a Weibull distribution, respectively, so that $a, b > 0$. The second line equality holds because that $\epsilon_t = y_t/\psi_t$ and $b\,\Gamma(1+1/a) = 1$ (unit mean constraint). Based on Equation (2.28) and (2.26), the conditional distribution pdf of y_t can be expressed as

$$f(y_t|\mathcal{F}_{t-1}, \underline{\theta}) = \frac{a}{y_t}\Big(\frac{y_t\,\Gamma(1+1/a)}{\psi_t}\Big)^a \exp\Big\{ -\Big(\frac{y_t\,\Gamma(1+1/a)}{\psi_t}\Big)^a \Big\}. \tag{2.29}$$

Hence, the conditional likelihood function used for model fitting evaluation can be expressed as

$$\begin{aligned} L(\underline{y}|\hat{\underline{\theta}}, y_1) &= \prod_{t=2}^{n} f(\epsilon_t|\mathcal{F}_{t-1}, \hat{\omega}, \hat{\alpha}, \hat{\beta}, \hat{a}) \\ &= \prod_{t=2}^{n} \Bigg\{ \frac{a(\hat{\omega} + \hat{\alpha}\ y_{t-1} + \hat{\beta}\psi_{t-1})}{y_t} \Big(\frac{y_t\,\Gamma(1+1/a)}{\hat{\omega} + \hat{\alpha}\ y_{t-1} + \hat{\beta}\psi_{t-1}}\Big)^a \\ &\qquad \times \exp\Big\{ -\Big(\frac{y_t\,\Gamma(1+1/a)}{\hat{\omega} + \hat{\alpha}\ y_{t-1} + \hat{\beta}\psi_{t-1}}\Big)^a \Big\} \Bigg\}, \end{aligned} \tag{2.30}$$

where $\hat{\underline{\theta}} = (\hat{\omega}, \hat{\alpha}, \hat{\beta}, \hat{a})^T$ is obtained from the quasi MLE based on the conditional likelihood function below

$$\begin{aligned} L^*(\underline{y}|\underline{\theta}, y_1) &= \prod_{t=2}^{n} f(y_t|\mathcal{F}_{t-1}, \omega, \alpha, \beta, a) \\ &= \prod_{t=2}^{n} \Bigg\{ \frac{a}{y_t}\Big(\frac{y_t\,\Gamma(1+1/a)}{\omega + \alpha\ y_{t-1} + \beta\psi_{t-1}}\Big)^a \exp\Big\{ -\Big(\frac{y_t\,\Gamma(1+1/a)}{\omega + \alpha\ y_{t-1} + \beta\psi_{t-1}}\Big)^a \Big\} \Bigg\}. \end{aligned} \tag{2.31}$$

From the problem description above, it is clear that it is GIC rather than AIC or TIC we should use for model evaluation. Equations (A.14) and (A.15) are needed to calculate the trace term $\mathrm{tr}\big\{R(\psi, \hat{G})^{-1}Q(\psi, \hat{G})\big\}$ in formula (2.7).

The empirical M-estimator $\hat{\underline{\theta}}$ is obtained as the solution of the equation

$$\sum_{t=2}^{n} \psi(y_t, \hat{\underline{\theta}}) \;=\; \frac{\partial \log L^*(\underline{y}|\hat{\underline{\theta}}, y_1)}{\partial \underline{\theta}} \;=\; \underline{0}, \tag{2.32}$$

where $L^*(\underline{y}|\underline{\theta}, y_1)$ is defined in Equation (2.25) for an EACD(1,1) model or (2.31) for a WACD(1,1) model. Furthermore, we have

$$\psi(y_t, \underline{\theta}) = \frac{\partial}{\partial \underline{\theta}} \; \log f(y_t|\mathcal{F}_{t-1}, \underline{\theta}), \tag{2.33}$$

and the conditional distribution model is defined by $f(\epsilon_t|\mathcal{F}_{t-1}, \underline{\theta})$ for EACD(1,1) in Equation (2.23) or for WACD(1,1) in Equation (2.28).

Therefore, the $R(\psi, \hat{G})$ and $Q(\psi, \hat{G})$ matrices can be expressed as

$$R(\psi, \hat{G}) = -\frac{1}{n} \sum_{t=2}^{n} \frac{\partial \psi(y_t, \underline{\theta})}{\partial \underline{\theta}} \Big|_{\underline{\theta} \,=\, \hat{\underline{\theta}}}, \tag{2.34}$$

and

$$Q(\psi, \hat{G}) = \frac{1}{n} \sum_{t=2}^{n} \psi(y_t, \underline{\theta}) \, \frac{\partial \log f(\epsilon_t|\mathcal{F}_{t-1}, \underline{\theta})}{\partial \underline{\theta}^T} \Big|_{\underline{\theta} \,=\, \hat{\underline{\theta}}}. \tag{2.35}$$

Now we are ready to calculate GIC for an ACD model. Following the effective model dimension concept in Equation (2.11)

$$\mathrm{GIC}(k) \;=\; -2 \, \log L(\underline{y}|\hat{\underline{\theta}}, y_1) + 2 \, \mathrm{EMD}, \tag{2.36}$$

where $\mathrm{EMD} \;=\; \mathrm{tr}\big\{R(\psi, \hat{G})^{-1} Q(\psi, \hat{G})\big\}$.

Based on Equation (2.8), the M-estimator version AIC score can also be calculated for an ACD model as

$$\mathrm{AIC}(k) \;=\; -2 \, \log L(\underline{y}|\hat{\underline{\theta}}, y_1) + 2 \, k, \tag{2.37}$$

where $L(\underline{y}|\hat{\underline{\theta}}, y_1)$ is defined by Equation (2.24) or (2.30).

We further define a *Pseudo-AIC* (AIC^*) as given below before presenting the model fitting results.

$$\mathrm{AIC}^*(k) \;=\; -2 \, \log L^*(\underline{y}|\hat{\underline{\theta}}, y_1) + 2 \, k, \tag{2.38}$$

where $L^*(\underline{y}|\hat{\underline{\theta}}, y_1)$ is defined by Equation (2.25) or (2.31). Equation (2.38) is an incorrect formula to calculate AIC for an ACD model. We calculate AIC^* to show how wrong it can be compared with the correct AIC score calculated from Equation (2.37).

Table 2.5 gives the model estimation results for an EACD(1,1) and a WACD(1,1) model in fitting the IBM share transaction data collected in early 1990s.

Based on Equations (2.36), (2.37), and (2.38), given the model estimation results in Table 2.5, the model evaluation results are summarized in Table 2.6. Parameters a and b are *shape* and *scale* respectively for a Weibull distribution and b is not a free parameter due to the unit mean constraint.

Table 2.5: Parameter estimates for fitted ACD models (IBM data)

Parameter (M-estimate)	EACD(1,1)	WACD(1,1)
$\hat{\omega}$	0.01079	0.01085
$\hat{\alpha}$	0.06111	0.06094
$\hat{\beta}$	0.9294	0.9292
$\hat{a}$	—	0.8841
$(\hat{b})$	—	0.9410

Table 2.6: AIC*, AIC, GIC and summary statistics for the fitted ACD models (IBM data)

Model	AIC*	AIC	GIC	EMD	mean(ϵ_t)	L-B($\hat{\epsilon}_t$) (lag=15)	L-B($\hat{\epsilon}_t^2$) (lag=15)
EACD(1,1) k=3	63908.8	65930.6	65934.4	4.92	1.000	55.8	35.3
WACD(1,1) k=4	62982.3	65312.6	65313.8	4.57	1.005	55.7	34.7

L-B statistic is a short name for Ljung-Box statistic[2] which is used to detect the autocorrelation of sample data. The L-B statistic is the most often used hypothesis test statistic in diagnostic check of a fitted ACD model. The remaining serial correlation in the residuals, measured both on the residuals and squared residuals, is considered as an important indicator for goodness-of-fit of the models. The R function `Box.test(...,type = "Ljung")` is used for calculating L-B statistic.

From Table 2.6, we notice that the difference between AIC and GIC is smaller with WACD(1,1) than with EACD(1,1) and GIC_{wacd} is smaller than GIC_{eacd}, too. This may indicates that (1) WACD(1,1) fits the data better and (2) therefore the discrepancy between k and EMD is smaller with WACD(1,1).

We also notice that AIC and GIC are in agreement with the conclusion that a WACD(1,1) model fits better to the data than an EACD(1,1) in this case. On the

[2]A detailed definition for L-B statistic can be found in Appendix A.3.

other hand, AIC* differs a lot from AIC in both models. We will talk about more on the relationship between AIC and GIC in Section 2.5.1. It is also fine in terms of the estimated residual means (they are all very close to one as they should).

However, the real concern comes from the L-B statistics results calculated from the estimated residual series (Table 2.6). In Figure 2.1, the empirical distribution of L-B statistics calculated from simulated unit mean random samples are plotted to check where the corresponding estimated residual (and squared residuals) L-B statistics are located. Here we are using the parametric bootstrap method (Efron and Tibshirani, 1993, p53) to draw statistical inference from a fitted model. If the estimated residual L-B statistics are well located within the range of the empirical distribution of the L-B statistics from random samples, we may assume that the iid condition for properly specifying the model likelihood functions has been reasonably satisfied.

In Figure 2.1, the top-left panel shows the histogram of L-B statistics (lag=15) calculated from 1000 random samples (each of size n=32965) of a unit mean exponential distribution. Based on the model evaluation results given in Table 2.6, the vertical dashed line is drawn to indicate the estimated residual L-B statistic of the fitted EACD(1,1) model. The bottom-left panel shows L-B statistics histogram calculated from the squared random sample data values and the vertical dashed line is the corresponding estimated squared residual L-B statistic. The right column panels display the same contents as in the left column panels but random sample are from a unit mean Weibull distribution (Weibull$(x|\hat{a},\hat{b})$ with $\hat{a} = 0.8841$ and $\hat{b} = 0.9410$ as given in Table 2.5) and the estimated residual L-B statistics are from the fitted WACD(1,1) model accordingly. Figure 2.1 shows that, for both models, the estimated residual L-B statistics (the vertical dashed lines) are far away from the range of those random sample L-B statistics distributions (the histograms); the estimated L-B statistics are at the upper end of the histograms for squared residuals. It is, therefore, unreasonable to assume that the estimated residuals are coming from a random sample. This implies the violation of the iid assumption condition. Therefore, those conditional likelihood functions (e.g. Equations (2.24), (2.25), (2.30), and (2.31)) which we have relied on for calculating AIC and GIC have been incorrectly specified. Even the parameter estimation results become questionable.

In this section we have shown how AIC and GIC may be calculated for a fitted ACD model although it is an unsuccessful attempt due to the violation of iid condition in residuals. It is the objectives of next chapter to show how to fit ACD models properly

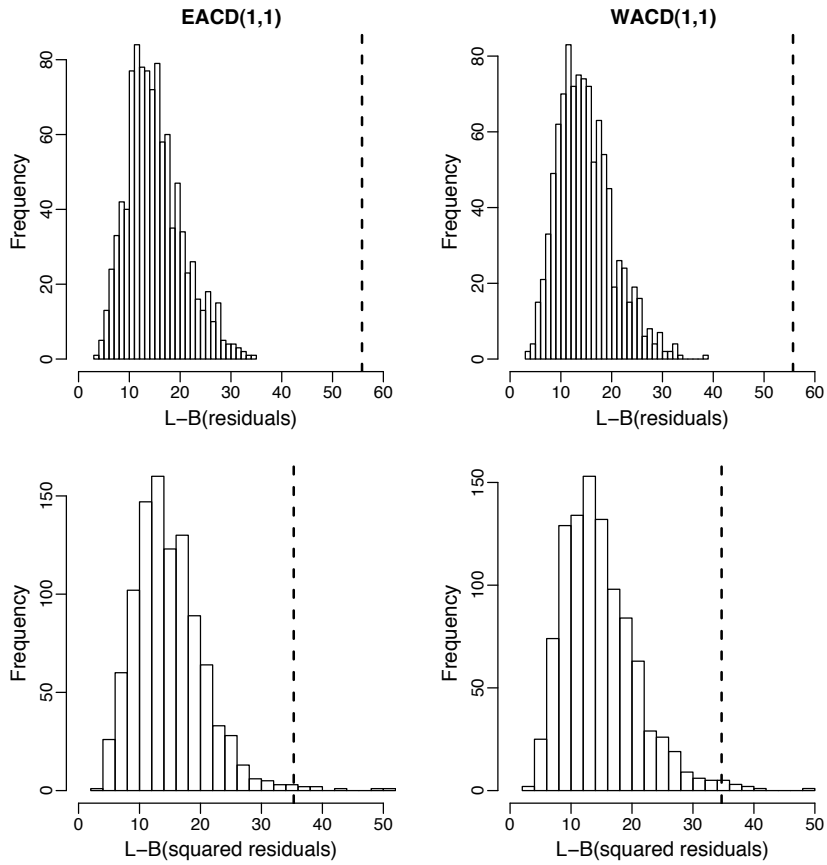

Figure 2.1: L-B statistics of estimated residual series versus the empirical distribution of L-B statistics of random samples

so that information-theoretic criteria can be applied correctly for model selection.

2.5 Discussion

2.5.1 Which one to use, AIC, TIC, or GIC

As we have shown in Section 2.3.1, AIC is a special case of TIC and GIC is an even more general criterion from which TIC can be derived. A logical conclusion becomes that we should always use GIC for model evaluation. However, from a practical standpoint, this may not be always the optimal choice or even necessary.

We assume that the large sample condition is satisfied so that the need to use AICc, instead of AIC, TIC, and GIC, is not in consideration in our following discussions.

Let us first consider the possible choice between AIC and TIC for model evaluation and selection. As we have shown in previous examples that the trace term EMD=

$\mathrm{tr}\big(K(\hat{\underline{\theta}})J(\hat{\underline{\theta}})^{-1}\big)$ is about equal to k for "good" models. What we mean by "good" models may be explained by Akaike (1974) in which Akaike stated that if the true distribution that generated the data exists near the specified parametric model, the bias associated with the sample log-likelihood of the model estimated through the empirical distribution of the truth can be approximated by the number of parameters.

When a fitted model is misspecified, Burnham and Anderson (2002) found that EMD's behaviour is unpredictable, i.e. EMD can be $= k$, $> k$, or $< k$. The change of AIC scores is independent of the pattern of EMD. Figure 2.2 and 2.3, which are adapted from Example 11 (p64) and Figure 5.3 (p125) in Konishi and Kitagawa (2008), provide two examples to illustrate this. For Figure 2.2, the data samples are generated from the true model

$$g(x) = (1-\epsilon)\phi(x|\mu_1,\sigma_1) + \epsilon\phi(x|\mu_2,\sigma_2) \qquad (0 \le \epsilon \le 1),$$

where ϵ denotes the mixing ratio. $\phi(x|\mu,\sigma)$ is the pdf of a normal distribution with mean μ and standard deviation σ and we have set $\mu_1 = \mu_2 = 0$; $\sigma_1 = 4$, $\sigma_2 = 1$. At each of the 20 different ϵ values, 200 simulated samples (n=1000) are generated. A normal distribution model is then fitted to each of the 200 simulated samples and AIC scores and EMDs are calculated. For calculating TIC, the EMD=$\mathrm{tr}\big(K(\hat{\underline{\theta}})J(\hat{\underline{\theta}})^{-1}\big)$ is defined in Equation(2.6). The EMD values are plotted in the upper panel. The solid line represents the mean EMDs and two dashed lines form a 95% quantile confidence band. The AIC scores are plotted in the lower panel with solid line for the means and dashed lines for 95% quantile confidence band. Figure 2.3 is produced in the same way as with Figure 2.2, but the data samples are generated from a different true model:

$$g(x) = (1-\epsilon)\big[0.3\lambda_1\exp(-\lambda_1 x)+0.7\lambda_2\exp(-\lambda_2 x)\big]+\epsilon\,\mathrm{Weibull}(x|shape,scale) \quad (0 \le \epsilon \le 1),$$

where ϵ denotes the mixing ratio and we have set $\lambda_1 = 0.5$, $\lambda_2 = 2$. The expression $\mathrm{Weibull}(x|shape,scale)$ represents the pdf of a Weibull distribution with $shape = 3$ and $scale = 1.5$. A mixed-exponential distribution model $p\lambda_1\exp(-\lambda_1 x) + (1-p)\lambda_2\exp(-\lambda_2 x)$ $(0 < p < 1)$ is then fitted to the simulated data and AIC scores and EMDs are calculated.

From Figure 2.2, in the upper panel we notice that EMD $= k = 2$ when the fitted model is correctly specified, i.e. when $\epsilon = 0$ or $\epsilon = 1$. The AIC score pattern in the lower panel shows a monotonic increase over the whole ϵ range. From Figure 2.3, again we observe that EMD $= k = 3$ when the fitted model is correctly specified, i.e.

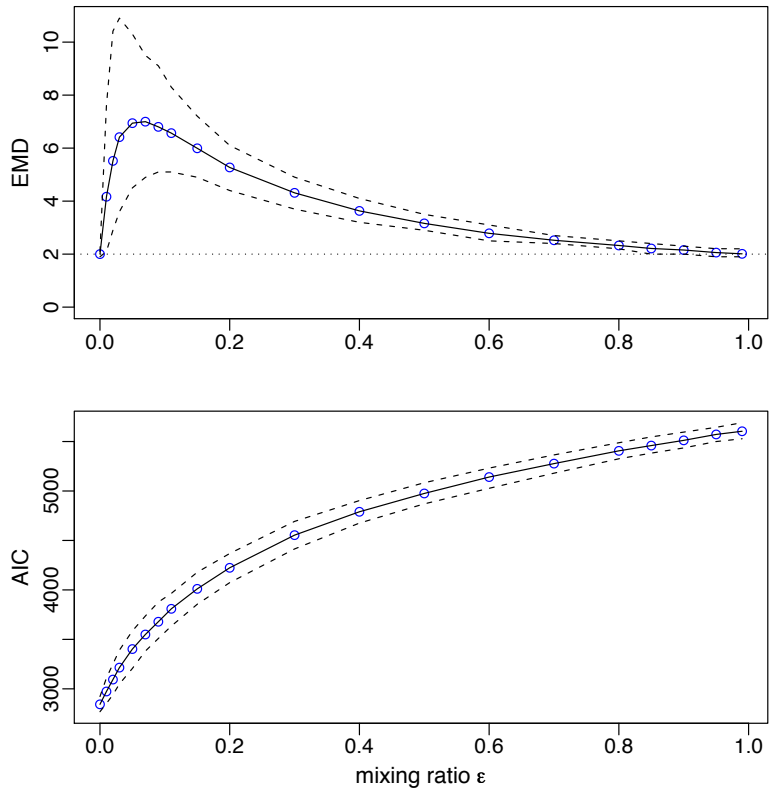

Figure 2.2: AIC versus TIC: A normal distribution model case

when $\epsilon = 0$ in the upper panel. However, both the upper panel and the lower panel in Figure 2.3 show complete different patterns from that in Figure 2.2. From both Figure 2.2 and 2.3 we notice that the absolute level of EMDs is very small compared with the absolute level of the AIC scores. Also, the sampling variation range of EMDs (less than 10) is much less than that of AIC scores (about a few hundreds).

If large samples are available, TIC might offer an improvement over AIC. However, it has been found that the estimation error of those two $k \times k$ matrices $K(\hat{\underline{\theta}})$ and $J(\hat{\underline{\theta}})$ can cause instability of the results of model selection (critical remarks on TIC, Section 2.3.3). When k gets large, $k > 10$ say, the computation of EMD can become a problem, too. Burnham and Anderson (2002), p435, reported that for all the cases they examined (in comparison of AIC versus TIC), if the model was less general than the truth (the real world case), they predominantly found that $\mathrm{tr}\big(K(\hat{\underline{\theta}})J(\hat{\underline{\theta}})^{-1}\big) < k$. Thus, use of AIC should then often lead to slightly more parsimonious models than the use of TIC. Our experience in using AIC and TIC agrees with what they

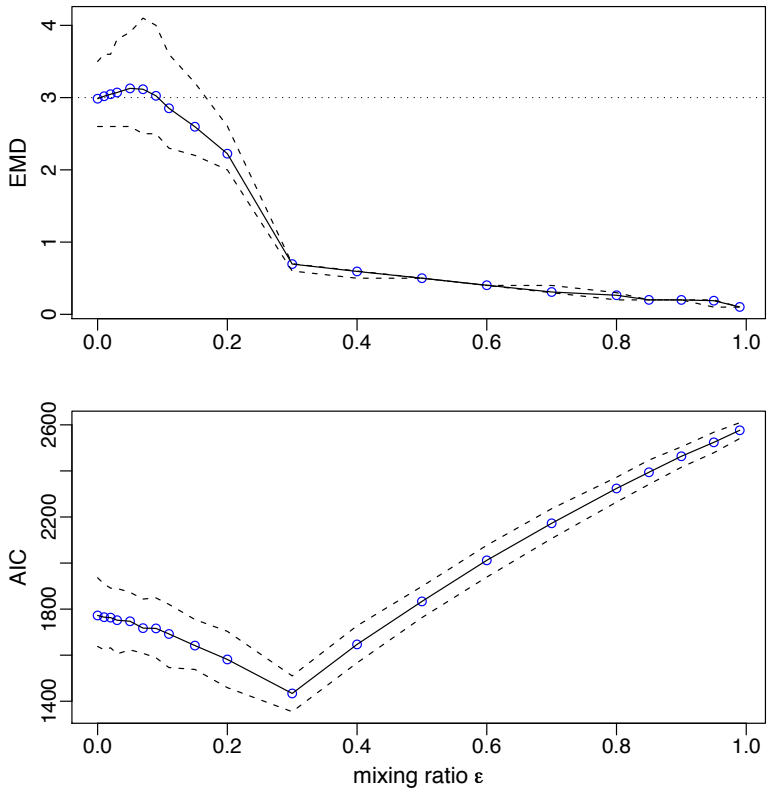

Figure 2.3: AIC versus TIC: A mixed-exponential distribution model case

reported. From a practical point of view, Konishi and Kitagawa (2008) argued that the AIC does not require any analytical derivation of the bias correction terms for individual problems and does not depend on the unknown probability distribution G, which removes fluctuations due to the estimation of the bias.

Theoretically, GIC is a general model evaluation criterion which extends the application domain to include the M-estimator framework. In practice, however, the calculation of the GIC trace term can be non-trivial. Because of the M-estimator version AIC (Equation 2.8) derived by Konishi and Kitagawa (2008), the research findings from comparing AIC against TIC are likely applicable for the comparison between AIC and GIC.

In principle, the application of using AIC for model evaluation should have been considerably limited by the strong true model inclusion assumption condition. However, from a practical point of view, we consider AIC to be a good proxy for TIC or GIC in model evaluation for the following reasons. By definition, the difference between

AIC and TIC (as well as between the M-estimator version AIC and GIC) is in their trace terms. Therefore, when the specified models are close to the true distribution, this implies that $k \approx \text{tr}\big(K(\underline{\hat{\theta}})J(\underline{\hat{\theta}})^{-1}\big)$ or $k \approx \text{tr}\Big\{R(\psi,\hat{G})^{-1}Q(\psi,\hat{G})\Big\}$; then AIC $\approx$ TIC or AIC $\approx$ GIC. In the case that a model is badly misspecified (the fitted model is far away from the true model), the sample log likelihood term (evaluated at MLE or at M-estimate) is in absolute domination over the trace term so that a bad model would not stand a chance to be selected. AIC and its generalized variants will select the same best model as long as at least there is one 'good' model, i.e. not all candidate models have almost the same 'distance' (within two AIC units variation, as a rule of thumb) from the true model.

To sum up, we reach the following conclusion. AIC is a practical and parsimonious implementation of TIC and GIC. The theoretical justification of the M-estimator version AIC has virtually removed the need to calculate TIC or GIC in model evaluation practice. Given required regularity conditions, it is always valid to apply AIC (MLE version or M-estimator version) for model evaluation in data analysis practice without worrying about the true model inclusion assumption. More real data empirical study cases are needed to test this conclusion.

2.5.2 Pearson's Q-statistic and bootstrap quantile-quantile plot

The χ^2 statistic and quantile-quantile plot are two often used tools to measure the similarities between distributions or goodness-of-fit.

The Q-statistic, as proposed by Karl Pearson in 1900, is defined as

$$\text{Q} = \sum_{i=1}^{m} \frac{(N_i - np_i)^2}{np_i}, \tag{2.39}$$

where N_i denotes the number of observations of the ith category and p_i is the probability of success (to obtain an observation which belongs to the ith category).

According to DeGroot (1986), the Q-statistic may be applied to any continuous distribution provided that the value of each expectation $np_i \geq 1.5$ for $i = 1, 2, ..., m$. Note that, given d unknown estimated parameters, the degrees of freedom for the asymptotic χ^2 distribution should then be $m - d - 1$ instead of $m - 1$ (see, e.g. Krishnamoorthy, 2006, p18). Since χ^2 statistics can be used to assess model fit without a specific alternative in mind (See e.g. Davison, 2003, p133), we consider using Q-statistic of Equation 2.39) as a measure to rank the goodness-of-fit of different models so that the

model evaluation results may be compared with those obtained by the information-theoretic approach. Note that there is a subtle but important difference here between the Q-statistic and χ^2 distribution. Unless the strong assumption condition, i.e. given the true distribution has been correctly specified by the model, is met, we should not assume that a Q-statistic follows an asymptotic χ^2 distribution. We will purposely avoid the traditional hypothesis testing approach but use Q-statistic as a goodness-of-fit measure to rank models. In real (observational) data analysis situations, we hold the view that ranking models (how good a model fits the observed data in a relative sense) makes more sense than a concept of testing 'adequacy'.

A specific example is given as follows. An iid sample of n = 200 is generated using the following R code

```
set.seed(101); xweibull = rweibull(200, shape=1.8, scale=1)
```

as the observed values.

Table 2.7: Measuring the goodness-of-fit by AIC scores and Pearson's Q-statistic

Fitted models	AIC	ΔAIC$_i$	W_i	Q
fitexp (k=1)	340.3	63.7	0.000	48.1
fitwei (k=2)	**276.6**	**0**	**0.913**	**7.07**
fitgam (k=2)	281.3	4.7	0.087	6.34
fitnorm (k=2)	315.9	39.3	0.000	27.6

Four probability distribution models: exponential$(x|r)$ (denoted by 'fitexp'), Weibull$(x|a,b)$ ('fitwei'), gamma$(x|a,b)$ ('fitgam'), and normal$(x|\mu,sd)$ ('fitnorm') are specified to fit to the generated sample data. Model estimation follows the routine MLE procedure and the results are: the *rate* parameter estimate for model fitexp is $\hat{r} = 1.167$; the *shape* and *scale* estimates for model fitwei are $\hat{a} = 1.662$ and $\hat{b} = 0.9588$; for model fitgam are $\hat{a} = 2.255$ and $\hat{b} = 0.3801$; and the *mean* and *standard deviation* estimates for model fitnorm are $\hat{\mu} = 0.8570$ and $\hat{sd} = 0.5277$. By approximately dividing observations into 10 equal number subgroups (m=10), the Q-statistics are calculated for each model. The model estimation and evaluation results are summarized in Table 2.7 (for reference, $\chi^2_{0.95,df=7} = 14.1$).

Note that AIC procedure conveys more statistical inference messages than that of the Q-statistic. According to the AIC scores, fitwei is identified as the best model (the bold line model). But, the Akaike W_i warns us that model fitgam is also supported

by the data significantly (evidenced also by ΔAIC which is less than 5). However, based on Q-statistic results, model fitgam is the best choice. Since Q-statistic does not take account of the sampling variation in fitting the data and the choice of how to subdivide the observations is somewhat subjective (and it affects the evaluation result), the Q-statistic is considered as a relatively rough (subject to sampling variation), but general and sensible method in evaluating models. This conclusion is supported by the bootstrap version qq-plots shown in Figure 2.4 below. The observed values are generated from the assumed true model: `set.seed(101); rweibull(200, shape=1.8,scale=1)`. For the fitted values, 1000 simulated samples (n=200) are generated from the fitted models (a) fitexp, (b) fitwei, (c) fitgam, and (d) fitnorm. At each of the 200 observed quantile values, the median, 2.5%, and 97.5% quantile values are calculated from the 1000 fitted values. The solid line represents the medians and two dashed lines form a 95% quantile confidence band.

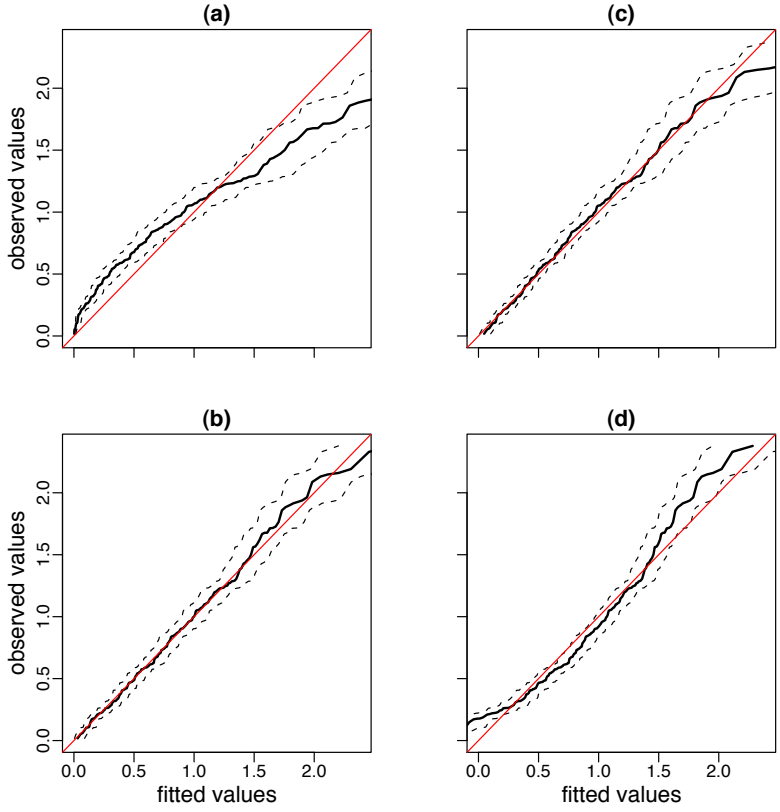

Figure 2.4: A parametric bootstrap version of quantile-quantile plot

A second way of confirming the information-theoretic approach's results is by qq-

plot. However, instead of just plotting any single realization from a fitted model against the observed sample values, we apply the parametric bootstrap idea (as we did in Section 2.5.1) to construct a 95% bootstrap quantile confidence band to check the goodness-of-fit.

Figure 2.4 shows the bootstrap version qq-plot results. Panel (b) shows that model fitwei fits the data perfectly (as it should) and a close look at panel (c) (model fitgam) identifies a slight lack-of-fit in the top-right corner. In panel (a) and (d), both models, fitexp and fitnorm, do not fit the data within the confidence bands. The AIC evaluation results in Table 2.7 are well backed by the bootstrap qq-plot results.

We will use Q-statistic and the bootstrap version qq-plot in checking the ACD model's goodness-of-fit results in next chapter.

2.6 Concluding remarks

Kullback-Leibler information is a unique overall measure of discrepancies between a fitted model and the true model which generates the observed sample. AIC, TIC, and GIC are derived as (asymptotically) unbiased estimators of K-L information. Therefore, they are K-L information-based criteria for model evaluation. Akaike's discovery of the formal relationship between Kullback-Leibler information and maximum likelihood combines model estimation and model evaluation under a single theoretical framework: optimization, which makes the information-theoretic criteria unique among all other model evaluation criteria or procedures. Under the general assumption that the true model is *unknown* and has *infinite dimension*, it can be theoretically justified that the information criteria, derived based on minimizing the relative expected K-L information, have the following unique properties: (1) the selected best model gives an optimal balance between the bias (model accuracy) and the variance (model complexity) which has the minimum predictive error among all the candidate models; (2) the selected best model changes as the sample size changes to allow a more complex model to be selected as more information is available when the sample size is getting larger.

Although the information-theoretic criteria make an automatic model selection process possible, it should be noted that the best way to ensure selecting a 'good' or 'useful' model is to reduce the number of candidate models fit to the data by thoughtful *a priori* model formulation. Once a group of reasonable candidate models are obtained, it is always valid to apply AIC for model evaluation within an M-estimator framework.

This makes the information-theoretic approach applicable to a vast range of statistical models and the computation issue for model evaluation is also facilitated considerably because of the constant trace term specification. We need more real data empirical study cases to test our research findings.

Efron (2000) has given the following comments on model selection and hypothesis testing approach:

> "Model selection is another area of statistical research where important development seem to be building up, but without a definitive breakthrough. ... The fact is that classic Fisherian estimation and testing theory are a good start, but not much more than that, on model selection. In particular, maximum likelihood estimation theory and model fitting do not account for the number of free parameters being fit, and that is why frequentist methods like Mallow's C_p, the Akaike Information Criterion, and Cross-validation have evolved."

In this sense, the information-theoretic approach has been justified as a valid and better alternative for model evaluation in general. Based on the theoretical justification and examples presented in this chapter, various ACD models will be evaluated and compared following an information-theoretic approach in the next chapter.

Chapter 3

Model selection for ACD models – an application

3.1 Introduction

High frequency financial data, which record transaction-by-transaction observations across time, have become more widely available in recent years due to advances in information technology. Such data provide market micro-structure information (O'Hara, 1995) in the form of transaction times and associated asset variables, such as stock price and volume, etc. These high-frequency data differ from more traditional financial time series recorded over discrete time intervals because they usually contain large numbers of irregularly spaced observations in (near) continuous time, and therefore require alternative stochastic models. Since high frequency data can be viewed as points in continuous time it follows that models based on stochastic point processes may be useful in a statistical study of high-frequency data (Bauwens and Hautsch, 2006).

The ACD model is a point process model which takes the interval specification as it was formulated in Engle and Russell (1998). The random durations between consecutive transactions are modelled based on an autoregressive structure as we have seen in Section 2.4.3. During the last decade, various extended ACD models were developed aiming to better describe the market micro-structure and better fit the error term distribution with the real stock transaction data.

Probably due to the inherent inability of hypothesis testing in model evaluation, the literature has so far devoted little attention to model comparison despite the availability of many competing ACD model specifications. There are several possible reasons for this. The first reason is that, because the model estimation of an ACD model fits within the quasi-MLE framework, so that the standard maximum likelihood theory is

no longer applicable to handle the related inferential issues. A second reason may be due to the inherent difficulty in testing the non-nested models. A more fundamental concern is that no known theory can justify what 'good' property a selected best model has in terms of the overall model performance, if a candidate model can ever be identified as the best model by following the hypothesis testing approach. Hence, nothing can generally be said about ranking models.

The most popular model diagnostic check in ACD model papers is to use the L-B statistic (Equation A.20) to test for serial dependence in the estimated residual (and squared residual) series. Luca and Gallo (2004) used quantile-quantile plots (qq-plots) as a visual check on model goodness-of-fit. Bauwens et al. (2004) compared a group of selected ACD models based on the evaluation of density forecasts and the calculated Pearson's Q-statistic (Equation 2.39). Bauwens et al. (2004) stated that they were aiming to compare ACD models using the criterion of predictive ability (p590), but they also acknowledged that their method is not useful for a comparison of misspecified models (p592). Meitz and Terasvirta (2006) introduced a number of new misspecification tests using the Lagrange Multiplier principle to evaluate ACD models. Because these articles followed the hypothesis testing approach for model evaluation, they all imply that a 'true' model (of finite dimension) exists and is measurable. Therefore they always aim to test whether a fitted model is 'adequate' or not compared to the 'true' model. This is fundamentally contrary to the general assumptions of the information-theoretic approach. Given a stock market situation, the trading process of a share can be affected by numerous factors, many of which may be unmeasurable. Therefore, the general assumptions backing up the hypothesis testing approach are unrealistic and we consider that the hypothesis testing approach of model evaluation for ACD models is theoretically inferior. So far, we have found no attempts to use the information-theoretic approach for model evaluation for ACD models.

Quasi-maximum likelihood (QMLE) implies 'model misspecification'. If we may interpret the typical hypothesis testing statement 'given the null hypothesis is true' as 'given the hypothesized distribution is correct', one should always question the validity of a hypothesis test under a misspecification situation. Under the general assumptions of the information-theoretic approach, any of our specified models is almost surely misspecified. As pointed out in Section 2.4.3, for an ACD model the log-likelihood function used for calculating the QMLE, i.e. for parameter estimation, is different from the log-likelihood function which defines the probabilistic property of the model.

It is the latter log-likelihood function that is used to define the *conditional distribution model* (as defined in Section 2.2) of an ACD process. Therefore, the model evaluation for candidate ACD models falls within an M-estimator framework. The ACD model literature has made no distinction between these important, but different, likelihood functions.

In this chapter, we fit various ACD models to two real stock transaction data sets (one is IBM data collected from the New York Stock Exchange and a second one is Darby data from the Kuala Lumpur Stock Exchange). Quasi-maximum likelihood is calculated for model estimation and the information-theoretic approach is followed for model evaluation and comparison. The structure of this chapter is as follows.

Section 3.2 deals with the description of the data sets, preparation of raw data, and exploratory data analysis (EDA). The treated data are ready for confirmatory analysis and EDA results are presented. Some technical details of raw data treatment are also discussed.

Section 3.3 specifies the fitted ACD models. There are six ACD models considered for analysis, four basic ACD models: EACD(1,1), EACD(2,2), WACD(1,1), and WACD(2,2); two mixed distribution ACD models: mixed exponential ACD(1,1) and mixed lognormal-gamma ACD(1,1). All these six ACD models are fitted to the two real data sets.

In order to minimize the serial dependence and to ensure the unit mean constraint with the estimated residual series, a penalty term is added to the QMLE function as the optimization objective function for parameter estimation. This is different from the estimation procedure employed in Engle and Russell (1998). In Section 3.4, the model estimation results are presented and some discussions are given on the model estimation details.

Section 3.5 presents the model evaluation results. AIC scores are calculated for model evaluation. Pearson's Q-statistic is also calculated as an extra evidence to cross check the information-theoretic criteria's evaluation results. First we evaluate and compare the four basic ACD models. In the second subsection, mixed lognormal-gamma ACD(1,1), a newly proposed model, is compared with existing ACD(1,1) models. In the third subsection, a two-stage ACD model fitting procedure is proposed. The Q-statistics and a bootstrap version of qq-plots are used to provide further empirical evidence to confirm the AIC model assessment results.

Section 3.6 gives the overall conclusion and summary of this chapter.

3.2 Data and exploratory analysis

3.2.1 Data description and preparation

The first data set selected for analysis is the IBM transaction data which were used in Engle and Russell (1998) and Zhang et al. (2001). The IBM data were collected for the period 01/11/1990 to 31/01/1991. These data are from the Trades, Orders Reports, and Quotes dataset constructed by Joel Hasbrouck and NYSE (New York Stock Exchange), and consist of 60328 transactions over 63 trading days. The raw data were down loaded from Tsay (2005)'s book webpage:
`gsbwww.uchicago.edu/fac/ruey.tsay/teaching/fts2` in September 2007. Each transaction record contains the trading date and a time stamp, measured in seconds after midnight, which is the time at which the transaction occurred. In addition to the date and transaction time, also included are the price and volume associated with each transaction, and bid and ask quotes at the time of transaction.

For the same reasons as specified in Engle and Russell (1998), observations dated 23/11/1990 and 27/12/1990 are deleted. Therefore, we obtain 59057 transactions at 52581 distinct transaction times for IBM data over 61 trading days. A similar data 'cleaning' procedure is followed so that only 'normal hour' durations are used for analysis. Raw data 'cleaning' details are as follows: (a) Only transactions recorded between 10am and 4pm are included; (b) the overnight waiting times are removed; (c) for each trading day, the first duration is the time interval between the first and second transaction times after 10am, i.e, the time intervals from exact 10am to the first transaction time after 10am are ignored. This is slightly different from what was treated in Engle and Russell (1998).

After all the adjustments to the raw transaction data, there are 51304 observations at 46060 distinct transaction times. This result is very close to what was obtained by Zhang et al. (2001) (51363 observations at 46120 distinct transaction times) and by Engle and Russell (1998) (46091 distinct-time duration observations). In order to focus on the model evaluation methodology issue and avoid the possible complexity caused by multiple transactions consideration, only the distinct transaction time series is analysed, i.e. sample size $n = 46060$ over a 61 trading-day period.

Raw transaction data for Sime Darby BHD, collected from the Kuala Lumpur Stock Exchange (KLSE), were retrieved from the Reuters Database via SIRCA (Securities Industry Research Centre of Asia-Pacific). Sime Darby BHD's stock is often one of the

top 10 most actively traded shares in KLSE in recent years. Like the IBM data, the raw data contain the detailed transaction information which includes transaction time, trading price and volume. One year (01/01/2006 - 31/12/2006) of inter-transaction duration observations of the Darby stock were selected for this study, and contained a total of 241 trading days over the year with no multiple transactions in time. Different from NYSE, there is a midday break on each trading day. The data were 'cleaned' by extracting only transactions recorded between 9am and 12:30pm and between 14:30pm to 17pm, removing overnight waiting times, and taking the first duration of each session (morning or afternoon) to be the time interval between the first and second transaction times after the session begins. After all the adjustments to the raw transaction data, this left 19,561 observations.

Engle and Russell (1998) found that intraday transactions tended to exhibit diurnal patterns, i.e. higher transaction frequency near the open and the close on a trading day, and this diurnal seasonal effect is statistically significant. A similar diurnal pattern is also identified with the Darby data. Figure 3.1 illustrates this issue for both IBM and Darby data. The upper panels show the inter-transaction durations by scatter plots. The solid line superimposed at the bottom part of the scatter plot represents the estimated mean duration times. The lower panels show the mean duration time patterns without the raw duration times in the background. The left column panel plots show a clear inverse U shape with the IBM inter-transaction duration mean levels. In the right column panels, Darby data's mean duration levels show a broken inverse U shape due to the midday break in KLSE. It is this diurnal mean duration level pattern that is treated as the diurnal seasonal effect in a inter-transaction duration time series.

The prepared raw duration data were then imported into the R statistical package (R Development Core Team, 2007) and the R function `supsmu` (the Friedman's super smoother function, Friedman, 1984) was used to calculate the diurnal seasonal effect $\phi(t_i)$ defined in Equation (3.1). The R manual recommends the parameter `span` takes a value between 0.2 and 0.4 for a good smoothing result. We tried a range of values and found a value of 0.2 gave a result that ensured the adjusted durations had significantly reduced the autocorrelation while keeping a mean close to unity as required.

3.2.2 Exploratory data analysis

We decided to use IBM and Darby shares trading data in this ACD model study mainly for two reasons: (1) IBM data has become a benchmark data set and so is well known;

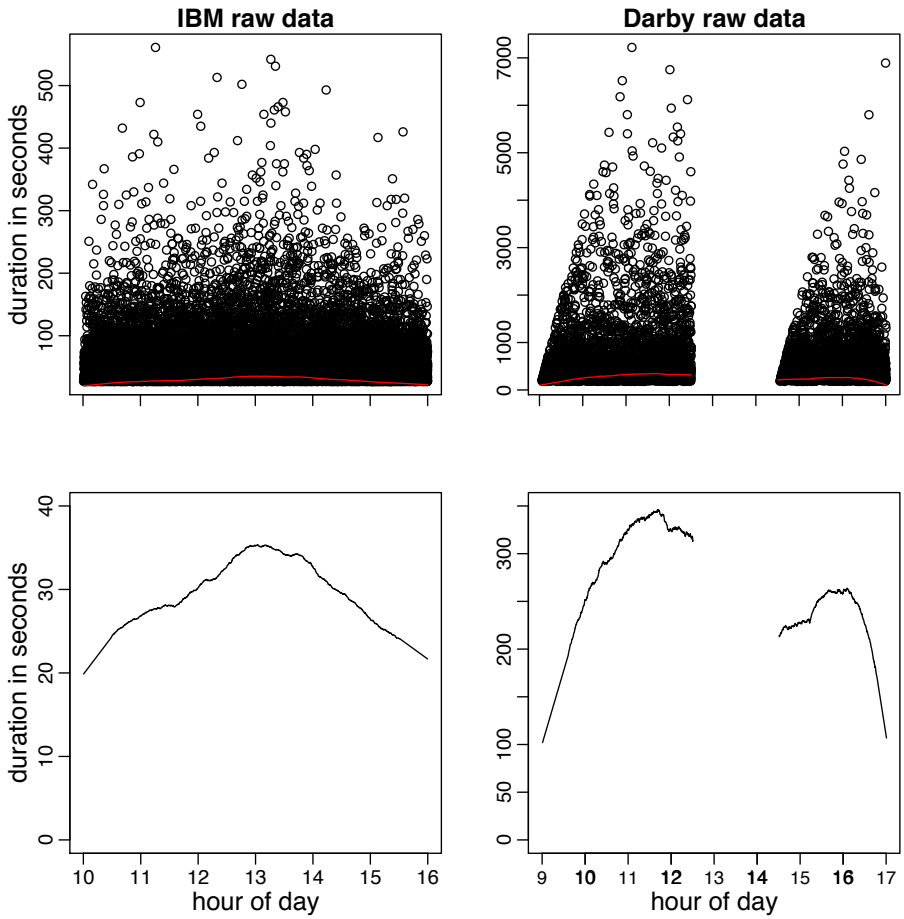

Figure 3.1: Transaction duration time-of-day pattern (duration time is measured in seconds)

(2) Darby data were collected recently from an emerging Asian financial market. The distinction between IBM data and Darby data makes it a good choice to test the information-theoretic approach in model evaluation with data sets collected from two different types of stock markets.

As initially suggested by Engle and Russell (1998), ACD models are usually fitted to diurnal seasonal effect adjusted inter-transaction durations. Let $\{t_i\}$ $(t_0 \leq t_1 \leq t_2 \leq ... \leq t_n)$ be a series of transaction times. Then $\Delta t_i = t_i - t_{i-1} \quad i = 1, 2, ..., n$ are the times between successive transactions, i.e. the inter-transaction durations. The adjusted inter-transaction durations are defined by

$$\Delta t_i^* = \Delta t_i / \phi(t_i) \qquad i = 1, 2, ..., n, \tag{3.1}$$

where $\phi(t_i)$ is the diurnal seasonal effect component of Δt_i obtained by using function

`supsmu` in R. Let us start the exploratory data analysis by examining the time plots of the adjusted durations Δt_i^*, as shown in Figure 3.2 and 3.3. In both figures, The upper panel shows the diurnal seasonal effect adjusted duration time plot. The lower panel shows the possible structural break within the sample. The horizontal dot line in the lower panel represents the overall mean level of the adjusted durations.

In Figure 3.2, the vertical dashed lines divide all observations into two parts: to the left-hand side are the durations observed during the first 50 trading days; to right are the observations of the last 11 trading days. The vertical dashed line is located at $n = 32965$. Likewise, the vertical dashed lines in Figure 3.3 divide all observations into two parts: observations from the first 210 trading days on the left and those from the last 31 trading days on the right. And the vertical dashed line is located at $n = 13031$.

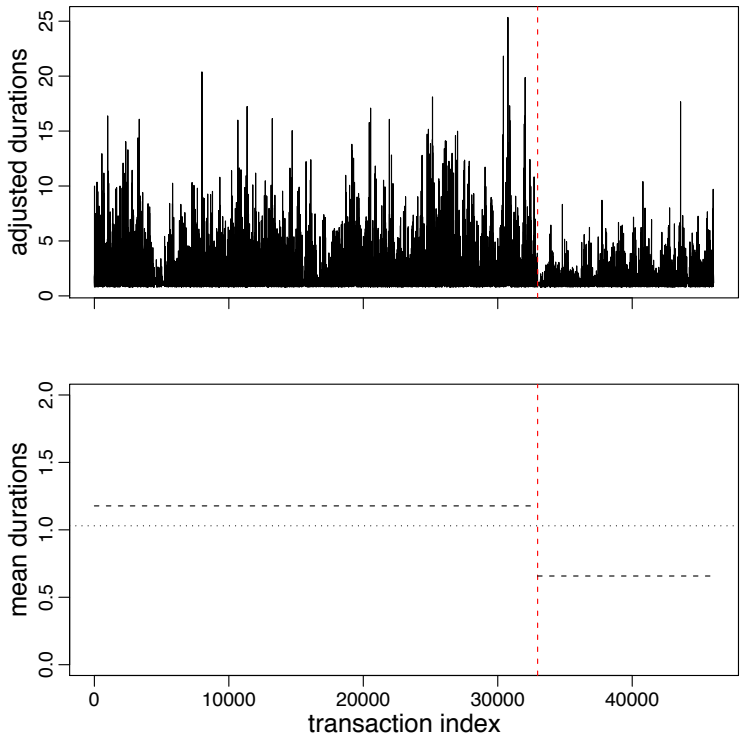

Figure 3.2: Time plot of standardized durations (IBM data) and structural break pattern

The lower panel in Figure 3.2 shows a clear structural break pattern (shifts in both mean levels and variances) with the IBM data, as noted by Zhang et al. (2001). Coincidently, similar structural break pattern is also identified with the Darby data as shown in Figure 3.3. More detailed numerical results for summarizing the shifts in

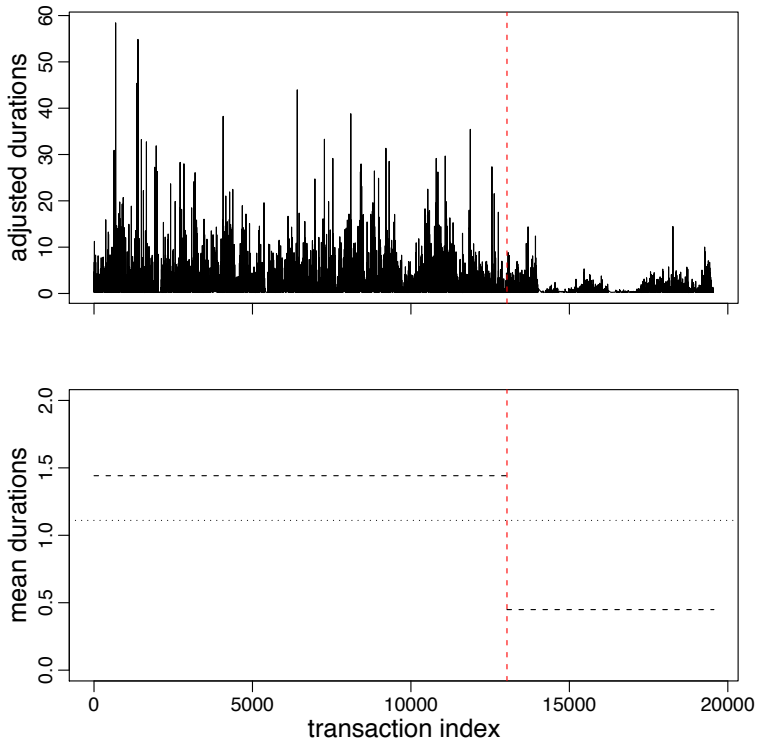

Figure 3.3: Time plot of standardized durations (Darby data) and structural break pattern

mean levels and variances are presented in Table 3.1. In order to avoid possible point process stationarity concern, we chose to use the first 50 trading days observations ($n = 32965$) from the IBM data and the first 210 trading days observations ($n = 13031$) from the Darby data as the sample data, respectively, for the subsequent analysis in this research.

The data preparation process starts all over again from the stock transaction times, $t_i, \quad i = 0, 1, 2, ..., n$. The adjusted duration time series are recalculated using Equation (3.1) based on the first 50 trading days observations ($n = 32965$) for the IBM data and on the first 210 days observations ($n = 13031$) for the Darby data. The key descriptive statistics of these two data sets are presented in Table 3.2.

Table 3.1: Numerical summary of shifts in mean levels and variances (diurnal seasonal effect adjusted inter-transaction durations)

	IBM data (61 trading days)	Darby data (241 trading days)
Full sample	(n= 1:46060)	(n = 1:19561)
Mean	1.030	1.110
Variance	2.010	6.038
First	50 trading days (n = 1:32965)	210 trading days (n = 1:13031)
Mean	1.178	1.442
Variance	2.460	8.345
Last	11 trading days (n=32966:46060)	31 trading days (n=13032:19561)
Mean	0.6577	0.4490
Variance	0.6857	0.7774

Table 3.2: Summary statistics for the standardized duration data

	IBM data (50 trading days)	Darby data (210 trading days)
sample size	32965	13031
Minimum	0.02444	0.01664
Median	0.5633	0.3158
Mean	1.030	1.136
Maximum	23.39	48.19
std. deviation	1.372	2.261
Ljung-Box	3902	553.1
Index of disp*	1.775	3.962

* Index of disp: Index of dispersion.

The IBM data is a mature stock long existed in NYSE whilst the Darby data is from an emerging Asian market. The IBM and Darby data have shown very different statistical properties. The most noticeable difference between these two stocks is the trading intensity, which for IBM is 32965 observations over a 50 trading-day period in the early 1990s, whilst for Darby it is 13031 observations over a 210 trading-day period in year 2006. The frequency distribution of the adjusted durations in both data is right-skewed and the Darby data have a much longer tail compared with the IBM data. There is a much higher autocorrelation among durations in IBM data than in the Darby series, indicated by a larger value of the Ljung-Box statistic. On the other hand, the Darby data have a higher standard deviation and a higher *index of dispersion*, indicating a higher level of clustering, than the IBM data.[1]

3.3 Fitted ACD models

Let $y_i \equiv \Delta t_i^*, \quad i = 1, 2, ..., n$, with the adjusted inter-transaction durations Δt_i^* is defined in Equation (3.1). Hence, $\{y_t\} = \{y_1, y_2, ..., y_n\}$ denotes the adjusted durations sample data for our analysis. Let $\psi_t = E(y_t|\mathcal{F}_{t-1})$ be the conditional expectation of the adjusted duration between the $(t-1)$st and tth transactions, where $\mathcal{F}_{t-1}$ is the information set available at the $(t-1)$st transaction. A general basic ACD(p,q) model is defined by

$$\begin{aligned} y_t &= \psi_t \, \epsilon_t \\ \psi_t &= \omega + \sum_{i=1}^{p} \alpha_i y_{t-i} + \sum_{i=1}^{q} \beta_i \psi_{t-i} \qquad t = 1, 2, ..., n, \end{aligned} \tag{3.2}$$

with $\omega > 0$, α_i, $\beta_i \geq 0 \; \forall \, i$, and ϵ_t is the multiplicative error term assumed to be iid with unit mean.

For a general ACD(p, q) model, denoting $t_0 = \max(p, q)$, the likelihood function of the durations $y_1, ..., y_n$ is given by:

$$f(\underline{y}|\underline{\theta}) = \left[\prod_{t=t_0+1}^{n} f(y_t|\mathcal{F}_{t-1}, \underline{\theta}) \right] \times f(\underline{y}_{t_0}|\underline{\theta}), \tag{3.3}$$

where $\underline{\theta}$ denotes the vector of model parameters, and n is the sample size. The marginal probability density function $f(\underline{y}_{t_0}|\underline{\theta})$ in Equation (3.3) is rather complicated for a general ACD model, and it is often ignored because its impact on the likelihood function

[1] The index of dispersion is defined by Equations (B.3) and (B.4).

diminishes as the sample size n increases (Tsay, 2005). Therefore, the following conditional likelihood function is used for parameter estimation:

$$L^*(\underline{y}|\underline{\theta}, y_{t_0}) = \prod_{t=t_0+1}^{n} f(y_t|\mathcal{F}_{t-1}, \underline{\theta}). \tag{3.4}$$

Because ACD model parameter estimation allows for misspecification and a unit-mean constraint is imposed on the error term, Engle and Russell (1998) stressed that the corresponding estimator fits within the *quasi-maximum likelihood* framework (see, e.g. White, 1982). More importantly, the conditional distribution model for an ACD model is defined by the specification of the error term ϵ_t, not y_t. Therefore, the likelihood function constructed for evaluating model fitting is different from Equation (3.4).

Based on Equation (3.2), an ACD(1,1) model can be expressed as

$$\begin{aligned} y_t &= \psi_t \, \epsilon_t \\ \psi_t &= \omega + \alpha_1 y_{t-1} + \beta_1 \psi_{t-1}, \end{aligned} \tag{3.5}$$

and an ACD(2,2) model can be expressed as:

$$\begin{aligned} y_t &= \psi_t \, \epsilon_t \\ \psi_t &= \omega + \alpha_1 y_{t-1} + \beta_1 \psi_{t-1} + \alpha_2 y_{t-2} + \beta_2 \psi_{t-2}. \end{aligned} \tag{3.6}$$

In published papers one seldom finds the use of higher order ACD models, e.g. an ACD(3,3) model, due to the principle of parsimony consideration. Our experience in analysing ACD models has also shown no real gain with employing higher order models in reducing the dependence and variance within the estimated residuals. Based on either econometric theory considerations (e.g. better understanding of the market structure or trading process) or statistical reasons (e.g. fitting a fat-tail distribution or nonlinear structure), various different ACD models were proposed by assuming different error term structures. In this section, six ACD models are specified for the purpose of model evaluation and comparison using the information-theoretic approach.

3.3.1 Basic ACD models

We first specify four basic ACD models. They are: EACD(1,1), EACD(2,2), WACD(1,1), and WACD(2,2). These basic ACD models were proposed by Engle and Russell (1998) and the traditional hypothesis tests were applied on these models for model evaluation. Obviously, the assessment of the goodness-of-fit of a fitted model is indirect

and not based on model's predictive ability. It is, therefore, necessary to apply the information-theoretic approach on these models to see how an ACD model can be evaluated differently.

Recall that from Section 2.4.3, an EACD(1,1) model is defined as an ACD(1,1) model with ϵ_t following a unit-mean exponential distribution. The conditional distribution model that describes an EACD(1,1) is

$$\begin{aligned} f(\epsilon_t|\mathcal{F}_{t-1},\underline{\theta}) &= \exp\Big\{-\frac{y_t}{\psi_t}\Big\} \\ &= \exp\Big\{-\frac{y_t}{\omega+\alpha_1\ y_{t-1}+\beta_1\psi_{t-1}}\Big\}, \end{aligned} \tag{3.7}$$

where $\underline{\theta} = (\omega, \alpha_1, \beta_1)^T$. Hence, based on Equation (3.5) and (3.7), the conditional likelihood function used for model fitting evaluation is given by

$$\begin{aligned} L(\underline{y}|\hat{\underline{\theta}}, y_1) &= \prod_{t=2}^{n} f(\epsilon_t|\mathcal{F}_{t-1},\hat{\underline{\theta}}) \\ &= \prod_{t=2}^{n} \exp\Big\{-\frac{y_t}{\hat{\omega}+\hat{\alpha}_1\ y_{t-1}+\hat{\beta}_1\psi_{t-1}}\Big\}. \end{aligned} \tag{3.8}$$

The QMLE $\hat{\underline{\theta}}$ is obtained based on the following conditional likelihood function

$$\begin{aligned} L^*(\underline{y}|\hat{\underline{\theta}}, y_1) &= \prod_{t=2}^{n} f(y_t|\mathcal{F}_{t-1},\hat{\underline{\theta}}) \\ &= \prod_{t=2}^{n} \Big\{\frac{1}{\hat{\omega}+\hat{\alpha}_1\ y_{t-1}+\hat{\beta}_1\psi_{t-1}} \exp\Big(-\frac{y_t}{\hat{\omega}+\hat{\alpha}_1\ y_{t-1}+\hat{\beta}_1\psi_{t-1}}\Big)\Big\}. \end{aligned} \tag{3.9}$$

In a similar way, for an EACD(2,2) model, based on Equation (3.6) and (3.7), the conditional likelihood function used for model fitting evaluation is given by

$$\begin{aligned} L(\underline{y}|\hat{\underline{\theta}}, y_1, y_2) &= \prod_{t=3}^{n} f(\epsilon_t|\mathcal{F}_{t-1},\hat{\underline{\theta}}) \\ &= \prod_{t=3}^{n} \exp\Big\{-\frac{y_t}{\psi_t}\Big\}, \end{aligned} \tag{3.10}$$

where $\psi_t = \hat{\omega} + \hat{\alpha}_1\ y_{t-1} + \hat{\beta}_1\psi_{t-1} + \hat{\alpha}_2\ y_{t-2} + \hat{\beta}_2\psi_{t-2}$.

And the conditional likelihood function for calculating QMLE $\hat{\underline{\theta}}$ is expressed as

$$\begin{aligned} L^*(\underline{y}|\hat{\underline{\theta}}, y_1, y_2) &= \prod_{t=3}^{n} f(y_t|\mathcal{F}_{t-1},\hat{\underline{\theta}}) \\ &= \prod_{t=3}^{n} \Big\{\frac{1}{\psi_t} \exp\Big(-\frac{y_t}{\psi_t}\Big)\Big\}, \end{aligned} \tag{3.11}$$

where $\psi_t = \hat{\omega} + \hat{\alpha}_1\ y_{t-1} + \hat{\beta}_1\psi_{t-1} + \hat{\alpha}_2\ y_{t-2} + \hat{\beta}_2\psi_{t-2}$ and $\underline{\hat{\theta}} = (\hat{\omega}, \hat{\alpha_1}, \hat{\beta}_1, \hat{\alpha_2}, \hat{\beta}_2)^T$.

The conditional distribution model describing a WACD model is

$$\begin{aligned} f(\epsilon_t|\mathcal{F}_{t-1}, \underline{\theta}) &= \frac{a\epsilon_t^{a-1}}{b^a}\exp\Big\{-(\frac{\epsilon_t}{b})^a\Big\} \\ &= \frac{a\psi_t}{y_t}\Big(\frac{y_t\,\Gamma(1+1/a)}{\psi_t}\Big)^a \exp\Big\{-\Big(\frac{y_t\,\Gamma(1+1/a)}{\psi_t}\Big)^a\Big\}, \end{aligned} \tag{3.12}$$

where a, b are the shape and scale parameters of a Weibull distribution, respectively, so that $a, b > 0$. The second line equality holds because that $\epsilon_t = y_t/\psi_t$ and $b\,\Gamma(1+1/a) = 1$ (unit mean constraint).

We have derived the corresponding likelihood functions for a WACD(1,1) model in Section 2.4.3 as follows. The conditional likelihood function used for model fitting evaluation is defined by

$$\begin{aligned} L(\underline{y}|\underline{\hat{\theta}}, y_1) &= \prod_{t=2}^{n} f(\epsilon_t|\mathcal{F}_{t-1}, \hat{\omega}, \hat{\alpha}_1, \hat{\beta}_1, \hat{a}) \\ &= \prod_{t=2}^{n}\Bigg\{\frac{\hat{a}(\hat{\omega} + \hat{\alpha}_1\ y_{t-1} + \hat{\beta}_1\psi_{t-1})}{y_t}\Big(\frac{y_t\,\Gamma(1+1/\hat{a})}{\hat{\omega} + \hat{\alpha}_1\ y_{t-1} + \hat{\beta}_1\psi_{t-1}}\Big)^{\hat{a}} \\ &\qquad \times \exp\Big\{-\Big(\frac{y_t\,\Gamma(1+1/\hat{a})}{\hat{\omega} + \hat{\alpha}_1\ y_{t-1} + \hat{\beta}_1\psi_{t-1}}\Big)^{\hat{a}}\Big\}\Bigg\}. \end{aligned} \tag{3.13}$$

The conditional likelihood function for obtaining the QMLE $\underline{\hat{\theta}} = (\hat{\omega}, \hat{\alpha}_1, \hat{\beta}_1, \hat{a})^T$, is defined by

$$\begin{aligned} L^*(\underline{y}|\underline{\hat{\theta}}, y_1) &= \prod_{t=2}^{n} f(y_t|\mathcal{F}_{t-1}, \hat{\omega}, \hat{\alpha}_1, \hat{\beta}_1, \hat{a}) \qquad (3.14) \\ &= \prod_{t=2}^{n}\Bigg\{\frac{\hat{a}}{y_t}\Big(\frac{y_t\,\Gamma(1+1/\hat{a})}{\hat{\omega} + \hat{\alpha}_1\ y_{t-1} + \hat{\beta}_1\psi_{t-1}}\Big)^{\hat{a}} \exp\Big\{-\Big(\frac{y_t\,\Gamma(1+1/\hat{a})}{\hat{\omega} + \hat{\alpha}_1\ y_{t-1} + \hat{\beta}_1\psi_{t-1}}\Big)^{\hat{a}}\Big\}\Bigg\}. \end{aligned}$$

Similarly, the corresponding likelihood functions for a WACD(2,2) model can be derived as follows: the conditional likelihood function used for model fitting evaluation is defined by

$$\begin{aligned} L(\underline{y}|\underline{\hat{\theta}}, y_1, y_2) &= \prod_{t=3}^{n} f(\epsilon_t|\mathcal{F}_{t-1}, \hat{\omega}, \hat{\alpha}_1, \hat{\beta}_1, \hat{\alpha}_2, \hat{\beta}_2, \hat{a}) \qquad (3.15) \\ &= \prod_{t=3}^{n}\Bigg\{\frac{\hat{a}\psi_t}{y_t}\Big(\frac{y_t\,\Gamma(1+1/\hat{a})}{\psi_t}\Big)^{\hat{a}} \exp\Big\{-\Big(\frac{y_t\,\Gamma(1+1/\hat{a})}{\psi_t}\Big)^{\hat{a}}\Big\}\Bigg\}, \end{aligned}$$

where $\psi_t = \hat{\omega} + \hat{\alpha}_1\ y_{t-1} + \hat{\beta}_1\psi_{t-1} + \hat{\alpha}_2\ y_{t-2} + \hat{\beta}_2\psi_{t-2}$; and the conditional likelihood

function for obtaining the QMLE $\hat{\underline{\theta}} = (\hat{\omega}, \hat{\alpha}_1, \hat{\beta}_1, \hat{\alpha}_2, \hat{\beta}_2, \hat{a})^T$, is defined by

$$\begin{aligned} L^*(\underline{y}|\hat{\underline{\theta}}, y_1, y_2) &= \prod_{t=3}^{n} f(y_t|\mathcal{F}_{t-1}, \hat{\omega}, \hat{\alpha}_1, \hat{\beta}_1, \hat{\alpha}_2, \hat{\beta}_2, \hat{a}) \\ &= \prod_{t=3}^{n} \Big\{ \frac{\hat{a}}{y_t} \Big(\frac{y_t\, \Gamma(1+1/\hat{a})}{\psi_t} \Big)^{\hat{a}} \exp\Big\{ -\Big(\frac{y_t\, \Gamma(1+1/\hat{a})}{\psi_t} \Big)^{\hat{a}} \Big\} \Big\}, \end{aligned} \tag{3.16}$$

where $\psi_t = \hat{\omega} + \hat{\alpha}_1\ y_{t-1} + \hat{\beta}_1 \psi_{t-1} + \hat{\alpha}_2\ y_{t-2} + \hat{\beta}_2 \psi_{t-2}$.

3.3.2 Two mixed-distribution ACD models

According to market micro-structure theories (O'Hara, 1995), there are informed and uninformed traders coexisting in a market and the interaction between the two types through information-revealing price formation processes is consistent with observed market behaviour. Luca and Gallo (2004) hence proposed the mixed Exponential ACD(1,1) model whose error term distribution is defined by

$$f(\epsilon_t|\mathcal{F}_{t-1}, \underline{\theta}) = p\, \frac{1}{\lambda_1} \exp\left(-\frac{\epsilon_t}{\lambda_1} \right) + (1-p) \frac{1}{\lambda_2} \exp\left(-\frac{\epsilon_t}{\lambda_2} \right), \tag{3.17}$$

where p is mixing probability of two exponential distributions such that $0 < p < 1$ and λ_1 and λ_2 are constrained by the unit mean condition $p\lambda_1 + (1-p)\lambda_2 = 1$.

Luca and Gallo (2004) argued that the weights of the mixture, p and $1-p$, can be interpreted as the probability of observing a transaction carried out by the informed and uninformed traders, respectively. The parameters λ_1 and λ_2 represent the instantaneous expected rate of transaction for informed and uninformed traders.

To satisfy the unit residual mean constraint condition, $p\lambda_1 + (1-p)\lambda_2 = 1$, we may let $\lambda_2 = (1 - p\lambda_1)/(1-p)$. A mixed exponential ACD(1,1) model can, therefore, be defined by the following likelihood functions. For defining the conditional distribution model, the likelihood function is defined by

$$\begin{aligned} L(\underline{y}|\hat{\underline{\theta}}, y_1) &= \prod_{t=2}^{n} f(\epsilon_t|\mathcal{F}_{t-1}, \hat{\omega}, \hat{\alpha}_1, \hat{\beta}_1, \hat{\lambda_1}, \hat{p}) \\ &= \prod_{t=2}^{n} \Big\{ \hat{p}\, \frac{1}{\hat{\lambda}_1} \exp\big(-\frac{y_t}{\psi_t \hat{\lambda}_1} \big) + \frac{(1-\hat{p})^2}{1-\hat{p}\hat{\lambda}_1} \exp\big(-\frac{y_t(1-\hat{p})}{\psi_t(1-\hat{p}\hat{\lambda}_1)} \big) \Big\}. \end{aligned} \tag{3.18}$$

And the likelihood function for parameter estimation is defined by

$$\begin{aligned} L^*(\underline{y}|\hat{\underline{\theta}}, y_1) &= \prod_{t=2}^{n} f(y_t|\mathcal{F}_{t-1}, \hat{\omega}, \hat{\alpha}_1, \hat{\beta}_1, \hat{\lambda_1}, \hat{p}) \\ &= \prod_{t=2}^{n} \Big\{ \frac{1}{\psi_t} \Big[\hat{p}\, \frac{1}{\hat{\lambda}_1} \exp\big(-\frac{y_t}{\psi_t \hat{\lambda}_1} \big) + \frac{(1-\hat{p})^2}{1-\hat{p}\hat{\lambda}_1} \exp\big(-\frac{y_t(1-\hat{p})}{\psi_t(1-\hat{p}\lambda_1)} \big) \Big] \Big\}, \end{aligned} \tag{3.19}$$

where $\psi_t = \hat{\omega} + \hat{\alpha}_1\ y_{t-1} + \hat{\beta}_1\psi_{t-1}$ and $\underline{\hat{\theta}} = (\hat{\omega}, \hat{\alpha_1}, \hat{\beta}_1, \hat{\lambda}_1, \hat{p})^T$.

As noted by Luca and Gallo (2004), the estimated residuals of an ACD model show a fat-tailed distribution. This is one major reason why EACD models did not fit the data well, an EACD model is bound to have a unit variance. A WACD model performs better because it allows variance (of the estimated residuals) to be more than one. As shown in Luca and Gallo (2004), a mixed exponential ACD model is more flexible to accommodate the variance. It is easy to show that further memory can be introduced into the error structure by using mixed lognormal and gamma densities. Furthermore, it is straightforward to check if this newly proposed mixed lognormal-gamma ACD model has a better performance based on the information-theoretic criteria. This is the motivation behind the proposal of a lognormal-gamma ACD(1,1) model.

The error term distribution of a lognormal-gamma ACD(1,1) model can be expressed as

$$f(\epsilon_t|\mathcal{F}_{t-1}, \underline{\theta}) = p\frac{1}{\sqrt{2\pi}\epsilon_t\sigma}\ \exp\Big[-\frac{(\log(\epsilon_t)-\mu)^2}{2\sigma^2}\Big] + (1-p)\frac{\epsilon_t^{a-1}}{\Gamma(a)b^a}\ \exp\Big(-\frac{\epsilon_t}{b}\Big) \quad (3.20)$$

where the mean is constrained to unity, i.e. $\mathrm{E}(\epsilon_t) = p\ \exp(\mu + \sigma^2/2) + (1-p)ab = 1$. Parameters μ and σ are the *log mean* and *log standard deviation* for the lognormal distribution and a, b are the *shape* and *scale* parameters for the gamma distribution, respectively.

To satisfy the unit residual mean constraint condition, $\mathrm{E}(\epsilon_t) = 1$, we may let $\mu = \log[1-(1-p)ab] - \log(p) - \sigma^2/2$. A mixed lognormal-gamma ACD(1,1) model can, therefore, be defined by the following likelihood functions. For defining the conditional distribution model, the likelihood function is expressed by

$$\begin{aligned} L(\underline{y}|\underline{\hat{\theta}}, y_1) &= \prod_{t=2}^{n} f(\epsilon_t|\mathcal{F}_{t-1}, \hat{\omega}, \hat{\alpha}_1, \hat{\beta}_1, \hat{\sigma}, \hat{a}, \hat{b}, \hat{p}) \quad (3.21)\\ &= \prod_{t=2}^{n}\Big\{\hat{p}\frac{\psi_t}{\sqrt{2\pi}y_t\hat{\sigma}}\ \exp\Big[-\frac{(\log(\frac{y_t}{\psi_t})-\hat{\mu})^2}{2\hat{\sigma}^2}\Big] + (1-\hat{p})\frac{\psi_t y_t^{\hat{a}-1}}{\Gamma(\hat{a})(\psi_t\hat{b})^{\hat{a}}}\ \exp\Big(-\frac{y_t}{\psi_t\hat{b}}\Big)\Big\}. \end{aligned}$$

And the likelihood function for parameter estimation can be derived as

$$\begin{aligned} L^*(\underline{y}|\underline{\hat{\theta}}, y_1) &= \prod_{t=2}^{n} f(y_t|\mathcal{F}_{t-1}, \hat{\omega}, \hat{\alpha}_1, \hat{\beta}_1, \hat{\sigma}, \hat{a}, \hat{b}, \hat{p}) \quad (3.22)\\ &= \prod_{t=2}^{n}\Big\{\hat{p}\frac{1}{\sqrt{2\pi}y_t\hat{\sigma}}\ \exp\Big[-\frac{(\log(\frac{y_t}{\psi_t})-\hat{\mu})^2}{2\hat{\sigma}^2}\Big] + (1-\hat{p})\frac{y_t^{\hat{a}-1}}{\Gamma(\hat{a})(\psi_t\hat{b})^{\hat{a}}}\ \exp\Big(-\frac{y_t}{\psi_t\hat{b}}\Big)\Big\}, \end{aligned}$$

where $\psi_t = \hat{\omega} + \hat{\alpha}_1\ y_{t-1} + \hat{\beta}_1\psi_{t-1}$ and $\underline{\hat{\theta}} = (\hat{\omega}, \hat{\alpha_1}, \hat{\beta}_1, \hat{\sigma}, \hat{a}, \hat{b}, \hat{p})^T$.

Similarly, AIC, AIC*, and GIC values can be calculated by adapting the Equations (3.25), (3.26), and (3.29), accordingly, for these two mixed distribution ACD(1,1) models. For simplicity, we will refer a mixed exponential ACD(1,1) model as mixedE-ACD(1,1) and a mixed lognormal-gamma ACD(1,1) model as mixedLG-ACD(1,1) throughout this thesis.

3.4 Model estimation

3.4.1 Some technical details

Recall that from Equation (3.2), a general basic ACD(p,q) model is defined by

$$\begin{aligned} y_t &= \psi_t \epsilon_t \\ \psi_t &= \omega + \sum_{i=1}^{p} \alpha_i y_{t-i} + \sum_{j=1}^{q} \beta_j \psi_{t-j} \qquad t = 1, 2, ..., n, \end{aligned}$$

with $\omega > 0,\ \alpha_i,\ \beta_j \geq 0\ \forall\, i, j$. Moreover, in order to ensure the stationarity and existence of the unconditional expected duration the model requires that $\sum_i \alpha_i + \sum_j \beta_j < 1$. These actually imply that $0 \leq \alpha_i,\ \beta_j < 1$.

Given a set of initial values for the estimated parameters $\underline{\hat{\theta}}_o$, initial values for ψ_1 and ψ_2 need to be determined to start the the estimation procedure. In the spirit of Engle and Russell (1998), the initial values for ψ_1 and ψ_2 are obtained in the following way. For the ACD(1,1) models $\psi_1 = \hat{\omega}_o/(1 - \hat{\beta}_{1o})$. For the ACD(2,2) models $\psi_1 = \hat{\omega}_o/(1 - \hat{\beta}_{1o} - \hat{\beta}_{2o})$ and $\psi_2 = \hat{\omega}_o + \hat{\alpha}_{1o} y_1 + \hat{\beta}_{1o}\psi_1$. We found that the choice of starting values had little effect on the final parameter estimates.

Model estimation is essentially an issue of optimization. In this study, a general-purpose optimization R function `optim`, which was introduced in Section 2.4.2, is used for parameter estimation. We have chosen to use the default method in `optim`, which is an implementation of that of Nelder and Mead (1965), so that the optimization process uses only function values and an analytic form of the first derivative of the objective function is not required.

In order to ensure the satisfaction of the model's stationarity and moment existence conditions, the following transformation is imposed on the searching range for possible parameter values in the optimization process.

$$\text{tsparam} = \log\left[\frac{\text{param} - lb}{ub - \text{param}}\right], \tag{3.23}$$

where 'param' stands for the original parameter value, 'tsparam' for the transformed parameter value, $lb =$ lower bound, and $ub =$ upper bound of the parameter value range. Hence,

$$\text{param} = lb + \frac{ub - lb}{1 + \exp(-\text{tsparam})}. \tag{3.24}$$

Therefore, while the transformed values can vary over the whole real number line (from $-\infty$ to $+\infty$), the original parameter values can only vary between a specified lower limit and an upper limit. For example, we may set $lb = 0$ and $ub = 1$ for α_i and β_i in the process of parameter estimation so that the condition $0 \leq \alpha_i, \beta_i < 1$ will always be satisfied.

Once an optimization process is initiated, the estimation results will be recorded when the process reaches its convergence state. If an optimization process reaches its iteration limit without convergence, the output results from the current run will be used as the input initial values for a second run of optimization. The optimization processes continue until the `convergence` signal is obtained.

Based on Equation (2.8), AIC can then be calculated by

$$\text{AIC}(k) = -2 \log L(\underline{y}|\hat{\underline{\theta}}, y_1) + 2k \tag{3.25}$$

for ACD(1,1) models, where $\hat{\underline{\theta}}$ is a $k \times 1$ M-estimate vector.

Accordingly, a *Pseudo-AIC* (AIC*) is defined by

$$\text{AIC}^*(k) = -2 \log L^*(\underline{y}|\hat{\underline{\theta}}, y_1) + 2\,k. \tag{3.26}$$

The QMLE, $\hat{\underline{\theta}}$, in Equation (3.25) and (3.26) is a solution of the conditional likelihood equation

$$\frac{\partial \log L^*(\underline{y}|\hat{\underline{\theta}}, y_1)}{\partial \hat{\underline{\theta}}} = \underline{0}, \tag{3.27}$$

for ACD(1,1) models. In the same way, AIC(k) and AIC*(k) can be calculated for ACD(2,2) models. Note that the likelihood functions $L(\underline{y}|\hat{\underline{\theta}}, y_1)$ and $L^*(\underline{y}|\hat{\underline{\theta}}, y_1)$ have been specified in Section 3.3 for various ACD models.

To obtain the result in Section 2.4.3, some special R functions have been written for solving Equation (2.32). The built-in R function `optim` is used as a minimizer function to find the numerical solution. The minimization objective function is the negative sample log-likelihood

$$-\log L^*(\underline{y}|\hat{\underline{\theta}}, y_1) \equiv -\sum_{t=2}^{n} \log f(y_t|\mathcal{F}_{t-1}, \hat{\underline{\theta}}). \tag{3.28}$$

Recall also that

$$\text{GIC}(k) \;=\; -2\,\log L(\underline{y}|\hat{\underline{\theta}}, y_1) + 2\,\text{EMD}, \tag{3.29}$$

where $\text{EMD} \;=\; \text{tr}\big\{R(\psi,\hat{G})^{-1}Q(\psi,\hat{G})\big\}$. The two $k\times k$ matrices, $R(\psi,\hat{G})$ and $Q(\psi,\hat{G})$ are defined by

$$R(\psi,\hat{G}) = -\frac{1}{n}\sum_{t=1}^{n}\frac{\partial\psi(y_t,\underline{\theta})}{\partial\underline{\theta}}\Big|_{\underline{\theta}\,=\,\hat{\underline{\theta}}},$$

and

$$Q(\psi,\hat{G}) = \frac{1}{n}\sum_{t=1}^{n}\psi(y_t,\underline{\theta})\,\frac{\partial\log f(\epsilon_t|\mathcal{F}_{t-1},\underline{\theta})}{\partial\underline{\theta}^T}\Big|_{\underline{\theta}\,=\,\hat{\underline{\theta}}},$$

where

$$\psi(y_t,\underline{\theta}) = \frac{\partial}{\partial\underline{\theta}}\;\log f(y_t|\mathcal{F}_{t-1},\underline{\theta}).$$

But as we pointed out in the discussion on the model evaluation results given in Table 2.6 and Figure 2.1, the estimated residuals using the above fitting procedure still contain significant autocorrelation. The concern about the violation of the independence condition with the estimated residual series makes us believe that the parameter estimation results could be unreliable due to the incorrect specification of the log-likelihood function $\log L^*(\underline{y}|\hat{\underline{\theta}}, y_1)$. Therefore, we decided to use a constrained log-likelihood function as the minimization objective function to ensure the remaining autocorrelation is minimized and the estimated residual mean is unity. In this way, the model estimation is still under an M-estimation framework.

The L-B statistic is used as a measure of autocorrelation in the estimated residuals. A penalty term in terms of estimated residual L-B statistic and the absolute value of the difference between the residual mean value and the unity is added to the negative log-likelihood function. The new minimization objective function is defined as (assuming an ACD(1,1) model)

$$-\log L^*(\underline{y}|\hat{\underline{\theta}}, y_1) + C(\underline{y}|\hat{\omega},\hat{\alpha}_1,\hat{\beta}_1), \tag{3.30}$$

where $C(\underline{y}|\hat{\omega},\hat{\alpha}_1,\hat{\beta}_1)$ is the penalty term which is defined by

$$C(\underline{y}|\hat{\omega},\hat{\alpha}_1,\hat{\beta}_1) \;=\; c_1[\text{L-B}(\hat{\underline{\epsilon}})] + c_2[\text{abs}(\text{mean}(\hat{\underline{\epsilon}}) - 1)], \tag{3.31}$$

where L-B statistic value is calculated using built-in R function `Box.test` and c_1, c_2 are two arbitrarily determined positive numbers.[2] The notation 'abs' stands for

[2]In the self-written R functions, the L-B statistic is calculated with the following R commands. `nj = round(log(n) + 0.5); residLB = Box.test (residX, lag = nj, type="Ljung")$statistic` where residLB $\equiv$ L-B($\hat{\underline{\epsilon}}$) and residX $\equiv \hat{\underline{\epsilon}}$. The two positive constants are set to be $c_1 = n/500$ and $c_2 = 10^5$, where n is the sample size.

'taking absolute value' operation and 'mean' for 'calculating sample mean' operation. The notation $\hat{\underline{\epsilon}} \equiv \{\hat{\epsilon}_t\}$ is used to represent the estimated residual series. Since $\hat{\epsilon}_t = y_t/\hat{\psi}_t$, the penalty term in the new minimization function is a function of parameter estimates $\hat{\omega}, \hat{\alpha}_1$, and $\hat{\beta}_1$ which are always greater than zero. Therefore, the parameter estimation for an ACD model is now implemented by minimizing the negative sample log-likelihood function $-\log L^*(\underline{y}|\hat{\underline{\theta}}, y_1)$ subject to a small residual L-B score and a unit mean constraint.

With the addition of the penalty term to the minimization objective function, a problem is created in calculating GIC scores. If we take a close look at the L-B statistic formula (Equation A.20), we realize that parameters are connected in a complicated way. We now have trouble calculating the EMD term in Equation (3.29) because it is very hard to express the $\psi(y_t, \underline{\theta})$ function analytically, and hence evaluating matrices $R(\psi, \hat{G})$ and $Q(\psi, \hat{G})$. Therefore, we turn to AIC for model evaluation for ACD models.

Equation (3.25) is valid for evaluating ACD models as we have justified the AIC usage under an M-estimator framework in Section 2.5.1. The model estimation results presented in the following section are obtained from optimization processes based on the minimization objective function (3.30). The calculated AIC scores for all fitted ACD models are reported in Section 3.5.

3.4.2 Model estimation results

A number of R functions have been written to perform the parameter estimation and calculation of AIC and TIC scores for fitted ACD models. Based on Equation (A.12) and the discussions in Section 2.4.3, ACD model parameter estimates are typical M-estimates (maximum likelihood-type estimates).

Tables 3.3 and 3.4 contain the parameter estimates for basic ACD models. In both tables, parameters a and b are shape and scale parameters respectively for a Weibull distribution and b is not a free parameter due to the unit mean constraint. The number of the free parameters is $k = 3$, $k = 5$, $k = 4$, and $k = 6$ for EACD(1,1), EACD(2,2), WACD(1,1), and WACD(2,2), respectively, due to the unit residual mean constraint. Correspondingly, M-estimates are presented for models mixedE-ACD(1,1) and mixedLG-ACD(1,1) in Table 3.5. In the table, parameter $\lambda_i \quad (i = 1, 2)$ is the mean for defining an exponential distribution and λ_2 is not a free parameter due to the unit mean constraint; parameters μ and σ are log mean and log standard deviation of a lognormal distribution, respectively; parameters a and b are shape and scale parameters

respectively for a gamma distribution; parameter μ is not a free parameter due to the unit mean constraint and p is the mixing probability parameter. The model dimensions are $k = 5$ and $k = 7$ for mixedE-ACD(1,1) and mixedLG-ACD(1,1), respectively. While the parameter estimates are obtained as the optimization processes reached convergence state, the justification of the stationarity and moment existence conditions in terms of the estimation results are given in Table 3.6.

Table 3.3: Parameter estimates for the fitted basic ACD models (IBM data)

Parameter (M-estimate)	EACD(1,1)	EACD(2,2)	WACD(1,1)	WACD(2,2)
$\hat{\omega}$	0.04666	0.05852	0.04832	0.05882
$\hat{\alpha_1}$	0.08106	0.09388	0.08205	0.09232
$\hat{\beta}_1$	0.8731	0.5150	0.8704	0.5348
$\hat{\alpha_2}$	—	0.01215	—	0.01065
$\hat{\beta}_2$	—	0.3215	—	0.3043
$\hat{a}$	—	—	0.8770	0.8784
$(\hat{b})$	—	—	0.9366	0.9375

Table 3.4: Parameter estimates for the fitted basic ACD models (Darby data)

Parameter (M-estimate)	EACD(1,1)	EACD(2,2)	WACD(1,1)	WACD(2,2)
$\hat{\omega}$	0.1207	0.2227	0.1234	0.2183
$\hat{\alpha_1}$	0.1217	0.09832	0.1207	0.09526
$\hat{\beta}_1$	0.7761	0.2197	0.7742	0.2057
$\hat{\alpha_2}$	—	0.1039	—	0.1041
$\hat{\beta}_2$	—	0.3876	—	0.4080
$\hat{a}$	—	—	0.6450	0.6454
$(\hat{b})$	—	—	0.7256	0.7261

The results presented in Table 3.6 are calculated based on parameter estimates given in Tables 3.3, 3.4, and 3.5. Expression $\hat{\omega}/(1 - \sum \hat{\alpha}_i - \sum \hat{\beta}_j)$ calculates the estimated unconditional expected durations. Compared with the the stationarity and moment existence conditions specified in the beginning of Section 3.4.1, it is clear all estimated parameters are valid. Notice that all ACD(1,1) models have very similar $\hat{\omega}, \hat{\alpha}_i, \hat{\beta}_j$ values and significantly different from those of ACD(2,2) models, with respect to each data set. We also observe that IBM data consistently shows a higher $\sum \hat{\alpha}_i + \sum \hat{\beta}_j$ value than Darby data (IBM$>$ 0.94 versus Darby $<$ 0.90). This indicates that the

Table 3.5: Parameter estimates of the mixed error distribution ACD models

	IBM	data	Darby	data
Parameter (M-estimate)	mixedE-ACD(1,1)	mixedLG-ACD(1,1)	mixedE-ACD(1,1)	mixedLG-ACD(1,1)
$\hat{\omega}$	0.04994	0.04712	0.1217	0.1372
$\hat{\alpha_1}$	0.08308	0.08202	0.1223	0.1263
$\hat{\beta_1}$	0.8678	0.8717	0.7746	0.7566
$\hat{\lambda_1}$	1.343	—	2.031	—
$(\hat{\lambda_2})$	0.3762	—	0.1944	—
$(\hat{\mu})$	—	-0.1241	—	-0.7116
$\hat{\sigma}$	—	0.8886	—	1.365
$\hat{a}$	—	2.402	—	3.678
$\hat{b}$	—	0.05155	—	0.01448
$\hat{p}$	0.645	0.738	0.439	0.794

Table 3.6: Examination of stationarity condition and estimated unconditional expected durations

	IBM	data	Darby	data
Models	$\sum \hat{\alpha}_i + \sum \hat{\beta}_j$	$\frac{\hat{\omega}}{(1-\sum \hat{\alpha}_i - \sum \hat{\beta}_j)}$	$\sum \hat{\alpha}_i + \sum \hat{\beta}_j$	$\frac{\hat{\omega}}{(1-\sum \hat{\alpha}_i - \sum \hat{\beta}_j)}$
EACD(1,1)	0.95416	1.018	0.8978	1.181
EACD(2,2)	0.94253	1.018	0.80952	1.169
WACD(1,1)	0.95245	1.016	0.8949	1.174
WACD(2,2)	0.94207	1.015	0.8131	1.168
mixedE-ACD(1,1)	0.95088	1.017	0.8969	1.180
mixedLG-ACD(1,1)	0.95372	1.018	0.8829	1.172

autoregressive structure in IBM data has a higher persistence property (long memory process). This model estimation result matches the exploratory data analysis results given in Table 3.2 which shows that IBM have much higher L-B statistic (L-B(IBM) = 3902 versus L-B(Darby) = 553.1). Furthermore, the difference in the estimated unconditional expected mean durations in Table 3.6 reflects the difference in the observed sample mean level in Table 3.2.

In the model estimation process, we noticed that the estimation of the autoregressive structure parameters (i.e. ω, α_i, and β_i) seems to be independent of the estimation of error term distribution parameters and is robust with respect to different data sets. The estimation of ω, α_i, and β_i is dominantly determined by the order of an ACD model and an ACD(2,2) model seems to give the optimal estimation in terms of L-B statistic. It is, therefore, natural to propose a two-step strategy for ACD model estimation to

achieve an optimal result, i.e. an ACD(2,2) autoregressive structure combined with different error term distributions. Table 3.7 gives the model estimation results of this two-step model estimation strategy. In the table, the autoregressive structure parameters, ω, α_i, and β_i, are estimated separately from those error term distribution parameters. Notation 'mixedE($\hat{\epsilon}_t$)' denotes a mixed exponential distribution model which fits the estimated residual series $\{\hat{\epsilon}_t\}$ and 'mixedLG' denotes a mixed lognormal-gamma distribution model. The estimated residual series $\{\hat{\epsilon}_t\}$ is generated based on the step one model estimation results given in Table 3.7. It is trivial to justify the validity of the parameter estimates reported in Table 3.7 in terms of ACD process stationarity and moment existence conditions.

Table 3.7: Parameter estimates of a two-step model estimation strategy based on an ACD(2,2) process

Parameter (M-estimate)	IBM data	Darby data	
$\hat{\omega}$	0.05357	0.2207	(step one estimation)
$\hat{\alpha_1}$	0.09027	0.08972	ACD(2,2) model
$\hat{\beta}_1$	0.5093	0.2234	
$\hat{\alpha_2}$	0.01416	0.1032	
$\hat{\beta}_2$	0.3337	0.3936	
$\hat{\lambda}_1$	1.391	2.059	(step two estimation)
$(\hat{\lambda}_2)$	0.4251	0.1986	mixedE($\hat{\epsilon}_t$)
$\hat{p}$	0.595	0.431	
$(\hat{\mu})$	-0.1117	-0.6921	(step two estimation)
$\hat{\sigma}$	0.8818	1.356	mixedLG($\hat{\epsilon}_t$)
$\hat{a}$	2.337	3.509	
$\hat{b}$	0.05421	0.01520	
$\hat{p}$	0.732	0.788	

There are a few advantages of this two-step model estimation strategy. It makes the fitted ACD models have the same estimated residual L-B statistic so that model evaluation results are more reliable. It makes the computation of GIC unnecessary because the model estimation problem reduces to a standard MLE procedure. The step one model estimation has a minimization objective function as

$$\big(\text{L-B}(\underline{\hat{\epsilon}}) - 14\big)^2 + 10^5 \big(\text{abs}(\text{mean}(\underline{\hat{\epsilon}}) - 1)\big), \tag{3.32}$$

where notation 'abs' and 'mean' has the same meaning as in Equation (3.31).[3] The

[3]The reason for minimizing L-B($\underline{\hat{\epsilon}}$) toward 14 rather than toward zero may be justified from Figure 2.1 and Figure 3.4. The L-B scores generated from an iid sample seems to center around 14.

second step estimation is simply just to estimate the error term distribution parameters using ordinary MLE procedure based on the estimated residual series produced from the first step estimation results. Therefore, we only need to calculate TIC for estimating the impact of the effective model dimension (EMD). Of course, the gain in this two-step model estimation strategy takes some cost, the increase in number of parameters in a model compared with an ACD(1,1) autoregressive structure.

3.5 Model evaluation and discussions

3.5.1 Model comparison among basic ACD models – a successful story

The model evaluation results presented in Tables 3.8 and 3.9 are calculated based on the model estimation results given in Table 3.3 and 3.4.

Before we discuss the model evaluation results given in Tables 3.8 and 3.9, we would like to make some remarks on the EACD(1,1) and WACD(1,1) model evaluation results compared with the results in Table 2.6. We notice that both L-B($\hat{\epsilon}_t$) (lag=15) and L-B($\hat{\epsilon}_t^2$) (lag=15) values reported in Table 3.8 have reduced substantially. If we further check these values against the empirical distribution of random sample L-B scores shown in Figure 2.1, we may consider that the remaining dependence structure in the estimated residuals have become insignificant. This improvement is due to the imposition of the penalty term $C(\underline{y}|\hat{\omega}, \hat{\alpha}_i, \hat{\beta}_j)$ in the minimization objective function (3.31). Correspondingly, we observe that the parameter estimates in Table 3.3 differ significantly from those obtained in Table 2.5.

Now let us turn to Tables 3.8 and 3.9. In both tables, $\text{AIC}(k) = -2 \log L(y_t|\underline{\hat{\theta}}) + 2k$ and $\text{AIC}^*(k) = -2(\log L^*(y_t|\hat{\theta}) + C(\underline{y}|\hat{\omega}, \hat{\alpha}_i, \hat{\beta}_j)) + 2k$, where $L(y_t|\underline{\hat{\theta}})$ and $L^*(y_t|\hat{\theta})$ are defined in Section 3.3, and $C(\underline{y}|\hat{\omega}, \hat{\alpha}_i, \hat{\beta}_j)$ is defined in Section 3.4.1. Symbol Q stands for Pearson's Q-statistic. For both IBM and Darby data, the AIC selected best model is WACD(1,1) (the bold line model) which is supported by the Pearson's Q-statistic assessment results. If we further examine the AIC difference values (ΔAIC_i), we notice that this is a single best model case (i.e. free of model averaging concern) because even the smallest ΔAIC_i values are much greater than 10.

If we compare these basic ACD models in terms of ACD(1,1) process against ACD(2,2) process, the model evaluation results show that the gain in better fitting the sample data is outweighed by the cost of increase in model complexity based on

Table 3.8: Model comparison among the fitted basic ACD models (IBM data)

Model	k	AIC*	ΔAIC_i^*	AIC	ΔAIC_i	$sd(\hat{\epsilon}_t)$	L-B($\hat{\epsilon}_t$) (lag=15)	L-B($\hat{\epsilon}_t^2$) (lag=15)	Q
EACD(1,1)	3	67649	2153	65925	1003	1.237	27.8	16.3	2228.4
EACD(2,2)	5	66489	993	65936	1014	1.235	19.5	14.5	2237.3
WACD(1,1)	**4**	**66652**	**1156**	**64922**	**0**	**1.239**	**28.2**	**16.5**	**1070.2**
WACD(2,2)	6	65496	0	64943	21	1.235	19.8	13.4	1075.0

Table 3.9: Model comparison among the fitted basic ACD models (Darby data)

Model	k	AIC*	ΔAIC_i^*	AIC	ΔAIC_i	$sd(\hat{\epsilon}_t)$	L-B($\hat{\epsilon}_t$) (lag=15)	L-B($\hat{\epsilon}_t^2$) (lag=15)	Q
EACD(1,1)	3	28471	6131	26062	5759	1.993	26.0	8.48	9759.3
EACD(2,2)	5	28093	5753	26068	5765	1.989	20.2	8.79	9783.2
WACD(1,1)	**4**	**22712**	**372**	**20303**	**0**	**1.993**	**26.0**	**8.58**	**1753.4**
WACD(2,2)	6	22340	0	20322	19	1.988	19.7	8.71	1760.4

AIC, e.g. EACD(1,1) has outperformed EACD(2,2) and WACD(1,1) has outperformed WACD(2,2). Notice that we will reach an exact opposite conclusions if using AIC* (pseudo-AIC) as the model evaluation criterion in this case. It is, therefore, important to realize that it is the specification of the residual distribution $f(\epsilon_t|\underline{\theta})$ defines the conditional distribution model for an ACD(p,q) process. Therefore, AIC should be calculated based on $\log L(y_t|\hat{\underline{\theta}})$ rather than $\log L^*(y_t|\hat{\theta}) + C(\underline{y}|\hat{\omega}, \hat{\alpha}_i, \hat{\beta}_j)$, which is used for parameter estimation, for an ACD model.

The estimation results for the standard deviation of the estimated residuals in Tables 3.8 and 3.9 seems very robust across different models. The estimated standard deviations are reduced to some extent compared with the observed standard deviations shown in Table 3.2, for both data sets. The Darby data have shown a higher estimated standard deviation than that of IBM data as it was with the observed standard deviations. The estimated residual means are all equal to unity (up to the third decimal places) due to the imposition of penalty term $C(\underline{y}|\hat{\omega}, \hat{\alpha}_i, \hat{\beta}_j)$. The only remaining minor concern is the small differences in the estimated L-B statistics between ACD(1,1) models and ACD(2,2) models. Therefore, only ACD(1,1) models are compared with each other in the following subsection.

Note that all L-B($\hat{\epsilon}_t^2$) values seem to be well located within the range of the empirical

random sample L-B statistic distributions (referring to the histograms of the bottom panels in Figure 2.1) although we did not model the squared residuals directly in the model estimation processes.

3.5.2 The mixed lognormal-gamma ACD model – a better fit model

Luca and Gallo (2004) reported that an mixedE-ACD(1,1) model well outperformed an EACD(1,1) and a WACD(1,1) model in their study. The model evaluation results given in Tables 3.10 and 3.11 provide two more cases to support their finding.

A careful reader may have noticed that the AIC scores for EACD(1,1) and WACD(1,1) models reported in Tables 3.10 and 3.11 are slightly larger than those reported in Tables 3.8 and 3.9. This is because when comparing an ACD(1,1) model with an ACD(2,2) model the calculation of AIC scores should only include $n-2$ observations with all models involved so that we evaluate models based on the same set of observations.

The model evaluation results presented in Tables 3.10 and 3.11 are calculated based on the model estimation results given in Tables 3.3, 3.4, and 3.5. This is a clear-cut decision-making situation that the newly proposed model mixedLG-ACD(1,1) is the best model for both IBM and Darby data sets. Model mixedLG-ACD(1,1) performs far better than other ACD(1,1) models examined based on both AIC scores and Q-statistics. However, we notice that the L-B($\hat{\epsilon}_t$) results are close to the upper end of the empirical random sample L-B statistic distribution range (referring to the histograms of the top panels in Figure 3.4). As the results reported in Tables 3.8 and 3.9 have shown, an ACD(2,2) model has a better performance than an ACD(1,1) model in terms of L-B statistics. That is why we proposed a two-step fitting procedure to achieve an optimal result by combining an ACD(2,2) process with a mixed distribution error term.

3.5.3 A two-stage ACD model fitting procedure

In the two-stage fitting procedure, the residual series are generated based on the estimation results obtained in the first step model estimation as reported in Table 3.7. Therefore, the second step model fitting becomes simply a probability distribution model problem within the MLE framework. Both AIC and TIC are reported in Tables 3.12 and 3.13 to examine the robustness of model evaluation results using AIC. In both tables, $\mathrm{AIC}(k) = -2\log L(y_t|\hat{\theta}) + 2k$ and $\mathrm{TIC}(k) = -2\log L(y_t|\hat{\theta}) + 2\,\mathrm{EMD}$, where $\mathrm{EMD} = \mathrm{tr}\big(K(\underline{\hat{\theta}})J(\underline{\hat{\theta}})^{-1}\big)$ as defined in Equation (2.6) and (2.11). The reported

Table 3.10: Model comparison between the basic ACD(1,1) models and mixed distribution ACD(1,1) models (IBM data)

Model	k	AIC*	ΔAIC_i^*	AIC	ΔAIC_i	$sd(\hat{\epsilon}_t)$	L-B($\hat{\epsilon}_t$) (lag=15)	L-B($\hat{\epsilon}_t^2$) (lag=15)	Q
EACD(1,1)	3	67649	3064	65934	3068	1.237	28.3	17.0	2228.4
WACD(1,1)	4	66652	2067	64930	2064	1.239	28.7	17.2	1070.2
mixedE-ACD(1,1)	5	66086	1501	64350	1484	1.239	29.2	17.4	698.7
mixedLG-ACD(1,1)	**7**	**64585**	**0**	**62866**	**0**	**1.238**	**28.3**	**17.5**	**112.5**

Table 3.11: Model comparison between the basic ACD(1,1) models and mixed distribution ACD(1,1) models (Darby data)

Model	k	AIC*	ΔAIC_i^*	AIC	ΔAIC_i	$sd(\hat{\epsilon}_t)$	L-B($\hat{\epsilon}_t$) (lag=15)	L-B($\hat{\epsilon}_t^2$) (lag=15)	Q
EACD(1,1)	3	28471	8411	26066	8432	1.993	26.0	8.48	9759.3
WACD(1,1)	4	22712	2652	20307	2673	1.993	26.0	8.58	1753.4
mixedE-ACD(1,1)	5	21629	1569	19220	1586	1.993	26.0	8.48	881.1
mixedLG-ACD(1,1)	**7**	**20060**	**0**	**17634**	**0**	**1.993**	**27.0**	**8.70**	**154.1**

AIC and TIC values show that the difference between ΔAIC and ΔTIC is insignificant ($\ll 2$): $\Delta AIC = 1464.1$ and $\Delta TIC = 1464.3$ for IMB data; $\Delta AIC = 1567.7$ and $\Delta TIC = 1568.7$ for Darby data. But note that the impact of EMD for Darby data is bigger than that of IBM data although both are negligible compared with the magnitude of the sample log-likelihood term. This result confirms our conclusion about the valid application of AIC drawn in Section 2.5.1.

Once again, the conclusion is straightforward. The combination of an ACD(2,2) process with a mixed lognormal-gamma distribution for error structure is the best model to fit the observed inter-transaction duration data. The model evaluation results are consistent between AIC and Pearson's Q-statistics. The remaining autocorrelation for both fitted models is examined visually by Figure 3.4. The top-left panel shows the histogram of L-B statistics (lag=15) calculated from 1000 random samples (each

of size $n = 32965$) of a unit mean mixed exponential distribution. The two dashed vertical lines indicate the estimated residual L-B statistics (lag=15) L-B($\hat{\epsilon}_t$) which are given in tables 3.12 and 3.13 for IBM and Darby data, respectively. The bottom-left panel shows L-B statistics (lag=15) histogram calculated from the squared sample data values and the dashed vertical lines are the corresponding estimated squared residual L-B statistics L-B($\hat{\epsilon}_t^2$). The right column panels display the same contents as in the left column panels but the L-B statistics histograms are based on a unit mean mixed lognormal-gamma distribution. The dashed lines are the same estimated residual L-B statistics as in the left column panels. Figure 3.4 confirms us that we may consider that the remaining dependence structure in the estimated residuals behaves like from a random sample. In Figures 3.5 and 3.6, the parametric bootstrap version qq-plots are drawn and the results also match the AIC model evaluation results very well.

Table 3.12: Evaluation of two mixed distribution models – a two-stage model fitting procedure (IBM data)

Model	k	AIC	TIC	EMD	sd($\hat{\epsilon}_t$)	L-B($\hat{\epsilon}_t$) (lag=15)	L-B($\hat{\epsilon}_t^2$) (lag=15)	Q
mixedE ($\hat{\epsilon}_t$)	2	64395.5	64395.8	2.15	1.235	20.3	16.9	720.6
mixedLG ($\hat{\epsilon}_t$)	4	62931.4	62931.5	4.09	1.235	20.3	16.9	91.7

Table 3.13: Evaluation of two mixed distribution models – a two-stage model fitting procedure (Darby data)

Model	k	AIC	TIC	EMD	sd($\hat{\epsilon}_t$)	L-B($\hat{\epsilon}_t$) (lag=15)	L-B($\hat{\epsilon}_t^2$) (lag=15)	Q
mixedE ($\hat{\epsilon}_t$)	2	19220.0	19221.9	2.94	1.984	19.7	8.82	911.0
mixedLG ($\hat{\epsilon}_t$)	4	17652.3	17653.2	4.45	1.984	19.7	8.82	142.5

Unlike an ordinary qq-plot, a bootstrap version qq-plot takes account of the sampling variation in assessing the model performance. In both Figures 3.5 and 3.6, 200 samples of the estimated residuals are generated from each of the two models and then each generated sample is sorted by ascending order. With respect to each of the observed residual quantile values, the median, 2.5% and 97.5% quantiles are calculated based on the 200 fitted values. The bold solid lines are the median quantile-quantile lines of the observed residual quantiles versus fitted quantiles. The two dashed lines

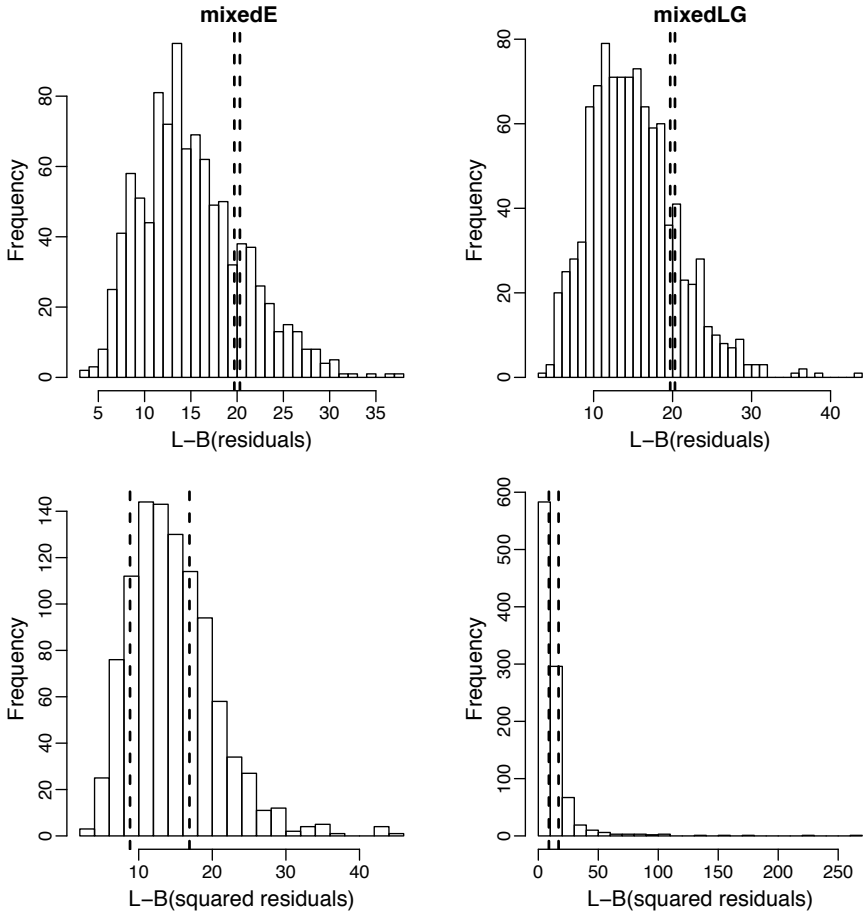

Figure 3.4: L-B statistics of estimated residual series versus the empirical distribution of L-B statistics of random samples

form the 95% bootstrap confidence quantile band. The diagonal solid lines are the perfect fit qq-lines. In order to examine the model fitting performance in more detail, the same qq-plot is expressed at three levels arranged by three panels in one column for each model as shown in Figures 3.5 and 3.6. The top panel shows a full range qq-plot with the $0-99\%$ quantile range part identified by two dotted lines at the bottom-left corner. In the middle panel, the identified $0-99\%$ quantile range part is enlarged to a full panel area and similarly, the $0-50\%$ quantile range part is identified by two dotted lines at the bottom-left corner. The $0-50\%$ quantile range part is then shown to a full panel area in the bottom panel. By examining these bootstrap qq-plots in details, we can easily identify which model fits which part of a data set well. By comparing Figure 3.5 with Figure 3.6, we immediately conclude that both models have

fitted IBM data better than fitted the Darby data which agree with the Q-statistics results. We also find that neither model has fitted the top 1% quantile part of the data satisfactorily. Most importantly, we have shown how Pearson's Q-statistics and bootstrap version of qq-plots consistently match AIC model evaluation results.

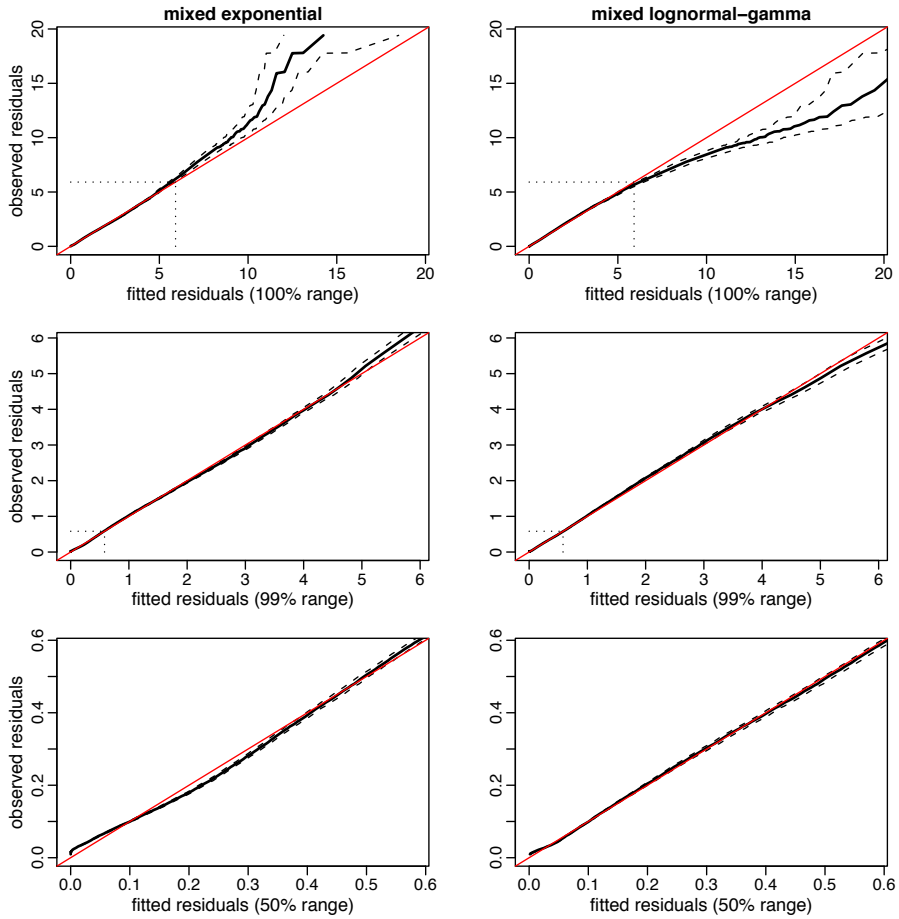

Figure 3.5: Parametric bootstrap version quantile-quantile plots (IBM data)

3.6 Summary and conclusion

ACD models provide an essential step in modelling market events, which can be modelled most realistically as an irregularly spaced time series. The same model fitting procedure for inter-transaction durations studied in this chapter can easily be extended to the analysis of time elapsed between any other types of financial events, e.g. the time between transactions corresponding to movements in share prices above or below a certain threshold, or to the accumulation of traded volume above a certain threshold.

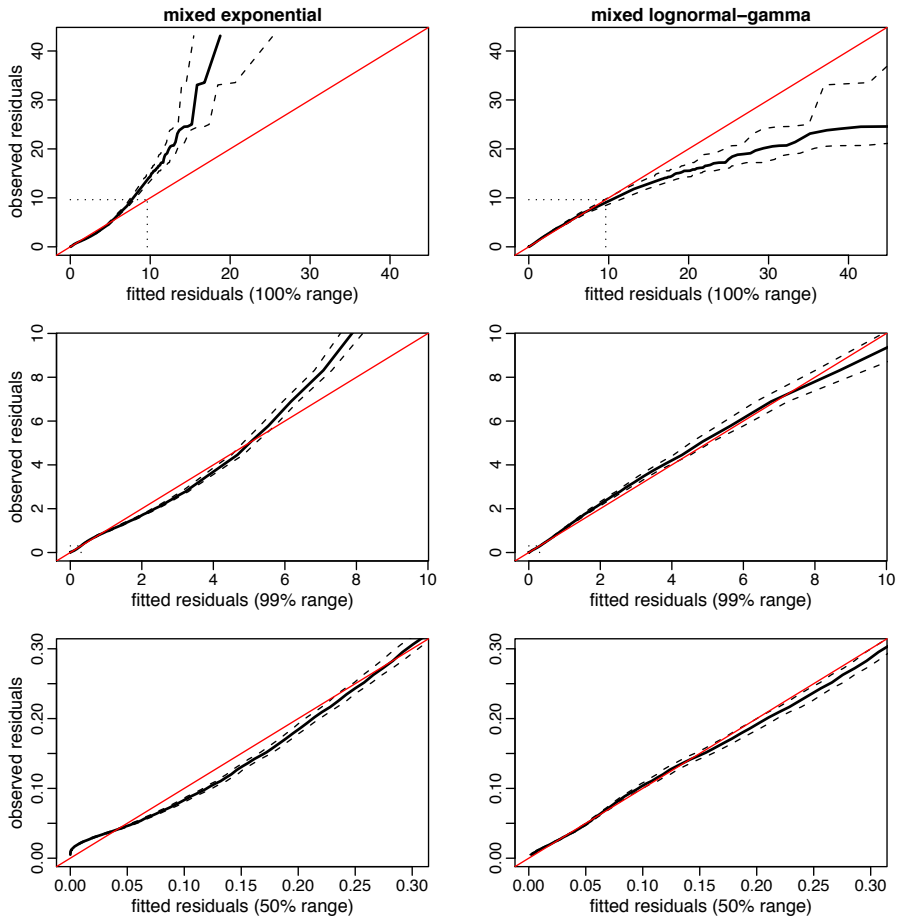

Figure 3.6: Parametric bootstrap version quantile-quantile plots (Darby data)

The extensive study of ACD models since its formal introduction in 1998 has resulted in many competing ACD model specifications. The necessity for model comparison between different ACD models is obvious. However, the traditional hypothesis testing approach can only provide limited and *ad hoc* solutions to the issue of evaluation of ACD models. In this chapter, we have shown how the information-theoretic approach can be applied to evaluate ACD models in an objective and efficient way. Not one p-value has been calculated because of its irrelevance to an information-theoretic approach.

In this chapter, four basic ACD models and two mixed distribution ACD models are specified to fit two real stock transaction data sets, the IBM data from a mature financial market NYSE and the Darby data from an emerging Asian market KLSE. We have applied AIC and TIC to evaluate and compare these specified ACD models. The

model evaluation results using Q-statistic and bootstrap qq-plot match nicely with the AIC and TIC assessment results. A two-step model estimation strategy is proposed to facilitate the model estimation procedure and may have provided a solution for selecting an ACD model with an optimal combination of the autoregressive structure and the probabilistic error term structure.

A primary concern at the model estimation stage with ACD models is the remaining dependence structure within the estimated residuals. This is the major reason for the proposal of a L-B statistic constrained QMLE minimization objective function. In this way, the ACD model parameter estimation is still within a M-estimation framework. Hence, the information-theoretic criteria are applicable for model evaluation.

From the point of view of shedding new light on the trading process or on the market micro-structure, the proposal of a new model, mixed lognormal-gamma ACD(1,1) may not be significant. However, the result that the mixed lognormal-gamma ACD(1,1) model outperforms easily all other specified ACD models in this study, based on AIC and TIC scores, emphasizes the capability and potential usage of the information-theoretic approach in model evaluation. This is something the hypothesis testing approach can never have achieved. Based on the same reason, we decided to only include basic ACD models and two mixed distribution ACD(1,1) models in this study rather than considering some more complicated existing ACD models.

The ACD model evaluation results are mainly achieved by applying the M-estimator version AIC in this study. These results support our conclusion made in Section 2.5.1 about the applicability of AIC procedure for model evaluation in general. Ideally, we would like to be able to compare the M-estimator AIC with GIC directly in model evaluation. This requires to overcome the computation difficulty in estimation of GIC's trace term which involves two $k \times k$ matrices. The key to this problem is to be able to get the reliable numerical estimates of the first and the second derivative functions of the optimization objective function for parameter estimation and this will be a topic of future research work.

Chapter 4

Stochastic rainfall models – a literature review

The following five chapters present the research results of the second topic of this thesis: point process models for fine-scale rainfall time series at a single site. The study described herein further develops the original Bartlett-Lewis Pulse (BLP) model proposed by Cowpertwait et al. (2007) in terms of model structure characterization, fitting procedure, and parameterisation.

For simplicity, Cowpertwait et al. (2007) is referred to as 'CIO2007' (stands for the paper written by Cowpertwait, Isham, and Onof in 2007 year) for the rest of this thesis and their fitted BLP model as 'the original BLP model'. The 60 years (01/01/1945 to 31/12/2004) of 5-minute rainfall series used in CIO2007 for fitting the original BLP model, provided by the New Zealand National Institute of Water and Atmospheric Research (NIWA), was recorded at Kelburn, a site near Wellington, New Zealand. Therefore, we refer this fine-scale rainfall data as 'the Kelburn data'. For a comparison with the results reported in CIO2007, the Kelburn data is selected for the study of point process rainfall models in this thesis.

4.1 Introduction

Fresh water is an essential resource and rainfall is the primary source of fresh water supply. The process of precipitation is a fundamental component of the hydrological cycle and it is a complex and delicately balanced process (Shaw, 1988). However, there is a need to model the process of precipitation for various applications, e.g. short term or long term forecasting, data reduction and generation of rainfall series, scientific understanding, etc. The deterministic approach (primarily based on atmospheric

physics) of modelling rainfall process is generally considered as inappropriate for the following reasons: (1) the system of governing laws is not strictly deterministic or the laws are not completely understood; (2) perfect measurement of the current state of the atmosphere and its environment is impossible (Stern and Coe, 1984, discussion part). Therefore, rainfall is treated as a stochastic process in many applications.

There are a vast range of rainfall models proposed for different application purposes in the literature. Following Cox and Isham (1994) and Onof et al. (2000), four broad types of rainfall models may be classified, namely, models of dynamic meteorology, multi-scaling models, empirical statistical models, and point process models. The multi-scaling models will not be reviewed here for they follow a completely different avenue in terms of the mathematical representation of rainfall. The multi-scaling models describe the spatial evolution of the rainfall process in a scale-independent fashion, e.g. multifractal cascades (Gupta and Waymire, 1993). For a review of multi-scaling models readers are referred to Foufoula-Georgiou and Krajewski (1995). The meteorological models, in which large systems of simultaneous nonlinear partial differential equations are specified to represent the physical processes, are generally used for weather forecasting. The meteorological models thus serve a different purpose from those rainfall models of interest here in which the aim is to produce realistic simulation of the observed rainfall series. Therefore, this literature review is confined to the empirical statistical models and point process models, i.e. stochastic rainfall models.

Because the precipitation process is essentially a space-time process, stochastic rainfall models can be further broadly divided into two classes: temporal (single site) models and spatial-temporal models.

As early as Le Cam (1961), the space-time rainfall process has been modelled analytically as a random field. In the first part of a series of three articles on the mathematical structure of rainfall representation, Waymire and Gupta (1981a) discussed the mathematical structure of rainfall in space and time. Whereas characterizing the highly variable temporal structure of rainfall at a single site is by no means straightforward, the challenge of modelling rainfall in both space and time is much more formidable. It is, therefore, not surprising that not as many spatial-temporal models (as temporal models) are found in the literature. Some examples of spatial-temporal models follow.

Rodriguez-Iturbe and Eagleson (1987) investigated the spatial and temporal structure of rainfall from storm events using point process techniques. Through detailed analysis of the space and time structure of a large number of storms observed by a 93-

station network in and around the 154-km^2 Walnut Gulch catchment in Arizona, USA, they demonstrated the feasibility of modelling the spatial and temporal structure of rainstorm events using mathematical multidimensional point process techniques. Cox and Isham (1988) developed a relatively simple model, in which storm centres arrive in a homogeneous three dimensional Poisson process (two space and one time). Following each storm origin, and at the same spatial location, cell origins occur in a temporal Bartlett-Lewis process (Section 1.1.1). The aim of this model was to extend the earlier temporal model developed in Rodriguez-Iturbe et al. (1987a) to a fully spatial-temporal process in such a way that the marginal temporal process at a single spatial location of the earlier model is preserved. This model can be used to generate rainfall series over a given area for a given time scale. Partly motivated by Cox and Isham (1988)'s work, Cowpertwait (1995) proposed a generalized spatial-temporal model of rainfall based on the Neyman-Scott process in which more than one type of rain cell can be present in the same storm. By fitting this spatial-temporal model, using two cell types, to rainfall data taken from six sites in the Thames basin, Cowpertwait (1995) has shown that the model is likely to be particularly useful in hydrological studies that use hourly data. Cowpertwait's (1995) model was implemented into a modelling package called RainSim (Burton et al., 2008). Ultimately, a full continuous space-time rainfall model enables the reproduction of most essential features of the rainfall process (Onof et al., 2000).

In the following review, we will concentrate on temporal models of rainfall modelling at short time scales up to and including daily series, since these relate to the models in this research. These models may be useful in typical hydrological applications that include the design of stormwater sewerage systems for which sub-hourly data are likely of particular importance, and the estimation of the design flood hydrograph, typically based on hourly to daily rainfall (Onof et al., 2000). We focus on the evolution of the Poisson-based point process models: from Poisson white noise model to the recent developments of temporal Poisson cluster process models. The Poisson white noise model is the simplest Poisson-based point process rainfall model, and thus forms a starting point for the review of the Poisson-based rainfall models. One objective of this research is to improve the model performance of the Bartlett-Lewis pulse (BLP) model. This literature review provides the background information about why and how the BLP model was developed and indicates a direction in which the model may be improved.

4.2 Overview of stochastic rainfall models

4.2.1 Rainfall measurement and characteristics

Precipitation is an intermittent process over continuous space and time, which is usually recorded as cumulative amounts over fixed time intervals and at fixed locations. Rainfall models are fitted to rainfall time series data, typically recorded at fixed time intervals, e.g. 1-hour or 24-hour, using a 127 mm diameter rainfall gauge. Over the last few decades, automatic recording gauges like a tipping-bucket rain gauge or a Dines tilting siphon rain gauge (see Figure 4.1) are used so that rainfall at a single spot can be recorded as a continuous trace or pluviograph (Shaw, 1988; Butler and Davies, 2004; Sansom, 1987, 1992). Modern information technology is applied to digitize the recording of raingauge charts since the late 1970s so that the resulting data are point rainfall totals at a resolution of 0.01 mm at one minute intervals (Acreman, 1990). Hence, fine-scale rainfall time series at various aggregation levels (1-min upwards) are widely available for analysis nowadays.

Figure 4.1: A Dines tilting siphon rain gauge. This picture is downloaded from R W Munro Limited web page www.munro-group.co.uk in 2010.

Rainfall is the end result of complex atmospheric processes interacting in space and time. There are a few important empirical characteristics associated with the rainfall series which should not be ignored in modelling rainfall.

Firstly, rainfall series exhibit clustering patterns: rainfall events are grouped between long dry spells. For example, a frontal weather system will usually give rise to

a sequence of events as it passes over an area. Therefore, a stochastic rainfall model should be formulated using this information (Kavvas and Delleur, 1981).

Secondly, rainfall data are collected in discrete time, and can be aggregated over a variety of time scales, with different effects being evident at different scales. In particular, statistical dependencies change with a change of scales, say from hours to days to weeks (Rodriguez-Iturbe et al., 1984). In general, the autocorrelations between rainfall amounts on successive intervals decrease as the interval length increases, e.g. daily series has a higher autocorrelation than the weekly or monthly series (Buishand, 1977). It is often important that a model can be used reliably to represent the main features of rainfall at scales other than those used in fitting to empirical data.

Thirdly, due to the recording instrument measurement accuracy limit and the data conversion consideration in the process of digitalization, trace value rainfalls may be recorded as zero. For example, a 'dry' 1-hour interval may be defined as an interval over which less than 0.1 mm rainfall has been recorded (Sansom, 1987, 1992). Because the estimate of the proportion of dry intervals is sensitive to the accuracy of the data measurements and the specified threshold value for defining a dry interval, using the probability of a dry interval in fitting the model to data can lead to bias. It is, therefore, reasonable to adopt a threshold proportion dry concept for the assessment of model performance. For example, a threshold of 0.8 mm was used for the Netherlands daily rainfall data in Buishand (1977); in climate change analysis a dry day threshold of 1.0 mm is typical (Burton et al., 2008); the threshold proportion dry (PD) notion has been applied in the assessment of the PD properties of the fitted models in Rodriguez-Iturbe et al. (1987b), Cowpertwait (1998) and CIO2007.

A final concern is the seasonality exhibited in rainfall series. The most obvious, and probably most significant one, is the annual cycle seasonal pattern. Such seasonal patterns imply non-stationarity, which cannot be ignored. A commonly accepted treatment is to pool the observations from the same month (or the same season) into subsamples and assume weak stationarity for the rainfall processes within the subsamples (see, e.g. Buishand, 1977; Rodriguez-Iturbe et al., 1987; Foufoula-Georgiou and Lettenmaier, 1987; Acreman, 1990; CIO2007). An alternative approach is to specify the model parameters as a function of time of year to take account of seasonal effects as treated in Stern and Coe (1984). The year to year variation in rainfall series will be treated as stationary because the models covered in this review have not been considered to represent the long term (beyond annual cyclicity) rainfall patterns. In the case

of fine-scale (i.e. sub-hourly) rainfall series, diurnal patterns may exist.

Rainfall data collected through radar and satellite measurements are beyond the research scope of this thesis. Readers are referred to Foufoula-Georgiou and Krajewski (1995) for more on radar and satellite rainfall data analysis.

4.2.2 Empirical statistical models and point process models

The temporal process of rainfall can be divided into the occurrence process and the depth process. Rainfall depths are associated with the occurrence of rainfall events (Waymire and Gupta, 1981a). Empirical statistical models essentially take the discrete time series approach to modelling rainfall. Generally, an empirical statistical model is fitted in two separate steps: the occurrence of rainfall is modelled first, followed by the modelling of the rainfall amounts and then two models are superimposed. Empirical statistical models are referred to as profile-based models by Cameron et al. (2000). Figure 4.2 shows a schematic of an empirical statistical rainfall model for discrete time intervals. A group of successive wet periods, W_i, $i = 1, 2, ...$, may be considered as a 'rainy event' (solid rectangles in the schematic) and a rainfall event is characterized by its duration (the length of the horizontal line segment), mean intensity (the height of the dashed line which is the mean of the depths within the same wet spell), and profile (the distribution of depths y_{ij}). The lengths of successive dry periods between wet spells (i.e. between rainfall events), D_i, $i = 1, 2, ...$, are treated as inter-event durations.

Typically, a Markov chain or an alternating renewal process is used to model the sequence of wet and dry spells. As suggested by Waymire and Gupta (1981a), a Markov chain is generally inadequate in representing the types of dependencies that are exhibited by rainfall. This seems supported by the empirical evidence in that, when fitting to hourly or sub-hourly data, a Markov chain approach ends up with a large number of model parameters due to the stronger serial correlation and longer memory of dependence structure among successive rainfall depths (Acreman, 1990). Kavvas and Delleur (1975) carried out a fairly detailed data analysis for 17 rainfall stations in Indiana, USA. They demonstrated that either a Poisson process model or a renewal process with gamma arrivals typically do not reproduce the empirical behaviour exhibited by temporal rainfall. Their main results were reproduced by Waymire and Gupta (1981c) using the probability generating functional for a Neyman-Scott process, which is a different, but more direct, approach. Compared with point process models, in

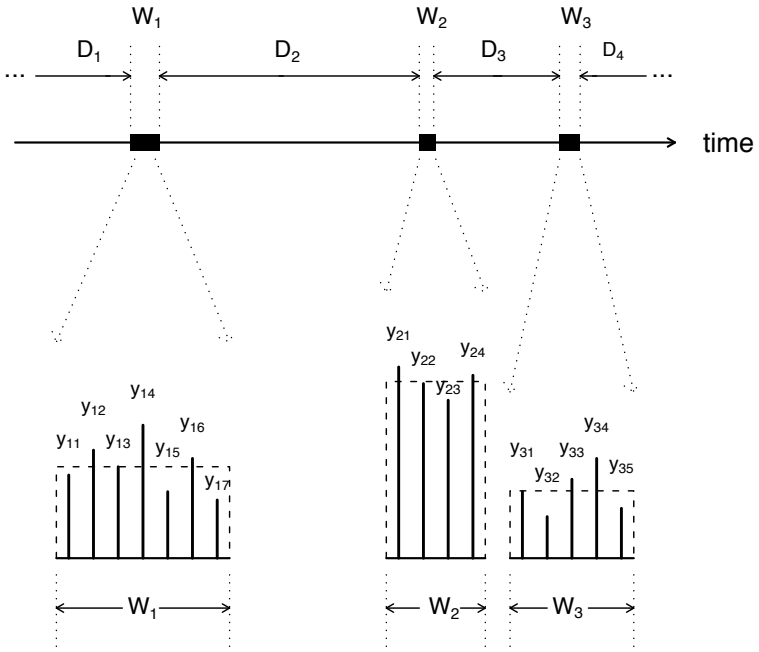

Figure 4.2: A schematic of an empirical statistical rainfall model. Symbols D_i, $i = 1, 2, ...$, represents the dry spells sequence and W_i, $i = 1, 2, ...$, represents the wet spells sequence.

which some basic physical (observed) structures are identified and then represented mathematically, a major disadvantage with empirical statistical models is the lack of physical motivation. In this sense, all empirical statistical models may be considered as *ad hoc*, i.e. fitting data to specific distributions, piecing together the resulting descriptions, and then coming up with a model that 'happens to work'. In Section 4.3, four empirical statistical models will be reviewed in more detail: one Markov chain case, two alternating renewal process cases, and one Markov renewal model case. More detailed review of empirical statistical models can be found in Srikanthan and McMahon (2001).

In the point process approach, we assume the existence of an underlying rainfall

generating mechanism evolving continuously in time. The rainfall process is treated as a marked point process (Section 1.1.1 and Appendix B.1). Therefore, a point process model, at least in part, represents the physical process of precipitation in its model specification, using a continuous time model for the occurrence of the rainfall events, and associating some random amount of rainfall with each event. With the specification of a small number of parameters from which other properties of the process can be deduced via the mathematical representation, a point process model can be used to provide insight into the hierarchical structure of the rainfall process. As will be demonstrated in Section 4.4.2, a point process model fits both the occurrence process and the associated rainfall depth process in one unified step because the rainfall process can be treated as a marked point process. The point process models (as with any other rainfall model) are inevitably idealized and some of the detailed assumptions are made, at least partly, for mathematical tractability. Whereas the models do not represent the detailed physical mechanisms, the parameters, at least in some broad sense, have a physical interpretation (Rodriguez-Iturbe et al., 1987a). Point process models perhaps provide the most physically realistic approach to continuous rainfall simulation with a minimal parametrisation (Cox and Isham, 1994).

In a simplest possible point process model formulation setting, a Poisson process is assumed to represent the occurrence process of rainfall events. We can then either attach an instantaneous pulse (burst) or a rectangular pulse to each of the events and each pulse is associated with a random amount of rainfall (rainfall depth for the instantaneous pulse case or intensity for the rectangular pulse case) (Rodriguez-Iturbe et al., 1984). Cameron et al. (2000) referred to the point process models as pulse-based models because of the way they characterize the rainfall process. The schematic of the Poisson white noise model (an instantaneous pulse model) is shown in panel (a) of Figure 4.3 and the Poisson rectangular pulse model is shown by Figure 4.4 (these two models will be described in more detail in Section 4.4). It is a well-known meteorologic observation that rainfall events occur in clusters as we have just reviewed in Section 4.2.1. It is also well-known in the literature that Poisson process is inadequate to represent the rainfall occurrence process because the events are assumed to be independent which cannot capture the dependence structure in observed rainfall series (Kavvas and Delleur, 1975; Waymire and Gupta, 1981a; Rodriguez-Iturbe et al., 1984). This lead to the development of Poisson cluster point process models.

In the formulation of point process models, subjective terminology such as storms,

cells, bursts, pulses, pulse durations, has been used to conceptualize the rainfall intensity process (Waymire and Gupta, 1981a; Rodriguez-Iturbe et al., 1984). In a Poisson cluster rainfall model, the term 'storm' is used to represent the rainfall generating mechanism, hence, storm origins can be interpreted as the arrival times of the leading edges of frontal rainfall areas or rainbands. The actual rainfall occurrences in time, which are conceptualized as rain cells, will be considered as a secondary random sequence conditional on and stemmed from the existing storms (Kavvas and Delleur, 1981; Cox and Isham, 1994). One of the simplest Poisson cluster model is the Neyman-Scott white noise model and its schematic is shown in panel (b) of Figure 4.3.

Poisson cluster models were originally developed by Neyman and Scott (1958) to model the spatial distribution of galaxies and were first applied to precipitation modelling by Le Cam (1961). Based on the work of Le Cam, Kavvas and Delleur (1981) applied a Neyman-Scott (NS) model to the rainfall occurrences on the time continuum. Kavvas and Delleur found that the NS process appeared to represent the clustering of daily rainfall occurrences in Indiana, USA. Rodriguez-Iturbe et al. (1984) compared the Poisson white noise model (Poisson process for occurrences associated with instantaneous pulses), the Poisson rectangular pulse model (Poisson process for occurrences associated with rectangular pulses), and the Neyman-Scott white noise model (Poisson cluster process for occurrences associated with instantaneous pulses). There are two major findings by Rodriguez-Iturbe et al. (1984): (1) the NS process describes rainfall occurrence process better than the Poisson process and this result confirms Kavvas and Delleur's (1981) conclusion; (2) the mathematical description of the rainfall process depends on the scale of measurement. Foufoula-Georgiou and Lettenmaier (1986) also examined these three models and they showed that these three models exhibit a lack of consistency at time-scales different from the one used for model fitting. Perhaps motivated by these research findings, Rodriguez-Iturbe et al. (1987a) first developed the Neyman-Scott rectangular pulse (NSRP) model and the Bartlett-Lewis rectangular pulses (BLRP) model in which a Poisson cluster process is associated with rectangular pulses to represent a rainfall process. Rodriguez-Iturbe et al. (1987b) presented a detailed empirical analysis of rainfall data taken from Denver, Colorado, USA. They compared the performance of the Poisson rectangular pulses, and the NSRP and BLRP models. They showed that the Poisson rectangular pulses model had a much poorer fit to the historical data than the cluster models.

Rodriguez-Iturbe et al.'s (1987a) work represents a remarkable progress in the de-

velopment of the point process models. The second order moments of the aggregated series were derived for the NSRP and the BLRP models. In addition, the probability of an arbitrary interval being dry was found for the BLRP model (Rodriguez-Iturbe et al., 1987a). It is the first time in the literature that an hourly rainfall series has been shown to be successfully modelled by a point process model in the sense that the main features (e.g. mean, variance, and the autocovariance function) of the observed series have been preserved at various levels of aggregation, from 1 hour to 24 hours (Rodriguez-Iturbe et al., 1987a,b). Rodriguez-Iturbe et al. (1987a) has become an influential paper that has set up the framework for subsequent development of Poisson cluster rainfall models over the past twenty plus years. Even the terminology and notation used in the formulation of various cluster rainfall models have followed a similar convention as used by Rodriguez-Iturbe et al. (1987a).

Much work on the theoretical basis for the development of stochastic rainfall models has been completed for many years. For example, a thorough treatment on Markov chains can be found in Feller (1968); functions which describe the statistical properties of a point process can be found in Cox and Lewis (1978); the general theory of point processes are outlined in Cox and Isham (1980) and their book includes a discussion of cluster point processes that include the Neyman-Scott process and the Bartlett-Lewis process. Waymire and Gupta (1981b,c) have presented a careful review of the theory of point processes and have illustrated their applicability to modelling rainfall and rainfall-driven hydrological processes. Point process models have dominated research in stochastic rainfall modelling in early 1980's.

In the following sections, after a review on four empirical statistical models, a number of important point process models, including NSRP and BLRP models, will be described in detail and the development of point process models after 1980s will be outlined. Definitions and some important properties of the Poisson process and Markov chain models can be found in Appendix B.1 and B.2 as background theory information. Readers may find them helpful for easy and better understanding of the description and specification of various stochastic models which will be reviewed in detail soon.

4.3 Some typical empirical statistical models

With the aim of generating daily rainfall sequences, Buishand (1977) analysed rainfall series of different climatic regions using alternating renewal processes in discrete time for the occurrence process. In the alternating renewal process, the daily rainfall data is considered as a sequence of alternating wet and dry spells of varying length. The wet and dry spells are assumed to be independent and the distributions may be different for wet and dry spells. Buishand (1977) found the zero truncated negative binomial distributions to be useful for the description of the wet and dry spell sequences. A distribution of rainfall amounts, fitted with shifted gamma distributions, on wet days is then superimposed on the binary process structure. With data series collected from different rainfall stations which are located in different climatic regions (e.g. some in the north Europe, some in the tropical region), Buishand (1977) has made a comprehensive and detailed analysis of the empirical characteristics of the collected daily rainfall data. For example, apart from what we have cited in the overview section, Buishand also found that: (1) there is no evidence for correlation between the rainfall amount on the first day of a wet spell and the length of the preceding dry spell; (2) the first and the last day of a wet spell have smaller means than the other wet days. the smallest mean is found for solitary wet days; (3) there is some evidence for serial correlation of successive rainfall amounts within a wet spell. Because of the diversity of his data sets, Buishand showed that the variability of the model performance in fitting to different type of data series was high. For example, he found that an alternating renewal process performed better than a binary discrete autoregressive moving average (DARMA) process in fitting the data from the Netherlands but the DARMA model looked more promising in tropical and monsoonal areas. Whereas Buishand (1977) has thoroughly developed the alternating renewal process modelling rainfall approach, he pointed out that in the case of long dry spells, the negative binomial distribution sometimes cannot fit the distribution or the length of these spells. He suggested, in such cases, using transition probabilities for the generation of the wet-day series instead of generating lengths of wet and dry spells.

Stern and Coe (1984) employed a Markov chain model for daily occurrences which represents a general approach to the analysis of daily rainfall data from a single site and similar ideas have been presented by other authors, both previously and subsequently. In this approach, rainfall occurrences are characterized by the binary process indicat-

ing whether or not it is raining on a particular day. Like all the empirical statistical rainfall models, the analysis is broken up into separate models for rainfall occurrences and rainfall amounts. In particular, Stern and Coe (1984) used a model with Markov chains fitted to the occurrence of rainy days, and a gamma distribution fitted to the rainfall amounts captured on wet days. The model was fitted to data taken from a 53-year record in Morogoro, Tanzania, and a 37-year record taken from Irbid, Jordan, as examples to assess the model performance. In the use of the models section, the authors restricted their attention to a first order, two-state Markov chain to describe the occurrence of rain process, although higher order Markov chains (Appendix B.2) were examined in the model fitting section. One special feature (which has been emphasized by the authors as an important advantage) is the possibility of fitting the model as a Generalized Linear Model (GLM, Nelder and Wedderburn, 1972) framework. The seasonal effects of the rainfall occurrence process was taken into account by incorporating a GLM model into the expressions for transition probabilities for the Markov chain. The seasonal effects of rainfall amount was taken into account through the mean of the gamma distribution which was allowed to depend on whether rain had occurred on the previous day, as determined by the Markov chain. Standard statistical tests, e.g. deviances and Pearson's Q-statistic, were used to assess the model goodness-of-fit. Depending on the result of these tests, for a given site, the model used between 20 to 50 parameters with seasonal effects taken into account. The performance of the model was shown to be good for the intended application, which was agricultural planning.

Foufoula-Georgiou and Lettenmaier (1987) stated that Markov chains have been found, in general, to be inadequate when modelling the clustering dependencies present in daily rainfall occurrences. Foufoula-Georgiou and Lettenmaier (1987), therefore, developed a Markov renewal model for rainfall occurrences in which the time between rainfall occurrences (in number of days) were represented by two different geometric distributions (a two-state process). The transition from one distribution to the other was governed by a Markov chain. The daily rainfall process is modelled by coupling the Markov renewal model with a mixed exponential distribution for the rainfall amount captured on rainy days. The authors defined a rainfall event as a rainy period of consecutive wet days. In the model, rainfall events were classified as either primary or secondary, where a primary event corresponded to the arrival of a front, and a secondary event corresponded to the occurrence of rainfall within the same frontal system. The model thus exhibits clustering as a result of this dependence structure.

This model is different from a Markov chain in that the probability of having a rainy day does not depend on the condition (wet or dry) of the previous day but rather on the number of days since the last rainfall event. The authors pointed out that the model behaves as a Markov chain during a rainy spell. The model uses four parameters for the rainfall occurrence process and then an additional three parameters for the depth (rainfall amount) process, before seasonal effects are taken into account. The model was applied to a 15 years daily rainfall series for Snoqualmie Falls, Washington, USA. The data were divided into five seasons and the parameters were estimated for each season using the maximum likelihood method. The model was found to be successful in preserving the short-term structure of the occurrence process, as well as the distributional properties of the seasonal rainfall amounts.

Acreman (1990) developed a model to generate hourly rainfall series. The observed dry spell sequence and wet spell sequence (Figure 4.2) have been treated as independent of each other and the Pareto and exponential distributions were fitted to the spell lengths of the two sequences, respectively. This is an alternating renewal process model for the rainfall occurrence process. Note that the author has used a continuous distribution to fit the discrete data – the spell lengths are measured in number of hours. Therefore, he had to make continuity corrections in the fitting process. The length of a wet spell, i.e. 'the duration of an event' as the author has termed it, was defined as an integer value defining the length of a continuous sequence of wet hours bounded on either side by at least one dry hour. Given an event duration, a gamma distribution, conditional on the event duration, was fitted to the total volume of rainfall observed in the rainfall event. In this way, the strong correlation evident between event depths and durations was taken into account. Furthermore, a beta distribution was used to obtain the rainfall profile (i.e. how rainfall depths were distributed over the duration) of rainfall events. Modelling the rainfall event profile was crucial for retaining the observed annual maximum rainfall property in the fitted model. The model was fitted to data from Farnborough, UK, and the data divided into summer and winter subsamples. There were a total of 22 parameters in the model and some of these parameters were dependent on the season, and others were constant throughout the year. The model was assessed by comparing the monthly rainfall totals and the annual maxima obtained from the simulation samples against the historical values. The results showed that the simulated monthly totals were consistently greater than the historical monthly totals for December to April, and consistently less than the historical values

for July to October. This implied that the data may need to be divided into more seasonal subsamples. The simulated annual maxima at the 1 hour aggregation level fitted the historical annual maxima well up to a 20-year return period[1]. The simulated annual maxima at the 24 hours aggregation level showed a slight overestimation when compared to the historical values, although it should be noted that the extreme values had not been used to fit the model.

4.4 Evolution of point process models

4.4.1 Instantaneous pulses models

The phrase 'white noise' is used for 'instantaneous pulses' or 'bursts' in the point process rainfall model literature. Given the occurrence of a rainfall event, the pulse depth (the rainfall amount of the event) is assumed to occur as an instantaneous burst. The pulse depth is treated as a random variable which is associated with each rainfall occurrence. The Poisson white noise model is the simplest possible point process rainfall model in which the rainfall occurrence process is assumed to follow a Poisson process with instantaneous pulses associated with each occurrence. Depths are aggregated to fixed time intervals which match historical rainfall series so that the discrepancy between the continuous time model representation and the observed discrete time series data is overcome. The rainfall occurrence process can be better represented by a Poisson cluster process, e.g. the Neyman-Scott process. The Neyman-Scott white noise model is formulated by assuming a Neyman-Scott process for rainfall occurrences with associated instantaneous pulses. A schematic of both models is shown in Figure 4.3.

The Neyman-Scott (NS) process is defined by a Poisson distribution of storm origins (cluster centres) in time, where a random number of rain cell origins are associated with each storm origin. The waiting times of the rain cell origins after the storm origin are assumed to be independent and identically distributed (usually an exponential distribution is assumed). Note that the NS process employs a hierarchical model structure to represent the actual rainfall occurrences. An NS process is schematically shown in panel (b) of Figure 4.3. The random depths of the instantaneous pulses are represented by the heights of pulses.

Kavvas and Delleur (1975) first used the Neyman-Scott white noise model for representing rainfall events. They derived the probability generating function for the

[1]The definition of the return period can be found in Section 6.4.3.

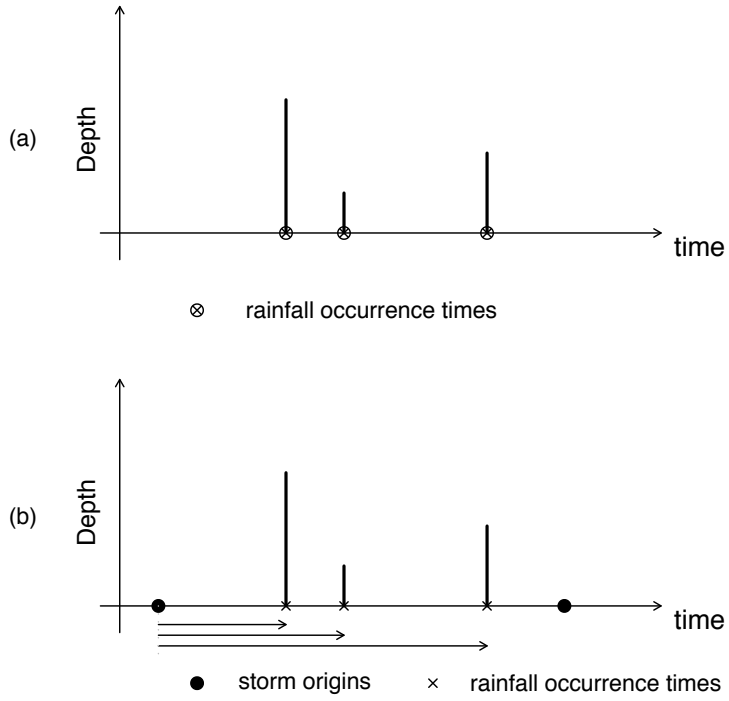

Figure 4.3: A schematic of (a) the Poisson white noise model; (b) the Neyman-Scott white noise model

occurrence process, and fitted the model to daily rainfall series in Indiana, USA. In an attempt to illustrate that a mathematical description of the rainfall process depends on the scale of measurements, Rodriguez-Iturbe et al. (1984) derived the analytical expressions for the mean, variance, and correlation function of the aggregated processes (at the hourly and daily time scales) for three models which had been used for comparison: the Poisson white noise model, the Neyman-Scott white noise model, and the Poisson rectangular pulse model. While the Neyman-Scott white noise model has been found to represent the rainfall occurrence process quite well, neither model can preserve the main features of the observed rainfall series at various time scales (Kavvas and Delleur, 1981; Rodriguez-Iturbe et al., 1984; Foufoula-Georgiou and Lettenmaier,

1986).

4.4.2 Poisson rectangular pulse model

The Poisson rectangular pulse model is the simplest model built from rectangular pulses associated with a Poisson-based stochastic process. A Poisson rectangular pulse model has two marks attached to a Poisson process: (i) a rainfall intensity X (the height of a rectangular pulse) and (ii) a lifetime L of a rainfall event (the length of a rectangular pulse). Figure 4.4 shows a schematic of the Poisson rectangular pulse model. The arrival times of rectangular pulses (the rainfall occurrence process), which are represented by symbol $\otimes$, follow a Poisson process with rate λ (per unit time). Different pulses may overlap to contribute to a single 'rainfall event' as shown in panel (a). The total intensity is the sum of intensities of all active pulses as shown in panel (b). Both X and L are random variables of some distributions of positive domain and we denote $\mathrm{E}(X) \equiv \mu_X$ and $\mathrm{E}(L) \equiv \mu_L$.

The Poisson rectangular pulse model has been studied and used by Rodriguez-Iturbe et al. (1984) and the moment properties (up to second order) were found. The derivations provide the background for the development of Poisson cluster rainfall models. Therefore, Rodriguez-Iturbe et al. (1987a) examined the Poisson rectangular pulse model in detail before the introduction of the Neyman-Scott rectangular pulse (NSRP) models and the Bartlett-Lewis rectangular pulse (BLRP) models.

It is helpful to outline the major results of the Poisson rectangular pulse model presented by Rodriguez-Iturbe et al. (1987a). These derivations illustrate how a continuous process rainfall model can be used to fit the observed discrete time rainfall series data. They also help us to understand how various Poisson cluster rainfall models can be further developed and evolved from one to another. The major model properties are outlined as follows.

Let $X_{t-u}(u)$ be the intensity at time t due to a point with origin at time $t-u$ (and so measured a time u later) and $N(t-u)$ be the corresponding counting process of the rectangular pulse occurrences. Then, the total intensity $Y(t)$ at time t is given by

$$Y(t) = \int_{u=0}^{\infty} X_{t-u}(u)\mathrm{d}N(t-u), \tag{4.1}$$

where we assume that $X_{t-u}(u)$ and $\mathrm{d}N(t-u)$ are independent of each other as noted in Equation (B.11). We have $X_{t-u}(u) = X$ with probability $S_L(u)$ or zero otherwise, where S_L is the survivor function for the lifetime L of a rainfall event. It follows that

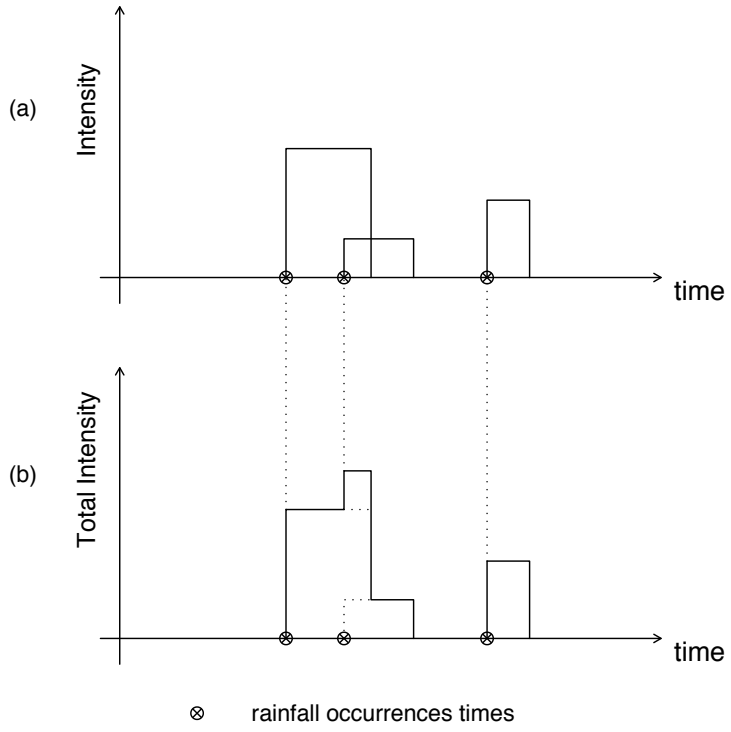

Figure 4.4: A schematic of the Poisson rectangular pulse model: (a) marked Poisson arrival process; (b) the total rainfall intensity.

$\mathrm{E}[X_{t-u}(u)] = \mu_X S_L(u)$ and $\int_0^\infty S_L(u)\mathrm{d}u = \mu_L$. Based on the Poisson process properties given in Appendix B.1.2, the mean follows as

$$\begin{aligned} \mathrm{E}[Y(t)] &= \int_0^\infty \mathrm{E}\big[X_{t-u}(u)\big]\mathrm{E}\big[\mathrm{d}N(t-u)\big] \\ &= \int_0^\infty \mu_X S_L(u)\lambda \mathrm{d}u \\ &= \lambda \mu_X \mu_L. \end{aligned} \tag{4.2}$$

The autocovariance function of $Y(t)$, denoted by $c_Y(\tau)$, is a major component in the derivation of a number of key statistical properties of the aggregated rainfall process that are used for model fitting and assessment. The autocovariance function of $Y(t)$

follows as

$$\begin{aligned} c_Y(\tau) &= \mathrm{Cov}\big[Y(t), Y(t+\tau)\big] \\ &= \lambda\big(\mu_X^2 + \sigma_X^2\big) \int_\tau^\infty S_L(u)\mathrm{d}u. \end{aligned} \tag{4.3}$$

The continuous time rainfall process is generally observed as an aggregated process in discrete time, i.e. measurements of temporal rainfall are made in accumulated amounts over non-overlapping time intervals. Let $Y_i^{(h)}$ be the total rainfall in a time interval $((i-1)h, ih]$. Thus, for all models $Y_i^{(h)}$ is given by

$$Y_i^{(h)} = \int_{(i-1)h}^{ih} Y(t)\mathrm{d}t, \quad i = 1, 2, ... \tag{4.4}$$

Equations for calculating the statistical properties of the aggregated process $\{Y_i^{(h)}\}$ were derived by Rodriguez-Iturbe et al. (1984) as

$$\left.\begin{array}{lcl} \mathrm{E}\big[Y_i^{(h)}\big] & = & h\mathrm{E}[Y(t)], \\ \mathrm{Var}\big[Y_i^{(h)}\big] & = & 2\int_0^h (h-u)c_Y(u)\mathrm{d}u, \\ \mathrm{Cov}\big[Y_i^{(h)}, Y_{i+k}^{(h)}\big] & = & \int_{-h}^h c_Y(kh+u)(h-|u|)\mathrm{d}u. \end{array}\right\} \tag{4.5}$$

In particular, let us assume that both X and L are exponential random variables with rate parameter ξ and η, respectively. Note that $\mu_X = \xi^{-1}$, $\mu_L = \eta^{-1}$, and $S_L(u) = e^{-\eta u}$. It follows immediately from Equations (4.2), (4.3), and (4.5) that

$$\left.\begin{array}{lcl} \mathrm{E}\big[Y_i^{(h)}\big] & = & \frac{h\lambda}{\xi\eta}, \\ \mathrm{Var}\big[Y_i^{(h)}\big] & = & \frac{4\lambda}{\xi^2\eta^3}(h\eta - 1 + e^{-h\eta}), \\ \mathrm{Cov}\big[Y_i^{(h)}, Y_{i+k}^{(h)}\big] & = & \frac{2\lambda}{\xi^2\eta^3}(e^{h\eta} + e^{-h\eta} - 2)e^{-k\eta}, \quad k \geq 1. \end{array}\right\} \tag{4.6}$$

Three parameters, λ, ξ, and η, are needed to specify a Poisson rectangular pulse model in this special case.

Note that, except for the Poisson process assumption for the storm origin generation mechanism, no distributional assumption has been made up to Equation (4.5). In fact, these derived properties (up to Equation 4.5) hold in all rectangular pulse models (Onof et al., 2000).

Rodriguez-Iturbe et al. (1987a) has also shown that the probability that an interval of length h is dry (proportion dry) for the Poisson rectangular pulse model is $\mathrm{P}\big[Y_i^{(h)} = 0\big] = e^{-\lambda(\mu_L + h)}$.

From the detailed description and review on the Poisson white noise model, the Neyman-Scott white noise model, and the Poisson rectangular pulse model, in the previous two subsections, we conclude the following.

A serious drawback with the Poisson white noise model and the Poisson rectangular pulse model is the Poisson arrival times assumption. Theoretically, the Poisson arrival times assumption implies the increments of the counting process over disjoint time intervals are independent and this does not fit in general for the description of rainfall occurrence processes (Waymire and Gupta, 1981a); empirical study showed that a model based on this assumption fails to fit hourly data over a range of aggregated levels (Rodriguez-Iturbe et al., 1987a).

The Poisson white noise model and the Neyman-Scott white noise model cannot represent the rainfall process adequately over a range of time scales because the instantaneous pulses characterization is too simple to describe the rainfall depths given the assumed occurrence process models.

These findings naturally lead to the idea of combining the NS process with a rectangular pulses characterization for the representation of the rainfall process.

4.4.3 Basic NSRP and BLRP models

The Neyman-Scott rectangular pulse (NSRP) models and the Bartlett-Lewis rectangular pulse (BLRP) models were first developed and examined in Rodriguez-Iturbe et al. (1987a,b) and Rodriguez-Iturbe et al. (1988). In the simplest NSRP model specification, it is assumed that (1) the storm origins arrive according to a Poisson process of rate λ; (2) each storm origin generates a random number of rain cells, no cell origin being located at the storm origin. The number of cells is geometric or Poisson (with mean μ_C) and the cell arrival times are independent and exponentially distributed (parameter β) from the storm origin; (3) the random intensity and duration of each cell are both exponential (parameters ξ and η respectively); (4) the intensity of each cell is constant throughout the rain cell lifetime; (5) the total intensity at a time point is the sum of the intensities of all active rain cells. It is also assumed that the different intensities and durations are mutually independent and independent of the Poisson process. Five parameters, $\{\lambda,\ \mu_C,\ \beta,\ \xi,\ \eta\}$, are needed to specify a basic NSRP model. The difference between a basic NSRP and a basic BLRP model is in the location of rain cell origins relative to storm origins. In a basic BLRP model, each storm origin generates a random number of rain cells with the arrival times following another Poisson process (rate β). The lifetime (duration of activity) of the storm is exponential (parameter γ) which entails that the number of cells is geometric ($\mu_C = 1+\beta/\gamma$). It is assumed that a cell origin is associated with the storm origin with a BLRP model in Rodriguez-Iturbe

et al. (1987a). As with the NSRP model, five parameters, $\{\lambda, \gamma, \beta, \xi, \eta\}$, are needed to specify a basic BLRP model. Schematics of both models are drawn in Figure 4.5.

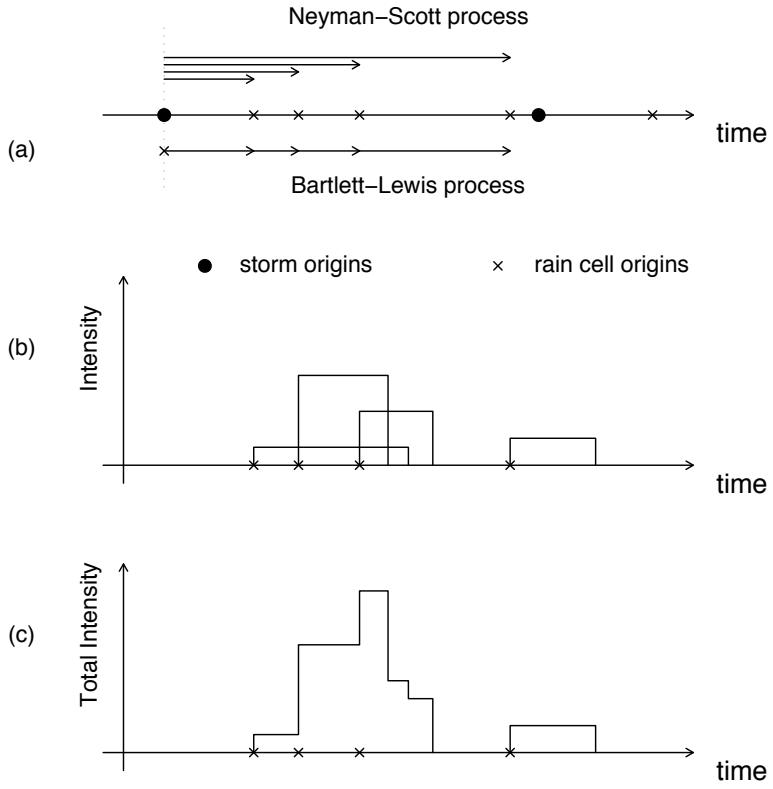

Figure 4.5: A schematic of the basic NSRP model and the basic BLRP model

The Neyman-Scott (NS) process and the Bartlett-Lewis (BL) process are depicted in panel (a) of Figure 4.5 and the differences are clear: a random number of cells occur around the storm origin with no cell origin being located at the storm origin for the NS process; a random number of cells occur sequentially from the storm origin with a cell origin being located at the storm origin for the BL process. The heights of rectangles in panel (b) represent the intensities of corresponding rectangular pulses. The total intensity in panel (c) at any point in time is the sum of all intensities contributed by all active rain cells. Therefore, in the NSRP and BLRP models, rainfall process is represented by the combination of a Poisson cluster process (for occurrence process)

with rectangular pulses (for rainfall depth process).

The NSRP model and the BLRP model are similar not only in their model specifications. Even the model performances of two equivalent models (one from each type) can hardly be distinguished empirically (Rodriguez-Iturbe et al., 1987b; Cox and Isham, 1994). Cowpertwait (1998) has shown that, up to second order, the two models essentially have equivalent statistical properties. Wheater et al. (2005) concluded that the differences between NSRP and BLRP models were probably negligible. One advantage of the NS process is that it can be, but not the BL process, generalized immediately to more than one dimension (Cox and Isham, 1980). On the other hand, the BL process is in some cases more mathematically tractable (Cox and Isham, 1994; CIO2007). Overall, no real difference in the performance of the two models has yet been found.

In the NSRP and BLRP models, the dry periods are not modelled explicitly, but as a consequence of making up the spaces between wet periods. Rodriguez-Iturbe et al. (1987a,b) successfully showed that the basic NSRP and BLRP models could preserve the first and second order properties of the depth distribution at different time scales. However, they also identified lack-of-fit to the historical proportion dry of periods.

Another challenge is in model fitting. The lack of satisfactory parameter estimation methods for point process models is partly due to the difficulty of finding likelihood functions for these models (Rodriguez-Iturbe et al., 1987a; Cox and Isham, 1994; Onof et al., 2000). The method of moments or its modified version are commonly used for parameter estimation with NSRP and BLRP models. For example, the method of fitting used in Rodriguez-Iturbe et al. (1988) was to equate selected model properties to their historical values, month by month, and to use remaining model properties and simulated values to check the adequacy of fit.

4.4.4 Developments from basic NSRP and BLRP models

Over the last two decades, a number of improvements on the basic NSRP and BLRP models have been made. Other properties such as the probability that a period is dry for the basic NSRP model was found by Cowpertwait (1991), and used to fit the basic NSRP model to a series of 10 years of British data.

Rodriguez-Iturbe et al. (1988) first proposed the random parameter model idea to improve the fit to the proportion dry. In the proposed modified BLRP model, the cell duration distribution parameter η, was allowed to vary randomly between storms. Subsequently, Entekhabi et al. (1989) introduced a modified NSRP model by allowing

the cell duration to vary for each storm according to a gamma distribution (which resulted in having an additional parameter in the model, i.e. six parameters in total). They fitted the model to the Denver summer rainfall data and found that the proportion of dry days derived from the model compared favourably to the historical values. More properties for the modified BLRP model (Rodriguez-Iturbe et al., 1988) were derived by Onof and Wheater (1993, 1994) and they reported a good fit to the proportion of dry periods at all time scales for British rainfall.

Motivated by the meteorological fact of the existence of high intensity convective rain cells within lower intensity large mesoscale areas of precipitation (e.g. Austin and Houze, 1972; Waymire and Gupta, 1981a; Cox and Isham, 1994), Cowpertwait (1994) proposed a generalized point process model for rainfall. In the model, a number n of cell types are used for the NSRP model with different exponential distributions for both cell durations and cell intensities. The model was fitted using two cell types (eight parameters per season) to hourly data from Farnborough, UK. In comparison with the equivalent basic NSRP model, the most significant improvement achieved with the generalized model was a good fit to the annual maximum aggregated at the 1 hour scale. Another way of improving the representation of rainfall based on the basic NSRP and BLRP specifications is to introduce an explicit dependence between cell duration and intensity within the model (Kakou and Onof, 1996). In addition, Cowpertwait (2004) used superposed independent processes for different storm types to characterize distinct types of precipitation which provides yet another modified NSRP model. An analytical expression for the third-order moment property was derived and used to obtain a better fit to extreme values (Cowpertwait, 1998).

Recently, CIO2007 introduced the Bartlett-Lewis Pulse (BLP) model for fine resolution rainfall series. In a BLP model, the pulse arrival process is composed of three Poisson processes, which should account for the rainfall variability over a wide range of time scales and be sufficient for most applications (Koutsoyiannis, 2002). Because of such a complex model structure, on the other hand, a closed form of likelihood function can not be readily obtained for BLP models, so the method of moments approach is adopted for parameter estimation. Potentially, superposition of BLP processes leads to a large number of model properties for fitting among which a subjective choice has to be made.

4.5 Concluding remarks

The empirical statistical models and the point process models are two popular classes of stochastic rainfall models. Both classes have been used for generating continuous rainfall series at a point that preserves the main statistical properties of observed rainfall series. Simplicity (in terms of model fitting) and the intuitive appeal have been counted as the positive points of the empirical approach because the observed discrete rainfall series represent the combined effect of several underlying mechanisms (Stern and Coe, 1984; Foufoula-Georgiou and Lettenmaier, 1987). The major drawback of the empirical models is that they are not based on any known physical process governing rainfall. One important contribution of Waymire and Gupta (1981a,b,c) is that they have shown, through theoretical justification and application examples, that rainfall must be considered in its space-time context and modelled on physical ground in terms of a random field. The single site point process models form the temporal part of this approach and many successful application cases have been reported since the late 1980s, especially for the sub-daily rainfall series.

In general, the specification of empirical statistical models follows an *ad hoc* procedure. By contrast, the point process models provide a systematic and consistent approach of the mathematical representation of rainfall. The Poisson cluster models have been successfully used to fit rainfall series at various aggregation levels, from 5 minutes to 24 hours.

Parameter estimation is a problem, for which a satisfactory solution has not yet been found, for both the NSRP and BLRP models. Onof et al. (2000) regard this as one of the major areas of possible future improvement.

In both the NSRP and BLRP models, rainfall intensity is a random variable that remains constant throughout the lifetime of a rain cell. Although this may seem unrealistic in continuous time, there are many examples in the literature where NSRP and BLRP models provide a good fit to rainfall data aggregated at the 1-hour level or higher (Cowpertwait, 1994; Verhoest et al., 1997; Onof et al., 2000). fine-scale empirical analysis showed that a typical observed wet spell lasts from a few minutes to a few hours, so that a rectangular pulse approximation is adequate for many practical applications that only require rainfall simulation results at higher aggregation levels. There is a need to develop a more realistic pulse structure for the analysis of fine scale rainfall data (e.g. at 5-minute level), which is required for applications like the design

of stormwater sewerage systems. One approach has been to simulate 1-hour series using a rectangular pulse model and then to disaggregate the series using another stochastic model (Cowpertwait et al., 1996b; Onof et al., 2005). In general, however, it is preferable to have a single model that is capable of representing series over all time scales, since this would provide a more complete description of the physical process as well as reduce model uncertainty. This is a primary motivation for the development of a Bartlett-Lewis Pulse (BLP) model by CIO2007.

Chapter 5

Data description and exploratory analysis

5.1 Introduction

The Kelburn data which has been analysed by CIO2007 is selected for the study of further developments of the BLP models in this thesis. The objectives of this chapter are: (1) to give a detailed description of the data set and justify the raw data treatment; (2) to examine the long term trend pattern and the seasonal patterns which may exist. For a proper comparison, we will follow CIO2007 to apply BLP models to the pooled calendar month subsamples of the Kelburn data. Seasonal patterns within the pooled calendar month subsamples are also checked.

5.2 Data description and preparation

The Kelburn data were originally recorded in the form of continuous raingauge charts from a Dine's tilting siphon rain gauge and then aggregated over 5-minute time intervals (Sansom, 1987, 1992). A typical section of the 60-year 5-minute Kelburn rainfall data file is displayed in Table 5.1. The first column contains 12 digits which fully describe the 5-minute intervals over which rainfall totals were recorded. The first 8 digits give the date information in the form of 'yyyymmdd'. The last 4 digits define the starting time of the 5-minute intervals in hours and minutes relative to midnight. The values in the second column are the recorded 5-minute rainfall totals in millimeters. The third column records the length of continuous rain in minutes within each 5-minute interval.

There are 6311520 rainfall observations (over 5-minute intervals) over a 60-year time span (including 15 leap years). A summary of different types of observations in the data set is given in Table 5.2. The upper box shows that, out of all 6311520

Table 5.1: A typical section of the 60-year 5-minute Kelburn rainfall data file

```
...
195006290900  -0.050    0
195006290905  -0.060    0
195006290910  -0.060    2
195006290915   0.060    5
195006290920   0.050    5
195006290925   0.060    5
195006290930   0.060    5
195006290935   0.060    5
195006290940   0.060    5
195006290945   0.060    5
195006290950   0.050    5
195006290955   0.040    2
195006291000   0.000    0
195006291005   0.000    0
...
```

observations, there are 5830445 observations (92.4%) being recorded as zero rainfall (dry intervals) or negative values and 481075 observations (7.6%) are wet intervals (with positive rainfalls). The middle box shows us the composition of the 5830445 observations (92.4%) specified in the upper box: 5807472 observations are dry intervals and 22973 observations are with negative values. The 22973 observations with negative values are further broken down into two groups in the lower box: observations with zero length of rainfall duration and those with positive duration.

From Table 5.2 we notice that there are negative rainfall records in the raw data and

Table 5.2: Summary of number of observations in the Kelburn data

6311520	=	5830445	+	481075
all observations		0 or -ve obs		+ve obs
100%		92.4%		7.6%
5830445	=	5807472	+	22973
0 or -ve obs		0 rainfall		-ve obs
100%		99.6%		0.4%
22973	=	22930	+	43
-ve obs		0 duration		+ve duration

Abbreviations: obs = observations; +ve = positive; -ve = negative.

these records have been treated as missing values in CIO2007. After a close examination of the raw data records, it was found that almost all negative rainfall records are accompanied with a zero length of rainfall duration (the information given in the third column of Table 5.1 and summarised in the lower box of Table 5.2). No evidence has been found that these negative observations are coming from any particular year(s) over the sample period. No negative observations have been found mixed within a consecutive positive records sequence. Furthermore, the negative records take a very small percentage in the total number of observations (0.36%). Therefore, we decide that it is reasonable to set all negative rainfall records to zero in this study. We assume that the impact of this raw data treatment on the final analysis results is negligible.

From Table 5.1, we notice that the minimum positive rainfall records are 0.010 mm and that any changes in recorded rainfall amount are always in multiples of 0.010 mm due to the digitization process accuracy limit. In order to eliminate the systematic error source in fitting an exponential distribution to rainfall depths, $X(t)$, we have shifted the positive records origin by 0.009999 before the fitting procedure and added it back in the simulation stage. Although it does happen that a rainfall process does not last for the full 5-minute for the beginning and end intervals of a wet spell, we will treat all recorded rainfall totals as spread out evenly over the corresponding 5-minute intervals.

Because of the measurement limit, the definition of a dry or a wet 5-minute interval is a relative concept in a practical sense. In the Kelburn data, any 5-minute interval in which less than 0.01 mm rainfall total was recorded has been treated as a dry interval (records were rounded down to zero). The length (duration) of a dry or wet spell is measured by the number of 5-minute intervals. For example, a wet spell with four consecutive intervals has a period length of $5 \times 4 = 20$ minutes. Figure 5.1 shows a schematic of a 5-minute rainfall data time plot. Symbols D_i and W_i represent the dry spells and wet spells, respectively. The unit length on the time dimension is 5 minutes. The dotted line is the rainfall measurement limit line at 0.01 mm. The height of the vertical bars represents the rainfall amounts in millimeters. If we regard the recorded rainfall amount as spread evenly over the 5-minute interval, the rainfall intensity concept applies, i.e. mm per 5-min. If, instead, we consider each wet 5-minute interval as a rainfall pulse, the height is referred to as the rainfall depth as drawn in Figure 5.1.

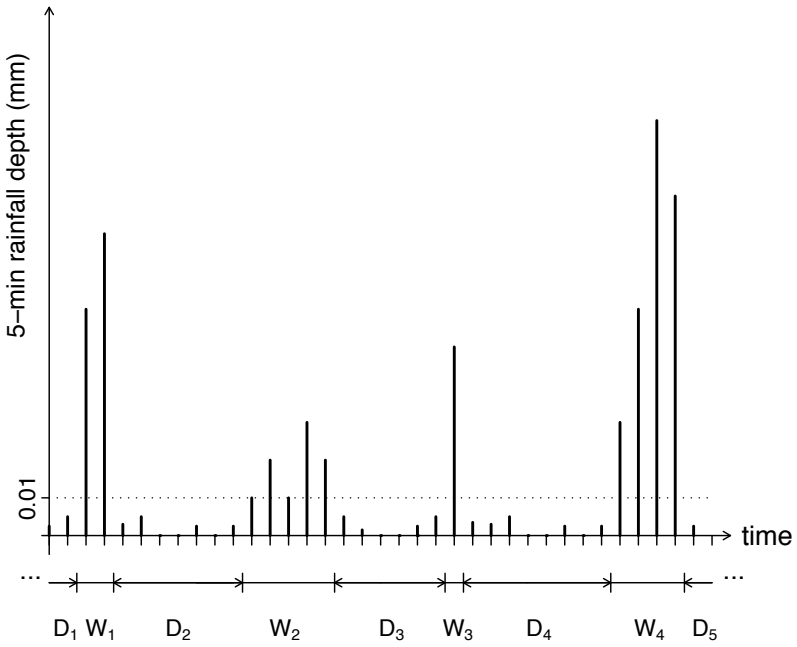

Figure 5.1: A schematic of a 5-minute rainfall time series plot

5.3 Exploratory data analysis

5.3.1 Long term trend

We begin the exploratory analysis by examining the rainfall mean level patterns. Figure 5.2 is an annual rainfall totals time plot of the Kelburn data. The points (circles) represent annual rainfall totals. The dashed line (smoothed line) shows the long term trend of the annual totals. The overall average annual rainfall over the period of 1945–2004 is 1223 mm which is indicated by the horizontal dotted line . The dashed line (smoothed line with smoother span equals 1/4) exhibits a long term annual rainfall fluctuation pattern. From the figure, some irregular 3–7-year long term rainfall patterns are apparent. For example, a wetter period pattern occurred in the late 1970s

and a drier period pattern occurred in the late 1980s. Probably the most noticeable long term pattern is that the dashed line has been consistently below the overall average annual rainfall level (the horizontal dotted line) since the early 1980s: a drier period lasted for more than 20 years. However, based on Figure 5.2, the long term trend (beyond annual cyclicity) concern over the full sample period (60 years) can not be justified.

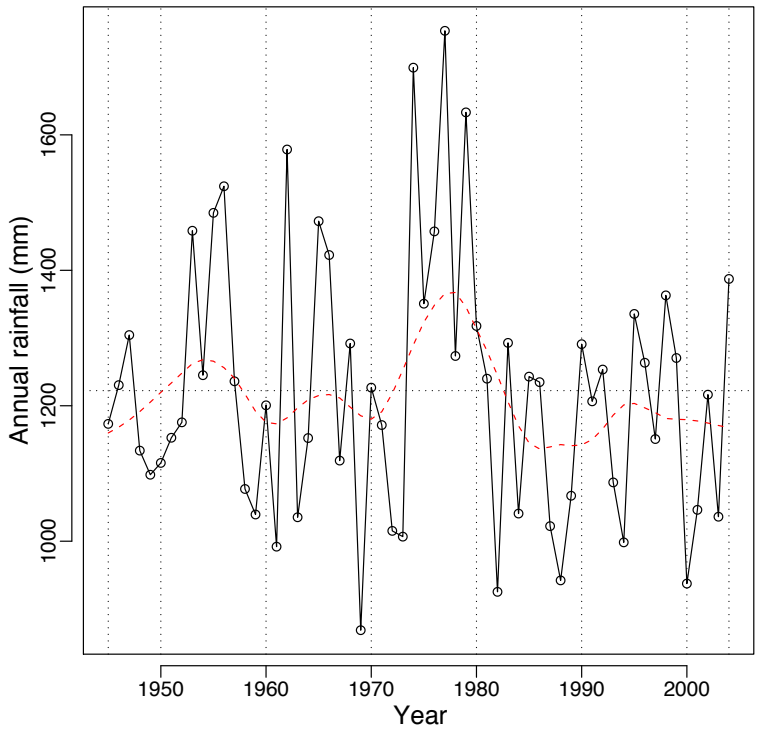

Figure 5.2: Annual rainfall totals: the Kelburn data (1945-2004)

Figure 5.3 examines the long term trend of the dry spell and wet spell durations over the 60 years sampling period. The solid lines represent the annual mean dry and wet spell durations in terms of the number of 5-minute intervals. The dashed lines show the long term trend of the annual mean durations. Both panels show a clear downward trend. For dry spell duration, the trend level decreased from about 20 hours (less than 250 5-minute intervals) down to about 10 hours (120 5-minute intervals) which is a very substantial reduction. The wet spell durations experienced the similar downward changes, from about 2 hours to less than 1 hour. The changes in both the

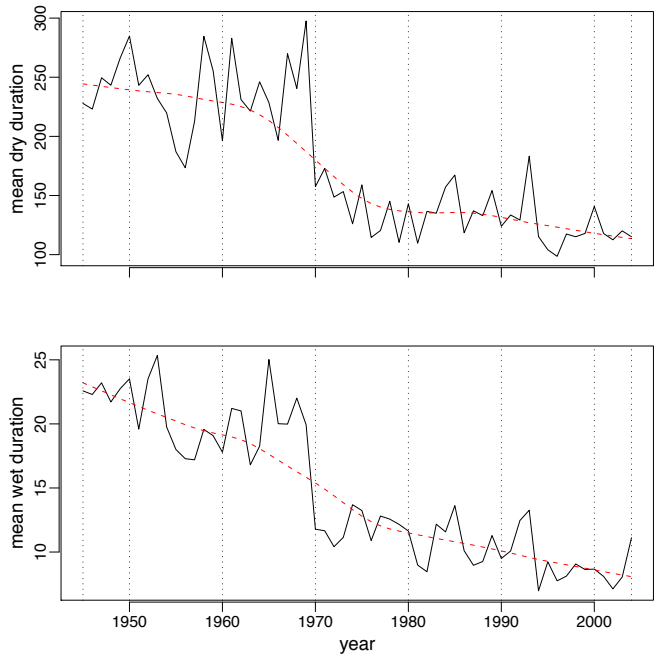

Figure 5.3: Long term patterns of mean duration of dry and wet spells: the Kelburn data (1945-2004)

dry spell and wet spell mean durations will have an impact on the long term pattern of proportion dry.

We observe a drier period from 1980s onwards in Figure 5.2. We may further identify which factor mainly caused this drier period based on Figure 5.4. In the upper panel, the solid line represents the annual mean proportion dry. In the lower panel, the solid line represents the annual mean depths. The dashed lines are the smoothed lines to show the long term trend patterns. In the upper panel, we notice a upwards trend of proportion dry by about 2% from 1945 to 2004. In the lower panel, we also notice an upward trend for the mean depth. Note that the gradient of the proportion dry upward trend line is bigger than that of the mean depth after 1980. It can be shown that there is a strong negative correlation between annual mean proportion dry level and rainfall total (-0.81) and a mild positive correlation between annual mean depth and rainfall total (+0.34). Therefore, it is mainly the effect of increase in proportion dry levels that caused the drier period since the 1980s.

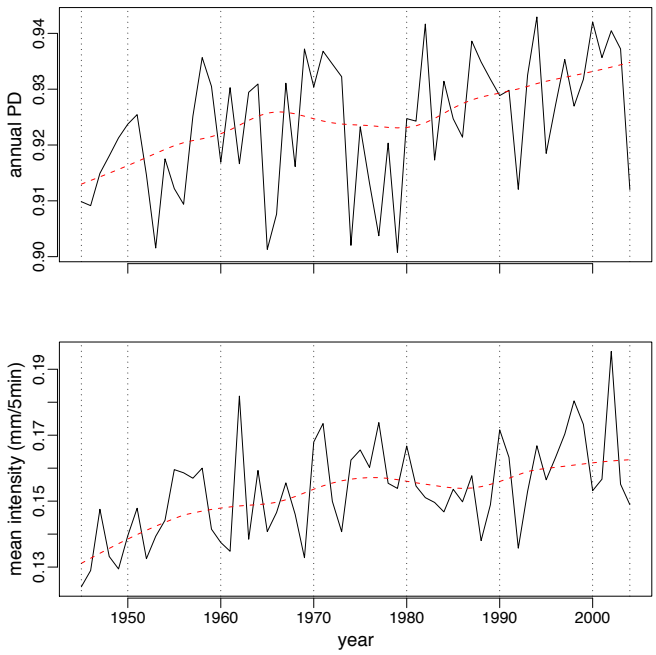

Figure 5.4: Long term patterns of proportion dry and mean depths: the Kelburn data (1945-2004)

5.3.2 Monthly seasonality and diurnal pattern

Figure 5.5 shows time plots of the monthly and daily rainfall averages. In the upper panel, the points (circles), connected by solid lines, represent mean monthly rainfall (averages over the 60 years of sampling period for each calendar month). The two dashed lines form an approximate 95% confidence band for the monthly mean rainfall. In the lower panel, the solid line represents the mean daily rainfall totals and the dashed line is the smoothed trend line to show the seasonal pattern of daily rainfall. The horizontal dotted lines in both panels indicate the overall mean rainfall levels. There is a 69.7% – 136% variation range around the overall mean in the monthly mean rainfall seasonal pattern and the approximate 95% confidence band only contains four (out of 12) points. Both panels exhibit a strong seasonal pattern indicating non-stationarity problem. The rainfall averages are higher during the winter season and lower in the summer season.

Because the Kelburn data are a sub-hourly fine resolution rainfall time series, this enables us to look at the diurnal pattern which is shown in Figure 5.6. The points (circles) represent mean hourly rainfall. The two dashed lines form an approximate 95%

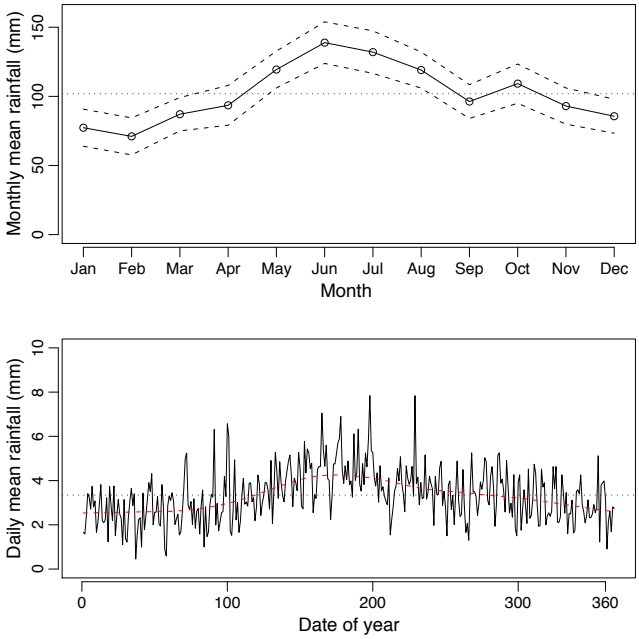

Figure 5.5: Seasonal rainfall patterns: the Kelburn data (1945-2004)

confidence band of the hourly means. The horizontal dotted line indicates the overall mean level of hourly rainfall. Note that the driest period appears in the afternoon hours and the wettest period appears in the early morning hours. The non-stationary feature of this diurnal pattern seems statistically significant: 11 out of 24 points are inside the confidence band and there is an 83.7% – 114% variation range around the overall mean. This implies that the diurnal rainfall changes can not be explained by sampling variation of a stationary series. However, in comparison with the seasonal pattern shown in Figure 5.5, there are higher proportion of points within the confidence band and the variation range is smaller.

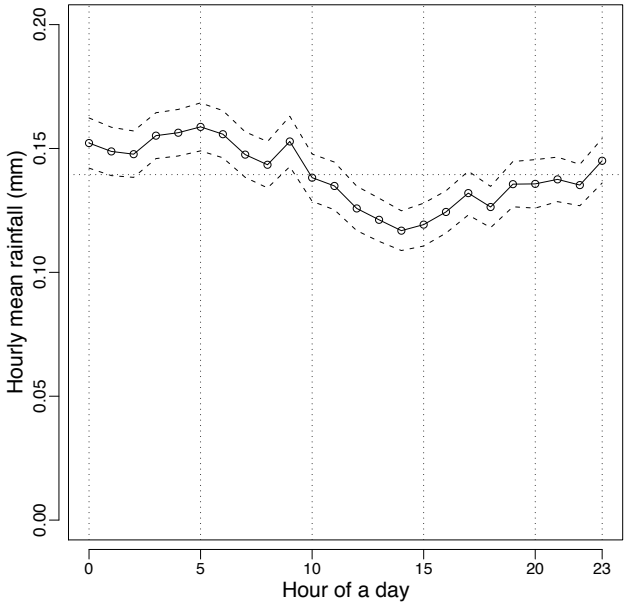

Figure 5.6: Diurnal rainfall pattern: the Kelburn data (1945-2004)

5.3.3 Numerical summary and discussion

The seasonal patterns and diurnal patterns of rainfall depths and proportion of dry periods (PD) can be examined numerically from Table 5.3 and 5.4. The rainfall depth is defined by the total rainfall (mm) recorded in 5-minute intervals. The seasonal pattern is examined by the variation from month to month within a year. The diurnal pattern is examined by the variation from hour to hour within a day and the hour of a day starting from midnight. Note that, in Table 5.3, the two highest proportion dry months and the two highest depth months happened together in January and February, two of New Zealand's summer months. For the winter season, the two lowest proportion dry months were June and July while the two lowest depth months were August and September. This implies that it was likely to have the most intensive rainfalls in January and February but the rainfall occurrences were less frequent than in the other months. On the other hand, although the rainfall occurrences were most frequent in June and July, the least intensive rainfalls were most likely to happen in August and September. We must be cautious about the interpretation of the diurnal proportion dry pattern. Both dry spells and wet spells are grouped by their starting times, but a dry or wet spell which started in this hour may well last for more than

Table 5.3: The seasonal patterns of mean depth and proportion dry

month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
depth	0.175	0.171	0.167	0.166	0.153	0.152	0.144	0.137	0.138	0.148	0.151	0.158
PD	0.951	0.947	0.940	0.936	0.911	0.894	0.898	0.902	0.921	0.918	0.928	0.942

Table 5.4: The diurnal patterns of mean depth and proportion dry

hour	0	1	2	3	4	5	6	7	8	9	10	11
depth	0.155	0.150	0.147	0.148	0.145	0.147	0.146	0.143	0.156	0.167	0.165	0.167
PD	0.921	0.920	0.920	0.922	0.934	0.940	0.940	0.948	0.941	0.917	0.936	0.938
hour	12	13	14	15	16	17	18	19	20	21	22	23
depth	0.162	0.162	0.158	0.157	0.155	0.155	0.146	0.152	0.151	0.150	0.145	0.147
PD	0.938	0.918	0.906	0.914	0.914	0.912	0.900	0.893	0.923	0.919	0.918	0.915

one hour. Therefore, the diurnal proportion dry patterns should only be interpreted in the sense of the dry spells starting times. Table 5.4 shows that dry spells started more often (i.e. the corresponding wet spells ended) in the early morning hours and started least often in the late afternoon hours. Since, on average, a wet spell lasts about one hour (Table 5.5), this implies that early morning hours are wetter and the late afternoon hours are drier which by and large match the conclusions made from examining Figure 5.6. The highest depth period appeared in the late morning hours.

We examine the empirical distributions of rainfall depths and the dry spell and wet spell durations in Table 5.5. The depth is defined as rainfall total (mm) per 5-minute given a wet 5-minute interval. Dry spell and wet spell durations are measured in the number of 5-minute intervals, e.g. the median of dry spells durations is 14.0 means that the median duration is $5 \times 14.0 = 70$ minutes. The distributions of rainfall depths, dry spell durations, and wet spell durations are all right-skewed. In particular, the distributions of depths and dry spell durations had very long right tails (mean >> median). On the other hand, the wet spell durations distribution was skewed to the right to a less extent.

Table 5.6 shows the lag 1 to lag 6 autocorrelations among dry and wet spell sequences in terms of their durations. Note that all correlation coefficients are small; the highest one is the lag 1 correlation for the pooled wet spell sequence (0.1160). In a rainfall process, dry spells and wet spells form an alternating sequence. From the cor-

Table 5.5: Empirical distribution of rainfall depths and dry or wet spell durations

	Minimum	1stQuartile	Median	Mean	3rdQuartile	Maximum
depth(mm)	0.0100	0.0400	0.0900	0.1525	0.1800	9.100
dryspell(× 5 min)	1.0	5.0	14.0	159.2	63.0	7302.0
wetspell(× 5 min)	1.00	3.00	6.00	13.14	12.00	495.00

Table 5.6: Autocorrelations among dry and wet spell sequences

	lag 1	lag 2	lag 3	lag 4	lag 5	lag 6
dry and wet spell alternating sequence	-0.04881	0.09129	-0.04815	0.08245	-0.04788	0.07036
dry spell sequence	0.04263	0.03527	0.02064	0.03986	0.03717	0.02176
wet spell sequence	0.1160	0.08273	0.06900	0.06053	0.05664	0.05584

relation coefficient values in the first row, we note that all odd lags are negative values and all even lags are positive values. The negative correlation at odd lags implies that after a long dry spell we may expect a relatively shorter wet spell, or alternatively, after a long wet spell we may expect a relatively shorter dry spell. On the other hand, the positive correlation at even lags implies that a long dry (wet) spell is more likely to be followed by a relatively longer dry (wet) spell. The positive correlation patterns are highlighted by pooling the dry (wet) spells in one sequence group as exhibited by the correlation coefficient values in row two (pooled dry spell sequence) and row three (pooled wet spell sequence). Note that the correlations among the consecutive wet spells are stronger than the dry spells and it seems that these low level, but significant positive correlations, decay at a very slow rate (i.e. exhibit long memory).

We denote the skewness and lag 1 autocorrelation statistics calculated at the 5-minute, 1 hour, and 24 hours aggregation levels by skew5m, skew1h, skew24h, ac5m, ac1h, and ac24h, respectively.[1] The results given by Table 5.7 confirm the results in the literature (e.g. Buishand, 1977), i.e. rainfall observations over longer intervals have less autocorrelation. By pooling the observations of the same month together for each of the 12 months over the 60-year sampling period, weak stationarity is assumed within each subsample. The skewness and lag 1 autocorrelation statistics are calculated for each subsample at different aggregation levels for comparison. From the table, rainfall observations at the 5-minute level consistently exhibit a much stronger autocorrela-

[1]The definition for skewness and lag 1 autocorrelation are given by Equation (6.7).

Table 5.7: Sample statistics of skewness and lag 1 autocorrelation for pooled subsamples by calendar month

month	skew5m	skew1h	skew24h	ac5m	ac1h	ac24h
Jan	23.55	12.10	5.030	0.8647	0.6123	0.2783
Feb	17.83	11.18	5.234	0.8356	0.6197	0.1854
Mar	19.50	11.56	4.436	0.8453	0.6164	0.1598
Apr	18.92	10.85	4.885	0.8338	0.5875	0.2146
May	15.78	10.82	4.182	0.8597	0.5926	0.2312
Jun	17.45	12.79	3.645	0.8464	0.6123	0.2415
Jul	12.74	8.064	3.602	0.8525	0.6647	0.2554
Aug	12.29	7.932	5.168	0.8596	0.6398	0.1786
Sep	13.02	8.538	3.397	0.8587	0.6250	0.1977
Oct	12.83	8.072	3.734	0.8637	0.6512	0.2090
Nov	22.46	12.40	4.130	0.8119	0.5986	0.2175
Dec	19.06	12.04	6.921	0.8674	0.6008	0.1434

tion than the hourly observations and likewise the autocorrelation of the hourly data is stronger than the daily data. The observed autocorrelation among rainfall series probably results from clustering of rainfall events.

While a significant seasonal pattern can be identified with the skewness statistics (low in the winter and high in the summer), the seasonal pattern with the lag 1 autocorrelation statistics is much less obvious. Note that, in the 'skew24h' column of Table 5.7, there is a big sudden rise for the August value compared with the June, July, September, and October observations. This makes the 24-hour skewness seasonal pattern unexpectedly different from those of 5-min skewness and 1-hour skewness seasonal patterns. By checking the raw data records, we found that there was a 116.88 mm daily rainfall recorded for the date of 7th August, 1963. This is the second highest daily rainfall record over the 60-year sample period. This is very unusual because intensive rains are more likely to occur in summer season; for example, the highest daily rainfall, 147.8 mm, was recorded in December. Exploratory analysis shows that without including this single day's record, the historical 24-hour skewness for August should have been 4.335 instead of 5.168. Such a high influence observation can have a significant impact on the model parameter estimation as the empirical study results show in Chapter 6 and Chapter 8.

5.4 Seasonality within the pooled calendar month subsamples

In CIO2007, the Kelburn data were divided into 12 pooled calendar month subsamples and the original BLP model was fitted to each of the 12 subsamples. In this way, the monthly seasonal effects were assumed to have been taken into account and nothing was mentioned about the concern of the diurnal effects. We will follow CIO2007 in this respect but we will examine explicitly how well the monthly seasonal effects and the diurnal effects have been taken into account after the pooled calendar month subsamples treatment.

Figure 5.7 is drawn to examine the rainfall total changes over the full sample period (60 years) for each calendar month. In each panel, 60 monthly rainfall total values (one from each year) are plotted as points (circles) and the solid line is the smoothed trend line. The trend line levels for summer months (December, January, and February) are consistently lower than those of the winter months (June, July, and August) as we have observed from Figure 5.5. No consistent or strong long term patterns have been identified within the pooled calendar month subsamples.

Figure 5.8 is drawn to examine the diurnal patterns by looking at the mean hour rainfall changes for each calendar month. In each panel, each year's hourly mean rainfall are plotted as points, i.e. 60 observations for each hour of day. Again, the solid line is the smoothed trend line. We notice that, from the figure, the changes in trend line levels represent the monthly seasonal pattern well, i.e. lower in the summer months and higher in the winter months. We have not observed any consistent or strong diurnal pattern.

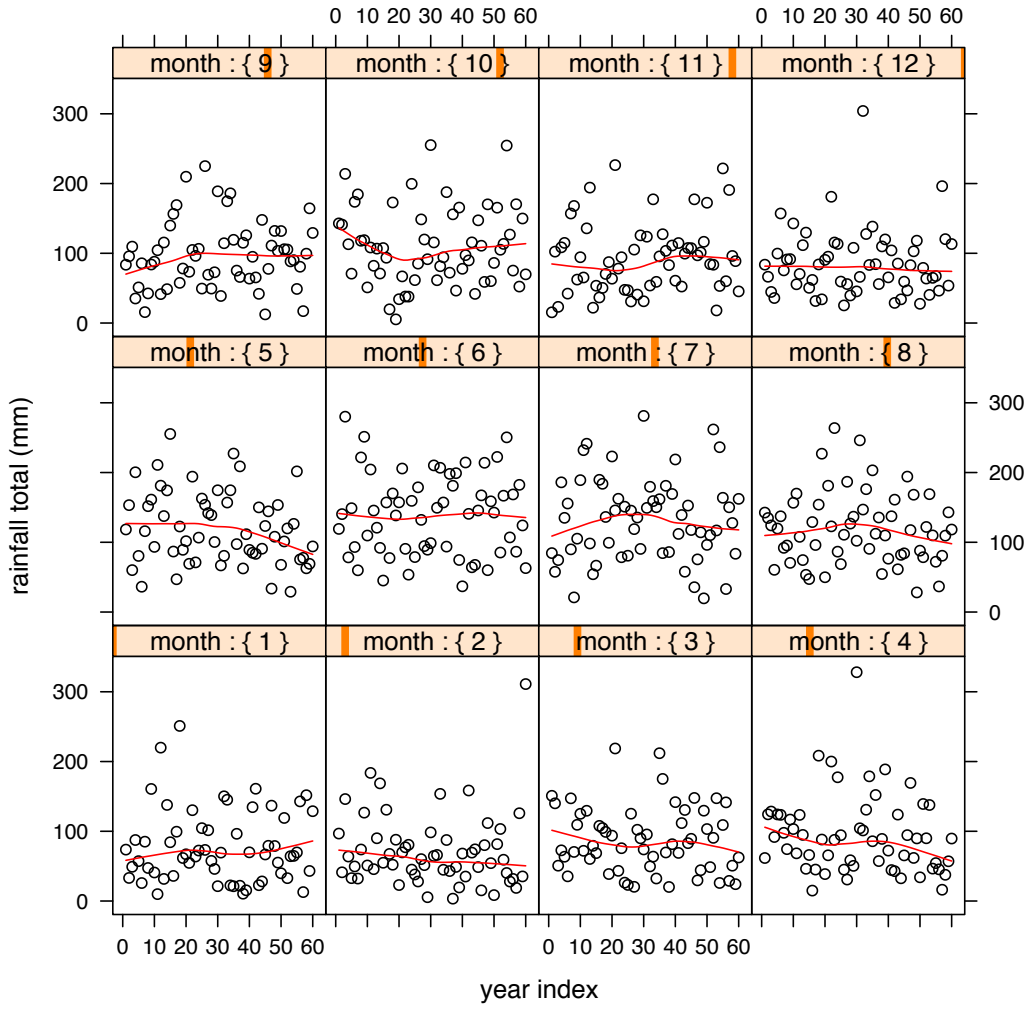

Figure 5.7: Monthly rainfall totals of pooled calendar month subsamples: the Kelburn data (1945-2004)

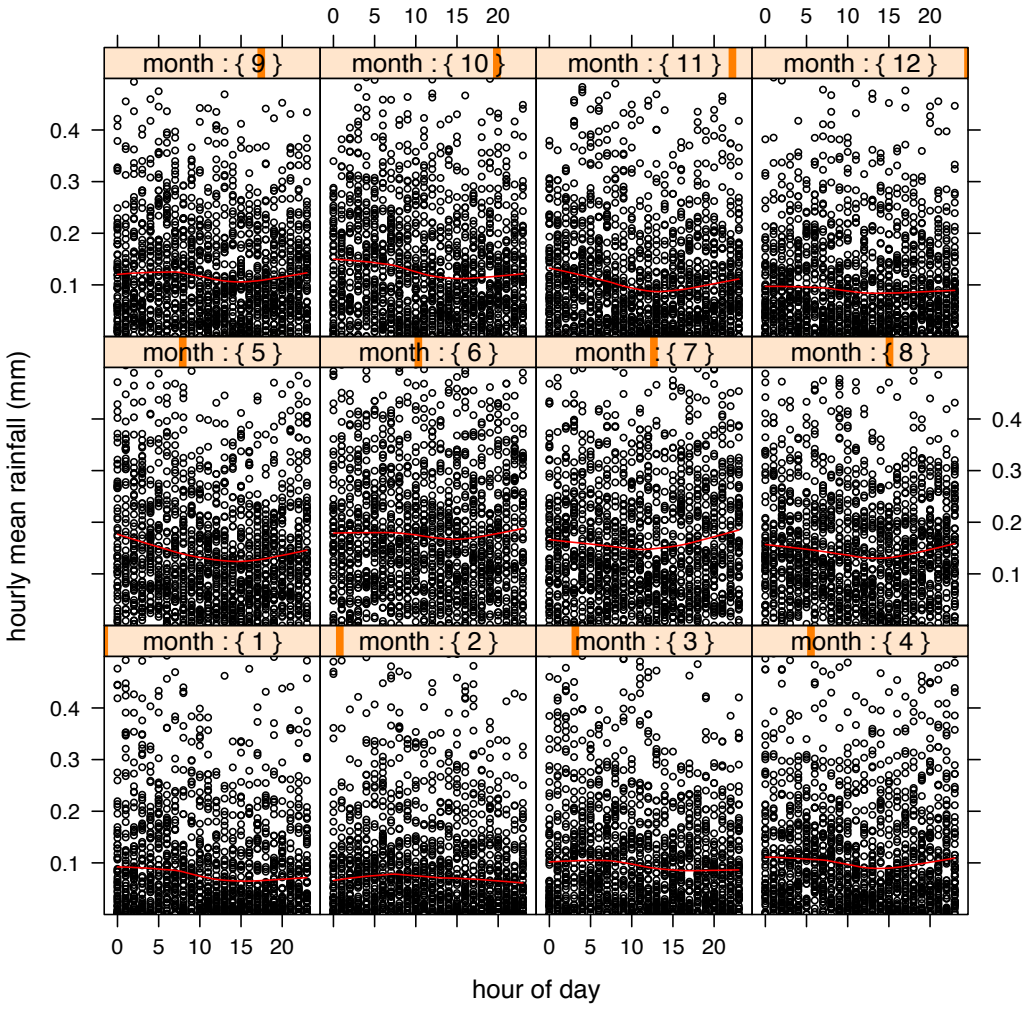

Figure 5.8: Hourly mean rainfall of pooled calendar month subsamples: the Kelburn data (1945-2004)

5.5 Concluding remarks

The results of an extensive exploratory analysis of the Kelburn data exhibit a strong seasonal pattern in terms of monthly mean rainfalls and a significant diurnal pattern in terms of hourly mean rainfalls. By dividing the full sample into 12 pooled calendar month subsamples, a further investigation on the long term trend and diurnal patterns reveals that the monthly seasonal pattern can be well captured by subsamples and no strong or consistent long term trend or diurnal patterns have been identified within subsamples. The weak stationarity assumption for pooled calendar month subsamples is supported by the exploratory analysis results obtained in this chapter.

An empirical statistical rainfall model was formulated and the model was tentatively fitted to a subsample of the Kelburn data – observations recorded over the period 1980 – 1999 (20 years). Although we have decided not to present the results in this thesis because the model still needs some refinement, preliminary research results did show that both the monthly seasonal patterns and the diurnal patterns can be accommodated quite well within one model. However, a set of substantially many parameters is required in this empirical model. For example, it needs about 26 parameters in total (with monthly seasonal effects taken into account) for model specification which is two to three times as many as a correspondent point process model (e.g. a BLP model) will need. An empirical statistical model for fine-scale rainfall data can be one of our future research topics.

An additional concern may arise for the pooled subsamples based on hour of day. By pooling observations which are from disjoint time intervals, we have implicitly assumed that observations of different subsamples are independent of each other. This may be a plausible assumption for the calendar month subsamples but is unlikely to hold for the hour of day subsamples. More empirical study is needed on the properties of the pooled subsamples based on the hour of day.

Chapter 6

Different BLP model specifications and empirical analysis

6.1 Introduction

As reviewed in Chapter 4, we know that Neyman-Scott and Bartlett-Lewis rectangular pulse (NSRP and BLRP) models of rainfall have been widely studied over the last two decades since their initial development by Rodriguez-Iturbe et al. (1987a). These models (or variants) perform well in relation to observed statistical properties of rainfall series over a range of time scales from 1-hour upwards. In order to provide a more realistic characterization of the precipitation process in analysis of fine-scale rainfall time series, the Bartlett-Lewis Pulse (BLP) model was developed and fitted to the Kelburn data, using the simplest BLP model specification, in CIO2007. The model fitting results reported in CIO2007 were encouraging, but more empirical work was needed. Several issues are identified for possible improvement with the original BLP as follows.

The instantaneous rainfall pulse generation process, the rain cell occurrences process, and the storm arrivals process are modelled by three Poisson processes, in a composite conditional structure, to characterize the physical rainfall occurrences process in a BLP model. The pulse generation processes are conditional on the active rain cells which in turn depend on the storm arrivals and durations. Such a nesting model formulation structure can capture the cluster rainfall occurrence patterns adequately but an analytic form of the model likelihood function is not available. As with most NSRP and BLRP models, the parameters in BLP models have been estimated using method of moments. This leads to a large number of possible parameter sets among which a subjective choice is made in determining the optimization objective function.

Therefore, it follows that an analysis of the fitting procedures is worthwhile.

In the original BLP model specification it was assumed that all pulse depths were independent although the model properties derived in CIO2007 does allow a BLP model to specify a within-cell depth dependence structure. The implementation of the within-cell depth dependence model specification should give a more realistic characterization of the rainfall occurrence processes so that a better model fit result is expected.

While CIO2007 reported excellent fitting results in terms of moment properties and extreme values (annual maxima at 1, 6, and 24 hours aggregation levels), a tendency for overestimation of the historical proportion of dry periods (proportion dry) values and lack-of-fit of extreme values for individual months were identified. The utilization of the derived model proportion dry properties needs to be investigated. Furthermore, a comprehensive model performance assessment analysis is necessary in comparing different model specifications.

In the current chapter, we report on a more detailed empirical investigation, which includes a study of model specification and fitting procedures, statistical properties, and model performance comparison by simulation. All statistical properties which have been modelled and some properties which have not been included in the fitting procedure are assessed in the simulation study.

The chapter is organised as follows. Section 6.2 gives a detailed description of the original model specification and the derived statistical properties of a BLP model. Different model specifications are proposed in this section. Section 6.3 discusses the fitting procedures. The issue of the inclusion of the proportion dry properties in model fitting is dealt with in this section. In Section 6.4, the parameter estimates and model assessment results from an extensive simulation study are presented and discussed. Section 6.5 summarizes and concludes the chapter.

The programming task for BLP model estimation and for generation of simulation samples based on a fitted BLP model is not trivial. By courtesy of the authors of CIO2007, the R programs (for model estimation) and the C programs (for generation of simulation samples) used in this study are adapted from the programs used in CIO2007.

6.2 BLP model specification and properties

In defining the BLP model, the terminology and notation in CIO2007 is used. As originally proposed by Rodriguez-Iturbe et al. (1987a), the continuous process of rainfall occurrences may be modelled in the following structure.

Storm origins arrive according to a Poisson process $\{T_i\}$ at rate λ, i.e. on average, there are λ storm origins per unit time interval. Each storm has a random lifetime of length D_i, where the D_i are independent and exponentially distributed with parameter γ. Therefore, each storm terminates at time $T_i + D_i$. During the lifetime of the ith storm, cell origins, denoted by $\{T_{ij}\}$, are generated in a Poisson process of rate β. This process terminates at the end of the storm lifetime, i.e. $T_i < T_{ij} \leq T_i + D_i$. The random durations of the cells, L_{ij}, are assumed to be exponentially distributed with parameter η. Both storms, and cells within storms, are assumed to evolve independently. The point process $\{T_{ij}\}$ is therefore a Bartlett-Lewis (BL) process (see, e.g. Cox and Isham, 1980, p77).

Conditional on the BL process for rainfall occurrences (the rainfall events starting times), if we further assume each rain cell has an independent random intensity, X_{ij}, that remains constant over the lifetime of the cell, we get the basic BLRP model (Figure 4.5). If instead, we further assume a Poisson process (with parameter ξ) which generates a sequence of pulses, T_{ijk}, from the ijth cell origin, we get a BLP model as first proposed in CIO2007. In a BLP model, each pulse T_{ijk} is associated with a random rainfall *depth*, X_{ijk}, generated from a positive-valued distribution, e.g. an exponential distribution (with parameter θ). We assume that the process of pulses terminates with the cell or storm lifetime, whichever is the sooner (i.e. at $\min(T_{ij} + L_{ij}, T_i + D_i)$). This BLP model formulation structure is shown by schematics in Figure 6.1 which is a reproduction of Figure 1 in CIO2007.

The BLP model has provided a more realistic mathematical characterization of the rainfall process which is vital in modelling the sub-hourly data. In a BLP model, the pulse arrival process is composed of three Poisson processes, which should account for the rainfall variability over a wide range of time scales and be sufficient for most applications (e.g. see Koutsoyiannis, 2002). Furthermore, the two storm types superposition BLP model specification has a sound physical basis – precipitation may contain a mixture of two broad types of rainfall: convective rainfall which is short-lived and very intense, and stratiform rainfall which is more prolonged but generally

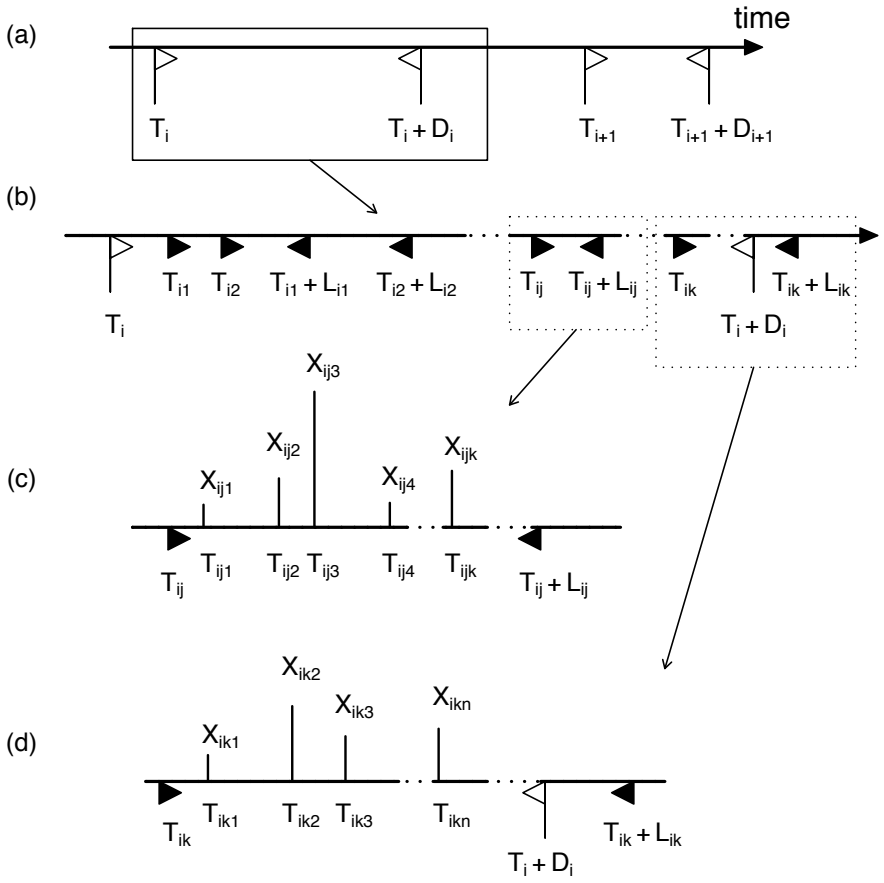

Figure 6.1: A schematic of the Bartlett-Lewis pulse (BLP) model: (a) the storm arrival process; (b) the cell birth and death process in storm i; (c) the pulse processes in cell j, storm i for a cell that terminates before the end of the storm, and (d) the pulse processes in cell k, storm i for a cell that terminating later than the end of the storm.

less intense (Austin and Houze, 1972; Cox and Isham, 1994).

Note that, for a basic BLRP model, we have assumed that there is a cell associated with the storm origin (Rodriguez-Iturbe et al., 1987a). Whereas, in the BLP model, there is no assumption that a cell origin occurs at the storm origin, or that a pulse occurs at a cell origin. Thus both storms and cells could have no rainfall. This is partly for mathematical convenience, but does affect the interpretation of the model parameters. In particular, λ is the rate at which storms occur, not just those that have non-zero rain, which is a suitable interpretation when rainfall is considered spatially, because a storm traversing a region may leave some points dry.

If we take X_{ijk} to be an exponential random variable, the BLP model has 6 parameters, which is one more than the basic BLRP model, summarized below:

λ – rate of storm origins; β – rate of cell origins; ξ – pulse arrival rate;
γ^{-1} – mean storm lifetime; η^{-1} – mean cell duration; θ – mean pulse depth.

Note that all these variables are assumed to be mutually independent in fitting the Kelburn data, although the BLP model specification framework does allow for within-cell depth dependence in CIO2007.

To account for distinct types of precipitation, e.g. convective and stratiform rain, independent BLP processes may be superposed, in which case subscripts are placed on the parameters (CIO2007). For example, λ_1 and λ_2 may represent the arrival rates for type 1 and type 2 processes of storms, corresponding to stratiform (or frontal) and convective weather systems.

Observed rainfall records are usually available in aggregated form, so for model fitting the BLP process needs to be aggregated to a discrete time series $\{Y_i^{(h)}\}$ given by:

$$Y_i^{(h)} = \int_{(i-1)h}^{ih} X(t)\mathrm{d}N(t) \tag{6.1}$$

where $X(t)$ is the depth of a pulse located at time t, and $N(t)$ is the counting process of pulse occurrences. Figure 6.2 shows a schematic of how the aggregated time series $\{Y_i^{(h)}\}$ is formed. In the figure, horizontal distances are measured in time unit; heights are measured in mm of rainfall depth or amount. Therefore, $Y_i^{(h)} = \sum X_{((i-1)h,ih]}$, where h is the time scale. For example, if $h = 1$ hour, then $\{Y_i\}$ will be an hourly precipitation time series. From Equation (6.1), properties up to third order, and the probability that an interval is dry (proportion dry, for short), were derived in CIO2007.

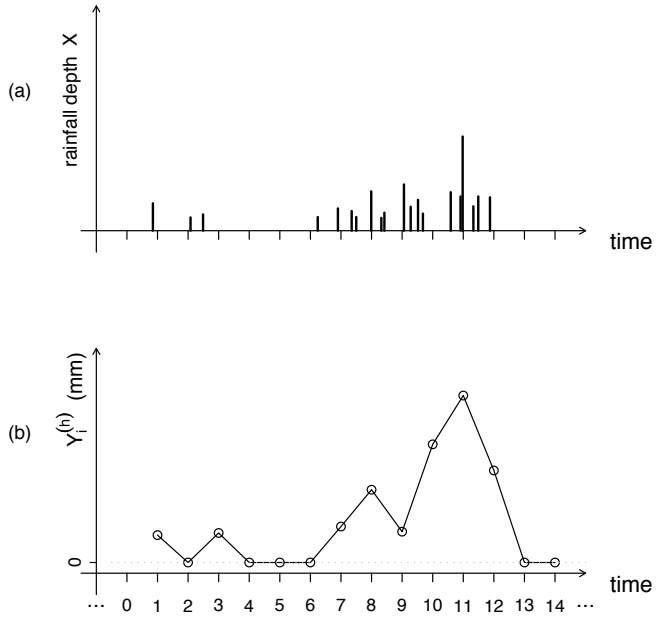

Figure 6.2: A schematic of the formation of the aggregated rainfall series: (a) the rainfall pulse depths generated from a BLP model; (b) the time plot of the aggregated rainfall time series $\{Y_i^{(h)}\}$.

The formulas of mean, variance, and autocovariance are cited from CIO2007 as given by Equations (6.2) to (6.4) below.

$$\mu(h) = \mathrm{E}\left[Y_i^{(h)}\right] = \lambda\mu_p\mu_X h = \frac{\lambda\beta\xi\mu_X h}{\gamma(\gamma+\eta)} \tag{6.2}$$

$$\mathrm{Var}\left[Y_i^{(h)}\right] = \lambda\mu_p\mathrm{E}(X^2)h + 2A\mu_X^2\phi(\gamma) + 2[B_1\mathrm{E}(X_{ijk}X_{ij\ell}) - B_2\mu_X^2]\phi(\gamma+\eta) \tag{6.3}$$

$$\mathrm{Cov}\left[Y_i^{(h)}, Y_{i+k}^{(h)}\right] = A\mu_X^2\psi(\gamma) + [B_1\mathrm{E}(X_{ijk}X_{ij\ell}) - B_2\mu_X^2]\psi(\gamma+\eta) \quad k \geq 1 \tag{6.4}$$

where $\mu_p = \beta\xi\{\gamma(\gamma+\eta)\}^{-1}$, $A = \lambda\mu_p\beta\xi/\eta$, $B_1 = \lambda\mu_p\xi$, $B_2 = A\gamma/(\gamma+2\eta)$, and $\phi(\alpha) = (e^{-\alpha h} - 1 + \alpha h)/\alpha^2$, $\psi(\alpha) = e^{-\alpha(k-1)h}(1-e^{-\alpha h})^2/\alpha^2$. The proportion dry $\mathrm{P}\left[Y_i^{(h)} = 0\right]$ is expressed in the following equation:

$$\begin{aligned} -\lambda^{-1}\log\ \mathrm{P}\left[Y_i^{(h)} = 0\right] &= \int_0^\infty \mathrm{d}x[e^{-\gamma x} - f(h,x) - \gamma\int_0^h \mathrm{d}u f(u,x)] \\ &\quad + \int_0^h \mathrm{d}x[1 - g(x) - \gamma\int_0^x \mathrm{d}u g(u)] \end{aligned} \tag{6.5}$$

where the functions f and g are given by:

$$\left.\begin{aligned} -\log\ f(u,x) &= \gamma(x+u) + \tfrac{\beta\xi}{\eta(\eta+\xi)}(1-e^{-\eta x})(1-e^{-(\eta+\xi)u}) \\ &\quad + \tfrac{\beta u\xi}{\eta+\xi} - \tfrac{\beta\xi}{(\eta+\xi)^2}(1-e^{-(\eta+\xi)u}) \\ -\log\ g(u) &= \gamma u + \tfrac{\beta u\xi}{\eta+\xi} - \tfrac{\beta\xi}{(\eta+\xi)^2}(1-e^{-(\eta+\xi)u}). \end{aligned}\right\} \tag{6.6}$$

The third moment formula is reproduced as Equation (B.12) in Appendix B.

In the above, the pulse depths may be taken to be independent exponential (θ) random variables, so that $\mathrm{E}(X_{ijk}) = \mu_X = \theta$, $\mathrm{E}(X_{ijk}X_{ij\ell}) = \mu_X^2 = \theta^2$, $\mathrm{E}(X_{ijk}X_{ij\ell}X_{ijn}) = \mu_X^3 = \theta^3$, and $\mathrm{E}(X_{ijk}^2 X_{ij\ell}) = 2\theta^3$ (expressions containing 3 pulse depths are needed in the third moment function). In the model specification that follows, the fitted BLP models obtained using this is referred to as "FIT-O", which stands for "original" fit (or "zero-change") since it is the model specification used in CIO2007.

However, in the model formulation, pulse depths may be dependent. A straightforward way to model this is to take pulse depths within the same cell to be exponential random variables with a mean that varies from cell to cell. This can be achieved by taking the density function of X_{ijk} to be $e^{-x/\mu_j}/\mu_j$ $(x > 0)$ where the mean of pulse depths in the jth cell, μ_j, is an exponential random variable with density $e^{-\mu_j/\theta}/\theta$ $(\mu_j > 0)$. By integrating over the conditional distribution, we can derive $\mathrm{E}(X_{ijk})$, $\mathrm{E}(X_{ijk}X_{ij\ell})$, $\mathrm{E}(X_{ijk}X_{ij\ell}X_{ijn})$, and $\mathrm{E}(X_{ijk}^2 X_{ij\ell})$, and the results can be substituted into the model equations above. Fitted BLP models obtained using this "conditional" exponential distribution will be referred to as "FIT-C". When the proportion dry is also included in the fitting procedure, the model specification will be referred to as "FIT-C-PD".

In a BLP model, the rainfall process is treated as a continuous time process and then aggregated to a discrete time series $\{Y_i^{(h)}\}$ as shown in Equation (6.1) and Figure 6.2. The same counting process structure for pulse occurrences $N(t)$ is assumed for both FIT-O and FIT-C. The only difference is in the characterization of the the depth of a rainfall pulse $X(t)$. Table 6.1 summarizes the differences in terms of the properties of $X(t)$.

Based on the results given in Table 6.1, it is not difficult to verify that FIT-C has introduced higher covariance and skewness given the same mean depth. Therefore, we may expect to observe some improvement in extreme value patterns from the simulation results. Note that the total number of parameters has remained unchanged with FIT-C as with FIT-O.

Table 6.1: Differences between FIT-O and FIT-C

	FIT-O	FIT-C
$\mathrm{E}(X)$	θ	θ
$\mathrm{E}(X^2)$	$2\theta^2$	$4\theta^2$
$\mathrm{E}(X^3)$	$6\theta^3$	$36\theta^3$
$\mathrm{E}(X_{ijk}X_{ij\ell})$	θ^2	$2\theta^2$
$\mathrm{E}(X_{ijk}^2X_{ij\ell})$	$2\theta^3$	$12\theta^3$
$\mathrm{E}(X_{ijk}X_{ij\ell}X_{ijn})$	θ^3	$6\theta^3$

Notation: $X(t) \equiv X \equiv X_{ij\ell}$; $\mathrm{E}(X) \equiv \mathrm{E}(X_{ijk}) \equiv \mu_X$.

The details of how the proportion dry property is included in the fitting procedure is given in the next section.

6.3 Fitting procedures

6.3.1 The original BLP model fitting procedure

For model fitting, it is convenient to define the following functions, which are formed from Equations (6.2)–(6.5):

$$\left.\begin{array}{lcl} \text{Coefficient of variation, } \nu(h) & = & \mathrm{E}\left[\{Y_j^{(h)}-\mu(h)\}^2\right]^{\frac{1}{2}}/\mu(h) \\ \text{Lag 1 autocorrelation, } \rho(h) & = & \mathrm{Corr}\left[Y_j^{(h)}, Y_{j+1}^{(h)}\right] \\ \text{Coefficient of skewness, } \kappa(h) & = & \mathrm{E}\left[\{Y_j^{(h)}-\mu(h)\}^3\right]/\mathrm{E}\left[\{Y_j^{(h)}-\mu(h)\}^2\right]^{\frac{3}{2}} \\ \text{Proportion dry, } \omega(h) & = & \mathrm{P}\left[Y_j^{(h)}=0\right] = e^{-\lambda I(\beta,\xi,\gamma,\eta)} \end{array}\right\} \tag{6.7}$$

where $I(\beta,\xi,\gamma,\eta)$ is the integral expression on the right-hand-side of equation (6.5) and

$$\mathrm{Corr}\left[Y_j^{(h)}, Y_{j+1}^{(h)}\right] = \frac{\mathrm{Cov}\left[Y_j^{(h)}, Y_{j+1}^{(h)}\right]}{\mathrm{Var}\left[Y_j^{(h)}\right]}.$$

There are many possible combinations of (6.7) that could be used to fit the BLP model to data. Denoting a model property above as f and the equivalent sample estimate as $\hat{f}$, the fitting procedure in CIO2007 is based on minimising the following sum of squares

$$\sum_{f\in F}\left\{\left(1-\frac{\hat{f}(h)}{f(h)}\right)^2+\left(1-\frac{f(h)}{\hat{f}(h)}\right)^2\right\} \tag{6.8}$$

where $F = \{\nu(h), \rho(h), \kappa(h) : h = 1/12, 1, 6, 24\}$. For a given set of sample properties $(\hat{f})$ minimising (6.8) returns a set of estimated parameters: $\{\hat{\lambda}_i, \hat{\beta}_i, \hat{\xi}_i, \hat{\gamma}_i, \hat{\eta}_i : i = 1, 2\}$ for two storm types. The scale parameter θ is then estimated from the 1 h sample mean rainfall $(\bar{x})$ using:

$$\hat{\theta} = \bar{x}\left[\frac{\hat{\lambda}_1\hat{\beta}_1\hat{\xi}_1}{\hat{\gamma}_1(\hat{\gamma}_1+\hat{\eta}_1)} + \frac{\hat{\lambda}_2\hat{\beta}_2\hat{\xi}_2}{\hat{\gamma}_2(\hat{\gamma}_2+\hat{\eta}_2)}\right]^{-1} \tag{6.9}$$

Note that the mean for the superposed BLP process is just obtained by adding the expectations in Equation (6.2) for two independent BLP processes corresponding to storms of type 1 process and type 2 process as it is specified in FIT-O. It is assumed, partly for parameter parsimony, that the distributions of the pulse depths are the same for different processes of storms. The estimation of θ in Equation (6.9), which occurs after the other parameters have been estimated, ensures the model has an exact fit to the sample mean (CIO2007).

The original BLP model is specified by assuming two types of storm processes superposed with a common mean pulse depth θ. Therefore, FIT-O has 11 parameters in total, $\{\lambda_i, \beta_i, \xi_i, \gamma_i, \eta_i, \theta : i = 1, 2\}$, and FIT-C has retained the same number of parameters. To make reference for different model specifications easy, we denote the original BLP model representing a single process of storms by FIT-O1 and the single process BLP model with conditional mean depth distribution by FIT-C1. It is clear that both FIT-O1 and FIT-C1 have six parameters in total $\{\lambda, \beta, \xi, \gamma, \eta, \theta\}$.

In (6.8), each model function (f) is divided by the equivalent value from the sample $(\hat{f})$ to ensure that each fitted property receives equal weight. The sum contains the two terms $f/\hat{f}$ and $\hat{f}/f$ to help ensure that the optimal solution is not biased (above or below the sample values) when exact fits to the estimates are not obtained.

The same fitting procedure of CIO2007, function (6.8) and Equation (6.9), is used when the pulse depths follow a conditional exponential distribution (FIT-C), and again we assume that $\theta_1 = \theta_2 = \theta$.

6.3.2 Inclusion of proportion dry property in model fitting

In order to include the proportion dry property in model fitting, we need to solve the computation problem first, because the integral expression $I(\beta, \xi, \gamma, \eta)$ involves some complicated integral calculations which can not be solved analytically. Therefore, Simpson's rule (e.g. Kreyszig (1993), p961) is applied to get a numerical solution for

$\mathrm{P}\big[Y_j^{(h)} = 0\big]$. Details and verification of the numerical solution in calculating the proportion dry for BLP models can be found in Appendix B.

For a superposed BLP model of two independent processes, $\mathrm{P}\big[Y_j^{(h)} = 0\big] = \mathrm{P}_1\big[Y_j^{(h)} = 0\big]\,\mathrm{P}_2\big[Y_j^{(h)} = 0\big]$, where $\mathrm{P}_1\big[Y_j^{(h)} = 0\big] = e^{-\lambda_1 I(\beta_1,\xi_1,\gamma_1,\eta_1)}$ and $\mathrm{P}_2\big[Y_j^{(h)} = 0\big] = e^{-\lambda_2 I(\beta_2,\xi_2,\gamma_2,\eta_2)}$. Once proportion dry values can be calculated for a fitted model, one obvious solution to include the proportion dry property in model fitting is to add the proportion dry property terms to the objective function (6.8) with $F = \{\nu(h), \rho(h), \kappa(h), \omega(h) : h = 1/12, 1, 6, 24\}$. Alternatively, a fitting procedure for FIT-C-PD is now to minimize the objective function (6.8) subject to the following condition:

$$\log \hat{\omega}(h) = -\hat{\lambda}_1 I(\hat{\beta}_1, \hat{\xi}_1, \hat{\gamma}_1, \hat{\eta}_1) - \hat{\lambda}_2 I(\hat{\beta}_2, \hat{\xi}_2, \hat{\gamma}_2, \hat{\eta}_2) \tag{6.10}$$

where $h = 24$, i.e. the 24 hours sample proportion dry $\hat{\omega}(24)$ is fitted. The constraint (6.10) ensures an exact match to the sample proportion dry. The inclusion of the proportion dry property (6.10) reduces the total number of parameters by one, in the minimization routine, because $\hat{\lambda}_1$ or $\hat{\lambda}_2$ can be expressed by other parameters in the model through Equation (6.10). The proportion dry is independent of the distribution of the pulse depth, and, furthermore, the sample estimate of the proportion dry tends to have a smaller standard error when compared to high-order moments. Hence, the addition of the proportion dry property is likely to lead to a more robust fitting procedure.

6.4 Empirical Analysis

6.4.1 Data and parameter estimates

Following CIO2007, the sample estimates ($\hat{f}$) in (6.8) were calculated for each calendar month in the series, by pooling all available data for the month over the 60-year record. For each month, the parameter estimates were obtained using the three fitting procedures: FIT-O, FIT-C and FIT-C-PD. The resulting estimates are shown in Table 6.2.

Eleven parameter estimates ($\hat{\lambda}_1, \hat{\beta}_1, \hat{\xi}_1, \hat{\gamma}_1, \hat{\eta}_1, \hat{\lambda}_2, \hat{\beta}_2, \hat{\xi}_2, \hat{\gamma}_2, \hat{\eta}_2, \hat{\theta}$) are listed by columns from left to right in Table 6.2. There are three sets of parameter estimates (i.e. three rows, one for each of the three model specifications in the order: FIT-O, FIT-C and FIT-C-PD) for each month.

Table 6.2: Parameter estimates of three BLP model specifications for the Kelburn rainfall series. The units are hour^{-1} for all estimates except θ which has units mm.

m	$\hat{\lambda}_1$	$\hat{\beta}_1$	$\hat{\xi}_1$	$\hat{\gamma}_1$	$\hat{\eta}_1$	$\hat{\lambda}_2$	$\hat{\beta}_2$	$\hat{\xi}_2$	$\hat{\gamma}_2$	$\hat{\eta}_2$	$\hat{\theta}$
1	0.00674	0.119	491	0.0421	0.709	0.000242	0.938	4960	0.181	6.46	0.00729
	0.00573	0.232	875	0.0469	0.500	0.00571	2.66	5610	1.64	5.38	0.00186
	0.0211	0.424	1770	0.244	0.366	0.000112	0.840	14600*	0.00983	6.36	0.000760
2	0.00798	0.531	181	0.110	1.26	0.000167	3.94	1300	0.308	8.84	0.0183
	0.00776	1.06	256	0.107	1.37	0.000262	13.7	1240	0.299	12.5	0.00685
	0.0308	0.192	262	0.283	0.103	0.00117	1.32	1310*	0.0656	5.04	0.00487
3	0.0117	0.344	263	0.113	0.885	0.000681	2.80	1960	0.429	7.00	0.0104
	0.0109	0.668	272	0.107	0.888	0.00117	6.18	1450	0.389	6.83	0.00494
	0.0270	0.400	210	0.268	0.258	0.000866	0.646	1590*	0.0499	4.76	0.00556
4	0.0123	0.233	276	0.0794	0.877	0.000443	1.21	2400	0.155	6.24	0.0104
	0.0131	0.348	366	0.0758	0.844	0.000743	2.87	2010	0.130	6.51	0.00422
	0.0329	0.117	503	0.112	0.417	0.000963	1.78	2760*	0.0748	5.70	0.00279
5	0.0180	0.146	195	0.0716	0.464	0.000986	0.564	1910	0.105	3.60	0.00933
	0.0211	0.451	178	0.114	0.598	0.000977	1.16	1150	0.0712	3.44	0.00577
	0.0340	0.213	178	0.134	0.320	0.00129	1.03	1160*	0.0627	3.36	0.00532
6	0.0186	0.075	90.8	0.0444	0.252	0.00151	0.434	1290	0.186	2.78	0.0162
	0.0186	0.332	80.8	0.0719	0.366	0.00156	0.552	858	0.0946	2.45	0.00952
	0.0393	0.388	56.8	0.176	0.139	0.000626	0.255	853*	0.0171	2.28	0.00943
7	0.0133	0.203	399	0.0516	0.503	0.00185	1.96	3140	0.188	5.70	0.00345
	0.0126	0.419	209	0.0509	0.592	0.00218	4.23	1120	0.174	5.25	0.00370
	0.0535	0.223	145	0.211	0.112	0.000771	0.846	1050*	0.0255	4.02	0.00517
8	0.0390	0.550	546	0.225	0.794	0.000915	2.19	4280	0.108	6.05	0.00231
	0.0342	1.58	281	0.233	1.19	0.000826	5.41	1550	0.107	5.90	0.00261
	0.0312	2.16	225	0.239	1.37	0.000688	5.21	1350*	0.103	5.75	0.00313
9	0.0159	0.337	198	0.0932	0.925	0.000889	2.59	1230	0.428	5.62	0.0101
	0.0152	0.690	242	0.0937	0.995	0.00117	5.79	1080	0.401	5.79	0.00444
	0.0307	0.361	203	0.179	0.377	0.00113	0.612	1140*	0.0361	4.01	0.00442
10	0.0138	0.351	430	0.0850	0.847	0.00103	2.59	2680	0.250	6.56	0.00448
	0.0190	0.265	328	0.107	0.333	0.00181	0.527	1800	0.0440	4.63	0.00314
	0.0369	0.260	395	0.203	0.212	0.00129	1.05	2170*	0.0448	4.78	0.00233
11	0.0136	0.196	133	0.0742	0.931	0.000140	3.43	1580	0.601	7.89	0.0245
	0.0128	0.406	149	0.0712	0.916	0.000440	7.15	1140	0.542	7.80	0.0102
	0.0356	0.0850	134	0.0857	0.389	0.000158	0.333	1110*	0.0104	4.46	0.0108
12	0.0132	0.730	470	0.166	1.51	0.000166	2.42	3810	0.160	6.66	0.00605
	0.0135	1.37	483	0.165	1.59	0.000194	6.58	2750	0.162	7.07	0.00319
	0.0214	1.27	1610	0.239	1.21	0.000850	5.03	9350*	0.151	5.79	0.000628

* For FIT-C-PD, ξ_2 is not a free parameter. $\xi_2[i] = \xi_1[i] \times ratio[i]$, $i = 1, ..., 12$ (month), where $ratio$ is a constant vector: $ratio$ = (8.2, 5.0, 7.6, 5.5, 6.5, 15.0, 7.2, 6.0, 5.6, 5.5, 8.3, 5.8).

Since a BLP model is conceptually a physical process based rainfall model, some intended physical meanings may be attached to model parameters for interpretation of what we may learn from the estimation results. Table 6.2 shows that $\hat{\lambda}_1 > \hat{\lambda}_2$ and $\hat{\xi}_1 \ll \hat{\xi}_2$ for all three models. This implies that process 2 has a lower storm generation rate with a much more frequent pulse generation than process 1. Because the mean depth $\hat{\theta}$, is the same for both processes, more frequent pulse generation is equivalent to higher rainfall intensity. Furthermore, in most months, process 1 storms last longer (on average) than process 2 storms (i.e. $\hat{\gamma}_1^{-1} > \hat{\gamma}_2^{-1}$) with FIT-O. This seems to imply that we may match the meteorological definition of stratiform rainfalls with process 1 and convective rainfalls with process 2. However, if we look at the FIT-C-PD results, process 1 storms last consistently shorter (on average) than process 2 storms (i.e. $\hat{\gamma}_1^{-1} < \hat{\gamma}_2^{-1}$). FIT-C sits somewhere in between FIT-O and FIT-C-PD. Therefore, the two distinct processes specified in a BLP model are rather arbitrary in a meteorological sense.

The parameter estimates $\hat{\lambda}_i$ and $\hat{\gamma}_i$ (i = 1,2) seem significantly correlated (Table 6.4). Note that FIT-C-PD consistently has higher $\hat{\lambda}_1$ values and lower $\hat{\gamma}_1^{-1}$ values than FIT-O and FIT-C. This implies that FIT-C-PD describes process 1 as storms that occurs more often but where each storm has a shorter mean duration. Although the resulting parameter estimates with FIT-C-PD are very different from FIT-O and FIT-C, it can be shown that the mean storm hours per month are reasonably close for all three models. Furthermore, they all have the same pattern in that the mean storm hours per month of process 1 is much larger than process 2 and the minimum of $\hat{\lambda}_1$ occurs in January.

The parameter estimates $\hat{\beta}_i$ and $\hat{\eta}_i$ (i = 1,2) show strong positive correlations (> 0.8). This implies that a higher cell generation rate (larger $\hat{\beta}_i$) tends to be accompanied by a shorter cell duration (smaller $\hat{\eta}_i^{-1}$). The parameters $\hat{\eta}_1$ and $\hat{\eta}_2$ are highly positively correlated and $\hat{\eta}_1^{-1} \gg \hat{\eta}_2^{-1}$. This means that process 1 has a much longer mean cell duration than process 2. The patterns are similar for all three model specifications but FIT-C-PD has a much bigger range of variation in terms of the ratio $\hat{\eta}_2/\hat{\eta}_1$.

Since $\hat{\gamma}^{-1}$ is intended to account for the mean storm lifetime and $\hat{\eta}^{-1}$ for mean cell duration, intuitively we expect to see that $\hat{\gamma}_i^{-1} > \hat{\eta}_i^{-1}$ (i = 1,2). However, we notice that $\hat{\gamma}_1^{-1} < \hat{\eta}_1^{-1}$ for February, March, June, and July with FIT-C-PD. A BLP model assumes that the process of pulses terminates with the cell or storm lifetime, whichever is the sooner. This setting has the effect of introducing dependence between cells within storms. Therefore, when $\hat{\gamma}_1^{-1} < \hat{\eta}_1^{-1}$ occurs, termination of pulses process before the end of a cell duration is highly likely to happen.

The parameter estimates $\hat{\xi}_1$ and $\hat{\xi}_2$ are so strongly positively correlated (> 0.93) that it seems that we should assume $\hat{\xi}_2 = \text{constant} \times \hat{\xi}_1$ in model estimation. The lowest rainfall intensity occurs in June (New Zealand winter) and the highest intensity occurs in January (New Zealand summer) for both processes.

By looking at the $\hat{\theta}$ values, which represent the mean depths associated with pulses, they are very small (less than 0.01 mm in most months) which suggests that BLP models may be able to capture a rainfall process even at droplets level. The values of $\hat{\theta}$ with FIT-C and FIT-C-PD are consistently much less than FIT-O.

In the model estimation process, the convergence status can be achieved in minimising (6.8) with different parameter sets. Different initial values or changes in parameter space limits can all result in different parameter estimates. Over-parametrisation occurs when the number of independent estimated parameters is more than the number of independent sample properties, and is likely to be the cause for having more than one sets of 'optimal' parameter estimates for the same objective function.

It is worth noting that there are substantial differences between the parameter estimates obtained from FIT-O and FIT-C. But the trade-off between parameters means that these do not necessarily reflect big changes in the model properties as will be shown in the following sections.

6.4.2 Moments

All figures in this chapter have the same layout for comparison between three different model specifications: FIT-O (left column plots), FIT-C (middle column plots) and FIT-C-PD (right column plots). Historical values are always represented by points (crosses) and model fitted values or simulation results represented by lines. Based on the estimated model parameters, 200 simulation samples are generated from FIT-O, FIT-C and FIT-C-PD respectively. Model performance in fitting the moments of rainfall totals, annual rainfall maxima patterns, and reproduction of proportion dry patterns are compared based on the simulation results. A parametric bootstrap approach (e.g. see Efron and Tibshirani, 1993, p53) is adopted for statistical reference. With respect to each observed sample statistic, there are 200 corresponding simulated values calculated from the simulation samples which are generated from the fitted model. The median, 2.5%, and 97.5% quantiles are calculated from each set of 200 simulated values so that a 95% parametric bootstrap confidence band can be constructed. The medians are represented by solid lines and the 95% quantile confidence bands are formed by two

dashed lines. If the observed statistics are well located in the middle of the confidence band, we conclude that the observed values may be considered as a realization of the random processes specified by the fitted model.

Figure 6.3 to Figure 6.8 show the comparisons of fitting the moment properties of rainfall totals. Figure 6.3 to Figure 6.6 show the simulation results of those statistical properties used in model fitting. All means have an exact fit and any deviation from the fitted model median lines is due to sampling variation. Most of CV fits are almost exact with only a few historical points deviating slightly from the fitted medians, e.g. January, February, and December with FIT-O at 5 min and 1 hour aggregation levels (Figure 6.3 and Figure 6.4) and January, February with FIT-C-PD at 24 hours level (Figure 6.6), etc.

The skewness plots in Figure 6.3 to Figure 6.5 indicate that there is some underestimation bias in fitting the sample skewness, i.e. fitted median lines are consistently under the historical skewness values with all three procedures. A reason for this is that sample time series observations are not independent so that the degrees of freedom used in calculating the sample skewness statistics are inflated. The exact fit of proportion dry at 24 hours level helps to adjust the error in skewness (Figure 6.6).

The proportion dry constraint (6.10) causes slight underestimation in lag 1 autocorrelation fitting at the 6 and 24 hours levels (Figure 6.5 and Figure 6.6). Most historical autocorrelations are well located within the 95% quantile confidence band. Only the January historical values at the 5-min level for FIT-C and FIT-C-PD are slightly deviated from the medians (Figure 6.3).

The statistical properties at the 30-min level (Figure 6.7) and the kurtosis at different aggregation levels (Figure 6.8) have been matched well, without being directly fitted. Most of the historical values are well located within the confidence bands. Only a few points are sitting on the boundary of the confidence bands or deviate significantly from the median lines, e.g. the lag 1 autocorrelation (Figure 6.7), and the kurtosis at the 5-min and 1 hour levels (Figure 6.8).

Overall, we conclude that all three models reproduce the moment properties well and that the sample moments and autocorrelations may be viewed as realisations from the fitted models. No notable differences are identified between three model specifications. The changes in model specification and fitting procedures have caused no noticeable negative impact in fitting the sample properties.

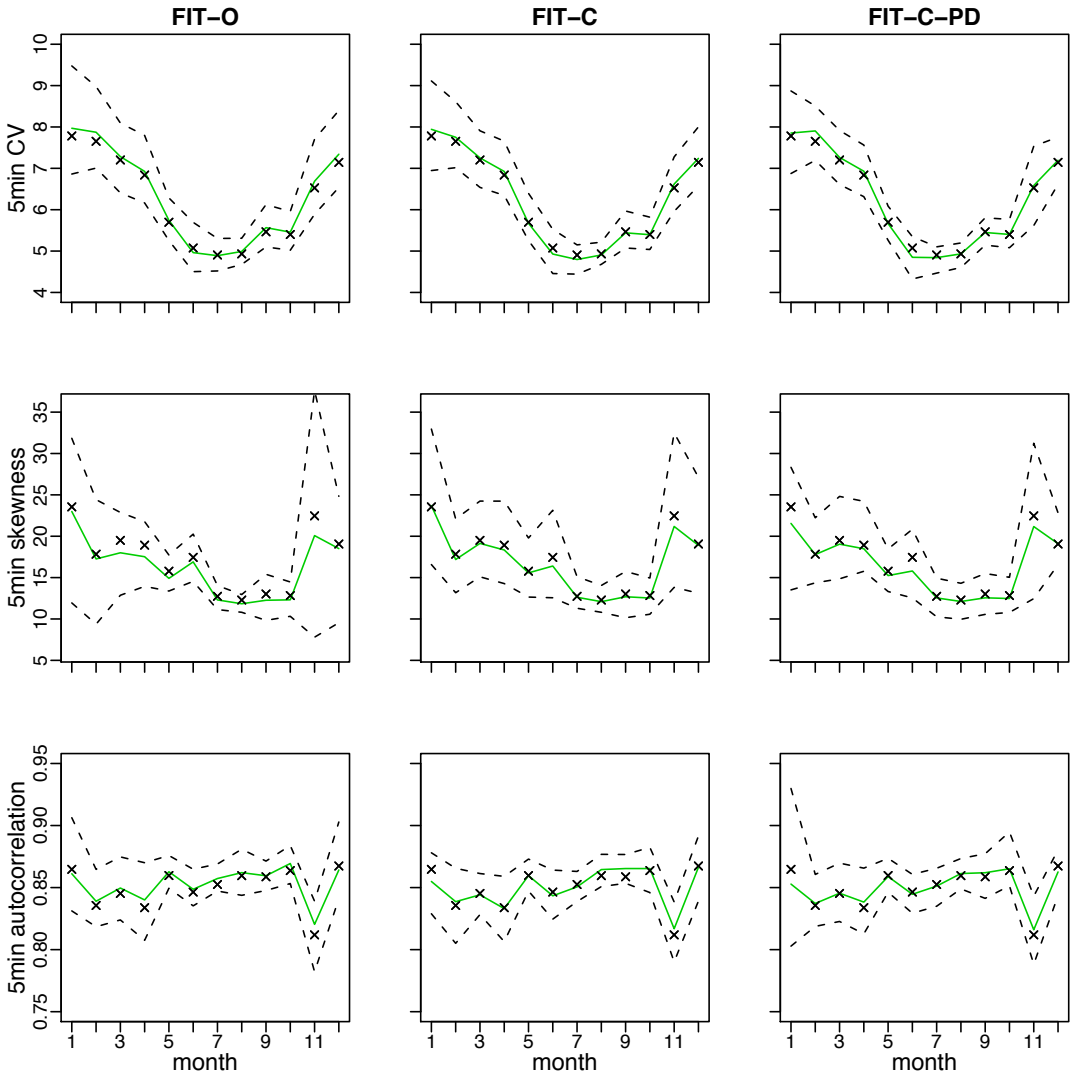

Figure 6.3: Model comparisons (properties used in model fitting): coefficient of variation, skewness, and lag 1 autocorrelation at the 5-min aggregation level.

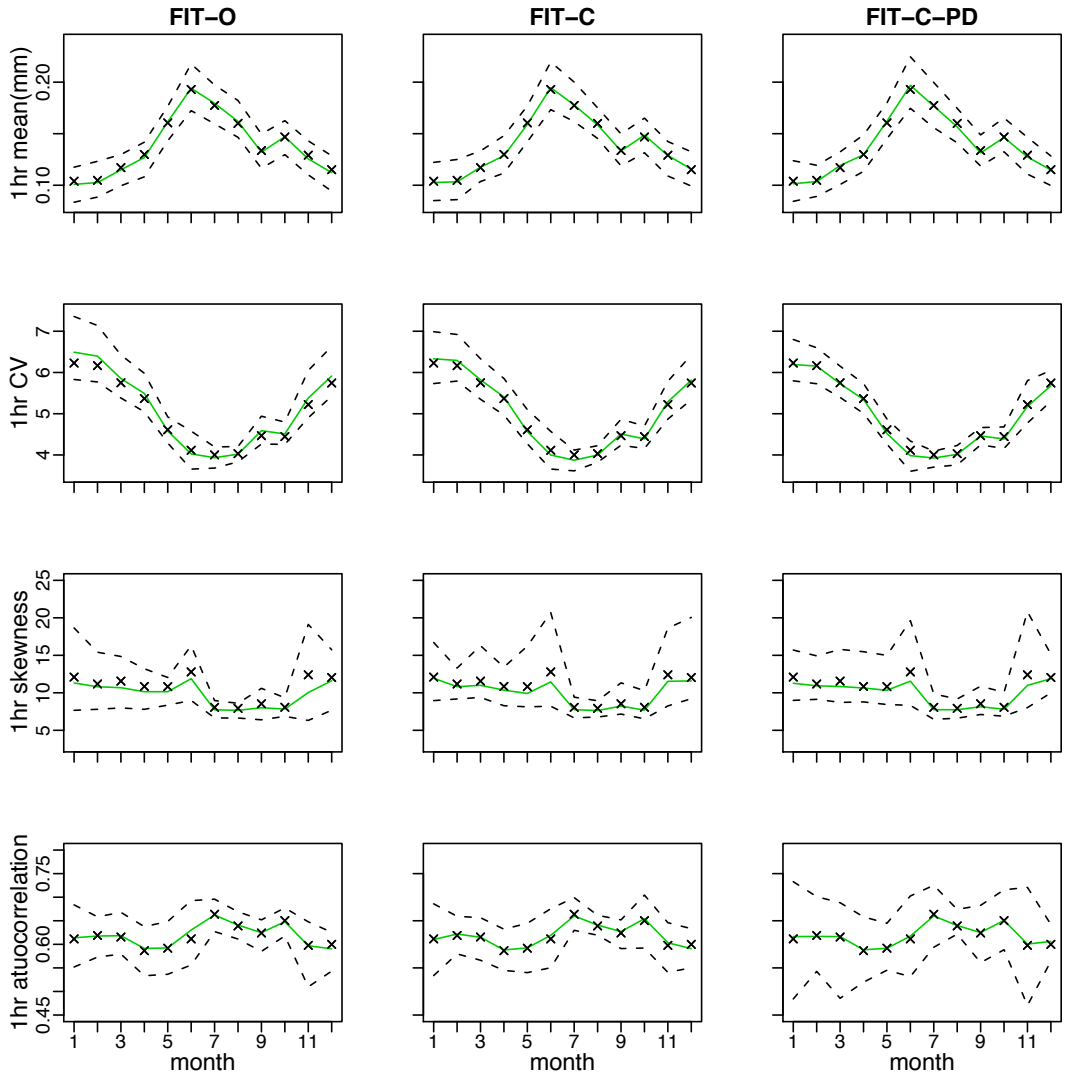

Figure 6.4: Model comparison (properties used in model fitting): mean, coefficient of variation, skewness, and lag 1 autocorrelation at the 1 hour aggregation level.

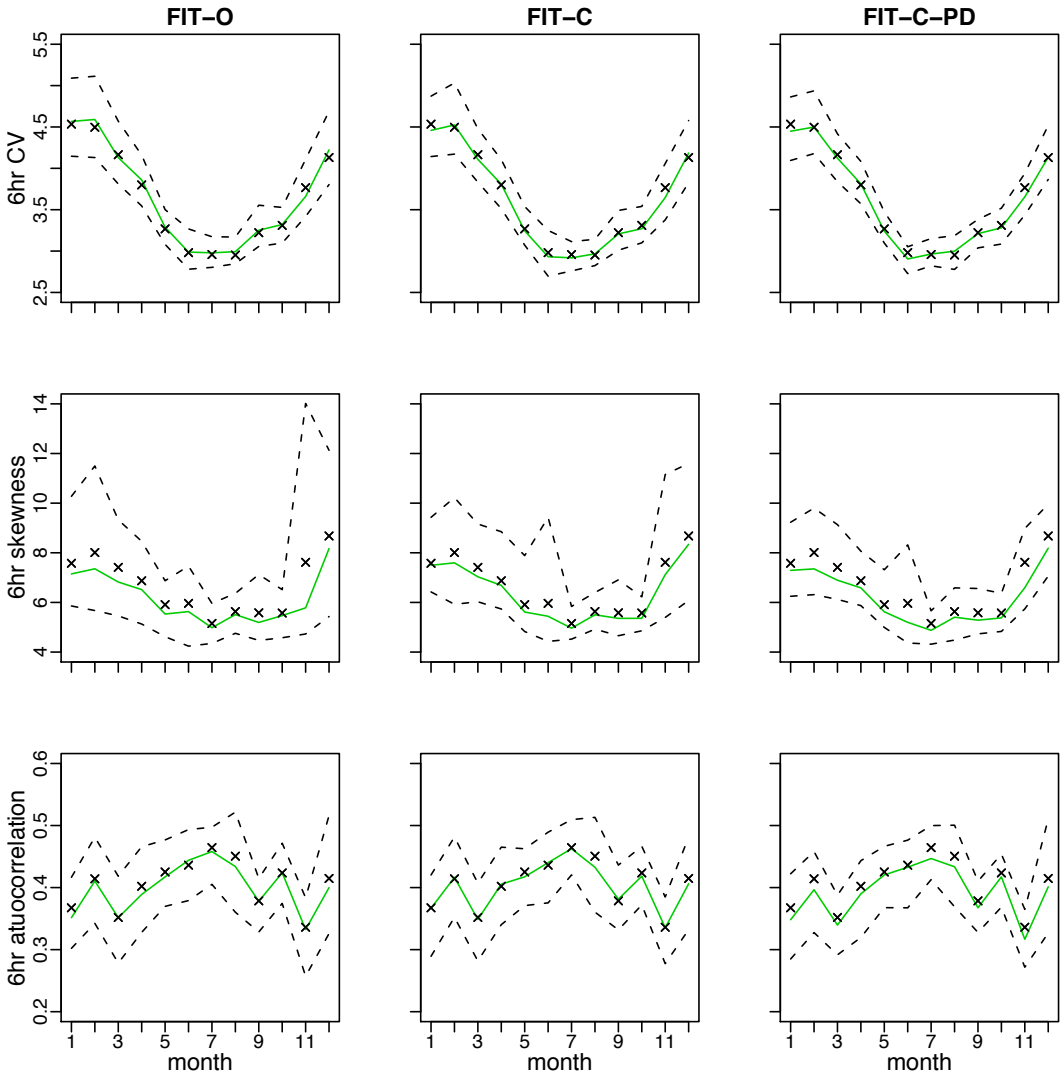

Figure 6.5: Model comparison (properties used in model fitting): coefficient of variation, skewness, and lag 1 autocorrelation at the 6 hours aggregation level.

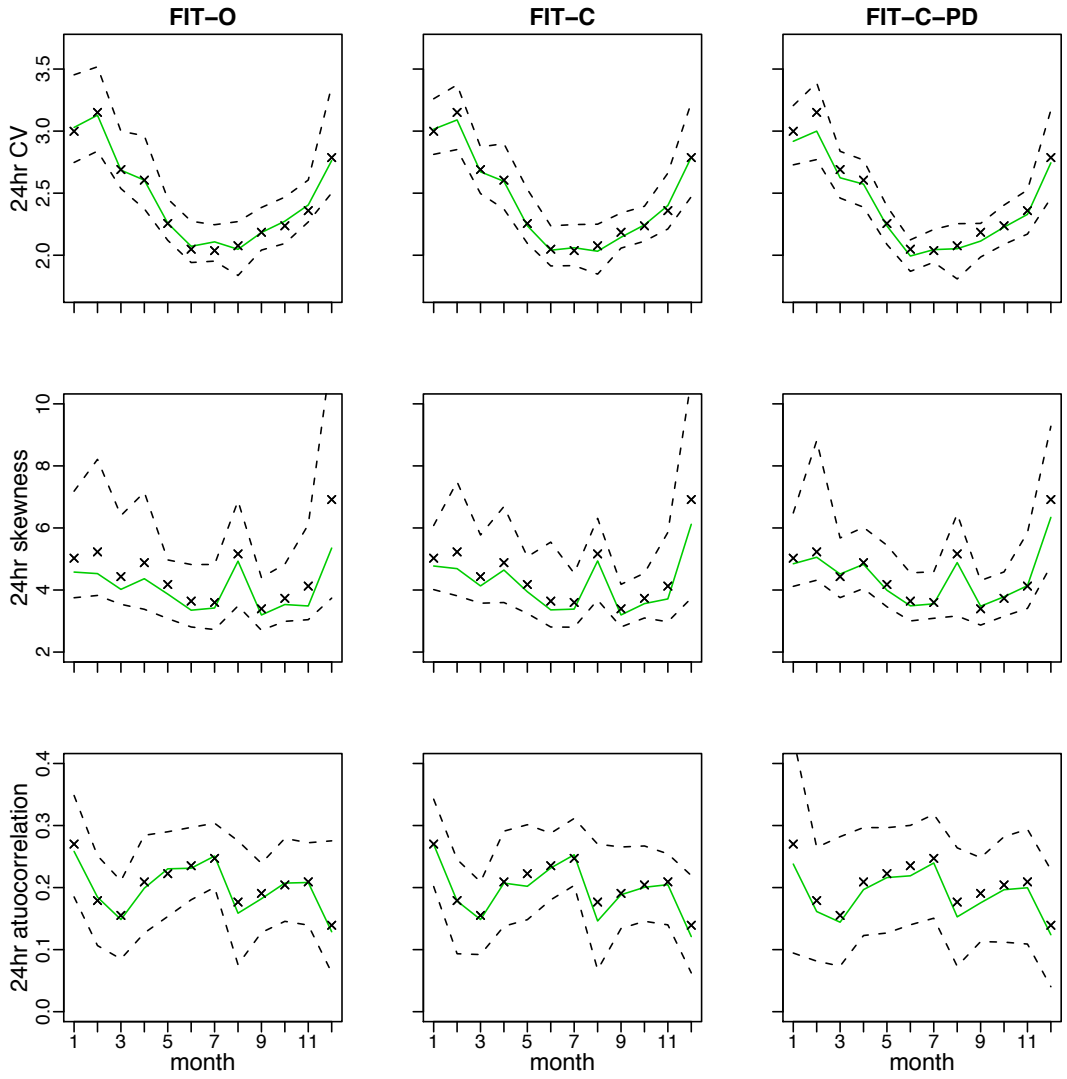

Figure 6.6: Model comparison (properties used in model fitting): coefficient of variation, skewness, and lag 1 autocorrelation at the 24 hours aggregation level.

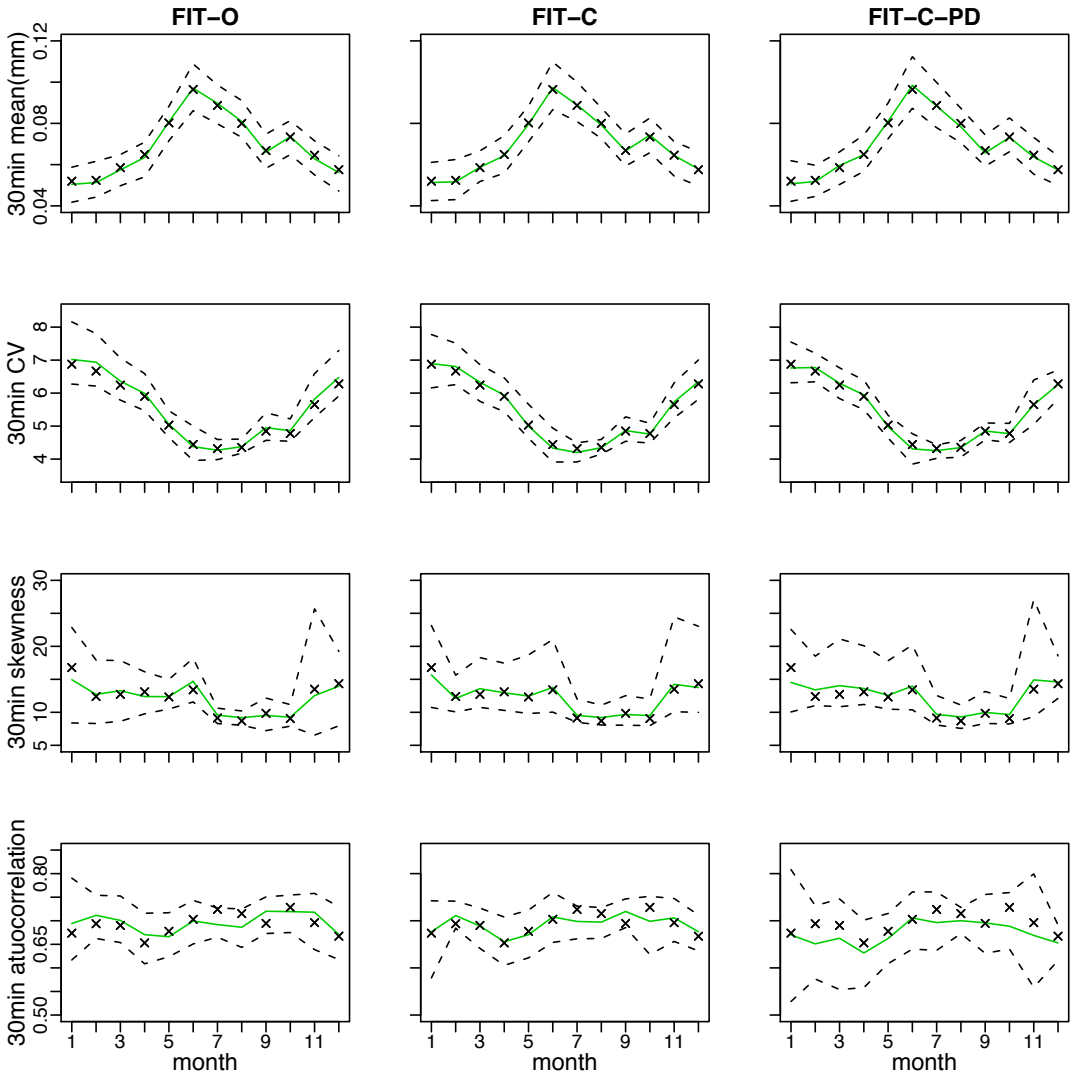

Figure 6.7: Model comparison (properties not used in model fitting): mean, coefficient of variation, skewness, and lag 1 autocorrelation at the 30-min aggregation level.

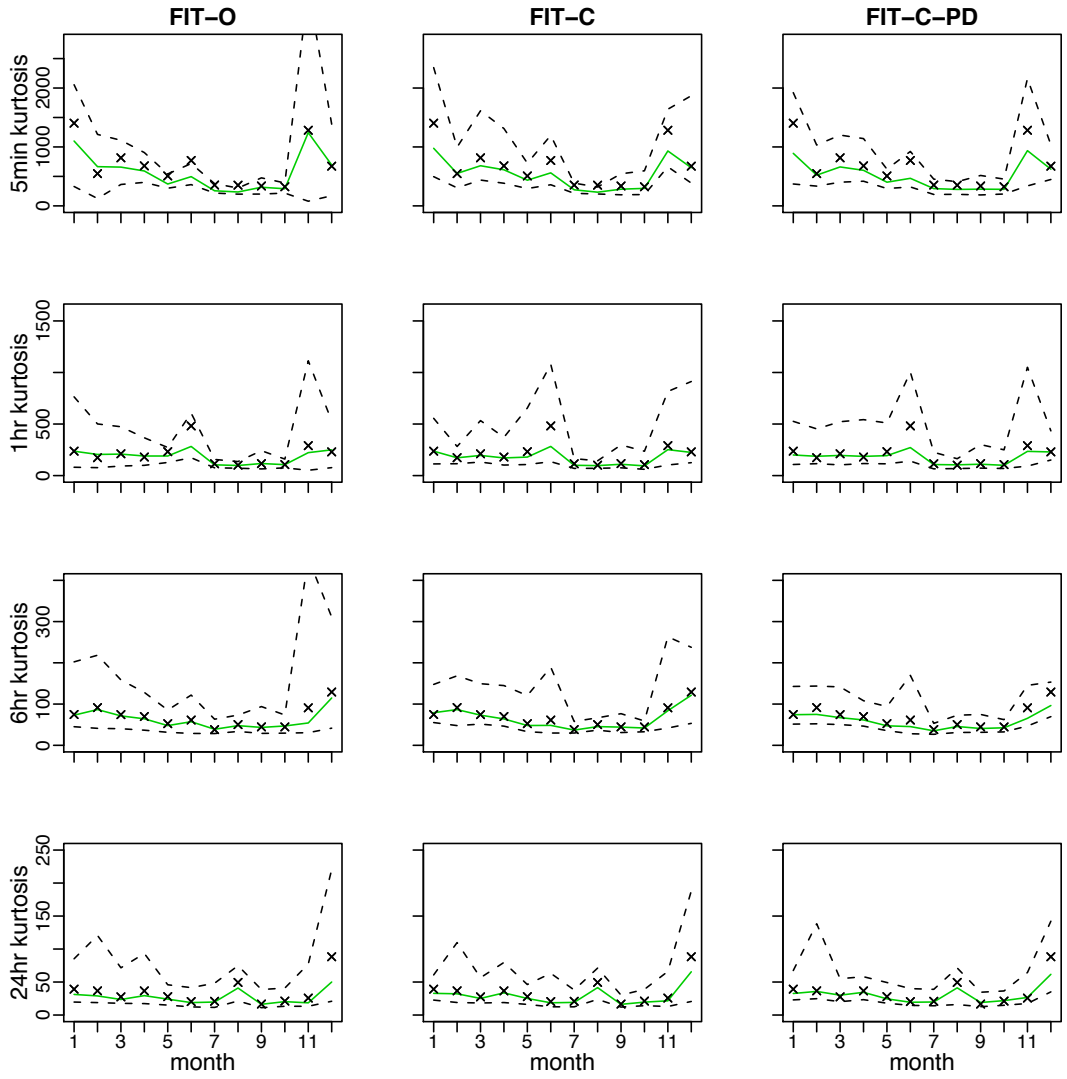

Figure 6.8: Model comparison (properties not used in model fitting): kurtosis at the 5-min, 1, 6, and 24 hours aggregation levels.

6.4.3 Extreme values

A BLP model is fitted to data without directly including the sample extreme value statistics in the fitting procedure. Extreme values of rainfall totals, e.g. annual maxima extracted at different time scale aggregation levels plotted against the reduced Gumbel variate, are compared in this section as follows. Figure 6.9 and Figure 6.10 present the annual maxima plots at five different aggregation time scales, 5-min, 1, 6, 12, and 24 hours. Figure 6.11 through to Figure 6.14 compare the annual maxima patterns for four individual months, December and January which are two summer months in New Zealand, and June and July for two winter months. In each figure, the 60 observed annual maxima (one for each year of the 60-year sampling period) are ordered and then plotted against the reduced Gumbel variate. The reduced Gumbel variate is calculated by $-\log(-\log((i-0.35)/60)), \quad i = 1, 2, ..., 60$. This procedure is equivalent to plotting the ordered annual maxima against $(i-0.35)/60, \quad i = 1, 2, ..., 60$ on Gumbel probability paper. If the scatter plot shows a linear pattern, it indicates that the observed extreme values can be modelled by a Gumbel distribution.[1] As we have described in the beginning of Section 6.4.2, the observed historical values represented by points (crosses) are plotted against the median lines and the 95% quantile confidence band obtained from the 200 simulation samples. The horizontal scales on the bottom left part of the plotting area indicate the corresponding *return period* positions. The return period is a concept defined to represent the frequency of the rainfall in hydrological analysis. An annual maximum rainfall event has a return period of T years if it has an equal or greater magnitude once every T years on average (Butler and Davies, 2004). For example, a rainfall event that a threshold rainfall level is exceeded on average five times in 100 years has a return period of 20 years. Annual maximum rainfall events are normally used to determine return period based on the assumption that the largest event in one year is statistically independent of the largest event in any other year.

It is obvious that, overall, FIT-C outperforms FIT-O in fitting the extreme values. With only one historical value marginally outside the 95% confidence band at 5-min level, we may conclude that FIT-C has fitted extreme values adequately at all aggregation time scales examined (Figure 6.9 and Figure 6.10).

FIT-O fits extreme values at the 5-min level slightly better than FIT-C-PD while

[1] For more about the reduced Gumbel variate, see e.g. Metcalfe, 1997, p92.

both procedures slightly underestimate the historical values (Figure 6.9). FIT-O shows a slight overestimation of historical extreme values at 12 hours aggregation level over the return period range of $5 < T < 50$ years (Figure 6.10). FIT-C-PD fits extreme values very well at 1, 6, 12, and 24 hours aggregation levels (Figure 6.9 and Figure 6.10).

Next we extend our study to investigate if there is any improvement using FIT-C or FIT-C-PD when the annual maxima patterns are broken down by month.

Preliminary research results have shown lack-of-fit, when using FIT-O, to extreme values for the individual months at finer aggregation levels , e.g. 5-min and 1 hour. Figure 6.11 to Figure 6.14 present the comparisons for 5-min and 1 hour levels. The results show that FIT-C has a clear improvement over FIT-O in fitting the individual month annual maxima. The improvement for the summer season is more significant than the winter season. Note that there are some 'step' patterns shown in the fitted median lines which matches the worst cases of deviation from the observed extreme values, e.g. in December for FIT-O and FIT-C (Figure 6.11), in January for FIT-O and FIT-C-PD (Figure 6.12). Each of these 'step' patterns matches with a very small $\hat{\lambda}_2$ value, less than 0.0003, and these all happened in the summer season (Table 6.2). Based on the estimated parameters, process 2 has a less frequent storm occurrences ($\hat{\lambda}_1 > \hat{\lambda}_2$) but higher rainfall intensity ($\hat{\xi}_1\hat{\theta} < \hat{\xi}_2\hat{\theta}$). Therefore, the extreme rainfalls are mostly affected by the process 2 rainfalls. It is the very small storm origins generation rate $\hat{\lambda}_2$, 3.3 storms every 10000 hours on average, caused the 'steps' of the fitted curves – either there is no rainfall (zero storm) or there is a lot (heavy rain when a storm is generated). From Table 6.2, we note that FIT-C has higher $\hat{\lambda}_2$ values than FIT-O in most months.

There is no obvious improvement when using FIT-C-PD instead of FIT-C. Overall, we may conclude that none of the models fit the 5-min monthly annual maxima completely adequate, but that FIT-C provides best overall fit.

6.4.4 Proportion dry

Historical proportion dry values (points) are plotted against fitted proportion dry values (solid lines) at three aggregation levels , $h = 1/12, 1, 24$ hours, which are shown in Figure 6.15. It is clear that both FIT-C and FIT-C-PD fit the proportion dry patterns much better than FIT-O in all cases examined. In particular, proportion dry at 24 hours has a perfect fit with FIT-C-PD because $\hat{\mathrm{P}}(Y_i^{(24)} = 0)$ is used as a constraint in the fitting procedure (see Equations (6.5) and (6.7)).

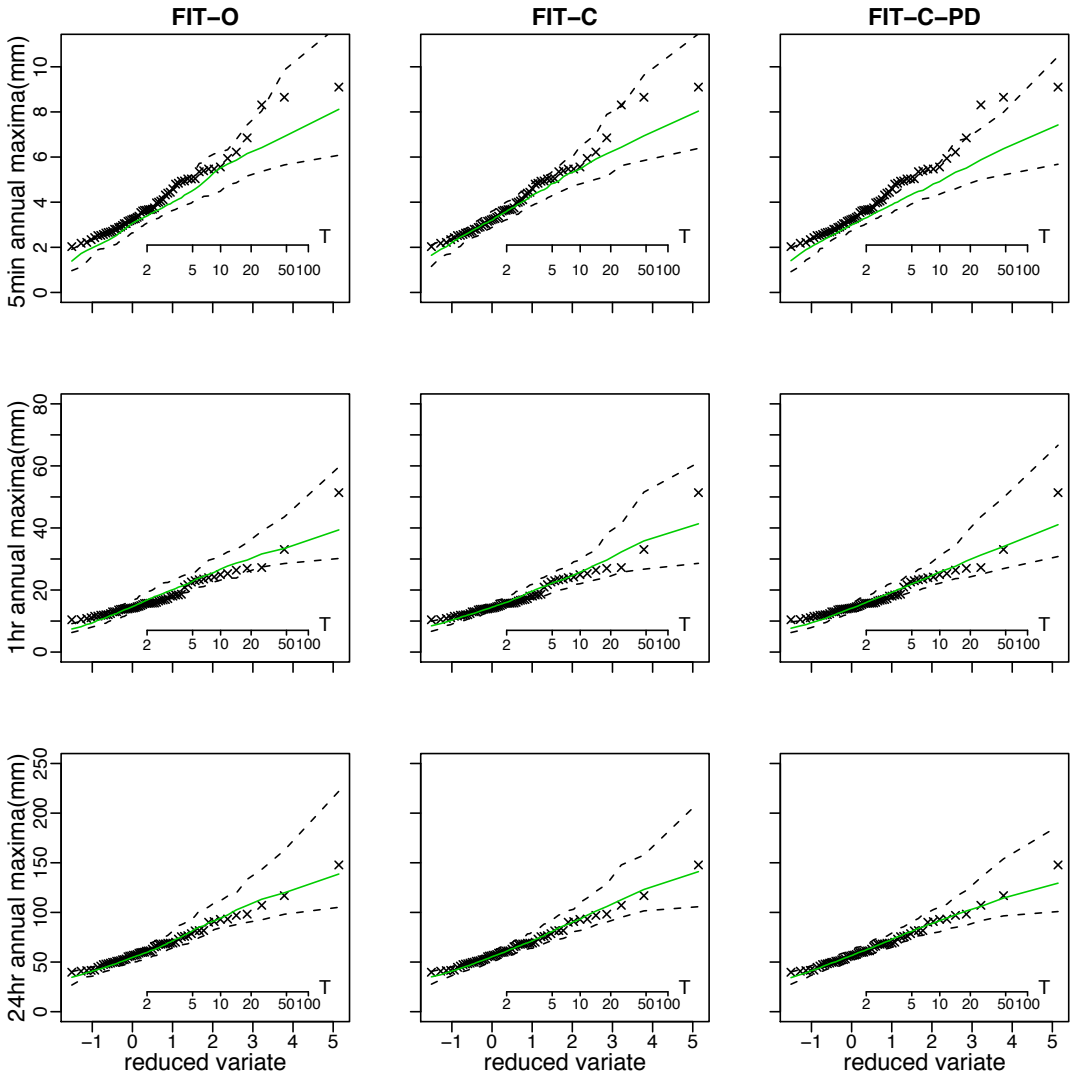

Figure 6.9: Model comparison: extreme value plots at the 5-min, 1, and 24 hours aggregation levels.

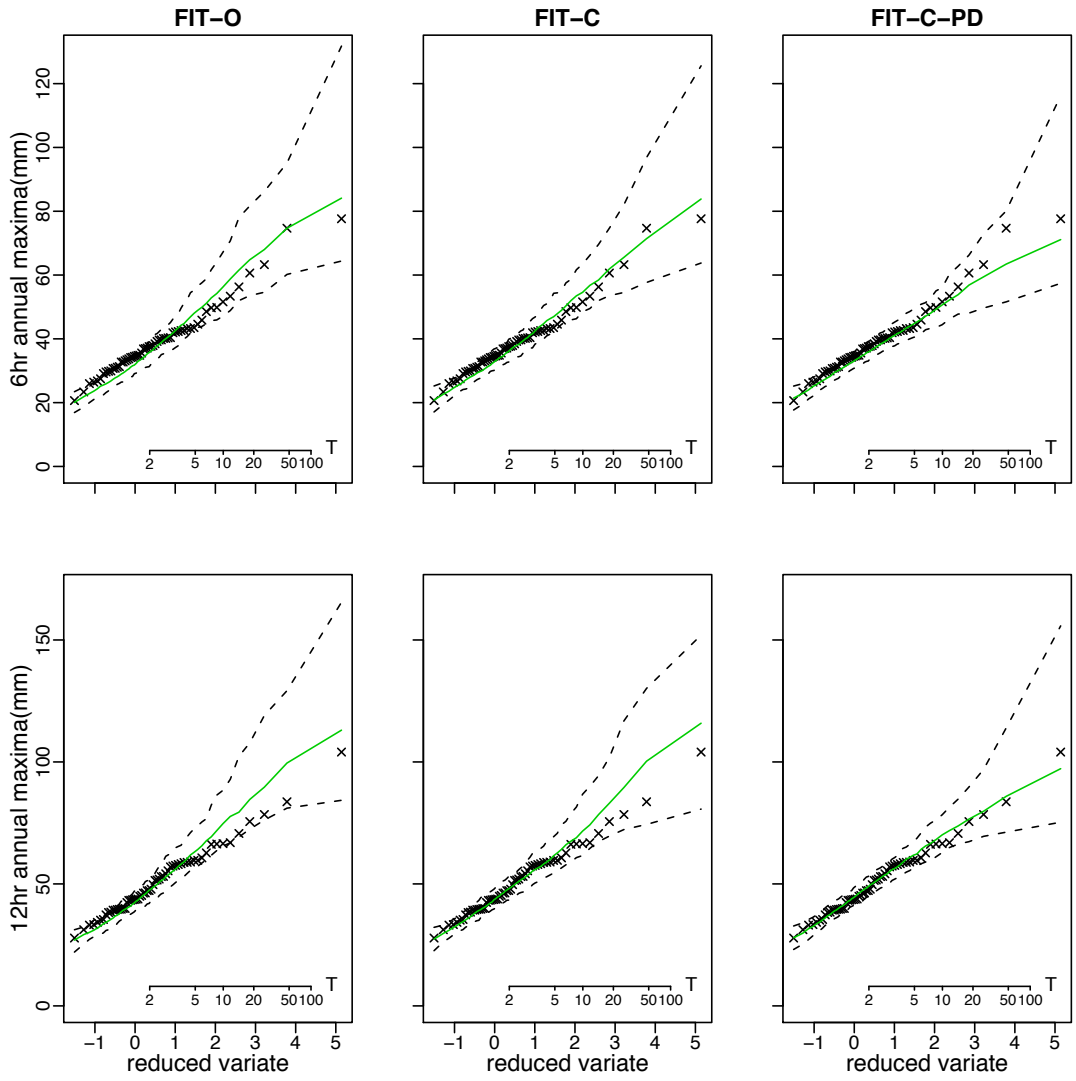

Figure 6.10: Model comparison: extreme value plots at the 6 and 12 hours aggregation levels.

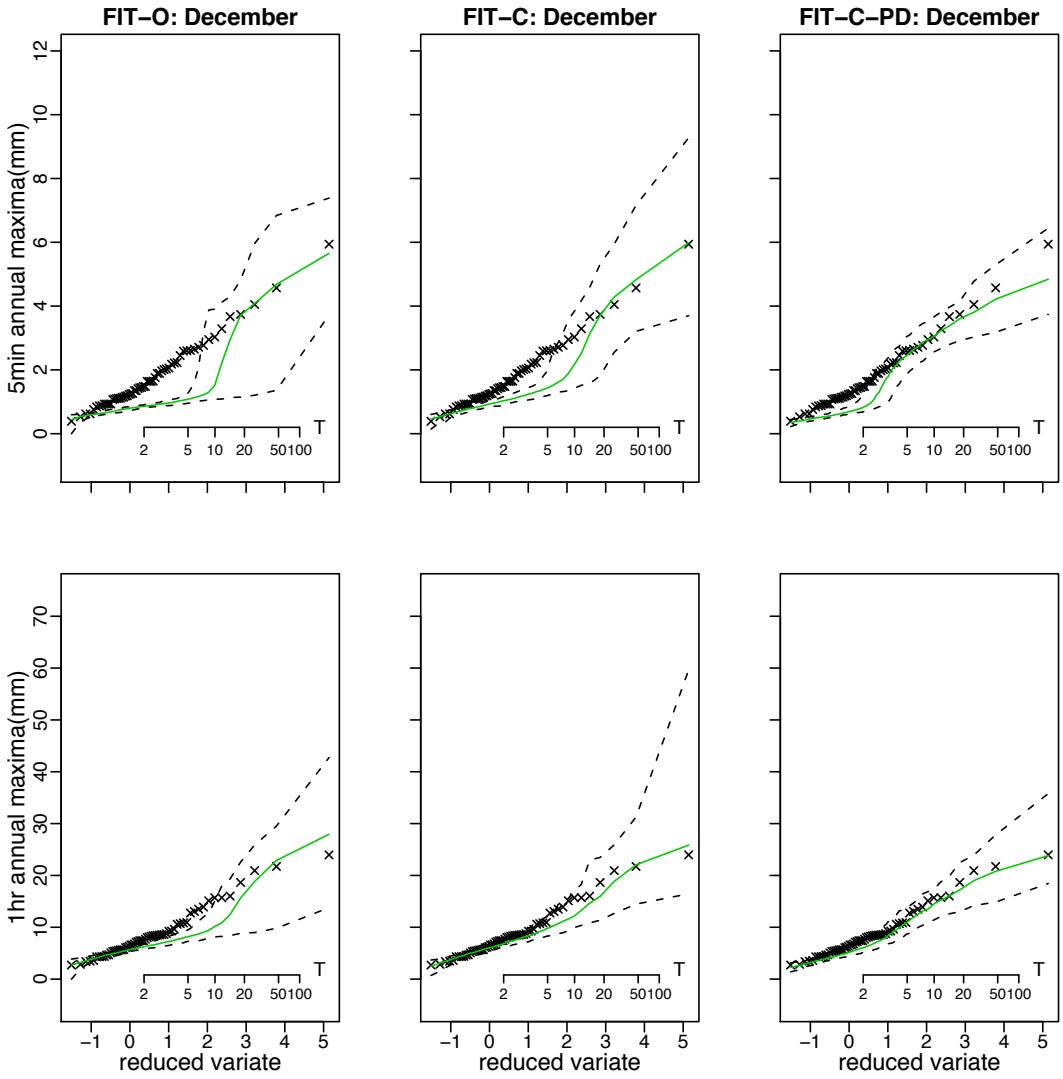

Figure 6.11: Model comparison: ordered December maxima at the 5-min and 1 hour aggregation levels.

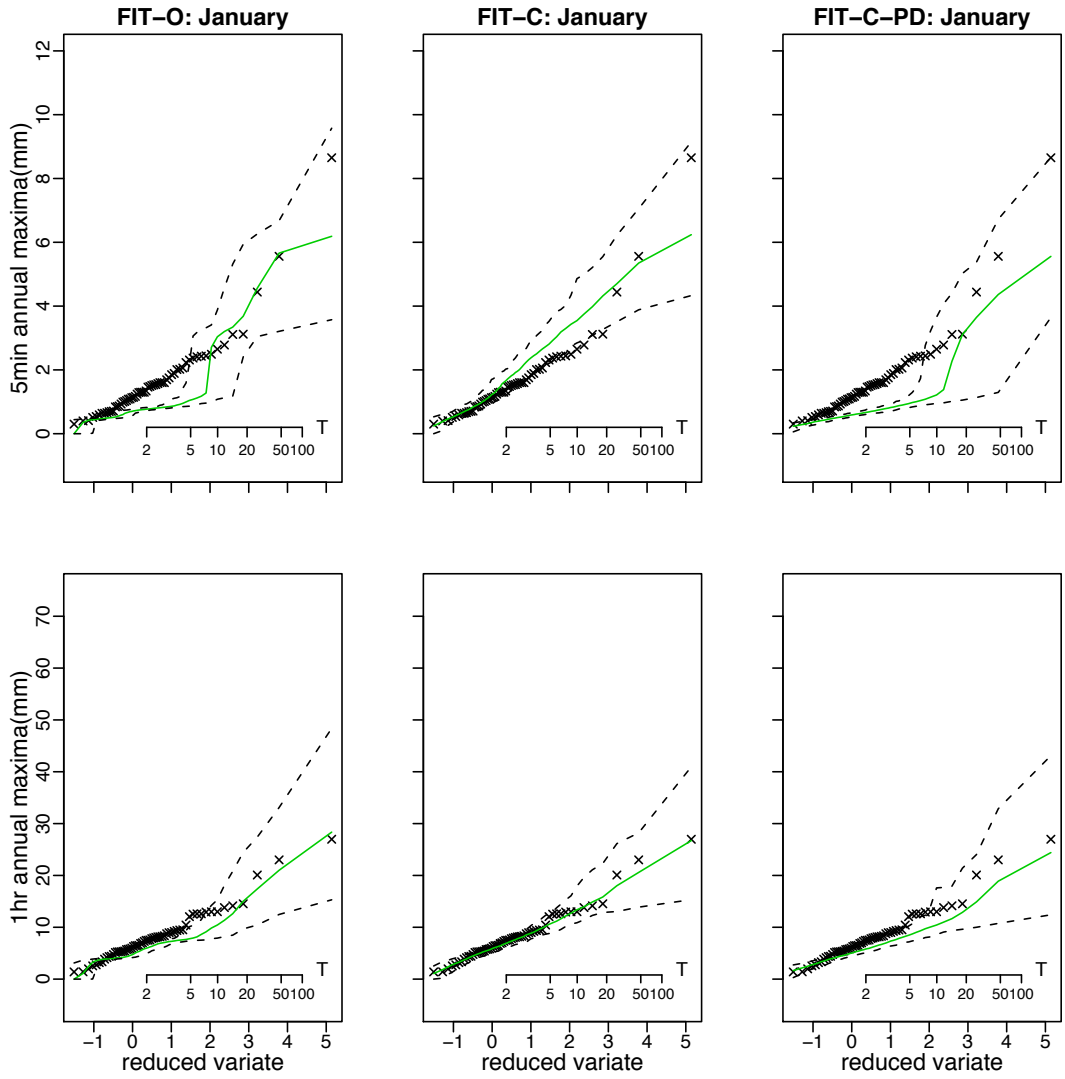

Figure 6.12: Model comparison: ordered January maxima at the 5-min and 1 hour aggregation levels.

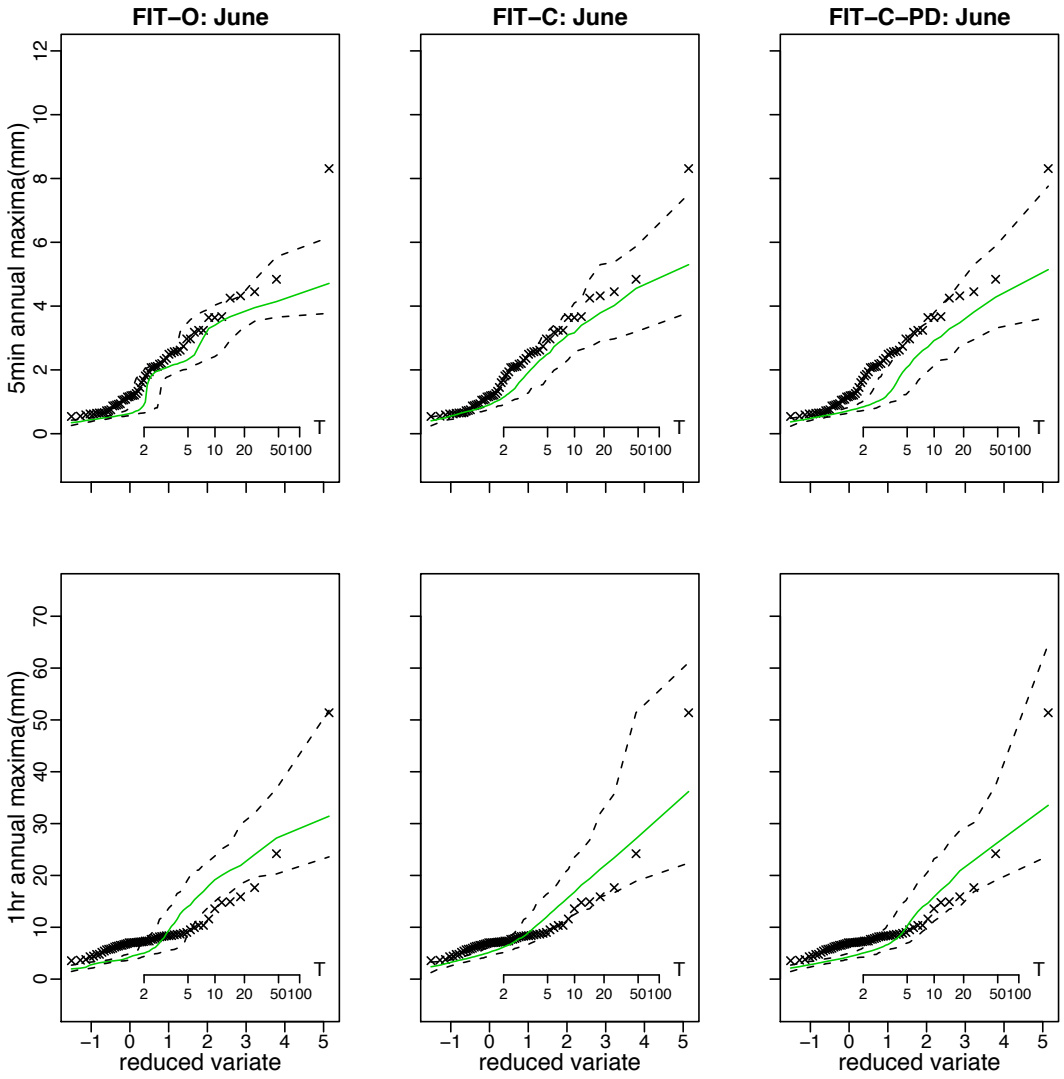

Figure 6.13: Model comparison: ordered June maxima at the 5-min and 1 hour aggregation levels.

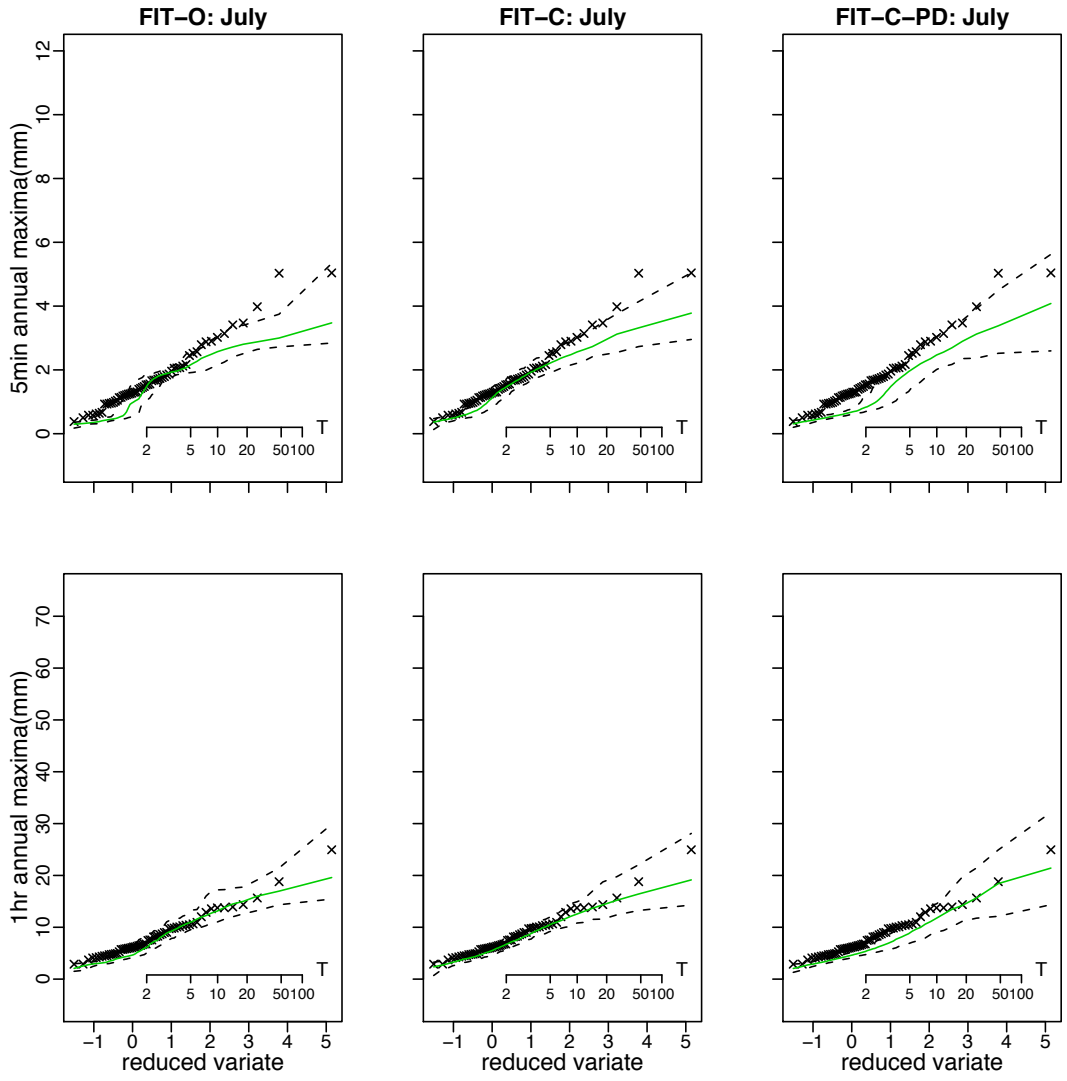

Figure 6.14: Model comparison: ordered July maxima at the 5-min and 1 hour aggregation levels.

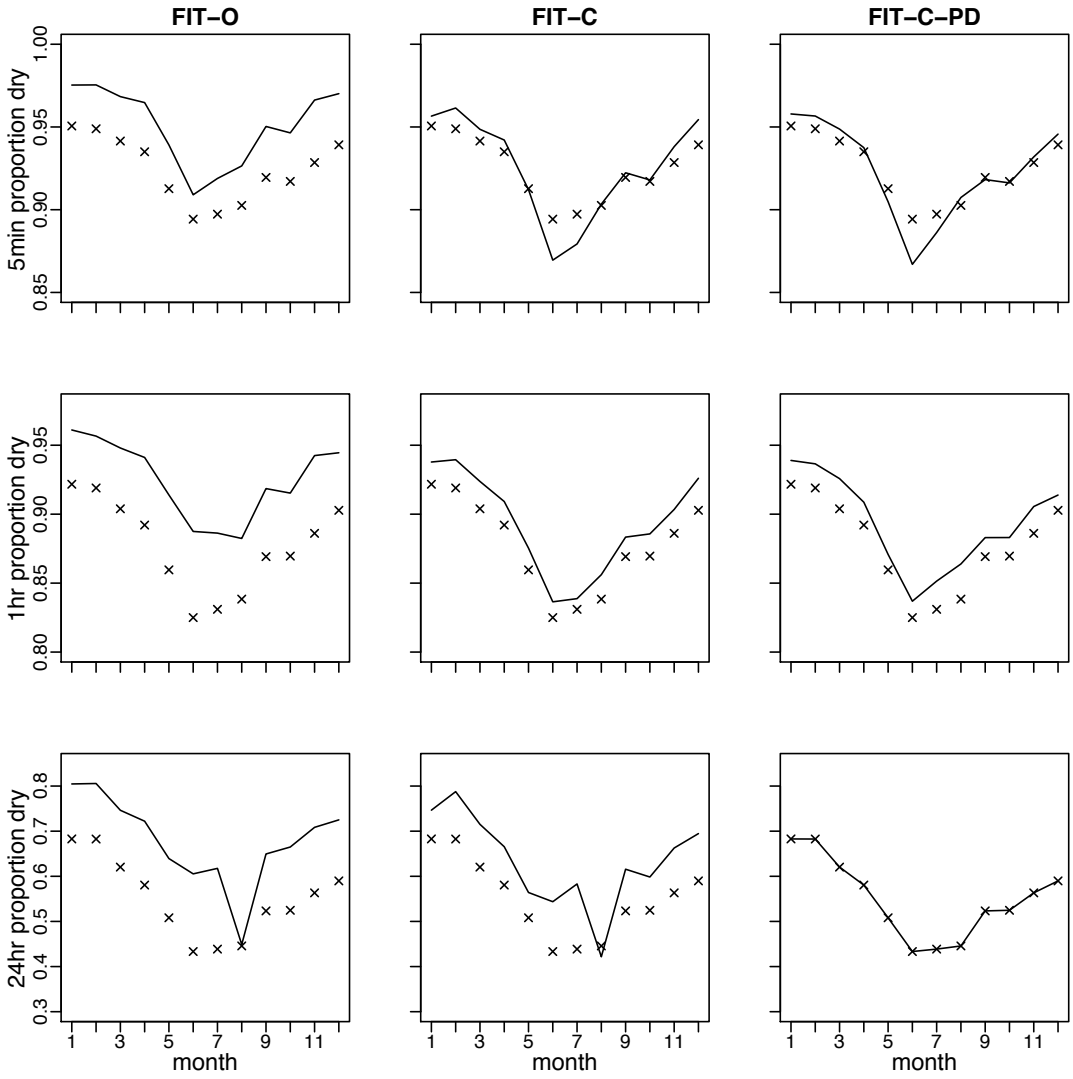

Figure 6.15: Model comparison: proportion dry patterns at the 5-min, 1 hour, and 24 hour aggregation levels.

Figure 6.16 to Figure 6.19 present the simulation results of the proportion of dry at various aggregation levels. The points and lines drawn in the figures follow the layout explained in the beginning of Section 6.4.2.

Systematic error exists when observed rainfall data are fitted by a theoretic probability distribution of positive domain due to the measurement accuracy of the gauges. If the measurement limit for a 5-min interval is 0.01 mm as in this study, any measurement less than 0.01 mm would appear as zero. It is, therefore, reasonable to assess model performance in fitting dry period proportions below some threshold. For practical purpose, if a rainfall total over a given time interval is below a specified threshold (small positive) value, δ, say, this interval may be considered as a 'dry period'. For example, $\mathrm{P}(Y_i^{(1)} < \delta)$ implies that all 1 hour intervals with rainfall totals recorded less than δ (mm) are treated as dry intervals. We compare the proportion dry fits based on different threshold values, namely, 0.01, 0.1 mm for the 5-min level (Figure 6.16) and 0.01, 0.1, 0.5, 2.0 mm for the 1 hour (Figure 6.17), and the 24 hours (Figure 6.18) levels.

From Figure 6.16, Figure 6.17, and Figure 6.18, we see that the fits based on FIT-C and FIT-C-PD improve as the threshold values increase. The simulations based on FIT-O of the proportion dry have poor fits to the observed values for the thresholds specified. The inclusion of a proportion dry constraint improves the proportion dry fittings at all time scale levels downwards but has little effect on higher time scale levels. For example, if we impose proportion dry constraint at the 1 hour aggregation level, we will have an exact proportion dry fit at 1 hour level and improved proportion dry fittings at levels like 30 min or 5 min but almost no effect on 6 or 24 hours proportion dry fitting results. This is the main reason for us to choose FIT-C-PD with proportion dry at 24 hours level and we have a perfect fit with FIT-C-PD as shown in Figure 6.18. Note also the sudden drops of the fitted August values in the plots of FIT-O and FIT-C columns in Figure 6.18 (also the bottom row of Figure 6.15) . This is most likely resulted from the parameter estimates affected by accommodating the sudden rise in the 24 hours historical skewness values as commented in Section 5.3.3. Compared with the perfect fit plots in the FIT-C-PD column, it is certain that the imposition of the proportion dry constraint (6.10) in the fitting procedure should have some impact on the parameter estimation. This may explain the slight (underfit) changes in the extreme value fit at the 5-min level with FIT-C-PD (Figure 6.9).

In Figure 6.19, thresholds of $\delta = 0.3$ mm for 1 hour intervals, $\delta = 0.8$ mm for 6

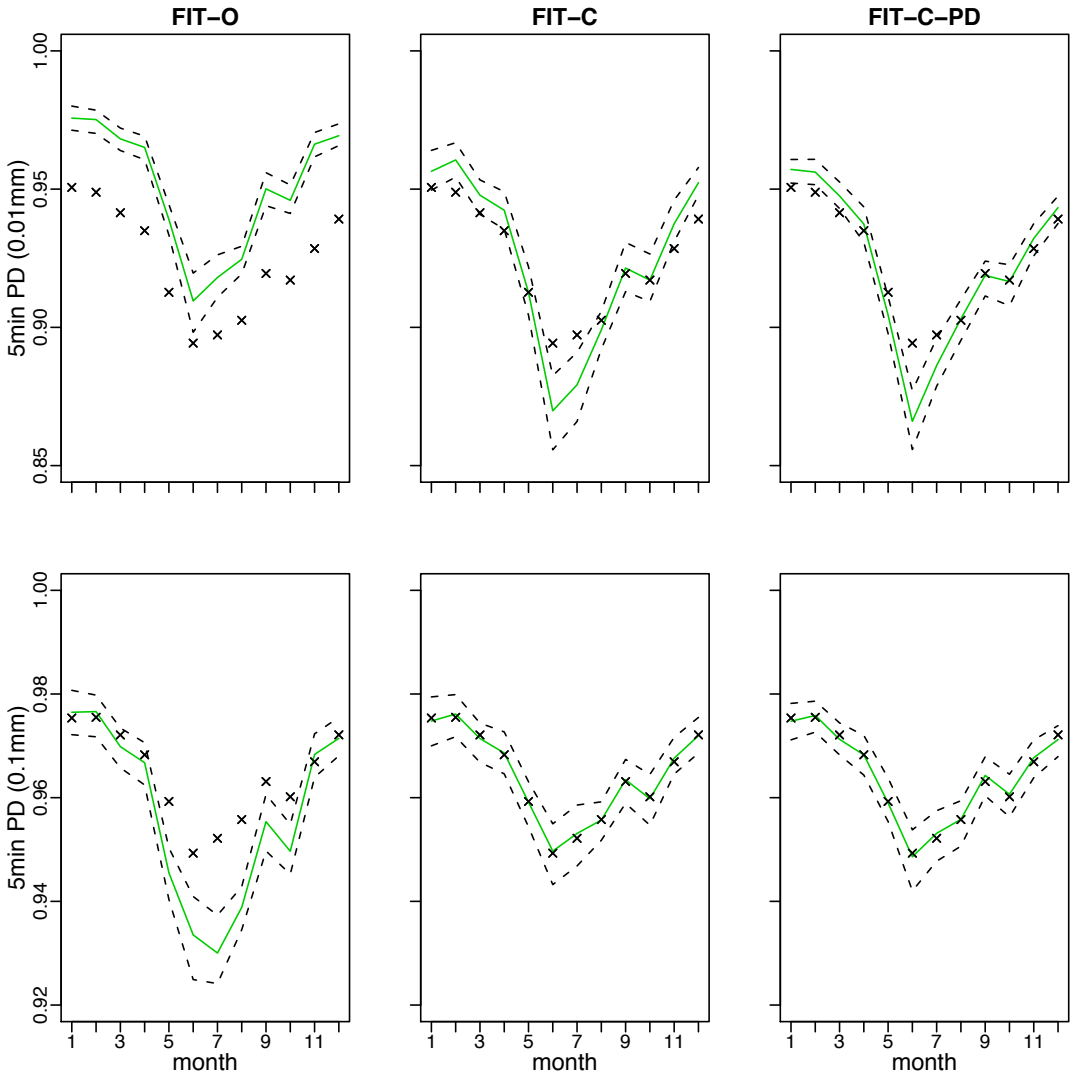

Figure 6.16: Model comparison: proportion dry patterns with different threshold values at the 5-min aggregation level.

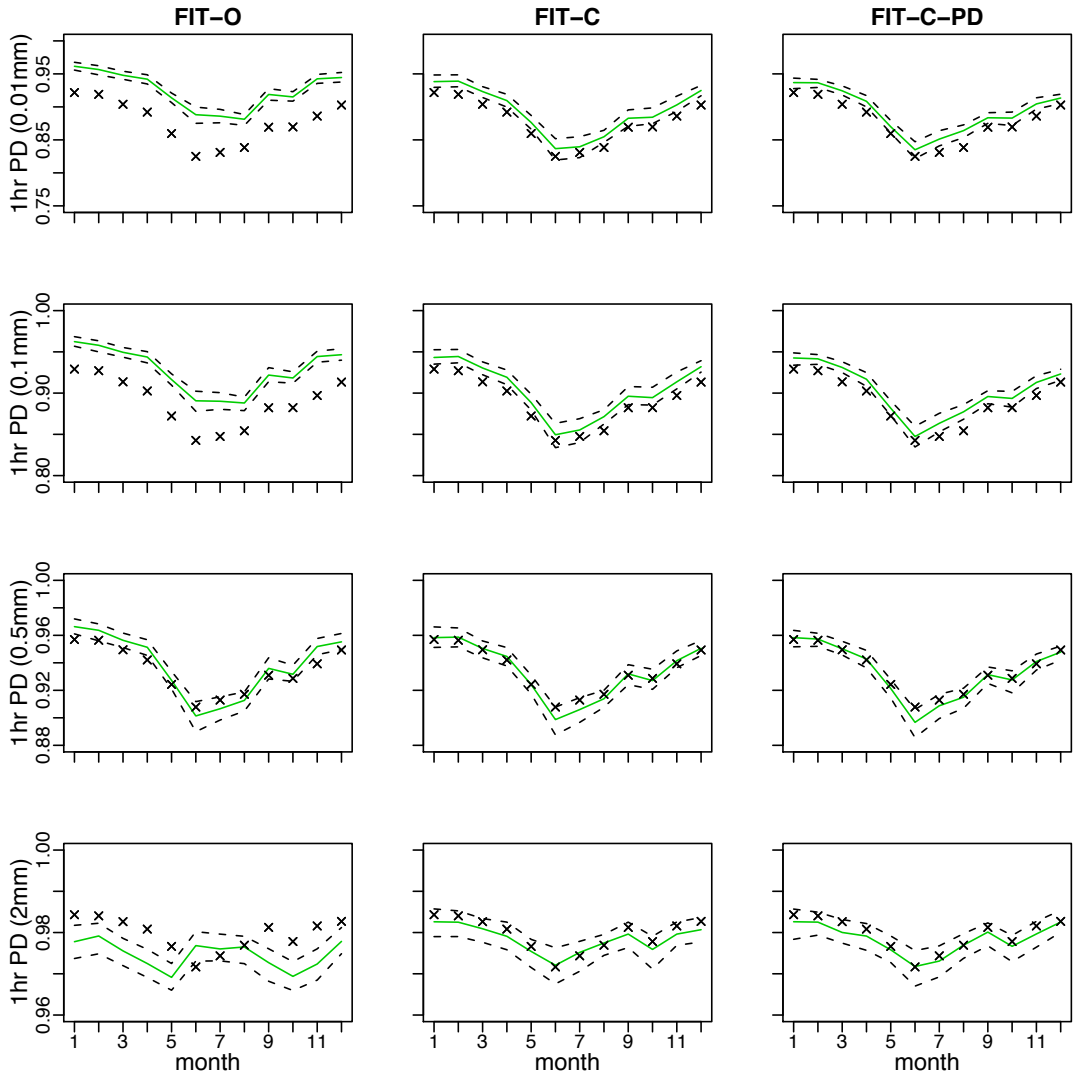

Figure 6.17: Model comparison: proportion dry patterns with different threshold values at the 1 hour aggregation level.

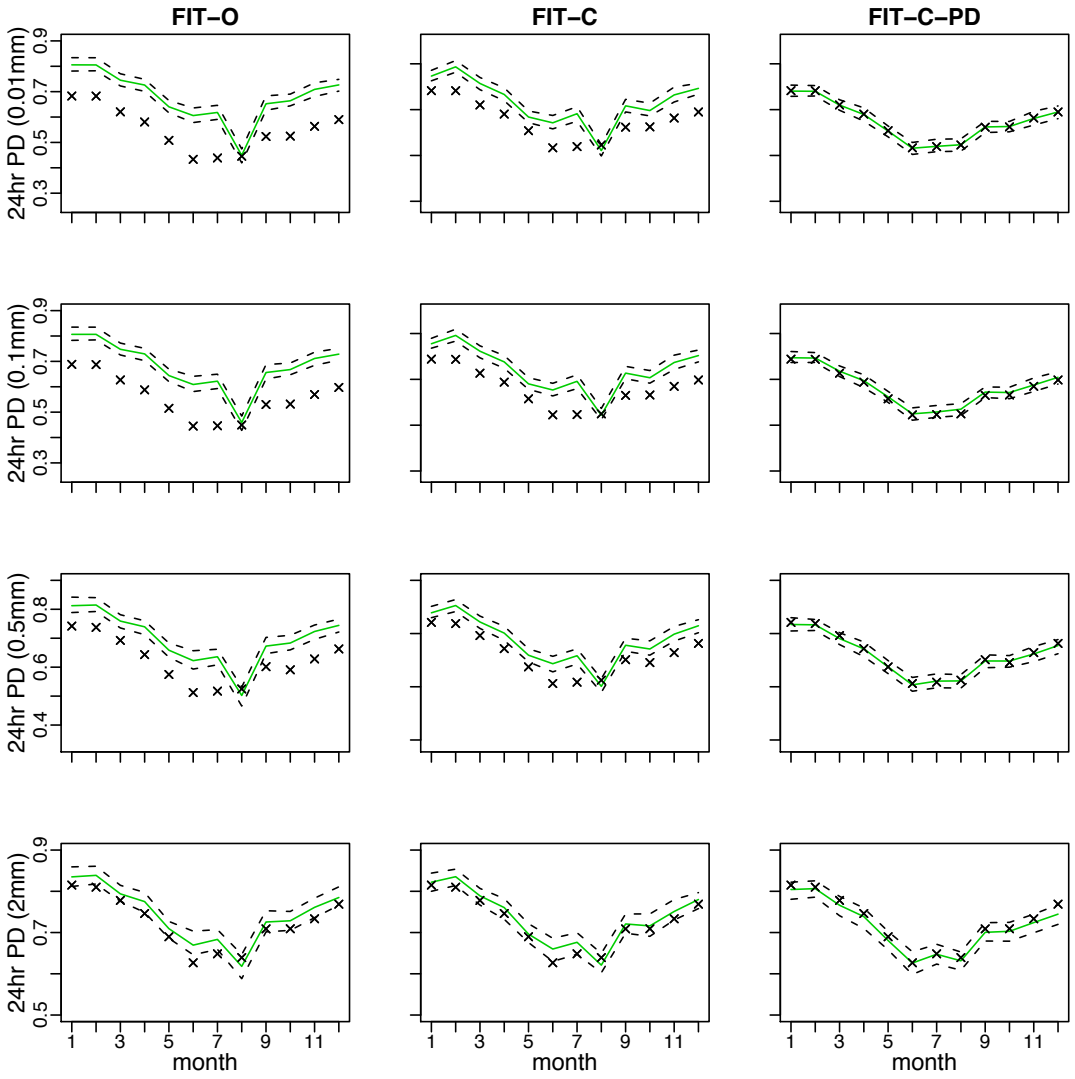

Figure 6.18: Model comparison: proportion dry patterns with different threshold values at the 24 hours aggregation level.

hours or 12 hours intervals, and $\delta = 1$ mm for 24 hours intervals, are used to compare the three model specifications. The simulation results show that most of the historical proportion dry values are well located within the fitted 95% quantile confidence band with FIT-C-PD , whereas almost all points are outside the confidence band with FIT-O. FIT-C improves on FIT-O but is not as good as FIT-C-PD at the 12 and 24 hours levels as expected.

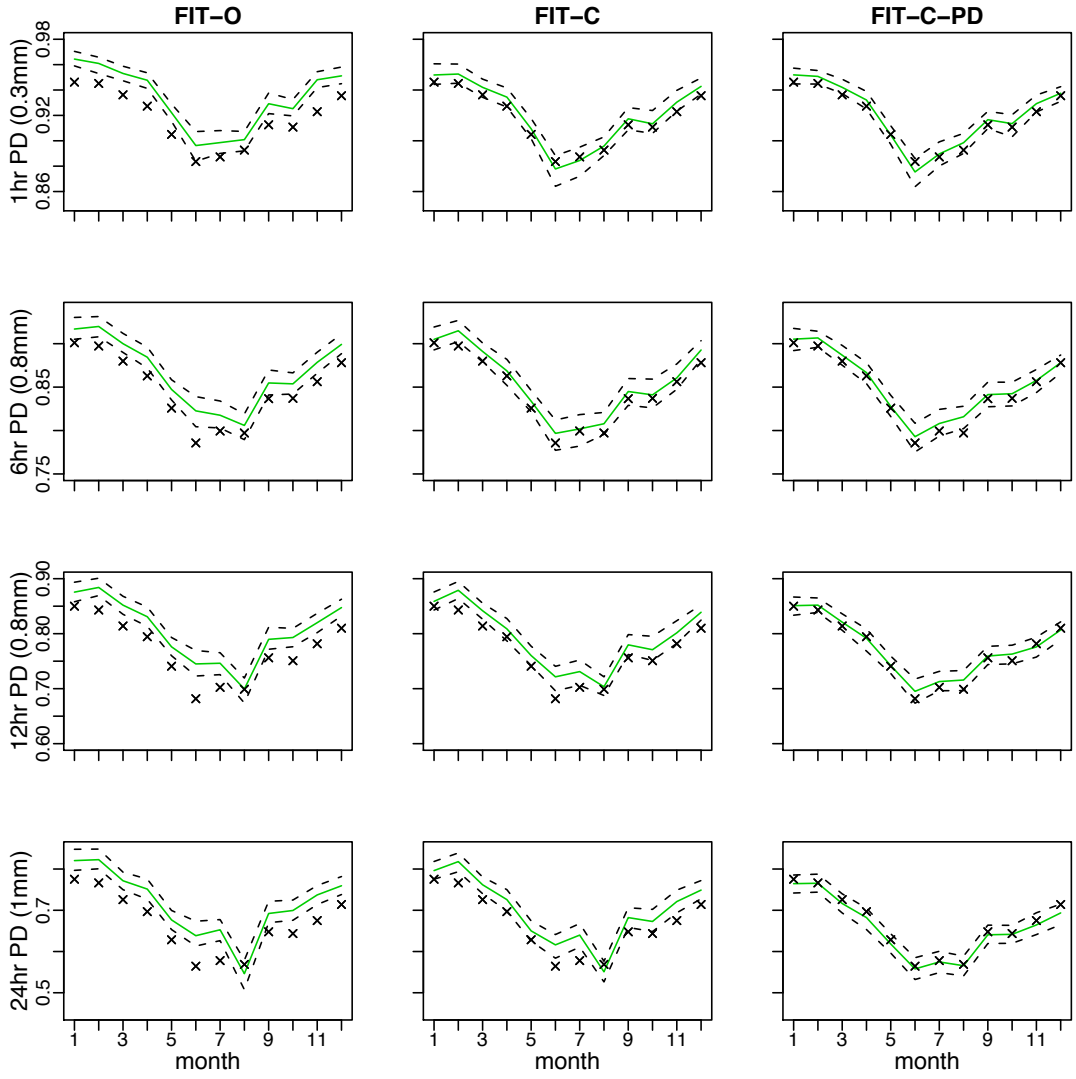

Figure 6.19: Model comparison: proportion dry by threshold values at different aggregation time scale levels.

6.4.5 Model parametrisation and fitting

Recall that three Poisson processes are involved in a BLP model in describing the rainfall pulse process and there are six parameters for a single-process BLP model: generation rate of storm origins, rate of rain cell origins, and pulse arrival rate are denoted by λ, β, and ξ, respectively; the mean storm lifetime, mean cell duration, and mean pulse depth are denoted by γ^{-1}, η^{-1}, and θ, respectively. In the model specifications of FIT-O and FIT-C , two distinct independent storm processes are superimposed and a common mean pulse depth is assumed. Hence, there are 11 parameters in total in each model as shown in Table 6.2

The determination of the the objective function (6.8) is based on the fitting procedure strategy of Cowpertwait et al. (1996a), which assumes that it is more desirable to fit a larger set of sample moments approximately rather than a smaller set exactly. In addition, intuitively and qualitatively, we believe that if the statistical properties at 5-min and 24 hours levels can be fitted well, those properties of in-between time scale levels (e.g. 1 hour, 6 hours, etc) should also fit well. This is how (6.8) has been determined, which includes 12 properties for estimating 10 parameters, $\{\lambda_i, \beta_i, \xi_i, \gamma_i, \eta_i : i = 1, 2\}$ (θ is estimated separately in the very last step). Two problems are found with this fitting procedure: (1) the minimization process is not robust (e.g. easy to hit the specified parameter space bounds or mis-fit badly); (2) many different but possibly equivalent optimal solutions may be obtained. The optimization results are sensitive to the specified parameter space bounds.

As defined in Section 6.3, the 12 sample properties included in (6.8) consist of three sample statistics (coefficient of variation, lag 1 autocorrelation, and coefficient of skewness) at four different aggregation levels (5-min, 1, 6, and 24 hours). The statistical properties of the same sample statistic, e.g. CV, at different aggregation levels, tend to highly correlate to each other. This results in a smaller number of independent components of constraints in the minimization objective function. If the number of independent components of constraints is less than the number of parameters for estimation, the problem of overparametrisation occurs. This is most likely the reason for the unstable and non-unique minimization results. In order to find evidence of overparametrisation, a basic Principal Components Analysis (PCA)[2] is made on those statistical properties which are included in (6.8).

[2]For more details about PCA, see, e.g. Manly (1994)

Table 6.3: Basic Principal Components Analysis result: importance of components

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12
sd	2.412	1.567	1.121	0.9542	0.7325	0.6737	0.5303	0.3867	0.2835	0.2038	0.1083	0.07706
prop of var	0.485	0.205	0.105	0.0759	0.0447	0.0378	0.0234	0.0125	0.0067	0.00346	0.00098	0.00049
cumu prop	0.485	0.689	0.794	0.8699	0.9146	0.9525	0.9759	0.9884	0.9951	0.9985	0.9995	1.0000

The 60 years of 5-min Kelburn data are divided into 12 consecutive 5-year subsamples and each subsample is further divided into 12 subsamples pooled by each month (144 subsamples in total and 12 subsamples for each month which come from different 5-year groups). In this way, each of the 12 properties has 12 replicates. The basic PCA ranks the importance of each principal component. The results are presented in Table 6.3. The abbreviations 'sd' stands for standard deviation; 'prop of var' for proportion of variance; 'cumu prop' for cumulative proportion.

The results show that those statistical properties used in (6.8) are highly correlated to each other. Six principal components can explain more than 95% of the variance expressed by those 12 statistical properties. Nine principal components can explain more than 99% of the total variance. This indicates that the number of independent constraints is less than the number of model parameters, which implies the model is over parameterised and the minimization procedure is therefore not robust. It is therefore possible to have more than one solution of 'optimal' parameter set. One direct solution for the overparametrisation problem is to reduce the number of model parameters. This issue will be further investigated in the next chapter.

Table 6.4 shows how parameter estimates of model FIT-C are correlated to one another. The correlation coefficients are calculated based on the parameter estimates of FIT-C given in Table 6.2. We concentrate on how the corresponding parameters of distinct storm processes are correlated and those linear correlation coefficients are highlighted by bold font. The most striking one is the correlation between $\hat{\xi}_1$ and $\hat{\xi}_2$ which is as high as 0.982. One obvious option to reduce the number of parameters is to assume that there is a deterministic linear regression relationship between $\hat{\xi}_1$ and $\hat{\xi}_2$, hence one can be expressed in terms of the other by a fitted linear regression function. However, simulation study shows that this treatment causes some lack-of-fit to the sample properties. Therefore, a more elaborate but less efficient way of specifying a

Table 6.4: Linear correlation between model parameters: FIT-C

	$\hat{\lambda}_1$	$\hat{\beta}_1$	$\hat{\xi}_1$	$\hat{\gamma}_1$	$\hat{\eta}_1$	$\hat{\lambda}_2$	$\hat{\beta}_2$	$\hat{\xi}_2$	$\hat{\gamma}_2$	$\hat{\eta}_2$
$\hat{\lambda}_1$	1.000	0.419	−0.391	0.728	−0.006	**−0.324**	−0.318	−0.385	−0.545	−0.420
$\hat{\beta}_1$	–	1.000	−0.026	0.877	0.851	−0.486	**0.568**	−0.112	−0.278	0.431
$\hat{\xi}_1$	–	–	1.000	−0.084	0.035	0.712	−0.068	**0.982**	0.768	0.071
$\hat{\gamma}_1$	–	–	–	1.000	0.588	−0.469	0.205	−0.150	**−0.413**	0.116
$\hat{\eta}_1$	–	–	–	–	1.000	−0.572	0.789	−0.066	−0.146	**0.710**
$\hat{\lambda}_2$	–	–	–	–	–	1.000	−0.412	0.772	0.804	−0.347
$\hat{\beta}_2$	–	–	–	–	–	–	1.000	−0.152	0.044	0.943
$\hat{\xi}_2$	–	–	–	–	–	–	–	1.000	0.818	−0.020
$\hat{\gamma}_2$	–	–	–	–	–	–	–	–	1.000	0.094
$\hat{\eta}_2$	–	–	–	–	–	–	–	–	–	1.000

Table 6.5: Ratio of pulse arrival rates: $\hat{\xi}_2/\hat{\xi}_1$

month	1	2	3	4	5	6	7	8	9	10	11	12
FIT-O	10.10	7.17	7.46	8.69	9.75	14.21	7.88	7.84	6.19	6.22	11.87	8.10
FIT-C	6.41	4.85	5.33	5.50	6.50	10.63	5.38	5.52	4.46	5.48	7.63	5.69
FIT-C-PD (fixed)	8.2	5.0	7.6	5.5	6.5	15.0	7.2	6.0	5.6	5.5	8.3	5.8

deterministic relationship between $\hat{\xi}_1$ and $\hat{\xi}_2$ is adopted in fitting model FIT-C-PD. A constant *ratio* vector is given in the third row in Table 6.5 so that $\hat{\xi}_2$ can be determined by $\hat{\xi}_2[i] = \hat{\xi}_1[i] \times ratio[i]$, $i = 1, ..., 12$ (month). The $\hat{\xi}_2$ to $\hat{\xi}_1$ ratios in row 1 and 2 are derived, respectively, based on the parameter estimates of FIT-O and FIT-C given in Table 6.2. The averages of the row 1 and row 2 values are used as the initial *ratio* values to start the model fitting process. The *ratio* vector is finally obtained by manual fine adjustment in the model fitting process. In this way, the negative impact on the model fit of FIT-C-PD is insignificant. However, such a treatment for reducing the number of model parameters is subjective and inefficient. Alternatively, we may assume those pair of parameters which have the least correlation to be the same for different processes, e.g. assuming a common storm generation rate or a common storm mean lifetime for both storm processes. These possibilities are investigated in the next two chapters.

In the process of selecting the best model specification and fitting procedure, we have tested a range of possible schemes and tried to identify the impact from each single change. We found that the increase of within cell depths dependence introduced

by assuming a conditional depth distribution has a major impact in improving annual maxima patterns over the lower return period range ($T < 3$ years). However, the further inclusion of a proportion dry constraint compromises the 5-min extreme values fit slightly (causing a slight underfit) while gaining improvement in fitting the proportion dry. The inclusion of the PD constraint (6.10) has imposed an ideal condition for the parameter estimation. The observed rainfall data actually adopt a threshold PD definition as we have discussed in detail in Section 6.4.4. This could be the reason for the slight compromise for the extreme values fit using FIT-C-PD at the 5-min level (Figure 6.9).

6.5 Summary and conclusion

Motivated for fitting the fine-scale (i.e. sub-hourly) rainfall series with more realistic rain cell profiles, the BLP model was developed by CIO2007 based on the original Bartlett-Lewis rainfall model. In this chapter we have presented the results of an empirical study of model specification and fitting procedures and the resulting model performance of BLP models in fitting a 5-min rainfall time series. Two different BLP model specifications, FIT-C, assuming a conditional mean depth distribution and FIT-C-PD with a proportion dry constraint at 24 hours level also included, are proposed in comparison with the original BLP model FIT-O. The total number of estimated parameters is unchanged with FIT-C. By assuming $\hat{\xi}_2 = \text{constant} \times \hat{\xi}_1$, FIT-C-PD has one estimated parameters less than FIT-O. FIT-C and FIT-C-PD have retained a good fit to moment properties as obtained using FIT-O. A satisfactory goodness-of-fit result to moment properties not used in fitting provides additional evidence for the adequacy of the BLP models.

By simulation, FIT-O shows a slight underfit to the 5-min extreme values and some slight overfit at 12 hours level. Other than these, FIT-O fits extreme values very well. FIT-C improves upon FIT-O. The annual maxima fitting results using FIT-C are adequate. Significant improvement is obtained in fitting the individual month annual maxima using FIT-C, although there are still some lack of fit cases. The improvement in fitting extreme values is due to the introduction of conditional mean depth distribution.

Proportion dry is a key feature in applications of a stochastic rainfall model. The previous work have shown that the basic NSRP and BLRP models have a tendency

to overestimate the observed proportion dry. For example, Cowpertwait et al. (1996a) included proportion dry directly in the objective function for parameter estimation to improve the model proportion dry performance; according to Frost et al. (2004), a basic NSRP model generally overestimates proportion dry even when the proportion dry is directly included in the fitting procedure. FIT-O also produces an overestimation in the observed proportion dry. In contrast, both FIT-C and FIT-C-PD have significantly improved the model proportion dry performance at all time scale levels examined, with or without assuming a threshold. FIT-C-PD further improves the fit to the proportion dry at higher time scale levels, e.g. 12 and 24 levels, from FIT-C. However, this gain is at a cost of slightly underfitting 5-min extreme values. If 5-min extreme value pattern is not of the major concern in the application, further improvement on the reproduction of proportion dry patterns may be achieved by imposing two proportion dry constraints in the fitting procedure. It is likely that FIT-C and FIT-C-PD's ability in reproduction of proportion dry patterns will satisfy the needs for many practical applications.

Chapter 7

BLP models with continuous distributions of storm types

7.1 Introduction

In Chapter 6, based on the empirical study results, we conclude that a different BLP model specification, FIT-C, significantly improved the model performance of the original BLP model, FIT-O, in fitting the annual maxima and proportion dry values. The improvements are due to the successful implementation of the within cell pulse depths dependence structure by assuming a conditional mean exponential distribution. On the other hand, the discussions on the fitting procedure reveal that further research is needed in optimization of the model parametrisation. In this chapter, we investigate an alternative way for the characterization of the rainfall process within the BLP framework and aim to improve upon FIT-C through the optimization of model parametrisation.

Cowpertwait (2010) proposed a Neyman-Scott model with continuous distributions of storm types. In this approach, individual storms are characterized by a continuum of storm types z, where z is a random variable of a continuous probability distribution, and model parameters are taken to be functions of z. This model setting allows for superposition of different types of storms within the same storm process and the parametrisation enables the exploration of how selected model parameters are associated. BLP models with continuous distributions of storm types are formulated in this chapter and the statistical properties up to third order are derived for several typical model specifications. Based on the derived properties of the fitted models, the continuous-storm-types BLP models are compared to the original BLP models in fitting the Kelburn data.

In what follows, Section 7.2 gives the details of the formulation of a BLP model with continuous distributions of storm types. Moment properties up to third order are derived in the general case and for several typical model specifications in particular. In Section 7.3, model comparison results between different single process BLP models are presented. BLP models with two storm processes superposed are compared in Section 7.4. Section 7.5 finishes this chapter with concluding remarks.

7.2 Model specification and properties

7.2.1 Formulation of a continuous-storm-types BLP model

In the original BLP model setting, the model parameters are constant with each individual storm of the same process while in reality storms are all different. Following a continuous-storm-types approach, model parameters are expressed as functions of a continuous random variable z with probability density function $f(z)$. Hence, governed by $f(z)$, each storm is characterized by different model parameters and occurs with probability $f(z)\mathrm{d}z$. Denote an individual storm as a type z storm and represent this single storm type rainfall process by a BLP model. Based on the original BLP model framework, a BLP model with continuous distributions of storm types can then be obtained by integration of all types of storms over the sample space of z to represent a rainfall process which consists of a continuum of storm types. The details of the formulation of a BLP model with continuous distributions of storm types follow.

Let $\{Y_{z,i}^{(h)}: \quad i = 1, 2, ...\}$ be the aggregated rainfall series sampled over an interval of length h due to a type z storm, so that $Y_{z,i}^{(h)} = \int_{(i-1)h}^{ih} X_z(t)\mathrm{d}N_z(t)$, where $X_z(t)$ is the depth of a pulse located at time t, and $N_z(t)$ is the counting process of the pulse occurrences.

Assuming a null state of knowledge about the distributions of storm types, an obvious and mathematically tractable choice is to take z to be a uniform random variable, $f(z) = 1; \quad 0 < z < 1$, in which an equal chance of occurrence for all types of storms is implied. Following the original BLP model and without loss of generality, we take $X(z)$ (the pulse depths due to type z storm) to be an independent exponential random variable with mean $\theta(z)$. Correspondingly, by expressing other model parameters as functions of z, a BLP model of type z storm is specified within the original BLP model framework as:

λ_z	– rate of type z storm origins;	$\beta(z)$	– rate of cell origins for type z storms;
$\xi(z)$	– pulse arrival rate for type z storms;	$\gamma^{-1}(z)$	– mean storm lifetime;
$\eta^{-1}(z)$	– mean cell duration;	$\theta(z)$	– mean pulse depth.

For a type z storm, the rate of the storm origins, λ_z, is determined by

$$\lambda_z \;=\; \lambda f(z)\mathrm{d}z, \tag{7.1}$$

where λ is a positive constant. Moment properties up to third order can be written out directly based on Table 6.1 and Equations (6.2), (6.3), (6.4), and (B.12), by substituting λ_z, $\beta(z)$, $\xi(z)$, $\gamma(z)$, $\eta(z)$, and $\theta(z)$ for λ, β, ξ, γ, η, and θ. To calculate the moment properties for a rainfall process that consists of all types of storms, we need to aggregate rainfalls contributed from each individual storms over the fixed length of time intervals. Details of moment properties derivation are given as follows.

7.2.2 Moment properties

The mean, variance, covariance, and the third moment of rainfalls in the ith interval of length h due to type z storm are denoted by $\mathrm{E}\left[Y_{z,i}^{(h)}\right]$, $\mathrm{Var}\left[Y_{z,i}^{(h)}\right]$, $\mathrm{Cov}\left[Y_{z,i}^{(h)}, Y_{z,i+k}^{(h)}\right]$, and $\mathrm{E}\left[(Y_{z,i}^{(h)})^3\right]$, respectively. The total rainfall contributed from all types of storms is conceptualized as the results of superposition of independent storms and each storm is characterized differently by the model parameters which are the functions of a random variable z. The total rainfall in the ith time interval due to the superposition of all storm types thus can be obtained as $Y_i^{(h)} = \int_z Y_{z,i}^{(h)}$. Properties, up to third order, of the aggregated continuous-storm-types process are the sum of the equivalent properties for each independent type z storm.

$$\left.\begin{aligned}
\mathrm{E}\left[Y_i^{(h)}\right] &= \int_z \mathrm{E}\left[Y_{z,i}^{(h)}\right] \\
\mathrm{Var}\left[Y_i^{(h)}\right] &= \int_z \mathrm{Var}\left[Y_{z,i}^{(h)}\right] \\
\mathrm{Cov}\left[Y_i^{(h)}, Y_{i+k}^{(h)}\right] &= \int_z \mathrm{Cov}\left[Y_{z,i}^{(h)}, Y_{z,i+k}^{(h)}\right] \\
\mathrm{E}\left[\left(Y_i^{(h)} - \mathrm{E}(Y_i^{(h)})\right)^3\right] &= \int_z \mathrm{E}\left[\left(Y_{z,i}^{(h)} - \mathrm{E}(Y_{z,i}^{(h)})\right)^3\right].
\end{aligned}\right\} \tag{7.2}$$

Note that, conceptually, the integral symbol $\int_z$ is used to represent the summation of the rainfall contributions from all independent storm processes characterized through

z. Although the total rainfall results are obtained by performing the ordinary integral calculations over the sample space of z with respect to $f(z)$ as we will give the formula soon, this is not a case of calculating the moment properties given a conditional distribution setting.

There are many possible choices for the model parameter functions and the original BLP model is recovered by setting the model parameters to be constants: $\beta(z) = \beta$, $\xi(z) = \xi$, $\gamma(z) = \gamma$, $\eta(z) = \eta$, and $\theta(z) = \theta$. Partly for mathematical tractability, in this study, we will concentrate on the investigation of the relationships between the model performance and the different parameter functions of $\beta(z)$, $\xi(z)$, and $\theta(z)$. Therefore, we set $\gamma(z) = \gamma$, $\eta(z) = \eta$ and the conditional pulse depth distribution (as defined in Section 6.2 for FIT-C) is assumed in the following model specification and property derivation steps.

Based on Table 6.1 and Equation (6.2), we can write

$$\mathrm{E}\left[Y_{z,i}^{(h)}\right] = \frac{\lambda_z \xi(z)\beta(z)\theta(z)h}{\gamma(z)(\gamma(z)+\eta(z))} = \frac{\lambda\xi(z)\beta(z)\theta(z)h}{\gamma(\gamma+\eta)} f(z)\,\mathrm{d}z. \tag{7.3}$$

By (7.2) and (7.3), the overall mean property $\mathrm{E}\left[Y_i^{(h)}\right]$ is obtained as (assuming pdf $f(z) = 1; \quad 0 < z < 1$)

$$\begin{aligned}
\mathrm{E}\left[Y_i^{(h)}\right] &= \int_{\Omega_z} \frac{\lambda\xi(z)\beta(z)\theta(z)h}{\gamma(\gamma+\eta)} f(z)\,\mathrm{d}z \\
&= \int_0^1 \frac{\lambda\xi(z)\beta(z)\theta(z)h}{\gamma(\gamma+\eta)}\,\mathrm{d}z \\
&= \frac{\lambda h}{\gamma(\gamma+\eta)} \int_0^1 \beta(z)\xi(z)\theta(z)\,\mathrm{d}z \\
&= \frac{\lambda h}{\gamma(\gamma+\eta)} W_{111},
\end{aligned} \tag{7.4}$$

where W_{111} is defined in (7.5) below.

To make the derived formulas easy to specify, we now define some notation for integral expressions which will be used for different BLP model specifications (assuming pdf $f(z) = 1; \quad 0 < z < 1$)

$$W_{ijk} = \int_{\Omega_z} \left[\beta(z)\right]^i \left[\xi(z)\right]^j \left[\theta(z)\right]^k f(z)\,\mathrm{d}z = \int_0^1 \left[\beta(z)\right]^i \left[\xi(z)\right]^j \left[\theta(z)\right]^k \mathrm{d}z, \tag{7.5}$$

where $i, j, k = 1$ or 2 or 3. For the derivation of moment properties up to third order, we need to evaluate the following 10 integrals for each BLP model: W_{111}, W_{112}, W_{113}, W_{122}, W_{123}, W_{133}, W_{222}, W_{223}, W_{233}, and W_{333}.

In a similar way, based on Equation (6.3), (6.4), (7.2) and Table 6.1, the variance and autocovariance formulas can be derived as

$$\mathrm{Var}\left[Y_i^{(h)}\right] = \frac{\lambda}{\gamma(\gamma+\eta)}\left[4hW_{112} + \frac{2\phi(\gamma)}{\eta}W_{222} + 4\phi(\gamma+\eta)W_{122} - \frac{2\gamma\phi(\gamma+\eta)}{\eta(\gamma+2\eta)}W_{222}\right], \tag{7.6}$$

and

$$\mathrm{Cov}\left[Y_i^{(h)}, Y_{i+k}^{(h)}\right] = \frac{\lambda}{\gamma(\gamma+\eta)}\left[\frac{\psi(\gamma)}{\eta}W_{222} + 2\psi(\gamma+\eta)W_{122} - \frac{\gamma\psi(\gamma+\eta)}{\eta(\gamma+2\eta)}W_{222}\right], \tag{7.7}$$

where $\phi(\alpha) = (e^{-\alpha h} - 1 + \alpha h)/\alpha^2$ and $\psi(\alpha) = e^{-\alpha(k-1)h}(1-e^{-\alpha h})^2/\alpha^2$, and W_{112}, W_{122}, W_{222} are defined in Equation (7.5).

Although algebraically it is more complex, the derivation of the third moment property is similar to that of the first and the second moment properties. The derivation and the final expression of the third moment about the mean is outlined as follows.

Based on the third moment expression given in Equation (B.12), the third moment about the origin for type z storm, $\mathrm{E}\left[(Y_{z,i}^{(h)})^3\right]$, can be obtained. It follows that the third moment about the mean for a particular storm can be obtained as

$$\mathrm{E}\left[\left(Y_{z,i}^{(h)} - \mathrm{E}(Y_{z,i}^{(h)})\right)^3\right] = \mathrm{E}\left[(Y_{z,i}^{(h)})^3\right] - 3\mathrm{E}\left[Y_{z,i}^{(h)}\right]\mathrm{Var}\left[Y_{z,i}^{(h)}\right] - \left(\mathrm{E}(Y_{z,i}^{(h)})\right)^3.$$

To count for all possible types of storms, the third moment about mean is given by

$$\begin{aligned}\mathrm{E}\left[\left(Y_i^{(h)} - \mathrm{E}(Y_i^{(h)})\right)^3\right] &= \int_0^1 \mathrm{E}\left[\left(Y_{z,i}^{(h)} - \mathrm{E}(Y_{z,i}^{(h)})\right)^3\right] \mathrm{d}z \\ &= \int_0^1 \left\{\mathrm{E}\left[(Y_{z,i}^{(h)})^3\right] - 3\mathrm{E}\left[Y_{z,i}^{(h)}\right]\mathrm{Var}\left[Y_{z,i}^{(h)}\right] - \left(\mathrm{E}(Y_{z,i}^{(h)})\right)^3\right\} \mathrm{d}z.\end{aligned} \tag{7.8}$$

With the help of the mathematical programing language package Maple for symbolic operation with algebras, after simplification, the final expression of the third moment about mean can be expressed as

$$
\begin{aligned}
&\mathrm{E}\left[\left(Y_i^{(h)} - \mathrm{E}(Y_i^{(h)})\right)^3\right] \\
&= \frac{6\lambda(6m_1W_{133} + 2m_2W_{233} + m_3W_{333})}{\gamma^4(2\eta+\gamma)(\eta+\gamma)^4\eta^2} \\
&\quad + \frac{6\lambda}{(\eta+\gamma)(2\eta+\gamma)^2}\Big[-\frac{4W_{233}}{\eta(\eta+\gamma)} + \frac{W_{333}}{\eta^2(3\eta+\gamma)}\Big] \\
&\qquad \times\left[h - \frac{3\eta+2\gamma}{(\eta+\gamma)(2\eta+\gamma)} + \frac{(2\eta+\gamma)e^{-(\eta+\gamma)h}}{\eta(\eta+\gamma)} - \frac{(\eta+\gamma)e^{-(2\eta+\gamma)h}}{\eta(2\eta+\gamma)}\right] \\
&\quad + \frac{6\lambda}{(\eta\gamma)(\eta+\gamma)^2}\Big[\frac{4W_{233}}{\gamma} - \frac{W_{333}}{\eta(2\eta+\gamma)}\Big] \\
&\qquad \times\left[h - \frac{\eta+2\gamma}{\gamma(\eta+\gamma)} + \frac{(\eta+\gamma)e^{-\gamma h}}{\eta\gamma} - \frac{\gamma e^{-(\eta+\gamma)h}}{\eta(\eta+\gamma)}\right] \\
&\qquad + \frac{72\lambda}{\gamma(\eta+\gamma)^2}W_{123}\left[h - \frac{1}{\eta+\gamma} + \frac{e^{-(\eta+\gamma)h}}{\eta+\gamma}\right] \\
&\quad + \frac{24\lambda}{\eta\gamma^2(\eta+\gamma)}W_{223}\left[h - \frac{1}{\gamma} + \frac{e^{-\gamma h}}{\gamma} - \frac{\gamma^2}{(\eta+\gamma)(2\eta+\gamma)}\left(h - \frac{1}{\eta+\gamma} + \frac{e^{-(\eta+\gamma)h}}{\eta+\gamma}\right)\right] \\
&\qquad + \frac{36\lambda h}{\gamma(\eta+\gamma)}W_{113}, \qquad (7.9)
\end{aligned}
$$

where W_{133}, W_{233}, W_{333}, W_{123}, W_{223}, W_{113} are defined in Equation (7.5) and

$$
\begin{aligned}
m_1 &= \gamma^3\eta^2\Big[\eta h e^{-(\eta+\gamma)h}(2\eta+3\gamma) + h(\gamma^2 e^{-(\eta+\gamma)h} + 2\eta^2) \\
&\qquad + \gamma h(3\eta+\gamma) - 2(2\eta+\gamma) + 2e^{-(\eta+\gamma)h}(2\eta+\gamma)\Big]; \\
m_2 &= 2\eta^2\gamma^3\Big[e^{-(\eta+\gamma)h}(\eta h + 2 + \gamma h) + (\eta h + \gamma h - 2)\Big]; \\
m_3 &= \gamma h e^{-\gamma h}(2\eta^4 + \gamma^4 + 7\eta^3\gamma) - \gamma^5 h e^{-(\eta+\gamma)h} + \gamma\eta^2 e^{-\gamma h}(14\eta + 18\gamma) \\
&\qquad + \eta\gamma^3(9\eta h + 5\gamma h + 10)e^{-\gamma h} - \eta\gamma(14\eta^2 + 18\eta\gamma + 10\gamma^2) \\
&\qquad - \gamma^4 e^{-(\eta+\gamma)h}(\eta h + 2) + \eta^3(2\eta\gamma h + 7\gamma^2 h - 4\eta) \\
&\qquad + 2e^{-\gamma h}\gamma^4 + 2\eta^4) + \eta\gamma^3 h(9\eta + 4\gamma).
\end{aligned}
$$

Note that, in the derivation process, all λ_z^2 terms are treated as zeros because of the involvement of $(\mathrm{d}z)^2$ terms.

Because of the complexity in the analytic expression of the formula for calculating proportion dry[1], the corresponding formula for a continuous-storm-types BLP model in the general case is not available.

[1]More details can be found in Appendix B.5.

7.2.3 Fitted models

We denote the original BLP model specification for a single process of storms by FIT-O1 (six parameters in total) and the corresponding conditional mean depth BLP model specification by FIT-C1. Using FIT-C1 as the base model, four continuous-storm-types BLP models are specified in terms of parameter functions $\beta(z)$, $\xi(z)$, and $\theta(z)$. The statistical properties of these four fitted continuous-storm-types BLP models can be obtained by evaluation of Equation (7.5) based on the specified parameter functions $\beta(z)$, $\xi(z)$, and $\theta(z)$ in particular.

In model FIT-C1-βzθz, the parameter functions $\beta(z)$, $\xi(z)$, and $\theta(z)$ are defined by

$$\left.\begin{array}{rcl} \beta(z) &=& \beta_0(1-z)+\beta_1 z \\ \xi(z) &=& \xi \\ \theta(z) &=& \theta z. \end{array}\right\} \tag{7.10}$$

This model specification enables us to examine whether storms with low rain cell generation rates on average tend to correspond to storms with low or high mean pulse depths. For example, as $z \to 0$, $\theta(z) \to 0$ and $\beta(z) \to \beta_0$; as $z \to 1$, $\theta(z) \to \theta$ and $\beta(z) \to \beta_1$. If the parameter estimation results show that $\hat{\beta}_1 < \hat{\beta}_0$, this implies that lower rain cell generation rates (hence less frequent rainfall) tends to be associated with more intensive rainfall (higher mean pulse depth). This may give a better understanding of the BLP model representation for rainfall processes. For similar reasons, three other continuous-storm-types BLP models are specified.

The functions $\beta(z)$, $\xi(z)$, and $\theta(z)$ for model FIT-C1-ξzθz are defined by

$$\left.\begin{array}{rcl} \beta(z) &=& \beta \\ \xi(z) &=& \xi_0(1-z)+\xi_1 z \\ \theta(z) &=& \theta z. \end{array}\right\} \tag{7.11}$$

This model tends to investigate the impact of the association between the pulse arrival rates and the mean pulse depths.

The functions $\beta(z)$, $\xi(z)$, and $\theta(z)$ for model FIT-C1-βzξz are defined by

$$\left.\begin{array}{rcl} \beta(z) &=& \beta_0(1-z)+\beta_1 z \\ \xi(z) &=& \xi_0(1-z)+\xi_1 z \\ \theta(z) &=& \theta. \end{array}\right\} \tag{7.12}$$

This model tends to investigate the impact of the association between the rain cell origins generation rates and the pulse arrival rates.

The functions $\beta(z)$, $\xi(z)$, and $\theta(z)$ for model FIT-C1-βzξzθz are defined by

$$\left.\begin{array}{rcl} \beta(z) & = & \beta_0(1-z)+\beta_1 z \\ \xi(z) & = & \xi_0(1-z)+\xi_1 z \\ \theta(z) & = & \theta z. \end{array}\right\} \tag{7.13}$$

This model tends to investigate how the model performance is affected when associating the rain cell generation rates, the pulse arrival rates, and the mean pulse depths all together through the continuous random variable z.

Equations (B.13), (B.14), (B.15), (B.16), (B.17) of Appendix B give the evaluation results of Equation (7.5) for models FIT-C1, FIT-C1-βzθz, FIT-C1-ξzθz, FIT-C1-βzξz, and FIT-C1-βzξzθz. Therefore, moment properties up to third order can be obtained for these models based on Equations (7.4), (7.6), (7.7), and (7.9). These fitted models are used to fit the Kelburn data using the original BLP model minimisation objective function (6.8). The model comparison results are presented in the next section.

7.3 Model comparisons: single BLP process

7.3.1 Parameter estimates

The resulting parameter estimates for model FIT-O1 and FIT-C1 are listed in Table 7.1. The resulting parameter estimates for model FIT-C1-βzθz and FIT-C1-ξzθz are presented in Tables 7.2 and 7.3, respectively. The resulting parameter estimates for model FIT-C1-βzξz and FIT-C1-βzξzθz are listed in Table 7.4. Among these six BLP models, FIT-C1-βzξzθz has the most model parameters, eight, which is three less than FIT-C but two more than FIT-C1 and FIT-O1.

By looking into the details of the parameter estimates in Table 7.1, we have the following findings. For most months of the year, FIT-C1 has higher generation rates of storm origins (larger $\hat{\lambda}$ values), higher generation rates of rain cell origins (larger $\hat{\beta}$ values), associated with shorter mean storm lifetimes (larger $\hat{\gamma}$ values) and shorter mean rain cell durations (larger $\hat{\eta}$ values), than FIT-O1. And FIT-C1 has higher mean pulse arrival rates associated with smaller mean pulse depths (larger $\hat{\xi}$ values but smaller $\hat{\theta}$ values) than FIT-O1. Two exceptions are with August and December. These are probably caused by the extreme rainfall events observed in August and December as we have noted in Section 5.3.3, because the within cell dependence structure implemented by the conditional mean pulse depths specification is more sensitive to the extreme rainfall events.

Table 7.1: Parameter estimates: for each month (m) the estimates are listed in the order of model FIT-O1 and FIT-C1. The units are hour^{-1} for all estimates except θ which has units mm.

m	$\hat{\lambda}$	$\hat{\beta}$	$\hat{\xi}$	$\hat{\gamma}$	$\hat{\eta}$	$\hat{\theta}$
1	0.00521	0.235	47.5	0.0663	1.31	0.153
	0.00635	0.383	56.8	0.0596	1.14	0.0509
2	0.00626	0.512	55.5	0.121	1.35	0.0988
	0.00773	0.68	65.2	0.109	1.05	0.0363
3	0.00906	0.405	50.5	0.136	1.39	0.123
	0.0118	0.519	60.4	0.118	1.12	0.0435
4	0.00753	0.432	49.5	0.104	1.54	0.129
	0.00945	0.573	58.2	0.0907	1.22	0.046
5	0.00898	0.477	60.5	0.0989	1.69	0.102
	0.0112	0.613	73.1	0.0853	1.34	0.0362
6	0.00939	0.463	52.6	0.0946	1.78	0.14
	0.0122	0.535	62.7	0.0768	1.39	0.0496
7	0.0106	0.492	51.2	0.0902	1.34	0.0799
	0.0129	0.635	60.1	0.0787	1.04	0.0295
8	0.0112	0.665	52.7	0.131	1.17	0.0648
	0.00366	0.0166	54.3	0.00213	0.289	0.0278
9	0.0133	0.358	60.7	0.109	1.31	0.0662
	0.0164	0.523	74.5	0.0976	1.11	0.0227
10	0.0112	0.469	57.4	0.109	1.27	0.0681
	0.0137	0.637	68.0	0.0971	1.01	0.0248
11	0.00891	0.228	41.5	0.0862	1.35	0.177
	0.0116	0.308	49.2	0.0717	1.13	0.0591
12	0.00655	0.708	56.0	0.159	1.39	0.103
	0.0209	0.0486	51.8	0.0581	0.32	0.0451

Table 7.2 shows that, except for December, the rain cell generation rates are negatively associated with the mean pulse depths, i.e. $\hat{\beta}_0 > \hat{\beta}_1$ and as $z \to 0$ $\theta(z) \to 0$ but $\beta(z) \to \hat{\beta}_0$. Compared with those resulting parameter estimates of FIT-C1 in Table 7.1, changes in August and December for FIT-C1-$\beta z\theta z$ are much smoother. Model FIT-C1-$\beta z\theta z$ has consistently higher pulse arrival rates ($\hat{\xi}$) than FIT-C1.

Note that the interpretation of parameters β_0, β_1, and θ in FIT-C1-$\beta z\theta z$ is different from parameters β and θ in the original BLP model (i.e. FIT-O1) or in FIT-C1 because of the continuous random variable z's involvement. With FIT-O1 and FIT-C1, the cell origins generation rates and the mean pulse depths are constants for any individual storm, whereas, the cell origins generation rates can be any value between $\hat{\beta}_0$ and $\hat{\beta}_1$, and the mean pulse depths can be any value between 0 and $\hat{\theta}$, governed by z, in FIT-

Table 7.2: Parameter estimates for each month (m): model FIT-C1-βzθz. The units are hour^{-1} for all estimates except θ which has units mm.

m	$\hat{\lambda}$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\xi}$	$\hat{\gamma}$	$\hat{\eta}$	$\hat{\theta}$
1	0.00820	1.07	0.000810	62.0	0.0628	1.15	0.0823
2	0.00960	1.50	0.629	71.5	0.115	1.2	0.0475
3	0.0149	1.44	0.000	66.3	0.122	1.13	0.0710
4	0.0112	1.87	0.255	64.1	0.0998	1.46	0.0666
5	0.0136	1.82	0.267	82.4	0.0917	1.54	0.0510
6	0.0152	1.37	0.116	69.1	0.0779	1.41	0.0747
7	0.0159	1.85	0.218	65.9	0.0829	1.18	0.0434
8	0.0181	3.79	0.679	71.3	0.132	1.71	0.0324
9	0.0207	1.47	0.000	83.8	0.101	1.13	0.0363
10	0.0170	1.40	0.502	74.6	0.102	1.11	0.0330
11	0.0151	0.804	0.00665	53.6	0.0731	1.12	0.0950
12	0.0144	0.350	0.949	67.9	0.141	1.01	0.0482

C1-βzθz. Therefore, those parameters which have been expressed as functions of z should have the similar interpretations in the following continuous-storm-types model specifications.

Table 7.3 shows that, with no exception, the pulse arrival rates are positively associated with the mean pulse depths, i.e. $\hat{\xi}_0 < \hat{\xi}_1$ and as $z \to 0$ $\theta(z) \to 0$ and $\xi(z) \to \hat{\xi}_0$. The higher pulse arrival rates associated with the higher mean pulse depths implies that the extreme heavy rainfall events are easier to be represented by this model specification. This is supported by the fact that the changes in parameter estimates from month to month is much smoother with model FIT-C1-ξzθz than FIT-C1. Compared with the resulting parameter estimates of FIT-C1-βzθz in Table 7.2, FIT-C1-ξzθz has consistently higher generation rates of storm origins but smaller mean pulse depths.

Table 7.4 shows that, in most months of year, the rain cell generation rates are negatively associated with the pulse arrival rates (i.e. as $z \to 0$, $\beta(z) \to \hat{\beta}_0$ and $\xi(z) \to \hat{\xi}_0$, but $\hat{\beta}_0 > \hat{\beta}_1$, $\hat{\xi}_0 < \hat{\xi}_1$) which implies that more frequent rainfalls tend to be less intensive. This pattern of parameter estimates is found in both FIT-C1-βzξz and FIT-C1-βzξzθz. The association of the mean pulse depths $\theta(z)$ into model specification causes a slight increase in the storm origins generation rates (slightly higher $\hat{\lambda}$ values) and a very small decrease in the mean storm lifetimes (tiny increase in $\hat{\gamma}$ values). On the other hand, the mean rain cell durations are caused to increase slightly (smaller $\hat{\eta}$ values). Again, the association patterns are consistent in the pulse arrival rates and

Table 7.3: Parameter estimates for each month (m): model FIT-C1-ξzθz. The units are hour^{-1} for all estimates except θ which has units mm.

m	$\hat{\lambda}$	$\hat{\beta}$	$\hat{\xi}_0$	$\hat{\xi}_1$	$\hat{\gamma}$	$\hat{\eta}$	$\hat{\theta}$
1	0.00939	0.450	56.4	62.5	0.0639	1.21	0.0626
2	0.0119	0.995	42.4	81.6	0.118	1.28	0.0404
3	0.0178	0.690	45.7	72.6	0.128	1.27	0.0503
4	0.0166	0.780	17.6	78.7	0.0987	1.43	0.0489
5	0.0210	0.805	11.1	106.0	0.0920	1.55	0.0359
6	0.0232	0.658	8.88	89.4	0.0820	1.54	0.0502
7	0.0230	0.896	15.9	82.2	0.0859	1.26	0.0308
8	0.0268	0.979	4.63	84.3	0.122	0.972	0.0262
9	0.0233	0.650	78.5	82.9	0.104	1.21	0.0275
10	0.0209	0.899	49.3	83.9	0.106	1.2	0.0279
11	0.0203	0.375	19.6	64.0	0.0778	1.23	0.0654
12	0.0165	0.897	5.31	92.1	0.144	1.08	0.0406

the exceptions all come from the relationships of the rain cell origins generation rates.

Unlike in the FIT-C case (superposition of two independent BLP processes), the optimization process for parameter estimation is now robust. With all these six models, the optimal parameter sets are unique and not sensitive to the initial values. The difficulties in the fitting procedure caused by overparametrisation no longer exist. However, the derived properties from these fitted models fit the historical values much poorer than FIT-C and the results are shown in the following subsection.

7.3.2 Derived properties

Given the parameter estimates, the statistical properties of coefficient of variation (CV), coefficient of skewness, and Lag 1 autocorrelation, as defined in (6.7), are derived for the BLP models FIT-O1, FIT-C1, FIT-C1-βzθz, FIT-C1-ξzθz, FIT-C1-βzξz, and FIT-C1-βzξzθz. The derived properties are compared to the historical values to check the goodness-of-fit of the fitted models. The comparison results are presented in Figure 7.1 through to Figure 7.4.

Figure 7.1 is an enlargement of the top-left plot of Figure 7.2 with graphical symbols explained. In all figures, the historical values are represented by points (crosses). The solid lines represent the derived properties values of model FIT-C. Because FIT-C has an almost perfect fit to the historical values, it is used here as a standard of a 'good' fit. The deviation from solid lines indicates a lack-of-fit of a fitted model. In these

Table 7.4: Parameter estimates: for each month (m) the estimates are listed in the order of model FIT-C1-βzξz and FIT-C1-βzξzθz. The units are hour^{-1} for all estimates except θ which has units mm.

m	$\hat{\lambda}$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\xi}_0$	$\hat{\xi}_1$	$\hat{\gamma}$	$\hat{\eta}$	$\hat{\theta}$
1	0.00800	1.10	0.000	0.00347	130	0.0634	1.22	0.0418
	0.00845	1.31	0.0000184	39.2	79.8	0.0648	1.18	0.072
2	0.0101	1.06	0.853	0.00152	110	0.116	1.25	0.0303
	0.0103	2.90	0.00151	18.6	105	0.117	1.15	0.0474
3	0.0145	1.49	0.00271	0.00129	138	0.123	1.21	0.0360
	0.0157	2.08	0.0000583	23.5	96.4	0.127	1.20	0.0572
4	0.0131	0.714	0.717	0.00117	97.3	0.0964	1.40	0.0385
	0.0132	2.33	0.000304	11.9	97.8	0.0966	1.30	0.0582
5	0.0149	0.930	0.660	0.00449	134	0.0895	1.50	0.0284
	0.0164	2.43	0.000434	7.54	136	0.089	1.41	0.0418
6	0.0146	1.19	0.283	0.00269	133	0.078	1.48	0.0383
	0.0184	1.94	0.0000250	6.47	116	0.0779	1.42	0.0575
7	0.0173	0.941	0.756	0.00130	104	0.0836	1.21	0.0243
	0.0183	2.60	0.00089	9.57	102	0.0832	1.12	0.0371
8	0.0231	0.350	1.12	0.00829	93.1	0.121	0.994	0.0216
	0.0278	0.776	1.09	5.00	83.9	0.123	1.02	0.0258
9	0.0200	1.49	0.0306	0.00262	168	0.101	1.18	0.0185
	0.0215	1.96	0.0000241	40.9	116	0.104	1.17	0.0301
10	0.0169	1.27	0.637	0.00663	123	0.103	1.16	0.0203
	0.0183	2.67	0.000160	21.9	110	0.104	1.11	0.0322
11	0.0147	0.843	0.0000600	0.00131	112	0.0742	1.20	0.0489
	0.0167	1.16	0.00000620	14.2	81.2	0.0766	1.18	0.0751
12	0.0137	0.432	0.984	0.00231	103	0.142	1.09	0.0334
	0.0164	0.930	0.880	5.25	92.3	0.143	1.08	0.0407

figures, dashed lines are used for representing FIT-O1 and FIT-C1 and dotted lines for those four continuous-storm-types models, each with different types of points as shown in the key of Figure 7.1. The fits of CV at the 5-min aggregation level are examined in Figure 7.1. It is clear that FIT-O1 has the worst overfit and FIT-C1 improves a lot but still is not as good as those continuous-storm-types models in most months. The dotted line of model FIT-C1-βzξzθz has a least deviation from the solid line. Note that the two exceptions occur in August and December, the two months with extreme rainfall events happened during the sample period as commented in Section 5.3.3. The fits of CV at the 5-min, 1, 6, and 24 hours levels are examined in Figure 7.2. The findings from the examination of the fits of CV at the 1, 6, and 24 hours levels are essentially the same as in the 5-min level case.

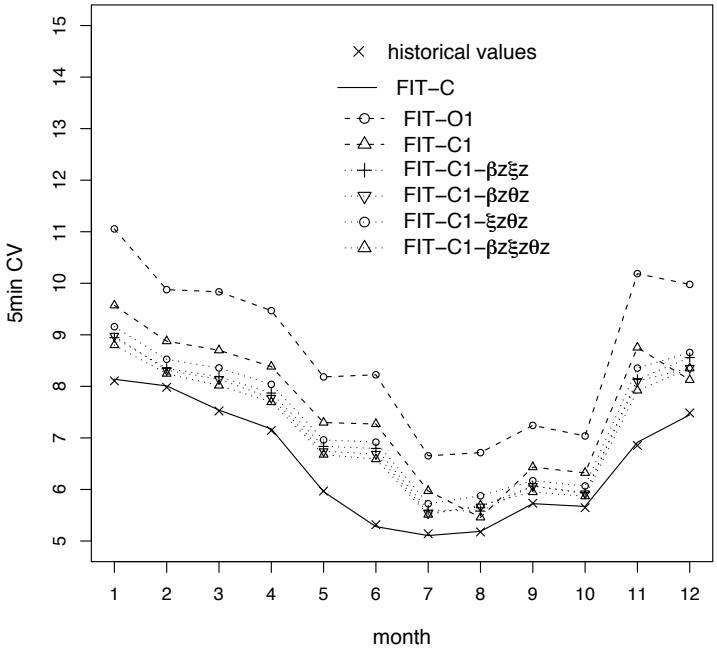

Figure 7.1: Model comparison (single process models): coefficient of variation aggregated at the 5-min level.

In Figures 7.3 and 7.4, we examine the fits of skewness and lag 1 autocorrelation at the 5-min, 1, 6, and 24 hours levels, respectively. Big underfit deviations appear at the 5-min levels. However, the gaps decrease as the aggregation time scales increase. Except for FIT-O1, all models have a reasonable good fit with the lag 1 autocorrelations at the 24 hours level. Again, FIT-O1 is identified to be the worst fit model. In the 5-min aggregation level cases, the performance of FIT-C1 can be identified as the second worst and FIT-C1-βzξzθz as the best. In other cases, the lack-of-fits are apparent but the patterns are harder to define. Note the two huge sudden rises of the FIT-C1 fits with the lag 1 autocorrelations at the 1 hours level (bottom-left plot of Figure 7.4). Once again, we identify the impact of the extreme rainfall events happened in August and December on the fitted model as noted before. Overall, we conclude that the continuous-storm-types models outperform FIT-O1 and FIT-C1 without taking into account of the number of models parameters. In Table 7.5, we compare these models numerically based on the minimisation objective function (6.8) and the number of model parameters are also compared.

Table 7.5: Model comparison in terms of minimisation function value and the total number of parameters: single process models

	FIT-O1	FIT-C1	FIT-C1-βzθz	FIT-C1-ξzθz	FIT-C1-βzξz	FIT-C1-βzξzθz
minimised value (total)	13.9	5.60	2.72	3.14	4.06	2.68
minimised value (maximum)	2.36	1.14	0.589	0.679	0.837	0.566
relative error % (maximum)	100	75.5	54.3	58.3	64.7	53.2
parameters	6	6	7	7	8	8

In the optimization process for parameter estimation, the objective function (6.8) is minimised. The minimised value represents the total squared relative errors due to the differences between the estimated (fitted) values and the observed (historical) values coming from the 12 properties included. The values of the first row in Table 7.5 are the sums of minimised values from 12 pooled monthly subsamples. The values of the second row are the minimised values of the worst fit month. The third row gives the values of the relative errors in percentages for the worst fit month, assuming all errors contributed from a single property. The bottom row gives the number of parameters of the fitted models. As shown in Figures 7.1 through to 7.4, the discrepancy errors are actually contributed to by more than one of the properties included in (6.8). Therefore, the relative error ($|1 - \hat{f}/f| \times 100\%$) actually allocated to each property should have been much smaller than the results given in the third row of Table 7.5. The maximal relative error values presented in Table 7.5 have been calculated based on the worst scenario.

Based on the minimisation values and the number of model parameters presented in Table 7.5, the following conclusions can be made. FIT-C1 improves upon FIT-O1 significantly and both models have six parameters. This implies that the rainfall processes are better represented by FIT-C1 due to the introduction of the conditional mean depths distribution. The continuous-storm-types models improve upon FIT-C1 significantly but at the cost of increasing the number of model parameters. Model FIT-C1-βzξzθz has the best goodness-of-fit to historical values among these six models but has the most number of parameters. Expressing the mean depth as a function $\theta(z) = \theta z$ does improve the model fit. However, none of the single process models examined fits the Kelburn data well using FIT-C as the standard.

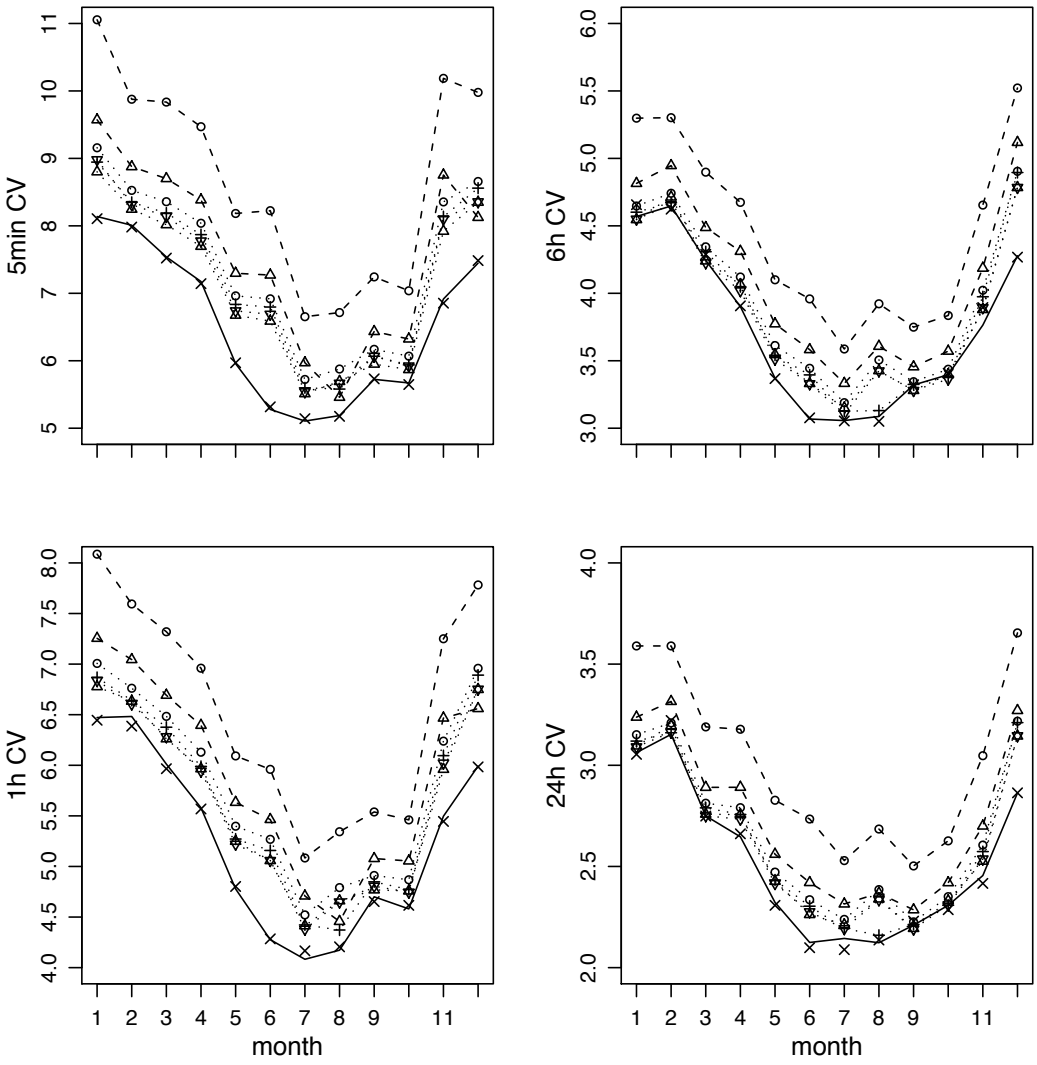

Figure 7.2: Model comparison (single process models): coefficient of variation aggregated at the 5-min, 1, 6, and 24 hours levels.

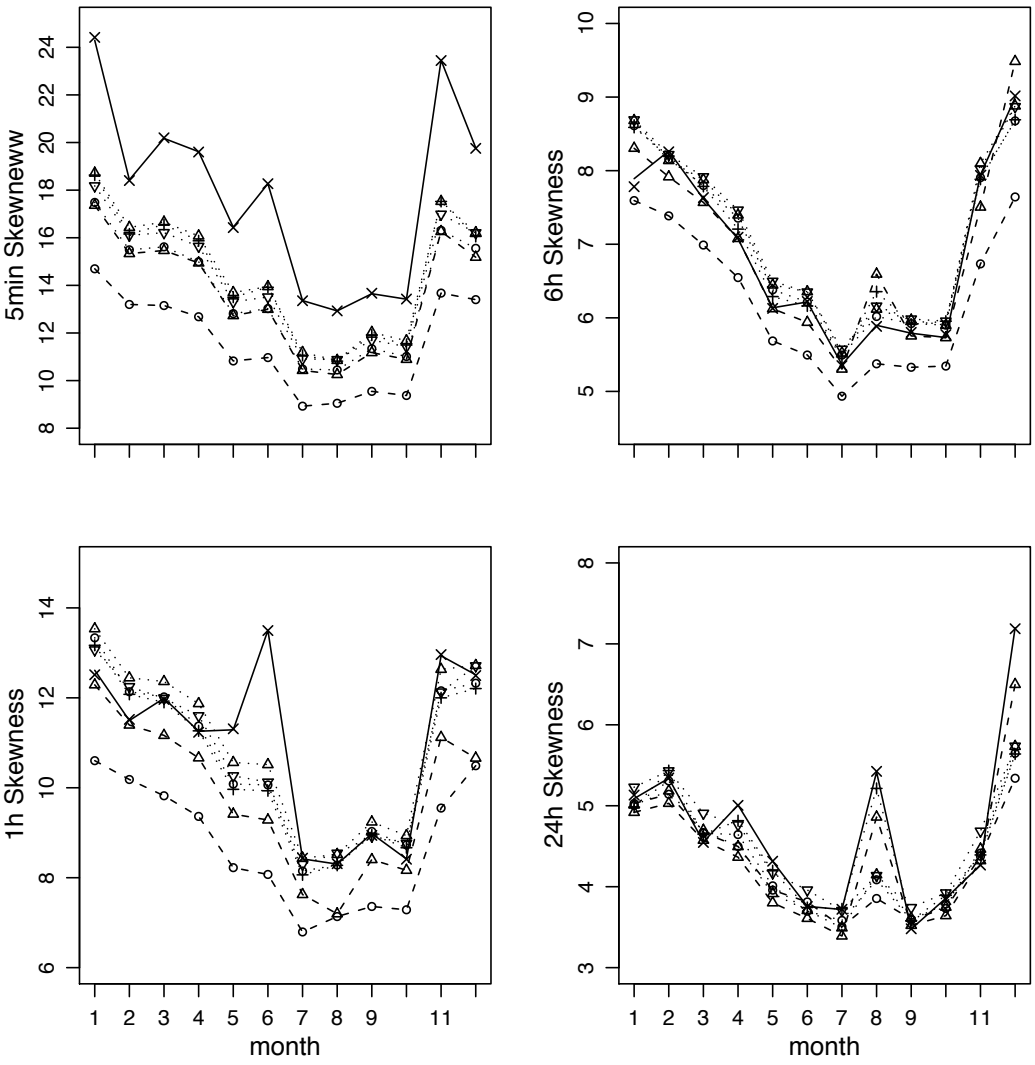

Figure 7.3: Model comparison (single process models): skewness aggregated at the 5-min, 1, 6, and 24 hours levels.

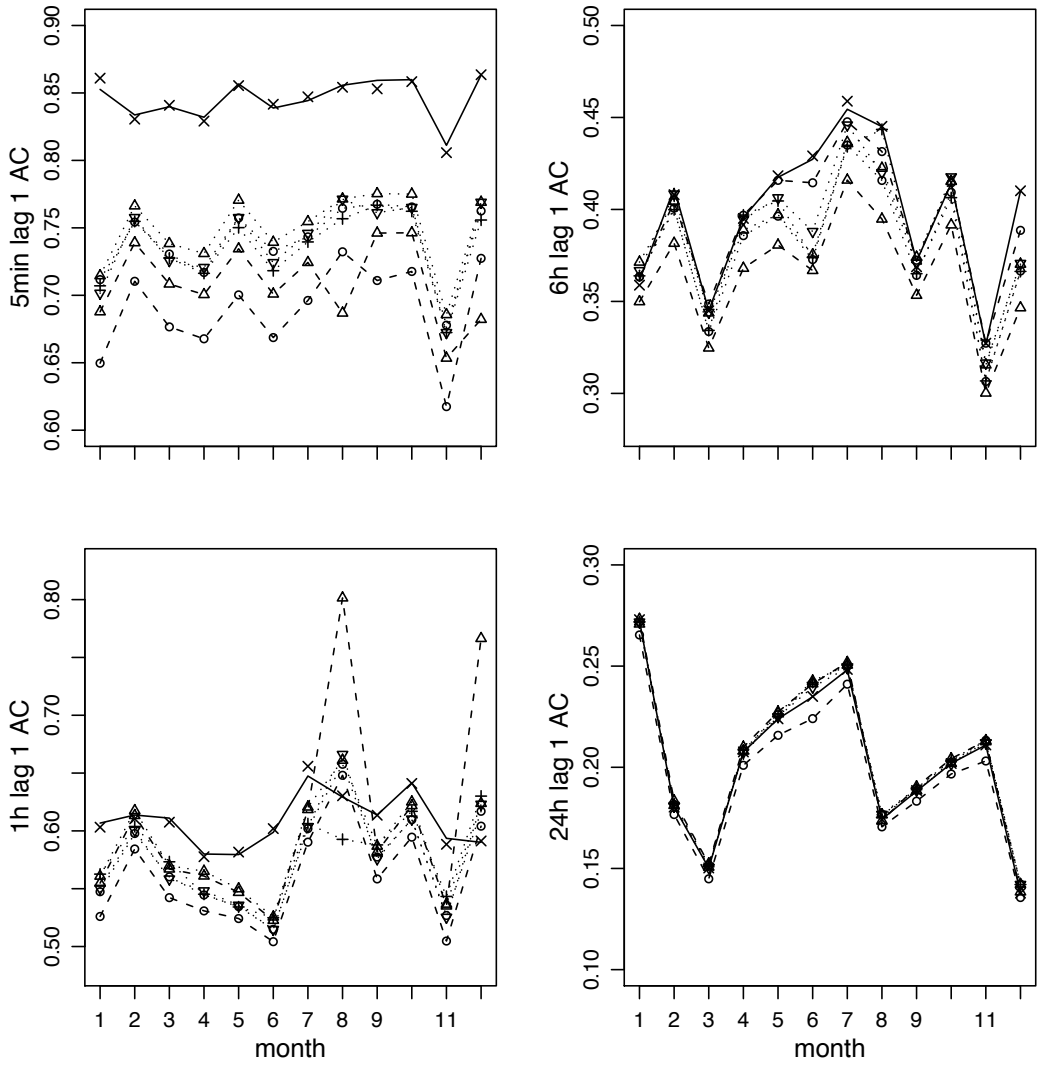

Figure 7.4: Model comparison (single process models): lag 1 autocorrelation aggregated at the 5-min, 1, 6, and 24 hours levels.

7.4 Model comparisons: superposition of two independent BLP processes

In this section, we want to investigate how much improvement of model fits can be achieved by increasing the number of parameters in a two-processes-superposed model setting as specified in FIT-C.

Using FIT-C as the standard goodness-of-fit model, three more BLP models are specified and compared. FIT-C-8 is a reduced FIT-C model in which a common β (the rain cell origins generation rate), a common γ (mean storm lifetime), and a common η (mean rain cell duration) are assumed in representing the two independent processes of storms. There are eight parameters in total with FIT-C-8, $\{\lambda_i, \xi_i, \beta, \gamma, \eta, \theta : i = 1, 2\}$, three less than the parameters in FIT-C, $\{\lambda_i, \beta_i, \xi_i, \gamma_i, \eta_i, \theta : i = 1, 2\}$. FIT-C-9 is constructed in a similar way by assuming a common β (mean rain cell generation rate) and a common γ (mean storm lifetime) based on FIT-C. Therefore, there are nine parameters in total with FIT-C-9, $\{\lambda_i, \xi_i, \eta_i, \beta, \gamma, \theta : i = 1, 2\}$. In FIT-C-10, a common γ (mean storm lifetime) is assumed based on FIT-C. There are 10 parameters in total with FIT-C-10, $\{\lambda_i, \beta_i, \xi_i, \eta_i, \gamma, \theta : i = 1, 2\}$. The difference between FIT-C-10 and FIT-C is just by one parameter. Therefore, we increase the number of model parameters by one each time, starting from FIT-C-8. We then can examine how much model fits can be improved in each step using FIT-C as the standard of a 'good' fit.

7.4.1 Parameter estimates

The resulting parameter estimates for model FIT-C-8, FIT-C-9, and FIT-C-10 are presented in Tables 7.6, 7.7, and 7.8 respectively.

Table 7.6 shows that, except for August and December, the parameter estimates $\hat{\gamma}$, $\hat{\eta}$, and $\hat{\theta}$ are close to those estimates presented in Table 7.4 for model FIT-C1-βzξz. On the other hand, the parameter estimates in Table 7.6 are very different from those presented in Table 6.2 for FIT-C. This implies that the model performance of FIT-C-8 may behave more like a single process model rather than a two-processes-superposed model. The parameter estimates of August and December in Table 7.6 are extraordinarily different from other months and this is the results of fitting to historical values involve extreme rainfall events as we have noted before (Sections 5.3.3, 6.4.4, 7.3.2).

Table 7.7 shows that the parameter estimates of FIT-C-9 have similar features to

Table 7.6: Parameter estimates for each month (m): model FIT-C-8. The units are hour^{-1} for all estimates except θ which has units mm.

m	$\hat{\lambda}_1$	$\hat{\xi}_1$	$\hat{\lambda}_2$	$\hat{\xi}_2$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\eta}$	$\hat{\theta}$
1	0.00494	25.1	0.00395	79.1	0.467	0.0647	1.28	0.0419
2	0.00838	64.8	0.000321	177	1.07	0.12	1.38	0.0277
3	0.0207	8.01	0.00813	85.8	0.746	0.13	1.37	0.0333
4	0.0113	45.5	0.00211	135	0.990	0.105	1.71	0.0293
5	0.0168	61.6	0.00328	248	1.18	0.101	2.05	0.0149
6	0.0870	35.0	0.00507	776	1.33	0.0988	2.63	0.00524
7	0.0157	61.0	0.00109	181	1.14	0.0906	1.52	0.0182
8	0.0792	695.0	0.00479	15600	6.94	0.154	5.91	0.000154
9	0.0120	32.2	0.0103	104	0.668	0.104	1.27	0.0183
10	0.0207	10.9	0.00944	97.6	0.963	0.107	1.30	0.0186
11	0.0149	45.0	0.00130	135	0.447	0.0828	1.40	0.0392
12	0.0155	1750.0	0.00172	10600	5.88	0.185	5.59	0.000431

those obtained with FIT-C for most months of year. January, August, and December are extraordinarily different in terms of $\hat{\xi}_1$, $\hat{\xi}_2$, and $\hat{\theta}$. Again, this is likely due to the extreme rainfall events. Note also the extreme small values of the storm origins generations rate estimates (process 2) $\hat{\lambda}_2$. These extraordinary or extreme parameter estimates indicate that we need more than nine parameters in a model to represent the Kelburn data well.

Parameter estimates presented in Table 7.8 are more similar to those given in Table 6.2 for FIT-C for most months of year. Noticeable differences are found with estimates for January, February, August and November. The sudden change in $\hat{\lambda}_1$ values in August is more substantial. Note the extremely small value $\hat{\lambda}_2 = 0.00000196$ in the November parameter estimates, which is accompanied by the largest mean rainfall depth estimate $\hat{\theta} = 0.0175$. This is similar to the pattern of November parameter estimates given in Table 6.2 which can cause substantial underestimation bias of the historical skewness at the simulation stage as noted in Section 6.4.2.

7.4.2 Derived properties

As in Section 7.3.2, the statistical properties of coefficient of variation (CV), coefficient of skewness, and Lag 1 autocorrelation, are derived for these reduced FIT-C models FIT-C-8, FIT-C-9, and FIT-C-10. The derived properties are plotted against the historical values to check the model fits in Figure 7.5 through to Figure 7.8.

Table 7.7: Parameter estimates for each month (m): model FIT-C-9. The units are hour^{-1} for all estimates except θ which has units mm.

m	$\hat{\lambda}_1$	$\hat{\xi}_1$	$\hat{\eta}_1$	$\hat{\lambda}_2$	$\hat{\xi}_2$	$\hat{\eta}_2$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\theta}$
1	0.00641	2560	0.526	0.00142	18700	6.20	0.207	0.0475	0.000684
2	0.00760	197	0.381	0.00383	1060	4.68	0.346	0.101	0.00725
3	0.0136	185	0.537	0.00173	1290	4.62	0.328	0.112	0.00862
4	0.0112	352	0.284	0.00734	2300	4.93	0.132	0.0620	0.00388
5	0.0132	183	1.42	0.00000170	5760	3.57	0.941	0.0970	0.00969
6	0.0180	81.3	0.507	0.00112	896	2.45	0.536	0.0838	0.00980
7	0.0158	153	0.265	0.00545	1010	3.69	0.174	0.0584	0.00621
8	0.0152	1880	5.36	0.0000298	17000	6.05	6.88	0.154	0.000633
9	0.0174	186	0.638	0.00182	1100	4.02	0.396	0.0943	0.00601
10	0.0143	276	0.396	0.00476	1600	4.49	0.328	0.0886	0.00376
11	0.0141	117	0.924	0.00000105	6550	4.90	0.310	0.0745	0.0175
12	0.00992	2870	4.83	0.000111	18000	5.74	5.81	0.185	0.000570

Table 7.8: Parameter estimates for each month (m): model FIT-C-10. The units are hour^{-1} for all estimates except θ which has units mm.

m	$\hat{\lambda}_1$	$\hat{\beta}_1$	$\hat{\xi}_1$	$\hat{\eta}_1$	$\hat{\lambda}_2$	$\hat{\beta}_2$	$\hat{\xi}_2$	$\hat{\eta}_2$	$\hat{\gamma}$	$\hat{\theta}$
1	0.00608	0.263	1690	0.586	0.0221	0.0113	13200	6.17	0.0512	0.00106
2	0.00895	0.222	242	0.326	0.00241	0.660	1160	4.85	0.0951	0.00618
3	0.0144	0.299	197	0.528	0.00121	0.548	1310	4.75	0.111	0.00811
4	0.0135	0.507	205	0.995	0.000246	2.36	1350	6.57	0.0965	0.00841
5	0.0224	0.220	196	0.438	0.00145	1.26	1190	3.49	0.0837	0.0051
6	0.0186	0.453	80.2	0.441	0.00123	0.522	880	2.44	0.0817	0.00969
7	0.0168	0.659	137	0.824	0.000238	2.02	1020	4.72	0.0848	0.00794
8	0.0792	0.149	883	0.627	0.00192	5.99	3570	6.16	0.127	0.000928
9	0.0182	0.475	177	0.825	0.000364	0.71	1280	4.51	0.0985	0.00705
10	0.0161	0.241	335	0.355	0.00308	0.608	1760	4.63	0.0847	0.00317
11	0.0143	0.314	115	0.928	0.00000196	0.447	4770	5.08	0.075	0.0175
12	0.0135	1.35	482	1.59	0.000207	6.71	2700	7.13	0.164	0.00319

Correspondingly, Figure 7.5 is an enlargement of the top-left plot of Figure 7.6 with graphical symbols explained. In all figures, again, the historical values are represented by points (crosses). The solid lines represent the derived properties values of model FIT-C and it is used as the standard of 'good' fit. In the following figures, dotted lines with circle points are for model FIT-C-8; dotted lines with upward triangle points for FIT-C-9; dotted lines with '+' sign points for FIT-C-10.

The fits of CV at the 5-min aggregation level are examined in Figure 7.5. The dotted lines representing model FIT-C-8 clearly deviate (overfit) from the solid lines and other two models are regarded as good as FIT-C. The fits of CV at the 1, 6, and 24 hours levels are examined in Figure 7.6. A few small deviation points are identified with FIT-C-9, e.g. August and December fits at the 1 hour level, August fit at the 24 hours level. Overall, we conclude that FIT-C-8 has the worst fits and other two fitted models are as good as FIT-C in most cases examined.

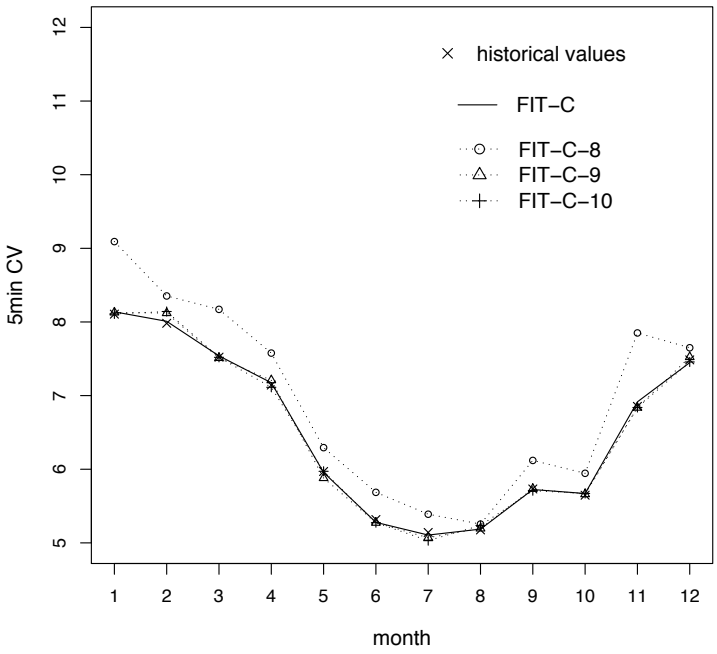

Figure 7.5: Model comparison (two processes superposed models): coefficient of variation aggregated at the 5-min level.

Figures 7.7 and 7.8 examine the fits of skewness and lag 1 autocorrelation at the 5-min, 1, 6, and 24 hours levels, respectively. Substantial underfit deviations appear in

the skewness fits at the 5-min, 1 hour levels and in the autocorrelation fits at the 5-min level with model FIT-C-8. In other cases, FIT-C-8 is also identified as the worst fit model although the deviations are to a lesser extent. Model FIT-C-10 fits the skewnesses and the autocorrelations as good as FIT-C in most cases. A few small deviating points are found in fitting the skewness at the 6 hours level and in fitting the lag 1 autocorrelation at the 1 hour level with FIT-C-10. Model FIT-C-9 improves upon FIT-C-8 substantially but is slightly worse than FIT-C-10. Overall, FIT-C-10 has the smallest deviation from FIT-C as we expected. We further examine the model fits in Table 7.9 based on the objective function minimisation values and the number of model parameters.

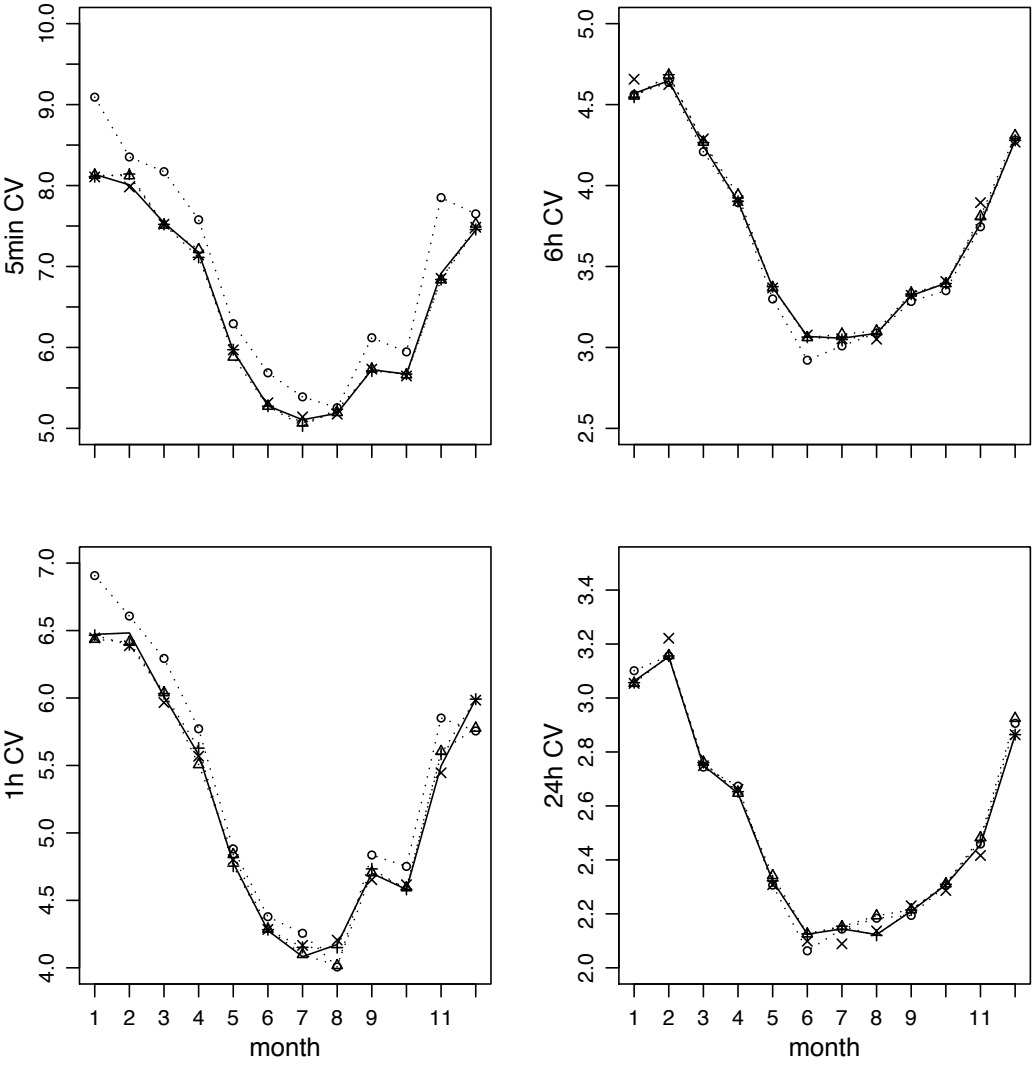

Figure 7.6: Model comparison (two processes superposed models): coefficient of variation aggregated at the 5-min, 1, 6, and 24 hours levels.

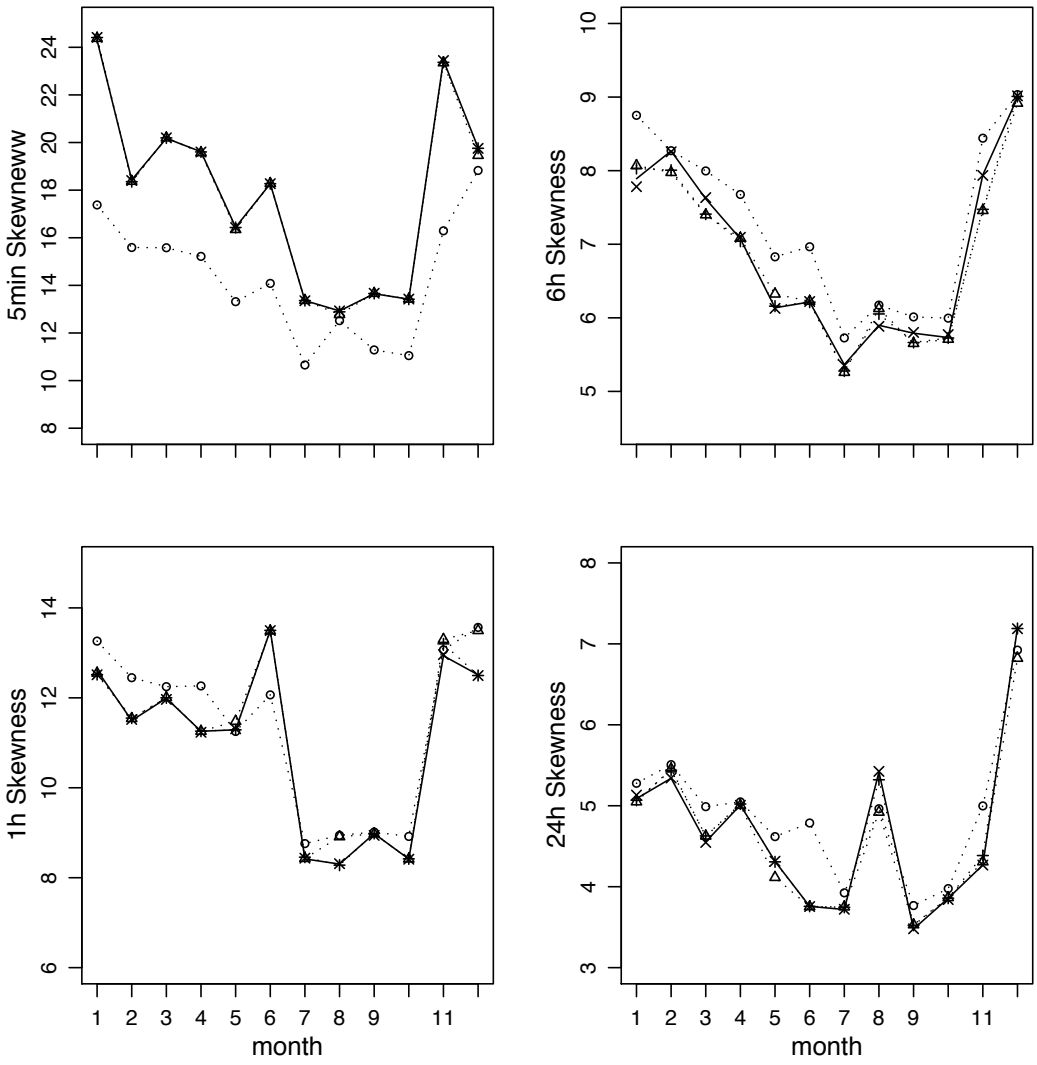

Figure 7.7: Model comparison (two processes superposed models): skewness aggregated at the 5-min, 1, 6, and 24 hours levels.

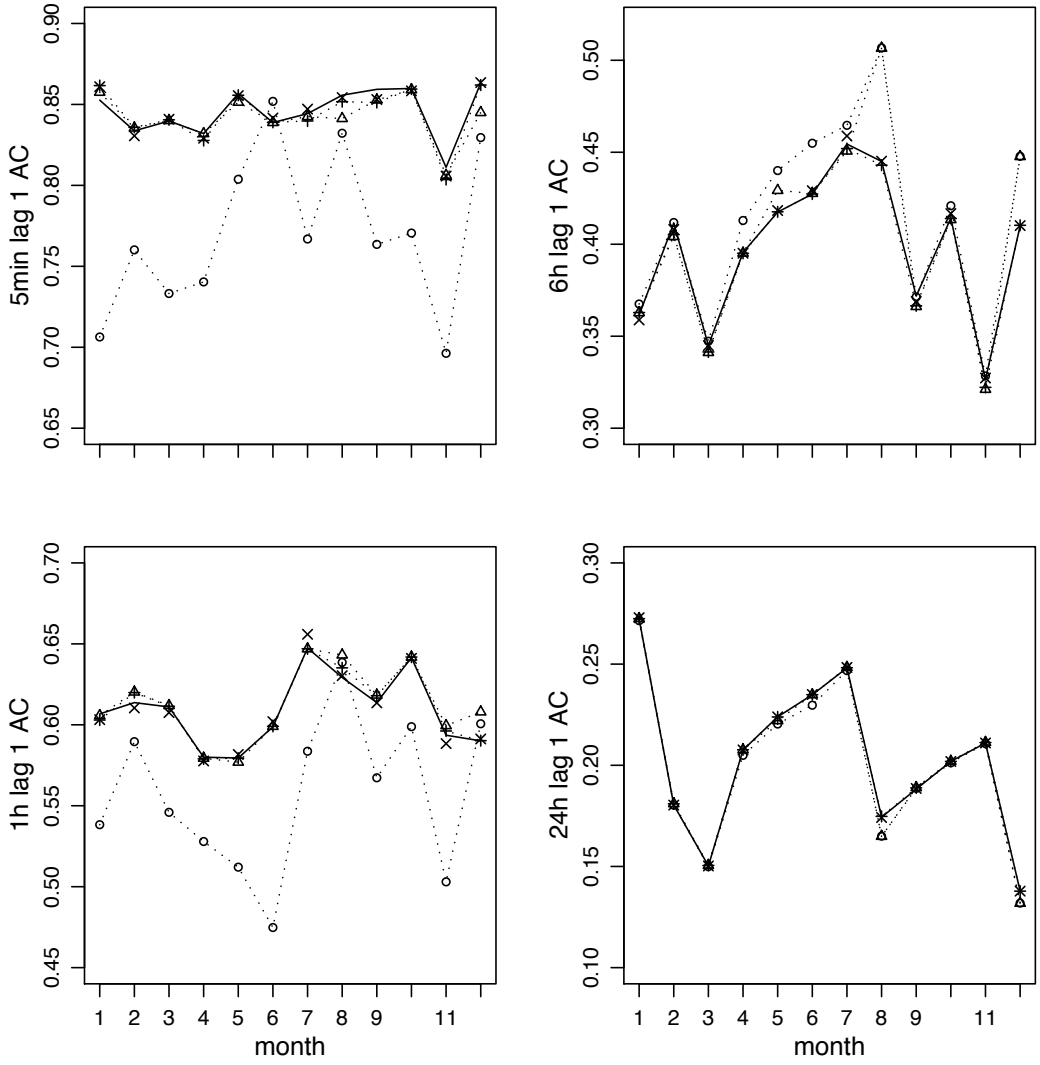

Figure 7.8: Model comparison (two processes superposed models): lag 1 autocorrelation aggregated at the 5-min, 1, 6, and 24 hours levels.

Table 7.9 is designed in the same way as in Table 7.5 for presenting the numerical results of model comparison. FIT-O and FIT-C are also included in Table 7.9 for comparison with those three reduced FIT-C models. From Table 7.9, we note that FIT-C-8 has achieved the same minimised value (total) as by model FIT-C1-βzξzθz (Table 7.5) and both models have the same number of parameters, eight. But these two models have been conceptualized differently. By increasing one more parameter in the model, FIT-C-9 makes a huge gain in the model fits upon FIT-C-8. The minimised total value decreases dramatically from 2.68 in FIT-C-8 to 0.170 in FIT-C-9 and the maximal relative error is reduced from 49.4% to 19.9%. As we have noted in Table 7.6, the model performance of FIT-C-8 is more like a single process BLP model. The model fit gain from FIT-C-9 to FIT-C-10 is also substantial, the maximal relative error reduced from 19.9% to 8.34%. The above comparison results suggest that it is necessary to have at least nine parameters in a BLP model to obtain a reasonable goodness-of-fit to the Kelburn data.

Measured by the minimised objective function values and the relative errors, FIT-O and FIT-C fit the Kelburn data better than FIT-C-10 but the difference is regarded as minor. The goodness-of-fit differences between FIT-O and FIT-C may be considered as numerically insignificant. However, as the simulation study results show in Chapter 6, FIT-C improves significantly upon FIT-O in fitting the extreme rainfalls and the proportion dry values. A simulation study is needed to compare FIT-C-10 and FIT-C in more detail to check what differences in model performance can be caused by a 4% difference in the maximal relative error of the derived properties.

Table 7.9: Model comparison in terms of minimisation function value and the total number of parameters: two processes superposed models

	FIT-C-8	FIT-C-9	FIT-C-10	FIT-O	FIT-C
minimised value (total)	2.68	0.170	0.0384	0.0118	0.0118
minimised value (maximum)	0.489	0.0795	0.0139	0.00345	0.00339
relative error % (maximum)	49.4	19.9	8.34	4.15	4.12
parameters	8	9	10	11	11

7.5 Concluding remarks

In this chapter, a BLP model with continuous distributions of storm types is formulated to represent a single process of storms which is characterized as consisting of a continuum of storm types. In the continuous-storm-types BLP process approach, model parameters are expressed as functions of a continuous random variable z. The original BLP model (i.e. FIT-O1) and FIT-C1 can be recovered as a special case by setting the model parameters to be constants. By taking z to be a uniform random variable and define parameter functions $\beta(z)$, $\xi(z)$, and $\theta(z)$, four typical continuous-storm-types BLP models are specified. The model performances are compared between FIT-O1, FIT-C1, and the four continuous-storm-types BLP models based on the derived model properties, the objective function minimisation values, and the number of parameters.

The comparison results show that, given one or two more parameters specified in a model, the continuous-storm-types BLP models substantially improve upon FIT-O1 and FIT-C1. However, all fitted models are far from adequate. They are very poor in fitting to the corresponding historical values.

A superposed BLP model (of two independent processes) with eight parameters can perform as good as the best one (with eight parameters) of those single process continuous-storm-types BLP models. A superposed BLP model (of two independent processes) with nine parameters can perform much better than the eight-parameter models (superposed or single process). Note that even though FIT-C1 substantially outperforms FIT-O1 (Table 7.5), the superposition model of two FIT-C1 processes (i.e. FIT-C) has show no improvement upon the superposition model of two FIT-O1 processes (i.e. FIT-O) based on the minimisation values (Table 7.9).

Therefore, the research results of this chapter suggest that superposition of two independent BLP processes is necessary for an adequate fit to the Kelburn data, although the continuous-storm types characterization improves model performance substantially in single process model cases. We need to compare the continuous-storm-types BLP models with FIT-C1 given the same number of parameters in one model. If a continuous-storm-types BLP model could outperform FIT-C1 given the same number of model parameters (i.e. six), the superposition of two such processes should then be investigated to compare the model performance against FIT-C.

Chapter 8

Superposition of continuous-storm-types BLP processes

In Chapter 7, we proposed an alternative BLP model characterization framework, the BLP model with continuous distributions of storm types. The model comparison results show that, given one or two more parameters specified in a model, the fitted continuous-storm-types BLP models outperform FIT-O1 and FIT-C1 substantially. The model comparison results in the cases of superposition of two independent BLP processes show that a superposed BLP model performs much better than a single process model and the number of model parameters is the dominant factor of model performance measured by the minimisation objective function values. We have also noted that, although FIT-C1 substantially outperforms FIT-O1, we can not distinguish FIT-C (the superposition of two independent FIT-C1 processes) from FIT-O (the superposition of two independent FIT-O1 processes) in model performance based on the minimisation values. These findings lead to the research work of this chapter: (1) we want to formulate some continuous-storm-types BLP models which should have the same number of model parameters as FIT-C1 (i.e. six) and compare these models with FIT-C1 (Section 8.1); (2) we will construct a superposed continuous-storm-types BLP model, based on the result of (1), and make comparison to FIT-C through simulation study (Sections 8.1 and 8.2); (3) the model performance of FIT-C-10 will also be examined in comparison with the superposed continuous-storm-types BLP model and FIT-C (Section 8.2). We are aiming to draw a conclusion about the effect of superposition of two continuous-storm-types BLP processes and the impact on model performances by reducing one parameter from FIT-C (Section 8.3) .

The R programs (for model estimation) and the C programs (for generation of simulation samples) used for the study of this chapter are obtained by corresponding modifications from those programs used in Chapter 6. For model performance assessment and comparison of different models, the methodology used in Chapter 6 is adopted, i.e. the layout of the tables and figures, and the interpretation of the model fitting and simulation results follows the same convention as we have explained in Chapter 6.

8.1 Superposition of two continuous-storm-types BLP processes

8.1.1 Continuous-storm-types BLP models of six parameters

We concentrate our study on the relationship between parameters β and θ so that we can examine how the frequency of rainfall events is associated with the rainfall fall intensity. From the results in Table 7.5 we note that model FIT-C1-βzθz substantially improves upon FIT-C1 (total minimisation values reduced from 5.60 to 2.72) but FIT-C1-βzθz has one more parameter than FIT-C1 (seven versus six). We note that, from the parameter estimates Table 7.2, the rain cell generation rates (β) are negatively associated with the mean pulse depths (θ). Therefore, an obvious six-parameter parametrisation is to set the parameter functions $\beta(z)$, $\xi(z)$, and $\theta(z)$ being defined by

$$\left.\begin{array}{rcl} \beta(z) & = & \beta(1-z) \\ \xi(z) & = & \xi \\ \theta(z) & = & \theta z, \end{array}\right\} \tag{8.1}$$

where z is a uniform random variable with pdf $f(z) = 1; \quad 0 < z < 1$. We denote this six-parameter continuous-storm-types BLP model by FIT-C1-βzθz(6). This model specification has the properties that all types of storms are assumed to occur with an equal probability and the association between β and θ is linear as in FIT-C1-βzθz. In particular, as $z \to 0$, $\beta(z) \to \beta$ and $\theta(z) \to 0$; as $z \to 1$, $\beta(z) \to 0$ and $\theta(z) \to \theta$.

Based on Equation (7.5), we have

$$\left.\begin{array}{rcl} W_{111} &=& \frac{1}{6}\beta\xi\theta \\ W_{112} &=& \frac{1}{12}\beta\xi\theta^2 \\ W_{222} &=& \frac{1}{30}\beta^2\xi^2\theta^2 \\ W_{122} &=& \frac{1}{12}\beta\xi^2\theta^2 \\ W_{233} &=& \frac{1}{60}\beta^2\xi^3\theta^3 \\ W_{333} &=& \frac{1}{140}\beta^3\xi^3\theta^3 \\ W_{123} &=& \frac{1}{20}\beta\xi^2\theta^3 \\ W_{223} &=& \frac{1}{60}\beta^2\xi^2\theta^3 \\ W_{113} &=& \frac{1}{20}\beta\xi\theta^3 \\ W_{133} &=& \frac{1}{20}\beta\xi^3\theta^3. \end{array}\right\} \tag{8.2}$$

Therefore, given (8.2), the statistical properties up to third order for FIT-C1-βzθz(6) can be obtained from Equations (7.4), (7.6), (7.7), and (7.9).

It can be shown that, from the parameter estimates of BLP models (e.g. Table 6.2), the majority of rainfalls (more than 80%) in the observed rainfall series are contributed from the more frequent but less intensive rainfall process (process 1 in our case). It is, therefore, reasonable to assume an exponential random variable z so that rainfall events that are more frequent but with lower intensity have a higher probability to occur. We denote such a model by FIT-C1-βzθzE and the parameter functions $\beta(z)$, $\xi(z)$, and $\theta(z)$ are defined by

$$\left.\begin{array}{rcl} \beta(z) &=& \beta e^{-z} \\ \xi(z) &=& \xi \\ \theta(z) &=& \theta z, \end{array}\right\} \tag{8.3}$$

where z is a unit mean exponential random variable with pdf $f(z) = e^{-z}; \quad 0 < z < +\infty$. To get the integral expressions equivalent to (8.2), we now need to integrate over

the range of $0 < z < +\infty$ with respect to $f(z)$ and we have

$$\left.\begin{array}{rcrcl}
W_{111} &=& \int_0^{+\infty} \beta(z)\xi(z)\theta(z)\, e^{-z}\mathrm{d}z &=& \frac{1}{4}\beta\xi\theta \\
W_{112} &=& \int_0^{+\infty} \beta(z)\xi(z)\big[\theta(z)\big]^2 e^{-z}\mathrm{d}z &=& \frac{1}{4}\beta\xi\theta^2 \\
W_{222} &=& \int_0^{+\infty} \big[\beta(z)\big]^2\big[\xi(z)\big]^2\big[\theta(z)\big]^2 e^{-z}\mathrm{d}z &=& \frac{2}{27}\beta^2\xi^2\theta^2 \\
W_{122} &=& \int_0^{+\infty} \big[\beta(z)\big]\big[\xi(z)\big]^2\big[\theta(z)\big]^2 e^{-z}\mathrm{d}z &=& \frac{1}{4}\beta\xi^2\theta^2 \\
W_{233} &=& \int_0^{+\infty} \big[\beta(z)\big]^2\big[\xi(z)\big]^3\big[\theta(z)\big]^3 e^{-z}\mathrm{d}z &=& \frac{2}{27}\beta^2\xi^3\theta^3 \\
W_{333} &=& \int_0^{+\infty} \big[\beta(z)\big]^3\big[\xi(z)\big]^3\big[\theta(z)\big]^3 e^{-z}\mathrm{d}z &=& \frac{3}{128}\beta^3\xi^3\theta^3 \\
W_{123} &=& \int_0^{+\infty} \big[\beta(z)\big]\big[\xi(z)\big]^2\big[\theta(z)\big]^3 e^{-z}\mathrm{d}z &=& \frac{3}{8}\beta\xi^2\theta^3 \\
W_{223} &=& \int_0^{+\infty} \big[\beta(z)\big]^2\big[\xi(z)\big]^2\big[\theta(z)\big]^3 e^{-z}\mathrm{d}z &=& \frac{2}{27}\beta^2\xi^2\theta^3 \\
W_{113} &=& \int_0^{+\infty} \big[\beta(z)\big]\big[\xi(z)\big]\big[\theta(z)\big]^3 e^{-z}\mathrm{d}z &=& \frac{3}{8}\beta\xi\theta^3 \\
W_{133} &=& \int_0^{+\infty} \big[\beta(z)\big]\big[\xi(z)\big]^3\big[\theta(z)\big]^3 e^{-z}\mathrm{d}z &=& \frac{3}{8}\beta\xi^3\theta^3.
\end{array}\right\} \tag{8.4}$$

Similarly, given (8.4), the statistical properties up to third order for FIT-C1-βzθzE can be obtained from Equations (7.4), (7.6), (7.7), and (7.9). In FIT-C1-βzθzE, those types of storms with more frequent rainfall events but less intensive are assumed to occur with higher probabilities as being determined by the exponential random variable pdf $f(z) = e^{-z}$. The association between β and θ is non-linear as specified in (8.3). Therefore, as $z \to 0$, $\beta(z) \to \beta$ and $\theta(z) \to 0$; as $z \to +\infty$, $\beta(z) \to 0$ and $\theta(z) \to +\infty$. The converging speed of the pdf $f(z) = e^{-z}$ is of a higher order than the diverging speed of $\theta(z)$ so that the evaluation of (8.4) converges.

The parameter estimates of FIT-C1-βzθz(6) and FIT-C1-βzθzE are given in Table 8.1. Both models specify the negative association relationships between the rain cell origins generation rates (β) and the mean pulse depths (θ) but through different continuous random variables. Therefore, we note the substantial differences and consistent patterns in parameter estimates $\hat{\beta}$ and $\hat{\theta}$ while $\hat{\gamma}$ and $\hat{\eta}$ do not differ much. Note also the substantial differences of the August and December $\hat{\lambda}$ values in comparison to other months $\hat{\lambda}$ values which are very close. This implies that the parametrisation is sensitive to the extreme rainfall events occurred in August and December as we commented in Section 5.3.3 and 7.3.

As specified in (8.1) and (8.3), both FIT-C1-βzθz(6) and FIT-C1-βzθzE are six-parameter continuous-storm-types BLP models. Model comparison results (Table 8.3) show that FIT-C1-βzθzE performs much better than FIT-C1-βzθz(6) (more on model comparison in Section 8.1.3). Therefore, FIT-C1-βzθzE is chosen to form a superposed continuous-storm-types BLP model.

Table 8.1: Parameter estimates: for each month (m) the estimates are listed in the order of model FIT-C1-βzθz(6) and FIT-C1-βzθzE. The units are hour^{-1} for all estimates except θ which has units mm.

m	$\hat{\lambda}$	$\hat{\beta}$	$\hat{\xi}$	$\hat{\gamma}$	$\hat{\eta}$	$\hat{\theta}$
1	0.00820	1.07	62.0	0.0628	1.15	0.0824
	0.00846	1.83	69.0	0.0699	1.17	0.0319
2	0.00963	1.93	71.6	0.112	1.08	0.0595
	0.00946	3.31	80.0	0.12	1.11	0.0233
3	0.0149	1.44	66.3	0.122	1.13	0.071
	0.0147	2.52	74.6	0.132	1.18	0.0276
4	0.0120	1.55	63.5	0.0930	1.22	0.0753
	0.0122	2.51	70.9	0.0987	1.23	0.0294
5	0.0145	1.6	82.1	0.0862	1.33	0.0580
	0.0149	2.45	97.0	0.0900	1.33	0.0216
6	0.0160	1.31	69.1	0.0759	1.35	0.0805
	0.0172	1.89	79.5	0.0777	1.33	0.0305
7	0.0165	1.70	65.6	0.0802	1.05	0.0486
	0.0168	2.69	73.2	0.0844	1.05	0.0191
8	0.0398	0.141	57.1	0.0494	0.275	0.0444
	0.0295	0.812	58.9	0.0975	0.346	0.0182
9	0.0207	1.47	83.8	0.101	1.13	0.0363
	0.0204	2.56	98.1	0.110	1.17	0.0136
10	0.0172	1.80	74.8	0.100	1.04	0.0406
	0.0170	3.09	83.9	0.107	1.08	0.0158
11	0.0153	0.806	53.6	0.0732	1.12	0.0957
	0.0161	1.29	60.2	0.0792	1.13	0.0368
12	0.0230	0.262	55.7	0.0868	0.345	0.0723
	0.0166	1.06	63.8	0.121	0.484	0.0283

8.1.2 A superposed continuous-storm-types BLP model

A superposed continuous-storm-types BLP model FIT-C-βzθzE is formulated by superposing two independent FIT-C1-βzθzE processes. The statistical properties up to third order for FIT-C-βzθzE can be obtained based on Equations (8.4), (7.4), (6.9), (7.6), (7.7), and (7.9). The parameter estimates are given in Table 8.2. FIT-C-βzθzE has the same number of model parameters as in FIT-C so that model comparison can be made on the same ground. Note that the interpretation of parameters β and θ in Table 8.2 is different from those in Table 6.2 for FIT-C because of the different parametrisation approach as we have noted in Section 7.3.

Table 8.2: Parameter estimates for each month (m): model FIT-C-βzθzE. The units are hour^{-1} for all estimates except θ which has units mm.

m	$\hat{\lambda}_1$	$\hat{\beta}_1$	$\hat{\xi}_1$	$\hat{\gamma}_1$	$\hat{\eta}_1$	$\hat{\lambda}_2$	$\hat{\beta}_2$	$\hat{\xi}_2$	$\hat{\gamma}_2$	$\hat{\eta}_2$	$\hat{\theta}$
1	0.0161	13.7	643	0.430	0.299	0.000576	3.44	11400	0.0209	5.87	0.000616
2	0.00690	4.44	818	0.0902	1.82	0.00158	60.5	2100	0.306	12.1	0.00199
3	0.0160	3.43	259	0.203	0.945	0.000447	3.32	1320	0.0312	8.4	0.00643
4	0.0207	2.33	726	0.173	1.07	0.000553	13.0	2920	0.0616	10.6	0.00251
5	0.0333	2.23	152	0.242	0.474	0.00253	5.02	683	0.0697	3.32	0.00589
6	0.0152	0.683	49.0	0.0540	0.115	0.00934	6.46	488	0.346	2.45	0.00839
7	0.00932	0.917	199	0.0488	0.288	0.0103	9.85	768	0.245	4.47	0.00385
8	0.0347	5.36	280	0.211	1.45	0.00077	26.8	996	0.104	6.95	0.00335
9	0.0143	3.15	470	0.0925	1.19	0.00342	44.9	1310	0.485	8.90	0.00223
10	0.0166	6.97	204	0.255	0.401	0.00320	3.04	1360	0.0372	4.77	0.00255
11	0.0145	1.40	150	0.0752	0.887	0.00112	36.4	807	0.848	9.78	0.0106
12	0.00986	3.51	962	0.231	0.135	0.00704	0.344	7910	0.0168	4.46	0.000666

8.1.3 Model comparison

As in Tables 7.5 and 7.9, we compare FIT-C1-βzθz(6), FIT-C1-βzθzE, and FIT-C-βzθzE based on the minimised objective function values in Table 8.3. The model fit results for FIT-C1, FIT-C1-βzθz, FIT-C-10, and FIT-C are reproduced from Tables 7.5 and 7.9.

The first four models in Table 8.3 are single process BLP models and the last three are superposed BLP models. The improvement from a single process model to a superposed model is huge, but at the cost of increase in the number of model parameters. FIT-C1-βzθz and FIT-C1-βzθz(6) follow the same parametrisation approach and FIT-C1-βzθz performs better as expected because it has one more parameter in the model. Obviously, FIT-C1-βzθzE, in which a non-linear relationship is specified for β (the cell origins generation rates) and θ (the mean pulse depths) through an exponential random variable z, has the best performance among the single process models. Note the substantial improvement made upon FIT-C1 by FIT-C1-βzθzE (the total minimisation value reduced from 5.60 to 1.86). It is worth noting, however, that FIT-C-βzθzE (superposition of two independent FIT-C1-βzθzE processes) does not improve upon FIT-C (superposition of two independent two FIT-C1 processes) in terms of the minimised objective function values (given that both minimised values are very close to zero). The same situation occurs to FIT-O1 and FIT-C1 (the total minimisation value reduced from 13.9 to 5.60, Table 7.5), and FIT-O and FIT-C (the same total minimisation value 0.0118, Table 7.9). However, we have justified that some significant improvements are achieved by FIT-C upon FIT-O according to the simulation results presented in Chap-

ter 6. Therefore, it is necessary to compare the model performances in more detail between FIT-C-βzθzE and FIT-C through simulation study. Both FIT-C-βzθzE and FIT-C outperform FIT-C-10 by about 4% in the maximum relative error, but both models have one parameter more than FIT-C-10. We also want to investigate the impact on model performances by reducing one parameter from FIT-C through simulation study.

Table 8.3: Model comparison in terms of minimisation function value and the total number of parameters

	FIT-C1	FIT-C1-βzθz	FIT-C1-βzθz(6)	FIT-C1-βzθzE	FIT-C-10	FIT-C-βzθzE	FIT-C
minimised value (total)	5.60	2.72	3.25	1.86	0.0384	0.0180	0.0118
minimised value (maximum)	1.14	0.589	0.718	0.359	0.0139	0.00369	0.00339
relative error % (maximum)	75.5	54.3	59.9	42.4	8.34	4.30	4.12
parameters	6	7	6	6	10	11	11

8.2 Comparison of three superposed BLP models by simulation

As in Chapter 6, all figures in the this section have the same layout for comparison between three different model specifications: FIT-C (left column plots), FIT-C-10 (middle column plots) and FIT-C-βzθzE (right column plots). Historical values are represented by points (crosses) and model fitted values or simulation results represented by lines. Solid line represents the medians of the corresponding statistics calculated from 200 simulation samples. Dashed lines form the corresponding 95% quantile confidence band from the simulation samples. Note that the simulation results of FIT-C are reproduced from Chapter 6's figures for convenience to draw conclusions.

8.2.1 Moments

Figure 8.1, 8.2, and 8.3 present the simulation results of moment properties at 5-min, 1 and 24 hours of aggregation levels, respectively.

Let us first examine the fitting results of the 1 hour means (top row panels in

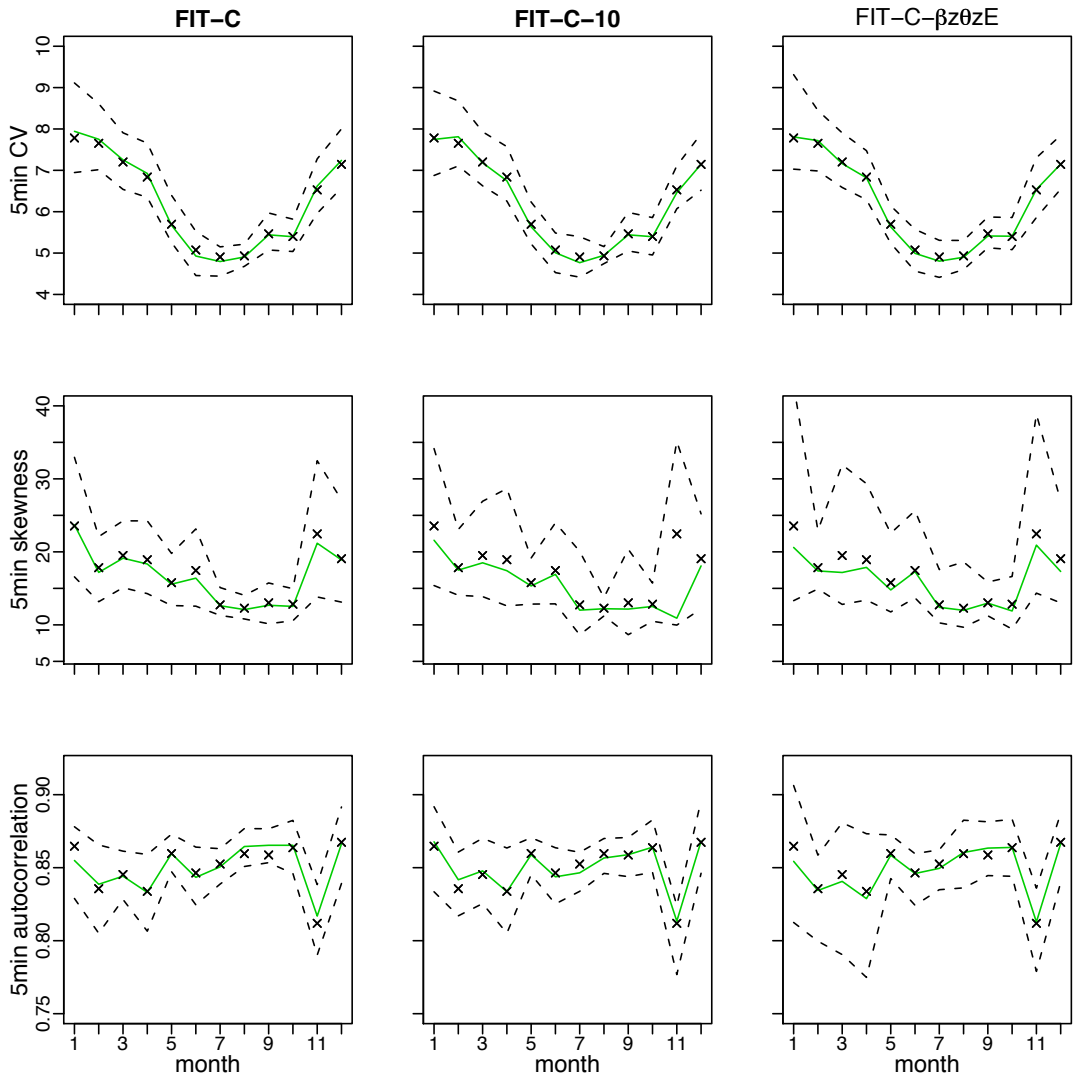

Figure 8.1: Model comparison: coefficient of variation, skewness, and lag 1 autocorrelation at the 5-min aggregation level.

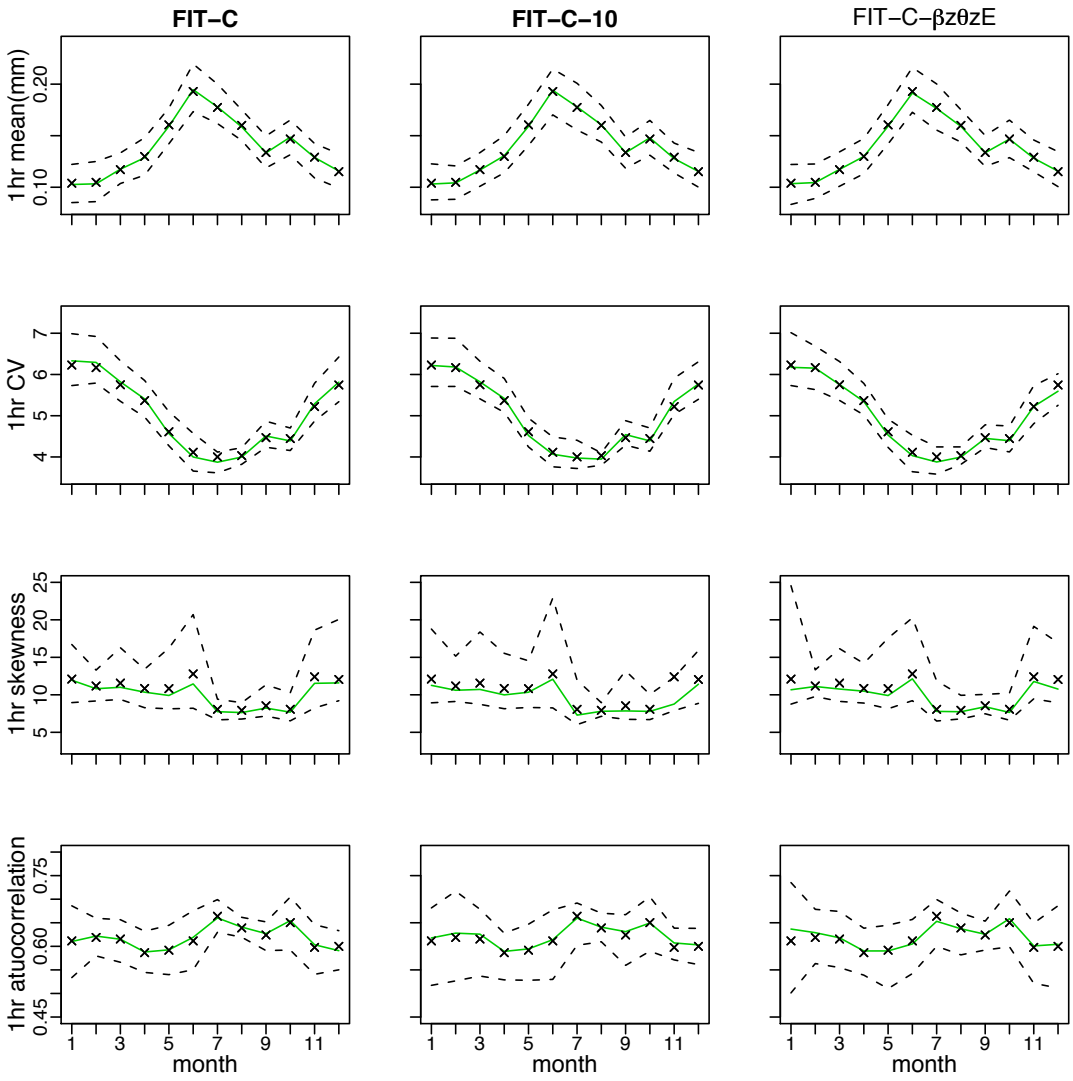

Figure 8.2: Model comparison: mean, coefficient of variation, skewness, and lag 1 autocorrelation at the 1 hour aggregation level.

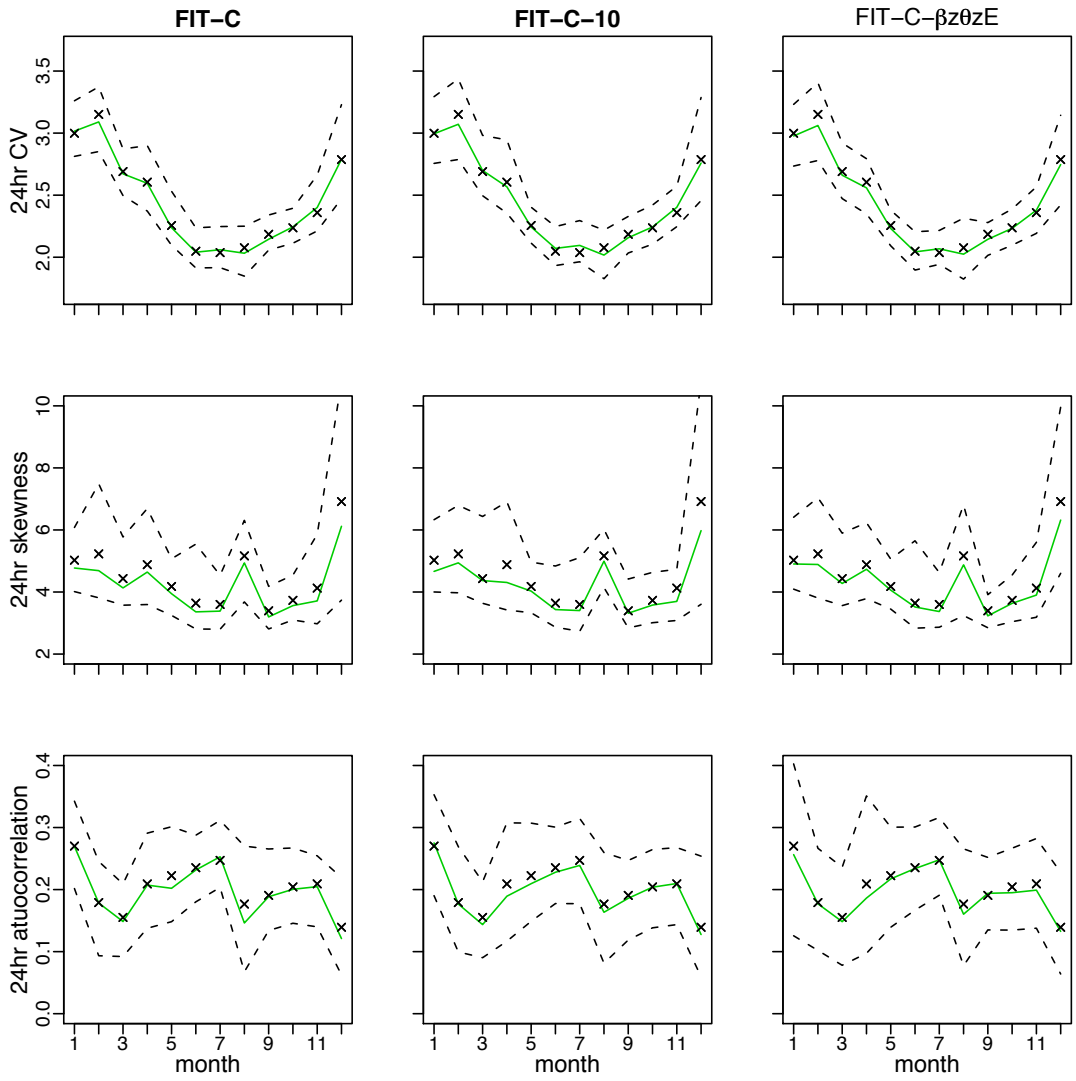

Figure 8.3: Model comparison: coefficient of variation, skewness, and lag 1 autocorrelation at the 24 hours aggregation level.

Figure 8.2) and of the coefficients of variation (top row panels in Figure 8.1 and 8.3, and second row panels in Figure 8.2). All historical values, except for one or two cases of small deviations from the median lines (e.g. February points at the 24 hours level), are well located within the 95% quantile confidence band. It is a clear-cut adequate fit case.

For the fitting of skewness (second row panels in Figure 8.1 and 8.3, and the third row panels in Figure 8.2), while we can still claim a good fit in general, a slight underestimation of the historical skewness values exists for all three model specifications just as we have identified and discussed in Section 6.4.2 and a few big deviation points are worth commenting. At the 5-min aggregation level (Figure 8.1), FIT-C-βzθzE and FIT-C-10 have slightly larger deviations (from the median lines) than FIT-C in January, March, and April. Note the big deviation point in November with FIT-C-10. This is caused by a very small $\hat{\lambda}_2$ value associated with a relatively large $\hat{\theta}$ value as mentioned in commenting Table 7.4. A big deviation point in November at the 1 hour aggregation level (Figure 8.2) is also identified with FIT-C-10. FIT-C-βzθzE slightly outperforms the other two models at the 24 hours level (Figure 8.3).

Overall, we conclude that no real difference is found between FIT-C and FIT-C-βzθzE in fitting the moment properties and both models fit the moment properties adequately. Significant lack-of-fit is found in November at the 5-min and 1 hour aggregation levels with FIT-C-10.

8.2.2 Extreme values

Figure 8.4 and 8.5 examine the annual maxima plots at five different aggregation time scales, 5-min, 1, 6, 12, and 24 hours. Figure 8.6 through to Figure 8.9 compare the annual maxima patterns for four individual months, December and January (two summer months), and June and July (two winter months). Same as in Chapter 6, ordered annual maxima plotted against the reduced Gumbel variate. The values from the historical series are shown as points (crosses). Solid line represents the medians of ordered annual maxima calculated from 200 simulation samples. Dashed lines form the corresponding 95% quantile confidence band.

In Chapter 6, we have concluded that FIT-C fits extreme values (annual maxima) adequately at all aggregation levels examined. When further breaking down the annual maxima patterns by month, we have found that FIT-C improves significantly upon the original BLP model FIT-O, although the results may not be considered as an adequate

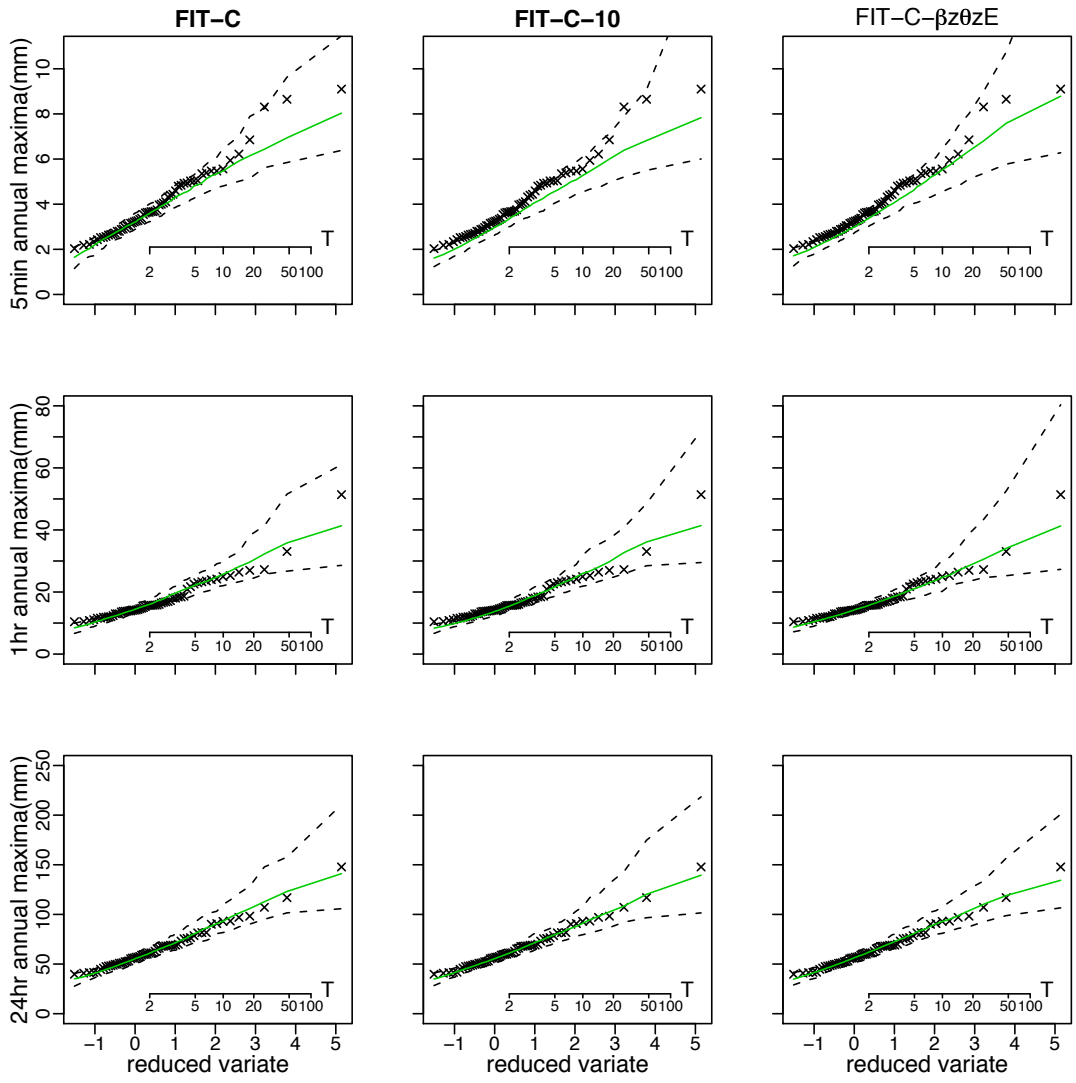

Figure 8.4: Model comparison: extreme value plots at the 5-min, 1, and 24 hours aggregation levels.

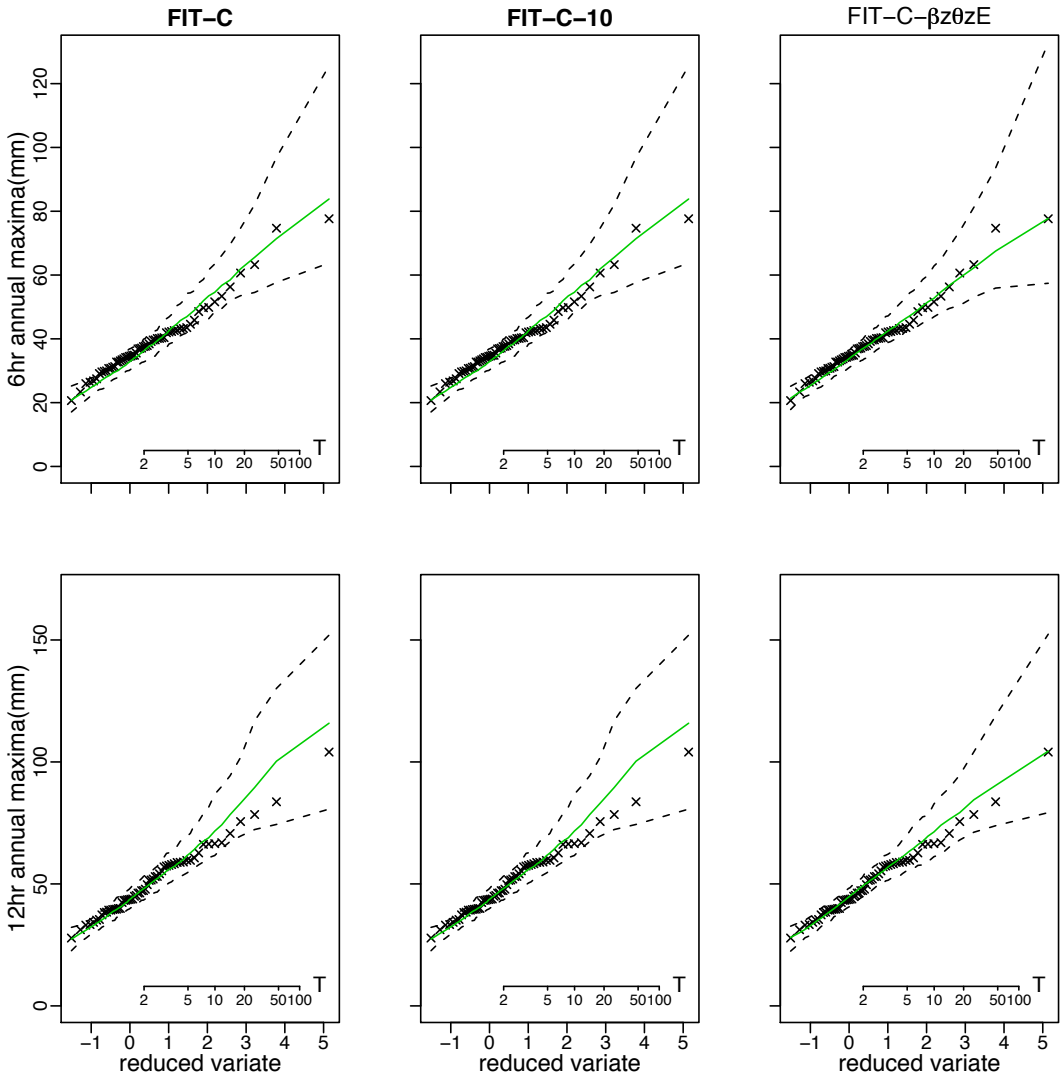

Figure 8.5: Model comparison: extreme value plots at the 6 and 12 hours aggregation levels.

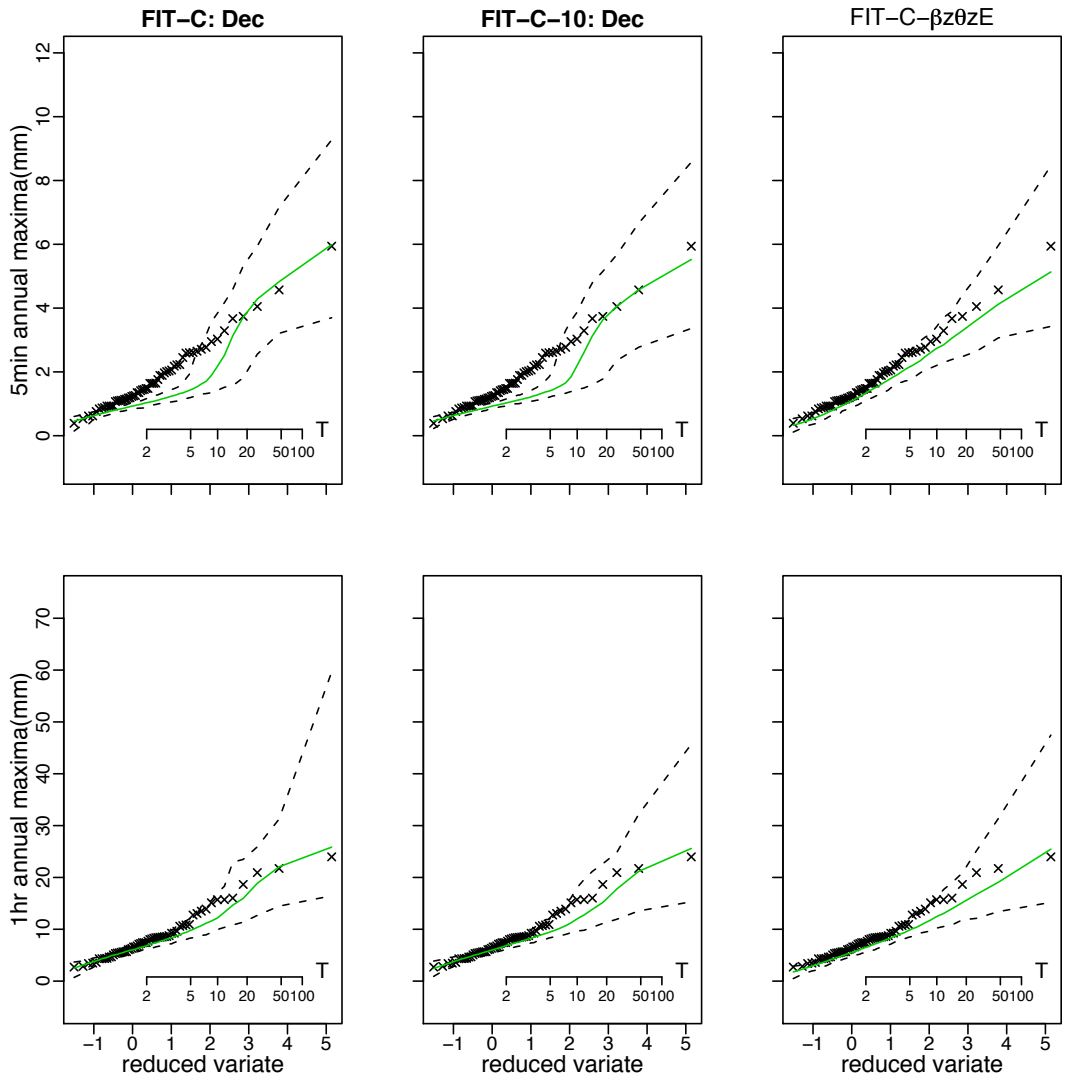

Figure 8.6: Model comparison: ordered December maxima at the 5-min and 1 hour aggregation levels.

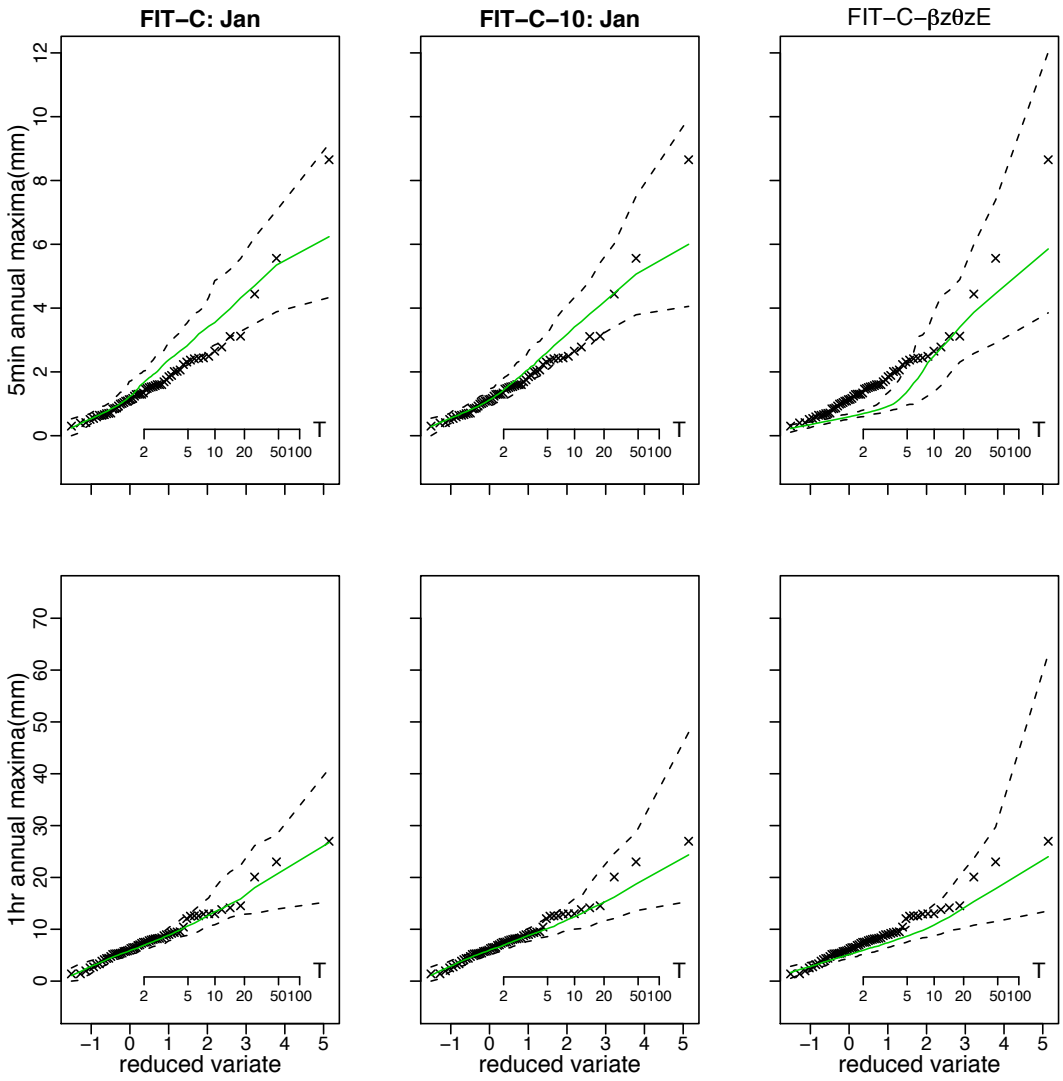

Figure 8.7: Model comparison: ordered January maxima at the 5-min and 1 hour aggregation levels.

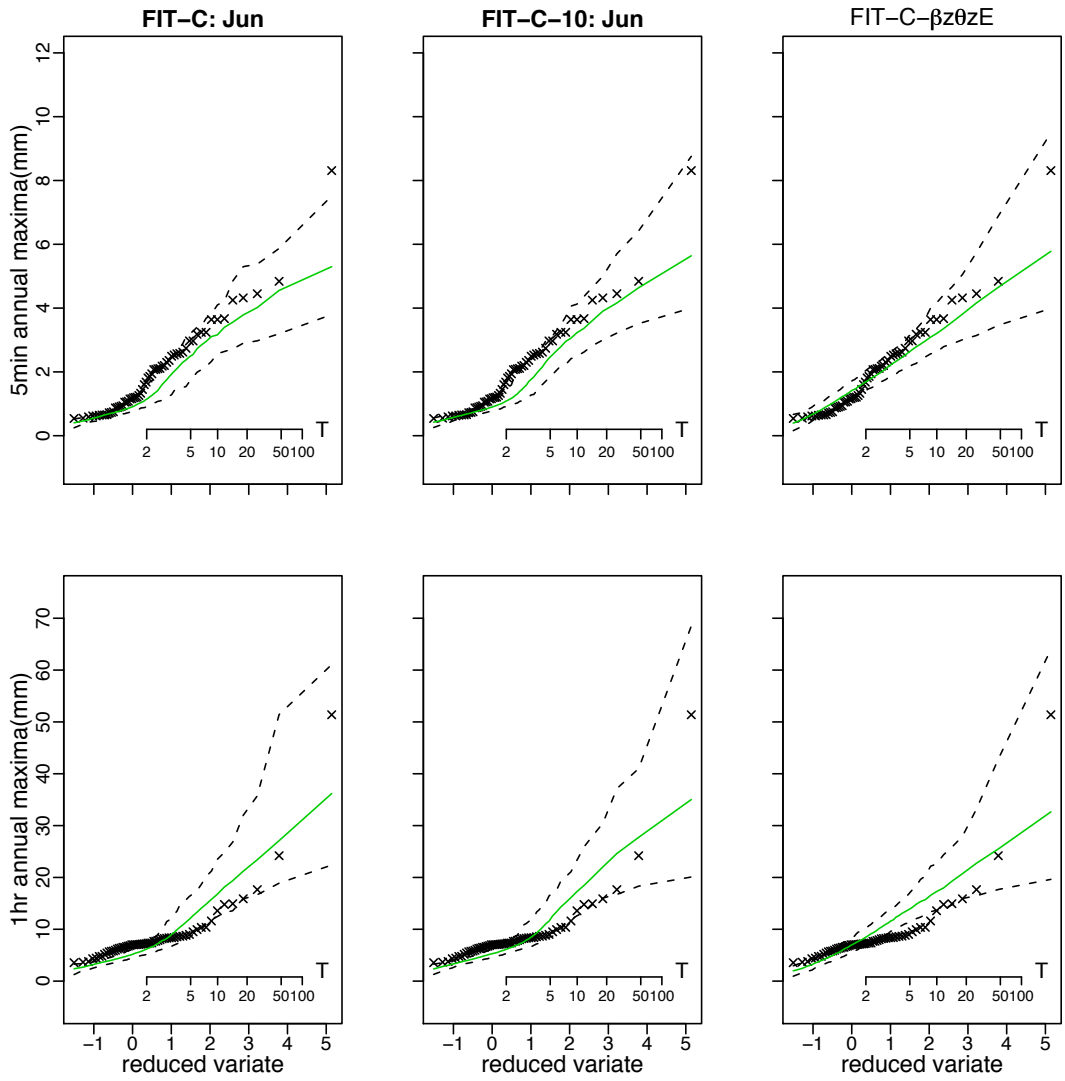

Figure 8.8: Model comparison: ordered June maxima at the 5-min and 1 hour aggregation levels.

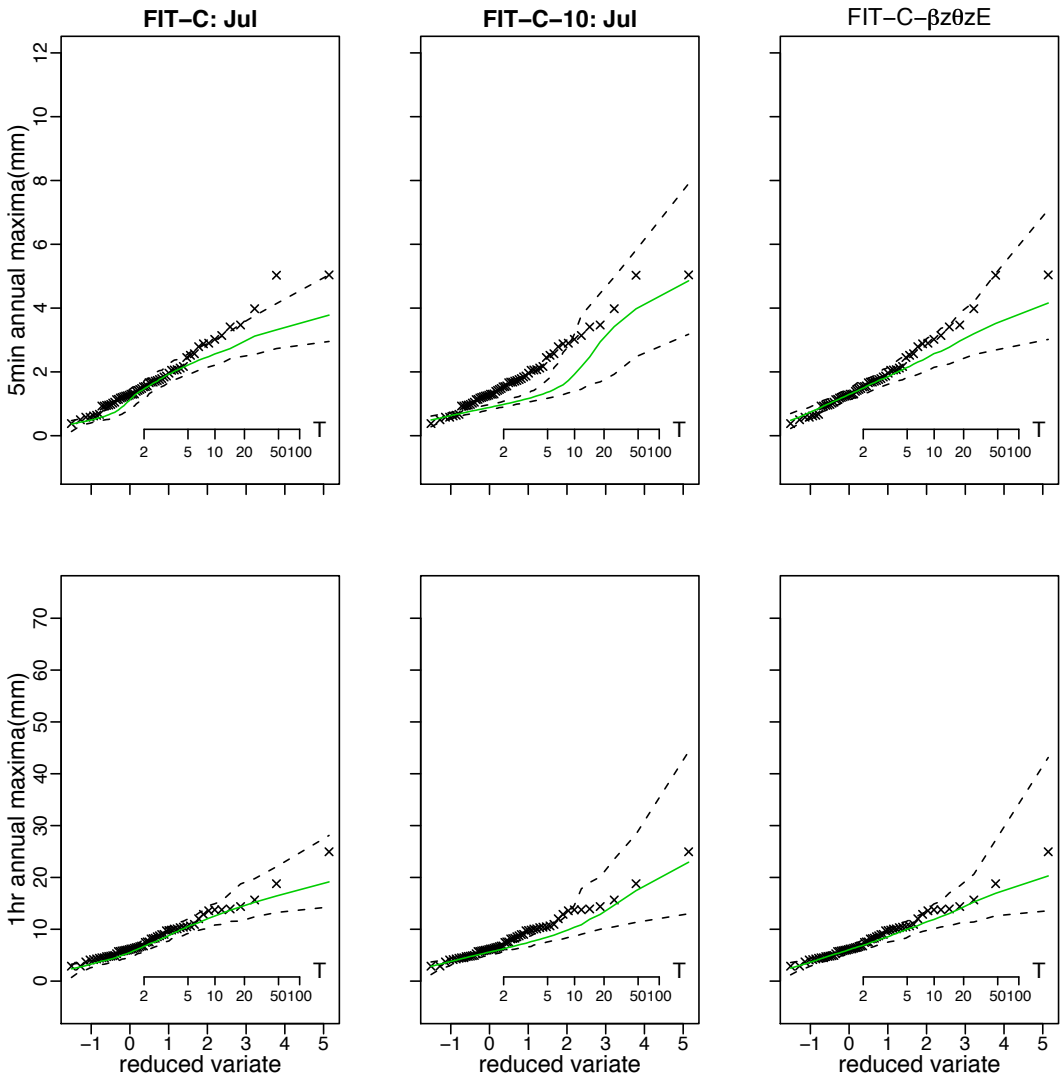

Figure 8.9: Model comparison: ordered July maxima at the 5-min and 1 hour aggregation levels.

fit. By examining Figure 8.4 to Figure 8.9, we compare FIT-C-10 and FIT-C-βzθzE with FIT-C in fitting the extremes in detail.

Figure 8.4 shows that, at the 5-min aggregation level (top row plots), all three models slightly underfit the historical values when the return period is over the range of $T > 5$ given that most points are within the confidence band and FIT-C-βzθzE has the best fit; over the range of $T \leq 5$, FIT-C has an adequate fit while both FIT-C-10 and FIT-C-βzθzE underfit slightly. All three models fit the extreme values at the 1 and 24 hours levels (middle and bottom row plots) equally well and have an adequate fit.

Figure 8.5 shows that, at the 6 hours aggregation level (top row plots), all three models fit the historical values equally well and have an adequate fit. At the 12 hours aggregation level (bottom row plots), FIT-C-βzθzE has the best and adequate fit while both FIT-C and FIT-C-10 slightly overfit the historical values over the return period range of $T > 10$.

Figure 8.6 shows that FIT-C-βzθzE has the best fit at the 5-min level although none of the three models fits the historical values adequately. At the 1 hour level, all three models perform equally well and all have a slightly underfit given that most points are within the confidence band.

Figure 8.7 shows that FIT-C has the best at the 5-min level while all three models deviate substantially from the historical values. At the 1 hour level, FIT-C has the best and adequate fit while both FIT-C-10 and FIT-C-βzθzE slightly underfit the historical values over the return period range of $T > 20$ given that most points are within the confidence band.

Figure 8.8 shows that FIT-C-βzθzE has the best and adequate fit at the 5-min level while both FIT-C and FIT-C-10 underfit the historical values. All three models deviate from the historical values substantially to the same degree at the 1 hour aggregation level.

Figure 8.9 shows that, at the 5-min aggregation level, FIT-C-βzθzE has the best fit and have an adequate fit when $T < 5$; FIT-C-10 has the worst lack-of-fit while both FIT-C and FIT-C-10 underfit the historical values. All three models fit the historical values equally well at the 1 hour aggregation level and the fits are considered as adequate.

Overall, we conclude that FIT-C-10 has the poorest performance and FIT-C-βzθzE slightly outperforms FIT-C in fitting the historical extreme rainfall values.

8.2.3 Proportion dry

Figure 8.10 and Figure 8.11 present the simulation results of proportion dry fitting. Figure 8.10 compares the model proportion dry fitting performance based on the threshold value $\delta = 0.01\ mm$ which is the measurement limit of Kelburn data[1]. In Figure 8.11, thresholds of $\delta = 0.05\ mm$ for the 5-min intervals, $\delta = 0.1\ mm$ for the 1 hour intervals, $\delta = 0.5\ mm$ for the 6 hours intervals, $\delta = 1.0\ mm$ for the 24 hours intervals, are used to compare the three fitted models. Presumably, this is a reasonable threshold specification for many applications. For example, a 'dry' 1-hour interval was defined using $\delta = 0.1\ mm$ when converting the continuous pluviograph rainfall records into discrete time series (Sansom, 1987, 1992); in climate change analysis, a dry day threshold of 1.0 mm is typical (Burton et al., 2008).

Figure 8.10 shows that, based on threshold value $\delta = 0.01\ mm$, none of the three models has an adequate fit to the historical PDs at all aggregation levels examined. The deviations at the higher aggregations levels (i.e. 6, 24 hours) are larger than the deviations at the lower levels (i.e. 5-min and 1 hour). For the 5-min intervals (top row plots), FIT-C-βzθzE has the worst fits in June and July. Apart from that, the deviations from the historical PDs are minor for all three models. For the 1 hour intervals (second row plots), the deviations from the historical PDs are minor for all three models, and FIT-C-βzθzE has the best and an almost adequate fit. For the 6, 24 hours intervals, FIT-C-βzθzE is found to be the best fit model and FIT-C-10 has a slightly worse lack-of-fit than FIT-C.

Figure 8.11 shows that all three models fit the PDs reasonably well under the specified threshold scheme. In particular, FIT-C-βzθzE has an almost adequate fit at the 5-min and 1 hour levels and outperform FIT-C and FIT-C-10. Although, at the 6 and 24 hours aggregation levels, FIT-C-βzθzE does not fit the historical PDs as good as at the lower aggregation levels, it is still identified as the best-fit model. The model performances of FIT-C and FIT-C-10 in fitting PDs are equivalent. In summary, we conclude that, though it is small, the improvement in fitting historical proportion dry values achieved by FIT-C-βzθzE upon FIT-C is significant and reducing one parameter from FIT-C by assuming a common mean storm lifetime has no significant impact.

[1]Recall that the definition and discussion on proportion dry based on threshold values are given in Section 6.4.4

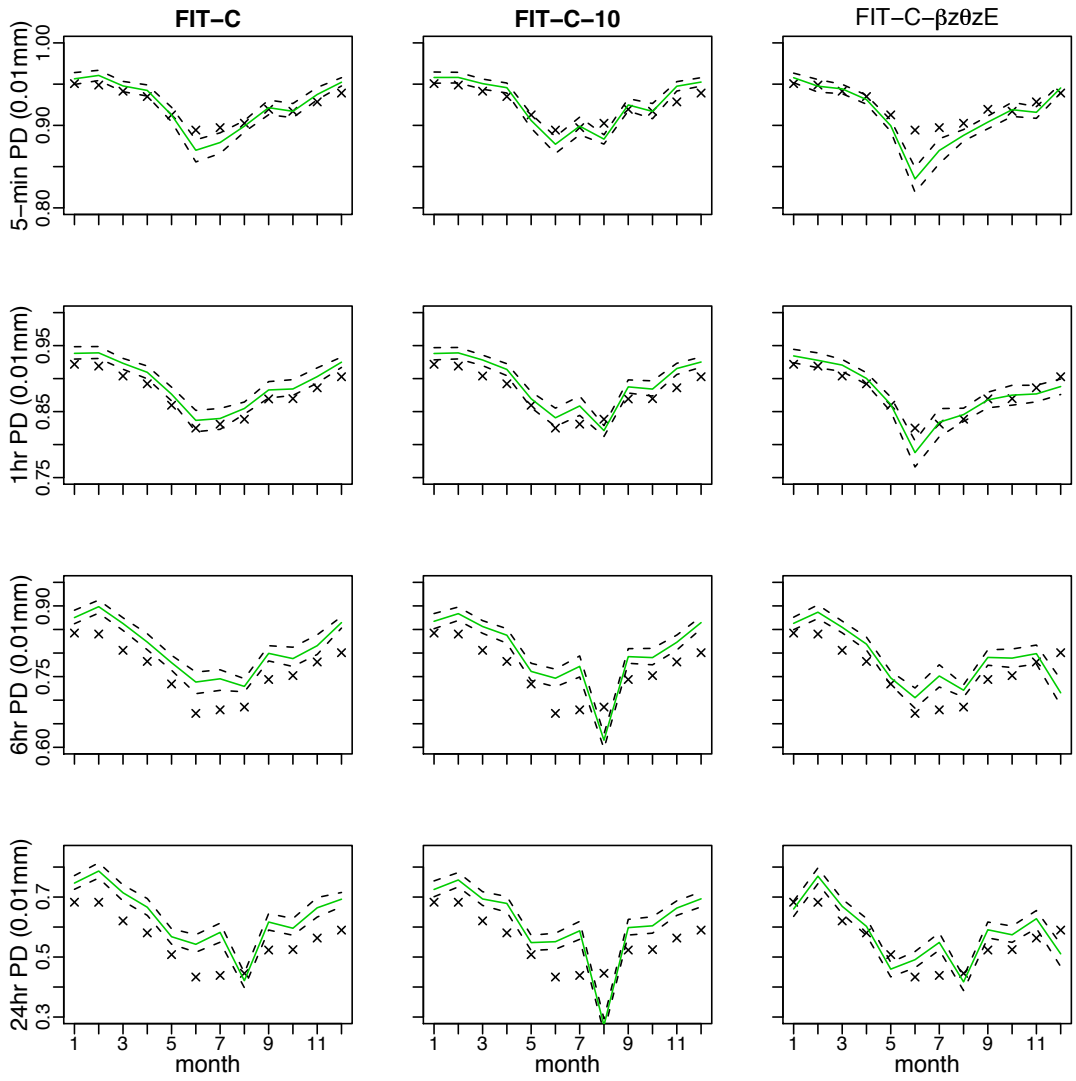

Figure 8.10: Model comparison: proportion dry patterns at the 5-min, 1, 6, and 24 hours aggregation levels.

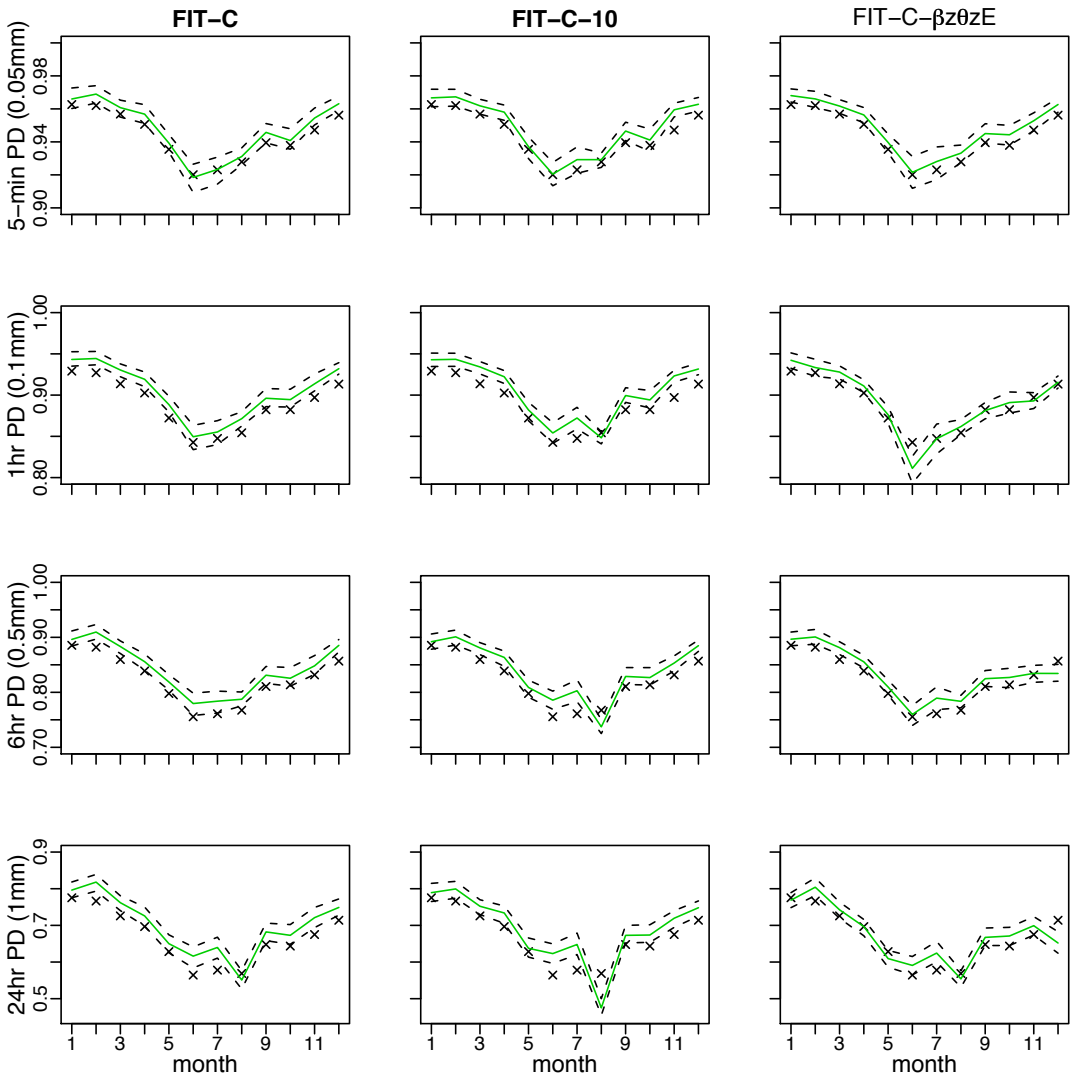

Figure 8.11: Model comparison: proportion dry by threshold values at different aggregation time scale levels.

8.3 Conclusions

The model comparison results presented in this chapter show that, given the same number of model parameters, a (single process) continuous-storm-types BLP model can improve upon FIT-C1 substantially. The better model performance of FIT-C1-βzθzE indicates that the rainfall process observed at Kelburn is better represented by a non-linear negative association between the cell origins generation rates (β) and the mean pulse depths (θ) through a unit mean exponential random variable. This implies that the assumption of equal occurrence probability for all types of storms is not realistic.

The fact that a substantial difference in single process models' performances only result in some small differences when two single processes are superposed accordingly, continuous-storm-types BLP model or not, suggests that superposition of distinct processes is the major factor in an adequate representation of the observed Kelburn rainfall series. However, superposition of distinct processes can easily lead to model over-parametrisation. The superposed continuous-storm-types BLP model proposed in this study, FIT-C-βzθzE, has achieved some small but real improvements in fitting the extreme rainfall values and the proportion of dry periods upon FIT-C. On the other hand, the overparametrisation problem remains with FIT-C-βzθzE as with FIT-C.

Reducing one parameter from FIT-C by assuming a common mean storm lifetime as specified in FIT-C-10, the negative impact on the model performances is real but slight. On the other hand, in FIT-C-10, no significant improvement can be identified to the problems caused by overparametrisation, e.g. parameter estimation results are sensitive to the initial values and the parameter space bounds, and the optimal parameter estimates set is not unique. The difficulty is that, model fits are poor when model parameters are less than nine but the parameter estimation is stable; model fits improve dramatically when the model parameters are increased to nine in a sensible way and get even better if they continue to increase, but the unstable parameter estimation process problems occur.

Finally, we give an explanation to why the superposed continuous-storm-types model FIT-C-βzθzE improves upon FIT-C so little whilst FIT-C1-βzθzE shows a big improvement upon FIT-C1, and why none of the single process BLP models (continuous-storm-types or not) fitted in this study are able to describe the Kelburn data adequately. The observed rainfall series is the results of a complex and delicately balanced

process in a random spatial-temporal field (Shaw, 1988; Waymire and Gupta, 1981a). For example, observed rainfall can come from rain drops formed at different heights in the sky above the rain gauge spot. Therefore, it is crucial that a fitted model is able to represent the overlaps of different storms. By examining the the estimated storm origins generation rates ($\hat{\lambda}$) and the estimated mean storm lifetimes ($\hat{\gamma}^{-1}$) of those fitted single process BLP models (Tables 7.1, 7.2, 7.3, 7.4, 8.1) we found that $\hat{\lambda} \ll \hat{\gamma}$, which implies that most storms end before the arrival of the next storm origins. It is this inability to represent the overlaps of different storms that causes the single process BLP models lack-of-fit to the Kelburn data. In contrast, a superposed BLP model can represent the overlaps of different storms well, although at the cost of using many more parameters. In our fitted continuous-storm-types BLP models, we have concentrated on expressing parameters β, ξ, and θ as functions of z and purposely avoided expressing parameters γ and η as functions of z due to the mathematical tractability concern. This limitation in the continuous-storm-types approach probably explains why even the best single process continuous-storm-types BLP model FIT-C1-βzθzE is not a contender to the superposed BLP models. FIT-C1-βzθzE improves upon FIT-C1 because of a better representation of the variation of different types of storms in terms of the rain cell origins generation rate (β) and the mean pulse depths (θ). Although the continuous-storm-types specification of FIT-C1-βzθzE does contribute to the model performance improvement, the impact is small in comparison with the impact due to the superposition of two independent BLP processes. That is why FIT-C-βzθzE (superposition of two FIT-C1-βzθzE processes) does improve upon FIT-C (superposition of two FIT-C1 processes) but the improvement is small. Note that when the parameter estimation minimisation objective function values are close to zero, a proper way to compare competitive models is by a simulation study.

The above conclusions point to the direction of future research. If we can express all six parameters (λ, β, ξ, γ, η, θ) as functions of a continuous random variable z in a sensible way, it is possible to specify a single process continuous-storm-types BLP model which may perform as well as a superposed BLP model. In addition, there are many choices for the distributions of z other than the uniform distribution and exponential distribution which we have explored in this study. It is possible that some other continuous distribution may have a better representation of the association relationships among model parameters. We leave these issues as future research topics.

Chapter 9

Summary and future work

9.1 Summary

Over the last decade, different ACD models have been specified for better understanding the market micro-structure of stock trading processes. However, the model comparison issue has been largely ignored due to the limitation of the traditional hypothesis testing approach in model evaluation (Bauwens et al., 2004; Meitz and Terasvirta, 2006). For the first time in the literature, ACD models have been evaluated and ranked by following the information-theoretic approach so that model comparison becomes straightforward.

The original BLP model, proposed by Cowpertwait et al. (2007) (referred to as 'CIO2007'), can be used to produce realistic simulation samples of fine-scale rainfall series which are required by many applications. In this thesis, a modified BLP model is proposed, in which a conditional mean exponential distribution is used to capture the within cell pulse depths dependence structure and a continuum of storm types within a process is assumed. Given the same number of model parameters, this modified model has significantly improved the performances of the original BLP model.

These are the two highlights of the thesis. We now summarize our main research results in the following two subsections.

9.1.1 Information-theoretic criteria and model evaluation for ACD models

Kullback-Leibler information measures the expected information loss when a statistical model $f(x|\underline{\theta})$ is used to approximate the full reality or truth $g(x)$ in terms of the difference of the expected log-likelihoods (Equation (2.10)). Akaike Information Criterion (AIC), Takeuchi Information Criterion (TIC) and Generalized Information Criterion

(GIC) are the Kullback-Leibler information based criteria for model evaluation, hence, are referred to as information-theoretic criteria. In addition to the traditional hypothesis testing approach, the information-theoretic approach provides a valid and more effective alternative for model evaluation.

Conceptually, a general information criterion (IC) can be defined as

$$\text{IC} = -(\text{log-likelihood}_{max}) + \text{EMD},$$

where $\text{log-likelihood}_{max}$ is the sample log-likelihood evaluated at some optimal estimate of $\underline{\theta}$ and EMD stands for 'Effective Model Dimension'. The sample log-likelihood as an estimator of the expected log-likelihood has a bias of order n^{-1}. Theoretically, EMD is the bias correction term which is essential for constructing an information criterion (Konishi and Kitagawa, 2008). Intuitively, EMD term may also be interpreted as a measure of the complexity of a fitted model. The EMD term has an impact to penalize the more complex model so that the principle of parsimony will be ensured.

In particular, AIC, TIC, and GIC are defined as

$$\text{AIC}(\mathbf{k}) \;=\; -\mathbf{2}\log \text{L}(\hat{\underline{\theta}}) + \mathbf{2}\,\mathbf{k},$$

$$\text{TIC}(\mathbf{k}) \;=\; -\mathbf{2}\log \text{L}(\hat{\underline{\theta}}) + \mathbf{2}\,\text{tr}\big(\mathbf{K}(\hat{\underline{\theta}})\mathbf{J}(\hat{\underline{\theta}})^{-\mathbf{1}}\big),$$

and

$$\text{GIC}(\mathbf{k}) \;=\; -\mathbf{2}\log \text{L}(\hat{\underline{\theta}}_{\mathbf{M}}) + \mathbf{2}\,\text{tr}\Big\{\mathbf{R}(\psi,\hat{\mathbf{G}})^{-\mathbf{1}}\mathbf{Q}(\psi,\hat{\mathbf{G}})\Big\}.$$

We may consider k, $\text{tr}\big(K(\hat{\underline{\theta}})J(\hat{\underline{\theta}})^{-1}\big)$, and $\text{tr}\big\{R(\psi,\hat{G})^{-1}Q(\psi,\hat{G})\big\}$ to be the EMD for AIC, TIC, and GIC, respectively. For a valid application, all three criteria require the large sample assumption. As a rule of thumb, a sample may be considered as 'large' if the number of observations per estimated parameter (i.e. the ratio n/k) is greater than or equal to 40 (Burnham and Anderson, 2002). AIC and TIC only works under an MLE framework but TIC extends AIC's applications to allow the misspecified candidate models. TIC can be derived as a special case of GIC and the application domain is further enlarged by including the robust estimation framework with GIC. No distributional assumption has been made in the derivation of AIC, TIC, or GIC.

Given assumption conditions specified in Section 2.3.1, AIC, TIC, and GIC are asymptotically unbiased estimators of the relative expected K-L information. The best model selected by the information-theoretic criteria is a 'good' model in the sense that:

(1) it is a parsimonious model which has the minimum predictive error among all candidate models so that it has an optimal balance between the model accuracy (bias) and the model complexity (variance);

(2) it is sample size dependent so that it allows a more complex model to be selected as more information is available when the sample size is getting larger;

(3) under a finite dimensional true model setting, a selected 'good' model may not necessarily be the true model due to the sampling variation.

Theoretically, if we would like to assume that the true model is contained in the specified models, AIC is valid to use for model evaluation even under a robust estimation framework. That the true model needs to be contained in the specified models is a very strong assumption which limits AIC applications considerably. However, in practice, AIC has been considered as a good proxy for TIC and GIC in general for the following reasons. If the specified models are close to the true distribution, $k \approx$ EMD, then, AIC $\approx$ TIC, or AIC $\approx$ GIC. In the case of a model being badly misspecified, the sample maximum likelihood term is in absolute domination over the bias correction (or the trace) term so that a bad model would not stand a chance to be selected. Therefore, we conclude that it is valid to use AIC for model evaluation in general because AIC is a practical and parsimonious implementation of TIC and GIC. If we have a strong reason to doubt a model evaluation result made by AIC, we need to calculate TIC or GIC for a cross-check. However, the calculation of the trace terms with TIC and GIC could be non-trivial. More research is needed to study the utilities of TIC and GIC in real data analysis.

In this thesis, four basic Autoregressive Conditional Duration (ACD) models and two mixed distribution ACD models are specified to fit two real inter-transaction durations data sets. It is the first time in the literature that ACD models have been evaluated using the information-theoretic criteria. Among the six evaluated candidate models, a newly proposed ACD(1,1) model with a mixed lognormal-gamma distribution error term is identified as the best AIC selected model. By following the information-theoretic approach, the comparison of ACD models with different autoregressive structures and error term structures is straightforward and the results can be interpreted.

9.1.2 Further developments of the original BLP model

There are different types of rainfall models in the literature. The empirical statistical models and the point process models are the two main groups of stochastic rainfall models which we have reviewed in detail. Our focus has been on the evolution of the Poisson-based point process models: from Poisson white noise model to the recent developments of temporal Poisson cluster process models, following the line of work developed in Rodriguez-Iturbe et al. (1984, 1987a,b).

The Neyman-Scott rectangular pulse (NSRP) model and the Bartlett-Lewis rectangular pulse (BLRP) model were first developed and studied by Rodriguez-Iturbe et al. (1987a,b). The NSRP and BLRP models (and their variants) have been widely studied over the last two decades, and many successful applications in modelling the rainfall series that are aggregated over time intervals of 1 hour or more were reported. The goal of these rainfall models is to produce the realistic simulation of continuous rainfall series which is required by many applications including: urban drainage system design, civil defence projects, ecosystem studies, and radio transmission study. Therefore, the model performance is assessed by how well the observed sample statistics can be retained by the simulation samples generated from a fitted model.

Rainfall data are typically available in the form of a time series sampled over fixed time intervals, e.g. daily or hourly observations. The observed rainfall series typically have clustering patterns, i.e. the index of dispersion is much greater than one.

The empirical statistical models follow a discrete time approach to model the rainfall process and have been used to fit daily and hourly rainfall series for many decades. While there were many reported successful cases of good fit to both moment properties and extreme values, an empirical statistical model was blamed for the lack of physical motivation in the model formulation as its main disadvantage. In contrast, the point process models, at least in part, represent the physical process of precipitation.

A point process model employs marked continuous-time point processes to fit the observed discrete rainfall time series. In a point process model, the dry periods between rainfall events (wet periods) are not modelled explicitly, but as a consequent result in making up the spaces between wet periods.

The research so far has not been able to justify any real difference in the empirical model performance between two equivalent NSRP and BLRP models. Bartlett-Lewis Pulse (BLP) model developed by CIO2007, is an extension of BLRP model for mod-

elling fine-scale rainfall series. In order to have a more realistic characterization of the rain cell profiles, BLP models replace the rectangular pulse profiles by a Poisson process of instantaneous pulse depths. A basic version of BLP model needs six parameters for model specification:

λ – rate of storm origins; γ^{-1} – mean storm lifetime; β – rate of cell origins;
ξ – pulse arrival rate; η^{-1} – mean cell duration; θ – mean pulse depth.

The combined effect of three Poisson processes should account for fluctuations over a wide range of scales and be sufficient for most applications (Koutsoyiannis, 2002).

We denote the original BLP model that represents a single process of storms by FIT-O1 and the single process BLP model with conditional mean depth distribution by FIT-C1. Both FIT-O1 and FIT-C1 have six parameters in total $\{\lambda, \beta, \xi, \gamma, \eta, \theta\}$. The original BLP model used to fit the Kelburn data in CIO2007, FIT-O, is formulated by superposing two independent FIT-O1 processes but assuming a common mean pulse depth θ. Therefore, FIT-O has 11 parameters in total, $\{\lambda_i, \beta_i, \xi_i, \gamma_i, \eta_i, \theta : i = 1, 2\}$. Likewise, model specification FIT-C is formulated by superposing two independent FIT-C1 processes.

CIO2007 reported an excellent fitting results in terms of moment properties and extreme values (annual maxima at 1, 6, and 24 hours aggregation levels). On the other hand, a tendency for overestimation of the historical proportion of dry periods (proportion dry) values and lack-of-fit of extreme values for individual months were identified for improvement. The model specification FIT-C is proposed in this thesis to implement the within cell pulse depths dependence structure as formulated in the original BLP model characterization framework (but not implemented in FIT-O). A comprehensive model performance comparison is made between FIT-O and FIT-C through simulation study. The simulation results show that FIT-C improves the fit to the observed proportion of dry periods significantly, whilst retaining a good fit to moment properties. FIT-C also achieves a better fit to annual extreme rainfall values at the 5-min and 12 hours aggregation levels whilst retaining a good fit at other aggregation levels, and significant improvement in the goodness-of-fit to extremes for the individual months. With a proportion dry constraint at 24 hours level also included, further improvements are achieved in fitting to the historical proportion dry values at higher time scale levels, e.g. 12 and 24 hours levels, upon FIT-C. However, this gain is at a cost of slightly underfitting 5-min extreme values.

Finally, an alternative approach for characterization of rainfall processes within BLP model framework is investigated in this thesis. A BLP model with continuous distributions of storm types is formulated to represent a single process of rainfalls which is characterized as consisting of a continuum of types of storms. Following the continuous-storm-types approach, model parameters are expressed as functions of a continuous distribution random variable z with probability density function $f(z)$. Hence, governed by $f(z)$, each storm is characterized by different model parameters and occurs with probability $f(z)\mathrm{d}z$, or at the rate of $\lambda f(z)\mathrm{d}z$. FIT-O1 and FIT-C1 are recovered by setting the model parameters to be constants: $\beta(z) = \beta$, $\xi(z) = \xi$, $\gamma(z) = \gamma$, $\eta(z) = \eta$, and $\theta(z) = \theta$.

For a continuous-storm-types BLP model, the statistical properties up to third order are derived and several models are specified with different parameter functions. By expressing the generation rates of cell origins (β) negatively associated with the mean pulse depths (θ) through a unit mean exponential random variable z, a continuous-storm-types BLP model FIT-C1-βzθzE is specified and has been identified as the best-fit model among single process BLP models assessed. In particular, FIT-C1-βzθzE substantially improves upon FIT-C1. Therefore, a superposed continuous-storm-types BLP model FIT-C-βzθzE is formulated (by superposing two independent FIT-C1-βzθzE processes) to compare with FIT-C in model performances through simulation. The model comparison results show that FIT-C-βzθzE have achieved some small but real improvements on FIT-C in fitting the extreme values and the proportion of dry. We conclude that the continuous-storm-types approach have a better representation of the rainfall processes than that of the original BLP model approach. However, because we did not express parameters γ^{-1} (mean storm lifetime) and η^{-1} (mean cell duration) as functions of z due to mathematical tractability concerns (they appear in the denominator part in the moment properties formulas and the integration is hard, if not impossible, to evaluate), this limits the ability of the fitted continuous-storm-types BLP models to represent the overlaps of different storms. Hence, while the improvement due to the continuous-storm-types characterization does exist, the contribution to the final results is small.

9.2 Future research topics

From this thesis, six promising topics of interest for future research are identified as listed below.

Topic 1 More empirical studies are needed to investigate utilities of TIC and GIC in real data analysis. These include: how to calculate TIC and GIC in a more efficient and reliable way using a popular statistical package, e.g. R; how model evaluation results could differ using AIC instead of TIC or GIC; what other statistical inferences an information-theoretic criterion can make apart from ranking models; robust estimation in fine-scale time series analysis. We should promote the use of the information-theoretic approach for model evaluation in statistical practice.

Topic 2 The stock inter-transaction duration series show strong clustering patterns just as the fine-scale rainfall data does. According to the market micro-structure theory, it is reasonable to classify all traders into two groups: the informed traders group and the non-informed traders group. If we consider transaction monetary volume (i.e. price times volume) as the mark variable associated with transaction times, The BLP models studied in this thesis may be appropriate for analysis of such financial marked point process data. By resorting to the autoregressive process to capture the clustering patterns, the ACD model has been criticized as it has no firm basis in economic theory. If a BLP type of point process model can be formulated to better model the stock transaction data, it is likely, at least in part, that the model parameters can be interpreted based on some economic theory.

Topic 3 One of the weak areas identified with the point process rainfall models is the model estimation method. The method of moments procedure involves subjective choices. One possible alternative is to use a Bayesian approach for model estimation, e.g. under a hierarchical model framework. Once the Bayesian model estimation can be implemented, the information-theoretic approach can be used for Bayesian model selection.

Topic 4 Although our current research has been confined to single site rainfall models, a good single-site temporal rainfall model obviously forms the foundation for the construction of a good spatial-temporal stochastic rainfall model. Therefore, BLP models developed in this thesis may be adapted into various Bartlett-Lewis

process type or Neyman-Scott process type spatial-temporal models for further development.

Topic 5 An empirical statistical model for fine-scale rainfall series has not been found in all stochastic rainfall model literature we have reviewed. In spite of the lack of physical motivation, empirical rainfall models have intuitive appeal in terms of model specification and parameter estimation. If some sort of systematic and more objective model specification procedure can be successfully developed, an optimization of parametrisation of an empirical rainfall model may lead to a parsimonious model specification as economic as a corresponding BLP model. In such a case, the empirical rainfall model may be considered as a valid alternative to Poisson cluster models for modelling the fine-scale rainfall data.

Topic 6 To investigate the possibility of expressing all six parameters (λ, β, ξ, γ, η, θ) as functions of a continuous random variable z in a sensible way, so that a single process continuous-storm-types BLP model, which involves less model parameters, can be a contender of a superposed BLP model.

Appendix A

Background theory and some related topics for information-theoretic criteria

A.1 Some standard results of maximum likelihood estimator

Let $\hat{\underline{\theta}}_n$ be the MLE based on n observations, i.e. $\hat{\underline{\theta}}_n$ is a solution of the likelihood equation

$$\frac{\partial \log L(\underline{\theta})}{\partial \underline{\theta}} = \sum_{r=1}^{n} \frac{\partial \log\, f(x_r|\underline{\theta})}{\partial \underline{\theta}} = \underline{0}, \tag{A.1}$$

where $\partial \log L(\underline{\theta})/\partial \underline{\theta}$ is a $k\times 1$ vector, the ith component of which is given by $\partial \log L(\underline{\theta})/\partial \theta_i$, and $\underline{0}$ is the $k \times 1$ zero vector, i.e. all the components of which are 0.

Conventionally, the $k \times k$ matrix formed by taking the second derivative of the log-likelihood function is called the *hessian* matrix and is given by

$$H(\underline{\theta}) = \frac{\partial^2 \log L(\underline{\theta})}{\partial \underline{\theta}\, \partial \underline{\theta}^T} = \sum_{r=1}^{n} \frac{\partial^2 \log f(x_r|\underline{\theta})}{\partial \underline{\theta}\, \partial \underline{\theta}^T}, \tag{A.2}$$

whose (i, j) element is given by

$$\sum_{r=1}^{n} \frac{\partial^2 \log f(x_r|\underline{\theta})}{\partial \theta_i \partial \theta_j} = \sum_{r=1}^{n} \frac{\partial}{\partial \theta_i} \Big\{ \frac{\partial \log f(x_r|\underline{\theta})}{\partial \theta_j} \Big\}, \qquad i, j = 1, 2, ..., k.$$

Some important MLE's properties can be derived from the maximum likelihood estimation and large sample theory (see, for example, DeGroot, 1986, Chapter 7; Davison, 2003, Chapter 4; or Konishi and Kitagawa, 2008, Chapter 3). These properties are summarized in (A) and (B) below which have been mainly adapted from Konishi and Kitagawa (2008).

(A) Assume that a random sample $\underline{x} = \{x_1, x_2, ..., x_n\}$ is generated from the density function $f(\underline{x}|\underline{\theta}_o)$ where $\underline{\theta}_o$ is a solution of

$$\int f(x|\underline{\theta}) \frac{\partial \log f(x|\underline{\theta})}{\partial \underline{\theta}} \mathrm{d}x = \underline{0}.$$

Then the following statements hold:

(i) The solution of Equation (A.1) $\hat{\underline{\theta}}_n$ converges in probability to $\underline{\theta}_o$ when $n \to +\infty$.

(ii) The MLE $\hat{\underline{\theta}}_n$ has asymptotic normality such that the distribution of $\sqrt{n}(\hat{\underline{\theta}}_n - \underline{\theta}_o)$ converges in distribution to the k-dimensional normal distribution $N_k(\underline{0}, \Sigma)$ with mean vector $\underline{0}$ and *variance-covariance matrix* $\Sigma = I(\underline{\theta}_o)^{-1}$, where the matrix $I(\underline{\theta}_o)$ is the *Fisher information matrix*, $I(\underline{\theta})$, evaluated at $\underline{\theta} = \underline{\theta}_o$.

Fisher information matrix, also called *expected information matrix*, is the mean information that the data will contain when collected. It is given by

$$\begin{aligned} I(\underline{\theta}) &= \int f(x|\underline{\theta}) \frac{\partial \log f(x|\underline{\theta})}{\partial \underline{\theta}} \frac{\partial \log f(x|\underline{\theta})}{\partial \underline{\theta}^T} \mathrm{d}x \\ &= -\int f(x|\underline{\theta}) \frac{\partial^2 \log f(x|\underline{\theta})}{\partial \underline{\theta}\, \partial \underline{\theta}^T} \mathrm{d}x. \end{aligned} \tag{A.3}$$

In order to derive the asymptotic normality in (A), the existence of $\underline{\theta}_o \in \Theta$ such that $g(x) = f(x|\underline{\theta}_o)$ is assumed. This is the typical assumption setting with the standard theory of maximum likelihood estimation. If we allow that data $\underline{x} = \{x_1, x_2, ..., x_n\}$ are generated from some unknown distribution $g(x)$ which may or may not be contained in the fitted models, modified asymptotic normality results can be obtained as in (B) below.

(B) Assume that the data are generated from some unknown distribution $g(x)$ and $\underline{\theta}_o$ is a solution of

$$\int g(x) \frac{\partial \log f(x|\underline{\theta})}{\partial \underline{\theta}} \mathrm{d}x = \underline{0}.$$

We need to define two more $k \times k$ matrices that correspond to the outer product form (the K matrix) and the second derivative form (the J matrix) of the Fisher information matrix $I(\underline{\theta})$ in Equation (A.3).

$$\begin{aligned} K(\underline{\theta}) &= \int g(x) \frac{\partial \log f(x|\underline{\theta})}{\partial \underline{\theta}} \frac{\partial \log f(x|\underline{\theta})}{\partial \underline{\theta}^T} \mathrm{d}x \\ &= \Big(\int g(x) \frac{\partial \log f(x|\underline{\theta})}{\partial \theta_i} \frac{\partial \log f(x|\underline{\theta})}{\partial \theta_j} \mathrm{d}x \Big), \end{aligned} \tag{A.4}$$

$$J(\underline{\theta}) = -\int g(x)\frac{\partial^2 \log f(x|\underline{\theta})}{\partial\underline{\theta}\,\partial\underline{\theta}^T}\,\mathrm{d}x.$$
$$= -\Big(\int g(x)\frac{\partial^2 \log f(x|\underline{\theta})}{\partial\theta_i\,\partial\theta_j}\,\mathrm{d}x\Big), \quad i,j = 1,2,...,k. \tag{A.5}$$

Now the following statements can be made with respect to the maximum likelihood estimator $\hat{\underline{\theta}}_n$:

(i) The MLE $\hat{\underline{\theta}}_n$ converges in probability to $\underline{\theta}_o$ when $n \to +\infty$.

(ii) The MLE $\hat{\underline{\theta}}_n$ has asymptotic normality such that the distribution of $\sqrt{n}(\hat{\underline{\theta}}_n - \underline{\theta}_o)$ converges in distribution to the k-dimensional normal distribution $N_k(\underline{0}, \Sigma)$ with mean vector $\underline{0}$ and variance-covariance matrix $\Sigma = J(\underline{\theta}_o)^{-1}K(\underline{\theta}_o)J(\underline{\theta}_o)^{-1}$, where the matrices $K(\underline{\theta}_o)$ and $J(\underline{\theta}_o)$ are two $k \times k$ matrices, defined by Equations (A.4) and (A.5), evaluated at $\underline{\theta} = \underline{\theta}_o$.

The matrix $\Sigma = J(\underline{\theta}_o)^{-1}K(\underline{\theta}_o)J(\underline{\theta}_o)^{-1}$ is the so-called *information sandwich variance matrix*. For more details about this sandwich variance matrix, see Davison (2003, p147), or White (1982).

The results in (B) are a generalization of the results in (A). In general, $K(\underline{\theta}) \neq J(\underline{\theta})$. If the distribution $g(x)$ that generated the data is included in the class of parametric models $\{f(x|\underline{\theta}); \underline{\theta} \in \Theta \subset R^k\}$, Equations (A.4) and (A.5) reduce to Equation (A.3). The asymptotic variance-covariance matrix for $\sqrt{n}(\hat{\underline{\theta}}_n - \underline{\theta}_o)$ becomes

$$\Sigma = J(\underline{\theta}_o)^{-1}K(\underline{\theta}_o)J(\underline{\theta}_o)^{-1}$$
$$= I(\underline{\theta}_o)^{-1}I(\underline{\theta}_o)I(\underline{\theta}_o)^{-1}$$
$$= I(\underline{\theta}_o)^{-1}. \tag{A.6}$$

A direct result from Equation (A.6) is that

$$\mathrm{tr}(K(\underline{\theta}_o)J(\underline{\theta}_o)^{-1}) = \mathrm{tr}(I(\underline{\theta}_o)I(\underline{\theta}_o)^{-1}) = \mathrm{tr}(\mathrm{I}_k) = k, \tag{A.7}$$

where tr denotes the matrix trace function and I_k is a $k \times k$ identity matrix. Since $\underline{\theta}$ is a k-dimensional parameter vector, k is the number of parameters estimated in a fitted model.

Note that the matrices $K(\underline{\theta}_o)$ and $J(\underline{\theta}_o)$ can be estimated by replacing the unknown probability distribution $G(x)$ or $g(x)$ by an empirical distribution function $\hat{G}(x)$ or $\hat{g}(x)$ based on the observed data as follows:

$$K(\hat{\underline{\theta}}) = \frac{1}{n}\sum_{r=1}^{n} \frac{\partial \log f(x_r|\underline{\theta})}{\partial \underline{\theta}} \frac{\partial \log f(x|\underline{\theta})}{\partial \underline{\theta}^T}\bigg|_{\hat{\underline{\theta}}}, \tag{A.8}$$

$$J(\hat{\underline{\theta}}) = -\frac{1}{n}\sum_{r=1}^{n} \frac{\partial^2 \log f(x_r|\underline{\theta})}{\partial \underline{\theta}\, \partial \underline{\theta}^T}\bigg|_{\hat{\underline{\theta}}}. \tag{A.9}$$

The $(i,j)^{th}$ elements of these matrices are

$$K_{ij}(\hat{G}) = \frac{1}{n}\sum_{r=1}^{n} \frac{\partial \log f(X_r|\underline{\theta})}{\partial \theta_i} \frac{\partial \log f(X_r|\underline{\theta})}{\partial \theta_j}\bigg|_{\hat{\underline{\theta}}} \quad \text{and} \tag{A.10}$$

$$J_{ij}(\hat{G}) = -\frac{1}{n}\sum_{r=1}^{n} \frac{\partial^2 \log f(X_r|\underline{\theta})}{\partial \theta_i\, \partial \theta_j}\bigg|_{\hat{\underline{\theta}}}, \quad i,j = 1,2,...,k, \tag{A.11}$$

respectively. Obviously, here $\hat{\underline{\theta}} \equiv \hat{\underline{\theta}}_n$ from Equation (A.1).

A.2 Some standard results of robust statistics

In order to define the Generalized Information Criterion (GIC), we need to describe the concepts of M-estimator (maximum-likelihood-type estimator) and *influence function* which belong to the robust statistics framework (Huber and Ronchetti, 2009). Konishi and Kitagawa (2008, Chapter 5) have given a full treatment on M-estimator and the influence function in connection with GIC based on a statistical functional approach. Since the objective here is to calculate the GIC from sample data, to avoid unnecessary technicalities we confine the definitions of M-estimator and influence function to be under the empirical distribution $\hat{G}$.

The empirical M-estimator $\hat{\underline{\theta}}_M$ is defined as the solution of the equation

$$\sum_{r=1}^{n} \psi(X_r, \hat{\underline{\theta}}_M) = \underline{0} \tag{A.12}$$

with ψ being some function on $\Omega \times \Theta$, where Ω is the sample space and Θ is the parameter space. Equation (A.12) has a structure similar to Equation (A.1). The maximum likelihood estimator can be considered as a special case of an M-estimator, corresponding to

$$\psi(x, \underline{\theta}) = \frac{\partial}{\partial \underline{\theta}} \log f(x|\underline{\theta}). \tag{A.13}$$

Now consider a statistical model $f(x|\hat{\underline{\theta}}_M)$ estimated using an M-estimator procedure based on n observations, and define

$$R(\psi, \hat{G}) = -\frac{1}{n}\sum_{r=1}^{n} \frac{\partial \psi(x_r, \underline{\theta})}{\partial \underline{\theta}} \Big|_{\underline{\theta} = \hat{\underline{\theta}}_M}, \tag{A.14}$$

and

$$Q(\psi, \hat{G}) = \frac{1}{n}\sum_{r=1}^{n} \psi(x_r, \underline{\theta}) \frac{\partial \log f(x_r|\underline{\theta})}{\partial \underline{\theta}^T} \Big|_{\underline{\theta} = \hat{\underline{\theta}}_M}. \tag{A.15}$$

If $\underline{\theta}$ is a k-dimensional parameter vector, both $R(\psi, \hat{G})$ and $Q(\psi, \hat{G})$ are $k \times k$ matrices. There is a similar structure relationship between Equations (A.14) and (A.9), and between Equations (A.15) and (A.8).

If model parameters are estimated using MLE so that Equation (A.13) holds, Equations (A.14) and (A.15) will become (see the definitions for $J(\hat{\underline{\theta}})$ and $K(\hat{\underline{\theta}})$ in Equations A.9 and A.8)

$$R(\psi, \hat{G}) \;=\; -\frac{1}{n}\sum_{r=1}^{n} \frac{\partial^2 \log f(x_r|\underline{\theta})}{\partial \underline{\theta}\, \partial \underline{\theta}^T} \Bigg|_{\hat{\underline{\theta}}} \;=\; J(\hat{\underline{\theta}}),$$

$$Q(\psi, \hat{G}) \;=\; \frac{1}{n}\sum_{r=1}^{n} \frac{\partial \log f(x_r|\underline{\theta})}{\partial \underline{\theta}} \frac{\partial \log f(x|\underline{\theta})}{\partial \underline{\theta}^T} \Bigg|_{\hat{\underline{\theta}}} \;=\; K(\hat{\underline{\theta}}),$$

so that

$$R(\psi, \hat{G})^{-1}\, Q(\psi, \hat{G}) \;=\; J(\hat{\underline{\theta}})^{-1} K(\hat{\underline{\theta}}). \tag{A.16}$$

From Smith and Teo (1989, p294), we have

$$\mathrm{tr}\big(J(\hat{\underline{\theta}})^{-1} K(\hat{\underline{\theta}})\big) \;=\; \mathrm{tr}\big(K(\hat{\underline{\theta}}) J(\hat{\underline{\theta}})^{-1}\big). \tag{A.17}$$

The empirical *influence function* $T_M^{(1)}(x; \hat{G})$ for the M-estimator is defined by (Konishi and Kitagawa, 2008, p116)

$$T_M^{(1)}(x; \hat{G}) = R(\psi, \hat{G})^{-1} \psi(x, \hat{\underline{\theta}}_M). \tag{A.18}$$

Therefore,

$$R(\psi, \hat{G})^{-1}\, Q(\psi, \hat{G}) \;=\; \frac{1}{n}\sum_{r=1}^{n} T_M^{(1)}(x_r; \hat{G}) \frac{\partial \log f(x_r|\underline{\theta})}{\partial \underline{\theta}^T} \Big|_{\underline{\theta} = \hat{\underline{\theta}}_M}. \tag{A.19}$$

As shown in Section 2.3.1, it is the influence function $T_M^{(1)}(x; \hat{G})$ that distinguishes between GIC and TIC due to the different model estimation methods.

A.3 Definition of the Ljung-Box statistic

We denote the L-B statistic by $Q(m)$ as a test statistic for the null hypothesis $H_o : \rho_1 = \ldots = \rho_m = 0$ against the alternative hypothesis $H_a : \rho_i \neq 0$ for some $i \in 1, \ldots, m$. The L-B statistic is defined by Ljung and Box (1978) (also see Tsay, 2005, p27):

$$Q(m) = n(n+2) \sum_{l=1}^{m} \frac{\hat{\rho}_l^2}{n-l} \tag{A.20}$$

where n is the sample size and $\hat{\rho}_l$ is the lag l sample autocorrelation of a time series y_t as defined by:

$$\hat{\rho}_l = \frac{\sum_{t=l+1}^{n}(y_t - \overline{y})(y_{t-l} - \overline{y})}{\sum_{t=1}^{n}(y_t - \overline{y})^2}, \qquad 0 \leq l < n-1.$$

Under certain conditions, $Q(m)$ is asymptotically a chi-squared random variable with m degrees of freedom.

A.4 AICc, the small sample version AIC

There is no universal formula for AIC in small sample cases because the derivation of AIC depends on large sample properties. The most widely used small sample version AIC is derived as:

$$\mathbf{AIC_c(k)} = -\mathbf{2}\log \mathrm{L}(\hat{\theta}) + \mathbf{2\,k}\frac{\mathbf{n}}{\mathbf{n-k-1}} \tag{A.21}$$

by Hurvich and Tsai (1989) adapted from results in Sugiura (1978). This formula has been proved valid for Gaussian (i.e. normally distributed) linear regression and ARMA models.

Let $\{(y_r, x_r);\ r = 1, 2, ..., n\}$ be n sets of data obtained in terms of the response variable Y and $p-1$ explanatory variables $\underline{X}$. The linear regression model with a normally distributed error term can be expressed as a conditional distribution model:

$$f(y_r|\underline{x}_r, \underline{\theta}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big\{ -\frac{(y_r - \underline{x}_r^T \underline{\beta})^2}{2\sigma^2} \Big\}, \quad r = 1, 2, ..., n,$$

where the unknown parameters in the model are $\underline{\theta} = (\underline{\beta}^T, \sigma^2)^T$ and $\underline{\beta} = (\beta_0, \beta_1, ..., \beta_{p-1})^T$.

The AIC for a linear regression model model is given by

$$\begin{aligned} \mathrm{AIC}(k) &= -2\ \log \mathrm{L}(\underline{\hat{\theta}}) + 2k \\ &= n\log(2\pi) + n\log(\hat{\sigma}^2) + n + 2(p+1), \end{aligned} \tag{A.22}$$

where $\hat{\sigma}^2$ is the MLE of the variance of the normal distribution of the error term and $k = p + 1$ is the number of estimated parameters.

A familiar form of AICc formula which is used more often in statistical text books is derived as

$$\begin{aligned} \text{AICc}(k) &= -2 \log \text{L}(\underline{\hat{\theta}}) + 2(p+1)\frac{n}{n-p-2} \\ &= n\log(2\pi) + n\log(\hat{\sigma}^2) + n + 2(p+1)\frac{n}{n-p-2}. \end{aligned} \tag{A.23}$$

If we compare Equation (A.21) with Equation (A.23) (both for defining AICc), one may realize that it can cause confusion without truly understanding what the k and p stand for.

Now we use an example to show when we should use AICc rather than AIC. In Crawley (2007), p353, a simplest regression model is used to fit a very small data set ($n = 9$) to show how AIC is calculated. The data set is $\underline{y} = \{12, 10, 8, 11, 6, 7, 2, 3, 3\}$. A global mean regression model in R is fitted to the data, i.e. $y_r = \beta_0 + \epsilon_r, \quad r = 1, 2, ..., 9$. It is easy to get $\hat{\sigma}^2 = 12.10$ (2dp) from the R output. Since we have $p = 1$ and $n = 9$, so AIC is calculated as

$$\text{AIC} = 9\log 2\pi + 9\log(12.10) + 9 + 2(1+1) = 51.98,$$

based on formula (A.22). However, $n = 9$ is a typical small sample case. We should use AICc rather than AIC for model evaluation. Based on formula (A.23)

$$\text{AICc} = 9\log 2\pi + 9\log(12.10) + 9 + 2(1+1)\frac{9}{9-1-2} = 53.98$$

which is substantially different from AIC.

Here arises a critical question: how large is a *large sample* so that AIC is appropriate to use? The answer lies with not only the absolute size of n, more importantly the ratio n/k i.e. the number of observations per estimated parameter. Burnham and Anderson (2002) recommend that we need to use AICc instead of AIC if $\frac{n}{k} < 40$. Their recommendation agrees with our experience of studying AIC.

Comparing AIC (Equation (2.5)) with AICc (Equation (A.21)), the real difference is in the penalty term. In Figure A.1, we examined how the penalty terms of AIC and AICc behave differently as the sample size changes (as the x-axis values) under three different scenarios, case 1: $k = 2$; case 2: $k = 3$; and case 3: $k = 4$. Sample size n changes over a range up to $n = 100$. The y-axis values *penalty score* are calculated

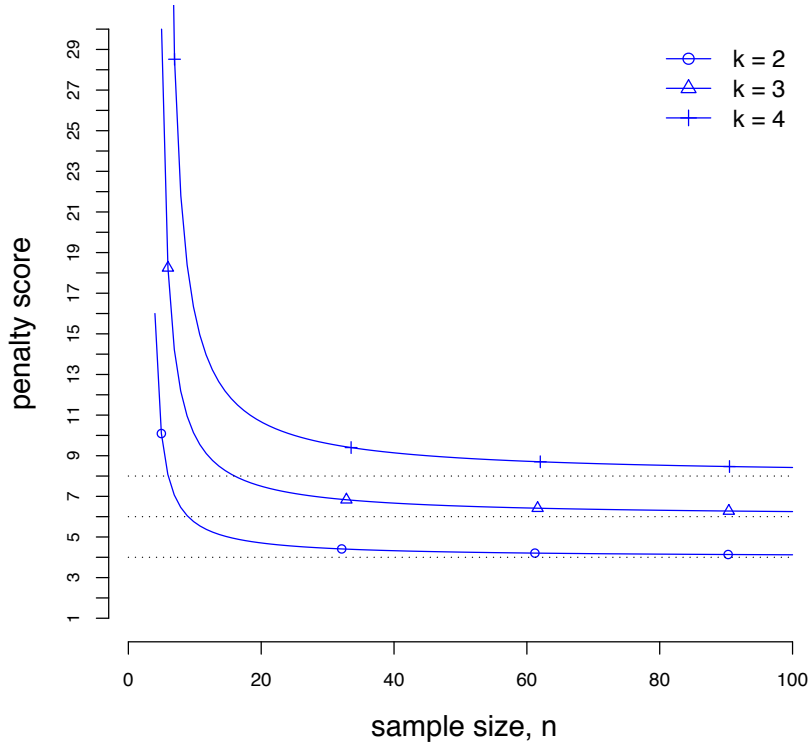

Figure A.1: Penalty term comparison: AIC versus AICc

based on the formulas: penalty score(AIC) $= 2k$ and penalty score(AICc) $= 2k\,\frac{n}{n-k-1}$ (the second term in Equations (2.5) and (A.21)). The solid line with circle points represents models with two estimated parameters ($k = 2$) of AICc penalty score; solid line with triangle points for $k = 3$ case; solid line with + signs for $k = 4$ case. The three horizontal dot lines are the corresponding AIC penalty scores. Figure A.1 shows us that all solid lines approach the dot lines closer and closer as n gets larger. Actually, the dot lines are the corresponding asymptotes of the solid lines since $\lim_{n\to\infty} 2k\,\frac{n}{n-k-1} = 2k$ for any fixed k. It is easy to check that the rule of using AICc if $\frac{n}{k} < 40$ is supported by Figure A.1.

Results from a further numerical investigation into AICc is given in Table A.1. Four normally distributed random samples are generated using R (by specifying `set.seed(101)`) and a normal distribution is fitted to each of the four samples. Using the R function `fitdistr` in MASS package, $\log \mathrm{L}(\underline{\hat{\theta}})$ where $\underline{\hat{\theta}} = (\hat{\mu}, \hat{\sigma^2})^T$, can be obtained from the list

Table A.1: Difference between AIC and AICc: when sample size is small

Normally distributed random sample	AIC	AICc	AICc – AIC	n/k
n=10	16.58	22.29	5.71	5
n=20	54.01	54.72	0.71	10
n=50	137.86	138.11	0.25	25
n=100	273.14	273.27	0.13	50

of the R output object. AIC and AICc is then calculated based on R output results and Equation (2.5) and (A.21). We can see that when $n = 100$ (hence $n/k = 50$) the difference between AIC and AICc is almost negligible, whereas the difference is relatively huge when $n = 10$.

In the case of $n/k < 40$, Burnham and Anderson (2002), p326, strongly recommended using AICc as an omnibus small-sample form of AIC unless a better form is known. Burnham and Anderson (2002) argues that since

$$\text{AICc}(k) = \text{AIC}(k) + \frac{2k(k+1)}{n-k-1}, \tag{A.24}$$

essentially AICc requires only that the likelihood function be proportional to a normal distribution in general. So long as the underlying distribution is unimodal and neither badly skewed nor heavy tailed, AICc will be a good approximation even at quite small sample sizes.

Note that we should not try to make a comparison between AIC (or AICc) scores in Table A.1. It is meaningless to compare AIC (or AICc) scores calculated from different data sets.

A.5 AIC, BIC, DIC and Bayesian model evaluation

Perhaps because of the nature of Bayesian statistics (a likelihood and prediction based approach which updating information conditional on an observed sample), it is popular in Bayesian data analysis to use information criteria for model evaluation (see, for example, Bernardo and Smith, 1993; Congdon, 2003; and Gelman et al., 2004). We concentrate our discussions on AIC, BIC, DIC, arguably the three most widely used information criteria in Bayesian model evaluation.

AIC and BIC have been defined in Equations (2.5) and (2.13), respectively. AIC and BIC use the same sample maximum likelihood term to measure the inaccuracy

of a model but have a different second term for penalizing the model complexity. Despite the similar formula structure, they are fundamentally different in various aspects. They are different in the general assumptions that underlie their derivation (an infinite-dimensional versus a finite true model); different in their optimization targets (minimization of the mean squared predictive error versus maximization of the marginal likelihood of the data); different in their model selection objective (a 'good' model in terms of an optimal balance of bias-variance trade-off versus a true model); different in statistical properties (efficient and weakly consistent versus strong model dimension consistency but not efficient). AIC is an unbiased estimator of the relative expected K-L information but BIC is not. The AIC selected 'best' model changes as the sample size changes which allows a more complex model to be selected as more information is available as the sample size getting larger. Hence, as $n \longrightarrow +\infty$ (no information loss about the true model is possible), AIC will finally select the true model (this is another expression of AIC's asymptotic consistency property). On the other hand, BIC selected 'best' model would be fixed regardless of the changes in sample size because BIC also assumes that the specified models must be fixed (you should not add in a new model or drop off any existing model). In general, an AIC selected best model is different from a BIC selected best model. Burnham and Anderson (2002), p295, made a comment that "But we are left with no theoretical basis to know what sort of parsimony the BIC model selection procedure has." Burnham and Anderson (2002) shown that AIC can be justified as a Bayesian model selection criterion. In practice, both AIC and BIC have been widely used for model evaluation in Bayesian or non-Bayesian cases.

In recent years, Deviance Information Criterion (DIC) has become more and more popular in Bayesian model evaluation. Deviances are most often discussed and used in regression models where there is a well-defined 'saturated model', i.e. a perfect fit model because each observation is described by a parameter. The *Deviance* of a model is defined as -2 times the log-likelihood: (see Gelman et al., 2004, or Claeskens and Hjort, 2008)

$$\text{deviance: } D(\underline{x}, \underline{\theta}) \;=\; -2 \log f(\underline{x}|\underline{\theta}). \tag{A.25}$$

$D(x, \underline{\theta})$ is proportional to the mean squared error if the model is normal with constant variance.

Let us denote a posterior distribution of $\underline{\theta}$ conditional on the observed sample by

$\pi(\underline{\theta}|\underline{x})$. A Bayes estimate of $\underline{\theta}$ is obtained by taking expectation over the posterior distribution

$$\overline{\underline{\theta}} = \mathrm{E}_{\pi(\underline{\theta}|\underline{x})}\,[\underline{\theta}].$$

A Bayesian version of the deviance difference is defined as

$$\mathrm{dd}(\underline{x}, \underline{\theta}, \overline{\underline{\theta}}) \;=\; D(\underline{x}, \underline{\theta}) - D(\underline{x}, \overline{\underline{\theta}}) \;=\; 2\left(\log f(\underline{x}|\overline{\underline{\theta}}) - \log f(\underline{x}|\underline{\theta})\right). \tag{A.26}$$

Therefore, the 'effective number of parameters' p_D, which is equivalent to the EMD concept defined in Equation (2.11), can be defined as

$$\begin{aligned} p_D \;&=\; \mathrm{E}_{\pi(\underline{\theta}|\underline{x})}\left[\mathrm{dd}(\underline{x}, \underline{\theta}, \overline{\underline{\theta}})\right] \\ &=\; \overline{D(\underline{x}, \underline{\theta})} - D(\underline{x}, \overline{\underline{\theta}}), \end{aligned} \tag{A.27}$$

where $\overline{D(\underline{x}, \underline{\theta})}$ is the posterior mean of $D(\underline{x}, \underline{\theta})$. DIC is then defined by

$$\begin{aligned} \mathrm{DIC} \;&=\; D(\underline{x}, \overline{\underline{\theta}}) + 2\,p_D \\ &=\; \overline{D(\underline{x}, \underline{\theta})} + p_D. \end{aligned} \tag{A.28}$$

DIC was developed by Spiegelhalter et al. (2002) as a generalization of AIC for Bayesian model evaluation.

Gelman et al. (2004) noted that the deviance has an important role in statistical model comparison because of its connection to the K-L information. The expected deviance (with respect to the true sampling distribution) equals 2 times the K-L information, up to a fixed constant. In the limiting case when the sample size approaches infinity, a model with the lowest K-L information, and hence the lowest expected deviance, will have the highest posterior probability. Gelman et al. (2004) further pointed out that the deviance is a standard summary of any discrepancy measure D which may measure the lack of fit of a model. Thus, Gelman et al. (2004), p181, concluded that "it seems reasonable to estimate expected deviance as a measure of overall model fit." Venables and Ripley (1996) noted that minimizing AIC will select the same 'best' model as minimizing a deviance based information criterion in the generalized linear models framework.

Burnham and Anderson (2002) noted that DIC does not use Bayes factors and behaves like AIC rather than like BIC. They also believed that DIC may have provided one solution in handling the models which include random effects. However, Burnham and Anderson (2002) considered the MCMC approach to model fitting as a disadvantage with DIC in comparison with AIC for model evaluation.

Claeskens and Hjort (2008) commented that DIC shares more similarities with the frequentist criterion AIC than the Bayesian BIC. They further shown that p_D in Equation (A.28) tends in probability to EMD $= \text{tr}\big(K(\underline{\hat{\theta}})J(\underline{\hat{\theta}})^{-1}\big)$ of TIC in Equation (2.6).

Overall, DIC is an information-theoretic-type Bayesian model selection criterion.

There is a small but important detail worth noting in Gelman et al. (2004). The authors had mentioned nothing at all about BIC through out the book, hence no any related references were cited. "Bayesian Data Analysis" as a very popular text book in its second edition, this must not be by accident. Our interpretation on this is, Gelman et al. did not consider BIC worth mentioning as a useful tool in Bayesian model checking (the word used by Gelman et al. for 'inference and model comparison'). This signals something similar to Burnham and Anderson (2002)'s comment on BIC as quoted earlier.

It is interesting to know that Akaike (1978) introduced a model selection criterion conceived from a Bayesian perspective for the problem of selecting a model in linear regression which is equivalent to BIC defined by Equation (2.13). In McQuarrie and Tsai (1998) and Chatfield (2004), this Akaike proposed Bayesian information criterion were mentioned as BIC and the criterion defined by Equation (2.13) was named as Schwarz Information Criterion (SIC) instead. Probably because Schwarz's derivation is more general than the usual linear regression, most literature nowadays refers Equation (2.13) as BIC and Akaike (1978)'s Bayesian information criterion is much less mentioned.

Appendix B

Background theory and some derived results for BLP models

B.1 Poisson process and some important properties

B.1.1 Definition

A *point process* series is often denoted by $\{Y_t, t \in T\}$. The subscript t, running over a suitable index set T. In this research, we concentrate on situations where t represents time. Therefore, if t for time index and the index set is $T = \mathbb{Z}_+ \equiv \{0, 1, 2, \dots\}$, $\{Y_t, t \in T\}$ is a discrete-time process; if $T = \mathbb{R}_+ \equiv [0, \infty)$, $\{Y_t, t \in T\}$ is a continuous-time process, or a process in continuous time.

The simplest stochastic point process is a *Poisson process*. Let $N(u, v)$ be the number of points in an interval (u, v), where $u < v$, and $\mathcal{H}_t$ denote the history of the process at time t. Then for a given positive constant λ (points per unit time), a Poisson process of rate λ is defined by

$$\Pr\{N(t, t+\delta) = 1 | \mathcal{H}_t\} = \lambda\delta + o(\delta), \tag{B.1}$$

$$\Pr\{N(t, t+\delta) > 1 | \mathcal{H}_t\} = o(\delta), \tag{B.2}$$

for all t, as $\delta \to 0^+$ (Cox and Isham, 1980).

Equations (B.1) and (B.2) express three important conditions for defining a Poisson process: (a) the number of occurrences in any two disjoint intervals must be independent of each other. Because any event defined by occurrences in (t, ∞) is independent of $\mathcal{H}_t$; (b) equation (B.1) requires that the probability of an occurrence in any arbitrarily small interval must be proportional to the length of that interval as $\delta \to 0^+$; (c) equation (B.2) virtually excludes the possibility of multiple simultaneous occurrences.

Equations (B.1) and (B.2) define a Poisson process by an intensity specification (Cox and Isham, 1980). Alternatively, a Poisson process can be defined either by an interval specification or by a counting specification. The interval specification states that the intervals between successive points are independently exponentially distributed with rate λ. The counting specification states that the number of points in any fixed interval of length t should have a Poisson distribution with mean λt. Therefore, a Poisson process may be treated as a discrete-time process in terms of its counting specification or as a continuous-time process in terms of its interval specification, depending on the situation. A Poisson process could be defined by any one of these three equivalent specifications. The Poisson process plays a central role in the theory of point processes as normal distribution plays in standard statistical data analysis and it usually serves as a starting point for the specification of a complex point process (Cox and Isham, 1980).

B.1.2 Some important properties

Different point processes may be distinguished by their clustering patterns by *index of dispersion* as defined in Cox and Isham (1980):

$$I_Y = \frac{\mathrm{Var}(Y)}{\{\mathrm{E}(Y)\}^2}, \tag{B.3}$$

or

$$I(t) = \frac{\mathrm{Var}\{N(t)\}}{\mathrm{E}\{N(t)\}}, \tag{B.4}$$

where Y is the interval variable between successive points and $N(t)$ is the count of the number of points in $(0, t]$ with the standard notation Var for variance and E for taking expectation on the interval variable or the count variable. By the definition of index of dispersion, it is clear that for a Poisson process $I_Y = 1$ or $I(t) = 1$ and it would be a deterministic case if $I_Y = 0$ or $I(t) = 0$. Therefore, often the Poisson process serves as a basis for qualitative comparison. If $0 < I_Y < 1$ or $0 < I(t) < 1$, such a point process is said to be under dispersed. On the other hand, If $I_Y > 1$ or $I(t) > 1$, a point process is said to be over dispersed. The higher the index of dispersion the more clustering in the process.

Let us denote

$$\delta N(t) = N(t, t+\delta) \to \mathrm{d}N(t) \text{ as } \delta \to 0^+.$$

Hence, $\mathrm{d}N(t) = 1$ if there is a point at time t; otherwise $\mathrm{d}N(t) = 0$. And the sum of point occurrences over the interval (u, v)

$$N(u, v) \;=\; \int_u^v \mathrm{d}N(t).$$

By assuming a stationary process with point occurrence rate λ, based on Equations (B.1) and (B.2), we have:

$$\begin{aligned} \mathrm{E}[N(0,t)] &= \mathrm{E}\Big[\int_0^t \mathrm{d}N(s)\Big] \\ &= \int_0^t \mathrm{E}\big[\mathrm{d}N(s)\big] \\ &= \int_0^t \lambda \mathrm{d}s \;=\; \lambda t, \end{aligned} \tag{B.5}$$

and

$$\begin{aligned} \mathrm{Var}[N(0,t)] &= \mathrm{Cov}\big[N(0,t), N(0,t)\big] \\ &= \int_0^t \int_0^t \mathrm{Cov}\big[\mathrm{d}N(t), \mathrm{d}N(s)\big]. \end{aligned} \tag{B.6}$$

The covariance in the integral is related to the *covariance density function* $c(u)$ which takes the form (Cox and Isham, 1980, p33):

$$c(u) \;=\; \lambda\delta(u) + \lambda h(u) - \lambda^2, \tag{B.7}$$

where $\delta(u)$ is the Dirac delta function[1] and $h(u)$ is the conditional intensity function with

$$h(u) \;=\; \lim_{\delta_1,\delta_2 \to 0^+} \frac{\Pr\big[N(u, u+\delta_2) > 0 | N(-\delta_1, 0) > 0\big]}{\delta_2}.$$

Therefore, $h(u)$ is a density for the probability of a point at time u given a point exists at time $t = 0$.

[1] An intuitively helpful definition of Dirac delta function is given by

$$\int_{-\infty}^{+\infty} \delta(x)\mathrm{d}x \;=\; 1. \tag{B.8}$$

This can be made rigorous by use of distribution theory (Chatfield, 2004).

The conditional intensity function can be calculated using the model hazard function. For a Poisson process of rate λ, therefore, $h(u) = \lambda$. Hence, the moments are (from Equation (B.5) to (B.8)):

$$\mathrm{E}[N(0,t)] \;=\; \lambda t, \tag{B.9}$$

and

$$\begin{aligned} \mathrm{Var}[N(0,t)] \;&=\; \int_0^t \int_0^t \lambda\delta(v-u)\mathrm{d}v\mathrm{d}u \\ &=\; \lambda \int_0^t \mathrm{d}u \;=\; \lambda t. \end{aligned} \tag{B.10}$$

Let $\mathrm{d}N(t)$ be a point process and $X(t)$ be any stochastic process.

$$Y(t) \;=\; X(t)\,\mathrm{d}N(t) \tag{B.11}$$

is a *marked point process*. In order to be able to derive properties of $Y(t)$, we usually require $X(t)$ and $\mathrm{d}N(t)$ to be independent.

B.2 Markov property and Markov chain

Let $\{X_n : n = 0, 1, 2, ...\}$ be a discrete stochastic process that takes on a finite or countable number of possible values. If $X_n = i$, then the process is said to be in state i at time n. The Markov property states that, given the past states $X_0, X_1, ..., X_{n-1}$ and the current state X_n, the conditional probability distribution of any future state X_{n+1} depends only on the current state:

$$\Pr(X_{n+1}|X_1, X_2, ..., X_n) = \Pr(X_{n+1}|X_n) \quad \text{for } n \geq 0.$$

A Markov chain is a discrete stochastic process with the Markov property that goes on forever. The concept of a Markov chain can be extended by defining higher order Markov chains. For example, a second order Markov chain is a discrete stochastic process with property

$$\Pr(X_{n+1}|X_1, X_2, ..., X_n) = \Pr(X_{n+1}|X_{n-1}, X_n) \quad \text{for } n \geq 1.$$

The changes of state of the process are called transitions and the probabilities associated with various state-changes are called transition probabilities. The elements of each row of a transition matrix should add up to unity.

For example, let $\{X_n\}$ be a daily rainfall occurrence process. Furthermore, let us define $X_n = 0$ if the nth day is dry and $X_n = 1$ if the nth day is wet. The transition matrix may be denoted by:

$$\begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix},$$

where $p_{11} = \Pr(X_n = 0|X_{n-1} = 0)$, $p_{12} = \Pr(X_n = 1|X_{n-1} = 0)$, $p_{21} = \Pr(X_n = 0|X_{n-1} = 1)$, and $p_{22} = \Pr(X_n = 1|X_{n-1} = 1)$. Therefore, p_{11} is the probability that the nth day is dry given the $(n-1)$st day is dry, and so on so forth. Obviously, $p_{11} + p_{12} = p_{21} + p_{22} = 1$. If we want to investigate what is the probability that the nth day is dry given the previous two days are dry, a second order Markov chain may need to be employed.

The probabilistic structure of a Markov chain is determined completely by its transition matrix and an initial probability distribution. For more about Markov chain, readers are referred to any of the standard text books on stochastic processes, e.g. see Feller (1968) or Norris (1997).

B.3 The third moment formula of the original BLP model

$$
\begin{aligned}
&\frac{\mathrm{E}((Y^{(h)})^3)}{\lambda\beta\xi^3} \\
&= \frac{6}{(\eta+\gamma)^2}\Big[\frac{\mathrm{E}(X_{ijk}X_{ij\ell}X_{ijn})}{\gamma(\eta+\gamma)} + \frac{\mathrm{E}(X_{ijk}X_{ij\ell})\mu_X\beta}{\gamma(\eta+\gamma)}\Big(\frac{1}{\eta}+\frac{\lambda}{\gamma(\eta+\gamma)}-\frac{\gamma}{\eta(2\eta+\gamma)}\Big) \\
&\quad -\frac{\mu_X^3\beta^2}{\eta(\eta+\gamma)(2\eta+\gamma)}\Big(\frac{1}{\eta}+\frac{\lambda}{\gamma(\eta+\gamma)}\Big)\Big]\Big[h-\frac{2}{\eta+\gamma}+\frac{2e^{-(\eta+\gamma)h}}{\eta+\gamma}+he^{-(\eta+\gamma)h}\Big] \\
&\qquad +\frac{6}{(\eta+\gamma)(2\eta+\gamma)}\Big[-\frac{2\mathrm{E}(X_{ijk}X_{ij\ell})\mu_X\beta}{\eta(\eta+\gamma)(2\eta+\gamma)}+\frac{\mu_X^3\beta^2}{\eta^2(2\eta+\gamma)(3\eta+\gamma)}\Big] \\
&\quad \times\Big[h-\frac{3\eta+2\gamma}{(\eta+\gamma)(2\eta+\gamma)}+\frac{(2\eta+\gamma)e^{-(\eta+\gamma)h}}{\eta(\eta+\gamma)}-\frac{(\eta+\gamma)e^{-(2\eta+\gamma)h}}{\eta(2\eta+\gamma)}\Big] \\
&\qquad +\frac{6\mu_X^3\beta^2}{\eta\gamma^3(\eta+\gamma)}\Big(\frac{1}{\eta}+\frac{\lambda}{\gamma(\eta+\gamma)}\Big)\Big[h-\frac{2}{\gamma}+\frac{2e^{-\gamma h}}{\gamma}+he^{-\gamma h}\Big] \\
&\quad +\frac{6}{\gamma(\eta+\gamma)}\Big[\frac{2\mathrm{E}(X_{ijk}X_{ij\ell})\mu_X\beta}{\eta\gamma(\eta+\gamma)}-\frac{\mu_X^3\beta^2}{\eta^2(\eta+\gamma)(2\eta+\gamma)}\Big] \\
&\qquad \times\Big[h-\frac{\eta+2\gamma}{\gamma(\eta+\gamma)}+\frac{(\eta+\gamma)e^{-\gamma h}}{\eta\gamma}-\frac{\gamma e^{-(\eta+\gamma)h}}{\eta(\eta+\gamma)}\Big] \\
&\quad +\frac{12}{\eta+\gamma}\Big[\frac{\mathrm{E}(X_{ijk}X_{ij\ell})\mu_X\lambda\beta}{\gamma^2(\eta+\gamma)^2}-\frac{\mu_X^3\lambda\beta^2}{\eta\gamma(\eta+\gamma)^2(2\eta+\gamma)}\Big] \\
&\qquad \times\Big[\frac{1}{2}h^2-\frac{h}{\eta+\gamma}+\frac{1}{(\eta+\gamma)^2}-\frac{e^{-(\eta+\gamma)h}}{(\eta+\gamma)^2}\Big] \\
&\quad +\frac{12\mu_X^3\lambda\beta^2}{\eta\gamma^3(\eta+\gamma)^2}\Big[\frac{1}{2}h^2-\frac{h}{\gamma}+\frac{1}{\gamma^2}-\frac{e^{-\gamma h}}{\gamma^2}\Big]+\frac{\mu_X^3\lambda^2\beta^2h^3}{\gamma^3(\eta+\gamma)^3} \\
&\quad +\frac{6\mathrm{E}(X_{ijk}^2X_{ij\ell})}{\xi\gamma(\eta+\gamma)^2}\Big[h-\frac{1}{\eta+\gamma}+\frac{e^{-(\eta+\gamma)h}}{\eta+\gamma}\Big] \\
&\quad +\frac{6\mathrm{E}(X^2)\mu_X\beta}{\xi\eta\gamma^2(\eta+\gamma)}\Big[h-\frac{1}{\gamma}+\frac{e^{-\gamma h}}{\gamma}-\frac{\gamma^2}{(\eta+\gamma)(2\eta+\gamma)}\Big(h-\frac{1}{\eta+\gamma}+\frac{e^{-(\eta+\gamma)h}}{\eta+\gamma}\Big)\Big] \\
&\quad +\frac{3\mathrm{E}(X^2)\mu_X\lambda\beta h^2}{\xi\gamma^2(\eta+\gamma)^2}+\frac{\mathrm{E}(X^3)h}{\xi^2\gamma(\eta+\gamma)}. \qquad \text{(B.12)}
\end{aligned}
$$

This third moment formula was derived for the original BLP model by CIO2007.

B.4 Parameter functions for different BLP models with continuous distributions of storm types

B.4.1 Model FIT-C1

The parameters are defined by

$$\left.\begin{array}{rcl} \beta(z) & = & \beta \\ \xi(z) & = & \xi \\ \theta(z) & = & \theta, \end{array}\right\}$$

where z is a uniform random variable with pdf $f(z) = 1; \quad 0 < z < 1$. Based on Equation (7.5), we have

$$\left.\begin{array}{rcl} W_{111} & = & \beta\xi\theta \\ W_{112} & = & \beta\xi\theta^2 \\ W_{222} & = & \beta^2\xi^2\theta^2 \\ W_{122} & = & \beta\xi^2\theta^2 \\ W_{233} & = & \beta^2\xi^3\theta^3 \\ W_{333} & = & \beta^3\xi^3\theta^3 \\ W_{123} & = & \beta\xi^2\theta^3 \\ W_{223} & = & \beta^2\xi^2\theta^3 \\ W_{113} & = & \beta\xi\theta^3 \\ W_{133} & = & \beta\xi^3\theta^3. \end{array}\right\} \tag{B.13}$$

B.4.2 Model FIT-C1-βzθz

The parameters are defined by

$$\left.\begin{array}{rcl} \beta(z) & = & \beta_0(1-z) + \beta_1 z \\ \xi(z) & = & \xi \\ \theta(z) & = & \theta z, \end{array}\right\}$$

where z is a uniform random variable with pdf $f(z) = 1; \quad 0 < z < 1$. Based on Equation (7.5), we have

$$\left.\begin{array}{rcl} W_{111} &=& \xi\theta(\frac{1}{6}\beta_0 + \frac{1}{3}\beta_1) \\ W_{112} &=& \xi\theta^2(\frac{1}{12}\beta_0 + \frac{1}{4}\beta_1) \\ W_{222} &=& \xi^2\theta^2(\frac{1}{30}\beta_0^2 + \frac{1}{10}\beta_0\beta_1 + \frac{1}{5}\beta_1^2) \\ W_{122} &=& \xi^2\theta^2(\frac{1}{12}\beta_0 + \frac{1}{4}\beta_1) \\ W_{233} &=& \xi^3\theta^3(\frac{1}{60}\beta_0^2 + \frac{1}{15}\beta_0\beta_1 + \frac{1}{6}\beta_1^2) \\ W_{333} &=& \xi^3\theta^3(\frac{1}{140}\beta_0^3 + \frac{1}{35}\beta_0^2\beta_1 + \frac{1}{14}\beta_0\beta_1^2 + \frac{1}{7}\beta_1^3) \\ W_{123} &=& \xi^2\theta^3(\frac{1}{20}\beta_0 + \frac{1}{5}\beta_1) \\ W_{223} &=& \xi^2\theta^3(\frac{1}{60}\beta_0^2 + \frac{1}{15}\beta_0\beta_1 + \frac{1}{6}\beta_1^2) \\ W_{113} &=& \xi\theta^3(\frac{1}{20}\beta_0 + \frac{1}{5}\beta_1) \\ W_{133} &=& \xi^3\theta^3(\frac{1}{20}\beta_0 + \frac{1}{5}\beta_1). \end{array}\right\} \tag{B.14}$$

B.4.3 Model FIT-C1-ξzθz

The parameters are defined by

$$\left.\begin{array}{rcl} \beta(z) &=& \beta \\ \xi(z) &=& \xi_0(1-z) + \xi_1 z \\ \theta(z) &=& \theta z, \end{array}\right\}$$

where z is a uniform random variable with pdf $f(z) = 1; \quad 0 < z < 1$. Based on Equation (7.5), we have

$$\left.\begin{array}{rcl} W_{111} &=& \beta\theta(\frac{1}{6}\xi_0 + \frac{1}{3}\xi_1) \\ W_{112} &=& \beta\theta^2(\frac{1}{12}\xi_0 + \frac{1}{4}\xi_1) \\ W_{222} &=& \beta^2\theta^2(\frac{1}{30}\xi_0^2 + \frac{1}{10}\xi_0\beta_1 + \frac{1}{5}\xi_1^2) \\ W_{122} &=& \beta\theta^2(\frac{1}{30}\xi_0^2 + \frac{1}{10}\xi_0\beta_1 + \frac{1}{5}\xi_1^2) \\ W_{233} &=& \beta^2\theta^3(\frac{1}{140}\xi_0^3 + \frac{1}{35}\xi_0^2\xi_1 + \frac{1}{14}\xi_0\xi_1^2 + \frac{1}{7}\xi_1^3) \\ W_{333} &=& \beta^3\theta^3(\frac{1}{140}\xi_0^3 + \frac{1}{35}\xi_0^2\xi_1 + \frac{1}{14}\xi_0\xi_1^2 + \frac{1}{7}\xi_1^3) \\ W_{123} &=& \beta\theta^3(\frac{1}{60}\xi_0^2 + \frac{1}{15}\xi_0\xi_1 + \frac{1}{6}\xi_1^2) \\ W_{223} &=& \beta^2\theta^3(\frac{1}{60}\xi_0^2 + \frac{1}{15}\xi_0\xi_1 + \frac{1}{6}\xi_1^2) \\ W_{113} &=& \beta\theta^3(\frac{1}{20}\xi_0 + \frac{1}{5}\xi_1) \\ W_{133} &=& \beta\theta^3(\frac{1}{140}\xi_0^3 + \frac{1}{35}\xi_0^2\xi_1 + \frac{1}{14}\xi_0\xi_1^2 + \frac{1}{7}\xi_1^3). \end{array}\right\} \tag{B.15}$$

B.4.4 Model FIT-C1-βzξz

The parameters are defined by

$$\left.\begin{array}{rcl}\beta(z) &=& \beta_0(1-z)+\beta_1 z\\ \xi(z) &=& \xi_0(1-z)+\xi_1 z\\ \theta(z) &=& \theta,\end{array}\right\}$$

where z is a uniform random variable with pdf $f(z)=1;\quad 0<z<1$. Based on Equation (7.5), we have

$$\left.\begin{array}{rcl}
W_{111} &=& \theta(\frac{1}{3}\beta_0\xi_0+\frac{1}{6}\beta_0\xi_1+\frac{1}{6}\beta_1\xi_0+\frac{1}{3}\beta_1\xi_1)\\
W_{112} &=& \theta^2(\frac{1}{3}\beta_0\xi_0+\frac{1}{6}\beta_0\xi_1+\frac{1}{6}\beta_1\xi_0+\frac{1}{3}\beta_1\xi_1)\\
W_{222} &=& \theta^2(\frac{1}{5}\beta_1^2\xi_1^2+\frac{1}{5}\beta_0^2\xi_0^2+\frac{1}{30}\beta_0^2\xi_1^2+\frac{2}{15}\beta_0\beta_1\xi_0\xi_1+\frac{1}{10}\beta_0^2\xi_0\xi_1+\frac{1}{10}\beta_0\beta_1\xi_0^2\\
&& +\frac{1}{10}\beta_0\beta_1\xi_1^2+\frac{1}{10}\beta_1^2\xi_0\xi_1+\frac{1}{30}\beta_1^2\xi_0^2)\\
W_{122} &=& \theta^2(\frac{1}{4}\beta_0\xi_0^2+\frac{1}{6}\beta_0\xi_0\xi_1+\frac{1}{12}\beta_0\xi_1^2+\frac{1}{12}\beta_1\xi_0^2+\frac{1}{6}\beta_1\xi_0\xi_1+\frac{1}{4}\beta_1\xi_1^2)\\
W_{233} &=& \theta^3(\frac{1}{60}\beta_0^2\xi_1^3+\frac{1}{6}\beta_1^2\xi_1^3+\frac{1}{60}\beta_1^2\xi_0^3+\frac{1}{10}\beta_0\beta_1\xi_0^2\xi_1+\frac{1}{10}\beta_0\beta_1\xi_0\xi_1^2+\frac{1}{6}\beta_0^2\xi_0^3\\
&& +\frac{1}{10}\beta_0^2\xi_0^2\xi_1+\frac{1}{20}\beta_0^2\xi_0\xi_1^2+\frac{1}{15}\beta_0\beta_1\xi_0^3+\frac{1}{15}\beta_0\beta_1\xi_1^3+\frac{1}{20}\beta_1^2\xi_0^2\xi_1+\frac{1}{10}\beta_1^2\xi_0\xi_1^2)\\
W_{333} &=& \theta^3(\frac{1}{7}\beta_1^3\xi_1^3+\frac{1}{140}\beta_0^3\xi_1^3+\frac{1}{140}\beta_1^3\xi_0^3+\frac{1}{7}\beta_0^3\xi_0^3+\frac{1}{14}\beta_0^3\xi_0^2\xi_1+\frac{1}{35}\beta_0^3\xi_0\xi_1^2\\
&& +\frac{1}{14}\beta_0^2\beta_1\xi_0^3+\frac{1}{35}\beta_0^2\beta_1\xi_1^3+\frac{1}{35}\beta_0\beta_1^2\xi_0^3+\frac{1}{14}\beta_0\beta_1^2\xi_1^3+\frac{1}{35}\beta_1^3\xi_0^2\xi_1+\frac{1}{14}\beta_1^3\xi_0\xi_1^2\\
&& +\frac{3}{35}\beta_0^2\beta_1\xi_0^2\xi_1+\frac{9}{140}\beta_0^2\beta_1\xi_0\xi_1^2+\frac{9}{140}\beta_0\beta_1^2\xi_0^2\xi_1+\frac{3}{35}\beta_0\beta_1^2\xi_0\xi_1^2)\\
W_{123} &=& \theta^3(\frac{1}{4}\beta_0\xi_0^2+\frac{1}{6}\beta_0\xi_0\xi_1+\frac{1}{12}\beta_0\xi_1^2+\frac{1}{12}\beta_1\xi_0^2+\frac{1}{6}\beta_1\xi_0\xi_1+\frac{1}{4}\beta_1\xi_1^2)\\
W_{223} &=& \theta^3(\frac{1}{5}\beta_1^2\xi_1^2+\frac{1}{5}\beta_0^2\xi_0^2+\frac{1}{30}\beta_0^2\xi_1^2+\frac{2}{15}\beta_0\beta_1\xi_0\xi_1+\frac{1}{10}\beta_0^2\xi_0\xi_1+\frac{1}{10}\beta_0\beta_1\xi_0^2\\
&& +\frac{1}{10}\beta_0\beta_1\xi_1^2+\frac{1}{10}\beta_1^2\xi_0\xi_1+\frac{1}{30}\beta_1^2\xi_0^2)\\
W_{113} &=& \theta^3(\frac{1}{3}\beta_0\xi_0+\frac{1}{6}\beta_0\xi_1+\frac{1}{6}\beta_1\xi_0+\frac{1}{3}\beta_1\xi_1)\\
W_{133} &=& \theta^3(\frac{1}{5}\beta_0\xi_0^3+\frac{3}{20}\beta_0\xi_0^2\xi_1+\frac{1}{10}\beta_0\xi_0\xi_1^2+\frac{1}{20}\beta_0\xi_1^3+\frac{1}{20}\beta_1\xi_0^3+\frac{1}{10}\beta_1\xi_0^2\xi_1\\
&& +\frac{3}{20}\beta_1\xi_0\xi_1^2+\frac{1}{5}\beta_1\xi_1^3).
\end{array}\right\} \quad \text{(B.16)}$$

B.4.5 Model FIT-C1-βzξzθz

The parameters are defined by

$$\left.\begin{array}{rcl}\beta(z) &=& \beta_0(1-z)+\beta_1 z\\ \xi(z) &=& \xi_0(1-z)+\xi_1 z\\ \theta(z) &=& \theta z,\end{array}\right\}$$

where z is a uniform random variable with pdf $f(z) = 1; \quad 0 < z < 1$. Based on Equation (7.5), we have

$$\left.\begin{aligned}
W_{111} &= \theta(\tfrac{1}{12}\beta_0\xi_0 + \tfrac{1}{12}\beta_0\xi_1 + \tfrac{1}{12}\beta_1\xi_0 + \tfrac{1}{4}\beta_1\xi_1)\\
W_{112} &= \theta^2(\tfrac{1}{30}\beta_0\xi_0 + \tfrac{1}{20}\beta_0\xi_1 + \tfrac{1}{20}\beta_1\xi_0 + \tfrac{1}{5}\beta_1\xi_1)\\
W_{222} &= \theta^2(\tfrac{1}{7}\beta_1^2\xi_1^2 + \tfrac{1}{105}\beta_0^2\xi_0^2 + \tfrac{1}{105}\beta_0^2\xi_1^2 + \tfrac{4}{105}\beta_0\beta_1\xi_0\xi_1 + \tfrac{1}{70}\beta_0^2\xi_0\xi_1 + \tfrac{1}{70}\beta_0\beta_1\xi_0^2\\
&\quad + \tfrac{1}{21}\beta_0\beta_1\xi_1^2 + \tfrac{1}{21}\beta_1^2\xi_0\xi_1 + \tfrac{1}{105}\beta_1^2\xi_0^2)\\
W_{122} &= \theta^2(\tfrac{1}{60}\beta_0\xi_0^2 + \tfrac{1}{30}\beta_0\xi_0\xi_1 + \tfrac{1}{30}\beta_0\xi_1^2 + \tfrac{1}{60}\beta_1\xi_0^2 + \tfrac{1}{15}\beta_1\xi_0\xi_1 + \tfrac{1}{6}\beta_1\xi_1^2)\\
W_{233} &= \theta^3(\tfrac{1}{252}\beta_0^2\xi_1^3 + \tfrac{1}{9}\beta_1^2\xi_1^3 + \tfrac{1}{504}\beta_1^2\xi_0^3 + \tfrac{1}{84}\beta_0\beta_1\xi_0^2\xi_1 + \tfrac{1}{42}\beta_0\beta_1\xi_0\xi_1^2 + \tfrac{1}{504}\beta_0^2\xi_0^3\\
&\quad + \tfrac{1}{210}\beta_0^2\xi_0^2\xi_1 + \tfrac{1}{168}\beta_0^2\xi_0\xi_1^2 + \tfrac{1}{315}\beta_0\beta_1\xi_0^3 + \tfrac{1}{36}\beta_0\beta_1\xi_1^3 + \tfrac{1}{84}\beta_1^2\xi_0^2\xi_1 + \tfrac{1}{24}\beta_1^2\xi_0\xi_1^2)\\
W_{333} &= \theta^3(\tfrac{1}{10}\beta_1^3\xi_1^3 + \tfrac{1}{840}\beta_0^3\xi_1^3 + \tfrac{1}{840}\beta_1^3\xi_0^3 + \tfrac{1}{840}\beta_0^3\xi_0^3 + \tfrac{1}{420}\beta_0^3\xi_0^2\xi_1 + \tfrac{1}{420}\beta_0^3\xi_0\xi_1^2\\
&\quad + \tfrac{1}{420}\beta_0^2\beta_1\xi_0^3 + \tfrac{1}{120}\beta_0^2\beta_1\xi_1^3 + \tfrac{1}{420}\beta_0\beta_1^2\xi_0^3 + \tfrac{1}{30}\beta_0\beta_1^2\xi_1^3 + \tfrac{1}{120}\beta_1^3\xi_0^2\xi_1 + \tfrac{1}{30}\beta_1^3\xi_0\xi_1^2\\
&\quad + \tfrac{1}{140}\beta_0^2\beta_1\xi_0^2\xi_1 + \tfrac{3}{280}\beta_0^2\beta_1\xi_0\xi_1^2 + \tfrac{3}{280}\beta_0\beta_1^2\xi_0^2\xi_1 + \tfrac{1}{40}\beta_0\beta_1^2\xi_0\xi_1^2)\\
W_{123} &= \theta^3(\tfrac{1}{140}\beta_0\xi_0^2 + \tfrac{2}{105}\beta_0\xi_0\xi_1 + \tfrac{1}{42}\beta_0\xi_1^2 + \tfrac{1}{105}\beta_1\xi_0^2 + \tfrac{1}{21}\beta_1\xi_0\xi_1 + \tfrac{1}{7}\beta_1\xi_1^2)\\
W_{223} &= \theta^3(\tfrac{1}{8}\beta_1^2\xi_1^2 + \tfrac{1}{280}\beta_0^2\xi_0^2 + \tfrac{1}{168}\beta_0^2\xi_1^2 + \tfrac{1}{42}\beta_0\beta_1\xi_0\xi_1 + \tfrac{1}{140}\beta_0^2\xi_0\xi_1 + \tfrac{1}{140}\beta_0\beta_1\xi_0^2\\
&\quad + \tfrac{1}{28}\beta_0\beta_1\xi_1^2 + \tfrac{1}{28}\beta_1^2\xi_0\xi_1 + \tfrac{1}{168}\beta_1^2\xi_0^2)\\
W_{113} &= \theta^3(\tfrac{1}{60}\beta_0\xi_0 + \tfrac{1}{30}\beta_0\xi_1 + \tfrac{1}{30}\beta_1\xi_0 + \tfrac{1}{6}\beta_1\xi_1)\\
W_{133} &= \theta^3(\tfrac{1}{280}\beta_0\xi_0^3 + \tfrac{3}{280}\beta_0\xi_0^2\xi_1 + \tfrac{1}{56}\beta_0\xi_0\xi_1^2 + \tfrac{1}{56}\beta_0\xi_1^3 + \tfrac{1}{280}\beta_1\xi_0^3 + \tfrac{1}{56}\beta_1\xi_0^2\xi_1\\
&\quad + \tfrac{3}{56}\beta_1\xi_0\xi_1^2 + \tfrac{1}{8}\beta_1\xi_1^3).
\end{aligned}\right\} \tag{B.17}$$

B.5 Verification of the numerical solution for calculating proportion dry

The proportion dry $\mathrm{P}\big[Y_i^{(h)} = 0\big]$ is expressed in the following equation:

$$\begin{aligned}
-\lambda^{-1}\log\ \mathrm{P}\big[Y_i^{(h)} = 0\big] &= \int_0^\infty \mathrm{d}x\big[e^{-\gamma x} - f(h,x) - \gamma\int_0^h \mathrm{d}u f(u,x)\big]\\
&\quad + \int_0^h \mathrm{d}x\big[1 - g(x) - \gamma\int_0^x \mathrm{d}u g(u)\big]
\end{aligned} \tag{B.18}$$

where the functions f and g are given by:

$$\left.\begin{aligned}
-\log\ f(u,x) &= \gamma(x+u) + \tfrac{\beta\xi}{\eta(\eta+\xi)}(1-e^{-\eta x})(1-e^{-(\eta+\xi)u})\\
&\quad + \tfrac{\beta u\xi}{\eta+\xi} - \tfrac{\beta\xi}{(\eta+\xi)^2}(1-e^{-(\eta+\xi)u})\\
-\log\ g(u) &= \gamma u + \tfrac{\beta u\xi}{\eta+\xi} - \tfrac{\beta\xi}{(\eta+\xi)^2}(1-e^{-(\eta+\xi)u}).
\end{aligned}\right\} \tag{B.19}$$

Note that Equations (B.18) and (B.19) are the reproduction of Equations (6.5) and (6.6) which are cited from CIO2007.

In order to express $\mathrm{P}\big[Y_i^{(h)} = 0\big]$ explicitly in a closed form, we need to solve an integral of the following form:

$$\int_0^x e^{e^x} \mathrm{d}x \;=\; -\int_1^\infty \frac{e^{te^x}}{t} \mathrm{d}t$$

which is called 'Exponential integral'. Only a numerical solution can be obtained for an 'Exponential integral' problem (e.g. Press et al., 1992, p222). Because both integrand functions involved in Equation (B.18), $e^{-\gamma x} - f(h,x) - \gamma \int_0^h \mathrm{d}uf(u,x)$ and $1 - g(x) - \gamma \int_0^x \mathrm{d}ug(u)$, are continuous smooth functions and converge to zero very fast, as shown by Figure B.1 and B.2, a simple solution is to apply Simpson's rule (e.g. Kreyszig, 1993, p961) to get a numerical solution for $\mathrm{P}\big[Y_j^{(h)} = 0\big]$.

To compute the integral $\mathrm{J} = \int_a^b w(x)\mathrm{d}x$, we divide the interval of integration $a \le x \le b$ into an *even* number of equal subintervals, say, into $n = 2m$ subintervals of length $D = (b-a)/2m$, with endpoints $x_0(=a), x_1, ..., x_{2m-1}, x_{2m}(=b)$. The Simpson's rule formula can then be expressed as

$$\int_a^b w(x)\mathrm{d}x \;\approx\; \frac{D}{3}(w_0 + 4w_1 + 2w_2 + 4w_3 + ... + 2w_{2m-2} + 4w_{2m-1} + w_{2m}), \qquad \text{(B.20)}$$

where $w_j = w(x_j)$.

Our Simpson's rule approximation algorithm for calculating the PD values for a superposed BLP model of two independent processes of storms is as follows. Let $W_1 = e^{-\gamma x} - f(h,x) - \gamma \int_0^h \mathrm{d}uf(u,x)$ and $W_2 = 1 - g(x) - \gamma \int_0^x \mathrm{d}ug(u)$. Then Equation (B.18) can be rewritten as

$$-\lambda^{-1} \log \mathrm{P}\big[Y_i^{(h)} = 0\big] \;=\; \int_0^\infty W_1 \mathrm{d}x + \int_0^h W_2 \mathrm{d}x,$$

where h is the aggregation time scale level.

In order to apply the Simpson's rule for approximation we need to examine how fast the integrand function W_1 approaches zero. This is done numerically in figures B.1 and B.2. These graphs are plotted based on the BLP model FIT-O parameter estimates. The left column panels show the process 1 storm cases and the right column panels show the process 2 storm cases. Based on figures B.1 and B.2, following approximations are made before applying the Simpson's rule formula (B.20).

For type 1 process of storms

$$-\lambda_1^{-1} \log \mathrm{P}_1\big[Y_i^{(h)} = 0\big] \;\approx\; \int_0^{300} W_1 \mathrm{d}x + \int_0^h W_2 \mathrm{d}x,$$

and for type 2 process of storms

$$-\lambda_2^{-1} \log \mathrm{P}_2\big[Y_i^{(h)} = 0\big] \;\approx\; \int_0^{120} W_1 \mathrm{d}x + \int_0^{h} W_2 \mathrm{d}x.$$

For applying the Simpson's rule formula (B.20), we need to determine how many subintervals to divide for integrals $\int_0^{300} W_1 \mathrm{d}x$, $\int_0^{120} W_1 \mathrm{d}x$, and $\int_0^{h} W_2 \mathrm{d}x$, respectively. Our study shows that following scheme works adequately. For integration interval $0 \leq x \leq 300$, we have set the number of subintervals $n = 2m = 150$; for integration interval $0 \leq x \leq 120$, $2m = 60$; and for integration interval $0 \leq x \leq h$, $2m = 30$. Once the numerical solutions for $\mathrm{P}_1\big[Y_i^{(h)} = 0\big]$ and $\mathrm{P}_2\big[Y_i^{(h)} = 0\big]$ are obtained, for a superimposed BLP model of two independent processes, the PD values can be calculated as

$$\mathrm{P}\big[Y_i^{(h)} = 0\big] \;=\; \mathrm{P}_1\big[Y_i^{(h)} = 0\big] \;\times\; \mathrm{P}_2\big[Y_i^{(h)} = 0\big].$$

Using the parameter estimates from model specification FIT-O, long simulation samples are generated and PD values at different aggregation time scale levels are calculated. The verification results of the Simpson's rule approximation algorithm of the fitted model's PD values against the simulation samples PD values are given in Figure B.3. The simulation results show that our Simpson's rule algorithm works adequately well in calculating the approximated PD values of a fitted BLP model. In all cases examined, we have observed that the fitted PD values (bold solid lines) are well located in the middle of the distribution ranges of the 30 simulation samples PD values (+ points) which are calculated based on the Simpson's rule approximation algorithm.

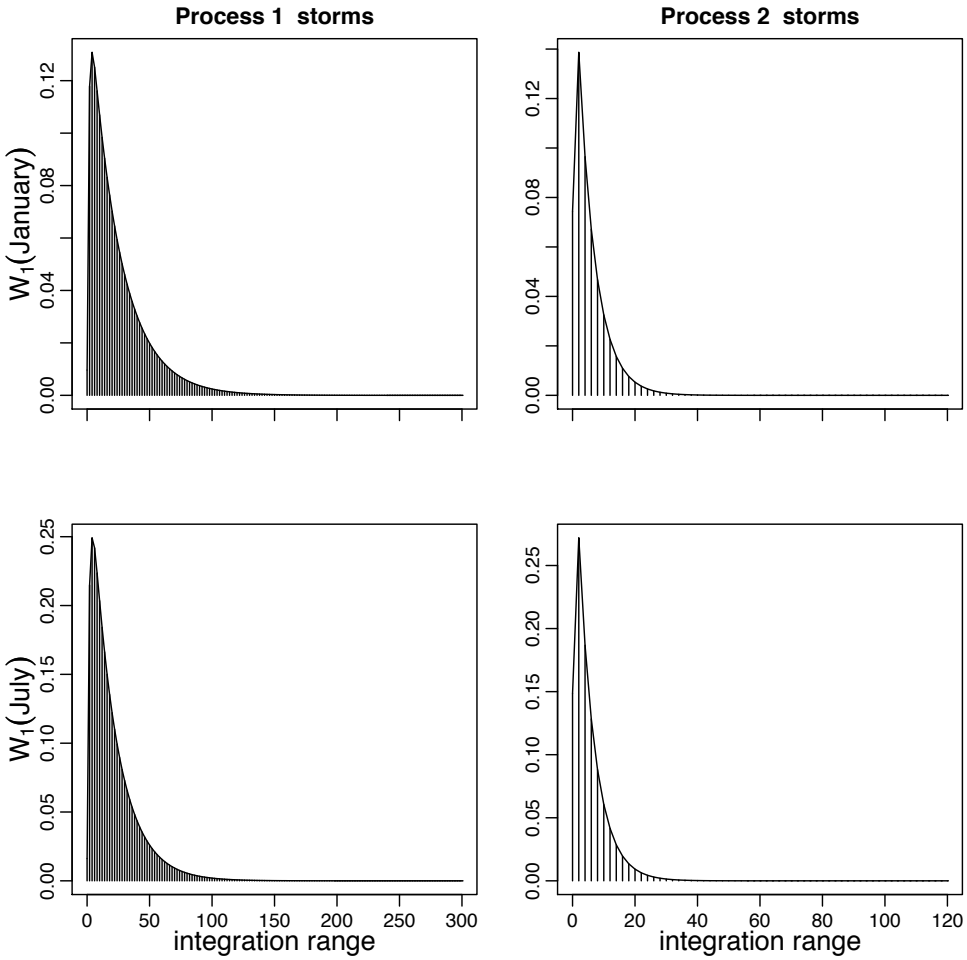

Figure B.1: Numerical examination of the integrand function $W_1 = e^{-\gamma x} - f(h,x) - \gamma \int_0^h \mathrm{d}u f(u,x)$, where h=5-min.

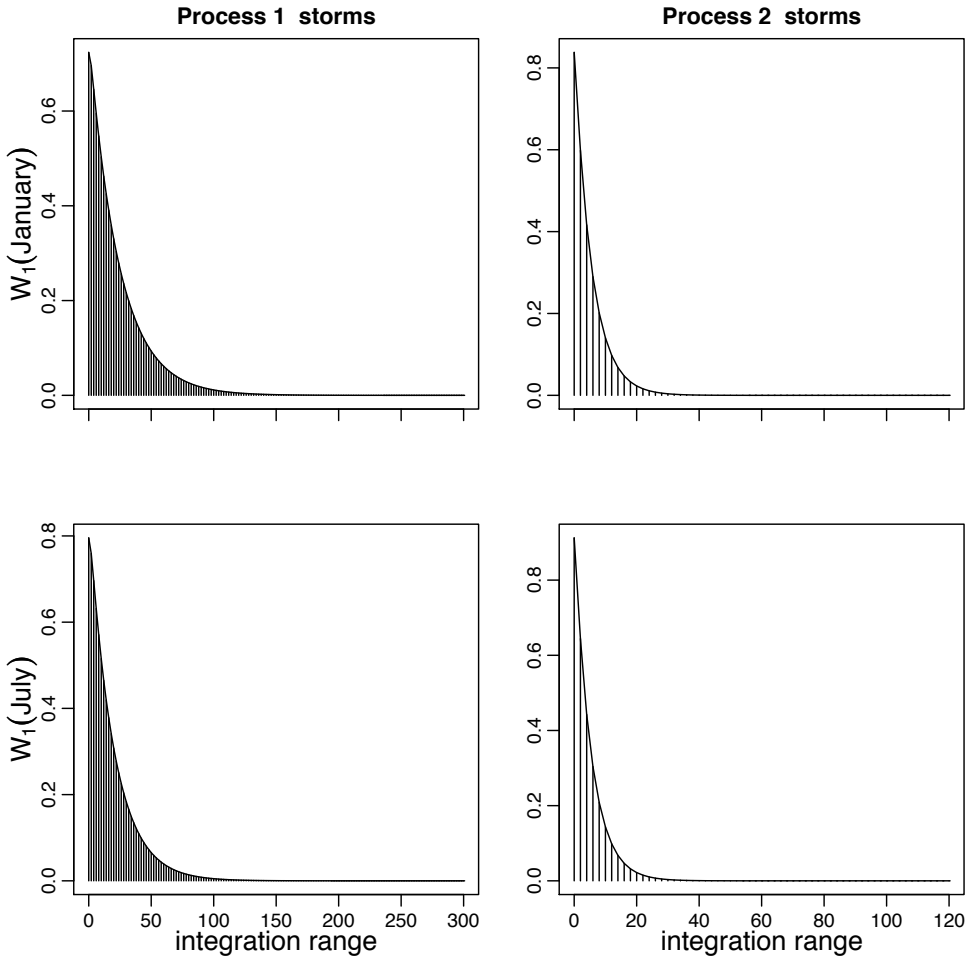

Figure B.2: Numerical examination of the integrand function $W_1 = e^{-\gamma x} - f(h,x) - \gamma \int_0^h \mathrm{d}u f(u,x)$, where h=24 hours.

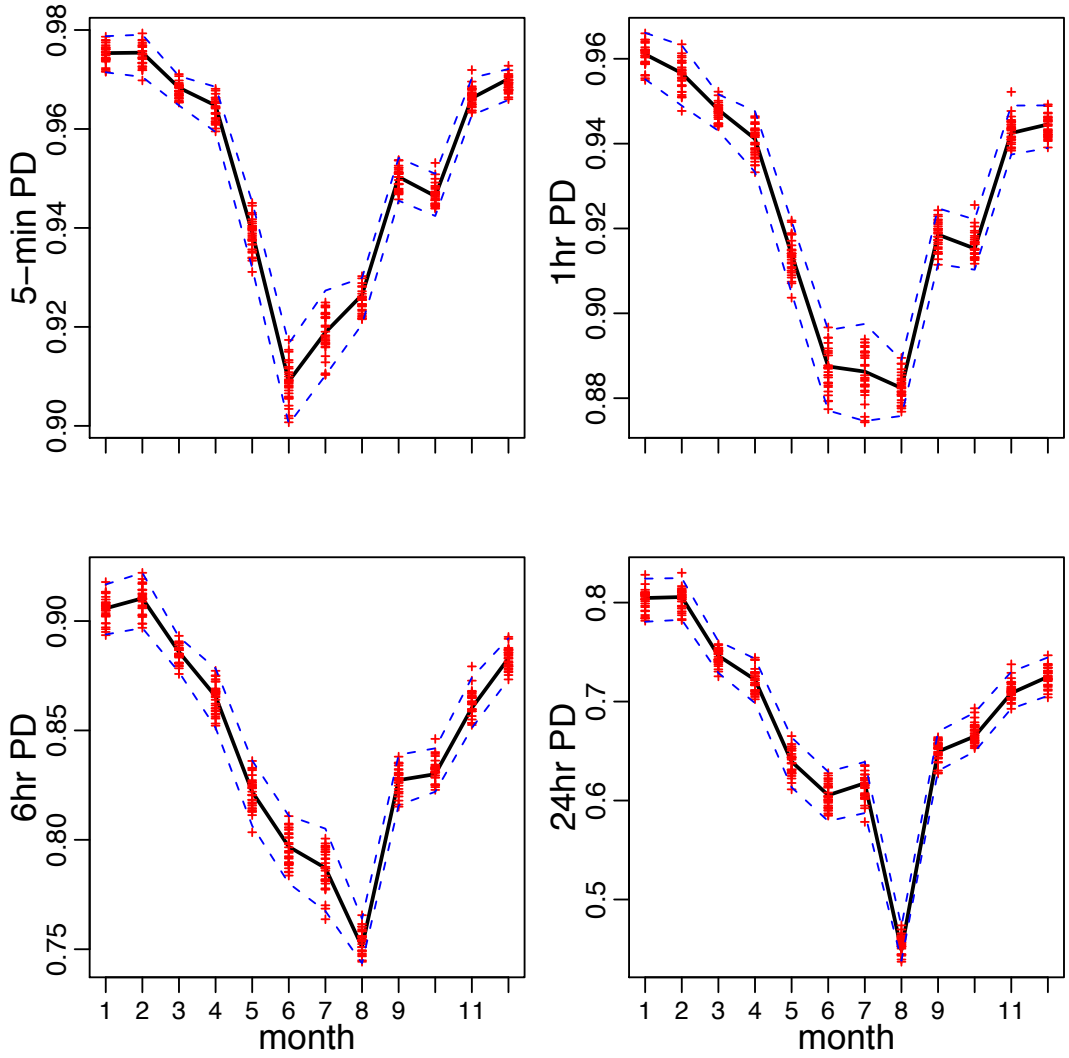

Figure B.3: Verification of the approximation of PD calculation using the Simpson's rule. The bold solid lines represent the PD values calculated from the Simpson's rule formula based on the parameter estimates of the fitted model, FIT-O. The PD values calculated from 30 simulation samples of FIT-O, each is of 100 years, are shown as points (+). The dashed lines form the 95% quantile confidence band are based on the simulated PD values.

References

Acreman, M. C. (1990). A simple stochastic model of hourly rainfall for Farnborough, England. *Hydrological Sciences*, 35:No.2, 119–148.

Akaike, H. (1969). Fitting autoregressive models for prediction. *Annuals of the Institute of Statistical Mathematics*, 21:243–247.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. *In B.N. Petrov & B.F. Csaki (Eds.), Second International Symposium on Information Theory,Academiai Kiado: Budapest*, pages 267–281.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, AC-19*, pages No.6, 716–723.

Akaike, H. (1978). A Bayesian analysis of the minimum AIC procedure. *Annals of the Institute of Statistical Mathematics*, pages 30, 9–14.

Applebaum, D. (2008). *Probability and Information: an integrated approach*. Cambridge University Press, second edition.

Austin, P. M. and Houze, R. A. (1972). Analysis of the structure of precipitation patterns in new england. *Journal of applied meteorology*, 11:926–935.

Bauwens, L., Giot, P., Gramming, J., and Veredas, D. (2004). A comparison of financial duration models via density forecasts. *International Institute of Forecasters*, 20:589–609.

Bauwens, L. and Hautsch, N. (2006). Modelling financial high frequency data using point processes. *CORE Discussion Paper*, 80.

Bernardo, J. M. and Smith, A. F. M. (1993). *Bayesian Theory*. John Wiley & Sons.

Bozdogan, H. (1987). Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, 52:No.3, 345–370.

Buishand, T. (1977). Stochastic modelling of daily rainfall sequences. *Mededlingen Landbouwhogeschool, Wageningen.*

Burkhardt, W. (1994). *First steps in Maple.* Springer-Verlag London Limited.

Burnham, K. and Anderson, D. (2002). *Model Selection and Multimodel Inference: a practical information-theoretic approach.* Springer.

Burton, A., Kilsby, C., Fowler, H., Cowpertwait, P., and O'Connell, P. (2008). Rainsim: A spatial-temporal stochastic rainfall modelling system. *Environmental Modelling & Software*, 23:1356–1369.

Butler, D. and Davies, J. W. (2004). *Urban Drainage.* Spon Press, 2nd edition.

Cameron, D., Beven, K., and Tawn, J. (2000). An evaluation of three stochastic rainfall models. *Journal of Hydrology*, 228:130–149.

Casella, G. and Berger, R. L. (2002). *Statistical Inference.* the Wadsworth Group, Thomson Learning Inc., 2nd edition.

Chatfield, C. (2004). *The Analysis of Time Series: an introduction.* CHAPMAN & HALL/CRC, 6th edition.

Claeskens, G. and Hjort, N. L. (2008). *Model selection and model averaging.* Cambridge University Press.

Congdon, P. (2003). *Applied Bayesian Modelling.* John Wiley & Sons Inc.

Cowpertwait, P. (1994). A generalized point process model of rainfall. *Proceedings of the Royal Society of London, Series A*, 447:23–37.

Cowpertwait, P. (1995). A generalized spatial-temporal model of rainfall based on a clustered point process. *Proceedings of the Royal Society of London, Series A*, 450:163–175.

Cowpertwait, P. (1998). A Poisson-cluster model of rainfall: high-order moments and extreme values. *Proceedings of the Royal Society of London, Series A*, 454:885–898.

Cowpertwait, P. (2004). Mixed rectangular pulses models of rainfall. *Hydrology and Earth System Sciences*, 8:993–1000.

Cowpertwait, P. (2010). A Neyman-Scott model with continuous distributions of storm types. *Australian and New Zealand Industrial and Applied Mathematics Journal*, 51:C97C108.

Cowpertwait, P., Isham, V., and Onof, C. (2007). Point process models of rainfall: developments for fine-scale structure. *Proceedings of the Royal Society of London, Series A*, 463:2569–2587.

Cowpertwait, P., O'Connell, P., Metcalfe, A., and Mawdsley, J. (1996a). Stochastic point process modelling of rainfall. I. Single-site fitting and validation. *Journal of Hydrology*, 175:17–46.

Cowpertwait, P., O'Connell, P., Metcalfe, A., and Mawdsley, J. (1996b). Stochastic point process modelling of rainfall. II. Regionalisation and disaggregation. *Journal of Hydrology*, 175:47–65.

Cowpertwait, P. S. P. (1991). Further developments of the Neyman-Scott clustered point process for modeling rainfall. *Water Resources Research*, 27:No.7, 1431–1438.

Cox, D. and Isham, V. (1988). A simple spatial-temporal model of rainfall. *Proceedings of the Royal Society of London, Series A*, 415:317–328.

Cox, D. R. and Isham, V. (1980). *Point Processes*. Chapman & Hall.

Cox, D. R. and Isham, V. (1994). Stochastic models of precipitation. *In: V. Barnett and K.F. Turkman (Eds.), Statistics for the Environment 2: Water Related Issues, Wiley, Chichester, UK, pp. 218.*

Cox, D. R. and Lewis, P. A. W. (1978). *The Statistical Analysis of Series of Events*. Chapman and Hall, London.

Crawley, M. J. (2007). *The R Book*. John Wiley & Sons, Ltd, The Atrium.

Davison, A. C. (2003). *Statistical Models*. Cambridge University Press.

DeGroot, M. H. (1986). *Probability and Statistics*. Addison-Wesley Publishing Company, 2nd edition.

Efron, B. (2000). Hypothesis testing in the 20th century with a special reference to testing with misspecified models. *Statistics For The 21st Century: methodologies for applications of the future, edited by C.R. Rao and G. J. Szekely, pp.109–144.*

Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap.* Chapman & Hall/CRC.

Engle, R. F. and Russell, J. R. (1998). Autoregressive conditional duration: A new model for irregularly spaced transaction data. *Econometrica*, 66:No.5, 1127–1162.

Entekhabi, D., Rodriguez-Iturbe, I., and Eagleson, P. S. (1989). Probabilistic representation of the temporal rainfall process by a modified neyman-scott rectangular pulses model: parameter estimation & validation. *Water Resources Research*, 25:No.2, 295–302.

Feller, W. (1968). *An Introduction to Probability Theory and its Applications, vol. 1.* John Wiley, New York.

Foufoula-Georgiou, E. and Krajewski, W. (1995). Recent advances in rainfall modeling, estimation, and forecasting. *Reviews of Geophysics*, pages 1125–1137.

Foufoula-Georgiou, E. and Lettenmaier, D. (1986). Compatibility of continuous rainfall occurrence models with discrete rainfall observations. *Water Resources Research*, 22:No.8, 1316–1322.

Foufoula-Georgiou, E. and Lettenmaier, D. P. (1987). A Markov renewal model for rainfall occurrences. *Water resources research*, 23:No.5, 875–884.

Friedman, J. (1984). A variable span smoother. *Technical Report No. 5, Laboratory for Computational Statistics, Department of Statistics, Standford University.*

Frost, A., Srikanthan, R., and Cowpertwait, P. (2004). Stochastic generation of point rainfall data at subdaily timescales: A comparison of DRIP and NSRP. *Technical Report, Report 04/9, Cooperative Research Centre for Catchment Hydrology.*

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004). *Bayesian Data Analysis.* Chapman & Hall/CRC, second edition.

Gupta, V. and Waymire, E. (1993). A statistical analysis of mesoscale rainfall as a random cascade. *Applied Meteorology*, 32:No.2, 251–267.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.

Heck, A. (1996). *Introduction to Maple*. Springer-Verlag New York, Inc., second edition.

Huber, P. J. and Ronchetti, E. M. (2009). *Robust statistics*. John Wiley & Sons, Ltd, second edition.

Hurvich, C. and Tsai, C. (1989). Regression and time series model selection in small samples. *Biometrika*, 76:297–307.

Kakou, A. and Onof, C. (1996). A point process model for rainfall with duration intensity dependence. *Annales Geophysicae, Suppl. II to vol. 14*.

Kavvas, M. L. and Delleur, J. W. (1975). The stochastic and chronologic structure of rainfall sequences c application to indiana. *Water Resources Research Center, Purdue University, West Lafayette, Iniana*.

Kavvas, M. L. and Delleur, J. W. (1981). A stochastic model of daily rainfall sequences. *Water Resources Research*, 17:No.4, 1151–1160.

Konishi, S. and Kitagawa, G. (1996). Generalized information criteria in model selection. *Biometrika*, 83:875–890.

Konishi, S. and Kitagawa, G. (2008). *Information criteria and statistical modeling*. Springer Science+Business Media, LLC.

Koutsoyiannis, D. (2002). The Hurst phenomenon and fractional Gaussian noise made easy. *Hydrological Sciences Journal*, 47:573–595.

Kreyszig, E. (1993). *Advanced Engineering Mathematics*. John Wiley & Sons, 7th edition.

Krishnamoorthy, K. (2006). *Handbook of Statistical Distributions with Applications*. Chapman & Hall/CRC Press.

Kullback, S. and Leibler, R. (1951). On information and sufficiency. *Annuals of Mathematical Statistics*, 22:79–86.

Lindsey, J. K. (1996). *Parametric Statistical Inference*. Oxford University Press Inc. New York.

Ljung, G. and Box, G. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65:No.2, 297–303.

Luca, G. D. and Gallo, G. M. (2004). Mixture processes for financial intradaily durations. *Studies in Nonlinear Dynamics & Econometrics*, 8:No.2, Article8.

Manly, B. F. (1994). *Multivariate statistical methods: a primer*. Chapman & Hall, second edition.

Le Cam, L. M. (1961). A stochastic description of precipitation. *In: Neyman, J. (Ed.), Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, vol. 3. University of California, Berkeley, California, pp. 165186.*

McQuarrie, A. D. R. and Tsai, C.-L. (1998). *Regression and Time Series Model Selection*. World Scientific, Sigapore.

Meitz, M. and Terasvirta, T. (2006). Evaluating models of autoregressive conditional duration. *Journal of Business & Economic Statistics*, 24:104–124.

Metcalfe, A. V. (1997). *Statistics in Civil Engineering*. John Wiley & Sons, Inc., New York.

Nelder, J. and Mead, R. (1965). A simplex algorithm for function minimization. *Computer Journal*, 7:308–313.

Nelder, J. A. and Wedderburn, R. W. M. (1972). General linear models. *Journal of the Royal Statistical Society*, pages 370–384.

Neyman, J. and Scott, E. L. (1958). Statistical approach to problems of cosmology. *Journal of the Royal Statistical Society, Series B*, 20:No.1, 1–43.

Norris, J. (1997). *Markov Chains*. Cambridge University Press.

O'Hara, M. (1995). *Market Microstructure Theory*. Blackwell Publishers Ltd.

Onof, C., Chandler, R. E., Kakou, A., Northrop, P., Wheater, H. S., and Isham, V. (2000). Rainfall modelling using Poisson-cluster processes: a review of developments. *Stochastic Environmental Research and Risk Assessment*, 14:384–411.

Onof, C., Townend, J., and Kee, R. (2005). Comparison of two hourly to 5-min rainfall disaggregators. *Atmospheric Research*, 77:176–187.

Onof, C. and Wheater, H. S. (1993). Modelling of British rainfall using a random parameter bartlett-lewis rectangular pulse model. *Hydrological Sciences Journal*, 149:67–95.

Onof, C. and Wheater, H. S. (1994). Improvements to the modelling of British rainfall using a modified random parameter Bartlett-Lewis rectangular pulse model. *Journal of Hydrology*, 157:177–195.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1992). *Numerical Recipes in C*. Cambridge University Press, 2nd edition.

R Development Core Team (2007). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Rissanen, J. (1978). Modeling by shortest data description. *Automatica*, 14:465–471.

Rodriguez-Iturbe, I., Cox, D., and Isham, V. (1987a). Some models for rainfall based on stochastic point processes. *Proc. R. Soc. Lond. A*, 410:269–288.

Rodriguez-Iturbe, I., Cox, D., and Isham, V. (1988). A point process model for rainfall: further developments. *Proc. R. Soc. Lond. A*, 417:283–298.

Rodriguez-Iturbe, I. and Eagleson, P. S. (1987). Mathematical models of rainstorm events in space and time. *Water Resources Research*, 23:No.1, 181–190.

Rodriguez-Iturbe, I., Gupta, V. K., and Waymire, E. (1984). Scale considerations in the modeling of temporal rainfall. *Water Resources Research*, 20:No.11, 1611–1619.

Rodriguez-Iturbe, I., Febres de Power, B., and Valdes, J. B. (1987b). Rectangular pulses point process models for rainfall: analysis of empirical data. *Journal of Geophysical Research*, 92:D8, 9645–9656.

Sansom, J. (1987). Digitising pluviographs. *Journal of Hydrology (N.Z.)*, 26:No.2.

Sansom, J. (1992). Breakpoint representation of rainfall. *Journal of Applied Meteorology*, 31:1514–1519.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6:No.2, 461–464.

Shaw, E. M. (1988). *Hydrology in Practice*. Chapman & Hall, second edition.

Shibata, R. (1980). Asymptotic efficient selection of the order of the model for estimating parameters of a linear process. *Annals of Statistics*, 8:147–164.

Shumway, R. H. and Stoffer, D. S. (2006). *The Analysis of Time Series An Introduction*. Springer Science+Business Media, LLC, 2nd edition.

Smith, D. J. and Teo, K. L. (1989). *Linear Algebra*. New Zealand Mathematical Society (Inc).

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of Royal Statistical Society*, pages 583–639.

Srikanthan, R. and McMahon, T. A. (2001). Stochastic generation of annual, monthly and daily climate data: a review. *Hydrology and Earth System Sciences*, 5:No.4, 653–670.

Stern, R. and Coe, R. (1984). A model fitting analysis of daily rainfall data. *Journal of the Royal Statistical Society, Series A (General)*, 147:No.1, 1–34.

Stone, M. (1977). An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion. *Journal of the Royal Statistical Society*, pages 44–47.

Sugiura, N. (1978). Further analysis of the data by Akaike's information criterion and the finite corrections. *Communications in Statistics Series*, pages No.1, 13–26.

Takeuchi, K. (1976). Distribution of information statistics and criteria for adequacy of models. *Mathematical Science*, 153:1218, in Japanese.

The GCC team (2007). *GCC: A free C language software package, compiler version 4.1.1.*

Tsay, R. S. (2005). *Analysis of Financial Time Series*. John Wiley & Sons, Inc., 2nd edition.

Venables, W. and Ripley, B. (1996). *Modern Applied Statistics with S-Plus*. Springer-Verlag New York, Inc., corrected fourth printing.

Verhoest, N., Troch, P., and Troch, F. D. (1997). On the applicability of Bartlett-Lewis rectangular pulses models in the modeling design storms at a point. *Journal of Hydrology*, 202:108–120.

Waymire, E. and Gupta, V. K. (1981a). The mathematical structure of rainfall representation, 1. a review of the stochastic rainfall models. *Water Resources Research*, 17:No.5, 1261–1272.

Waymire, E. and Gupta, V. K. (1981b). The mathematical structure of rainfall representation, 2. a review of the theory of point processes. *Water Resources Research*, 17:No.5, 1273–1285.

Waymire, E. and Gupta, V. K. (1981c). The mathematical structure of rainfall representation, 3. some applications of the point process theory to rainfall processes. *Water Resources Research*, 17:No.5, 1287–1294.

Wedderburn, R. W. M. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika*, 61:439–447.

Wheater, H., Chandler, R., Onof, C., Isham, V., Bellone, E., Yang, C., Lekkas, D., Lourmas, G., and Segond, M.-L. (2005). Spatial-temporal rainfall modelling for flood risk estimation. *Stoch. Environ. Res. Risk Assess.*, 19:403–416.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, 50:No.1, 1–25.

Wood, S. N. (2006). *Generalized additive models: an introduction with R*. Chapman & Hall/CRC.

Zhang, M., Russell, J., and Tsay, R. S. (2001). A nonlinear autoregressive conditional duration model with applications to financial transaction data. *Journal of Econometrics*, 104:179–207.

Index