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