PARAMETER-DRIVEN COUNT TIME SERIES MODELS

NAWWAL AHMAD BUKHARI

FACULTY OF SCIENCE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2018

PARAMETER-DRIVEN COUNT TIME SERIES MODELS

NAWWAL AHMAD BUKHARI

DISSERTATION SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE

INSTITUTE OF MATHEMATICAL SCIENCES
FACULTY OF SCIENCE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2018

UNIVERSITI MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: (I.C./Passport No.:)

Registration/Matric No.:

Name of Degree:

Title of Project Paper/Research Report/Dissertation/Thesis ("this Work"):

Field of Study:

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature Date:

Subscribed and solemnly declared before,

Witness's Signature Date:

Name:
Designation:

PARAMETER-DRIVEN COUNT TIME SERIES MODELS

ABSTRACT

Time series data involving counts are commonly encountered in many different fields including insurance industry, economics, medicine, communications, epidemiology, hydrology and meteorology. In this study, a parameter-driven count time series model with three different distributions that are Poisson, zero-inflated Poisson and negative binomial was developed. A key property of our model is that the distributions of the observed count data are independent, conditional on the latent process, although the observations are correlated marginally. The first part of the study derives the explicit solutions of the moment properties (mean, variance, skewness and kurtosis) of the distributions together with their respective autocovariance and autocorrelation functions, up to the *i*th order. The empirical study shows that the derivation fits the theoretical results of an autocorrelation function. The estimation of parameter is in the second part of this study. Since the proposed model is non-linear and non-Gaussian, Monte Carlo Expectation Maximization (MCEM) algorithm with the aid of particle filtering and particle smoothing methods are applied to approximate the integrals in the E-step of the algorithm. The proposed model are illustrated with simulated data and an application on Malaysia dengue data. Simulation shows that MCEM algorithm and particle method are useful for the parameter estimation of the Poisson model. In addition, Poisson model fits better in terms of Akaike information criterion (AIC) and log-likelihood when compared with several models including model from Yang et al. (2015).

Keywords: Dengue data, integer time series, Sequential Monte Carlo, State-space models.

MODEL BILANGAN SIRI MASA BERPACUKAN PARAMETER

ABSTRAK

Data siri masa yang melibatkan bilangan lazimnya ditemui di dalam pelbagai bidang termasuk industri insurans, ekonomi, perubatan, komunikasi, epidemiologi, hidrologi dan meteorologi. Di dalam kajian ini, kami mengolah satu model bilangan siri masa berpacukan parameter bersama tiga taburan berbeza, iaitu Poisson, Poisson sifar melambung dan binomial negatif. Satu ciri penting model berpacukan parameter adalah tidak bersandar dan bersyarat ke atas proses terpendam taburan, walaupun cerapan berkolerasi marginal. Bahagian pertama kajian ini menerbitkan penyelesaian yang jelas (min, varians, kepencongan dan kurtosis) bagi setiap taburan berserta fungsi autokovarians dan autokorelasi, sehingga tertib ke-i. Kajian empirikal menunjukkan terbitan tersebut berpadanan dengan teori fungsi autokorelasi. Anggaran parameter adalah bahagian kedua kajian ini. Oleh kerana model yang yang diolah adalah tidak linear dan tidak Gaussian, kami menggunakan algoritma MCEM, di mana kaedah penapisan dan pelicinan zarah digunakan bagi menganggar pengkamiran di dalam langkah algoritma-E tersebut. Model yang diolah dipersembahkan melalui data simulasi dan aplikasi ke atas data denggi Malaysia. Simulasi kami menunjukkan algoritma MCEM dan kaedah zarah membantu mendapatkan anggaran parameter bagi model Poisson. Sebagai tambahan, model Poisson yang dicadangkan lebih berpadanan dari segi AIC dan log-kebarangkalian jika dibandingkan dengan beberapa model lain termasuk model yang di perkenalkan oleh Yang et al. (2015).

Kata Kunci: Data denggi, integer siri masa, Monte Carlo berurutan, model ruang keadaan.

ACKNOWLEDGEMENTS

In completing this thesis, there are several people I would like to convey my deepest gratitude to. First and foremost to my supervisors, Dr. Koh You Beng and Prof. Dr. Ibrahim bin Mohamed. Their guidance and support has helped me in completing this study. The amount of knowledge, understanding and dedication that both of them has shown me has been truly inspiring for me to push forward and work hard for my Master's degree.

Secondly, without the financial support from Kementerian Pengajian Tinggi Malaysia which offered me a scholarship for Master's degree, this work would have not been possible. Next, I would also like to thank the staff of the Institute of Mathematical Sciences for the support they have given me over the past few years.

Last but not least, I would like to thank my family for their unconditional love and support throughout my two years of graduate school. And also to all my friends especially those in the postgraduate room, thank you for all your helps and supports throughout this journey.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

iid	:	identically and independently distributed.
ACD	:	autoregressive conditional duration.
AR	:	autoregressive.
ARMA	:	autoregressive moving average.
DLM	:	dynamic linear model.
DSOE	:	dual source of error.
EM	:	Expectation Maximization.
GARCH	:	Generalized autoregressive conditionally heteroscedastic.
GLM	:	generalized linear models.
HMM	:	hidden Markov model.
MCEM	:	Monte Carlo expectation-maximization.
MLE	:	maximum likelihood estimation.
NB	:	negative binomial.
RCA	:	random coefficient autoregressive.
SCD	:	stochastic conditional duration.
SIS	:	sequential importance sampling.
SMC	:	Sequential Monte Carlo.
SSOE	:	single source of error.
SV	:	stochastic volatility.
ZIP	:	zero inflated Poisson.

CHAPTER 1: INTRODUCTION

1.1 Background

In general, there are two categories of time series models: observation-driven models and parameter-driven models (Cox, 1981) which differs in account of autocorrelation. The temporal correlation in between observations for observation-driven models are modelled directly from the function of previous responses. However, parameter-driven models contradict in a form where an unobserved latent process is employed to narrate the serial correlation. Presumably, the observations are to be independently distributed, conditioning on the latent process. Compared to observation-driven models, the concept behind parameter-driven models is more appealing. However, parameter estimation in parameter-driven models is often impossible (Chan & Ledolter, 1995; Durbin & Koopman, 2000).

Time series of nonnegative counts have become available in various fields including actuarial science, computer science, economics, epidemiology, finance, hydrology, and meteorology.

The U.S. polio incidence counts, which was applied by Zeger & Qaqish (1988) initially, serves as a benchmark data set in numerous count time series papers. With a sample size $N = 168$, it is a monthly data from January 1970 to December 1983. The data set has been employed widely, e.g Chan & Ledolter (1995), Kuk & Cheng (1997), Davis et al. (1999, 2000), Fahrmeir & Tutz (2001), Fokianos (2001), Jung & Liesenfeld (2001), Davis & Wu (2009) and Zhu (2011) with different conclusions as to the significance of each paper.

The time series model has been widely used in explaining the performance of financial time series. For example, modelling a physician expenditure by Shumway & Stoffer (1982) and the analysis of pound-dollar daily exchange rates in Kim & Stoffer (2008).

Time series model has been used in discussing motorway casualties around the world. For example Bhattacharyya & Layton (1979) on the effectiveness of seat belt legislation, Harvey & Fernandes (1989) on the effect of seat belt legislation on British road casualties, Michener & Tighe (1992) discussed the effect of increasing speed limit on fatal highway accidents, Harvey & Durbin (1986) and Johansson (1996) on the effect of speed limit on minor injuries and vehicle damage.

1.1.1 Observation-driven Continuous-Valued Time Series Models

Transaction and quote data is an example of analysis financial intraday data which has attract a massive interest among researchers. This is because, this type of data is irregular spaced, dissimilar to closing price with evenly space time. Another additional element, the intervals between events are randomly distributed.

To study the dynamic structure of the adjusted durations x_i with $x_i = t_i - t_{i-1}$, where t_i is the time of the ith transaction, Engle & Russell (1998) proposed the autoregressive conditional duration (ACD) model. ϕ_i is the conditional expectation of the adjusted duration between the $(i-1)$th and the ith trade. The basic ACD model is defined as

$$x_i = \phi_i \epsilon_i, \tag{1.1}$$

$$\phi_i = \mathrm{E}\left[x_i | F^x_{t_i-1}\right] \tag{1.2}$$

where ϵ_i are the identically and independently distributed (iid) nonnegative random variables with density function $f(.)$ and unit mean, and $F^x_{t_i-1}$ is the information available up to $(i-1)$th trade, ϵ_i also assumed to be independent of $F^x_{t_i-1}$.

Other continuous-valued time series models defined under observation-driven model is

random coefficient autoregressive (RCA);

$$y_t = (\theta + b_t)y_{t-1} + \epsilon_t \tag{1.3}$$

where b_t and ϵ_t are uncorrelated zero mean processes with unknown variance σ_b^2 and variance $\sigma_\epsilon^2(\theta)$ with unknown parameter θ.

The nonlinear case of an RCA model was discussed in Tjøstheim (1986), which was referred to as doubly stochastic time series model,

$$y_t = \theta_t f\left(t, F_{t-1}^y\right) + \epsilon_t, \tag{1.4}$$

where $\{\theta + b_t\}$ in (1.3) is replaced by a more general stochastic sequence $\{\theta_t\}$ and y_{t-1} is replaced by a function of the past, $f(t, F_{t-1}^y)$. θ_t is a moving average sequence of the form

$$\theta_t = \theta + a_t + a_{t-1}, \tag{1.5}$$

where a_t consists of square integrable independent random variables with mean zero and variance σ_a^2.

1.1.2 Parameter-driven Continuous-Valued Time Series Models

A random parameter model is equal to a dual source of error (DSOE), in which both measurement and state equations contain a source of randomness (Feigin et al., 2008). For example the stochastic volatility (SV) model which was first introduced by Taylor (1982).

In the standard SV model structure, the probability function $f(r|h)$ is used to generate the returned data, r. h is an unobserved vector of volatilities with a probabilistic structures $f(h|\theta)$, where θ is a vector of parameters. In the standard form of the model, volatility is

modelled as an autoregressive (AR) process,

$$h_t = \phi h_{t-1} + w_t \tag{1.6}$$

and the returns are given by

$$r_t = \beta \exp\left(\frac{h_t}{2}\right)\epsilon_t, \tag{1.7}$$

where w_t and ϵ_t are independent process.

Equation (1.7) is linearized by taking the logarithm of the squared returns, which yields the equation

$$y_t = \alpha + h_t + v_t, \tag{1.8}$$

where $y_t = \log(r_t^2)$, $\alpha = \log(\beta^2) + E\left(\log(\epsilon_t^2)\right)$ and $v_t = \log(\epsilon_t^2) - E\left(\log(\epsilon_t^2)\right)$.

Equations (1.6) and (1.8) are the standard univariate SV model which results in a linear, non-Gaussian, state-space model for which (1.8) is the observation equation and (1.6) is the state equation.

Kim & Stoffer (2008) retains the state equation in (1.6) but the observation in equation (1.8) is changed to

$$y_t = \alpha + x_t + v_t, \tag{1.9}$$

$$v_t = I_t z_{t1} + (1 - I_t) z_{t0} - \mu\pi, \tag{1.10}$$

with z_{t0} ~iidN$(0, R_0)$, z_{t1} ~iidN(μ, R_1), I_t is an indicator variable and π is an unknown mixing probability; i.e. $\Pr(I_t = 1) = \pi = 1 - \Pr(I_t = 0)$. An additional term $\mu\pi$ is added into the equation making v_t a zero mean variable.

1.2 Problem statement

This research starts with the parameter-driven Poisson model proposed by Feigin et al. (2008). As this research developed, we propose two new parameter-driven models using zero-inflated Poisson distribution and negative Binomial distribution. We later derived the moment of properties for these models.

1.3 Objectives

The objectives of the study are:

a. Derive the higher order moments of parameter-driven Poisson model proposed by Feigin et al. (2008).
b. Propose new parameter-driven models using zero-inflated Poisson distribution and negative Binomial distribution and derive the moment properties of these models.
c. Obtain parameter estimation of the Poisson model.
d. Apply the Poisson model on real data set.

1.4 Research outline

This research focuses on several parameter-driven count time series models. It is outlined as follows:

Chapter two gives the literature review on count times series models and the differences between observation-driven and parameter-driven models. The reviews on several estimation methods upon these models are also presented as well.

Chapter three discusses the properties of a parameter-driven count time series models. In this chapter, we proposed new parameter-driven models using different conditional discrete distributions such as Poisson, zero-inflated Poisson and negative binomial. For each models, we derived their moment properties up to the fourth moment, kurtosis. The simulation study of each models are presented as well.

Chapter four focused on the estimation of the Poisson model. The step-by-step derivations for the MCEM algorithm are shown in this chapter. The simulation study results are also presented in this chapter.

Chapter five is the application of this model on real-life data. We consider the Malaysia dengue data and the application was discussed later on this chapter.

Chapter six is the conclusion of this research. We summarised our findings and provides some suggestions on expanding this research in the future.

CHAPTER 2: LITERATURE REVIEW

2.1 Count Time Series Models

Count time series have been widely used in various fields these days. Cameron & Trivedi (2013) is the pioneer in discussing the models for time series of counts. Fahrmeir & Tutz (2001) discussed several models within the framework of generalized linear models (GLM) followed by Fokianos & Kedem (2002). McKenzie (2003) provided a complete historical development of the field and the extensive class of discrete-valued time series models with a distinct insistence on models based on thinning mechanisms.

Generating sequence of dependent random variates is crucial. However, it relies heavily on a particular marginal distribution and correlation structure, hence simulation is essential in evaluating count time series model (Phatarfod & Mardia, 1973). Box et al. (2015) is the first to study the simple linear time series models in Gaussian case. Tong (1990) has broaden the approach to the non-linear and non-Gaussian cases.

Recently, a number of work on count time series modelling can be found in the literature. Ahmad & Francq (2016) looked at estimating the mean parameter of a count time series models using the quasi-maximum likelihood estimation method. They applied the method on integer-valued autoregressive and generalized autoregressive conditional heteroscedasticity models. For low-count time seriese models, Drovandi & McCutchan (2016) used Bayesian approach in making inferences on the parameters and model selection. The approach does not involve between-model proposals required in reversible jump Markov chain Monte Carlo, and does not rely on weak approximations. In a hidden markov set up, Sebastian et al. (2018) proposed a method of analysing count time series data taking into account the dependency between states from which more counts are reported and the transition between states due to some spatial condition. Möller et al. (2018) extended

the binomial AR model for autocorrelated counts designed for different varieties of zero patterns. Specifically, they consider the case when there is a large number of zeroes in a finite support $\{0, 1, ..., n\}$ with a fixed upper limit $n \in N$ are missing. They handle the issue using the hidden markov model and illustrated using monetary policy decisions by the National Bank of Poland.

Based on Cox (1981), there are two classes of models for count data: observation-driven and parameter-driven models. It is carried out by focusing on methods to interprate the data and to forecast empirically. Many papers have also been published in comparing the performance of observation- and parameter-driven models on different types of data. Hasan et al. (2016) compared the two models using the zero-inflated Poisson models for longitudinal count data. They incorporate the serial correlation to overcome the problem of separating zeroes and positive responses in observation-driven model and latent process with random effects for parameter-driven model respectively. The performance is examined via simulation while the models are illustrated using hospital utilization data. They found out that the standard errors for the observation-driven model are significantly underestimated that might lead to misleading conclusion. Koopman et al. (2016) investigated the observation- and parameter-driven models for predicting the time-varying parameters. Via simulation, they showed that the observation-driven model and correctly specified parameter-driven have similar predictive accuracy. Catania & Nonejad (2016) investigated the leverage effect in generating the density forecase of equity returns using the observation-driven and parameter-driven models. The comparison of the two models considered are carried out on a large number of financial time series. The authors considered the observatin-driven t-EGARCH(1,1) (see Nelson, 1991), Beta-t-EGARCH (see Harvey, 2013) and SPEGARCH(1,1) (see Pascual et al., 2006) models. In additon, the use the stocahstic votality model (SV) as proposed by Kim et al. (1998). The problem

is applied on the Dow Jones and S& P 500 data. The result indicates the Beta-t-EGARCH is the preferred model for the problem considered. In the next two sections, we will look at the observation- and parameter-driven models.

2.1.1 Observation-driven models

An observation-driven model is where the serial correlation is modelled directly via lagged values of the count variable, with strategies adopted to ensure that the integer nature of the data is preserved (see Al-Osh & Alzaid, 1987; McKenzie, 1988). For an overview of the literature in this area, refer Zeger & Qaqish (1988), MacDonald & Zucchini (1997), McKenzie (2003) and Kedem & Fokianos (2005). Different oservation-driven models can be found in the literature. The most popular such model is the GARCH models introduced by Engle (1982) and Bollerslev (1986). Its strength lies on the derivation of its closed form likelihood function which enables fast and simple estimation. However, in some other cases such as the existence of leverage effect, the model does not perform well (see Hansen & Lunde, 2005). Other models include the generalized autoregressive score (GAS) proposed by Creal et al. (2013). The maximum likelihood estimation is also straightforward but require different mechanism to update the parameters.

The serial correlation in the observation-driven model is based on the lagged values of the count variable. Feigin et al. (2008) classified this type of model as single source of error (SSOE). The initial values of latent parameter, conditional on lagged values of the counts are deterministic.

Conditioning on the lagged value of its stochastic and time-varying parameter λ_t, the count time series variable y_t, at time t, is assumed to have a Poisson distribution. The

model is defined as

$$y_t|\lambda_t \sim \text{Poisson}(\lambda_t), \tag{2.1}$$

$$\lambda_t = \lambda + \phi\lambda_{t-1} + \alpha(y_{t-1} - \lambda_{t-1}), \tag{2.2}$$

for $t = 2, 3, ..., T$ where the restrictions $\lambda > 0, \phi \geq \alpha \geq 0$ and $\phi \leq 1$ are imposed.

Poisson is the most common deviates in the studies on count time series. Nevertheless, analytical evidences indicates that in many situations, a count time series is better modelled with non-Poisson deviates (Davis & Wu, 2009; Zhu, 2011).

Let X_t be a time series of counts. We assume that X_t conditional on F_{t-1}, the random variables $X_1, ..., X_n$ are independent and the conditional distribution of X_t is specified by a negative binomial (NB) distribution.

$$X_t|F_{t-1} \sim \text{NB}(r, p_t), \tag{2.3}$$

where r is a positive number and p_t satisfies the model

$$\frac{1 - p_t}{p_t} = \lambda_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i} + \sum_{j=1}^{q} \beta_j \lambda_{t-j} \tag{2.4}$$

with $\alpha_0 > 0, \alpha \geq 0, \beta_j \geq 0, p \geq 1$ and $q \geq 0$.

Observation-driven models with deviates from the one-parameter exponential family, which includes Poisson distribution, binomial distribution (with known number of trials), and NB distribution (with known number of failures) was discussed in Davis & Liu (2012).

Many researchers employed a serial dependence of the observed counts through a state process, such models are known as GLM. The distribution of the observed count is specified conditioning on the state process, and the it is often takes the form of the

conditional expectation. However, if the state variable is determined by the history of the observed counts and states, then the model is characterized as observation-driven.

Even when prior observations are unobservable, the choice of initial values does not affect the approximation of the likelihood function. Furthermore, it derives a reliable consistency and asymptotic normality of the conditional maximum likelihood estimation (MLE).

2.1.2 Parameter-driven models

In a parameter-driven model, the dependence among observations is introduced indirectly through a latent process such as a hidden Markov chain as in Leroux & Puterman (1992), or a latent stationary autoregressive process as described in Zeger & Qaqish (1988), Chan & Ledolter (1995) and Yang et al. (2015). Oh & Lim (2001) applied a practical simulation-based method for Bayesian analysis of the parameter-driven model for time series Poisson data. One main feature of a parameter-driven model is that the distribution of the observed count data is independent, conditional on the latent process, although the observations are correlated marginally, (see Frühwirth-Schnatter & Wagner, 2006). A common parameter-driven model is the stochastic volatility model (SV) introduced by Taylor (2007). Several authors have shown that SV perform better than GARCH-type models (see Jacquier et al., 2002; Kim et al., 1998). However, the likelihod function is not in closed form and hence require computational techniques to estimate the parameters (see Kim et al., 1998; Koopman et al., 2016). Dunsmuir & He (2017) proposed a parameter-driven model for binomial time series where the logic of success probability is modelled as a linear function of observed explanatory variables and a stationary latent Gaussian process.The purpose is to test the serial dependence in the series. The asymptotic theory for the estimation of the parameters of the models using the generalized linear model estimating function is described in Dunsmuir & He (2017).

According to Feigin et al. (2008) the DSOE model is defined as (2.1), but with

$$\lambda_t = h(x_t), \tag{2.5}$$

$$x_t = a + \kappa x_{t-1} + \eta_t; \text{where } \eta_t \; iid \mathrm{N}(0, \sigma_\eta^2), \tag{2.6}$$

for $t = 1, 2, ..., T$, where $h(x_t)$ is any function that maps x_t into the positive space of λ_t and the stationary restriction $|\kappa| < 1$ is imposed.

Conditional on a stationary latent process ϵ_t, y_t is a sequence of independent counts with mean and variance given by

$$u_t = \mathrm{E}(y_t|\epsilon_t) = e^{x_t'\beta}, \tag{2.7}$$

$$w_t = \mathrm{var}(y_t|\epsilon_t) = u_t. \tag{2.8}$$

With $\mathrm{E}(\epsilon_t) = 1$ and $\mathrm{cov}(\epsilon_t, \epsilon_{t+\tau}) = \sigma^2\rho_\epsilon(\tau)$, the marginal moments of y_t are

$$\mu_t = \mathrm{E}(y_t) = e^{x_t'\beta}, \tag{2.9}$$

$$v_t = \mathrm{var}(y_t) = \mu_t + \sigma^2\mu_t^2, \tag{2.10}$$

$$\rho_y(t, \tau) = \mathrm{corr}(y_t, y_{t+\tau}) = \frac{\rho_\epsilon(\tau)}{\left[\{1 + (\sigma^2\mu_t)^{-1}\}\{1 + (\sigma^2\mu_{t+\tau})\right]^{\frac{1}{2}}}. \tag{2.11}$$

From equation (2.9), $\log \mu_t = c + x_t'\beta$. Note that, assuming a stationary autoregressive latent process does not specify any distributional assumption on the latent process.

On the other hand, Chan & Ledolter (1995) proposed a stationary Gaussian AR(1) latent process, $W_t = \rho W_{t-1} + \epsilon_t$, where ϵ_t is iid $\mathrm{N}(0, \sigma_\eta^2)$. Given W_t, the observation Y_t are

generated independently from a Poisson distribution with mean λ_t satisfying

$$\log \lambda_t = \alpha' U_t + W_t. \tag{2.12}$$

Hay et al. (2001) proposed the same model with W forms a stationary and invertible Gaussian ARMA(p, q) process which may be written in terms of multivariate normal distribution

$$W \sim \text{MVN}(\mathbf{0}, \Sigma) \tag{2.13}$$

where $\mathbf{0}$ is the $n \times 1$ mean vector of zeroes and Σ is the covariance matrix.

Frühwirth-Schnatter & Wagner (2006) modelled $y_t|\lambda_t \sim$ Poisson (λ_t), where λ_t depends on covariates $Z_t = (Z_t^{(1)}, Z_t^{(2)})$ through fixed model parameter α and time varying model parameters β_t;

$$\lambda_t \sim \exp\left[\left(Z_t^{(1)}\right)' \alpha + \left(Z_t^{(2)}\right)' \beta_t\right]. \tag{2.14}$$

Yang et al. (2013) introduced an autoregressive model for count time series with excess zeroes. Y_t is conditionally distributed as zero inflated Poisson with parameter λ_t and ω_t with probability mass function defined as follows:

$$f_{Y_t}(y_t|F_{t-1}; \boldsymbol{\theta}) = \omega_t I_{y_t=0} + (1 - \omega_t)\exp\left[\frac{(-\lambda_t)\lambda_t^{y_t}}{y_t!}\right]. \tag{2.15}$$

with $\boldsymbol{\theta}$ is the unknown parameters. The proposed zero inflated Poisson (ZIP) autoregressive

model in which the parameter λ_t and ω_t are modelled as follows

$$\log(\lambda_t) = \eta_t = x_{t-1}^{\mathsf{T}}\beta, \tag{2.16}$$

$$\log\left[\frac{\omega_t}{(1-\omega_t)}\right] = \xi = z_{t-1}^{\mathsf{T}}\gamma. \tag{2.17}$$

where $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)^{\mathsf{T}}$ and $\boldsymbol{\gamma} = (\gamma_1, ..., \gamma_q)^{\mathsf{T}}$ are parameters and $x_{t-1} = (x_{(t-1)1}, ..., x_{(t-1)p})^{\mathsf{T}}$, $z_{t-1} = (z_{(t-1)1}, ..., z_{(t-1)q})^{\mathsf{T}}$ denote vectors of past explanatory covariates.

In Yang et al. (2013), Y_t is conditionally distributed as zero inflated negative binomial with parameters k and ω defined as follows:

$$f_{Y_t}(y_t|F_{t-1};\boldsymbol{\theta}) = \omega I_{y_t=0} + (1-\omega)\frac{\Gamma(k+y_t)}{\Gamma(k)y_t!}p_t^k(1-p_t)^{y_t}. \tag{2.18}$$

where $p_t = \dfrac{k}{k+\lambda_t}$ is the probability of success in a NB distribution and λ_t is an intensity parameter that linked to the latent state z_t through a log-linear model,

$$\log\lambda_t = \log w_t + x_t^{\mathsf{T}}\beta + z_t. \tag{2.19}$$

Here, x_t is a set of explanatory variables, β is the vector regression coefficients and $\log w_t$ denotes an offset variable. Meanwhile, z_t is a stationary autoregressive process of order p, AR(p) such that

$$z_t = \phi_1 z_{t-1} + ... + \phi_p z_{t-p} + \epsilon_t \tag{2.20}$$

where ϵ_t is a Gaussian white noise process with mean 0 and variance σ^2, $\phi = (\phi_1, ..., \phi_p)^{\mathsf{T}}$ is a p-dimensional vector that consists of the autoregressive coefficients of z_t.

Recently, Tang & Cavanaugh (2018) discussed the state-space models for binomial time

series with excess zeroes.

The contrast between the observation-driven and the parameter-driven models for count data is comparable to the contrast between a generalized autoregressive conditionally heteroscedastic (GARCH) model and a stochastic volatility model for financial returns (Kim et al., 1998). The canonical model in this class for regularly spaced data is

$$y_t = \beta e^{h_t/2}\epsilon_t, \text{where } t \geq 1, \tag{2.21}$$

$$h_{t+1} = \mu + \phi(h_t - \mu) + \sigma_n \eta_t, \tag{2.22}$$

$$h_1 \sim \mathrm{N}\left(\mu, \frac{\sigma^2}{1-\phi^2}\right), \tag{2.23}$$

where y_t is the mean corrected at time t, h_t is the log volatility which is assumed to follow as a stationary process ($|\phi| < 1$) with h_1 drawn from the stationary distribution, ϵ_t and η_t are uncorrelated standard normal white noise shocks. The parameter β plays the role of the constant scaling factor and can be thought of as the modal instantaneous volatility, ϕ as the persistence in the volatility, and σ_n the volatility of the log-volatility. Model in (2.21) can be transformed into a linear model by taking the logarithm of the squares of observations

$$\log y_t^2 = h_t + \log \epsilon_t^2. \tag{2.24}$$

For the same data type, it imitates the differences between the ACD model (Engle & Russell, 1998) and the alternative stochastic conditional duration (SCD) (Bauwens & Veredas, 2004; Strickland et al., 2006; Thavaneswaran et al., 2015).

As other non-Gaussian settings, the credits of the observation-driven and parameter-driven approaches to model counts remain an open empirical question. With an additional source of random error, the latter approach may yield more flexibility than the former

approach. On the other hand, the estimation of the parameter-driven models require the use of some form of computationally intensive simulation methodology (Chan & Ledolter, 1995; Durbin & Koopman, 2012; Frühwirth-Schnatter & Wagner, 2006; Jung et al., 2006).

The response distribution in count time series is often considered as non-Gaussian. Since Kalman filter and Kalman smoother is a conventional method for parameter estimation, it is unreliable to applied on a state space setting. Therefore, we resort to Monte Carlo methods, and construct a Monte Carlo expectation-maximization (MCEM) algorithm based on particle filter (Gordon et al., 1993) and particle smoother (Godsill et al., 2004). A similar MCEM algorithm has been proposed by Kim & Stoffer (2008) to fit SV models.

2.2 Estimation

2.2.1 Maximum Likelihood Estimation (MLE)

MLE is a method for estimating the parameters of a statistical model given observations, by finding the parameter values that maximize the likelihood function. To solve the nonlinear equations which result from differentiating the likelihood function, we applied MLE techniques with the aid of scoring (Gupta & Mehra, 1974).

Several examples on the feasibility of these methods for several specific cases were testified by Ledolter (1979) and Goodrich & Caines (1979). Harvey & Phillips (1979) and Jones (1980) studied the MLE of the parameters in an autoregressive moving average (ARMA) process.

2.2.2 EM Algorithm

A number of unappealing characteristics obtained from the likelihood methods applied in the above references can be avoided using the Expectation Maximization (EM) algorithm.

Due to the fact that the response distribution is Non-Gaussian, The marginal likelihood of the observed data $y_{1:n} = (y_1, ..., y_n)^\intercal$ cannot be expressed empirically. Hence, it is extremely

challenging to apply direct maximization of the marginal likelihood. Despite using gradient-based methods, we resort to EM algorithm, a popular approach for calculating MLE on models involved in missing data and/or unobservable latent variables.

Dempster et al. (1977) was the first to introduce EM algorithm, and applied the algorithm to several examples including missing value situations, applications to grouped, censored or truncated data, finite mixture models, variance component estimation, hyperparameter estimations, iteratively reweighted least squares and factor analysis.

The corrections in the subsequent iterations commonly involve calculating the inverse of the matrix of second order partial. However, with a substantial number of parameters, it can be rather large. On the other hand, the EM steps constantly increase the likelihood and it is guaranteed to converge to a stationary point for an exponential family (Wu, 1983).

Shumway & Stoffer (1982) introduced an EM algorithm to estimate parameters based on the Kalman filtering and smoothing techniques for data with missing values. A comprehensive discussion of the Kalman methods can be found in the Shumway & Stoffer (1982).

In the E-step of the algorithm, the conditional expectation of $\log L(\boldsymbol{\theta})$ given the observed data $y_{1:n}$, is given by

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(j)}) = E\left\{\log L(\boldsymbol{\theta})|y_{1:n}, \boldsymbol{\theta}^{(j)}\right\}.$$

Due to the orthogonal decomposition of the complete data log-likelihood, the M-step of the algorithm is a clear-cut. In the M-step, partial derivatives are applied to maximize $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(j)})$.

The likelihood would be immensely simplified if the latent process is observable. Hence, EM algorithm can be used to maximize the likelihood function. However, the E step is normally difficult since the conditional distribution of the latent process given the count data is complicated. It is impossible to compute the E-step manually. Therefore, we resort

to Markov Chain sampling techniques to implement the E step.

2.2.3 MCEM Algorithm

The modified scheme was called MCEM algorithm, proposed by Wei & Tanner (1990). The EM and MCEM algorithms provide a minimal amount of information to the data analyst.

The MCEM algorithm in a complex genetic model situation was used by Guo & Thompson (1991). For some early surveys on Markov Chain sampling methods, see Besag & Green (1993), Smith & Roberts (1993) and Tierney (1994).

One significant feature of the EM algorithm is that the likelihood of the observed data always increases along an EM sequence. For other convergence properties of EM algorithm, see Dempster et al. (1977) and Wu (1983). Under suitable regularity conditions, an MCEM sequence will converge to the maximizer of the likelihood of the observed data.

2.2.4 Particle Methods

If the data are modelled by a linear Gaussian state-space model, it is possible to derive an exact analytical expression to compute the evolving sequence of posterior distributions. This recursion is the well known Kalman Filter. If the data are modelled as a partially observed finite state-space Markov Chain, it is possible to obtain analytical solution, also known as hidden Markov model (HMM) filter (Harrison & West, 1999; Vidoni, 1999).

However, real data can be very complex, typically involving elements of non-Gaussian, high dimension and non-linear. The problem appears including Bayesian filtering, optimal (linear) filtering, stochastic filtering and on-line inference and learning. Many approximation methods, such as the extended Kalman Filter, Gaussian sum approximations and grid-based filters have been proposed to overcome the problem. These methods however are either fail to take into account all statistical features of the process or are too

difficult to implement computationally (see Doucet et al., 2001).

Sequential Monte Carlo (SMC) methods are a set of simulation-based methods which provide a convenient and attractive approach to computing the posterior distribution. Recently, there has been tremendous of scientific papers on SMC methods and their applications. Several closely related algorithms are bootstrap filters, condensation filters, Monte Carlo filters, interacting particle approximations, survival of the fittest and the one we are using is the particle filters.

Since their introduction in 1993 by Gordon et al. (1993), particle filters have become a very popular class of numerical methods for the solution of optimal estimation in non-linear and non-Gaussian models. In comparison with standard approximation methods, such as extended Kalman Filter, the principle advantage of particle methods is that they do not rely on any local linearisation technique or any crude functional approximation (Doucet & Johansen, 2009). Johansen (2009) shows that any particle filter can be implemented using a computational framework.

Particle methods for filtering and smoothing have become the most common example of SMC algorithms. The key concept of particle methods is to approximate the conditional density of the latent states given the observed data using sequential importance sampling (SIS) and resampling. SIS is the SMC method that forms the basis of the particle methods. The general concept of particle filtering and smoothing for state-space models can be found in Kim & Stoffer (2008).

CHAPTER 3: PARAMETER-DRIVEN COUNT TIME SERIES MODELS

In this chapter, we focus on the parameter-driven models or dynamic models for count time series. We explain briefly the autoregressive (AR) model with it stationary property and the autocorrelation function (ACF). We first review the state-space model (Kalman & Bucy, 1961) for normally distributed data. In time series literature, such a model is often called the dynamic linear model (DLM). After reviewing DLM, we proposed a class of dynamic models for count time series with three different distributions, that is Poisson, zero-inflated Poisson and negative binomial.

3.1 Autoregressive (AR) model

In the AR model, the current value of the variable is defined as a function of its previous values plus an error term. In other words, the dependent variable, y_t, is taken as the function of the time lagged-values of itself such that

$$y_t = \alpha + \kappa_1 y_{t-1} + \kappa_2 y_{t-2} + \ldots + \kappa_n y_{t-n} + \eta_t, \tag{3.1}$$

where α and κ_i $(i = 1, 2, ..., n)$ are parameters to be estimated and η_t is the error term assumed iid with mean 0 and variance, σ_η^2.

Model (3.1) is the general form of the n^{th} order AR process where n denotes the number of lagged terms of the dependent variable.

We now introduce a backward-shift operator B, which shifts time one step back, such that $By_t = y_{t-1}$; or in general $B^k y_t = y_{t-k}$ for $k = 1, 2, ..., n$. Using this notation, we can rewrite Model (3.1) as

$$\left(1 - \kappa_1 B - \kappa_2 B^2 - \ldots - \kappa_n B^n\right) y_t = \alpha + \eta_t.$$

Note that the application of the backward-shift operator on a constant (which is the same for all t) results in the constant itself ($B^n\alpha = \alpha$) see Abraham & Ledolter (2009).

3.1.1 Stationary process

For stationarity, the probability distribution at any times $t_1, t_2, ..., t_n$ must be the same as the probability distribution at times $t_1 + k, t_2 + k, ..., t_n + k$, where k is an arbitrary shift along the time axis. In other words, the marginal distribution does not depend on time, which in turn implies $E(y_t) = E(y_{t-1}) = E(y_{t-2}) = ... = E(y_{t-n}) = \mu_y$ and $\sigma^2(y_t) = \sigma^2(y_{t-1}) = \sigma^2(y_{t-2}) = ... = \sigma^2(y_{t-n}) = \sigma_y^2$.

In Model (3.1), for $n = 1$, the stationary condition for the AR(1) model is $|\kappa_1| < 1$. The mean and variance of the model are shown in Equation (3.2) and (3.3) respectively.

$$\begin{aligned} \mu_y &= E(y_t) \\ &= E(\alpha + \kappa y_{t-1}) \\ \mu_y &= \frac{\alpha}{1-\kappa} \end{aligned} \tag{3.2}$$

$$\begin{aligned} \sigma_y^2 &= \text{Var}(y_t) \\ &= \text{Var}(\alpha + \kappa y_{t-1} + \eta_t) \\ \sigma_y^2 &= \kappa^2\sigma_y^2 + \sigma_\eta^2 \\ &= \frac{\sigma_\eta^2}{1-\kappa^2} \end{aligned} \tag{3.3}$$

3.1.2 Autocorrelation function (ACF)

The stationary condition implies that the mean and the variance of the process are constant and that the autocovariance (ACV);

$$\gamma_k = \text{Cov}(y_t, y_{t-k}) = E(y_t - \mu_y)(y_{t-k} - \mu_y), \tag{3.4}$$

and the autocorrelations

$$\rho_k = \frac{\text{Cov}(y_t, y_{t-k})}{\sqrt{\sigma^2(y_t)\sigma^2(y_{t-k})}} = \frac{\text{Cov}(y_t, y_{t-k})}{\sigma^2(y_t)}. \tag{3.5}$$

The ACF plays a major role in modelling the dependencies among observations since it characterizes, together with the process mean $E(y_t)$ and $\text{Var}(y_t)$, the stationary stochastic process.

3.2 State-space model

The state space model or dynamic linear model (DLM), in its basic form, employs an order one autoregressive, AR(1), as the state equation,

$$x_t = \alpha + \kappa x_{t-1} + \eta_t, \tag{3.6}$$

for $t = 1, 2, ..., n$ with a Gaussian white noise process, $\eta_t \sim iid\ \text{N}(0, \sigma_\eta^2)$ and the stationary restriction $|\kappa| < 1$. Equation (3.6) can be written iteratively as

$$\begin{aligned}
x_t &= \alpha + \kappa(\alpha + \kappa x_{t-2} + \eta_{t-1}) + \eta_t \\
&= \alpha + \kappa\alpha + \kappa^2 x_{t-2} + \kappa\eta_{t-1} + \eta_t \\
&= \alpha + \kappa\alpha + \kappa^2(\alpha + \kappa x_{t-3} + \eta_{t-2}) + \kappa\eta_{t-1} + \eta_t \\
&= \sum_{j=0}^{j=i-1} (\kappa^j \alpha) + \kappa^i x_{t-i} + \sum_{j=0}^{j=i-1} (\kappa^j \eta_{t-j}) \\
&= \frac{\alpha(1-\kappa^i)}{1-\kappa} + \kappa^i x_{t-i} + \sum_{j=0}^{j=i-1} (\kappa^j \eta_{t-j}).
\end{aligned} \tag{3.7}$$

In the DLM, we assume the initial state x_0 is normally distributed with mean μ_0 and variance σ_0^2. By the stationary assumption, we have

$$x_t|x_{t-1} \sim \mathrm{N}(\alpha + \kappa x_{t-1}, \sigma_\eta^2). \tag{3.8}$$

for $t = 1, 2, ..., n$. The observation y_t, conditioning on the current state x_t, is assumed to be independently distributed as illustrated in Figure 3.1.

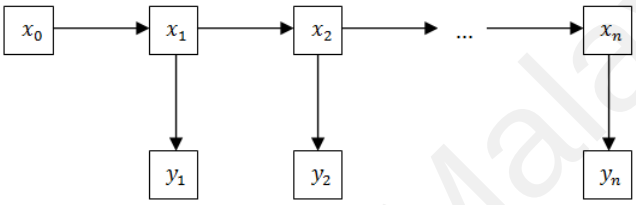


Figure 3.1: Graphical illustration of the state evolution and data generation in the dynamic linear model.

We denote that x_t is normally distributed such that $x_t \sim \mathrm{N}(\mu_x, \sigma_x^2)$ with $\mu_x = \dfrac{\alpha}{1-\kappa}$ and $\sigma_x^2 = \dfrac{\sigma_\eta^2}{1-\kappa^2}$. Hence, the moment generating function (MGF) of x_t is $M_x(r) = E[e^{x_t r}] = e^{r\left(\mu_x + \frac{1}{2}\sigma_x^2\right)}$.

3.3 Dynamic Poisson model

Let $y_1, y_2, ..., y_n$ be a sequence of count data, observed at discrete, evenly spaced time points. Given x_t, the model are assumed to be independently distributed as Poisson random variables with parameter λ_t, written as:

$$Y_t|x_t \sim \mathrm{Poisson}(\lambda_t),$$

with $\lambda_t = e^{x_t}$ where x_t is defined as Equation (3.6). In other words, the conditional probability function of Y_t can be written as

$$P(Y_t = y|x_t) = \frac{e^{-\lambda_t}\lambda_t^y}{y!}. \tag{3.9}$$

for $t = 1, 2, ..., n$. The dynamic Poisson model can be written in the following state-space form:

$$x_t|x_{t-1} \sim \mathrm{N}(\alpha + \kappa x_{t-1}, \sigma_\eta^2), \tag{3.10}$$

$$Y_t|x_t \sim \mathrm{Poisson}(\lambda_t), \tag{3.11}$$

for $t = 1, 2, ..., n$ and $\lambda_t = e^{x_t}$. The initial state of x_0 is assumed to be normally distributed with mean μ_0 and variance σ_0^2. The conditional moment generating function (CMGF) of Poisson is

$$E[e^{Y_t r}|x_t] = e^{\lambda_t(e^r - 1)}. \tag{3.12}$$

Since Y_t is condition on x_t with mean λ_t, hence

$$E(Y_t|x_t) = \lambda_t,$$

$$\mathrm{Var}(Y_t|x_t) = \lambda_t,$$

$$\gamma(Y_t|x_t) = \frac{E(Y_t - \lambda_t)^3}{[\mathrm{Var}(Y_t)]^{3/2}} = \frac{1}{\sqrt{\lambda_t}},$$

$$K(Y_t|x_t) = \frac{E(Y_t - \lambda_t)^4}{[\mathrm{Var}(Y_t)]^2} = 3 + \frac{1}{\lambda_t},$$

where γ and K denote skewness and kurtosis respectively. Given the CMGF of the model in Equation (3.12), hence the MGF of Y, $M_Y(r)$ is

$$M_Y(r) = E\left[E\left(e^{Y_t r}|x_t\right)\right] = E[e^{\lambda_t(e^r-1)}] = M_{\lambda_t}(e^r - 1), \tag{3.13}$$

where M_{λ_t} is the MGF of λ_t. Since $\lambda_t = e^{x_t}$ and $x_t \sim iid$ N $(0, \sigma_\eta^2)$, hence the distribution of λ_t is log-Normal with parameter $(0, \sigma_\eta^2)$. The first four moments of λ_t are given below.

$$\begin{aligned}
E(\lambda_t) &= E(e^{x_t}) = e^{\left(\mu_x + \frac{1}{2}\sigma_x^2\right)} = \zeta, \\
E(\lambda_t^2) &= E(e^{2x_t}) = e^{\left(2\mu_x + 2\sigma_x^2\right)} = \zeta^2 e^{\sigma_x^2}, \\
E(\lambda_t^3) &= E(e^{3x_t}) = e^{\left(3\mu_x + \frac{9}{2}\sigma_x^2\right)} = \zeta^3 e^{3\sigma_x^2}, \\
E(\lambda_t^4) &= E(e^{4x_t}) = e^{\left(4\mu_x + 8\sigma_x^2\right)} = \zeta^4 e^{6\sigma_x^2}.
\end{aligned}$$

Given Equation (3.13), the first four raw moments of Y_t are

$$\begin{aligned}
E(Y_t) &= M_Y'(r)|_{r=0} = M_{\lambda_t}'(e^r - 1)e^r|_{r=0} \\
&= E(\lambda_t) = E(e^{x_t}) = e^{\mu_x + \frac{1}{2}\sigma_x^2} \\
&= \zeta, \\
E(Y_t^2) &= \mathrm{M}_Y''(r)|_{r=0} = \mathrm{M}_{\lambda_t}''(e^r - 1)e^{2r}|_{r=0} + \mathrm{M}_{\lambda_t}'(e^r - 1)e^r|_{r=0} = \mathrm{M}_{\lambda_t}''(0)e^0 + \mathrm{M}_{\lambda_t}'(0)e^0 \\
&= E(\lambda_t^2) + E(\lambda_t) = E(e^{2x_t}) + E(e^{x_t}) = e^{2\mu_x + 2\sigma_x^2} + e^{\mu_x + \frac{1}{2}\sigma_x^2} \\
&= \zeta^2 e^{\sigma_x^2} + \zeta,
\end{aligned}$$

$$E(Y_t^3) = \mathrm{M}_Y'''(r)|_{r=0} = \mathrm{M}_{\lambda_t}'''(e^r-1)e^{3r}|_{r=0} + 3\mathrm{M}_{\lambda_t}''(e^r-1)e^{2r}|_{r=0} + \mathrm{M}_{\lambda_t}'(e^r-1)e^r|_{r=0}$$

$$= \mathrm{M}_{\lambda_t}'''(0)e^0 + 3\mathrm{M}_{\lambda_t}''(0)e^0 + \mathrm{M}_{\lambda_t}'(0)e^0$$

$$= E(\lambda_t^3) + 3E(\lambda_t^2) + E(\lambda_t) = E(e^{3x_t}) + 3E(e^{2x_t}) + E(e^{x_t})$$

$$= e^{3\mu_x+\frac{9}{2}\sigma_x^2} + 3e^{2\mu_x+2\sigma_x^2} + e^{\mu_x+\frac{1}{2}\sigma_x^2}$$

$$= \zeta^3 e^{3\sigma_x^2} + 3\zeta^2 e^{\sigma_x^2} + \zeta,$$

$$E(Y_t^4) = \mathrm{M}_Y''''(r)|_{r=0}$$

$$= \mathrm{M}_{\lambda_t}''''(e^r-1)e^{4r}|_{r=0} + 6\mathrm{M}_{\lambda_t}'''(e^r-1)e^{3r}|_{r=0} + 7\mathrm{M}_{\lambda_t}''(e^r-1)e^{2r}|_{r=0} + \mathrm{M}_{\lambda_t}'(e^r-1)e^r|_{r=0}$$

$$= \mathrm{M}_{\lambda_t}''''(0)e^0 + 6\mathrm{M}_{\lambda_t}'''(0)e^0 + 7\mathrm{M}_{\lambda_t}''(0)e^0 + \mathrm{M}_{\lambda_t}'(0)e^0$$

$$= E(\lambda_t^4) + 6E(\lambda_t^3) + 7E(\lambda_t^2) + E(\lambda_t)$$

$$= E(e^{4x_t}) + 6E(e^{3x_t}) + 7E(e^{2x_t}) + E(e^{x_t})$$

$$= e^{4\mu_x+8\sigma_x^2} + 6e^{3\mu_x+\frac{9}{2}\sigma_x^2} + 7e^{2\mu_x+2\sigma_x^2} + e^{\mu_x+\frac{1}{2}\sigma_x^2}$$

$$= \zeta^4 e^{6\sigma_x^2} + 6\zeta^3 e^{3\sigma_x^2} + 7\zeta^2 e^{\sigma_x^2} + \zeta.$$

From the derivation of the raw moments above, we can derive the first four central moments of Y_t;

- Mean

$$E(Y_t) = M_Y'(r)|_{r=0} = M_{\lambda_t}'(e^r-1)e^r|_{r=0}$$

$$= E(\lambda_t) = E(e^{x_t})$$

$$= e^{\mu_x+\frac{1}{2}\sigma_x^2} = \zeta. \tag{3.14}$$

- Variance

$$\begin{aligned}\text{Var}(Y_t) &= E(Y_t^2) - [E(Y_t)]^2 \\ &= \text{M}_Y''(0) - [\text{M}_Y'(0)]^2 \\ &= e^{2\mu_x + 2\sigma_x^2} + e^{\mu_x + \frac{1}{2}\sigma_x^2} - e^{2\mu_x + \sigma_x^2} \\ &= \zeta^2 \left(e^{\sigma_x^2} - 1\right) + \zeta. \end{aligned} \tag{3.15}$$

- Skewness

$$\begin{aligned}\gamma(Y_t) &= \frac{E[Y_t - E(Y_t)]^3}{[\text{Var}(Y_t)]^{3/2}} \\ &= \frac{E[Y_t^3] - 3E(Y_t)\sigma^2(Y_t) - [E(Y_t)]^3}{[\text{Var}(Y_t)]^{3/2}} \\ &= \frac{\text{M}_Y'''(0) - 3\text{M}_Y'(0)\text{Var}(Y_t) - [\text{M}_Y'(0)]^3}{[\text{Var}(Y_t)]^{3/2}} \\ &= \frac{\zeta^3 \left(e^{3\sigma_x^2} - 3e^{\sigma_x^2} + 2\right) + 3\zeta^2 \left(e^{\sigma_x^2} - 1\right) + \zeta}{\left[\zeta^2 \left(e^{\sigma_x^2} - 1\right) + \zeta\right]^{3/2}}. \end{aligned} \tag{3.16}$$

- Kurtosis

$$\begin{aligned}K(Y_t) &= \frac{E[Y_t - E(Y_t)]^4}{[\text{Var}(Y_t)]^2} \\ &= \frac{E(Y_t^4) - 4E(Y_t)E(Y_t^3) + 6E(Y_t^2)[E(Y_t)]^2 - 4E(Y_t)[E(Y_t)]^3 + [E(Y_t)]^4}{[\text{Var}(Y_t)]^2} \\ &= \frac{\text{M}_Y''''(0) - 4\text{M}_Y'(0)\text{M}_Y'''(0) + 6\text{M}_Y''(0)[\text{M}_Y'(0)]^2 - 4\text{M}_Y'(0)[\text{M}_Y'(0)]^3 + [\text{M}_Y'(0)]^4}{[\text{Var}(Y_t)]^2} \\ &= \frac{\zeta^4 \left(e^{6\sigma_x^2} - 4e^{3\sigma_x^2} + 6e^{\sigma_x^2} - 3\right) + \zeta^3 \left(6e^{3\sigma_x^2} - 12e^{\sigma_x^2} + 6\right) + \zeta^2 \left(7e^{\sigma_x^2} - 4\right) + \zeta}{\left[\zeta^2 \left(e^{\sigma_x^2} - 1\right) + \zeta\right]^2}. \end{aligned} \tag{3.17}$$

The autocovariance function is defined as Equation (3.4), and we have

$$\begin{aligned} E(Y_t Y_{t-i}) &= E\{E[(Y_t Y_{t-i})|\lambda_t, \lambda_{t-i}]\} \\ &= E[E[(Y_t|\lambda_t)E(Y_{t-i}|\lambda_{t-i})] \\ &= E(\lambda_t \lambda_{t-i}) \\ &= E(e^{x_t} e^{x_{t-i}}). \end{aligned} \tag{3.18}$$

We know that

$$\mu_x = \frac{\alpha}{1-\kappa} \quad \text{and} \quad \sigma_x^2 = \frac{\sigma_\eta^2}{1-\kappa^2},$$

therefore,

$$\alpha = \mu_x(1-\kappa) \quad \text{and} \quad \sigma_\eta^2 = \sigma_x^2(1-\kappa^2).$$

By substituting α, σ_η^2 and x_t from Equation (3.7) into Equation (3.18), we have

$$\begin{aligned} E(Y_t Y_{t-i}) &= E(e^{x_t} e^{x_{t-i}}) \\ &= E(e^{\frac{\alpha(1-\kappa^i)}{1-\kappa} + \kappa^i x_{t-i} + \sum_{j=0}^{j=i-1}(\kappa^j \eta_{t-j})} e^{x_{t-i}}) \\ &= E\left\{ exp\left[\frac{\alpha(1-\kappa^i)}{1-\kappa} + (\kappa^i + 1)x_{t-i} + \sum_{j=0}^{j=i-1}(\kappa^j \eta_{t-j}) \right] \right\} \\ &= e^{\mu_x(1-\kappa^i)} e^{\mu_x(\kappa^i+1) + \frac{1}{2}\sigma_x^2(\kappa^i+1)^2} e^{\frac{1}{2}(1-\kappa^{2i})\sigma_x^2} \\ &= exp\left[\mu_x(1-\kappa^i) + \mu_x(\kappa^i+1) + \frac{1}{2}\sigma_x^2(\kappa^i+1)^2 + \frac{1}{2}(1-\kappa^{2i})\sigma_x^2 \right] \\ &= exp\left[2\mu_x + (1+\kappa^i)\sigma_x^2 \right]. \end{aligned}$$

Therefore, the autocovariance and ACF of the dynamic Poisson model are;

$$\begin{aligned}
\text{Cov}(Y_t, Y_{t-i}) &= E(Y_t Y_{t-i}) - E(Y_t)E(Y_{t-i}) \\
&= e^{2\mu_x + (1+\kappa^i)\sigma_x^2} - \left[e^{\mu_x + \frac{1}{2}\sigma_x^2}\right]^2 \\
&= e^{2\mu_x + (1+\kappa^i)\sigma_x^2} - \left[e^{2\mu_x + \sigma_x^2}\right] \\
&= \zeta^2 \left[e^{\kappa^i \sigma_x^2} - 1\right]. \qquad (3.19)
\end{aligned}$$

$$\begin{aligned}
\rho_i &= \frac{\text{Cov}(Y_t, Y_{t-i})}{\sqrt{\sigma^2(Y_t)\sigma^2(Y_{t-i})}} \\
&= \frac{\zeta^2 \left(e^{\kappa^i \sigma_x^2} - 1\right)}{\zeta^2 \left(e^{\sigma_x^2} - 1\right) + \zeta} \\
&= \frac{(e^{\kappa^i \sigma_x^2} - 1)}{(e^{\sigma_x^2} - 1) + \frac{1}{\zeta}}. \qquad (3.20)
\end{aligned}$$

3.3.1 Simulation study

Larger samples are often employed in quantitative research. A basic rule is that increasing the sample size increases its reliability. We generate a data set of length 100,000 and calculate the standard central moment and its autocovariance and autocorrelation.

We first generate the latent variables x_t according to the Equation (3.10) for $t = 1, 2, \ldots, 100,000$. For a given value of x_t, we generate the observed variable y_t according to the Equation (3.11), where $\lambda_t = e^{(x_t)}$. Two sets of simulation studies are featured in this section with $\sigma_\eta = 0.01$ and 0.2. Since we did not have the exact true parameter vector, we applied several value for each parameter. That is $\alpha = -0.2, 0, 0.2$ and $\kappa = -0.5, 0, 0.5$.

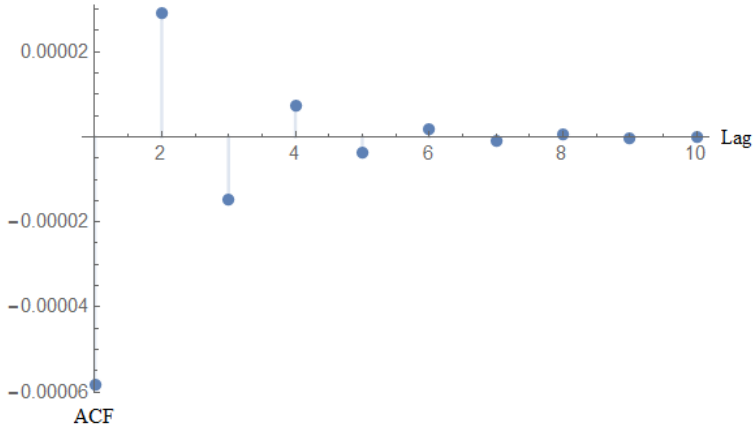
Tables 3.1 and 3.2 shows the generated and true values of the moments structures for dynamic Poisson model together with the 1st order ACF and ACV. The MSE values and biases are small.

Table 3.1: Generated data and true values for the moment structures with $\sigma_\eta = 0.01$.

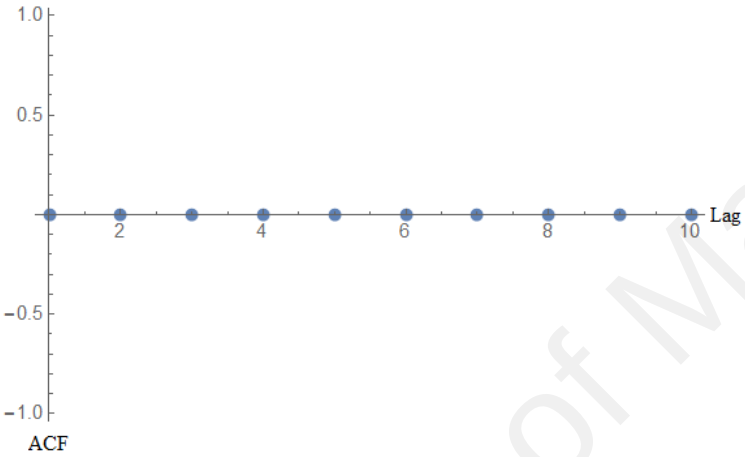
		$\alpha = -0.2$			$\alpha = 0$			$\alpha = 0.2$		
		Generated	True	MSE	Generated	True	MSE	Generated	True	MSE
$\kappa = -0.5$	μ_{Y_t}	0.8754	0.8752	9.38E-06	1.0001	1.0001	7.78E-06	1.1426	1.1427	9.56E-06
	$\sigma^2_{Y_t}$	0.8760	0.8753	2.39E-05	1.0003	1.0002	2.82E-05	1.1418	1.1429	2.54E-05
	γ_{Y_t}	1.0710	1.0691	1.03E-04	1.0011	1.0002	1.05E-04	0.9350	0.9357	9.34E-05
	K_{Y_t}	4.1534	4.1432	3.42E-03	4.0068	4.0006	3.06E-03	3.8770	3.8759	2.45E-03
	ACV(1)	0.0035	-0.0001	1.26E-05	0.0009	-0.0001	1.04E-06	0.0023	-0.0001	5.98E-06
	ACF(1)	0.0040	-0.0001	1.65E-05	0.0009	-0.0001	1.03E-06	0.0020	-0.0001	4.58E-06
$\kappa = 0$	μ_{Y_t}	0.8187	0.8188	9.69E-06	1.0000	1.0001	1.46E-05	1.2213	1.2215	1.14E-05
	$\sigma^2_{Y_t}$	0.8180	0.8188	1.87E-05	1.0004	1.0002	3.07E-05	1.2216	1.2216	3.70E-05
	γ_{Y_t}	1.1026	1.1053	8.56E-05	1.0016	1.0001	1.08E-04	0.9041	0.9050	8.26E-05
	K_{Y_t}	4.2106	4.2218	2.62E-03	4.0105	4.0005	2.76E-03	3.8161	3.8192	2.24E-03
	ACV(1)	0.0005	0.0008	7.10E-08	0.0034	-0.0017	2.57E-05	0.0039	-0.0030	4.76E-05
	ACF(1)	0.0006	0.0009	1.17E-07	0.0034	-0.0017	2.57E-05	0.0032	-0.0025	3.22E-05
$\kappa = 0.5$	μ_{Y_t}	0.6704	0.6704	7.71E-06	1.0002	1.0001	9.56E-06	1.4919	1.4919	1.03E-05
	$\sigma^2_{Y_t}$	0.6709	0.6704	1.74E-05	1.0007	1.0002	2.37E-05	1.4919	1.4922	5.22E-05
	γ_{Y_t}	1.2231	1.2215	1.56E-04	0.9990	1.0002	9.15E-05	0.8196	0.8189	9.37E-05
	K_{Y_t}	4.4992	4.4924	5.14E-03	3.9958	4.0006	2.69E-03	3.6753	3.6709	1.97E-03
	ACV(1)	-0.0009	8.06E-05	9.62E-07	-0.0040	0.0001	1.67E-05	0.0044	0.0001	1.84E-05
	ACF(1)	-0.0014	0.0001	2.31E-06	-0.0040	0.0001	1.69E-05	0.0029	0.0001	8.03E-06

Table 3.2: Generated data and true values for the moment structures with $\sigma_\eta = 0.2$.

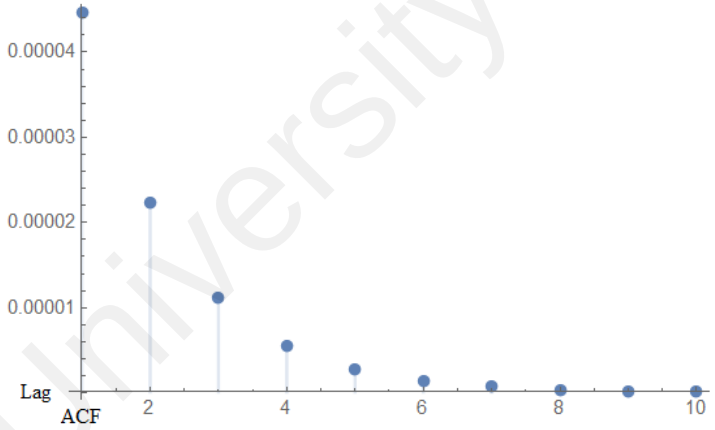
		$\alpha = -0.2$			$\alpha = 0$			$\alpha = 0.2$		
		Generated	True	MSE	Generated	True	MSE	Generated	True	MSE
$\kappa = -0.5$	μ_{Y_t}	0.8988	0.8988	8.73E-06	1.0280	1.0270	1.14E-05	1.1734	1.1753	1.62E-05
	$\sigma^2_{Y_t}$	0.9435	0.9431	3.40E-05	1.0859	1.0848	4.74E-05	1.1734	1.1753	1.62E-05
	γ_{Y_t}	1.1350	1.1337	1.68E-04	1.0660	1.0712	9.48E-05	1.0130	1.0135	1.23E-04
	K_{Y_t}	4.4109	4.4059	5.84E-03	4.2614	4.2697	3.21E-03	4.1469	4.1512	3.75E-03
	ACV(1)	-0.0200	-0.0313	1.70E-04	-0.0282	-0.0278	4.85E-08	-0.0331	-0.0362	9.76E-06
	ACF(1)	-0.0211	-0.0225	9.86E-06	-0.0259	-0.0256	6.73E-08	-0.0265	-0.0290	6.03E-06
$\kappa = 0$	μ_{Y_t}	0.8355	0.8353	8.30E-06	1.0201	1.0202	9.18E-06	1.2454	1.2461	1.47E-05
	$\sigma^2_{Y_t}$	0.8646	0.8637	2.68E-05	1.0620	1.0627	3.55E-05	1.3089	1.3094	6.06E-05
	γ_{Y_t}	1.1498	1.1506	1.18E-04	1.0526	1.0525	1.12E-04	0.9659	0.9650	1.29E-04
	K_{Y_t}	4.3986	4.4112	4.29E-03	4.200	4.1964	3.26E-03	4.0274	4.0213	3.62E-03
	ACV(1)	-0.0004	-0.0049	2.05E-05	-0.0022	-0.0060	1.42E-05	0.0025	-0.0022	2.19E-05
	ACF(1)	-0.0005	-0.0057	2.68E-05	-0.0020	-0.0057	1.34E-05	0.0019	-0.0017	1.27E-05
$\kappa = 0.5$	μ_{Y_t}	0.6884	0.6884	7.12E-06	1.0271	1.0270	1.16E-05	1.5324	1.5532	4.51E-04
	$\sigma^2_{Y_t}$	0.7145	0.7194	4.58E-05	1.0855	1.0848	4.30E-05	1.6603	1.6607	8.43E-05
	γ_{Y_t}	1.2748	1.2741	1.44E-04	1.0718	1.0712	1.21E-04	0.9108	0.9116	8.12E-05
	K_{Y_t}	4.7462	4.7414	5.47E-03	4.2711	4.2697	4.32E-03	3.9574	3.9592	2.55E-03
	ACV(1)	0.0130	0.0128	1.42E-06	0.0245	0.0285	1.60E-05	0.0585	0.0634	2.40E-05
	ACF(1)	0.0195	0.0179	2.44E-06	0.0225	0.0263	1.42E-05	0.0354	0.0382	7.58E-06



(a)

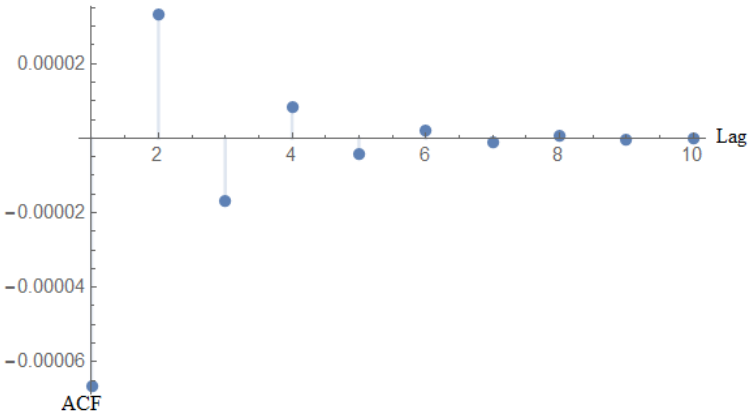


(b)

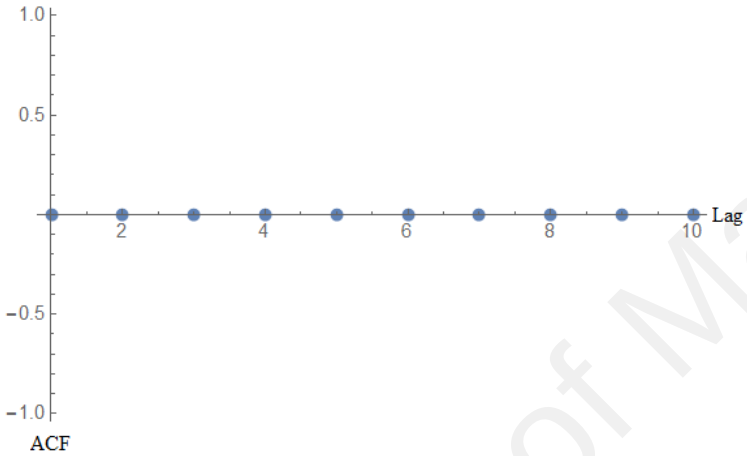


(c)

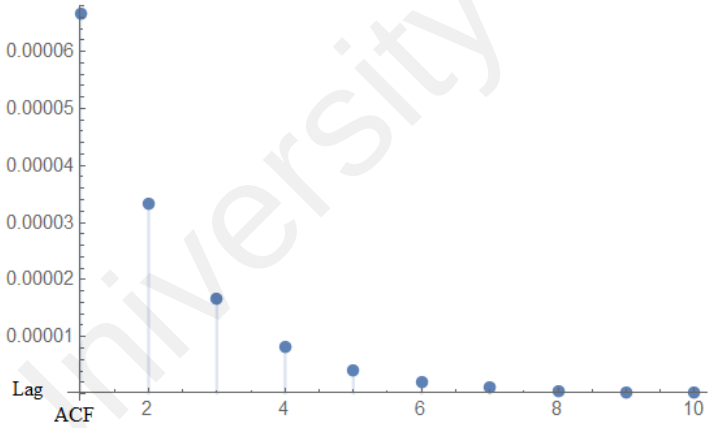
Figure 3.2: ACF plots for the Poisson model with (a) $\alpha = -0.2$, $\kappa = -0.5$ and $\sigma_\eta = 0.01$, (b) $\alpha = -0.2$, $\kappa = 0$ and $\sigma_\eta = 0.01$ and (c) $\alpha = -0.2$, $\kappa = 0.5$ and $\sigma_\eta = 0.01$.



(a)

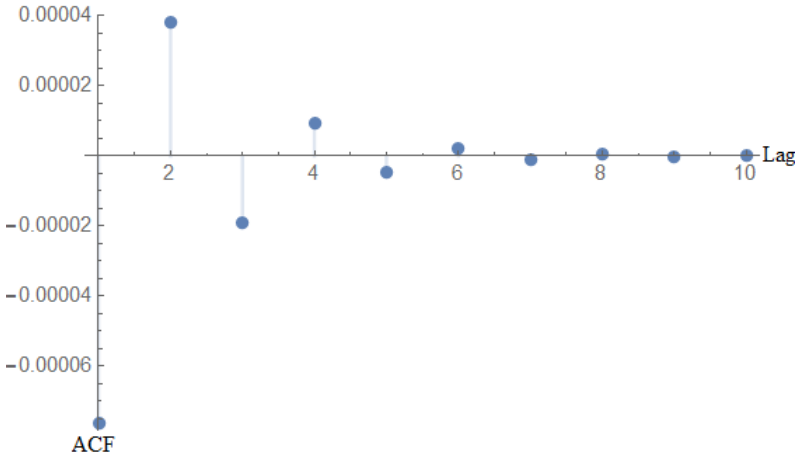


(b)

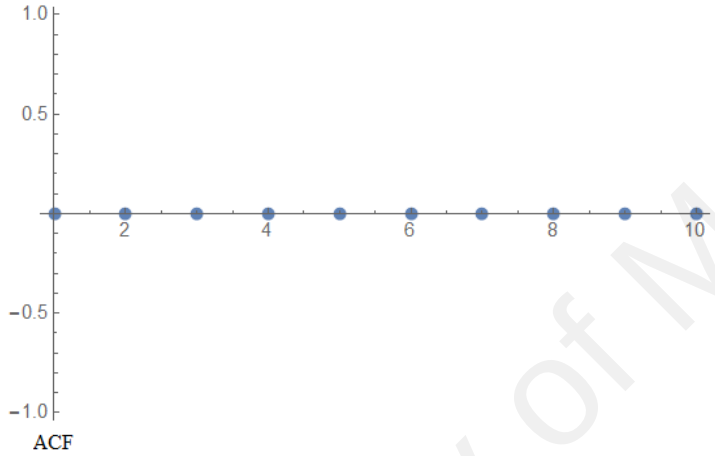


(c)

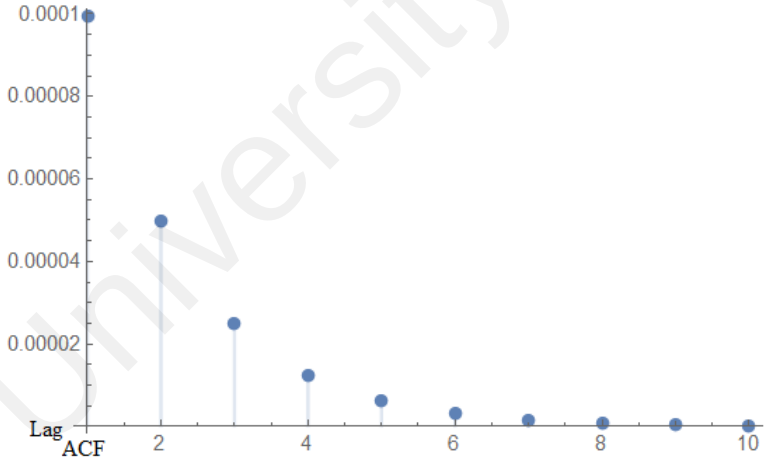
Figure 3.3: ACF plots for the Poisson model with (a) $\alpha = 0$, $\kappa = -0.5$ and $\sigma_\eta = 0.01$, (b) $\alpha = 0$, $\kappa = 0$ and $\sigma_\eta = 0.01$ and (c) $\alpha = 0$, $\kappa = 0.5$ and $\sigma_\eta = 0.01$.



(a)

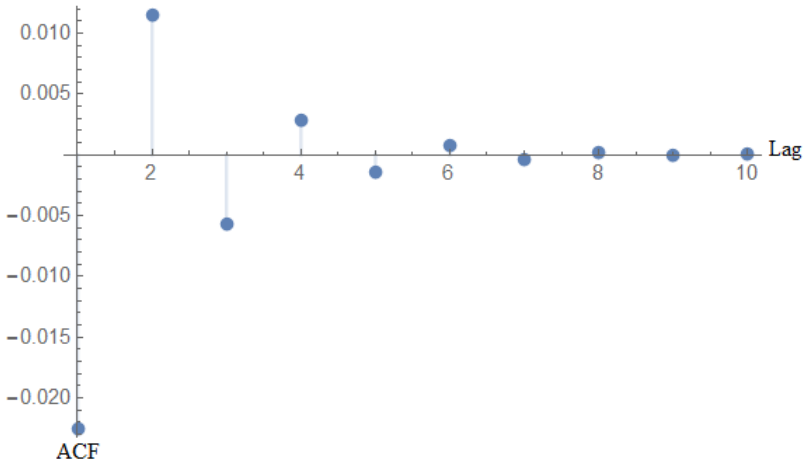


(b)

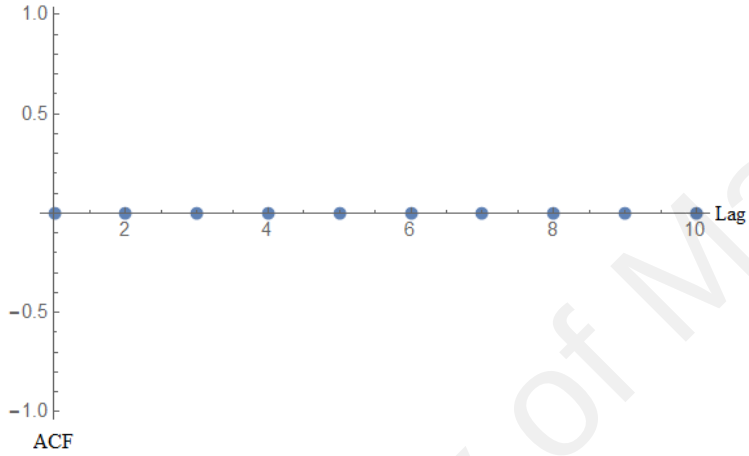


(c)

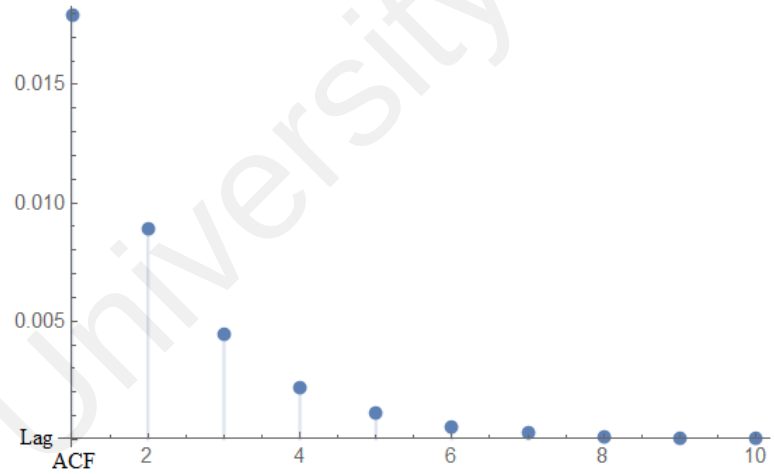
Figure 3.4: ACF plots for the Poisson model with (a) $\alpha = 0.2$, $\kappa = -0.5$ and $\sigma_\eta = 0.01$, (b) $\alpha = 0.2$, $\kappa = 0$ and $\sigma_\eta = 0.01$ and (c) $\alpha = 0.2$, $\kappa = 0.5$ and $\sigma_\eta = 0.01$.



(a)

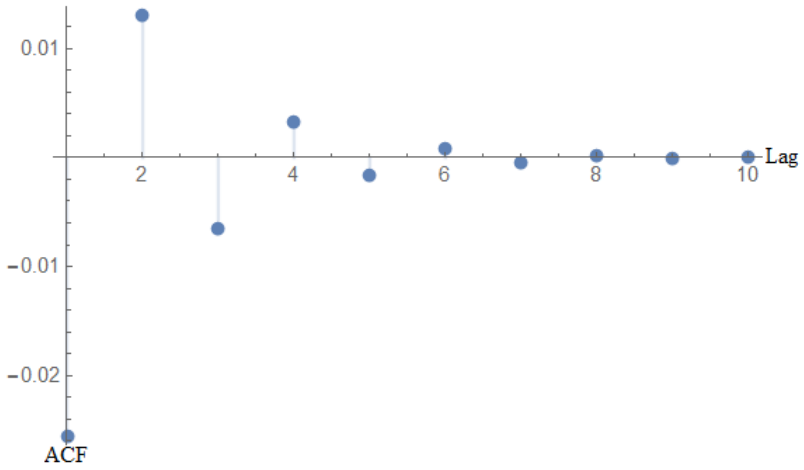


(b)

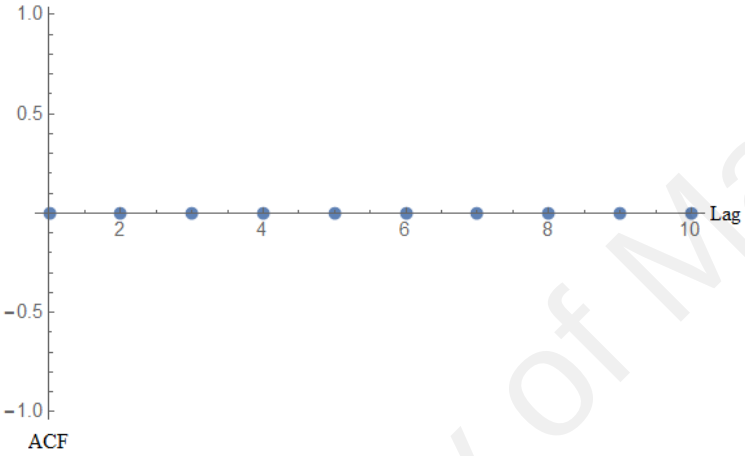


(c)

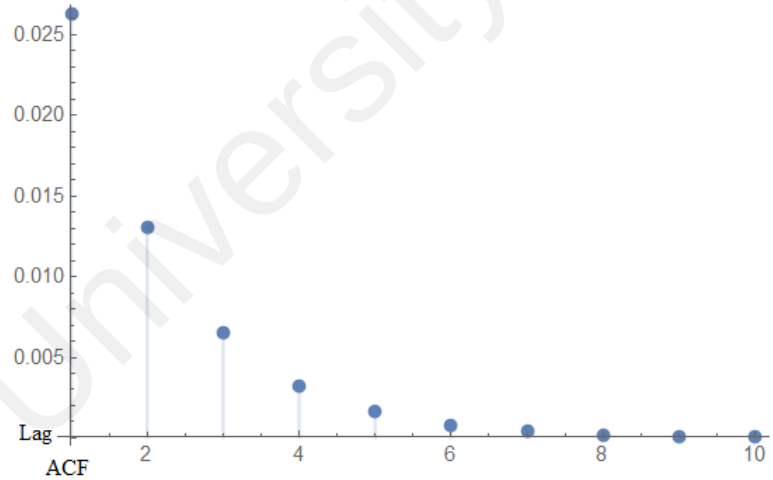
Figure 3.5: ACF plots for the Poisson model with (a) $\alpha = -0.2$, $\kappa = -0.5$ and $\sigma_\eta = 0.2$, (b) $\alpha = -0.2$, $\kappa = 0$ and $\sigma_\eta = 0.2$ and (c) $\alpha = -0.2$, $\kappa = 0.5$ and $\sigma_\eta = 0.2$.



(a)

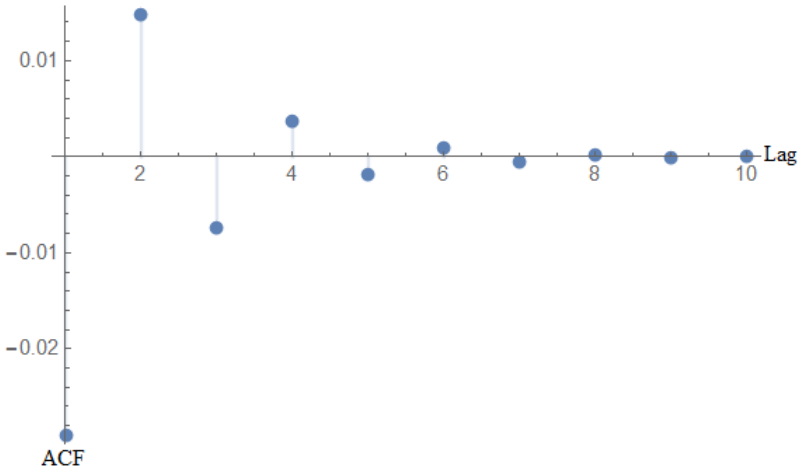


(b)

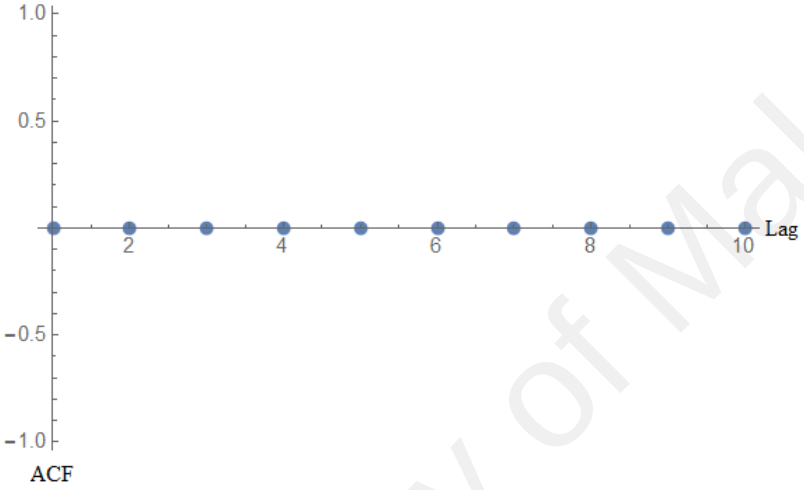


(c)

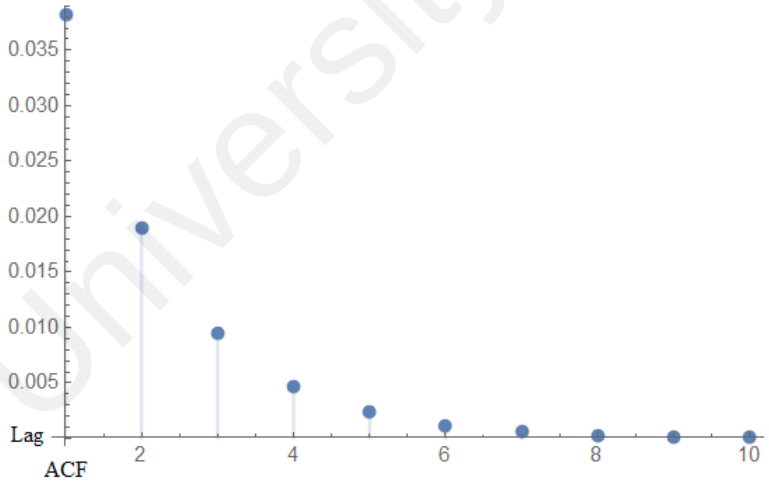
Figure 3.6: ACF plots for the Poisson model with (a) $\alpha = 0$, $\kappa = -0.5$ and $\sigma_\eta = 0.2$, (b) $\alpha = 0$, $\kappa = 0$ and $\sigma_\eta = 0.2$ and (c) $\alpha = 0$, $\kappa = 0.5$ and $\sigma_\eta = 0.2$.



(a)



(b)



(c)

Figure 3.7: ACF plots for the Poisson model with (a) $\alpha = 0.2$, $\kappa = -0.5$ and $\sigma_\eta = 0.2$, (b) $\alpha = 0.2$, $\kappa = 0$ and $\sigma_\eta = 0.2$ and (c) $\alpha = 0.2$, $\kappa = 0.5$ and $\sigma_\eta = 0.2$.

When $\kappa = 0,5$, the ACF values are positive and dies down exponencially while when $\kappa = -0.5$, the ACF oscillates between between negative and positive values. Similar results are observed when different values of σ_η is considered. ACF values when $\sigma_\eta = 0.2$ is smaller to compare with the ACF value of $\sigma_\eta = 0.01$. This is because ACF is directly proportional to σ_η value as shown in Equation (3.5). Based on all the graphs in Fig. (3.2) to Fig. (3.7), we can say that the our derivation is correct and follows the theoretical results.

3.4 Dynamic zero-inflated Poisson (ZIP) model

The model is defined as $Y_t|x_t \sim \text{ZIP}[(1-\omega)\lambda_t]$ with conditional probability function

$$P(Y_t = y|x_t) = \omega I_{(y=0)} + (1-\omega)\frac{e^{-\lambda_t}\lambda_t^y}{y!}, \tag{3.21}$$

and $\lambda_t = e^{x_t}$ and x_t is defined as Equation (3.6). $I_{(y=0)}$ is an indicator takes the value 1 when $y = 0$ and 0 otherwise, while the zero-inflation parameter ω are treated as constant. We can rewrite the dynamic ZIP model in the form;

$$x_t|x_{t-1} \sim \text{N}(\alpha + \kappa x_{t-1}, \sigma_\eta^2), \tag{3.22}$$

$$Y_t|x_t \sim \text{ZIP}[(1-\omega)\lambda_t]. \tag{3.23}$$

The cumulative MGF of the model is;

$$\begin{aligned} E[e^{Y_t r}|x_t] &= \sum_{y=0}^{\infty} e^{yr}\left[\omega I_{(y=0)} + (1-\omega)\frac{e^{-\lambda_t}\lambda^y}{y!}\right] \\ &= \omega + (1-\omega)e^{\lambda(e^r-1)} \end{aligned}$$

The conditional moments of the zero inflated Poisson distribution are given by;

$$
\begin{aligned}
E(Y_t|x_t) &= (1-\omega)\lambda_t, \\
\sigma^2(Y_t|x_t) &= (1-\omega)\lambda_t(1+\lambda_t\omega), \\
\gamma(Y_t|x_t) &= \frac{\lambda_t(1-\omega)\left[1+3\lambda_t\omega+\lambda_t^2\omega(2\omega-1)\right]}{[(1-\omega)\lambda_t(1+\lambda_t\omega)]^{3/2}}, \\
K(Y_t|x_t) &= \frac{\lambda_t(1-\omega)\left[1+6\lambda_t^2\omega^2+\lambda_t(3+4\omega)+\lambda_t^3\omega(1-3\omega+3\omega^2)\right]}{[(1-\omega)\lambda_t(1+\lambda_t\omega)]^2}.
\end{aligned}
$$

The initial state of x_0 is assumed to be normally distributed with mean, μ_0, and variance, σ_0^2. The MGF of Y, $M_Y(r)$, is

$$
M_Y(r) = E[\omega + (1-\omega)e^{\lambda_t(e^r-1)}] = \omega + (1-\omega)M_{\lambda_t}(e^r-1). \tag{3.24}
$$

From Equation (3.24), we can derive the raw moment properties of the dynamic ZIP model

$$
\begin{aligned}
E(Y_t) &= \mathrm{M}'_Y(r)|_{r=0} = (1-\omega)\mathrm{M}'_{\lambda_t}(e^r-1)e^r|_{r=0} = (1-\omega)\mathrm{M}'_{\lambda_t}(0)e^0 \\
&= (1-\omega)E(\lambda_t) = (1-\omega)E(e^{x_t}) = (1-\omega)e^{\mu_x+\frac{1}{2}\sigma_x^2} \\
&= (1-\omega)\zeta \\
E(Y_t^2) &= \mathrm{M}''_Y(r)|_{r=0} = (1-\omega)\mathrm{M}''_{\lambda_t}(e^r-1)e^{2r}|_{r=0} + (1-\omega)\mathrm{M}'_{\lambda_t}(e^r-1)e^r|_{r=0} \\
&= (1-\omega)\left[\mathrm{M}''_{\lambda_t}(e^r-1)e^{2r} + \mathrm{M}'_{\lambda_t}(e^r-1)e^r\right]|_{r=0} \\
&= (1-\omega)\left[\mathrm{M}''_{\lambda_t}(0)e^0 + \mathrm{M}'_{\lambda_t}(0)e^0\right] \\
&= (1-\omega)\left[E(\lambda_t^2) + E(\lambda_t)\right] = (1-\omega)\left[E(e^{2x_t}) + E(e^{x_t})\right] \\
&= (1-\omega)\left[e^{2\mu_x+2\sigma_x^2} + e^{\mu_x+\frac{1}{2}\sigma_x^2}\right] \\
&= (1-\omega)\zeta\left[\zeta e^{\sigma_x^2} + 1\right]
\end{aligned}
$$

$$E(Y_t^3) = \mathrm{M}_Y'''(r)|_{r=0} = (1-\omega)\left[\mathrm{M}_{\lambda_t}'''(e^r-1)e^{3r} + 3\mathrm{M}_{\lambda_t}''(e^r-1)e^{2r} + \mathrm{M}_{\lambda_t}'(e^r-1)e^r\right]|_{r=0}$$

$$= (1-\omega)\left[\mathrm{M}_{\lambda_t}'''(0)e^0 + 3\mathrm{M}_{\lambda_t}''(0)e^0 + \mathrm{M}_{\lambda_t}'(0)e^0\right]$$

$$= (1-\omega)\left[E(\lambda_t^3) + 3E(\lambda_t^2) + E(\lambda_t)\right] = (1-\omega)\left[E(e^{3x_t}) + 3E(e^{2x_t}) + E(e^{x_t})\right]$$

$$= (1-\omega)\left[e^{3\mu_x+\frac{9}{2}\sigma_x^2} + 3e^{2\mu_x+2\sigma_x^2} + e^{\mu_x+\frac{1}{2}\sigma_x^2}\right]$$

$$= (1-\omega)\zeta\left[\zeta^2 e^{3\sigma_x^2} + 3\zeta e^{\sigma_x^2} + 1\right]$$

$$E(Y_t^4) = \mathrm{M}_Y''''(r)|_{r=0}$$

$$= (1-\omega)\left[\mathrm{M}_{\lambda_t}''''(e^r-1)e^{4r} + 6\mathrm{M}_{\lambda_t}'''(e^r-1)e^{3r} + 7\mathrm{M}_{\lambda_t}''(e^r-1)e^{2r} + \mathrm{M}_{\lambda_t}'(e^r-1)e^r\right]|_{r=0}$$

$$= (1-\omega)\left[\mathrm{M}_{\lambda_t}''''(0)e^0 + 6\mathrm{M}_{\lambda_t}'''(0)e^0 + 7\mathrm{M}_{\lambda_t}''(0)e^0 + \mathrm{M}_{\lambda_t}'(0)e^0\right]$$

$$= (1-\omega)\left[E(\lambda_t^4) + 6E(\lambda_t^3) + 7E(\lambda_t^2) + E(\lambda_t)\right]$$

$$= (1-\omega)\left[E(e^{4x_t}) + 6E(e^{3x_t}) + 7E(e^{2x_t}) + E(e^{x_t})\right]$$

$$= (1-\omega)\left[e^{4\mu_x+8\sigma_x^2} + 6e^{3\mu_x+\frac{9}{2}\sigma_x^2} + 7e^{2\mu_x+2\sigma_x^2} + e^{\mu_x+\frac{1}{2}\sigma_x^2}\right]$$

$$= (1-\omega)\zeta\left[\zeta^3 e^{6\sigma_x^2} + 6\zeta^2 e^{3\sigma_x^2} + 7\zeta e^{\sigma_x^2} + \zeta\right]$$

From the derivations of the raw moments above, we can find the central moment of the dynamic ZIP model.

- Mean

$$E(Y_t) = \mathrm{M}_Y'(0)$$

$$= (1-\omega)\zeta. \tag{3.25}$$

- Variance

$$\begin{aligned}\sigma^2(Y_t) &= E(Y_t^2) - [E(Y_t)]^2 \\ &= \mathrm{M}_Y''(0) - [\mathrm{M}_Y'(0)]^2 \\ &= (1-\omega)\zeta\left[\zeta e^{\sigma_x^2} - (1-\omega)\zeta + 1\right]. \end{aligned} \tag{3.26}$$

- Skewness

$$\begin{aligned}\gamma(Y_t) &= \frac{E[Y_t - E(Y_t)]^3}{[\sigma^2(Y_t)]^{3/2}} \\ &= \frac{E[Y_t^3] - 3E(Y_t)\sigma^2(Y_t) - [E(Y_t)]^3}{[\sigma^2(Y_t)]^{3/2}} \\ &= \frac{\mathrm{M}_Y'''(0) - 3\mathrm{M}_Y'(0)\sigma^2(Y_t) - [\mathrm{M}_Y'(0)]^3}{[\sigma^2(Y_t)]^{3/2}} \\ &= \frac{2(1-\omega)^3\zeta^3 - 3(1-\omega)^2\left[\zeta^3 e^{\sigma_x^2} + \zeta^2\right] + (1-\omega)\left[\zeta^4 e^{3\sigma_x^2} + 3\zeta^2 e^{\sigma_x^2} + \zeta\right]}{\left[(1-\omega)\zeta\left[\zeta e^{\sigma_x^2} - (1-\omega)\zeta + 1\right]\right]^{\frac{3}{2}}} \end{aligned} \tag{3.27}$$

- Kurtosis

$$\begin{aligned}K(Y_t) &= \frac{E[Y_t - E(Y_t)]^4}{[\sigma^2(Y_t)]^2} \\ &= \frac{E(Y_t^4) - 4E(Y_t)E(Y_t^3) + 6E(Y_t^2)[E(Y_t)]^2 - 4E(Y_t)[E(Y_t)]^3 + [E(Y_t)]^4}{[\sigma^2(Y_t)]^2} \\ &= \frac{\mathrm{M}_Y''''(0) - 4\mathrm{M}_Y'(0)\mathrm{M}_Y'''(0) + 6\mathrm{M}_Y''(0)[\mathrm{M}_Y'(0)]^2 - 4\mathrm{M}_Y'(0)[\mathrm{M}_Y'(0)]^3 + [\mathrm{M}_Y'(0)]^4}{[\sigma^2(Y_t)]^2} \\ &= \frac{7(1-\omega)\zeta^2 e^{\sigma_x^2} + 6(1-\omega)\zeta^3 e^{3\sigma_x^2} + (1-\omega)\zeta + (1-\omega)\zeta^4 e^{6\sigma_x^2}}{\left[(1-\omega)\zeta\left[\zeta e^{\sigma_x^2} - (1-\omega)\zeta + 1\right]\right]^2} \\ &\quad - \frac{4(1-\omega)\zeta\left[3(1-\omega)\zeta^2 e^{\sigma_x^2} + (1-\omega)\zeta + (1-\omega)\zeta^3 e^{3\sigma_x^2}\right]}{\left[(1-\omega)\zeta\left[\zeta e^{\sigma_x^2} - (1-\omega)\zeta + 1\right]\right]^2} \\ &\quad + \frac{6(1-\omega)^3\zeta^2\left[\zeta + \zeta^2 e^{\sigma_x^2}\right] + 3\zeta^4(1-\omega)^4}{\left[(1-\omega)\zeta\left[\zeta e^{\sigma_x^2} - (1-\omega)\zeta + 1\right]\right]^2} \end{aligned} \tag{3.28}$$

The derivations of the autocovariance and autocorrelation for the dynamic ZIP model is quite similar to the dynamic Poisson model. Following the working step in Equation (3.18), we have

$$\begin{aligned} E(Y_tY_{t-i}) &= E\{E[(Y_tY_{t-i})|\lambda_t, \lambda_{t-i}]\} \\ &= E\left[E[(Y_t|\lambda_t)E(Y_{t-i}|\lambda_{t-i})]\right] \\ &= (1-\omega)^2 E(e^{x_t}e^{x_{t-i}}). \end{aligned} \tag{3.29}$$

By substituting α, σ_η^2 and x_t from Equation (3.7) into Equation (3.29), we have

$$\begin{aligned} E(Y_tY_{t-i}) &= (1-\omega)^2 E\left[e^{x_t}e^{x_{t-i}}\right] \\ &= (1-\omega)^2 E\left[e^{\frac{\alpha(1-\kappa^i)}{1-\kappa}+\kappa^i x_{t-i}+\sum_{j=0}^{j=i-1}(\kappa^j\eta_{t-j})}e^{x_{t-i}}\right] \\ &= (1-\omega)^2 E\left\{exp\left[\frac{\alpha(1-\kappa^i)}{1-\kappa}+(\kappa^i+1)x_{t-i}+\sum_{j=0}^{j=i-1}(\kappa^j\eta_{t-j})\right]\right\} \\ &= (1-\omega)^2 e^{\mu_x(1-\kappa^i)}e^{\mu_x(\kappa^i+1)+\frac{1}{2}\sigma_x^2(\kappa^i+1)^2}e^{\frac{1}{2}(1-\kappa^{2i})\sigma_x^2} \\ &= (1-\omega)^2 exp\left[\mu_x(1-\kappa^i)+\mu_x(\kappa^i+1)+\frac{1}{2}\sigma_x^2(\kappa^i+1)^2+\frac{1}{2}(1-\kappa^{2i})\sigma_x^2\right] \\ &= (1-\omega)^2 exp\left[2\mu_x+(1+\kappa^i)\sigma_x^2\right]. \end{aligned}$$

Hence, the autocovariance and autocorrelation of the dynamic ZIP model are;

$$\begin{aligned}
\text{Cov}(Y_t, Y_{t-i}) &= E(Y_t Y_{t-i}) - E(Y_t)E(Y_{t-i}) \\
&= (1-\omega)^2 e^{2\mu_x + (1+\kappa^i)\sigma_x^2} - \left[(1-\omega)e^{\mu_x + \frac{1}{2}\sigma_x^2}\right]^2 \\
&= (1-\omega)^2 e^{2\mu_x + (1+\kappa^i)\sigma_x^2} - \left[(1-\omega)^2 e^{2\mu_x + \sigma_x^2}\right] \\
&= (1-\omega)^2 \zeta^2 \left[e^{\kappa^i \sigma_x^2} - 1\right]. \qquad (3.30)
\end{aligned}$$

$$\begin{aligned}
\rho_i &= \frac{\text{Cov}(Y_t, Y_{t-i})}{\sqrt{\sigma_{Y_t}^2 \sigma_{Y_{t-i}}^2}} \\
&= \frac{(1-\omega)^2 \zeta^2 \left[e^{\kappa^i \sigma_x^2} - 1\right]}{(1-\omega)\zeta^2 \left[e^{\sigma_x^2} - (1-\omega)\right] + (1-\omega)\zeta}. \qquad (3.31)
\end{aligned}$$

3.4.1 Simulation study

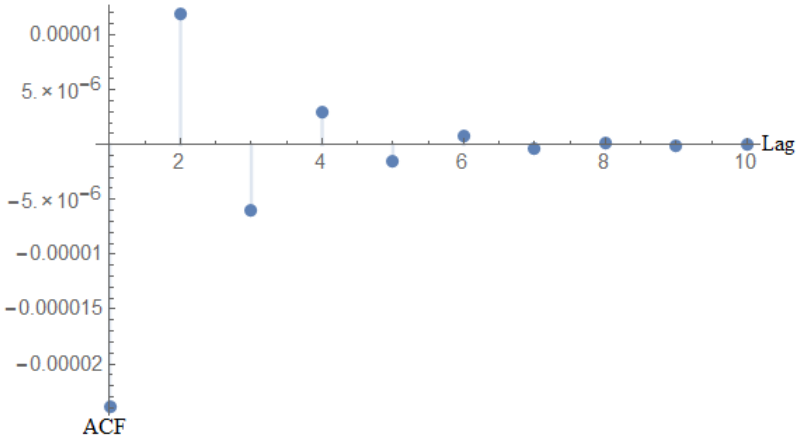
We repeat the steps in Section 3.3.1 where we generate a data of size 100,000 and calculate the standard central moment and its autocovariance and autocorrelation. Table 3.3 and Table 3.4 show the generated and true value of the moment structures for the dynamic ZIP model. For this model, ω is kept constant at 0.5. ACV(1) and ACF(1) are the first order autocovariance and autocorrelation respectively. From the MSE results, we can conclude that our derivation is correct based on the comparisons of the true value and the empirical value from the generated data. In general, the absolute differences are small. Smaller differences could be obtained by increasing the sample size. When $\kappa = 0.5$, all lags are positive and dies down exponentially, while when $\kappa = -0.5$, the lags oscillates between negative and positive values.

Table 3.3: Generated data and true values for the moment structures with $\sigma_\eta = 0.01$.

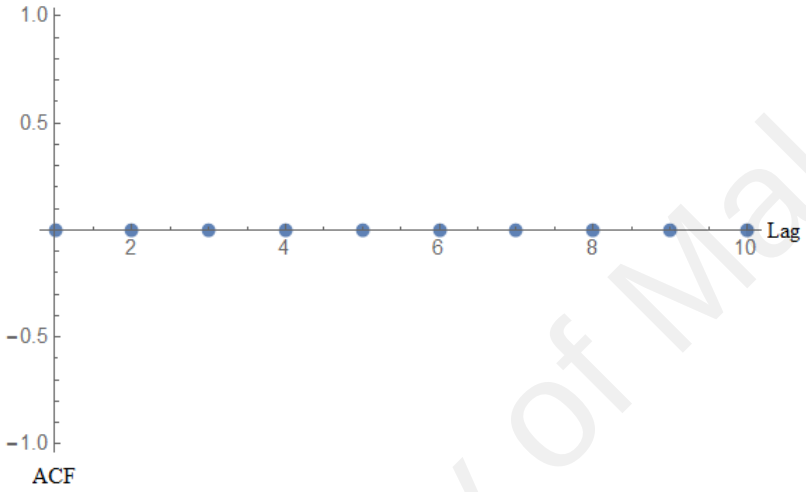
		$\alpha = -0.2$			$\alpha = 0$			$\alpha = 0.2$		
		Generated	True	MSE	Generated	True	MSE	Generated	True	MSE
$\kappa = -0.5$	μ_{Y_t}	0.4375	0.4376	5.71E-06	0.5000	0.5000	7.45E-06	0.5712	0.5714	9.75E-06
	$\sigma^2_{Y_t}$	0.6287	0.6297	2.70E-05	0.7502	0.7501	2.84E-05	0.8977	0.8979	4.53E-05
	γ_{Y_t}	2.0282	2.0285	2.98E-04	1.9251	1.9247	2.61E-04	1.8254	1.8231	1.69E-04
	K_{Y_t}	7.3106	7.3085	1064E-02	6.7805	6.7788	1.35E-02	6.2979	6.2806	8.28E-03
	ACV(1)	-1.88E-04	-1.28E-05	3.09E-08	-2.00E-03	-1.67E-05	4.09E-06	-3.00E-03	-2.18E-05	8.46E-06
	ACF(1)	-3.01E-04	-2.39E-05	7.68E-08	-2.64E-03	-2.67E-05	7.11E-06	-3.24E-03	-2.96E-05	1.07E-05
$\kappa = 0$	μ_{Y_t}	0.4093	0.4094	6.15E-06	0.5001	0.5000	7.09E-06	0.6109	0.6107	1.07E-05
	$\sigma^2_{Y_t}$	0.5768	0.5770	2.34E-05	0.7503	0.7501	3.45E-05	0.9844	0.9838	4.37E-05
	γ_{Y_t}	2.0825	2.0814	2.08E-04	1.9228	1.9246	1.40E-04	1.7728	1.7730	1.92E-04
	K_{Y_t}	7.5895	7.5859	1.37E-02	6.7641	6.7785	8.13E-03	6.0450	6.0426	8.78E-03
	ACV(1)	1.47E-03	-1.97E-05	2.22E-06	-4.86E-04	0.0032	1.32E-05	7.07E-04	-0.0044	2.62E-05
	ACF(1)	2.59E-03	-3.45E-05	6.68E-06	-6.41E-04	0.0042	2.33E-05	7.12E-04	-0.0044	2.60E-05
$\kappa = 0.5$	μ_{Y_t}	0.3349	0.3352	5.39E-06	0.5002	0.5000	6.28E-06	0.7455	0.7460	1.53E-05
	$\sigma^2_{Y_t}$	0.4473	0.4476	1.52E-05	0.7504	0.7501	2.81E-05	1.3019	1.3026	7.23E-05
	γ_{Y_t}	2.2459	2.2455	3.27E-04	1.9253	1.9247	1.99E-04	1.6256	1.6252	1.21E-04
	K_{Y_t}	8.4769	8.4753	2.27E-02	6.7902	6.7788	1.18E-02	5.3720	5.3722	3.83E-03
	ACV(1)	2.59E-03	7.49E-06	6.67E-06	-1.49E-03	-1.67E-05	2.26E-06	3.84E-03	3.71E-05	1.45E-05
	ACF(1)	5.77E-03	1.91E-05	3.30E-05	-1.98E-03	-2.67E-05	4.01E-06	2.95E-03	3.62E-05	8.51E-06

Table 3.4: Generated data and true values for the moment structures with $\sigma_\eta = 0.2$.

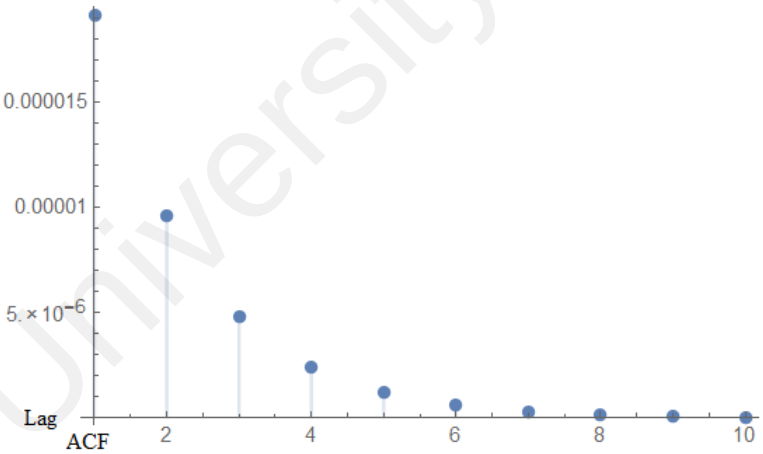
		$\alpha = -0.2$			$\alpha = 0$			$\alpha = 0.2$		
		Generated	True	MSE	Generated	True	MSE	Generated	True	MSE
$\kappa = -0.5$	μ_{Y_t}	0.4492	0.4494	5.15E-06	0.5137	0.5135	8.94E-06	0.5867	0.5868	1.22E-05
	$\sigma^2_{Y_t}$	0.6727	0.6735	2.97E-05	0.8057	0.8061	4.22E-05	0.9692	0.9688	6.37E-05
	γ_{Y_t}	2.0881	2.0894	2.96E-04	1.9876	1.9907	2.64E-04	1.8955	1.8947	1.74E-04
	K_{Y_t}	7.7184	7.7261	1.79E-02	7.1879	7.2069	1.69E-02	6.7247	6.7196	9.66E-03
	ACV(1)	-0.0060	-0.0053	5.59E-07	-0.0051	-0.0069	3.42E-06	-0.0113	-0.0091	4.80E-06
	ACF(1)	-0.0090	-0.0095	2.06E-07	-0.0063	-0.0105	1.79E-05	-0.0115	-0.0116	1.23E-08
$\kappa = 0$	μ_{Y_t}	0.4178	0.4176	5.08E-06	0.5101	0.5101	8.55E-06	0.6231	0.6230	7.31E-06
	$\sigma^2_{Y_t}$	0.6064	0.6063	2.09E-05	0.7918	0.7915	4.45E-05	1.0428	1.0429	5.78E-05
	γ_{Y_t}	2.1236	2.1244	2.60E-04	1.9724	1.9733	2.60E-04	1.8272	1.8279	1.75E-04
	K_{Y_t}	7.8837	7.8875	1.84E-02	7.0822	7.0913	1.47E-02	6.3615	6.3669	9.09E-03
	ACV(1)	-0.0026	-0.0026	2.50E-10	-0.0043	-0.0043	1.77E-09	-0.0019	-0.0019	2.06E-09
	ACF(1)	-0.0043	-0.0043	6.31E-13	-0.0054	-0.0054	1.74E-10	-0.0019	-0.0019	5.18E-10
$\kappa = 0.5$	μ_{Y_t}	0.3444	0.3442	3.79E-06	0.5137	0.5135	7.10E-06	0.7661	0.7661	1.51E-05
	$\sigma^2_{Y_t}$	0.4757	0.4757	1.56E-05	0.8071	0.8061	3.42E-05	1.4166	1.4172	1.09E-04
	γ_{Y_t}	2.2956	2.2968	3.08E-04	1.9927	1.9907	2.00E-04	1.7094	1.7093	1.40E-04
	K_{Y_t}	8.8678	8.88720	2.31E-02	7.2235	7.2069	1.17E-02	5.8416	5.8336	7.31E-03
	ACV(1)	0.0022	0.0032	1.03E-06	0.0073	0.0071	2.60E-08	0.0095	0.0159	4.06E-05
	ACF(1)	0.0046	0.0078	1.01E-05	0.0090	0.0108	3.07E-06	0.0067	0.0145	6.09E-05



(a)

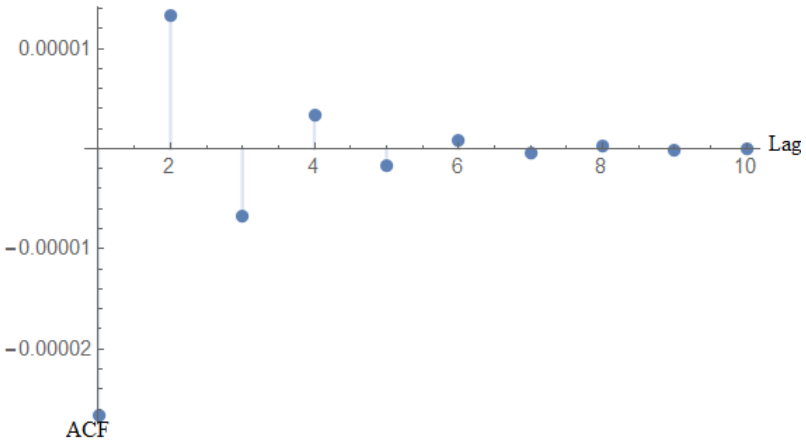


(b)

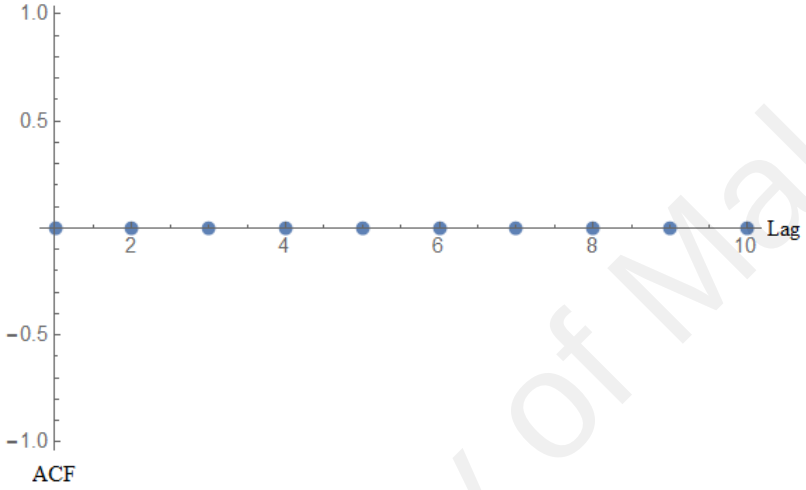


(c)

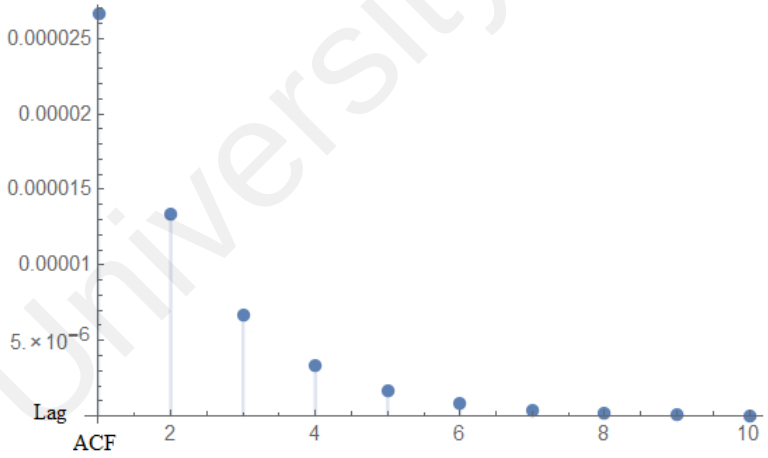
Figure 3.8: ACF plots for the ZIP model with (a) $\alpha = -0.2$, $\kappa = -0.5$ and $\sigma_\eta = 0.01$, (b) $\alpha = -0.2$, $\kappa = 0$ and $\sigma_\eta = 0.01$, and (c) $\alpha = -0.2$, $\kappa = 0.5$ and $\sigma_\eta = 0.01$.



(a)

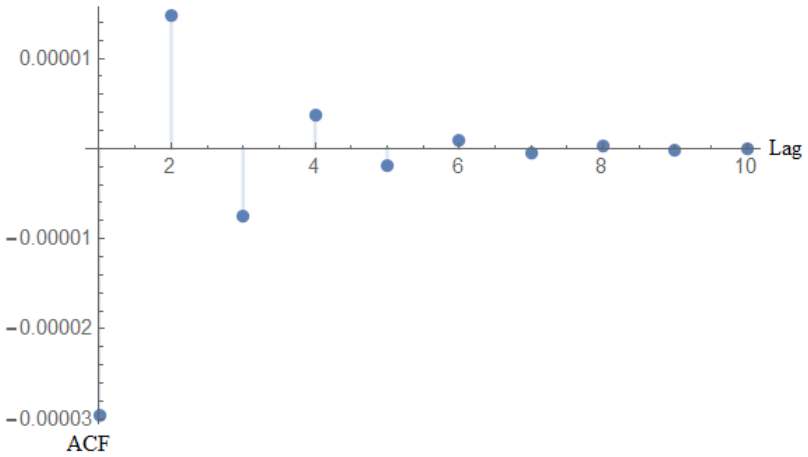


(b)

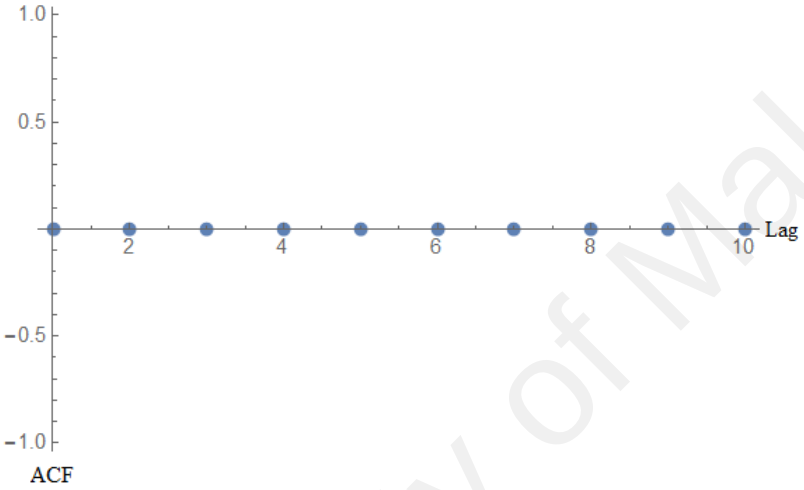


(c)

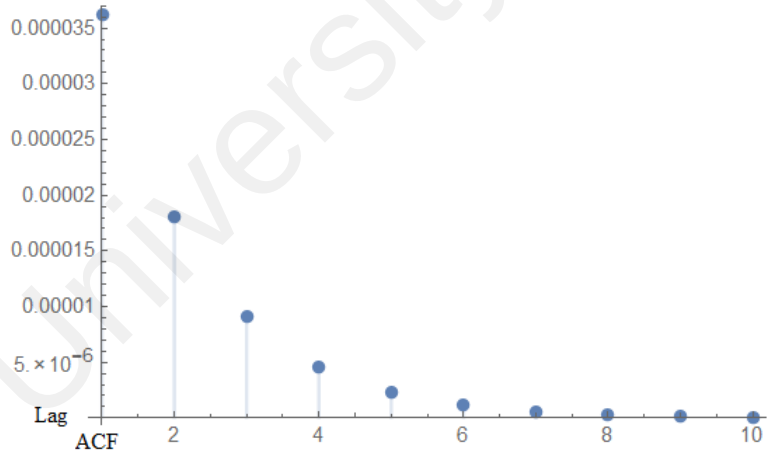
Figure 3.9: ACF plots for the ZIP model with (a) $\alpha = 0$, $\kappa = -0.5$ and $\sigma_\eta = 0.01$, (b) $\alpha = 0$, $\kappa = 0$ and $\sigma_\eta = 0.01$, and (c) $\alpha = 0$, $\kappa = 0.5$ and $\sigma_\eta = 0.01$.



(a)

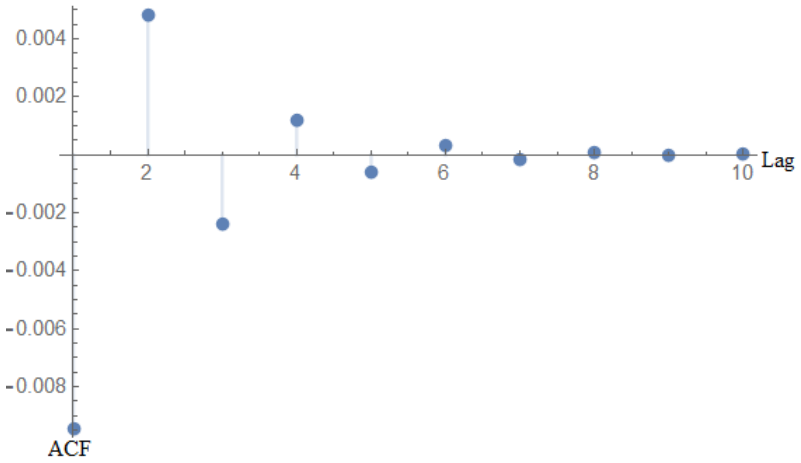


(b)

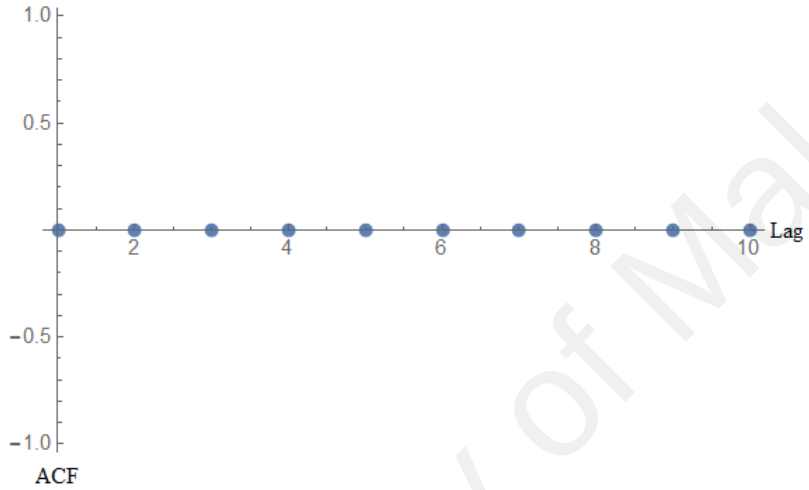


(c)

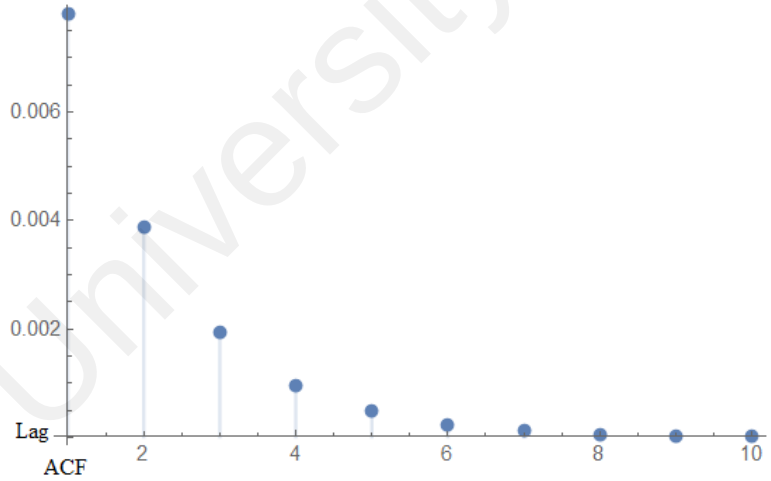
Figure 3.10: ACF plots for the ZIP model with (a) $\alpha = 0.2$, $\kappa = -0.5$ and $\sigma_\eta = 0.01$, (b) $\alpha = 0.2$, $\kappa = 0$ and $\sigma_\eta = 0.01$, and (c) $\alpha = 0.2$, $\kappa = 0.5$ and $\sigma_\eta = 0.01$.



(a)

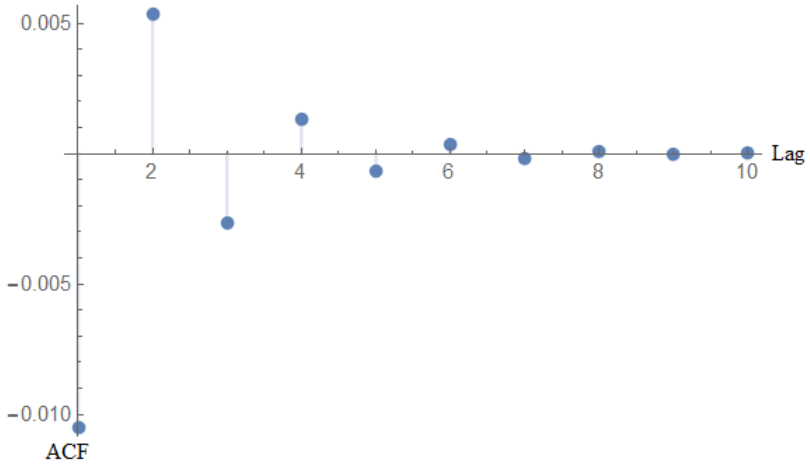


(b)



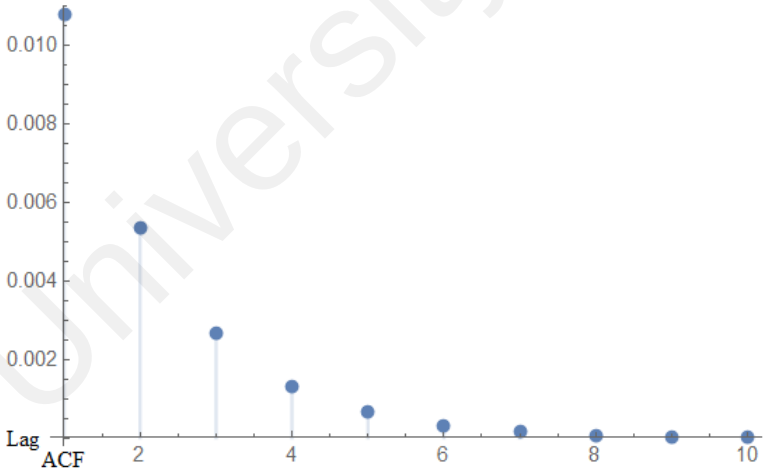
(c)

Figure 3.11: ACF plots for the ZIP model with (a) $\alpha = -0.2$, $\kappa = -0.5$ and $\sigma_\eta = 0.2$, (b) $\alpha = -0.2$, $\kappa = 0$ and $\sigma_\eta = 0.2$, and (c) $\alpha = -0.2$, $\kappa = 0.5$ and $\sigma_\eta = 0.2$.



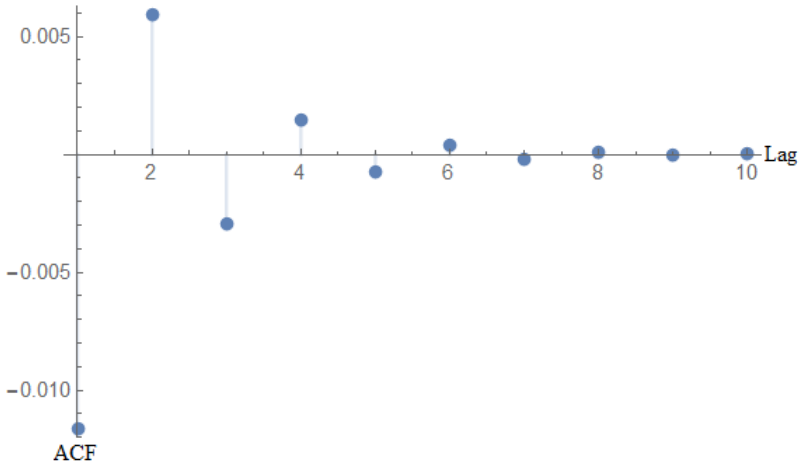
(a)

(b)

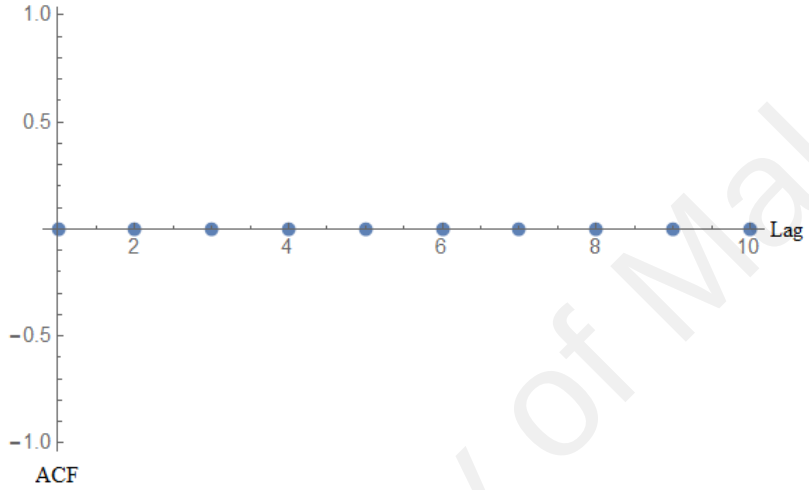


(c)

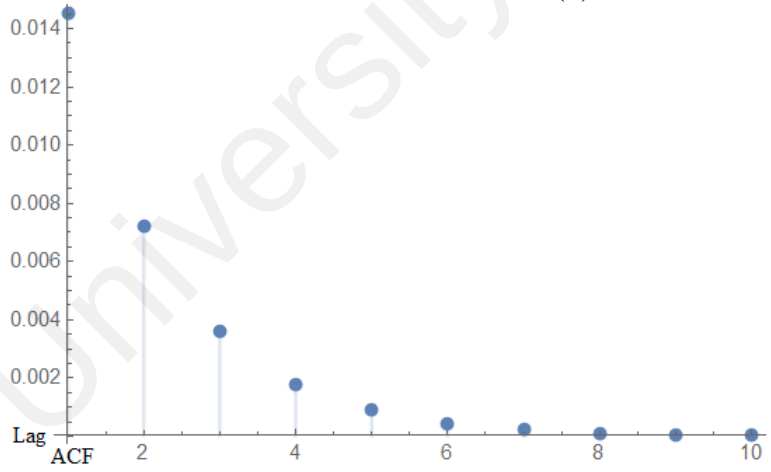
Figure 3.12: ACF plots for the ZIP model with (a) $\alpha = 0$, $\kappa = -0.5$ and $\sigma_\eta = 0.2$, (b) $\alpha = 0$, $\kappa = 0$ and $\sigma_\eta = 0.2$, and (c) $\alpha = 0$, $\kappa = 0.5$ and $\sigma_\eta = 0.2$.



(a)



(b)



(c)

Figure 3.13: ACF plots for the ZIP model with (a) $\alpha = 0.2$, $\kappa = -0.5$ and $\sigma_\eta = 0.2$, (b) $\alpha = 0.2$, $\kappa = 0$ and $\sigma_\eta = 0.2$, and (c) $\alpha = 0.2$, $\kappa = 0.5$ and $\sigma_\eta = 0.2$.

When $\kappa = 0.5$, all lags are positive and dies down exponentially, while when $\kappa = -0.5$, the lags oscillates between negative and positive values. Based on all the graphs in Fig. (3.8) to Fig (3.13), we can say that the our derivation is correct and follows the theoretical results.

3.5 Dynamic negative binomial model

We establish a class of dynamic models for overdispersed count time series, based on the negative binomial distribution. The probability mass function of the observation y_t, conditioning on the current state x_t, is defined as

$$Y_t|x_t \sim \text{NB}(P_t, k),$$

with conditional probability function

$$P(Y_t = y|x_t) = \begin{pmatrix} y+k-1 \\ k-1 \end{pmatrix} {P_t}^k (1-P_t)^y, \tag{3.32}$$

where P_t is the probability of success in a negative binomial distribution at time t, and $\lambda_t = \dfrac{k(1-P_t)}{P_t}$ where k is the number of trial. λ_t is assumed to be modelled as $\lambda_t = e^{x_t}$ and x_t is defined as Equation (3.6).

Similar to the earlier dynamic models, we can rewrite the dynamic negative binomial model in the following form:

$$x_t|x_{t-1} \sim \text{N}(\alpha + \kappa x_{t-1}, \sigma_\eta^2), \tag{3.33}$$

$$Y_t|x_t \sim \text{NB}(P_t, k). \tag{3.34}$$

The initial state of x_0 is assumed to be normally distributed with mean, μ_0, and variance,

σ_0^2. The CMGF of the model is given by

$$E[e^{Y_t r}|x_t] = \left[\frac{P_t}{(1-P_t)e^t}\right]^r.$$

The conditional moments of the negative binomial distribution are given by:

$$\begin{aligned} E(Y_t|x_t) &= \lambda_t, \\ \sigma^2(Y_t|x_t) &= \lambda_t + \tau\lambda_t^2, \\ \gamma(Y_t|x_t) &= \frac{1+2\lambda_t\tau}{\sqrt{\lambda_t(1+\lambda_t\tau)}}, \\ K(Y_t|x_t) &= \frac{1+3\lambda_t^2(1+2\tau)+\lambda_t(3+6\tau)}{\lambda_t(\tau\lambda_t+1)}, \end{aligned}$$

where $\tau = \frac{1}{k}$. The derivation of the central moments for dynamic NB model are quite simple.

- Mean

$$E(Y_t) = E[E(Y_t|x_t)] = E(\lambda_t) = e^{\mu_x + \frac{1}{2}\sigma_x^2} = \zeta. \tag{3.35}$$

- Variance

$$\begin{aligned} \text{Var}(Y_t) &= \text{Var}[E(Y_t|x_t)] + E[\text{Var}(Y_t|x_t)] \\ &= \text{Var}[\lambda_t] + E[\tau\lambda_t^2 + \lambda_t] \\ &= (1+\tau)e^{2\mu_x+2\sigma_x^2} - e^{2\mu_x+\sigma_x^2} + e^{\mu_x+\frac{1}{2}\sigma_x^2} \\ &= \zeta - \zeta^2[1 - e^{\sigma_x^2}(1+\tau)]. \end{aligned} \tag{3.36}$$

- Skewness

$$\begin{aligned}\gamma(Y_t) &= \frac{E\left[Y_t - E\left(Y_t\right)\right]^3}{\left[\mathrm{Var}(Y_t)\right]^{3/2}} \\ &= \frac{e^{\mu_x+\frac{1}{2}\sigma_x^2} + 2e^{3\mu_x+\frac{3}{2}\sigma_x^2} + 3e^{2\mu_x+2\sigma_x^2} + e^{3\mu_x+\frac{9}{2}\sigma_x^2} + 3e^{2\mu_x+\sigma_x^2}\tau + 3e^{3\mu_x+\frac{9}{2}\sigma_x^2}\tau}{\left[(1+\tau)e^{2\mu_x+2\sigma_x^2} - e^{2\mu_x+\sigma_x^2} + e^{\mu_x+\frac{1}{2}\sigma_x^2}\right]^{3/2}} \\ &\quad + \frac{2e^{3\mu_x+\frac{9}{2}\sigma_x^2}\tau^2 - 3e^{\mu_x+\frac{1}{2}\sigma_x^2}\left(e^{\mu_x+\frac{1}{2}\sigma_x^2} + e^{2\mu_x+\sigma_x^2} + e^{2\mu_x+\sigma_x^2}\tau\right)}{\left[(1+\tau)e^{2\mu_x+2\sigma_x^2} - e^{2\mu_x+\sigma_x^2} + e^{\mu_x+\frac{1}{2}\sigma_x^2}\right]^{3/2}} \\ &= \frac{\zeta - 3\zeta^2 + 2\zeta^3 - 3\zeta^3 e^{\sigma_x^2}(1+\tau) + 3\zeta^4(1+\tau) + \zeta^3 e^{3\sigma_x^2}(1+3\tau+2\tau^2)}{\left[(1+\tau)\zeta^2 e^{\sigma_x^2} - \zeta^2 + \zeta\right]^{3/2}}.\end{aligned} \tag{3.37}$$

- Kurtosis

$$\begin{aligned}K(Y_t) &= \frac{E[Y_t - E(Y_t)]^4}{[\mathrm{Var}(Y_t)]^2} \\ &= \frac{e^{\mu_x+\frac{1}{2}\sigma_x^2} - 4e^{2\mu_x+\sigma_x^2} + 6e^{3\mu_x+\frac{3}{2}\sigma_x^2} - 3e^{4\mu_x+2\sigma_x^2} + 7e^{2\mu_x+2\sigma_x^2}(1+\tau)}{\left[(1+\tau)e^{2\mu_x+2\sigma_x^2} - e^{2\mu_x+\sigma_x^2} + e^{\mu_x+\frac{1}{2}\sigma_x^2}\right]^2} \\ &\quad - \frac{12e^{3\mu_x+\frac{5}{2}\sigma_x^2}(1+\tau) + 6e^{4\mu_x+3\sigma_x^2}(1+\tau) + 6e^{3\mu_x+\frac{9}{2}\sigma_x^2}(1+3\tau+2\tau^2)}{\left[(1+\tau)e^{2\mu_x+2\sigma_x^2} - e^{2\mu_x+\sigma_x^2} + e^{\mu_x+\frac{1}{2}\sigma_x^2}\right]^2} \\ &\quad - \frac{4e^{4\mu_x+5\sigma_x^2}(1+3\tau+2\tau^2) + e^{4\mu_x+8\sigma_x^2}(1+6\tau+11\tau^2+6\tau^3)}{\left[(1+\tau)e^{2\mu_x+2\sigma_x^2} - e^{2\mu_x+\sigma_x^2} + e^{\mu_x+\frac{1}{2}\sigma_x^2}\right]^2} \\ &= \frac{\zeta - 4\zeta^2 + 6\zeta^3 - 3\zeta^4 + 7\zeta^3 e^{\sigma_x^2}(1+\tau) - 12\zeta^3 e^{\sigma_x^2}(1+\tau) + 6\zeta^4 e^{\sigma_x^2}(1+\tau)}{\left[(1+\tau)\zeta^2 e^{\sigma_x^2} - \zeta^2 + \zeta\right]^2} \\ &\quad + \frac{6\zeta^3 e^{3\sigma_x^2}(1+3\tau+2\tau^2) - 4\zeta^4 e^{3\sigma_x^2}(1+3\tau+2\tau^2)}{\left[(1+\tau)\zeta^2 e^{\sigma_x^2} - \zeta^2 + \zeta\right]^2} \\ &\quad + \frac{\zeta^4 e^{6\sigma_x^2}(1+6\tau+11\tau^2+6\tau^3)}{\left[(1+\tau)\zeta^2 e^{\sigma_x^2} - \zeta^2 + \zeta\right]^2}.\end{aligned} \tag{3.38}$$

The autocovariance of the dynamic NB model is equal to the autocovariance of the dynamic Poisson model. However these two models differ in the autocorrelation function due to the presence of $\tau = \frac{1}{k}$, where k is the number of trial.

$$Cov(Y_t, Y_{t-i}) = \zeta^2 \left[e^{\kappa^i \sigma_x^2} - 1 \right], \tag{3.39}$$

$$\rho_i = \frac{\left[e^{\kappa^i \sigma_x^2} - 1 \right]}{\left[1 - \zeta + \zeta e^{\sigma_x^2}(1 + \tau) \right]^2}. \tag{3.40}$$

3.5.1 Simulation study

With k fixed to 100, we generate a data of 100,000 and calculate the standard central moments and its autocovariance and autocorrelation. Tables 3.5 and 3.6 show the generated and true values of the moment structures for the dynamic NB model. ACV(1) and ACF(1) are the first order autocovariance and autocorrelation respectively. From the MSE results, we can conclude that our derivation is correct based on the comparisons of the true value and the empirical value from the generated data. In general, the absolute differences are small. Smaller differences could be obtained by increasing the sample size.

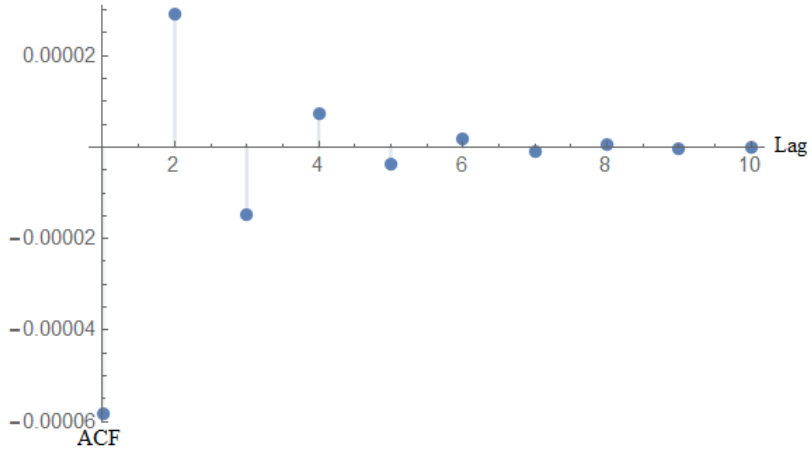
When $\kappa = 0.5$, all lags are positive and dies down exponentially, while when $\kappa = -0.5$, the lags oscillates between negative and positive values. Based on all the graphs in Figure 3.14 to Figure 3.19, we can say that the derivation is correct and follows the theoretical results.

Table 3.5: Generated data and true values for the moment structures with $\sigma_\eta = 0.01$.

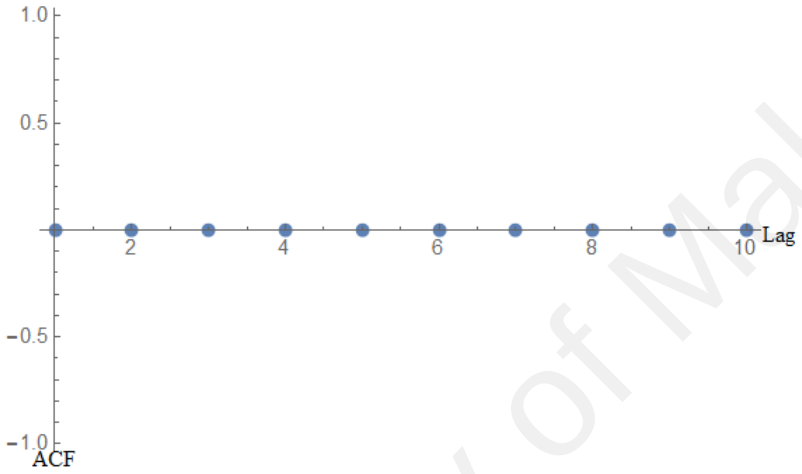
		$\alpha = -0.2$			$\alpha = 0$			$\alpha = 0.2$		
		Generated	True	MSE	Generated	True	MSE	Generated	True	MSE
$\kappa = -0.5$	μ_{Y_t}	0.8747	0.8752	6.84E-04	1.0012	1.0001	1.20E-03	1.1407	1.1427	1.19E-03
	$\sigma^2_{Y_t}$	0.8833	0.8830	1.69E-03	1.0092	1.0102	3.37E-03	1.1561	1.1559	4.10E-03
	γ_{Y_t}	1.0858	1.0831	1.30E-02	1.0039	1.0151	1.02E-02	0.9457	0.9517	9.27E-03
	K_{Y_t}	4.2146	4.1933	4.17E-01	3.9946	4.0507	2.55E-01	3.8803	3.9259	2.10E-01
	ACV(1)	2.76E-02	-5.11E-05	7.63E-04	-1.44E-04	6.67E-05	6.04E-09	-3.85E-02	-8.70E-05	1.48E-03
	ACF(1)	3.24E-02	-5.83E-05	1.05E-03	-1.41E-04	-6.67E-05	5.55E-09	-3.62E-02	-7.62E-05	1.31E-03
$\kappa = 0$	μ_{Y_t}	0.8172	0.8188	8.95E-04	1.0034	1.0001	1.18E-03	1.2221	1.2215	1.34E-03
	$\sigma^2_{Y_t}$	0.8323	0.8255	2.13E-03	1.0114	1.0102	3.39E-03	1.2290	1.2365	4.68E-03
	γ_{Y_t}	1.1270	1.1188	8.84E-03	0.9980	1.0151	1.12E-02	0.9117	0.9215	1.07E-02
	K_{Y_t}	4.2474	4.2719	2.54E-01	3.9687	4.0506	2.57E-01	3.8343	3.8693	2.13E-01
	ACV(1)	2.90E-02	-3.74E-03	1.07E-03	-6.27E-03	-1.01E-02	1.47E-05	-1.06E-02	1.30E-02	5.54E-04
	ACF(1)	3.46E-02	-4.20E-03	1.51E-03	-6.98E-0.3	-1.00E-02	9.60E-06	-8.82E-03	1.06E-02	3.78E-04
$\kappa = 0.5$	μ_{Y_t}	0.6731	0.6704	1.05E-03	1.0032	1.0001	1.02E-03	1.5017	1.4919	1.76E-03
	$\sigma^2_{Y_t}$	0.6788	0.6749	2.25E-03	1.0153	1.0102	3.51E-03	1.5265	1.5145	6.34E-03
	γ_{Y_t}	1.2225	1.2338	1.61E-02	1.0080	1.0151	1.15E-02	0.8197	0.8372	8.77E-03
	K_{Y_t}	4.4776	4.5425	5.04E-01	4.0211	4.0507	2.68E-01	3.6469	3.7211	1.66E-01
	ACV(1)	-2.32E-02	3.00E-05	5.39E-04	-2.82E-02	6.67E-05	8.01E-04	-3.40E-02	1.48E-04	1.17E-03
	ACF(1)	-3.18E-02	4.47E-05	1.01E-03	-2.72E-02	6.67E-05	7.41E-04	-2.30E-02	9.94E-05	5.33E-04

Table 3.6: Generated data and true values for the moment structures with $\sigma_\eta = 0.2$.

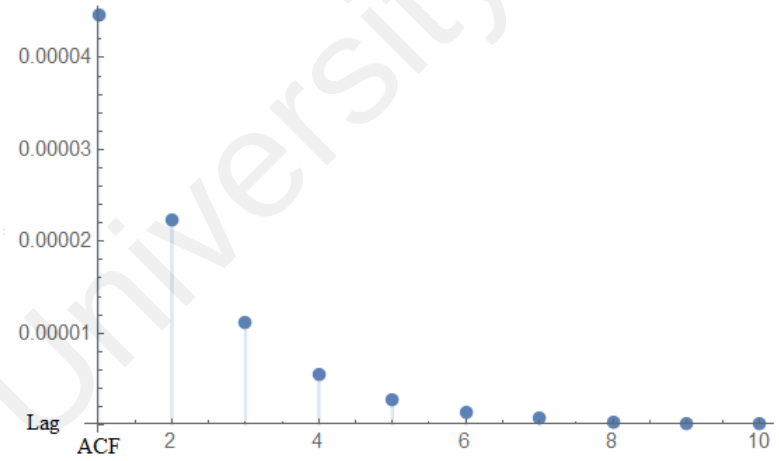
		$\alpha = -0.2$			$\alpha = 0$			$\alpha = 0.2$		
		Generated	True	MSE	Generated	True	MSE	Generated	True	MSE
$\kappa = -0.5$	μ_{Y_t}	0.9009	0.8988	1.03E-03	1.0351	1.0270	1.20E-03	1.1788	1.1735	1.03E-03
	$\sigma^2_{Y_t}$	0.9534	0.9516	3.06E-03	1.1038	1.0959	4.32E-03	1.2756	1.2635	5.67E-03
	γ_{Y_t}	1.1499	1.1490	1.13E-02	1.0632	1.0876	1.34E-02	1.0314	1.0311	1.25E-02
	K_{Y_t}	4.4717	4.4663	3.97E-01	4.2227	4.3312	4.31E-01	4.1796	4.2139	3.11E-01
	ACV(1)	-0.0402	-0.0213	3.57E-04	0.0161	-0.0278	1.93E-03	-0.0433	-0.0362	5.09E-05
	ACF(1)	-0.0436	-0.0225	4.43E-04	0.0141	-0.0256	1.57E-03	-0.0312	-0.0290	4.99E-06
$\kappa = 0$	μ_{Y_t}	0.8345	0.8353	9.90E-04	1.0228	1.0202	1.40E-03	1.2485	1.2461	1.26E-03
	$\sigma^2_{Y_t}$	0.8676	0.8710	1.95E-03	1.0801	1.0735	3.64E-03	1.3284	1.3256	4.86E-03
	γ_{Y_t}	1.1491	1.1651	1.36E-02	1.0612	1.0685	9.94E-03	0.9742	0.9828	1.16E-02
	K_{Y_t}	4.3733	4.4685	4.48E-01	4.2108	4.2548	3.31E-01	4.0797	4.0810	3.48E-01
	ACV(1)	-0.0145	-0.0548	1.62E-03	0.0228	0.0016	4.48E-04	0.0121	0.0051	4.84E-05
	ACF(1)	-0.0157	-0.0576	1.75E-03	0.0205	0.0016	3.58E-04	0.0090	0.0039	2.62E-05
$\kappa = 0.5$	μ_{Y_t}	0.6815	0.6884	8.39E-04	1.0241	1.0270	1.09E-03	1.5321	1.5321	1.98E-02
	$\sigma^2_{Y_t}$	0.7090	0.7194	2.45E-03	1.0966	1.0959	3.83E-02	1.6907	1.6855	9.97E-03
	γ_{Y_t}	1.2808	1.2874	1.56E-02	1.0806	1.0876	1.35E-02	0.9383	0.9317	1.44E-02
	K_{Y_t}	4.7548	4.8001	5.45E-01	4.2766	4.3312	3.96E-01	4.0562	4.0247	3.33E-01
	ACV(1)	-0.1088	0.0128	5.59E-04	0.0290	0.0285	2.16E-07	-0.0021	0.0634	4.29E-03
	ACF(1)	-0.0170	0.0179	1.22E-03	0.0259	0.0263	1.37E-07	-0.0013	0.0382	1.56E-03



(a)

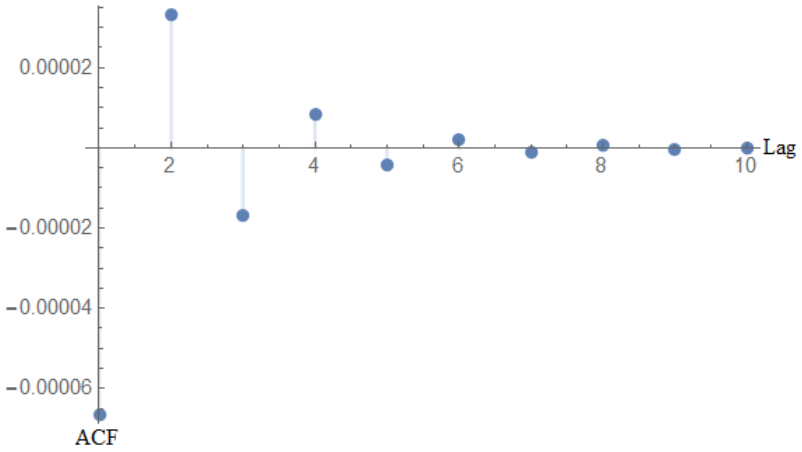


(b)

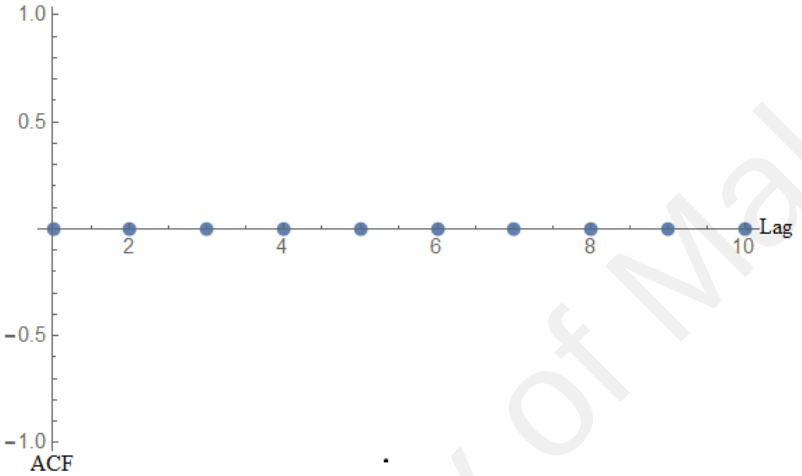


(c)

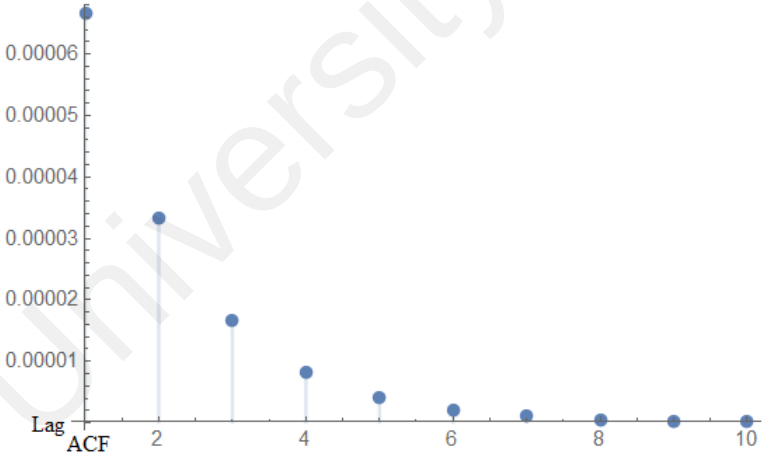
Figure 3.14: ACF plots for the NB model with (a) $\alpha = -0.2$, $\kappa = -0.5$ and $\sigma_\eta = 0.01$, (b) $\alpha = -0.2$, $\kappa = 0$ and $\sigma_\eta = 0.01$, and (c) $\alpha = -0.2$, $\kappa = 0.5$ and $\sigma_\eta = 0.01$.



(a)

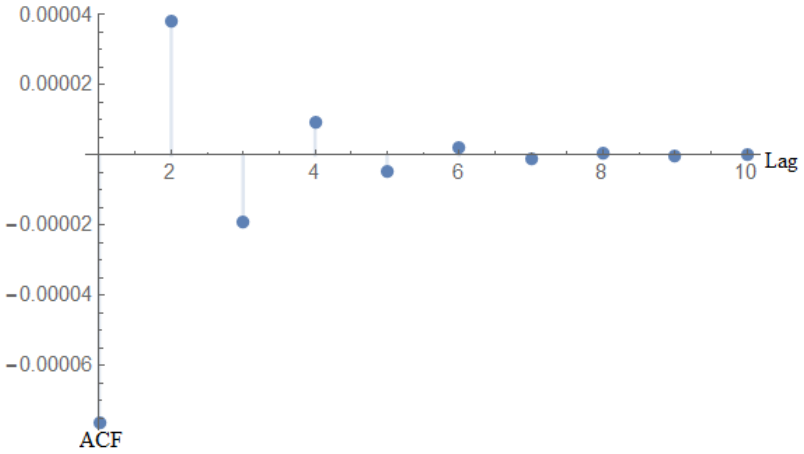


(b)

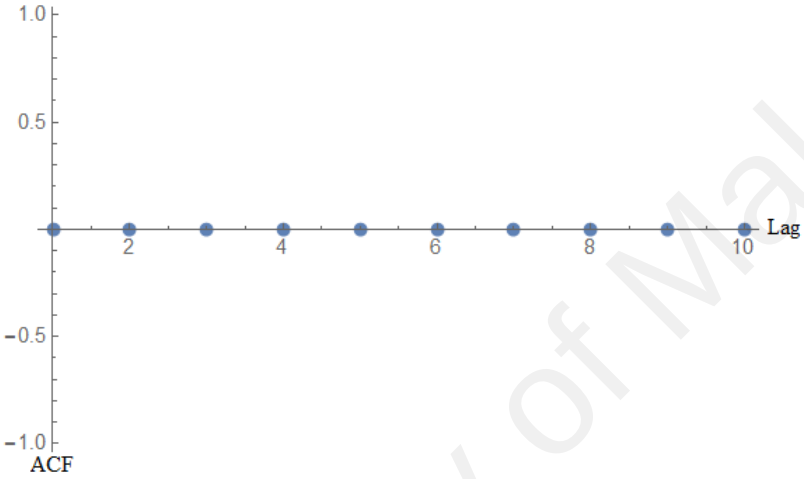


(c)

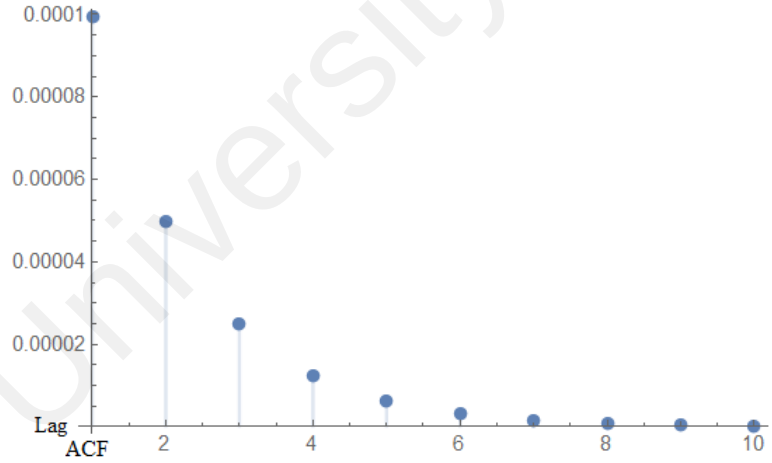
Figure 3.15: ACF plots for the NB model with (a) $\alpha = 0$, $\kappa = -0.5$ and $\sigma_\eta = 0.01$, (b) $\alpha = 0$, $\kappa = 0$ and $\sigma_\eta = 0.01$, and (c) $\alpha = 0$, $\kappa = 0.5$ and $\sigma_\eta = 0.01$.



(a)

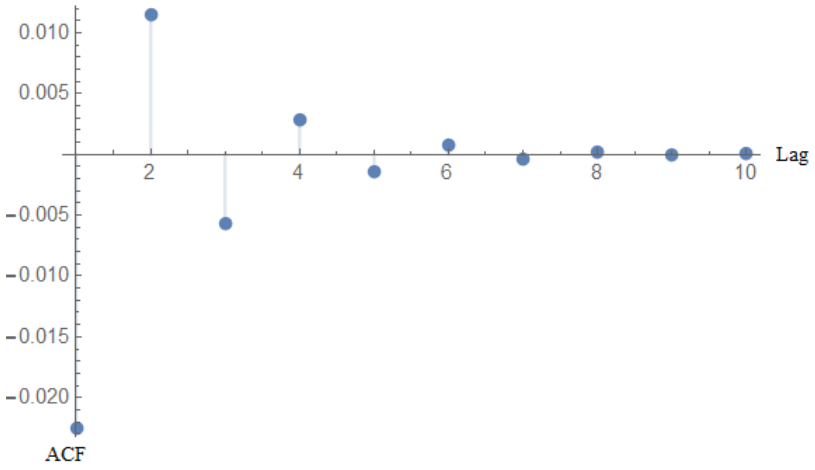


(b)

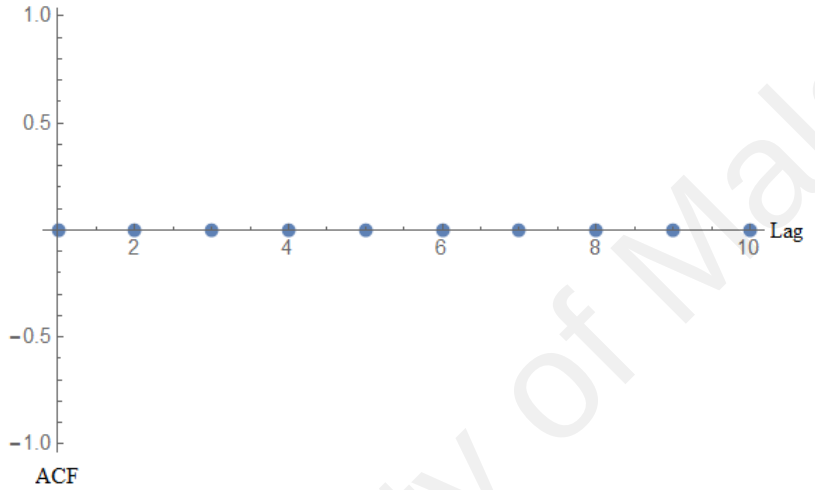


(c)

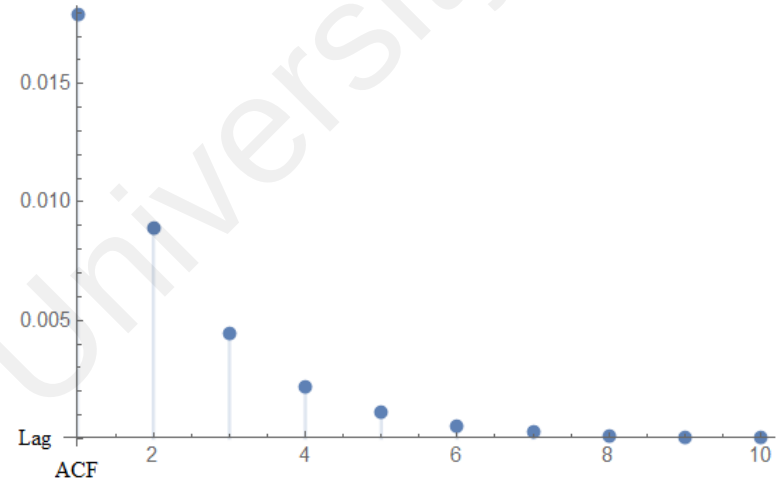
Figure 3.16: ACF plots for the NB model with (a) $\alpha = 0.2$, $\kappa = -0.5$ and $\sigma_\eta = 0.01$, (b) $\alpha = 0.2$, $\kappa = 0$ and $\sigma_\eta = 0.01$, and (c) $\alpha = 0.2$, $\kappa = 0.5$ and $\sigma_\eta = 0.01$.



(a)

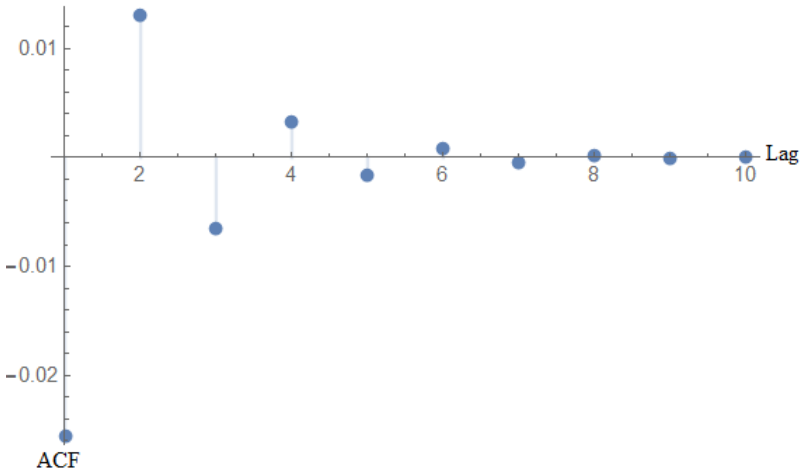


(b)

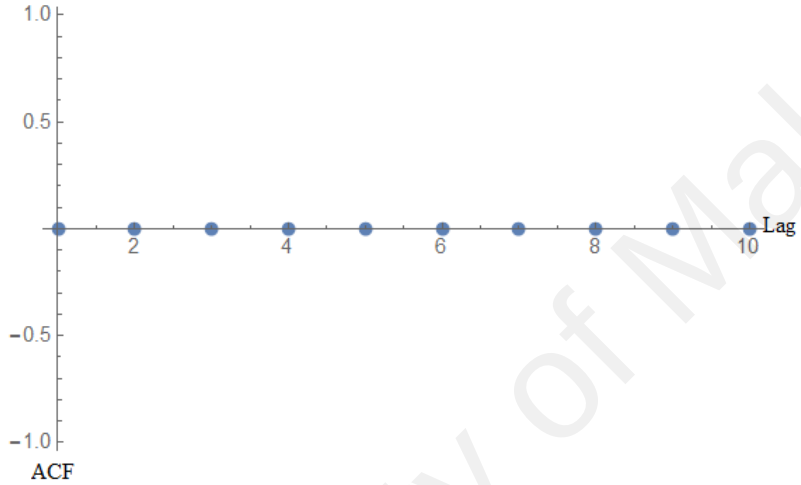


(c)

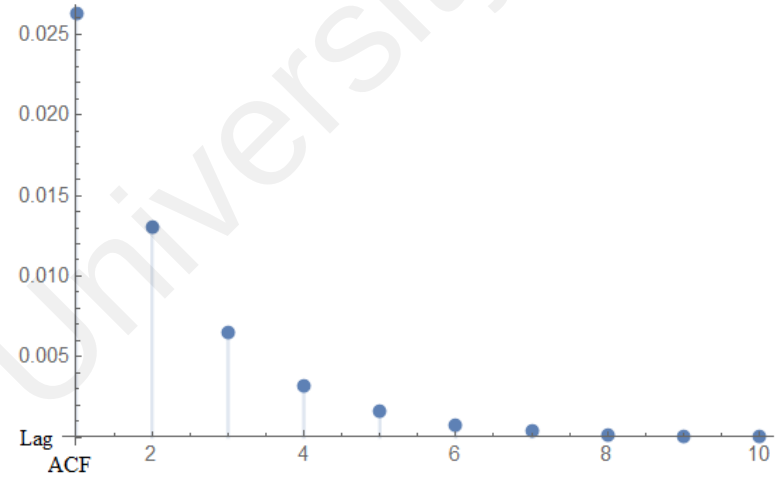
Figure 3.17: ACF plots for the NB model with (a) $\alpha = -0.2$, $\kappa = -0.5$ and $\sigma_\eta = 0.2$, (b) $\alpha = -0.2$, $\kappa = 0$ and $\sigma_\eta = 0.2$, and (c) $\alpha = -0.2$, $\kappa = 0.5$ and $\sigma_\eta = 0.2$.



(a)

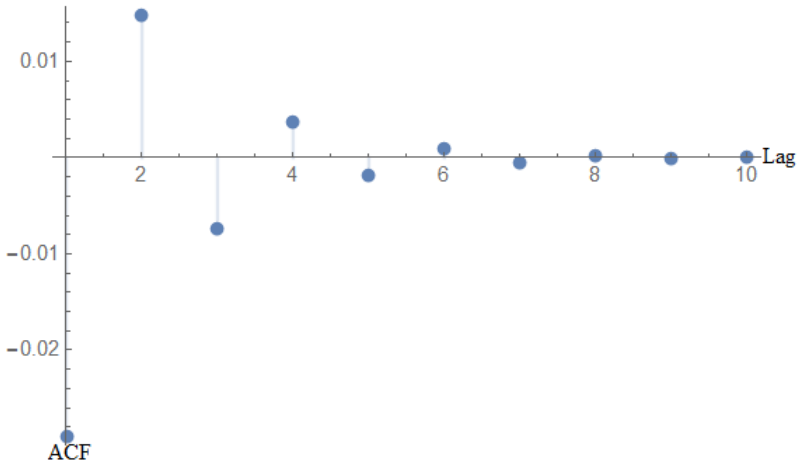


(b)

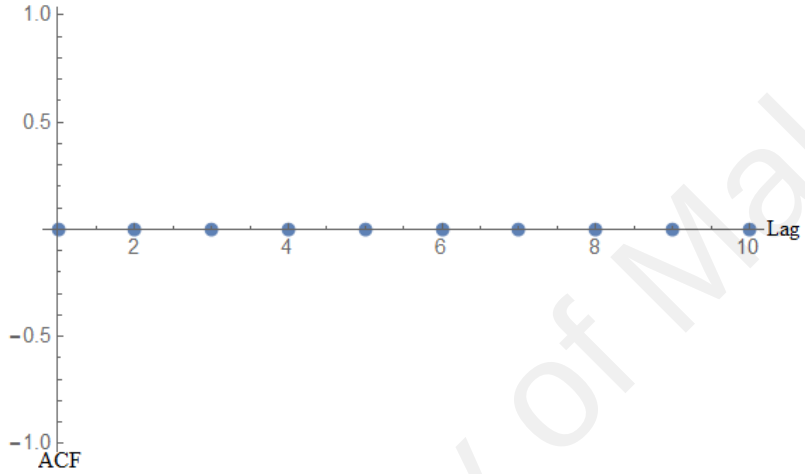


(c)

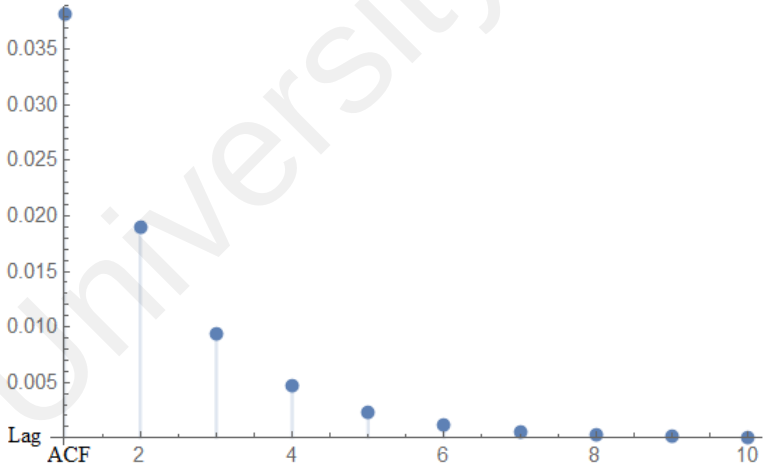
Figure 3.18: ACF plots for the NB model with (a) $\alpha = 0$, $\kappa = -0.5$ and $\sigma_\eta = 0.2$, (b) $\alpha = 0$, $\kappa = 0$ and $\sigma_\eta = 0.2$, and (c) $\alpha = 0$, $\kappa = 0.5$ and $\sigma_\eta = 0.2$.



(a)



(b)



(c)

Figure 3.19: ACF plots for the NB model with (a) $\alpha = 0.2$, $\kappa = -0.5$ and $\sigma_\eta = 0.2$, (b) $\alpha = 0.2$, $\kappa = 0$ and $\sigma_\eta = 0.2$, and (c) $\alpha = 0.2$, $\kappa = 0.5$ and $\sigma_\eta = 0.2$.

3.6 Summary

From the derivations of the moment properties of each dynamic models, the mean of Poisson is equal to the mean of negative binomial. However, from the variance of both models, the negative binomial model is more suitable for an overdispersed data.

From the ACF plot of each dynamic linear models with several parameter values ($\alpha = -0.2, 0$ and 0.2, $\kappa = -0.5, 0$ and 0.5 and $\sigma_\eta = 0.2$ and 1), we can summarize that our derivation fits the theoretical results. For example, if we consider AR(1) with negative coefficient ($\kappa = -0.5$), then the ACF will be dies down exponentially with changing of sign from negative at lag 1, positive lag 2, negative lag 3 and so on.

CHAPTER 4: ESTIMATION

In this chapter, due to time constraint, we focus only on the Poisson parameter-driven model or the dynamic Poisson model. In the previous chapter, we have reviewed the state-space model (Kalman & Bucy, 1961) for normally distributed data. To estimate the model parameters, we devise a Monte Carlo EM (MCEM) algorithm, where particle filtering and smoothing methods (Gordon et al., 1993; Godsill et al., 2004) are employed to approximate high-dimensional integrals.

4.1 MCEM algorithm

Due to the fact that the response distribution is non-Gaussian, the marginal likelihood of the observed data $y_{1:n} = (y_1, ..., y_n)^\intercal$ cannot be expressed empirically. In addition, the random effect, x_t are autocorrelated. Thus, it is troublesome to apply direct maximization of the marginal likelihood. Instead of using gradient-based method (Newton-Raphson), alternatively, we resort to another popular approach for calculating maximum likelihood estimators for models comprise of missing data and unobservable latent variables, the EM algorithm (Dempster et al., 1977).

In the dynamic linear model (DLM), we assume the initial state x_0 is normally distributed with mean μ_0 and variance σ_0^2. The observation y_t conditioning on the current state x_t, defined as Equations (3.10) and (3.11).

Assuming the latent processes $x_{0:n} = (x_0, ..., x_n)^\intercal$ is observable, we decompose the complete data likelihood as follows:

$$L(\boldsymbol{\theta}) = f(x_{0:n}, y_{1:n}) = f(x_{0:n}) f(y_{1:n}|x_{0:n}) = f(x_0) \prod_{t=1}^{n} f(x_t|x_{t-1}) \prod_{t=1}^{n} f(y_t|x_t). \qquad (4.1)$$

Here $\boldsymbol{\theta} = (\alpha, \kappa, \sigma_\eta)^\intercal$ is the vector of unknown parameters.

Due to the orthogonal decomposition in (4.1), the complete data log-likelihood (based on Chan & Ledolter (1995), conditional on the initial latent variable x_0 and up to an additive constant) is given by

$$\log L(\boldsymbol{\theta}) = -\frac{n}{2}\log\left(\sigma_\eta^2\right) - \left[\frac{\sum_{t=1}^{n}\left(x_t - \alpha - \kappa x_{t-1}\right)^2}{2\sigma_\eta^2}\right] - \sum_{t=1}^{n}\left[e^{x_t} - y_t x_t\right].$$

To implement the EM algorithm, we need to compute the expectation of $\log L(\boldsymbol{\theta})$ given the observed data $y_{1:n}$. In the MCEM algorithm, we first compute the E-step of the EM algorithm, where the conditional expectation of $\log L(\boldsymbol{\theta})$ given the observed data $y_{1:n}$, is given by

$$\begin{aligned} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(j)}) &= E\left\{\log L(\boldsymbol{\theta})|y_{1:n}, \boldsymbol{\theta}^{(j)}\right\} \\ &= -\frac{n}{2}\log\left(\sigma_\eta^2\right) - \left[\frac{\sum_{t=1}^{n}\left(A_t^{(j)} - 2\alpha B_t^{(j)} - 2\kappa C_t^{(j)} + 2\alpha\kappa D_t^{(j)} + \alpha^2 + \kappa^2 E_t^{(j)}\right)}{2\sigma_\eta^2}\right] \\ &\quad - \sum_{t=1}^{n}\left[F_t^{(j)} - y_t B_t^{(j)}\right], \end{aligned}$$

where

$$\begin{aligned} A_t^{(j)} &= E\left(x_t^2|y_{1:n}, \boldsymbol{\theta}^{(j)}\right), \\ B_t^{(j)} &= E\left(x_t|y_{1:n}, \boldsymbol{\theta}^{(j)}\right), \\ C_t^{(j)} &= E\left(x_t x_{t-1}|y_{1:n}, \boldsymbol{\theta}^{(j)}\right), \end{aligned}$$

$$D_t^{(j)} = E\left(x_{t-1}|y_{1:n}, \boldsymbol{\theta}^{(j)}\right),$$
$$E_t^{(j)} = E\left([x_{t-1}]^2\,|y_{1:n}, \boldsymbol{\theta}^{(j)}\right),$$
$$F_t^{(j)} = E\left(e^{x_t}|y_{1:n}, \boldsymbol{\theta}^{(j)}\right).$$

Given the state-space representation in Equations (3.10) and (3.11), the M-step of the algorithm is much easier than the direct maximization of the marginal likelihood of $y_{1:n}$. In the M-step, we applied the partial derivatives to maximize $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(j)})$:

$$\frac{\partial Q}{\partial \alpha} = \frac{1}{\sigma_\eta^2}\sum_{t=1}^{n}\left[B_t^{(j)} - \kappa D_t^{(j)} - \alpha\right]$$
$$\frac{\partial Q}{\partial \kappa} = \frac{1}{\sigma_\eta^2}\sum_{t=1}^{n}\left[C_t^{(j)} - \alpha D_t^{(j)} - \kappa E_t^{(j)}\right]$$
$$\frac{\partial Q}{\partial \sigma_\eta} = -\frac{n}{\sigma_\eta} + \frac{1}{\sigma_\eta^3}\sum_{t=1}^{n}\left[A_t^{(j)} - 2\alpha B_t^{(j)} - 2\kappa C_t^{(j)} + 2\alpha\kappa D_t^{(j)} + \alpha^2 + \kappa^2 E_t^{(j)}\right]$$

At the jth iteration, closed-form solutions exist to update $\alpha^{(j+1)}$, $\kappa^{(j+1)}$ and ${\sigma_\eta}^{(j+1)}$:

$$\alpha^{(j+1)} = -\sum_{t=1}^{n}\left(\frac{B_t^{(j)}E_t^{(j)} - C_t^{(j)}D_t^{(j)}}{{D_t^{(j)}}^2 - nE_t^{(j)}}\right)$$
$$\kappa^{(j+1)} = -\sum_{t=1}^{n}\left(\frac{nC_t^{(j)} - B_t^{(j)}D_t^{(j)}}{{D_t^{(j)}}^2 - nE_t^{(j)}}\right)$$
$${\sigma_\eta}^{(j+1)} = \sqrt{\frac{1}{n}\sum_{t=1}^{n} w_t}$$

where $w_t = A_t^{(j)} - 2\alpha^{(j+1)}B_t^{(j)} - 2\kappa^{(j+1)}C_t^{(j)} + 2\alpha^{(j+1)}\kappa^{(j+1)}D_t^{(j)} + \left(\alpha^{(j+1)}\right)^2 + \left(\kappa^{(j+1)}\right)^2 E_t^{(j)}$.

We compute the observed information matrix $\mathbf{I}_o(\boldsymbol{\theta})$ from the MLE obtained through MCEM algorithm. According to the missing information principle from Louis's formula (Louis, 1982):

$$\mathbf{I}_o(\boldsymbol{\theta}) = \mathbf{I}_c(\boldsymbol{\theta}) - \mathbf{I}_m(\boldsymbol{\theta}), \tag{4.2}$$

in which the complete information matrix $\mathbf{I}_c(\boldsymbol{\theta})$ and the missing data information matrix $\mathbf{I}_m(\boldsymbol{\theta})$ are defined as follows:

$$\mathbf{I}_c(\boldsymbol{\theta}) = E\left(-\frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}|y_{1:n}\right),$$
$$\mathbf{I}_m(\boldsymbol{\theta}) = E\left(-\frac{\partial \log L(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial \log L(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^\top}|y_{1:n}\right) - E\left(\frac{\partial \log L(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}|y_{1:n}\right)E\left(\frac{\partial \log L(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^\top}|y_{1:n}\right).$$

The first order derivatives of $\log L(\boldsymbol{\theta})$are given by:

$$\frac{\partial \log L(\boldsymbol{\theta})}{\partial\alpha} = \frac{\sum_{t=1}^{n}(x_t - \alpha - \kappa x_{t-1}^{(i)})}{\sigma_\eta^2},$$
$$\frac{\partial \log L(\boldsymbol{\theta})}{\partial\kappa} = \frac{\sum_{t=1}^{n}(x_{t-1}^{(i)})(x_t - \alpha - \kappa x_{t-1}^{(i)})}{\sigma_\eta^2,}$$
$$\frac{\partial \log L(\boldsymbol{\theta})}{\partial\sigma_\eta} = \frac{-n}{\sigma_\eta} + \frac{\sum_{t=1}^{n}(x_t - \alpha - \kappa x_{t-1}^{(i)})^2}{\sigma_\eta^3},$$

and the second-order derivatives are:

$$\frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial\alpha\partial\alpha} = -\frac{1}{\sigma_\eta^2},$$
$$\frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial\kappa\partial\kappa} = \frac{-\sum_{t=1}^{n}\left[x_{t-1}^{(i)}\right]^2}{\sigma_\eta^2},$$
$$\frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial\sigma_\eta\partial\sigma_\eta} = \frac{n}{\sigma_\eta^2} - \frac{3\sum_{t=1}^{n}(x_t - \alpha - \kappa x_{t-1}^{(i)})^2}{\sigma_\eta^4},$$

$$\frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial\alpha\partial\kappa} = \frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial\kappa\partial\alpha} = \frac{-\sum_{t=1}^{n} x_{t-1}^{(i)}}{\sigma_\eta^2},$$

$$\frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial\alpha\partial\sigma_\eta} = \frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial\sigma_\eta\partial\alpha} = \frac{-2\sum_{t=1}^{n}(x_t - \alpha - \kappa x_{t-1}^{(i)})}{\sigma_\eta^3},$$

$$\frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial\kappa\partial\sigma_\eta} = \frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial\sigma_\eta\partial\kappa} = \frac{-2\sum_{t=1}^{n}(x_{t-1}^{(i)})(x_t - \alpha - \kappa x_{t-1}^{(i)})}{\sigma_\eta^3}.$$

In the next section, we approximate the conditional expectations in $\mathbf{I}_c(\boldsymbol{\theta})$ and $\mathbf{I}_m(\boldsymbol{\theta})$ with the used of particle methods.

The stopping rule based on log-likelihood change is dubious, which differ from the exact EM algorithm for linear and Gaussian state-space models. It is not guaranteed to increase at each iteration (Shumway & Stoffer, 1982). Hence, we require to analyse the trace plots of parameter estimates for stabilization.

4.2 Particle methods

We applied the particle filtering on our dynamic Poisson model by generating $x_{0|0}^{(i)} \sim N(\mu_0, \sigma_0)$ for a given μ_0 and σ_0. Next, we perform the following steps for $t = 1, ..., n$ to produce a set of N particles ($i = 1, 2, ..., N$) at each time point t.

4.2.1 Particle filtering

(F.1) Generate $x_{t|t-1}^{(i)} \sim N(\alpha + \kappa x_{t-1|t-1}, \sigma_\eta^2)$

(F.2) Compute the filtering weights

$$q_{t|t-1}^{(i)} = \frac{\left[\lambda_{t|t-1}^{(i)}\right]^{y_t} \exp\left\{-\lambda_{t|t-1}^{(i)}\right\}}{y_t!}$$

where $\lambda_{t|t-1}^{(i)} = e^{x_{t|t-1}^{(i)}}$.

(F.3) Based on the preceding filtering weights, generate $x_{t|t}^{(i)}$ by resampling $x_{t|t-1}^{(i)}$ with replacement.

As a by product of the filtering algorithm, the likelihood of the observed data can be calculated from

$$\sum_{t=1}^{n}\left(\frac{1}{N}\sum_{i=1}^{N}q_{t|t-1}^{(i)}\right).$$

4.2.2 Particle smoothing

In the particle smoothing step, we first choose $x_{n|n}^{(r)} = x_{n|n}^{(i)}$ with probability $q_{n|n-1}^{(i)}$ for $r = 1, 2, ..., R$. Then, for $t = n-1, n-2, ..., 1$:

(S.1) Calculate the smoothing weights

$$q_{t|n}^{(i)} \propto q_{t|t-1}^{(i)} \exp\left\{-\frac{1}{2\sigma_\eta^2}\left(x_{t+1|n}^{(i)} - \alpha - \kappa x_{t|t}^{(i)}\right)^2\right\}$$

(S.2) Choose $x_{t|n}^{(r)} = x_{t|t}^{(i)}$ with probability $q_{t|n}^{(i)}$.

We repeat the the preceding steps for $r = 1, 2, ..., R$. The non-linear and non-Gaussian extensions of the Kalman filter and the Kalman smoother are known as the forward-filtering and backward-smoothing procedures.

4.3 A simulation study

A simulation study was conducted to evaluate the performance of the MCEM algorithm. We consider time series data of length 500, which are simulated from the Poisson model stated in Equations (3.22) and (3.23). Selected values for α, κ and σ_η has been chosen based on stationary process.

When fitting the model, the number of particles, N, and the number of replications, R, are chosen to be 500. The algorithm is stopped after 500 iterations. Tables 4.1, 4.2

and 4.3 provide the mean of the last 100 parameter estimate values and the empirical standard errors of the estimates (in parentheses). The biases of the estimates are close to zero indicating the good performance of the estimators proposed in this study.

Table 4.1: Parameter estimates (standard errors) for the proposed model with $\sigma_\eta = 0.01$

			α			
			-0.2	0	0.2	1.0
κ	-0.5	α	-0.1997 (0.0050)	0.0005 (0.0005)	0.2026 (0.0054)	1.0029 (0.0259)
		κ	-0.4958 (0.0383)	-0.5149 (0.0386)	-0.4877 (0.0391)	-0.4983 (0.0387)
		σ_η	0.0100 (0.0003)	0.0099 (0.0003)	0.0100 (0.0003)	0.0100 (0.0003)
	0	α	-0.1936 (0.0086)	0.0070 (0.0006)	0.2040 (0.0093)	1.0030 (0.0452)
		κ	-0.0023 (0.0447)	0.0020 (0.0447)	0.0133 (0.0446)	0.0020 (0.0449)
		σ_η	0.0100 (0.0003)	0.0102 (0.0003)	0.0101 (0.0003)	0.0099 (0.0003)
	0.5	α	-0.1827 (0.0140)	0.0134 (0.0013)	0.2052 (0.0162)	0.9774 (0.0768)
		κ	0.5087 (0.0380)	0.5299 (0.0375)	0.5191 (0.0377)	0.5117 (0.0384)
		σ_η	0.0101 (0.0003)	0.0101 (0.0003)	0.0101 (0.0003)	0.0100 (0.0003)
	0.9	α	-0.1277 (0.0077)	0.0534 (0.0042)	0.0221 (0.0135)	0.1762 (0.0135)
		κ	0.9064 (0.0058)	0.9078 (0.0066)	0.9892 (0.0067)	0.9892 (0.0067)
		σ_η	0.0097 (0.0004)	0.0098 (0.0003)	0.0113 (0.0003)	0.0113 (0.0004)

Table 4.2: Parameter estimates (standard errors) for the proposed model with $\sigma_\eta = 0.2$

			α			
			-0.2	0	0.2	1.0
κ	-0.5	α	-0.0230 (0.0870)	-0.0851 (0.0869)	0.1888 (0.0865)	1.0296 (0.0860)
		κ	-0.5148 (0.0435)	-0.5639 (0.0435)	-0.3225 (0.0432)	-0.5035 (0.0431)
		σ_η	0.3426 (0.0120)	0.3048 (0.0119)	0.1016 (0.0114)	0.2147 (0.0105)
	0	α	-0.2704 (0.0720)	-0.0545 (0.0722)	0.2231 (0.07212)	0.9530 (0.0702)
		κ	0.0953 (0.0357)	0.0645 (0.0358)	0.0914 (0.03577)	0.0173 (0.0348)
		σ_η	0.2487 (0.0081)	0.2505 (0.0082)	0.2453 (0.0081)	0.2321 (0.0076)
	0.5	α	-0.2808 (0.0601)	-0.0564 (0.0598)	0.1856 (0.0610)	1.0529 (0.0599)
		κ	0.6914 (0.0298)	0.5559 (0.0296)	0.4455 (0.0302)	0.4799 (0.0296)
		σ_η	0.1917 (0.0061)	0.1934 (0.0062)	0.1916 (0.0062)	0.2072 (0.0059)
	0.9	α	-0.1658 (0.0546)	-0.0055 (0.0567)	0.1382 (0.0548)	1.0202 (0.0561)
		κ	0.9141 (0.0270)	0.9225 (0.0280)	0.9299 (0.0277)	0.8976 (0.0387)
		σ_η	0.2148 (0.0056)	0.1702 (0.0055)	0.2050 (0.0053)	0.1992 (0.0042)

Table 4.3: Parameter estimates (standar errors) for the proposed model with $\sigma_\eta = 1$

			α			
			-0.2	0	0.2	1.0
κ	-0.5	α	-0.2406 (0.0770)	-0.0835 (0.0774)	0.2599 (0.0776)	0.8982 (0.0760)
		κ	-0.3911 (0.0382)	-0.2901 (0.0383)	-0.6098 (0.0384)	-0.5502 (0.0376)
		σ_η	1.0784 (0.0089)	1.1288 (0.0089)	1.0723 (0.0089)	1.0793 (0.0086)
	0	α	-0.3546 (0.0066)	-0.0122 (0.0066)	0.3502 (0.0056)	0.9582 (0.0067)
		κ	-0.0010 (0.0049)	0.333 (0.0050)	0.0279 (0.0052)	0.0649 (0.0042)
		σ_η	1.0328 (0.0095)	1.0789 (0.0071)	0.9981 (0.0065)	1.0222 (0.0062)
	0.5	α	-0.1933 (0.0749)	0.0280 (0.0743)	0.2097 (0.0747)	1.1293 (0.0736)
		κ	0.5728 (0.0371)	0.4580 (0.0367)	0.5758 (0.0370)	0.5831 (0.0364)
		σ_η	1.0580 (0.0085)	1.0104 (0.0084)	0.9639 (0.0084)	1.1442 (0.0083)
	0.9	α	-0.0773 (0.0056)	-0.0356 (0.0019)	0.3384(0.0031)	
		κ	0.9141(0.0032)	0.9024 (0.0020)	0.8561 (0.0013)	
		σ_η	0.9491 (0.0094)	0.9633 (0.0075)	1.0136 (0.0051)	

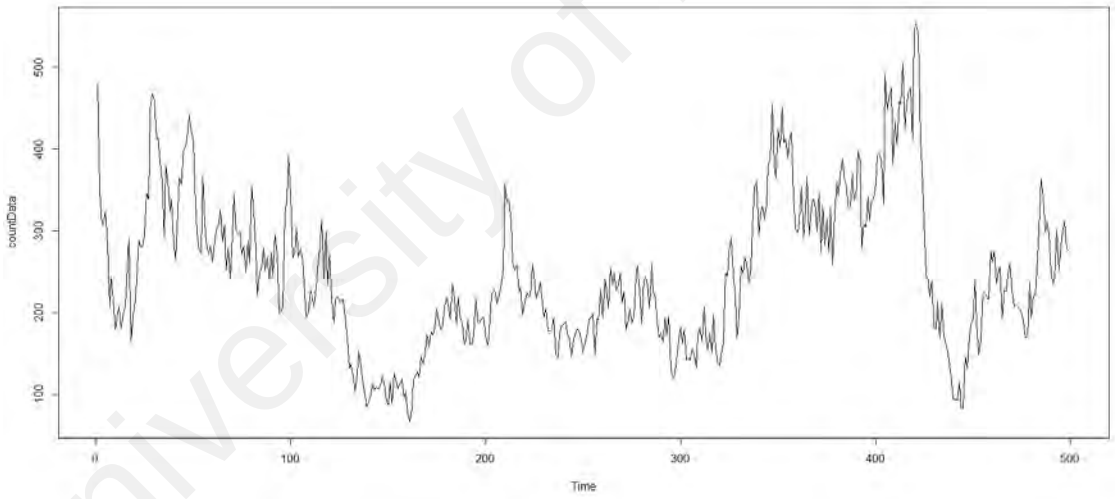


Figure 4.1: Time series plot for $\alpha = 0.2$, $\kappa = 0.9$ and $\sigma_\eta = 0.1$.

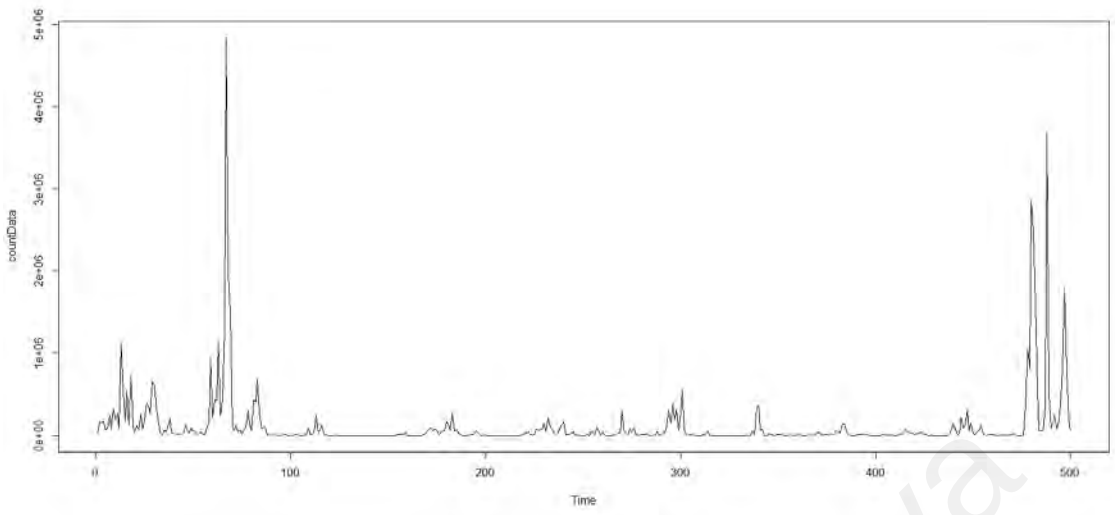


Figure 4.2: Time series plot for $\alpha = 1$, $\kappa = 0.9$ and $\sigma_\eta = 1$.

An error occured when we set $\alpha = 1$, $\kappa = 0.9$ and $\sigma_\eta = 1$ as our true values. Figure 4.1 shows the time series plot for $\alpha = 0.2$, $\kappa = 0.9$ and $\sigma_\eta = 0.1$ while Figure 4.2 is the time series plot for the error output. The plot in Figure 4.2 shows several unusual peaks with an extreme value (up to 5×10^6). This lead to the fact that there is too few positive probabilities in the sample, and our method unable to solve such problem.

4.4 Summary

Based on our simulation results, the estimates value obtained are close to the true value with small standard error. For well-behaved time series data, we found that the proposed model works well and should be considered in future works.

CHAPTER 5: APPLICATION AND DISCUSSION

5.1 Data

We consider the Malaysia dengue data to illustrate our methodology. According to iDengue remote sensing website, the number of dengue cases in Malaysia has increases rapidly in 2014 (108,698 cases) and 2015 (120,836 cases) compare to that in 2012 (21,900 cases) and 2013 (43,346 cases).

Dengue fever is a mosquito-borne tropical disease which is passed on by several species of mosquito, mostly of the *Aedes* type. A single mosquito can produce 90 million mosquitoes in 60 days. Dengue is transferred in two ways: i) when a mosquito bites an infected person, then bites another; and ii) when an infected mosquito lays eggs, passing the disease to its offspring.

According to World Health Organization, 40% of the world's population is at risk from dengue. That is over 2.5 billion people. Dengue is the leading source of serious illness and death among children in parts of Asia and Latin America. There is currently no treatment for dengue. However, early detection and access to proper medical care can save lives.

For this study, our focus is on the most populated state in Malaysia (Selangor), the capital city of Malaysia (Wilayah Persekutuan Kuala Lumpur) and Wilayah Persekutuan Putrajaya. In 2010, Selangor recorded 13,086 cases from 38,943 cases in Malaysia. The number of reported cases however decrease to 7,799 and 9,113 in 2011 and 2012 respectively. In 2013 and 2014, 50% of the total cases reported occur in Selangor. This data consists of weekly counts of dengue cases from February 2010 to September 2013 and the time series plot of the data is shown in Figure 5.1.

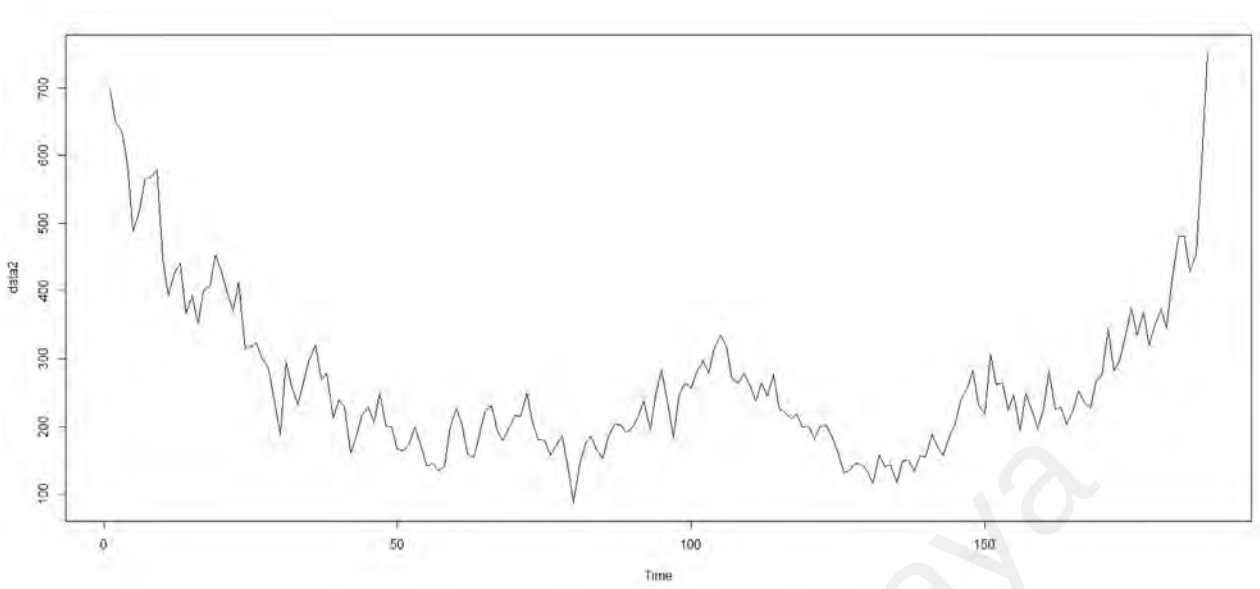


Figure 5.1: The time series plot of total deangue cases in Selangor, Kuala Lumpur and Putrajaya.

Working on the summary statistics, the data has mean of 264.777 cases per week with variance of 14091.400. While the skewness and kurtosis are 1.600 and 5.732 respectively.

5.2 Application and discussion

When fitting the data to the dynamic Poisson model, we set the number of particles, N and the number of replications, R equals to 500 and we stopped the algorithm after 500 iterations.

Since the propose model is non-Gaussian and non-linear model, in order to solve the initial values problem, we ran the program several times and each time with the adjusted initial values. The last fitted value of the model parameters is then used as an adjusted initial value for the next run. We found that the adjusted initial value improved the performance of the estimation in term of the smallest AIC and loglikelihood. Therefore, we chose $\alpha = 0.17$, $\kappa = 0.97$ and $\sigma_\eta = 0.11$ as our final initial value which yields the result in Table 5.1.

Table 5.1: Parameter estimates with several models fit to the Malaysia dengue data.

	Std. Po	Std. NB	Dzim(Po)	Proposed Model
Mean	264.777	264.777		
	(1.1867)	(7.7525)		
r		6.354		
		(0.6540)		
β_0			5.592925	
ϕ_1			0.958195	
α				0.219069
				(0.1157)
κ				0.960257
				(0.0209)
σ			0.112546	0.1125604
				(0.0062)
AIC	10081.698	2263.014	1879.018	1878.187
Log-likehood	-5041.849	-1133.507	-936.5089	-936.0937

Table 5.1 shows the comparison of parameter estimates for four models. The second column shows the results for standard Poisson model and followed by standard negative binomial model in column three. The fourth column is the output from Yang et al. (2015) model and the final column shows the results from dynamic Poisson model. The standard errors are stated in the bracket under each parameters. However, we are unable to calculate the standard errors for Yang et al. (2015) model. Among these four models, our proposed model has the smallest AIC and loglikelihood.

By substituting the α, κ and σ_η values in Table 5.1, we can solve Equations (3.14) to (3.17). The comparisons between the empirical values and the estimated values for the moment properties of the data set is shown in Table 5.2. Since Yang et al. (2015) did not provide any moment properties, hence we are unable to calculate the moment properties for the particular model.

Table 5.2: Comparisons between the empirical values and the estimated values of the moment properties of Malaysia dengue data.

	Empirical value	Estimated value		
		Std. Po	Std. NB	Proposed Model
Mean	264.777	264.777	264.777	268.648
Variance	14091.400	264.777	11298.3	13014
Skewness	1.600	0.0615	0.793482	1.31973
Kurtosis	5.732	3.0038	3.94438	6.26001

From the results in Table 5.2, the moment properties of our proposed model is the closest to the empirical value of the data than the other standard models. Since the estimated mean of a Poisson and negative binomial distribution based on MLE is the sample mean, hence that explains why the estimated mean of the standard Poisson and standard negative binomial are equal to the empirical mean. In our case, although the mean of the proposed model is slightly different, the variance, skewness and kurtosis of our model are better than the other two models. This shows that the proposed model is a better fitted model for the Malaysia dengue data.

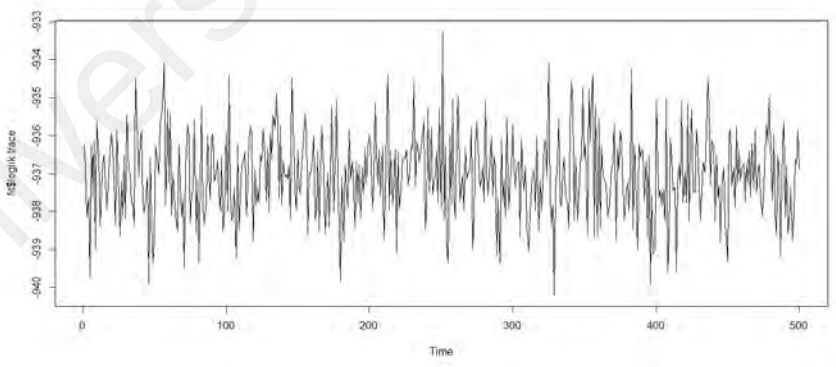


Figure 5.2: Trace plot of log-likelihood for proposed model fit to dengue data.

As illustrated in Figure 5.2, our loglikelihood stabilize around -933 and -940 as it approach MLE. The stopping rule of MCEM algorithm has been discussed in Section 3.

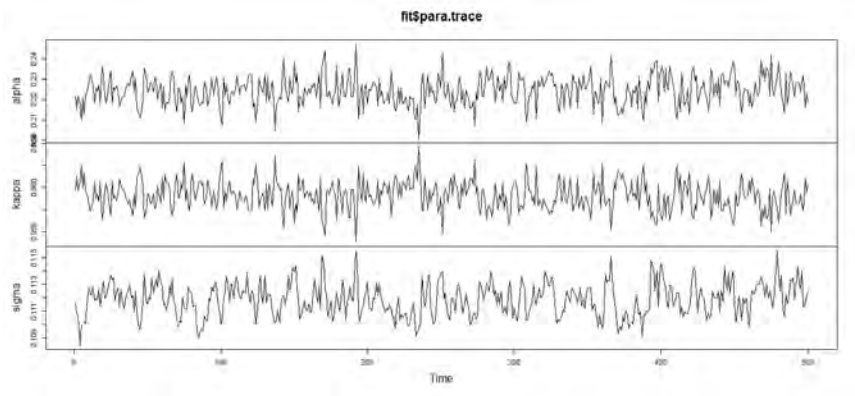


Figure 5.3: Trace plots of scaled changes in parameter estimates from starting values.

Trace plots illustrating the changes for the parameter estimated are presented in Figure 5.3. Note that our parameters are still changing substantially within a reasonable tolerance level. In Figure 5.3, the initial value are $\alpha = 0.17$, $\kappa = 0.97$ and $\sigma = 0.11$, which is the adjusted final initial value.

5.3 Summary

In this chapter, we implemented our proposed model on Malaysia dengue data. The modelling procedures are compared with that of Yang et al. (2015). Our model is simpler and easier to employ. Furthermore, we obtained a better output with smaller AIC and closest moment statistical properties compared to the other models.

CHAPTER 6: CONCLUSION AND FURTHER RESEARCH

6.1 Conclusion

In this research, we suggested a class of state-space models to model the time series of counts. All simulated and real examples were presented to exhibit the proposed methodology.

Based on the results in the simulation study (Chapter 3):

1. The autocorrelations function are positive if $\kappa > 0$.
2. The autocorrelations function are positive if the parameter restriction $\kappa < 0$ and i is an integer hold.
3. The autocorrelation functions is dominated by an exponential decay as $t \to \infty$ for Property 1 and 2.
4. When i is an odd integer for $k < 0$, the autocorrelaton function oscillates and decay quickly with increasing t.

We implemented the modelling procedures introduced in Chapter 4 and compared it with that of Yang et al. (2015) in Chapter 5. We found that, our proposed model was not only simpler and easier to employ, but also gain a better output with smaller AIC.

Based on the application results on the Malaysia dengue data in Chapter 5, our model has the smallest AIC and closest moment statistical properties when compared to three other models.

6.2 Further Research

We have obtained the characterization of the first four moments for three parameter-driven models with AR(1) latent process. With these results, we are able to see what an estimated parameter-driven model implies. These results can be extended to other parameter-driven model with different latent stochastic process.

When dealing with a non-linear model, the application of an effective stopping criteria for the MCEM algorithm is demanding. There are several stopping criteria generally used in literature. For example, when the score vector is close to zero or when the log-likelihood starts to converge, we can terminate the algorithm. In future work, the expansion of a better and automated stopping rule is an objective that can be considered.

We wish to applied the parameter estimation of these three models with an application on a real life data set. With the rise of dengue and Zika cases in Malaysia recently, we can fit this model to the current data in order that more drastic measures may be taken. Further research on this issue is needed.

REFERENCES

Abraham, B., & Ledolter, J. (2009). *Statistical methods for forecasting* (Vol. 234). John Wiley & Sons.

Ahmad, A., & Francq, C. (2016). Poisson QMLE of count time series models. *Journal of Time Series Analysis*, *37*(3), 291–314.

Al-Osh, M., & Alzaid, A. A. (1987). First-order integer-valued autoregressive (INAR (1)) process. *Journal of Time Series Analysis*, *8*(3), 261–275.

Bauwens, L., & Veredas, D. (2004). The stochastic conditional duration model: A latent variable model for the analysis of financial durations. *Journal of Econometrics*, *119*(2), 381–412.

Besag, J., & Green, P. J. (1993). Spatial statistics and Bayesian computation. *Journal of the Royal Statistical Society: Series B (Methodological)*, *55*(1), 25–37.

Bhattacharyya, M., & Layton, A. P. (1979). Effectiveness of seat belt legislation on the Queensland road toll—an Australian case study in intervention analysis. *Journal of the American Statistical Association*, *74*(367), 596–603.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*(3), 307–327.

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.

Cameron, A. C., & Trivedi, P. K. (2013). *Regression analysis of count data* (Vol. 53). Cambridge university press.

Catania, L., & Nonejad, N. (2016). Density forecasts and the leverage effect: Some evidence from observation and parameter-driven volatility models. *arXiv preprint arXiv:1605.00230*.

Chan, K., & Ledolter, J. (1995). Monte Carlo EM estimation for time series models involving counts. *Journal of the American Statistical Association*, *90*(429), 242–252.

Cox. (1981). Statistical analysis of time series: Some recent developments. *Scandinavian Journal of Statistics*, *8*(2), 93–115.

Creal, D., Koopman, S. J., & Lucas, A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, *28*(5), 777–795.

Davis, R., Dunsmuir, W. T., & Wang, Y. (1999). Modeling time series of count data. *Statistics Textbooks and Monographs*, *158*, 63–114.

Davis, R., Dunsmuir, W. T., & Wang, Y. (2000). On autocorrelation in a Poisson regression model. *Biometrika*, *87*(3), 491–505.

Davis, R., & Liu, H. (2012). Theory and inference for a class of nonlinear models with application to time series of counts. *Statistica Sinica*, *26*, 1673–1707.

Davis, R., & Wu, R. (2009). A negative binomial model for time series of counts. *Biometrika*, *96*(3), 735–749.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (methodological)*, *39*(1), 1–38.

Doucet, A., De Freitas, N., & Gordon, N. (2001). *Sequential Monte Carlo methods in practice*. Springer.

Doucet, A., & Johansen, A. M. (2009). A tutorial on particle filtering and smoothing: Fifteen years later. *Handbook of Nonlinear Filtering*, *12*(3), 656–704.

Drovandi, C. C., & McCutchan, R. A. (2016). Alive SMC2: Bayesian model selection for low-count time series models with intractable likelihoods. *Biometrics*, *72*(2), 344–353.

Dunsmuir, W., & He, J. (2017). Marginal estimation of parameter driven Binomial time series models. *Journal of Time Series Analysis*, *38*(1), 120–144.

Durbin, J., & Koopman, S. J. (2000). Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *62*(1), 3–56.

Durbin, J., & Koopman, S. J. (2012). *Time series analysis by state space methods* (Vol. 38). Oxford University Press.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*, *50*(4), 987–1007.

Engle, R. F., & Russell, J. R. (1998). Autoregressive conditional duration: A new model for irregularly spaced transaction data. *Econometrica*, *66*(5), 1127–1162.

Fahrmeir, L., & Tutz, G. (2001). *Multivariate statistical modelling based on generalized linear models*. Springer.

Feigin, P. D., Gould, P., Martin, G. M., & Snyder, R. D. (2008). Feasible parameter regions for alternative discrete state space models. *Statistics & Probability Letters*, *78*(17), 2963–2970.

Fokianos, K. (2001). Truncated poisson regression for time series of counts. *Scandinavian Journal of Statistics*, *28*(4), 645–659.

Fokianos, K., & Kedem, B. (2002). *Regression model for time series analysis.* Wiley Interscience.

Frühwirth-Schnatter, S., & Wagner, H. (2006). Auxiliary mixture sampling for parameter-driven models of time series of counts with applications to state space modelling. *Biometrika*, *93*(4), 827–841.

Godsill, S. J., Doucet, A., & West, M. (2004). Monte Carlo smoothing for nonlinear time series. *Journal of the American Statistical Association*, *99*(1), 156–168.

Goodrich, R., & Caines, P. (1979). Linear system identification from nonstationary cross-sectional data. *IEEE Transactions on Automatic Control*, *24*(3), 403–411.

Gordon, N. J., Salmond, D. J., & Smith, A. F. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings F (Radar and Signal Processing)*, *140*(2), 107–113.

Guo, S. W., & Thompson, E. A. (1991). Monte Carlo estimation of variance component models for large complex pedigrees. *Mathematical Medicine and Biology*, *8*(3), 171–189.

Gupta, N., & Mehra, R. (1974). Computational aspects of maximum likelihood estimation and reduction in sensitivity function calculations. *IEEE Transactions on Automatic Control*, *19*(6), 774–783.

Hansen, P. R., & Lunde, A. (2005). A forecast comparison of volatility models: does anything beat a garch (1, 1)? *Journal of applied econometrics*, *20*(7), 873–889.

Harrison, J., & West, M. (1999). *Bayesian forecasting & dynamic models*. Springer New York.

Harvey, A. C. (2013). *Dynamic models for volatility and heavy tails: with applications to financial and economic time series* (Vol. 52). Cambridge University Press.

Harvey, A. C., & Durbin, J. (1986). The effects of seat belt legislation on British road casualties: A case study in structural time series modelling. *Journal of the Royal Statistical Society. Series A (General)*, *149*(3), 187–227.

Harvey, A. C., & Fernandes, C. (1989). Time series models for count or qualitative observations. *Journal of Business & Economic Statistics*, *7*(4), 407–417.

Harvey, A. C., & Phillips, G. D. (1979). Maximum likelihood estimation of regression models with autoregressive-moving average disturbances. *Biometrika*, *66*(1), 49–58.

Hasan, M. T., Huda, S., & Sneddon, G. (2016). A comparative study of observation- and parameter-driven zero-inflated Poisson models for longitudinal count data. *Communications in Statistics-Simulation and Computation*, *45*(10), 3643–3659.

Hay, J. L., Pettitt, A. N., et al. (2001). Bayesian analysis of a time series of counts with covariates: An application to the control of an infectious disease. *Biostatistics*, *2*(4), 433–444.

Jacquier, E., Polson, N. G., & Rossi, P. E. (2002). Bayesian analysis of stochastic volatility models. *Journal of Business & Economic Statistics*, *20*(1), 69–87.

Johansen, A. M. (2009). SMCTC: Sequential Monte Carlo in C++. *Journal of Statistical Software*, *30*(6), 1–41.

Johansson, P. (1996). Speed limitation and motorway casualties: a time series count data regression approach. *Accident Analysis & Prevention*, *28*(1), 73–87.

Jones, R. H. (1980). Maximum likelihood fitting of ARMA models to time series with missing observations. *Technometrics*, *22*(3), 389–395.

Jung, R. C., Kukuk, M., & Liesenfeld, R. (2006). Time series of count data: Modeling, estimation and diagnostics. *Computational Statistics & Data Analysis*, *51*(4), 2350–2364.

Jung, R. C., & Liesenfeld, R. (2001). Estimating time series models for count data using efficient importance sampling. *Advances in Statistical Analysis*, *4*(85), 387–407.

Kalman, R. E., & Bucy, R. S. (1961). New results in linear filtering and prediction theory. *Journal of Basic Engineering*, *83*(1), 95–108.

Kedem, B., & Fokianos, K. (2005). *Regression models for time series analysis* (Vol. 488). John Wiley & Sons.

Kim, S., Shephard, N., & Chib, S. (1998). Stochastic volatility: Likelihood inference and comparison with ARCH models. *The Review of Economic Studies*, *65*(3), 361–393.

Kim, S., & Stoffer, D. S. (2008). Fitting stochastic volatility models in the presence of irregular sampling via particle methods and the EM algorithm. *Journal of Time Series Analysis*, *29*(5), 811–833.

Koopman, S. J., Lucas, A., & Scharth, M. (2016). Predicting time-varying parameters with parameter-driven and observation-driven models. *Review of Economics and Statistics*, *98*(1), 97–110.

Kuk, A. Y., & Cheng, Y. W. (1997). The Monte Carlo Newton-Raphson algorithm. *Journal of Statistical Computation and Simulation*, *59*(3), 233–250.

Ledolter, J. (1979). A recursive approach to parameter estimation in regression and time series models. *Communications in Statistics-Theory and Methods*, *8*(12), 1227–1245.

Leroux, B. G., & Puterman, M. L. (1992). Maximum-penalized-likelihood estimation for independent and Markov-dependent mixture models. *Biometrics*, *48*(2), 545–558.

Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society; Series B (Methodological)*, *44*(2), 226–233.

MacDonald, I. L., & Zucchini, W. (1997). *Hidden Markov and other models for discrete-valued time series* (Vol. 110). CRC Press.

McKenzie, E. (1988). Some ARMA models for dependent sequences of Poisson counts. *Advances in Applied Probability*, *20*(4), 822–835.

McKenzie, E. (2003). Ch. 16. Discrete variate time series. *Handbook of Statistics*, *21*, 573–606.

Michener, R., & Tighe, C. (1992). A Poisson regression model of highway fatalities. *The American Economic Review*, *82*(2), 452–456.

Möller, T. A., Weiß, C. H., Kim, H.-Y., & Sirchenko, A. (2018). Modeling zero inflation in count data time series with bounded support. *Methodology and Computing in Applied Probability*, *20*(2), 589–609.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society*, *59*(2), 347–370.

Oh, M.-S., & Lim, Y. B. (2001). Bayesian analysis of time series Poisson data. *Journal of Applied Statistics*, *28*(2), 259–271.

Pascual, L., Romo, J., & Ruiz, E. (2006). Bootstrap prediction for returns and volatilities in garch models. *Computational Statistics & Data Analysis*, *50*(9), 2293–2312.

Phatarfod, R., & Mardia, K. (1973). Some results for dams with Markovian inputs. *Journal of Applied Probability*, *10*(1), 166–180.

Sebastian, T., Jeyaseelan, V., Jeyaseelan, L., Anandan, S., George, S., & Bangdiwala, S. I. (2018). Decoding and modelling of time series count data using Poisson hidden Markov model and Markov ordinal logistic regression models. *Statistical methods in medical research*.

Shumway, R., & Stoffer, D. (1982). An approach to time series smoothing and forecasting using the EM algorithm. *Journal of Time Series Analysis*, *3*(4), 253–264.

Smith, A. F., & Roberts, G. O. (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Methodological)*, *55*(1), 3–23.

Strickland, C. M., Forbes, C. S., & Martin, G. M. (2006). Bayesian analysis of the stochastic conditional duration model. *Computational Statistics & Data Analysis*, *50*(9), 2247–2267.

Tang, F., & Cavanaugh, J. E. (2018). State-space models for Binomial time series with excess zeros. In *Time series analysis and applications.* InTech.

Taylor, S. J. (1982). Financial returns modelled by the product of two stochastic processes-a study of the daily sugar prices 1961-75. *Time series analysis: theory and practice*, *1*, 203–226.

Taylor, S. J. (2007). *Modelling financial time series* (2nd ed.). World Scientific.

Thavaneswaran, A., Ravishanker, N., & Liang, Y. (2015). Generalized duration models and optimal estimation using estimating functions. *Annals of the Institute of Statistical Mathematics*, *67*(1), 129–156.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *Annals of Statistics*, *22*(4), 1701–1728.

Tjøstheim, D. (1986). Some doubly stochastic time series models. *Journal of Time Series Analysis*, *7*(1), 51–72.

Tong, H. (1990). *Non-linear time series: A dynamical system approach*. Oxford University Press.

Vidoni, P. (1999). Exponential family state space models based on a conjugate latent process. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *61*, 213–221.

Wei, G. C., & Tanner, M. A. (1990). A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *Journal of the American Statistical Association*, *85*(411), 699–704.

Wu, C. J. (1983). On the convergence properties of the EM algorithm. *The Annals of Statistics*, *11*(1), 95–103.

Yang, M., Cavanaugh, J. E., & Zamba, G. K. (2015). State-space models for count time series with excess zeros. *Statistical Modelling*, *15*(1), 70–90.

Yang, M., Zamba, G. K., & Cavanaugh, J. E. (2013). Markov regression models for count time series with excess zeros: A partial likelihood approach. *Statistical Methodology*, *14*, 26–38.

Zeger, S. L., & Qaqish, B. (1988). Markov regression models for time series: A quasi-likelihood approach. *Biometrics*, *44*(4), 1019–1031.

Zhu, F. (2011). A negative binomial integer-valued GARCH model. *Journal of Time Series Analysis*, *32*(1), 54–67.