OBSERVATION-DRIVEN REGRESSION MODELS FOR TIME SERIES OF COUNTS

TASLIM S. MALLICK

Observation-driven Regression Models for Time Series of Counts

by

©Taslim S. Mallick

A thesis submitted to the School of Graduate Studies in partial fulfillment of the requirement for the Degree of Master of Science in Statistics

Department of Mathematics and Statistics
Memorial University of Newfoundland

July, 2004

St. John's Newfoundland Canada

Your file Votre référence
ISBN: 0-494-02354-6
Our file Notre référence
ISBN: 0-494-02354-6

Abstract

There are many situations in practice where one may encounter time series of counts. For example, one may require to analyse a time series of number of tourists or time series of number of patients for a particular disease mainly for the purpose of forecasting of a future count. The analysis of this type of time series of counts is, however, not adequately addressed in the literature. One of the main difficulties in analysing such a time series is the problem of modelling the autocorrelations of the count responses recorded sequentially. In this thesis, we first use an observation-driven correlation model for both stationary and non-stationary Poisson count data. When count responses are subject to overdispersion, one may use a time series of negative binomial counts to analyse such overdispersed and correlated data. There exists a random effects based parameter-driven approach to model this type of time series of negative binomial counts. This approach, however, has some pitfalls as it is difficult to interpret the correlations of observations through the correlations of the random effects. As a remedy, following McKenzie (1986, *Adv. Appl. Probab.*) we use an observation-driven correlation model to fit correlated negative binomial stationary data. Next we generalize this to the non-stationary data.

As far as the estimation of parameters is concerned, we follow Sutradhar (2003, *Statistical Science*) and use a generalized quasilikelihood approach for the estimation of the associated regression parameter. The overdispersion and correlation parameters are estimated by using the well-known method of moments. This estimation approach yields consistent estimates for all parameters of the model. This consistency property is examined through a simulation study for stationary and non-stationary Poisson as well as negative binomial count data. The estimation method is illustrated by using a real life data that was earlier analyzed by Zeger (1988, *Biometrika*). We have also developed the formulae for forecasting a future count for the stationary Poisson and negative binomial time series. The performance of the forecasting functions is also examined through a simulation study.

Acknowledgements

I am grateful to my supervisor Professor B.C. Sutradhar for his constant encouragement, valuable comments, and helpful suggestions in completing this thesis. It was indeed a great privilege to work on this important problem in the area of discrete time series analysis, which was suggested by Professor Sutradhar.

I sincerely acknowledge the financial support provided by the School of Graduate Studies and Department of Mathematics and Statistics in the form of Graduate Fellowships and Teaching Assistantships. Further, I wish to thank the Department for providing us a friendly atmosphere and the necessary facilities to complete my program in particular.

I also thank Dr. Larry Bauer and Dr. Garry Sneddon of Memorial University of Newfoundland for taking the time to serve as examiners of this thesis. Their constructive comments and suggestions served to improve the quality of the thesis.

I am also grateful to my parents, brother and sister for their eternal love, emotional support and encouragement during this program.

It is my great pleasure to thank my friends and well-wishers who directly or indirectly encouraged and helped me in the M.Sc. program and contributed to this dissertation.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background of the Problem

Analysis of time series for continuous data has been extensively reviewed in the literature. This type of time series is very common in practice. For example, economic time series of export or import data, climatology time series of rainfall or temperature data, and biomedical time series of blood pressure. However there are many other situations where one may encounter time series of discrete count data. Examples of interest include time series of tourist data, time series of the number of storms in a particular region, and the time series of the number of individuals having a particular disease. The analysis of this type of discrete time series of count data is however not adequately addressed in the literature.

Cox (1981) characterized two classes of models of time-dependent data: observation-driven and parameter-driven models. In an observation-driven model, the conditional distribution of the present observation is specified as a function of past observations. In parameter-driven models, autocorrelation is introduced through a latent process. Some authors, such as Jacobs and Lewis (1978a, 1978b, 1983), McKenzie (1986), Al-Osh and Aly (1992) have dealt with observation-driven time series of binary and/or count data, whereas Davis et al (2003) have dealt with observation-driven time series of count data. Some other authors such as Zeger (1988), Harvey and Fernandes

(1989), Kulendran and King (1997), Davis et al (2000) have dealt with parameter-driven time series of counts.

Jacobs and Lewis (1978a, 1978b, 1983) have modeled the discrete time series by exploiting certain mixture principles. They introduced a discrete mixed autoregressive moving average (DARMA) process to provide a scheme for modelling a stationary sequence of dependent discrete random variables. These authors have mainly concentrated on the modelling of binary time series data. McKenzie(1986) has however pointed out that DARMA is usually overparametrized for practical use and sometimes too specialized in correlation structure. Moreover it is not clear how the time series of counts can be generated and analysed based on the DARMA approach of Jacobs and Lewis (1978a, 1978b, 1983).

Turning back to McKenzie (1986, 1988), this author in addition to binary time series, has modeled the time series of negative binomial and geometric marginals through a 'thinning' operation. However he did not include any inference procedure for the estimation of the parameters and his study was confined to the stationary non-regression problems. Al-Osh and Aly (1992) studied an AR(1) (autoregressive of order 1) model for time series of counts which is similar to that of McKenzie (1986, 1988).

As opposed to the observation driven models, Zeger (1988) discussed parameter driven time series models for count data that include overdispersion. In this approach, the count observations are assumed to follow a Poisson distribution conditional on certain random effects. Furthermore it is assumed that these random effects are correlated with a specified mean, variance and autocorrelation structure. This approach consequently makes the count observations to be correlated unconditionally but nothing can be said about the marginal distribution of the count response. This is because no marginal distribution was assumed for the random effect, although in his simulation study he has used a log-normal distribution for the random effects for convenience. One of the major difficulties of this modeling is that it is not easy to interpret the correlation of the data (see Jowaheer and Sutradhar (2002)), whereas in practice it may be appropriate to consider Gaussian type AR(1), MA(1) (moving average of order 1) or equicorrelation structure for the count data, at least in the stationary setup. As an illustration of the model, Zeger (1988) used monthly data of polio counts, reported by the US Center of Disease Control for the years 1970-83, in order to identify the trend and seasonal effect of the data.

Harvey and Fernandes (1989) modelled the time series of count data in a different manner. They introduced a hyper-parameter into the model through which the parameters of the underlying distributions become dependent successively with respect to time. This approach eventually yields a negative binomial conditional distribution, whereas it seems to be more appropriate to use a marginal negative binomial distribution for the data. Moreover their estimation for the variance component (overdispersion) of the model was done by exploiting an approximate likelihood function which was developed based on an independence assumption for the data separated by lag 2 or more. This is a questionable approximation as there is no reason why lag 2 or higher order lag correlations will be insignificant. Also in their approach, it is extremely difficult to find and interpret the unconditional correlations of the data.

Kulendran and King (1997) analysed count data in the presence of trend and seasonal components. They used a Gaussian distribution to model the correlations of the count data without challenging the discrete nature of the data. Thus the estimate may not be consistent and reliable.

Davis et al (2000) discussed a random effect driven approach for time series of Poisson counts. As an extension of Zeger's (1988) approach, they have used the extended method of moments for the estimation of the variance component and lag correlation parameters of the random effect. For the estimation of the regression parameter, they have ignored the correlations of the data and obtained a likelihood estimator, whereas Zeger(1988) used the quasilikelihood approach. Note that, as the correlation of the data was ignored in regression parameter estimation, this type of estimator will naturally be inefficient. Further note that, one of the major problems of this modelling is the computation and estimation of the actual correlations which was also the case in Zeger's(1988) approach.

Recently, Davis et al (2003) have used an observation driven modelling for the time series of count data. More specifically, conditional on the past observations, they have assumed a Poisson distribution for the current random count. This appears to be similar to the conditional approach of Harvey and Fernandes(1989), where again finding the unconditional correlation and interpretation of such correlation is a problem.

Most of the studies discussed above modeled the correlations in time series of counts through the introduction of correlated random effects. In this approach, while it is easy to understand or interpret the correlations of the random effects, it is however not easy to understand or interpret the correlations of the data. To be specific, the unconditional correlations may have complicated forms which may not be easy to interpret. For this reason, Jowaheer and Sutradhar (2002) have proposed certain observation-driven correlation models for the count data in a longitudinal set up. These models are quite similar to the Gaussian autocorrelation models. Consequently it is easy to compute the correlations and also easy to interpret them. The statistical analysis of count data generated by this type of observation-driven models is not however adequately addressed in the literature. The objective of the present thesis, as given below, is to model the correlation of a time series of counts through observation-driven relationship and discuss the inferences such as estimation and forecasting based on such modelling.

1.2 Objective of The Thesis

Analysing time series of counts is an important research topic. The analysis of such data, especially when they are subject to overdispersion, has been hampered due to the difficulty of modelling their correlation structure. Some authors such as Zeger (1988) modelled the correlations through the correlations of random effects. This however makes the estimation and interpretation of the model (including the correlation structure) difficult. One of the main objectives of the thesis is to develop an observation-driven correlation model for non-stationary negative binomial data. This we have shown in Chapter 5. Next, the parameters of such a model are estimated by using a consistent estimation approach, combining the well-known generalized quasi-likelihood (GQL) and moment method. Another main objective of the thesis is to develop an appropriate function for forecasting a future count.

We now provide the specific objectives of the thesis as follows. In Chapter 2, we discuss some existing time series models for count data along with their limitations. In Chapter 3, we fit Poisson models for time series of count data for stationary and non-stationary cases. Note that the stationary Poisson model was originally discussed by McKenzie (1988), whereas Sutradhar, Jowaheer and Rao (2003) have developed a non-stationary Poisson model. The parameters of such models are estimated by using the GQL approach. The estimation performances are examined by a simulation study. We also examine the forecasting aspects under the stationary model.

In Chapter 4, a parameter-driven model is reviewed for time series of negative binomial count data. As mentioned before, in Chapter 5, we discuss observation-driven modelling of negative binomial time series of counts and compare with the parameter-driven model to be discussed in Chapter 4. We perform a simulation study to fit the model and apply the non-stationary model to a real life data of polio counts. Finally, in Chapter 6, we make some comments and discuss about the scope of further research in this area.

Chapter 2

Existing Time Series Models for Count Data and Their Limitations

When the responses of a count random variable are recorded for a long period of time, they form a time series of counts. It is then most likely that these counts will be autocorrelated. The modelling of the correlations for count data, specially when count data follow a non-stationary series, is however not easy. This problem of modelling naturally affects the analysis of the time series of counts. This is because the inference for the mean level and the variations in data can not be efficiently achieved without knowing the autocorrelation structure of the data. Some authors, such as Zeger (1988), Harvey and Fernandes (1989) and Davis et al (2000) considered the parameter-driven models for time series of count data. In this approach, the correlated random effects are used to generate correlated count data. The interpretation of such correlations is however not easy. Some other authors, such as McKenzie (1986) and Davis et al (2003) have considered observation-driven models for time series of counts. In this approach, it is easy to interpret correlations for the purpose of mean and variance calculation. But in general, these authors did not pursue the inference problem.

In Section 2.1 below, we discuss in detail the parameter-driven model discussed by Zeger (1988).

2.1 Semi-parametric Poisson Mixed Model: A Parameter-driven Marginal Approach

Let y_t be the observed count, and x_t be the associated vector of covariates at time t. If θ_t is the random effect at time t such that $E(\theta_t) = 1$, $var(\theta_t) = \alpha$, and $cov(\theta_t, \theta_{t+l}) = \alpha\rho_\theta(l)$, then conditional on the random effects θ_t, it has been assumed in Zeger (1988) that y_t has same mean and variance function given by

$$E(Y_t \mid \theta_t) = \mu_t = e^{x_t'\beta}\theta_t = m_t\theta_t = var(Y_t \mid \theta_t),$$

where $m_t = e^{x_t'\beta}$. Note that this property of equal mean and variance is characterized by the well known Poisson distribution. One may then obtain the unconditional marginal mean, variance, and autocorrelation function of y_t as

$$E(Y_t) = E_{\theta_t}E(Y_t \mid \theta_t) = E_{\theta_t}(m_t\theta_t) = m_t, \tag{2.1}$$

$$\begin{aligned} V(Y_t) &= E_{\theta_t}var(Y_t \mid \theta_t) + var_{\theta_t}E(Y_t \mid \theta_t) \\ &= E_{\theta_t}(m_t\theta_t) + var_{\theta_t}(m_t\theta_t) \\ &= m_t + \alpha m_t^2. \end{aligned} \tag{2.2}$$

Therefore, although the distribution of y_t is unknown (as the distribution of θ_t is unknown) the distribution of y_t accommodates overdispersion indexed by α. Next, under the assumption that the observations are independent conditional on the random effects $\theta_1, \cdots, \theta_T$, one may show by similar calculations as in (2.2) that

$$\begin{aligned} cov(Y_t, Y_{t+l}) &= E_{\theta_t}cov(Y_t, Y_{t+l} \mid \theta_t) + cov_{\theta_t}\left[E(Y_t \mid \theta_t), E(Y_{t+l} \mid \theta_{t+l})\right] \\ &= cov_{\theta_t}(m_t\theta_t, m_{t+l}\theta_{t+l}) \\ &= m_t m_{t+l}\alpha\rho_\theta(l). \end{aligned} \tag{2.3}$$

Therefore, the lag l unconditional correlation, denoted by $\rho_y(l)$ has the form given by

$$\rho_y(l) = \frac{\rho_\theta(l)}{\sqrt{\left(1+\frac{1}{\alpha m_t}\right)\left(1+\frac{1}{\alpha m_{t+l}}\right)}}. \quad (2.4)$$

Note that the lag l autocorrelation (2.4) of the responses $y_1, \cdots, y_T$ is a function of the lag l autocorrelation of the random effects $\theta_1, \cdots, \theta_t$. Thus the interpretation of $\rho_y(l)$ based on a complicated function of $\rho_\theta(l)$, becomes difficult. For example, Zeger (1988) in his simulation study assumed Gaussian autocorrelation structure for $\log\theta_t$, $t = 1, \cdots, T$ which leads the responses $y_1, \cdots, y_T$ to be correlated in a special way. The finding of $\rho_\theta(l)$ based on $\rho_{log\theta}(l)$ is however not easy. This makes the autocorrelation of $\rho_y(l)$ difficult to compute and understand. Furthermore, it is not easy to see the range of the correlations defined by $\rho_y(l)$ in (2.4). This is because the range of $\rho_y(l)$ does not only depend on the range of $\rho_\theta(l)$, rather it also depends on the values of α and m_t. It is therefore clear that the modelling of the correlations of the observations based on the above random effect approach has serious limitations. Further note that the correlation formula $\rho_y(l)$ should play an important role in forecasting the future count observations. But as the formula of $\rho_y(l)$ in (2.4) may not represent the correlations of the repeated observations, the forecasting consequently will be adversely affected. This problem, however, will not arise when an observation driven correlation process is adopted to model such correlations (see Jowaheer and Sutradhar (2002)).

2.1.1 Estimation of Parameters

Even though $\rho_y(l)$ has serious limitations, we now briefly discuss the estimation approach used by Zeger (1988) for β, α, and $\rho_\theta(l)$ parameters. Note that $\rho_\theta(l)$ along with β and α determines the correlations of the count observations.

For the estimation of the regression parameter β, Zeger generalized the quasi-likelihood estimating equation and obtained an iterative weighted and filtered least squares algorithm for time series data. The quasilikelihood estimating equation for the time series data has the form

$$\frac{\partial m'}{\partial \beta} V^{-1}(y - m) = 0,$$

where $y' = (y_1, \ldots, y_T)$, $m' = (m_1, \ldots, m_T)$, and $V = var(y) = A + \alpha A R_\theta A$, where $A = diag(m_1, \ldots, m_T)$ and R_θ is the correlation matrix of the random effects. To avoid the complexity of the inversion of the V matrix, he approximated V by $V_R = D^{\frac{1}{2}} R(w) D^{\frac{1}{2}}$, where $D = diag(m_t + \alpha m_t^2)$ and $R(w)$ is some 'working' correlation matrix. Therefore, replacing V by V_R in the quasilikelihood estimating equation and using the fact that $V_R^{-1} \simeq D^{-\frac{1}{2}} L' L D^{-\frac{1}{2}}$, L being the matrix which applies the autoregressive filter, he proposed the iterative weighted and filtered least squares equation for β given by

$$\hat{\beta}_R^{(j+1)} = \left[\left(LD^{-\frac{1}{2}} \frac{\partial m}{\partial \beta} \right)' \left(LD^{-\frac{1}{2}} \frac{\partial m}{\partial \beta} \right) \right]^{-1} \left(LD^{-\frac{1}{2}} \frac{\partial m}{\partial \beta} \right)' \left(LD^{-\frac{1}{2}} Z \right), \tag{2.5}$$

where $Z = (\partial \mu' / \partial \beta) \beta + (y - \mu)$.

With regard to the construction of the $R(w)$ matrix, Zeger (1988) used a special working matrix to replace the correlation matrix with general elements $\rho_\theta(l)$. This approach has problems as it is not clear how one can compute $\rho_\theta(l)$. Moreover the efficiency of β may be lost to a significant extent when such an approximation is used.

For the estimation of the variance component α and autocorrelation parameter $\rho_\theta(l)$ of the random effects, moment estimation was used. The estimating equations are

$$\hat{\alpha} = \frac{\sum_{t=1}^{T} \left[(y_t - \hat{m_t})^2 - \hat{m_t} \right]}{\sum_{t=1}^{T} \hat{m_t}^2}, \tag{2.6}$$

$$\hat{\rho}_\theta(l) = \frac{\sum_{t=l+1}^{T} \left[(y_t - \hat{m_t}) \left(y_{t-l} - \hat{m_{t-l}} \right) \right]}{\hat{\alpha} \sum_{t=l+1}^{T} \hat{m_t} \hat{m_{t-l}}}. \tag{2.7}$$

Note that $\hat{\alpha}$ obtained by (2.6) has limitations as $\hat{\alpha}$ can be negative. Also it is not clear what range restriction is needed for $\rho_\theta(l)$. This means that the range restriction for $\rho_y(l)$ is not clear al all. This may cause problems in the inversion of the covariance matrix which is needed to estimate β.

2.2 Dynamic Poisson Models: A Parameter-driven Conditional Approach

Marginal Model:

Suppose that conditional on θ_t, y_t has the Poisson distribution given by

$$f(y_t \mid \theta_t) = \frac{e^{-\theta_t}\theta_t^{y_t}}{y_t!}. \tag{2.8}$$

Note that this Poisson distribution was also indirectly used by Zeger (1988). Next, to find the marginal first and second order moments of $y_1, \cdots, y_T$, Zeger (1988) did not assume any specific multivariate distribution for $\theta_1, \cdots, \theta_T$; rather he assumed that $\theta_1, \cdots, \theta_T$ are dependent with a general autocorrelation structure. This approach is referred to as the marginal approach.

Conditional Model:

There exists an alternative approach where a distribution is assumed for θ_t conditional on the history $y_1, \cdots, y_{t-1}$ and associated covariates. For example, Harvey and Fernandes (1989) (see also Settimi and Smith (2000, p.139-140)) assumed that θ_t conditional on y_{t-1} is distributed as $G(a_t, b_t)$, i.e.

$$f(\theta_t \mid y_{t-1}) = \frac{e^{-b_t\theta_t}\theta_t^{a_t-1}}{\Gamma(a_t)b_t^{-a_t}}, \theta_t > 0 \tag{2.9}$$

This leads to the conditional distribution of y_t given y_{t-1} as

$$\begin{aligned} f(y_t \mid y_{t-1}) &= \int_0^\infty f(y_t \mid \theta_t) f(\theta_t \mid y_{t-1})\, d\theta_t \\ &= \frac{\Gamma(y_t + a_t)}{\Gamma(a_t)y_t!} (b_t)^{a_t} (1 + b_t)^{-(y_t+a_t)}, \end{aligned} \tag{2.10}$$

which is the probability density function (*pdf*) of a negative binomial variable. Further note that the pdf in (2.10) may be re-expressed as as

$$f(y_t \mid y_{t-1}) = \frac{\Gamma(y_t + a_t)}{\Gamma(a_t)y_t!} \left(\frac{1}{1+b_t}\right)^{y_t} \left(1 - \frac{1}{1+b_t}\right)^{a_t}, \tag{2.11}$$

which is widely used in the literature. This pdf form in (2.11) is denoted by $NB\left(a_t, \frac{1}{b_t}\right)$. It can be shown that the mean and variance function of y_t given y_{t-1} have the formulae:

$$E\left(Y_t \mid y_{t-1}\right) = \frac{a_t}{b_t} = r_t$$

$$V\left(Y_t \mid y_{t-1}\right) = r_t + \frac{1}{a_t} r_t^2,$$

(Johnson and Kotz (1992, p.199)) where $\frac{1}{a_t}$ may be considered as overdispersion index parameter.

Predictive Distribution:

Sometimes it may be of interest to find the predictive distribution of θ_t given y_t. This can be obtained by using the well-known Bayesian approach. More specifically it can be shown that

$$\theta_t \mid y_t \sim G\left(a_t^*, b_t^*\right),$$

where $a_t^* = a_t + y_t$ and $b_t^* = 1 + b_t$.

Conditional Model for the Regression Case:

Suppose that $\mu_t = m_t\theta_t$, where $m_t = e^{x_t^{'}\beta}$ and θ_t is the random effect as defined above. In this case, the pdf of y_t conditional on θ_t is written as

$$f\left(y_t \mid \theta_t\right) = \frac{e^{-\mu_t}\mu_t^{y_t}}{y_t!}.$$

Further suppose that the distribution of θ_t given y_{t-1} is the same as before. That is, θ_t has the gamma distribution given by $\theta_t \mid y_{t-1} \sim G(a_t, b_t)$ as in (2.9). It then follows that

$$\begin{aligned} f\left(y_t \mid y_{t-1}\right) &= \int_0^\infty f\left(y_t \mid \theta_t\right) f\left(\theta_t \mid y_{t-1}\right) d\theta_t \\ &= \int_0^\infty \frac{e^{-m_t\theta_t}(m_t\theta_t)^{y_t}}{y_t!} \frac{e^{-b_t\theta_t}\theta_t^{a_t-1}}{\Gamma(a_t)b_t^{-a_t}} d\theta_t \end{aligned}$$

$$
\begin{aligned}
&= \frac{b_t^{a_t} m_t^{y_t}}{\Gamma(a_t) y_t!} \frac{\Gamma(y_t + a_t)}{(b_t + m_t)^{y_t + a_t}} \\
&= \frac{\Gamma(y_t + a_t)}{\Gamma(a_t) y_t!} \left(\frac{b_t}{m_t}\right)^{a_t} \left(1 + \frac{b_t}{m_t}\right)^{-(y_t + a_t)} \\
&= \frac{\Gamma(y_t + a_t)}{\Gamma(a_t) y_t!} \left(\frac{1}{1 + \frac{b_t}{m_t}}\right)^{y_t} \left(1 - \frac{1}{1 + \frac{b_t}{m_t}}\right)^{a_t}. \qquad (2.12)
\end{aligned}
$$

The form stated in (2.12) is matched with that in Jowaheer and Sutradhar (2002) and accordingly denoted by $NB\left(a_t, \frac{m_t}{b_t}\right)$ with mean and variance

$$
\begin{aligned}
E\left(Y_t \mid y_{t-1}\right) &= \frac{a_t}{b_t} m_t = r_t^* \\
V\left(Y_t \mid y_{t-1}\right) &= r_t^* + \frac{1}{a_t} r_t^{*2}.
\end{aligned}
$$

2.2.1 Estimation of Parameters

Note that it is extremely difficult to write the exact likelihood function of the data. As a remedy, Harvey and Fernandes (1989) used an approximate log-likelihood function given by

$$
\begin{aligned}
\mathcal{L}(w, \beta) &= log \prod_{t=1}^{T} f\left(y_t \mid y_{t-1}\right) \\
&= \sum \left[log\Gamma(y_t + a_t) - log\Gamma(a_t) - log(y_t!) + a_t log\left(\frac{b_t}{m_t}\right) \right. \\
&\quad \left. - (y_t + a_t) log\left(1 + \frac{b_t}{m_t}\right) \right] \qquad (2.13)
\end{aligned}
$$

and suggested the estimation of the parameters involved. To be specific, using the recurrence relation $a_t = wa_{t-1}$ and $b_t = wb_{t-1}$ such that $a_0 = b_0 = 1$, the likelihood function may be maximized to estimate w and β.

In their approach, it is very difficult to find the unconditional correlation of the data. Moreover their likelihood function was based on the conditional distribution which is based on independence assumption for the data separated by lag 2 or more. This approximation may not yield efficient estimates for the parameters.

2.3 Modelling Poisson Counts: An Indirect Observation-driven Approach

To discuss Davis et al's (2003) model, we turn back to Zeger's (1988) model reviewed in section 2.1. When a Poisson distribution is used for y_t conditional on $\theta_t = e^{\gamma_t}$, Zeger's (1988) model can be written as

$$y_t \mid \theta_t \sim P\left(\mu_t = e^{x_t'\beta+\gamma_t}\right). \quad (2.14)$$

Next, to recognize the correlation structure of $y_1, \cdots, y_t, \cdots, y_T$, Zeger assumed that $corr(\theta_t, \theta_{t-l}) = \rho_\theta(l)$, where $\rho_\theta(l)$ denotes the lag l correlation among θ's. Note that in a simulation study Zeger considered a Gaussian ARMA(p,q) type correlation process for $\gamma_1, \cdots, \gamma_T$. More specifically, one may write the ARMA(p,q) process for γ_t as

$$\gamma_t = \phi_1\gamma_{t-1} + \cdots + \phi_p\gamma_{t-p} + e_t + \psi_1 e_{t-1} + \cdots + \psi_q e_{t-q}, \quad (2.15)$$

where e_t's are independently and identically distributed (iid) $N(0, \alpha)$. This is a random-effect driven process as far as the generation of the data $y_1, \cdots, y_T$ is concerned. As mentioned in section 2.1, it is not easy to describe the unconditional correlation structure based on such latent process.

Davis et al (2003) have used a model similar to (2.14)-(2,15) but unlike Zeger (1988) they used $e_t = \frac{y_t-\mu_t}{\mu_t^{-\lambda}}$ for $\lambda \in (0,1]$ in (2.14), where y_t's are counts and wrote the model as

$$y_t \mid y_{t-q}, \cdots, y_{t-1}, \gamma_{t-p}, \cdots, \gamma_{t-1} \sim P\left(\mu_t = e^{x_t'\beta+\gamma_t}\right) \quad (2.16)$$

and have referred to this model (2.16) as an observation-driven correlation model. Note that as

$$\gamma_t = \phi_1\gamma_{t-1} + \cdots + \phi_p\gamma_{t-p} + e_t + \psi_1 e_{t-1} + \cdots + \psi_q e_{t-q}$$

and $e_t = \frac{y_t-\mu_t}{\mu_t^{-\lambda}}$, where e_t now does not have normal distribution (although e_t has zero mean). It is not at all clear how to find the unconditional correlation structure for

$y_1, \cdots, y_t, \cdots, y_T$ based on (2.16). Thus this type of observation-driven model modified directly from the parameter-driven model appears to have serious limitations. To be specific, it is difficult to find out the unconditional marginal distribution of y_t from (2.16) and it is much more difficult to derive and interpret the unconditional correlation structure of the data. As opposed to this type of artificial observation-driven model, we, following McKenzie (1986), will discuss a Gaussian type observation-driven model for count data in section 2.4.

It should also be mentioned that a model similar to (2.16) of Davis et al (2003) was considered by Honore and Kyriazidou (2000) in the context of repeated binary data. More specifically, Honore and Kyriazidou (2000) considered

$$p\left(y_t = 1 \mid x_t, y_1, \cdots, y_{t-1}\right) = \frac{e^{x_t^{'}\beta + \gamma y_{t-1}}}{1 + e^{x_t^{'}\beta + \gamma y_{t-1}}}, \tag{2.17}$$

where γ may be referred to as the AR(1) type dependence parameter. It is however not clear how to compute the correlation of the data $y_1, \cdots, y_t, \cdots, y_T$ from the non-linear implicit relationship between y_t and the past response. This hampers the estimation of β and γ under the present time series set up for discrete data.

2.3.1 Estimation of Parameters

As far as the estimation of parameters $\delta = \left(\beta^T, \phi^T, \psi^T\right)^T$ is concerned, Davis et al (2003) used a conditional maximum likelihood approach. To be specific, they wrote the log-likelihood function exploiting the conditional density (2.16), which is given by

$$\mathcal{L}(\delta) = \sum_{t=1}^{n} \left[y_t log\mu_t - \mu_t\right], \tag{2.18}$$

where $log\mu_t = x_t^{'}\beta + \gamma_t$ with $\gamma_t = \phi_1\gamma_{t-1} + \cdots + \phi_p\gamma_{t-p} + e_t + \psi_1 e_{t-1} + \cdots + \psi_q e_{t-q}$, and $e_t = \frac{y_t - \mu_t}{\mu_t^{-\lambda}}$. Note that as $e_t = \frac{y_t - \mu_t}{\mu_t^{-\lambda}}$, y_t being a count observation, it is not easy to find the correct correlation structure for $\{\gamma_t\}$. Consequently, even though one may obtain likelihood estimates for the parameters involved, the appropriateness of such an approximate likelihood model to fit the data seems to be questionable.

2.4 A Direct Observation-driven model For Time Series of Counts

McKenzie (1988) constructed a model for time series with Poisson marginals. He used a Gaussian type autoregressive model to establish a direct relationship between y_t and the past observations. More specifically, the observation y_t at time t has the relationship with y_{t-1}, for example, as

$$y_t = \rho * y_{t-1} + d_t, \tag{2.19}$$

where y_{t-1} has Poisson distribution with parameter $m^* = e^{\beta_1}$, i.e., $y_{t-1} \sim P(m^*)$. In (2.19), ρ is a constant scale parameter satisfying the range restriction $0 < \rho < 1$. Moreover, for given y_{t-1}, $\rho * y_{t-1}$ in (2.19) is computed through a binomial thinning operation. To be specific, $\rho * y_{t-1}$ is the sum of y_{t-1} binary observations, where each observation is generated with probability ρ. Therefore we can write

$$\rho * y_{t-1} = \sum_{j=1}^{y_{t-1}} b_j(\rho) = z_{t-1} \tag{2.20}$$

with $Pr\left[b_j(\rho) = 1\right] = \rho$ and $Pr\left[b_j(\rho) = 0\right] = 1 - \rho$. It then follows that, conditional on y_{t-1}, z_{t-1} has the binomial distribution. We denote this binomial distribution as $B(y_{t-1}, \rho)$. Furthermore if we assume that $d_t \sim P(m^*(1-\rho))$ and is independent of z_{t-1}, it may be shown that $y_t \sim P(m^*)$. It also follows that $E(y_t y_{t-l}) = m^* \rho^l + {m^*}^2$, yielding the lag-$l$ correlation between y_t and y_{t-l} as ρ^l, which is characterized by the property of the Gaussian AR(1) model. Note that as opposed to the range restriction $0 < \rho < 1$ in (2.19), ρ has the range restriction $-1 < \rho < 1$ in the Gaussian AR(1) model.

Note that the time series of counts with negative binomial marginals was also studied by McKenzie (1986). For details on the analysis of negative binomial time series, see Chapter 4 in the present thesis.

2.4.1 Estimation of Parameters

McKenzie (1986, 1988) considered direct observation-driven models to generate Poisson and negative binomial counts which are very useful in generating data in practice. Note that he did not however consider any statistical inference such as estimation of parameters for time series of counts, whether Poisson or negative binomial.

Chapter 3

Observation-driven Model for Stationary and Non-stationary Poisson Time Series

In this chapter, we will discuss modelling time series of Poisson counts, estimation of the parameters associated in such a model as well as the forecasting of a future count. With regard to the modelling of the count responses, we first assume that the count observations of a time series are marginally distributed as Poisson. Next, we introduce an observation-driven correlation process so that marginally each of the counts follows a Poisson distribution, but, they jointly follow a multivariate Poisson distribution. Note that the inference procedure to be adapted in this chapter however will not require the specific form of the multivariate Poisson distribution.

In notation, suppose that $y_1, \cdots, y_t, \cdots, y_T$ is a time series of counts and $x_t = (x_{t1}, \cdots, x_{tp})'$ is a vector of p-dimensional covariates associated with y_t. We assume that y_t is distributed as Poisson, i.e.,

$$f(y_t) = \frac{e^{-m_t} m_t^{y_t}}{y_t!}, y_t = 0, 1, 2, \cdots \tag{3.1}$$

where, $m_t = e^{x_t' \beta}$. Here $E(Y_t) = V(Y_t) = m_t$. It is clear that the mean and variance are time dependent, which indicates that the series is non-stationary by nature. As far as the dependence of observations of this time series is concerned, following the correlation structure proposed by McKenzie (1988) for stationary counts, recently, Sutradhar, Jowaheer and Rao (2003) have developed a correlation structure for the non-stationary Poisson counts in a longitudinal set up. We however confine our study to the time series set up and generalize McKenzie's (1988) stationary model to the non-stationary case.

In section 3.2, we deal with the estimation of the regression parameter β as well as the correlation parameter to be introduced in section 3.1. More specifically, a generalized quasilikelihood (GQL) approach following Sutradhar (2003, section 3) (see also Zeger (1988)) will be discussed for the estimation of the regression parameter β. The correlation parameter will be consistently estimated by using the method of moments. An independence assumption-based 'working' GQL approach also will be discussed for β estimation. The performance of the estimators will be examined through a simulation study in section 3.3.

In section 3.4, we discuss the forecasting of a future count based on the fitted model to be discussed in section 3.2. To be specific, it is shown that the forecasting performance becomes better when the underlying model is fitted by using the GQL approach rather than the independence-assumption based 'working' GQL approach.

We now turn back to the modelling of the correlation structure for stationary and non-stationary Poisson time series following McKenzie (1988).

3.1 Correlation Structure for Poisson Time Series

3.1.1 Stationary Model

Recall from section 2.4 that when count responses $y_1, \cdots, y_t, \cdots, y_T$ are generated by a stationary AR(1) type model (2.19), the expectation, variance and the correlation are given by

$$E(Y_t) = m^* = e^{\beta_1},$$

$$V(Y_t) = m^* = e^{\beta_1},$$

and

$$corr(Y_t, Y_{t-l}) = \rho^l = \rho_l.$$

Here ρ_l indicates the correlation between two observations lag-l apart which is time independent. Let $C^*(\rho)$ denote the $T \times T$ correlation matrix of all observations $y_1, \cdots, y_t, \cdots, y_T$. It then follows that for stationary AR(1) Poisson counts this $C^*(\rho)$ may be written as

$$C^*(\rho) = \begin{pmatrix} 1 & \rho & \cdots & \rho^{T-1} \\ \rho & 1 & \cdots & \rho^{T-2} \\ \vdots & \vdots & & \vdots \\ \rho^{T-1} & \rho^{T-2} & \cdots & 1 \end{pmatrix}, \tag{3.2}$$

where $0 < \rho < 1$. Note that this stationary correlation structure (3.2) is not suitable for a Poisson time series with marginal distributional property given by (3.1). This is because (3.1) provides non-stationary mean and variance for the count responses. Note that Sutradhar, Jowaheer and Rao (2003) have introduced a non-stationary Poisson AR(1) model in the longitudinal set up.

In the following subsection we review in brief the non-stationary Poisson AR(1) model discussed in Sutradhar, Jowaheer and Rao (2003) in the time series set up. In particular we provide the correlation structure for such non-stationary time series which will be a modification of the stationary correlation structure given in (3.2).

3.1.2 Non-stationary Model

For convenience, here we rewrite the AR(1) relationship (2.19) among Poisson counts, which is given by

$$y_t = \rho * y_{t-1} + d_t \tag{3.3}$$

Note however that under the non-stationary case one requires non-stationary marginal distributions for y_{t-1} as well as for d_t, which is different than the stationary case. More specifically, in the non-stationary case it is assumed that $y_{t-1} \sim P(m_{t-1})$ and $d_t \sim P(m_t - \rho m_{t-1})$. Note that as in the stationary case we assume here that y_{t-1}

and d_t are independent and $\rho * y_{t-1}$ is the same binomial thinning operation as in (2.19). This yields that $y_t \sim P(m_t)$. It then follows that

$$E(Y_t) = m_t$$

$$V(Y_t) = m_t,$$

but one requires special attention to compute the lag covariances.

Further note that since d_t in (3.3) is Poisson, the mean parameter must be positive, i.e., $m_t - \rho m_{t-1} > 0$, which implies $\rho < \frac{m_t}{m_{t-1}}$. Moreover since ρ is a probability parameter, ρ must satisfy the restriction $0 < \rho < min\left(\frac{m_2}{m_1}, \cdots, \frac{m_T}{m_{T-1}}, 1\right)$ in the non-stationary time series set up.

As far as the covariance of the counts $y_1, \cdots, y_t, \cdots, y_T$ are concerned, they may be directly obtained from (3.3) by using the binomial thinning operation explained in (2.20). To be specific, the lag-l $(l = 1, 2, \cdots, T-1)$ covariance can be obtained as follows by induction.

lag-1 covariance :

$$\begin{aligned}
cov\left(Y_t, Y_{t-1}\right) &= E(Y_tY_{t-1}) - E(Y_t)E(Y_{t-1}) \\
&= E_{Y_{t-1}}E_{Y_t}\left(Y_tY_{t-1} \mid y_{t-1}\right) - m_tm_{t-1} \\
&= E_{Y_{t-1}}\left[Y_{t-1}E_{Y_t}(Y_t \mid y_{t-1})\right] - m_tm_{t-1} \\
&= E_{Y_{t-1}}\left[Y_{t-1}(\rho Y_{t-1} + m_t - \rho m_{t-1})\right] - m_tm_{t-1} \\
&= \rho E_{Y_{t-1}}\left(Y_{t-1}^2\right) + \left(m_t - \rho m_{t-1}\right)E_{Y_{t-1}}(Y_{t-1}) - m_tm_{t-1} \\
&= \rho\left(m_{t-1} + m_{t-1}^2\right) + m_{t-1}(m_t - \rho m_{t-1}) - m_tm_{t-1} \\
&= \rho m_{t-1} + m_tm_{t-1} - m_tm_{t-1} \\
&= \rho m_{t-1} \qquad (3.4)
\end{aligned}$$

lag-2 covariance :

$$cov\,(Y_t, Y_{t-2}) = E(Y_tY_{t-2}) - E(Y_t)E(Y_{t-2}),$$

where

$$\begin{aligned}
E(Y_tY_{t-2}) &= E_{Y_{t-2}}E_{Y_{t-1}}E_{Y_t}\left(Y_tY_{t-2} \mid y_{t-1}, y_{t-2}\right) \\
&= E_{Y_{t-2}}E_{Y_{t-1}}\left[Y_{t-2}E_{Y_t}(Y_t \mid y_{t-1}, y_{t-2})\right] \\
&= E_{Y_{t-2}}E_{Y_{t-1}}\left[Y_{t-2}\left\{\rho\left(Y_{t-1} \mid y_{t-2}\right) + m_t - \rho m_{t-1}\right\}\right] \\
&= E_{Y_{t-2}}\left[\rho Y_{t-2}E_{Y_{t-1}}\left(Y_{t-1} \mid y_{t-2}\right)\right] + (m_t - \rho m_{t-1})E_{Y_{t-2}}(Y_{t-2}) \\
&= E_{Y_{t-2}}\left[\rho Y_{t-2}\left(\rho Y_{t-2} + m_{t-1} - \rho m_{t-2}\right)\right] + \left[m_{t-2}(m_t - \rho m_{t-1})\right] \\
&= \rho^2 E_{Y_{t-2}}(Y_{t-2}^2) + \rho(m_{t-1} - \rho m_{t-2})E_{Y_{t-2}}(Y_{t-2}) + m_{t-2}(m_t - \rho m_{t-1}) \\
&= \rho^2(m_{t-2} + m_{t-2}^2) + \rho m_{t-2}(m_{t-1} - \rho m_{t-2}) + m_{t-2}(m_t - \rho m_{t-1}) \\
&= \rho^2 m_{t-2} + m_t m_{t-2}
\end{aligned}$$

Consequently, one obtains

$$cov\,(Y_t, Y_{t-2}) = \rho^2 m_{t-2}. \tag{3.5}$$

Next, by similar calculations i.e., by induction, one obtains the lag-l covariance as

$$cov(Y_t, Y_{t-l}) = \rho^l m_{t-l} \tag{3.6}$$

It then follows that the lag-l correlation between y_t and y_{t-l} has the form

$$\begin{aligned}
\rho_y(l) &= \frac{cov(Y_t, Y_{t-l})}{\sqrt{V(Y_t)V(Y_{t-l})}} \\
&= \frac{\rho^l m_{t-l}}{\sqrt{m_t m_{t-l}}}
\end{aligned}$$

$$= \rho^{l}\sqrt{\frac{m_{t-l}}{m_t}}. \tag{3.7}$$

Therefore the autocorrelation matrix $C(\rho)$, for the non-stationary Poisson AR(1) counts $y_1, \cdots, y_T$ has the form

$$C(\rho) = \begin{pmatrix} 1 & \rho\sqrt{\frac{m_1}{m_2}} & \cdots & \rho^{T-1}\sqrt{\frac{m_1}{m_T}} \\ \rho\sqrt{\frac{m_1}{m_2}} & 1 & \cdots & \rho^{T-2}\sqrt{\frac{m_2}{m_T}} \\ \vdots & \vdots & & \vdots \\ \rho^{T-1}\sqrt{\frac{m_1}{m_T}} & \rho^{T-2}\sqrt{\frac{m_2}{m_T}} & \cdots & 1 \end{pmatrix}, \tag{3.8}$$

which is different than the stationary correlation matrix given in (3.2). More specifically, for the non-stationary case the correlation between two time points t and t' $(t < t')$ is $\rho_{|t-t'|} = \rho^{|t-t'|}\sqrt{\frac{m_t}{m_{t'}}}$, whereas $\rho_{|t-t'|} = \rho^{|t-t'|}$ in the stationary case. Note that in the stationary case, $m_t = m_{t'}$ and $C(\rho)$ reduces to $C^*(\rho)$ of (3.2).

3.2 Statistical Inference: Estimation of Parameters

The model (3.3) involves two unknown parameters: (i) β, the p-dimensional vector of regression parameters, and (ii) ρ, the autocorrelation parameter. In time series analysis, both β and ρ parameters are important to consider. To be specific, the regression parameter β plays an important role in obtaining the deterministic pattern of the series. This is because, the mean and variance of the data are $m_t = e^{x_t'\beta}$, which are functions of β. Next, for the purpose of forecasting a future count, it is necessary to know the deterministic pattern as well as the correlation parameter involved in the model. This is because, the forecasting is usually made by regressing the future response on the present and past observations. This type of regression function involves the correlation parameter. Thus in this section, we discuss the estimation of both β and ρ and in section 3.4 we will discuss the forecasting aspects.

In order to estimate β consistently and efficiently, we follow Sutradhar (2003, section 3) and use the Generalized Quasilikelihood (GQL) approach in the present time series case (see also Zeger(1988)). As far as the estimation of the correlation parameter ρ is concerned, we estimate this parameter consistently by using the traditional method of moments. Note, however, that when correlations are small, a 'working' GQL approach using $\rho = 0$ may perform better than the GQL approach based on the true correlation structure. This is because, when ρ is small, the estimation variability in estimating ρ in the GQL approach may yield a poorer (in the sense of mean squared errors) estimate of β as compared to the independence assumption based ($\rho = 0$) GQL approach where it is not necessary to estimate ρ.

3.2.1 Generalized Quasilikelihood Estimation for β

Recall that $m = (m_1, \cdots, m_T)$ is the T-dimensional mean vector of $y = (y_1, \cdots, y_T)$ and $\Sigma(\rho) = (\sigma_{tt'})$ is the covariance matrix of y. Here $\sigma_{tt} = V(Y_t)$ and $\sigma_{tt'} = cov(Y_t, Y_{t'})$. Note that $\sigma_{tt'}$ and hence $\Sigma(\rho)$ matrix is a function of ρ.

For known ρ, one may write the QL estimating equation for β as

$$\frac{\partial m'}{\partial \beta} \Sigma(\rho)^{-1}(y - m) = 0. \tag{3.9}$$

The GQL (generalized quasilikelihood) estimate of β is then computed by solving the estimating equation

$$\frac{\partial m'}{\partial \beta} \Sigma(\hat{\rho})^{-1}(y - m) = 0, \tag{3.10}$$

where $\hat{\rho}$ is obtained by method of moments as showed in the next subsection. Let $\hat{\beta}_{GQL}$ be the GQL estimator of β obtained from (3.10). Note that the computation of $\hat{\beta}_{GQL}$ is usually done by an iterative method. More specifically, $\hat{\beta}_{GQL}$ is obtained by using the iterative equations

$$\hat{\beta}_{GQL}(m+1) = \hat{\beta}_{GQL}(m) + \left[\left(\frac{\partial m'}{\partial \beta} \Sigma(\hat{\rho})^{-1} \frac{\partial m}{\partial \beta'} \right)^{-1} \frac{\partial m'}{\partial \beta} \Sigma(\hat{\rho})^{-1}(y - m) \right]_{\hat{\beta}_{GQL}(m)}, \tag{3.11}$$

where $\hat{\beta}_{GQL}(m)$ is the value of β at the m-th iteration and the expression in the square bracket is evaluated at $\hat{\beta}_{GQL}(m)$. Further note that, as $m = (m_1, \cdots, m_T)'$

and $\frac{\partial m_t}{\partial \beta} = (x_{t1}m_t, \cdots, x_{tp}m_t)'$, the iterative equation (3.11) may be re-expressed as

$$\hat{\beta}_{GQL}(m+1) = \hat{\beta}_{GQL}(m) + \left[\left(X'A\Sigma(\hat{\rho})^{-1}AX\right)^{-1} X'A\Sigma(\hat{\rho})^{-1}(y-m)\right]_{\hat{\beta}_{GQL}(m)}, \quad (3.12)$$

where $A = diag(m_1, \cdots, m_T)$, and

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ \vdots & \vdots & & \vdots \\ x_{t1} & x_{t2} & \cdots & x_{tp} \\ \vdots & \vdots & & \vdots \\ x_{T1} & x_{T2} & \cdots & x_{Tp} \end{pmatrix} = \begin{pmatrix} x_1' \\ \vdots \\ x_t' \\ \vdots \\ x_T' \end{pmatrix}.$$

3.2.2 Moment Equation for ρ

For a given estimated value of β, we can obtain a moment estimate of the autocorrelation parameter, $\hat{\rho}$, which is consistent for ρ. This moment estimator is obtained as

$$\hat{\rho} = \frac{\sum_{t=1}^{T-1} \tilde{y}_t \tilde{y}_{t+1}/(T-1)}{\sum_{t=1}^{T} \tilde{y}_t^2/T}, \quad (3.13)$$

(see Sutradhar (2003)) where $\tilde{y}_t = \frac{y_t - m_t}{\sqrt{m_t}}$.

Next, the estimate obtained by using (3.13) is used in (3.12) to obtain an improved estimate of β. The improved estimate of β is then used in (3.13) to obtain an improved estimate of ρ. This cycle of iteration continues until convergence, i.e., when old and new values of the estimates are reasonably very close.

3.2.3 Simplified Formula for $\hat{\beta}_{GQL}$ Under Stationary Case

Estimation under the stationary model may be carried out in the similar fashion by replacing m_t with $m^* = e^{\beta_1}$ for all $t = 1, 2, \cdots, T$ in (3.9). In fact, unlike the non-stationary case, we can obtain a closed form expression for $\hat{\beta}_{GQL,1}$ in this stationary case. More specifically, under the stationary case $m = (m_1, \cdots, m_T)'$ in (3.9) reduces to $m = m^*(1, \cdots, 1)'$, $\Sigma(\rho)$ in (3.9) reduces to $\Sigma^*(\rho) = A^{*^{1/2}}C^*(\rho)A^{*^{1/2}}$, where $A^* =$

$diag(m^*, \cdots, m^*)$, $C^*(\rho)$ is defined in (3.2). Consequently

$$\Sigma^*(\rho)^{-1} = A^{*^{-1/2}} \left[C^*(\rho)\right]^{-1} A^{*^{-1/2}},$$

where

$$A^{*^{-1/2}} = \begin{pmatrix} 1/\sqrt{m^*} & 0 & \cdots & 0 \\ 0 & 1/\sqrt{m^*} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1/\sqrt{m^*} \end{pmatrix},$$

and

$$[C^*(\rho)]^{-1} = \frac{1}{1-\rho^2} \begin{pmatrix} 1 & -\rho & 0 & \cdots & 0 & 0 & 0 \\ -\rho & 1+\rho^2 & -\rho & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -\rho & 1+\rho^2 & -\rho \\ 0 & 0 & 0 & \cdots & 0 & -\rho & 1 \end{pmatrix},$$

(Kendal and Stuart (1968, p.473-474)) yielding

$$\frac{\partial m'}{\partial \beta_1} \Sigma^*(\rho)^{-1} = \begin{bmatrix} (1-\rho) & (1-\rho)^2 & \cdots & (1-\rho)^2 & (1-\rho) \end{bmatrix}.$$

Next the estimating equation $\frac{\partial m'}{\partial \beta} \Sigma(\rho)^{-1}(y - m) = 0$ in (3.9) reduces to

$$\begin{aligned} & \frac{\partial[m^{*'} 1_T]}{\partial \beta_1} \Sigma^*(\rho)^{-1}(y - m^* 1_T) \\ = \;& (1-\rho)(y_1 - m^*) + (1-\rho)^2 \left[(y_2 - m^*) + \cdots + (y_{T-1} - m^*)\right] + (1-\rho)(y_T - m^*) \\ = \;& (1-\rho)(y_1 + y_T - 2m^*) + (1-\rho)^2 \left[\sum_{t=2}^{T-1} y_t - (T-2)m^*\right] \\ = \;& (1-\rho)(y_1 + y_T) + (1-\rho)^2 \sum_{t=2}^{T-1} y_t - 2m^*(1-\rho) - m^*(1-\rho)^2(T-2) \qquad (3.14) \end{aligned}$$

By setting (3.14) to zero, we get the equation for $\hat{\beta}_{QL,1}$ as

$$2m^*(1-\rho) + m^*(1-\rho)^2(T-2) = (1-\rho)(y_1 + y_T) + (1-\rho)^2 \sum_{t=2}^{T-1} y_t,$$

yielding

$$\hat{m}^* = e^{\hat{\beta}_{QL,1}} = \frac{y_1 + y_T + (1-\rho)\sum_{t=2}^{T-1} y_t}{T - \rho(T-2)},$$

that is,

$$\hat{\beta}_{GQL,1} = log\left[\frac{y_1 + y_T + (1-\hat{\rho})\sum_{t=2}^{T-1} y_t}{T - \hat{\rho}(T-2)}\right]. \tag{3.15}$$

3.2.4 Independence Assumption Based 'Working' GQL Approach for β Estimation

In this approach one use $\rho = 0$ in estimating β even though the counts collected in sequence are dependent. More specifically, the regression parameter β is estimated by using the 'working' GQL estimating equation

$$\frac{\partial m'}{\partial \beta} A^{-1}(y - m) = 0, \tag{3.16}$$

where $A = diag\{var(Y_1), \cdots, var(Y_T)\} = diag\{m_1, \cdots, m_T\}$. Note that the estimating equation (3.16) is a special case of (3.9) as $\Sigma(\rho)$ yields the A matrix when $\rho = 0$. Let $\hat{\beta}_{GQL}(I)$ be the solution of (3.16) in this special case which may be obtained by using the iterative equation

$$\hat{\beta}_{GQL}(I)[m+1] = \hat{\beta}_{GQL}(I)[m] + \left[\left(X'AX\right)^{-1} X'(y - m)\right]_{\hat{\beta}_{GQL}(I)[m]}, \tag{3.17}$$

where $\hat{\beta}_{GQL}(I)[m]$ is the value of β at the m-th iteration.

Note that in estimating β_1 under the stationary model, it follows from (3.15) that $\hat{\beta}_{GQL,1}(I)$ has the closed form expression given by

$$\hat{\beta}_{GQL,1}(I) = log\left[\frac{\sum_{t=1}^{T} y_t}{T}\right] = log(\bar{y}), \tag{3.18}$$

which is just a moment estimator.

3.3 Performance of the GQL Estimation Approach: A Simulation Study

Note that the observations in a time series are recorded in sequence of time and in practice one usually deals with a large time series. In this section, we conduct a simulation study to examine the performance of the GQL estimation approach discussed in section 3.2, where β is obtained by solving a GQL estimating equation and ρ is estimated by the method of moments. We also examine the performance of an independence assumption based 'working' GQL approach for the estimation of β. This approach will be referred to as the GQL(I) approach.

3.3.1 Simulation Design

Non-stationary Case

In our simulation study, we consider $T = 100$, where T is the length of the time series. Further we consider a regression set up where each of the T observations is assumed to be influenced by a number of covariates. Similar to Zeger (1988), we consider $p = 5$. That is the observations will be effected by 5 covariates, namely, linear trend and sine and cosine pairs at the annual and semi-annual frequencies, which may be thought of as a measure of seasonal effects. More specifically, for $t = 1, \cdots, T$, the mean of the count response y_t at time t has been regressed on the vector of covariates $x_t = (x_{t1}, \cdots, x_{t5})'$, where $x_{tu} = x^*_{tu} - \bar{x}^*_u$, $\bar{x}^*_u = \frac{\sum_{t=1}^T x^*_{tu}}{T}$ for $u = 1, \cdots, 5$ with

$$(x^*_{t1}, \cdots, x^*_{tu}, \cdots, x^*_{t5})' = [t/1000, cos(2\pi t/12), sin(2\pi t/12), cos(2\pi t/6), sin(2\pi t/6)]' .$$

As far as the generation of count observations is concerned, y_t has been generated by using the binomial thinning based relationship (3.3) where we consider the regression parameters representing the effects of trend and seasonal components as $\beta_1 = -0.005$, $\beta_2 = 0.5$, $\beta_3 = -0.5$, $\beta_4 = 0.5$, and $\beta_5 = -0.5$ respectively. These parameter values are chosen from Zeger (1988), even though we do not necessarily have to consider these values only. Note that for the selection of the correlation parameter ρ, one must make sure that the restriction $\rho < \frac{m_t}{m_{t-1}}$ is satisfied for all $t = 2, \cdots, 100$, where m_t's ($t = 1, \cdots, 100$) were calculated by using $m_t = e^{x_t' \beta}$. With x_t's as described above, it was found that for the selected covariates, ρ satisfies the restriction $0 < \rho < 0.36$. Consequently, for convenience, we have selected $\rho = 0.3$ for the present non-stationary Poisson AR(1) data.

Stationary Case

In the stationary case it is sufficient to consider the intercept covariate. Thus, we choose $m^* = e^{\beta_1}$. As far as the values of ρ is concerned, we can choose any ρ between 0 and 1. To examine the effect of small and large correlations in the estimation, we have chosen ρ = 0.3, 0.5, 0.7, 0.9. By using $\beta_1 = 0.5$ and a selected value of ρ, we have generated the stationary count observations by using the binomial thinning based relationship (2.19). Note that this relationship (2.19) is naturally much simpler than that of (3.3) in generating the count data.

3.3.2 Comparison of GQL and GQL(I) Estimation Approaches Under Stationary Poisson Model

In this subsection, we examine the performances of the GQL and GQL(I) estimation approaches in estimating the regression function. Note that for the estimation of the constant regression function, GQL(I) approach does not require any estimation of ρ (as $\rho = 0$ is used), whereas ρ is estimated by the method of moments under the GQL approach. As far as the simulation size is concerned, we conduct 1000 simulations and compute the simulated means (SM) and simulated standard errors (SSE) of the GQL and GQL(I) estimators of β obtained by (3.15) and (3.18) respectively. These results are reported in Table 3.1. As we mentioned earlier, the correlation parameter ρ in the GQL approach is estimated from (3.13) by using $m_t = m^* = e^{\beta_1}$. We also computed the SM and SSE of the moment estimator of ρ in the same Table 3.1.

Table 3.1: Comparison of GQL and GQL(I) estimators for the intercept parameter $\beta = 0.5$ by simulated means (SM), simulated standard errors (SSE) and simulated mean squared errors (SMSE) for selected values of true correlation parameter ρ under stationary Poisson AR(1) model

		Estimates		
ρ	Statistic	$\hat{\beta}_{GQL}(I)$[1]	$\hat{\rho}$	$\hat{\beta}_{GQL}$
0.3	SM	0.505	0.287	0.509
	SSE	0.106	0.098	0.106
	SMSE	0.011	0.010	0.011
0.5	SM	0.508	0.470	0.517
	SSE	0.134	0.095	0.135
	SMSE	0.018	0.010	0.019
0.7	SM	0.496	0.643	0.519
	SSE	0.191	0.086	0.193
	SMSE	0.036	0.011	0.037
0.9	SM	0.462	0.814	0.516
	SSE	0.356	0.084	0.345
	SMSE	0.128	0.014	0.119

[1] GQL(I): GQL with 'working' independence correlation matrix I.

Note that for convenience we have also computed the simulated mean squared errors (SMSE) for the estimators of regression as well as correlation parameters. These are also reported in the same Table 3.1 along with SM and SSE.

It is clear from Table 3.1 that the GQL estimator for β_1 performed well for all values of ρ, although the simulated standard errors get larger as ρ increases. For example, in estimating the true parameter $\beta_1 = 0.5$ we obtained $\hat{\beta}_{GQL,1} = 0.509$ with SSE 0.106 (i.e., SMSE = 0.011) when $\rho = 0.3$, and $\hat{\beta}_{GQL,1} = 0.516$ with SSE 0.345 (SMSE = 0.119) when $\rho = 0.9$. Note that for this change in values of ρ, the β_1 estimate changed slightly, whereas a large change occurred in standard error. With regard to the estimation of the correlation parameter ρ, the results in Table 3.1 show that as the true value of ρ gets larger the simulated estimates became more biased. For example, when $\rho = 0.3$, we obtained $\hat{\rho} = 0.287$ with SE 0.098, but for $\rho = 0.9$, we obtained $\hat{\rho} = 0.814$ with SSE 0.084.

With regard to the estimation performance of the independence assumption based GQL(I) approach, it is clear from Table 3.1 that this approach yields estimates with the same or smaller mean squared errors when the correlation is small. To be specific, $\hat{\beta}_{GQL,1}(I)$ has almost the same MSE as compared to that of $\hat{\beta}_{GQL,1}$ for correlation values up to 0.7. For example, when $\rho = 0.7$ we obtain $\hat{\beta}_{GQL,1}(I) = 0.496$ with SMSE = 0.036 which is approximately the same as the SMSE associated with $\hat{\beta}_{GQL,1}$ (SMSE = 0.037). One possible reason for $\hat{\beta}_{GQL,1}$ performing almost the same as $\hat{\beta}_{GQL,1}(I)$ for small ρ is that there occurs some estimation variability associated with the ρ estimation in GQL approach, whereas GQL(I) approach does not require any estimation for ρ. For large correlation values such as $\rho = 0.9$, the GQL approach (with SMSE = 0.119) appears to perform better than the GQL(I) approach (with SMSE = 0.128). It is, therefore, clear that there will be no loss or a little loss in efficiency in estimating β by using the simpler GQL(I) approach for any ρ, especially for the designs we have considered in the thesis. Note that as mentioned earlier, we are however interested in both β and ρ parameters in time series analysis. This is particularly important for the forecasting of a future count, which we deal with in section 3.4.

3.3.3 Comparison of GQL, GQL(I) and GQL(C^*) Estimation Approaches Under Non-stationary Poisson Model

Recall from subsection 3.1.2 that the correlation parameter ρ is restricted to $0 < \rho < min\left(\frac{m_2}{m_1}, \cdots, \frac{m_T}{m_{T-1}}, 1\right)$, whereas in the stationary case ρ has the range $0 < \rho < 1$. This is because in the non-stationary case the covariate values are time dependent. Note that this range restriction for ρ must be satisfied in estimating this parameter as well as the regression parameter $\beta = (\beta_1, \cdots, \beta_5)'$. For this purpose, using the parameter values as well as the values of the covariates as explained under simulation design, we have computed the true restriction for ρ which was found to be $0 < \rho < 0.36$. As mentioned under simulation design, for convenience we considered $\rho = 0.3$.

Now for the estimation of β by the iterative equation (3.12) under the non-stationary case and for the estimation of ρ by using (3.13), we have computed the SM and SSE based on 1000 simulations both for β and ρ estimators. These statistics are reported in Table 3.2 under the GQL method. For example, by using the GQL approach we have obtained $\hat{\beta}_{GQL,4} = 0.489$ with SSE 0.141 (SMSE = 0.020), and the SM of $\hat{\rho}$ was found to be 0.268 with SSE 0.112 (SMSE = 0.014). It appears that $\beta_4 = 0.5$ and $\rho = 0.3$ parameters are estimated considerably well by the GQL approach. Note however that the GQL approach has yielded almost unbiased estimates for all regression parameters, but the standard errors appear to be quite large in some cases.

Next, as it was found under the stationary case that the GQL(I) approach performed almost the same as the GQL approach in estimating β, we have also used this simpler GQL(I) approach under the non-stationary case. Furthermore, we have used another version of the GQL approach where we use $\Sigma(\rho) = \Sigma^*(\rho) = A^{1/2}C^*(\rho)A^{1/2}$ in (3.12) with $C^*(\rho)$ as a stationary correlation matrix defined in (3.2). We refer to this approach as the GQL(C^*) approach.

The SM, SSE and SMSE for estimators of the components of β and the estimator of ρ (whenever appropriate) under the GQL(I) and GQL(C^*) approaches are also reported in Table 3.2. It is clear from the table that both GQL(I) and GQL(C^*) perform better than the GQL approach in estimating the regression parameters. In particular, both of these approaches yielded smaller standard errors and smaller mean squared errors. Thus, for the estimation of the regression parameters, we recommend the use of the simpler GQL(I) or GQL(C^*) approach. Note that between these two approaches GQL(I) is naturally much simpler as it does not require the estimation of ρ at all. But as mentioned earlier, the estimation of ρ is important for the purpose of forecasting which is however beyond the scope of the present thesis for the non-stationary case.

Table 3.2: Simulated means (SM), simulated standard errors (SSE) and simulated mean squared errors (SMSE) of estimates for the regression and correlation parameter for true correlation parameter $\rho = 0.3$ under non-stationary Poisson AR(1) model with $\beta_1 = -0.005$, $\beta_2 = 0.5$, $\beta_3 = -0.5$, $\beta_4 = 0.5$, and $\beta_5 = -0.5$.

		Estimates					
Method	Statistic	$\hat{\beta}_{GQL,1}$	$\hat{\beta}_{GQL,2}$	$\hat{\beta}_{GQL,3}$	$\hat{\beta}_{GQL,4}$	$\hat{\beta}_{GQL,5}$	$\hat{\rho}$
GQL	SM	-0.004	0.515	-0.446	0.489	-0.500	0.268
	SSE	0.017	0.592	0.878	0.141	0.507	0.112
	SMSE	0.000	0.351	0.774	0.020	0.257	0.014
$GQL(C^*)$[1]	SM	-0.004	0.495	-0.472	0.491	-0.512	0.267
	SSE	0.004	0.168	0.179	0.138	0.141	0.110
	SMSE	0.000	0.028	0.033	0.019	0.020	0.013
$GQL(I)$[2]	SM	-0.004	0.500	-0.476	0.486	-0.512	–
	SSE	0.004	0.170	0.178	0.139	0.141	–
	SMSE	0.000	0.029	0.032	0.020	0.020	–

[1] GQL(C^*): GQL with 'working' correlation matrix $C^*(\rho)$.
[2] GQL(I): GQL with 'working' independence correlation matrix I.

3.4 Statistical Inference: Forecasting

In this section we discuss the forecasting aspect for the stationary Poisson AR(1) time series. For simplicity we consider one step ahead forecasting. For this purpose, we provide the formula for the forecasting function under the stationary Poisson AR(1) model as follows. Following the forecasting technique under the Gaussian set up, the forecasting function in the present set up can be directly obtained from the binomial thinning based relationship

$$y_t = \rho * y_{t-1} + d_t = \sum_{j=1}^{y_{t-1}} b_j(\rho) + d_t,$$

(see eqn(2.19)), where $y_{t-1} \sim P(m^*)$ and $d_t \sim P(m^*(1-\rho))$ with $m^* = e^{\beta_1}$. To be specific, the formula for the forecast function is computed as

$$\begin{aligned} E(Y_t \mid y_{t-1}) &= \rho y_{t-1} + m^* - \rho m^* \\ &= m^* + \rho(y_{t-1} - m^*). \end{aligned} \tag{3.19}$$

Note that in this section, we examine through a simulation study the performance of the forecasting function (3.19) for the cases when the model is fitted by both GQL and GQL(I) approaches. Thus, under the GQL approach, the forecasting function (3.19) is estimated by

$$\hat{y}_{t,GQL} = \hat{E}(Y_t \mid y_{t-1}) = \hat{m}^* + \hat{\rho}(y_{t-1} - \hat{m}^*), \tag{3.20}$$

where $\hat{m}^* = e^{\hat{\beta}_{GQL,1}}$, whereas under the GQL(I) approach the forecasting function (3.19) is obtained by

$$\hat{y}_{t,GQL}(I) = \hat{\hat{m}}^*, \tag{3.21}$$

where $\hat{\hat{m}}^* = e^{\hat{\beta}_{GQL,1}(I)}$.

For the purpose of comparing the performance of the forecasting formulae (3.20) and (3.21), we conduct 1000 simulations and generate $T = 101$ observations in each simulation. Here we use the first 100 observations in each simulation to estimate β_1 and ρ under the GQL estimation and only β_1 under the GQL(I) approach. Then using $\hat{\beta}_{GQL,1}$ and $\hat{\rho}$ in (3.20) and $\hat{\beta}_{GQL,1}(I)$ in (3.21) we obtain the forecasted values for the 101-*th* observation. To see the performances of the forecasting functions (2.20) and (2.21) we exhibit y_{101}, $\hat{y}_{101,GQL}$ and $\hat{y}_{101,GQL}(I)$ in Figure A.1 for the first and last 50 simulations, when $\rho = 0.3$. Note that as we are forecasting the future count, we have also considered the integer approximation forecasting. That is, in each simulation, we converted the real values of $\hat{y}_{101,GQL}$ and $\hat{y}_{101,GQL}(I)$ to the corresponding integer values, and these values along with true y_{101} are plotted in Figure A.2 for the same correlation $\rho = 0.3$. It is clear from both figures A.1 and A.2 that the forecasting based on the GQL approach performs better than the GQL(I) approach. This is because the forecasting values based on the GQL approach appears to be closer to the true values of y_{101} than the forecasting values produced by the GQL(I) approach. This is not surprising as the GQL approach incorporates the correlation parameter in forecasting the future count, whereas GQL(I) always uses $\rho = 0$. For $\rho = 0.9$, we have done similar forecasting, and forecasted values based on the GQL and GQL(I) approaches are plotted in Figure A.3 (corresponding to Figure A.1) and Figure A.4

Table 3.3: Simulated mean forecast (SMF), simulated standard error (SSE) of the forecasted values and forecasted mean squared error (FMSE) for $\beta_1 = 0.5$ and selected values of true correlation parameter ρ under stationary Poisson AR(1) model using (i) 'working' independence (ii) correct correlation structure.

ρ	Simulated average y_{101}	Statistic	GQL(I) Approach	GQL Approach
0.3	3	SMF	2	2
		SSE	0.173	0.409
		FMSE	5.512	4.452
0.5	3	SMF	2	2
		SSE	0.232	0.662
		FMSE	7.275	5.379
0.7	4	SMF	2	2
		SSE	0.301	0.877
		FMSE	10.560	7.277
0.9	4	SMF	2	2
		SSE	0.538	1.133
		FMSE	11.345	7.679

(corresponding to Figure A.2). When these Figures A.3 and A.4 are compared with Figure A.1 and A.2 respectively, it is clear that GQL approach performs much better as the correlation gets larger.

Note that in order to get an overall idea about the forecasting performance of (3.20) and (3.21), we have also computed the forecasted mean squared errors (FMSE) given by

$$FMSE = \frac{\sum_{s=1}^{1000}(y_{101,s} - \hat{y}_{101,s})^2}{1000} \tag{3.22}$$

under the GQL and GQL(I) approaches. These FMSEs under the GQL and GQL(I) approaches along with the simulated average of y_{101} are given in Table 3.3. It is clear from the table that the FMSEs based on the GQL approach are uniformly smaller than the FMSEs based on the GQL(I) approach. This in turn shows that the correlation structure plays an important role in modelling the time series of counts.

Chapter 4

Analysis of Stationary and Non-stationary Time Series for Negative Binomial Counts: A Parameter-driven Approach

Recall that in the last chapter we analyzed time series of counts when they follow Poisson distribution marginally. Under this Poisson model the data exhibit the same mean and variance. But in practice there may, however, be situations when variance in the count time series may be greater than the mean. If this happens, it may be reasonable to assume that the count responses are marginally overdispersed. This type of time series where each response is subject to overdispersion may be modelled through correlated negative binomial distribution. Zeger (1988) has analysed similar time series of counts where correlations are assumed to be generated through a latent process. See also Harvey and Fernandes (1989) for a similar analysis, where the conditional distribution of counts was modelled using a negative binomial distribution.

In this chapter, following Zeger (1988) we discuss in brief a parameter-driven correlation model for the time series of negative binomial counts. To be specific, the negative binomial correlation model is generated by using a Poisson-gamma mixed model. This is shown in section 4.1 for non-stationary count data and in 4.2 for the stationary count data. Note that on top of regression and correlation parameters of the Poisson model, there is an additional overdispersion parameter in the negative binomial model. Moreover the correlation parameter under the parameter-driven model explains the correlations of the random effects rather than the correlations of the observations.

In section 4.3, we deal with the estimation of the parameters associated in such a parameter-driven model for the stationary and non-stationary models. In section 4.4, we conduct a simulation study to examine the performance of the estimation approaches to be discussed in section 4.3. Note however that it is not easy to derive the forecasting function in this parameter-driven set up, which is a drawback of the parameter-driven approach. The forecasting issues however will be revisited in the next chapter under the observation-driven correlation model.

4.1 Non-stationary Negative Binomial Counts

Suppose that $\mu_t^* = m_t\theta_t = e^{x_t'\beta+\gamma_t}$. Also suppose that conditional on the random effect θ_t, the count observation y_t follow Poisson distribution with mean μ_t^*, i.e.,

$$f(y_t \mid \theta_t) = \frac{e^{-\mu_t^*}\mu_t^{*y_t}}{y_t!}, y_t = 0, 1, \cdots, \tag{4.1}$$

Now if θ_t has gamma distribution denoted by $G(\frac{1}{\alpha}, \frac{1}{\alpha})$ of the form:

$$f(\theta_t) = \frac{1}{\Gamma(\frac{1}{\alpha})\alpha^{1/\alpha}}(\theta_t)^{\frac{1}{\alpha}-1}e^{-\frac{\theta_t}{\alpha}}, \tag{4.2}$$

then, one may obtain that $y_t \sim NB(\frac{1}{\alpha}, \alpha m_t)$, i.e., the marginal distribution is negative binomial with the following form:

$$f(y_t) = \frac{\Gamma(\frac{1}{\alpha}+y_t)}{\Gamma(\frac{1}{\alpha})y_t!}\left(\frac{\alpha m_t}{1+\alpha m_t}\right)^{y_t}\left(1-\frac{\alpha m_t}{1+\alpha m_t}\right)^{1/\alpha}. \tag{4.3}$$

Accordingly, $E(Y_t) = m_t$ and $V(Y_t) = m_t + \alpha m_t^2$. It is clear that the distribution of y_t accommodates overdispersion indexed by α.

Recall from Chapter 2 that Zeger (1988) did not assume any distribution for θ_t, but he assumed $\{\theta_t\}$ to be a non-negative time series with $E(\theta_t) = 1$, $V(\theta_t) = \alpha$ and $cov(\theta_t, \theta_{t-l}) = \alpha\rho_\theta(l)$. One of the negative consequences of this assumption associated with θ_t with no distribution is that one can not derive the marginal and conditional distributions of the responses, which naturally hampers the efforts for forecasting, an important issue in time series. To avoid these complications, suppose that θ_t follows a gamma distribution marginally as in (4.2), and we attempt to generate an autocorrelation structure with lag l correlation $\rho_\theta(l) = corr(\theta_t, \theta_{t-l})$ without losing this gamma marginal distribution. More specifically, for this we follow McKenzie (1988) and relate θ_t with θ_{t-1} as

$$\theta_t = A_t\theta_{t-1} + B_t, \tag{4.4}$$

where θ_{t-1} has gamma distribution, i.e., $\theta_{t-1} \sim G(\frac{1}{\alpha}, \frac{1}{\alpha})$, and A_t has the beta distribution, $A_t \sim Be(\lambda, \frac{1}{\alpha} - \lambda)$, yielding $A_t\theta_{t-1} \sim G(\lambda, \frac{1}{\alpha})$. Here the beta density of A_t has the form

$$\frac{1}{B(\lambda, \frac{1}{\alpha} - \lambda)} A_t^{\lambda-1}(1 - A_t)^{\frac{1}{\alpha}-\lambda-1},$$

where $B(\lambda, \frac{1}{\alpha} - \lambda) = \frac{\Gamma(\lambda)\Gamma(\frac{1}{\alpha}-\lambda)}{\Gamma(\frac{1}{\alpha})}$. Next under the assumption that $B_t \sim G(\frac{1}{\alpha} - \lambda, \frac{1}{\alpha})$, it follows that θ_t has $G(\frac{1}{\alpha}, \frac{1}{\alpha})$ distribution which is exactly the same as in (4.2). Consequently, $E(\theta_t) = 1$ and $V(\theta_t) = \alpha$. Next, by some algebra it can be shown from (4.4) that

$$cov(\theta_t, \theta_{t-l}) = \alpha\rho_\theta(l) = \alpha(\alpha\lambda)^l = \alpha\rho^{*l}, \tag{4.5}$$

where $\rho^* = \rho_\theta(1) = \alpha\lambda$ is the lag-1 correlation of $\{\theta_t\}$. Note that the gamma and Poisson mixture produces the negative binomial marginal distribution for y_t given by (4.3) with $E(Y_t) = m_t$ and $V(Y_t) = m_t + \alpha m_t^2$, but the covariance structure of y_t is determined by the covariance structure of θ_t as follows:

$$\begin{aligned} cov(Y_t, Y_{t-l}) &= E_{\theta_t} cov(Y_t, Y_{t-l} \mid \theta_t) + cov_{\theta_t}\left[E(Y_t \mid \theta_t), E(Y_{t-l} \mid \theta_{t-l})\right] \\ &= cov_{\theta_t}(m_t\theta_t, m_{t-l}\theta_{t-l}) \end{aligned}$$

$$= m_t m_{t-l} \alpha \rho_\theta(l)$$

$$= m_t m_{t-l} \alpha \rho^{*l}. \tag{4.6}$$

It then follows that

$$corr(Y_t, Y_{t-l}) = \rho_y(l) = \frac{m_t m_{t-l} \alpha \rho^{*l}}{\sqrt{(m_t + \alpha m_t^2)(m_{t-l} + \alpha m_{t-l}^2)}}. \tag{4.7}$$

It is now clear from (4.7) that unlike the Gaussian case, the negative binomial observations have a complicated correlation structure. Furthermore, even though we have used the gamma distribution for the random effects resulting in the correlation structure (4.7), it is not easy to derive the necessary joint and conditional distributions of the count responses for the purpose of forecasting.

4.2 Stationary Negative Binomial Counts

The parameter-driven model for stationary negative binomial counts can be obtained directly by using the relationship (4.4) and substituting $m_t = e^{x_t^{'}\beta} = e^{\beta_1} = m^*$ in (4.1). In the stationary case, we have $E(Y_t) = m^*$, $V(Y_t) = m^* + \alpha m^{*2}$, and $cov(Y_t, Y_{t-l}) = m^{*2}\alpha\rho^{*l}$. This yields the lag-$l$ correlations of the responses as $corr(Y_t, Y_{t-l}) = \frac{\rho^{*l}}{1+\frac{1}{\alpha m^*}}$.

4.3 Estimation of Parameters

4.3.1 Non-stationary Case

The parameter-driven model described above involves 3 unknown parameters: (i) β, the p-dimensional vector of regression parameters (ii) ρ^*, the autocorrelation parameter of the random effects, and (iii) α, the overdispersion parameter. We estimate the main parameter β by using the GQL approach and the other two parameters α and ρ^* by method of moments. Recall that ρ^*, the autocorrelation of the random effects, determines the autocorrelation of the observations y_t in a complicated way. It is in general true that even if ρ^* is large, the correlations of the observations are usually small which is determined by (4.7).

GQL Estimator of β

The GQL estimating equation for β is similar to that of (3.12). More specifically, following (3.12) we now obtain the estimate of β by using the iterative equation as

$$\begin{aligned}\hat{\beta}_{GQL}(m+1) &= \hat{\beta}_{GQL}(m) + \Big[\left(X'A\Sigma(\hat{\rho}^*,\hat{\alpha})^{-1}AX\right)^{-1} \\ &\qquad X'A\Sigma(\hat{\rho}^*,\hat{\alpha})^{-1}(y-m)\Big]_{\hat{\beta}_{GQL}(m)}, \end{aligned} \tag{4.8}$$

where y is now a vector of negative binomial responses, m representing its mean vector and A has the same notation as in section 3.2. Note that here the covariance matrix Σ is function of both ρ^* and α as given by (4.6), where in (3.12) Σ is a function of the observation correlation, ρ only.

The correlation parameter ρ^* and overdispersion parameter α involved in the iterative equation may be estimated by the method of moments as given below.

Moment Equation for α

Note that both the variance and lag-1 covariance of y_t's are functions of α. Consequently, a moment equation for α can be obtained by equating the following expression

$$E\left[(y_t-m_t)^2+(y_t-m_t)(y_{t-1}-m_{t-1})\right] = m_t+\alpha m_t^2+\alpha\rho^* m_t m_{t-1}.$$

It then follows that the moment estimator of α has the following form

$$\hat{\alpha} = \frac{\frac{\sum_{t=1}^T (y_t-\hat{m}_t)^2}{T} + \frac{\sum_{t=2}^T (y_t-\hat{m}_t)(y_{t-1}-\hat{m}_{t-1})}{T-1} - \frac{\sum_{t=1}^T \hat{m}_t}{T}}{\frac{\sum_{t=1}^T \hat{m}_t{}^2}{T} + \hat{\rho}^* \frac{\sum_{t=2}^T \hat{m}_t\hat{m}_{t-1}}{T-1}}. \tag{4.9}$$

Moment Equation for ρ^*

Similar to the Gaussian time series, the sample lag-1 correlation for the observation $\{y_t\}$ can be obtained as

$$\hat{\rho}_y(1) = \frac{c\hat{o}v(y_t, y_{t-1})}{\hat{V}(y_t)} = \frac{\sum_{t=2}^{T}(y_t - \hat{m}_t)(y_{t-1} - \hat{m}_{t-1})/(T-1)}{\sum_{t=1}^{T}(y_t - \hat{m}_t)^2/T}.$$

Now by equating this sample correlation with its population counterpart, we can obtain the moment estimator of ρ^* by solving

$$E[\hat{\rho}_y(1)] \simeq \frac{E\left[\sum_{t=2}^{T}(y_t - \hat{m}_t)(y_{t-1} - \hat{m}_{t-1})/(T-1)\right]}{E\left[\sum_{t=1}^{T}(y_t - \hat{m}_t)^2/T\right]} = \frac{\sum_{t=2}^{T}\alpha\rho^* m_t m_{t-1}/(T-1)}{\sum_{t=1}^{T}(m_t + \alpha m_t^2)/T}.$$

To be specific, the moment estimator of ρ^* has the form

$$\hat{\rho}^* = \frac{\hat{\rho}_y(1)\sum_{t=1}^{T}(\hat{m}_t + \alpha\hat{m}_t^2)/T}{\alpha\sum_{t=2}^{T}\hat{m}_t\hat{m}_{t-1}/(T-1)}. \qquad (4.10)$$

4.3.2 Stationary Case

In the stationary case, we use the same moment equation for α and ρ^* by replacing $m_t = m^* = e^{\beta_1}$ in (4.9) and (4.10) respectively. However we obtain a simplified expression for the GQL estimator of β_1. Assuming $\Sigma^{-1} = S$, the QL estimating equation for β_1 in the stationary case becomes

$$\begin{aligned} \frac{\partial[m^{*'}1_T]}{\partial\beta_1}S(y - m^*1_T) &= 0 \\ \Rightarrow m^*1_T'S(y - m^*1_T) &= 0 \end{aligned}$$

It then follows that

$$\begin{aligned} m^* &= \frac{\sum_{t=1}^{T}\sum_{t'=1}^{T}S_{tt'}y_{t'}}{\sum_{t=1}^{T}\sum_{t'=1}^{T}S_{tt'}} \\ \Rightarrow \hat{\beta}_{QL,1} &= log\left[\frac{\sum_{t=1}^{T}\sum_{t'=1}^{T}S_{tt'}y_{t'}}{\sum_{t=1}^{T}\sum_{t'=1}^{T}S_{tt'}}\right], \end{aligned} \qquad (4.11)$$

where $S_{t,t'}$ is the $(t,t')th$ element of the $S = \Sigma^{-1}$ matrix. Note that the computation of $\hat{\beta}_{GQL,1}$ by (4.11) requires $\hat{\rho}^*$ from (4.10) and $\hat{\alpha}$ from (4.9) for the S matrix.

4.4 Performance of GQL Estimation Approach: A Simulation Study

4.4.1 Simulation Design

Non-stationary Case

Recall that in Chapter 3 we generated Poisson time series based on the observation-driven model (3.3). Thus $T = 100$ observations were generated so that they have the lag-1 correlation ρ. Now to generate $T = 100$ negative binomial counts, we first generate $\theta_1, \cdots, \theta_t, \cdots, \theta_{100}$ following the observation-driven relationship (4.4). To be specific, to generate these 100 values for the random effects we need to know the values of ρ^* and α, where ρ^* is the lag-1 correlation of θ's and α is a scale parameter. We consider $\rho^* = 0.5$, 0.7 and 0.9 and $\alpha = 0.1$, 0.5, 0.75 and 1 in the simulation study. Note that the variance of y_t gets affected by α and the correlations among y_t's are a function of both α and ρ^*.

Next, for the realized value of θ_t, we generate a Poisson observation y_t with mean $\mu_t^* = m_t\theta_t$ where $m_t = e^{x_t'\beta}$. These Poisson observations are in fact negative binomial observations as θ_t was generated from a marginal gamma distribution. It is clear that the generation of these responses require x_t and β. We consider $\beta = (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5)' = (-0.005, 0.5, -0.5, 0.5, -0.5)$ as in the Poisson simulation study under the non-stationary Poisson model. As far as the covariates are concerned, we consider the same x_t as in section 3.3.1 for $t = 1, \cdots, 100$.

Stationary Case :

In the stationary case, $m_t = m^* = e^{\beta_1}$ for all t. Here, we consider $\beta_1 = 0.5$. For the given value of θ_t (generated based on ρ^* and α) we now generate Poisson y_t with mean $m_t = m^*$. These observations are then stationary negative binomial observations. As far as the values of α and ρ^* are concerned, we use the same values for them as in the non-stationary case.

4.4.2 Estimation Performance

Note that in order to have some sense about the correlations of the negative binomial responses as the functions of α and ρ^* (ρ^* being the correlation of random effects), we first report the lag-1, lag-2 and lag-3 correlations of the data in Table 4.1 when the observations were generated based on selected values of α and ρ^*. It is clear that even if ρ^* is large as $\rho^* = 0.9$, the correlations of the data appear to be small. For example, when $\alpha = 1$ and $\rho^* = 0.9$, the negative binomial responses exhibit only 0.47 as lag-1 correlation for the non-stationary case and this correlation value 0.429 for the stationary case. These results show that one has to be careful in interpreting the estimate of ρ^* under the present set up, as it is not actually the correlation of the data.

We now turn back to the performance of the GQL approach in estimating β and the performance of moment approaches in estimating α and ρ^*. More specifically, we obtain the GQL estimate of β by (4.8) and moment estimates of α and ρ^* by (4.9) and (4.10) respectively. Note that when α is small such as $\alpha = 0.1$ and ρ^* is close to the boundary such as $\rho^* = 0.95$, the moment method may yield inadmissible estimates such as $\hat{\alpha} < 0$ and $\hat{\rho}^* > 1$. But when α gets larger, the number of cases with negative $\hat{\alpha}$ become smaller. For example, when $\alpha = 0.1$,, there were 327 negative $\hat{\alpha}$ values out of 1000 for the case $\rho^* = 0.5$, but when α was increased to 0.5, there were 19 negative $\hat{\alpha}$ values. In order to determine the extent of this problem with the moment method for small α, we report the simulated means (SM), simulated standard errors (SSE), simulated median (SMD) and simulated mean absolute deviation (SMAD) in Table 4.2 for β, α and ρ^* estimates, when $\alpha = 0.1$ only, under two scenarios: first, by using only those simulated estimates which were obtained based on the restriction $\hat{\rho}^* < 1$, and second, by using simulated estimates with $\hat{\alpha} > 0$ and $\hat{\rho}^* < 1$. Note that even though SM and SSE are standard measures to understand the behavior of the estimates, it was however found in the simulation study that the estimates $\hat{\alpha}$ and $\hat{\rho}^*$ usually exhibit skewed distributions which makes these SM and SSE measure less meaningful. Consequently, we have also incorporated the SMD and SMAD, where

Table 4.1: Lag-1, 2 and 3 correlations of observations for selected values of ρ^* (lag-1 correlation of random effects) and α under stationary and non-stationary set up.

α	ρ^*	$\rho_y(l)$	Stationary Case	Non-stationary Case
0.1	0.5	$\rho_y(1)$	0.057	0.314
		$\rho_y(2)$	0.025	0.021
		$\rho_y(3)$	0.005	-0.161
	0.7	$\rho_y(1)$	0.080	0.324
		$\rho_y(2)$	0.051	0.030
		$\rho_y(3)$	0.030	-0.150
	0.9	$\rho_y(1)$	0.090	0.339
		$\rho_y(2)$	0.080	0.041
		$\rho_y(3)$	0.064	-0.142
0.5	0.5	$\rho_y(1)$	0.201	0.332
		$\rho_y(2)$	0.087	0.044
		$\rho_y(3)$	0.022	-0.104
	0.7	$\rho_y(1)$	0.268	0.386
		$\rho_y(2)$	0.173	0.090
		$\rho_y(3)$	0.111	-0.075
	0.9	$\rho_y(1)$	0.314	0.427
		$\rho_y(2)$	0.273	0.139
		$\rho_y(3)$	0.232	-0.038
0.75	0.5	$\rho_y(1)$	0.240	0.322
		$\rho_y(2)$	0.107	0.048
		$\rho_y(3)$	0.047	-0.085
	0.7	$\rho_y(1)$	0.334	0.403
		$\rho_y(2)$	0.212	0.110
		$\rho_y(3)$	0.132	-0.051
	0.9	$\rho_y(1)$	0.386	0.451
		$\rho_y(2)$	0.328	0.170
		$\rho_y(3)$	0.275	-0.002
1.0	0.5	$\rho_y(1)$	0.273	0.334
		$\rho_y(2)$	0.118	0.066
		$\rho_y(3)$	0.041	-0.067
	0.7	$\rho_y(1)$	0.369	0.416
		$\rho_y(2)$	0.230	0.129
		$\rho_y(3)$	0.143	-0.028
	0.9	$\rho_y(1)$	0.429	0.470
		$\rho_y(2)$	0.364	0.190
		$\rho_y(3)$	0.304	0.025

for simulation size 1000, SMAD of any estimator τ_s under *s-th* simulation is

$$SMAD = \frac{Median|\tau_s - Median(\tau_s)|}{0.6745}, s = 1, 2, \cdots, 1000.$$

The results in Table 4.2 show that the GQL approach produces reasonable estimates for the regression components even though $\hat{\alpha}$ was negative in many situations. As far as the moment estimate of α is concerned, $\hat{\alpha}$ appears to underestimate α which is not surprising as a significant number of simulations yielded $\hat{\alpha} < 0$. Also the moment estimator $\hat{\rho}^*$ underestimated ρ^*. The estimates represented by the sample median were however found to be better than the estimate of ρ^* represented by the mean.

Next, in order to see the improvement in estimates of α and ρ^*, we have excluded those simulations which had $\hat{\alpha} < 0$, and computed the SM, SSE, SMD and SMAD for all estimators as in the above for the case of $\alpha = 0.1$ only. These results are reported at the bottom part of the same table. It is clear from this part of the table that $\hat{\alpha}$ and $\hat{\rho}^*$ are now less biased as compared to the estimates under the first scenario.

Note that as mentioned earlier, the results in Table 4.2 show the magnitude of the problem in estimating α and ρ^* by the moment method when α is small. In other words, these results in Table 4.2 show that moment method may not be practical for the case when α is small.

We now examine the performance of the same estimation procedure (i.e. GQL for β and moment method for α and ρ) for the cases when α is considerably larger such as $\alpha = 0.5$, 0.75 and 1.0. The simulation results are reported in Table 4.3. Note that the distributions of $\hat{\alpha}$ and $\hat{\rho}^*$ were found to be symmetric in general for all cases with $\alpha = 0.5$, 0.75 and 1.0. We have, therefore, reported the SM and SSE only. The results of the table show that the GQL estimates of β and moment estimates of ρ^* perform well in general. For example, when $\alpha = 0.75$ and $\rho^* = 0.7$, the moment estimate of ρ^* was found to be $\hat{\rho}^* = 0.672$ and the estimates of the components of β were $\hat{\beta}_{GQL,1} = -0.005$, $\hat{\beta}_{GQL,2} = 0.497$, $\hat{\beta}_{GQL,3} = -0.493$, $\hat{\beta}_{GQL,4} = 0.486$, $\hat{\beta}_{GQL,5} = -0.482$ which are almost unbiased for the corresponding true values $\beta_1 = -0.005$, $\beta_2 = 0.5$, $\beta_3 = -0.5$, $\beta_4 = 0.5$, and $\beta_5 = -0.5$.

Table 4.2: Simulated means(SM), simulated standard errors (SSE), simulated median (SMD) and simulated mean absolute deviation (SMAD) of GQL estimates of the regression parameter and moment estimates of correlation parameter ρ^* and overdispersion parameter α, for selected values of ρ^* and $\alpha = 0.1$ only under non-stationary negative binomial model with $\beta_1 = -0.005$, $\beta_2 = 0.5$, $\beta_3 = -0.5$, $\beta_4 = 0.5$, and $\beta_5 = -0.5$.

			Estimates						
α	ρ^*	Statistic	$\hat{\beta}_{GQL,1}$	$\hat{\beta}_{GQL,2}$	$\hat{\beta}_{GQL,3}$	$\hat{\beta}_{GQL,4}$	$\hat{\beta}_{GQL,5}$	$\hat{\alpha}$	$\hat{\rho}^*$
0.1	0.5	SM	-0.005	0.500	-0.490	0.484	-0.488	0.059	0.331
		SSE	0.004	0.188	0.327	0.301	0.210	0.169	0.626
		SMD	-0.005	0.501	-0.484	0.489	-0.494	0.044	0.450
		SMAD	0.004	0.148	0.174	0.152	0.143	0.104	0.741
	0.7	SM	-0.005	0.501	-0.497	0.494	-0.497	0.058	0.371
		SSE	0.005	0.338	0.665	0.760	0.516	0.175	0.616
		SMD	-0.005	0.490	-0.484	0.483	-0.491	0.041	0.522
		SMAD	0.004	0.143	0.167	0.150	0.139	0.107	0.634
	0.9	SM	-0.005	0.495	-0.494	0.488	-0.494	0.047	0.378
		SSE	0.006	0.167	0.186	0.172	0.176	0.232	0.641
		SMD	-0.005	0.494	-0.488	0.489	-0.493	0.025	0.598
		SMAD	0.005	0.141	0.165	0.144	0.147	0.104	0.521
			Estimates (after deleting negative $\hat{\alpha}$'s)						
0.1	0.5	SM	-0.005	0.503	-0.493	0.485	-0.492	0.129	0.342
		SSE	0.004	0.207	0.381	0.352	0.236	0.148	0.638
		SMD	-0.005	0.506	-0.489	0.491	-0.504	0.087	0.537
		SMAD	0.004	0.147	0.168	0.152	0.139	0.091	0.678
	0.7	SM	-0.005	0.508	-0.506	0.506	-0.507	0.131	0.451
		SSE	0.005	0.403	0.809	0.927	0.625	0.155	0.595
		SMD	-0.005	0.490	-0.489	0.491	-0.495	0.093	0.695
		SMAD	0.004	0.144	0.170	0.148	0.133	0.084	0.378
	0.9	SM	-0.005	0.496	-0.509	0.500	-0.505	0.141	0.535
		SSE	0.007	0.184	0.196	0.187	0.192	0.224	0.561
		SMD	-0.005	0.493	-0.511	0.501	-0.497	0.089	0.907
		SMAD	0.005	0.148	0.160	0.143	0.147	0.084	0.068

Table 4.3: Simulated means(SM) and simulated standard errors (SSE) of GQL estimates of the regression parameter β, moment estimates of correlation parameter ρ^* and overdispersion parameter α, for selected values of ρ^* and α under non-stationary negative binomial model with $\beta_1 = -0.005$, $\beta_2 = 0.5$, $\beta_3 = -0.5$, $\beta_4 = 0.5$, and $\beta_5 = -0.5$.

			Estimates						
α	ρ^*	Statistic	$\hat{\beta}_{GQL,1}$	$\hat{\beta}_{GQL,2}$	$\hat{\beta}_{GQL,3}$	$\hat{\beta}_{GQL,4}$	$\hat{\beta}_{GQL,5}$	$\hat{\alpha}$	$\hat{\rho}^*$
0.5	0.5	SM	-0.005	0.496	-0.494	0.486	-0.494	0.364	0.432
		SSE	0.006	0.188	0.219	0.170	0.173	0.283	0.438
	0.7	SM	-0.005	0.500	-0.487	0.486	-0.497	0.355	0.647
		SSE	0.007	0.183	0.214	0.165	0.166	0.323	0.376
	0.9	SM	-0.005	0.501	-0.490	0.486	-0.474	0.369	0.745
		SSE	0.009	0.157	0.182	0.152	0.147	0.491	0.398
0.75	0.5	SM	-0.005	0.492	-0.467	0.482	-0.483	0.556	0.444
		SSE	0.006	0.209	0.234	0.188	0.181	0.443	0.374
	0.7	SM	-0.005	0.497	-0.493	0.486	-0.482	0.502	0.672
		SSE	0.007	0.215	0.239	0.178	0.170	0.413	0.320
	0.9	SM	-0.004	0.492	-0.487	0.481	-0.481	0.518	0.808
		SSE	0.011	0.175	0.209	0.151	0.158	0.731	0.276
1.0	0.5	SM	-0.004	0.490	-0.488	0.478	-0.483	0.688	0.463
		SSE	0.007	0.238	0.247	0.197	0.204	0.531	0.324
	0.7	SM	-0.005	0.493	-0.466	0.461	-0.476	0.692	0.687
		SSE	0.012	0.239	0.647	0.558	0.318	0.635	0.272
	0.9	SM	-0.005	0.495	-0.487	0.487	-0.488	0.629	0.843
		SSE	0.017	0.209	0.216	0.175	0.177	0.902	0.197

As far as the performance of the moment approach is concerned for the estimate of α, $\hat{\alpha}$ always underestimates the true α value. This simulation study therefore suggests that one requires some bias correction for the estimation of α so that $\hat{\alpha}$ gets larger in general. If this happens, one then will not anticipate negative estimates for α when α is small such as $\alpha = 0.1$. This bias correction is however beyond the scope of the present thesis.

Furthermore, we have also conducted a simulation study to examine the performance of the GQL and moment estimates for the stationary negative binomial data. The results for selected α and ρ^* with $\beta_1 = 0.5$ are given in Table 4.4. It is clear that in general the GQL and moment estimates perform well in estimating the parameters, although for the cases with small α, $\hat{\rho}^*$ does not quite produce unbiased estimate. More specifically, biases in all 3 parameters seem to increase with ρ^* and to a lesser extent with α.

In summary, the GQL approach performs well both for non-stationary and stationary negative binomial data. The moment estimates of α and ρ^* work well in the stationary case, whereas this moment approach performs well for ρ^* only under the non-stationary case, suggesting some bias correction for α estimates.

Table 4.4: Simulated means(SM), simulated standard errors (SSE), simulated median (SMD) and simulated mean absolute deviation (SMAD) for GQL estimates of β_1, moment estimates of α and ρ^* with $\beta_1 = 0.5$ for selected values of α and ρ^* under stationary negative binomial AR(1) model.

			Estimates		
α	ρ^*	Statistic	$\hat{\beta}_{GQL,1}$	$\hat{\alpha}$	$\hat{\rho}^*$
0.1	0.5	SM	0.501	0.122	0.317
		SSE	0.095	0.138	0.649
		SMD	0.502	0.094	0.457
		SMAD	0.094	0.106	0.731
	0.7	SM	0.491	0.119	0.427
		SSE	0.114	0.125	0.615
		SMD	0.495	0.097	0.651
		SMAD	0.119	0.102	0.443
	0.9	SM	0.491	0.112	0.458
		SSE	0.155	0.116	0.635
		SMD	0.492	0.091	0.813
		SMAD	0.163	0.105	0.204
0.5	0.5	SM	0.497	0.465	0.487
		SSE	0.150	0.205	0.282
		SMD	0.499	0.441	0.495
		SMAD	0.159	0.201	0.276
	0.7	SM	0.480	0.445	0.656
		SSE	0.185	0.202	0.255
		SMD	0.476	0.417	0.674
		SMAD	0.179	0.183	0.293
	0.9	SM	0.469	0.422	0.768
		SSE	0.294	0.315	0.257
		SMD	0.481	0.370	0.889
		SMAD	0.296	0.222	0.090

(Table: 4.4 contd...)

α	ρ^*	Statistic	Estimates $\hat{\beta}_{GQL,1}$	$\hat{\alpha}$	$\hat{\rho}^*$
0.75	0.5	SM	0.478	0.701	0.473
		SSE	0.169	0.281	0.239
		SMD	0.469	0.662	0.477
		SMAD	0.164	0.253	0.237
	0.7	SM	0.476	0.688	0.659
		SSE	0.219	0.309	0.223
		SMD	0.472	0.639	0.671
		SMAD	0.209	0.274	0.247
	0.9	SM	0.429	0.606	0.791
		SSE	0.361	0.366	0.205
		SMD	0.438	0.528	0.882
		SMAD	0.372	0.292	0.101
1.0	0.5	SM	0.484	0.925	0.474
		SSE	0.182	0.349	0.209
		SMD	0.488	0.867	0.472
		SMAD	0.182	0.311	0.209
	0.7	SM	0.465	0.888	0.658
		SSE	0.255	0.420	0.202
		SMD	0.475	0.820	0.672
		SMAD	0.249	0.327	0.202
	0.9	SM	0.427	0.826	0.813
		SSE	0.407	0.668	0.176
		SMD	0.434	0.663	0.878
		SMAD	0.392	0.386	0.106

Chapter 5

Analysis of Stationary and Non-stationary Time Series for Negative Binomial Counts: An Observation-driven Approach

Recall that in Chapter 4 we have shown how to analyze a negative binomial time series, where an observation of the series was generated from a Poisson distribution conditional on a random effect. The observations become correlated through the correlation structure of the random effects. As discussed, this type of correlated negative binomial data do not follow any Gaussian type correlation pattern; rather, the correlations of the responses are complicated and difficult to interpret. As a remedy here we introduce an observation-driven negative binomial time series which yields Gaussian type correlation structure for the stationary data. Note that this type of observation-driven negative binomial time series model was first introduced by McKenzie (1986) for the stationary case. McKenzie (1986) however did not deal with a non-stationary model. Furthermore, we refer to Jowaheer and Sutradhar (2002) for the use of observation-driven negative binomial correlation model under the longitudinal set up.

In this chapter, we mainly concentrate on the time series modelling of non-stationary negative binomial data. We discuss the estimation of parameters involved in such a time series in section 5.2. The forecasting aspect under the stationary model is given in section 5.4. Note that the present negative binomial time series model may be considered as an extension of the correlation model (3.3) developed for the non-stationary Poisson data in Chapter 3.

We now turn back to the modelling of the negative binomial time series with correlations generated based on certain observation-driven process.

5.1 Correlation Structure for Negative Binomial Time Series

5.1.1 Stationary Model

As we mentioned earlier, in the observation-driven approach, responses are dependent in a natural way, i.e., this dependency may be specified as a function of past observations. Following McKenzie (1986), one may relate y_t with y_{t-1} in the following way

$$y_t = \eta_t * y_{t-1} + d_t, \tag{5.1}$$

where η_t has beta distribution, i.e., $\eta_t \sim Be\left(\frac{\rho}{\alpha}, \frac{1-\rho}{\alpha}\right)$ and y_{t-1} has the negative binomial distribution (4.3) with parameters $1/\alpha$ and αm^* ($m^* = e^{\beta_1}$), i.e., $y_{t-1} \sim NB(1/\alpha, \alpha m^*)$. Also in (5.1), $d_t \sim NB(\frac{1-\rho}{\alpha}, \alpha m^*)$. Under these distributional assumptions, y_t follows the same negative binomial distribution with parameters $1/\alpha$ and αm^*, i.e. $y_t \sim NB(1/\alpha, \alpha m^*)$. It then follows that $E(Y_t) = m^*$ and $V(Y_t) = m^* + \alpha m^{*2} = v$ (say), for $t = 1, \cdots, T$. Moreover, the relationship in (5.1) yields an AR(1) type correlation structure for which $\rho_y(l) = \rho^l = \rho_l$ is the lag-l ($l = 1, 2, \cdots, T-1$) autocorrelation between y_t and y_{t-l} (see Jowaheer and Sutradhar (2002)). It is clear from above that the mean, variance and autocorrelation function are time independent, therefore the series may be considered as stationary. It then follows that the correlation matrix of such a stationary negative binomial model (5.1) is the same as (3.2) of stationary Poisson AR(1), but the autocovariance structure of $y = (y_1, \cdots, y_t, \cdots, y_T)'$ would be different as the variances are different under two models. More specifically, here

$$\Sigma^*(\rho, \alpha) = A^{*^{1/2}} C^*(\rho) A^{*^{1/2}},$$

where $C^*(\rho)$ defined as in (3.2), but $A^* = diag(v, \cdots, v)$ with $v = m^* + \alpha m^{*2}$.

Note that unlike the stationary case, the correlation structure for the non-stationary negative binomial model would be different than that of the Poisson model. In the following subsection, we describe the generalization of the correlation structure of the non-stationary Poisson model to the non-stationary negative binomial model.

5.1.2 Non-stationary Negative Model

Note that for both the stationary and non-stationary Poisson models, one uses the same relationship (2.19) or (3.3) (i.e. $y_t = \rho * y_{t-1} + d_t$). The difference between the models lies in the assumptions made for d_t and y_{t-1}. Similarly, one must have the same relationship,

$$y_t = \eta_t * y_{t-1} + d_t \tag{5.2}$$

for both stationary and non-stationary cases, but the assumptions for the non-stationary case will be different than the stationary case given in (5.1) in section 5.1.1. For the non-stationary case, we assume that $y_{t-1} \sim NB(1/\alpha, \alpha m_{t-1})$ and $d_t \sim NB(N, P)$ with

$$N = \frac{(m_t - \rho m_{t-1})^2}{\alpha(m_t^2 - \rho m_{t-1}^2)}, P = \frac{\alpha(m_t^2 - \rho m_{t-1}^2)}{(m_t - \rho m_{t-1})} \tag{5.3}$$

whereas in the stationary case y_{t-1} and d_{t-1} have simpler distributional assumptions, namely $y_{t-1} \sim NB(1/\alpha, \alpha m^*)$ and $d_t \sim NB(\frac{1-\rho}{\alpha}, \alpha m^*)$. As far as the distribution of η_t is concerned, it has the same distribution under both stationary and non-stationary models, namely $\eta_t \sim Be\left(\frac{\rho}{\alpha}, \frac{1-\rho}{\alpha}\right)$. It then follows that $y_t \sim NB(1/\alpha, \alpha m_t)$. Also it follows that $E(Y_t) = m_t$ and $V(Y_t) = m_t + \alpha m_t^2 = v_t$ (say).

Note that in the present non-stationary case ρ is not bounded by 0 and 1 any-more. Rather, ρ must satisfy a restriction similar to that of the Poisson case. More specifically, to find the exact restriction we observe that (Johnson and Kotz (1992, p.199))

$$E(d_t) = NP = m_t - \rho m_{t-1}, \tag{5.4}$$

and

$$\begin{aligned} V(d_t) &= NPQ, Q = 1 + P \\ &= m_t + \alpha m_t^2 - \rho(m_{t-1} + \alpha m_{t-1}^2). \end{aligned} \tag{5.5}$$

Here as the negative binomial parameters associated with d_t must have to be positive in order to generate positive mean and variance, one uses the restrictions $m_t - \rho m_{t-1} > 0$ and $m_t^2 - \rho m_{t-1}^2 > 0$ simultaneously. These restrictions provide the range for ρ as $0 < \rho < min\left\{\frac{m_2}{m_1}, \cdots, \frac{m_T}{m_{T-1}}, \frac{m_2^2}{m_1^2}, \cdots, \frac{m_T^2}{m_{T-1}^2}, 1\right\}$.

Computation for Lag Covariances

Note that all possible lag covariances can be obtained in the manner similar to that of the non-stationary Poisson case. These are, in general, obtained by induction.

Lag-1 covariance

$$cov(Y_t, Y_{t-1}) = E(Y_t Y_{t-1}) - E(Y_t)E(Y_{t-1}),$$

where

$$\begin{aligned} E(Y_t Y_{t-1}) &= E_{Y_{t-1}} E_{Y_t} (Y_t Y_{t-1} \mid y_{t-1}) \\ &= E_{Y_{t-1}} [Y_{t-1} E_{Y_t}(Y_t \mid y_{t-1})] \\ &= E_{\eta_t} E_{Y_{t-1}} [Y_{t-1} E_{Y_t}(Y_t \mid y_{t-1}, \eta_t)] \\ &= E_{\eta_t} E_{Y_{t-1}} [Y_{t-1}(\eta_t Y_{t-1} + m_t - \rho m_{t-1})] \end{aligned}$$

$$
\begin{aligned}
&= E_{\eta_t} E_{Y_{t-1}} \left[\eta_t Y_{t-1}^2 + Y_{t-1}(m_t - \rho m_{t-1})\right] \\
&= E_{\eta_t} \left[\eta_t E_{Y_{t-1}}(Y_{t-1}^2) + (m_t - \rho m_{t-1}) E_{Y_{t-1}}(Y_{t-1})\right] \\
&= E_{\eta_t} \left[\eta_t (m_{t-1} + \alpha m_{t-1}^2 + m_{t-1}^2) + m_{t-1}(m_t - \rho m_{t-1})\right] \\
&= \rho(m_{t-1} + \alpha m_{t-1}^2 + m_{t-1}^2) + m_t m_{t-1} - \rho m_{t-1}^2 \\
&= \rho(m_{t-1} + \alpha m_{t-1}^2) + m_t m_{t-1}
\end{aligned}
$$

Therefore

$$
cov(Y_t, Y_{t-1}) = \rho(m_{t-1} + \alpha m_{t-1}^2) = \rho v_{t-1}, \tag{5.6}
$$

where v_{t-1} is the variance of y_{t-1}.

lag-2 covariance

$$
cov(Y_t, Y_{t-2}) = E(Y_t Y_{t-2}) - E(Y_t)E(Y_{t-2}),
$$

where

$$
\begin{aligned}
E_{Y_t}(Y_t Y_{t-2}) &= E_{\eta_t} E_{Y_{t-2}} E_{Y_{t-1}} E_{Y_t}(Y_t Y_{t-2} \mid y_{t-1}, y_{t-2}, \eta_t) \\
&= E_{\eta_t} E_{Y_{t-2}} \left[Y_{t-2} E_{Y_{t-1}} E_{Y_t}(Y_t \mid y_{t-1}, y_{t-2}, \eta_t)\right] \\
&= E_{\eta_t} E_{Y_{t-2}} \left[Y_{t-2} E_{Y_{t-1}} \{\eta_t (Y_{t-1} \mid y_{t-2}, \eta_t) + m_t - \rho m_{t-1}\}\right] \\
&= E_{\eta_t} E_{Y_{t-2}} \left[Y_{t-2} \eta_t E_{Y_{t-1}}(Y_{t-1} \mid y_{t-2}, \eta_t) + Y_{t-2}(m_t - \rho m_{t-2})\right] \\
&= E_{\eta_t} E_{Y_{t-2}} \left[Y_{t-2} \eta_t (\eta_{t-1} Y_{t-2} + m_{t-1} - \rho m_{t-2}) + Y_{t-2}(m_t - \rho m_{t-1})\right] \\
&= E_{\eta_t} E_{Y_{t-2}} \left[\eta_t \eta_{t-1} Y_{t-2}^2 + \eta_t Y_{t-2}(m_{t-1} - \rho m_{t-2}) + Y_{t-2}(m_t - \rho m_{t-1})\right] \\
&= E_{\eta_t} \left[\eta_t \eta_{t-1} E_{Y_{t-2}}(Y_{t-2}^2) + \eta_t (m_{t-1} - \rho m_{t-2}) E_{Y_{t-2}}(Y_{t-2}) \right. \\
&\qquad \left. + (m_t - \rho m_{t-1}) E_{Y_{t-2}}(Y_{t-2})\right]
\end{aligned}
$$

$$
\begin{aligned}
&= E_{\eta_t}\Big[\eta_t\eta_{t-1}(m_{t-2} + \alpha m_{t-2}^2 + m_{t-2}^2) + \eta_t m_{t-2}(m_{t-1} - \rho m_{t-2}) \\
&\quad + m_{t-2}(m_t - \rho m_{t-1})\Big] \\
&= (m_{t-2} + \alpha m_{t-2}^2 + m_{t-2}^2)E_{\eta_t}(\eta_t\eta_{t-1}) + m_{t-2}(m_{t-1} - \rho m_{t-2})E_{\eta_t}(\eta_t) \\
&\quad m_{t-2}(m_t - \rho m_{t-1}) \\
&= \rho^2(m_{t-2} + \alpha m_{t-2}^2) + m_t m_{t-2}
\end{aligned}
$$

Therefore

$$
\begin{aligned}
cov(Y_t, Y_{t-2}) &= \rho^2(m_{t-2} + \alpha m_{t-2}^2) + m_t m_{t-2} - m_t m_{t-2} \\
&= \rho^2(m_{t-2} + \alpha m_{t-2}^2) \\
&= \rho^2 v_{t-2}, \qquad (5.7)
\end{aligned}
$$

where v_{t-2} is the variance of y_{t-2}.

lag-3 covariance

$$cov(Y_t, Y_{t-3}) = E_{Y_t}(Y_t Y_{t-3}) - E(Y_t)E(Y_{t-3}),$$

where

$$E_{Y_t}(Y_t Y_{t-3}) = E_{\eta_t} E_{Y_{t-3}} E_{Y_{t-2}} E_{Y_{t-1}} E_{Y_t}(Y_t Y_{t-3} \mid y_{t-1}, y_{t-2}, y_{t-3}, \eta_t).$$

Now by similar calculation as for the lag 2 covariance, it can be shown that

$$cov(Y_t, Y_{t-3}) = \rho^3(m_{t-3} + \alpha m_{t-3}^2) = \rho^3 v_{t-3}.$$

In general, the lag-l covariance has the form

$$cov(Y_t, Y_{t-l}) = \rho^l(m_{t-l} + \alpha m_{t-l}^2) = \rho^l v_{t-l}. \qquad (5.8)$$

Once the lag covariances are computed, the autocorrelations can be obtained by using the formula

$$\rho_y(l) = \frac{cov(Y_t, Y_{t-l})}{\sqrt{(V(Y_t)V(Y_{t-l}))}} = \frac{\rho^l v_{t-l}}{\sqrt{v_t v_{t-l}}} = \rho^l \sqrt{\frac{v_{t-l}}{v_t}},$$

which is similar to (3.7) derived under the non-stationary Poisson case. Therefore the autocorrelation matrix $C(\rho)$, for the non-stationary negative binomial counts $y_t, \cdots, y_T$ has the form

$$C(\rho) = \begin{pmatrix} 1 & \rho\sqrt{\frac{v_1}{v_2}} & \cdots & \rho^{T-1}\sqrt{\frac{v_1}{v_T}} \\ \rho\sqrt{\frac{v_1}{v_2}} & 1 & \cdots & \rho^{T-2}\sqrt{\frac{v_2}{v_T}} \\ \vdots & \vdots & & \vdots \\ \rho^{T-1}\sqrt{\frac{v_1}{v_T}} & \rho^{T-2}\sqrt{\frac{v_2}{v_T}} & \cdots & 1 \end{pmatrix}, \tag{5.9}$$

which is similar to the correlation structure (3.8) under the non-stationary Poisson case.

5.2 Statistical Inference: Estimation of Parameters

Recall that in Chapter 4 we used the GQL approach to estimate β and a moment approach to estimate α and ρ^*, where ρ^* was the first lag correlation for the random effects. As opposed to ρ^*, in the present observation-driven process, ρ is the first lag correlation for the stationary negative binomial data. In this section, we use the same GQL approach but the lag-l correlation in the correlation matrix is obtained from $\rho_y(l) = \rho^l\sqrt{\frac{v_{t-l}}{v_t}}$ for the non-stationary case and $\rho_y(l) = \rho^l$ for the stationary case. On the other hand α and ρ are estimated by method of moments. Note that here ρ is the correlation parameter for stationary negative binomial responses whereas ρ^* (used in parameter-driven model) is never a direct correlation parameter for Poisson observation. Further note that for the non-stationary negative binomial model, ρ is a scale parameter which is proportional to the correlation of responses. The proposed

GQL approach for the estimation of β under the non-stationary negative binomial model is discussed in section 5.2.1. In section 5.2.1 we also provide the moment approach for estimating α and ρ.

5.2.1 Estimation of Parameters Under Non-stationary Negative Binomial Model

The GQL Estimator of β

Similar to that of the non-stationary Poisson model, we use the iterative equation

$$\begin{aligned}\hat{\beta}_{GQL}[m+1] &= \hat{\beta}_{GQL}[m] + \Big[\left(X' A\Sigma(\hat{\rho},\hat{\alpha})^{-1}AX\right)^{-1} \\ & \qquad X' A\Sigma(\hat{\rho},\hat{\alpha})^{-1}(y-m)\Big]_{\hat{\beta}_{GQL}[m]}, \qquad (5.10)\end{aligned}$$

where y is a vector of negative binomial responses, m represents its mean vector, and A has the same notation as in section 3.2. We define the estimator of β obtained from (5.10) as $\hat{\beta}_{GQL}$.

Moment Equation for ρ

Suppose that

$$\tilde{y}_t = \frac{y_t - E(Y_t)}{\sqrt{V(Y_t)}} = \frac{y_t - m_t}{\sqrt{v_t}}$$

yielding $E(\tilde{Y}_t) = 0$, and $V(\tilde{Y}_t) = 1$. Next it follows from (5.6) that $cov(Y_t, Y_{t-1}) = \rho(m_{t-1} + \alpha m_{t-1}^2) = \rho v_{t-1}$, which is a function of ρ parameter, yielding the lag 1 correlation as $corr(\tilde{Y}_t, \tilde{Y}_{t-1}) = \frac{cov(Y_t,Y_{t-1})}{\sqrt{v_t v_{t-1}}} = \rho\sqrt{\frac{v_{t-1}}{v_t}}$. Consequently,

$$E\left[\frac{\sum_{t=2}^{T}\tilde{Y}_t\tilde{Y}_{t-1}/(T-1)}{\sum_{t=1}^{T}\tilde{Y}_t^2/(T)}\right] \simeq \frac{E\left[\sum_{t=2}^{T}\tilde{Y}_t\tilde{Y}_{t-1}/(T-1)\right]}{E(\sum_{t=1}^{T}\tilde{Y}_t^2/T)} = \rho\frac{\sum_{t=2}^{T}\sqrt{v_{t-1}/v_t}}{T-1},$$

producing

$$\hat{\rho} = \frac{\sum_{t=2}^{T}\tilde{y}_t\tilde{y}_{t-1}/(T-1)}{\sum_{t=1}^{T}\tilde{y}_t^2/T}\frac{1}{\frac{\sum_{t=2}^{T}\sqrt{v_{t-1}/v_t}}{T}}, \qquad (5.11)$$

which is a moment estimator of ρ.

Moment Equation for α

As $V(Y_t) = E(Y_t - m_t)^2 = m_t + \alpha m_t^2$, the moment equation of α may be obtained by solving the equation $\frac{\sum_{t=1}^{T} E(y_t - m_t)^2}{T} = \frac{\sum_{t=1}^{T}(m_t + \alpha m_t^2)}{T}$. To be specific, one obtains

$$\hat{\alpha} = \frac{\sum_{t=1}^{T}\left[(y_t - \hat{m}_t)^2 - \hat{m}_t\right]}{\sum_{t=1}^{T} \hat{m}_t^{\,2}}. \tag{5.12}$$

Using the GQL estimate of β obtained from (5.10), we compute $\hat{\rho}$ and $\hat{\alpha}$ from (5.11) and (5.12) respectively and use them in (5.10) again to obtain an improved estimate of β. This cycle of iteration continues until convergence.

Note that so far we have estimated β by using the GQL approach. As an alternative, one may, however, like to use the independence assumption based GQL approach, which is much simpler. This 'working' independence based estimating equation for β may be obtained from (5.10) under a non-stationary model by using $\Sigma = diag[\sigma_{11}, \sigma_{22}, \cdots, \sigma_{TT}]$. We now compute α from (5.12) by replacing $\hat{m}_t$ with $\hat{m}_t(I) = e^{x_t' \hat{\beta}_{GQL}(I)}$. More specifically,

$$\hat{\alpha}(I) = \frac{\sum_{t=1}^{T}\left[(y_t - \hat{m}_t(I))^2 - \hat{m}_t(I)\right]}{\sum_{t=1}^{T} \hat{m}_t(I)^2}.$$

Further note that for the stationary model, we may similarly obtain $\hat{\beta}_{GQL,1}(I)$ and $\hat{\alpha}(I)$ by using $\Sigma = diag\{v_1, \cdots, v_T\}$ with $v_t = m^* + \alpha {m^*}^2$ for all $t = 1, \cdots, T$ in the estimating equations for β_1 and α. More specifically, for $\hat{\alpha}(I)$ we use $\hat{m}_t(I) = \hat{m}^* = e^{\hat{\beta}_{GQL,1}(I)}$.

5.2.2 Estimation of Parameters under the Stationary Negative Binomial Model

Under the stationary model we need to estimate β_1, where $m^* = e^{\beta_1}$, as well as α and ρ. Note that unlike the computation of $\hat{\beta}_{GQL}$ in the non-stationary case, one may obtain a closed form expression for the GQL estimate of β_1. To be specific, $\hat{\beta}_{GQL,1}$ has the formula

$$\hat{\beta}_{GQL,1} = log\left[\frac{y_1 + y_T + (1-\hat{\rho})\sum_{t=2}^{T-1} y_t}{T - \hat{\rho}(T-2)}\right].$$

For the estimation of α in the stationary case, we use (5.12) and replace $\hat{m}_t = \hat{m}^* = e^{\hat{\beta}_{GQL,1}}$, for all $t = 1, \cdots, T$.

Note that the correlation parameter ρ is very important in time series, particularly in the case of forecasting. To estimate this parameter, we use the moment method which is given below.

Moment Equation for ρ

Following Jowaheer and Sutradhar (2002), the correlation parameter ρ may be estimated by the method of moments as

$$\hat{\rho} = \frac{\sum_{t=1}^{T-1} \tilde{y}_t \tilde{y}_{t+1}/(T-1)}{\sum_{t=1}^{T} \tilde{y}_t^2/T},$$

where $\tilde{y}_t = \frac{y_t - m^*}{\sqrt{v}}$ and $v = m^* + \alpha {m^*}^2$. Note that in the time series set up, T is usually large. Nevertheless one may possibly improve this formula for moderately large or small T, as follows:

In the stationary negative binomial case, we have $E(Y_t) = e^{\beta_1} = m^*$, $V(Y_t) = m^* + \alpha {m^*}^2 = v$ and $cov(Y_t, Y_{t-l}) = \rho^l v$. It then follows that

$$\begin{aligned} V(\bar{Y}) &= \frac{1}{T^2}\left[\sum_t V(Y_t) + 2\sum_{t<l} cov(Y_t, Y_l)\right] \\ &= \frac{v}{T} + \frac{2v}{T^2}\left[(T-1)\rho + (T-2)\rho^2 + \cdots + \rho^{T-1}\right] \end{aligned} \quad (5.13)$$

Now

$$E\left[\sum_{t=1}^{T}(Y_t - \bar{Y})^2\right] = E\sum\left[(Y_t - m^*) - (\bar{Y} - m^*)\right]^2 = Tv - TV(\bar{Y})$$

$$\Rightarrow E\left[\frac{\sum_{t=1}^{T}(Y_t - \bar{Y})^2}{T}\right] = v - \frac{v}{T} - \frac{2v}{T} g_1(\rho), \quad (5.14)$$

where $g_1(\rho) = \left(1 - \frac{1}{T}\right)\rho + \left(1 - \frac{2}{T}\right)\rho^2 + \cdots + \left(1 - \frac{t-1}{T}\right)\rho^{T-1}$. Note that for large T,

$$E\left[\frac{\sum_{t=1}^{T}(Y_t - \bar{Y})^2}{T}\right] = v = m^* + \alpha {m^*}^2. \quad (5.15)$$

Suppose $\frac{T}{T-1} \approx 1$ for large T. It then follows that $\bar{Y}_1 = \frac{Y_2+\cdots+Y_T}{T-1} = \bar{Y} - \frac{Y_1}{T-1}$, and $\bar{Y}_2 = \frac{Y_1+\cdots+Y_{T-1}}{T-1} = \bar{Y} - \frac{Y_T}{T-1}$. Therefore

$$\begin{aligned}
E\left[\sum_{t=2}^{T}(Y_t-\bar{Y})(Y_{t-1}-\bar{Y})\right] &= \sum_{t=2}^{T}E(Y_t-m^*)(Y_{t-1}-m^*)-(T-1)E(\bar{Y}-m^*) \\
&\quad (\bar{Y}_2-m^*)-(T-1)E(\bar{Y}-m^*)(\bar{Y}_1-m^*) \\
&\quad +(T-1)E(\bar{Y}-m^*)^2 \\
&= \sum_{t=2}^{T}E(Y_t-m^*)(Y_{t-1}-m^*)+E\left[Y_1(\bar{Y}-m^*)\right] \\
&\quad +E\left[Y_T(\bar{Y}-m^*)\right]-(T-1)E(\bar{Y}-m^*)^2 \\
&= \rho(T-1)(m^*+\alpha m^{*^2})-(T-1)V(\bar{Y}) \\
&\quad +E\left[Y_1(\bar{Y}-m^*)\right]+E\left[Y_T(\bar{Y}-m^*)\right],
\end{aligned}$$

where $E\left[Y_1(\bar{Y}-m^*)\right] = E\left[Y_T(\bar{Y}-m^*)\right] = (m^*+\alpha m^{*^2})g_2(\rho)$, $g_2(\rho) = 1+\rho+\cdots+\rho^{T-1}$. Therefore we have

$$E\left[\frac{\sum_{t=2}^{T}(Y_t-\bar{Y})(Y_{t-1}-\bar{Y})}{T-1}\right] = \rho v - \frac{v}{T} - \frac{2v}{T}g_1(\rho) + \frac{2v}{T-1}g_2(\rho). \quad (5.16)$$

Note that for large T,

$$E\left[\frac{\sum_{t=2}^{T}(Y_t-\bar{Y})(Y_{t-1}-\bar{Y})}{T-1}\right] = \rho v,$$

which may be used to obtain an initial estimate of ρ as

$$\rho_0 = \frac{\sum_{t=2}^{T}(Y_t-\bar{Y})(Y_{t-1}-\bar{Y})/(T-1)}{v} = \frac{\sum_{t=2}^{T}(Y_t-\bar{Y})(Y_{t-1}-\bar{Y})/(T-1)}{\sum_{t=1}^{T}(Y_t-\bar{Y})^2/T} \quad (5.17)$$

Consequently, taking the leading ρ from (5.16), we find an improved estimator for ρ as

$$\hat{\rho} = \frac{\sum_{t=2}^{T}(Y_t-\bar{Y})(Y_{t-1}-\bar{Y})/(T-1)+v/T+2vg_1(\rho_0)/T-2vg_2(\rho_0)/T(T-1)}{v}.$$

Now by (5.14), one obtains

$$\hat{\rho} = \frac{\sum_{t=2}^{T}(Y_t - \bar{Y})(Y_{t-1} - \bar{Y})/(T-1) + v/T + 2vg_1(\rho_0)/T - 2vg_2(\rho_0)/T(T-1)}{\sum_{t=1}^{T}(Y_t - \bar{Y})^2/T + v/T + 2vg_1(\rho_0)/T}. \tag{5.18}$$

5.3 Performance of the GQL Estimation Approach: A Simulation Study

In this section, we conduct a simulation study to examine performance of the GQL approach and the independence assumption based 'working' GQL approach in estimating regression parameter under non-stationary model. We also examine this performance for the stationary model. Note that in all cases, the α and ρ parameters are estimated by appropriate moment estimating equations from section 5.2.

5.3.1 Simulation Design

Non-stationary Case

For the non-stationary negative binomial time series, we choose the same design matrix as that of the non-stationary Poisson case discussed in section 3.3. Next, the count observations y_t for $t = 1, \cdots, T$ have been generated by using the beta-binomial thinning based relationship (5.2), where the regression parameters representing the effects of trend and seasonal components were chosen as $\beta_1 = -0.005$, $\beta_2 = 0.5$, $\beta_3 = -0.5$, $\beta_4 = 0.5$, and $\beta_5 = 0.5$ respectively. As in the Poisson case, while selecting the correlation parameter ρ, one must make sure that the restriction $0 < \rho < min\left\{\frac{m_2}{m_1}, \cdots, \frac{m_T}{m_{T-1}}, \frac{m_2^2}{m_1^2}, \cdots, \frac{m_T^2}{m_{T-1}^2}, 1\right\}$ is satisfied for all $t = 2, \cdots, T$. It was found that for the selected covariates, ρ satisfies the restriction $0 < \rho < 0.13$. We have selected $\rho = 0.12$ for convenience. As far as the values of overdispersion parameter α is concerned, we consider α as 0.1, 0.5, 0.9, and 1.5.

Stationary Case

In the stationary case, we consider only the intercept parameter $\beta_1 = 0.5$. We generate the count responses y_t for $t = 1, \cdots, T$ from beta-binomial thinning based relationship (5.1) for the stationary case. The correlation parameter ρ can take any values within $0 < \rho < 1$ in the stationary case. For the simulation study we choose $\rho = 0.3$, 0.5, 0.7, and 0.9 to examine the effect of high and low correlation on the estimation of β. On the other hand α was chosen as 0.1, 0.3, 0.5, 0.7 and 0.9.

5.3.2 Comparison of the GQL and GQL(I) Approaches Under the Stationary Negative Binomial Model

The purpose of the simulation study is to examine the performance of the GQL estimation approach in estimating the regression function, while ρ and α are estimated by the method of moments. For this we conduct 1000 simulations and report the simulated means (SM) and simulated standard errors (SSE) of the GQL and GQL(I) estimators of β obtained by using the formulae in section 5.2.2. These results are reported in Table 5.1. We also computed the SM and SSE of the moment estimators of α obtained from (5.12) and ρ obtained from (5.18) which are shown in the same Table 5.1. For convenience, we also report the simulated mean squared errors (SMSE) along with these values.

We observe from Table 5.1 that the GQL estimator of the intercept parameter β_1 and moment estimator of α are negatively affected by strong correlations for any selected values of α. The biases associated with $\hat{\beta}_{GQL,1}$ and $\hat{\alpha}$ appear to be significant when the count observations are strongly correlated. For example, for $\alpha = 0.3$, $\hat{\beta}_{GQL,1} = 0.475$ with SSE 0.131 when $\rho = 0.3$, but when ρ gets as large as 0.9, we obtain $\hat{\beta}_{GQL,1} = 0.399$ with SSE 0.437. Since $\beta_1 = 0.5$, the β_1 estimates show that $\hat{\beta}_{GQL,1}$ is highly biased when ρ is large. Similarly, in estimating $\alpha = 0.5$, we obtained $\hat{\alpha} = 0.478$ with SSE 0.218 for $\rho = 0.3$, but $\hat{\alpha} = 0.403$ with SSE 0.605 for $\rho = 0.9$. This again shows that $\hat{\alpha}$ also gets largely biased when ρ gets larger. As far as the

Table 5.1: Simulated means(SM), simulated standard errors (SSE) and simulated mean squared errors (SMSE) for GQL and $GQL(I)$ estimates of β_1, moment estimates of α and ρ with $\beta_1 = 0.5$ and selected values of α and ρ under the stationary negative binomial AR(1) model.

			Estimates				
α	ρ	Statistic	$\hat{\beta}_{GQL,1}(I)^1$	$\hat{\alpha}(I)^1$	$\hat{\rho}$	$\hat{\beta}_{GQL,1}$	$\hat{\alpha}$
0.1	0.3	SM	0.488	0.115	0.286	0.485	0.117
		SSE	0.115	0.092	0.107	0.115	0.094
		SMSE	0.013	0.009	0.012	0.013	0.009
	0.5	SM	0.488	0.111	0.479	0.482	0.114
		SSE	0.149	0.104	0.099	0.148	0.109
		SMSE	0.022	0.011	0.010	0.022	0.012
	0.7	SM	0.474	0.116	0.676	0.467	0.123
		SSE	0.199	0.122	0.084	0.200	0.135
		SMSE	0.040	0.015	0.008	0.041	0.019
	0.9	SM	0.450	0.117	0.867	0.434	0.151
		SSE	0.379	0.166	0.060	0.383	0.276
		SMSE	0.146	0.028	0.005	0.151	0.079
0.3	0.3	SM	0.478	0.292	0.288	0.475	0.296
		SSE	0.131	0.164	0.113	0.131	0.167
		SMSE	0.018	0.027	0.013	0.018	0.028
	0.5	SM	0.474	0.270	0.471	0.469	0.277
		SSE	0.174	0.179	0.114	0.174	0.185
		SMSE	0.031	0.033	0.014	0.031	0.035
	0.7	SM	0.478	0.252	0.669	0.471	0.265
		SSE	0.223	0.222	0.094	0.223	0.239
		SMSE	0.050	0.052	0.010	0.050	0.058

[1] I: 'Working' Independence

(Table 5.1 Contd....)

α	ρ	Statistic	Estimates $\hat{\beta}_{GQL,1}(I)$	$\hat{\alpha}(I)$	$\hat{\rho}$	$\hat{\beta}_{GQL,1}$	$\hat{\alpha}$
0.3	0.9	SM	0.415	0.214	0.859	0.399	0.268
		SSE	0.434	0.293	0.066	0.437	0.415
		SMSE	0.196	0.093	0.006	0.201	0.173
0.5	0.3	SM	0.479	0.473	0.283	0.476	0.478
		SSE	0.145	0.214	0.119	0.145	0.218
		SMSE	0.021	0.047	0.014	0.022	0.048
	0.5	SM	0.490	0.455	0.476	0.484	0.465
		SSE	0.178	0.252	0.115	0.180	0.259
		SMSE	0.032	0.066	0.014	0.033	0.068
	0.7	SM	0.467	0.427	0.665	0.458	0.446
		SSE	0.265	0.299	0.101	0.264	0.321
		SMSE	0.071	0.095	0.011	0.071	0.106
	0.9	SM	0.390	0.336	0.856	0.371	0.403
		SSE	0.456	0.464	0.071	0.457	0.605
		SMSE	0.220	0.242	0.007	0.225	0.375
0.7	0.3	SM	0.490	0.662	0.284	0.487	0.669
		SSE	0.153	0.264	0.125	0.153	0.269
		SMSE	0.024	0.071	0.016	0.024	0.073
	0.5	SM	0.469	0.634	0.469	0.464	0.647
		SSE	0.198	0.299	0.120	0.198	0.311
		SMSE	0.040	0.094	0.015	0.041	0.100
	0.7	SM	0.467	0.610	0.663	0.457	0.640
		SSE	0.270	0.397	0.107	0.268	0.440
		SMSE	0.074	0.166	0.013	0.074	0.197

(Table 5.1 Contd....)

			Estimates				
α	ρ	Statistic	$\hat{\beta}_{GQL,1}(I)$	$\hat{\alpha}(I)$	$\hat{\rho}$	$\hat{\beta}_{GQL,1}$	$\hat{\alpha}$
0.7	0.9	SM	0.401	0.465	0.854	0.374	0.585
		SSE	0.511	0.505	0.078	0.513	0.773
		SMSE	0.271	0.310	0.008	0.279	0.611
0.9	0.3	SM	0.481	0.853	0.283	0.477	0.862
		SSE	0.168	0.329	0.124	0.167	0.333
		SMSE	0.029	0.110	0.016	0.028	0.112
	0.5	SM	0.475	0.811	0.470	0.469	0.829
		SSE	0.220	0.353	0.127	0.219	0.373
		SMSE	0.049	0.133	0.017	0.049	0.144
	0.7	SM	0.460	0.719	0.658	0.450	0.748
		SSE	0.286	0.428	0.108	0.283	0.468
		SMSE	0.083	0.216	0.013	0.083	0.242
	0.9	SM	0.363	0.633	0.849	0.328	0.805
		SSE	0.543	0.661	0.087	0.543	0.991
		SMSE	0.314	0.508	0.010	0.324	0.991

estimation of ρ is concerned, we observe that the moment estimator for ρ performs well in all cases. For example, for the true $\rho = 0.9$ and $\alpha = 0.1$, we obtained $\hat{\rho} = 0.867$ with SSE 0.060 and for true $\alpha = 0.9$, $\hat{\rho} = 0.849$ with SSE 0.087.

Recall from Chapter 3 that the 'working' independence approach GQL(I) performs well in the Poisson AR(1) model. We have also used this GQL(I) for the negative binomial time series. With regard to the estimation performance of independence assumption based GQL(I) approach, it is clear from Table 5.1 that this estimator is almost equally efficient as compared to that of GQL for almost all values of ρ. For example, when $\alpha = 0.7$ and $\rho = 0.3$, we obtained $\hat{\beta}_{GQL,1}(I) = 0.490$ with SMSE = 0.024 which is the same as the SMSE associated with $\hat{\beta}_{GQL,1}$ (SMSE = 0.024). Similarly, for $\alpha = 0.7$ and $\rho = 0.9$, the SMSE (=0.271) associated with $\hat{\beta}_{GQL,1}(I)$ is almost the same as the SMSE (= 0.279) associated with $\hat{\beta}_{GQL,1}$.

5.3.3 Comparison of the GQL and GQL(I) Approaches Under the Non-stationary Negative Binomial Model

Note that under the non-stationary negative binomial model, the GQL and GQL(I) estimates for β were obtained from (5.10) by using $\Sigma(\alpha, \rho)$ matrix based on (5.9) and $\Sigma(\alpha, \rho) = diag(\sigma_{11}, \cdots, \sigma_{TT})$, respectively. The α and ρ parameters were estimated by using the moment estimating equations from (5.12) and (5.11) respectively under the GQL approach and α (as $\rho = 0$) was estimated by $\hat{\alpha}(I)$ from section 5.2.1. The simulated results are reported in Table 5.2.

It is clear from Table 5.2 that the GQL approach produces almost unbiased estimates for the regression parameters in the non-stationary case, whereas the SSE gets larger when the value of overdispersion parameter α increases. For example, for $\beta_2 = 0.5$, we obtained $\hat{\beta}_{GQL,2} = 0.493$ with SSE 0.154 for $\alpha = 0.1$, but for $\alpha = 1.5$, $\hat{\beta}_{GQL,2} = 0.467$ with larger SSE 0.245. With regard to the estimation of α and ρ parameters, the moment estimators were found to underestimate the corresponding true values. However the biases associated with them appear not to be significant. For example, we obtained $\hat{\rho} = 0.101$ with SSE 0.074 for $\alpha = 0.1$, and $\hat{\rho} = 0.089$ with SSE 0.065 for $\alpha = 1.5$, where the true value of ρ is $\rho = 0.12$. This value of ρ is chosen to satisfy the range restriction which is determined by the design matrix. Note that for other designs it is quite possible to consider large values of ρ.

As far as the estimation of α is concerned, we obtained $\hat{\alpha} = 0.089$ with SSE 0.082 when the true value $\alpha = 0.1$. Also for true $\alpha = 1.5$, the estimate of α was found to be $\hat{\alpha} = 1.110$ with SSE 0.787.

Next, to examine the performance of the GQL(I) under the non-stationary model, we observe from Table 5.2 that this approach yields the estimates of the regression parameters with the same or smaller mean squared errors as compared to the GQL estimates (where ρ is present as $\rho = 0.12$). For example, when $\alpha = 0.1$ and $\beta_2 = 0.5$, we obtained $\hat{\beta}_{GQL,2}(I) = 0.494$ with SMSE $= 0.024$ which is same as the SMSE associated $\hat{\beta}_{GQL,2}$. However the performance of $\hat{\beta}_{GQL}(I)$ was found to be better than $\hat{\beta}_{GQL}$ as α gets larger. For example, for $\alpha = 1.5$, the SMSE associated with $\hat{\beta}_{GQL,2}$ was

Table 5.2: Simulated means(SM), simulated standard errors (SSE) and simulated mean squared errors (SMSE) of estimates of the regression parameter β, correlation parameter ρ and overdispersion parameter α for true $\rho = 0.12$ and selected values of α under the non-stationary negative binomial model with $\beta_1 = -0.005$, $\beta_2 = 0.5$, $\beta_3 = -0.5$, $\beta_4 = 0.5$, and $\beta_5 = -0.5$.

			Estimates						
Method	α	Statistic	$\hat{\beta}_{GQL,1}$	$\hat{\beta}_{GQL,2}$	$\hat{\beta}_{GQL,3}$	$\hat{\beta}_{GQL,4}$	$\hat{\beta}_{GQL,5}$	$\hat{\alpha}$	$\hat{\rho}$
GQL	0.1	SM	-0.004	0.493	-0.474	0.480	-0.505	0.089	0.101
		SSE	0.004	0.154	0.171	0.149	0.151	0.082	0.074
		SMSE	0.000	0.024	0.030	0.023	0.023	0.007	0.006
	0.5	SM	-0.004	0.497	-0.459	0.462	-0.521	0.373	0.098
		SSE	0.005	0.260	0.197	0.250	0.341	0.272	0.072
		SMSE	0.000	0.068	0.040	0.064	0.117	0.090	0.006
	0.9	SM	-0.004	0.491	-0.450	0.477	-0.523	0.651	0.094
		SSE	0.005	0.210	0.214	0.215	0.191	0.423	0.073
		SMSE	0.000	0.044	0.048	0.047	0.037	0.241	0.006
	1.5	SM	-0.004	0.467	-0.453	0.474	-0.508	1.110	0.089
		SSE	0.009	0.245	0.267	0.237	0.268	0.787	0.065
		SMSE	0.000	0.061	0.073	0.057	0.072	0.771	0.005
$GQL(I)$[1]	0.1	SM	-0.004	0.494	-0.475	0.478	-0.505	0.043	–
		SSE	0.004	0.154	0.171	0.143	0.151	0.117	–
		SMSE	0.000	0.024	0.030	0.021	0.023	0.017	–
	0.5	SM	-0.004	0.492	-0.461	0.471	-0.512	0.368	–
		SSE	0.005	0.179	0.194	0.175	0.181	0.265	–
		SMSE	0.000	0.032	0.039	0.031	0.033	0.088	–
	0.9	SM	-0.004	0.491	-0.451	0.473	-0.524	0.649	–
		SSE	0.005	0.212	0.214	0.214	0.190	0.414	–
		SMSE	0.000	0.045	0.048	0.047	0.037	0.234	–
	1.5	SM	-0.004	0.465	-0.459	0.470	-0.515	1.104	–
		SSE	0.006	0.239	0.251	0.237	0.223	0.783	–
		SMSE	0.000	0.058	0.065	0.057	0.050	0.770	–

[1] $GQL(I)$: GQL with 'working' independence correlation matrix I.

found to be 0.061, which is larger than the SMSE associated with $\hat{\beta}_{GQL,2}(I)$ (= 0.058). With regard to the estimation performance of α under the 'working' independence assumption, we found that both $\hat{\alpha}$ and $\hat{\alpha}(I)$ are equally efficient. This is because, the estimator α does not involve the correlation parameter ρ, i.e. it depends only on β for which the estimates are almost same under both approaches. These results, therefore, show that in general GQL(I) performs better in estimating β and α. This however does not mean that GQL can be avoided. This is because, as we discuss in the next section, the estimation for ρ is quite important with regard to the forecasting of a future count.

5.4 Statistical Inference: Forecasting

In order to examine the forecasting performance of the GQL based approach, in this section, we consider only the forecasting aspect for the stationary negative binomial AR(1) time series. For simplicity here also we consider one step ahead forecasting as in the Poisson case. The forecasting function can be obtained from the beta-binomial thinning based relationship

$$y_t = \eta_t * y_{t-1} + d_t = \sum_{j=1}^{y_{t-1}} b_j(\eta_t) + d_t,$$

where $\eta_t \sim Be\left(\frac{\rho}{\alpha}, \frac{1-\rho}{\alpha}\right)$, $y_{t-1} \sim NB(1/\alpha, \alpha m^*)$ and $d_t \sim NB(\frac{1-\rho}{\alpha}, \alpha m^*)$. Therefore the formula for the forecast function is computed as

$$\begin{aligned} E(Y_t \mid y_{t-1}) &= E_{\eta_t} E_{y_t}(Y_t \mid y_{t-1}, \eta_t) \\ &= E_{\eta_t}[\eta_t y_{t-1} + m^*(1-\rho)] \\ &= \rho y_{t-1} + m^*(1-\rho)] \\ &= m^* + \rho(y_{t-1} - m^*), \end{aligned} \quad (5.19)$$

which is exactly same as in the stationary Poisson AR(1) case, but the estimates of m^* and ρ will be different as y_t's are negative binomial counts. Consequently, the forecasting function under the GQL approach is estimated by

$$\hat{y}_{t,GQL} = \hat{E}(Y_t \mid y_{t-1}) = \hat{m}^* + \hat{\rho}(y_{t-1} - \hat{m}^*) \tag{5.20}$$

with $\hat{m}^* = e^{\hat{\beta}_{GQL,1}}$, whereas under the GQL(I) approach

$$\hat{y}_{t,GQL}(I) = \hat{\hat{m}}^*, \tag{5.21}$$

where $\hat{\hat{m}}^* = e^{\hat{\beta}_{GQL,1}(I)}$.

For the purpose of comparing the performances of the forecasting formulae (5.20) and (5.21), we conduct 1000 simulations and generate $T = 101$ observations in each simulation. We use the first 100 observations to obtain $\hat{\beta}_{GQL,1}$ and $\hat{\beta}_{GQL,1}(I)$ and use them in (5.20), (5.21) respectively to forecast 101-*th* observation. Similar to the Poisson case, we plot y_{101}, $\hat{y}_{101,GQL}$ and $\hat{y}_{101,GQL}(I)$ for first and last 50 simulations in each figure (given in Appendix B). We have also used the integer approximation and all of these has been done for $\rho = 0.3$, 0.9 and $\alpha = 0.5$, 0.9. It is clear from the figures that the forecasting based on GQL approach always performs better as compared to that of GQL(I). This is because, the forecasting values based on GQL appears to be closer to the true value than that of based on GQL(I). The performance of GQL appears to be much better when ρ is large.

Next, to get an overall idea about the forecasting performance of GQL and GQL(I) based forecasting functions, we have also computed the forecasted mean squared errors (FMSE) which is given in (3.22). The FMSEs under the GQL and GQL(I) approaches along with the simulated average of y_{101} are given in Table 5.3. It is clear from the table that the FMSEs under GQL approach are always smaller than the FMSEs associated with GQL(I). Hence it is evident that the use of correlation structure is very important in modelling the time series of counts, particularly as ρ increases-which is, of course, not surprising.

Table 5.3: Simulated mean forecast (SMF), simulated standard error (SSE) of the forecasted values and forecasted mean squared error (FMSE) for $\beta_1 = 0.5$ and selected values of true correlation parameter ρ under the stationary negative binomial AR(1) model using (i) 'working' independence (ii) correct correlation structure.

α	ρ	Simulated average y_{101}	Statistic	GQL(I) Approach	GQL Approach
0.3	0.3	2	SMF	2	2
			SSE	0.207	0.543
			FMSE	2.571	2.390
	0.5	2	SMF	2	2
			SSE	0.288	0.766
			FMSE	2.505	1.847
	0.7	2	SMF	2	2
			SSE	0.367	1.130
			FMSE	2.661	1.384
	0.9	2	SMF	2	2
			SSE	0.680	1.350
			FMSE	2.306	0.380
0.5	0.3	2	SMF	2	2
			SSE	0.239	0.587
			FMSE	2.712	2.520
	0.5	2	SMF	2	2
			SSE	0.291	0.928
			FMSE	3.030	2.216
	0.7	2	SMF	2	2
			SSE	0.436	1.256
			FMSE	3.182	1.756
	0.9	2	SMF	2	2
			SSE	0.717	1.471
			FMSE	2.760	0.517

(Table 5.3 contd....)

α	ρ	Simulated average y_{101}	Statistic	GQL(I) Approach	GQL Approach
0.7	0.3	2	SMF	2	2
			SSE	0.253	0.607
			FMSE	3.851	3.606
	0.5	2	SMF	2	2
			SSE	0.331	0.976
			FMSE	3.383	2.750
	0.7	2	SMF	2	2
			SSE	0.451	1.373
			FMSE	3.744	1.820
	0.9	2	SMF	2	2
			SSE	0.815	1.651
			FMSE	3.615	0.734
0.9	0.3	2	SMF	2	2
			SSE	0.270	0.685
			FMSE	4.051	3.699
	0.5	2	SMF	2	2
			SSE	0.381	0.956
			FMSE	4.463	3.505
	0.7	2	SMF	2	2
			SSE	0.476	1.260
			FMSE	3.695	1.947
	0.9	2	SMF	2	2
			SSE	0.830	1.736
			FMSE	4.188	1.123

5.5 A Numerical Illustration: Re-analysis of the Polio Count Data

In this section, we have fitted the observation-driven negative binomial model to the time series of the monthly number of cases of poliomyelitis reported by the U.S. Centers for Disease Control for the years 1970-1983. Here total number of observations is $T = 168$. Note that this data was first analysed by Zeger (1988) and then by Davis et al (2000) both by using their proposed parameter-driven models. For the purpose of comparison, we have used the same regression variables as in Zeger (1988). Consequently, we have regressed the monthly number of polio cases on a linear trend as well as sine, cosine pairs at annual and semi-annual frequencies to reveal the evidence of seasonality. More specifically, we use

$$x_t = [1, t^{'}/1000, cos(2\pi t^{'}/12), sin(2\pi t^{'}/12), cos(2\pi t^{'}/6), cos(2\pi t^{'}/6)]^{'},$$

where $t^{'} = (t - 73)$ is used to locate the intercept term at January 1976, for $t = 1, \cdots, 168$. The mean and variance of the data was found to be 1.33 and 3.48 respectively, which indicates the presence of overdispersion in the data. Therefore it is appropriate to use the negative binomial model to analyse such data. More specifically, we use the observation-driven correlated negative binomial model given by

$$y_t = \eta_t * y_{t-1} + d_t,$$

(see eqn.(5.2)), which is a beta-binomial thinning based relationship.

The above model was fitted by applying the GQL approach discussed in section 5.2. More specifically, the vector of regression effects (of x_t), $\beta = (\beta_1, \beta_2, \cdots, \beta_6)^{'}$, was estimated by (5.10), whereas α and ρ were estimated by (5.12) and (5.11) respectively. Note that in applying (5.10), (5.11) and (5.12) iteratively, we first start with initial values $\alpha = 0.1$ and $\rho = 0.1$ in estimating β by (5.10). It was found that in 4 iterations $\hat{\beta}_{GQL}$ produced converged estimates of β as shown under the 6-*th* column in Table 5.4. We then use this $\hat{\beta}_{GQL}$ in (5.12) and (5.11) and obtained $\hat{\alpha}$ and $\hat{\rho}$ which are

Table 5.4: Comparison of the estimates (Est.) based on observation-driven approach with Zeger's (1988) and Davis et al's (2000) estimates for the polio data.

	Parameter-driven models				Observation-driven Model			
	Zeger's method		Davis et al's method		1-cycle based Estimates		Converged Estimates	
Parameters	Est.	SE	Est.	SE	Est.	SE	Est.	SE
Intercept (β_1)	0.17	0.13	0.207	0.075	0.207	0.075	0.212	0.129
Trend$\times 10^{-3}$ (β_2)	-4.35	2.68	-4.799	1.399	-4.797	1.403	-3.876	2.539
$\cos(2\pi t/12)$ (β_3)	-0.11	0.16	-0.149	0.097	-0.150	0.097	-0.133	0.172
$\sin(2\pi t/12)$ (β_4)	-0.48	0.17	-0.532	0.109	-0.532	0.109	-0.490	0.169
$\cos(2\pi t/6)$ (β_5)	0.20	0.14	0.169	0.098	0.170	0.099	0.165	0.149
$\sin(2\pi t/6)$ (β_6)	-0.41	0.14	-0.432	0.101	-0.431	0.101	-0.404	0.150
α	0.77	–	–	–	0.755	–	0.807	–
$\rho_y(1)$	0.25	–	–	–	0.239	–	0.239	–
$\rho_y(2)$	–	–	–	–	0.204	–	0.206	–

also reported in the 6-*th* column. Note that these estimates are referred to as 1-cycle based estimates. Further, we have continued the cycle of iterations until convergence for estimates of all parameters. It was found that convergence was achieved in 3 cycles of iteration. These converged results are shown in the 8-*th* column in Table 5.4. The estimates of β, α and ρ (wherever applicable) from Zeger (1988) and Davis et al (2000) are reproduced in columns 2 and 4 respectively.

From Table 5.4 it is clear that the one cycle based estimates based on our proposed observation-driven model are almost identical to that of Davis et al (2000) along with the associated estimated standard errors. As far as the overdispersion parameter is concerned, the present moment approach produced similar estimate as in Zeger (1988). Also, the estimates of the lag-1 correlation of observations are almost the same. Note that although our 1-cycle based regression estimates are close to those of Davis et al's (2000), our final converged estimates along with their standard errors appear to be different. The converged estimates and their standard errors are reported in columns 8 and 9.

Note that the proposed observation-driven model is easily fitted to the data and we can have estimates of lag correlations directly and easily from the responses, whereas in Zeger's (1988) approach, these correlations are computed based on a complicated function. By the same token it is extremely difficult to obtain the lag correlations of observations in Davis et al's (2000) method. These correlations are not available.

Chapter 6

Concluding Remarks

6.1 General Remarks

In analysing any time series, discrete or continuous, it is necessary to (1) model the correlation structure of the series, (2) estimate the parameters of the model, and (3) forecast the future observations. The modelling of time series of (discrete) count data, unlike the modelling of continuous time series, is not adequately addressed in the literature. The modelling becomes much more difficult when count data are non-stationary and subject to overdispersion. In this thesis, we have introduced an observation-driven non-stationary negative binomial model to interpret the correlations of the data. This we have done following the existing models (see McKenzie (1988), Al-Osh and Aly (1992), Sutradhar, Jowaheer and Rao (2003)) for stationary Poisson and negative binomial data. It is much easier to interpret the correlation of the data based on such models as compared to certain parameter-driven models (Zeger (1988), Davis et al (2000), Harvey and Fernandes (1989)). Further we have shown that it is much simpler to forecast a future count based on proposed models as compared to the parameter-driven models. In fact it is quite difficult to develop a forecasting function based on parameter-driven processes.

As far as the estimation of parameters of the proposed model is concerned, we have used the generalized quasilikelihood (GQL), independence assumption based

GQL approach GQL(I) and a special working correlation matrix based GQL approach GQL(C^*). The simulation studies show that in most of the cases GQL(I) performs better than the other approaches. However this finding does not mean that one can ignore the correlation parameter. This is because the correlation parameter plays an important role in forecasting future counts. With regard to forecasting based on the stationary model, GQL based forecasting was found to be uniformly better than the GQL(I) based forecasting.

While modelling negative binomial counts through a parameter-driven approach, we have encountered the problems of having estimates in inadmissible regions, especially for overdispersion and correlation parameters. In contrast, we have found that the observation-driven process based correlations can be easily computed and parameters can be easily estimated.

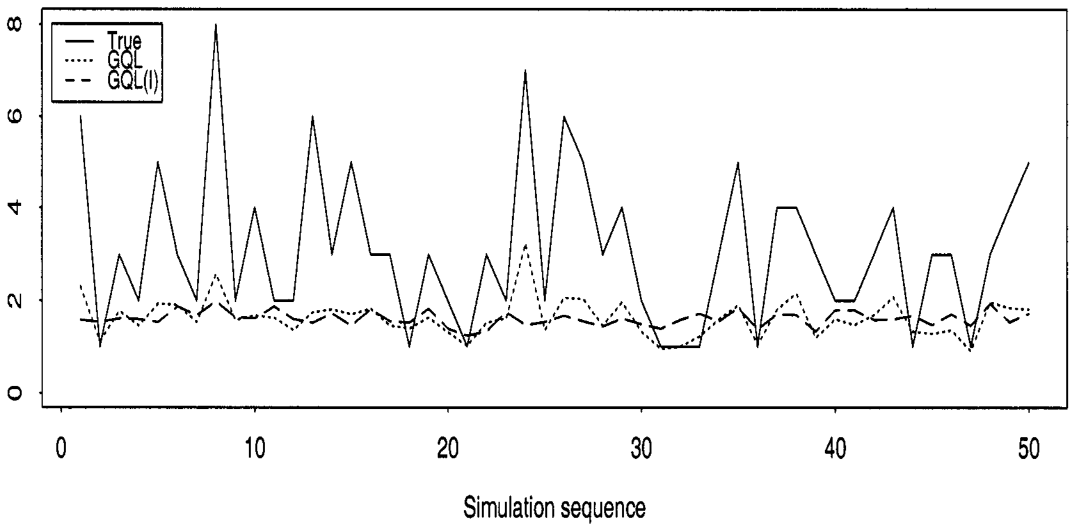
We have also applied the proposed extended non-stationary model to fit the polio count data considered by Zeger (1988) and Davis et al (2000). Model fitting performance was found to be almost the same as compared to that of Zeger and Davis et al. This shows the usefulness of the proposed simpler observation-driven model in fitting negative binomial time series of counts.

6.2 Proposal for Further Research

Note that the overdispersion as well as autocorrelation parameters were computed by the method of moments. It was found that the moment estimate of ρ was almost unbiased whereas the moment approach produced biased estimate for the overdispersion parameter. A better estimation method, specially in estimating the overdispersion parameter may be of future interest.

Appendix A

Figures: Poisson Based Forcasting



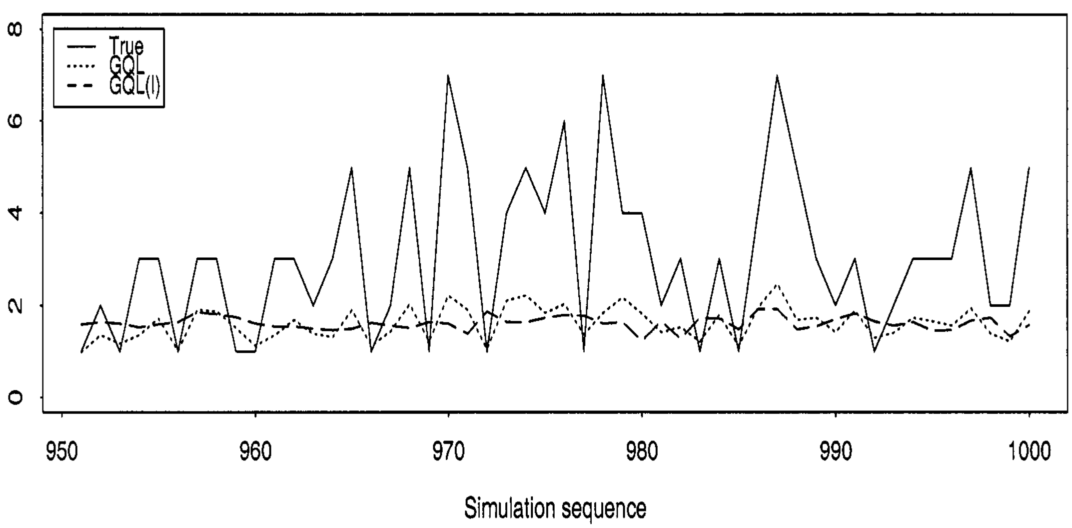
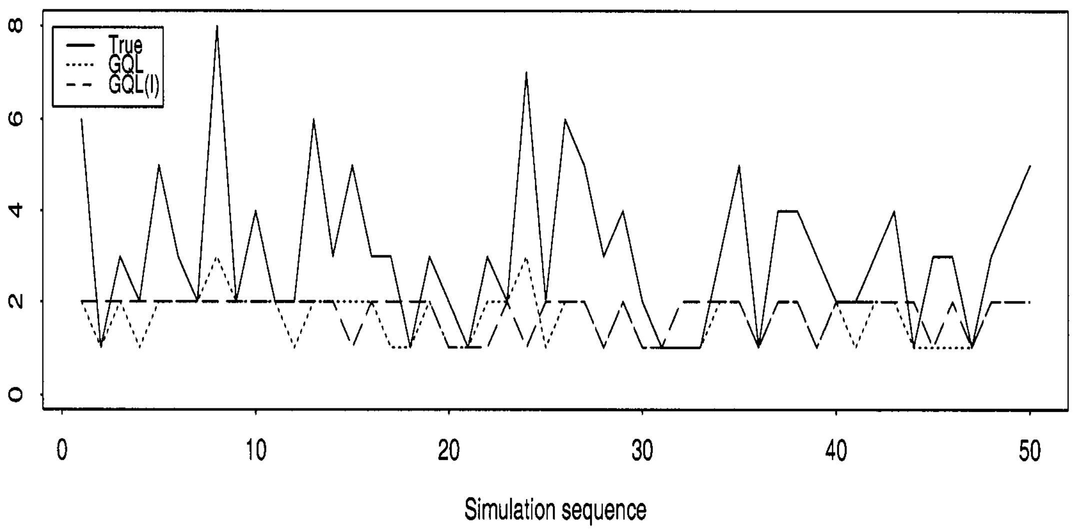


Figure A.1: Poisson model based 1-step ahead forcasted value for $\rho = 0.3$



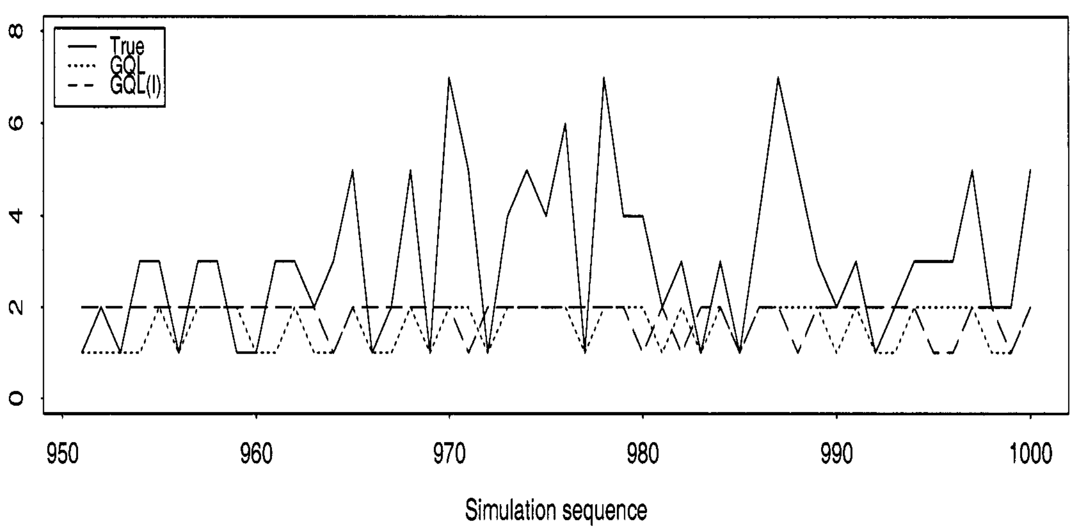
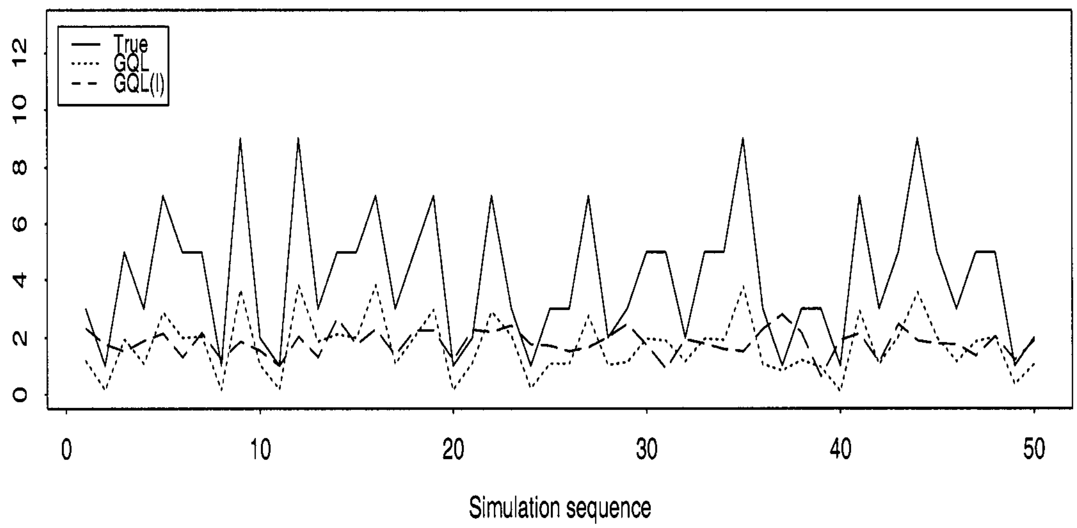


Figure A.2: Poisson model based 1-step ahead forcasted value (rounded to integer) for $\rho = 0.3$



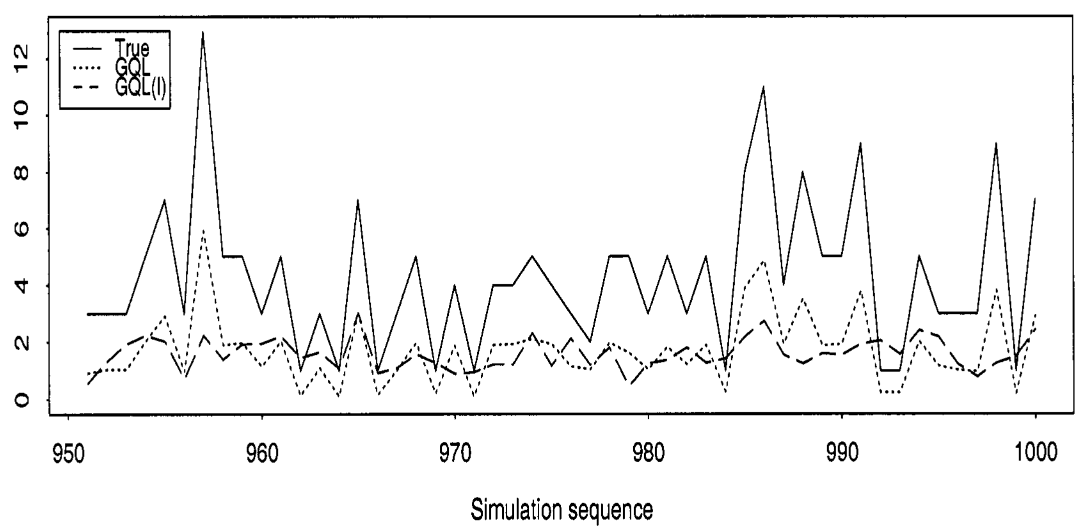
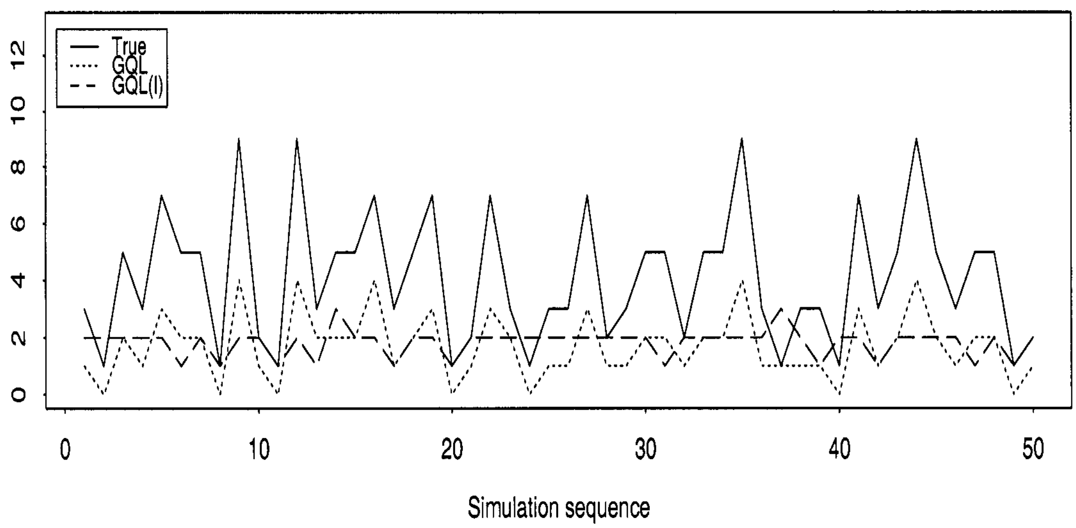


Figure A.3: Poisson model based 1-step ahead forcasted value for $\rho = 0.9$



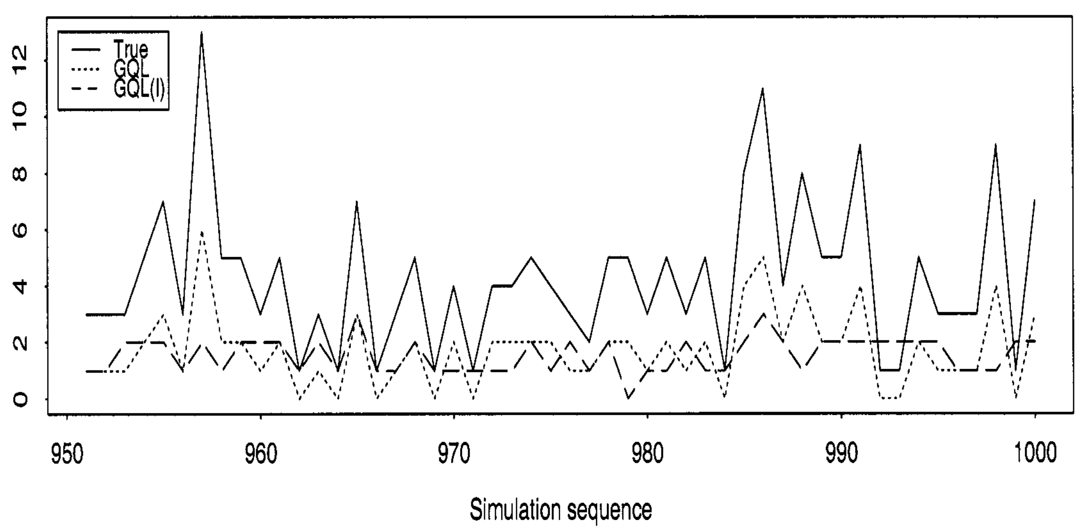
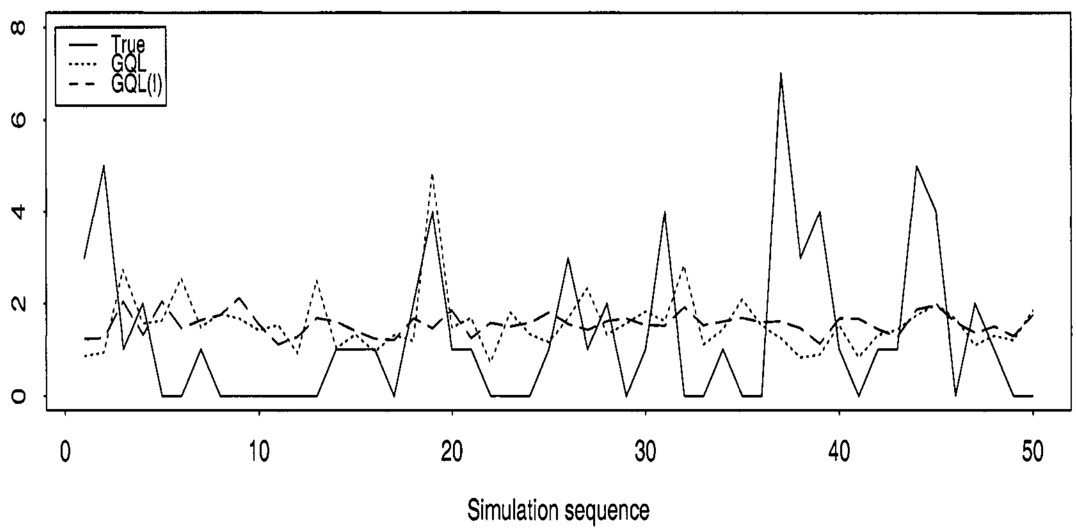


Figure A.4: Poisson model based 1-step ahead forcasted value (rounded to integer) for $\rho = 0.9$

Appendix B

Figures: Negative Binomial Based Forecasting



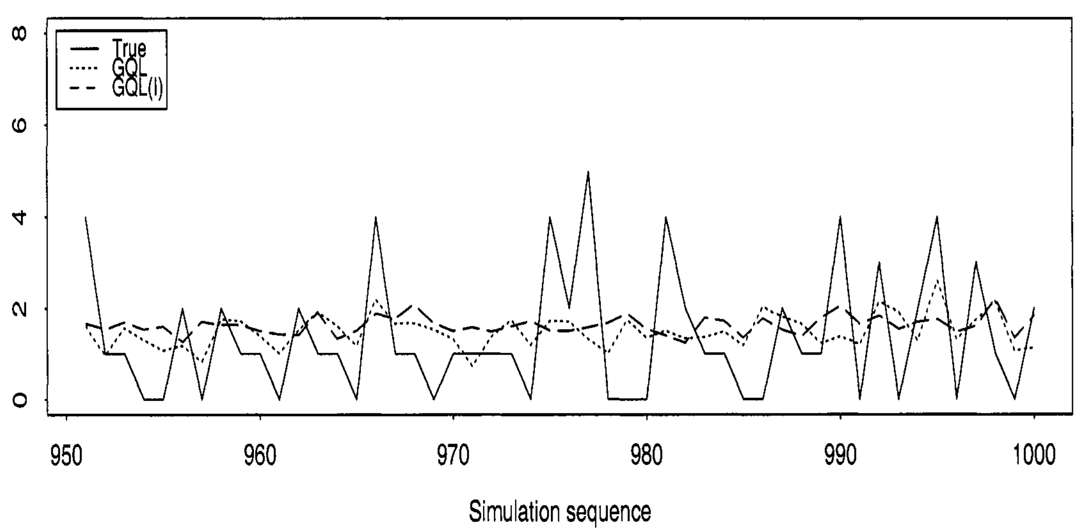
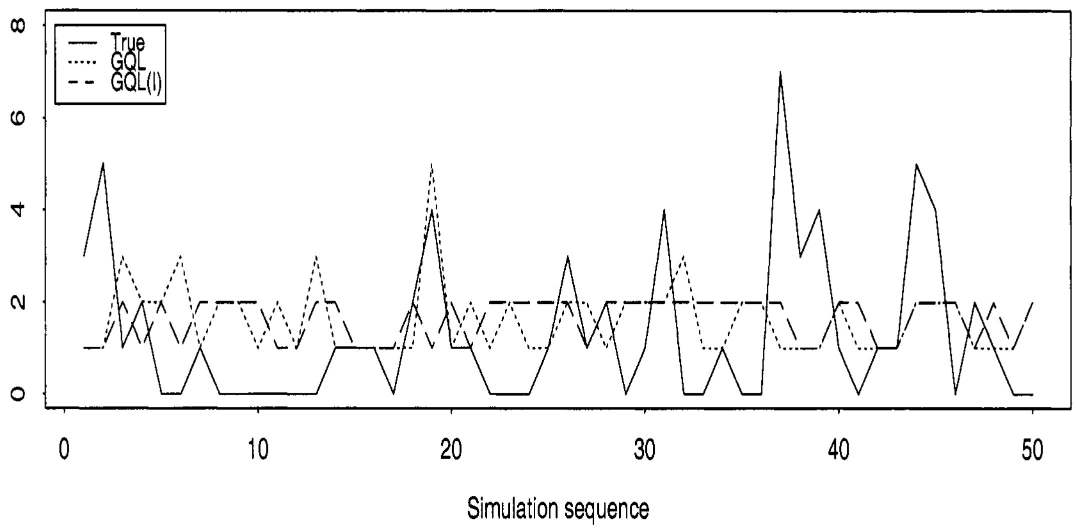


Figure B.1: Negative binomial model based 1-step ahead forcasted value for $\alpha = 0.5$, $\rho = 0.3$



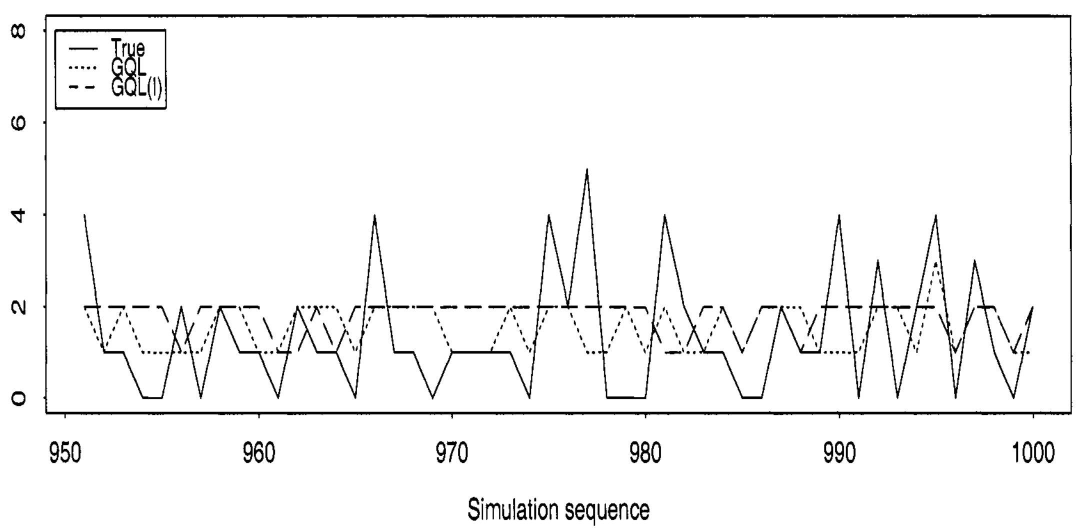


Figure B.2: Negative binomial model based 1-step ahead forcasted value (rounded to integer) for $\alpha = 0.5$, $\rho = 0.3$

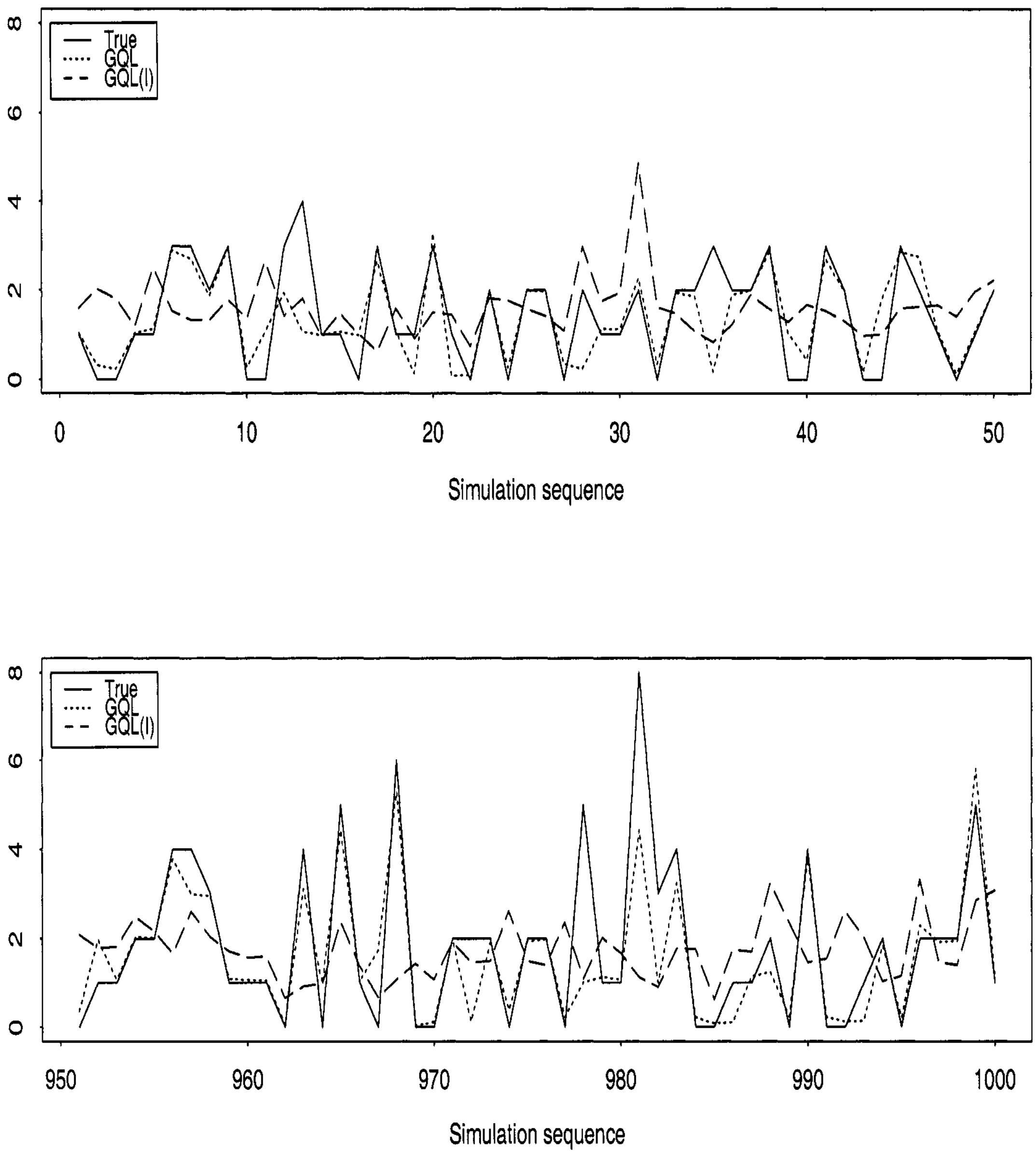


Figure B.3: Negative binomial model based 1-step ahead forcasted value for $\alpha = 0.5$, $\rho = 0.9$

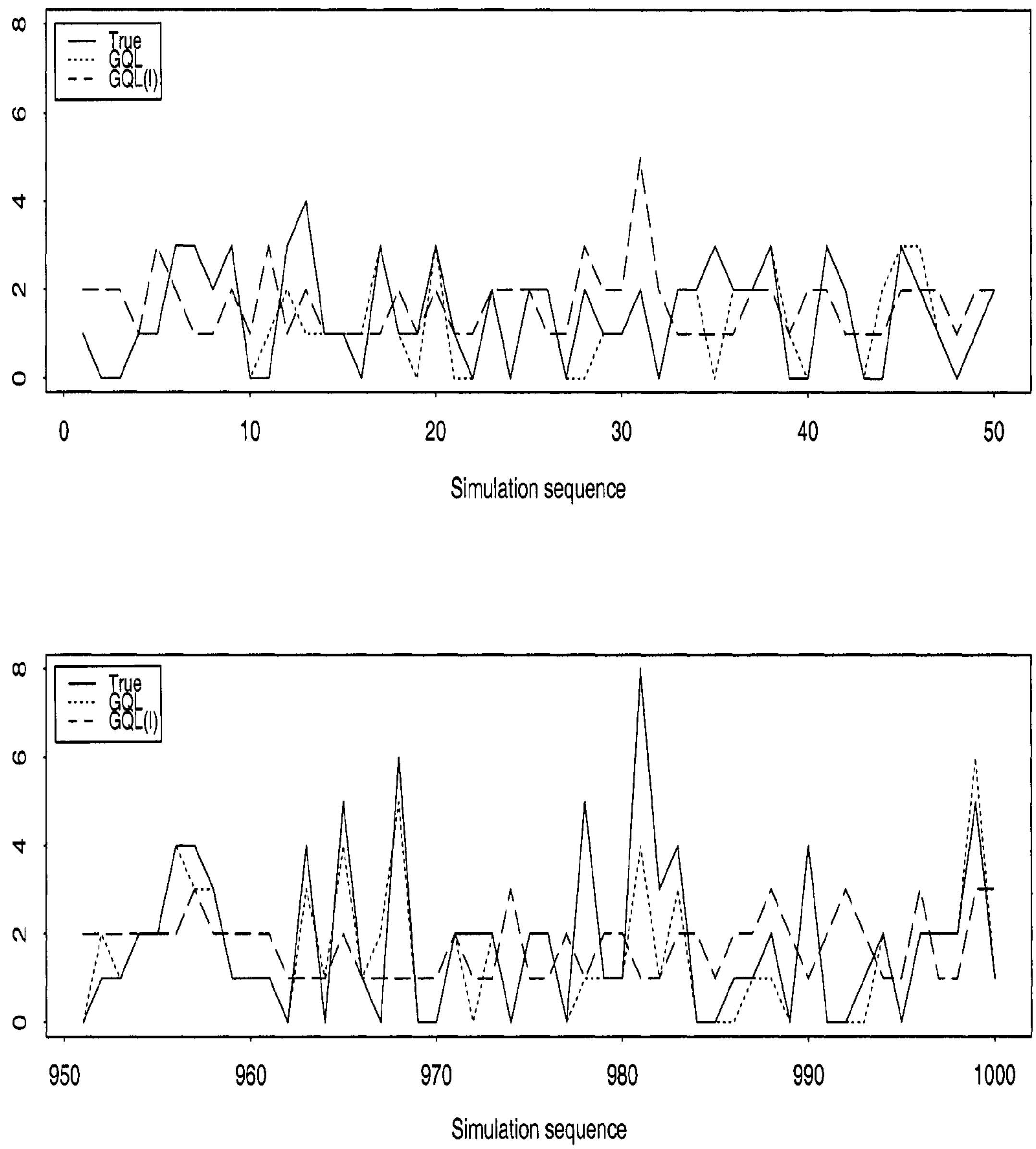
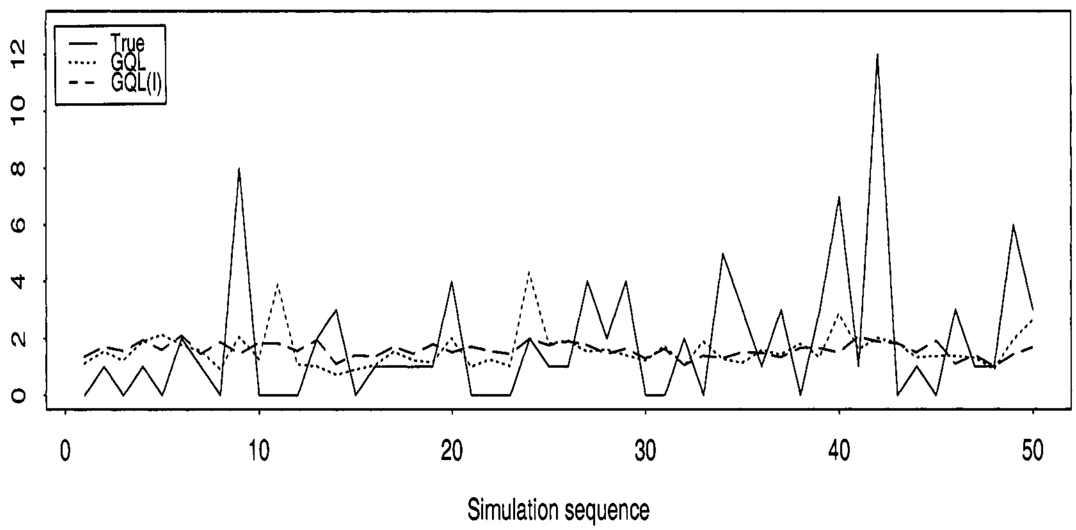


Figure B.4: Negative binomial model based 1-step ahead forcasted value (rounded to integer) for $\alpha = 0.5$, $\rho = 0.9$



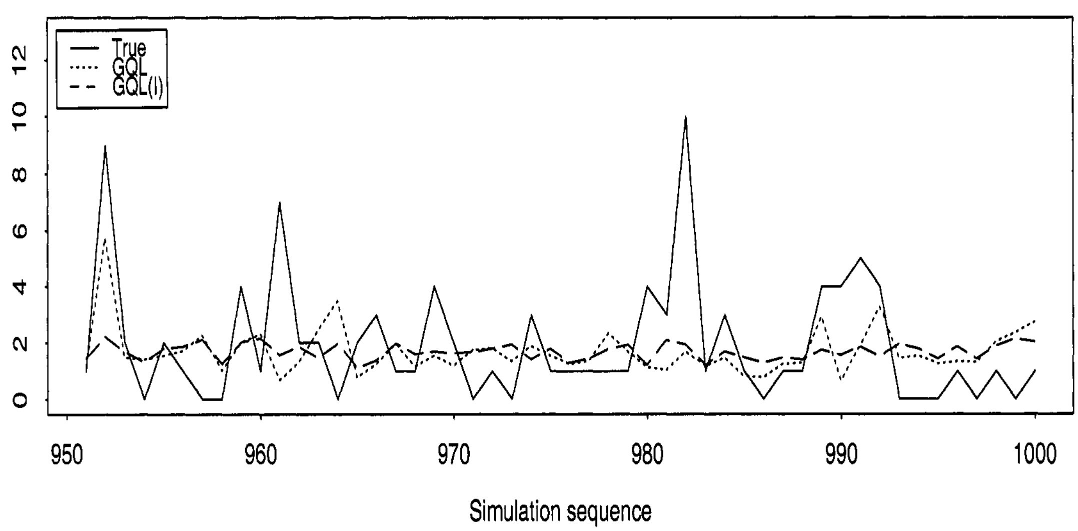


Figure B.5: Negative binomial model based 1-step ahead forcasted value for $\alpha = 0.9$, $\rho = 0.3$

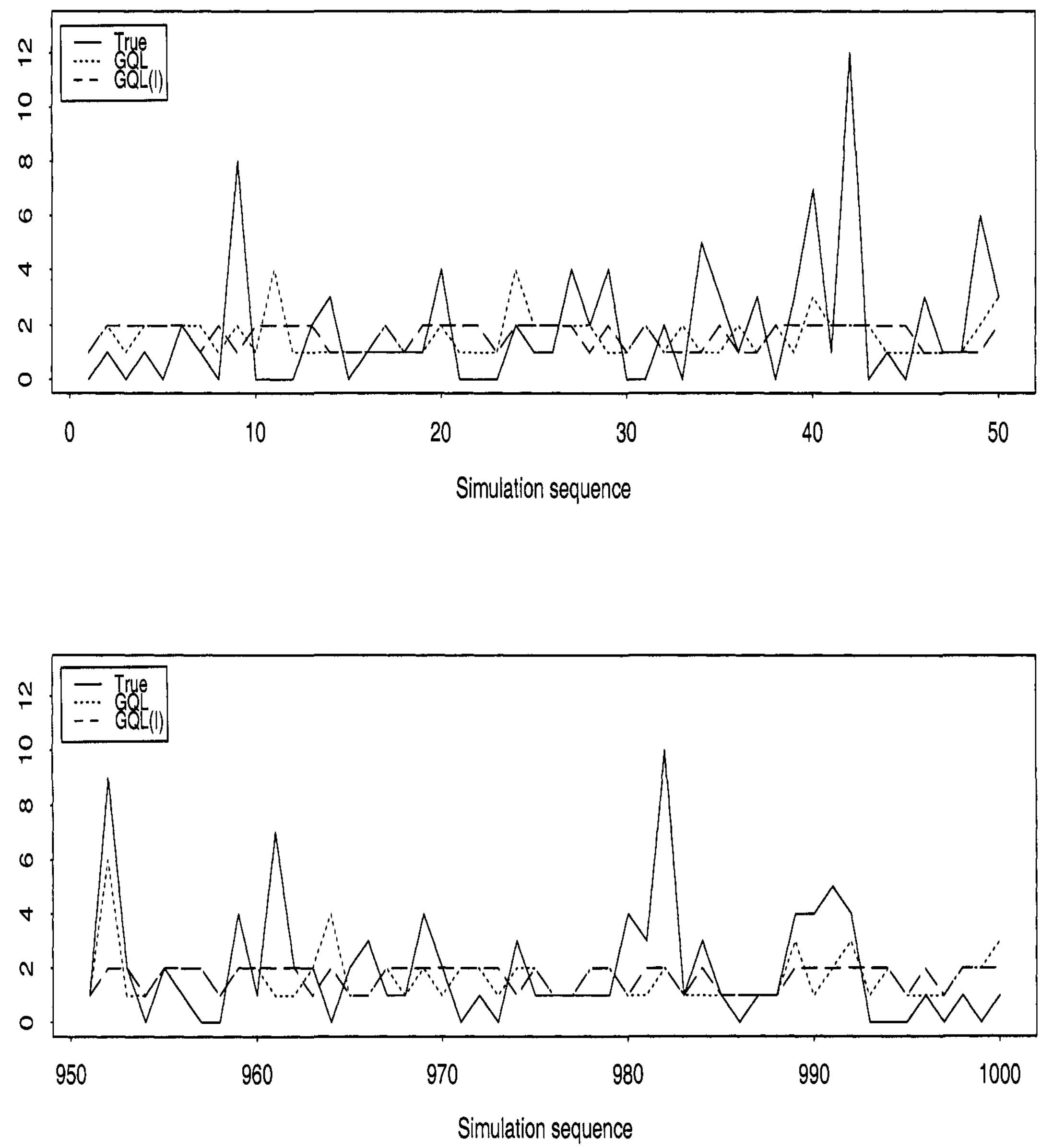
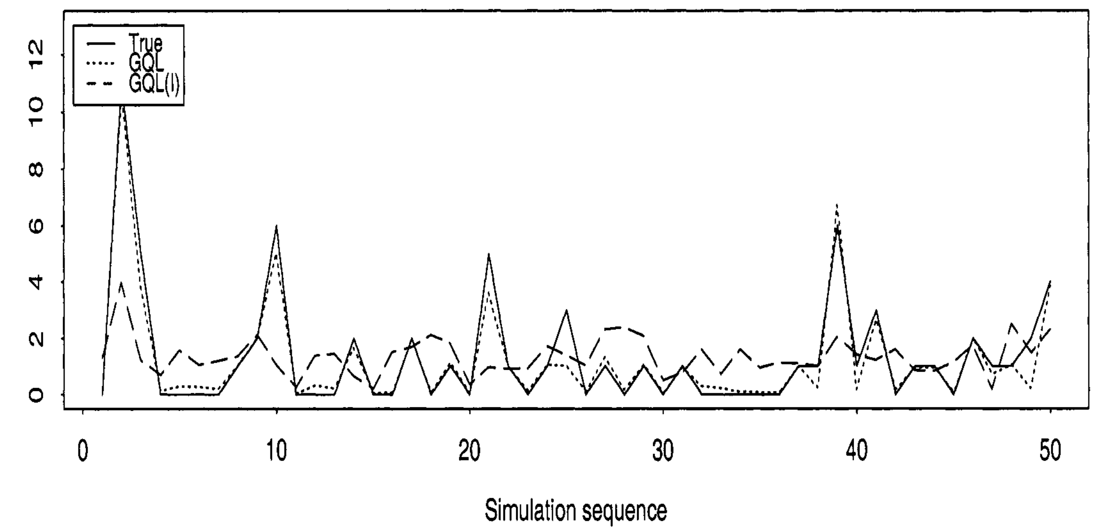


Figure B.6: Negative binomial model based 1-step ahead forcasted value (rounded to integer) for $\alpha = 0.9$, $\rho = 0.3$



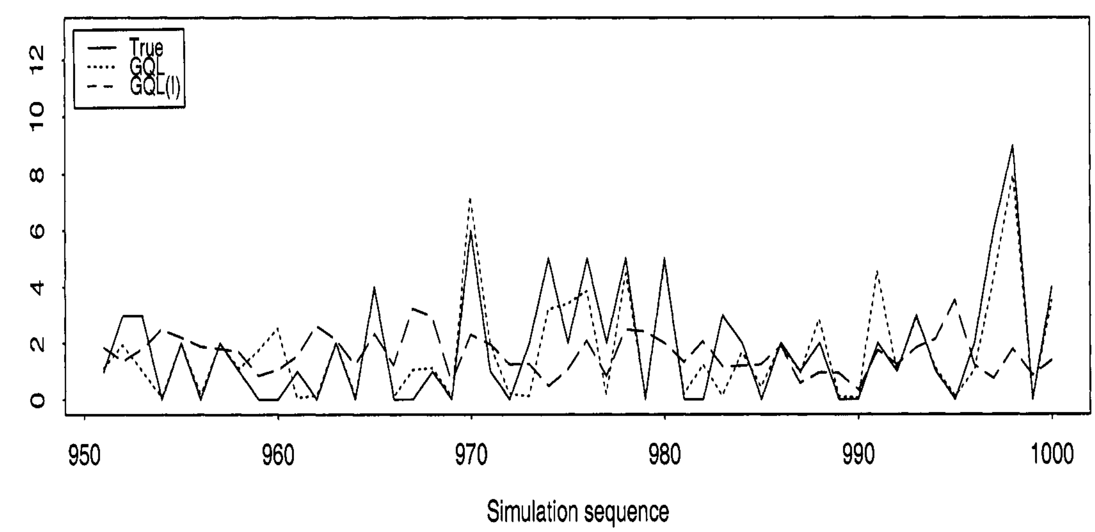
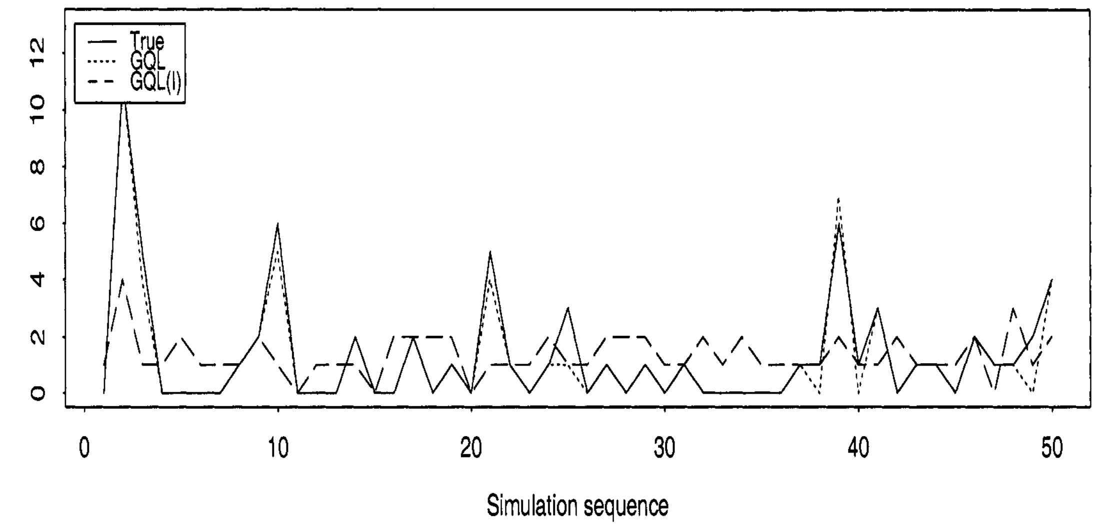


Figure B.7: Negative binomial model based 1-step ahead forcasted value for $\alpha = 0.9$, $\rho = 0.9$



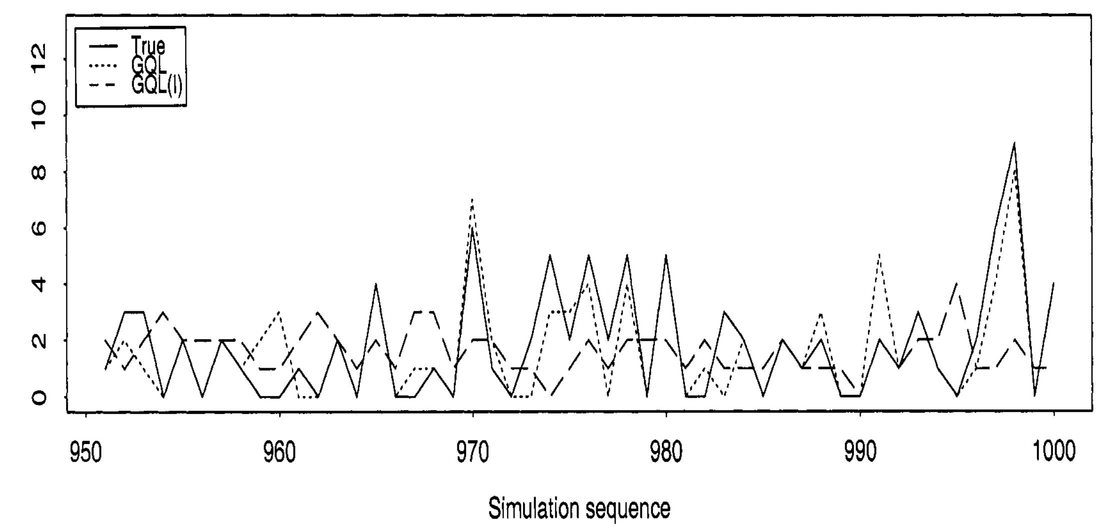


Figure B.8: Negative binomial model based 1-step ahead forcasted value (rounded to integer) for $\alpha = 0.9$, $\rho = 0.9$

Bibliography

[1] Al-Osh, M.A. and Aly, A.A. (1992). First Order Autoregressive Time Series with Negative Binomial and Geometric Marginals. *Commun. Statist.-Theory Meth.*, **21**, 2483-2492.

[2] Cox, D.R. (1981). Statistical Analysis of Time Series, Some Recent Developments. *Scand. J. Statist.*, **8**, 93-115.

[3] Davis, R.A., Dunsmuir, W.T.M. and Streett, S.B. (2003). Observation-driven Models for Poisson Counts. *Biometrika*, **90**, 777-790.

[4] Davis, R.A., Dunsmuir, W.T.M. and Wang, Y. (2000). On Autocorrelation in a Poisson Regression Model. *Biometrika*, **87**, 491-505.

[5] Harvey, A.C. and Fernandes, C. (1989). Time Series Models for Count or Qualitative Observations. *Journal of Business and Economic Statistics*, **7**, 407-417.

[6] Honore, B.E. and Kyriazidou, E. (2000). Panel data-discrete choice models with lagged dependent variables. *Econometrica*, **89**, 389-399.

[7] Jacobs, P.A. and Lewis, P.A.W. (1978a). Discrete Time Series Generated by Mixtures I: Correlational and Runs Properties. *J. R. Statist. Soc. B*, **40**, 94-105.

[8] Jacobs, P.A. and Lewis, P.A.W. (1978b). Discrete Time Series Generated by Mixtures II: Asymptotic Properties. *J. R. Statist. Soc. B*, **40**, 222-228.

[9] Jacobs, P.A. and Lewis, P.A.W. (1983). Stationary Discrete Autoregressive-moving Average Time Series Generated by Mixtures. *J. Time Series Anal.*, **4**, 19-36.

[10] Johnson, N.L. and Kotz, S. (1992). *Univariate Discrete Distributions.* John Wiley & Sons. Inc.

[11] Jowaheer, V. and Sutradhar, B.C. (2002). Analysis of Longitudinal Count Data with Overdispersion. *Biometrika*, **89**, 389-399.

[12] Kendal, M.G. and Stuart, A. (1968). *The Advanced Theory of Statistics, Vol. 3.* Charles Griffin & Company Limited.

[13] Kulendran, N. and King, M.L. (1997). Forecasting International Quarterly Tourist flows using Error Correction and Time Series Models. *International Journal of Forecasting*, **13**, 319-327.

[14] McKenzie, E. (1988). Some ARMA Models for Dependent Sequences of Poisson Counts. *Adv. in Appl. Probab.*, **20**, 822-835.

[15] McKenzie, E. (1986). Autoregressive Moving Average Processes with Negative Binomial and Geometric Marginal Distributions. *Adv. Appl. Probab.*, **18**, 679-705.

[16] Settimi, R. and Smith, J.Q. (2000). A Comparison of Approximate Bayesian Forecasting Methods for Non-Gaussian Time Series. *Journal of Forecasting*, **19**, 135-148.

[17] Sutradhar, B.C. (2003). An Overview on Regression Models for Discrete Longitudinal Responses. *Statistical Science*, **18**, 377-393.

[18] Sutradhar, B.C., Jowaheer, V. and Rao, R.P. (2003). On Longitudinal Mixed Models for Non-stationary Count data. *Submitted for Publication.*

[19] Zeger, S.L. (1988). A Regression Model for Time Series of Counts. *Biometrika*, **75**, 621-629.