Inference on Autoregressive Moving Average Models for Count Data

by

© **Syeda Fateha Akter**

A thesis submitted to the School of Graduate Studies in partial fulfillment of the requirements for the degree of Master of Science.

Department of Mathematics and Statistics

Memorial University

January 2025

St. John's, Newfoundland and Labrador, Canada

Abstract

In the analysis of count time series at equally spaced intervals with covariate information, Poisson Autoregressive (AR) or Integer-Valued Autoregressive (INAR) models have been widely discussed in the literature, with their fundamental properties and estimation methods thoroughly explored. However, when time series data exhibits both long-term dependencies (autocorrelation) and moving average effects, capturing both of these elements is essential for more effective modeling and forecasting. To address this, we introduce autoregressive moving average (ARMA) models of order (1,1) for count time series. We first consider the case where the offspring random variable follows a Bernoulli distribution, meaning that each individual in the population at time $t-1$ can produce only one or zero offspring at time t. Additionally, we extend this model to incorporate the possibility of any individual producing multiple offspring at a given time point, resulting in a binomial offspring random variable. We derive the key properties of these models, present methods for parameter estimation and forecasting function. The performance of the proposed methods are assessed through simulation studies.
Keywords: moving average model, autoregressive moving average model, generalized quasi likelihood method, generalized method of moments.

Acknowledgements

First, I want to express my honor and gratitude to Allah for enabling me to accomplish this thesis research.

I am truly thankful to my supervisors, Professor Alwell Oyet and Professor Asokan Variyath, for their support, guidance, and financial assistance throughout this journey. Their encouragement, helpful suggestions, and careful guidance were key in shaping this research. Their knowledge and constructive feedback have played a huge role in my academic growth and in completing this thesis. I appreciate the time and effort they dedicated to helping me reach my goals.

I also want to thank Dr. Zhaozhi Fan and Dr. Alex Shestopaloff for introducing me to various statistical concepts in their courses. I am grateful to Dr. Jahrul Alam for guiding me in writing a manuscript-based thesis.

I sincerely acknowledge the financial support I received through graduate and teaching assistantships from the School of Graduate Studies and the Department of Mathematics and Statistics. I also want to thank the administrative staff of the Department of Mathematics and Statistics for their continuous help and support in the computer lab.

I deeply appreciate my friends' unwavering support and encouragement during this journey. Their kind words and willingness to stand by me through tough times and joyful moments have meant so much to me.

Finally, I am incredibly grateful to my parents and family for their unconditional love, support, and encouragement.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Consider a sequence of n random variables $Y_{t_1}, Y_{t_2}, ..., Y_{t_n}$ indexed by time, where $t_1 < t_2 < t_3 < ... < t_n$. Let $y_{t_1}, y_{t_2}, ..., y_{t_n}$ be observed values of the random variables $\{Y_t, t \in \mathbb{Z}\}$. The random variable Y_t is said to be nth order weakly stationary if the joint moments of Y_t up to order n does not depend on time. When observations $y_{t1}, y_{t2}, ..., y_{t_n}$ are measured at specific time points $t_1, t_2, ..., t_n$, respectively ($t_1 < t_2 < ... < t_n$), then this set of observations is called time series data (Box et al., 2015). The time intervals between observations can either be equal or unequal. Since the focus is on time series data with equal intervals, some commonly used time series models assuming equal time spacing are discussed. Furthermore, depending on the nature of the observed values, time series data are classified into different types, such as continuous time series, count time series, and multinomial time series (Chatfield, 1975).

1.1 Models for Continuous Time Series Data

Continuous time series data refers to sequential observations collected over a period of time, where each data point can take any value within a specified range. The concept of correlation in time series data resulting from lagged linear relationships led to the development of autoregressive (AR) and autoregressive moving average (ARMA) models, as detailed by Whittle (1951).

1.1.1 Autoregressive Model

In AR(p) model the current value of the time series, y_t is considered as a linear combination of p most recent past values of itself and a random term w_t that incorporates everything new at the series at time point t that is not explained by the past values (Box et al., 2015; Chuang, 1991; Shumway et al., 2000; Teräsvirta, 1994). Let y_t be a stationary process with mean μ and $\tilde{y}_t = y_t - \mu$ be the mean deleted process. Additionally, w_t is a white noise process with mean 0 and variance σ^2. A *pth* order AR process, $\tilde{y}_t$, has the form,

$$\tilde{y}_t = \phi_1 \tilde{y}_{t-1} + \phi_2 \tilde{y}_{t-2} + ... + \phi_p \tilde{y}_{t-p} + w_t. \tag{1.1}$$

Notice that in equation (1.1), $\tilde{y}_t$ is modeled as a linear combination of past history of $\tilde{y}_t$ and a random component, w_t. Thus, $\tilde{y}_t$ is a random variable. Here, ϕ_i is the parameter of the model for $i = 1, 2, \ldots, p$, representing the influence of the i-th lagged value on the current value y_t. Some basic properties of the model are:

- $E(y_t) = \mu$.
- $v(y_t) = \gamma(0) = \sum_{j=1}^{p} \phi_j \gamma(j) + \sigma^2$, where $\gamma(j)$ is the autocovariance function.
- $cov(y_t, y_{t-k}) = \gamma(k) = \sum_{j=1}^{p} \phi_j \gamma(|k-j|)$, where the lag $k \geq 1$.
- $corr(y_t, y_{t-k}) = \rho(k) = \sum_{j=1}^{p} \phi_j \rho(|k-j|)$, where the lag $k \geq 1$. It can be observed that the autocorrelation function (ACF), denoted by $\rho(k)$, satisfies a difference equation. This difference equation for the lag-k ACF is commonly referred to as the Yule-Walker difference equation (Box et al., 2015).

1.1.2 Moving Average Model

In MA(q) model the current value of the time series, y_t is considered as a linear combination of lagged white noise processes (Box et al., 2015; Chuang, 1991; Shumway et al., 2000). A mean deleted *qth* order MA process, $\tilde{y}_t$, has the form,

$$\tilde{y}_t = w_t - \theta_1 w_{t-1} - \theta_2 w_{t-2} - ... - \theta_q w_{t-q}.$$

Here, θ_i is the parameter of the model for $i = 1, 2, \ldots, p$, representing the influence of the i-th lagged white noise process on the current value of $\tilde{y}_t$. Some basic properties of the model are:

- $E(y_t) = \mu$.
- $v(y_t) = \gamma(0) = \sigma^2 \sum_{j=0}^{q} {\theta_j}^2$, where $\theta_0 = -1$.
- $cov(y_t, y_{t-k}) = \gamma(k) = \sigma^2 \sum_{i=0}^{q} (\theta_i \theta_{i+k} - \theta_k)$, where $\theta_0 = -1$.
- $corr(y_t, y_{t-k}) = \rho(k) = \dfrac{\sum_{i=0}^{q} \theta_i \theta_{i+k}}{\sum_{j=0}^{q} {\theta_j}^2}$, where $\theta_0 = -1$.

1.1.3 Autoregressive Moving Average Model

In ARMA model, the AR and the MA components are combined to consider both the effects of past values and past random terms of a time series (Box et al., 2015; Chuang, 1991; Shumway et al., 2000). A mean deleted ARMA(p,q) model has the form,

$$\tilde{y}_t = \phi_1 \tilde{y}_{t-1} + \phi_2 \tilde{y}_{t-2} + ... + \phi_p \tilde{y}_{t-p} + w_t - \theta_1 w_{t-1} - \theta_2 w_{t-2} - ... - \theta_q w_{t-q}.$$

Here, ϕ_i and θ_i are the parameters of the model for $i = 1, 2, \ldots, p$, representing the influence of the i-th lagged value on the current value y_t and i-th lagged white noise process on the current value y_t respectively. When $q = 0$, an ARMA(p,0) model reduces to the AR(p) process. Similarly, when $p = 0$, an ARMA(0,q) model reduces to the MA(q) process. So, MA and AR models are special cases of ARMA(p,q) model. The simplest example of the ARMA(p,q) model is the ARMA(1,1) model defined as,

$$\tilde{y}_t = \phi_1 \tilde{y}_{t-1} + w_t - \theta_1 w_{t-1}.$$

Some basic properties of the ARMA(1,1) model are:

- $E(y_t) = \mu$.
- $v(y_t) = \gamma(0) = \dfrac{{\theta_1}^2 - 2\theta_1\phi_1 + 1}{1 - {\phi_1}^2}\sigma^2$.

- $cov(y_t, y_{t-k}) = \gamma(k) = \begin{cases} \phi_1\gamma(0) - \theta_1\sigma^2, & \text{for } k = 1, \\ {\phi_1}^k\gamma(0), & \text{for } k > 1. \end{cases}$

- $\text{corr}(y_t, y_{t-k}) = \rho(k) = \begin{cases} \phi_1 - \dfrac{\theta_1\sigma^2}{\gamma(0)}, & \text{for } k = 1, \\ {\phi_1}^k, & \text{for } k > 1. \end{cases}$

From the ACF of the ARMA(1,1) model, it can be observed that the ACF behaves like that of the AR(1) model after lag 1 (Brockwell & Davis, 2002; Hyndman, 2018). In general, it can be shown that the ACF of ARMA(p,q) model behaves like that of the AR(p) model after lag q (Brockwell & Davis, 2002; Hyndman, 2018).

1.2 Models for Count Time Series Data

Count data consists of non-negative integer values that represent the number of occurrences of an event (Czado et al., 2009; Cameron & Trivedi, 2013). Early researchers who have studied count time series have often used Poisson Autoregressive models of order 1 (Al-Osh & Alzaid, 1987; Sutradhar, 2003, 2011; Oyet & Sutradhar, 2013). Al-Osh and Alzaid (1987) introduced a model for a stationary sequence of integer-valued random variables with lag-one dependence, naming it the integer-valued autoregressive process of order one (INAR(1) process). In their proposed INAR(1) process binomial thinning operation, denoted by *, was used to replace scalar multiplication of discrete random variables (Steutel & Van Harn, 1979; Steutel et al., 1983; Puig & Valero, 2007). Specifically, if Y is a discrete random variable taking non-negative integer values, then the binomial thinning operation is defined as,

$$\rho * Y = \sum_{j=1}^{Y} b_j(\rho), \tag{1.2}$$

where $b_j(\rho)$ is an identically and independently distributed binary random variable with $P[b_j(\rho) = 1] = \rho = 1 - P[b_j(\rho) = 0]$ and $\rho \in [0, 1]$. In each model, thinning operations are assumed to be independent of previous history of the process. Al-Osh and Alzaid (1987) demonstrated that the distribution properties of the INAR(1) process are similar to those of the AR(1) model for continuous data. They also discussed the estimation of parameters using the Yule-Walker estimators, the conditional least squares estimators and the maximum likelihood estimators.

1.2.1 Autoregressive Model

Zhang and Oyet (2014) extended the Poisson AR(1) model to the AR(2) model. They studied the properties of the model and proposed a GQL approach for estimating the model parameters.

Let $\boldsymbol{x}_t = (x_{t_1}, x_{t_2}, ..., x_{t_p})^T$ be a p-dimensional covariate vector measured at time t, along with the count y_t for a single community and $\boldsymbol{\beta}$ be the $(p \times 1)$ vector of covariate effect parameter. Let d_t represent the immigration variable, indicating the number of individuals entering the community from different regions at time t. Based on the study conducted by Zhang and Oyet (2014), for a single community a AR(p) model for count data can be written as,

$$y_r = \sum_{l=1}^{r-1} \rho_l * y_{r-l} + d_r,\ r = 2, 3, ..., p \tag{1.3}$$

$$y_t = \sum_{l=1}^{p} \rho_l * y_{t-l} + d_t,\ t = p+1, p+2, ..., T, \tag{1.4}$$

where $y_1 \sim Poi(\mu_1 = exp(\boldsymbol{x}_1{}^T\boldsymbol{\beta}))$ and, d_t and y_{t-l} are independent. In particular, when $p = 1$, the AR(p) model reduces to the AR(1) model and by using binomial thinning operation the model is given by

$$y_t = \sum_{j=1}^{y_{t-1}} b_j(n_t, \rho) + d_t, \tag{1.5}$$

where $b_j \sim Binomial(n_t, \rho)$ with assumptions $y_1 \sim Poi(\mu_1 = exp(\boldsymbol{x}_1{}^T\boldsymbol{\beta}))$, $d_t \sim Poi(\mu_t - n_t\rho\mu_{t-1})$ for $t = 2, 3, ..., T$ and, d_t and y_{t-1} are independent. Here $0 < \rho < 1$. The basic properties of the AR model, satisfying the condition $\rho < \min\left(\frac{\mu_t}{n_t\mu_{t-1}}, 1\right)$, are as follows,

i. $\mu_t = exp(\boldsymbol{x}_t{}^T\boldsymbol{\beta})$,

ii. $\sigma_{tt} = \mu_t - n_t\rho^2\mu_{t-1} + n_t{}^2\rho^2\sigma_{t-1,t-1}$,

iii. $cov(y_t, y_{t-k}) = \left(\prod_{l=0}^{k-1} n_{t-l}\right) \rho^k \sigma_{t-k,t-k},$

iv. $corr(y_t, y_{t-k}) = \left(\prod_{l=0}^{k-1} n_{t-l}\right) \rho^k \sqrt{\frac{\sigma_{t-k,t-k}}{\sigma_{tt}}}.$

However, when $n_t = 1$, variable b_j then reduces to a binary variable leading y_{it} to follow a Poisson AR model with $E(y_t) = v(y_t) = \mu_t = exp(\boldsymbol{x}_t{}^T\boldsymbol{\beta}))$ (Oyet & Sutradhar, 2013).

1.2.2 Moving Average Model

A moving average (MA) model of order q can be defined as,

$$y_t = \sum_{l=1}^{q} \rho_l * d_{t-l} + d_t. \tag{1.6}$$

for $t = 2, 3, ..., T$ (McKenzie, 1988; Brännäs & Hall, 2001; Weiß, 2008). In particular, when $q = 1$, the MA(q) model reduces to the MA(1) model, given by

$$y_t = \rho * d_{t-1} + d_t, \tag{1.7}$$

assuming that $y_1 = d_1 \sim Poi(\mu_1 = exp(\boldsymbol{x}_1{}^T\boldsymbol{\beta}))$ and $d_t \sim Poi\left(\sum_{u=o}^{t-1}(-\rho)^u \mu_{u-t}\right)$ for all $t = 2, 3, ..., T$ (McKenzie, 1988). The mean and variance of the MA model are equal, $E(y_t) = v(y_t) = \mu_t = exp(\boldsymbol{x}_t{}^T\boldsymbol{\beta})$. The correlation between y_t and y_k is defined as,

$$corr(y_t, y_k) = \begin{cases} \dfrac{\rho\left[\sum_{t=0}^{min(t,k)-1}(-\rho)^t \mu_{min(t,k)-t}\right]}{\sqrt{\mu_t \mu_k}}, & \text{for } |t-k| = 1, \\ 0, & \text{for otherwise.} \end{cases}$$

In MA model, ρ must satisfy the condition $0 < \rho < min[1, \rho_{20}, ..., \rho_{t0}, ..., \rho_{T0}]$, where ρ_{t0} is the solution of $\sum_{u=o}^{t-1}(-\rho)^u \mu_{u-t} = 0$.

1.3 Generalized Quasi Likelihood Method

Several methods have been proposed in the literature for estimating the parameters of models for count data. Liang and Zeger (1986) proposed generalized estimating equation (GEE) approach where a "working" correlation matrix is used to estimate the covariate parameter. However, Crowder (1995) demonstrated that the asymptotic properties of the estimator can undergo a breakdown due to the uncertainty in the definition of the "working" correlation matrix. Later, Sutradhar and Das (1999) demonstrated that the independence assumption based quasi-likelihood (QL) approach is more efficient than the GEE approach considering situations where GEE provides consistent estimators for the covariate parameters. Sutradhar (2003) extended the work of Wedderburn (1974) to the general stationary setup and proposed a generalized quasi-likelihood (GQL) method to estimate the time independent covariates effect of the models for binary or count data. Furthermore, Sutradhar et al. (2010) demonstrated that the stationary correlation based estimation approach may lead to inefficient regression estimates for time dependent count data. They suggested a generalized quasi-likelihood (GQL) approach based on a true non-stationary correlation structure.

In this GQL approach the estimated lag correlation of the responses are used to construct the correlation structure. The correlation structure can be written as follows

$$C(\rho) = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} & \cdots & \rho_{1T} \\ \rho_{21} & 1 & \rho_{23} & \cdots & \rho_{2T} \\ \vdots & \vdots & \vdots & & \vdots \\ \rho_{T1} & \rho_{T2} & \rho_{T3} & \cdots & 1 \end{bmatrix} \tag{1.8}$$

where $\rho_{m,n}$ represents the correlation between y_m and y_n for $m = 1, 2, ..., T$ and $n = 1, 2, ..., T$. Let, $E(y_t) = \mu_t$, $v(y_t) = \sigma_{tt}$ and $A = \text{diag}(\sigma_{11}, \sigma_{22}, ..., \sigma_{TT})$. As $\Sigma(\rho)$ denote the variance covariance matrix of $\boldsymbol{y}$, $\Sigma(\rho) = A^{\frac{1}{2}} C(\rho) A^{\frac{1}{2}}$. Then GQL estimating equation for regression parameter, $\boldsymbol{\beta}$, is given by,

$$\boldsymbol{X}^T \boldsymbol{A} \Sigma^{\mathbf{-1}}(\rho) \, (\boldsymbol{y} - \boldsymbol{\mu}) = \mathbf{0}, \tag{1.9}$$

where $\boldsymbol{y} = (y_1, y_2, ..., y_T)^T$, $\boldsymbol{\mu} = (\mu_1, \mu_2, ..., \mu_T)^T$ and $\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_T)^T$. By using (1.9), estimate of $\boldsymbol{\beta}$ can be obtained assuming all $\rho_{m,n}$ are known. Newton-raphson iterative approach is used to solve the equation in (1.9).

1.4 Motivation

Al-Osh and Alzaid (1987) studied an AR model and discussed the Yule-Walker, Conditional Least Squares (CLS) estimates, and MLE estimates of model parameters. Some researchers examined AR processes as branching processes and their applications (Winnicki, 1988; Wei & Winnicki, 1990; Du Jin-Guan, 1991; Sutradhar et al., 2010; Weiß, 2015). However, in practice, one may encounter count data that is not suitable for AR models. Such data may exhibit both autocorrelation and moving average effects, making it crucial to account for both aspects to improve modeling and forecasting accuracy. McKenzie (1988) proposed an ARMA model but did not provide any estimation methods. To the best of our knowledge, there are no studies in the existing literature that investigate the properties, estimation methods, and forecasting functions of ARMA models for count data based on Poisson immigration variables. Therefore, in this thesis, we extend the work of Al-Osh and Alzaid (1987) and McKenzie (1988) to the estimation of ARMA(1,1) model parameters.

The objectives of our research are as follows:

i. develop the basic properties of stationary and non-stationary cases for proposed Poisson ARMA (1,1) model for binary offspring and ARMA (1,1) model with binomial offspring,

ii. compare the proposed ARMA (1,1) models for count data with ARMA (1,1) for continuous data to check for similarities,

iii. estimate the parameters of the models using Generalized Quasi Likelihood method and Generalized Method of Moments,

iv. derive asymptotic distribution of covariate effect parameter, $\boldsymbol{\beta}$, and

v. derive forecasting function for ARMA (1,1) model for binary offspring and ARMA (1,1) model with binomial offspring.

1.5 Main Contributions and Outline of Thesis

In this section, we will discuss the primary contributions of this thesis, highlighting the key findings, methodologies, and theoretical advancements made throughout our research.

First, we proposed an ARMA(1,1) model with binary offspring. The fundamental properties of this model are derived and summarized in Theorem 2.2.1. Additionally, we present the Generalized Quasi-Likelihood (GQL) estimation method for estimating the covariate effect parameter $\boldsymbol{\beta}$, along with the Generalized Method of Moments (GMM) for estimating the parameters ρ_1 and ρ_2. These methods were found to perform well in the simulation study. The asymptotic distribution of the covariate effect parameter $\boldsymbol{\beta}$ is derived and presented in Theorem 2.2.2. Finally, we derived the l-step ahead forecasting function, including the mean and variance of the forecast error.

We also proposed an ARMA(1,1) model with binomial offspring. The fundamental properties of this model are derived and summarized in Theorem 2.3.1. Furthermore, we present the GQL estimation method for estimating the parameter $\boldsymbol{\beta}$, along with the GMM for estimating the parameters ρ_1 and ρ_2. We derive the l-step ahead forecasting function, as well as the mean and variance of the forecast error. Additionally, we consider a special case where the number of offspring produced by an individual during a given time period does not depend on time. The fundamental properties of this model are derived and summarized in Theorem 2.3.2. We again present the GQL estimation method for estimating the parameter $\boldsymbol{\beta}$, along with the GMM for estimating the parameters ρ_1 and ρ_2. These methods perform well in the simulation study. We derived the l-step ahead forecasting function, along with the mean and variance of the forecast error.

This thesis is written in manuscript form. In Chapter 2, ARMA(1,1) models for count data are discussed in detail. Section 2.2 explores the ARMA(1,1) model for count data with binary offspring, while Section 2.3 covers the ARMA(1,1) model for count data with binomial offspring. The conclusion of the research is presented in Chapter 3, summarizing the key findings and contributions of the study. It also outlines some recommendations for future research as extensions of this thesis.

Chapter 2

Autoregressive Moving Average Models for Count Data

Abstract

When equally spaced time series of counts is observed along with covariate information at each time point several authors have discussed the analysis of such data with Poisson Autoregressive or Integer valued Autoregressive (INAR) models. The basic properties and estimation of these INAR models are well documented in the literature. When time series data exhibits both autocorrelation and moving average effects, it is imperative to account for both of these aspects for improving modelling and forecasting effectiveness. Consequently, we consider autoregressive moving average models of order (1,1) for count time series with covariate information. First, we consider binary offspring random variable so that each member of a family at time $t-1$ is allowed to only a single offspring to the population at time t and propose an ARMA(1,1) model with binary offspring. We also extend our proposed model by allowing the possibility of an individual producing more than one offspring in a time point and propose an extended ARMA(1,1) model with binomial offspring. We derive the basic properties of the models and discuss the estimation of the model parameters. The performance of our proposed methods are examined through simulation studies.

Keywords: moving average model, autoregressive moving average model, generalized quasi likelihood method, generalized method of moments.

2.1 Introduction

Several research are available for Poisson autoregressive model where a specified characteristic of interest is considered to be correlated with the observations of previous time points.

Al-Osh and Alzaid (1987) introduced a model for a stationary sequence of integer-valued random variables with lag-one dependence, naming it the integer-valued autoregressive process of order one (INAR(1) process). The binomial thinning operation, denoted by $*$, was used to replace scalar multiplication of discrete random variables (Steutel et al., 1983). Specifically, if Y is a discrete random variable taking non-negative integer values and $\rho \in [0, 1]$, then the binomial thinning operation is defined as,

$$\rho * Y = \sum_{j=1}^{Y} b_j(\rho),$$

where $b_j(\rho)$ is an identically and independently distributed binary random variable with $P[b_j(\rho) = 1] = \rho = 1 - P[b_j(\rho) = 0]$. In each model, thinning operations are independent of previous history of the process. In their paper, they demonstrated that the distribution properties of the INAR(1) process are similar to those of the AR(1) model for continuous data. They also discussed the estimation of parameters using the Yule-Walker estimators, the conditional least squares estimators and the maximum likelihood estimators.

Later, Mckenzie (1988) proposed a family of models for discrete time processes where Poisson distribution is considered as the marginal distribution. In his paper, he discussed AR(1), MA(1), MA(q) and ARMA(1,q) models. According to Mckenzie (1988), using a common innovation process, $\{W_t\}$, the ARMA(1,q) model can be defined as,

$$\text{The AR(1) component: } y_t = \rho * y_{t-1} + W_t.$$

$$\text{The MA}(q) \text{ component: } X_t = y_{t-q} + \sum_{j=1}^{q} b_j(\rho) W_{t-j}.$$

Oyet and Sutradhar (2013) modeled infectious disease data collected over a short period of time using a branching process with immigration and provided consistent estimates of the parameters in their proposed model. Zhang and Oyet (2014) extended this work by proposing a second order longitudinal dynamic model based on a second order branching process with immigration. However, the effect of the random term over the past time periods was not considered.

In this paper, we propose autoregressive moving average models of order (1,1) for count data. It is an extension of the work of Al-Osh and Alzaid (1987) and McKenzie (1988). The proposed Poisson ARMA(1,1) model with binary offspring is discussed in Section 2.2, with basic properties and parameter estimation methods outlined in Sections 2.2.1 and 2.2.2, respectively. The asymptotic distribution of the GQL estimate is presented in Section 2.2.3, followed by a simulation study evaluating the performance of the estimation methods in Section 2.2.4 and the forecasting function in Section 2.2.5. Additionally, we introduce an extended ARMA(1,1) model with binomial offspring in Section 2.3, where the properties and parameter estimation methods are discussed in Sections 2.3.1 and 2.3.2, respectively, and the forecasting function for this model is outlined in Section 2.3.3. A special case of the model is presented in Section 2.3.4, and Section 2.3.5 demonstrates a simulation study to assess the performance of the estimation methods for the model parameters. Finally, conclusion and potential future research are provided in Section 2.4.

2.2 ARMA Model for Count Data with Binary Offspring

Let y_t be the total number of individuals at time t with a specified characteristics of interest. Let $\boldsymbol{x}_t = (\boldsymbol{x}_{t1}, \boldsymbol{x}_{t2}, \ldots, \boldsymbol{x}_{tp})^T$ be a p-dimensional vector of covariates at time t, and the correlation parameters be ρ_1 and ρ_2 such that $0 < \rho_1 < 1$ and $0 < \rho_2 < 1$. Additionally, let d_t represent the immigration variable, indicating the number of individuals entering the population from different regions at time t. Then, our proposed version of ARMA(1,1) model is given by

$$y_t = \rho_1 * y_{t-1} + \rho_2 * d_{t-1} + d_t, \tag{2.1}$$

In this model, we assume that an individual can produce either one offspring or none, with the offspring variable, b_{1j}, represented as a binary random variable with probability ρ_1. Furthermore, b_{2j} is regarded as a binary variable with a probability of ρ_2, capturing the effect of immigration variables from previous time periods. Then proposed model in (2.1) can be written as,

$$y_t = \sum_{j=1}^{y_{t-1}} b_{1j}(\rho_1) + \sum_{j=1}^{d_{t-1}} b_{2j}(\rho_2) + d_t. \tag{2.2}$$

We make the following assumptions about our proposed ARMA(1,1) model in (2.2),

Assumption 1.1. $y_0 = 0$.

Assumption 1.2. $y_1 \sim \text{Poi}(\mu_1 = exp(\boldsymbol{x}_1{}^T \boldsymbol{\beta}))$ and $d_1 \sim \text{Poi}(\mu_1 = exp(\boldsymbol{x}_1{}^T \boldsymbol{\beta}))$.

Assumption 1.3. $d_t \sim \text{Poi}(\mu_t - \rho_1 \mu_{t-1} - \rho_2 \mu_{d_{t-1}})$, for all $t = 2, 3, ..., T$.

Assumption 1.4. d_t and y_{t-1} are independent for $t = 2, 3, ..., T$.

Here, $0 < \rho_1 < 1$ and $0 < \rho_2 < 1$. Again, $\mu_t = exp(\boldsymbol{x}_t{}^T \beta)$, where $\boldsymbol{x}_t = (x_{t_1}, x_{t_2}, ..., x_{t_p})^T$ and $\boldsymbol{\beta} = (\beta_1, \beta_2, ..., \beta_p)^T$ for $t = 1, 2, ..., T$. The mean of Poisson must be non-negative. Therefore, for $t = 2, 3, \ldots, T$,

$$\rho_1 < min\left\{\frac{\mu_t}{\mu_{t-1}} - \rho_2 \frac{\mu_{d_{t-1}}}{\mu_{t-1}}, 1\right\}, \quad \text{for fixed} \quad \rho_2. \tag{2.3}$$

For stationary case, when $\boldsymbol{x}_t = \boldsymbol{x}, \mu_t = \mu$,

$$\rho_1 < min\left\{1 - \rho_2 \frac{\mu_{d_{t-1}}}{\mu}, 1\right\}, \quad \text{for fixed} \quad \rho_2. \tag{2.4}$$

2.2.1 Basic Properties

Based on assumption 1.2, $E(y_1) = v(y_1) = \mu_1$. Additionally, $E(d_1) = v(d_1) = \mu_1$. By taking successive expectations, and using conditional mean and variance, the mean and variance of y_t is obtained for $t = 1, 2, \ldots, T$. The following Theorem 2.2.1 outlines the basic properties of the model in 2.2. Its proof is given in the Appendix A.1.

Theorem 2.2.1 *Let y_t be the number of individuals at time t, and d_t be the immigration variable at time t for $t = 1, 2, ..., T$. Consider a binary offspring variable for* *the binomial thinning operation in model (2.2), which satisfies Assumptions 1.1-1.4. Then,* $y_t \sim Poi(\mu_t = exp(\boldsymbol{x}_t{}^T \boldsymbol{\beta}))$*, with the following properties:*

i. The covariance between y_t *and* y_{t-k} *is given by:*

$$cov(y_t, y_{t-k}) = \begin{cases} (\rho_1 + \rho_2)\mu_{t-1} - \rho_2(\rho_1\mu_{t-2} + \rho_2\mu_{d_{t-2}}), & \textit{for } k = 1, \\ \rho_1^k \mu_{t-k}, & \textit{for } k > 1. \end{cases}$$

ii. The correlation between y_t *and* y_{t-k} *is given by:*

$$corr(y_t, y_{t-k}) = \begin{cases} (\rho_1 + \rho_2)\sqrt{\dfrac{\mu_{t-1}}{\mu_t}} - \rho_1\rho_2\dfrac{\mu_{t-2}}{\sqrt{\mu_t\mu_{t-1}}} - \rho_2^2\dfrac{\mu_{d_{t-2}}}{\sqrt{\mu_t\mu_{t-1}}}, & \textit{for } k = 1, \\ \rho_1^k\sqrt{\dfrac{\mu_{t-k}}{\mu_t}}, & \textit{for } k > 1. \end{cases}$$

Stationary Case: In particular, when y_t do not depend on time t for $t = 1, 2, ..., T$, $\boldsymbol{x}_t = \boldsymbol{x}$ leads to $\mu_t = \mu$, and $\text{corr}(y_t, y_{t-k}) = \begin{cases} \rho_1 + \rho_2 - \rho_1\rho_2 - \rho_2^2\dfrac{\mu_{d_{t-2}}}{\mu_t}, & \text{for } k = 1, \\ \rho_1^k, & \text{for } k > 1. \end{cases}$

The expression of autocorrelation for $k > 1$ in Theorem 2.2.1 is same as the autocorrelation structure of AR(1) model for continuous data. This is also apparent in the stationary case. It can be concluded that for $k > 1$, ARMA(1,1) model behaves like AR(1) model whereas for continuous process ACF of ARMA(1,1) model also behaves like AR(1) model after lag 1 (Al-Osh & Alzaid, 1987). This is one of the many similarities between the properties of ARMA(1,1) for count data and ARMA(1,1) for continuous data.

2.2.2 Estimation of Parameters

In this section, estimation methods to estimate the parameters of Poisson ARMA(1,1) model with binary offspring are developed. To estimate the parameter $\boldsymbol{\beta}$ a Generalized Quasi-Likelihood (GQL) method is used assuming parameters ρ_1 and ρ_2 are fixed. Again, Generalized Method of Moments (GMM) is used to estimate parameters ρ_1 and ρ_2 for fixed $\boldsymbol{\beta}$.

2.2.2.1 Estimation of Parameter $\boldsymbol{\beta}$

Let the $(T \times P)$ matrix of covariates be defined as $\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_T)^\top$, the response vector $\boldsymbol{y} = (y_1, y_2, \ldots, y_T)^\top$, and the mean vector $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_T)^\top$. Let $\boldsymbol{\Sigma}$ be the variance-covariance matrix of $\boldsymbol{y}$. In estimating $\boldsymbol{\beta}$ using the GQL method, we assume that ρ_1 and ρ_2 are known. The expected value of y_t is given by $E(y_t) = \mu_t = e^{\boldsymbol{x}_t^\top \boldsymbol{\beta}}$ for all $t = 1, 2, 3, \ldots, T$. The GQL estimating equation for $\boldsymbol{\beta}$ is then given by

$$\frac{\partial \boldsymbol{\mu}^T}{\partial \boldsymbol{\beta}} \Sigma^{-1} (\boldsymbol{y} - \boldsymbol{\mu}) = \mathbf{0}$$

$$or, \boldsymbol{X}^T \boldsymbol{U} \Sigma^{\mathbf{-1}} (\boldsymbol{y} - \boldsymbol{\mu}) = \mathbf{0}, \tag{2.5}$$

where, $\boldsymbol{U}_{(T \times T)} = diag(\mu_1, \mu_2, \ldots, \mu_T)^T$. The symmetric covariance matrix Σ in terms of ρ_1 and ρ_2, can be defined as,

$$\Sigma = \begin{bmatrix} \mu_1 & (\rho_1+\rho_2)\mu_1 & \rho_1^2\mu_1 & \cdots & \rho_1^{T-1}\mu_1 \\ & \mu_2 & \begin{array}{c}(\rho_1+\rho_2)\mu_2 \\ +\rho_2(\rho_1\mu_1 + \rho_2\mu_{d_1})\end{array} & \cdots & \rho_1^{T-2}\mu_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ & & & & \begin{array}{c}(\rho_1+\rho_2)\mu_{T-1} \\ -\rho_2(\rho_1\mu_{T-2} + \rho_2\mu_{d_{T-2}}) \\ \mu_T\end{array} \end{bmatrix}$$

By using the Newton-Raphson iterative approach the GQL estimating equation can be solved as (Wedderburn, 1974),

$$\hat{\boldsymbol{\beta}}_{(r+1)} = \hat{\boldsymbol{\beta}}_{(r)} + \left[\boldsymbol{X}^T \boldsymbol{U} \Sigma^{-1} \boldsymbol{U} \boldsymbol{X}\right]^{-1} \left[\boldsymbol{X}^T \boldsymbol{U} \Sigma^{-1} (\boldsymbol{y} - \boldsymbol{\mu})\right]_{\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}_{(r)}}. \tag{2.6}$$

Here, $\hat{\boldsymbol{\beta}}_{(r)}$ is the estimated value of $\boldsymbol{\beta}$ at rth iteration. The GQL estimate is a consistent estimate since $E\left(\boldsymbol{X}^\top \boldsymbol{U}\ \boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{\mu})\right) = 0$ (Zhang & Oyet, 2014).

2.2.2.2 GMM Estimation of Parameters ρ_1 and ρ_2

Let us define, S_{tt}, $S_{t,t+1}$ and $S_{t,t+2}$ to be the standardized sample variance, standardized sample lag 1 autocovariance and standardized sample lag 2 autocovariance respectively. Then,

$$S_{tt} = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{y_t-\mu_t}{\sigma_t}\right)^2,$$

$$S_{t,t+1} = \frac{1}{T-1}\sum_{t=1}^{T-1}\left(\frac{y_t-\mu_t}{\sigma_t}\right)\left(\frac{y_{t+1}-\mu_{t+1}}{\sigma_{t+1}}\right),$$

$$\text{and} \quad S_{t,t+2} = \frac{1}{T-2}\sum_{t=1}^{T-2}\left(\frac{y_t-\mu_t}{\sigma_t}\right)\left(\frac{y_{t+2}-\mu_{t+2}}{\sigma_{t+2}}\right).$$

Using first order approximation and assuming higher orders are negligible, since $E(S_{tt}) = 1$, the moment equations are,

$$\frac{S_{t,t+1}}{S_{tt}} = E\left(\frac{S_{t,t+1}}{S_{tt}}\right) = E(S_{t,t+1}) \tag{2.7}$$

$$\text{or,} \quad \frac{S_{t,t+1}}{S_{tt}} = \sum_{t=1}^{T-1}\frac{1}{T-1}\left[(\rho_1+\rho_2)\sqrt{\frac{\mu_t}{\mu_{t+1}}} - \rho_1\rho_2\frac{\mu_{t-1}}{\sqrt{\mu_t\mu_{t+1}}} - {\rho_2}^2\frac{\mu_{d_{t-1}}}{\sqrt{\mu_t\mu_{t+1}}}\right] \tag{2.8}$$

$$\text{and} \quad \frac{S_{t,t+2}}{S_{tt}} = E\left(\frac{S_{t,t+2}}{S_{tt}}\right) = E(S_{t,t+2}) \tag{2.9}$$

$$\text{or,} \quad \frac{S_{t,t+2}}{S_{tt}} = \sum_{t=1}^{T-2}\frac{1}{T-2}{\rho_1}^2\sqrt{\frac{\mu_t}{\mu_{t+2}}}. \tag{2.10}$$

By solving the moment equations and using Newton-Raphson iterative approach, the estimate of ρ_1 and ρ_2 can be obtained as,

$$\hat{\rho}_1 = \left(\frac{S_{t,t+2}}{S_{tt}}\left[\frac{1}{T-2}\sum_{t=1}^{T-2}\sqrt{\frac{\mu_t}{\mu_{t+2}}}\right]^{-1}\right)^{1/2} \tag{2.11}$$

$$\hat{\rho}_{2_{(r+1)}} = \hat{\rho}_{2_{(r)}} - \left[\frac{\partial g(\rho_1,\rho_2)}{\partial \rho_2}\right]^{-1} g(\rho_1,\rho_2)\Bigg|_{\rho_1=\hat{\rho}_{1_{(r)}},\rho_2=\hat{\rho}_{2_{(r)}}} \tag{2.12}$$

where,

$$g(\rho_1,\rho_2) = \frac{S_{t,t+1}}{S_{tt}} - \sum_{t=1}^{T-1} \frac{1}{T-1}\left[(\rho_1+\rho_2)\sqrt{\frac{\mu_t}{\mu_{t+1}}} - \rho_1\rho_2\frac{\mu_{t-1}}{\sqrt{\mu_t\mu_{t+1}}} - {\rho_2}^2\frac{\mu_{d_{t-1}}}{\sqrt{\mu_t\mu_{t+1}}}\right]$$

$$\frac{\partial g(\rho_1,\rho_2)}{\partial\rho_2} = -\sum_{t=1}^{T-1}\frac{1}{T-1}\left[\sqrt{\frac{\mu_t}{\mu_{t+1}}} - \rho_1\frac{\mu_{t-1}}{\sqrt{\mu_t\mu_{t+1}}} - 2\rho_2\frac{\mu_{d_{t-1}}}{\sqrt{\mu_t\mu_{t+1}}}\right].$$

Here, $\hat{\rho}_{1_{(r)}}$ and $\hat{\rho}_{2_{(r)}}$ are the estimated values of ρ_1 and ρ_2 at the rth iteration respectively.

2.2.3 Asymptotic Distribution of GQL Estimate $\boldsymbol{\beta}$

The following Theorem 2.2.2 outlines the asymptotic distribution of $\hat{\boldsymbol{\beta}}$. Proof of Theorem 2.2.2 is given in the Appendix A.2.

Theorem 2.2.2 *Let the GQL (Generalized Quasi-Likelihood) estimating function for $\boldsymbol{\beta}$ be given by*

$$\boldsymbol{X}^T\boldsymbol{U}\Sigma^{-1}(\boldsymbol{y}-\boldsymbol{\mu}) = \boldsymbol{X}^T\boldsymbol{U}\left[\boldsymbol{A}^{\frac{1}{2}}\boldsymbol{C}(\rho)\boldsymbol{A}^{\frac{1}{2}}\right]^{-1}(\boldsymbol{y}-\boldsymbol{\mu}) = \sum_{j=1}^{T}\sum_{i=1}^{T}\frac{\boldsymbol{x}_j\mu_j q_{ji}(y_i-\mu_i)}{\sigma_j\sigma_i}$$

where $\boldsymbol{A}^{\frac{1}{2}} = diag(\sigma_1,\sigma_2,\ldots,\sigma_T)^{\frac{1}{2}}$ and $\boldsymbol{C}(\rho)^{-1} = (\boldsymbol{q}_1,\boldsymbol{q}_2,\ldots,\boldsymbol{q}_T)^T$. Here, q_i is the inverse of $(i-1)$th lag autocorrelation of $\boldsymbol{y}$ for $i = 1,2,...,T$.
As $T\to\infty$, the estimator $\hat{\boldsymbol{\beta}}$ asymptotically follows a Gaussian distribution with mean $\boldsymbol{\beta}$ and covariance matrix,

$$cov(\hat{\beta}) = R^* = \left[\boldsymbol{X}^T\boldsymbol{U}\Sigma^{-1}\boldsymbol{U}\boldsymbol{X}\right]^{-1}.$$

2.2.4 Forecasting

One of the primary goals of time series analysis is to predict future values of the series. This section focuses on the approach to forecasting, where the l step ahead forecast of y_t is denoted by $\hat{y}_{t+l}$. Here, t is the forecast origin and l is the lead time. After estimating the model parameters, the l step ahead forecast for y_t can be derived as: $\hat{y}_{t+l} = y_t(l) = E(y_{t+l}|y_{t+l-1})$ (Brockwell & Davis, 2002; Freeland & McCabe, 2004;

Sutradhar, 2008). The 1 step ahead forecast of y_t can be obtained as

$$y_t(1) = E(y_{t+1}|y_t) = E_{d_t}\left(E(y_{t+1}|y_t, d_t)\right).$$

Now using equation (A.2) from Appendix A, it can written as

$$E\left(y_{t+1}|y_t, d_t\right) = \rho_1 y_t + \rho_2 d_t + E(d_{t+1})$$

$$\begin{aligned} E_{d_t}\left(E\left(y_{t+1}|y_t, d_t\right)\right) &= \rho_1 y_t + \rho_2 E(d_t) + E(d_{t+1}) \\ &= \rho_1 y_t + \rho_2 \mu_{d_t} + \mu_{t+1} - \rho_1 \mu_t - \rho_2 \mu_{d_t} \\ &= \mu_{t+1} + \rho_1 (y_t - \mu_t). \end{aligned}$$

Therefore, $y_t(1) = E(y_{t+1}|y_t) = \mu_{t+1} + \rho_1(y_t - \mu_t)$, where $y_t = y_t(0)$. The forecast error can be obtained as

$$\begin{aligned} e_t(1) &= y_{t+1} - y_t(1) \\ &= y_{t+1} - \mu_{t+1} - \rho_1(y_t - \mu_t). \end{aligned}$$

Now, $E(e_t(1)) = E(y_{t+1}) - E(y_t(1)) = \mu_{t+1} - \mu_{t+1} - \rho_1(\mu_t - \mu_t) = 0$. This implies that the 1 step ahead forecast is unbiased. To calculate the variance of the forecast error, the following conditional variance is used.

$$\begin{aligned} v(e_t(1)) &= E_{d_t}\left(v(e_t(1)|d_t)\right) + v_{d_t}\left(E(e_t(1)|d_t)\right), \\ \text{where, } E(e_t(1)|y_t, d_t) &= E(y_{t+1}|y_t, d_t) - E(y_t(1)|y_t, d_t) \\ &= \rho_1 y_t + \rho_2 d_t + \mu_{d_{t+1}} - \mu_{t+1} - \rho_1(y_t - \mu_t) \\ &= \rho_2 d_t + \mu_{d_{t+1}} - \mu_{t+1} + \rho_1 \mu_t. \end{aligned}$$

$$\begin{aligned} \text{Consequently, } E(e_t(1)|d_t) &= E_{y_t}(E(e_t(1)|y_t, d_t)) \\ &= E_{y_t}(\rho_2 d_t + \mu_{d_{t+1}} - \mu_{t+1} + \rho_1 \mu_t) \\ &= \rho_2 d_t + \mu_{d_{t+1}} - \mu_{t+1} + \rho_1 \mu_t \end{aligned}$$

$$
\begin{aligned}
\text{Again, } v(e_t(1)|d_t) &= E_{y_t}\left(v(e_t(1)|y_t, d_t)\right) + v_{y_t}\left(E(e_t(1)|y_t, d_t)\right) \\
&= E_{y_t}\left(v(y_{t+1} - y_t(1)|y_t, d_t)\right) + v_{y_t}\left(E(e_t(1)|y_t, d_t)\right) \\
&= E_{y_t}\left(v(y_{t+1}|y_t, d_t)\right) + v_{y_t}\left(E(e_t(1)|y_t, d_t)\right) \\
&= E_{y_t}\left(\rho_1(1-\rho_1)y_t + \rho_2(1-\rho_2)d_t + \mu_{t+1} - \rho_1\mu_t - \rho_2\mu_{d_t}\right) \\
&\quad + v_{y_t}(\rho_2 d_t + \mu_{d_{t+1}} - \mu_{t+1} + \rho_1\mu_t) \\
&\quad [\text{Using (A.6)}] \\
&= \mu_{t+1} - \rho_1^2\mu_t + \rho_2(1-\rho_2)d_t - \rho_2\mu_{d_t} \\
\text{Hence, } v(e_t(1)) &= E_{d_t}\left(v(e_t(1)|d_t)\right) + v_{d_t}\left(E(e_t(1)|d_t)\right)
\end{aligned}
$$

$$
\begin{aligned}
&= E_{d_t}\left(\mu_{t+1} - \rho_1^2\mu_t + \rho_2(1-\rho_2)d_t - \rho_2\mu_{d_t}\right) \\
&\quad + v_{d_t}(\rho_2 d_t + \mu_{d_{t+1}} - \mu_{t+1} + \rho_1\mu_t) \\
&= \mu_{t+1} - \rho_1^2\mu_t - \rho_2^2\mu_{d_t} + \rho_2^2\mu_{d_t} \\
&= \mu_{t+1} - \rho_1^2\mu_t
\end{aligned}
$$

2.2.5 Simulation Study

A simulation study was conducted to visualize the sample paths for the Poisson ARMA(1,1) model with binary offspring for count data and Normal ARMA(1,1) model for continuous data in stationary setup. An R code was developed for this purpose. Data was generated from the proposed Poisson ARMA(1,1) model in (2.2), assuming 2nd order weak stationarity (i.e.$\mu_t = \mu$) with $T = 100$, $\boldsymbol{\beta} = (0.2, 0.6)^T$, $\rho_2 = 0.4$ and $\boldsymbol{x} = (3, 1.5)^T$. For fixed $\rho_2 = 0.4$, we used (2.4) to obtain $\rho_1 = 0.5$. Now, ARMA(1,1) model for continuous data can be written as,

$$
y_t = (1 - \phi_1)\mu_c + \phi_1 y_{t-1} + a_t - \theta_1 a_{t-1}, \tag{2.13}
$$

where $a_t \overset{iid}{\sim} \mathrm{N}(0, 1)$. Using ϕ_1=0.5 and $\mu = \mu_c = v(y_t) = 4.48$, the value of θ_1 is obtained as $\theta_1 = -0.1086$. Using these values in (2.13), data were generated from the Normal ARMA(1,1) model for 100 time points. Figure 2.1 shows the sample paths of y_t's generated from the continuous Normal ARMA(1,1) model overlaid on the sample path of data generated from Poisson ARMA(1,1) model. The sample paths for both models were constructed using the same mean and variance. However, a wider spread is observed in the data generated from the Poisson ARMA(1,1) model compared to the ARMA(1,1) model for continuous data.

Furthermore, an extensive simulation study was carried out to evaluate the performance of the estimation methods for $\boldsymbol{\beta}$, ρ_1 and ρ_2. R codes were developed to carry out the simulation study. Data was generated for different combinations of $\boldsymbol{\beta}$, ρ_1 and ρ_2, while time points $T = 250$, $T = 500$ and $T = 700$ were used to examine the effect of time lengths on the estimates. Two covariates x_{t1} and x_{t2} were considered in this study, where for $t = 1, 2, ..., T$,

$$x_{t1} \sim Binomial(1, 0.5) \quad \text{and,} \quad x_{t2} = \frac{t}{T}$$

Once generated, it was fixed for all simulations. Due to the restrictions on ρ_1 and ρ_2

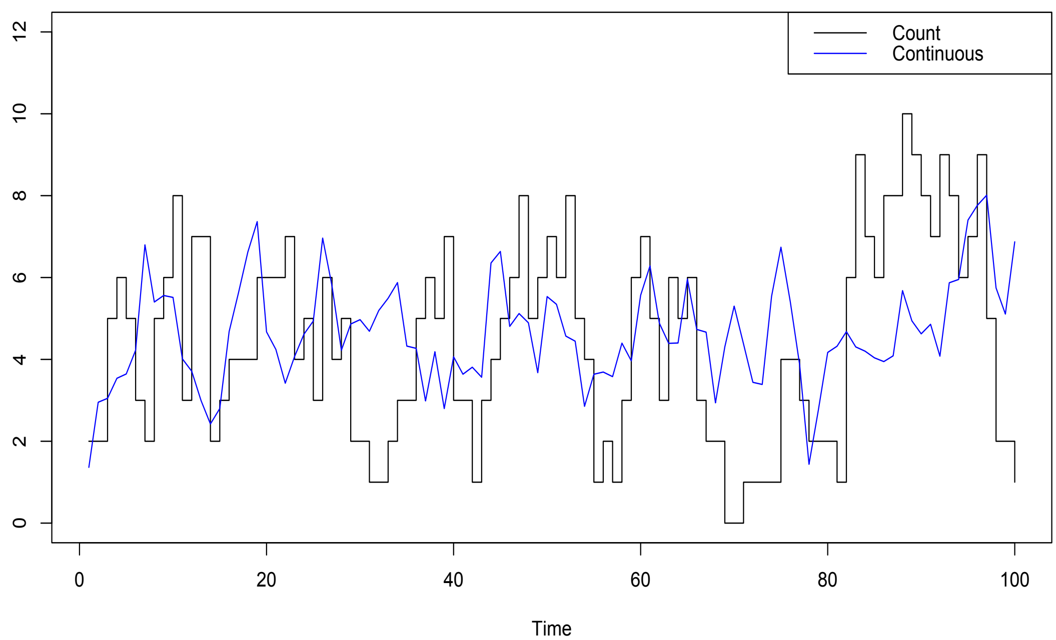

Figure 2.1: A plot of data generated from Poisson ARMA(1,1) model for count data (black step) overlaying sample paths from Normal ARMA(1,1) model for continuous data (blue line)

outlined in (2.3), only a narrow range of values can be chosen for these parameters. After some trial and error, two values for ρ_2 were fixed, and the corresponding ρ_1

values were calculated satisfying the condition in (2.3). For each combination of $\boldsymbol{\beta}$, ρ_1 and ρ_2, we generated y_1 and d_1 from $Poi(\mu_1)$, and d_t from $Poi(\mu_t - \rho_1\mu_{t-1} - \rho_2\mu_{d_{t-1}})$ for $t = 2, 3, ..., T$. Finally, using y_1 and d_t's, y_t's were generated from the proposed model (2.2) for $t = 2, 3, ..., T$.

Initial values of $\boldsymbol{\beta} = (0, 0)$, $\rho_1 = 0$ and $\rho_2 = 0$ were used to estimate the parameters using equations (2.6), (2.11) and (2.12). Iteration for equations (2.6) and (2.12) were continued until convergence was achieved. This procedure was repeated 1000 times for fixed values of $\boldsymbol{\beta}$, ρ_1 and ρ_2. The average estimates, $\hat{\boldsymbol{\beta}}$, $\hat{\rho}_1$ and $\hat{\rho}_2$, from the 1000 simulations are presented in Table 2.1. Additionally, the standard error of each estimated parameter was calculated from the 1000 simulations to show the dispersion around the mean.

Table 2.1 shows that the GQL method perfoms well to estimate $\boldsymbol{\beta}$, while the GMM method effectively estimates ρ_1 and ρ_2. As T increases, the estimates approach the true values, and the standard error decreases. For example, when T=700, with $\boldsymbol{\beta} = (0.2, 0.3)$, $\rho_1 = 0.3$ and $\rho_2 = 0.2$, the estimates are, $\hat{\boldsymbol{\beta}} = (0.200, 0.298)$, $\hat{\rho}_1 = 0.353$ and $\hat{\rho}_2 = 0.172$, whereas for T=250, the estimates are $\hat{\boldsymbol{\beta}} = (0.164, 0.417)$, $\hat{\rho}_1 = 0.331$ and $\hat{\rho}_2 = 0.132$. Moreover, it can be observed that when $T = 250$, the estimated values are not as close to the true values. However, for $T = 500$ and $T = 700$, the GQL and GMM methods provide better estimates of the model parameters.

2.3 ARMA Model for Count Data with Binomial Offspring

In this section, we propose an ARMA(1,1) model assuming that an individual can produce more that one offspring. Assuming that an individual can produce n_t offspring at time point t, for $t = 1, 2, ..., T$, an extended autoregressive moving average model of order (1,1) can be defined as,

$$y_t = \sum_{j=1}^{y_{t-1}} b_{1j}(n_t, \rho_1) + \sum_{j=1}^{d_{t-1}} b_{2j}(\rho_2) + d_t. \tag{2.14}$$

	True Values			Parameter Estimates					
T	$\boldsymbol{\beta}$	ρ_1	ρ_2	$\hat{\boldsymbol{\beta}}$	$SE(\hat{\boldsymbol{\beta}})$	$\hat{\rho}_1$	$SE(\hat{\rho}_1)$	$\hat{\rho}_2$	$SE(\hat{\rho}_2)$
250	(0,0)	0.3	0.2	(-0.063,0.102)	(0.002,0.003)	0.472	0.001	0.063	0.002
500				(0.060,-0.104)	(0.001,0.002)	0.339	0.001	0.164	0.002
700				(0.004,-0.035)	(0.001,0.001)	0.352	0.001	0.184	0.002
250		0.4	0.3	(-0.066,0.023)	(0.002,0.003)	0.568	0.001	0.090	0.002
500				(0.047,-0.058)	(0.001,0.002)	0.438	0.001	0.193	0.002
700				(-0.012,-0.024)	(0.001,0.002)	0.482	0.001	0.166	0.001
250	(0.3,0)	0.3	0.2	(0.272,-0.034)	(0.002,0.004)	0.482	0.002	0.063	0.002
500				(0.318,-0.063)	(0.001,0.002)	0.345	0.002	0.149	0.002
700				(0.284,-0.004)	(0.001,0.002)	0.374	0.001	0.153	0.002
250		0.4	0.3	(0.247,0.035)	(0.002,0.004)	0.562	0.001	0.110	0.002
500				(0.323,-0.022)	(0.001,0.002)	0.460	0.001	0.166	0.001
700				(0.282,-0.041)	(0.001,0.002)	0.458	0.001	0.197	0.001
250	(0.2,0.3)	0.3	0.2	(0.164,0.417)	(0.002,0.003)	0.331	0.002	0.132	0.003
500				(0.211,0.247)	(0.001,0.002)	0.331	0.002	0.151	0.002
700				(0.200,0.298)	(0.001,0.002)	0.353	0.001	0.172	0.002
250		0.4	0.3	(0.182,0.276)	(0.002,0.004)	0.549	0.001	0.111	0.002
500				(0.224,0.248)	(0.001,0.002)	0.461	0.001	0.163	0.002
700				(0.179,0.288)	(0.001,0.002)	0.477	0.001	0.191	0.002
250	(0.3, 0.5)	0.3	0.2	(0.267,0.460)	(0.002,0.003)	0.482	0.001	0.075	0.002
500				(0.325,0.491)	(0.001,0.002)	0.339	0.002	0.152	0.002
700				(0.317,0.470)	(0.001,0.002)	0.363	0.001	0.149	0.002
250		0.4	0.3	(0.274,0.409)	(0.004,0.006)	0.542	0.002	0.137	0.002
500				(0.318,0.493)	(0.001,0.002)	0.466	0.001	0.160	0.002
700				(0.302,0.502)	(0.001,0.002)	0.465	0.001	0.191	0.002

Table 2.1: Comparison of True and Estimated Parameter Values of Poisson ARMA(1,1) model with binary offspring for different combination of $\boldsymbol{\beta}$, ρ_1, ρ_2 and T values

In the model (2.14), we assume that the offspring variable $b_{1j} \sim Binomial(n_t, \rho_1)$ and the variable $b_2 j \sim Binomial(1, \rho_2)$.. Assumptions about ARMA(1,1) model for count data with binomial offspring in (2.14) are,

Assumption 2.1. $y_0 = 0$.

Assumption 2.2. $y_1 \sim \text{Poi}(\mu_1 = exp(\boldsymbol{x}_1{}^T\boldsymbol{\beta}))$ and $d_1 \sim \text{Poi}(\mu_1 = exp(\boldsymbol{x}_1{}^T\boldsymbol{\beta}))$.

Assumption 2.3. $d_t \sim \text{Poi}(\mu_t - \rho_1 n_t \mu_{t-1} - \rho_2 \mu_{d_{t-1}})$, for all $t = 2, 3, ..., T$.

Assumption 2.4. d_t and y_{t-1} are independent for $t = 2, 3, ..., T$.

Here, $0 < \rho_1 < 1$, $0 < \rho_2 < 1$ and, $\mu_t = exp(\boldsymbol{x}_t{}^T\boldsymbol{\beta})$, where $\boldsymbol{x}_t = (x_{t_1}, x_{t_2}, ..., x_{t_p})^T$ and $\boldsymbol{\beta} = (\beta_1, \beta_2, ..., \beta_p)^T$ for $t = 1, 2, ..., T$. The mean of Poisson must be non-negative. Therefore, for $t = 2, 3, \ldots, T$,

$$\rho_1 < min\left\{\frac{\mu_t}{n_t\mu_{t-1}} - \rho_2\frac{\mu_{d_{t-1}}}{\mu_{t-1}}, 1\right\}, \quad \text{for fixed} \quad \rho_2. \tag{2.15}$$

In particular, for stationary case, when $\boldsymbol{x}_t = \boldsymbol{x}, \mu_t = \mu$,

$$\rho_1 < min\left\{\tfrac{1}{n_t} - \rho_2\tfrac{\mu_{d_{t-1}}}{\mu}, 1\right\}, \text{ for fixed } \rho_2.$$

2.3.1 Basic Properties

Basic properties of ARMA(1,1) model for count data with binomial offspring are obtained by using conditional expectation and conditional variance formulas. The following Theorem 2.3.1 outlines the basic properties of the model in (2.14). Its proof is given in the Appendix A.3.

Theorem 2.3.1 *Let y_t be the number of individuals at time t, and d_t be the immigration variable at time t for $t = 1, 2, ..., T$. If a binomial offspring variable for the binomial thinning operation in model (2.14) is considered and model (2.14) satisfies Assumptions 2.1-2.4,*

$$i. E(y_t) = \mu_t = exp(\boldsymbol{x}_t{}^T\boldsymbol{\beta}), \quad \textit{for all} \quad t = 1, 2, 3, ..., T.$$

$$ii. v(y_t) = \begin{cases} \mu_1, & \textit{for } t = 1, \\ \mu_2 + \rho_1^2 n_2(n_2 - 1)\mu_1, & \textit{for } t = 2, \\ \mu_t + \rho_1^2 n_t(n_t - 1)\mu_{t-1} & \\ + \sum_{l=1}^{t-2}\left[\rho_1^{2(l+1)} n_{t-l}(n_{t-l} - 1)\left(\prod_{j=0}^{l-1} n_{t-j}^2\right)\mu_{t-(l+1)}\right], & \textit{for } t = 3, .., T. \end{cases}$$

iii. The covariance between y_t and y_{t-k} is given by:

$$cov(y_t, y_{t-k}) = \begin{cases} n_t\rho_1\sigma_{t-1,t-1} + \rho_2\sigma_{d_{t-1},d_{t-1}}, & \textit{for } k = 1, \\ \left(\prod_{l=0}^{k-1} n_{t-l}\right)\rho_1^k\sigma_{t-k,t-k}, & \textit{for } k > 1. \end{cases}$$

iv. The correlation between y_t and y_{t-k} is given by:

$$corr(y_t, y_{t-k}) = \begin{cases} \rho_1 n_t \sqrt{\dfrac{\sigma_{t-1,t-1}}{\sigma_{t,t}}} + \rho_2 \dfrac{\sigma_{d_{t-1},d_{t-1}}}{\sqrt{\sigma_{t,t}\sigma_{t-1,t-1}}}, & \textit{for } k = 1, \\ \left(\prod_{l=0}^{k-1} n_{t-l}\right) \rho_1^k \sqrt{\frac{\sigma_{t-k,t-k}}{\sigma_{t,t}}}, & \textit{for } k > 1. \end{cases}$$

2.3.2 Estimation of Parameters

In this section, Generalized Quasi-Likelihood (GQL) method to estimate the parameter $\boldsymbol{\beta}$ and Generalized Method of Moments (GMM) to estimate parameters ρ_1 and ρ_2 of proposed extended ARMA(1,1) model with binomial offspring are discussed.

2.3.2.1 Estimation of Parameter $\boldsymbol{\beta}$

Let $\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_T)^\top$ be the $(T \times P)$ matrix of covariates. Let us define the response vector as $\boldsymbol{y} = (y_1, y_2, \ldots, y_T)^\top$ and the mean vector as $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_T)^\top$. Let $\boldsymbol{\Sigma}$ represent the variance-covariance matrix of $\boldsymbol{y}$. In estimating $\boldsymbol{\beta}$ with the GQL method, we assume that ρ_1 and ρ_2 are known. The expected value of y_t is expressed as $E(y_t) = \mu_t = exp(\boldsymbol{x}_t^\top \boldsymbol{\beta})$ for $t = 1, 2, \ldots, T$. The GQL estimating equation for $\boldsymbol{\beta}$ is same as the one used for the model in (2.2). The only difference lies in the covariance structure. In this case, the symmetric covariance matrix is given by,

$$\Sigma = \begin{bmatrix} \mu_1 & n_2\rho_1\sigma_{11} + \rho_2\sigma_{d_1,d_1} & n_3\rho_1^2\sigma_{11} & \cdots & \prod_{l=0}^{T-2} n_{T-l}\rho_1^{T-1}\sigma_{11} \\ & \mu_2 + \rho_1^2 n_2(n_2-1)\mu_1 & n_3\rho_1\sigma_{22} + \rho_2\sigma_{d_2,d_2} & \cdots & \prod_{l=0}^{T-3} n_{T-l}\rho_1^{T-2}\sigma_{22} \\ & & & & \vdots \\ & & & & \begin{array}{c} \mu_T + \rho_1^2 n_T(n_T-1)\mu_{T-1} + \\ \sum_{l=1}^{T-2} \left[\rho_1^{2(l+1)} n_{T-l}(n_{T-l}-1) \right. \\ \left. \left(\prod_{j=0}^{l-1} n_{T-j}^2\right) \mu_{T-(l+1)}\right] \end{array} \end{bmatrix}$$

Therefore, by using Newton-Raphson iterative approach the GQL estimating equation can be solved as (Wedderburn, 1974),

$$\hat{\boldsymbol{\beta}}_{(r+1)} = \hat{\boldsymbol{\beta}}_{(r)} + \left[\boldsymbol{X}^T \boldsymbol{U} \Sigma^{-1} \boldsymbol{U} \boldsymbol{X}\right]^{-1} \left[\boldsymbol{X}^T \boldsymbol{U} \Sigma^{-1} \left(\boldsymbol{y} - \boldsymbol{\mu}\right)\right]_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}_{(r)}}. \tag{2.16}$$

Here, $\hat{\boldsymbol{\beta}}_{(r)}$ is the estimated value of $\boldsymbol{\beta}$ at rth iteration. The GQL estimate is a consistent estimate since $E\left(\boldsymbol{X}^\top \boldsymbol{U}\ \boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{\mu})\right) = 0$ (Zhang & Oyet, 2014).

2.3.2.2 GMM Estimation of Parameters ρ_1 and ρ_2

The GMM estimation of parameters ρ_1 and ρ_2 in the ARMA(1,1) model for count data with binomial offspring is similar to the approach discussed in Section 2.2.2.2. The moment equations are similar to those in (2.7) and (2.9). By solving these moment equations for a fixed $\boldsymbol{\beta}$ and applying the Newton-Raphson iterative method, the estimates of the parameters ρ_1 and ρ_2 in model (2.14) can be obtained as follows:

$$\hat{\rho}_1 = \left(\frac{S_{t,t+2}}{S_{tt}} \left[\frac{1}{T-2} \sum_{t=1}^{T-2} n_t n_{t+1} \sqrt{\frac{\sigma_{t,t}}{\sigma_{t+2,t+2}}} \right]^{-1} \right)^{1/2} \tag{2.17}$$

$$\hat{\rho}_{2_{(r+1)}} = \hat{\rho}_{2_{(r)}} - \left[\frac{\partial f(\rho_1, \rho_2)}{\partial \rho_2} \right]^{-1} f(\rho_1, \rho_2) \Bigg|_{\rho_1 = \hat{\rho}_{1_{(r)}}, \rho_2 = \hat{\rho}_{2_{(r)}}} \tag{2.18}$$

where,

$$f(\rho_1, \rho_2) = \frac{S_{t,t+1}}{S_{tt}} - \sum_{t=1}^{T-1} \frac{1}{T-1} \left[\rho_1 n_{t+1} \sqrt{\frac{\sigma_{t,t}}{\sigma_{t+1,t+1}}} + \rho_2 \frac{\sigma_{d_t,d_t}}{\sqrt{\sigma_{t,t}\sigma_{t+1,t+1}}} \right]$$

$$\frac{\partial g(\rho_1, \rho_2)}{\partial \rho_2} = - \sum_{t=1}^{T-1} \frac{1}{T-1} \frac{\sigma_{d_t,d_t}}{\sqrt{\sigma_{t,t}\sigma_{t+1,t+1}}}.$$

Here, $\hat{\rho}_{1_{(r)}}$ and $\hat{\rho}_{2_{(r)}}$ are the estimated values of ρ_1 and ρ_2 at the rth iteration respectively.

2.3.3 Forecasting

This section focuses on the approach to forecasting, where l step ahead forecast of y_t is denoted by $\hat{y}_{t+l}$. Here, t is the forecast origin and l is the lead time. After estimating the model parameters, the l step ahead forecast for y_t can be derived as: $\hat{y}_{t+l} = y_t(l) = E(y_{t+l}|y_{t+l-1})$ (Brockwell & Davis, 2002; Freeland & McCabe, 2004; Sutradhar, 2008). The 1 step ahead forecast of y_t can be obtained as

$$y_t(1) = E(y_{t+1}|y_t) = E_{d_t}\left(E\left(y_{t+1}|y_t, d_t\right)\right).$$

Now using equation (A.10) from Appendix A, it can written as

$$\begin{aligned} E\left(y_{t+1}|y_t, d_t\right) &= \rho_1 n_{t+1} y_t + \rho_2 d_t + E(d_{t+1}) \\ E_{d_t}\left(E\left(y_{t+1}|Y_t, d_t\right)\right) &= \rho_1 n_{t+1} y_t + \rho_2 E(d_t) + E(d_{t+1}) \\ &= \rho_1 n_{t+1} y_t + \rho_2 \mu_{d_t} + \mu_{t+1} - \rho_1 n_{t+1} \mu_t - \rho_2 \mu_{d_t} \\ &= \mu_{t+1} + \rho_1 n_{t+1}(y_t - \mu_t). \end{aligned}$$

Therefore, $y_t(1) = E(y_{t+1}|y_t) = \mu_{t+1} + \rho_1 n_{t+1}(y_t - \mu_t)$, where $y_t = y_t(0)$. The forecast error can be obtained as

$$\begin{aligned} e_t(1) &= y_{t+1} - y_t(1) \\ &= y_{t+1} - \mu_{t+1} - \rho_1 n_{t+1}(y_t - \mu_t). \end{aligned}$$

Now, $E(e_t(1)) = E(y_{t+1}) - E(y_t(1)) = \mu_{t+1} - \mu_{t+1} - \rho_1 n_{t+1}(\mu_t - \mu_t) = 0$. This implies that the 1 step ahead forecast is unbiased. To calculate the variance of the forecast error, the following conditional variance is used.

$$\begin{aligned} v(e_t(1)) &= E_{d_t}\left(v(e_t(1)|d_t)\right) + v_{d_t}\left(E(e_t(1)|d_t)\right) \\ \text{where, } E(e_t(1)|y_t, d_t) &= E(y_{t+1}|y_t, d_t) - E(y_t(1)|y_t, d_t) \\ &= \rho_1 n_{t+1} y_t + \rho_2 d_t + \mu_{d_{t+1}} - \mu_{t+1} - \rho_1 n_{t+1}(y_t - \mu_t) \\ &= \rho_2 d_t + \mu_{d_{t+1}} - \mu_{t+1} + \rho_1 n_{t+1} \mu_t. \\ \text{Consequently, } E(e_t(1)|d_t) &= E_{y_t}(E(e_t(1)|y_t, d_t)) \\ &= E_{y_t}(\rho_2 d_t + \mu_{d_{t+1}} - \mu_{t+1} + \rho_1 n_{t+1} \mu_t) \\ &= \rho_2 d_t + \mu_{d_{t+1}} - \mu_{t+1} + \rho_1 n_{t+1} \mu_t \end{aligned}$$

$$\begin{aligned}
\text{Again, } v(e_t(1)|d_t) &= E_{y_t}\left(v(e_t(1)|y_t, d_t)\right) + v_{y_t}\left(E(e_t(1)|y_t, d_t)\right) \\
&= E_{y_t}\left(v(y_{t+1} - y_t(1)|y_t, d_t)\right) + v_{y_t}\left(E(e_t(1)|y_t, d_t)\right) \\
&= E_{y_t}\left(v(y_{t+1}|y_t, d_t)\right) + v_{y_t}\left(E(e_t(1)|y_t, d_t)\right) \\
&= E_{y_t}\left(\rho_1(1-\rho_1)n_{t+1}y_t + \rho_2(1-\rho_2)d_t + \mu_{t+1} - \rho_1 n_{t+1}\mu_t - \rho_2\mu_{d_t}\right) \\
&\quad + v_{y_t}(\rho_2 d_t + \mu_{d_{t+1}} - \mu_{t+1} + \rho_1 n_{t+1}\mu_t) \\
&\quad \text{[Using A.14]} \\
&= \mu_{t+1} - \rho_1^2 n_{t+1}\mu_t + \rho_2(1-\rho_2)d_t - \rho_2\mu_{d_t}. \\
\text{Hence, } v(e_t(1)) &= E_{d_t}\left(v(e_t(1)|d_t)\right) + v_{d_t}\left(E(e_t(1)|d_t)\right) \\
&= E_{d_t}\left(\mu_{t+1} - \rho_1^2 n_{t+1}\mu_t + \rho_2(1-\rho_2)d_t - \rho_2\mu_{d_t}\right) \\
&\quad + v_{d_t}(\rho_2 d_t + \mu_{d_{t+1}} - \mu_{t+1} + \rho_1 n_{t+1}\mu_t) \\
&= \mu_{t+1} - \rho_1^2 n_{t+1}\mu_t - \rho_2^2\mu_{d_t} + \rho_2^2\mu_{d_t} \\
&= \mu_{t+1} - \rho_1^2 n_{t+1}\mu_t.
\end{aligned}$$

2.3.4 Special Case (When $n_t = n$)

If the total number of offspring at each time point does not depend on time, $n_t = n$, then the ARMA(1,1) model with binomial offspring in (2.14) can be written as,

$$y_t = \sum_{j=1}^{y_{t-1}} b_{1j}(n, \rho_1) + \sum_{j=1}^{d_{t-1}} b_{2j}(\rho_2) + d_t. \tag{2.19}$$

In this case the assumptions about ARMA(1,1) model with binomial offspring would be,

Assumption 3.1. $y_0 = 0$.

Assumption 3.2. $y_1 \sim \text{Poi}(\mu_1 = exp(\boldsymbol{x}_1{}^T\beta))$ and $d_1 \sim \text{Poi}(\mu_1 = exp(\boldsymbol{x}_1{}^T\beta))$.

Assumption 3.3. $d_t \sim \text{Poi}(\mu_t - \rho_1 n\mu_{t-1} - \rho_2\mu_{d_{t-1}})$, for all $t = 2, 3, ..., T$.

Assumption 3.4. d_t and y_{t-1} are independent for $t = 2, 3, ..., T$.

Here, $0 < \rho_1 < 1$, $0 < \rho_2 < 1$ and, $\mu_t = exp(\boldsymbol{x}_t{}^T\beta)$, where $\boldsymbol{x}_t = (x_{t_1}, x_{t_2}, ..., x_{t_p})^T$. The mean of Poisson must be non-negative. Therefore, for $t = 2, 3, \ldots, T$,

$$\rho_1 < min\left\{\frac{\mu_t}{n\mu_{t-1}} - \rho_2\frac{\mu_{d_{t-1}}}{\mu_{t-1}}, 1\right\}, \quad \text{for fixed} \quad \rho_2. \tag{2.20}$$

In particular, for stationary case, when $\boldsymbol{x}_t = \boldsymbol{x}, \mu_t = \mu$,

$$\rho_1 < min\left\{\tfrac{1}{n} - \rho_2\tfrac{\mu_{d_{t-1}}}{\mu}, 1\right\}, \text{ for fixed } \rho_2.$$

2.3.4.1 Basic properties

Basic properties of ARMA(1,1) model for count data with binomial offspring (when $n_t = n$) are obtained by using conditional expectation and conditional variance formulas. The following Theorem 2.3.2 states the basic properties of the model in (2.19). Its proof is given in the Appendix A.4.

Theorem 2.3.2 *Let y_t be the number of individuals at time t, and d_t be the immigration variable at time t for $t = 1, 2, ..., T$. If a binomial offspring variable for the binomial thinning operation in model (2.19) is considered and model (2.19) satisfies Assumptions 3.1-3.4,*

$$i. E(y_t) = \mu_t = exp(\boldsymbol{x}_t{}^T\boldsymbol{\beta}), \quad \textit{for all} \quad t = 1, 2, 3, ..., T.$$

$$ii. v(y_t) = \begin{cases} \mu_1, & \textit{for } t = 1, \\ \mu_t + n(n-1) + \sum_{l=1}^{t-1} \rho_1^{2(l)} n^{2(l-1}\mu_{t-l}, & \textit{for } t = 2, 3, .., T. \end{cases}$$

$$iii. \;\; cov(y_t, y_{t-k}) = \begin{cases} n\rho_1\sigma_{t-1,t-1} + \rho_2\sigma_{d_{t-1},d_{t-1}}, & \textit{for } k = 1, \\ (n\rho_1)^k\sigma_{t-k,t-k}, & \textit{for } k > 1. \end{cases}$$

$$iv. \;\; corr(y_t, y_{t-k}) = \begin{cases} n\rho_1\sqrt{\dfrac{\sigma_{t-1,t-1}}{\sigma_{t,t}}} + \rho_2\dfrac{\sigma_{d_{t-1},d_{t-1}}}{\sqrt{\sigma_{t,t}\sigma_{t-1,t-1}}}, & \textit{for } k = 1, \\ (n\rho_1)^k\sqrt{\dfrac{\sigma_{t-k,t-k}}{\sigma_{t,t}}}, & \textit{for } k > 1. \end{cases}$$

2.3.4.2 Estimation of Parameter $\boldsymbol{\beta}$, ρ_1 and ρ_2

A new simplified covariance structure is obtained for this special case. The symmetric covariance matrix is given by,

$$\Sigma = \begin{bmatrix} \mu_1 & n\rho_1\sigma_{11} + \rho_2\sigma_{d_1,d_1} & (n\rho_1)^2\sigma_{11} & \dots & (n\rho_1)^{T-1}\sigma_{11} \\ & \mu_2 - n\rho_1^2\mu_1 + n^2\rho_1^2\sigma_{11} & n\rho_1\sigma_{22} + \rho_2\sigma_{d_2,d_2} & \dots & (n\rho_1)^{T-2}\sigma_{22} \\ & & & & \vdots \\ & & & & \mu_T - n\rho_1^2\mu_{T-1} + n^2\rho_1^2\sigma_{T-1,T-1} \end{bmatrix}$$

Similar to the GQL estimation method discussed in Section 2.3.3.1, the estimation of the regression parameter $\boldsymbol{\beta}$ can be obtained by using Newton-Raphson iterative approach for known ρ_1 and ρ_2. The GQL estimating equation can be solved as (Wedderburn, 1974),

$$\hat{\boldsymbol{\beta}}_{(r+1)} = \hat{\boldsymbol{\beta}}_{(r)} + \left[\boldsymbol{X}^T\boldsymbol{U}\Sigma^{-1}\boldsymbol{U}\boldsymbol{X}\right]^{-1}\left[\boldsymbol{X}^T\boldsymbol{U}\Sigma^{-1}(\boldsymbol{y}-\boldsymbol{\mu})\right]_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}_{(r)}} \tag{2.21}$$

Here, $\hat{\boldsymbol{\beta}}_{(r)}$ is the estimated value of $\boldsymbol{\beta}$ at the rth iteration. Furthermore, by using the GMM for known $\boldsymbol{\beta}$ the estimates of ρ_1 and ρ_2 can be obtained as,

$$\hat{\rho}_1 = \left(\frac{S_{t,t+2}}{S_{tt}}\left[\frac{1}{T-2}\sum_{t=1}^{T-2} n^2\sqrt{\frac{\sigma_{t,t}}{\sigma_{t+2,t+2}}}\right]^{-1}\right)^{1/2} \tag{2.22}$$

$$\hat{\rho}_{2(r+1)} = \hat{\rho}_{2(r)} - \left[\frac{\partial f(\rho_1,\rho_2)}{\partial\rho_2}\right]^{-1} f(\rho_1,\rho_2)\Bigg|_{\rho_1=\hat{\rho}_{1(r)},\rho_2=\hat{\rho}_{2(r)}} \tag{2.23}$$

where,

$$f(\rho_1,\rho_2) = \frac{S_{t,t+1}}{S_{tt}} - \sum_{t=1}^{T-1}\frac{1}{T-1}\left[\rho_1 n\sqrt{\frac{\sigma_{t,t}}{\sigma_{t+1,t+1}}} + \rho_2\frac{\sigma_{d_t,d_t}}{\sqrt{\sigma_{t,t}\sigma_{t+1,t+1}}}\right]$$

$$\frac{\partial g(\rho_1,\rho_2)}{\partial\rho_2} = -\sum_{t=1}^{T-1}\frac{1}{T-1}\frac{\sigma_{d_t,d_t}}{\sqrt{\sigma_{t,t}\sigma_{t+1,t+1}}}$$

Here, $\hat{\rho}_{1(r)}$ and $\hat{\rho}_{2(r)}$ are the estimated value of ρ_1 and ρ_2 at rth iteration respectively.

2.3.4.3 Forecasting

In this section forecasting function is derived for the ARMA(1,1) model for count data with binomial offspring when the number of offspring at a given time does not depend on time. Let us denote the l step ahead forecast of y_t as $\hat{y}_{t+l}$. Here, t is the forecast origin and l is the lead time. After obtaining the estimated value of the model parameters, the l step ahead forecast for y_t can be derived as: $\hat{y}_{t+l} = y_t(l) = E(y_{t+l}|y_{t+l-1})$ (Brockwell & Davis, 2002; Freeland & McCabe, 2004; Sutradhar, 2008). Accordingly, the 1 step ahead forecast of y_t can be obtained as

$$y_t(1) = E(y_{t+1}|y_t) = E_{d_t}\left(E\left(y_{t+1}|Y_t, d_t\right)\right).$$

Now using equation (A.10) from Appendix A, it can be written as

$$\begin{aligned} E\left(y_{t+1}|y_t, d_t\right) &= \rho_1 n_{t+1} y_t + \rho_2 d_t + E(d_{t+1}) \\ E_{d_t}\left(E\left(y_{t+1}|Y_t, d_t\right)\right) &= \rho_1 n y_t + \rho_2 E(d_t) + E(d_{t+1}) \\ &= \rho_1 n y_t + \rho_2 \mu_{d_t} + \mu_{t+1} - \rho_1 n_{t+1} \mu_t - \rho_2 \mu_{d_t} \\ &= \mu_{t+1} + \rho_1 n(y_t - \mu_t). \end{aligned}$$

Therefore, $y_t(1) = E(y_{t+1}|y_t) = \mu_{t+1} + \rho_1 n(y_t - \mu_t)$, where $y_t = y_t(0)$. The forecast error can be obtained as

$$\begin{aligned} e_t(1) &= y_{t+1} - y_t(1) \\ &= y_{t+1} - \mu_{t+1} - \rho_1 n(y_t - \mu_t). \end{aligned}$$

Now, $E(e_t(1)) = E(y_{t+1}) - E(y_t(1)) = \mu_{t+1} - \mu_{t+1} - \rho_1 n(\mu_t - \mu_t) = 0$. Consequently, it can be said that the 1 step ahead forecast is unbiased. To calculate the variance of the forecast error, the following conditional variance is used.

$$\begin{aligned} v(e_t(1)) &= E_{d_t}\left(v(e_t(1)|d_t)\right) + v_{d_t}\left(E(e_t(1)|d_t)\right), \\ \text{where, } E(e_t(1)|y_t, d_t) &= E(y_{t+1}|y_t, d_t) - E(y_t(1)|y_t, d_t) \\ &= \rho_1 n_{t+1} y_t + \rho_2 d_t + \mu_{d_{t+1}} - \mu_{t+1} - \rho_1 n(y_t - \mu_t) \\ &= \rho_2 d_t + \mu_{d_{t+1}} - \mu_{t+1} + \rho_1 n \mu_t. \end{aligned}$$

$$
\begin{aligned}
\text{Consequently, } E(e_t(1)|d_t) &= E_{y_t}(E(e_t(1)|y_t, d_t)) \\
&= E_{y_t}(\rho_2 d_t + \mu_{d_{t+1}} - \mu_{t+1} + \rho_1 \mu_t) \\
&= \rho_2 d_t + \mu_{d_{t+1}} - \mu_{t+1} + \rho_1 n \mu_t.
\end{aligned}
$$

$$
\begin{aligned}
\text{Again, } v(e_t(1)|d_t) &= E_{y_t}\left(v(e_t(1)|y_t, d_t)\right) + v_{y_t}\left(E(e_t(1)|y_t, d_t)\right) \\
&= E_{y_t}\left(v(y_{t+1} - y_t(1)|y_t, d_t)\right) + v_{y_t}\left(E(e_t(1)|y_t, d_t)\right) \\
&= E_{y_t}\left(v(y_{t+1}|y_t, d_t)\right) + v_{y_t}\left(E(e_t(1)|y_t, d_t)\right) \\
&= E_{y_t}\left(\rho_1(1-\rho_1) n y_t + \rho_2(1-\rho_2) d_t + \mu_{t+1} - \rho_1 n \mu_t - \rho_2 \mu_{d_t}\right) \\
&\quad + v_{y_t}(\rho_2 d_t + \mu_{d_{t+1}} - \mu_{t+1} + \rho_1 n \mu_t) \\
&\quad [\text{Using A.22}] \\
&= \mu_{t+1} - \rho_1^2 n_{t+1} \mu_t + \rho_2(1-\rho_2) d_t - \rho_2 \mu_{d_t}. \\
\text{Hence, } v(e_t(1)) &= E_{d_t}\left(v(e_t(1)|d_t)\right) + v_{d_t}\left(E(e_t(1)|d_t)\right) \\
&= E_{d_t}\left(\mu_{t+1} - \rho_1^2 n \mu_t + \rho_2(1-\rho_2) d_t - \rho_2 \mu_{d_t}\right) \\
&\quad + v_{d_t}(\rho_2 d_t + \mu_{d_{t+1}} - \mu_{t+1} + \rho_1 n \mu_t) \\
&= \mu_{t+1} - \rho_1^2 n_{t+1} \mu_t - \rho_2^2 \mu_{d_t} + \rho_2^2 \mu_{d_t} \\
&= \mu_{t+1} - \rho_1^2 n \mu_t.
\end{aligned}
$$

2.3.5 Simulation Study

A simulation study is carried out to evaluate the performance of the estimation methods for the parameters $\boldsymbol{\beta}$, ρ_1 and ρ_2 of the ARMA(1,1) model with binomial offspring. R codes were developed to carry out the simulation study. Data was generated for different combinations of $\boldsymbol{\beta}$, ρ_1 and ρ_2, while time points $T = 250$, $T = 500$ and $T = 700$ were used to examine the effect of time lengths on the estimates. We consider, $n = 3$, allowing the possibility of an individual producing at most 3 offspring at a given time point. Two covariates x_{t1} and x_{t2} were considered in this study, where for $t = 1, 2, ..., T$,

$$x_{t1} \sim Binomial(1, 0.5) \quad \text{and,} \quad x_{t2} = \frac{t}{T}.$$

Due to the restrictions on ρ_1 and ρ_2, only a narrow range of values can be chosen for these parameters. After some trial and error, two values for ρ_2 were fixed, and the corresponding ρ_1 values were calculated satisfying the condition in (2.20). For each combination of $\boldsymbol{\beta}$, ρ_1 and ρ_2, we generated y_1 and d_1 from $Poi(\mu_1)$, and d_t from $Poi(\mu_t - n\rho_1\mu_{t-1} - \rho_2\mu_{d_{t-1}})$ for $t = 2, 3, ..., T$. Finally, using y_1 and d_t's, y_t's were generated from the proposed model (2.19) for $t = 2, 3, ..., T$. Once generated, data was fixed for all simulations. Initial values of $\boldsymbol{\beta} = (0, 0)$, $\rho_1 = 0$ and $\rho_2 = 0$ were used to estimate the parameters using equations (2.21), (2.22) and (2.23). Iteration for equations (2.21) and (2.23) were continued until convergence. This procedure was repeated 1000 times for fixed values of $\boldsymbol{\beta}$, ρ_1 and ρ_2. The average estimates, $\hat{\boldsymbol{\beta}}$, $\hat{\rho_1}$ and $\hat{\rho_2}$, from the 1000 simulations are presented in Table 2.2. Additionally, the standard

	True Values			Parameter Estimates					
T	$\boldsymbol{\beta}$	ρ_1	ρ_2	$\hat{\boldsymbol{\beta}}$	$SE(\hat{\boldsymbol{\beta}})$	$\hat{\rho}_1$	$SE(\hat{\rho}_1)$	$\hat{\rho}_2$	$SE(\hat{\rho}_2)$
250	(0.2,0.3)	0.18	0.15	(0.136,0.371)	(0.002,0.005)	0.211	0.0005	0.105	0.002
500				(0.198,0.346)	(0.002,0.004)	0.189	0.0005	0.128	0.002
700				(0.184,0.267)	(0.001,0.003)	0.195	0.0004	0.125	0.001
250		0.20	0.12	(0.186,0.075)	(0.002,0.006)	0.184	0.0007	0.092	0.002
500				(0.169,0.415)	(0.001,0.004)	0.207	0.0005	0.092	0.002
700				(0.201,0.316)	(0.001,0.003)	0.200	0.0004	0.083	0.002
250	(0.25, 0.35)	0.18	0.15	(0.248,0.416)	(0.002,0.005)	0.187	0.0007	0.136	0.003
500				(0.281,0.287)	(0.001,0.003)	0.192	0.0005	0.139	0.002
700				(0.222,0.354)	(0.001,0.003)	0.199	0.0005	0.156	0.002
250		0.20	0.12	(0.257,0.178)	(0.002,0.006)	0.196	0.0007	0.096	0.002
500				(0.286,0.298)	(0.002,0.004)	0.203	0.0005	0.144	0.003
700				(0.222,0.351)	(0.001,0.003)	0.214	0.0004	0.123	0.002
250	(0.3, 0.5)	0.18	0.15	(0.301,0.481)	(0.002,0.005)	0.177	0.0008	0.258	0.004
500				(0.269,0.596)	(0.001,0.003)	0.201	0.0005	0.185	0.003
700				(0.276,0.484)	(0.001,0.003)	0.182	0.0005	0.234	0.002
250		0.20	0.12	(0.259,0.526)	(0.004,0.007)	0.211	0.0007	0.180	0.003
500				(0.292,0.555)	(0.001,0.004)	0.212	0.0005	0.160	0.002
700				(0.286,0.601)	(0.001,0.004)	0.212	0.0004	0.181	0.002

Table 2.2: Comparison of True and Estimated Parameter Values of ARMA(1,1) model with binomial offspring for different combination of $\boldsymbol{\beta}$, ρ_1, ρ_2 and T values

error of each estimated parameter was calculated from the 1000 simulations to show the dispersion around the mean.

Table 2.2 shows that the GQL method perfoms well to estimate $\boldsymbol{\beta}$, while the GMM method effectively estimates ρ_1 and ρ_2. As T increases, the estimates approach the true values, and the standard error decreases. For example, when T=700, with $\boldsymbol{\beta} = (0.2, 0.3)$, $\rho_1 = 0.18$ and $\rho_2 = 0.15$, the estimates are, $\hat{\boldsymbol{\beta}} = (0.184, 0.267)$, $\hat{\rho_1} = 0.195$ and $\hat{\rho_2} = 0.125$, whereas for T=250, the estimates are $\hat{\boldsymbol{\beta}} = (0.136, 0.371)$,

$\hat{\rho}_1 = 0.211$ and $\hat{\rho}_2 = 0.105$. Moreover, it can be observed that when $T = 250$, the estimated values are not as close to the true values. Additionally, it can be observed that the standard error of the estimates are close to zero. As T increases the standard errors of the estimates decrease. Therefore, it can be said that GQL and GMM methods perform well in estimating the model parameters.

2.4 Concluding Remarks

In this paper, we have developed ARMA models for count data that incorporate the impact of past immigration variables on count data. We have proposed an ARMA(1,1) model with a binary offspring and extended the model by considering a binomial offspring. We derived the basic properties of these models. We found that the lag k autocorrelation function of the model satisfied a Yule-Walker type difference equation similar to that of the model for continuous data after lag 1. Specifically, we found that the lag k ACF ($k > 1$) followed that of the AR(1) model for count data. This is a pattern commonly found in ARMA(1,1) model for continuous time series. Additionally, we have discussed the GQL approach for estimating the covariate effect parameter and the GMM approach for estimating the correlation index parameters. The results of the simulation study have showed that the GQL and GMM approaches performed well in estimating the parameters of the models. Our findings indicate that the ARMA(1,1) model with binary offspring and the ARMA(1,1) model with binomial offspring are useful for count data influenced by past immigration variables, providing flexibility for modeling different types of count data. We have also derived forecasting function for ARMA(1,1) model with binary offspring and ARMA(1,1) model with binomial offspring..

Chapter 3

Discussion of Results and Future Work

In this research, we proposed ARMA models that incorporate the effect of past immigration variables. First, we consider a simple scenario where an individual can produce either one offspring or none at a given time, resulting in the offspring variable being binary. We also consider a binary immigration variable and propose a Poisson ARMA(1,1) model with binomial offspring. The basic properties of the model are derived. We also investigate the stationary case and derive the basic properties for this situation. The acf of y_t satisfies a Yule-Walker type difference equation. From Theorem 2.2.1 it can be observed that the acf of y_t at lag k is given by,

$$corr(y_t, y_{t-k}) = \rho_1,\ corr(y_{t-1}, y_{t-k}) = \rho_1^k \sqrt{\frac{\mu_{t-k}}{\mu_t}},\ \text{ for } k > 1.$$

For stationary case, when y_t do not depend on time t for $t = 1, 2, ..., T$, the acf of y_t at lag k is, $corr(y_t, y_{t-k}) = \rho_1^k,$ for $k > 1$. Hence, in both stationary and non-stationary cases, the acf in the Poisson ARMA(1,1) model with binary offspring behaves like the acf of a Poisson AR(1) model after lag 1. This is the same property evident in the ARMA(1,1) model for continuous data outlined in Section 1.1.3.

We then apply the GQL and GMM methods to estimate the model parameters. Through simulation studies, these methods are found to be performing well in estimating the parameters $\boldsymbol{\beta}$, ρ_1, and ρ_2. A forecasting function for ARMA(1,1) model with binary offspring is developed to predict future values of y_t.

Later, we extend our model to consider a binomial offspring variable, allowing the possibility that an individual can produce n_t offspring at a given time. We propose an ARMA(1,1) model with binomial offspring and derive the basic properties and estimation methods for the model parameters, considering both scenarios where the maximum number of offspring (n_t) depends on time and where it remains constant (n) over time. From Theorem 2.2.1 and 2.3.1, it can be observed that the mean of y_t for ARMA(1,1) model with binomial offspring is same as the mean of y_t for ARMA(1,1) model with binary offspring. The variance of y_t for ARMA(1,1) model with binomial offspring is,

$$v(y_t) = \begin{cases} \mu_1, & \text{for } t = 1, \\ \mu_2 + \rho_1^2 n_2(n_2 - 1)\mu_1, & \text{for } t = 2, \\ \mu_t + \rho_1^2 n_t(n_t - 1)\mu_{t-1} & \\ + \sum_{l=1}^{t-2} \left[\rho_1^{2(l+1)} n_{t-l}(n_{t-l} - 1) \left(\prod_{j=0}^{l-1} n_{t-j}^2\right) \mu_{t-(l+1)}\right], & \text{for } t = 3, .., T. \end{cases}$$

when $n_t = 1$, the variance of y_t reduces to μ_t, which is the same as that of the ARMA(1,1) model with binary offspring. This is also evident for covariance and correlation between y_t and y_{t-k} for lag k, indicating that the ARMA(1,1) model with binomial offspring is a generalization of the ARMA(1,1) model with binary offspring. From Theorem 2.3.1, it can be observed that the acf of y_t for ARMA(1,1) model with binomial offspring is,

$$corr(y_t, y_{t-k}) = \left(\prod_{l=0}^{k-1} n_{t-l}\right) \rho_1^k \sqrt{\frac{\sigma_{t-k,t-k}}{\sigma_{t,t}}}, \text{ for } k > 1,$$

which is same as the acf of y_t for AR(1) model for count data, as outlined in Section 1.2.1. Therefore, the acf in the ARMA(1,1) model with binomial offspring behaves like the acf of a AR(1) model with binomial offspring after lag 1. This property is also observed when $n_t = n$. In this case, $corr(y_t, y_{t-k}) = (n\rho_1)^k \sqrt{\frac{\sigma_{t-k,t-k}}{\sigma_{t,t}}}$, for $k > 1$.

A simulation study is conducted to evaluate the performance of the GQL and GMM methods for the ARMA(1,1) model with binomial offspring when the offspring variable, $b_{1j} \sim bin(n, \rho_1)$. We observe that the GQL method performs well in estimating the covariate parameter $\boldsymbol{\beta}$, while the GMM method is effective in estimating ρ_1 and ρ_2. We derive forecasting functions for ARMA(1,1) model with binomial offspring considering both situations where the maximum number of offspring (n_t) depends on time and where it remains constant (n) over time.

In conclusion, our proposed models can be used for count data when the immigration variable from previous time points affects the number of individuals at the present time point. Additionally, our proposed models offer researchers the option to choose between models for binary offspring variables or binomial offspring variables, depending on the type of count data.

There are several opportunities for further research in this area. The next step could involve identifying an appropriate model for the available count data. Another area for exploration is the use of a mixed model approach that incorporates random effects. This would allow for the inclusion of unobservable factors that may influence the time series data, providing a more comprehensive understanding of the underlying dynamics. Future research could also extend the current model to higher-order Poisson ARMA models, enabling more complex autocorrelation structures to be captured in the count data, potentially leading to more precise and comprehensive modeling.

References

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

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

Brännäs, K., & Hall, A. (2001). Estimation in integer-valued moving average models. *Applied Stochastic Models in Business and Industry*, *17*(3), 277–291.

Brockwell, P. J., & Davis, R. A. (2002). *Introduction to time series and forecasting*. Springer.

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

Chatfield, C. (1975). *The analysis of time series: theory and practice*. Springer.

Chuang, A. (1991). *Time series analysis: univariate and multivariate methods*. Taylor & Francis.

Crowder, M. (1995). On the use of a working correlation matrix in using generalised linear models for repeated measures. *Biometrika*, *82*(2), 407–410.

Czado, C., Gneiting, T., & Held, L. (2009). Predictive model assessment for count data. *Biometrics*, *65*(4), 1254–1261.

Du Jin-Guan, L. Y. (1991). The integer-valued autoregressive (inar (p)) model. *J. Time Ser. Anal*, *12*, 129–142.

Freeland, R. K., & McCabe, B. P. (2004). Forecasting discrete valued low count time series. *International Journal of Forecasting*, *20*(3), 427–434.

Hyndman, R. (2018). *Forecasting: principles and practice*. OTexts.

Liang, K.-Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, *73*(1), 13–22.

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

Oyet, A. J., & Sutradhar, B. C. (2013). Longitudinal modeling of infectious disease. *Sankhya B*, *75*, 319–342.

Puig, P., & Valero, J. (2007). Characterization of count data distributions involving additivity and binomial subsampling. *Bernoulli*, 544–555.

Shumway, R. H., Stoffer, D. S., & Stoffer, D. S. (2000). *Time series analysis and its applications* (Vol. 3). Springer.

Steutel, F. W., & Van Harn, K. (1979). Discrete analogues of self-decomposability and stability. *The Annals of Probability*, 893–899.

Steutel, F. W., Vervaat, W., & Wolfe, S. J. (1983). Integer-valued branching processes with immigration. *Advances in applied probability*, *15*(4), 713–725.

Sutradhar, B., Oyet, A. J., & Gadag, V. G. (2010). On quasi-likelihood estimation for branching processes with immigration. *Canadian Journal of Statistics*, *38*(2), 290–313.

Sutradhar, B. C. (2003). An overview on regression models for discrete longitudinal responses. *Statistical Science*, *18*(3), 377–393.

Sutradhar, B. C. (2008). On forecasting counts. *Journal of Forecasting*, *27*(2), 109–129.

Sutradhar, B. C. (2011). *Dynamic mixed models for familial longitudinal data* (Vol. 18). Springer.

Sutradhar, B. C., & Das, K. (1999). Miscellanea. on the efficiency of regression estimators in generalised linear models for longitudinal data. *Biometrika*, *86*(2), 459–465.

Teräsvirta, T. (1994). Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the american Statistical association*, *89*(425), 208–218.

Wedderburn, R. W. (1974). Quasi-likelihood functions, generalized linear models, and the gauss—newton method. *Biometrika*, *61*(3), 439–447.

Wei, C., & Winnicki, J. (1990). Estimation of the means in the branching process with immigration. *The Annals of Statistics*, 1757–1773.

Weiß, C. H. (2008). Serial dependence and regression of poisson inarma models. *Journal of Statistical Planning and Inference*, *138*(10), 2975–2990.

Weiß, C. H. (2015). A poisson inar (1) model with serially dependent innovations. *Metrika*, *78*(7), 829–851.

Whittle, P. (1951). *Hypothesis testing in time series analysis*. Uppsala: Almquist &

Wiksell.

Winnicki, J. (1988). Estimation theory for the branching process with immigration. *Contemp. Math*, *80*, 301–322.

Zeger, S. L. (1988). A regression model for time series of counts. *Biometrika*, *75*(4), 621–629.

Zhang, C., & Oyet, A. J. (2014). Second order longitudinal dynamic models with covariates: estimation and forecasting. *Metrika*, *77*(7), 837–859.

Appendix A

Proof of Theorems

In this Appendix proof of Theorem 2.2.1, Theorem 2.2.2, Theorem 2.3.1 and Theorem 2.3.2 are discussed.

A.1 Proof of Theorem 2.2.1

Based on the Assumption 1.2 stated in section 2.2, $E(y_1) = v(y_1) = \mu_1$. Additionally, $E(d_1) = v(d_1) = \mu_1$. From Assumption 1.3, $E(d_t) = v(d_t) = \mu_t - \rho_1\mu_{t-1} - \rho_2\mu_{d_{t-1}}$ for $t = 2, 3, ..., T$.

Mean: Let, $\mu_t = exp(\boldsymbol{x}_t{}^T\beta)$ be the mean of y_t and μ_{d_t} be the mean of d_t. We know,

$$E\left(y_t\right) = E_{d_{t-1},y_{t-1}}\left(E\left(y_t|y_{t-1}, d_{t-1}\right)\right) \tag{A.1}$$

By taking conditional expectation on (2.1) we get,

$$\begin{aligned}
E\left(y_t|y_{t-1}, d_{t-1}\right) &= \rho_1 y_{t-1} + \rho_2 d_{t-1} + E(d_t) && \text{(A.2)}\\
E_{y_{t-1}}\left(E\left(y_t|y_{t-1}, d_{t-1}\right)\right) &= \rho_1 E(y_{t-1}) + \rho_2 d_{t-1} + E(d_t) && \text{(A.3)}\\
E_{d_{t-1},y_{t-1}}\left(E\left(y_t|y_{t-1}, d_{t-1}\right)\right) &= \rho_1 E(y_{t-1}) + \rho_2 E(d_{t-1}) + E(d_t) && \text{(A.4)}\\
&= \rho_1 E(y_{t-1}) + \rho_2 E(d_{t-1}) + \mu_t - \rho_1\mu_{t-1} - \rho_2\mu_{d_{t-1}} && \text{(A.5)}
\end{aligned}$$

Therefore,

$$E(y_t) = \rho_1 E(y_{t-1}) + \rho_2 E(d_{t-1}) + \mu_t - \rho_1\mu_{t-1} - \rho_2\mu_{d_{t-1}}$$

For $t = 2$,

$$\begin{aligned} E(y_2) &= \rho_1 E(y_1) + \rho_2 E(d_1) + \mu_2 - \rho_1\mu_1 - \rho_2\mu_{d_1} \\ &= \rho_1\mu_1 + \rho_2\mu_1 + \mu_2 - \rho_1\mu_1 - \rho_2\mu_1 \\ &= \mu_2 \end{aligned}$$

For $t = 3$,

$$\begin{aligned} E(y_3) &= \rho_1 E(y_2) + \rho_2 E(d_2) + \mu_3 - \rho_1\mu_2 - \rho_2\mu_{d_2} \\ &= \rho_1\mu_2 + \rho_2(\mu_2 - \rho_1\mu_1 - \rho_2\mu_{d_1}) + \mu_3 - \rho_1\mu_2 - \rho_2(\mu_2 - \rho_1\mu_1 - \rho_2\mu_{d_1}) \\ &= \mu_3 \end{aligned}$$

Through the method of mathematical induction, if $E(y_{t-1}) = \mu_{t-1},\ E(d_{t-1}) = \mu_{d_{t-1}}$ then $E(y_t) = \rho_1\mu_{t-1} + \rho_2\mu_{d_{t-1}} + \mu_t - \rho_1\mu_{t-1} - \rho_2\mu_{d_{t-1}} = \mu_t$.
Therefore, $E(y_t) = \mu_t = exp(\boldsymbol{x}_t{}^T\boldsymbol{\beta})$ for all $t = 1, 2, 3, ..., T$

Variance: Let, $v(y_t) = \sigma_{t,t}$ and $v(d_t) = \sigma_{d_t,d_t}$ for all $t = 1, 2, ..., T$. From Assumption 1.4 of the model in 2.2, $cov(y_{t-1}, d_t) = 0$. Now by taking conditional variance on (2.1),

$$\begin{aligned} v(y_t|y_{t-1}, d_{t-1}) &= y_{t-1}v(b_{1j}) + d_{t-1}v(b_{2j}) + v(d_t) \\ &= \rho_1(1-\rho_1)y_{t-1} + \rho_2(1-\rho_2)d_{t-1} + \mu_t - \rho_1\mu_{t-1} - \rho_2\mu_{d_{t-1}}. \end{aligned} \tag{A.6}$$

Let, $v(y_{t-1}) = \sigma_{t-1,t-1}$ and $v(d_{t-1}) = \sigma_{d_{t-1},d_{t-1}}$. Then

$$\begin{aligned} v(y_t|d_{t-1}) &= E_{y_{t-1}}\left(v\left(y_t|y_{t-1}, d_{t-1}\right)\right) + v_{y_{t-1}}\left(E\left(y_t|y_{t-1}, d_{t-1}\right)\right) \\ &= E_{y_{t-1}}(\rho_1(1-\rho_1)y_{t-1} + \rho_2(1-\rho_2)d_{t-1} + \mu_{dt}) \\ &+ v_{y_{t-1}}(\rho_1 y_{t-1} + \rho_2 d_{t-1} + \mu_t - \rho_1\mu_{t-1} - \rho_2\mu_{d_{t-1}}) \\ &= \rho_1(1-\rho_1)\mu_{t-1} + \rho_2(1-\rho_2)d_{t-1} + \mu_{dt} + \rho_1^2\sigma_{t-1,t-1}. \end{aligned}$$

Now,

$$\begin{aligned} v\left(y_t\right) &= E_{d_{t-1}}\left(v\left(y_t|\ d_{t-1}\right)\right) + v_{d_{t-1}}\left(E\left(y_t|\ d_{t-1}\right)\right) \\ &= E_{d_{t-1}}\left(\rho_1(1-\rho_1)\mu_{t-1} + \rho_2(1-\rho_2)d_{t-1} + \mu_{d_t} + \rho_1^2\sigma_{t-1,t-1}\right) \\ &+ v_{d_{t-1}}\left(\mu_t + \rho_2 d_{t-1} - \rho_2\mu_{d_{t-1}}\right) \end{aligned}$$

$$= \rho_1(1-\rho_1)\mu_{t-1} + \rho_2(1-\rho_2)\mu_{d_{t-1}} + \mu_t - \rho_1\mu_{t-1} - \rho_2\mu_{d_{t-1}} + \rho_1^2\sigma_{t-1,t-1} + \rho_2^2 v(d_{t-1})$$
$$= \mu_t - \rho_1^2\mu_{t-1} - \rho_2^2\mu_{d_{t-1}} + \rho_1^2\sigma_{t-1,t-1} + \rho_2^2\sigma_{d_{t-1},d_{t-1}}$$
$$= \mu_t - \rho_1^2\mu_{t-1} + \rho_1^2\sigma_{t-1,t-1} \text{ [Since from Assumption 1.2, } \mu_{d_{t-1}} = \sigma_{d_{t-1},d_{t-1}}].$$

Therefore, we get the following expression for variance of y_t,

$$v(y_t) = \sigma_{t,t} = \mu_t - \rho_1^2\mu_{t-1} + \rho_1^2\sigma_{t-1,t-1}. \tag{A.7}$$

Again, $v(y_1) = \mu_1$. For $T = 2, 3, ...T$, using the equation in (A.7),
when $t = 2$,

$$\begin{aligned} v(y_2) &= \mu_2 - \rho_1^2\mu_1 + \rho_1^2\sigma_{1,1} \\ &= \mu_2 - \rho_1^2\mu_1 + \rho_1^2\mu_1 \\ &= \mu_2. \end{aligned}$$

when $t = 3$,

$$\begin{aligned} v(y_3) &= \mu_3 - \rho_1^2\mu_2 + \rho_1^2\sigma_{2,2} \\ &= \mu_3 - \rho_1^2\mu_2 + \rho_1^2 mu_2 \\ &= \mu_3. \end{aligned}$$

If we calculate for $t = 4, 5, ...$ and so on, assume that, $\sigma_{t-1,t-1} = \mu_{t-1}$ then,

$$\begin{aligned} v(y_t) &= \mu_t - \rho_1^2\mu_{t-1} + \rho_1^2\sigma_{t-1,t-1} \\ &= \mu_t - \rho_1^2\mu_{t-1} + \rho_1^2\mu_{t-1} \\ &= \mu_t. \end{aligned}$$

So, $v(y_t) = \mu_t = exp(\boldsymbol{x_t}^T\beta)$ for all $t = 1, 2, 3, ..., T$. Since mean and variance are equal, $\{y_t\}$ is a Poisson ARMA(1,1) process.

Covariance and Correlation: To derive the lag k covariance between y_t and y_{t-k}, we use the following formula as,

$$\begin{aligned} cov(y_t, y_{t-k}) &= cov\left(\sum_{j=1}^{y_{t-1}} b_{1j}(\rho_1), y_{t-k}\right) + cov\left(\sum_{j=1}^{d_{t-1}} b_{2j}(\rho_2), y_{t-k}\right) + cov\left(d_t, y_{t-k}\right) \\ &= cov\left(\sum_{j=1}^{y_{t-1}} b_{1j}(\rho_1), y_{t-k}\right) + cov\left(\sum_{j=1}^{d_{t-1}} b_{2j}(\rho_2), y_{t-k}\right) \text{ [since } cov(d_t, y_{t-k}) = 0]. \end{aligned}$$

Considering the first part of the above formula we get,

$$\begin{aligned} cov\left(\sum_{j=1}^{y_{t-1}} b_{1j}(\rho_1), y_{t-k}\right) &= E\left[cov\left(\sum_{j=1}^{y_{t-1}} b_{1j}(\rho_1), y_{t-k}|y_{t-1}, d_{t-1}, y_{t-k}\right)\right] \\ &\quad + cov\left[E\left(\sum_{j=1}^{y_{t-1}} b_{1j}(\rho_1)|y_{t-1}, d_{t-1}, y_{t-k}\right), E\left(y_{t-k}|y_{t-1}, d_{t-1}, y_{t-k}\right)\right] \\ &= cov(\rho_1 y_{t-1}, y_{t-k}) \\ &= \rho_1 cov(y_{t-1}, y_{t-k}). \end{aligned}$$

Similarly,

$$cov\left(\sum_{j=1}^{d_{t-1}} b_{2j}(\rho_2), y_{t-k}\right) = \rho_2 cov(d_{t-1}, y_{t-k}). \tag{A.8}$$

The above relationship in (A.8) holds only when $k = 1$. Otherwise, when $k > 1$, $cov\left(\sum_{j=1}^{d_{t-1}} b_{2j}(\rho_2), y_{t-k}\right) = 0$. Therefore, when $k = 1$, we get,

$$\begin{aligned} cov(y_t, y_{t-k}) &= \rho_1 cov(y_{t-1}, y_{t-k}) + \rho_2 cov(d_{t-1}, y_{t-k}) \\ &= \rho_1 cov(Y_{t-1}, Y_{t-1}) + \rho_2 cov(d_{t-1}, y_{t-1}) \\ &= \rho_1 \sigma_{t-1,t-1} + \rho_2 cov\left(d_{t-1}, \sum_{j=1}^{y_{t-2}} b_{1j}(\rho_1) + \sum_{j=1}^{d_{t-2}} b_{2j}(\rho_2) + d_{t-1}\right) \\ &= \rho_1 \sigma_{t-1,t-1} + \rho_2 \sigma_{d_{t-1},d_{t-1}} \\ &= \rho_1 \mu_{t-1} + \rho_2(\mu_{t-1} - \rho_1 \mu_{t-2} - \rho_2 \mu_{d_{t-2}}) \\ &= (\rho_1 + \rho_2)\mu_{t-1} + \rho_2(\rho_1 \mu_{t-2} + \rho_2 \mu_{d_{t-2}}). \end{aligned}$$

When $k > 1$,

$$\begin{aligned} cov(y_t, y_{t-k}) &= \rho_1 cov(y_{t-1}, y_{t-k}) \\ &= \rho_1^2 cov(y_{t-2}, y_{t-k}) \\ &= \rho_1^3 cov(y_{t-3}, y_{t-k}) \text{ leading to} \\ cov(y_t, y_{t-k}) &= \rho_1^k cov(y_{t-k}, y_{t-k}) = \rho_1^k \mu_{t-k}. \end{aligned}$$

In general, for $t = 1, 2, ..., T$ we get the following expression for covariance between y_t and y_{t-k} for lag $k = 1, 2, .., T-1$

$$\text{cov}(y_t, y_{t-k}) = \begin{cases} (\rho_1 + \rho_2)\mu_{t-1} - \rho_2(\rho_1 \mu_{t-2} + \rho_2 \mu_{dt-2}), & \text{for } k = 1, \\ \rho_1^k \mu_{t-k}, & \text{for } k > 1. \end{cases}$$

By using the following formula for correlation,

$$corr(y_t, y_{t-k}) = \frac{cov(y_t, y_{t-k})}{\sqrt{v(y_t)}\sqrt{v(y_{t-k})}},$$

and plugging in the expressions for respective covariance and variances, we get correlation between y_t and y_{t-k} given by,

$$\text{corr}(y_t, y_{t-k}) = \begin{cases} (\rho_1 + \rho_2)\sqrt{\dfrac{\mu_{t-1}}{\mu_t}} - \rho_1 \rho_2 \dfrac{\mu_{t-2}}{\sqrt{\mu_t \mu_{t-1}}} - \rho_2^2 \dfrac{\mu_{d_{t-2}}}{\sqrt{\mu_t \mu_{t-1}}}, & \text{for } k = 1, \\ \rho_1^k \sqrt{\frac{\mu_{t-k}}{\mu_t}}, & \text{for } k > 1. \end{cases}$$

A.2 Proof of Theorem 2.2.2

From the GQL estimation method discussed in section 2.2.2.1 we get the following estimating equation,

$$\boldsymbol{X}^T \boldsymbol{U} \Sigma^{-1} (\boldsymbol{y} - \boldsymbol{\mu}) = 0$$

$$\text{or, } \boldsymbol{X}^T \boldsymbol{U} (A^{\frac{1}{2}} C(\rho) A^{\frac{1}{2}})^{-1} (\boldsymbol{y} - \boldsymbol{\mu}) = 0,$$

$$\text{or, } \boldsymbol{X}^T \boldsymbol{U} A^{-\frac{1}{2}} (C(\rho))^{-1} A^{-\frac{1}{2}} (\boldsymbol{y} - \boldsymbol{\mu}) = 0,$$

where $\boldsymbol{A}^{\frac{1}{2}} = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_T)^{\frac{1}{2}}$ and $\boldsymbol{C}(\rho)$ is defined in (1.8).Also, let us consider, $(C(\rho))^{-1} = Q$, then

$$(C(\rho))^{-1} = \begin{bmatrix} q_{11} & q_{12} & \ldots & q_{1T} \\ q_{21} & q_{22} & \ldots & q_{2T} \\ \vdots & \vdots & & \vdots \\ q_{T1} & q_{T2} & \ldots & q_{TT} \end{bmatrix} \text{ and } \boldsymbol{U}A^{-\frac{1}{2}} = \begin{bmatrix} \frac{\mu_1}{\sigma_1{}^{\frac{1}{2}}} & 0 & \ldots & 0 \\ 0 & \frac{\mu_2}{\sigma_2{}^{\frac{1}{2}}} & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & \frac{\mu_T}{\sigma_T{}^{\frac{1}{2}}} \end{bmatrix}$$

Combining these two we get,

$$\begin{aligned} \boldsymbol{U}A^{-\frac{1}{2}}(C(\rho))^{-1} &= \begin{bmatrix} \frac{\mu_1}{\sigma_1{}^{\frac{1}{2}}} & 0 & \ldots & 0 \\ 0 & \frac{\mu_2}{\sigma_2{}^{\frac{1}{2}}} & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & \frac{\mu_T}{\sigma_T{}^{\frac{1}{2}}} \end{bmatrix} \begin{bmatrix} q_{11} & q_{12} & \ldots & q_{1T} \\ q_{21} & q_{22} & \ldots & q_{2T} \\ \vdots & \vdots & & \vdots \\ q_{T1} & q_{T2} & \ldots & q_{TT} \end{bmatrix} \\ &= \begin{bmatrix} \frac{\mu_1 q_{11}}{\sigma_1{}^{\frac{1}{2}}} & \frac{\mu_1 q_{12}}{\sigma_1{}^{\frac{1}{2}}} & \ldots & \frac{\mu_1 q_{1T}}{\sigma_1{}^{\frac{1}{2}}} \\ \frac{\mu_2 q_{21}}{\sigma_2{}^{\frac{1}{2}}} & \frac{\mu_2 q_{22}}{\sigma_2{}^{\frac{1}{2}}} & \ldots & \frac{\mu_2 q_{2T}}{\sigma_2{}^{\frac{1}{2}}} \\ \vdots & \vdots & & \vdots \\ \frac{\mu_T q_{T1}}{\sigma_T{}^{\frac{1}{2}}} & \frac{\mu_T q_{T2}}{\sigma_T{}^{\frac{1}{2}}} & \ldots & \frac{\mu_T q_{TT}}{\sigma_T{}^{\frac{1}{2}}} \end{bmatrix} \end{aligned}$$

This leads to,

$$\boldsymbol{U}A^{-\frac{1}{2}}(C(\rho))^{-1}A^{-\frac{1}{2}}(\boldsymbol{y}-\boldsymbol{\mu}) = \begin{bmatrix} \frac{\mu_1 q_{11}}{\sigma_1{}^{\frac{1}{2}}} & \frac{\mu_1 q_{12}}{\sigma_1{}^{\frac{1}{2}}} & \dots & \frac{\mu_1 q_{1T}}{\sigma_1{}^{\frac{1}{2}}} \\ \frac{\mu_2 q_{21}}{\sigma_2{}^{\frac{1}{2}}} & \frac{\mu_2 q_{22}}{\sigma_2{}^{\frac{1}{2}}} & \dots & \frac{\mu_2 q_{2T}}{\sigma_2{}^{\frac{1}{2}}} \\ \vdots & \vdots & & \vdots \\ \frac{\mu_T q_{T1}}{\sigma_T{}^{\frac{1}{2}}} & \frac{\mu_T q_{T2}}{\sigma_T{}^{\frac{1}{2}}} & \dots & \frac{\mu_T q_{TT}}{\sigma_T{}^{\frac{1}{2}}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma_1{}^{\frac{1}{2}}} & 0 & \dots & 0 \\ 0 & \frac{1}{\sigma_2{}^{\frac{1}{2}}} & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \frac{1}{\sigma_T{}^{\frac{1}{2}}} \end{bmatrix} \begin{bmatrix} y_1-\mu_1 \\ y_2-\mu_2 \\ \vdots \\ y_T-\mu_T \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\mu_1 q_{11}(y_1-\mu_1)}{\sigma_1{}^{\frac{1}{2}}\sigma_1{}^{\frac{1}{2}}} & \frac{\mu_1 q_{12}(y_2-\mu_2)}{\sigma_1{}^{\frac{1}{2}}\sigma_2{}^{\frac{1}{2}}} & \dots & \frac{\mu_1 q_{1T}(y_T-\mu_T)}{\sigma_1{}^{\frac{1}{2}}\sigma_T{}^{\frac{1}{2}}} \\ \frac{\mu_2 q_{21}(y_1-\mu_1)}{\sigma_2{}^{\frac{1}{2}}\sigma_1{}^{\frac{1}{2}}} & \frac{\mu_2 q_{22}(y_2-\mu_2)}{\sigma_2{}^{\frac{1}{2}}\sigma_2{}^{\frac{1}{2}}} & \dots & \frac{\mu_2 q_{2T}(y_T-\mu_T)}{\sigma_2{}^{\frac{1}{2}}\sigma_T{}^{\frac{1}{2}}} \\ \vdots & \vdots & & \vdots \\ \frac{\mu_T q_{T1}(y_1-\mu_1)}{\sigma_T{}^{\frac{1}{2}}\sigma_1{}^{\frac{1}{2}}} & \frac{\mu_T q_{T2}(y_2-\mu_2)}{\sigma_T{}^{\frac{1}{2}}\sigma_2{}^{\frac{1}{2}}} & \dots & \frac{\mu_T q_{TT}(y_T-\mu_T)}{\sigma_T{}^{\frac{1}{2}}\sigma_T{}^{\frac{1}{2}}} \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{i=1}^{T} \frac{\mu_1 q_{1i}(y_i-\mu_i)}{\sigma_1{}^{\frac{1}{2}}\sigma_i{}^{\frac{1}{2}}} \\ \vdots \\ \sum_{i=1}^{T} \frac{\mu_T q_{Ti}(y_i-\mu_i)}{\sigma_T{}^{\frac{1}{2}}\sigma_i{}^{\frac{1}{2}}} \end{bmatrix}$$

Finally we can write,

$$\boldsymbol{X}^T\boldsymbol{U}A^{-\frac{1}{2}}(C(\rho))^{-1}A^{-\frac{1}{2}}(\boldsymbol{y}-\boldsymbol{\mu}) = \begin{bmatrix} X_1 & X_2 & \dots & X_T \end{bmatrix} \begin{bmatrix} \sum_{i=1}^{T} \frac{\mu_1 q_{1i}(y_i-\mu_i)}{\sigma_1{}^{\frac{1}{2}}\sigma_i{}^{\frac{1}{2}}} \\ \vdots \\ \sum_{i=1}^{T} \frac{\mu_T q_{Ti}(y_i-\mu_i)}{\sigma_T{}^{\frac{1}{2}}\sigma_i{}^{\frac{1}{2}}} \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{i=1}^{T} \frac{X_1\mu_1 q_{1i}(y_i-\mu_i)}{\sigma_1{}^{\frac{1}{2}}\sigma_i{}^{\frac{1}{2}}} & \dots & \sum_{i=1}^{T} \frac{X_T\mu_T q_{Ti}(y_i-\mu_i)}{\sigma_T{}^{\frac{1}{2}}\sigma_i{}^{\frac{1}{2}}} \end{bmatrix}$$

$$= \sum_{j=1}^{T}\sum_{i=1}^{T} \frac{X_j\mu_j q_{ji}(y_i-\mu_i)}{\sigma_j{}^{\frac{1}{2}}\sigma_i{}^{\frac{1}{2}}}$$

Here, X_t is a vector of p covariates at time point t for $t = 1, 2, ..., T$. Now according to the work conducted by Zeger (1988), since we can rewrite the GQL estimating function of $\boldsymbol{\beta}$ as summations over T, as $T \to \infty$, then $\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, R^*)$. The covariance matrix, R^* can be defined as

$$R^* = \left[\frac{\partial \boldsymbol{\mu}^T}{\partial \boldsymbol{\beta}} \Sigma^{-1} \frac{\partial \boldsymbol{\mu}^T}{\partial \boldsymbol{\beta}}\right] = \left[\boldsymbol{X}^T \boldsymbol{U} \Sigma^{-1} \boldsymbol{U} \boldsymbol{X}\right]^{-1}.$$

A.3 Proof of Theorem 2.3.1

Based on the Assumption 2.2 outlined in section 2.3, $E(y_1) = v(y_1) = \mu_1$. Additionally, $E(d_1) = v(d_1) = \mu_1$. From Assumption 2.3, $E(d_t) = v(d_t) = \mu_t - n_t\rho_1\mu_{t-1} - \rho_2\mu_{d_{t-1}}$ for $t = 2, 3, ..., T$.

Mean: Let, $\mu_t = exp(\boldsymbol{x}_t{}^T\beta)$ be the mean of y_t and μ_{dt} be the mean of d_t. We know,

$$E\left(y_t\right) = E_{d_{t-1},y_{t-1}}\left(E\left(y_t|y_{t-1}, d_{t-1}\right)\right) \tag{A.9}$$

By taking conditional expectation on 2.14 we get,

$$E\left(y_t|Y_{t-1}, d_{t-1}\right) = \rho_1 n_t y_{t-1} + \rho_2 d_{t-1} + E(d_t) \tag{A.10}$$

$$E_{y_{t-1}}\left(E\left(y_t|y_{t-1}, d_{t-1}\right)\right) = \rho_1 n_t E(y_{t-1}) + \rho_2 d_{t-1} + E(d_t) \tag{A.11}$$

$$E_{d_{t-1},y_{t-1}}\left(E\left(y_t|y_{t-1}, d_{t-1}\right)\right) = \rho_1 n_t E(y_{t-1}) + \rho_2 E(d_{t-1}) + E(d_t) \tag{A.12}$$

$$= \rho_1 n_t E(y_{t-1}) + \rho_2 E(d_{t-1}) + \mu_t - \rho_1 n_t \mu_{t-1} - \rho_2 \mu_{d_{t-1}} \tag{A.13}$$

Therefore,

$$E(y_t) = \rho_1 n_t E(y_{t-1}) + \rho_2 E(d_{t-1}) + \mu_t - \rho_1 n_t \mu_{t-1} - \rho_2 \mu_{d_{t-1}}$$

For $t = 2$,

$$\begin{aligned} E(y_2) &= \rho_1 n_2 E(y_1) + \rho_2 E(d_1) + \mu_2 - \rho_1 n_2 \mu_1 - \rho_2 \mu_{d_1} \\ &= \rho_1 n_2 \mu_1 + \rho_2 \mu_1 + \mu_2 - \rho_1 n_2 \mu_1 - \rho_2 \mu_1 \\ &= \mu_2 \end{aligned}$$

For $t = 3$,

$$\begin{aligned} E(y_3) &= \rho_1 n_3 E(y_2) + \rho_2 E(d_2) + \mu_3 - \rho_1 n_3 \mu_2 - \rho_2 \mu_{d_2} \\ &= \rho_1 n_2 \mu_2 + \rho_2(\mu_2 - \rho_1 n_2 \mu_1 - \rho_2 \mu_{d_1}) + \mu_3 - \rho_1 n_3 \mu_2 - \rho_2(\mu_2 - \rho_1 n_2 \mu_1 - \rho_2 \mu_{d_1}) \\ &= \mu_3 \end{aligned}$$

Through the method of mathematical induction, if $E(y_{t-1}) = \mu_{t-1}$, $E(d_{t-1}) = \mu_{d_{t-1}}$ then $E(y_t) = \rho_1 n_t \mu_{t-1} + \rho_2 \mu_{d_{t-1}} + \mu_t - \rho_1 n_t \mu_{t-1} - \rho_2 \mu_{d_{t-1}} = \mu_t$.
Therefore, $E(y_t) = \mu_t = exp(\boldsymbol{x}_t{}^T \beta)$ for all $t = 1, 2, 3, ..., T$

Variance: Let, $v(y_t) = \sigma_{t,t}$ and $v(d_t) = \sigma_{d_t,d_t}$ for all $t = 1, 2, ..., T$. Additionally, $v(d_t) = E(d_t) = \mu_t - \rho_1 n_t \mu_{t-1} - \rho_2 \mu_{d_{t-1}}$. From Assumption 2.4 of the model in (2.14), $cov(y_{t-1}, d_t) = 0$. Now by taking conditional variance on 2.14

$$\begin{aligned} v(y_t|y_{t-1}, d_{t-1}) &= y_{t-1} v(b_{1j}) + d_{t-1} v(b_{2j}) + v(d_t) \\ &= \rho_1(1-\rho_1) n_t y_{t-1} + \rho_2(1-\rho_2) d_{t-1} + \mu_t - \rho_1 n_t \mu_{t-1} - \rho_2 \mu_{d_{t-1}} \end{aligned} \tag{A.14}$$

Let, $v(Y_{t-1}) = \sigma_{t-1,t-1}$ and $v(d_{t-1}) = \sigma_{d_{t-1},d_{t-1}}$. Then,

$$\begin{aligned} v(y_t|d_{t-1}) &= E_{y_{t-1}}\left(v\left(y_t|y_{t-1}, d_{t-1}\right)\right) + v_{y_{t-1}}\left(E\left(y_t|y_{t-1}, d_{t-1}\right)\right) \\ &= E_{y_{t-1}}(\rho_1(1-\rho_1) n_t y_{t-1} + \rho_2(1-\rho_2) d_{t-1} + \mu_{dt}) \\ &+ v_{y_{t-1}}(\rho_1 n_t y_{t-1} + \rho_2 d_{t-1} + \mu_t - \rho_1 n_t \mu_{t-1} - \rho_2 \mu_{d_{t-1}}) \\ &\qquad \text{[using (A.10)]} \\ &= \rho_1(1-\rho_1) n_t \mu_{t-1} + \rho_2(1-\rho_2) d_{t-1} + \mu_{dt} + \rho_1^2 n_t^2 \sigma_{t-1,t-1} \end{aligned}$$

Now,

$$\begin{aligned} v\left(y_t\right) =& E_{d_{t-1}}\left(v\left(y_t|\ d_{t-1}\right)\right) + v_{d_{t-1}}\left(E\left(Y_t|\ d_{t-1}\right)\right) \\ =& E_{d_{t-1}}\left(\rho_1(1-\rho_1) n_t \mu_{t-1} + \rho_2(1-\rho_2) d_{t-1} + \mu_{dt} + \rho_1^2 n_t^2 \sigma_{t-1,t-1}\right) \\ & + v_{d_{t-1}}\left(\mu_t + \rho_2 d_{t-1} - \rho_2 \mu_{d_{t-1}}\right) \ \text{[using (A.11)]} \\ =& \rho_1(1-\rho_1) n_t \mu_{t-1} + \rho_2(1-\rho_2) \mu_{d_{t-1}} + \mu_t - \rho_1 n_t \mu_{t-1} - \\ & \rho_2 \mu_{d_{t-1}} + \rho_1^2 n_t^2 \sigma_{t-1,t-1} + \rho_2^2 v(d_{t-1}) \\ =& \mu_t - \rho_1^2 n_t \mu_{t-1} - \rho_2^2 \mu_{d_{t-1}} + \rho_1^2 n_t^2 \sigma_{t-1,t-1} + \rho_2^2 \sigma_{d_{t-1},d_{t-1}} \end{aligned}$$

$$=\mu_t - \rho_1^2 n_t \mu_{t-1} + \rho_1^2 n_t^2 \sigma_{t-1,t-1}$$

[Since from Assumption 2.2, $\mu_{d_{t-1}} = \sigma_{d_{t-1},d_{t-1}}$]

Therefore, we get the following expression for variance of y_t,

$$v(y_t) = \sigma_{t,t} = \mu_t - \rho_1^2 n_t \mu_{t-1} + \rho_1^2 n_t^2 \sigma_{t-1,t-1} \tag{A.15}$$

Again, $v(y_1) = \mu_1$. For $T = 2, 3, ...T$, using equation (A.15) we get,
when $t = 2$,

$$\begin{aligned} v(y_2) &= \mu_2 - \rho_1^2 n_2 \mu_1 + \rho_1^2 n_2^2 v(y_1) \\ &= \mu_2 - \rho_1^2 n_2 \mu_1 + \rho_1^2 n_2^2 \mu_1 \\ &= \mu_2 + \rho_1^2 n_2 (n_2 - 1) \mu_1 \end{aligned}$$

when $t = 3$,

$$\begin{aligned} v(y_3) &= \mu_3 - \rho_1^2 n_3 \mu_2 + \rho_1^2 n_3^2 v(y_2) \\ &= \mu_3 - \rho_1^2 n_3 \mu_2 + \rho_1^2 n_3^2 (\mu_2 + \rho_1^2 n_2 (n_2 - 1) \mu_1) \end{aligned}$$

when $t = 4$,

$$\begin{aligned} v(y_4) &= \mu_4 - \rho_1^2 n_4 \mu_3 + \rho_1^2 n_4^2 v(y_3) \\ &= \mu_4 - \rho_1^2 n_4 \mu_3 + \rho_1^2 n_4^2 (\mu_3 + \rho_1^2 n_3 (n_3 - 1) \mu_2 + \rho_1^4 n_3 n_2 (n_2 - 1) \mu_1) \\ &= \mu_4 + \rho_1^2 n_4 (n_4 - 1) \mu_3 + \rho_1^4 n_4^2 n_3 (n_3 - 1) \mu_2 + \rho_1^6 n_4^2 n_3^2 n_2 (n_2 - 1) \mu_1 \end{aligned}$$

If we calculate for $t = 5, 6, ...$ and so on we get a general expression for the variance of y_t which is given by,

$$v(y_t) = \begin{cases} \mu_1, & \text{for } t = 1 \\ \mu_2 + \rho_1^2 n_2 (n_2 - 1) \mu_1, & \text{for } t = 2 \\ \mu_t + \rho_1^2 n_t (n_t - 1) \mu_{t-1} + & \\ \displaystyle\sum_{l=1}^{t-2} \left[\rho_1^{2(l+1)} n_{t-l} (n_{t-l} - 1) \left(\prod_{j=0}^{l-1} n_{t-j}^2 \right) \mu_{t-(l+1)} \right], & \text{for } t = 3, .., T \end{cases}$$

Covariance and Correlation: To derive the lag k covariance between y_t and y_{t-k}, we use the following formula as, for $k = 1, 2, ..., T-1$,

$$
\begin{aligned}
cov(y_t, y_{t-k}) &= cov\left(\sum_{j=1}^{y_{t-1}} b_{1j}(n_t, \rho_1), y_{t-k}\right) + cov\left(\sum_{j=1}^{d_{t-1}} b_{2j}(\rho_2), y_{t-k}\right) + cov(d_t, y_{t-k}) \\
&= cov\left(\sum_{j=1}^{y_{t-1}} b_{1j}(n_t, \rho_1), y_{t-k}\right) + cov\left(\sum_{j=1}^{d_{t-1}} b_{2j}(\rho_2), y_{t-k}\right) \ \text{[since } cov(d_t, y_{t-k}) = 0\text{]}
\end{aligned}
$$

Considering the first part of the above formula we get, The covariance and correlation between y_t and y_{t-k} for $t = 1, 2, ..., T$ are obtained as,

$$
\begin{aligned}
cov\left(\sum_{j=1}^{y_{t-1}} b_{1j}(n_t, \rho_1), y_{t-k}\right) &= E\left[cov\left(\sum_{j=1}^{y_{t-1}} b_{1j}(n_t, \rho_1), y_{t-k} | y_{t-1}, d_{t-1}, y_{t-k}\right)\right] \\
&+ cov\left[E\left(\sum_{j=1}^{y_{t-1}} b_{1j}(n_t, \rho_1) | y_{t-1}, d_{t-1}, y_{t-k}\right), E\left(y_{t-k} | y_{t-1}, d_{t-1}, y_{t-k}\right)\right] \\
&= cov(\rho_1 n_t y_{t-1}, y_{t-k}) \\
&= \rho_1 n_t cov(y_{t-1}, y_{t-k}).
\end{aligned}
$$

Similarly, for the second part we get the same expression as in (A.8) given by,

$$
cov\left(\sum_{j=1}^{d_{t-1}} b_{2j}(\rho_2), y_{t-k}\right) = \rho_2 cov(d_{t-1}, y_{t-k}). \tag{A.16}
$$

The above relationship in (A.16) holds only when $k = 1$. Otherwise, when $k > 1$, $cov\left(\sum_{j=1}^{d_{t-1}} b_{2j}(\rho_2), y_{t-k}\right) = 0$. Therefore, when $k = 1$, we get,

$$
\begin{aligned}
cov(y_t, y_{t-k}) &= \rho_1 n_t cov(y_{t-1}, y_{t-k}) + \rho_2 cov(d_{t-1}, y_{t-k}) \\
&= \rho_1 n_t cov(y_{t-1}, y_{t-1}) + \rho_2 cov(d_{t-1}, y_{t-1}) \\
&= \rho_1 n_t \sigma_{t-1,t-1} + \rho_2 cov\left(d_{t-1}, \sum_{j=1}^{y_{t-2}} b_{1j}(\rho_1) + \sum_{j=1}^{d_{t-2}} b_{2j}(\rho_2) + d_{t-1}\right) \\
&= \rho_1 n_t \sigma_{t-1,t-1} + \rho_2 \sigma_{d_{t-1}, d_{t-1}}
\end{aligned}
$$

When $k > 1$,

$$\begin{aligned} cov(y_t, y_{t-k}) &= \rho_1 n_t cov(y_{t-1}, y_{t-k}) \\ &= \rho_1^2 n_t n_{t-1} cov(y_{t-2}, y_{t-k}) \\ &= \rho_1^3 n_t n_{t-1} n_{t-2} cov(y_{t-3}, y_{t-k}) \text{ leading to} \\ cov(y_t, y_{t-k}) &= \left(\prod_{l=0}^{k-1} n_{t-l}\right) \rho_1^k cov(y_{t-k}, y_{t-k}) = \left(\prod_{l=0}^{k-1} n_{t-l}\right) \rho_1^k \sigma_{t-k} \end{aligned}$$

In general, for $t = 1, 2, ..., T$ we get the following expression for covariance between y_t and y_{t-k} for lag $k = 1, 2, .., T - 1$

$$\text{cov}(y_t, y_{t-k}) = \begin{cases} n_t \rho_1 \sigma_{t-1,t-1} + \rho_2 \sigma_{d_{t-1},d_{t-1}}, & \text{for } k = 1, \\ \left(\prod_{l=0}^{k-1} n_{t-l}\right) \rho_1^k \sigma_{t-k,t-k}, & \text{for } k > 1. \end{cases}$$

By using the following formula for correlation,

$$corr(y_t, y_{t-k}) = \frac{cov(y_t, y_{t-k})}{\sqrt{v(y_t)}\sqrt{v(y_{t-k})}}$$

and plugging in the expressions for respective covariance and variances, we get correlation between y_t and y_{t-k} given by,

$$\text{corr}(y_t, y_{t-k}) = \begin{cases} \rho_1 n_t \sqrt{\dfrac{\sigma_{t-1,t-1}}{\sigma_{t,t}}} + \rho_2 \dfrac{\sigma_{d_{t-1},d_{t-1}}}{\sqrt{\sigma_{t,t}\sigma_{t-1,t-1}}}, & \text{for } k = 1, \\ \left(\prod_{l=0}^{k-1} n_{t-l}\right) \rho_1^k \sqrt{\frac{\sigma_{t-k,t-k}}{\sigma_{t,t}}}, & \text{for } k > 1. \end{cases}$$

A.4 Proof of Theorem 2.3.2

Based on the Assumption 3.2 outlined in section 2.3.4, $E(y_1) = v(y_1) = \mu_1$. Additionally, $E(d_1) = v(d_1) = \mu_1$. From Assumption 3.3, $E(d_t) = v(d_t) = \mu_t - n\rho_1\mu_{t-1} - \rho_2\mu_{d_{t-1}}$ for $t = 2, 3, ..., T$.

Mean: Let, $\mu_t = exp(\boldsymbol{x}_t{}^T\beta)$ be the mean of y_t and μ_{dt} be the mean of d_t. We know,

$$E\left(y_t\right) = E_{d_{t-1},y_{t-1}}\left(E\left(y_t|y_{t-1}, d_{t-1}\right)\right) \tag{A.17}$$

By taking conditional expectation on (2.19) we get,

$$E\left(y_t|Y_{t-1},d_{t-1}\right) = \rho_1 n y_{t-1} + \rho_2 d_{t-1} + E(d_t) \tag{A.18}$$

$$E_{y_{t-1}}\left(E\left(y_t|y_{t-1},d_{t-1}\right)\right) = \rho_1 n E(y_{t-1}) + \rho_2 d_{t-1} + E(d_t) \tag{A.19}$$

$$E_{d_{t-1},y_{t-1}}\left(E\left(y_t|y_{t-1},d_{t-1}\right)\right) = \rho_1 n E(y_{t-1}) + \rho_2 E(d_{t-1}) + E(d_t) \tag{A.20}$$

$$= \rho_1 n E(y_{t-1}) + \rho_2 E(d_{t-1}) + \mu_t - \rho_1 n \mu_{t-1} - \rho_2 \mu_{d_{t-1}} \tag{A.21}$$

Therefore,

$$E(y_t) = \rho_1 n E(y_{t-1}) + \rho_2 E(d_{t-1}) + \mu_t - \rho_1 n \mu_{t-1} - \rho_2 \mu_{d_{t-1}}$$

For $t = 2$,

$$\begin{aligned} E(y_2) &= \rho_1 n E(y_1) + \rho_2 E(d_1) + \mu_2 - \rho_1 n \mu_1 - \rho_2 \mu_{d_1} \\ &= \rho_1 n \mu_1 + \rho_2 \mu_1 + \mu_2 - \rho_1 n \mu_1 - \rho_2 \mu_1 \\ &= \mu_2 \end{aligned}$$

For $t = 3$,

$$\begin{aligned} E(y_3) &= \rho_1 n E(y_2) + \rho_2 E(d_2) + \mu_3 - \rho_1 n \mu_2 - \rho_2 \mu_{d_2} \\ &= \rho_1 n \mu_2 + \rho_2(\mu_2 - \rho_1 n \mu_1 - \rho_2 \mu_{d_1}) + \mu_3 - \rho_1 n \mu_2 - \rho_2(\mu_2 - \rho_1 n \mu_1 - \rho_2 \mu_{d_1}) \\ &= \mu_3 \end{aligned}$$

Through the method of mathematical induction, if $E(y_{t-1}) = \mu_{t-1}$, $E(d_{t-1}) = \mu_{d_{t-1}}$ then $E(y_t) = \rho_1 n \mu_{t-1} + \rho_2 \mu_{d_{t-1}} + \mu_t - \rho_1 n \mu_{t-1} - \rho_2 \mu_{d_{t-1}} = \mu_t$.
Therefore, $E(y_t) = \mu_t = exp(\boldsymbol{x}_t{}^T \beta)$ for all $t = 1, 2, 3, ..., T$.

Variance: Let, $v(y_t) = \sigma_{t,t}$ and $v(d_t) = \sigma_{d_t,d_t}$ for all $t = 1, 2, ..., T$. Additionally, $v(d_t) = E(d_t) = \mu_t - \rho_1 n \mu_{t-1} - \rho_2 \mu_{d_{t-1}}$. From Assumption 3.4 of the model in (2.19), $cov(y_{t-1}, d_t) = 0$. Now by taking conditional variance on (2.19),

$$\begin{aligned} v(y_t|y_{t-1},d_{t-1}) &= y_{t-1} v(b_{1j}) + d_{t-1} v(b_{2j}) + v(d_t) \\ &= \rho_1(1-\rho_1) n y_{t-1} + \rho_2(1-\rho_2) d_{t-1} + \mu_t - \rho_1 n \mu_{t-1} - \rho_2 \mu_{d_{t-1}}. \end{aligned} \tag{A.22}$$

Let, $v(Y_{t-1}) = \sigma_{t-1,t-1}$ and $v(d_{t-1}) = \sigma_{d_{t-1},d_{t-1}}$. Then,

$$\begin{aligned} v(y_t|d_{t-1}) &= E_{y_{t-1}}\left(v\left(y_t|y_{t-1}, d_{t-1}\right)\right) + v_{y_{t-1}}\left(E\left(y_t|y_{t-1}, d_{t-1}\right)\right) \\ &= E_{y_{t-1}}(\rho_1(1-\rho_1)ny_{t-1} + \rho_2(1-\rho_2)d_{t-1} + \mu_{dt}) \\ &+ v_{y_{t-1}}(\rho_1 ny_{t-1} + \rho_2 d_{t-1} + \mu_t - \rho_1 n\mu_{t-1} - \rho_2\mu_{d_{t-1}}) \\ &\qquad\qquad\qquad\qquad\qquad\qquad \text{[using (A.18)]} \\ &= \rho_1(1-\rho_1)n\mu_{t-1} + \rho_2(1-\rho_2)d_{t-1} + \mu_{dt} + \rho_1^2 n^2\sigma_{t-1,t-1}. \end{aligned}$$

Now,

$$\begin{aligned} v\left(y_t\right) =& E_{d_{t-1}}\left(v\left(y_t|\ d_{t-1}\right)\right) + v_{d_{t-1}}\left(E\left(Y_t|\ d_{t-1}\right)\right) \\ =& E_{d_{t-1}}\left(\rho_1(1-\rho_1)n\mu_{t-1} + \rho_2(1-\rho_2)d_{t-1} + \mu_{dt} + \rho_1^2 n^2\sigma_{t-1,t-1}\right) \\ & + v_{d_{t-1}}\left(\mu_t + \rho_2 d_{t-1} - \rho_2\mu_{d_{t-1}}\right) \text{ [using (A.19)]} \\ =& \rho_1(1-\rho_1)n\mu_{t-1} + \rho_2(1-\rho_2)\mu_{d_{t-1}} + \mu_t - \rho_1 n\mu_{t-1} - \\ & \rho_2\mu_{d_{t-1}} + \rho_1^2 n^2\sigma_{t-1,t-1} + \rho_2^2 v(d_{t-1}) \\ =& \mu_t - \rho_1^2 n\mu_{t-1} - \rho_2^2\mu_{d_{t-1}} + \rho_1^2 n^2\sigma_{t-1,t-1} + \rho_2^2\sigma_{d_{t-1},d_{t-1}} \\ =& \mu_t - \rho_1^2 n\mu_{t-1} + \rho_1^2 n^2\sigma_{t-1,t-1} \\ & \text{[Since from Assumption 3.2, } \mu_{d_{t-1}} = \sigma_{d_{t-1},d_{t-1}}\text{].} \end{aligned}$$

Therefore, we get the following expression for variance of y_t,

$$v(y_t) = \sigma_{t,t} = \mu_t - \rho_1^2 n\mu_{t-1} + \rho_1^2 n^2\sigma_{t-1,t-1}. \tag{A.23}$$

Again, $v(y_1) = \mu_1$. For $T = 2, 3, ...T$, using the equation in (A.23) we get, when $t = 2$,

$$\begin{aligned} v(y_2) &= \mu_2 - \rho_1^2 n\mu_1 + \rho_1^2 n^2 v(y_1) \\ &= \mu_2 - \rho_1^2 n\mu_1 + \rho_1^2 n^2\mu_1 \\ &= \mu_2 + \rho_1^2 n(n-1)\mu_1 \end{aligned}$$

when $t = 3$,

$$\begin{aligned} v(y_3) &= \mu_3 - \rho_1^2 n\mu_2 + \rho_1^2 n^2 v(y_2) \\ &= \mu_3 + n(n-1)(\rho_1^2\mu_2 + \rho_1^4 n^2 \mu_1) \end{aligned}$$

when $t = 4$,

$$\begin{aligned} v(y_4) &= \mu_4 - \rho_1^2 n\mu_3 + \rho_1^2 n^2 v(y_3) \\ &= \mu_4 - \rho_1^2 n\mu_3 + \rho_1^2 n^2(\mu_3 + n(n-1)(\rho_1^2\mu_2 + \rho_1^4 n^2\mu_1)) \\ &= \mu_4 + n(n-1)(\rho_1^2\mu_3 + \rho_1^4 n^2\mu_2 + \rho_1^6 n^3\mu_1) \end{aligned}$$

If we calculate for $t = 5, 6, ...$ and so on we get a general expression for the variance of y_t which is given by,

$$v(y_t) = \begin{cases} \mu_1, & \text{for } t = 1, \\ \mu_t + n(n-1) + \sum_{l=1}^{t-1} \rho_1^{2(l)} n^{2(l-1} \mu_{t-l}, & \text{for } t = 2, 3, .., T. \end{cases}$$

Covariance and Correlation: To derive the lag k covariance between y_t and y_{t-k}, we use the following formula as, for $k = 1, 2, ..., T-1$,

$$\begin{aligned} cov(y_t, y_{t-k}) &= cov\left(\sum_{j=1}^{y_{t-1}} b_{1j}(n, \rho_1), y_{t-k}\right) + cov\left(\sum_{j=1}^{d_{t-1}} b_{2j}(\rho_2), y_{t-k}) + cov(d_t, y_{t-k}\right) \\ &= cov\left(\sum_{j=1}^{y_{t-1}} b_{1j}(n, \rho_1), y_{t-k}\right) + cov\left(\sum_{j=1}^{d_{t-1}} b_{2j}(\rho_2), y_{t-k}\right) \text{ [since } cov(d_t, y_{t-k}) = 0] \end{aligned}$$

Considering the first part of the above formula we get, The covariance and correlation between y_t and y_{t-k} for $t = 1, 2, ..., T$ are obtained as,

$$\begin{aligned} cov\left(\sum_{j=1}^{y_{t-1}} b_{1j}(n, \rho_1), y_{t-k}\right) &= E\left[cov\left(\sum_{j=1}^{y_{t-1}} b_{1j}(n, \rho_1), y_{t-k}|y_{t-1}, d_{t-1}, y_{t-k}\right)\right] \\ &+ cov\left[E\left(\sum_{j=1}^{y_{t-1}} b_{1j}(n, \rho_1)|y_{t-1}, d_{t-1}, y_{t-k}\right), E\left(y_{t-k}|y_{t-1}, d_{t-1}, y_{t-k}\right)\right] \end{aligned}$$

$$= cov(\rho_1 n y_{t-1}, y_{t-k})$$
$$= \rho_1 n cov(y_{t-1}, y_{t-k})$$

Similarly, for the second part we get the same expression as in (A.8) given by

$$cov\left(\sum_{j=1}^{d_{t-1}} b_{2j}(\rho_2), y_{t-k}\right) = \rho_2 cov(d_{t-1}, y_{t-k}) \tag{A.24}$$

The above relationship in (A.24) holds only when $k = 1$. Otherwise, when $k > 1$, $cov\left(\sum_{j=1}^{d_{t-1}} b_{2j}(\rho_2), y_{t-k}\right) = 0$. Therefore, when $k = 1$, we get,

$$\begin{aligned}
cov(y_t, y_{t-k}) &= \rho_1 n cov(y_{t-1}, y_{t-k}) + \rho_2 cov(d_{t-1}, y_{t-k}) \\
&= \rho_1 n cov(y_{t-1}, y_{t-1}) + \rho_2 cov(d_{t-1}, y_{t-1}) \\
&= \rho_1 n \sigma_{t-1,t-1} + \rho_2 cov\left(d_{t-1}, \sum_{j=1}^{y_{t-2}} b_{1j}(\rho_1) + \sum_{j=1}^{d_{t-2}} b_{2j}(\rho_2) + d_{t-1}\right) \\
&= \rho_1 n \sigma_{t-1,t-1} + \rho_2 \sigma_{d_{t-1},d_{t-1}}.
\end{aligned}$$

When $k > 1$,

$$\begin{aligned}
cov(y_t, y_{t-k}) &= \rho_1 n cov(y_{t-1}, y_{t-k}) \\
&= \rho_1^2 n^2 cov(y_{t-2}, y_{t-k}) \\
&= \rho_1^3 n^3 cov(y_{t-3}, y_{t-k}) \text{ leading to} \\
cov(y_t, y_{t-k}) &= \prod_{l=0}^{k-1} (n\rho_1)^k cov(y_{t-k}, y_{t-k}) = \prod_{l=0}^{k-1} (n\rho_1)^k \sigma_{t-k,t-k}.
\end{aligned}$$

In general, for $t = 1, 2, ..., T$ we get the following expression for covariance between y_t and y_{t-k} for lag $k = 1, 2, .., T-1$

$$\text{cov}(y_t, y_{t-k}) = \begin{cases} n\rho_1 \sigma_{t-1,t-1} + \rho_2 \sigma_{d_{t-1},d_{t-1}}, & \text{for } k = 1, \\ \prod_{l=0}^{k-1} (n\rho_1)^k \sigma_{t-k,t-k}, & \text{for } k > 1. \end{cases}$$

By using the following formula for correlation,

$$corr(y_t, y_{t-k}) = \frac{cov(y_t, y_{t-k})}{\sqrt{v(y_t)}\sqrt{v(y_{t-k})}},$$

and plugging in the expressions for respective covariance and variances, we get correlation between y_t and y_{t-k} given by,

$$\operatorname{corr}(y_t, y_{t-k}) = \begin{cases} \rho_1 n \sqrt{\dfrac{\sigma_{t-1,t-1}}{\sigma_{t,t}}} + \rho_2 \dfrac{\sigma_{d_{t-1},d_{t-1}}}{\sqrt{\sigma_{t,t}\sigma_{t-1,t-1}}}, & \text{for } k = 1, \\ \displaystyle\prod_{l=0}^{k-1} (n\rho_1)^k \sqrt{\dfrac{\sigma_{t-k,t-k}}{\sigma_{t,t}}}, & \text{for } k > 1. \end{cases}$$