Fitting Bivariate Integer-valued AR(1) Models with Negative Binomial, and Normal Innovations to Count Data with Seasonality

Moses Luri

Abstract

In this study, I demonstrate one way of eliminating seasonality in count time series data using an actual data set of hepatitis A incidents in two Australian states. We then fit the deseasonalized count data (or stabilized residuals) to two bivariate integer value auto regressive models of order one (BINAR(1)); the first with negative binomial innovations, and the second with normal innovations. I obtain and compare maximum likelihood estimates, and diagnostics of the two BINAR(1) models showing that the model that assumes bivariate negative binomial innovations fits the deseasonalized data better than the one that assumes bivariate normal innovations. I therefore conclude that the approach employed in our study to adjust for seasonality in count time series data is relevant for preserving the count structure in deseasonalizing count data; and provides an alternative to using normal approximations.

Keywords: Bivarivate integer-value AR(1), count time series, seasonality, deseasonalized data, negative binomial innovations, normal innovations, maximum likelihood estimation, diagnostics

Contents

1 Introduction

Many natural phenomena occur that yield multivariate count data. These data observed over time produce multivariate count time series data. An example of this, is the number of people admitted each week to the emergency rooms (ERs) of three different hospitals, in one year. [2] and [3], cited in [5] proposed the integer autoregressive time series models for fitting count time series data. And following this, both [5] and [4] developed and studied bivariate integer-value autoregressive models of first order (BINAR(1)).

A particularly important problem arises when a count time series data set that is of interest contains trend and/or seasonal component(s). When this happens, such trend and/or seasonality component(s) must first be removed before any modeling can be done on the data. In general, removing a trend and/or seasonality in a data set requires some form of time series decomposition (but more on this later).

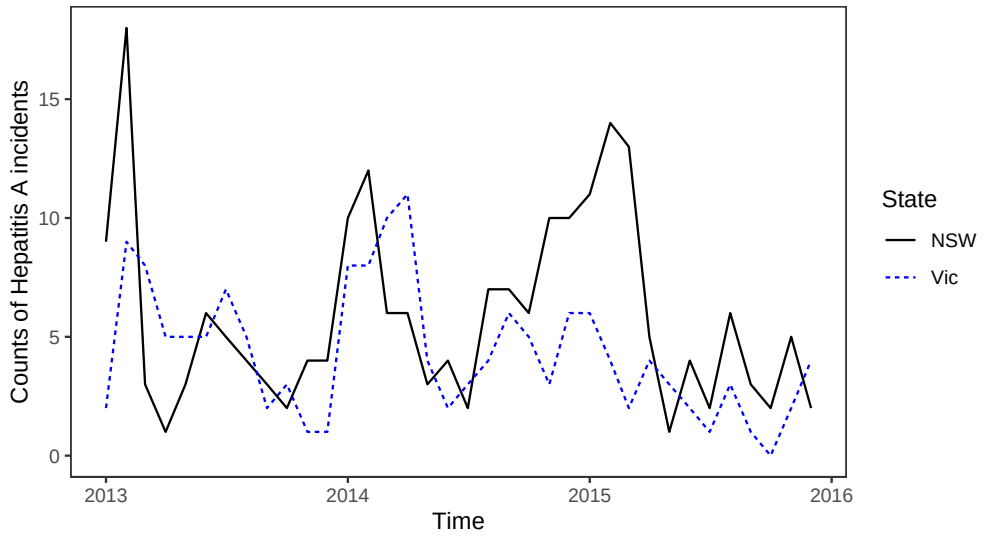
To put things in context, Figure 1 below shows time series plots of monthly hepatitis A incidents in two Australian states: New South Wales (NSW) and Victoria (Vic), for the period January 2013 through December 2015. [4] estimated a BINAR(1) model using this data set. They identify overdispersion as an issue and therefore assumed a bivariate negative binomial error component in their BINAR(1) estimation to address this problem. The authors however did not address the issue of seasonality that appear to be present in the data; a seasonal pattern can be observed, with spikes in hepatitis A incidents occurring every Australian summer (February).

This study therefore focuses on two things: first, remove the seasonal component in the hepatitis A data set; and second, use the seasonally-adjusted (or deseasonalized) data to re-fit the BINAR(1) model with negative binomial innovation (error component) of [4], and compare that model fit with a second BINAR(1) estimation that assumes normal innovations. The broad relevance of this study that it demonstrates one way of handling seasonality in count time series data. Thus, potentiabelly providing an alternative to the use of normal approximations for fitting seasonally-adjusted residuals of count time series data.

The remainder of this study is organized as follows. In Section 2, we present the steps, and remove seasonality in the hepatitis A data set. In Section 3, we describe general specifications

of BINAR(1) processes with negative binomial and normal innovations, and obtain conditional likelihood functions required to fit the respective BINAR(1) models. In Section 4, we present and compare the maximum likelihood estimation results and perform model fit diagnostics for the two models considered. We provide concluding remarks in Section 5.

Figure 1: Time series of hepatitis A incidents in two Australian states



2 Eliminating Seasonality in Count Time Series Data

The first step in dealing with seasonality in any time series is to determine the underlying model that can be used to decompose the observed data. Decomposition models are typically additive or multiplicative, but can also take other forms such as pseudo-additive [1]. An additive model is appropriate when the amplitude of both the seasonal and irregular variations in a time series do not change as the level of the trend rises or falls. While a multiplicative model is appropriate when the amplitude of both the seasonal and irregular variations increase as the level of the trend rise. We use an additive specification in this study, since that is better suited for our hepatitis A data set.

The general additive model expresses the observed time series (X_t) as the sum of three independent components: the seasonal s_t (or repeating short-term cycle in the series), the trend m_t (increasing or decreasing value in the series), and the irregular i_t (or random variation in the series), see [1]. That is,

$$X_t = s_t + m_t + i_t \tag{1}$$

where $E(i_t) = 0$, $s_{t+p} = s_t$ and $\sum_{i=1}^{p} s_i = 0$. p is the period of the seasonal component (which is 12 in the case our hepatitis A data set). For convenience, we index the time series by year and month as follows: $x_{l,k}$, $l = 1, 2, 3$, $k = 1, ..., 12$ denote the number of hepatitis A cases reported for month k in year l.

[1] notes that if the trend component in a data is small (as in the hepatitis A data) it is not unreasonable assume that trend component of Equation (1) is constant for each year. So that, since $\sum_{k=1}^{12} s_k = 0$, we obtain the natural unbiased estimate

$$\hat{m} = \frac{1}{12} \sum_{k=1}^{12} x_{l,k}, \tag{2}$$

while for s_k, $1, ..., 12$ we obtain the estimates,

$$\hat{s}_k = \frac{1}{3} \sum_{l=1}^{3} (x_{l,k} - \hat{m}_l), \tag{3}$$

which then satisfies the requirement that $\sum_{k=1}^{12} s_k = 0$. The seasonal component in Equation (1) is therefore removed by simply computing $x_{l,k} - \hat{s}_k$ to arrive at a deseasonalized time series. Notice that the deseasonalized data is no longer in count form, and so some adjustment is necessary to restore the count structure. We do this by adding the smallest observation in the deseasonalized data to all observations, and round off the result to the nearest integer. We adjust further by adding a constant positive integer value to the rounded values. In summary, we compute

$$x^d_{l,k} = \min_{l^*}(x_{l^*,k} - \hat{s}_k) + (x_{l,k} - \hat{s}_k) + c \tag{4}$$

where $x^d_{l,k}$ is the deseasonalized count data and c is a positive integer.

Figure 2 shows the deseasonalized hepatitis A data for the two Australian states [1], while Figure 3 displays the autocorrelation functions (ACF) and partial autocorrelation functions (PACF) for the

[1]The plots in Figure 2 exclude March 2015 values for both states, because they exhibit possible outlier behavior (see Figure 8 in appendix).

deseasonalized hepatitis A data. Observe that the deseasonalized time series in both states exhibit significant first order autocorrelation, hence, fitting bivariate AR(1) models is appropriate.

Figure 2: Deseasonalized time series plot of hepatitis A incidents in two Australian states

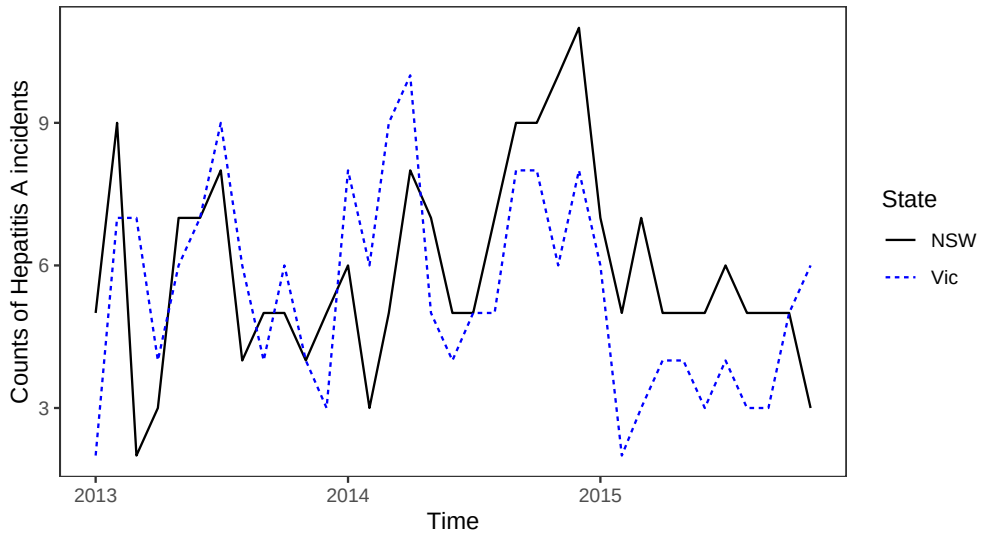
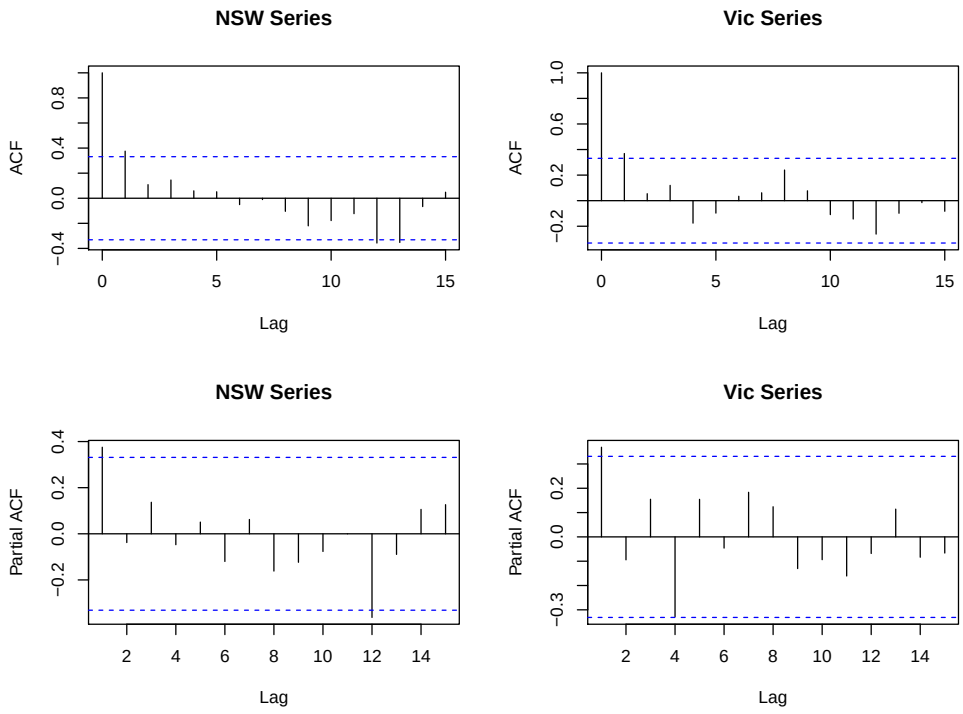


Figure 3: ACF and PACF plots of deseasonalized hepatitis A time series for two Australian states



3 Bivariate INAR(1) Process with Negative Binomial and Normal Innovations

[4] provides a detailed derivation and analysis of the joint conditional likelihood function required to obtain maximum likelihood estimates of the bivariate INAR(1) process with negative binomial innovations. We therefore recommend reading Section 2 of [4] alongside this study. In what follows below, we present a general description of the BINAR(1) model, including the conditional likelihood function needed to compute parameter estimates and numerical diagnostics.

3.1 Bivariate AR(1) Process

For $t = 1, 2, ...$, let $\boldsymbol{Y_t} = (Y_{1,t}, Y_{2,t})'$ and $\boldsymbol{R_t} = (R_{1,t}, R_{2,t})'$ be real-valued random vectors, and let α_1 and α_2 be scalar constants belonging to the unit interval $[0, 1]$. A bivariate autoregressive process of order 1 is given by

$$\boldsymbol{Y_t} = A\boldsymbol{Y_{t-1}} + \boldsymbol{R_t} = \begin{bmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{bmatrix} \begin{bmatrix} Y_{1,t-1} \\ Y_{2,t-1} \end{bmatrix} + \begin{bmatrix} R_{1,t} \\ R_{2,t} \end{bmatrix}. \tag{5}$$

From Equation (5), one can extract

$$Y_{j,t} = \alpha_j Y_{j,t-1} + R_{j,t} \quad \text{for} \quad j = 1, 2. \tag{6}$$

The components of the random vector $\boldsymbol{R_t}$ are often referred to as innovations. Suppose $\boldsymbol{R_1}, \boldsymbol{R_2}, ...$ are iid such that $E(R_{j,t}) = \mu_j \in \mathbb{R}$ and variance $\text{Var}(R_{j,t}) = \sigma_j^2 \in \mathbb{R}_+$ where $j = 1, 2$. Then, from

(6) we have that

$$E(Y_{j,t}) = \frac{\mu_j}{1-\alpha_j} \tag{7}$$

$$\text{Var}(Y_{j,t}) = \frac{\sigma_j^2}{1-\alpha_j^2} \tag{8}$$

$$\text{Cov}(Y_{j,t}, Y_{j,t+h}) = \alpha_j^h \ \text{Var}(Y_{j,t})$$

$$\text{Corr}(Y_{j,t}, Y_{j,t+h}) = \alpha_j^h, \quad \text{for} \quad h = 0, 1, 2, ...$$

Notice that the correlation between the elements of $\boldsymbol{Y_t}$ originate from correlations within $\boldsymbol{R_t}$. In particular, it can be shown that covariance and correlation between $Y_{1,t}$ and $Y_{2,t+h}$ are respectively analogously defined as

$$\text{Cov}(Y_{1,t+h}, Y_{2,t}) = \frac{\alpha_1^h}{1-\alpha_1\alpha_2} \ \text{Cov}(R_{1,t}, R_{2,t}), \tag{9}$$

$$\text{Corr}(Y_{1,t+h}, Y_{2,t}) = \frac{\alpha_1^h\sqrt{(1-\alpha_1^2)(1-\alpha_2^2)}}{(1-\alpha_1\alpha_2)\sigma_1\sigma_2} \ \text{Cov}(R_{1,t}, R_{2,t}), \quad \text{for} \quad h = 0, 1, 2, ... \tag{10}$$

The covariance $\text{Cov}(R_{1,t}, R_{2,t})$, or equivalently, $\text{Cov}(Y_{1,t+h}, Y_{2,t})$ and $\text{Corr}(Y_{1,t+h}, Y_{2,t})$ depend on distribution of the random innovation vector $\boldsymbol{R} = (R_1, R_2)'$. In this study we consider the cases where random innovations follow negative binomial and normal distributions.

3.2 Bivariate Negative Binomial Distribution

Following [4], suppose $\boldsymbol{R}_t$ follows a bivariate negative binomial (NB) distribution. We can derive the covariance structure $\boldsymbol{R}_t$ as follows.

Let θ be a Gamma random variable with shape parameter $1/\beta$ and scale parameter 1. Given θ, and assume that R_1 and R_2 are independently distributed Poission random variables with means $\lambda_1\beta\theta$ and $\lambda_2\beta\theta$, respectively. The joint probability mass function of $\boldsymbol{R} = (R_1, R_2)'$ is

$$f_{\boldsymbol{R}}(r_1, r_2|\beta) = \frac{(r_1 + r_2 + \beta^{-1})}{r_1!r_2!\Gamma(\beta^{-1})} \frac{\lambda_1\lambda_2\beta^{-(\beta^{-1})}}{(\lambda_1 + \lambda_2 + \beta^{-1})^{r_1+r_2+\beta^{-1}}} \tag{11}$$

for $r_1, r_2 = 0, 1, 2, ...$ According to [4], the marginal distribution of R_i, for $i = 1, 2$, is NB($\eta_1 =$

$1/\beta, \eta = \lambda_i$) and $\text{Cov}(R_1, R_2) = \beta\lambda_1\lambda_2$, which completes the Equations (9) and (10). Refer to [4] for the joint conditional distribution of $\boldsymbol{R}_t$ given $\boldsymbol{R}_{t-1} = \boldsymbol{r}$.

3.3 Bivariate Normal Distribution

Let R_1 and R_2 be two normally distributed random variables, $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ respectively. The random vector $\boldsymbol{R} = (R_1, R_2)'$ is a bivariate normal random variable with mean and covariance parameter restrictions:

1. $\mu_1, \mu_2, \in \mathbb{R}$
2. $\sigma_1, \sigma_2 \in \mathbb{R}_+$
3. $\sigma_{12} = \sigma_{21} = \rho\sigma_1\sigma_2$, where $\sigma_{12} = \sigma_{21}$ and ρ are respectively the covariance and correlation between R_1 and R_2
4. $\sigma_{12} \leq \sigma_1\sigma_2$ (Cauchy - Schwartz)

The joint probability density function of $\boldsymbol{R} = (R_1, R_2)'$ is

$$f_{\boldsymbol{R}}(r_1, r_2|\mu_1, \mu_2, \sigma_1, \sigma_2, \rho) = \frac{\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(r_1-\mu_1)^2}{\sigma_1^2} + \frac{(r_2-\mu_2)^2}{\sigma_2^2} - \frac{2\rho(r_1-\mu_1)(r_2-\mu_2)}{\sigma_1\sigma_2}\right]\right\}}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \tag{12}$$

where $r_1, r_2 \in (-\infty, \infty)$ and the mean and covariance matrix of $\boldsymbol{R} = (R_1, R_2)'$ are respectively defined as $\boldsymbol{\mu} = (\mu_1, \mu_2)$ and $\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_2\sigma_1\rho & \sigma_2^2 \end{bmatrix}$ — a positive semi-definite matrix. In this case, $\text{Cov}(R_1, R_2) = \sigma_1\sigma_2\rho$.

3.4 Joint Conditional Distribution of Y_t Given $Y_{t-1} = y_{t-1}$

The joint conditional distribution of $\boldsymbol{Y_t}$ given $\boldsymbol{Y_{t-1}} = \boldsymbol{y_{t-1}}$ is fundamental to obtain the likelihood function required to obtain estimates for the parameter vector $\boldsymbol{\theta} = (\alpha_1, \alpha_2, \mu_1, \mu_2, \sigma_1, \sigma_2, \rho)'$. The

joint conditional pdf of $\boldsymbol{Y}_t$ given $\boldsymbol{Y}_{t-1} = \boldsymbol{y_{t-1}}$ is

$$f_{\boldsymbol{Y}}(\boldsymbol{Y_t}|\boldsymbol{Y_{t-1}}, \boldsymbol{\theta}) = \frac{1}{2\pi|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}\left[\boldsymbol{Y_t} - (\boldsymbol{AY_{t-1}} + \boldsymbol{\mu})\right]' \Sigma^{-1} \left[\boldsymbol{Y_t} - (\boldsymbol{AY_{t-1}} + \boldsymbol{\mu})\right]\right\} \tag{13}$$
$$= \frac{1}{2\pi|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}\boldsymbol{R_t'}\Sigma^{-1}\boldsymbol{R_t}\right\}$$

The mean and variance of $Y_{j,t}|y_{j,t-1}$ are respectively,

$$E(Y_{j,t}|y_{j,t-1}) = \alpha_j y_{j,t-1} + \mu_j, \quad \text{Var}(Y_{j,t}|y_{j,t-1}) = \text{Var}(R_j) = \sigma_j^2 \quad \text{for} \quad j = 1, 2.$$

The conditional covariance between $Y_{1,t}$ and $Y_{2,t}$ is

$$\text{Cov}(Y_{1,t}, Y_{2,t}|y_{t-1}) = \text{Cov}(R_{1,t}, R_{2,t}) = \sigma_1\sigma_2\rho$$

3.5 Maximum Likelihood Estimation

Let $(y_1, ..., y_t)$ be all observations up to and including those at time t. The conditional likelihood function at time t has the form

$$\mathcal{L}(\boldsymbol{\theta}|y_1, ..., y_t) = \prod_{i=1}^{t} f_{\boldsymbol{Y}}(\boldsymbol{Y_t}|\boldsymbol{Y_{t-1}}, \boldsymbol{\theta}) \tag{14}$$

where $\boldsymbol{y_0}$ is a specified initial value. In estimating $\boldsymbol{\theta}$, one may skip the specification of $\boldsymbol{y_0}$, and start measuring likelihood contributions at time $i = 2$ rather than $i = 1$, and then maximize $\mathcal{L}(\boldsymbol{\theta}|y_1, ..., y_t)$ with respect to $\boldsymbol{\theta}$.

3.6 Diagnostics

Here, I describe diagnostics for assessing the goodness of fit. Typically, in model fitting, this is accomplished by means of residual analysis [5]. The classical definition of a residual is the difference between the observed and fitted value. The conditional residual of $Y_{j,t}|y_{j,t-1}$ is

$$\hat{R}_{j,t} = Y_{j,t} - \hat{\alpha}_j Y_{j,t-1} - \hat{\mu}_j \tag{15}$$

where $\hat{\alpha}_j$ and $\hat{\mu}_j$ are the maximum likelihood estimates of α_j and μ_j.

4 Numerical Results

Using the seasonally-adjusted hepatitis A data presented in Figure 2, I obtain maximum likelihood estimates (base on the general specification in Equation (14) for two bivarite AR(1) models: the first with negative binomial innovations, and the second with normal innovations. I report and compare the log-likelihood and AIC values, as well as dignostic plots of residuals [2] for the two model fits.

First, a total of five parameters are estimated for the BINAR(1) model fit with negative binomial innovations, and seven parameters for the model with normal innovations. The log-likelihood and AIC values for the BINAR(1) model fit with negative binomial innovations are respectively, (-139.2525) and (278.505), but (-136.0692) and (286.1384) respectively for the model fit with normal innovations. Thus, the model fit with negative binomial innovations fits the data better than the one with normal innovations based on AIC.

The residual and Ljung-Box test plots in Figures 4 and 5 show that there is no lack of fit in either of the two fitted models. The residuals of the fit with negative binomial innovations however, are less spread out than the residuals of the fit with normal innovations. Finally, the ACF and PACF plots of residuals of the two model fits are presented in Figures 6 and 7. Observe that none to ACF and PACF plots exhibits signs of auto-correlations in the residuals, hence satisfying the assumption of independence.

5 Conclusion

Adjusting for seasonality (and/or trends) in count time series using available classical (or traditional) methods can result in stationary residuals that are effectively non-count in nature [3]. A common approach then, is to model the data as an approximation of some continuous distribution (often normal). The method illustrated in this study shows an alternative way of keeping the count

[2] Residuals are computed using Equation (15).

[3] Applying the method of differencing to remove seasonality and/or trends in count time series may preserve the count structure in the final stationary data. However, differencing generally results in data loss at the head of a data set; and may therefore not be suitable for small data sets.

structure in count time series that is seasonalized and/or trended. Most importantly, I show that keeping the count structure delivers a better model fit than using a continuous approximation.

Figure 4: Residual and Ljung-Box test plots for BINAR(1) model fit with negative binomial innovations to deseasonalized hepatitis A data for two Australian States

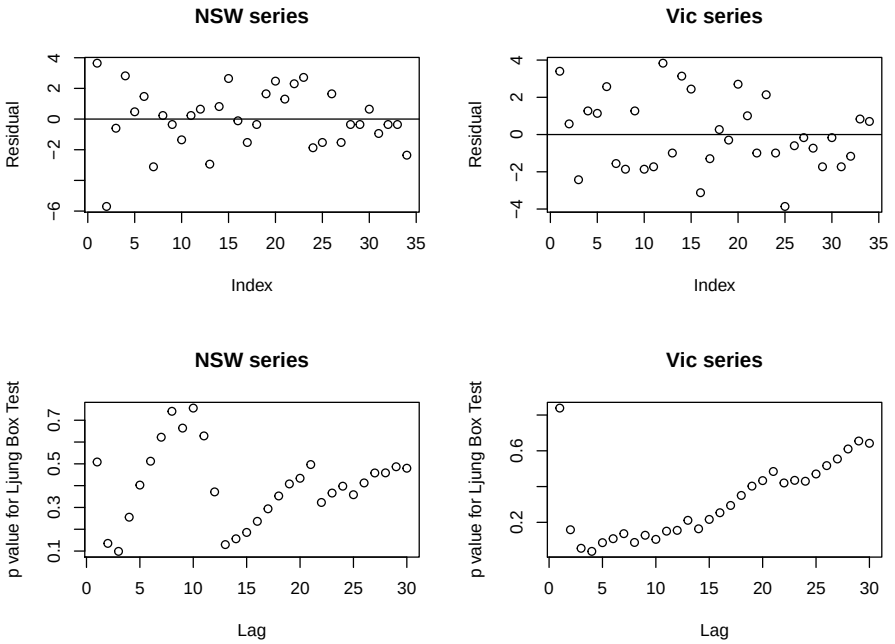


Figure 5: Residual and Ljung-Box test plots of BINAR(1) model fit with normal innovations to deseasonalized hepatitis A data for two Australian States

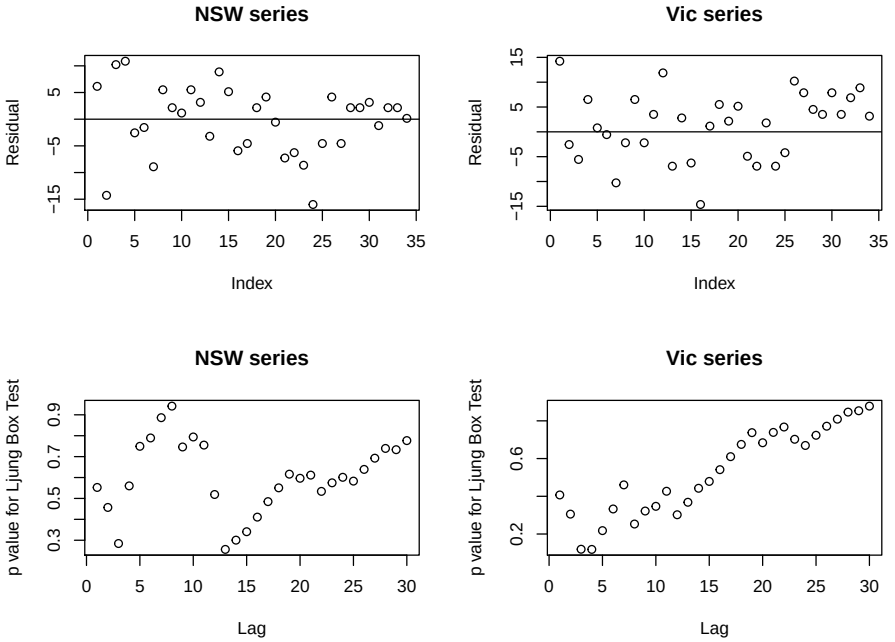


Figure 6: ACF and PACF plots of residuals of BINAR(1) model fit with negative binomial innovations to deseasonalized hepatitis A data for two Australian States

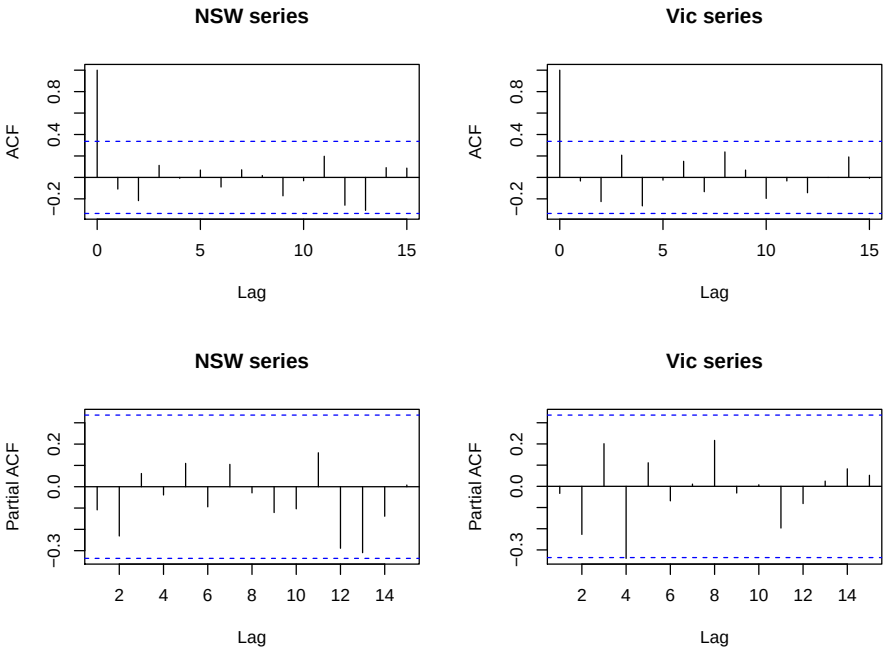
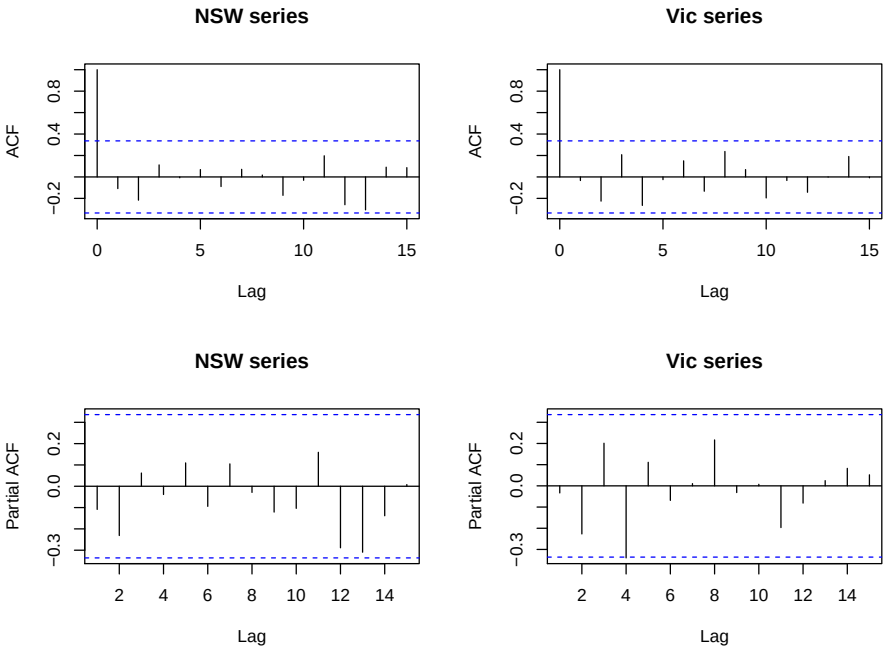


Figure 7: ACF and PACF plots of residuals of BINAR(1) model fit with normal innovations to deseasonalized hepatitis A data for two Australian States



References

[1] Peter J Brockwell and Richard A Davis. *Introduction to time series and forecasting*. springer, 2016.

[2] Ed McKenzie. "Some simple models for discrete variate time series 1". In: *JAWRA Journal of the American Water Resources Association* 21.4 (1985), pp. 645–650.

[3] MA Al-Osh and Aus A Alzaid. "First-order integer-valued autoregressive (INAR (1)) process". In: *Journal of Time Series Analysis* 8.3 (1987), pp. 261–275.

[4] Francis G Pascual and Sherzod B Akhundjanov. "Monitoring a bivariate INAR (1) process with application to Hepatitis A". In: *Communications in Statistics-Theory and Methods* (2019), pp. 1–23.

[5] Xanthi Pedeli and Dimitris Karlis. "A bivariate INAR (1) process with application". In: *Statistical modelling* 11.4 (2011), pp. 325–349.

Appendices

Figure 8: Deseasonalized time series plots of hepatitis A incidents in two Australian states

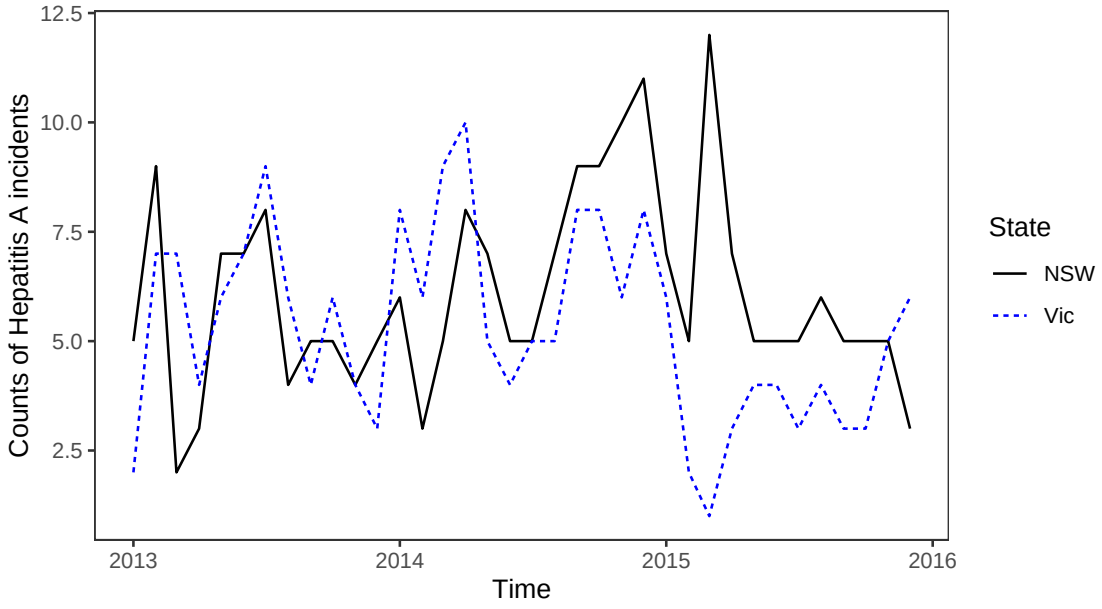


Table 1: The Hepatitis A data for two Australian states

.

Period	NSW	Vic	Period	NSW	Vic
Jan-13	9	2	Jul-14	2	3
Feb-13	18	9	Aug-14	7	4
Mar-13	3	8	Sep-14	7	6
Apr-13	1	5	Oct-14	6	5
May-13	3	5	Nov-14	10	3
Jun-13	6	5	Dec-14	10	6
Jul-13	5	7	Jan-15	11	6
Aug-13	4	5	Feb-15	14	4
Sep-13	3	2	Mar-15	13	2
Oct-13	2	3	Apr-15	5	4
Nov-13	4	1	May-15	1	3
Dec-13	4	1	Jun-15	4	2
Jan-14	10	8	Jul-15	2	1
Feb-14	12	8	Aug-15	6	3
Mar-14	6	10	Sep-15	3	1
Apr-14	6	11	Oct-15	2	0
May-14	3	4	Nov-15	5	2
Jun-14	4	2	Dec-15	2	4

Source: The NNDSS.

Table 2: The Deseasonalized Hepatitis A data for two Australian states

.

Period	NSW	Vic	Period	NSW	Vic
Jan-13	5	5	Jul-14	5	5
Feb-13	9	7	Aug-14	7	5
Mar-13	2	9	Sep-14	9	8
Apr-13	3	9	Oct-14	9	8
May-13	7	10	Nov-14	10	6
Jun-13	7	11	Dec-14	11	8
Jul-13	8	7	Jan-15	7	6
Aug-13	4	5	Feb-15	5	2
Sep-13	5	12	Mar-15	12*	1*
Oct-13	5	7	Apr-15	7	3
Nov-13	4	5	May-15	5	4
Dec-13	5	5	Jun-15	5	4
Jan-14	6	5	Jul-15	5	3
Feb-14	3	6	Aug-15	6	4
Mar-14	5	5	Sep-15	5	3
Apr-14	8	5	Oct-15	5	3
May-14	7	5	Nov-15	5	5
Jun-14	5	3	Dec-15	3	6

* Exhibits possible outlier behavior, hence are eliminated from the plots in figure 2