Poisson Autoregressive and Moving-Average Models for Forecasting Non-stationary Seasonal Time Series of Tourist Counts in Mauritius

Vandna Jowaheer[1,4], Naushad Ali Mamode Khan[2] and Yuvraj Sunecher[3]
[1,2]University of Mauritius, Reduit, Mauritius
[3]University of Technology, Pointe-Aux-Sables, Mauritius
[4]Corresponding author: Vandna Jowaheer, e-mail: vandnaj@uom.ac.mu

Abstract
This paper aims at developing Poisson autoregressive and moving-average models while incorporating transcendental covariates corresponding to the seasonal impulses for non-stationary time series of large counts. These models are implemented to analyse the non-stationary monthly time series of tourist arrivals in Mauritius over the period of twenty five years (Dec 1985 – Dec 2010). The regression effects of the covariates are estimated using efficient generalized quasi-likelihood approach and the correlation parameters are consistently estimated using the method of moments. The optimal model selection for forecasting is made using the diagnostics based on portmanteau type statistic. The forecasting model is also validated using out-sample series over the years 2011 and 2012.
Key Words: Count time series, non-stationarity, Poisson models, seasonal variations.

1. Introduction

Tourist arrival counts have been forecasted in the literature either using the econometric regression models or Gaussian integrated autoregressive - moving average (ARIMA) models. For example, traditional regression models for forecasting tourism data have been discussed by Bicak et al (2005) among many others. Gaussian ARIMA models have been used by Cho (2001). Seasonal ARIMA (SARIMA) models are discussed by Loganathan and Yahaya (2010) among others.
However, the tourist arrivals are count responses correlated over time in time series set up. These counts tend to be non-stationary over years and are subject to seasonal variations within a year. The general regression method may be suitable for stationary time series data but is grossly inappropriate for non-stationary time series [Kulendran and Witt (2001)]. Some advanced regression models such as error correction model (ECM), time varying parameter model (TVP), autoregressive distributed lag model (ADLM) and vector autoregressive (VAR) models have also been employed by some authors. A comparative study of these regression models is done by Li et al. (2006 b). Even these dynamic regression models are not suitable to model count time series. Similarly, Gaussian ARIMA and SARIMA models would work well only for continuous non-stationary responses. Hence, the forecasts of tourist arrivals based on these two models may not be reliable.
There exist two broad approaches to analyse time dependent count data: parameter driven and observation driven. Under the parameter driven approach used by Zeger (1988), the correlations are modeled using a latent effect. This approach, in fact produces very complicated correlation structure which is quite difficult to interpret. On the contrary, under observation-driven approach, the correlations among the repeated observations are modeled based on the assumption that the current response is a function of past responses and the correlation structure arising there from is much simpler to interpret. For count data, this approach provides integer-valued AR and MA correlation models which have been studied by Mckenzie (1988) for stationary time series of counts. Sutradhar (2003) has discussed this type of stationary correlation structures for Poisson counts in longitudinal set up and developed the method of generalized quasi-likelihood (GQL) which is highly efficient in estimating the regression parameters. Sutradhar (2008) has extended the AR(1) correlation structure to analyse non-stationary time series of negative-binomial counts which include non-stationary Poisson AR(1) model as a special case. However, these models

have so far been used to analyse small count responses where as monthly responses of tourist arrivals are large count values.

In this paper, we construct Poisson AR(1) and MA(1) models for non-stationary monthly time series of tourist arrival counts while including transcendental covariates to model seasonal variations. Methods for estimating model parameters, diagnostics and forecasting equations are briefly discussed. The analysis of time series of tourist arrivals in Mauritius over the period December 1985 to September 2012 is performed using these models and methods. Model diagnostics and validation are then done to select the best fitting model.

2. Correlation Models for Non-stationary Poisson Counts

Let $y_t \ (t = 1,\ldots,T)$ be the count response at time *t* and x_t be the *p*-dimensional vector associated with y_t, designed to accommodate trend and seasonality of the responses over time. Let $y_1 \sim P(\mu_1)$ where $\mu_1 = e^{x_1'\beta}$, β being *p*-dimensional vector of regression effects.

Poisson AR(1) model

Following Sutradhar (2008), non-stationary Poisson AR(1) model is specified as

$$y_t = \rho^* y_{t-1} + d_t \quad ; \quad t = 2,\ldots,T. \qquad (2.1)$$

where $y_{t-1} \sim P(\mu_{t-1})$ and the random errors $d_t \sim P(\mu_t - \rho\mu_{t-1})$ with $\mu_t = e^{x_t'\beta}$, d_t and y_{t-1} are independent. Also '*' denotes the binomial thinning operation such that $\rho^* y_{t-1} \mid y_{t-1}, \rho \sim B(y_{t-1}, \rho)$. Consequently,

$$E(Y_t) = V(Y_t) = \mu_t \quad ; \quad t = 1,\ldots,T. \qquad (2.2)$$

For $t = 2,\ldots,T$ and $l = 1,\ldots,T-1$,

$$Cov(Y_t, Y_{t-l}) = \rho^l \mu_{t-l} \quad ; \quad 0 < \rho < \min\left[1, \frac{\mu_t}{\mu_{t-1}}\right], \qquad (2.3)$$

Thus the l^{th} $(l = 1,\ldots,T-1)$ lag correlation, ρ_l can be calculated as

$$\rho_l = Corr(Y_t, Y_{t-l}) = \rho^l \sqrt{\frac{\mu_{t-l}}{\mu_t}} \quad \text{for } t = 2,\ldots,T. \qquad (2.4)$$

Poisson MA(1) model

Non-stationary Poisson MA(1) model equation obtained as an extension of stationary Poisson MA(1) model proposed by Sutradhar (2003) is given by

$$y_t = \rho^* d_{t-1} + d_t \ ; \quad t = 2,\ldots,T. \qquad (2.5)$$

where $d_t \sim P\left(\frac{\mu_t}{1+\rho}\right)$ with $\mu_t = e^{x_t'\beta}$, d_t and y_{t-1} are independent.

Also '*' denotes the same binomial thinning operation as in AR(1) case.

Hence,

$$E(Y_t) = \nu_t = E_{d_{t-1}} E[\rho^* d_{t-1} \mid d_{t-1}] + E[d_t] = [\rho\mu_{t-1} + \mu_t]/(1+\rho) \qquad (2.6)$$

$$\begin{aligned} Var(Y_t) &= Var_{d_{t-1}} E[\rho^* d_{t-1} \mid d_{t-1}] + E_{d_{t-1}} Var[\rho^* d_{t-1} \mid d_{t-1}] + Var[d_t] \\ &= Var_{d_{t-1}}[\rho d_{t-1}] + E_{d_{t-1}}[\rho(1-\rho) d_{t-1}] + [\mu_t/(1+\rho)] \\ &= [\rho\mu_{t-1} + \mu_t]/(1+\rho). \end{aligned} \qquad (2.7)$$

$$Cov(Y_t, Y_{t-l}) = \begin{cases} \rho\mu_{t-l} / (1+\rho) & \text{for } l = 1 \\ 0 & \text{for } l > 1 \end{cases} ; \ \max[-\mu_t / \mu_{t-1}] < \rho < 1. \qquad (2.8)$$

Consequently, the l^{th} $(l = 1, ..., T-1)$ lag correlation, ρ_l can be calculated as

$$\rho_l = \begin{cases} [\rho\mu_{t-l} / (1+\rho)] / [v_t v_{t-l}]^{1/2} & \text{for } l = 1 \\ 0 & \text{for } l > 1 \end{cases}. \qquad (2.9)$$

3. Estimation of Parameters

Regression parameter β is estimated by using GQL approach (Sutradhar, 2003) and correlation parameter ρ is estimated by the method of moments for the models provided in Section 2.

GQL estimating equation for β

Let u be the $T \times 1$ mean vector and Σ be $T \times T$ covariance matrix of y. Then GQL estimating equation is given by

$$\frac{\partial u'}{\partial \beta} \Sigma^{-1} (y - u) = 0 \qquad (2.10)$$

$\hat{\beta}$ can be obtained by solving (2.10) iteratively such that

$$\hat{\beta}_{r+1} = \hat{\beta}_r + (W'\Sigma^{-1}W)^{-1}(W'\Sigma^{-1}(y-u)). \qquad (2.11)$$

For AR(1) model, $W = MX$, u and Σ are obtained from equations (2.2) and (2.3).
Under MA(1) model, u and Σ are obtained from equations (2.6) and (2.8) and $W = AX + BZ$ where

$X = (x_1, ..., x_T)'$, $M = \text{diag}(\mu_1, \mu_2, ..., \mu_T)$,

$$Z = (1_p, x_1, ..., x_{T-1}), \ A = \text{diag}\left(\mu_1, \frac{\mu_2}{1+\rho}, ..., \frac{\mu_T}{1+\rho}\right) \ \& \ B = \text{diag}\left(0, \frac{\rho\mu_1}{1+\rho}, ..., \frac{\rho\mu_{T-1}}{1+\rho}\right).$$

As $T \to \infty$, $T^{1/2}(\hat{\beta} - \beta) \sim N\left(0, \ T(W'\Sigma^{-1}W)^{-1}\right)$.

Moment estimating equations for ρ

Under AR(1) model,

$$\hat{\rho} = \sum_{t=2}^{T} \frac{\tilde{y}_t \, \tilde{y}_{t-1}}{\tilde{y}_t^2} \frac{T}{\sum_{t=2}^{T} (u_{t-1} / u_t)^{1/2}} \quad \text{where } \tilde{y}_t = \frac{y_t - u_t}{\sqrt{u_t}} ; u_t = \mu_t. \qquad (2.12)$$

Under MA(1) model, ρ is obtained by iteratively solving the moment equation,

$$\frac{a(\rho)}{b(\rho)} = \frac{\rho}{1+\rho} c(\rho) \qquad (2.13)$$

where $u_t = v_t$ and

$$a(\rho) = \frac{1}{T-1} \sum_{t=2}^{T} \frac{(y_t - u_t)(y_{t-1} - u_{t-1})}{\sqrt{u_t} \qquad \sqrt{u_{t-1}}}, \ b(\rho) = \frac{1}{T} \sum_{t=1}^{T} \left[\frac{y_t - u_t}{\sqrt{u_t}}\right]^2, \ c(\rho) = \frac{1}{T-1} \sum_{t=2}^{T} \frac{\mu_{t-1}}{\sqrt{u_t \, u_{t-1}}}$$

Under both models, $\hat{\rho}$ must satisfy the corresponding range restrictions.

4. Model Diagnostics and Forecasting Equations

Model diagnosis is done using a portmanteau type criterion based on the fitted residuals. Let $\hat{r}_t = \dfrac{y_t - \hat{u}_t}{\hat{u}_t^{1/2}}$ be the estimated standardized residual under a given model. Then

$$\hat{\rho}_l = \frac{\sum_{t=1}^{T-l} \left(\hat{r}_t \, \hat{r}_{t+l} \right) / T - l}{\sum_{t=1}^{T} \hat{r}_t^2 / T}. \tag{2.14}$$

For a large lag K, say K = 25, we can construct a residual squared distance (RSD) such that $RSD = \sum_{l=1}^{k} \hat{\rho}_l^{\,2}$ (2.15)

where $T(RSD) \sim \chi^2_{k-1}$.

A model satisfying this criterion of goodness is deemed acceptable and in case all the models are acceptable, the model with minimum RSD is recommended for forecasting. In case the models have almost same RSD values, they are compared on the basis of Mean Absolute Deviation (MAD) and Root Mean Squared Error (RMSE) criteria. Here,

MAD $= \dfrac{1}{T}\sum_{t=1}^{T} |e_t|$ and RMSE $= \sqrt{\dfrac{1}{T}\sum_{t=1}^{T} e_t^2}$, where $e_t = y_t - \hat{y}_t$.

The model with least values of MAD and RMSE is selected for forecasting.

Forecasting Equations

Forecasting equations can be obtained by calculating expected values of future observations. Under AR(1) model forecasting equations based on equations (2.1) to (2.4) are

$$\hat{y}_{t+k} = \begin{cases} \hat{\mu}_1 & ; t = 1 \,\&\, k = 0 \\ \hat{\rho}(y_t - \hat{\mu}_t) + \hat{\mu}_{t+1} & ; k = 1 \,\&\, t = 1, 2, ..., T \\ \hat{\rho}(\hat{y}_{t+k-1} - \hat{\mu}_{t+k-1}) + \hat{\mu}_{t+k} & ; t = T \,\&\, k = 2,, 3, ... \end{cases}. \tag{2.16}$$

Under MA(1) model forecasting equations based on equations (2.5) to (2.9) are

$$\hat{y}_{t+k} = \begin{cases} \hat{\mu}_1 / (1 + \hat{\rho}) & ; t = 1 \,\&\, k = 0 \\ [\hat{\mu}_{t+k} + \hat{\rho}\hat{\mu}_{t+k-1}] / (1 + \hat{\rho}) & ; t = 1, 2, ..., T \,\&\, k = 1, 2, ... \end{cases} \tag{2.17}$$

5. Analysis of Tourist Arrival Counts in Mauritius

The monthly counts of tourist arrivals in Mauritius over the period Dec 1985 – Sep 2012 are analysed. The time series plot (Figure 1) of these data show an upward linear trend as well as seasonal variations. The monthly patterns within a year are repeated every year. The minimum number of tourist arrivals is observed during the month of June and the maximum number during December every year. Moreover, the series exhibit non-stationarity as indicated by sample auto-correlation function (acf) plot (Figure 2).

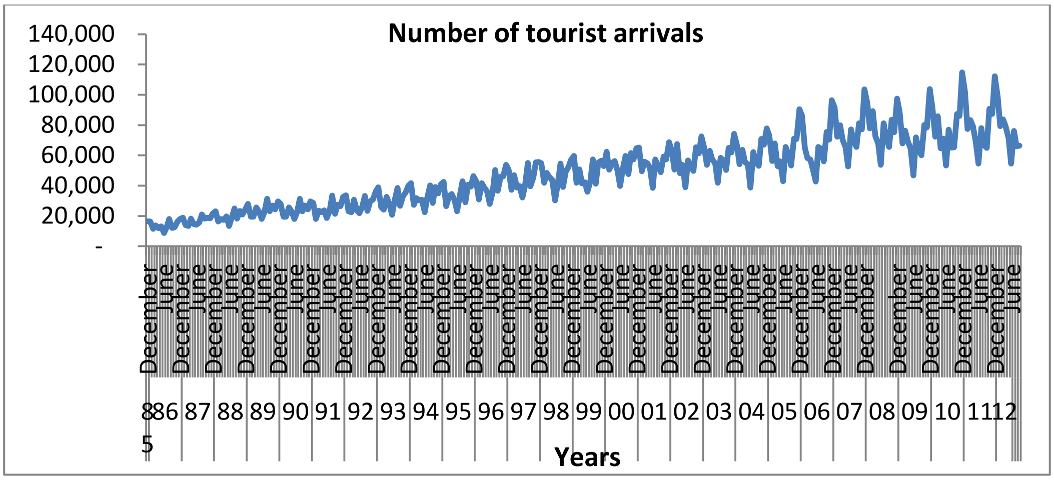

Figure 1: Monthly tourist arrivals in Mauritius from Dec 1985 to Sept 2012

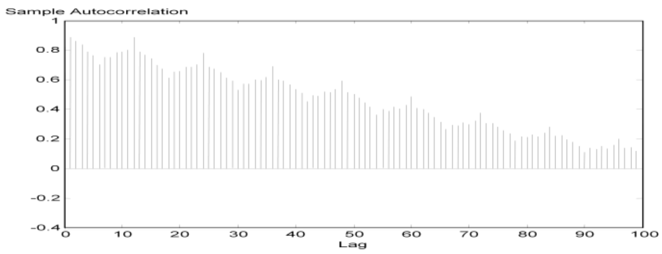

Figure2: Sample Autocorrelation Function

To analyse these series we fit the models discussed in Section 2. The covariates are so designed as to model the linear trend as well as the seasonality. To be specific, time count corresponds to the linear trend and sine and cosine pairs at annual and semi-annual frequencies correspond to the seasonal patterns. Hence,

$$x_t = \left[1,\ t,\ \cos\left(\frac{2\pi t}{12}\right),\ \sin\left(\frac{2\pi t}{12}\right),\ \cos\left(\frac{2\pi t}{6}\right),\ \sin\left(\frac{2\pi t}{6}\right)\right]'$$

The models are fitted to the time series over the period Dec1985 – Dec 2010. The series over the period Jan 2011 – Sep 2012 are used to validate the models. The estimates of the parameters resulting from the fitting of AR(1) and MA(1) models are shown in Table 1. RSD, MAD and RMSE values are also displayed in the table.

Using RSD values from Table 1, the goodness of fit statistics for AR(1) and for MA(1) models are 18.782 and 35.999 respectively both of which are lesser than the tabulated $\chi^2_{0.05}(24) = 36.415$. Hence, both models are acceptable for forecasting. However, the T(RSD) value for MA(1) model is very close to tabulated chi-squared value. Comparing on the basis of MAD and RMSE values, AR(1) model is a better fitting model and therefore should provide better forecasts as compared to MA(1) model. Next, under both models, we obtain the forecasted tourist counts over the time period Jan 2011 to Sep 2012 and compare them with observed counts in order to validate the models. Forecasts are obtained in two ways: First, we use estimated counts stepwise to forecast future values. Second, we forecast the future values by sequentially updating the models while adding an observed value at each step. The MAD and RMSE values computed from the application of AR(1) and MA(1) models provided in Table 2 confirm that the forecasts obtained under sequential AR(1) model are the best.

Table 1: Output resulting from the fitting of AR(1) and MA(1) models

	AR(1)		MA(1)	
Parameters	**Estimates**	**SE**	**Estimates**	**SE**
β_1	9.8452	0.0008	9.6386	0.0005
β_2	0.0052	0.0001	0.0061	0.0001
β_3	0.1351	0.0004	0.1019	0.0004
β_4	0.0739	0.0004	0.0230	0.0005
β_5	-0.0092	0.0004	-0.0296	0.0005
β_6	0.0281	0.0004	-0.0194	0.0004
ρ	0.1653	-	0.6047	-
RSD	0.0624	-	0.1196	-
MAD	5779.20	-	6993.00	-
RMSE	7271.06	-	9233.97	-

Table 2: MAD and RMSE values for tourist counts from Jan 2011 to Sep 2012.

Forecasting Model	MAD	RMSE
Stepwise AR(1)	15704.00	18375.62
Sequential AR(1)	11677.24	14617.16
Stepwise MA(1)	57862.00	60504.40
Sequential MA(1)	21637.19	24093.00

6. Conclusion

Two non-stationary models: Poisson AR(1) and MA(1) are developed to analyse and forecast the monthly time series of tourist counts in Mauritius recorded over past 25 years. These models take into account the linear trend, seasonality and non-stationarity of time correlated tourist counts. Models are fitted and diagnostics are performed to select the better model for forecasting. Models are also validated using out-sample data. It is concluded that Poisson AR(1) model provides better forecasts than Poisson MA(1) model.

References

1. Bicak H. A., Altinay M. & Jenkins H. (2005). “Forecasting tourism demand of North Cyprus”. *Journal of Hospitality and Leisure Marketing* 12: 87-99.
2. Cho V. (2001). “Tourism Forecasting and its relationship with leading economic indicators”. *Hospitality and Tourism Research* 25(4): 399-420.
3. Kulendren N. & Witt S. F. (2001). “Cointegration versus least squares regression”. *Annals of Tourism Research* 28: 291-311.
4. Li G., Wong K. K. F., Song H. & Witt S. F. (2006 b). “Tourism demand forecasting: A time varying parameter error correction model”. *Journal of Travel Research* 45: 175-185.
5. Loganathan N. & Yahaya I. (2010). “Forecasting international tourism demand in Malaysia using Box-Jenkins Sarima applications”. *South Asian Journal of Tourism and Heritage* 3(2): 50-60.
6. McKenzie E. (1988). “Some ARMA models for dependent sequences of Poisson counts”. *Advances in Applied Probability* 20: 822-835.
7. Sutradhar B. C. (2003). “An overview on regression models for discrete longitudinal responses”. *Statistical Science* 18: 377-393.
8. Sutradhar B. C. (2008). “On forecasting counts”. *Forecasting* 27: 109-129.
9. Zeger S. L. (1988). “A regression model for time series of counts”. *Biometrika* 75: 621-629.