A Periodic Conditional Poisson Model

Yi Zhang, V Samaranayake

Department of Mathematics, University Missouri Science and Technology, Rolla, MO 65409, USA

Abstract

A periodic version of the autoregressive conditional Poisson model (ACP), introduced by Heinen [1] in 2003 , is proposed. In the ACP model, the conditional mean of the Poisson process at a given time is assumed to follow a formulation that links it to past counts and past means. The proposed Periodic Autoregressive Conditional Poisson (PACP) model assumes that the data are generated by Poisson process whose conditional mean follows an ACP model with parameters that varies seasonally. Such models would be more appropriate when modeling count data series exhibiting conditional heteroskedastic behavior that varies from season to season. Properties of the model are investigated, and an alternative format of the model is presented to make it comparable to a vector ARMA process. A Monte Carlo simulation study, that employs the maximum likelihood method to estimate the parameters, shows an accurate estimation of the parameters with a relatively small Monte Carlo standard error. The simulation study also investigated the use of AIC and BIC criteria to differentiate between periodic and non-periodic cases with promising results. An analysis of a simulated data is used to illustrate the importance of identifying the true structure of time series count data with periodic behavior and potential for the wide uses of such models.

Key Words: count data, discrete time series, seasonality, conditional heteroscedasticity, time varying parameters

1 Introduction

Advanced data collection technologies are generating numerous time series of count data that exhibit periodic behavior. Examples of such time series range from the number of transactions per minute involving a given stock to the number of hourly clicks on a website. While traditional approaches such as Poisson regression can handle many of these time series, some count series exhibit clustering of high counts, similar to volatility clustering found in stock return series. For example, such clustering is seen in incident counts of common infectious deceases, where a high prevalence of the decease during the recent past gives rise to higher counts during the next data gathering period. Among count data models, the autoregressive conditional Poisson (ACP) model proposed by Heinen [1] allows for such behavior, specifically because the structure of the ACP formulation is very similar to the

Generalized Autoregressive Conditional Heteroscedastic (GARCH) processes that are used to model economic data with volatility clustering. In addition, the ACP model also makes it possible to analyze discrete correlated data with over-dispersion. However, the ACP model is not structured to capture periodic behavior inherent in some count data series. Thus, we proposed a generalized version of the ACP model, namely the Periodic Autoregressive Conditional Poisson (PACP) model, which could accommodate such characteristics.

2 Review of Models for Time Series Count Data

Markov chains are one way to deal with count data [2]. This method requires the definition of the (stationary) transition probabilities between all possible outcomes that the random variable could generate. However, a sequence $X_1, X_2, \dots$ of random variables taking values in some set is a Markov chain if the conditional distribution of X_{n+1}, given $X_1, X_2, \dots, X_n$, depends on X_n only, which would limit the use of this model because it ignores dependence on values prior to time *n*. Also, when the number of possible outcomes grows very large, this model is no longer easily tractable and its parameter estimation becomes cumbersome.

An alternative is to use a hidden Markov model. These models are a modified application of Markov chains as it is assumed that an underlying unobserved state of the system, determined by a Markov process, changes in time. The system's present state should determine the distribution of observations at the current time [3]. Sebastian et al. [4] developed the Markov ordinal logistic regression model with the transition probability defined as $P(Y_t|Y_{t-1}, \mathbf{Z}) = \dfrac{e^{\alpha_i - \mathbf{Z}'\beta}}{1 - e^{\alpha_i - \mathbf{Z}'\beta}}$, where Y_i represents the states of Markov chain, while Z denotes some known covariates. Cooper et al. [5] proposed a so-called 'structured hidden Markov model' for the epidemic process that intuitively follows a hidden Markov chain process since patients communicate with each other and the epidemic process usually gives out routine surveillance data that could often be partially observed. 'Structured' implies that a simple transition model is driving the underlying Markov chain. However, the need to determine the order of the Markov chain before applying the model accounts for one obvious drawback of this type of model. Another inevitable problem of such models is that the variability of the outcomes may be small.

Another branch of methods developed from the application of the Markov Chain is the binomial thinning process proposed by McKenzie [6] as a simple model to deal with discrete variate time series problems. The thinning operator takes the sum of $X_i, i = 1, 2, \dots, n$ identically independent Bernoulli random variables, each of which takes value 1 with probability α and 0 with probability $1 - \alpha$. The data generating mechanism is modeled similar as an AR process in the sense that the current count is dependent on the number of Bernoulli random variables given by the previous count. For example, Poisson AR(1) process is constructed as $X_n = \alpha * X_{n-1} + W_n$, where X_n and W_n are both Poisson process with means θ and $\theta(1 - \alpha)$ respectively. The thinning operator $*$ is defined as follows: $\alpha * X_{n-1}$ denotes the number of successes observed from X_{n-1} Bernoulli trail with success probability α. Geometric AR(1), negative binomial AR(1), binomial AR(1) and compound correlated bi-variate Poisson distribution were also proposed. They also investigated the seasonality problem in counts data, and the seasonal mean was set as $\mu_n = a\cos\omega n + b\sin\omega n$, similarly the innovation mean was set as $\omega_n = A\cos\omega n + B\sin\omega n$, where $A = a - \alpha(a\cos\omega - b\sin\omega)$ and $B = b - \alpha(a\sin\omega + b\cos\omega)$. Zhu and Joe [7] further modified the model, incorporating covariates to the mean of the stationary Markov time series allowing time varying components. Also, they extend the model and the structure to mimic an AR(2) model to solve higher order dependence structure contained in time series count data. The problem with this kind of model is that the seasonal pattern embedded in this fashion is not flexible enough to model data demonstrating complex periodic components that may require the inclusion of a large number of trigonometric functions to model the cyclical behavior with reasonable accuracy.

Many count data models are based on the use of the Poisson distribution. The Poisson regression model, as the basic count regression model, is well described in the book by Cameron and Trivedi [8]. In this approach it is assumed that y_i, an independent observation from a Poisson distribution, given the vector of regressor $\mathbf{x}_i$, has a density function $f(y_i|\mathbf{x}_i) = \dfrac{e^{-\mu_i}{\mu_i}^{y_i}}{y_i!}, y_i = 0, 1, 2, 3, \ldots$. The relationship between the mean and regressors is shown by the link function $\mu_i = exp(\mathbf{x}_i'\beta)$. It is worth noting that empirical data usually shows more variation than can be accounted by such Poisson models.

In order to account for unobserved heterogeneity and the correlation of events in the observed data, Winkelmann [9] derived several compound Poisson models, taking additional unobserved heterogeneity into consideration by letting $\lambda_i = exp(x_i\beta + \epsilon_i) = exp(x_i\beta)\mu_i$. They applied the model to labor mobility data and their results illustrate the necessity to allow for the generalizations of the standard Poisson regression model. Coxe et al. [10] provided a clear review of some appropriate regression models applicable to count data. Starting with standard Poisson regression model, two variants of Poisson regression, negative binomial regression and over-dispersed Poisson regression are gradually formed to handle the over-dispersion phenomenon. A comparison among those models, using a simulated data set of drinks consumed by university students on Saturday night [11], demonstrates the strengths and weaknesses of these models.

In order to generalize the relationship between the mean and the variance imposed by the Poisson regression models, Linden and Mantyniemi [12] utilized a negative binomial formulation. In their approach, two parameters are introduced to accommodate different 'quadratic mean–variance relationships' [12]. They expressed the probability mass function of the random variable *X* as $P(X = x|r,p) = \dfrac{\Gamma(x+r)}{x!\Gamma(r)}p^r(1-p)^x$, with the expectation (theoretical mean) $\mu = \dfrac{r(1-p)}{p}$ and variance $\sigma = \dfrac{r(1-p)}{p^2}$. Based on this setup, parameters *r* and *p* could be solved from their relationship with the mean and the variance, as $r = \dfrac{\mu^2}{\sigma^2 - \mu}$ and $p = \dfrac{1}{\omega + \theta\mu}$. The negative binomial distribution allows more flexible parameterization, which could be used to represent multiple types of over-dispersed Poisson processes. By establishing a quadratic function of the mean to describe the variance, $\sigma^2 = \omega\mu + \theta\mu^2$, diverse relationships between the mean and variance can be obtained by varying the two over-dispersion parameters ω and θ as long as the condition $\sigma^2 > \mu$ is satisfied. Scenarios where over-dispersion might happen due to factors such as sampling, environmental dissimilarity, or flocking behavior, were exemplified using bird migration data showing a high level of over-dispersion. In this study, the negative binomial distribution, using well-selected over-dispersion parameters, appropriately represented the mean-variance relationships in the considered scenarios. However, distinct assumptions about mean-variance relationships could lead to completely different coefficients which need careful identification and interpretation.

The class of discrete-valued time series models analogous to Gaussian ARMA models were advocated by Jacobs and Lewis [3]. The data generating process they assumed was a probabilistic linear combination of independently and identically distributed discrete random variables.[13] Two simple stationary processes of discrete random variables, DARMA (p, N+1) and NDARMA (p, N), whose first-order marginal distributions are arbitrarily chosen, are listed in [14]. One major drawback of the DARMA model is that even if the distribution of independent identically distributed variables is continuous, the sequence has a high density around a single value.

The integer-valued generalized autoregressive conditional heteroskedastic (INGARCH) (p,q) process was proposed by Rene [15] to model integer-valued data with Poisson deviates. In the article, important conditions for the existence of the mentioned process are discussed. When it comes to the situation $p = 1$, $q = 1$, such an integer-valued GARCH process is in essence a standard ARMA (1, 1) model. The asymptotic properties of the maximum likelihood estimates of model parameters were studied. Numbers of people infected by Campylobacterosis (a bacteria caused disease) over a certain period was analyzed by the observation driven model, and a one-step ahead forecast was also provided.

After the above INGARCH Poisson model was proposed, a negative binomial version was built on this structure. Zhu [16] developed the negative binomial integer GARCH (NBINGARCH) model and discussed some properties of it. By letting $\frac{1-p_t}{p_t} = \lambda_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i} + \sum_{j=1}^{q} \beta_j \lambda_{t-j}$, the negative binomial INGARCH could deal with problems that occur when fitting over-dispersed data by a Poisson based INGARCH model. The model could also handle potential extreme observations.

Another common feature in time series count data is the excessive zeros. The zero-inflated Poisson model [17] was introduced to deal with this type of data. The model assumes excessive zeros is the outcome from two processes: one a binary process with probability π of getting a zero and another as an ordinary Poisson process. Later, Yang [18] linked the parameters of the rate of the Poisson process and the probability of the binary process with exogenous explanatory covariates. Möller [19] extended the binomial thinning AR(1) process by adding a bounded support to accommodate different zero-inflated data types. They assumed the observation at time t is generated from two thinning processes $X_t = \alpha \circ X_{t-1} + \beta \circ (n - X_{t-1})$, where α and β are the probabilities of getting a one in a Bernoulli trial and n is the maximum value that X_t could take.

3 Proposed Periodic ACP Models

In the article by Heinen [1], an Autoregressive Conditional Poisson (ACP) model that handles count data from a Poisson distribution with an autoregressive mean was developed.

Let $\mathcal{F}_t$ represent all information available before and including time *t*. Conditioning on the past information, the count present follows a Poisson distribution with a mean μ_t related to the past,

$$N_t|\mathcal{F}_t \sim P(\mu_t),$$

where the mean has an autoregressive conditional intensity structure inspired by conditional variance in GARCH [20] model of Bollerslev [21]. In the ACP model,

$$E[N_t|\mathcal{F}_t] = \mu_t = \omega + \sum_{j=1}^{p} \alpha_j N_{t-j} + \sum_{j=1}^{q} \beta_j \mu_{t-j}, \tag{3.1}$$

under the condition that all of $\alpha_j's$,$\beta_j's$ and ω are positive.

However, there are empirical count data series that demonstrate periodic characteristics. Therefore, a periodically varying coefficient autoregressive conditional Poisson model would be more appropriate under such circumstance. Thus, we generalize the ACP Model to a periodic autoregressive conditional Poisson model(PACP), which provides more flexibility when modeling periodic count data.

To define the desired structure, let $\{N_t : t \in \mathbb{N}\}$ be the time series of interest, with N_t denoting the count at time t. We assume that t falls into one of s periods that recur in a periodic fashion, and let $s(t)$ denote the period to which t belongs. Denoting the $\sigma-$ algebra generated by $\{N_i : i \leq t\}$ as $\mathcal{F}_t$. assume that

$$N_t|\mathcal{F}_t \sim P(\lambda_t)$$

with the mean having a time varying structure

$$\lambda_t = \omega_{s(t)} + \sum_{i=1}^{q} \alpha_{is(t)} N_{t-i} + \sum_{j=1}^{p} \beta_{js(t)} \lambda_{t-j}, \tag{3.2}$$

where $\omega_s's$, $\alpha_{is(t)}'s$, $\beta_{js(t)}'s$ and $i = 1, 2, , \ldots, q$, $j = 1, 2, \ldots, p$ are positive for all values of $s(t)$. Note the $s(t)$ represents the corresponding stage of the periodic cycle at time *t*. Note that the above formulation parallels that of a periodic GARCH(*p*,*q*) process [21].

4 Some Properties of the Model

Following the derivations by Bollerslev [21], the equation 3.1 could be rewritten as

$$N_t = \epsilon_t + \omega_{s(t)} + \sum_{i=1}^{max(p,q)} (\alpha_{is(t)} + \beta_{is(t)}) N_{t-i} - \sum_{j=1}^{p} \beta_{js(t)} \epsilon_{t-j}, \tag{4.1}$$

where $\epsilon_t = N_t - \lambda_t$.

Now, model in Equation (3.2) could be interpreted as an ARMA(max(*p*,*q*),*p*) process instead of a GARCH(*p*,*q*) process. Some properties of ϵ_t are discussed below. We have

$$E(\epsilon_t) = E(N_t - \lambda_t) = E((N_t - \lambda_t | \mathcal{F}_{t-1})) = E(\lambda_t - \lambda_t) = 0,$$

and

$$\begin{aligned}
Var(\epsilon_t) &= Var\left(E(\epsilon_t|\mathcal{F}_{t-1})\right) + E\left(Var(\epsilon_t|\mathcal{F}_{t-1})\right) \\
&= Var\left(0\right) + E\left(Var(\epsilon_t|\mathcal{F}_{t-1})\right) \\
&= Var\left(0\right) + E\left(Var(N_t - \lambda_t|\mathcal{F}_{t-1})\right) \\
&= E\left(Var(N_t - \lambda_t|\mathcal{F}_{t-1})\right) \\
&= E\left(Var(N_t|\mathcal{F}_{t-1})\right) \\
&= E\left(\lambda_t\right) = \mu_t.
\end{aligned}$$

We considered two cases: (1) Periodic data with a single observation within each period and (2) Periodic data with multiple observations within each period. For the first case, a Vector ARMA form of the time series is derived analogous to [21], and this form is derived in the appendix.

5 Likelihood Function, Score, Hessian and Parameter Estimation

Let $\underline{\theta} \equiv (\omega_s, \alpha_{is(t)}, \beta_{is(t)})$ for $i = 1, \ldots, q$, $j = 1, \ldots, p$, represent all parameters in the Periodic ACP model. The conditional log-likelihood function for the model could be written as the sum of log-likelihood for each observation from different periods. Thus we have

$$l_{\underline{T}}(\underline{\theta}) = \sum_{t=1}^{T} (-\lambda_t(\underline{\theta}) + N_t \log \lambda_t(\underline{\theta}) - \log(N_t!)). \tag{5.1}$$

The corresponding score function and Hessian matrix are

$$\frac{\partial l_T}{\partial \theta} = \sum_{t=1}^{T} -\frac{\partial \lambda_t}{\partial \theta} + \frac{N_t}{\lambda_t}\left(\frac{\partial \lambda_t}{\partial \theta}\right),$$

$$\frac{\partial^2 l_T}{\partial \theta^2} = \left(-\frac{N_t}{\lambda_t^2}\right)\left(\frac{\partial \lambda_t}{\partial \theta}\right)\left(\frac{\partial \lambda_t}{\partial \theta}\right)',$$

where

$$\frac{\partial \lambda_t}{\partial \theta} = V_t' + \sum_{j=1}^{p} \beta_{js(t)} \frac{\partial \lambda_{t-i}}{\partial \theta},$$

and

$$V_t = [1, N_{t-1}, N_{t-2}, \ldots, N_{t-q}, \lambda_{t-1}, \lambda_{t-2}, \ldots, \lambda_{t-p}].$$

6 The Monte-Carlo Simulation Study

We conducted a Monte-Carlo simulation study to investigate how well the PACP model parameters are estimated by the MLE procedure. A simulation study was also performed to investigate the use of AIC and BIC criteria to differentiate between periodic and non-periodic cases.

The maximum likelihood method is utilized to estimate the parameters of PACP Model. The log likelihood function is defined as (5.1)

The properties of estimates were studied across different combinations of parameters using 3,000 simulation runs for each combination. Bias and Monte Carlo standard error were computed for each of the parameter combination. In order to eliminate the artifacts arising out of initial conditions, the first 240 time series data points were discarded.

6.1 Case of a single observation within a period

The simulated data were generated from a Periodic ACP process with 2 periods and only a single data point within each period. Two different parameter sets were used for analysis. For Table 1, the true parameter sets is $\omega_1 = 3$, $\omega_2 = 5$, $\alpha_1 = 0.1$, $\alpha_2 = 0.3$, $\beta = 0.1$, while for Table 2 the true parameter sets is $\omega_1 = 10$, $\omega_2 = 8$, $\alpha_1 = 0.25$, $\alpha_2 = 0.35$, $\beta = 0.2$. For each combination of parameter sets, sample size T=500 and T=1000 were considered. Note that the time series lengths T=500 and 1,000 are comparable to the lengths of series of day and night counts of a given phenomenon over a few years. Maximum likelihood estimation results from 3,000 simulations based on the above sample sizes with a single observation within a period are reported in Tables 1 and 2.

Table 1: Maximum likelihood estimation results from 3,000 simulations based on different sample sizes (single observation within a period); parameter Set 1.

Parameters	True Value	T=500			T=1,000		
		Estimates	Bias	SE	Estimates	Bias	SE
ω_1	3.0000	2.7783	-0.2217	0.0194	2.9019	-0.0981	0.0142
ω_2	5.0000	4.8543	-0.1457	0.0130	4.9279	-0.0721	0.0097
α_1	0.1000	0.0989	-0.0011	0.0009	0.0983	-0.0017	0.0007
α_1	0.3000	0.2985	-0.0015	0.0015	0.2990	-0.0010	0.0011
β	0.1000	0.1345	0.0345	0.0030	0.1166	0.0166	0.0021

Table 2: Maximum likelihood estimation results from 3,000 simulations based on different sample sizes (single observation within a period); parameter Set 2.

Parameters	True Value	T=500			T=1,000		
		Estimates	Bias	SE	Estimates	Bias	SE
ω_1	10.0000	10.2290	0.2290	0.0394	10.0930	0.0930	0.0285
ω_2	8.0000	8.2398	0.2398	0.0407	8.0681	0.0681	0.0298
α_1	0.2500	0.2484	-0.0016	0.0012	0.2500	0.0000	0.0008
α_2	0.3500	0.3477	-0.0023	0.0012	0.3511	0.0011	0.0008
β	0.2000	0.1889	-0.0111	0.0023	0.1949	-0.0051	0.0017

6.2 Case of multiple observations within a period

The simulated data were generated from a Periodic ACP process with 4 seasons and each season having 90 data points. Three different parameter sets were used for analysis. For Table 3, the true parameter set is $\omega_1 = 2$, $\omega_2 = 5$, $\omega_3 = 3$, $\omega_4 = 4$, $\alpha_1 = 0.1$, $\alpha_2 = 0.05$, $\alpha_3 = 0.2$, $\alpha_4 = 0.2$, $\beta = 0.1$. For Table 4 the true parameter set is $\omega_1 = 10$, $\omega_2 = 7$, $\omega_3 = 5$, $\omega_4 = 12$, $\alpha_1 = 0.2$, $\alpha_2 = 0.3$, $\alpha_3 = 0.1$, $\alpha_4 = 0.3$, $\beta = 0.3$. For Table 5 the true parameter set is $\omega_1 = 4$, $\omega_2 = 6$, $\omega_3 = 5$, $\omega_4 = 4$, $\alpha_1 = 0.1$, $\alpha_2 = 0.2$, $\alpha_3 = 0.1$, $\alpha_4 = 0.2$, $\beta = 0.4$. For each combination of parameter sets, sample size T=540 and T=1,080 were considered. These are approximately equivalent to the numbers of observations for a 1.5-year and 3-year daily count data respectively. Maximum likelihood estimation results from 3,000 simulations based on different sample sizes are reported in Table 3-5.

Table 3: Maximum likelihood estimation results from 3,000 simulations based on different sample sizes (multiple observations within a period); parameter Set 1.

Parameters	True Value	T=540			T=1,080		
		Estimates	Bias	SE	Estimates	Bias	SE
ω_1	2	1.974	-0.026	0.007	1.991	-0.009	0.006
ω_2	5	4.87	-0.13	0.017	4.939	-0.061	0.013
ω_3	3	3.008	0.008	0.013	3.001	0.001	0.01
ω_4	4	4.018	0.018	0.017	3.997	-0.003	0.013
α_1	0.1	0.093	-0.007	0.001	0.096	-0.004	0.001
α_2	0.05	0.056	0.006	0.001	0.053	0.003	0.001
α_3	0.2	0.181	-0.019	0.002	0.192	-0.008	0.001
α_4	0.2	0.181	-0.019	0.002	0.193	-0.007	0.001
β	0.1	0.117	0.017	0.003	0.107	0.007	0.002

Table 4: Maximum likelihood estimation results from 3,000 simulations based on different sample sizes (multiple observations within a period); parameter Set 2.

Parameters	True Value	T=540			T=1,080		
		Estimates	Bias	SE	Estimates	Bias	SE
ω_1	10.000	10.651	0.651	0.043	10.338	0.338	0.030
ω_2	7.000	7.597	0.597	0.036	7.299	0.299	0.025
ω_3	5.000	5.250	0.250	0.020	5.120	0.120	0.013
ω_4	12.000	13.103	1.103	0.058	12.400	0.400	0.033
α_1	0.200	0.190	-0.010	0.001	0.191	-0.009	0.001
α_2	0.300	0.288	-0.012	0.001	0.292	-0.008	0.001
α_3	0.100	0.093	-0.007	0.001	0.094	-0.006	0.001
α_4	0.300	0.284	-0.016	0.002	0.294	-0.006	0.001
β	0.300	0.278	-0.022	0.002	0.292	-0.008	0.001

Table 5: Maximum likelihood estimation results from 3,000 simulations based on different sample sizes (multiple observations within a period); parameter Set 3.

Parameters	True Value	T=540			T=1,080		
		Estimates	Bias	SE	Estimates	Bias	SE
ω_1	4.000	4.476	0.476	0.025	4.264	0.264	0.018
ω_2	6.000	6.856	0.856	0.042	6.456	0.456	0.029
ω_3	5.000	5.654	0.654	0.033	5.327	0.327	0.022
ω_4	4.000	4.738	0.738	0.033	4.343	0.343	0.022
α_1	0.100	0.093	-0.007	0.001	0.095	-0.005	0.001
α_2	0.200	0.194	-0.006	0.001	0.197	-0.003	0.001
α_3	0.100	0.088	-0.012	0.001	0.095	-0.005	0.001
α_4	0.200	0.178	-0.022	0.002	0.193	-0.007	0.001
β	0.400	0.347	-0.053	0.003	0.373	-0.027	0.002

From the simulation results, it is clear that the maximum likelihood estimate method gives a relatively small bias (10%) and low Monte Carlo standard error. Note that the estimation bias is smaller and standard error is lower for the larger sample size, in both the single observation per period and multiple observations per period cases. This demonstrates the MLE is a viable method for estimating the parameters of the suggested Periodic Autoregressive Poisson Model and that larger sample sizes produce more accurate estimates.

7 Model Selection

To examine whether AIC and/or BIC are a good criteria to distinguish the true structure of the count data, a small scale Monte Carlo simulation study was performed. All statistics reported here are calculated from N=3,000 replications and each replication having sample size T=1,080. In order to avoid artifacts created by initial conditions, the first 360 time series data points were discarded.

Mean AIC is averaged from AIC values for each of the replications and the percentage in the brackets indicates the proportion of simulation runs that yielded a smaller AIC value for the corresponding model.

Table 6 shows results for the case when the data were generated from an ACP process with true parameters $\omega = 2, \alpha = 0.1, \beta = 0.15$. Both ACP and Periodic ACP Model were filled to the data. The AIC values for ACP Model are lower than those for the PACP model 94 out of 100 times, which suggests AIC performs well in identifying the true structure of the time series. BIC does not work in this situation since Periodic ACP model gives out same estimates for each season as the ACP model. AIC put more penalty on large parameters sets than BIC, thus it has better perfomance than BIC.

Table 6: ACP and PACP model selection by AIC criteria with simulated time series data with ACP as the data generating process.

Parameters	ACP Model	Periodic ACP Model
ω_1	1.8626	1.9312
ω_2	-	1.9271
ω_3	-	1.9307
ω_4	-	1.9337
α_1	0.099266	0.095937
α_2	-	0.09726
α_3	-	0.096276
α_4	-	0.095989
β	0.15582	0.1314
Mean AIC	3955.77(94.233%)	3961.795(5.767%)

Table 7 shows results when a Periodic ACP process with true parameters $\omega_1 = 2, \omega_2 = 5, \omega_3 = 3, \omega_4 = 4, \alpha_1 = 0.1, \alpha_2 = 0.05, \alpha_3 = 0.2, \alpha_4 = 0.2, \beta = 0.1$ is the underlying structure producing the count data. Both ACP model and Periodic ACP model were used to fit the data. In this case, AIC and BIC all show their strong ability to select the right structure. Notice that when there is periodicity in the count data, an ordinary ACP model gives out estimates of α and β with their sum close to one, suggesting near non-stationarity, raising questions about the appropriateness of the model.

Table 7: ACP and PACP model selection by AIC and BIC criteria with simulated time series data generated under a PACP model.

Parameters	ACP Model	Periodic ACP Model
ω_1	0.15944	1.9978
ω_2	-	4.9635
ω_3	-	3.005
ω_4	-	3.9987
α_1	0.15236	0.094843
α_2	-	0.051539
α_3	-	0.19369
α_4	-	0.19514
β	0.81353	0.10489
Mean AIC	4726.275 (0%)	4605.73 (100%)
Mean BIC	4783.136 (0%)	4650.593 (100%)

8 Visualization of simulated data and estimated intensity process

The application of the proposed PACP model is demonstrated using a simulated data set. The data is generated from a Periodic ACP process with 4 seasons and each season has 90 data points. The parameter set is $\omega_1 = 2$, $\omega_2 = 5$, $\omega_3 = 3$, $\omega_4 = 4$, $\alpha_1 = 0.1$, $\alpha_2 = 0.05$, $\alpha_3 = 0.2$, $\alpha_4 = 0.2$, $\beta = 0.1$.

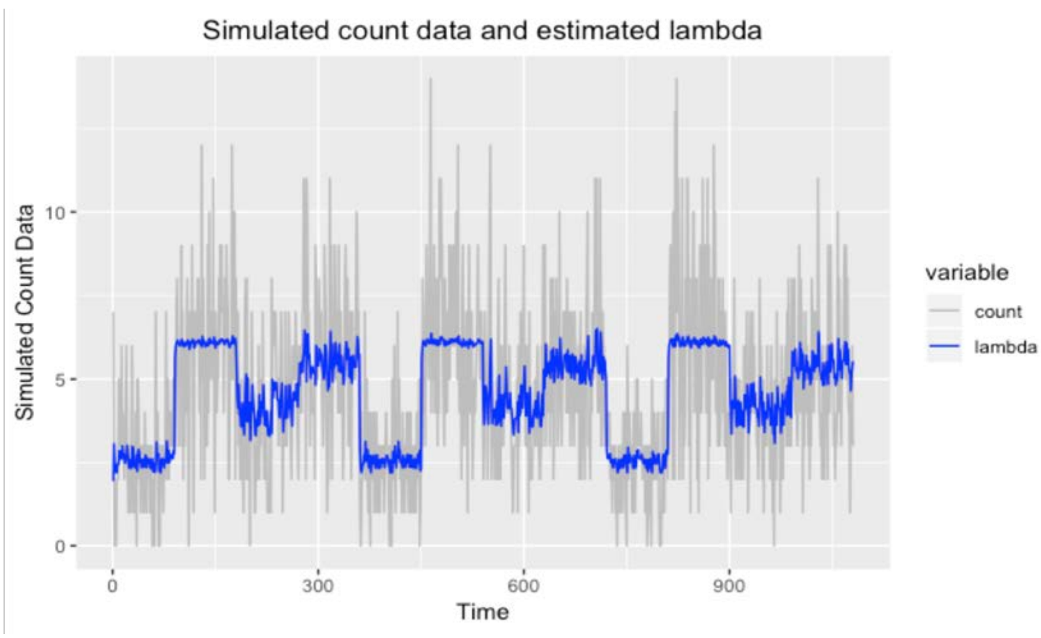

Figure 1: Simulated time series count data and the estimated intensity λ_t process

As shown in Figure 1, the grey line represents the simulated data while the blue line indicates the estimated underlying periodic process. The results shows the PACP model captures the cyclical movement of the process well.

9 Conclusion

The model provided here is a natural generalization of the Autoregressive Conditional Periodic Poisson model, which allows periodicity to be taken into consideration when modeling count time series. The reported simulation results in Section 6 show that the MLE provides reasonable estimates of the model parameters of the PACP model. We also studied the utility of using AIC and BIC criteria in determining if the underlying data generating process is ACP or PACP. Results suggest that the use of AIC criteria is a trustworthy way to differentiate between the underlying ACP or PACP structure. In addition, the simulated data illustrates there is indeed the necessity to generalize the original ACP model to accommodate periodic component in the count data.

References

[1] A. Heinen, “Modelling time series count data: an autoregressive conditional poisson model,” 2003.

[2] S. Chib and R. Winkelmann, “Markov chain monte carlo analysis of correlated count data,” *Journal of Business & Economic Statistics*, vol. 19, no. 4, pp. 428–435, 2001.

[3] I. L. MacDonald and W. Zucchini, *Hidden Markov and other models for discrete-valued time series*, vol. 110. CRC Press, 1997.

[4] T. Sebastian, V. Jeyaseelan, L. Jeyaseelan, S. Anandan, S. George, and S. I. Bangdiwala, “Decoding and modelling of time series count data using poisson hidden markov model and markov ordinal logistic regression models,” *Statistical methods in medical research*, vol. 28, no. 5, pp. 1552–1563, 2019.

[5] B. Cooper and M. Lipsitch, “The analysis of hospital infection data using hidden markov models,” *Biostatistics*, vol. 5, no. 2, pp. 223–237, 2004.

[6] E. McKenzie, “Some simple models for discrete variate time series 1,” *JAWRA Journal of the American Water Resources Association*, vol. 21, no. 4, pp. 645–650, 1985.

[7] R. Zhu and H. Joe, “Modelling count data time series with markov processes based on binomial thinning,” *Journal of Time Series Analysis*, vol. 27, no. 5, pp. 725–738, 2006.

[8] A. C. Cameron and P. K. Trivedi, *Regression analysis of count data*, vol. 53. Cambridge university press, 1998.

[9] R. Winkelmann and K. F. Zimmermann, “Recent developments in count data modelling: theory and application,” *Journal of economic surveys*, vol. 9, no. 1, pp. 1–24, 1995.

[10] S. Coxe, S. G. West, and L. S. Aiken, “The analysis of count data: A gentle introduction to poisson regression and its alternatives,” *Journal of personality assessment*, vol. 91, no. 2, pp. 121–136, 2009.

[11] T. DeHart, H. Tennen, S. Armeli, M. Todd, and G. Affleck, “Drinking to regulate negative romantic relationship interactions: The moderating role of self-esteem,” *Journal of Experimental Social Psychology*, vol. 44, no. 3, pp. 527–538, 2008.

[12] A. Lindén and S. Mäntyniemi, “Using the negative binomial distribution to model overdispersion in ecological count data,” *Ecology*, vol. 92, no. 7, pp. 1414–1421, 2011.

[13] P. A. Jacobs and P. A. Lewis, “Discrete time series generated by mixtures. i: Correlational and runs properties,” *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 40, no. 1, pp. 94–105, 1978.

[14] P. A. Jacobs and P. A. Lewis, “Stationary discrete autoregressive-moving average time series generated by mixtures,” *Journal of Time Series Analysis*, vol. 4, no. 1, pp. 19–36, 1983.

[15] R. Ferland, A. Latour, and D. Oraichi, "Integer-valued garch process," *Journal of Time Series Analysis*, vol. 27, no. 6, pp. 923–942, 2006.

[16] F. Zhu, "A negative binomial integer-valued garch model," *Journal of Time Series Analysis*, vol. 32, no. 1, pp. 54–67, 2011.

[17] D. Lambert, "Zero-inflated poisson regression, with an application to defects in manufacturing," *Technometrics*, vol. 34, no. 1, pp. 1–14, 1992.

[18] M. Yang, G. K. Zamba, and J. E. Cavanaugh, "Markov regression models for count time series with excess zeros: A partial likelihood approach," *Statistical Methodology*, vol. 14, pp. 26–38, 2013.

[19] T. A. Möller, C. H. Weiß, H.-Y. Kim, and A. Sirchenko, "Modeling zero inflation in count data time series with bounded support," *Methodology and Computing in Applied Probability*, vol. 20, no. 2, pp. 589–609, 2018.

[20] T. Bollerslev, "Generalized autoregressive conditional heteroskedasticity," *Journal of econometrics*, vol. 31, no. 3, pp. 307–327, 1986.

[21] T. Bollerslev and E. Ghysels, "Periodic autoregressive conditional heteroscedasticity," *Journal of Business & Economic Statistics*, vol. 14, no. 2, pp. 139–151, 1996.

A Appendix

A.1 ACP extension

In this section, we derive the reformulation of the original set up equation for our model to an integer valued seasonal ARMA(maxp,q,p) process. The count data has a stochastic process satisfies

$$\lambda_t = \omega_{s(t)} + \sum_{i=1}^{q} \alpha_{is(t)} N_{t-i} + \sum_{j=1}^{p} \beta_{js(t)} \lambda_{t-j}.$$

Equivalently, it could be rewritten as an integer valued ARMA(maxp,q,p) model

(i) Case1 $p \leq q$.

$$N_t + \lambda_t = N_t + \omega_{s(t)} + \sum_{i=1}^{q} \alpha_{is(t)} N_{t-i} + \sum_{j=1}^{p} \beta_{js(t)} \lambda_{t-j},$$

$$N_t = N_t - \lambda_t + \omega_{s(t)} + \sum_{i=1}^{q} \alpha_{is(t)} N_{t-i} + \sum_{j=1}^{p} \beta_{js(t)} \lambda_{t-j},$$

$$N_t = N_t - \lambda_t + \omega_{s(t)} + \sum_{i=1}^{q} \alpha_{is(t)} N_{t-i} + \sum_{j=1}^{p} \beta_{js(t)} N_{t-j} - \sum_{j=1}^{p} \beta_{js(t)} N_{t-j} + \sum_{j=1}^{p} \beta_{js(t)} \lambda_{t-j},$$

$$N_t = N_t - \lambda_t + \omega_{s(t)} + \sum_{i=1}^{q} \alpha_{is(t)} N_{t-i} + \sum_{j=1}^{p} \beta_{js(t)} N_{t-j} - \left(\sum_{j=1}^{p} \beta_{js(t)} N_{t-j} - \sum_{j=1}^{p} \beta_{js(t)} \lambda_{t-j} \right),$$

$$N_t = \underbrace{N_t - \lambda_t}_{\epsilon_t} + \omega_{s(t)} + \sum_{i=1}^{q} \alpha_{is(t)} N_{t-i} + \sum_{j=1}^{p} \beta_{js(t)} N_{t-j} - \underbrace{\left(\sum_{j=1}^{p} \beta_{js(t)} N_{t-j} - \sum_{j=1}^{p} \beta_{js(t)} \lambda_{t-j} \right)}_{\sum_{j=1}^{p} \beta_{js(t)} \epsilon_{t-j}},$$

$$N_t = \epsilon_t + \omega_{s(t)} + \sum_{i=1}^{q} \alpha_{is(t)} N_{t-i} + \sum_{j=1}^{p} \beta_{js(t)} N_{t-j} - \sum_{j=1}^{p} \beta_{js(t)} \epsilon_{t-j},$$

$$N_t = \epsilon_t + \omega_{s(t)} + \sum_{i=1}^{q} \alpha_{is(t)} N_{t-i} + \sum_{j=1}^{p} \beta_{js(t)} N_{t-j} - \sum_{j=1}^{p} \beta_{js(t)} \epsilon_{t-j}.$$

Choose $\beta_{js(t)} \equiv 0$ for $j > p$, then,

$$N_t = \epsilon_t + \omega_{s(t)} + \sum_{i=1}^{q} \alpha_{is(t)} N_{t-i} + \sum_{j=1}^{p} \beta_{js(t)} N_{t-j} - \sum_{j=1}^{p} \beta_{js(t)} \epsilon_{t-j},$$

$$N_t = \epsilon_t + \omega_{s(t)} + \sum_{i=1}^{q} \alpha_{is(t)} N_{t-i} + \sum_{j=1}^{q} \beta_{js(t)} N_{t-j} - \sum_{j=1}^{p} \beta_{js(t)} \epsilon_{t-j} \;\mid\; \beta_{js(t)} \equiv 0, \quad \forall p \leq q,$$

$$N_t = \epsilon_t + \omega_{s(t)} + \sum_{i=1}^{q}(\alpha_{is(t)} + \beta_{is(t)})N_{t-i} - \sum_{j=1}^{p}\beta_{js(t)}\epsilon_{t-j},$$

$$N_t = \epsilon_t + \omega_{s(t)} + \sum_{i=1}^{max(p,q)}(\alpha_{is(t)} + \beta_{is(t)})N_{t-i} - \sum_{j=1}^{p}\beta_{js(t)}\epsilon_{t-j}.$$

(2) Case2 $p > q$. Let

$$N_t + \lambda_t = N_t + \omega_{s(t)} + \sum_{i=1}^{q}\alpha_{is(t)}N_{t-i} + \sum_{j=1}^{p}\beta_{js(t)}\lambda_{t-j},$$

$$N_t = N_t - \lambda_t + \omega_{s(t)} + \sum_{i=1}^{q}\alpha_{is(t)}N_{t-i} + \sum_{j=1}^{p}\beta_{js(t)}\lambda_{t-j},$$

$$N_t = N_t - \lambda_t + \omega_{s(t)} + \sum_{i=1}^{q}\alpha_{is(t)}N_{t-i} + \sum_{j=1}^{p}\beta_{js(t)}N_{t-j} - \sum_{j=1}^{p}\beta_{js(t)}N_{t-j} + \sum_{j=1}^{p}\beta_{js(t)}\lambda_{t-j},$$

$$N_t = N_t - \lambda_t + \omega_{s(t)} + \sum_{i=1}^{q}\alpha_{is(t)}N_{t-i} + \sum_{j=1}^{p}\beta_{js(t)}N_{t-j} - \left(\sum_{j=1}^{p}\beta_{js(t)}N_{t-j} - \sum_{j=1}^{p}\beta_{js(t)}\lambda_{t-j}\right),$$

$$N_t = \underbrace{N_t - \lambda_t}_{\epsilon_t} + \omega_{s(t)} + \sum_{i=1}^{q}\alpha_{is(t)}N_{t-i} + \sum_{j=1}^{p}\beta_{js(t)}N_{t-j} - \underbrace{\left(\sum_{j=1}^{p}\beta_{js(t)}N_{t-j} - \sum_{j=1}^{p}\beta_{js(t)}\lambda_{t-j}\right)}_{\sum_{j=1}^{p}\beta_{js(t)}\epsilon_{t-j}},$$

$$N_t = \epsilon_t + \omega_{s(t)} + \sum_{i=1}^{q}\alpha_{is(t)}N_{t-i} + \sum_{j=1}^{p}\beta_{js(t)}N_{t-j} - \sum_{j=1}^{p}\beta_{js(t)}\epsilon_{t-j},$$

$$N_t = \epsilon_t + \omega_{s(t)} + \sum_{i=1}^{q}\alpha_{is(t)}N_{t-i} + \sum_{j=1}^{p}\beta_{js(t)}N_{t-j} - \sum_{j=1}^{p}\beta_{js(t)}\epsilon_{t-j}.$$

Now choose $\alpha_{is(t)} \equiv 0$ for $i > q$. Then,

$$N_t = \epsilon_t + \omega_{s(t)} + \sum_{i=1}^{q}\alpha_{is(t)}N_{t-i} + \sum_{j=1}^{p}\beta_{js(t)}N_{t-j} - \sum_{j=1}^{p}\beta_{js(t)}\epsilon_{t-j},$$

$$N_t = \epsilon_t + \omega_{s(t)} + \sum_{i=1}^{p}\alpha_{is(t)}N_{t-i} + \sum_{j=1}^{p}\beta_{js(t)}N_{t-j} - \sum_{j=1}^{p}\beta_{js(t)}\epsilon_{t-j} \mid \alpha_{is(t)} \equiv 0, \quad \forall i \geq q,$$

$$N_t = \epsilon_t + \omega_{s(t)} + \sum_{i=1}^{p}(\alpha_{is(t)} + \beta_{is(t)})N_{t-i} - \sum_{j=1}^{p}\beta_{js(t)}\epsilon_{t-j},$$

$$N_t = \epsilon_t + \omega_{s(t)} + \sum_{i=1}^{max(p,q)}(\alpha_{is(t)} + \beta_{is(t)})N_{t-i} - \sum_{j=1}^{p}\beta_{js(t)}\epsilon_{t-j}.$$

For both cases, they all could be rewritten as an ARMA(max(p,q),p) process.

A.2 Develop VARMA Form (Four seasons without repeated observation)

In this section, we derive the Vector ARMA form of the periodic count time series from the integer valued seasonal ARMA(maxp,q,p) process. Note that we can write

$$\begin{bmatrix}1 & -(\alpha_{14}+\beta_{14}) & 0 & 0\\ 0 & 1 & -(\alpha_{13}+\beta_{13}) & 0\\ 0 & 0 & 1 & -(\alpha_{12}+\beta_{12})\\ 0 & 0 & 0 & 1\end{bmatrix}\begin{bmatrix}N_{4\tau}\\ N_{4\tau-1}\\ N_{4\tau-2}\\ N_{4\tau-3}\end{bmatrix}=\begin{bmatrix}\omega_4\\ \omega_3\\ \omega_2\\ \omega_1\end{bmatrix}+$$

$$\begin{bmatrix}0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ (\alpha_{11}+\beta_{11}) & 0 & 0 & 0\end{bmatrix}\begin{bmatrix}N_{4(\tau-1)}\\ N_{4(\tau-1)-1}\\ N_{4(\tau-1)-2}\\ N_{4(\tau-1)-3}\end{bmatrix}+\begin{bmatrix}1 & -\beta_{14} & 0 & 0\\ 0 & 1 & -\beta_{13} & 0\\ 0 & 0 & 1 & -\beta_{12}\\ 0 & 0 & 0 & 1\end{bmatrix}\begin{bmatrix}\epsilon_{4\tau}\\ \epsilon_{4\tau-1}\\ \epsilon_{4\tau-2}\\ \epsilon_{4\tau-3}\end{bmatrix}+\begin{bmatrix}0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ -\beta_{11} & 0 & 0 & 0\end{bmatrix}\begin{bmatrix}\epsilon_{4(\tau-1)}\\ \epsilon_{4(\tau-1)-1}\\ \epsilon_{4(\tau-1)-2}\\ \epsilon_{4(\tau-1)-3}\end{bmatrix},$$

Multiplying both sides by the inverse matrix given below,

$$\begin{bmatrix}1 & \alpha_{14}+\beta_{14} & (\alpha_{13}+\beta_{13})(\alpha_{14}+\beta_{14}) & (\alpha_{12}+\beta_{12})(\alpha_{13}+\beta_{13})(\alpha_{14}+\beta_{14})\\ 0 & 1 & \alpha_{13}+\beta_{13} & (\alpha_{12}+\beta_{12})(\alpha_{13}+\beta_{13})\\ 0 & 0 & 1 & \alpha_{12}+\beta_{12}\\ 0 & 0 & 0 & 1\end{bmatrix}.$$

We obtain the Vector ARMA form:

$$\begin{bmatrix}N_{4\tau}\\ N_{4\tau-1}\\ N_{4\tau-2}\\ N_{4\tau-3}\end{bmatrix}=\begin{bmatrix}\omega_1(\alpha_{12}+\beta_{12})(\alpha_{13}+\beta_{13})(\alpha_{14}+\beta_{14})+\omega_2(\alpha_{13}+\beta_{13})(\alpha_{14}+\beta_{14})+\omega_3(\alpha_{14}+\beta_{14})+\omega_4\\ \omega_1(\alpha_{12}+\beta_{12})(\alpha_{13}+\beta_{13})+\omega_2(\alpha_{13}+\beta_{13})+\omega_3\\ \omega_1(\alpha_{12}+\beta_{12})+\omega_2\\ \omega_1\end{bmatrix}$$

$$+\begin{bmatrix}(\alpha_{11}+\beta_{11})(\alpha_{12}+\beta_{12})(\alpha_{13}+\beta_{13})(\alpha_{14}+\beta_{14}) & 0 & 0 & 0\\ (\alpha_{11}+\beta_{11})(\alpha_{12}+\beta_{12})(\alpha_{13}+\beta_{13}) & 0 & 0 & 0\\ (\alpha_{11}+\beta_{11})(\alpha_{12}+\beta_{12}) & 0 & 0 & 0\\ \alpha_{11}+\beta_{11} & 0 & 0 & 0\end{bmatrix}\begin{bmatrix}N_{4(\tau-1)}\\ N_{4(\tau-1)-1}\\ N_{4(\tau-1)-2}\\ N_{4(\tau-1)-3}\end{bmatrix}$$

$$+\begin{bmatrix}1 & \alpha_{14} & \alpha_{13}(\alpha_{14}+\beta_{14}) & \alpha_{12}(\alpha_{13}+\beta_{13})(\alpha_{14}+\beta_{14})\\ 0 & 1 & \alpha_{13} & \alpha_{12}(\alpha_{13}+\beta_{13})\\ 0 & 0 & 1 & \alpha_{12}\\ 0 & 0 & 0 & 1\end{bmatrix}\begin{bmatrix}\epsilon_{4\tau}\\ \epsilon_{4\tau-1}\\ \epsilon_{4\tau-2}\\ \epsilon_{4\tau-3}\end{bmatrix}$$

$$+\begin{bmatrix}-\beta_{11}(\alpha_{12}+\beta_{12})(\alpha_{13}+\beta_{13})(\alpha_{14}+\beta_{14}) & 0 & 0 & 0\\ -\beta_{11}(\alpha_{12}+\beta_{12})(\alpha_{13}+\beta_{13}) & 0 & 0 & 0\\ -\beta_{11}(\alpha_{12}+\beta_{12}) & 0 & 0 & 0\\ -\beta_{11} & 0 & 0 & 0\end{bmatrix}\begin{bmatrix}\epsilon_{4(\tau-1)}\\ \epsilon_{4(\tau-1)-1}\\ \epsilon_{4(\tau-1)-2}\\ \epsilon_{4(\tau-1)-3}\end{bmatrix}.$$

Now take the expectation of both sides and assume that the observations from the same period have the same expected value, to get the expectation of each observation from each season:

$$\begin{cases} E(\lambda_1) = \frac{\omega 1+\omega 2(\alpha_{14}+\beta_{14})+\omega 3(\alpha_{13}+\beta_{13})(\alpha_{14}+\beta_{14})+\omega 4(\alpha_{12}+\beta_{12})(\alpha_{13}+\beta_{13})(\alpha_{14}+\beta_{14})}{1-(\alpha_{11}+\beta_{11})(\alpha_{12}+\beta_{12})(\alpha_{13}+\beta_{13})(\alpha_{14}+\beta_{14})} \\ E(\lambda_2) = \frac{\omega 1(\alpha_{11}+\beta_{11})(\alpha_{12}+\beta_{12})(\alpha_{13}+\beta_{13})+\omega 2+\omega 3(\alpha_{13}+\beta_{13})+\omega 4(\alpha_{12}+\beta_{12})(\alpha_{13}+\beta_{13})}{1-(\alpha_{11}+\beta_{11})(\alpha_{12}+\beta_{12})(\alpha_{13}+\beta_{13})(\alpha_{14}+\beta_{14})} \\ E(\lambda_3) = \frac{\omega 1(\alpha_{11}+\beta_{11})(\alpha_{12}+\beta_{12})+\omega 2(\alpha_{11}+\beta_{11})(\alpha_{12}+\beta_{12})(\alpha_{14}+\beta_{14})+\omega 3+\omega 4(\alpha_{12}+\beta_{12})}{1-(\alpha_{11}+\beta_{11})(\alpha_{12}+\beta_{12})(\alpha_{13}+\beta_{13})(\alpha_{14}+\beta_{14})} \\ E(\lambda_4) = \frac{\omega 1(\alpha_{11}+\beta_{11})+\omega 2(\alpha_{11}+\beta_{11})(\alpha_{14}+\beta_{14})+\omega 3(\alpha_{11}+\beta_{11})(\alpha_{13}+\beta_{13})(\alpha_{14}+\beta_{14})+\omega 4}{1-(\alpha_{11}+\beta_{11})(\alpha_{12}+\beta_{12})(\alpha_{13}+\beta_{13})(\alpha_{14}+\beta_{14})} \end{cases}$$

Now let $a_{11} + b_{11} = A$, $a_{12} + b_{12} = B$, $a_{13} + b_{13} = C$ and $a_{14} + b_{14} = D$, the expectation of observation of each season could be rewritten as:

$$\begin{cases} E(\lambda_1) : \frac{\omega_1+D\omega_2+CD\omega_3+BCD\omega_4}{1-ABCD} \\ E(\lambda_2) : \frac{ABC\omega_1+\omega_2+C\omega_3+BC\omega_4}{1-ABCD} \\ E(\lambda_3) : \frac{AB\omega_1+ABD\omega_2+\omega_3+B\omega_4}{1-ABCD} \\ E(\lambda_4) : \frac{A\omega_1+AD\omega_2+ACD\omega_3+\omega_4}{1-ABCD} \end{cases}.$$