# Autoregressive Conditional Heteroscedastic Hidden Markov Model

Yi Zhang, V Samaranayake

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

**Abstract**

With the rapid development of sensors and other data-gathering devices, high-frequency time series of count data have become common. Such series commonly exhibit conditional dependence of the parameters of the data generating process (DGP) to past values of the counts and parameter values. The Autoregressive Conditional Poisson (ACP) formulation is one model developed to describe the underlying data generating mechanism of such processes. In ACP Models, the mean of the Poisson process is assumed to be a linear function of past means and past counts through a GARCH type model. In this formulation, it is assumed that the parameters of the model that connects the conditional mean to past values remain constant over time. One generalization is to accommodate seasonal variations in one or more of these parameters, but in some empirical processes, the changes in the parameters may not occur systematically but according to a latent process. The proposed model addresses such a scenario where the Poisson intensity is modeled using an ARCH type formulation with select parameters taking different values based on the state defined by a hidden Markov chain. The application of the proposed model is illustrated using a synthetic and a real-life data set.

**Key Words:** count data, discrete time series, regime change, conditional heteroscedasticity, time-varying parameters

## 1 Introduction

High-frequency count data time series have become ubiquitous in many fields due to the rapid development of sensor and information gathering and storage capabilities. Many of these time series show temporal dependence, and some show patterns that may signal changes in the underlying data-generating mechanism. In other words, such series exhibit "regime" changes in the underlying data generating process. Examples of such series could be found in diverse areas of applications such as epidemiology and finance, and the effective modeling of such processes can provide valuable insight into the core mechanisms generating the counts. The Poisson Hidden Markov Model (P-HMM) is one of the formulations researchers commonly utilize to model count data processes that show possible shifts in the underlying data generating process. It was first developed to model time series of epileptic seizure counts (Albert, 1991), but has been used in many other empirical situations. This model assumes that a hidden Markov process determines the parameters of the Poisson process that generates the count data. Since within a given state, the counts generated from the Poisson process are assumed to be independent, the consistent serial dependence seen in some empirical series cannot be modeled adequately by a P-HMM. Thus, an autoregressive conditional Poisson hidden Markov model (ACP-HMM) is proposed to accommodate the serial dependence, the clustering of high and low counts, and at the same time, account for possible shifts in the underlying data generating process. It could be seen as a combination of a P-HMM model and an autoregressive structure, which admits the existence of several underlying mechanisms that switch back and forth while at the same time capturing the strong correlation among time-series observations. It also provides a way to model the clustering of high counts, using a formulation similar to the generalized conditional autoregressive heteroscedastic model proposed by Bollerslev (Bollerslev, 1986).

## 2 Review of Models for Time Series Count Data

The Poisson distribution is commonly employed to model count data. Poisson regression extends the utility of this distribution by allowing its mean to depend on exogenous variables (Consul & Famoye, 1992; Coxe, West, & Aiken, 2009; Hilbe, 2014). This is achieved via a function that links the Poisson mean to a linear combination of the collection of exogenous variables. Such regression models, however, are not able to handle data with over-dispersion unless additional parameters are included. More importantly, Poisson regression models do not accommodate serial dependence when applied to time series data, unless one or more of the exogenous variables are themselves serially dependent.

One class of models that are widely used to model correlated time series is the Autoregressive Moving Average (ARMA) models and their analogous discrete version: discrete autoregressive moving average (DARMA) models (Jacobs & Lewis, 1983), which are proposed for count data with arbitrarily chosen marginal distributions. By coupling two simple stationary processes, the discrete autoregressive of order $p$ (DAR($p$)) and the discrete moving average of order $q$ (DMA($q$)), the DARMA($p, q$) can be generated as a mixture of the two and can be further simplified into a single equation as the NDARMA($p, q$) (see Jacobs & Lewis, 1983). The problem of the DARMA model is that a single value might have a high density around it when the sequence is generated from such structures. They also do not allow for negative autocorrelations. McKenzie (1985, 1986, 1987, 1988) and Al-Osh and Alzaid (1987) developed a series of models that handle dependent sequences of Poisson counts for equally spaced count data time series by taking advantage of the ARMA process. One set of these models is the integer-valued autoregressive (INAR) Poisson models (Al-Osh & Alzaid, 1987; McKenzie, 1985, 1988). The authors considered a set of unobserved components that capture the important feature of the data that show a short-range dependence. McKenzie (1988) also developed integer-valued moving average (INMA) processes, where binomial thinning was used to replace scalar multiplication for discrete random variables. Davis, Dunsmuir, and Streett (2003) introduced a general linear autoregressive moving average (GLARMA) model for time series count data. This is an observation-driven model, which assumes the distribution of the current count, conditional on the past information follows a Poisson distribution. The model establishes a linear relationship between the logarithm of the conditional mean of the Poisson process and explanatory variables with martingale differences via an ARMA structure. In order to provide a more flexible solution to the overdispersion problem, Febritasari, Wardhani, and Sa'adah (2019) combined negative binomial distribution with generalized linear autoregressive moving average (GLARMA) models in time series. Alternatively, Zhu (2006) proposed a class of models for non-stationary time series based on binomial thinning, which can incorporate trends or covariates, and also allows higher-order Markov dependence. For additional variations and details about thinning models, the reader is referred to the review papers by Weiß (2008) and Scotto, Weiß, and Gouveia (2015). Shephard (1999) proposed a model where the conditional intensity (mean) of a Poisson process is linearly dependent on the previous observation and its expectation. Since the Poisson distribution could take care of integer data and the GARCH formulation could incorporate the influence of previous observations, this model parallels a GARCH process. Heinen (2003) further generalized the model to accommodate arbitrary lags and named it as Autoregressive Conditional Poisson (ACP) model. Their formulation is as follows: given the count time series $X_t : t \in \mathbb{N}$, and the $\sigma$ field $F_{t-1}$ generated by the set $\{X_t : i < t\}$,

$$\begin{aligned} X_t|\mathbb{F}_{t-1} &\sim Pois(\lambda_t), \quad for\ \forall t \in \mathbb{N} \\ \lambda_t &= \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i} + \sum_{j=1}^{q} \beta_j \lambda_{t-j}, \end{aligned} \tag{2.1}$$

where $\alpha_0 > 0, \alpha_i \geq 0, \beta_j \geq 0$ for $i = 1, \ldots, p, j = 1, \ldots, q.$ with the restriction that $0 \leq \sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1$. The conditions required for stationarity is discussed, and mean and variance are provided in this paper. Heinen, however, derived the theoretical properties of the ACP model only for the case $p = q = 1$. Building on Heinen's work, Ghahramani and Thavaneswaran (2009) extended the theoretical properties of the ACP model to a higher order, and showed that the mean structure could be re-parameterized as an ARMA process and the moment properties of the ACP models were deduced in details. Around the same time, Ferland, Latour, and Oraichi (2006) introduced an integer-valued generalized autoregressive conditional heteroskedastic (INGARCH) $(p, q)$ model with Poisson deviates,which in essence, is the same as the ACP model. Zhu (2011) developed a negative binomial integer-valued GARCH model, aiming to handle overdispersion and extreme observations. Subsequently, Zhu (2012) introduced a class of generalized Poisson integer-valued GARCH models, which can account for both overdispersion and underdispersion. Chen, So, Li, and Sriboonchitta (2016) proposed an autoregressive conditional negative binomial model for time series of counts in which the standard negative binomial parameters were re-parameterized with one of the new parameters modeled using a conditional autoregressive formulation similar to the ACP.

The above time series models do not allow for changes in the model from one mode to another, or in other words, they do not allow the time series to switch between different "regimes". Markov chain based models represent a general class of formulations which could handle time series with such characteristics (Chib & Winkelmann, 2001; Raftery, 1985). The regime switching is accomplished via a Markov chain with fixed transition probabilities. However, the above Markov chain based models build a direct relationship between regimes and observations, thus the distribution of $X_t$ depends only on $X_{t-1}$ and they tend to be overparameterized for most empirical data as there are usually many possible outcomes.

The original concept of the hidden Markov models was advocated by Buam et al. (Baum & Eagon, 1967; Baum & Petrie, 1966). In this approach, it is assumed that an unobserved latent state changes over time according to a Markov process, and one or more of the model's parameters change with the state. As mentioned before, the Poisson hidden Markov model (P-HMM) was first developed and applied to a time series of daily epileptic seizure counts (Albert, 1991). This model allows the mean of the Poisson distribution to change according to an underlying two-state Markov chain. The EM algorithm was employed to compute the estimates. Results obtained in this study showed that the P-HMM outperformed Poisson regression models. However, the P-HMM model would not allow the previous count to have a direct influence on the current one, which is a drawback in situations where there is high autocorrelation, a common phenomenon in time series data, thus limiting the wide usage of such models. In a previous paper (Zhang & Samaranayake, 2019), we proposed a periodic version of the autoregressive conditional Poisson (ACP) model by allowing the coefficients of the ACP model to vary periodically. Some empirical time series, however, suggest the possibility of irregular changes that imply regime switching cannot be explained by a periodic model. Thus we propose the autoregressive conditional Poisson hidden Markov model (ACP-HMM), which in a sense, is a combination of Poisson hidden Markov model and autoregressive conditional Poisson model. It would accommodate time series with irregular regime switching patterns while taking care of serial dependence.

# 3 Proposed ACP-HMM Models

Let a sequence of discrete random variables $\{S_t : t \in \mathbb{N}\}$ be a Markov chain with $m$ possible states and a transition probability matrix $\mathbf{\Gamma}(t) = \{\gamma_{ij}^s(t)\}$, $i, j = 1, 2, \ldots, m$, where $\gamma_{ij}^s(t) = P(S_{s+t} = j | S_s = i)$. In most cases, it is enough to use homogeneous Markov chains, where $\gamma_{ij}^s$ does not depend on $s$. Unless there is an explicit indication, it is assumed from here on that the Markov chain under discussion is a homogeneous one with transition probabilities denoted by $\gamma_{ij}$.

Note that the traditional Poisson Hidden Markov model would not account for autocorrelations between the counts. Thus, the autoregressive conditional Poisson hidden Markov model proposed here is more appropriate to fit correlated count data where one or more parameters of the ACP model change according to the state of a hidden Markov process.

We are now ready to define the proposed ACP-HMM process. Let $\{X_t : t \in \mathbb{N}\}$ denote observed time series of count data, with $X_t$ representing the count at time $t$. It is assumed that the mean of the Poisson process at time $t$ is propagated through an ACP structure, whose parameters take values based on the state of the underlying hidden Markov chain. Let $S_t$ denote the state of the Markov chain to which $t$ belongs and denote the $\sigma$ -algebra generated by $\{X_i, S_i : i \leq t\}$ as $F_t$. Given the past information $F_t$

$$X_t | F_t \sim Poisson(\lambda_t)$$

where $\lambda_t$ is a time-varying parameter defined by

$$\lambda_{t,S_t} = \omega_{S_t} + \sum_{i=1}^{q} \alpha_{i,S_t} X_{t-i}, \tag{3.1}$$

where $\omega_{S_t}$, $\alpha_{i,St}$, $i = 1, 2, , \dots, q$, are positive for all values of $S_t$.

Note that the above formulation parallels that of an ARCH($q$) process, but with the parameters varying with the state of the Markov chain. A simpler model, where the $\alpha_{i,S_t}$ remain constant across all states $S_t$ can be adopted and may be sufficient to model some empirical count data series. Note that the standard ACP model parallels a GARCH process, and we have employed the simpler ARCH formulation to avoid estimation issues posed by a more complex model.

## 4 Some Properties of the Model

Denote the unconditional probabilities of a Markov chain at time $t$ by $P(S_t = i)$, and probabilities of all possible outcomes at time $t$ as row vector $\mathbf{u}_t = \big(P(S_t = 1), P(S_t = 2), \dots, P(S_t = m)\big), t \in \mathbb{N}$, with $m$ representing the number of Markov states. Let $\mathbf{\Gamma}(t) = \{\gamma_{ij}(t)\}$, $i, j = 1, 2, \dots, m$, where $\gamma_{ij}(t) = P(S_{s+t} = j|S_s = i)$, and the mean of the Poisson process generated by $\lambda_{t,S_t} = \omega_{S_t} + \sum_{i=1}^{q} \alpha_{i,S_t} X_{t-i}$. In order to express the expected mean and variance of observation $X_t$ by vector and matrix calculations, define the row vector of means of the Poisson process under different states as $\boldsymbol{\lambda}_t = \big(\lambda_t(S_t = 1), \lambda_t(S_t = 2), \dots, \lambda_t(S_t = m)\big)$. Let $\boldsymbol{\delta} = \mathbf{u}_1 = \big(P(S_1 = 1), P(S_1 = 2), \dots, P(S_1 = m)\big)$ be the initial distribution of the Markov chain. Then, the $\mathbf{u}_{t+1}$ could be deduced from relation $\mathbf{u}_{t+1} = \mathbf{u}_t \mathbf{\Gamma}_t$. We restrict the scope of the study here to the homogeneous Markov chain model, and thus $\mathbf{\Gamma}(t)$ will be abbreviated as $\mathbf{\Gamma}$. So we have,

$$\mathbf{u}_t = \mathbf{u}_{t-1}\mathbf{\Gamma} = \boldsymbol{\delta}\mathbf{\Gamma}^{t-1},$$

$$E(X_t) = \mathbf{u}_t \boldsymbol{\lambda_t} = \boldsymbol{\delta}\mathbf{\Gamma}^{t-1}\boldsymbol{\lambda}'_t,$$

$$\begin{aligned} Var(X_t) &= E[Var(X_t|S_t)] + Var[E(X_t|S_t)], \\ &= E(X_t) + Var\big(E(X_t|S_t)\big), \\ &= E(X_t) + E[\big(E(X_t|S_t)\big)^2] - [E\big(E(X_t|S_t)\big)]^2, \\ &= E(X_t) + \sum_{S_t=i}^{m} \big(\lambda_t(S_t = i)\big)^2 P(S_t = i) - [E(E(X_t|S_t))]^2, \\ &= \boldsymbol{\delta}\mathbf{\Gamma}^{t-1}\boldsymbol{\lambda}'_t + \boldsymbol{\delta}\mathbf{\Gamma}^{t-1}(\boldsymbol{\lambda}^2_t)' - (\boldsymbol{\delta}\mathbf{\Gamma}^{t-1}\boldsymbol{\lambda}'_t)^2. \end{aligned} \tag{4.1}$$

where $\boldsymbol{\lambda_t}^2$ means squaring each element of vector $\boldsymbol{\lambda_t}$

## 5 Likelihood Function and Parameter Estimation

Let $\underline{\theta} \equiv (\boldsymbol{\delta}, \mathbf{\Gamma}, \omega_{S_t}, \alpha_{i,S_t})$ for $i = 1, 2, \dots, q$, represent all parameters in hidden Markov auto-regressive conditional Poisson model. The log-likelihood function for the model is given by

$$l_{\underline{T}}(\underline{\theta}) = log\Big(P(\boldsymbol{X}_T = \boldsymbol{x}_T)\Big) = log\big(\boldsymbol{\delta}\boldsymbol{P}(x_1)\boldsymbol{\Gamma}\boldsymbol{P}(x_2)\cdots\boldsymbol{\Gamma}\boldsymbol{P}(x_T)\mathbf{1'}\big), \tag{5.1}$$

where $\boldsymbol{\delta}$ is the initial distribution and

$$\boldsymbol{P}(x_t) = \begin{bmatrix} p_1(x_t) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \vdots & p_m(x_t) \end{bmatrix}, \tag{5.2}$$

$$p_i(x_t) = P(X_t = x_t | C_t = i),$$

$$\boldsymbol{\Gamma} = \begin{bmatrix} \gamma_{11} & \cdots & \gamma_{1m} \\ \vdots & \ddots & \vdots \\ \gamma_{m1} & \cdots & \gamma_{mm} \end{bmatrix}.$$

For the discrete case, elements in the likelihood function become progressively smaller as $t$ increases, and scaling the forward probabilities is a common way to avoid underflow. Thus we have,

$$\boldsymbol{\beta}_0 = \boldsymbol{\delta}\boldsymbol{P}(x_1),$$

$$\boldsymbol{\beta}_t = \boldsymbol{\beta}_{t-1}\boldsymbol{\Gamma}\boldsymbol{P}(x_t), \quad for\, t = 2, 3, \ldots, T.$$

$$\boldsymbol{\phi}_0 = \boldsymbol{\delta},$$

$$\boldsymbol{\phi}_t = \frac{\boldsymbol{\beta}_t}{\omega_t},$$

$$\omega_t\boldsymbol{\phi}_t = \omega_{t-1}\boldsymbol{\phi}_{t-1}\boldsymbol{B}_t,$$

where

$$\omega_t = \sum_i \beta_t(i) = \boldsymbol{\beta}_t\mathbf{1'},$$

$$\omega_0 = \boldsymbol{\delta}\mathbf{1'}.$$

Thus the scaled log likelihood function would be

$$log(L_T) = \sum_{t=1}^{T} log\Big(\frac{\omega_t}{\omega_{t-1}}\Big) = \sum_{t=1}^{T} log(\boldsymbol{\phi}_{t-1}\boldsymbol{B}_t\mathbf{1'}). \tag{5.3}$$

Note that the EM algorithm could also be derived and used. However, the solutions to the EM algorithm do not have a closed-form; thus, the maximum likelihood estimation method gives better estimates.

## 6 The Monte-Carlo Simulation Study

We conducted a Monte-Carlo simulation study to investigate the performance of maximum likelihood estimators of ACP-HMM model with the log likelihood function defined in (5.3). A simulation study was also used to investigate the use of AIC and BIC criteria to differentiate between highly correlated count data series and regular Poisson HMM processes.

The properties of estimates were studied across different combinations of parameters using 1,000 simulation runs for each combination. Bias, Monte Carlo standard error and mean absolute deviation were computed for each of the parameter combination sets. In order to eliminate the artifacts arising out of initial conditions, the first 240 time series data points were discarded. For the simulation of the parameter set, sample size T=1,440 were considered, which is comparable to the lengths of one-minute count data series observed over one day or the length of daily count data observed over a period of four years.

We provide the parameter sets used in the simulation study here before we move to details for each case.

**Case 1.** Time series of count data with 2 states and 2 lags are generated. Each state has the same lag coefficients.

$$\mathbf{\Gamma} = \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix}, \omega_1 = 20, \omega_2 = 10, \alpha_1 = 0.1, \alpha_2 = 0.2.$$

**Case 2.** Time series of count data with 3 states and 2 lags are generated. Each state has the same lag coefficients.

$$\mathbf{\Gamma} = \begin{bmatrix} 0.8 & 0.15 & 0.05 \\ 0.15 & 0.75 & 0.1 \\ 0.05 & 0.15 & 0.8 \end{bmatrix}, \omega_1 = 20, \omega_2 = 13, \omega_3 = 8, \alpha_1 = 0.2, \alpha_2 = 0.1.$$

**Case 3.** Time series of count data with 2 states and 1 lag are generated. Each state has a different value for the single lag coefficient.

$$\mathbf{\Gamma} = \begin{bmatrix} 0.75 & 0.25 \\ 0.2 & 0.8 \end{bmatrix}, \omega_1 = 10, \omega_2 = 20, \alpha_{1,1} = 0.3, \alpha_{1,2} = 0.2.$$

**Case 4.** Time series of count data with 3 states and 1 lag are generated. Each state has a different value for the single lag coefficient.

$$\mathbf{\Gamma} = \begin{bmatrix} 0.7 & 0.25 & 0.05 \\ 0.15 & 0.7 & 0.15 \\ 0.05 & 0.2 & 0.75 \end{bmatrix}, \omega_1 = 3, \omega_2 = 8, \omega_3 = 4, \alpha_{1,1} = 0.1, \alpha_{1,2} = 0.5, \alpha_{1,3} = 0.3.$$

### 6.1 Case 1. Time series of count data with 2 states and 2 lags are generated. Each state has the same lag coefficients.

The simulated data (listed in Table 1) were generated from an ACP-HMM process with 2 states and 2 lags. Each state has the same lag coefficients.

**Table 1:** Maximum likelihood estimation results from 1,000 simulations based on different number of Markov states ($m$ = 2 states, $q$ = 2).

| Parameter | True Coefficient | Estimates | MSE | MAD |
|---|---|---|---|---|
| p11 | 0.7 | 0.70083 | 0.000973 | 0.031194 |
| p12 | 0.3 | 0.29917 | 0.000973 | 0.031194 |
| p21 | 0.4 | 0.39909 | 0.001284 | 0.035827 |
| p22 | 0.6 | 0.60091 | 0.001284 | 0.035827 |
| $\omega_1$ | 20 | 20.099 | 0.796033 | 0.892207 |
| $\omega_2$ | 10 | 10.08 | 0.641469 | 0.800917 |
| $\alpha_1$ | 0.1 | 0.097745 | 0.00064 | 0.025289 |
| $\alpha_2$ | 0.2 | 0.19806 | 0.001001 | 0.031631 |

### 6.2 Case 2. Time series of count data with 3 states and 2 lags are generated. Each state has the same lag coefficients.

The simulated data (listed in Table 2) were generated from an ACP-HMM process with 3 states and 2 lags. Each state has the same lag coefficients.

**Table 2:** Maximum likelihood estimation results from 1,000 simulations based on different number of Markov states ($m$ = 3 states, $q$ = 2).

| Parameter | True Coefficient | Estimates | MSE | MAD |
|---|---|---|---|---|
| p11 | 0.8 | 0.75734 | 0.0259 | 0.069293 |
| p12 | 0.15 | 0.1596 | 0.0179 | 0.075879 |
| p13 | 0.05 | 0.083057 | 0.00789 | 0.053554 |
| p21 | 0.15 | 0.16917 | 0.0155 | 0.061574 |
| p22 | 0.75 | 0.71248 | 0.03 | 0.090094 |
| p23 | 0.1 | 0.11835 | 0.0166 | 0.06625 |
| p31 | 0.05 | 0.088715 | 0.0154 | 0.063711 |
| p32 | 0.15 | 0.19528 | 0.0292 | 0.094614 |
| p33 | 0.8 | 0.716 | 0.0561 | 0.113975 |
| $\omega_1$ | 20 | 18.887 | 7.21 | 1.907206 |
| $\omega_2$ | 13 | 12.018 | 6.58 | 1.725269 |
| $\omega_3$ | 8 | 7.3377 | 2.88 | 1.068537 |
| $\alpha_1$ | 0.2 | 0.2117 | 0.00142 | 0.029035 |
| $\alpha_2$ | 0.1 | 0.13157 | 0.00556 | 0.050574 |

### 6.3 Case 3. Time series of count data with 2 states and 1 lag are generated. Each state has a different value for the single lag coefficient.

The simulated data (listed in Table 3) were generated from an ACP-HMM process with 2 states and 1 lag. Each state has a different value for the single lag coefficient.

**Table 3:** Maximum likelihood estimation results from 1,000 simulations based on different number of Markov states ($m$ = 2 states, $q$ = 1).

| Parameter | True Coefficients | Estimates | MSE | MAD |
|---|---|---|---|---|
| p11 | 0.75 | 0.74789 | 0.001008 | 0.02426 |
| p12 | 0.25 | 0.25211 | 0.001008 | 0.02426 |
| p21 | 0.2 | 0.20461 | 0.001951 | 0.025965 |
| p22 | 0.8 | 0.79539 | 0.001951 | 0.025965 |
| $\omega_1$ | 10 | 9.9861 | 0.470935 | 0.536199 |
| $\omega_2$ | 20 | 20.056 | 1.107474 | 0.833189 |
| $\alpha_{1,1}$ | 0.3 | 0.30087 | 0.001984 | 0.034581 |
| $\alpha_{1,2}$ | 0.2 | 0.19915 | 0.001885 | 0.032861 |

### 6.4 Case 4. Time series of count data with 3 states and 1 lag are generated. Each state has a different value for the single lag coefficient.

The simulated data (listed in Table 4) were generated from an ACP-HMM process with 3 states and 1 lag. Each state has a different value for the single lag coefficients.

**Table 4:** Maximum likelihood estimation results from 1,000 simulations based on different number of Markov states ($m$ = 3 states, $q$ = 1).

| Parameter | True Coefficient | Estimates | MSE | MAD |
|---|---|---|---|---|
| p11 | 0.7 | 0.68446 | 0.010054 | 0.067461 |
| p12 | 0.25 | 0.2108 | 0.011515 | 0.07506 |
| p13 | 0.05 | 0.10475 | 0.018946 | 0.090441 |
| p21 | 0.15 | 0.15906 | 0.006665 | 0.060033 |
| p22 | 0.7 | 0.68963 | 0.004557 | 0.041534 |
| p23 | 0.15 | 0.15131 | 0.007579 | 0.065768 |
| p31 | 0.05 | 0.098007 | 0.014336 | 0.080578 |
| p32 | 0.2 | 0.18475 | 0.009468 | 0.073481 |
| p33 | 0.75 | 0.71724 | 0.012719 | 0.075851 |
| $\omega_1$ | 3 | 2.8684 | 0.24099 | 0.314269 |
| $\omega_2$ | 8 | 8.2234 | 1.2147 | 0.592089 |
| $\omega_3$ | 4 | 4.2615 | 1.4266 | 0.830676 |
| $\alpha_{1,1}$ | 0.1 | 0.23549 | 0.088734 | 0.246055 |
| $\alpha_{1,2}$ | 0.5 | 0.42867 | 0.067414 | 0.204649 |
| $\alpha_{1,3}$ | 0.3 | 0.3104 | 0.037574 | 0.173398 |

Based on the simulation results, it can be seen that the maximum likelihood estimation method provides relatively good estimates with low MSE and MAD in most cases. For cases 1-3, the estimates are quite accurate, and the MADs remain low. When the number of states is higher than two, and the ACP portion of the model has a complicated structure (Case 4) with both omega and alpha varying with the states, the MSE and MAD values become relatively large for some of the parameters. A similar situation is observed in Case 2 (with three states), where the MSE associated with the intercept parameters increased, similar to that in the Poisson HMM setup. Overall, results demonstrate that the MLE is a promising method for estimating the parameters of the suggested autoregressive conditional Poisson hidden Markov model (ACP-HMM), especially when the number of states is two.

# 7 Model Selection

To illustrate the importance of selecting the correct structure for the underlying data generating process and also to examine if AICc and/or BIC are good criteria to distinguish the true generating process, a small-scale Monte Carlo simulation study was performed. All statistics reported here are calculated from N=1,000 replications, and each replication having a sample size T=1,440. In order to avoid artifacts created by initial conditions, the first 240 time series data points were discarded.

Mean AICc and BIC are calculated from AICc and BIC values for all replications, and the percentage represents the proportion of simulation runs that yielded a smaller AICc or BIC value for the corresponding model.

Table 5 shows results for the case when the data were generated from an ACP-HMM process with true parameters:

$$\mathbf{\Gamma} = \begin{bmatrix} 0.7 & 0.3 \\ 0.2 & 0.8 \end{bmatrix}, \omega_1 = 5, \omega_2 = 20, \alpha_1 = 0.3, \alpha_2 = 0.1.$$

**Table 5:** Poisson HMM and ACP-HMM selection by AICc and BIC criteria with simulated time series data with small $\alpha$'s that do not differ much.

| Parameter | True Coefficient | ACP-HMM | Poisson HMM |
|---|---|---|---|
| p11 | 0.7 | 0.70224 | 0.73894 |
| p12 | 0.3 | 0.29776 | 0.26106 |
| p21 | 0.2 | 0.20249 | 0.15565 |
| p22 | 0.8 | 0.79751 | 0.84435 |
| $\omega_1$ | 5 | 5.0237 | 12.21 |
| $\omega_2$ | 20 | 20.019 | 29.896 |
| $\alpha_1$ | 0.3 | 0.29968 | - |
| $\alpha_2$ | 0.1 | 0.10057 | - |
| mean AICc | | 10181(100%) | 10792(0%) |
| mean BIC | | 10213(100%) | 10827(0%) |

Both the ACP-HMM and the Poisson HMM Model were utilized to fit the data. The average AICc and BIC values for ACP-HMM are lower than those for the Poisson HMM model and the right structure is always preferred, which suggests AICc and BIC performs well in identifying the true structure of the time series.

Table 6 shows results when an ACP-HMM process is the underlying structure producing the count data with true parameters:

$$\mathbf{\Gamma} = \begin{bmatrix} 0.7 & 0.3 \\ 0.2 & 0.8 \end{bmatrix}, \omega_1 = 2.5, \omega_2 = 9, \alpha_1 = 0.8, \alpha_2 = 0.1.$$

Both ACP-HMM model and Poisson HMM model were utilized to fit the data. In this case, AICc and BIC again showed their ability to select the right structure. Note that the simulation results show that if the data generating process of a count data time series has an autoregressive conditional heteroscedastic structure, and its parameters are governed by a hidden Markov process, then the regular Poisson HMM provides a poor fit, especially when one or more of the $\alpha$'s are high.

**Table 6:** Poisson HMM and ACP-HMM selection by AICc and BIC criteria with simulated time series data with one large $\alpha$ and the other small.

| Parameter | True Coefficient | ACP-HMM | Poisson HMM |
|---|---|---|---|
| p11 | 0.7 | 0.71857 | 0.92571 |
| p12 | 0.3 | 0.28143 | 0.074288 |
| p21 | 0.2 | 0.20091 | 0.073842 |
| p22 | 0.8 | 0.79909 | 0.92616 |
| $\omega_1$ | 2.5 | 2.5661 | 43.837 |
| $\omega_2$ | 9 | 8.9031 | 68.342 |
| $\alpha_1$ | 0.8 | 0.79107 | - |
| $\alpha_2$ | 0.1 | 0.096571 | - |
| mean AICc | | 10461(100%) | 12650(0%) |
| mean BIC | | 10493(100%) | 12686(0%) |

# 8 Visualization of Data and Model Structure using a Synthetic Data Set

The graph demonstrates the proposed ACP-HMM model with data generated from the parameter set:

$$\mathbf{\Gamma} = \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix}, \omega_1 = 20, \omega_2 = 10, \alpha_1 = 0.1, \alpha_2 = 0.2.$$

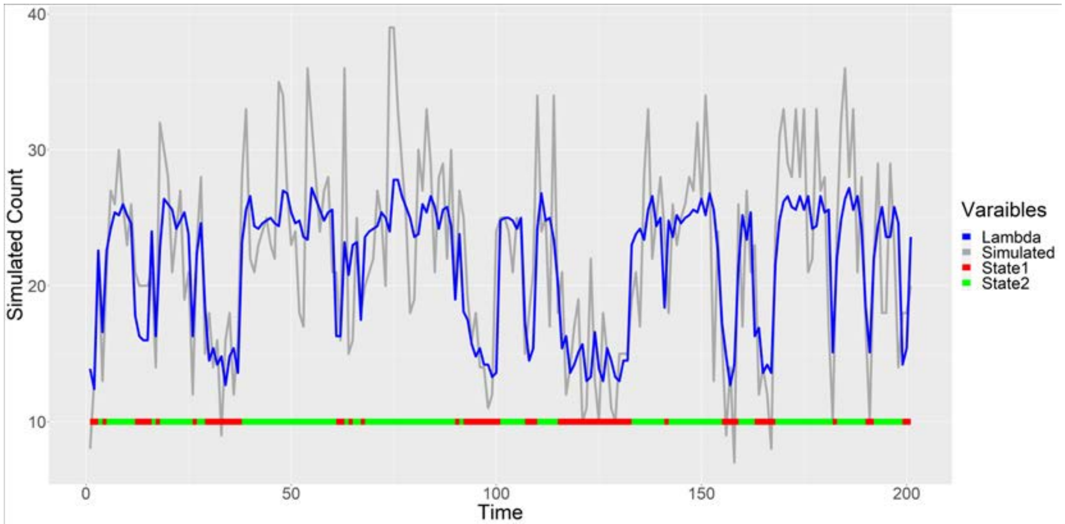


**Figure 1:** Simulated Time Series Count Data, and the underlying $\lambda_{S_t}$ and states.

As shown in Figure 1, the grey line represents the simulated data while the blue line indicates the underlying mean of the Poisson process. The red line represents the underlying process is at State 1 while the green line indicates the process is at State 2.

# 9 Application to a Real-Life Data Set

Figure 2 illustrates the daily number of deaths in Evora, Portugal, from 01/01/1996 to 12/31/2007. The sample mean equals 6.119, and the variance is 7.483. There seems to be an irregular periodicity present in this time series.

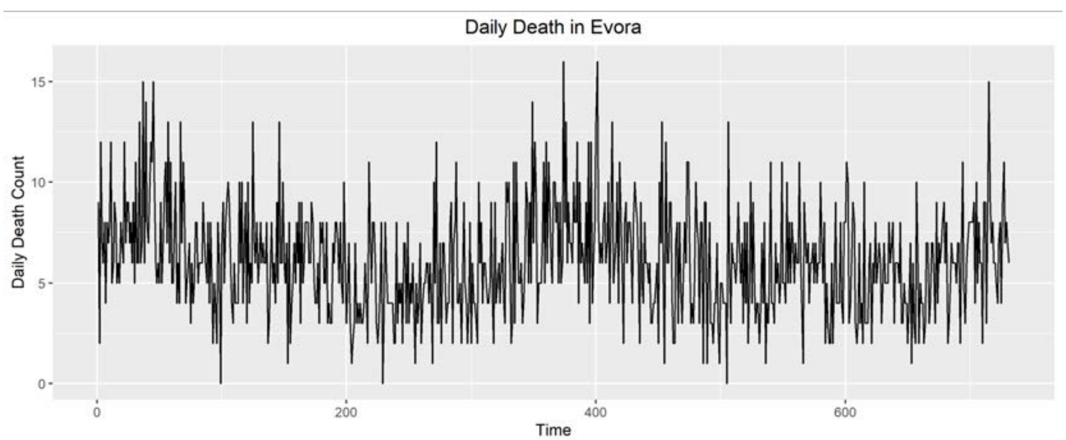


**Figure 2:** Daily Death in Evora from 01/01/1996 to 12/31/2007.

The autocorrelation function plot of the count data (as shown in Figure 3) suggests there is autocorrelation in the count data, and hence ACP structure is better than a regular Poisson model. Some irregular periodicity is also observed, hence the motivation for fitting a HMM could be seen.

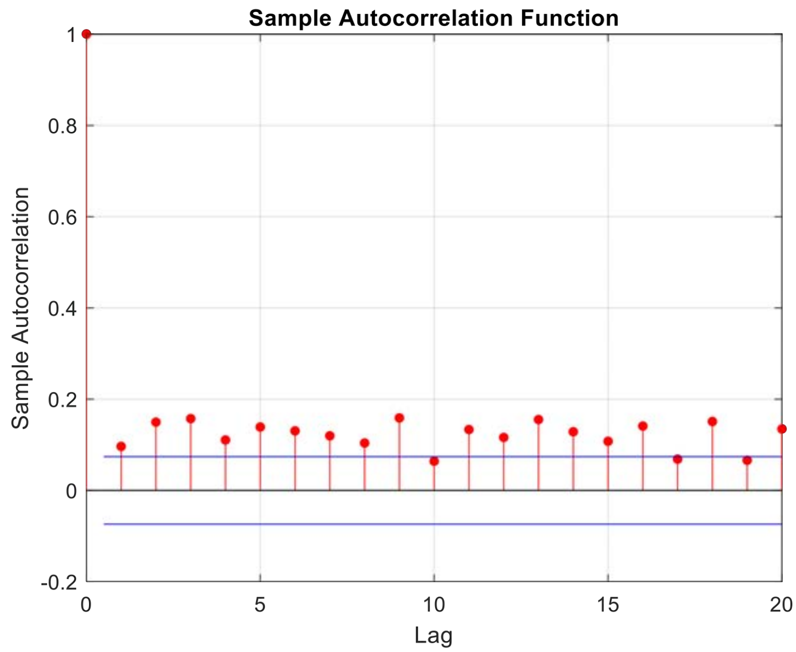


**Figure 3:** Autocorrelation of Daily Death Count Data.

**Table 7:** Daily Death Data fitted by Poisson HMM and ACP-HMM with AICc provided.

| Parameter | ACP-HMM | Poisson HMM |
|---|---|---|
| p11 | 0.9954 | 0.999999 |
| p12 | 0.0046 | 0.000001 |
| p21 | 0.0127 | 0.0422 |
| p22 | 0.9873 | 0.9578 |
| $\omega_1$ | 5.5312 | 6.123 |
| $\omega_2$ | 6.8751 | 33.0782 |
| $\alpha_{1,1}$ | 0.0001 | - |
| $\alpha_{1,2}$ | 0.1086 | - |
| AICc | 3445.2 | 3533.7 |

Daily Death Data is fitted by Poisson HMM and ACP-HMM with 2 states and their corresponding AICc's are provided in Table 7. Poisson HMM gives a large mean for the second Poisson process, and the transition matrix gives extreme probabilities, most probably due to the model's inability to account for the autocorrelation present in the data.

# 10 Conclusion

The model provided here is a natural generalization of the Poisson hidden Markov model, with a generalization made to take the influence of previous observations into consideration when modeling autocorrelated count time series. The reported simulation results in Section 6 show that the MLE method provides reasonable estimates of the model parameters of the ACP-HMM model. However, when the parameter set gets larger, the errors of the estimators increases , possibly because the hypersurface defining the likelihood function grows more complex. Investigating the utility of AICc and BIC criteria in determining the true structure of the count data shows promising results. Finally, we use a real-life data set to illustrate the importance of the proposed model in situations where evidence for regime switch and autocorrelations are both apparent, which is not an uncommon phenomenon in time series of count data.

# References

Albert, P. S. (1991). A two-state markov mixture model for a time series of epileptic seizure counts. *Biometrics*, 1371–1381.

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

Baum, L. E., & Eagon, J. A. (1967). An inequality with applications to statistical estimation for probabilistic functions of markov processes and to a model for ecology. *Bulletin of the American Mathematical Society*, *73*(3), 360–363.

Baum, L. E., & Petrie, T. (1966). Statistical inference for probabilistic functions of finite state markov chains. *The annals of mathematical statistics*, *37*(6), 1554–1563.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, *31*(3), 307–327.

Chen, C. W., So, M. K., Li, J. C., & Sriboonchitta, S. (2016). Autoregressive conditional negative binomial model applied to over-dispersed time series of counts. *Statistical Methodology*, *31*, 73–90.

Chib, S., & Winkelmann, R. (2001). Markov chain monte carlo analysis of correlated count data. *Journal of Business & Economic Statistics*, *19*(4), 428–435.

Consul, P., & Famoye, F. (1992). Generalized poisson regression model. *Communications in Statistics-Theory and Methods*, *21*(1), 89–109.

Coxe, S., West, S. G., & Aiken, L. S. (2009). The analysis of count data: A gentle introduction to poisson regression and its alternatives. *Journal of personality assessment*, *91*(2), 121–136.

Davis, R. A., Dunsmuir, W. T., & Streett, S. B. (2003). Observation-driven models for poisson counts. *Biometrika*, *90*(4), 777–790.

Febritasari, P., Wardhani, N. W. S., & Sa'adah, U. (2019). Generalized linier autoregressive moving average (glarma) negative binomial regression models with metropolis hasting algorithm. In *Iop conference series: Materials science and engineering* (Vol. 546, p. 052023).

Ferland, R., Latour, A., & Oraichi, D. (2006). Integer-valued garch process. *Journal of Time Series Analysis*, *27*(6), 923–942.

Ghahramani, M., & Thavaneswaran, A. (2009). On some properties of autoregressive conditional poisson (acp) models. *Economics Letters*, *105*(3), 273–275.

Heinen, A. (2003). Modelling time series count data: an autoregressive conditional poisson model.

Hilbe, J. M. (2014). *Modeling count data*. Cambridge University Press.

Jacobs, P. A., & Lewis, P. A. (1983). Stationary discrete autoregressive-moving average time series generated by mixtures. *Journal of Time Series Analysis*, *4*(1), 19–36.

McKenzie, E. (1985). Some simple models for discrete variate time series 1. *JAWRA Journal of the American Water Resources Association*, *21*(4), 645–650.

McKenzie, E. (1986). Autoregressive moving-average processes with negative-binomial and geometric marginal distributions. *Advances in Applied Probability*, 679–705.

McKenzie, E. (1987). Innovation distributions for gamma and negative binomial autoregressions. *Scandinavian journal of statistics*, 79–85.

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

Raftery, A. E. (1985). A model for high-order markov chains. *Journal of the Royal Statistical Society: Series B (Methodological)*, *47*(3), 528–539.

Scotto, M. G., Weiß, C. H., & Gouveia, S. (2015). Thinning-based models in the analysis of integer-valued time series: a review. *Statistical Modelling*, *15*(6), 590–618.

Shephard, N. (1999). *Bin models for trade-by-trade data. modelling the number of trades in a fixed interval of time* (Tech. Rep.). Mimeo. Presented at EC2, Madrid.

Weiß, C. H. (2008). Thinning operations for modeling time series of counts—a survey. *AStA Advances in Statistical Analysis*, *92*(3), 319.

Zhang, Y., & Samaranayake, V. (2019). A periodic conditional poisson model. *In JSM Proceedings, Business and Economic Statistics Section*, 2918-2935.

Zhu, F. (2011). A negative binomial integer-valued garch model. *Journal of Time Series Analysis*, *32*(1), 54–67.

Zhu, F. (2012). Modeling overdispersed or underdispersed count data with generalized poisson integer-valued garch models. *Journal of Mathematical Analysis and Applications*, *389*(1), 58–71.

Zhu, R., & Joe, H. (2006). Modelling count data time series with markov processes based on binomial thinning. *Journal of Time Series Analysis*, *27*(5), 725–738.