PoARX models for count time series

PoARX models for count time series

Jamie Halliday and Georgi N. Boshnakov

University of Manchester, United Kingdom
jamie.halliday@manchester.ac.uk

Abstract. This paper introduces multivariate Poisson autoregressive models with exogenous covariates (PoARX) for modelling multivariate time series of counts. We state conditions for a PoARX process to be stationary and ergodic before proposing a computationally efficient procedure for estimation of parameters by the method of inference functions (IFM) and stating asymptotic normality of these estimators. Lastly, we demonstrate an application to count data for the number of people entering and exiting a building, and show how the different aspects of the model combine to produce a strong predictive model. We conclude by listing directions for future work.

Keywords: Multivariate time series, Count data, Prediction, Copula

1 Introduction

The abundance of data brought about by the digital revolution has increased the availability of time series of counts. Such data appear in many areas, including statistics, econometrics, and the social and physical sciences. The most popular distribution for modelling count data is the Poisson distribution, which has attractive properties and is in some respects the count analogue of the Gaussian distribution. One restrictive property of the Poisson distribution however is that the mean and the variance are equal – this is rarely observed in applications.

For time series data, the correlation between observations provides additional challenges. The classic example of time series models is the ARMA model, which has multivariate extensions. A fruitful approach, employed in ARCH and GARCH models [6,3], uses a separate equation to model directly the dependence of the variance on the past. In order to improve the predictive accuracy, the aforementioned models have been augmented with additional exogenous covariates. However, these models do not make specific provision for the non-negativity and integer-valued nature of count data. An integer-valued analogue of the GARCH model, called INGARCH [7], uses Poisson deviates rather than normal innovations to combat these issues. Furthermore, a Poisson model for integer-valued time series has been proposed, called the Poisson autoregressive model [8], which has an autoregressive feedback mechanism for the mean. Subsequently, a class of dynamic Poisson models allowing for exogenous covariates was suggested called PARX [1]. Whilst the Poisson distribution has been widely used for univariate count models, multivariate generalisations have been relatively sparse so far. Recently, a summary of multivariate (Poisson) distributions for count data has been made [15], including multivariate extensions of a parametric (Poisson) distribution and copula modelling using univariate (Poisson) marginal distributions.

In this article, we use a copula to extend the (univariate) PARX model [1] to multivariate count time series. This approach is flexible and tractable. Use of covariates in the Poisson model offers clear potential for better modelling and, by including the time series covariates, we allow over-dispersed data to be considered by our model. Implementation in R [24] is available in the developmental package PoARX [11].

This paper is organised as follows. Section 2 introduces the univariate and multivariate PoARX model, giving stationarity and ergodicity conditions. In Section 3 we discuss estimation of parameters by the method of inference functions (IMF) [17] and asymptotic results for the resulting estimators are stated. We consider prediction in Section 4, looking at the generating functions for future horizons. Then we demonstrate an application of the PoARX model in Section 5 by analysing a bivariate time series of count data for number of people entering and exiting a building on the University of California, Irvine (UCI) campus [14]. Exogenous covariates, such as the occurrence of a meeting or conference are included in the model to aid predictive accuracy. We summarise our findings in Section 6 and outline suggestions for future work.

2 The multivariate PoARX model

In this section we present the new class of models, introducing the necessary background material about the univariate PoARX model and copulas before generalising to higher dimensions. For the purpose of this article we focus on using Frank's copula to capture dependence between time series, but any suitable copula could be used. We use Frank's copula because in two dimensions it can account for both positive and negative dependence by allowing the dependence parameter to take values along the entire real line except zero.

2.1 The univariate PoARX model

First, a note on terminology – [1] use the abbreviation PARX for this model but we prefer PoARX since it seems to suggest more clearly "Poisson" and avoids confusion with other meanings of "P" in similar abbreviations. For example, PAR is often used to mean periodic autoregression.

Let $\{Y_t;\ t = 1, 2, \dots\}$ denote an observed time series of counts, so that $Y_t \in \{0, 1, 2, \dots\}$ for all $t = 1, 2, \dots$. Further, let $x_{t-1} \in \mathbb{R}^r$ denote a vector of additional covariates considered for inclusion in the model. We say that $\{Y_t\}$ is a univariate PoARX(p,q) process and write $\{Y_t\} \sim \mathrm{PoARX}_1(p, q)$, if its dynamics can be written as follows:

$$\begin{aligned} Y_t \mid \mathcal{F}_{t-1} &\sim \mathrm{Poisson}(\lambda_t), \\ \lambda_t &= \omega + \sum_{l=1}^{p} \alpha_l Y_{t-l} + \sum_{l=1}^{q} \beta_l \lambda_{t-l} + \eta \cdot x_{t-1}, \end{aligned} \tag{1}$$

where Poisson(λ) denotes a Poisson distribution with parameter λ, $\mathcal{F}_{t-1}$ denotes the σ-field of past knowledge, $\sigma\{Y_{1-p}, \ldots, Y_{t-1}, \lambda_{1-q}, \ldots, \lambda_{t-1}, x_1, \ldots, x_{t-1}\}$, $\omega \geq 0$ is an intercept term, $\{\alpha_1, \ldots, \alpha_p\}$ and $\{\beta_1, \ldots, \beta_q\}$ are non-negative autoregressive coefficients, and η is a vector of non-negative coefficients for the exogenous covariates. Thus, the model for the intensity, λ_t, uses the past p values of the process, the past q values of the intensity and the covariates.

In order to ensure that the process is stationary and ergodic with polynomial moments of a given order, we place two further restrictions on the model [1]. Firstly, the autoregressive coefficients must obey the following condition,

$$\sum_{i=1}^{\max\{p,q\}} (\alpha_i + \beta_i) < 1. \tag{2}$$

Additionally, we require that each component of the exogenous covariates, denoted $x_t(k)$ to avoid confusion later, follows a Markov structure, that is,

$$x_t(k) = g(x_{t-1}(k), \ldots, x_{t-m}(k); \epsilon_t), \qquad k = 1, \ldots, r, \tag{3}$$

for some $m > 0$ and some function $g(\boldsymbol{x}, \epsilon)$ with vector $\boldsymbol{x}$ independent of the observed Y_t and unobserved λ_t, and with ϵ_t an i.i.d. error term.

2.2 Copulas

Copulas provide a well-defined approach to model multivariate data, with the dependence structure considered separately from the univariate margins [17]. Copula theory can be attributed to a theorem stating that any multivariate distribution can be represented as a function of its marginals [25]. Estimation of the copula is relatively straightforward as a two-stage procedure[16]. First the univariate margins are fitted to respective parameters before the copula fit to find the value of the dependence parameter. An important class of copulas are called Archimedian copulas [23], which can be constructed easily from a generator function $\varphi(\cdot)$.

Frank's copula [23] is one example of such a copula. As mentioned, the dependence parameter can take any value except zero in the two-dimensional case ($\rho \in \mathbb{R}\backslash\{0\}$). In higher dimensions, however, the dependence parameter is limited to values in $(0, \infty)$. In any case the limit as $\rho \to 0$ corresponds to independence. Using a subscript ρ to denote the case of Frank's copula, we write

$$C_\rho(u_1, \ldots u_{\mathrm{K}}) = \varphi_\rho^{[-1]} \left(\sum_{i=1}^{\mathrm{K}} \varphi_\rho(u_i) \right), \tag{4}$$

where the generator function is given by

$$\varphi_\rho(t) = -\log \left(\frac{\exp(-\rho t) - 1}{\exp(-\rho) - 1} \right), \tag{5}$$

and its inverse

$$\varphi_\rho^{[-1]}(t) = \varphi_\rho^{-1}(t) = -\frac{1}{\rho}\log\left(1 + \exp(-t)(\exp(-\rho) - 1)\right). \tag{6}$$

For discrete random variables the copula is no longer unique due to the presence of stepwise marginal distribution functions [18]. Despite this issue, copula models are still valid constructions for discrete distributions [9]. Additionally, evidence has been provided that suggests there are fewer identification problems when the marginal distributions are conditioned non-trivially upon covariates [27]. Joint probabilities are computed as rectangle probabilities.

2.3 The multivariate PoARX model

Let $\{Y_t = (Y_t^1, \ldots, Y_t^{\mathrm{K}}),\ t = 1, 2, \ldots\}$ be a multivariate time series and let $\{x_{t-1}^j = (x_{t-1}^j(1), \ldots, x_{t-1}^j(r))^\top,\ j = 1, 2, \ldots, \mathrm{K}\}$ be the matrix of exogenous covariates associated with Y_t. We say that $\{Y_t\}$ is a PoARX process and write $\{Y_t\} \sim \mathrm{PoARX}_{\mathrm{K}}(p, q)$, if each of the component time series is a univariate PoARX process and the joint conditional distribution is a copula Poisson. Let the intensities of PoARX processes be $\{\lambda_t^j;\ t = 1, 2 \ldots,\ j = 1, \ldots, \mathrm{K}\}$ and be denoted using $\lambda_t = \left(\lambda_t^1, \ldots \lambda_t^{\mathrm{K}}\right)$.

Let $\mathcal{D}(\lambda^1, \ldots, \lambda^{\mathrm{K}}; \rho)$ be a multivariate distribution based on Frank's copula (Equations (4)–(6)) with marginal distributions $\mathrm{Poisson}(\lambda^1)$, ..., $\mathrm{Poisson}(\lambda^{\mathrm{K}})$ and dependency parameter ρ. Before stating the entire behaviour of the multivariate model, the distribution function corresponding to $\mathcal{D}(\lambda^1, \ldots, \lambda^{\mathrm{K}}; \rho)$ is

$$F(y; \lambda, \rho) = C_\rho(F_1(y^1; \lambda^1), \ldots, F_{\mathrm{K}}(y^{\mathrm{K}}; \lambda^{\mathrm{K}})), \tag{7}$$

where $F_1, \ldots, F_{\mathrm{K}}$ are the distribution functions of the Poisson marginals, i.e.

$$F_j(x; \mu) = \sum_{k=0}^{x} \mathrm{e}^{-\mu}\frac{\mu^k}{k!}, \qquad j = 1, \ldots, \mathrm{K}.$$

The conditional distribution of Y_t is a Frank's copula distribution

$$Y_t \mid \mathcal{F}_{t-1} \sim \mathcal{D}(\lambda_t^1, \ldots \lambda_t^{\mathrm{K}}; \rho), \tag{8a}$$

where $\mathcal{F}_{t-1}$ denotes the σ-field defined by all previous observations and exogenous covariates, $\sigma\{Y_{1-p}, \ldots, Y_{t-1}, \lambda_{1-q}, \ldots, \lambda_{t-1}, x_1, \ldots, x_{t-1}\}$, where each term contains information on all components of the time series. As before, the dynamics of the components of Y_t are specified by the equations:

$$Y_t^j \mid \mathcal{F}_{t-1} \sim \mathrm{Poisson}(\lambda_t^j), \qquad j = 1, \ldots, \mathrm{K}; \tag{8b}$$

$$\lambda_t^j = \omega^j + \sum_{l=1}^{p} \alpha_l^j Y_{t-l}^j + \sum_{l=1}^{q} \beta_l^j \lambda_{t-l}^j + \eta^j \cdot x_{t-1}^j, \qquad j = 1, \ldots, \mathrm{K}; \tag{8c}$$

where $\alpha_l^j, \beta_l^j \geq 0$ denote coefficients for the past values of the observations and intensities respectively, η^j denotes the vector of (non-negative) coefficients for the exogenous covariates, and $\omega^j \geq 0$ denotes an (optional) intercept term. For each univariate process, the two conditions in Equations (2) and (3) must hold.

2.4 Properties of multivariate PoARX

Properties such as stationarity and ergodicity for PoARX models can be based upon the univariate results [1] developed using τ-weak dependence. For notational ease, impose the simpler Markov structure found below,

$$x_t^j(k) = g^j\left(x_{t-1}^j(k); \epsilon_t^j\right), \qquad j = 1, \ldots, \mathrm{K}, \qquad k = 1, \ldots, r. \tag{9}$$

However, the statements hold for the more general structure found in Equation (3). We also make three assumptions similar to those found for the univariate model [1].

Assumption 1 (Markov) The innovations ϵ_t^j and Poisson processes $N_t^j(\cdot)$ are i.i.d. for all $j = 1, 2, \ldots, \mathrm{K}$.

Assumption 2 (Exogenous stability)

$$\mathrm{E}\left|\left|g^j\left(x^j; \epsilon_t^j\right) - g^j\left(\tilde{x}^j; \epsilon_t^j\right)\right|\right|^s \leq \kappa \left|\left|x^j - \tilde{x}^j\right|\right|^s$$

for some $\kappa < 1$ and $\mathrm{E}\left|\left|g^j\left(0; \epsilon_t^j\right)\right|\right|^s < \infty$ for all $j = 1, 2, \ldots, \mathrm{K}$, for some $s \geq 1$.

Assumption 3 (PoARX stability) $\sum_{i=1}^{\max(p,q)} \alpha_i^j + \beta_i^j < 1$, for each $j = 1, 2, \ldots, \mathrm{K}$.

In the formulae below the operator vec has its usual meaning. For a matrix A, $\mathrm{vec}(A)$ is a (column) vector obtained by stacking the columns of A on top of each other. As a shorthand, $\mathrm{vec}(A_1, \ldots, A_m)$ is equivalent to the more verbose $\mathrm{vec}(\mathrm{vec}(A_1), \ldots, \mathrm{vec}(A_m))$.

Theorem 1. *Under Assumptions 1 – 3 and the Markov assumption in Equation* (9)*, there exists a weakly dependent stationary and ergodic solution,* $X_t^* = \mathrm{vec}\left((Y_t^*, \lambda_t^*, x_{t-1}^*)\right)$*, to Equations* (8)*. The solution is such that* $\mathrm{E}\left(||X_t^*||^s\right) < \infty$*, where* $s \geq 1$ *is found in Assumption 2,* $Y_t^* = (Y_t^{*1}, \ldots Y_t^{*\mathrm{K}})^\top$ *and* $\lambda_t^* = (\lambda_t^{*1}, \ldots \lambda_t^{*\mathrm{K}})^\top$ *are* K*-vectors, and* $x_{t-1}^* = (x_{t-1}^{*1}, \ldots x_{t-1}^{*\mathrm{K}})^\top$ *is a* $\mathrm{K} \times r$ *matrix.*

Proof. See the preprint [12].

A consequence of Theorem 1 is that it allows PoARX models to use the (weak) law of large numbers (LLN) for stationary and ergodic processes. To ensure the correct analysis of asymptotic behaviour, we need to be able to use the LLN for any initialisation, rather than a set of fixed initial values. Lemma 1 extends the LLN to hold for this case.

Lemma 1. *Let* $X_t = \mathrm{vec}\left((Y_t, \lambda_t, x_{t-1})^\top\right)$ *be a process satisfying the equation* $X_t = F(X_{t-1}; \xi_t)$ *where* ξ_t *are i.i.d,* $\mathrm{E}\,||F(x; \xi_t) - F(\tilde{x}; \xi_t)||^s \leq \kappa\,||x - \tilde{x}||^s$*, and* $\mathrm{E}\,||F(0; \xi_t)||^s < \infty$*. For any function* $h(x)$ *satisfying:*

(i). $||h(x)||^{1+\delta} \leq M(1 + ||x||^s)$ *for some* $M, \delta > 0$*,*

(ii). for some $c > 0$ there exists $L_c > 0$ such that $||h(x) - h(\tilde{x})|| \le L_c \, ||x - \tilde{x}||$ for $||x - \tilde{x}|| < c$,

it holds that

$$\frac{1}{T}\sum_{t=1}^{T} h(X_t) \xrightarrow{P} \mathrm{E}\left(h(X_t^*)\right), \quad \text{as } T \to \infty.$$

Proof. See [21], or apply the main result from [22].

3 Estimation

We consider the model specified by Equations (8), where we denote the unknown parameters by ϑ. Then with $\alpha^j = \left(\alpha_1^j, \dots, \alpha_p^j\right)^\top$, $\beta^j = \left(\beta_1^j, \dots, \beta_q^j\right)^\top$, and $\eta^j = \left(\eta_1^j, \dots, \eta_r^j\right)^\top$,

$$\begin{aligned} \vartheta &= \left(\omega^1, (\alpha^1)^\top, (\beta^1)^\top, (\eta^1)^\top, \dots, \omega^{\mathrm{K}}, (\alpha^{\mathrm{K}})^\top, (\beta^{\mathrm{K}})^\top, (\eta^{\mathrm{K}})^\top, \rho\right)^\top, \\ &= \left((\theta^1)^\top, \dots, (\theta^{\mathrm{K}})^\top, \rho\right), \end{aligned}$$

where $\theta^j \in \Theta^j \subset [0, \infty)^{1+p+q+r}$.

The probability mass function of the copula PoARX model, derived from the cumulative mass function as rectangle probabilities is

$$\begin{aligned} &\Pr(Y_t^1 = y_t^1, \dots, Y_t^{\mathrm{K}} = y_t^{\mathrm{K}}) \\ &\quad = \sum_{l_1=0}^{1} \cdots \sum_{l_{\mathrm{K}}=0}^{1} (-1)^{l_1+\cdots+l_{\mathrm{K}}} C_\rho\left(F_1(y_t^1 - l_1; \lambda_t^1), \dots, F_{\mathrm{K}}(y_t^{\mathrm{K}} - l_{\mathrm{K}}, \lambda_t^{\mathrm{K}})\right), \end{aligned}$$

with $C_\rho(\cdot)$ representing Frank's copula and $F_j(\cdot)$ the Poisson distribution function for $j = 1, \dots, \mathrm{K}$. The conditional log-likelihood for ϑ given the multivariate observations $y_1, \dots, y_n$ with initial values y_0 and λ_0 (denoted by the σ-field $\mathcal{F}_0$) is given by the following.

$$l(\vartheta) = \sum_{t=1}^{n} \log\left(\Pr((y_t^1, \dots y_t^{\mathrm{K}})^\top \mid \mathcal{F}_{t-1}; \vartheta)\right) = \sum_{t=1}^{n} l_t(\vartheta).$$

With the large dimension of ϑ it is computationally more feasible to use a two-stage procedure known as the method of inference functions (IFM) [17]. We estimate the marginal parameters separately from the dependence parameter, hence reducing the dimension of the unknown parameters in each maximisation process. The marginal log-likelihood for θ^j is written as

$$l^j(\theta^j) = \sum_{t=1}^{n} \log\left(\Pr(y_t^j \mid \mathcal{F}_{t-1}; \theta^j)\right) = -\lambda_t^j + y_t^j \log(\lambda_t^j) - \log(y_t^j!), \qquad (10)$$

with λ_t^j calculated using Equation (8c). Before we state the asymptotic result we impose two further conditions [1] on the parameters and the exogenous covariates.

Assumption 4 The space of possible parameters for each marginal distribution j, Θ^j, is compact for all $j = 1, \ldots, \mathrm{K}$. This means that for all $\theta^j = (\omega^j, \alpha^j, \beta^j, \eta^j) \in \Theta^j$, $\beta_i^j \leq \beta_i^{j,U}$, for each $i = 1, \ldots, q$, and $\omega^j \geq \omega_L^j$ for some constants $\omega_L^j > 0$ and $\beta_i^{j,U} > 0$ with $\sum_{i=1}^q \beta_i^{j,U} < 1$.

Assumption 5 The polynomials $A^j(z) := \sum_{i=1}^p \alpha_{0,i}^j z^i$ and $B^j(z) := 1 - \sum_{i=1}^q \beta_{0,i}^j z^i$ have no common roots; and for any $a \neq 0$ and $g \neq 0$, $\sum_{i=1}^p a_i Y_{t-i}^{*j} + \sum_{i=1}^r g_i x_{i,t}^{*j}$ has a non-degenerate distribution. This should be true for each $j = 1, \ldots, \mathrm{K}$.

Theorem 2. *Suppose that Assumptions 1 – 5 hold with $s \geq 2$ and the true value of ϑ is denoted by ϑ_0. Then ϑ is consistent and if $\vartheta \in \operatorname{int} \Theta$,*

$$\sqrt{n}(\tilde{\vartheta} - \vartheta_0) \xrightarrow{d} \mathcal{N}(0, V), \tag{11}$$

where V is a valid covariance matrix.

Proof. See preprint [12]. Details of V can be found there.

4 Forecasting

Forecasting with PoARX models is to some extent similar to the forecasting of GARCH-X processes [13]. Predictions for the intensities can be obtained recursively using Equation (8c) and the property $\mathrm{E}(Y_t^j \mid \mathcal{F}_{t-1}) = \lambda_t^j$. This procedure also gives point predictions for the process. However, there is substantial difference when predictive distributions are required.

One-step ahead forecasts at time t of the intensities $\lambda_{t+1}^j, \ldots, \lambda_{t+h-1}^j$, given information $\mathcal{F}_t$, parameters θ^j, and covariates x_t results in a known value of $\lambda_{t+1 \mid t}^j$ using Equation (8c). By the specifications of the model, the one-step ahead marginal predictive distributions are Poisson with predicted intensities $\lambda_{t+1 \mid t}^j$. The joint predictive distribution is obtained by substituting the predicted intensities in Equation (7).

For multi-step-ahead forecasts, the procedure is not so straightforward. Firstly, the computation of the h-step-ahead forecast at time t assumes that the exogenous covariates $x_t, \ldots, x_{t+h-1}$ are known. In practice, these will often need to be replaced by their own forecasts or projections. This is not a problem when the covariates are leading indicators, see the example in Section 5. With a slight abuse of notation we use $\lambda_{t+h \mid t}^j$ to represent the "intensity for horizon h conditional on $\mathcal{F}_t$ and $x_t, \ldots, x_{t+h-1}$". We let this knowledge be denoted by the σ-field $\mathcal{G}_t$. We show below that the predictive distributions for $h \geq 2$ are not necessarily Poisson using a characteristic function-type approach [4]. Since the Poisson distribution is discrete, we use conditional probability generating functions.

The probability generating functions can be calculated as follows, starting with $h = 2$. For a time series Y_t following a PoARX process with intensity λ_t, we can write $\lambda_{t+2\,|\,t} = c_{t+2} + \alpha_1 y_{t+1}$, where c_{t+2} is measurable w.r.t. $\mathcal{G}_t$. In the derivation below we will need the following result:

$$\begin{aligned}
\mathrm{E}(\exp\left((-1+z)\alpha_1 y_{t+1}\right) \mid \mathcal{G}_t) &= \sum_{k=0}^{\infty} \frac{\lambda_{t+1}^k}{k!} \exp\left(-\lambda_{t+1}\right) \exp\left((-1+z)\alpha_1 k\right) \\
&= \exp\left(\lambda_{t+1}(-1 + \mathrm{e}^{(-1+z)\alpha_1})\right).
\end{aligned}$$

The 2-step ahead forecast has the following generating function ($P_2(z)$ depends also on t but we omit that to keep the notation transparent):

$$\begin{aligned}
P_2(z) &= \mathrm{E}(z^{Y_{t+2}} \mid \mathcal{G}_t) = \mathrm{E}(\mathrm{E}(z^{Y_{t+2}} \mid \mathcal{G}_{t+1}) \mid \mathcal{G}_t) \\
&= \mathrm{E}(\exp\left((-1+z)\lambda_{t+2}\right) \mid \mathcal{G}_t) \\
&= \exp\left((-1+z)c_{t+2}\right) \mathrm{E}(\exp\left((-1+z)\alpha_1 y_{t+1}\right) \mid \mathcal{G}_t) \\
&= \begin{cases} \exp\left((-1+z)c_{t+2}\right) & \text{if } \alpha_1 = 0, \\ \exp\left((-1+z)c_{t+2}\right) \exp\left(\lambda_{t+1}(-1 + \exp\left(-1+z\right)\alpha_1)\right) & \text{if } \alpha_1 \neq 0. \end{cases}
\end{aligned}$$

We can see that if $\alpha_1 \neq 0$, then $P_2(z)$ is not Poisson, by the uniqueness property of generating functions. The joint distribution can be obtained by computing analogously the joint probability generating functions.

For $h > 2$ the above calculation can be extended by repeatedly using the property of the iterated conditional expectation. It can also be expressed recursively as follows:

$$\begin{aligned}
P_h(z) &= \mathrm{E}(z^{Y_{t+h}} \mid \mathcal{G}_t) = \mathrm{E}(\mathrm{E}(z^{Y_{t+h}} \mid \mathcal{G}_{t+1}) \mid \mathcal{G}_t) \\
&= \mathrm{E}(P_{h-1}(z) \mid \mathcal{G}_t).
\end{aligned}$$

Clearly, for $h \geq 2$ the forecast distribution is not necessarily Poisson. Nevertheless, using iterated conditional expectations we have that

$$\begin{aligned}
\mathrm{E}(Y_{t+h} \mid \mathcal{G}_t) &= \mathrm{E}(\mathrm{E}(Y_{t+h} \mid \mathcal{G}_{t+h-1}) \mid \mathcal{G}_t) \\
&= \mathrm{E}(\lambda_{t+h} \mid \mathcal{G}_t).
\end{aligned}$$

Therefore, we can generate h-step ahead forecast of the intensity with the following equation,

$$\lambda_{t+h\,|\,t} = \omega + \sum_{l=1}^{p} \alpha_l Y_{t+h-l\,|\,t} + \sum_{l=1}^{q} \beta_l \lambda_{t+h-l\,|\,t} + \eta \cdot x_{t+h-1}. \qquad (12)$$

where

$$Y_{t+k\,|\,t} = \begin{cases} \lambda_{t+k\,|\,t} & \text{if } k > 0, \\ y_{t+k} & \text{if } k \leq 0. \end{cases}$$

Prediction intervals can be obtained by computing the probabilities from the probability generating functions discussed above. Since these are probably feasible only for small horizons, simulation would be a more practical alternative.

To obtain a prediction interval for Y^j_{t+h}, simulate a trajectory of the PoARX time series until time $t+h$, resulting in one simulated value Y^j_{t+h}. Repeating this process B times allows access to the quantiles from which we can obtain a prediction interval for the time series. Simulating a joint predictive region is an area for further work and not discussed here.

5 Applications

We illustrate the use of PoARX models with a data set originally used for event detection [14]. The computations were done with R [24] using the implementation of the PoARX models in package PoARX [11].

5.1 Data

The data contains counts of the estimated number of people that entered and exited a building over thirty-minute intervals of a UCI campus building. Counts were recorded by an optical sensor at the front door starting from the end of 23/07/2005 until the end of 05/11/2005. The data has periodic tendencies but is also influenced by events within the building causing an influx of traffic. Originally, the data was used to build a novel event detection framework under a Bayesian scheme.

We will estimate the number of people entering ($\mathrm{N}^{\mathrm{I}}(\mathrm{t})$) and exiting ($\mathrm{N}^{\mathrm{O}}(\mathrm{t})$) the building using the Poisson distribution, as found in its original use [14]. The basis of model predictions will be the lagged values of the observations and mean value, as well as some exogenous covariates. These covariates are all indicator variables, representing the following. The first is a "weekday" indicator, that takes value 1 when the day is Monday–Friday. This corresponds to an uplift for working days. The second indicator is a "daytime" indicator, taking value 1 when the time is between 07:30 and 19:30, representing an uplift in the traffic during working hours. The third indicator is associated with the presence of an event occurring. For the flow count into the building, the variable takes the value 1 when an event will occur in the next hour. For the flow out of the building, the variable takes the value 1 in the hour after an event finished. These represent the arrival and departure of people coming to the building for the event. We will investigate whether the use of Frank's copula improves the predictions.

5.2 Estimation and prediction

We fit four types of models to the data in an attempt to find the best predictive model. Models 1 and 2 contain no covariates whilst Models 3 and 4 contain all covariates. Additionally, Models 1 and 3 are modelled using independent PoARX distributions whilst Models 2 and 4 fit the joint distribution using Frank's copula. To assess the quality, we used the log score [2] and training was implemented using 5-fold cross validation [26]. We also removed some observations to use as an independent test set.

For analysis, the lagged values chosen differed slightly for each time series. For the number of people entering the building ($N^I(t)$), we chose to use 4 lagged values for the observations (lags 1, 2, 48, 336) and 1 lagged value for the means (lag 1). Lagged values from the previous 2 observations represent the flow of people within the last hour, whilst the lag of 48 corresponds to the same time point on the previous day, and 336 to the same time point on the same day in the previous week. For $N^O(t)$ we used the same 4 lagged values for the observations (lags 1, 2, 48, 336) but included an extra lag for the mean values (lags 1, 48). These were chosen based on the cross-validated log scores.

In Table 1 we present the cross-validated (training) log score and the log score from the independent (test) data for each of the four models.

Table 1: Model training scores from cross-validated fit on 4000 observations

Model number	Training log score	Test log score
1	-15444	-4184
2	-15411	-4182
3	-25088	-4190
4	-16856	-4164

Looking solely at the training score, we notice that Model 2 appears to be the best model, while Model 1 is second. It seems as though the addition of the covariates weakens the fit of the model, despite the parameters of the relevant models being significantly greater than zero, statistically speaking. Furthermore, using this metric, we deduce that the use of Frank's copula improves the predictions compared to those using the independence assumption. The smallest score and therefore the worst performance is found in the results from Model 3. This model contains covariates along with the independence assumption.

However, since we are building a predictive model we also assess the predictive capabilities on data not used in the training process, to combat any overfitting. On the external data it seems as though the first three models have similar scores whilst Model 4 produces a significant improvement in predictive accuracy. This would suggest that the combination of the time series aspects, the covariates and the multivariate modelling produces the most accurate out-of-sample predictions for this kind of data. Decomposing the improvements, we start with Models 1 and 2. There is a very small increase in performance by removing the independence assumption and using Frank's copula, but perhaps this is not worth the extra complexity gained from using a copula model. From Model 1 it appears as though adding covariates does not increase either the model fit or the predictive accuracy. As mentioned earlier, one reason for this could be the violation of the assumption of independence due to the common covariates. It is only with the combination of the covariates and the copula model that we find the best performance. To the best of our knowledge, no tests for probabilistic forecast performance exist. Formally showing that the forecasts from Model 4 are best is an area of further work, with tests such as the Diebold–Mariano test not applicable for nested models [5] or probabilistic forecasts.

6 Conclusion

We introduced the multivariate PoARX model as an extension of the univariate PoARX model. Using previously established properties of the univariate PoARX model and copulas, we stated results for stability and asymptotic normality of estimates obtained via the method of inference functions [17]. Our discussion on forecasting, especially predictive distributions for horizons larger than one, seems novel even for univariate PoARX models. In particular, it is important to point out that the predictive distributions for lags greater than one are not Poisson.

In the example in Section 5 we illustrated the use of bivariate PoARX models for modelling the counts of the number of people entering and exiting a building, using lagged values and covariates. The results of adding covariates and a copula structure separately did not suggest that either addition significantly improved the model, but the combination of time series covariates, exogenous covariates, and the dependence structure produced the best predictions. However, we acknowledge that the log score was an arbitrary choice and that any strictly proper scoring rule would have sufficed (see [10] for more details). We feel that the analysis provides material for further thought and work on model evaluation for count data time series models.

There is much scope for further work in this area. The choice of Frank's copula was made for its ability to model a negative dependence in two dimensions, but any copula could be used. Another suggestion for change would be to consider distributions other than Poisson. We are considering the possibility of using the renewal count distributions [20] implemented in the R package Rcountr [19]. Combining these renewal distributions with the ideas found in this paper could lead to a fascinating new family of count time series models.

References

1. A. Agosto, G. Cavaliere, D. Kristensen, and A. Rahbek. Modelling corporate defaults: Poisson autoregression with exogenous covariates (PARX). *Journal of Empirical Finance*, 38:640 – 663, 2016.
2. J. E. Bickel. Some comparisons among quadratic, spherical, and logarithmic scoring rules. *Decision Analysis*, 4(2):49 – 65, 2007.
3. T. Bollerslev. Generalised autoregressive conditional heteroscedasticity. *Journal of Econometrics*, 31:307 – 327, 1986.
4. G. N. Boshnakov. Analytic expressions for predictive distributions in mixture autoregressive models. *Statistical & Probability Letters*, 79(15):1704–1709, 2009.
5. F. X. Diebold. Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of Diebold–Mariano tests. *Journal of Business & Economic Statistics*, 33(1), 2015.

6. R. F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4):987 – 1008, 1982.
7. R. Ferland, A. Latour, and D. Oraichi. Integer-valued GARCH processes. *Journal of Time Series Analysis*, 27(6):923 – 942, 2006.
8. K. Fokianos, A. Rahbek, and D. Tjøstheim. Poisson autoregression. *Journal of the American Statistical Association*, 104(488):1430 – 1439, 2009.
9. C. Genest and J. G. Nešlehová. A primer on copulas for discrete data. *The ASTIN Bulletin*, 37:475 – 515, 2007.
10. T. Gneiting and A. E. Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007.
11. J. Halliday and G. N. Boshnakov. *PoARX: Fit PoARX models to multivariate time series*, 2018. R package version 0.3.2 (under development, to be published on CRAN).
12. J. Halliday and G. N. Boshnakov. PoARX modelling for multivariate count time series. *eprint arXiv:1806.04892*, 2018.
13. P. R. Hansen, Z. Huang, and H. H. Shek. Realised GARCH: A joint model for returns and realised measures of volatility. *Journal of Applied Econometrics*, 27:877 – 906, 2012.
14. A. Ihler, J. Hutchins, and P. Smyth. Adaptive event detection with time-varying Poisson processes. In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.*, pages 207 – 216. ACM Press, 2006.
15. D. I. Inouye, E. Yang, G. I. Allen, and P. Ravikumar. A review of multivariate distributions for count data derived from the Poisson distribution. *Wiley Interdisciplinary Reviews: Computational Statistics*, 9(3), 2017.
16. H. Joe. *Multivariate models and dependence concepts*. Monographs on Statistics and Applied Probability. Chapman & Hall Ltd, 1997.
17. H. Joe. Asymptotic efficiency of the two-stage estimation method for copula-based models. *Journal of Multivariate Analysis*, 94:401 – 419, 2005.
18. H. Joe. *Dependence Modeling with Copulas*. New York: Chapman and Hall/CRC., 2014.
19. T. Kharrat and G. N. Boshnakov. *Countr: Flexible univariate count models based on renewal processes*, 2016. R package version 3.2.8.
20. T. Kharrat, G. N. Boshnakov, I. G. McHale, and R. Baker. Flexible regression models for count data based on renewal processes: The Countr package (to appear). *Journal of Statistical Software*, 2018.
21. D. Kristensen and A. Rahbek. Quasi-likelihood estimation of multivariate GARCH models: A weak dependence approach. Working Papers, 2015.
22. A. M. Lindner and A. Szimayer. A limit theorem for copulas, 2005. urn:nbn:de:bvb:19-epub-1802-0.
23. R. B. Nelsen. *An Introduction to Copulas*. New York: Springer, Second edition, 2006.
24. R Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2017.
25. A. Sklar. Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de statistique de l'Université de Paris*, 8:229 – 231, 1959.
26. M. Stone. Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society B*, 36(2):111 – 147, 1974.
27. P. K. Trivedi and D. Zimmer. A note on identification of bivariate copulas for discrete count data. *Econometrics*, 5(10), 2017.