A Semi-parametric Integer-valued Autoregressive Model with Covariates*

Yao Rao[1], David Harris[2] and Brendan McCabe[1]
[1]Management School, The University of Liverpool, Liverpool, UK;
[2]The Department of Economics, The University of Melbourne, Australia

December 29, 2021

Abstract

We consider a low count data INAR (Integer Autoregressive Regression) model in which the arrivals are modelled non-parametrically and are allowed to contain covariates. Accommodating possible covariates is important as exogenous variability, such as seasonality, often needs to be catered for. The main challenge is to maintain the axiomatic properties of the arrivals non-parametric mass function while, at the same time, incorporating covariates directly into the associated probabilities. Compared with models that impose standard distributions such as Poisson or Negative Binomial for the arrivals, our approach is more flexible and provides a general arrival specification. The dependence structure is parametric and uses the standard binomial thinning operator. The parameters are estimated by the Maximum Likelihood. Monte Carlo simulations show that our proposed model performs very well with good finite sample results. Two empirical issues are addressed where incorporating covariates is a prerequisite for successful modelling. The first incorporates seasonal covariates into a semi-parametric model for forecasting the numbers of claimants of wage loss benefits in the logging industry in British Columbia, Canada. The second investigates if macro-economic indicators in an economy may be useful in predicting the number of bank failures in the US financial sector.

Keywords: *count data time-series; covariates; semi-parametric;integer autoregressive model*

1 Introduction

There has been an increasing interest in modelling discrete-valued time series, usually composed of counts of certain events or objects in specified time intervals; see the book of Weiß (2018) for a comprehensive introduction. A wide range of applications has arisen in various areas and contexts such as in the social sciences, queueing systems, experimental biology, environmental processes, economics, finance, ecology, epidemiology, international tourism demand and statistical control processes.

Cox (1981) suggested that models with time-varying parameters could be thought of as either parameter or observation-driven. In observation-driven models, the current value of the parameter is a function of the lagged dependent variable while in parameter-driven models, parameters are governed by some specified stochastic process. These ideas have percolated into the

*We thank the associate editor and three anonymous referees for their constructive and insightful comments. We are also grateful to Andy Tremayne for helpful comments and discussions. The usual disclaimer applies.
Correspondence: Yao Rao, Management School, The University of Liverpool, Chatham Street, Liverpool, L69 7ZH, U.K. Email: Y.Rao@liv.ac.uk.

count time series literature where, for example, Zeger (1988) introduced the benchmark count data model in the class of parameter-driven specifications. A few methods, such as, but not limited to, those given in West, Harrison and Migon (1985), Durbin and Koopman (1997, 2000), Shephard and Pitt (1997), have been used for parameter-driven modeling. On the other hand, the observation-driven approach is taken by Zeger and Qaqish (1988), Li (1994), Benjamin *et al.* (2003), Davis *et al.* (2003) and Zheng *et al.* (2015). It is known that parameter-driven models are generally straightforward in their interpretation of the effects of covariates on the observed count process, however, these models require considerable computational efforts in parameter estimation and forecasting (see Durbin and Koopman (2000), Jung and Liesenfeld (2001)). Compared with this, observation-driven models benefit from straightforward estimation and forecasting while interpretation of the effect of covariates can be challenging. In addition, the above mentioned approaches are parametric and often invoke the exponential family as a convenient distributional class. Unfortunately, the exponential family is not a very rich source of discrete distributions being essentially limited to the categorical, Poisson and some very restricted Negative Binomial variables. The current paper is a hybrid; it is similar in spirit to the parameter-driven approach in that an underlying stochastic process is utilized (non-parametrically) but at the same time the dependence structure is observation-driven.

The integer autoregressive (INAR) class model has been extensively studied in the literature (for example, see the survey by Weiß (2008a), Jung *et al.* (2005) and Freeland and McCabe (2005)) and has been widely applied to fields such as economics, finance, marketing, environmental studies and so on (see McCabe *et al.* (2011)). One of the most prominent features is that the INAR model explains the current count by thinning (reducing) the lagged value of the count and adding an integer valued arrivals process (see equation (4) for the model specification). Thus, this model, while nonlinear, has a similar structure to, and shares properties with the standard real-valued autoregressive model e.g., a linear conditional expectation and geometric decay of the autocorrelation function. However, the use of binomial thinning operator in the INAR process always leads to an *integer* value for the observed series after recursion and that is not the case when the real-valued autoregressive model is employed.

The arrivals process in the INAR model is typically modelled parametrically in the literature, usually relying on Poisson or Negative Binomial distributions. Nevertheless, it is commonly accepted that the use of standard parametric forms, without corroborating information, often provides only very imprecise approximations to the underlying statistical distributions when, for example, outliers, fat tails and even negative skewness cannot be discounted. Hence, Drost *et al.* (2009) considered a semi-parametric joint estimation of the INAR coefficients and the innovation distribution where no specific distributional assumption is made for the arrivals. Based on this, Jentsch and Weiß (2019) estimate jointly the INAR coefficients and the distribution of the innovations using bootstrap procedure. Our paper is similar to Drost *et al.*(2009) in that we adopt the non-parametric, distribution-free perspective for the arrivals which allows for a more robust and flexible arrivals process and does not impose any, possibly inappropriate, distributional assumptions. However, we allow the arrivals to contain possible covariates in our model and follow a different estimation approach. Indeed, some simulations presented below, suggest that there are measurable advantages in adopting the approach as an alternative to a misspecified parametric model when the DGP is arbitrary and even when a properly specified parametric alternative is available, there is little loss.

In many cases, covariates play a very important role either in constructing a well-specified model for forecasting purposes or in assessing the effect of a change in the covariate on the count process being modelled. Covariates are usually incorporated parametrically; for example, when using Poisson(λ) arrivals it is customary to set $\lambda = \exp(\mathbf{x}_t'\beta)$ to introduce covariates $\mathbf{x}_t$. Freeland (1998) incorporates covariates parametrically into the thinning process as well as the arrivals and gives identification conditions. In applications of the INAR model though, and including the ones presented here, it is often more natural to consider the effects of covariates via the arrivals process rather than the thinning operator. In particular, since we model the arrivals process non-parametrically, the challenge is to incorporate covariates directly into the arrivals probabilities themselves whilst maintaining their probabilistic properties. To the best of our knowledge, this non-parametric construction with covariates is novel. This construction also leads to the most general disturbance specification in the INAR literature. We use the acronym S-INARX to denote the semi-parametric INAR model with covariates.

The remainder of the paper is structured as follows. We introduce the model and outline some of its properties in Section 2. Estimation methods for model parameters are discussed in Section 3 and Section 4 examines the small-sample properties via Monte Carlo simulations. Covariates, both dummy and continuous, are considered in the simulations. Section 5 provides a robustness check for using our model. Section 6 studies two empirical applications, producing forecast distributions for low count times series, i.e., predicting the number of benefit claimants in British Columbia, Canada and the number of bank failures in the U.S. financial sector. Section 7 concludes the paper.

2 Models with Covariates

In the first subsection, we consider how covariates might be directly incorporated nonparametrically, into the probability mass function (pmf) of a discrete random variable with the support on $\{0, 1, 2, ..., M\}$ which bypass distributional assumptions and parametric forms. Of course, this incorporation needs to be constructed in such a way that the probabilities remain non-negative and sum to unity. In the subsequent subsection, this construction is embedded as the arrivals disturbance in an INAR model.

2.1 Discrete Variables with Covariates

The construction starts with a latent variable which contains an unobserved continuous variable u_t^* which incorporates the covariate vector $\mathbf{x}_t$. Specifically, we assume that

$$u_t^* = \mathbf{x}_t'\beta + \varepsilon_t, \quad \varepsilon_t \sim \text{i.i.d. } N(0,1), \tag{1}$$

where $\mathbf{x}_t$ excludes a constant term. The normality assumption is not essential and is used mainly for concreteness here. The discrete variables u_t are defined in terms of the continuous latent u_t^* by

$$u_t = \begin{cases} 0, & \text{if } u_t^* \leq \gamma_1 \\ 1, & \text{if } \gamma_1 < u_t^* \leq \gamma_2 \\ 2, & \text{if } \gamma_2 < u_t^* \leq \gamma_3 \\ \vdots & \\ M, & \text{if } \gamma_M < u_t^* \end{cases} \tag{2}$$

where $\gamma_1, \ldots, \gamma_M$ are suitable threshold parameters and M is the upper limit of the support. Thus, the pmf for u_t is given, for $r = 0, 1, \ldots, M$, by

$$\begin{aligned} P(u_t = r \mid \mathbf{x}_t) &= p_{r|t}^u \\ &= \Phi(\gamma_{r+1} - \mathbf{x}_t'\beta) - \Phi(\gamma_r - \mathbf{x}_t'\beta) \end{aligned} \tag{3}$$

where implicitly $\gamma_0 = -\infty$, $\gamma_{M+1} = \infty$ and Φ is the distribution function of the standard normal distribution which follows from the normal assumption in equation (1). By construction these probabilities lie in $[0, 1]$ and sum to unity for all values of $\mathbf{x}_t$ and t. At first glance, the ordered probit specification appears parametric, and it is parametric in how the covariates enter the individual arrivals probabilities. But the probit merely plays the role of a link function as in the generalized linear models (GLM) literature. The set-up retains the semi-parametric structure across the arrivals probabilities and their support, as there is no distributional assumption (e.g., Poisson or Binomial etc.) introduced across these probabilities.

2.2 The INAR model with Covariates (S-INARX)

Consider the first order integer autoregressive, INAR(1), model (Al-Osh and Alzaid (1987)), and McKenzie (1985)) for count data,

$$y_t = \alpha \circ y_{t-1} + u_t, \quad t = 1, 2, ..., T \tag{4}$$

where the operator $\circ$, known as binomial thinning, is to ensure that only integer values of $\alpha \circ y_{t-1}$ will occur. (For other specifications of thinning operations, see e.g. the survey in Weiß (2008a)). The most common thinning operator is binomial thinning which is defined as

$$\alpha \circ y_{t-1} = \sum_{i=1}^{y_{t-1}} B_{it}$$

where $B_{1t}, B_{2t}, ..., B_{y_{t-1}t}$ are independently and identically distributed Bernoulli random variables with $P(B_{it} = 1) = 1 - P(B_{it} = 0) = \alpha$.

The distribution of the discrete arrivals u_t is to be treated semi-parametrically and will also depend on some vector of (observable) covariates $\mathbf{x}_t$. Specifically, the construction the previous sub-section described in (2) and (3) is used to specify u_t. The variable M and the $\gamma's$, are treated as unknown parameters to be estimated which may or may not be of intrinsic interest. Given low count data and using INAR model with arrivals specified by such a u_t (with some moderately sized α), the value of M is not expected to be prohibitively large. For example, in the simulations below, a value of $M = 2$ is sufficient to generate a sample of observations y_t with a maximum value in the region of 10 for large enough α. Thus, the process described in (2) is a device that allows covariates to impact directly on the arrivals probabilities themselves and at the same time preserves their properties and the dynamic properties of the model. It also allows the significance of β to be tested. The interpretation of the covariate structure of the model is straightforward. If an element $x_{j,t}$ of $\mathbf{x}_t$ has a positive coefficient β_j then larger values of $x_{j,t}$ predict larger values of u_t^* and hence generally shift the arrivals distribution up. More specifically, we can derive the marginal effect of a change in $\mathbf{x}_t$ on the conditional mean as well as the probabilities associated with the y_t's.

2.3 Marginal Effects

From the model in (4), we can write the conditional expectation,

$$E(y_t|y_{t-1}) = \alpha y_{t-1} + E(u_t) \tag{5}$$

where

$$E(u_t) = \sum_{r=0}^{M} r[\Phi(\gamma_{r+1} - \mathbf{x}_t'\beta) - \Phi(\gamma_r - \mathbf{x}_t'\beta)]$$

Consider the univariate case for ease of notation, when the covariate is a continuous variable, the marginal effects for the conditional mean can be derived as

$$\frac{\partial E(y_t|y_{t-1})}{\partial x_t} = \frac{\partial E(u_t)}{\partial x_t} = -\beta \sum_{r=0}^{M} r[\phi(\gamma_{r+1} - x_t\beta) - \phi(\gamma_r - x_t\beta)] \tag{6}$$

where ϕ is probability density function of the normal distribution. It is also interesting to know how the probabilities themselves vary with the change of covariate. When $u_t \in \{0, 1, 2, ..., M\}$, we have, using $p^u_{r|t} = \Phi(\gamma_{r+1} - x_t\beta) - \Phi(\gamma_r - x_t\beta)$, the convolution

$$\begin{aligned} P(y_t|y_{t-1}) &= \Pr(Y_t = y_t|Y_{t-1} = y_{t-1}) \\ &= \sum_{r=\max(0,(y_t-M))}^{\min(y_t,y_{t-1})} \mathrm{Bi}(r; y_{t-1}, \alpha) \cdot p^u_{(y_t-r)|t} \end{aligned} \quad (7)$$

from which we can calculate derivatives with respect to x_t. When x_t is a $0-1$ dummy the change in probabilities is given by

$$\sum_{r=\max(0,(y_t-M))}^{\min(y_t,y_{t-1})} \mathrm{Bi}(r; y_{t-1} - r, \alpha) \left[\Phi(\gamma_{r+1}) - \Phi(\gamma_r) - \{\Phi(\gamma_{r+1} - \beta) - \Phi(\gamma_r - \beta)\}\right]$$

In practice, M, $\{\gamma_r\}$, β and α need to be estimated as described in Section 3.

3 Estimation

For the estimation, the log-likelihood (conditional on y_1 and $\mathbf{x}_t$) for the S-INARX model is calculated by a standard convolution argument and is given by

$$\begin{aligned} &\ell_T(\alpha, \beta, \gamma(M), M) \\ &= \sum_{t=2}^{T} \log \sum_{r=\max(0,(y_t-M))}^{\min(y_t,y_{t-1})} \mathrm{Bi}(r; y_{t-1}, \alpha) \cdot p^u_{(y_t-r)|t} \\ &= \sum_{t=2}^{T} \log \sum_{r=\max(0,(y_t-M))}^{\min(y_t,y_{t-1})} \mathrm{Bi}(r; y_{t-1}, \alpha) \cdot \left[\Phi(\gamma_{(y_t-r)+1} - \mathbf{x}_t'\beta) - \Phi(\gamma_{(y_t-r)} - \mathbf{x}_t'\beta)\right] \end{aligned}$$

where $\mathrm{Bi}(x; n, p)$ are the usual binomial probabilities. The parameters to be estimated are M, $\gamma(M) = (\gamma_1, \ldots, \gamma_M)$ and (α, β). In situations where one is exclusively interested in assessing the covariate effect, only $\hat{\beta}$ and the associated t-statistic are of interest *per se*. To estimate the parameters, starting values are needed so we set $\alpha = 0$ and $\beta = \mathbf{0}$. Then, for $M_y = \max_t y_t$, define the sample proportions

$$\widetilde{p}^u_r = T^{-1} \sum_{t=1}^{T} 1_r(y_t) \quad r = 0, 1, ..., M_y.$$

Setting $p^u_0 = \Phi(\gamma_1)$, so that $\gamma_1 = \Phi^{-1}(p^u_0)$, an estimated value of γ_1 can be obtained from $\tilde{\gamma}_1 = \Phi^{-1}(\widetilde{p}^u_0)$. Then, according to equation (3), we have $p^u_1 = \Phi(\gamma_2) - \Phi(\gamma_1)$, which implies $\tilde{\gamma}_2 = \Phi^{-1}(\widetilde{p}^u_1 + \widetilde{p}^u_0) = \Phi^{-1}(\widetilde{p}_1 + \Phi(\tilde{\gamma}_1))$ and from here we can estimate $\tilde{\gamma}_2$. The same procedure may be carried out for the remaining $\tilde{\gamma}'s$ to get $\tilde{\gamma}_r$, $r = 0, 1, ..., M_y$. Note these initial values are particularly suited to the value of the thinning α being relatively small.

Using these starting values, one could then attempt to maximize the likelihood of $\ell_T(\alpha, \beta, \gamma(M)$, $M)$ by searching over a grid of $M \in [\max(y_t - y_{t-1}), M^+]$, say, and optimizing $\ell_T(\alpha, \beta, \gamma(M), M)$ at each step with respect to α, β, γ. Then use the estimates that correspond to the largest likelihood in the set. This may be a feasible Maximum Likelihood (ML) procedure in some empirical analyses but time-consuming for large T and in simulation experiments. As an alternative, since $M_y = \max_t y_t$ is the largest number of probabilities one could ever hope to estimate semi-parametrically, it is possible to replace M in the likelihood by M_y to obtain the function $\ell_T(\alpha, \beta, \gamma(M_y), M_y)$. This function may be treated as a pseudo-likelihood and its first order conditions provide suitable estimating equations for α, β and $\gamma(M_y)$. Hence, using the starting values above, maximizing $\ell_T(\alpha, \beta, \gamma(M_y), M_y)$ yields $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\gamma}$ leading to $\hat{p}^u_{r|t} : r = 0, 1, ..., M_y$ where

$$\hat{p}^u_{r|t} = \Phi(\hat{\gamma}_{r+1} - \mathbf{x}'_t\hat{\beta}) - \Phi(\hat{\gamma}_r - \mathbf{x}'_t\hat{\beta}). \tag{8}$$

The value M is then estimated as the smallest $\hat{M}$ such that $\sum_{r=0}^{\hat{M}} \hat{p}^u_r \geq 1 - \varepsilon_\pi$, for some small tolerance ε_π with $\hat{p}^u_r = T^{-1} \sum_{t=1}^{T} \hat{p}^u_{r|t}$. The simulations below show that this approach to estimation works well providing good estimates for $\hat{\alpha}$, $\hat{\beta}$, $\hat{M}$ and $\hat{\gamma}\left(\hat{M}\right)$. In particular $\hat{\beta}_j$ and its standard error, as required by the t-statistic to test the impact of a covariate, are well estimated. The parameter restriction imposed in the estimation is $\alpha \in (0, 1)$ and we use logit transformation $\alpha = (1 + e^{-a})^{-1}$ to guarantee this. In addition, to ensure that the estimates of $\hat{\gamma}\left(\hat{M}\right)$ monotonically increase, we do the following: firstly, define $\gamma_1^* = \gamma_1$, $\gamma_2^* = \sqrt{\gamma_2 - \gamma_1}$, $\gamma_3^* = \sqrt{\gamma_3 - \gamma_2}, ..., \gamma_{M+1}^* = \sqrt{\gamma_{M+1} - \gamma_M}$ as initial values where $\gamma_1, ...\gamma_{M+1}$ can be pre-specified providing $\gamma_1 < \gamma_2 < \gamma_3 < ... < \gamma_{M+1}$; then apply Maximum Likelihood estimation as described above and denote the estimated values as $\hat{\gamma}_1^*, \hat{\gamma}_2^*, \hat{\gamma}_3^*, ..., \hat{\gamma}_{M+1}^*$; after that, we square these estimates except for $\hat{\gamma}_1^*$ and obtain $\hat{\gamma}_1^*, (\hat{\gamma}_2^*)^2, (\hat{\gamma}_3^*)^2, ..., (\hat{\gamma}_{M+1}^*)^2$; finally, take the cumulative sum for each estimates so that the final estimates of $\hat{\gamma}\left(\hat{M}\right)$ will be $\hat{\gamma}_1 = \hat{\gamma}_1^*, \hat{\gamma}_2 = \hat{\gamma}_1^* + (\hat{\gamma}_2^*)^2, \hat{\gamma}_3 = \hat{\gamma}_1^* + (\hat{\gamma}_2^*)^2 + (\hat{\gamma}_3^*)^2, ..., \hat{\gamma}_{M+1} = \hat{\gamma}_1^* + (\hat{\gamma}_2^*)^2 + (\hat{\gamma}_3^*)^2 + ...(\hat{\gamma}_{M+1}^*)^2$.

4 Monte Carlo Study

In this section, we conduct Monte Carlo experiments to examine the finite sample performance of the proposed estimator in our model (as described in the last paragraph of Section 3). The simulations are based on 1000 replications and for sample sizes $T = 100, 200$ and 500. We choose $\alpha = \{0.2, 0.5, 0.8\}$. The latent variable u_t^* is constructed according to equation (1) for a given covariate x_t and the arrivals u_t and INARX observations y_t are generated by equation (2) and (4) subsequently. The threshold parameters are set to be $\gamma_1 = 0$ and $\gamma_2 = 1.3862$ which implies the arrivals distribution will have support $\{0, 1, 2\}$ and hence $M = 2$. The unreported case for a larger M, for example $M = 5$, is also simulated and the results are qualitatively similar to the the case when $M = 2$ reported in Tables 2 and 3. For all the covariates used in our simulations, the arrivals support remains $\{0, 1, 2\}$ albeit with different associated probabilities.

Before turning attention to other parameters, we first assess how well the parameters M, γ_1 and γ_2 are estimated. We model seasonal variation by assuming the covariate x_t is a 12-month seasonal dummy variable taking values 0 and 1; specifically, the first six months of x_t is set to 0 followed by the remaining six months taking the value 1 and this pattern is replicated to complete the sample. The coefficient β is set to be 1. Table 1 summarizes the distribution of the estimated M given the true value is 2 as in the DGP. Based on the settings above, we calculate the percentage of times, out of 1000 replications, that $\hat{M}$ takes different values. The results show that $\hat{M}$ performs extremely well, for example, over 96.9% of times, M is correctly estimated when $T = 200$ and M is consistently estimated[1] as the sample size increases. Figure

[1]Since M is an integer valued, the estimator $\hat{M}$ identifies M correctly and does not reside in an small interval close to M as in the case of a real valued parameter.

1 and 2 plot the densities of the estimated γ_1 and γ_2 with different sample sizes and α. Both figures clearly show that, $\hat{\gamma}_1$and $\hat{\gamma}_2$ converge to the true values of 0 and 1.3862 respectively as the sample size increases. Figure 3 provides sample density plots (solid line) for the t-values of the estimated β for different sample sizes with $\alpha = 0.2$. It is evident that the density is very close to the standard normal distribution (dotted line) even for smaller sample sizes.

Table 1. Distribution (percentage of times) of $\hat{M}$ when $M = 2$

	$\alpha = 0.2$				$\alpha = 0.5$				$\alpha = 0.8$			
$\hat{M}$	1	**2**	3	4	1	**2**	3	4	1	**2**	3	4
$T = 200$	0.0	**96.9**	2.9	0.2	0.0	**100**	0.0	0.0	0.5	**99.5**	0.0	0.0
$T = 500$	0.0	**99.8**	0.2	0.0	0.0	**100**	0.0	0.0	0.0	**100**	0.0	0.0
$T = 1000$	0.0	**99.9**	0.1	0.0	0.0	**100**	0.0	0.0	0.0	**100**	0.0	0.0

[Figure 1. Density plot for estimated γ_1 for different α]

[Figure 2. Density plot for estimated γ_2 for different α]

[Figure 3. The distribution of $t(\hat{\beta}_1)$ when $\alpha = 0.2$ in comparison with standard normal distribution]

To evaluate the performance of other parameters estimation, we calculate the bias and Root Mean Squared Error (RMSE). We consider two cases for covariates x_t : (1) when covariates x_t is seasonal dummy; (2) x_t is a continuous variable. The simulation results are summarized in Tables 2 and 3 respectively. In all tables, columns of $m(.)$ represents the mean value for each estimated parameter and $\sum_0^2 \hat{p}_r^u$ in the last column is the value

$$\sum_{r=0}^{2} \left\{ \sum_{j=1}^{Nreps} \left[\sum_{t=1}^{T} \hat{p}_{r|t}^{u[j]} / T \right] \right\} / Nreps \tag{9}$$

where $\hat{p}_{r|t}^{u[j]}$ is the $\hat{p}_{r|t}^{u}$ of the jth replication. Under the DGP described above, the value of $\sum_{r=0}^{2} \sum_{t=1}^{T} p_{r|t}^{u}/T$ adds up to 1. Hence we would expect that the sum in (9), comprising of three estimated probabilities, would be very close to 1 as well, despite the fact that for any data set, $\sum_{t=1}^{T} p_{r|t}^{u}/T$, $r = 1, ..., M_y$ in total are calculated. Again, the simulations confirm this, i.e., the implied estimate of M is very close to 2.

With the seasonal dummy covariates x_t, the estimation results are reported in Table 2(a) for the estimation of α and β and Table 2(b) for the estimation of p_r^u. We find a monotonic gain in $RMSE$ of estimated α, β and the p_r^u's when the sample size T increases for any fixed α and this pattern applies across all the α values. On the other hand, at any given T, the estimation of all the parameters is better when α is smaller. We also consider cases with other forms of dummy covariates x_t, including: (a) x_t is generated to follow a Bernoulli distribution with $P(x_t = 1) = p_x$; (b) x_t is assumed to alternate between 0 and 1 at each time period; (c) x_t takes $\{0, 1, 2, 3, 4\}$ in sequence and then repeats over the time. The patterns, for any given

T or α, are similar to our reported experiments. The results from these settings are provided in the supplementary appendix.

Next, we investigate estimation performance when the covariate x_t is a continuous variable. All the other parameters are kept unchanged except a simple continuous trigonometric model is considered in order to mimic the effect the seasonality, i.e., x_t is defined by $\beta_1 \sin(2\pi t/12) + \beta_2 \cos(2\pi t/12)$. The coefficients β_1 and β_2 are both set to be 1. The simulation results are summarized in Table 3(a) and 3(b). As the results for $\hat{\beta}_1, \hat{\beta}_2$ are similar, only those for $\hat{\beta}_1$ are reported. It is clear that $\hat{\alpha}$ and $\hat{\beta}_1$ are very close to their true values. Furthermore, when the sample size T goes up, $RMSE$ decreases monotonically and this applies for all α values. The estimation of $p^u_{r|t}$ also performs well as reflected by the small values of the bias and $RMSE$.

In conclusion, our proposed model performs very well in all the settings above including both dummy and continuous covariates x_t. In general, at any fixed α, there is a monotonic gain in bias and $RMSE$ for all the estimated parameters as the sample size increases. And such gains apply to all α values. This is consistent with expectations as estimation typically becomes more accurate when T gets larger. On the other hand, for any given T, the bias and $RMSE$ of estimated β_1 and p^u_r increase with α which may also be expected since with less thinning in y_t, M_y tends to be larger and hence more probability parameters are estimated in the arrivals. For the estimation of α, the change in $RMSE$ is not monotonic. However, in general, $RMSE$ decreases with α as an overall pattern.

Table 2(a). Bias and RMSE for $\hat{\alpha}$ and $\hat{\beta}$ with a dummy covariate
($x_t = \{0,0,0,0,0,0,1,1,1,1,1,1,...\}$, 12 month seasonal dummy)

α	T	$m(\hat{\alpha})$	$Bias(\hat{\alpha})$	$RMSE(\hat{\alpha})$	$m(\hat{\beta})$	$Bias(\hat{\beta})$	$RMSE(\hat{\beta})$
0.2	100	0.1951	−0.0049	0.0781	1.0430	0.0435	0.4504
	200	0.1955	−0.0045	0.0548	1.0070	0.0068	0.3143
	500	0.1979	−0.0022	0.0332	0.9957	−0.0043	0.1918
0.5	100	0.4857	−0.0143	0.0806	1.0830	0.0833	0.6940
	200	0.4920	−0.0080	0.0557	1.0320	0.0320	0.3511
	500	0.4971	−0.0029	0.0346	1.0210	0.0206	0.2059
0.8	100	0.7850	−0.0150	0.0510	1.3780	0.3780	1.9202
	200	0.7937	−0.0063	0.0316	1.0870	0.0869	0.5714
	500	0.7972	−0.0028	0.0173	1.0290	0.0293	0.2478

Table 2(b). Bias and RMSE for $\hat{p}^u_{r|t}$ with a dummy covariate
($x_t = \{0,0,0,0,0,0,1,1,1,1,1,1,...\}$, 12 month seasonal dummy)

α	T	$m(\hat{p}^u_0)$	$Bias(\hat{p}^u_0)$	$RMSE(\hat{p}^u_0)$	$m(\hat{p}^u_1)$	$Bias(\hat{p}^u_1)$	$RMSE(\hat{p}^u_1)$	$m(\hat{p}^u_2)$	$Bias(\hat{p}^u_2)$	$RMSE(\hat{p}^u_2)$	$\sum_0^2 \hat{p}^u_r$
0.2	100	0.3840	0.0018	0.0022	0.3123	−0.0012	0.0013	0.3028	−0.0016	0.0016	0.9991
	200	0.3832	0.0010	0.0014	0.3129	−0.0005	0.0005	0.3037	−0.0007	0.0013	0.9998
	500	0.3834	−0.0001	0.0012	0.3126	−0.0008	0.0008	0.3040	0.0009	0.0012	1.0000
0.5	100	0.3733	−0.0089	0.0089	0.3134	−0.0001	0.0034	0.3131	0.0088	0.0092	0.9999
	200	0.3777	−0.0044	0.0044	0.3137	0.0002	0.0015	0.3086	0.0042	0.0044	1.0000
	500	0.3802	−0.0033	0.0034	0.3150	0.0016	0.0018	0.3048	0.0017	0.0022	1.0000
0.8	100	0.3611	−0.0211	0.0200	0.3098	−0.0037	0.0141	0.3291	0.0248	0.0283	1.0000
	200	0.3698	−0.0124	0.0141	0.3159	0.0024	0.0093	0.3144	0.0100	0.0100	1.0000
	500	0.3772	−0.0064	0.0064	0.3138	0.0004	0.0030	0.3091	0.0059	0.0066	1.0000

Table 3(a). Bias and RMSE for $\hat{\alpha}$ and $\hat{\beta}_1$ with a continuous covariate
$(x_t = \beta_1 \sin(2\pi t/12) + \beta_2 \cos(2\pi t/12))$

α	T	$m(\hat{\alpha})$	$Bias(\hat{\alpha})$	$RMSE(\hat{\alpha})$	$m(\hat{\beta}_1)$	$Bias(\hat{\beta}_1)$	$RMSE(\hat{\beta}_1)$
0.2	100	0.2271	0.0271	0.0990	0.9128	−0.0872	0.2815
	200	0.2311	0.0311	0.0742	0.8717	−0.1283	0.2297
	500	0.2393	0.0392	0.0577	0.8641	−0.1359	0.1818
0.5	100	0.5234	0.0233	0.0816	0.9298	−0.0702	0.3146
	200	0.5295	0.0295	0.0587	0.8800	−0.1200	0.2420
	500	0.5361	0.0361	0.0502	0.8567	−0.1433	0.1965
0.8	100	0.7984	−0.0016	0.0432	0.9610	−0.039	0.9408
	200	0.8038	0.0038	0.0261	0.8801	−0.1199	0.2577
	500	0.8078	0.0078	0.0187	0.8717	−0.1283	0.1983

Table 3(b). Bias and RMSE for $\hat{p}^u_{r|t}$ with a continuous covariate
$(x_t = \beta_1 \sin(2\pi t/12) + \beta_2 \cos(2\pi t/12))$

α	T	$m(\hat{p}^u_0)$	$Bias(\hat{p}^u_0)$	$RMSE(\hat{p}^u_0)$	$m(\hat{p}^u_1)$	$Bias(\hat{p}^u_1)$	$RMSE(\hat{p}^u_1)$	$m(\hat{p}^u_2)$	$Bias(\hat{p}^u_2)$	$RMSE(\hat{p}^u_2)$	$\sum_0^2 \hat{p}^u_r$
0.2	100	0.5132	0.0184	0.0303	0.2521	−0.0070	0.0106	0.2339	−0.0122	0.0261	0.9992
	200	0.5214	0.0188	0.0339	0.2511	−0.0053	0.0109	0.2274	−0.0136	0.0295	0.9999
	500	0.5220	0.0209	0.0344	0.2519	−0.0049	0.0109	0.2261	−0.0159	0.0306	1.0000
0.5	100	0.5153	0.0206	0.0311	0.2546	−0.0045	0.0090	0.2300	−0.0162	0.0291	0.9999
	200	0.5314	0.0287	0.0401	0.2456	−0.0107	0.0146	0.2230	−0.0180	0.0326	1.0000
	500	0.5329	0.0318	0.0437	0.2470	−0.0098	0.0149	0.2200	−0.0220	0.0373	1.0000
0.8	100	0.4988	0.0041	0.0288	0.2533	−0.0058	0.0138	0.2478	0.0017	0.0178	1.0000
	200	0.5100	0.0074	0.0297	0.2551	−0.0012	0.0098	0.2349	−0.0062	0.0242	1.0000
	500	0.5187	0.0176	0.0321	0.2488	−0.0081	0.0128	0.2325	−0.0095	0.0249	1.0000

5 Robustness check

In this section, we investigate the performance of our proposed S-INARX (semi-parametric INARX) model under misspecification. Three cases are considered: (a) the arrivals are generated as described in Section 4 and both S-INARX and P-INARX (parametric INARX) are estimated. Our intention is to assess the penalty, if any, that arises from using a (misspecified) parametric model when the DGP is semi-parametric and to compare the performance with that of using the S-INARX model; (b) the arrivals are generated from a single Poisson distribution P_1 with parameter $\lambda_1 + \beta x_t$ and the resultant u_t is used as a parametric DGP. We compare the performance of the S-INARX model with that of a Poisson based parametric P-INARX model. The idea is to investigate the price that may be paid for using the semi-parametric S-INARX model when there is a well specified parametric alternative available; (c) the arrivals are generated using a mixture of two Poisson distributions P_1 and P_2 with means $\lambda_1^* = \lambda_1 + \beta x_t$ and $\lambda_2^* = \lambda_2 + \beta x_t$ respectively to approximate an arbitrary semi-parametric DGP. It is of interest to see how S-INARX and P-INARX perform under this setting as the DGP corresponds to neither model.

One of the primary reasons for entertaining structural time series models for counts (e.g. the INARX class) is to produce forecasts and so it seems natural to compare competing models via their one-step ahead forecast distributions. A standard log scoring rule is used to assess how well the forecast distributions perform. It is expected that our S-INARX model will have better performance in the first case above and in the second case, one hopes that the S-INARX model would have a forecasting performance that is not too inferior to that of the parametric P-INARX model. We choose $\lambda_1 = 1$, $\lambda_2 = 2$, $\beta = 1$, $\alpha = 0.2$ and the mixing probability $p = 0.3$ in the simulations. Table 4 summarizes the results.

It is clear that, if the underlying distribution of the arrivals is as described in case (a) above, the S-INARX performs better than P-INARX when $T = 100$ and then with sample size increases, the P-INARX actually has larger log score than that of S-INARX. This relative performance might be expected as there are more parameters to estimate using the S-INARX when the sample size is large and M_y tends to be greater. For case (b), when the underlying distribution of the arrivals is indeed Poisson and semi-parametric estimation is used, not much is lost compared with using the (correct) parametric specification except when the sample size is small, i.e. $T = 100$. With the sample size increases, the performance of S-INARX improves significantly. Finally, if the underlying distribution of the arrivals is mixed Poisson as in case (c), then the log score associated with the S-INARX method is larger than that of the log score using P-INARX estimation. Therefore, our conclusion is that the performance of the proposed S-INARX approach is less affected by model misspecification when compared with that of parametric P-INARX estimation.

Table 4. One-step ahead forecast distributions performance

Semi-parametric INARX (S-INARX)/Parametric-INARX (P-INARX)				
		S-INARX DGP	Poisson DGP	Mixed Poisson DGP
T		log score	log score	log score
100	S-INARX	−0.3891	−1.5790	−0.5988
	P-INARX	−0.5121	−0.5748	−0.7537
200	S-INARX	−0.5075	−0.4507	−0.5037
	P-INARX	−0.2486	−0.5748	−0.5720
500	S-INARX	−0.4937	−0.5101	−0.5154
	P-INARX	−0.2500	−0.5171	−0.5698

6 Empirical Analysis

In this section, we discuss two empirical prediction applications using our proposed model, incorporating a dummy and a continuous covariate respectively[2]. In particular, we highlight the role of nonparametric forecast distributions. It is very common in prediction settings to use the mean of the forecast distribution with it being optimal in the Mean Square Error sense. In the case of low counts, this practice would invariably lead to incoherent forecasts not contained in the integer-based support of the variable under study and would require some ad hoc rounding procedure to ensure credibility. A possible solution might be to use the median of the forecast distribution with it being optimal in the Absolute Error sense. That the median forecast may itself not be very informative is exemplified by the following two situations. Let X be a discrete variable that takes two values. In the first scenario, let the distribution of X be given by $P(X=0) = 1 - P(X=1) = 0.50$ while in the second let $P(X=0) = 1 - P(X=5) = 0.90$. In both cases, the median of X is 0. However, in the second situation, there is almost twice the probability of observing a zero making that outcome far more likely. Thus, the use of a single, even coherent, summary measure of a forecast distribution may be quite uninformative and possibly misleading.

It is clearly more instructive to give the full probability distribution for all the values in the support. In addition, should information be available to quantify the losses related to the individual projected numbers, i.e. a Loss Function be available, the forecast distribution allows the associated risk function to be calculated. From there, if a single number is required, for ease in communication, say, it is then possible to use the minimax forecast, for example. Thus, there is a potential triple benefit from the nonparametric distribution approach: both inappropriate distributional assumptions and possibly misleading singleton forecasts may be avoided and the forecast distribution can facilitate better decision making.

6.1 Cuts Data Application

In this application, we analyze some time series count data, obtained from the Workers Compensation Board (WCB) of British Columbia, Canada, using our proposed semi-parametric model. This data was originally studied by Freeland (1998) and later by Freeland and McCabe (2004), Zhu and Joe (2006) amongst others. The dataset consists of monthly counts of claimants collecting Short Term Wage Loss Benefit (STWLB) for injuries received in the workplace. In the selected data set, all the claimants are male, between the ages of 35 and 54, work in the logging industry and reported their claim to the Richmond, BC service delivery location. This data set consists of 120 observations starting in January 1985 and ending in December 1994. The data series we use here are claimants whose injuries are cuts related. The thinning parameter α can be interpreted as the probability that an existing claimant will continue to collect benefit in the next month and arrivals are new claimants.

Figure 4 shows the time series plot of the cuts data. The mean of the data is 6.13 and the variance is 11.80 with the smallest observation being 1 and the largest 21 in July 1987. This indicates some type of over-dispersion in the data. The July 1987 observation is by some distance larger than the preponderance of the data and in a parametric analysis, fitting a negative binomial for example, would often be treated as an outlier and possibly be deleted from the data set to ensure a good parametric fit. However, there is no need to do so here and we estimate the semi-parametric model described in Section 3. There is also an apparent seasonality pattern in this plot which is also confirmed by the sample autocorrelation function (ACF) in Figure 5 and the summary in Table 5. Such seasonality is not surprising in this

[2] The exposition in this section is based on Freeland (1998) and Freeland and McCabe (2004)

context as we expect fewer claims in the winter months and more in the summer months in the logging industry.

[Figure 4. Time series plot of the cuts data from 1985 to 1994]

[Figure 5. Sample autocorrelation function of cuts numbers]

Table 5. Mean of the counts in each month over 1985 to 1994

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Average	4.2	3.8	4.6	4.9	7.0	7.1	8.5	7.5	7.2	7.2	7.2	4.4

We chose to model the seasonal variation, firstly, by adding a mix of sine and cosine terms to the arrival process and also by using a dummy variable for winter and summer. In the former case, the term $\beta_0 + \beta_1 \sin(2\pi t/12) + \beta_2 \cos(2\pi t/12)$ replaces $\mathbf{x}_t'\beta$ in equation (1) and is then incorporated into our model in (4). In the dummy variable case, all months from December to May are set to 0 and others are set to be 1 and the term $\beta_0 + \beta_1 D_t$ is used as $\mathbf{x}_t'\beta$ in the arrival process. The estimation of α in (4) as well as the parameters β_0, $\beta_{1,}$ (and also β_2 in the sin/cos case) are of interest. To estimate the parameters, we use the pseudo Maximum Likelihood estimation technique described in Section 3. We also plot, in Figure 6 and 7, the ACF of the standardized Pearson residuals (i.e. deviations from the model conditional mean) from fitting both the above models which evidence the adequacy of our models. The results for the estimated parameters are reported in Table 6 and $\hat{M} = 14$ in both models. Our analysis shows the seasonal regression is statistically significant at the 5% level in both cases which confirms that the seasonality covariates do explain some variation in the cuts data. The estimated α is not affected by the choice of seasonal model. When we compare the estimate $\hat{\alpha}$ obtained here using the semi-parametric approach with trigonometric covariates with that of Freeland and McCabe (2004), $\hat{\alpha} = 0.406$, who incorporated the covariates parametrically using a Poisson distribution, we see little difference suggesting the original analysis remains robust and not particularly sensitive to the parametric distributional forms used. The misspecification tests used in the original study also support this conclusion.

[Figure 6. ACF of standardized Pearson residuals in seasonal dummy model]

[Figure 7. ACF of standardized Pearson residuals in seasonal dummy model]

With suitable re-estimation of the parameters for the appropriate sub-samples, and including the seasonal dummy covariate, we can make a forecast of the number of claimants for future months. This is an important managerial issue as it provides valuable information for decision makers to allocate resources as well as for financial planning. For example, assume that the current time is August 1991 and we wish to predict a value for September 1991. Using the sin/cos model, we report in Table 7 that the highest probabilities are assigned to the numbers 8, 9 and 7 while in fact, the true number is 7 for September 1991. For the low winter season, consider December 1993; the highest probability is allocated to 4 and 5 cases and indeed 4 cases did actually occur. Results, not reported here, show similar pattern in the dummy variable case. The overall picture is also consistent with the expectation that fewer claims occur in the winter months and more in the summer months in the logging industry.

Table 6. Parameter estimates based on cuts data

Panel A: sin/cos covariate				
	$\hat{\alpha}$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$
estimate	0.4361	3.4012	-0.5721	-0.9268
s.e	0.1655	0.2675	0.2732	0.2827
t-stat	2.6356	12.7169	-2.0941	-3.2780

Panel B: winter and summer dummy variable covariate			
	$\hat{\alpha}$	$\hat{\beta}_0$	$\hat{\beta}_1$
estimate	0.4406	2.55418	1.2740
s.e	0.02141	0.03926	0.05136
t-stat	20.5749	65.0623	24.8043

Table 7. Probability distributions forecast for *Sep* 1991 and *Dec* 1993 using seasonal dummy covariate

$y_{Sep,1991} = \mathbf{7}; y_{Aug,1991} = 11$		$y_{Dec,1993} = \mathbf{4}; y_{Nov,1993} = 6$	
$P(y_t = 3 \vert y_{t-1})$	0.0021	$P(y_t = 1 \vert y_{t-1})$	0.0117
$P(y_t = 4 \vert y_{t-1})$	0.0117	$P(y_t = 2 \vert y_{t-1})$	0.0561
$P(y_t = 5 \vert y_{t-1})$	0.0421	$P(y_t = 3 \vert y_{t-1})$	0.1244
$P(y_t = 6 \vert y_{t-1})$	0.1059	$P(y_t = 4 \vert y_{t-1})$	**0.1797**
$P(y_t = 7 \vert y_{t-1})$	**0.1900**	$P(y_t = 5 \vert y_{t-1})$	**0.1951**
$P(y_t = 8 \vert y_{t-1})$	**0.2436**	$P(y_t = 6 \vert y_{t-1})$	0.1651
$P(y_t = 9 \vert y_{t-1})$	**0.2187**	$P(y_t = 7 \vert y_{t-1})$	0.1086
$P(y_t = 10 \vert y_{t-1})$	0.1309	$P(y_t = 8 \vert y_{t-1})$	0.0598
$P(y_t = 11 \vert y_{t-1})$	0.0469	$P(y_t = 9 \vert y_{t-1})$	0.0350
$P(y_t = 12 \vert y_{t-1})$	0.0077	$P(y_t = 10 \vert y_{t-1})$	0.0263

6.2 Bank Failures

Monitoring financial systems is one of the key tasks of regulatory authorities and has typically focused on bank-specific, industry-specific and macroeconomic determinants of bank failure with the objective of understanding which factors are the most useful and significant in the prediction of banking crises. There has been an increasing interest in studying bank failures, and in particular, to detect any relationship between bank failure and macroeconomic or financial indicators. For instance, Demirgüç-Kunt and Detragiache (1998) underline that elements of the macroeconomic environment, such GDP growth, excessively high real interest rates and high inflation, significantly increase the likelihood of systemic banking crises. Similarly, output growth in Kaminsky and Reinhart (1999), Louzis *et al.* (2012), and low unemployment rates in Louzis *et al.* (2012) and Ghosh (2015) are found to be negatively associated with bank failures, as a dynamic economy usually enjoys a buoyant housing sector, while accommodating monetary policies encourage banks to offer more loans. The Buch *et al.* (2010) model includes GDP growth, inflation, the Federal Funds rate, house price inflation and a set of factors summarizing conditions in the banking sector.

This subsection aims to provide an empirical analysis of U.S. commercial banks failures using our proposed S-INARX model. A framework to monitor and forecast the occurrence of bank failures in US is important in particularly for institutions such as the Federal Deposit Insurance Corporation (FDIC). Our framework contributes to the literature as follows: Firstly, instead of using aggregated annual data on bank failures, we study monthly failure occurrences. This higher frequency data has the advantage of providing an attractive monitoring strategy for regulatory authorities, by giving a finer, real-time tool for detecting any developing financial vulnerabilities in a timely manner. Secondly, in contrast to studying measures such as bank failure rates and the failure ratio that are usually of a continuous nature, we look at the numbers of bank failures which are of more practical and direct relevance and therefore more natural to use. Finally, our model can provide predictions for the entire probability distribution that are more informative than those based on mean effects. It is often the case that the policy makers are interested not only in the average effect of a policy, but also in its distributional consequences and this can be very useful information for monitoring and further analysis.

In the S-INARX framework, the contagion risk of past bank failures is captured by the thinning component of the model while the impact of the covariates is included in the nonparametric arrival specification thus avoiding the need to make, possibly contentious, distributional assumptions. The data, i.e. the monthly number of failed banks, is extracted from reports that cover FDIC insured institutions. The number of observations is 84, covering bank failures from January 2012 to December 2018. We take the data up to December 2017 as our in-sample observations and the remainder are used for out-of sample prediction.

Figure 8 provides a time-series plot of the banking failure numbers. The average is 1.5833 while the variance is 3.5985 which indicates overdispersion in the distribution of failure numbers. While other factors, including bank or industry specific ones, can also be considered, we will focus on the link between macroeconomics variables and bank failures. Hence, following the literature, we investigate, as explanatory variables, the most commonly used macroeconomic indicators, i.e., GDP growth rate, unemployment rate, inflation and the Federal fund rate in U.S. The data were obtained from the FRED Economic data base and the OECD data resource. Then we fit the model (4) by including all the above macroeconomic indicators in the arrival process (1) and estimate the coefficients of these covariates as well as α, as in the previous Section.

[Figure 8. Plot of the number of bank failures from 2012 to 2017]

Table 8 summarizes the estimation results from fitting the S-INARX model to the monthly occurrence of bank failures; $\hat{\alpha}$ is the estimated thinning parameter and $\hat{\beta}_0$ and $\hat{\beta}_1, \hat{\beta}_2$ are respectively the intercept and slope coefficients in the covariate estimation using the inflation and the Fed fund rate as covariates. Other variables e.g. the unemployment rate and GDP growth rate are found insignificant in the estimation, and therefore have been removed from the regression. Plots and the ACF of the residuals from this model suggest they may be treated as noise. Again, the ACF of the standardized Pearson residuals from fitting this model is provided in Figure 9 and suggests the adequacy of this model. For the estimation of the parameters, $\hat{M}$ = 6, $\hat{\beta}_2$ is negative which is consistent with our expectation since the lower the interest rate, the less margin that banks earn which may encourage more risky loans; taken together, these would increase the risk of bank failure. The positive impact of inflation can be explained by the fact that when inflation increases, the value of money would fall and hence discourages people's saving and investment. More broadly, as it is known that banks' profit will be affected during an economic downturn, increasing inflation, which is generally considered bad for the economy, would add more risk to the banks position. The t-statistics in Table 8 confirm that both inflation and the Fed fund rate play a significant role in modelling monthly bank failure numbers.

[Figure 9. ACF of standardized pearson residuals of using bank data]

Table 8. Parameter estimates using bank data

	$\hat{\alpha}$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$
estimate	0.2096	0.1798	1.0120	-2.369
s.e	0.1167	0.5481	0.3505	0.8157
t-stat	2.0609	0.3281	2.8890	-2.9050

With the re-estimated parameters, we can now make a probability forecast of the number of bank failures for future months. We choose two examples: The first one is based on our last in-sample data point of December 2017, i.e., $y_{t-1} = 1$, and use this make a one-step ahead out-of-sample prediction distribution for January 2018; In the second example, we choose a larger base observation, i.e., $y_{t-1} = 4$, the number of August 2013, and use this to make a distribution forecast for bank failures in September 2013. The results are summarized in Table 9. Using the covariates of inflation and Federal fund rate, we find that the highest probability of failure is assigned to the value of 0 while in fact, the true number is 0 in the first example. In the second example the highest probabilities are allocated to outcomes 1 and 2 and, indeed, 2 cases did actually occur. It is clear from the probability distribution in the second column of Table 9 that only the outcomes 0 and 1 need to be seriously considered as possibilities. On the other hand, the distribution in column four, shows that any of the outcomes from 1 to 4 has an associated non negligible probability of occurring, with outcomes 1 and 2 deemed the most likely. The larger entropy associated with the distribution is apparent. This sort of information is not readily apparent to regulatory authorities when using single summary measures alone.

Table 9. Nonparametric Forecast probability distributions for *Jan* 2018 and *Sep* 2013

$y_{Jan,2018}= \mathbf{0}$; $y_{Dec,2017}= 1$		$y_{Sep,2013}= \mathbf{2}$; $y_{Aug,2013}= 4$	
$P(y_t= 0\|y_{t-1})$	**0.6452**	$P(y_t= 0\|y_{t-1})$	0.0830
$P(y_t= 1y_{t-1})$	0.2800	$P(y_t= 1\|y_{t-1})$	**0.2564**
$P(y_t= 2\|y_{t-1})$	0.0524	$P(y_t= 2\|\|y_{t-1})$	**0.2242**
$P(y_t= 3\|y_{t-1})$	0.0116	$P(y_t= 3\|y_{t-1})$	0.1382
$P(y_t= 4\|y_{t-1})$	0.0047	$P(y_t= 4\|y_{t-1})$	0.1144
$P(y_t= 5\|y_{t-1})$	0.0009	$P(y_t= 5\|y_{t-1})$	0.0646
$P(y_t= 6\|y_{t-1})$	0.0004	$P(y_t= 6\|y_{t-1})$	0.0615

We can also look at the marginal effects as described in Section 2.3. The marginal effect on the conditional mean can be obtained from equation (6) for any given value of x_t. Again, we consider two examples with the first one in June 2012 when 7 banks actually failed, which represents the possible aftermath of financial crisis. The second example we choose is July 2016 when there were no bank failures reflecting more normal times. Given the estimated $\hat{\gamma}'s$ and $\hat{\beta}'s$, we find from Table 10 that, at an inflation baseline of 1.6639 in June 2012, a small increase therein would suggest that the conditional mean of the failed bank numbers will increase by around 0.7708. Similarly, at a baseline inflation of 0.8271 in July 2016, the conditional mean of failed bank numbers will increase by 0.4573 reflecting the environment being strong. On the other hand, when the Fed fund rate (FDR) is at 0.16 as in June 2012, the conditional mean of the failed bank numbers will decrease by around 1.8039 as compared with 1.0702 when the FDR is at a relatively high value 0.39 for July 2016. These numbers are consistent with what we expect viz. that the change in the number of bank failures under normal economic conditions is not as sensitive as that during a financial crisis and economic downturn.

Table 10. Marginal effect on the conditional mean of failed banks

Date	Inflation	Marginal Effect	Fed fund rate	Marginal Effect
Jun 2012	1.6639	0.7708	0.16	−1.8039
Jul 2016	0.8271	0.4573	0.39	−1.0702

7 Conclusion

We propose a semi-parametric, S-INARX, approach to include covariates in the arrivals of the INARX (integer autoregressive regression with covariates) model which has frequently been used in count data analysis. Our approach adopts the non-parametric, distribution-free perspective for the disturbances which allows for a more robust and flexible arrival process than in parametric models that often impose, possibly inappropriate, distributional assumptions such as Poisson or negative binomial. In contrast to the typical non-parametric situation where a function (infinite dimensional) is to be estimated, our arrivals mass function is discrete and finite dimensional and while it is true that there are a larger number of parameters to be estimated, in comparison with a parametric distributional approach, the excess is not expected to be large when modelling low counts. A larger number of parameters does impose a greater computational cost, but the Monte Carlo studies indicate that the burden is modest. In addition, to help explain the behavior of the dependant variable of interest, our model incorporates exogenous covariates directly into the arrivals probabilities whilst maintaining their probabilistic properties. This offers a more natural alternative for considering the effects of covariates than incorporation via the thinning process. To the best of our knowledge, our construction with covariates is novel and leads to the most general disturbance specification in the INAR literature.

Monte Carlo studies show that our proposed method works very well in finite samples. We further provide a robustness check with the conclusion that the performance of our proposed S-INARX model is less affected by misspecification when compared with that of a parametric counterpart, the P-INARX model. Simulations show that the S-INARX model outperforms P-INARX when the underlying distribution of the arrivals is not parametric while not being inferior when arrivals are actually parametric. This is particularly true when low counts are considered in which M is not too large. However, it is a perennial problem in forecasting that the future may differ quite substantially from the past in unknown ways. Of course, this issue is likely to be much more acute in long-range forecasting than in the short-term one-step ahead framework considered here. However, in ongoing forecasting processes, accuracy checks are continually made with model updates and improvements regularly implemented. In the current context, this would include assessments of the possible future values of M. In addition, when a sequence of one-step ahead forecasts is made the model itself would usually be re-estimated as new observations become available thereby providing updates for M.

Finally, two empirical problems have been studied in the paper. The first concerns semi-parametric probability distribution forecasts for a low count times series of benefit claimants where covariates need to be incorporated to account for seasonality; the second investigates the feasibility of using macroeconomic indicators as covariates when trying to predict the monthly number of bank failures. Further, whilst using the semiparametric approach, we have shown that one-step ahead predictions for the entire probability distribution can be produced which can be more informative for policy makers and regulatory authorities in monitoring and decision-making processes.

References

Al-Osh, M. A., and Alzaid, A. A. (1987) First-order integer valued autoregressive (INAR(1)) process. *Journal of Time Series Analysis*, 8, 261-75.

Benjamin, M. A., Rigby, R. A., and Stasinopoulos, D. M. (2003) Generalized autoregressive moving average models. *Journal of the American Statistical Association*, 98, 214–223

Buch, Claudia M., Eickmeier, S and Prieto, E. (2010) Macroeconomic Factors and Micro-Level Bank Risk. CESifo Working Paper Series 3194, CESifo Group Munich.

Cox, D. R. (1981) Statistical Analysis of Time Series: Some Recent Developments. *Scandinavian Journal of Statistics.* 8(2), 93–115.

Davis, R. A, Dunsmuir, W. T. M and Streett, S. B. (2003) Observation-Driven Models for Poisson Counts. *Biometrika*. vol 90, No.4, 777-790.

Demirgüç-Kunt, A. and Detragiache, E. (1998) The determinants of banking crises in developing and developed countries. IMF Staff papers, Palgrave Macmillan, vol 45(1), 81-109.

Drost, F.C., van den Akker, R., Werker, B.J.M. (2009) Efficient estimation of auto-regression parameters and innovation distributions for semiparametric integer-valued AR(p) models. *Journal of the Royal Statistical Society*, Series B 71(2), 467–485.

Durbin, J., and Koopman, S. J. (1997) Monte Carlo Maximum Likelihood Estimation for Non-Gaussian State-Space Models. *Biometrika*, 84, 669-684.

Durbin, J. and Koopman, S. J. (2000) Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives. *Journal of Royal Statistic Society B* 62, 3–56.

Freeland, R. K. (1998) Statistical Analysis of Discrete Time series with Application to the Analysis of Workers' Compensation Claims Data. Ph.D. Thesis, Management Science Division, Faculty of Commerce and Business Administration, University of British Columbia.

Freeland, R. K and McCabe, B. P. M. (2004) Forecasting discrete valued low count time series. *International Journal of Forecasting*. 3; 427-434.

Freeland, R. K and McCabe, B. P. M. (2005) Asymptotic properties of CLS estimators in the Poisson AR(1) model. *Statistics & Probability Letters*. 73:147–153.

Ghosh, A. (2015) Banking-industry specific and regional economic determinants of nonperforming loans: Evidence from US states. *Journal of Financial Stability*, 20: 93–104.

Jentsch, C., Weiß, C. H. (2019) Bootstrapping INAR models. *Bernoulli* 25(3), 2359-2408.

Joe, H. (1996) Time series models with univariate margins in the convolution-closed infinitely divisible class. *Journal of Applied Probability,* 33, 664-677.

Jung, R.C. and Liesenfeld, R. (2001) Estimating time series models for count data using efficient importance sampling. *Allgemeines Statistisches Archive*, 85, 387-407.

Jung, R. C., Ronning, G, and Tremayne, A. R. (2005) Estimation in conditional first order autoregression with discrete support. Statistical Papers. 46:195–224.

Kaminsky, G and Reinhart, C. M. (1999) The twin crises: The causes of banking and balance-of-payments problems. *The American Economic Review*, 89:473–500.

Li, W. K. (1994) Time Series Models Based on Generalized Linear Models: Some Further Results. *Biometrics*, 50, 506-511.

Louzis, D. P., Vouldis, A. T. and Metaxas, V. L. (2012) Macroeconomic and bank-specific determinants of non-performing loans in Greece: A comparative study of mortgage, business and consumer loan portfolios. *Journal of Banking & Finance*, 36(4): 1012–1027.

McCabe, B. P. M., Martin, G. M. and Harris, D. (2011) Efficient probabilistic forecasts for counts. *Journal of the Royal Statistical Society: Series B*. vol 73, (2), 253-272.

McKenzie, E. (1985) Some simple models for discrete variate time series. *Journal of the American Water Resources Association*, vol. 21, no. 4, 645-650.

Shephard, N., and Pitt, M. K. (1997) Likelihood Analysis of Non-Gaussian Measurement Time Series. *Biometrika*, 84, 653-667.

Weiß, C. H. (2008a) Thinning operations for modelling time series of counts-a survey. *Advances in Statistical Analysis*, 92, 319-41.

Weiß, C. H. (2018) *An introduction to discrete-Valued time series.* John Wiley & Sons Ltd.

West, M., Harrison, P. J., and Migon, H. S. (1985) Dynamic Generalized Linear Models and Bayesian Forecasting. *Journal of the American Statistical Association*, 80, 7

Zeger, S. L. (1988) A regression model for time series of counts.*Biometrika*, 75, 621-629

Zeger, S. L., and Qaqish, B. (1988) Markov Regression Models for Time Series: A Quasi-Likelihood Approach. *Biometrics*, 44,1019-1032

Zheng, T. G, Xiao, H and Chen, R. (2015) Generalized ARMA models with martingale difference errors. *Journal of Econometrics*, 189(2), 492-506.

Zhu, R and Joe, H. (2006) Modelling count data times series with Markov Processes Based on Binomial Thinning. *Journal of Time Series Analysis.* vol 27, Issue 5, 725–738.