# Supplementary Material:
# On PMF-Forecasting for Count Processes

Annika Homburg[*], Christian H. Weiß[*,†], Layth C. Alwan[‡], Gabriel Frahm[*], Rainer Göb[§]

## Contents



### Abbreviations and Acronyms

i. i. d. = independent and identically distributed
r. v. = random variable
DGP = data-generating process
PMF = probability mass function
CDF = cumulative distribution function
ACF = autocorrelation function
MSE = mean squared error
Bin = binomial
Poi = Poisson
ZIP = zero-inflated Poisson
NB = negative binomial
INAR = integer-valued autoregressive
INARCH = integer-valued autoregressive conditional heteroscedastic
BinAR = binomial autoregressive
BinARCH = binomial autoregressive conditional heteroscedastic
ll-Poi-AR = log-linear Poisson autoregressive

[*]Department of Mathematics and Statistics, Helmut Schmidt University, 22008 Hamburg, Germany
[†]Corresponding author. E-Mail: weissc@hsu-hh.de
[‡]Sheldon B. Lubar School of Business, University of Wisconsin-Milwaukee, Milwaukee, WI, USA
[§]Institute of Mathematics, Department of Statistics, University of Würzburg, Germany

## S.1 Summary of Considered Count DGPs

In what follows, we summarize those models that were used as a DGP in our simulation studies, see Section S.2 for further details. These models belong either to the group of thinning-based models or the group of regression models. The respective definition and relevant properties are briefly listed below. More details and references on these and further count time series models can be found in the book by

Weiß, C.H. (2018) *An Introduction to Discrete-Valued Time Series.*
John Wiley & Sons, Inc., Chichester.

Table 1: Relevant count distributions and their PMF.

| | |
|---|---|
| $\text{Bin}(n,\pi)$ with $n\in\mathbb{N},\pi\in(0,1)$ | $P(X=x) \;=\; \binom{n}{x}\cdot\pi^x\cdot(1-\pi)^{n-x}\qquad$ for $0\le x\le n$ |
| $\text{Poi}(\lambda)$ with $\lambda>0$ | $P(X=x) \;=\; e^{-\lambda}\,\frac{\lambda^x}{x!}\qquad$ for $x\in\mathbb{N}_0$ |
| $\text{NB}(n,\pi)$ with $n\in(0,\infty),\pi\in(0,1)$ | $P(X=x) \;=\; \binom{n+x-1}{x}\cdot(1-\pi)^x\cdot\pi^n\qquad$ for $x\in\mathbb{N}_0$ |
| $\text{ZIP}(\lambda,\omega)$ with $\lambda>0,\omega\in[0,1)$ | $P(X=x) \;=\; \mathbb{1}(x=0)\cdot\omega+(1-\omega)\,e^{-\lambda}\,\frac{\lambda^x}{x!}\qquad$ for $x\in\mathbb{N}_0$ |

### S.1.1 Thinning-based Models

The thinning-based models have AR-like DGPs, where the AR model's multiplications are substituted by the integer-valued random operation of binomial thinning: For $\alpha\in(0,1)$ and a count r. v. $X$, it is defined by requiring $\alpha\circ X\big|X\ \sim\text{Bin}(X,\alpha)$, see Table 1. The following models assume that all thinnings are executed independently of each other, of the i. i. d. count innovations $(\epsilon_t)$, and of past observations.

**INAR(1) model:** Model recursion $X_t=\alpha\circ X_{t-1}+\epsilon_t$ with $\mu_\epsilon=\text{E}[\epsilon_t]$ and $\sigma_\epsilon^2:=\text{V}[\epsilon_t]$. Mean $\mu=\text{E}[X_t]$, variance $\sigma^2=\text{V}[X_t]$, and ACF $\text{r}(k)=\text{Corr}[X_t,X_{t-k}]$, respectively, are given by

$$\mu=\frac{\mu_\epsilon}{1-\alpha}\ ,\quad I=\frac{\sigma^2}{\mu}=\frac{\frac{\sigma_\epsilon^2}{\mu_\epsilon}+\alpha}{1+\alpha},\quad\text{and}\quad \text{r}(k)=\alpha^k.$$

The model constitutes a Markov chain with transition probabilities $p(x|x_T)=p(X_{T+1}=x|X_T=x_T)$ given by

$$p(x|x_T)\;=\;\textstyle\sum_{s=0}^{\min\{x,x_T\}}\binom{x_T}{s}\alpha^s(1-\alpha)^{x_T-s}\cdot P(\epsilon_t=x-s).$$

It is referred to as Poi-, NB-, or ZIP-INAR(1) model, respectively, if $\epsilon_t$ follows a Poisson, negative binomial, or zero-inflated Poisson distribution (see Table 1).

**INAR(2) model:** Model recursion $X_t=\alpha_1\circ_t X_{t-1}+\alpha_2\circ_t X_{t-2}+\epsilon_t$ with $\alpha_1+\alpha_2<1$, constitutes a second-order Markov process with transition probabilities

$$\begin{aligned}p(x|x_T,x_{T-1})\;=\;&\textstyle\sum_{j_1=0}^{\min\{x,x_T\}}\sum_{j_2=0}^{\min\{x-x_T,x_{T-1}\}}\\ &\tbinom{x_T}{j_1}\,\alpha_1^{j_1}\,(1-\alpha_1)^{x_T-j_1}\cdot\tbinom{x_{T-1}}{j_2}\,\alpha_2^{j_2}\,(1-\alpha_2)^{x_{T-1}-j_2}\cdot P(\epsilon_t=x-j_1-j_2).\end{aligned}$$

The ACF satisfies $\text{r}(1)=\alpha_1/(1-\alpha_2)$, and $\text{r}(k)=\alpha_1\,\text{r}(k-1)+\alpha_2\,\text{r}(k-2)$ for $k\ge 2$.

**BinAR(1) model** for bounded range $\{0,\ldots,n\}$ with some $n\in\mathbb{N}$.

Let $\pi\in(0,1)$ and $\alpha\in\Big(\max\{-\frac{\pi}{1-\pi},-\frac{1-\pi}{\pi}\},1\Big)$ and define $\beta:=\pi(1-\alpha)$ and $\gamma:=\beta+\alpha$. The BinAR(1) model recursion is

$$X_t\ =\ \gamma\circ X_{t-1}\ +\ \beta\circ(n-X_{t-1})\qquad \text{with } X_0\sim \mathrm{Bin}(n,\pi).$$

It constitutes a Markov chain with marginal distribution $\mathrm{Bin}(n,\pi)$, and with ACF $\mathrm{r}(k)=\alpha^k$, and with transition probabilities

$$p(x|x_T)\ =\ \sum_{m=\max\{0,x+x_T-n\}}^{\min\{x,x_T\}} \tbinom{x_T}{m}\tbinom{n-x_T}{x-m}\,\gamma^m(1-\gamma)^{x_T-m}\,\beta^{x-m}(1-\beta)^{n-x_T+m-x}.$$

### S.1.2 Regression Models

We consider the following AR-type INARCH models.

**Poi-INARCH(1) model:** Model recursion $X_t|X_{t-1},\ldots\ \sim\ \mathrm{Poi}(\beta+\alpha\,X_{t-1})$ with $\beta>0$ and $\alpha\in(0,1)$. Mean, variance, and ACF, respectively, are given by

$$\mu=\frac{\beta}{1-\alpha}\ ,\quad \sigma^2=\frac{\mu}{1-\alpha^2},\quad \text{and}\quad \mathrm{r}(k)=\alpha^k.$$

This model constitutes a Markov chain with transition probabilities

$$p(x|x_T)\ =\ \exp(-\beta-\alpha\,x_T)\,\frac{(\beta+\alpha\,x_T)^x}{x!}.$$

**Poi-INARCH(2) model:** Model recursion $X_t|X_{t-1},\ldots\ \sim\ \mathrm{Poi}(\beta+\alpha_1\,X_{t-1}+\alpha_2\,X_{t-2})$ with $\alpha_1+\alpha_2<1$ and ACF like for the INAR(2) model. The transition probabilities are

$$p(x|x_T,x_{T-1})\ =\ \exp(-\beta-\alpha_1\,x_T-\alpha_2\,x_{T-1})\,\frac{(\beta+\alpha_1\,x_T+\alpha_2\,x_{T-1})^x}{x!}.$$

**BinARCH(1) model:** Model recursion $X_t|X_{t-1},\ldots\ \sim\ \mathrm{Bin}\Big(n,\beta+\alpha\,\frac{X_{t-1}}{n}\Big)$ with $\beta,\beta+\alpha\in(0,1)$ and transition probabilities

$$p(x|x_T)\ =\ \binom{n}{x}\,\big(\beta+\alpha\,\tfrac{x_T}{n}\big)^x\,\big(1-\beta-\alpha\,\tfrac{x_T}{n}\big)^{n-x}.$$

## S.2 Simulation Study

For the DGPs described in Section S.1 and for each corresponding scenario according to Table 2, we simulated 1,000 time series and fitted the respective model to the data. Here, we used the method of moments together with the moment formulae provided by Section S.1. Then, the PMF forecasts (or CDF forecasts, respectively) were computed according to the formulae for the transition probabilities in Section S.1. These PMF or CDF forecasts were used to compute the different types of MSE described in Section 3:

- global MSEs $\|\widehat{\boldsymbol{p}}-\widehat{\boldsymbol{p}}_0\|^2 = \sum_{x=0}^{\infty}(\widehat{p}_x-\widehat{p}_{0,x})^2$ (coherent PMF), $\|\widehat{\boldsymbol{f}}-\widehat{\boldsymbol{f}}_0\|^2$ (coherent CDF), $\|\widehat{\boldsymbol{p}}_{\mathrm{a}}-\widehat{\boldsymbol{p}}_0\|^2$ (approximate PMF), and $\|\widehat{\boldsymbol{f}}_{\mathrm{a}}-\widehat{\boldsymbol{f}}_0\|^2$ (approximate CDF);
- local MSEs $\sum_{x=0}^{\infty}(\widehat{p}_x-\widehat{p}_{0,x})^2\,\mathbb{1}(\widehat{f}_{0,x}\leq 0.25)$ (lower-25% tail MSE for coherent PMF) and $\sum_{x=0}^{\infty}(\widehat{p}_x-\widehat{p}_{0,x})^2\,\mathbb{1}(\widehat{f}_{0,x}\geq 0.90)$ (upper-10% tail MSE for coherent PMF), and the respective tail versions for approximate and CDF forecasts.

Besides looking at the MSE values for the coherent and approximate forecasting, respectively, we computed also the difference between the approximate and coherent MSE, such as $\|\widehat{\boldsymbol{p}}_{\mathrm{a}}-\widehat{\boldsymbol{p}}_0\|^2-\|\widehat{\boldsymbol{p}}-\widehat{\boldsymbol{p}}_0\|^2$, where a value $>0$ implies that the approximate MSE was larger than the coherent one.

Table 2: Scenarios for different DGPs of simulation study, with 1,000 replications each.

| |
|---|
| Means $\mu\in\{1,1.075,\ldots,9.925,10\}$ for unbounded counts,<br>upper bounds $n\in\{10,\ldots,130\}$ and probability $\pi\in\{0.15,0.45\}$ for bounded counts. |
| Dispersion ratios $I\in\{1.4,2.4\}$ if considering overdispersion. |
| Dependence parameter $\alpha$ in $\{0.33,0.55,0.8\}$ (ACF at lag 1),<br>and $\alpha_2\in\{0.25,0.35,0.45\}$ as well as $\alpha_1=\alpha\,(1-\alpha_2)$ for AR(2)-like models. |
| Sample sizes $T\in\{75,250,2500\}$. |

Either the 1,000 MSE values per scenario themselves, or the 1,000 MSE differences, were analyzed by using a lean type of boxplot:

- The median is plotted as a black dot,
- the quartiles are connected by a thick grey line, and
- the 10%- and 90%-quantiles are connected by a thin black line.

These boxplots are then plotted against increasing mean $\mu$. The results are summarized as `pdf`-files in the folder `Full_Results`. There, we distinguish between subfolders for PMF and CDF forecasting, between the simple normal approximation (i. e., without continuity correction) and the continuity-corrected one, and between boxplots for the MSE values and for the MSE differences.