Received: 13 April 2020 | Revised: 6 July 2020 | Accepted: 16 September 2020
DOI: 10.1002/for.2729

RESEARCH ARTICLE

WILEY

A performance analysis of prediction intervals for count time series

Annika Homburg[1] | Christian H. Weiß[1] | Layth C. Alwan[2] | Gabriel Frahm[1] | Rainer Göb[3]

[1]Department of Mathematics and Statistics, Helmut Schmidt University, Hamburg, Germany

[2]Sheldon B. Lubar School of Business, University of Wisconsin–Milwaukee, Milwaukee, Wisconsin

[3]Institute of Mathematics, Department of Statistics, University of Würzburg, Würzburg, Germany

Correspondence
Christian H. Weiß, Department of Mathematics and Statistics, Helmut Schmidt University, 22008 Hamburg, Germany.
Email: weissc@hsu-hh.de

Funding information
Deutsche Forschungsgemeinschaft, Grant/Award Number: 394832307

Abstract

One of the major motivations for the analysis and modeling of time series data is the forecasting of future outcomes. The use of interval forecasts instead of point forecasts allows us to incorporate the apparent forecast uncertainty. When forecasting count time series, one also has to account for the discreteness of the range, which is done by using coherent prediction intervals (PIs) relying on a count model. We provide a comprehensive performance analysis of coherent PIs for diverse types of count processes. We also compare them to approximate PIs that are computed based on a Gaussian approximation. Our analyses rely on an extensive simulation study. It turns out that the Gaussian approximations do considerably worse than the coherent PIs. Furthermore, special characteristics such as overdispersion, zero inflation, or trend clearly affect the PIs' performance. We conclude by presenting two empirical applications of PIs for count time series: the demand for blood bags in a hospital and the number of company liquidations in Germany.

KEYWORDS

coherent forecasting, count time series, estimation error, Gaussian approximation, prediction interval

1 | INTRODUCTION

One of the major motivations for doing a time series analysis is to enable the forecasting of future outcomes of the underlying process. Often, this is done by computing point forecast values, but these might be misleading because of masking uncertainty and pretending spurious accuracy. These problems are avoided if computing interval forecasts instead. Furthermore, a *prediction interval* (PI) also allows us to prepare different strategies for the range of possible outcomes implied by the PI (Chatfield, 1993). There has been much research regarding PIs for continuous-valued time series; see, for example, the review articles by Chatfield (1993) and de Gooijer and Hyndman (2006). In the present work, however, we are interested in an important class of discrete-valued time series, namely *count time series*, which consist of quantitative observations from the set of nonnegative integers, $\mathbb{N}_0 = \{0, 1, \ldots\}$. Count time series have attracted the interest of researchers and practitioners over recent years (Weiß, 2018). Several models of count time series have been proposed in the literature (see Appendix A for a small selection), and they have been used in diverse application scenarios ranging from health to business (see also Section 4). There have been a few articles about PIs for common types of count distribution (without a time aspect); see the review by Hahn and Nelson (1973) for early works in this field, and the articles by Wang (2008), Krishnamoorthy and Peng (2011), and Bejleri and Nandram (2018) for more recent contributions and references. In view of our later investigations, it is worth pointing out that many of these PIs rely on Gaussian approximations of the actual count distribution, but articles on PIs for count time series are rare. Lambert (1997) and Mukhopadhyay and Sathish (2018) developed predictive-likelihood-based PIs for generalized autoregressive moving-average (ARMA) models, and data applications were reported by Freeland and McCabe (2004) and Bejleri and Nandram (2018). The article by Silva, Pereira, and Silva (2009) proposed a Bayesian PI for the Poisson integer-valued autoregressive process of order 1 (abbreviated "INAR(1)") due to McKenzie (1985), and it also presented some performance analyses. However, a comprehensive study of PI performance for various types of count processes $(X_t)_{t \in \mathbb{Z} = \{\ldots,-1,0,1,\ldots\}}$ is yet missing.

Remark 1. Some authors propose another way of generating "trustworthy" coherent forecasts, where the outcome of the prediction is the full forecast distribution; see McCabe and Martin (2005), Snyder, Ord, and Beaumont (2012) and Kolassa (2016). This approach can also be combined with Bayesian forecast methodology, as demonstrated by McCabe and Martin (2005) for the aforementioned INAR(1) models. On the one hand, the full forecast distribution is more informative than a PI, because it allows to judge, for example, whether some counts are more likely to be observed than others. On the other hand, it may overstrain the practitioner to capture all this information and to draw appropriate conclusions from them, whereas the well-established concept of PIs allows for an intuitive interpretation. So both forecasting concepts have their pros and cons, and in this article we focus on coherent forecasting by PIs.

A comprehensive performance analysis of PIs for count processes constitutes the first main objective of this article. In addition, we shall distinguish between PIs that explicitly rely on a count time series model (these are referred to as "coherent PIs") and PIs that build upon a Gaussian ARMA approximation. The details are described in Section 2. Notwithstanding the rich literature on forecasting count processes, ARMA approximations of count time series are still popular among practitioners for several reasons. First, if the count values are large, then Gaussian ARMA models may serve as suitable approximating models; such approximations continue to be standard practice (Chintalapudi, Battineni, & Amenta, 2020). Studies that use ARMA approximations for high count series rarely, if ever, come with warnings of their inappropriateness for low count series, leaving the general practitioner with the impression of universal application of these approximations. Second, the reviews of forecasting methods found in many textbooks and in the literature (e.g., Rahardja, 2020) focus exclusively on the continuous-based methods with no mention of the unique challenges of integer-based forecasting. Finally, these approximations are actually encouraged in many software platforms; for example, with Orcale's ARIMA forecasting within Crystal Ball, there is an option for the user to round the forecast values (Oracle, 2017, p. 48). Therefore, as an important second main objective, we compare the performance of such approximate PIs to the performance of the coherent PIs. It turns out that the Gaussian approximations perform substantially worse than the coherent PIs being based on an exact count model. The detailed results of our analyses are presented in Section 3. There, we start with the aforementioned Poisson INAR (1) model (Section 3.1) as our baseline model, and we extend the analysis to the following types of data-generating process (DGP):

- INAR(1) processes with overdispersion and zero inflation (Section 3.2);
- INARCH(1) processes as alternative AR(1)-like count DGPs (Section 3.3);
- higher order autoregressions within the INAR and INARCH family (Section 3.4);
- processes of bounded counts from these families (Section 3.5) — that is, where the range of generated counts is bounded from above by some given threshold value $n \in \mathbb{N}$;
- nonstationary count processes with seasonality or trend (Section 3.6).

A brief summary of definition and properties of the considered count time series models is provided by Appendix A. Our analyses are illustrated with selected figures and tables. Further results can be found in Supporting Information Supplement S as well as at https://www.hsu-hh.de/mathstat/en/research/projects/forecastingrisk. Section 4 investigates PIs being computed for two real-data examples: a count time series regarding the demand for blood bags in a hospital, and another one about the daily numbers of company liquidations in Germany. Finally, we conclude in Section 5.

2 | INTERVAL FORECASTS FOR COUNT TIME SERIES

2.1 | Coherent and approximate PIs

Given the count time series $x_1, \dots, x_T$ up to time $T \in \mathbb{N} = \{1, 2, \dots\}$, the aim is to compute a PI $[x_l^*, x_u^*]$ for the count X_{T+h} to be observed at time $T+h$, where $h \in \mathbb{N}$ denotes the forecast horizon. Note that it would be more correct to write $[x_{l,T+h}^*, x_{u,T+h}^*]$, because the actual PI depends both on T and on h. But to simplify the reading, we suppress the subscript "$T+h$" in the sequel. The PI $[x_l^*, x_u^*]$ is to be computed such that a given *coverage level* p_{cov} is ensured:

$$P(x_l^* \leq X_{T+h} \leq x_u^* | x_T, \dots, x_1) \geq p_{\text{cov}}.$$

Note that we also have to include the case of exceeding p_{cov}, because for a discrete random variable (r.v.) one can usually not meet the intended coverage level exactly. In our performance analyses in Section 3, we consider the choice $p_{\text{cov}} = 0.90$ for illustration; that is, the true coverage should be at least 90%.

Since $X_1, X_2, \dots$ is a count process, we are actually concerned with a finite prediction set; that is, it is possible to find integers $0 \leq x_l \leq x_u < \infty$ such that $[x_l^*, x_u^*] \cap \mathbb{N}_0 = \{x_l, \dots, x_u\}$. Therefore, from now on, our aim is to find such integer-valued bounds $0 \leq x_l \leq x_u < \infty$ with

$$P(x_l \leq X_{T+h} \leq x_u | x_T, \dots, x_1) \geq p_{\text{cov}}.$$

These bounds are determined based on the forecast distribution derived for $X_{T+h}|x_T, \dots, x_1$. If we uniquely set $x_l \equiv 0$, we refer to the PI as being *upper-sided*, whereas it is referred to as *two-sided* if also $x_l > 0$ is possible.

We compare the performance of *coherent* PIs (i.e., if a count model is used for $X_{T+h}|x_T, \dots, x_1$, either the true or a fitted one) with that of *approximate* PIs. The latter are computed by assuming a Gaussian approximation to the distribution of $X_{T+h}|x_T, \dots, x_1$. This distinction and terminology are borrowed from the point forecasting of count processes; see Freeland and McCabe (2004) and Homburg, Weiß, Alwan, Frahm, and Göb (2019) for further details. For the approximate PIs, we consider an approximating Gaussian process $Y_1, Y_2, \dots$ Since this is continuously distributed, the PIs can be chosen to meet p_{cov} exactly; that is, the exceedance of p_{cov} can be avoided under a Gaussian model assumption. Thus we compute a corresponding PI $[y_l^*, y_u^*]$ for Y_{T+h}, given that $y_T = x_T, \dots, y_1 = x_1$, such that

$$P(y_l^* \leq Y_{T+h} \leq y_u^* | x_T, \dots, x_1) = p_{\text{cov}}.$$

Then, we derive the resulting approximate integer-valued prediction set as $\{y_l, \dots, y_u\} = [y_l^*, y_u^*] \cap \mathbb{N}_0$ like before. The integer bounds can be computed as $y_l = \text{ceiling}(y_l^*)$ and $y_u = \text{floor}(y_u^*)$. Note that we have $y_l^* \leq X_{T+h} \leq y_u^*$ iff $y_l \leq X_{T+h} \leq y_u$.

2.2 | Computation of interval forecasts

For the sake of readability, in this and the next section, we suppress the time dependence of the r.v. X to be forecasted (and again of the integer-valued bounds $0 \leq x_l \leq x_u$ of the PI). The DGP behind X depends on some parameters the true values of which are summarized in the parameter vector $\boldsymbol{\theta}$. If forecasting based on a fitted model, we denote the corresponding estimate by $\hat{\boldsymbol{\theta}}$. Analogously, the parameter values of the Gaussian approximation are denoted by $\boldsymbol{\vartheta}$, the corresponding estimate by $\hat{\boldsymbol{\vartheta}}$.

Example 1. The first type of DGP to be considered in Section 3 is the *Poisson INAR(1) process*$(X_t)_{\mathbb{Z}}$ proposed by McKenzie (1985); see Appendix A for details. The Poi-INAR(1) DGP is fully specified by fixing the parameter values of $\boldsymbol{\theta} = (\mu, \alpha)$. Because of the Poisson's equidispersion property, the variance σ^2 has to equal the mean μ in this case. Since this limitation is often violated in practice, we shall also consider INAR(1) processes having innovations from a negative binomial (NB) or zero-inflated Poisson (ZIP) distribution; see Section 3.2. Then, the observations exhibit overdispersion (i.e., $\sigma^2 > \mu$), and we have to include a third parameter in $\boldsymbol{\theta}$. We use the dispersion ratio $I = \sigma^2/\mu$ for this purpose.

A continuous counterpart to the INAR(1) model is given by the Gaussian AR(1) model:

$$Y_t - \mu_Y = \phi(Y_{t-1} - \mu_Y) + \varepsilon_t \text{ with i.i.d. } \varepsilon_t \sim \text{N}(0, \sigma_\varepsilon^2). \quad (1)$$

It is often used by practitioners to approximate INAR (1)-like count models because of the similar autocorrelation structure. The Gau-AR(1) model is fully specified if the values of the marginal mean μ_Y, the variance σ_Y^2, and the autocorrelation parameter ϕ have been fixed; that is, $\boldsymbol{\vartheta} = (\mu_Y, \sigma_Y^2, \phi)$.

An *upper-sided* coherent PI for a target coverage level p_{cov} is determined by setting $x_l = 0$ and x_u equal to the p_{cov}-quantile of the forecast distribution for X; that is, $x_u = \min\{u \in \mathbb{N}_0 | P_{\hat{\theta}}(X \leq u) \geq p_{\text{cov}}\}$. For the upper-sided approximate PI, we also set the lower bound $y_l = 0$. Then, we first compute y_u^* as the p_{cov}-quantile of the Gaussian approximate forecast distribution. The resulting integer-valued approximate upper bound y_u follows as $y_u = \text{floor}(y_u^*)$; that is, the integer r.v. X exceeds y_u^* iff it exceeds y_u. At this point, let us recall that the quantiles for a continuously distributed r.v. can be chosen such that they meet the intended quantile level exactly. For a discrete r.v., however, the computed quantiles usually exceed the nominal quantile level. This discreteness effect has to be kept in mind also when determining a discrete two-sided PI.

The *two-sided* coherent PI for a target coverage level p_{cov} is determined by the following algorithm:

1. First, compute the largest integer $L \in \mathbb{N}_0$ such that $P_{\hat{\theta}}(X < L) \leq 1 - p_{\text{cov}}$.
2. Then, for all $l = 0, \ldots, L$, compute the smallest integer $u = u(l)$ such that $P_{\hat{\theta}}(l \leq X \leq u) \geq 1 - p_{\text{cov}}$.
3. Among the $L+1$ resulting PIs, choose the one(s) having minimal length.
4. If there exist several PIs $[x_{l,i}, x_{u,i}]$ of minimal length, then choose $[x_l, x_u]$ as the one with greatest coverage:

$$P_{\hat{\theta}}(X \in [x_l, x_u]) = \max_i P_{\hat{\theta}}(X \in [x_{l,i}, x_{u,i}]).$$

This algorithm also allows for intervals with $x_l = 0$ (so actually upper-sided intervals) if they happen to be the optimal choice in the sense of the algorithm. Note that because of the discreteness, even if the true values of the model parameters would be known, it is usually not possible to meet p_{cov} exactly. Therefore, the algorithm also allows to exceed p_{cov} but ensures a PI of minimal length (step 3). The motivation behind the step 4 (if it comes into effect at all) is to choose the greatest coverage "for the same price."

The two-sided approximate PI is again computed based on a Gaussian approximation. Assuming a Gaussian model, in turn, the common approach is to determine the PI's limits such that p_{cov} is reached exactly. So, we first compute y_l^* as the Gaussian $\frac{1-p_{\text{cov}}}{2}$-quantile, and y_u^* as the $\frac{1+p_{\text{cov}}}{2}$-quantile. Then, we define the integer-valued approximate bounds by $y_l = \text{ceiling}(y_l^*)$ and $y_u = \text{floor}(y_u^*)$, as explained at the end of Section 2.1.

Example 2. Let us continue Example 1. The h-step-ahead conditional forecast distribution of the INAR(1) process can be computed based on the convolution of the binomial distribution with the innovations' distribution; see Appendix A for details. In the particular case of the Poi-INAR(1) model, it becomes

$$\begin{aligned} P(X_{T+h} &= x | X_T = x_T) \\ &= \sum_{s=0}^{\min\{x, x_T\}} \binom{x_T}{s} (\alpha^h)^s (1-\alpha^h)^{x_T - s} \\ &\cdot \frac{e^{-\mu(1-\alpha^h)}}{(x-s)!} \left[\mu(1-\alpha^h)\right]^{x-s}. \end{aligned} \tag{2}$$

If approximating the Poi-INAR(1) by the Gau-AR (1) process, the h-step-ahead conditional distribution is computed via

$$\begin{aligned} Y_{T+h} | Y_T &= y_T \\ &\sim \mathrm{N}\left[\alpha^h y_T + \mu_Y (1-\alpha^h), \sigma_Y^2 (1-\alpha^{2h})\right]. \end{aligned} \tag{3}$$

In Section 3, we assume the model parameters to be unknown, so they are estimated for both the true model (Equation (2)) and the approximate model (Equation (3)). We use Yule–Walker estimation for this purpose, and the approximating Gauss model is directly fitted to the count data. In Supporting Information Supplement S.1, we also consider the known-parameter case to extract the pure effect of discreteness on the PIs' performance.

When working on our two major objectives — the performance analysis of coherent PIs for count processes as well as the comparison to the performance of approximate PIs — it turned out that the main findings can already be derived for the case of the forecast horizon being $h = 1$. Therefore, in view of a concise presentation, we restrict the main article to the choice $h = 1$. In Supporting Information Supplement S.4, however, we also provide some results regarding the forecast horizon $h > 1$ for completeness.

2.3 | Performance evaluation of interval forecasts

To analyze the performance of the computed PIs, and thus to get an idea about the effect of approximation and/or estimation error, we consider several performance metrics. For every approximate or estimated model, we compute the *true coverage* for each PI $[x_l, x_u]$ — that is, the true probability of X falling into that interval. The interval $[x_l, x_u]$ satisfies the given coverage requirement if $P_\theta(X \in [x_l, x_u]) \geq p_{\text{cov}}$.

Related to these coverages, we define several "overall performance metrics." Let $\mathbb{1}(A)$ denote the indicator function, taking the value 1 if A holds true, and 0 otherwise. For a given scenario (later, we shall be more precise about the actual meaning of "scenario"), we determine the set of all simulated coverage values, say $\{c_1, c_2, \ldots\}$, and compute the following sample statistics:

- the "shortfall rate" — that is, the proportion of coverages not satisfying the coverage requirement (relative frequencies of $c_i < p_{\text{cov}}$; computed as the mean of all $\mathbb{1}(c_i < p_{\text{cov}})$);

- the "average shortfall" — that is, the average amount of falling below p_{cov} (the mean of $c_i - p_{\text{cov}}$ given that $c_i < p_{\text{cov}}$ — that is, the mean of all $(c_i - p_{\text{cov}})\mathbb{1}(c_i < p_{\text{cov}})$ divided by the mean of all $\mathbb{1}(c_i < p_{\text{cov}})$);
- the "average exceedance" — that is, the average amount of exceeding p_{cov} (computed as the mean of all $(c_i - p_{\text{cov}})\mathbb{1}(c_i > p_{\text{cov}})$ divided by the mean of all $\mathbb{1}(c_i > p_{\text{cov}})$);
- the sample standard deviation among all c_i.

The first two performance metrics are considered most important, because we do not want to have any PI violating the coverage requirement. We therefore aim at a shortfall rate being zero, and if there are shortfalls the average about these shortfalls should be close to zero. But it would also be nice to meet the nominal coverage as close as possible; that is, the average exceedance should also be close to zero. Finally, if the standard deviation among all c_i equaled zero, this would imply that all intervals have exactly the same coverage value; that is, we would have a stable coverage performance within the considered scenario. We have to recall, however, that for discrete count data it is usually impossible to meet the given coverage requirement exactly, so a certain extent of exceedance and variation among the realized coverages is natural. Modifying the parametrization of the given DGP, also the set of attainable coverage values will change. Thus, at least except for some artificial scenarios, both the average exceedance and the standard deviation will be truly positive. Such discreteness effects are studied in more detail in Supporting Information Supplement S.1.

The above performance metrics (being related to the true coverage) are considered as most important for practice. Therefore, the analyses presented in the main article focus on these metrics. Nevertheless, we also computed a few further performance metrics, namely the interval length as well as an asymmetry measure for two-sided PIs. The obtained results are briefly discussed in Supporting Information Supplements S.2 and S.3, respectively. Another option for performance evaluation would be to use the "interval score" proposed by Gneiting and Raftery (2007).

3 | PERFORMANCE OF INTERVAL FORECASTS

For each parameter combination of $(\boldsymbol{\theta}, T)$ (out of >20,000 such combinations), and for the forecast horizon $h = 1$ (the case $h > 1$ is briefly discussed in Supporting Information Supplement S.4), we simulated 1,000 count time series. So, altogether, more than 20 million count time series were simulated and analyzed, which was only possible by the intensive use of parallel computing. The relevant (true or approximate) model was fitted to the data, and the interval forecasts were computed based on the latest observations in each time series. Then, we evaluated their performances using the metrics presented in Section 2.3 (see Supplements S.2 and S.3 for a discussion of additional performance metrics). Having 1,000 replications per $(\boldsymbol{\theta}, T)$, the standard error is <0.01 for each coverage value. The choice of DGPs and their parametrizations was made as in Homburg et al. (2019), who studied the performance of point forecasts.

3.1 | Poi-INAR(1) DGP

Here, and in subsequent sections, we use a lean type of "boxplot" for the visual performance analysis, showing the 10%, 25%, 50%, 75%, and 90% quantiles of the 1,000 simulated true coverages per scenario. In each of these "modified boxplots," the median is represented by a black square, and the lower and upper quartiles are connected by a thick gray line. The 10% and 90% quantiles, in turn, are connected by a thin black line. Note that because of the discreteness of count data, it may happen that some (or even all) of these quantiles agree with each other (also see Supporting Information Supplement S.1). Thus some types of lines might sometimes not be visible.

Figures 1 and 2 show boxplots of the true coverages of the upper-sided and two-sided PIs, respectively, for a given choice of (α, T). The boxplots refer to different choices of μ: each boxplot comprises the coverages resulting from 1,000 simulated Poi-INAR(1) time series under estimation uncertainty, and these boxplots are plotted against the mean μ. The last observation x_T, as it is used for computing the PI, is chosen as the respective last observation of each simulated time series.

In the respective top row of Figures 1 and 2, referring to the coherent PIs, we observe that only a few boxplots violate the dotted line corresponding to the nominal coverage $p_{\text{cov}} = 0.90$; that is, despite estimation uncertainty, only a small fraction of PIs does not satisfy the given coverage requirement. This differs considerably from the case of approximate PIs (respective bottom row of Figures 1 and 2), which show that a large fraction of intervals has a coverage $<p_{\text{cov}}$. Furthermore, the variation among the actually attained coverages is much larger in the case of approximate PIs. So there is not only a high risk of obtaining less coverage than required if using an approximate PI, but there is also much more uncertainty about the actual true coverage. It can also be seen that an increased amount of autocorrelation causes more variation in the coverage values. Regarding the marginal mean μ, we observe that the boxplots for the coherent PIs are closer to p_{cov} for $\mu \geq 4$ than for $\mu < 4$, both for the

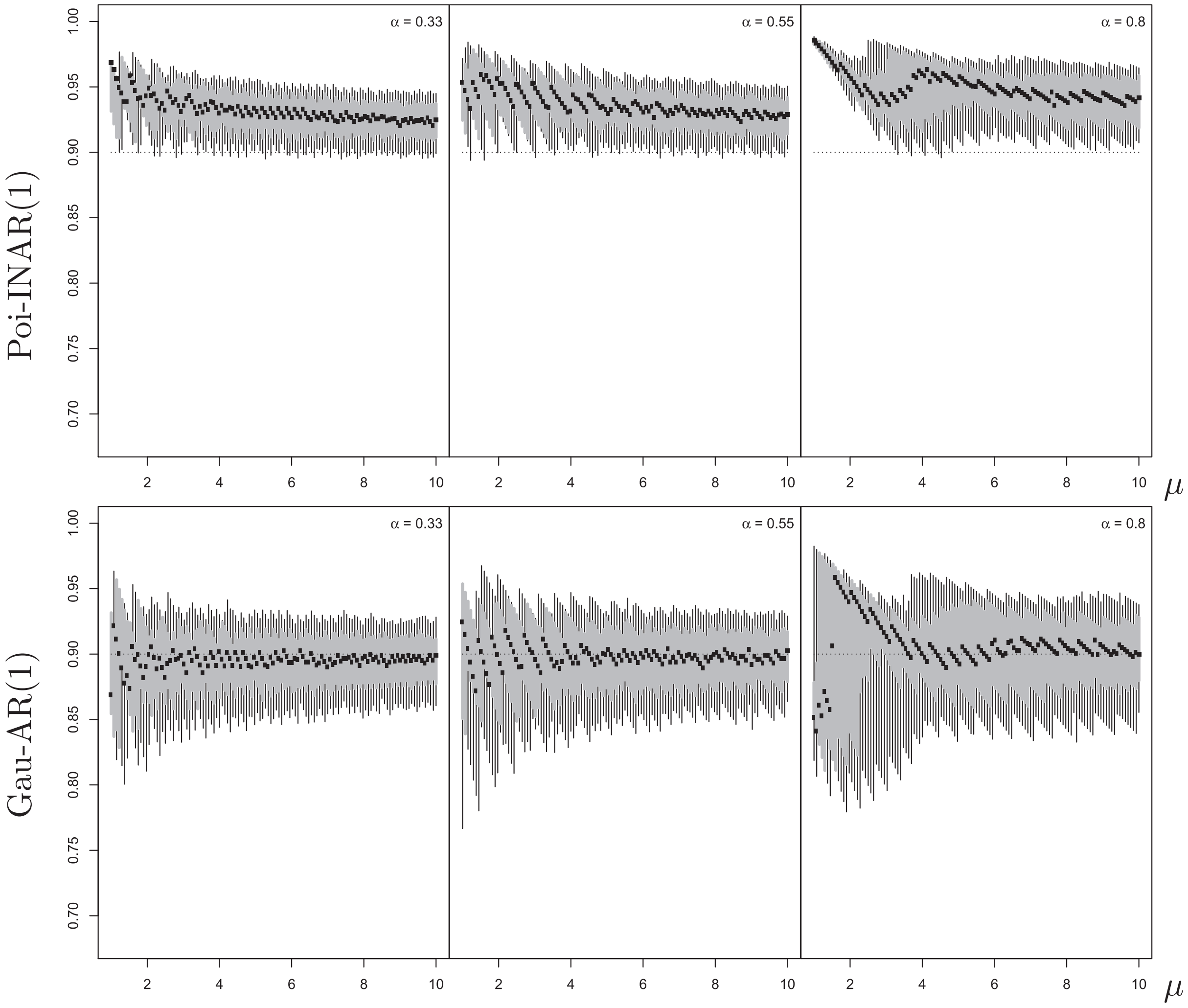

FIGURE 1 True coverage of coherent (top row) and approximate (bottom row) upper-sided 90% PIs against μ, DGP Poi-INAR(1) with different α, sample size $T = 250$, and forecast horizon $h = 1$. Boxplots for 1,000 simulated coverages

one- and two-sided cases. Thus strong exceedances of p_{cov} mainly occur for low counts. For the approximate PIs, an analogous distinction holds with respect to the variation among the coverages, but the fraction of intervals violating the coverage requirement does not seem to improve with increasing μ.

At this point, let us have a look at the overall performance metrics introduced in Section 2.3. For a Poi-INAR (1) DGP, these are computed by fixing (α, T) and by applying the respective sample statistic to the coverages resulting from all values of μ and all simulation runs. For example, the top left plot in Figure 1, where $(\alpha, T) = (0.33, 250)$, relies on 121 different values of μ, and on 1,000 simulations for each μ, that is, on altogether 121,000 simulated coverages. If one now computes the fraction of those 121,000 coverages being $<p_{\text{cov}}$, one obtains the number ≈ 0.0906, as shown in part (a) of Table 1. To make it simple, one number in Tables 1 and 2 always corresponds to one plot.

Part (a) in Tables 1 and 2 shows that the approximate PIs lead to a large shortfall rate, and this rate does not improve with increasing sample size. This differs from the case of coherent PIs, where this fraction clearly tends towards 0 for increasing T. Also, the average shortfall values in part (b) are much larger in absolute value for the approximate than for the coherent PIs. Thus the coherent PIs lead to increasingly less severe shortfalls than the approximate ones. On the other hand, the upper-sided coherent PIs in particular show a stronger average exceedance of the target coverage p_{cov} than the approximate PIs do (see also Figure 1), so they tend to be more conservative. In the two-sided case (see part (c) in Table 2), the average exceedances are quite similar for both types of PI. Finally, the standard deviations in part (d) of Tables 1 and 2 are always much larger for the approximate than for the coherent PIs, so there is more uncertainty regarding the actual coverage level.

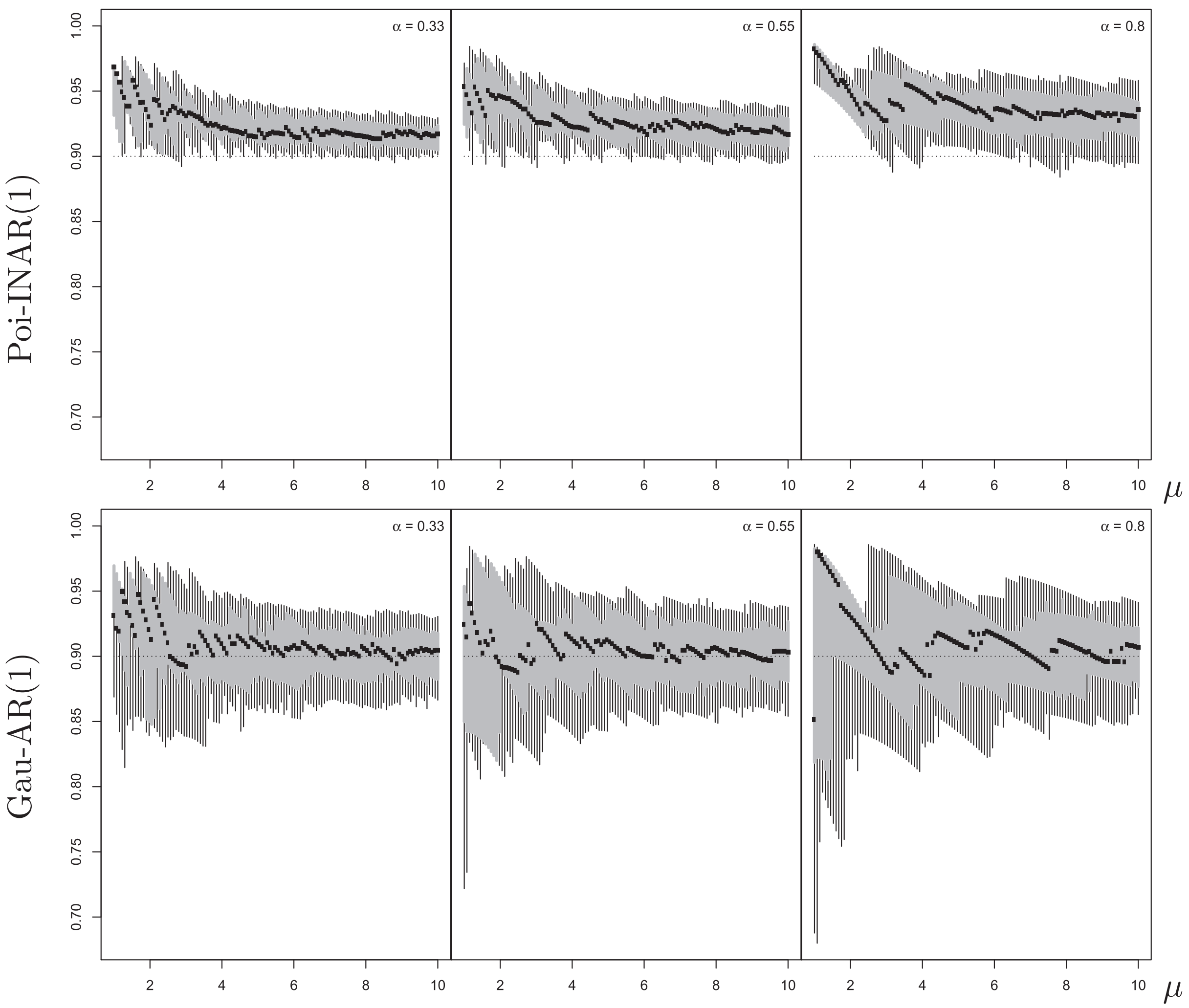

FIGURE 2 True coverage of coherent (top row) and approximate (bottom row) two-sided 90% PIs against μ, DGP Poi-INAR(1) with different α, sample size $T = 250$, and forecast horizon $h = 1$. Boxplots for 1,000 simulated coverages

TABLE 1 Performance metrics based on true coverages of coherent (columns "Coh") or approximate (columns "Gau") upper-sided 90% PIs, DGP Poi-INAR(1) with different (α, T) and forecast horizon $h = 1$, computed from all simulation runs for all levels of μ

	$\alpha = 0.33$		$\alpha = 0.55$		$\alpha = 0.8$		$\alpha = 0.33$		$\alpha = 0.55$		$\alpha = 0.8$	
T	**Coh**	**Gau**	**Coh**	**Gau**	**Coh**	**Gau**	**Coh**	**Gau**	**Coh**	**Gau**	**Coh**	**Gau**
	(a) Shortfall rate						**(b) Average shortfall**					
75	0.1610	0.5428	0.1315	0.5119	0.0702	0.4483	−0.0233	−0.0438	−0.0235	−0.0251	−0.0248	−0.0611
250	0.0906	0.5496	0.0780	0.5140	0.0511	0.4518	−0.0120	−0.0312	−0.0125	−0.0143	−0.0007	−0.0453
2,500	0.0280	0.5530	0.0259	0.5186	0.0210	0.4569	−0.0036	−0.0257	−0.0040	−0.0150	−0.0050	−0.0389
	(c) Average exceedance						**(d) Standard deviation**					
75	0.0376	0.0264	0.0422	0.0297	0.0564	0.0384	0.0305	0.0468	0.0312	0.0529	0.0313	0.0644
250	0.0340	0.0205	0.0379	0.0241	0.0498	0.0329	0.0218	0.0335	0.0232	0.0387	0.0257	0.0486
2,500	0.0316	0.0176	0.0352	0.0214	0.0455	0.0308	0.0177	0.0273	0.0193	0.0318	0.0226	0.0424

Let us draw up a first interim balance. If accounting for the discrete nature of count data while constructing a PI, recall the algorithm in Section 2.2, the PI is determined to get a coverage being ≥ p_{cov}. For a Poi-INAR(1) DGP, it turned out that the estimation uncertainty causes violations of this coverage requirement only very rarely. If, in contrast, treating the count data as being normally distributed, the corresponding procedure for computing PIs leads to frequent and strong shortfalls of p_{cov}. Furthermore, the variation among the realized coverages is much larger than in the coherent case, so the reliability of the approximate PIs is rather low. Hence the approximate PIs have a considerably worse performance than the coherent ones in the case of a Poi-INAR (1) DGP.

3.2 | INAR(1) DGPs with overdispersion

In this section, we still consider INAR(1) DGPs, but now having either NB- or ZIP-distributed innovations (instead of Poisson ones as in Section 3.1; see also Appendix A). Thus the DGPs now exhibit overdispersion instead of equidispersion. The parameters of both DGPs were chosen such that the observations have the same dispersion ratio $I=\sigma^2/\mu$. Despite this unique extent of overdispersion, there is a fundamental difference between the NB- and ZIP-INAR(1) DGPs: While the NB-distribution causes a "regular" type of overdispersion (PMF flattened compared to the Poi-distribution), the ZIP's overdispersion is caused by an isolated additional point mass in zero. The illustrative results presented here show a dispersion ratio of $I=2.4$ (similar to the data example presented in Section 4.1); that is, the variance is more than twice the mean (strong dispersion).

If computing coherent PIs based on a fitted NB- or ZIP-INAR(1) model, respectively, the results in Tables 3–6 do not show a notable difference between NB versus ZIP. But compared to the Poisson case in Tables 1 and 2, we see a clear increase in the shortfall rates; see part (a). Also the average shortfall is often increased, especially for $\alpha \leq 0.55$; see part (b). Thus the additional overdispersion as well as the additional parameter to be estimated lead to increased estimation uncertainty, which deteriorates the performance of the coherent PIs. Increasing the sample size T, in turn, clearly improves the performance.

The performance of the Gauss approximations is not only affected by the extent, but also by the type of overdispersion. For the NB-DGP with its "regular overdispersion," the corresponding Gauss approximation does clearly better than the one for the Poi-DGP. Especially in the two-sided case (Table 4 vs. Table 2), the shortfall rates and standard deviations are lower if applying the Gauss approximationto the NB-DGP rather than to the Poi-DGP. Interestingly, the two-sided NB-approximate PIs are often superior to the coherent ones in terms of the shortfall rate (but worse regarding the standard deviation), whereas the upper-sidedNB-approximate PIs perform clearly worse than their coherent counterparts (Table 3). In contrast, the average exceedance of the two-sided NB-approximate PIs is much larger than in the coherent case. The overfulfillment of the confidence requirement by the two-sided NB-approximate PIs is caused by the fact that the Gaussian approximation's forecast distribution tends to exhibit more dispersion than the NB-INAR(1)'s one for strong overdispersion (such as $I=2.4$); see Supporting Information Supplement S.2 for more details. As a consequence, the two-sided approximate PIs are chosen too large, causing less shortfall but more intense exceedance, as observed in Table 4 (as well as in Figure 4 below).

Regarding the ZIP-INAR(1) DGP, the Gaussian approximation performs particularly badly for the upper-sided PIs (Table 5), and the shortfall rates notably increase with increasing T. The two-sided approximate PIs, although not having such extremely large shortfall rates, perform worse than the coherent PIs with respect

TABLE 2 Performance metrics based on true coverages of coherent (columns "Coh") or approximate (columns "Gau") two-sided 90% PIs, DGP Poi-INAR(1) with different (α, T) and forecast horizon $h=1$, computed from all simulation runs for all levels of μ

	$\alpha=0.33$		$\alpha=0.55$		$\alpha=0.8$		$\alpha=0.33$		$\alpha=0.55$		$\alpha=0.8$	
T	**Coh**	**Gau**	**Coh**	**Gau**	**Coh**	**Gau**	**Coh**	**Gau**	**Coh**	**Gau**	**Coh**	**Gau**
	(a) Shortfall rate						**(b) Average shortfall**					
75	0.1646	0.4724	0.1582	0.4860	0.0997	0.4597	−0.0180	−0.0409	−0.0201	−0.0458	−0.0276	−0.0533
250	0.0833	0.4200	0.1019	0.4603	0.0909	0.4453	−0.0088	−0.0299	−0.0109	−0.0366	−0.0173	−0.0440
2,500	0.0231	0.3875	0.0353	0.4555	0.0397	0.4441	−0.0025	−0.0260	−0.0032	−0.0335	−0.0060	−0.0409
	(c) Average exceedance						**(d) Standard deviation**					
75	0.0279	0.0322	0.0347	0.0352	0.0515	0.0396	0.0249	0.0465	0.0280	0.0521	0.0329	0.0601
250	0.0262	0.0284	0.0306	0.0319	0.0420	0.0367	0.0188	0.0367	0.0212	0.0438	0.0263	0.0515
2,500	0.0252	0.0265	0.0277	0.0303	0.0350	0.0355	0.0164	0.0332	0.0176	0.0407	0.0205	0.0487

TABLE 3 Performance metrics based on true coverages of coherent (columns "Coh") or approximate (columns "Gau") upper-sided 90% PIs, DGP NB-INAR(1) with different (α, T), dispersion ratio $I = 2.4$, and forecast horizon $h = 1$, computed from all simulation runs for all levels of μ

	$\alpha = \mathbf{0.33}$		$\alpha = \mathbf{0.55}$		$\alpha = \mathbf{0.8}$		$\alpha = \mathbf{0.33}$		$\alpha = \mathbf{0.55}$		$\alpha = \mathbf{0.8}$	
T	**Coh**	**Gau**	**Coh**	**Gau**	**Coh**	**Gau**	**Coh**	**Gau**	**Coh**	**Gau**	**Coh**	**Gau**
	(a) Shortfall rate						**(b) Average shortfall**					
75	0.3270	0.5589	0.3206	0.5069	0.2441	0.3420	−0.0256	−0.0358	−0.0257	−0.0366	−0.0289	−0.0395
250	0.2046	0.5628	0.1993	0.4837	0.1645	0.2738	−0.0125	−0.0212	−0.0123	−0.0208	−0.0148	−0.0208
2,500	0.0625	0.5888	0.0640	0.4698	0.0565	0.1990	−0.0038	−0.0130	−0.0034	−0.0121	−0.0045	−0.0100
	(c) Average exceedance						**(d) Standard deviation**					
75	0.0265	0.0217	0.0271	0.0242	0.0316	0.0307	0.0315	0.0390	0.0330	0.0437	0.0391	0.0539
250	0.0210	0.0145	0.0216	0.0174	0.0257	0.0249	0.0183	0.0235	0.0184	0.0253	0.0214	0.0294
2,500	0.0173	0.0096	0.0183	0.0131	0.0217	0.0214	0.0105	0.0141	0.0110	0.0160	0.0129	0.0184

TABLE 4 Performance metrics based on true coverages of coherent (columns "Coh") or approximate (columns "Gau") two-sided 90% PIs, DGP NB-INAR(1) with different (α, T), dispersion ratio $I = 2.4$, and forecast horizon $h = 1$, computed from all simulation runs for all levels of μ

	$\alpha = \mathbf{0.33}$		$\alpha = \mathbf{0.55}$		$\alpha = \mathbf{0.8}$		$\alpha = \mathbf{0.33}$		$\alpha = \mathbf{0.55}$		$\alpha = \mathbf{0.8}$	
T	**Coh**	**Gau**	**Coh**	**Gau**	**Coh**	**Gau**	**Coh**	**Gau**	**Coh**	**Gau**	**Coh**	**Gau**
	(a) Shortfall rate						**(b) Average shortfall**					
75	0.3806	0.2609	0.3587	0.2582	0.2533	0.2060	−0.0253	−0.0294	−0.0246	−0.0324	−0.0277	−0.0370
250	0.2312	0.1306	0.2204	0.1446	0.1874	0.1277	−0.0117	−0.0168	−0.0116	−0.0204	−0.0150	−0.0231
2,500	0.0657	0.0524	0.0692	0.0809	0.0730	0.0742	−0.0035	−0.0135	−0.0033	−0.0171	−0.0047	−0.0185
	(c) Average exceedance						**(d) Standard deviation**					
75	0.0242	0.0334	0.0253	0.0352	0.0305	0.0395	0.0308	0.0353	0.0314	0.0391	0.0368	0.0463
250	0.0193	0.0296	0.0199	0.0313	0.0241	0.0355	0.0175	0.0219	0.0177	0.0247	0.0211	0.0281
2,500	0.0159	0.0283	0.0166	0.0300	0.0190	0.0341	0.0102	0.0149	0.0105	0.0183	0.0123	0.0205

TABLE 5 Performance metrics based on true coverages of coherent (columns "Coh") or approximate (columns "Gau") upper-sided 90% PIs, DGP ZIP-INAR(1) with different (α, T), dispersion ratio $I = 2.4$, and forecast horizon $h = 1$, computed from all simulation runs for all levels of μ

	$\alpha = \mathbf{0.33}$		$\alpha = \mathbf{0.55}$		$\alpha = \mathbf{0.8}$		$\alpha = \mathbf{0.33}$		$\alpha = \mathbf{0.55}$		$\alpha = \mathbf{0.8}$	
T	**Coh**	**Gau**	**Coh**	**Gau**	**Coh**	**Gau**	**Coh**	**Gau**	**Coh**	**Gau**	**Coh**	**Gau**
	(a) Shortfall rate						**(b) Average shortfall**					
75	0.2584	0.5859	0.2779	0.6791	0.3066	0.6461	−0.0285	−0.0445	−0.0292	−0.0475	−0.0311	−0.0455
250	0.1517	0.6086	0.1619	0.7490	0.1936	0.7250	−0.0142	−0.0313	−0.0143	−0.0337	−0.0175	−0.0337
2,500	0.0475	0.6267	0.0505	0.8036	0.0622	0.7861	−0.0041	−0.0249	−0.0041	−0.0272	−0.0055	−0.0288
	(c) Average exceedance						**(d) Standard deviation**					
75	0.0335	0.0262	0.0333	0.0236	0.0313	0.0222	0.0348	0.0465	0.0356	0.0475	0.0372	0.0496
250	0.0281	0.0190	0.0278	0.0154	0.0264	0.0176	0.0213	0.0321	0.0216	0.0308	0.0233	0.0319
2,500	0.0249	0.0144	0.0246	0.0100	0.0229	0.0171	0.0140	0.0248	0.0139	0.0222	0.0136	0.0253

TABLE 6 Performance metrics based on true coverages of coherent (columns "Coh") or approximate (columns "Gau") two-sided 90% PIs, DGP ZIP-INAR(1) with different (α, T), dispersion ratio $I = 2.4$, and forecast horizon $h = 1$, computed from all simulation runs for all levels of μ

	$\alpha = 0.33$		$\alpha = 0.55$		$\alpha = 0.8$		$\alpha = 0.33$		$\alpha = 0.55$		$\alpha = 0.8$	
T	**Coh**	**Gau**	**Coh**	**Gau**	**Coh**	**Gau**	**Coh**	**Gau**	**Coh**	**Gau**	**Coh**	**Gau**
	(a) Shortfall rate						**(b) Average shortfall**					
75	0.3110	0.3405	0.2921	0.3401	0.2791	0.3851	−0.0280	−0.0374	−0.0255	−0.0340	−0.0296	−0.0337
250	0.1695	0.2510	0.1642	0.2685	0.1817	0.3665	−0.0125	−0.0223	−0.0119	−0.0227	−0.0163	−0.0228
2,500	0.0477	0.2023	0.0462	0.2226	0.0638	0.3473	−0.0033	−0.0151	−0.0033	−0.0185	−0.0049	−0.0189
	(c) Average exceedance						**(d) Standard deviation**					
75	0.0298	0.0386	0.0290	0.0355	0.0293	0.0307	0.0345	0.0450	0.0321	0.0418	0.0356	0.0422
250	0.0251	0.0350	0.0243	0.0313	0.0238	0.0243	0.0200	0.0322	0.0191	0.0309	0.0214	0.0287
2,500	0.0227	0.0335	0.0215	0.0292	0.0202	0.0209	0.0135	0.0270	0.0128	0.0265	0.0121	0.0239

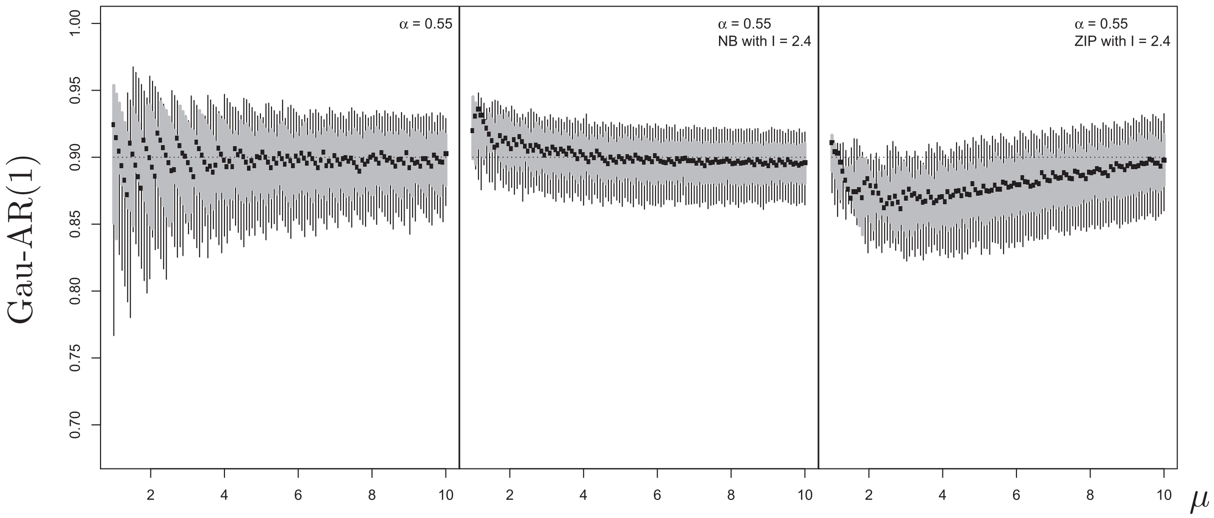

FIGURE 3 True coverage of approximate upper-sided 90% PIs against μ, for DGPs Poi-INAR(1), NB-INAR(1), and ZIP-INAR(1) (from left to right) with $\alpha = 0.55$, sample size $T = 250$, and forecast horizon $h = 1$. Boxplots for 1,000 simulated coverages, dispersion ratio $I = \sigma^2/\mu$ equals either 1 (Poi) or 2.4 (NB, ZIP)

to all types of performance metric in Table 6. This difference in performance between NB and ZIP is also illustrated by Figures 3 and 4. Furthermore, it can be seen that the actual mean μ has a strong effect on the performance of the approximate PIs. In the ZIP-case, for example, the shortfall is particularly large for $\mu \leq 8$ (upper-sided) or $\mu \leq 4$ (two-sided), respectively.

Compared to our conclusions in Section 3.1, we note a strong effect of overdispersion on the PIs' performance (see also Supporting Information Supplements S.2 and S.3 for a further discussion). The coherent PIs show more frequent shortfalls because of the additional uncertainty. However, the effect on the approximate PIs is even stronger and very different, depending on the actual type of overdispersion. The PIs' performance for the ZIP-DGP with its isolated point mass in zero deteriorates severely with respect to all performance criteria. For the two-sided PIs in the NB-case (if the extent of overdispersion is sufficiently large), we may also observe an overfulfillment of the coverage requirement; that is, the approximate PIs are chosen unnecessarily large, thus leading to strong exceedances of p_{cov}.

3.3 | Poi-INARCH(1) DGP

The INAR(1) model considered in Sections 3.1 and 3.2 is probably the most well-known model for count time series. Its basic idea (see also Appendix A.1 for further details) is to directly adapt the basic AR(1) recursion to

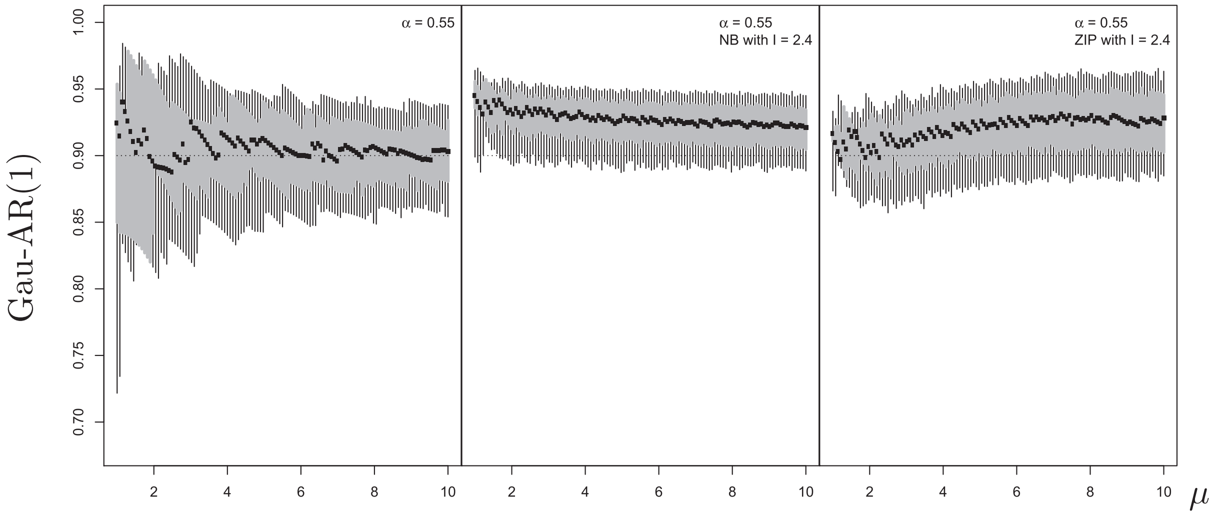

FIGURE 4 True coverage of approximate two-sided 90% PIs against μ, for DGPs Poi-INAR(1), NB-INAR(1), and ZIP-INAR(1) (from left to right) with $\alpha = 0.55$, sample size $T = 250$, and forecast horizon $h = 1$. Boxplots for 1,000 simulated coverages, dispersion ratio $I = \sigma^2/\mu$ equals either 1 (Poi) or 2.4 (NB, ZIP)

TABLE 7 Performance metrics based on true coverages of coherent (columns "Coh") or approximate (columns "Gau") two-sided 90% PIs, DGP Poi-INARCH(1) with different (α, T) and forecast horizon $h = 1$, computed from all simulation runs for all levels of μ

	$\alpha = 0.33$		$\alpha = 0.55$		$\alpha = 0.8$		$\alpha = 0.33$		$\alpha = 0.55$		$\alpha = 0.8$	
T	**Coh**	**Gau**	**Coh**	**Gau**	**Coh**	**Gau**	**Coh**	**Gau**	**Coh**	**Gau**	**Coh**	**Gau**
	(a) Shortfall rate						**(b) Average shortfall**					
75	0.1411	0.4802	0.1445	0.4668	0.1540	0.4340	−0.0188	−0.0415	−0.0213	−0.0564	−0.0271	−0.0918
250	0.0652	0.4300	0.0669	0.4220	0.0792	0.4127	−0.0089	−0.0314	−0.0098	−0.0480	−0.0117	−0.0883
2,500	0.0180	0.3981	0.0177	0.4026	0.0265	0.4016	−0.0027	−0.0280	−0.0026	−0.0449	−0.0030	−0.0834
	(c) Average exceedance						**(d) Standard deviation**					
75	0.0264	0.0357	0.0273	0.0459	0.0318	0.0642	0.0240	0.0493	0.0264	0.0649	0.0334	0.0954
250	0.0260	0.0326	0.0266	0.0435	0.0294	0.0638	0.0177	0.0407	0.0193	0.0568	0.0233	0.0918
2,500	0.0250	0.0305	0.0253	0.0420	0.0282	0.0631	0.0158	0.0369	0.0166	0.0539	0.0196	0.0878

the count-data case by replacing the involved multiplication by an integer-valued substitute, the binomial thinning operation. This, however, is (by far) not the only thinning-based approach in the literature; see Weiß (2018) for a brief survey. Alternative thinning operations (and corresponding time series models) can be constructed by using nonbinomial distributions instead. A popular example is given by the so-called "negative-binomial thinning operation" (and the resulting "NGINAR(1) model") proposed by Ristić, Bakouch, and Nastić (2009), which is based on a conditional NB-distribution. Although using a different thinning operation, the required computations for obtaining PIs are similar to the INAR(1) case, we just have to modify the formula for the transition probabilities in Appendix A.1. The Poi-INARCH(1) model considered in this section, which is another common choice in practice for AR(1)-like count DGPs, might also be understood as being a thinning-based model, by using a "Poisson thinning" operation. But it is more appropriate to classify it as a regression model, as done in Appendix A.2.

In our analyses, it turned out that the upper- and two-sided PIs perform nearly the same for this type of DGP (in analogy to the Poi-INAR(1) case presented in Section 3.1). Therefore, we restrict the following discussion to the two-sided case. In Table 7, we observe clearly more variation in the coverage values (see part (d)) as well as more frequent and more severe shortfalls (see parts (a,b)) for the approximate PIs than for the coherent ones. In contrast to the Poi-INAR(1) DGP in Section 3.1 (see Table 2), the coherent PIs of the Poi-INARCH (1) DGP show a more stable performance if increasing α.

However, we can neither observe a uniquely better nor a worse performance if comparing these two models.

This differs from the case of approximate PIs, which perform clearly worse for the Poi-INARCH(1) DGP: Although the shortfall rate tends to be lower than for the Poi-INAR(1) DGP, the comparison of Tables 2 and 7 shows more extreme average shortfall and exceedance values as well as a notable increase in standard deviation among the coverage values; see also Figure 5. This discrepancy becomes more severe for increasing autocorrelation. Therefore, the use of approximate PIs must be discouraged even more for the Poi-INARCH(1) DGP than for the Poi-INAR(1) DGP, whereas the coherent PIs perform quite similarly for both types of DGP.

3.4 | Higher-order Poi-INAR and Poi-INARCH DGPs

To examine the effect of a higher AR-order on the PIs' performance, we consider the Poi-INAR(2) and Poi-INARCH(2) DGP in this section; see Appendix A for details. They are extensions of the Poi-INAR(1) DGP discussed in Section 3.1 and the Poi-INARCH(1) DGP of Section 3.3, respectively. These second-order models are approximated by their corresponding continuous counterpart, the Gau-AR(2) model. Both types of PI are computed based on the last two observations x_T, x_{T-1}, and these are chosen as the respective last observations of each simulated time series. For comparability with the first-order DGPs, we fixed the lag-1 ACF at the same levels of α as in the previous section; that is, $\rho(1)=\alpha$. Then, we considered different choices of α_2 and computed $\alpha_1=\alpha(1-\alpha_2)$. The subsequent discussion is limited to the case of two-sided PIs, as the upper-sided PIs perform quite similarly.

While the increased AR order has no clear effect on the performance of the coherent PIs, the performance of the approximate PIs is adversely affected by an increase in the additional AR parameter α_2, causing more extreme average shortfall and standard deviation values. For this reason, we display the results for $\alpha_2=0.45$, the largest α_2-value in our simulations. Looking at Tables 8 and 9, the approximate PIs of the second-order DGPs exhibit a more extreme deviation from p_{cov} than the coherent PIs do with respect to all performance metrics: The shortfall rates are increased, shortfall and exceedance are more extreme on average, and the standard deviations are increased.

If we compare the coherent PIs' metrics of the Poi-INAR(2) DGP in Table 8 with those of the Poi-INAR (1) DGP in Table 2, more estimation uncertainty due to the additional parameter may be noticed in places, but disappears quickly with a growing sample size T. The increase in model order does not seem to impair the performance of the coherent PIs. For high autocorrelation, the second-order model even produces less and less extreme shortfall than the Poi-INAR(1) model. The approximate PIs of the Poi-INAR(2) DGP, in contrast, show a severe deterioration in their performance with growing autocorrelation parameter α, or with growing α_2. Both shortfall and exceedance become more extreme on average, and also the standard deviation among the coverages increases. For example, the average shortfall for $\alpha=0.8$ and $T=250$ in Table 8 is (in absolute value) almost twice as large as in the Poi-INAR(1) case in Table 2, and average exceedance as well as standard deviation are increased by more than 50%.

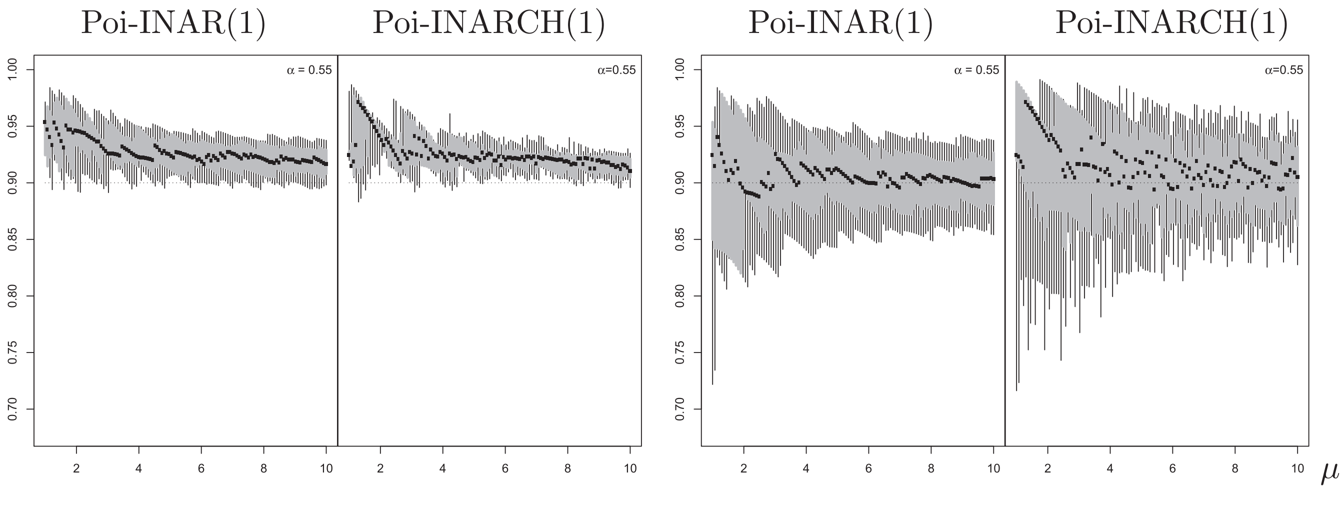

FIGURE 5 True coverage of coherent (left two plots) and approximate (right two plots) two-sided 90% PIs against μ, Poi-INARCH (1) DGP versus Poi-INAR(1) DGP with $\alpha=0.55$, sample size $T=250$, and forecast horizon $h=1$. Boxplots for 1,000 simulated coverages

TABLE 8 Performance metrics based on true coverages of coherent (columns "Coh") or approximate (columns "Gau") two-sided 90% PIs, DGP Poi-INAR(2) with different (α, T), $\alpha_2 = 0.45$, and forecast horizon $h = 1$, computed from all simulation runs for all levels of μ

	$\alpha = 0.33$		$\alpha = 0.55$		$\alpha = 0.8$		$\alpha = 0.33$		$\alpha = 0.55$		$\alpha = 0.8$	
T	**Coh**	**Gau**	**Coh**	**Gau**	**Coh**	**Gau**	**Coh**	**Gau**	**Coh**	**Gau**	**Coh**	**Gau**
	(a) Shortfall rate						**(b) Average shortfall**					
75	0.1495	0.4780	0.1431	0.4850	0.1429	0.4632	−0.0220	−0.0512	−0.0237	−0.0609	−0.0339	−0.0856
250	0.0882	0.4355	0.0788	0.4469	0.0750	0.4307	−0.0110	−0.0414	−0.0109	−0.0524	−0.0120	−0.0809
2,500	0.0294	0.4175	0.0258	0.4336	0.0287	0.4176	−0.0031	−0.0379	−0.0032	−0.0500	−0.0035	−0.0798
	(c) Average exceedance						**(d) Standard deviation**					
75	0.0334	0.0381	0.0357	0.0452	0.0424	0.0588	0.0288	0.0574	0.0307	0.0678	0.0391	0.0912
250	0.0296	0.0349	0.0314	0.0433	0.0364	0.0589	0.0211	0.0484	0.0221	0.0601	0.0262	0.0856
2,500	0.0268	0.0336	0.0288	0.0422	0.0322	0.0588	0.0178	0.0452	0.0190	0.0578	0.0227	0.0838

TABLE 9 Performance metrics based on true coverages of coherent (columns "Coh") or approximate (columns "Gau") two-sided 90% PIs, DGP Poi-INARCH(2) with different (α, T), $\alpha_2 = 0.45$, and forecast horizon $h = 1$, computed from all simulation runs for all levels of μ

	$\alpha = 0.33$		$\alpha = 0.55$		$\alpha = 0.8$		$\alpha = 0.33$		$\alpha = 0.55$		$\alpha = 0.8$	
T	**Coh**	**Gau**	**Coh**	**Gau**	**Coh**	**Gau**	**Coh**	**Gau**	**Coh**	**Gau**	**Coh**	**Gau**
	(a) Shortfall rate						**(b) Average shortfall**					
75	0.1830	0.4725	0.1964	0.4681	0.2281	0.4460	−0.0256	−0.0564	−0.0293	−0.0701	−0.0357	−0.1023
250	0.0915	0.4303	0.0954	0.4250	0.1110	0.4011	−0.0123	−0.0480	−0.0128	−0.0628	−0.0156	−0.0999
2,500	0.0254	0.4114	0.0264	0.4072	0.0300	0.3804	−0.0033	−0.0453	−0.0035	−0.0609	−0.0049	−0.0997
	(c) Average exceedance						**(d) Standard deviation**					
75	0.0279	0.0451	0.0294	0.0528	0.0356	0.0671	0.0299	0.0644	0.0340	0.0772	0.0438	0.1046
250	0.0265	0.0427	0.0276	0.0511	0.0328	0.0674	0.0203	0.0569	0.0216	0.0703	0.0281	0.0998
2,500	0.0255	0.0418	0.0265	0.0503	0.0305	0.0671	0.0166	0.0543	0.0176	0.0683	0.0229	0.0985

If comparing the metrics of the Poi-INARCH(2) DGP in Table 9 to those of Poi-INARCH(1) in Table 7, we notice a deterioration in most metrics for the approximate as well as the coherent PIs. In the coherent case, the additional estimation uncertainty causes, for example, a notable increase in both shortfall rate and average shortfall. With an increasing sample size T, though, this effect is damped and the metrics of the second-order model approach those in Table 7. However, this is not the case for the approximate PIs. These exhibit a greater extent of shortfall and exceedance, and more standard deviation among the coverage values.

To sum up, an increased AR-order for the count DGPs has a small but inconsistent effect on the performance of the coherent PIs. While the performance might even improve for Poi-INAR DGPs (especially if highly autocorrelated), we note a certain deterioration for Poi-INARCH DGPs. This is plausible in view of the different data-generating mechanisms behind these families; see also remark 4.1.7 in Weiß (2018): The INAR family tends to produce runs of certain count values, which is obviously advantageous for forecasting purposes, whereas INARCH DGPs tend to vivid fluctuations. For the approximate PIs, in contrast, the effect of an increased AR-order is quite homogeneous across the different model families: The performance becomes considerably worse with respect to both shortfalls and exceedances. Thus we can only reaffirm our advice of Sections 3.1 and 3.3 to not use approximate PIs for Poi-INAR and INARCH DGPs.

3.5 | DGPs for bounded counts

Our performance analyses also cover the case where the generated count time series have the bounded range $\{0, \ldots, n\}$ with a given $n \in \mathbb{N}$. As the bounded-counts counterpart to the Poi-INAR(1) and INARCH(1) model, respectively, we consider the BinAR(1) and BinARCH(1) model as described in Appendix A. Note that the bounded counts' mean μ now also depends on the actual value for the upper bound n: If $\pi = \beta/(1-\alpha) \in (0,1)$ denotes the "success" probability, then $\mu = n\,\pi$. In our simulations, we have chosen the success probability either as $\pi = 0.15$ ("rare event") or as $\pi = 0.45$ (nearly symmetric marginal distribution). But in contrast to the previous analyses, it is now not possible to choose arbitrary mean levels $\mu > 0$. Instead, we increased the integer-valued upper bound n by increments of 1, leading to $\mu = n\,\pi$. As a result, the summarized performance metrics are not directly comparable with the previous ones, and they are also not between $\pi = 0.15$ and $\pi = 0.45$.

It turns out that both the coherent and the approximate PIs for the BinAR(1) DGP show a very similar performance as in the Poi-INAR(1) case. In particular, the approximate PIs again perform considerably worse than the coherent ones. The rare-event case $\pi = 0.15$ is even more similar to the Poisson case than the near-to-symmetry case $\pi = 0.45$, which is not surprising in view of the Poisson limit theorem. If going from $\pi = 0.15$ to $\pi = 0.45$, the overall shortfall rates usually increase slightly. But the most striking pattern is a more pronounced effect of the discreteness (both for the coherent and the approximate PIs). This is illustrated by Figure 6, which refers to the two-sided intervals for $\alpha = 0.55$ and sample size $T = 250$.

Also for the BinARCH(1) model, the coherent PIs perform way better than their approximations. Comparing Figure 7 to Figure 6, we notice that the coherent PIs of the BinARCH(1) DGP tend to show less and less extreme

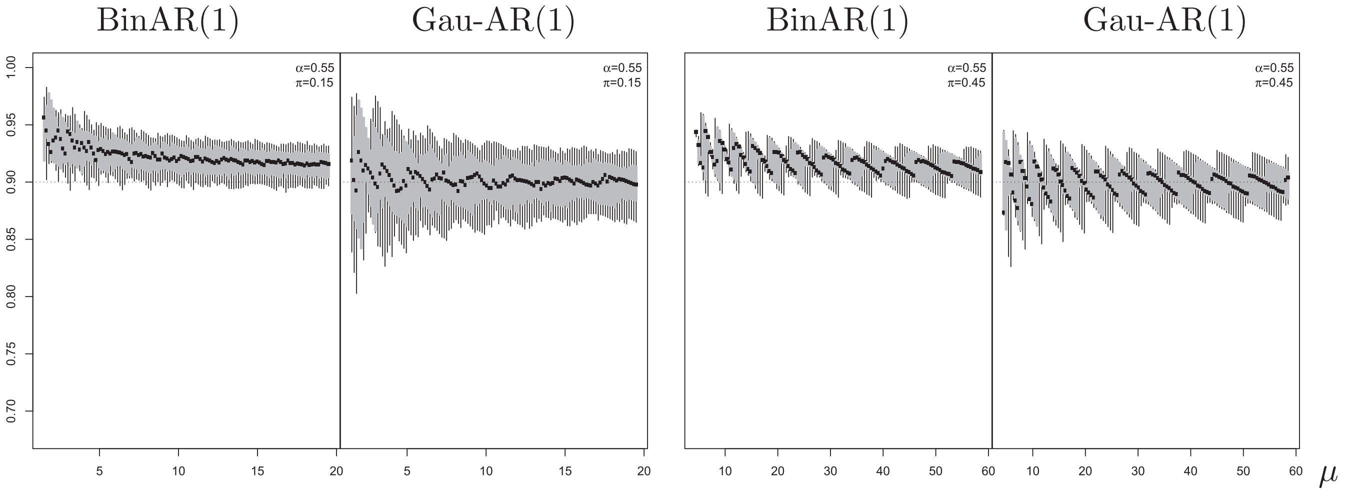

FIGURE 6 True coverage of coherent ("BinAR(1)") and approximate ("Gau-AR(1)") two-sided 90% PIs against μ, DGP BinAR(1) with $\pi = 0.15$ (left block) or $\pi = 0.45$ (right block), $\alpha = 0.55$, sample size $T = 250$, and forecast horizon $h = 1$. Boxplots for 1,000 simulated coverages

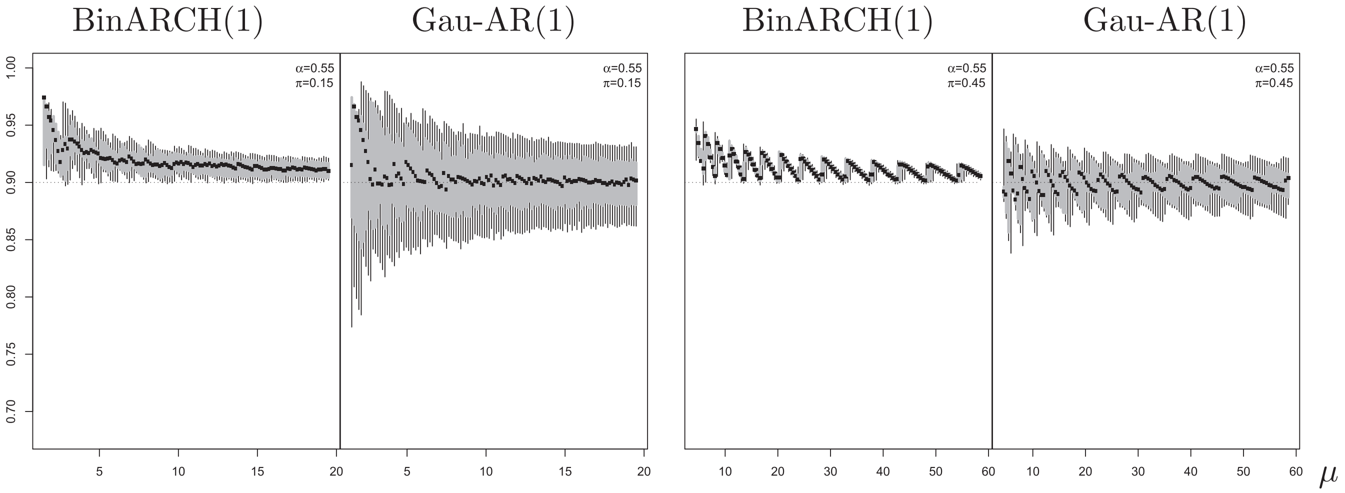

FIGURE 7 True coverage of coherent ("BinARCH(1)") and approximate ("Gau-AR(1)") two-sided 90% PIs against μ, DGP BinARCH (1) with $\pi = 0.15$ (left block) or $\pi = 0.45$ (right block), $\alpha = 0.55$, sample size $T = 250$, and forecast horizon $h = 1$. Boxplots for 1,000 simulated coverages

shortfall. This is confirmed if comparing the metrics in Tables 10 and 11, where it can also be seen that the average exceedance as well as standard deviation values are lower for the BinARCH(1) DGP. In case of a nearly symmetric marginal distribution ($\pi = 0.45$), no major difference between the performance of the approximate PIs can be noticed. For $\pi = 0.15$, however, the performance of the BinARCH(1)'s approximate PIs becomes considerably worse, analogously to the Poi-INARCH(1) case in Section 3.3. This deterioration manifests itself in terms of increased values of the average shortfall, the exceedance, and the standard deviation, especially for large α.

To recapitulate the bounded-counts case, the conclusions of Sections 3.1 and 3.3 for the two types of Poi-DGP apply in an analogous manner also to the respective Bin-DGPs, especially if being concerned with rare events. For nearly symmetrically distributed counts, however, we have a stronger discreteness pattern; that is, small changes in DGP parametrization might cause large changes in performance. In any case, the approximate PIs show a considerably worse shortfall behavior than the coherent PIs, and for the BinARCH(1) DGP, also the exceedance is worse. So we strongly advise against using approximate PIs for bounded-count DGPs.

3.6 | Nonstationary DGPs

The models considered in Sections 3.1–3.5 are suitable for stationary count time series. In practice, however, one is sometimes concerned with the forecasting of nonstationary count time series, which exhibit seasonality, trend, or other forms of nonstationarity. Again, there are many options of how to model nonstationary count time series. Motivated by the data example in Section 4.2, we examine the case of the ll-Poi-AR(1) DGP as our final simulation experiment, which can be equipped with seasonality or trend (see Appendix A.2 for further information). But models exist also for more sophisticated forms of nonstationarity, such as the random-environment

TABLE 10 Performance metrics based on true coverages of coherent (columns "Coh") or approximate (columns "Gau") two-sided 90% PIs, DGP BinAR(1) with $\pi = 0.15$, different (α, T), and forecast horizon $h = 1$, computed from all simulation runs for all levels of μ

	$\alpha = 0.33$		$\alpha = 0.55$		$\alpha = 0.8$		$\alpha = 0.33$		$\alpha = 0.55$		$\alpha = 0.8$	
T	**Coh**	**Gau**	**Coh**	**Gau**	**Coh**	**Gau**	**Coh**	**Gau**	**Coh**	**Gau**	**Coh**	**Gau**
	(a) Shortfall rate						**(b) Average shortfall**					
75	0.1834	0.5220	0.1734	0.5126	0.1190	0.4679	−0.0153	−0.0356	−0.0201	−0.0380	−0.0283	−0.0429
250	0.0979	0.4818	0.1158	0.4770	0.1142	0.4501	−0.0073	−0.0236	−0.0106	−0.0264	−0.0174	−0.0318
2,500	0.0295	0.4527	0.0396	0.4562	0.0516	0.4432	−0.0020	−0.0178	−0.0032	−0.0213	−0.0056	−0.0281
	(c) Average exceedance						**(d) Standard deviation**					
75	0.0225	0.0256	0.0300	0.0274	0.0482	0.0311	0.0211	0.0397	0.0260	0.0425	0.0335	0.0482
250	0.0210	0.0207	0.0255	0.0228	0.0371	0.0268	0.0158	0.0287	0.0188	0.0318	0.0255	0.0375
2,500	0.0199	0.0183	0.0222	0.0207	0.0289	0.0249	0.0137	0.0237	0.0148	0.0274	0.0178	0.0344

TABLE 11 Performance metrics based on true coverages of coherent (columns "Coh") or approximate (columns "Gau") two-sided 90% PIs, DGP BinARCH(1) with $\pi = 0.15$, different (α, T), and forecast horizon $h = 1$, computed from all simulation runs for all levels of μ

	$\alpha = 0.33$		$\alpha = 0.55$		$\alpha = 0.8$		$\alpha = 0.33$		$\alpha = 0.55$		$\alpha = 0.8$	
T	**Coh**	**Gau**	**Coh**	**Gau**	**Coh**	**Gau**	**Coh**	**Gau**	**Coh**	**Gau**	**Coh**	**Gau**
	(a) Shortfall rate						**(b) Average shortfall**					
75	0.1446	0.5231	0.1533	0.5104	0.1716	0.4802	−0.0139	−0.0361	−0.0147	−0.0419	−0.0194	−0.0584
250	0.0654	0.4825	0.0667	0.4766	0.0707	0.4679	−0.0062	−0.0241	−0.0062	−0.0321	−0.0073	−0.0515
2,500	0.0175	0.4540	0.0172	0.4552	0.0173	0.4666	−0.0014	−0.0190	−0.0016	−0.0278	−0.0019	−0.0496
	(c) Average exceedance						**(d) Standard deviation**					
75	0.0192	0.0273	0.0199	0.0335	0.0219	0.0479	0.0186	0.0410	0.0202	0.0486	0.0255	0.0673
250	0.0192	0.0228	0.0198	0.0300	0.0212	0.0462	0.0145	0.0304	0.0154	0.0400	0.0177	0.0604
2,500	0.0192	0.0203	0.0197	0.0286	0.0207	0.0455	0.0133	0.0259	0.0142	0.0367	0.0157	0.0586

INAR models dating back to Nastić, Laketa, and Ristić (2016). The ll-Poi-AR(1) model is approximated by the Gaussian regression model with ARMA innovations (Brockwell & Davis, 2016), which reduces to the ordinary ARMA model in the absence of seasonality and trend, and which has a linear trend and harmonic oscillation like the considered ll-Poi-AR(1) model. Since, now, the mean varies over time we no longer evaluate our simulation results by plots against the mean, but present tabulated values for selected scenarios. Thus Table 12 shows the averaged performance metrics of 1,000 simulated two-sided PIs for a representative parameter setting ($\gamma_0 = 1,\ T = 250,\ \alpha \in \{0.33, 0.55, 0.8\}$). The columns of the table are now labeled by the trend parameter γ_1, and the rows by the seasonality parameters (γ_2, γ_3).

Table 12 shows that the increase in trend (γ_1) or seasonality (γ_2, γ_3) generally causes the coherent PIs to fall short more often (see part (a)) and more severely (see part (b)). Seasonality also causes more variation of the coverage values (see part (d)), while increasing trend does not have such an effect. But, as can be seen from the supplementary simulation results, all performance metrics improve with increasing T, and also with increasing intercept γ_0. In particular, the average exceedance reduces with increasing γ_0, which is in line with the effect of a growing mean μ in the previous sections.

TABLE 12 Performance metrics based on true coverages of coherent (columns "Coh") or approximate (columns "Gau") two-sided 90% PIs, DGP ll-Poi-AR(1) with $\gamma_0 = 1$, $T = 250$, different $(\gamma_1, \gamma_2, \gamma_3)$, different α, and forecast horizon $h = 1$, computed from 1,000 simulation runs per scenario

	$\gamma_1 = 0$		$\gamma_1 = 0.001$		$\gamma_1 = 0.002$		$\gamma_1 = 0$		$\gamma_1 = 0.001$		$\gamma_1 = 0.002$	
(γ_2, γ_3)	**Coh**	**Gau**	**Coh**	**Gau**	**Coh**	**Gau**	**Coh**	**Gau**	**Coh**	**Gau**	**Coh**	**Gau**
$\alpha = 0.33$												
	(a) Shortfall rate						**(b) Average shortfall**					
(0, 0)	0.0910	0.4410	0.1200	0.6430	0.1530	0.8600	−0.0080	−0.0351	−0.0123	−0.0427	−0.0108	−0.0555
(0.1, 0.2)	0.1080	0.4220	0.1420	0.5740	0.1680	0.7240	−0.0170	−0.0545	−0.0185	−0.0623	−0.0161	−0.0781
(0.2, 0.6)	0.1210	0.4280	0.1300	0.5000	0.1750	0.5740	−0.0135	−0.0816	−0.0196	−0.1028	−0.0150	−0.1070
	(c) Average exceedance						**(d) Standard deviation**					
(0, 0)	0.0357	0.0465	0.0266	0.0261	0.0291	0.0259	0.0206	0.0461	0.0198	0.0462	0.0189	0.0462
(0.1, 0.2)	0.0342	0.0475	0.0295	0.0385	0.0269	0.0307	0.0248	0.0596	0.0240	0.0620	0.0212	0.0661
(0.2, 0.6)	0.0358	0.0574	0.0319	0.0549	0.0269	0.0534	0.0249	0.0785	0.0248	0.0911	0.0230	0.0965
$\alpha = 0.55$												
	(a) Shortfall rate						**(b) Average shortfall**					
(0, 0)	0.0360	0.3460	0.0730	0.6120	0.1330	0.7590	−0.0198	−0.0556	−0.0097	−0.0589	−0.0134	−0.0687
(0.1, 0.2)	0.0970	0.4450	0.1370	0.5370	0.1230	0.6650	−0.0161	−0.0642	−0.0173	−0.0779	−0.0140	−0.0807
(0.2, 0.6)	0.0910	0.4020	0.1190	0.4800	0.1770	0.5530	−0.0141	−0.0889	−0.0185	−0.1030	−0.0163	−0.1187
	(c) Average exceedance						**(d) Standard deviation**					
(0, 0)	0.0340	0.0441	0.0285	0.0435	0.0285	0.0383	0.0238	0.0580	0.0166	0.0639	0.0195	0.0663
(0.1, 0.2)	0.0333	0.0526	0.0312	0.0435	0.0274	0.0406	0.0245	.0688	0.0243	0.0752	0.0201	0.0752
(0.2, 0.6)	0.0346	0.0588	0.0313	0.0560	0.0281	0.0541	0.0233	0.0840	0.0242	0.0934	0.0233	0.1049
$\alpha = 0.80$												
	(a) Shortfall rate						**(b) Average shortfall**					
(0, 0)	0.0050	0.3610	0.1440	0.5800	0.0640	0.5830	−0.0349	−0.0793	−0.0129	−0.0764	−0.0121	−0.1064
(0.1, 0.2)	0.0930	0.3980	0.0960	0.5290	0.1080	0.5930	−0.0187	−0.0850	−0.0181	−0.0936	−0.0154	−0.0995
(0.2, 0.6)	0.1070	0.3710	0.0990	0.4890	0.1210	0.5370	−0.0168	−0.1084	−0.0183	−0.1195	−0.0184	−0.1275
	(c) Average exceedance						**(d) Standard deviation**					
(0, 0)	0.0253	0.0509	0.0348	0.0581	0.0289	0.0449	0.0180	0.0759	0.0253	0.0815	0.0182	0.0919
(0.1, 0.2)	0.0368	0.0596	0.0302	0.0555	0.0271	0.0501	0.0247	0.0834	0.0227	0.0922	0.0217	0.0953
(0.2, 0.6)	0.0369	0.0646	0.0320	0.0599	0.0290	0.0585	0.0264	0.0990	0.0244	0.1090	0.0240	0.1170

While all these effects of trend or seasonality are quite moderate regarding the coherent PIs, they lead to severe deterioration of the approximate PIs' performance. Increasing trend strongly affects the shortfall performance according to parts (a,b) — for example, with shortfall rates up to 86%. Increasing seasonality leads to heavily increased average shortfalls and standard deviations (see parts (b,d)). Therefore, the use of approximate models for computing integer-valued PIs has to be discouraged also if being concerned with nonstationary DGPs.

4 | EMPIRICAL INVESTIGATIONS

In this section, we analyze the performance of PIs for two real-data examples of count time series: the stationary demand counts data in Section 4.1 and the nonstationary liquidation counts data in Section 4.2.

4.1 | Demand for RBC O+ transfusion blood bags

Our first data application refers to the daily demand for RBC O+ transfusion blood bags in a regional hospital in southeastern Wisconsin between June 2009 and January 2010. The data set was originally published by Alwan, Xu, Yao, and Yue (2016) and further analyzed by Alwan and Weiß (2017). The full time series is of length 240 and does not exhibit any indications of nonstationarity. It has an AR(1)-like ACF and exhibits a strong degree of overdispersion. Therefore, Alwan and Weiß (2017) used the NB-INAR(1) model for these data. Since we want to use the data for illustrating PI computations, we split the full data into a training sample, which consists of the first $T = 150$ observations, and into a test sample consisting of the remaining observations $x_{T+1}, \ldots, x_{T+90}$. The latter are used for computing out-of-sample forecasts. But let us start with the training data $x_1, \ldots, x_T$. They have the sample mean 3.047, variance 7.575 (overdispersion by factor ≈ 2.49), and an AR(1)-like ACF with lag-1 value 0.288. As in Alwan and Weiß (2017), we use the NB-INAR(1) model for the data (recall Section 3.2), and consider a Gau-AR(1) model for approximation. Based on both models, we compute the one-step-ahead PIs ($p_{\text{cov}} = 0.90$), and we compare them, among others, to the actual outcomes x_t. The PIs constitute in-sample forecasts for $t = 2, \ldots, T$, and out-of-sample forecasts for $t = T + 1, \ldots, T + 90$. For illustration, Figure 8 shows a plot of the obtained upper-sided PIs, which can be interpreted as expressing some kind of worst-case prediction of the demand for blood bags. For readability, the upper limits of the PIs were shifted upwards by 0.2 units (coherent PI) and 0.4 units (approximate PI), respectively. It can be seen that the coherent PIs' upper limits are never smaller than the approximate PIs' ones.

Table 13 summarizes some performance results for the in-sample PIs. Let us first look at the realized in-sample coverage rates — that is, at the fraction of observations $x_2, \ldots, x_T$ falling within their actual PI. For the coherent PIs, these rates are 0.899 (upper-sided) and 0.906 (two-sided), so the coverage requirement $p_{\text{cov}} = 0.90$ is almost perfectly met in both cases. For the approximate PIs, however, we observe a divergent performance. The approximate upper-sided PIs lead to the rate 0.866, which falls considerably short of p_{cov}. The approximate two-sided PIs, in contrast, have the rate 0.919, which exceeds both p_{cov} and the corresponding rate of the coherent PIs (0.906). Analogous conclusions can be drawn from the average failures shown in Table 13, and also from the out-of-sample results, although some differences are less pronounced there. Thus the upper-sided approximate PIs tend to be too short (see also Figure 8), whereas the two-sided ones tend to be too large. This contradictory behavior appears plausible in view of our findings in

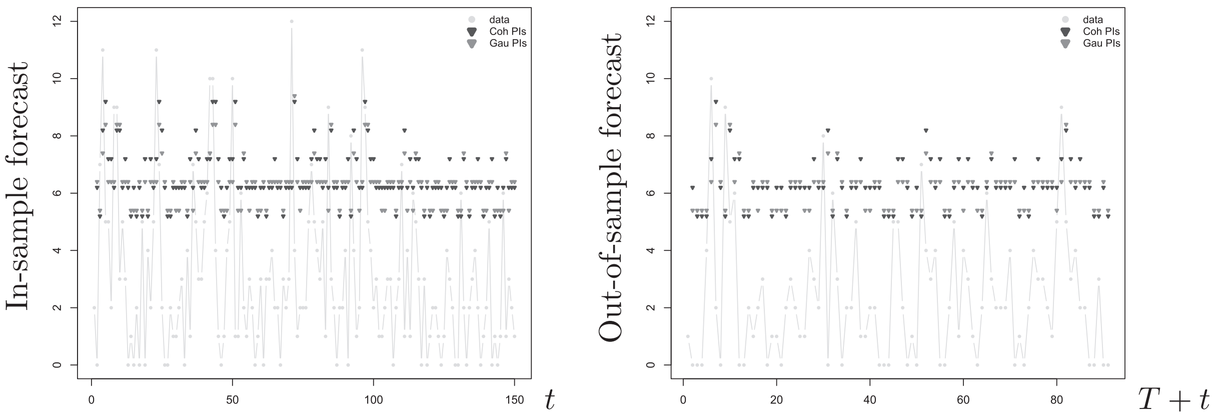

FIGURE 8 Plot of upper-sided in-sample and out-of-sample PIs for demand counts data

TABLE 13 Performance of in-sample and out-of-sample PIs for demand counts data (Section 4.1)

In-sample forecasting	**Coh**	**Gau**
Upper-sided PIs		
Coverage rate — i.e., mean of $\mathbb{1}(x_t \in [0, x_u])$	0.899	0.866
Average failure — i.e., mean of $\lvert x_t - x_u \rvert \mathbb{1}(x_t > x_u)$	0.268	0.342
Two-sided PIs		
Coverage rate — i.e., mean of $\mathbb{1}(x_t \in [x_l, x_u])$	0.906	0.919
Average failure — i.e., mean of $\lvert x_t - x_u \rvert \mathbb{1}(x_t > x_u) + \lvert x_t - x_l \rvert \mathbb{1}(x_t < x_l)$	0.262	0.208
Out-of-sample forecasting	**Coh**	**Gau**
Upper-sided PIs		
Coverage rate — i.e., mean of $\mathbb{1}(x_t \in [0, x_u])$	0.933	0.933
Average failure — i.e., mean of $\lvert x_t - x_u \rvert \mathbb{1}(x_t > x_u)$	0.156	0.178
Two-sided PIs		
Coverage rate — i.e., mean of $\mathbb{1}(x_t \in [x_l, x_u])$	0.922	0.933
Average failure — i.e., mean of $\lvert x_t - x_u \rvert \mathbb{1}(x_t > x_u) + \lvert x_t - x_l \rvert \mathbb{1}(x_t < x_l)$	0.167	0.122

Section 3.2, where we observed an analogous discrepancy for NB-INAR(1) DGPs with dispersion ratio $I = 2.4$ (in our data, we have $\hat{I} \approx 2.49$).

Certainly, we do not know the true model behind the demand count time series, but let us take the fitted coherent model as the benchmark. The approximate upper-sided PIs differ from the coherent ones in 39% of all cases (in-sample and out-of-sample), and according to the fitted NB-INAR(1) model we would always be concerned with a shortfall case (see also Figure 8). The approximate two-sided PIs differ from the coherent ones in 59% of all cases, but now these deviations would mainly be classified as exceedances with respect to the fitted NB-INAR(1) model (41% agreement, 58% exceedance, 1% shortfall).

4.2 | Company liquidations in Germany

The *Federal Gazette* (*Bundesanzeiger*; https://www.bundesanzeiger.de) is the central platform of the German Federal Ministry of Justice, where business disclosures such as liquidation announcements are published (among other things). We focus on the number of company liquidations among general commercial partnerships and limited partnerships ("GmbH & Co. KG") per working day in Germany. The data used for model fitting are from the period June 1, 2007, to September 30, 2019. The time series has a length of $T = 3157$ working days. The time series plot in Figure 9 shows clear signs of seasonality and trend, so it appears reasonable to use log-linear regression models (Appendix A.2) for these data. The analysis of the periodogram in Figure 9 implies the inclusion of harmonic terms for annual, semiannual and quarterly seasonality (periods 256, 128, and 64, respectively). Further data analyses showed the need also to include a linear and quadratic trend in t as well as an AR(1) component. Owing to the strong overdispersion, a conditional NB distribution has to be used — that is, altogether an ll-NB-AR(1) model following $X_t | X_{t-1}, \ldots \sim \mathrm{NB}\left(1, \frac{n}{M_t + n}\right)$ (see Appendix A.2). As in Section 3.6, we approximate this model by a Gaussian regression model with AR(1) innovations, including the same trend and harmonic components as the coherent model. It has to be noted, however, that the fitted ll-NB-AR(1) model is not a perfect choice for the data. A residual analysis indicates that the inclusion of further harmonic and autoregressive

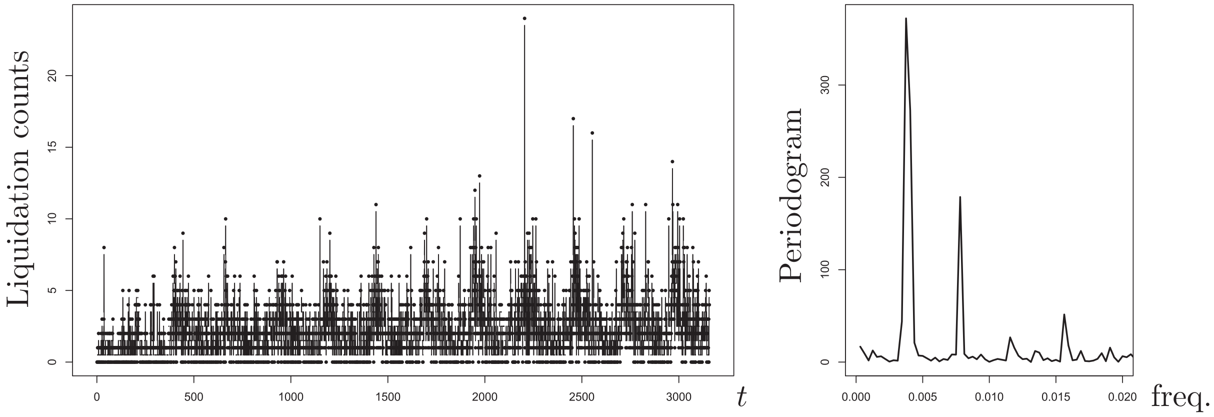

FIGURE 9 Time series plot of liquidation counts data, and detail of periodogram

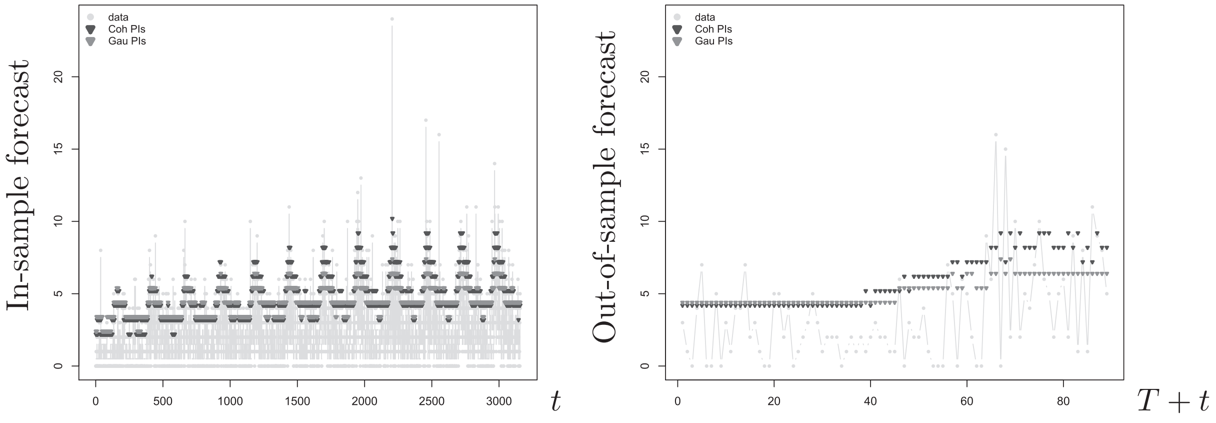

FIGURE 10 Plot of upper-sided in-sample and out-of-sample PIs for liquidation counts data

terms is advisable for improving the model. But for illustrative purposes, we want to keep the model manageable.

Besides an in-sample forecasting of the data $x_2, \dots, x_T$, we also do an out-of-sample forecasting again, by taking further liquidations data $X_{T+1}, \dots, X_{T+89}$ for the period October 1, 2019, to January 31, 2020, from the *Federal Gazette*. The determined coherent and approximate PIs are then compared to the respective realizations x_t. It turns out that the coherent two-sided PIs reduce to upper-sided PIs without exception, whereas we have about 18% (44%) of truly two-sided approximate PIs for the in-sample (out-of-sample) period. In Figure 10, we show the upper-sided PIs for both periods for illustration, where the plotted upper limits are again shifted upwards for readability.

TABLE 14 Performance of in-sample and out-of-sample PIs for liquidation counts data (Section 4.2)

In-sample forecasting	**Coh**	**Gau**
Upper-sided PIs		
Coverage rate — i.e., mean of $\mathbb{1}(x_t \in [0,x_u])$	0.934	0.906
Average failure — i.e., mean of $\lvert x_t - x_u \rvert \mathbb{1}(x_t > x_u)$	0.128	0.194
Two-sided PIs		
Coverage rate — i.e., mean of $\mathbb{1}(x_t \in [x_l,x_u])$	0.934	0.931
Average failure — i.e., mean of $\lvert x_t - x_u \rvert \mathbb{1}(x_t > x_u) + \lvert x_t - x_l \rvert \mathbb{1}(x_t < x_l)$	0.128	0.138
Out-of-sample forecasting	**Coh**	**Gau**
Upper-sided PIs		
Coverage rate — i.e., mean of $\mathbb{1}(x_t \in [0,x_\mathrm{u}])$	0.820	0.775
Average failure — i.e., mean of $\lvert x_t - x_u \rvert \mathbb{1}(x_t > x_u)$	0.438	0.685
Two-sided PIs		
Coverage rate — i.e., mean of $\mathbb{1}(x_t \in [x_l,x_u])$	0.820	0.708
Average failure — i.e., mean of $\lvert x_t - x_u \rvert \mathbb{1}(x_t > x_u) + \lvert x_t - x_l \rvert \mathbb{1}(x_t < x_l)$	0.438	0.674

Detailed performance analyses are summarized in Table 14. The in-sample coverage rate of the coherent PIs exceeds p_{cov}, whereas the out-of-sample one falls short of p_{cov}. The latter is commonly observed in empirical studies (see section 6 in Chatfield, 1993) and often caused by modeling issues. Therefore, a refinement of the coherent model seems advisable for future forecasting applications; see also the above discussion. But here our aim is to analyze the differences between coherent and approximate forecasting, so we continue with the current model fits. The coverage rates of the approximate PIs are always lower than those of the coherent ones, which is particularly problematic for the out-of-sample forecasts. This is supported by the right-hand part of Figure 10, where the upper limits of the approximate PIs appear to be too low between $T+40$ and $T+89$. Also, the average failure rates are larger for the approximate PIs throughout. If taking the fitted ll-NB-AR(1) model as the benchmark, then a large fraction of the computed approximate PIs is classified as suffering from shortfall, namely about 38% (54%) of the upper-sided PIs for the in-sample (out-of-sample) period, and 22% (45%) of the two-sided PIs. So, again, we note that the approximate Gaussian approach leads to considerably different inference from the coherent approach.

5 | CONCLUSIONS

We analyzed the performance of coherent PIs for various types of count processes. For the Poisson and binomial DGPs, the coherent PIs rarely fall short of the actual coverage requirement, and exceedances thereof are mitigated with increasing mean μ. However, increased dispersion of the DGP, resulting either from overdispersed distributions such as the negative binomial one, or from higher order INARCH processes, deteriorate the PIs' performance. The same occurs in the presence of seasonality and trend. But the performance of PIs is always considerably worse if these are computed based on a Gaussian approximation of the actual DGP. To some part, this is caused by the different way of constructing PIs for continuous data rather than for discrete data, which generally results in a strong shortfall tendency of the approximate PIs. But besides this systematic difference, we also observed that extraordinary features such as overdispersion, zero inflation, and trend have more drastic effects on the approximate PIs than on the coherent ones.

At this point, let us also have a look at the findings of Homburg et al. (2019) regarding *point* forecasts for count processes. For the central point forecasts (conditional median), Homburg et al. observed that the coherent point forecasts are almost unaffected by estimation error. Approximate point forecasts perform considerably worse but at least improve with increasing mean or decreased autocorrelation. This differs from our findings regarding PIs for count processes. In fact, there are more analogies to the performance of the noncentral point forecasts in Homburg et al., which is plausible as the PIs also rely on outer quantiles. Although Homburg et al. used a rather different (risk-related) performance criterion, they also note a visible effect of the estimation uncertainty even on the coherent forecasts, and in particular a much worse performance of the approximate forecasts throughout. So we must clearly confirm their overall conclusion that "the practice of discretizing Gaussian ARIMA forecasts for count time series is strongly discouraged."

For future research, one should try to develop approaches for incorporating the apparent estimation uncertainty into the forecasting procedure. Also the approach of Remark 1, where the full forecast distribution is computed as the predictor, deserves further attention. A comprehensive performance study as well as a comparison to approximate methods appear to be important for evaluating the practicality of this approach.

ACKNOWLEDGMENTS

The authors thank the two referees for their useful comments on an earlier draft of this article. This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), Projektnummer 394832307. Open access funding enabled and organized by Projekt DEAL.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available at https://www.hsu-hh.de/mathstat/en/research/projects/forecastingrisk.

ORCID

Annika Homburg https://orcid.org/0000-0001-7529-0389
Christian H. Weiß https://orcid.org/0000-0001-8739-6631
Layth C. Alwan https://orcid.org/0000-0002-5653-3098
Gabriel Frahm https://orcid.org/0000-0001-7507-730X

REFERENCES

Alwan, L. C., & Weiß, C. H. (2017). INAR implementation of newsvendor model for serially dependent demand counts. *International Journal of Production Research*, *55*(4), 1085–1099.

Alwan, L. C., Xu, M., Yao, D. Q., & Yue, X. (2016). The dynamic newsvendor model with correlated demand. *Decision Sciences*, *47*(1), 11–30.

Bejleri, V., & Nandram, B. (2018). Bayesian and frequentist prediction limits for the Poisson distribution. *Communications in Statistics: Theory and Methods*, *47*(17), 4254–4271.

Brockwell, P. J., & Davis, R. A. (2016). *Introduction to time series and forecasting* (3rd ed.). Cham, Switzerland: Springer.

Chatfield, C. (1993). Calculating interval forecasts. *Journal of Quality Technology*, *5*(4), 178–188.

Chintalapudi, N., Battineni, G., & Amenta, F. (2020). COVID-19 Virus outbreak forecasting of registered and recovered cases after sixty day lockdown in Italy: A data driven model approach. *Journal of Microbiology, Immunology and Infection*, *53*(3), 396–403.

de Gooijer, J. G., & Hyndman, R. J. (2006). 25 years of time series forecasting. *International Journal of Forecasting*, *22*(3), 443–473.

Freeland, R. K., & McCabe, B. P. M. (2004). Forecasting discrete valued low count time series. *International Journal of Forecasting*, *20*, 427–434.

Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, *102*(477), 359–378.

Hahn, G. J., & Nelson, W. (1973). A survey of prediction intervals and their applications. *Journal of Business and Economic Statistics*, *11*(2), 121–135.

Homburg, A., Weiß, C. H., Alwan, L. C., Frahm, G., & Göb, R. (2019). Evaluating approximate point forecasting of count processes. *Econometrics*, *7*(3), 30.

Kolassa, S. (2016). Evaluating predictive count data distributions in retail sales forecasting. *International Journal of Forecasting*, *32*(3), 788–803.

Krishnamoorthy, K., & Peng, J. (2011). Improved closed-form prediction intervals for binomial and Poisson distributions. *Journal of Statistical Planning and Inference*, *141*(5), 1709–1718.

Lambert, P. (1997). Predictions in overdispersed series of counts using an approximate predictive likelihood. *Journal of Forecasting*, *16*(3), 195–207.

McCabe, B. P. M., & Martin, G. M. (2005). Bayesian predictions of low count time series. *International Journal of Forecasting*, *21* (2), 315–330.

McKenzie, E. (1985). Some simple models for discrete variate time series. *Water Resources Bulletin*, *21*(4), 645–650.

Mukhopadhyay, S., & Sathish, V. (2018). Predictive likelihood for coherent forecasting of count time series. *Journal of Forecasting*, *38*(3), 222–235.

Nastić, A. S., Laketa, P. N., & Ristić, M. M. (2016). Random environment integer-valued autoregressive process. *Journal of Time Series Analysis*, *37*(2), 267–287.

Oracle (2017). Oracle® Crystal Ball: Predictor user's guide. Release 11.1.2.4.850. https://docs.oracle.com/cd/E57185_01/CBPUG/CBPUG.pdf

Rahardja, D. (2020). Statistical methodological review for time-series data. *Journal of Statistics and Management Systems*, *21* (1), 189–199.

Ristić, M. M., Bakouch, H. S., & Nastić, A. S. (2009). A new geometric first-order integer-valued autoregressive (NGINAR(1)) process. *Journal of Statistical Planning and Inference*, *139*(7), 2218–2226.

Silva, N., Pereira, I., & Silva, M. E. (2009). Forecasting in INAR (1) model. *REVSTAT*, *24*, 171–176.

Snyder, R. D., Ord, J. K., & Beaumont, A. (2012). Forecasting the intermittent demand for slow-moving inventories: A modelling approach. *International Journal of Forecasting*, *28*(2), 485–496.

Wang, H. (2008). Coverage probability of prediction intervals for discrete random variables. *Computational Statistics and Data Analysis*, *53*(1), 17–26.

Weiß, C. H. (2018). *An introduction to discrete-valued time series*. Chichester, UK: Wiley.

AUTHOR BIOGRAPHIES

Annika Homburg is a research assistant and PhD student at the Department Mathematics and Statistics at the Helmut Schmidt University in Hamburg, Germany. She researches in the field of discrete time series with a focus on coherent forecasting and evaluating continuous approximations to count data models.

Christian H. Weiß is a Professor in the Department of Mathematics and Statistics at the Helmut Schmidt University in Hamburg, Germany. He got his doctoral degree in mathematical statistics from the University of Würzburg, Germany. His research areas include time series analysis, statistical quality control, and computational statistics. He is an author of several textbooks and published his work in international scientific journals such as Bernoulli, Journal of the American Statistical Association, Journal of Multivariate Analysis, Journal of Quality Technology, Journal of the Royal Statistical Society, and Journal of Time Series Analysis.

Layth C. Alwan is Professor of Supply Chain/Operations Management and Business Statistics, Sheldon B. Lubar School of Business, University of Wisconsin-Milwaukee. He received his PhD in operations management and business statistics from the University of Chicago Booth School of Business. His research interests include includes supply chain/operations management, forecasting, business analytics, and statistical process control. He is an author of several textbooks in the areas of business statistics and quality control. His work has been published in leading journals including Decision Sciences, European Journal of Operational Research, International Journal of Production Research, IIE Transactions, Journal of Business and Economic Statistics, Journal of Royal Statistical Society, and Production and Operations Management.

Gabriel Frahm is Professor of Applied Stochastics and Risk Management at the Helmut Schmidt University in Hamburg. He completed his diploma studies in business administration and worked as a research assistant at the University of Cologne. He got his doctoral degree in statistics and also wrote his habilitation thesis in statistics and econometrics in Cologne. His research interests cover capital-market theory and portfolio optimization, financial mathematics and econometrics, game and decision theory, copulas and extreme-value theory, robust covariance matrices, random-matrix theory, and missing-data analysis. His contributions have been published in international scientific journals like, e.g., the Journal of Econometrics, the Journal of Multivariate Analysis, and the International Journal of Theoretical and Applied Finance, etc.

Rainer Göb is a Professor of Statistics at the Institute of Matematics of the University of Würzburg, Germany. His research interest is in the industrial application of statistics, in particular statistical sampling, process monitoring, predictive analytics, risk analysis, and the standardisation of statistical methods. He is the chair of the subcommittee "Acceptance Sampling" at the Technical Committee 69 "Application of Statistical Methods" of the International Organization for Standardization.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Homburg A, Weiß CH, Alwan LC, Frahm G, Göb R. A performance analysis of prediction intervals for count time series. *Journal of Forecasting*. 2021;40:603–625. https://doi.org/10.1002/for.2729

APPENDIX A: SUMMARY OF CONSIDERED COUNT DGPS

In what follows, we summarize those count time series models which were used as a DGP in our numerical and simulation studies. These models belong to either 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 (2018).

A.1 | Thinning-based models

The considered thinning-based models have an AR-like data-generating mechanism, 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 $\alpha \circ X | X \sim \text{Bin}(X, \alpha)$. The following models assume that all thinnings are executed independently of each other, and independently of the innovations as well as of past observations.

INAR(1) model: Model recursion $X_t = \alpha \circ X_{t-1} + \epsilon_t$, where innovations (ϵ_t) are independent and identically distributed (i.i.d.) count r.v.s with $\mu_\epsilon = \text{E}$ and variance $\sigma_\epsilon^2 := \text{var}$.

Mean $\mu = \text{E}$, variance $\sigma^2 = \text{var}$, and autocorrelation function (ACF) $\rho(k) = \text{corr}$, respectively, are given by

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

Constitutes a Markov chain with transition probabilities

$$\begin{aligned} p(x|x_T) &= p(X_{T+1} = x | X_T = x_T) \\ &= \sum\nolimits_{s=0}^{\min\{x, x_T\}} \binom{x_T}{s} \alpha^s (1-\alpha)^{x_T - s} \cdot P(\epsilon_t = x - s). \end{aligned}$$

Referred to as Poi-, NB-, or ZIP-INAR(1) model, respectively, if ϵ_t follows Poisson, negative binomial, or zero-inflated Poisson distribution.

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

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

ACF satisfies $\rho(1) = \alpha_1/(1-\alpha_2)$, and $\rho(k) = \alpha_1 \rho(k-1) + \alpha_2 \rho(k-2)$ for $k \geq 2$.

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

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

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

Constitutes a Markov chain with marginal distribution $\text{Bin}(n, \pi)$, and with ACF $\rho(k) = \alpha^k$. The transition probabilities are

$$\begin{aligned} p(x|x_T) \ &= \sum\nolimits_{m=\max(0, x + x_T - n)}^{\min(x, x_T)} \binom{x_T}{m} \binom{n - x_T}{x - m} \\ & \gamma^m (1-\gamma)^{x_T - m} \beta^{x-m} (1-\beta)^{n - x_T + m - x}. \end{aligned}$$

A.2 | Regression models

We consider the AR-type INARCH models (integer-valued autoregressive conditional heteroskedasticity) as well as the log-linear Poisson AR(1) model (ll-Poi-AR(1) model).

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

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

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 $X_t|X_{t-1},\ldots \sim \text{Poi}(\beta+\alpha_1 X_{t-1}+\alpha_2 X_{t-2})$ with $\alpha_1+\alpha_2<1$. ACF like for INAR(2) model, transition probabilities

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

BinARCH(1) model: Model recursion $X_t|X_{t-1},\ldots \sim \text{Bin}\left(n,\beta+\alpha\frac{X_{t-1}}{n}\right)$ with $\beta,\beta+\alpha\in(0,1)$. Transition probabilities

$$p(x|x_T)=\binom{n}{x}\left(\beta+\alpha\frac{x_T}{n}\right)^x\left(1-\beta-\alpha\frac{x_T}{n}\right)^{n-x}.$$

ll-Poi-AR(1) model with linear trend and harmonic oscillation (period p, angular frequency $\omega=2\pi/\text{p}$). Model recursion $X_t|X_{t-1},\ldots \sim \text{Poi}(M_t)$ with

$$\ln M_t \quad = \overbrace{\gamma_0+\gamma_1 t+\gamma_2\cos(\omega t)+\gamma_3\sin(\omega t)}^{=:\ln\mu_t} + \alpha_1(\ln(X_{t-1}+1)-\ln(\mu_{t-1}+1)).$$

Additional dispersion can be incorporated by using a conditional NB distribution: the **ll-NB-AR(1) model** relies on the recursion $X_t|X_{t-1},\ldots \sim \text{NB}\left(1,\frac{n}{M_t+n}\right)$, where the parameter n controls the dispersion level.