



Article

Analysis and Forecasting of Risk in Count Processes

Annika Homburg ¹, Christian H. Weiß ^{1,*}, Gabriel Frahm ¹, Layth C. Alwan ² and Rainer Göb ³

¹ Department of Mathematics and Statistics, Helmut Schmidt University, 22043 Hamburg, Germany; homburga@hsu-hh.de (A.H.); frahm@hsu-hh.de (G.F.)

² Sheldon B. Lubar School of Business, University of Wisconsin-Milwaukee, Milwaukee, WI 53211, USA; alwan@uwm.edu

³ Department of Statistics, Institute of Mathematics, University of Würzburg, 97070 Würzburg, Germany; goeb@mathematik.uni-wuerzburg.de

* Correspondence: weissc@hsu-hh.de; Tel.: +49-40-6541-2779

Abstract: Risk measures are commonly used to prepare for a prospective occurrence of an adverse event. If we are concerned with discrete risk phenomena such as counts of natural disasters, counts of infections by a serious disease, or counts of certain economic events, then the required risk forecasts are to be computed for an underlying count process. In practice, however, the discrete nature of count data is sometimes ignored and risk forecasts are calculated based on Gaussian time series models. But even if methods from count time series analysis are used in an adequate manner, the performance of risk forecasting is affected by estimation uncertainty as well as certain discreteness phenomena. To get a thorough overview of the aforementioned issues in risk forecasting of count processes, a comprehensive simulation study was done considering a broad variety of risk measures and count time series models. It becomes clear that Gaussian approximate risk forecasts substantially distort risk assessment and, thus, should be avoided. In order to account for the apparent estimation uncertainty in risk forecasting, we use bootstrap approaches for count time series. The relevance and the application of the proposed approaches are illustrated by real data examples about counts of storm surges and counts of financial transactions.

Keywords: count time series; expected shortfall; expectiles; Gaussian approximation; mid quantiles; tail conditional expectation; value at risk



Citation: Homburg, Annika, Christian H. Weiß, Gabriel Frahm, Layth C. Alwan, and Rainer Göb. 2021. Analysis and Forecasting of Risk in Count Processes. *Journal of Risk and Financial Management* 14: 182. <https://doi.org/10.3390/jrfm14040182>

Academic Editor: Zbigniew Palmowski

Received: 26 March 2021

Accepted: 14 April 2021

Published: 16 April 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

A risk is associated with the possible occurrence of an adverse event, such as a loss in a financial context or a loss caused by a natural disaster. The corresponding probability distribution is used to prepare for the case when this undesirable situation actually arises (Klüppelberg et al. 2014; McNeil et al. 2015). A risk measure maps this probability distribution onto a real number that is then used for a risk assessment (Emmer et al. 2015). There are a couple of different risk measures that are commonly used in applications, all having their advantages and disadvantages, see Emmer et al. (2015) for a recent survey and evaluation. In what follows here, we consider risk measures in the context of a count random variable (r. v.) X , i.e., with the range consisting of non-negative integers from the set $\mathbb{N}_0 = \{0, 1, \dots\}$. More precisely, we distinguish between unbounded counts (having full \mathbb{N}_0 as their range) and bounded counts (having the range $\{0, \dots, n\}$ with given upper bound $n \in \mathbb{N} = \{1, 2, \dots\}$). We assume that increasing counts go along with an increasing “loss,” e.g., X may express the number of defects in a (low volume) manufacturing context, the number of infections in disease monitoring, or the number of insolvencies in an economic study, see also the data examples discussed in Sections 2.3 and 5. Therefore, the risk measures to be presented in Section 2 focus on the upper tail of the count distribution. Such risk measures for count r. v., and their estimation from independent and identically distributed (i. i. d.) sample data, were investigated by Göb (2011). Homburg (2020) used risk measures to evaluate the goodness of an approximating count distribution (among

other such measures), and Weiß and Testik (2019) applied them to the discrete run length distribution of a control chart as a tool for performance analysis.

In this article, we do not concentrate on a count r. v. as such, but rather on count time series x_1, \dots, x_T stemming from some autocorrelated count process $(X_t)_{t \in \mathbb{Z}}$ with $\mathbb{Z} = \{\dots, -1, 0, 1, \dots\}$ (the data-generating process, DGP), see Weiß (2018) for a comprehensive survey. This should not be confused with Lévy processes, which are often used in financial applications (Cont and Tankov 2004), where the jumps evolve in continuous time and can be real valued. A well-known Lévy process is the Poisson point process, which counts the number of points along the time axis. However, the jumps generated by a Poisson point process do not occur at fixed points in time. Further, Lévy processes have stationary and (mutually) independent increments. By contrast, the count processes considered here evolve in discrete time, have a discrete range, consisting of natural numbers, and may exhibit serial dependence between successive count observations. These attributes are not captured by Lévy processes. Furthermore, the field of potential applications is not limited to financial time series, but covers many further areas such as those mentioned in the previous paragraph. To our knowledge, the study of correlated count processes in relationship to risk measures is a unique contribution to the risk research literature.

The available time series data are used for model fitting, and the fitted model is taken as the base for forecasting the risk at some future time $T + h$ with forecast horizon $h \in \mathbb{N}$. Here, we distinguish between two scenarios. If we fit a count time series model to these data, and if we compute the risk forecasts by explicitly accounting for the integer nature of the counts, then we refer to them as *coherent risk forecasts*, picking up the terminology in Freeland and McCabe (2004) (The use of “coherent” in the context of risk forecasts should not be confused with its use within “coherent risk measures” in the sense of Artzner et al. (1999). A coherent risk forecast in the sense of Freeland and McCabe (2004) accounts for the integer nature of counts, but it need not rely on a coherent risk measure.). One may certainly ask why one should do it any differently at all. In practice, however, it is still common to ignore the integer nature of autocorrelated count time series and to apply, e.g., the ordinary Gaussian autoregressive moving-average (ARMA) models (see, for example, Rahardja 2020). Such an approach might be motivated by readily available implementations for ARMA model fitting in nearly all statistical software packages, or it might be caused by inadequate communication of count time series models. If fitting a Gaussian time series model to the counts x_1, \dots, x_T , and if using this model to compute the risk forecast for time $T + h$ (computed in the Gaussian way), we refer to it as an *approximate risk forecast* in the sequel. The relevant definitions and computations for both coherent and approximate risk forecasts are described in Section 2. Although approximate risk forecasting might appear attractive at a first glance, because of simplified computations and readily available software implementations (e. g., Chan and Nadarajah 2019), it is not clear whether the obtained risk forecasts are indeed competitive to the coherent risk forecasts or they lead to a serious misjudgment of the actual risk. The studies of Homburg et al. (2019, 2020) about point and interval forecasting of count time series may serve as a warning that approximate approaches might end up in rather misleading results. Therefore, besides a detailed analysis of coherent risk forecasting for various count processes in the presence of estimation uncertainty, we also investigate the performance of approximate risk forecasting in comparison to the coherent forecasting approach.

The outline of this article is as follows. First, we propose and discuss the relevant risk measures for counts, and we compare their properties to the case in which the same types of risk measure are applied to real-valued r. v., see Section 2. These analyses are then used to explain the findings in our comprehensive simulation study, where the performance of both coherent and approximate risk forecasts under estimation uncertainty is investigated for diverse types of count processes. The general design of our simulation study as well as our approaches for performance evaluation are presented in Section 3, whereas the obtained results are discussed in Section 4. We consider quite different kinds of count processes, unbounded vs. bounded counts, various types of autoregressive DGPs as well

as non-stationary count processes having seasonality and trend (a brief summary of the considered models is provided by Appendix A). Our analyses in Section 4 are illustrated by means of selected figures and tables, while the full simulation results are offered as supplemental material at <https://www.hsu-hh.de/mathstat/en/research/projects/forecastingrisk> (accessed on 26 March 2021). In Section 5, we apply the coherent risk measures to count time series about transactions of structured products at financial markets. We use appropriate types of time series bootstrapping in order to account for the apparent estimation uncertainty, and we also study the possible effect of model misspecification. Finally, we conclude our work in Section 6.

2. Risk Measures for Count Random Variables

In what follows, we first introduce five different types of risk measures for count r. v., see Section 2.1 as well as Table 1 for a brief summary. When applying these measures to continuously distributed r. v., two of them agree with other measures, i.e., the number of different risk measures reduces to three, see Table 2. Not only do the number and computation of risk measures differ between the discrete count case and the continuous Gaussian case, there are also considerable differences in their properties, which are discussed in Section 2.2. We conclude with a first data example in Section 2.3, which serves as a motivation for the subsequent analyses.

Table 1. Computation of risk measures presented in Section 2.1 for a count random variable X , where the required PMFs are provided by Tables A1 and A2 in the Appendix A.

$$\text{VaR}_\rho = \min \{x \in \mathbb{N}_0 \mid P(X \leq x) \geq \rho\}, \quad \text{see (1), where } P(X \leq x) = \sum_{y=0}^x P(X = y).$$

$$\text{TCE}_\rho = \frac{E[X] - E[X \mathbb{1}(X < \text{VaR}_\rho)]}{1 - P(X < \text{VaR}_\rho)}, \quad \text{see (2), where } E[X \mathbb{1}(X < \text{VaR}_\rho)] = \sum_{x=0}^{\text{VaR}_\rho - 1} x \cdot P(X = x).$$

$$\text{ES}_\rho = \text{TCE}_\rho + (\text{TCE}_\rho - \text{VaR}_\rho) \frac{\rho - P(X < \text{VaR}_\rho)}{1 - \rho}, \quad \text{see (4).}$$

EVaR_ρ is computed as the solution to Equation

$$\rho (E[X] - e) - (1 - 2\rho) E[(e - X) \mathbb{1}(X \leq e)] = 0, \quad \text{see (8),}$$

where $E[(e - X) \mathbb{1}(X \leq e)] = \sum_{x=0}^{\lfloor e \rfloor} (e - x) P(X = x)$.

$$\text{MVar}_\rho = \begin{cases} 0 & \text{if } \rho < \pi_0, \\ \kappa k + (1 - \kappa)(k + 1) & \text{if } \rho = \kappa \pi_k + (1 - \kappa) \pi_{k+1} \text{ for some } 0 < \kappa \leq 1, \\ n & \text{if } \rho > \pi_n, \end{cases} \quad \text{see (9).}$$

2.1. Definition of Risk Measures

Let X be a count r. v. referring to an undesirable event (the “loss,” e.g., the number of defects, of infections, or of insolvencies). The relevant types of probability mass function (PMF) are summarized in Tables A1 and A2 in the Appendix A. For a fixed risk level $\rho \in (0, 1)$, the most basic risk measure is the *value at risk* (VaR_ρ), which is defined to be the lower ρ th quantile of the distribution of X ,

$$\text{VaR}_\rho = \min \{x \in \mathbb{N}_0 \mid P(X \leq x) \geq \rho\}. \tag{1}$$

VaR_ρ is interpreted as a threshold that is only exceeded in at most $(1 - \rho) \cdot 100\%$ of all cases. Later, we compute the VaR_ρ (as well as any other risk measure considered here) from the conditional distribution of X_{T+h} , given the past x_T, \dots, x_1 , but to keep our notation simple, we now just write X . (We do not further stress the term “conditional” here to avoid confusion with another risk measure, which is sometimes referred to as the “conditional VaR” (see GÖB 2011; Rockafellar and Uryasev 2002) and which corresponds to the tail conditional expectation or expected shortfall, respectively, in our terminology, given that

the cumulative distribution function (CDF) of X is continuous.) Note that since X is a count r. v., also its VaR can only take non-negative integer values. This differs from the remaining risk measures to be considered here, which can take arbitrary positive real numbers as their outcome.

Table 2. Computation of risk measures presented in Section 2.1 for the case of a normally distributed random variable $Y \sim N(\mu, \sigma^2)$, i.e., where $(Y - \mu)/\sigma$ follows the standard normal distribution $N(0, 1)$. Here, $\phi(x)$ denotes the density function of $N(0, 1)$, and $\Phi(x)$ its CDF.

$VaR_\rho = MVaR_\rho = \mu + \sigma \Phi^{-1}(\rho),$	where $\Phi^{-1}(\rho)$ provides the ρ -quantile of $N(0, 1)$.
$TCE_\rho = ES_\rho = \mu + \sigma \frac{\phi(\Phi^{-1}(\rho))}{1 - \rho}.$	
E VaR_ρ is computed as the solution to Equation	
$\rho(\mu - e) + (1 - 2\rho)(\mu - e)\Phi\left(\frac{e - \mu}{\sigma}\right) - (1 - 2\rho)\sigma\phi\left(\frac{e - \mu}{\sigma}\right) = 0.$	

If $X \geq VaR_\rho$ happens, which is referred to as a shortfall, the actual loss will often be larger than the threshold value VaR_ρ itself (i.e., $X > VaR_\rho$ with probability > 0). Therefore, more refined risk measures, obtained by averaging VaRs or exceedances thereof, have been proposed in the literature to express the “typical loss” one has to prepare for if an exceedance of the VaR occurs, see [Chan and Nadarajah \(2019\)](#) for a survey. [Artzner et al. \(1999\)](#) propose to use the *tail conditional expectation* (TCE_ρ), which is defined as the conditional mean loss in the case of reaching or exceeding the VaR_ρ . More precisely,

$$TCE_\rho = E[X | X \geq VaR_\rho] = \frac{E[X] - E[X \mathbb{1}(X < VaR_\rho)]}{1 - P(X < VaR_\rho)}, \tag{2}$$

where the indicator function $\mathbb{1}(A)$ takes the value 1 (0) if A is true (false). The last expression in (2) is used for computing the value of TCE_ρ , with $E[X^k \mathbb{1}(X < VaR_\rho)]$ being equal to the finite sum $\sum_{x=0}^{VaR_\rho-1} x^k \cdot P(X = x)$. Note that the event “ $X = VaR_\rho$ ” is included in the condition (2), although some authors use the condition “ $X > VaR_\rho$ ” instead ([Göb 2011](#)). For the discrete count r. v. considered here, this makes a difference. For bounded counts, the condition “ $X \geq VaR_\rho$ ” is advantageous if VaR_ρ equals the upper bound n of the range $\{0, \dots, n\}$. In addition, note that (2) implies that $TCE_\rho \geq VaR_\rho$.

An alternative risk measure related to the VaR_ρ is the *expected shortfall* (ES_ρ), proposed independently by [Acerbi and Tasche \(2002\)](#) and [Rockafellar and Uryasev \(2002\)](#), which is defined by averaging VaRs as

$$ES_\rho = \frac{1}{1 - \rho} \int_\rho^1 VaR_u du. \tag{3}$$

As shown in Proposition 3.2 of [Acerbi and Tasche \(2002\)](#), the ES_ρ can be computed from VaR_ρ and TCE_ρ as follows:

$$ES_\rho = TCE_\rho + (TCE_\rho - VaR_\rho) \frac{\rho - P(X < VaR_\rho)}{1 - \rho}. \tag{4}$$

This shows that $ES_\rho \geq TCE_\rho \geq VaR_\rho$. While $P(X < VaR_\rho) < \rho$ by (1) for discrete r. v., we have that $P(X < VaR_\rho) = \rho$ for continuously distributed r. v.. So ES_ρ and TCE_ρ necessarily coincide in the continuous case (also see Table 2). In the discrete case, it may happen for bounded counts that VaR_ρ agrees with the upper bound n , in which case ES_ρ

and TCE_ρ agree with n as well. Otherwise, we have the strict inequality $ES_\rho > TCE_\rho > VaR_\rho$. Note that (4) can be rewritten as

$$TCE_\rho = \underbrace{\frac{1-\rho}{1-P(X < VaR_\rho)}}_{\leq 1} ES_\rho + \left(1 - \frac{1-\rho}{1-P(X < VaR_\rho)}\right) VaR_\rho, \tag{5}$$

which shows that TCE_ρ is a weighted mean of ES_ρ and VaR_ρ . Later in Section 2.2, Equation (5) shall be helpful for explaining some properties of TCE_ρ . The application of the risk measures VaR_ρ , ES_ρ , and TCE_ρ to count r. v. was also considered by G6b (2011) and Homburg (2020), but both using a modified definition of TCE_ρ compared to (2). The integer-valued VaR_ρ was also investigated by Homburg et al. (2019) as a non-central point forecast for count time series.

Recently, *expectiles* have also been considered as an another type of risk measure, see Bellini and di Bernardino (2017) and the references therein. Expectiles were developed by Newey and Powell (1987) as an alternative to quantiles. In analogy to the VaR_ρ according to (1), we define the *expectile*- VaR_ρ ($EVaR_\rho$) as the ρ -expectile of X . The latter, in turn, is defined to be the (unique) minimizer of the following type of asymmetric quadratic loss, provided that the second-order moment of X exist:

$$EVaR_\rho = \arg \min_e E[\rho(X - e)^2 \mathbb{1}(X > e) + (1 - \rho)(e - X)^2 \mathbb{1}(X \leq e)]. \tag{6}$$

By contrast, the ρ -quantile (i.e., the VaR_ρ) minimizes $E[\rho(X - q) \mathbb{1}(X > q) + (1 - \rho)(q - X) \mathbb{1}(X \leq q)]$ in q (provided that $E[X]$ exists). While VaR_ρ is integer-valued for count r. v., the $EVaR_\rho$ can take any positive real number. The $EVaR_\rho$ is also uniquely characterized as the solution to the following equation (commonly referred to as the “first-order condition”):

$$\rho E[(X - e) \mathbb{1}(X > e)] = (1 - \rho) E[(e - X) \mathbb{1}(X \leq e)]. \tag{7}$$

Equation (7) shows that the $EVaR_\rho$ can be interpreted as that threshold value e where the ratio of the mean exceedance to the mean shortfall, $E[(X - e) \mathbb{1}(X > e)]/E[(e - X) \mathbb{1}(X \leq e)]$, equals the odds ratio $(1 - \rho)/\rho$. For computations, it is advantageous to rewrite (7) as

$$\rho (E[X] - e) - (1 - 2\rho) E[(e - X) \mathbb{1}(X \leq e)] = 0, \tag{8}$$

where the truncated mean is computed as the finite sum $E[(e - X) \mathbb{1}(X \leq e)] = \sum_{x=0}^{\lfloor e \rfloor} (e - x) P(X = x)$. Equation (8) also shows that the 0.5-expectile is just equal to the mean $E[X]$. A critical comparison of all aforementioned risk measures is provided by Emmer et al. (2015).

Finally, let us return to the VaR_ρ defined by (1). For the discrete count r. v. considered here, VaR_ρ behaves fundamentally different to the other considered risk measures: While TCE_ρ , ES_ρ , and $EVaR_\rho$ are real-valued in general, VaR_ρ takes values from the set of non-negative integers. This does not only constitute an additional difficulty if working with a Gaussian approximation (see Table 2), it also causes problems if estimating the VaR_ρ from given data. For example, we have a degenerate asymptotic distribution for the sample VaR_ρ , see Theorem 6 in Jentsch and Leucht (2016). For these reasons, we propose to consider also the following modification of VaR_ρ , which we refer to as the *mid*- VaR_ρ ($MVaR_\rho$). It is defined as the ρ -mid-quantile of X , a concept dating back to Parzen (1997) and further investigated by Ma et al. (2011) and Jentsch and Leucht (2016). Let $\pi_k = P(X < k) + P(X = k)/2$ (“mid-probability”) for $k = 0, 1, \dots$, and let n denote the upper limit of the range of X (with $n = \infty$ for unbounded counts). Then, we have that

$$MVaR_\rho = \begin{cases} 0 & \text{if } \rho < \pi_0, \\ \kappa k + (1 - \kappa)(k + 1) & \text{if } \rho = \kappa \pi_k + (1 - \kappa) \pi_{k+1} \text{ for some } 0 < \kappa \leq 1, \\ n & \text{if } \rho > \pi_n. \end{cases} \tag{9}$$

The last case is only relevant for bounded counts, i.e., if $n < \infty$. By using linear interpolation in (9), $MVaR_\rho$ is positive and real-valued, and much more suitable for statistical inference (Jentsch and Leucht 2016; Ma et al. 2011). For a continuously distributed r. v., we set $MVaR_\rho = VaR_\rho$, in analogy to the equality $ES_\rho = TCE_\rho$. Summaries of the discussed risk measures are provided by Tables 1 and 2.

Remark 1. We are well aware that many further risk measures have been proposed in the literature. For example, several risk measures are related to the three-parameter family proposed by Stone (1973), taking the form $E[(X - a)^k \mathbb{1}(X \geq b)]$ with $k \geq 0$ and $a, b \geq 0$ (for count r. v.). In fact, also the numerator of the TCE in (2), $E[X \mathbb{1}(X \geq VaR_\rho)]$, has to be mentioned in this context, which is obtained by setting $k = 1$, $a = 0$, and $b = VaR_\rho$. Other examples are the semi-deviations discussed by Ruszczyński (2010, p. 242) in the context of dynamic risk measures, i.e., $E[(X - \mu)^k \mathbb{1}(X \geq \mu)]^{1/k}$. To keep the scope of the present article manageable, we refrain from including further risk measures into our study than those five defined before. But we point out that also these Stone measures are easily computed in analogy to (2), i.e., as $E[(X - a)^k] - E[(X - a)^k \mathbb{1}(X < b)]$, where the last term only requires to compute a finite sum.

For the sake of a concise presentation, we also dispense with including additional cost considerations, as it was done, for example, by Alwan and Weiß (2017). We may interpret the VaR_ρ as assigning constant costs to the risk event $X \geq VaR_\rho$, whereas the TCE_ρ assumes linearly increasing costs. Power cost schemes could be achieved using adapting the aforementioned Stone approach.

2.2. Some Properties of Risk Measures for Counts

Let us have a first look at some properties of the five risk measures for counts from Section 2.1 (see Table 1), in contrast to their normal counterparts from Table 2. We start with the normal case, as this is more familiar to most practitioners. The first graph in Figure 1 plots the risk measures for a normal distribution with mean μ and variance $\sigma^2 = \mu$ (the Poisson distribution discussed afterwards has the same equidispersion property), where the risk level is given by $\rho = 0.95$. It can be seen that all risk measures are continuous and strictly increasing functions in μ , with $ES_{0.95} = TCE_{0.95} > VaR_{0.95} = MVaR_{0.95}$. For the presented scenario, all these measures are also greater than $EVaR_{0.95}$. Here and in the remaining plots of Figure 1, we also recognize increasing spread between the different measures for increasing μ , which is due to the variance being proportional to μ in all scenarios.

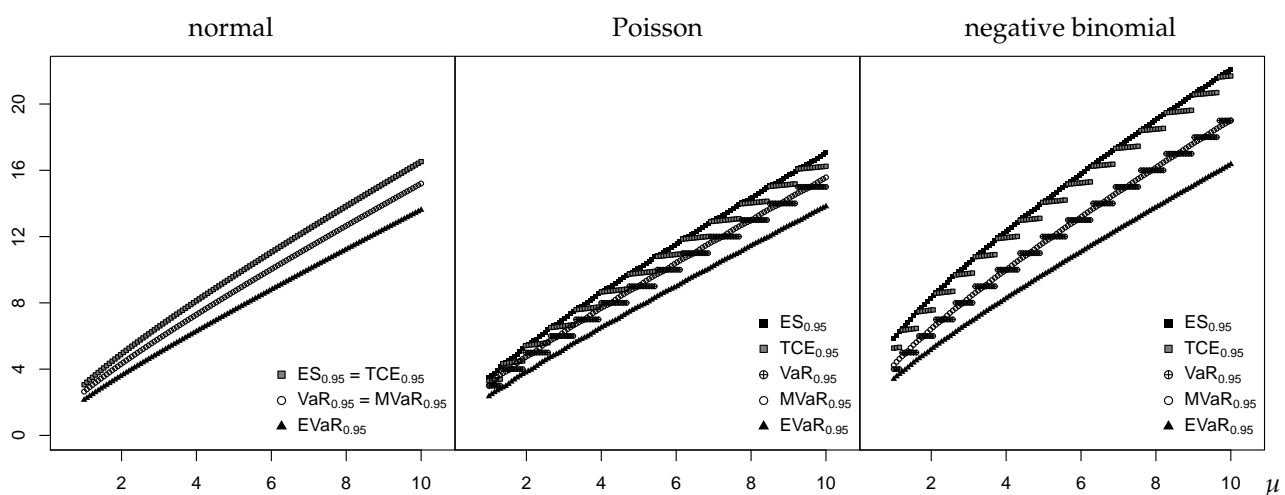


Figure 1. Different risk measures for distributions $N(\mu, \mu)$, $Poi(\mu)$, and $NB(\mu/(I - 1), 1/I)$ with $I = 2.4$, respectively, plotted against mean μ .

Next, we turn to the case of the Poisson (Poi) distribution with mean μ , $Poi(\mu)$, which is the most common distribution for unbounded counts (see Table A1 in Appendix A for computational details). As is well known, the shape of the $Poi(\mu)$ distribution approaches

that of the normal distribution $N(\mu, \mu)$ with increasing μ . So it is natural to expect that also the risk measures computed for $\text{Poi}(\mu)$ are close to those for $N(\mu, \mu)$. This, however, is not true in general as can be seen from the second graph in Figure 1. First, we recognize that $\text{ES}_{0.95} > \text{TCE}_{0.95}$ in contrast to the normal case, and also $\text{VaR}_{0.95} \neq \text{MVar}_{0.95}$ in those cases where $\text{MVar}_{0.95}$ does not take an integer value (recall the discussion in Section 2.1). Second, $\text{VaR}_{0.95}$ is not continuous but a piecewise constant function in μ , taking only integer outcomes, whereas the remaining measures are strictly increasing in the Poisson mean μ . Third, the $\text{TCE}_{0.95}$ behaves rather different from the normal case. It is only a piecewise continuous function in μ , which can be explained by Equation (5): $\text{TCE}_{0.95}$ is a weighted mean of the continuous $\text{ES}_{0.95}$ and the piecewise constant $\text{VaR}_{0.95}$. In addition, the actual values of $\text{TCE}_{0.95}$ are often closer to those of $\text{VaR}_{0.95}$ than of $\text{ES}_{0.95}$, which is opposite to the normal case with $\text{ES}_{0.95} = \text{TCE}_{0.95} > \text{VaR}_{0.95}$. This happens if the weight $\frac{1-\rho}{1-P(X < \text{VaR}_\rho)}$ in (5) is much smaller than 1, i.e., if $P(X < \text{VaR}_\rho)$ is much smaller than ρ . The reason for this phenomenon is given by the rather low dispersion of the Poi distribution (equidispersion): The upper tail of its PMF declines quickly towards 0, so we have rather large gaps between the attainable tail probabilities. Finally, there are also a few analogies between the normal and the Poi-case: $\text{ES}_{0.95}$, $\text{MVar}_{0.95}$, and $\text{EVar}_{0.95}$ behave quite similarly at a first glance. So, while there appears to be some chance that $\text{ES}_{0.95}$, $\text{MVar}_{0.95}$, and $\text{EVar}_{0.95}$ relying on a normal approximation might be close to their actual Poisson values, we cannot expect that such an approximation works well for $\text{TCE}_{0.95}$ and $\text{VaR}_{0.95}$ (Homburg et al. (2019) investigate discrete quantile forecasts and apply an additional discretization (ceiling) to the approximate Gaussian quantiles. Nevertheless, the approximation quality turns out to be very poor.).

In the third scenario shown in Figure 1, we consider counts stemming from a negative binomial (NB) distribution with mean μ and dispersion index $I = \sigma^2/\mu = 2.4$, see Table A1. So, while the Poi distribution is equidispersed ($I = 1$), the NB distribution exhibits overdispersion. This additional dispersion has the following consequences: We have a stronger discrepancy between $\text{ES}_{0.95}$, $\text{MVar}_{0.95}$, and $\text{EVar}_{0.95}$, and the measures $\text{TCE}_{0.95}$ and $\text{VaR}_{0.95}$ are now clearly separated. In fact, $\text{TCE}_{0.95}$ is now more close to $\text{ES}_{0.95}$ than to $\text{VaR}_{0.95}$, so these measures behave more similar to the normal case in this respect. This is explained by (5) and by the additional probability mass in the NB's upper tail, which implies that $P(X < \text{VaR}_\rho)$ is closer to ρ than in the Poi-case.

2.3. Data Example: Storm Surges in Norderney

At this point, let us look at a first real-world application. Storm surges have threatened the German coasts since time immemorial, with increasing economic damage because of an intensified use of the coastal areas, see Jensen and Müller-Navarra (2008) for a comprehensive survey. The website "Storm Surge Monitor" at <https://www.sturmflut-monitor.de/index.php.en> (accessed on 15 December 2020) provides information about the temporal development of storm surges during the last decades at different locations along the North and the Baltic Sea. As an illustrative example, let us analyze the last hundred seasons of storm surges in Norderney (an island in the German Bight, North Sea), where a season extends from July of the previous year to June of the current year. The count time series x_1, \dots, x_T with $T = 100$ provides the numbers of storm surges for the seasons 1921–2020, see the plot in Figure 2a. The corresponding sample autocorrelation function (ACF) does not show significant deviations from 0, so it appears reasonable to model the data as being i. i. d. Their sample mean equals $\hat{\mu} = 1.91$, their sample dispersion index $\hat{I} \approx 1.704$. So we are concerned with a considerable degree of overdispersion such that the NB distribution appears to be a plausible candidate model. The PMF $P(X = x)$ of the fitted NB(2.713, 0.587) distribution is compared to the sample PMF in Figure 2b, showing a reasonable agreement. Thus, we shall model the storm surge counts as being i. i. d. according to NB(2.713, 0.587).

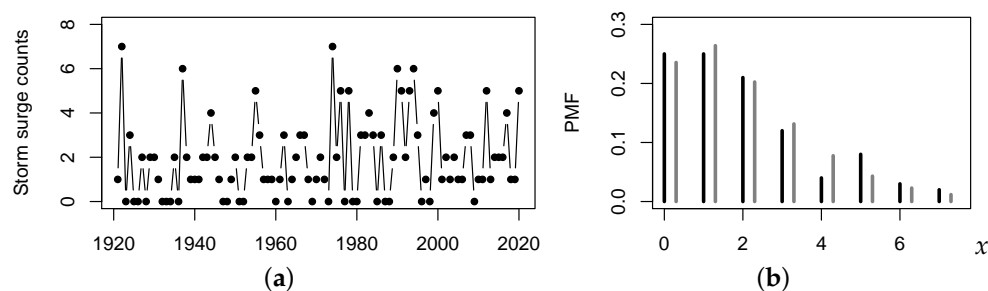


Figure 2. Plot of the storm surge counts (per season) in (a), their sample PMF (black) together with an NB-fit (grey) in (b).

Since the potential for economic damages (loss) increases with an increasing number of storm surges, we focus our risk analysis on the upper tail of the storm surges' distribution. For convenience, the risk level is again set at $\rho = 0.95$. We start by computing the five risk measures for counts from Section 2.1 for the fitted NB-model (coherent risk forecasts). Here, the $\text{VaR}_{0.95}$ takes the value 5, i.e., in at most 5% of all seasons, we expect to observe more than five storm surges. The $\text{MVar}_{0.95}$ can be understood as a refinement of the discrete $\text{VaR}_{0.95}$, leading to the value ≈ 5.533 . If having a season with ≥ 5 storm surges, then we actually expect $\text{TCE}_{0.95} \approx 6.025$ storm surges on average according to the NB-fit, whereas the $\text{ES}_{0.95}$ leads to the even larger value ≈ 6.825 . The $\text{EVar}_{0.95}$, in contrast, relying on the concept of expectiles, leads to a clearly lower risk value, namely ≈ 4.444 . These risk assessments might be used for defining appropriate preparations and countermeasures in practice. But here, they shall serve as the starting point for further statistical analyses.

First, we check what would have happened if we would have used a fitted normal distribution for computing the risk measures, recall Table 2. Then, $\text{VaR}_{0.95} = \text{MVar}_{0.95} \approx 4.877$, i.e., these approximate risk forecasts would be lower than their coherent counterparts. The same happens for $\text{TCE}_{0.95} = \text{ES}_{0.95} \approx 5.631$ as well as for $\text{EVar}_{0.95} \approx 3.967$. So, using the Gaussian approximation, the risk is always judged lower than if using the coherent NB-model. For the considered application, such a risk underrating might lead to, e.g., insufficient capital reserves for dealing with the economic damages caused by the storm surges. Given these discrepancies, the question arises whether risk predictions based on a Gaussian approximation generally tend to underrate the risk, or whether we have found only sporadic evidence in this data example. To answer this question, we present results from a comprehensive simulation study later in Section 4.

Second, even if using the more “conservative” coherent risk forecasts, one has to recall that these rely on a fitted model with $T = 100$ observations. The question arises to what extent the apparent estimation uncertainty affects the computed risk forecasts. Let us investigate this question with a bootstrap experiment. We generated 1000 i. i. d. samples from the $\text{NB}(2.713, 0.587)$ distribution (parametric bootstrap), fitted again a NB-model to each of the bootstrap samples, and computed the five risk measures based on these model fits. So for each risk measure, 1000 bootstrap replicates are available. Part (a) of Figure 3 shows boxplots of the replicated risk forecasts, which exhibit a lot of dispersion. Note that the boxplots of VaR and (slightly weakened) of TCE look degenerate because of the discreteness pattern already discussed in Section 2.2. So for practice, it appears to be advisable to account for the apparent estimation uncertainty before defining the countermeasures. Since the risk measures express some kind of “worst-case” scenario, a reasonable solution could be to compute an upper quantile from the bootstrapped risk forecasts and to use this quantile for decision making. The table in Figure 3b presents such quantiles for the levels $\gamma = 0.90, 0.95, 0.99$ for illustration. The task of accounting for the effect of estimation uncertainty on the coherent risk forecasts is further investigated in Section 5 below.

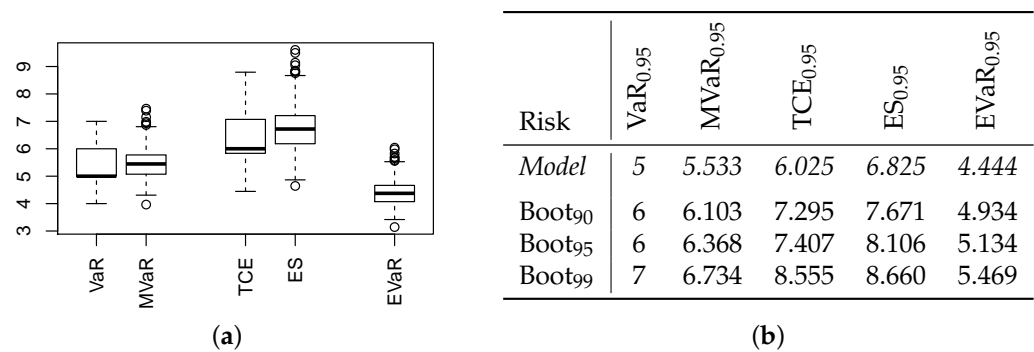


Figure 3. Boxplots of 1000 bootstrap replicates of VaR_{0.95}, MVaR_{0.95}, TCE_{0.95}, ES_{0.95}, and EVaR_{0.95} in (a), their model-based forecasts (italic font) as well as their bootstrapped γ -quantile (Boot _{γ}) in (b).

3. Simulation Study and Performance Evaluation

While the preliminary investigations of Section 2.2 provided some useful insights of the theoretical properties of the different risk measures under certain scenarios, we now turn to our main focus of this research, the performance of risk forecasting for count time series under estimation uncertainty. For the considered type of count DGP (X_t), and for a given time series x_1, \dots, x_T thereof, we fit either the actual count time series model to the data (coherent modeling), or a corresponding Gaussian time series model (approximate modeling), in analogy to Section 2.3. For the risk measure R of interest (one of the measures discussed in Section 2) and for the forecast horizon $h \in \mathbb{N}$, we compute

- the true risk R_{T+h} at time $T + h$, given x_1, \dots, x_T , by using the DGP’s true model;
- the corresponding coherent risk forecast value $\hat{R}_{T+h,c}$ by using the coherent model fit;
- and the approximate risk forecast value $\hat{R}_{T+h,a}$ based on the approximate model fit.

Then, we look at the differences between the respective forecast and true risk value, i.e., at

$$\hat{R}_{T+h,c} - R_{T+h} \quad \text{and} \quad \hat{R}_{T+h,a} - R_{T+h}, \quad \text{respectively.}$$

The first of these differences, $\hat{R}_{T+h,c} - R_{T+h}$, expresses the effect of estimation error during coherent modeling, whereas $\hat{R}_{T+h,a} - R_{T+h}$ is the combined effect of both estimation and approximation error. Generally, these differences should be close to zero to indicate a reasonable forecast precision, but also the sign of the deviation is important: If $\hat{R}_{T+h,c} - R_{T+h} < 0$ (risk underrating), then the risk at time $T + h$ is judged to be too low, which might have severe consequences in practice. For example, if the counts X_t express the number of infections by a severe disease and if the forecast value $\hat{R}_{T+h,c}$ is used to determine the capacity of the emergency room, then risk underrating may lead to the death of patients because of insufficient medical care. More generally, risk underrating can lead to the capital reserve being chosen too low, also see the discussion in Kerkhof et al. (2010) as well as the storm surge example in Section 2.3. Risk overrating (i.e., $\hat{R}_{T+h,c} - R_{T+h} > 0$), in contrast, can be understood as a conservative risk forecast. For the disease example, it may cause unnecessary costs (too much capacity of the emergency room) but does not endanger human life. Since deviations are unavoidable if parameters are estimated, it is plausible to get a balanced proportion of risk over- and underrating (50:50). Here, the frequency of overrating (plus possibly exact matches) is just the counterfrequency to underrating, i.e., it is sufficient to look at one of these frequencies. But in view of the above discussion about possible consequences of risk under- and overrating, our primary focus is on risk underrating here. Thus, both the actual extent of the deviations $\hat{R}_{T+h,c} - R_{T+h}$ as well as the tendency to negative deviations (i.e., if $\hat{R}_{T+h,c} - R_{T+h} < 0$ happens more often than with frequency 50%) are important performance criteria.

In our simulation study to be presented in Section 4, we generated 1000 count time series for each considered scenario, see Table 3. The model fitting was done by the method of moments, using the moment formulae provided by Appendix A. For each time series,

we computed all coherent and approximate risk forecasts (according to Table 2 in the approximate case, and by plugging-in Table A2 into Table 1 in the coherent case) as well as the corresponding deviations to the respective true risk values. For convenience, the risk level was set to $\rho = 0.95$ and the forecast horizon to $h = 1$. The resulting deviations are analyzed in various ways. As a visual tool, we compute lean types of boxplots for each set of 1000 deviations. These boxplots show the median of the deviations as a black dot, the quartiles are connected by a thick grey line, and the 10%- and 90%-quantiles by a thin black line. Then, having fixed the remaining parameters, we plot these boxplots against increasing values of the mean μ . The resulting graphs give insight into the absolute extent of the deviations and they allow us to judge how frequently and how severely we are concerned with a risk underrating.

Table 3. Scenarios for different DGPs of simulation study, with 1000 replications each.

Means $\mu \in \{1, 1.075, \dots, 9.925, 10\}$ for unbounded counts, upper bounds $n \in \{10, \dots, 130\}$ and probability $\pi \in \{0.15, 0.45\}$ for bounded counts.
Dispersion ratios $I \in \{1.4, 2.4\}$ if considering overdispersion.
Dependence parameter α in $\{0.33, 0.55, 0.8\}$ (ACF at lag 1), 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 $T \in \{75, 250, 2500\}$.

These visual inspections are supported by tables (the full set of tables is presented in the supplemental material) showing the following summarizing statistics:

- The relative frequency of risk underrating for each scenario, calculated as the sample mean of $\mathbb{1}(\hat{R}_{T+h,\cdot} < R_{T+h})$, and
- the mean severity of risk underrating, calculated as the sample mean of $(\hat{R}_{T+h,\cdot} - R_{T+h}) \mathbb{1}(\hat{R}_{T+h,\cdot} < R_{T+h})$ divided by the mean of $\mathbb{1}(\hat{R}_{T+h,\cdot} < R_{T+h})$.

These statistics are calculated for fixed parameters α , I , T etc., recall Table 3, but across all 121 different values of μ or n , respectively, and all 1000 replications, i.e., by averaging across 121×1000 deviation statistics.

4. Performance of Risk Forecasts for Count Processes

In this section, we present selected results from our simulation study regarding the performance of coherent and approximate risk forecasting for count processes, while the full set of tables with summarizing statistics is provided in the supplemental material. The general design of our simulations and analyses are described in Section 3, the investigated risk measures in Section 2, and the considered count time series models in Appendix A. The full simulation results are available at <https://www.hsu-hh.de/mathstat/en/research/projects/forecastingrisk> (accessed on 26 March 2021).

4.1. INAR(1) Count DGPs

The first type of DGP to be discussed follows the INAR(1) model (integer-valued autoregressive, see Appendix A.1 for details), which is probably the most well-known model for count time series. We consider both the equidispersed Poi-INAR(1) model (having Poisson-distributed innovations) as well as the overdispersed NB- and ZIP-INAR(1) models (having NB- and ZIP-distributed innovations, respectively), where ZIP abbreviates the zero-inflated Poisson distribution. Thus, the INAR(1) models are parametrized by the triple (μ, α, I) , see Table 3. Note that for the ZIP model, overdispersion is caused by an isolated point mass in zero, while the NB's overdispersion is caused by a flattened PMF compared to the Poi-case (also recall Figure 2b). It will be shown in this section that the approximate risk forecasts perform considerably worse than their coherent counterparts, and that additional overdispersion leads to a more or less disastrous performance.

Figure 4 shows the lean type of boxplots described in Section 3 regarding the deviations for the five risk measures of Section 2. Here, the INAR(1)-DGP is either equidispersed (Poi-case, i.e., $I = 1$) or overdispersed with $I = 2.4$ (NB- and ZIP-case), and the risk forecasts rely on a coherent model fit. Figure 5 is built up in the same way, but using a Gaussian AR(1) approximation for risk forecasting. Figures 4 and 5 are limited to the case of medium dependence ($\alpha = 0.55$) and sample size $T = 250$, but further scenarios are covered by the summarizing statistics in Tables 4 and 5 (the printed values for $\alpha = 0.55$ and $T = 250$ correspond to the graphs in Figures 4 and 5), and by the supplemental material. Let us start with the coherent risk forecasts in Figure 4. In contrast to the other risk measures introduced in Section 2, the coherent VaR forecast is the only one being discrete-valued. As a result, the boxplots of deviations either reduce to a single dot (so the large majority of VaR forecasts exactly matches the true VaR), or the lines extend to ± 1 (i.e., the true VaR is missed by ± 1); a similar pattern was observed for the VaR's boxplot in Figure 3b. The remaining risk measures are real-valued, so we will hardly observe an exact match. Instead, it would be plausible to get a balanced proportion of risk over- and underrating (50:50). The boxplots of the coherent ES, MVaR, and EVaR forecasts are indeed roughly symmetric around 0, with ES showing more dispersion than MVaR and EVaR. The largest dispersion is observed for the overdispersed NB-INAR(1) DGP, while the ZIP-INAR(1) DGP shows clearly less dispersion though still more than in the equidispersed Poi-case. Finally, the boxplots of TCE in Figure 4 show a conspicuous pattern. In the Poi-case, they nearly look the same as for the VaR and differ notably from the ES boxplots, which agrees with our findings in Sections 2.2 and 2.3. A similar pattern is observed in the ZIP case, which is plausible as the ZIP's upper tail is that of a Poi distribution. For the NB-INAR(1) DGP, in contrast, the TCE's boxplots of deviations are more similar to the ES than to the VaR case, again being plausible in view of Section 2.2. Generally, due to the higher variation among the TCE and ES values, these may produce risk forecasts deviating more from the corresponding true risk values than EVaR and MVaR.

If we compare next Figure 4 to Figure 5, it becomes clear that risk forecasting based on a Gaussian approximation might be very misleading. While in the Poi-case, the boxplots vary roughly around a constant value (although this constant is negative for VaR, ES, MVaR, and EVaR, i.e., we have a tendency to risk underrating), they show a severely non-constant behavior (against μ) in both overdispersion scenarios (this effect is already apparent for a more moderate extent of overdispersion such as $I = 1.4$). In the NB-case, we have a strong risk underrating throughout (As discussed by [Homburg et al. \(2019, p. 13\)](#), the normal distribution's upper tail is too light compared to the NB-case, explaining the strong tendency to risk underrating. It is clear that the underrating would even be intensified if we would consider a superlinear cost scheme, recall Remark 1.). By contrast, the ZIP case turns from underrating to overrating with increasing μ (these observations are in line with the analysis of the ceiled VaR in Section 4.1 of [Homburg et al. \(2019\)](#)). This behaviour is more extreme for TCE and ES than for the MVaR and EVaR. For VaR and TCE, we also have an additional discreteness pattern.

Finally, let us look at the summarizing statistics in Tables 4 and 5, recall Section 3 for their definition. Generally, to save some space, we provide such tables solely in the supplemental material. But since this is the first time where we refer to such tables, we show the tables for $I = 2.4$ in the main manuscript for illustration. The discrete coherent VaR forecasts have high frequencies of exact matches (especially in the Poi-case), going along with the low frequencies of underrating in Table 4. If, however, an underrating happens, it is at least -1 due to the integer nature of the coherent VaR forecast, explaining the values ≤ -1 in Table 5. Because of this discreteness feature, it is difficult to directly compare these values to the corresponding statistics of the real-valued approximate VaR forecasts. But it is apparent from Table 4 that the Gaussian VaR forecasts underrate the true VaR with high frequency in most cases.

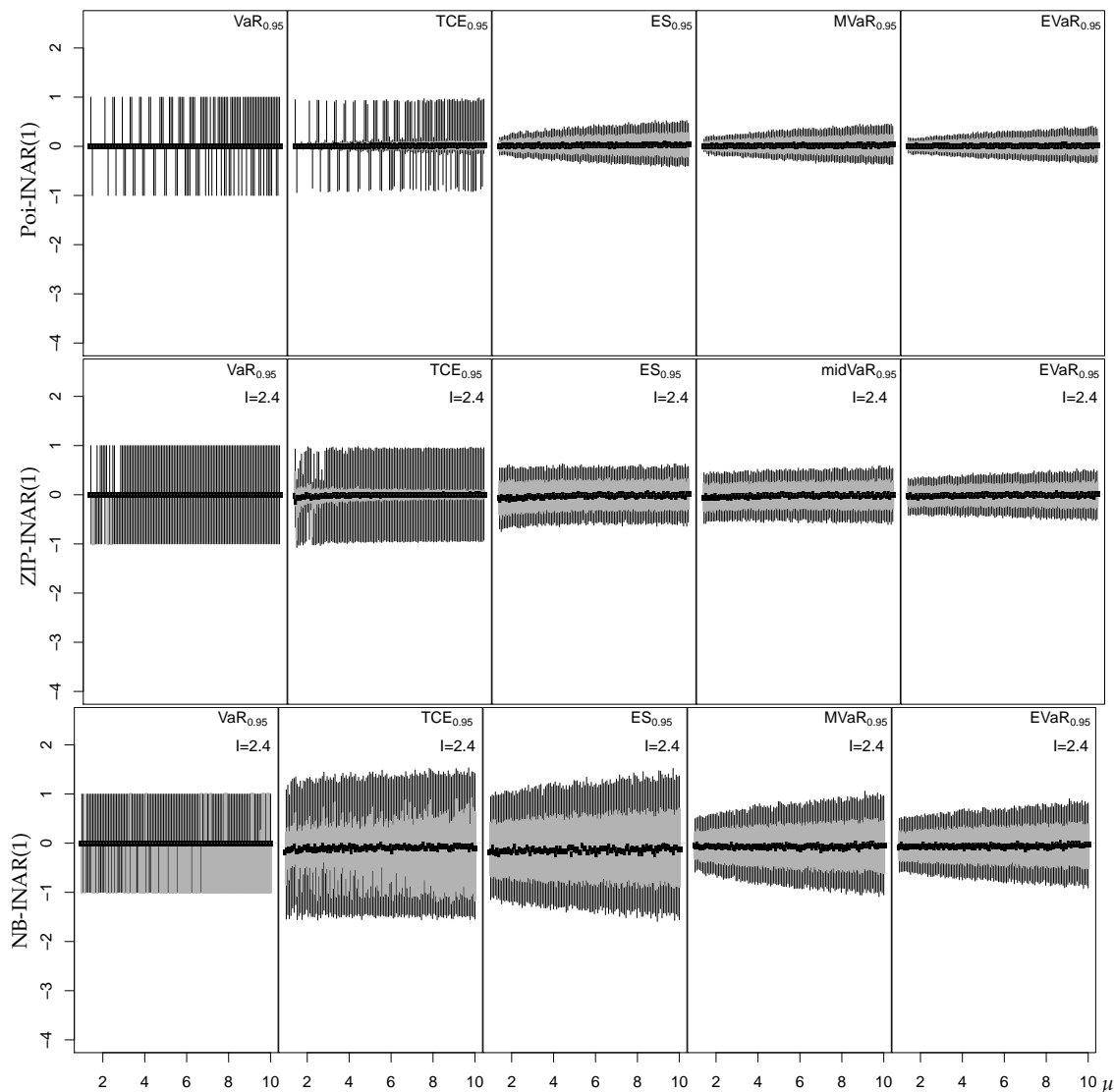


Figure 4. Boxplots of deviations for $VaR_{0.95}$, $TCE_{0.95}$, $ES_{0.95}$, $MVaR_{0.95}$, and $EVaR_{0.95}$. One-step-ahead forecasts based on coherent model fit for Poi-, ZIP-, and NB-INAR(1) DGP with $\alpha = 0.55$ and sample size $T = 250$.

The remaining risk measures are real-valued. Their coherent forecasts usually show a rather well-balanced proportion of risk over- and underrating (close to 50:50, see Table 4), being quite similar for the different risk measures. Exceptions happen for strong dependence ($\alpha = 0.8$) and sample sizes $T \leq 250$, then underratings are clearly less frequent than 50% in the Poi-case (conservative risk forecasting), whereas their frequency is increased in the overdispersion scenarios. The underrating frequencies of the approximate risk forecasts are usually much larger than those of their coherent counterparts (often $\geq 80\%$), with a few exceptions for the TCE. The differences between coherent and Gaussian approximation are, for most cases, made more profound with increasing sample size. Compared to the coherent model, the approximate risk forecasts do not only lead to a more frequent but also a more severe underrating, see Table 5, especially in the presence of overdispersion (and in the ZIP case, also more pronounced with increasing autocorrelation). The mean severities of the coherent TCE, ES, MVaR, and EVaR forecasts in Table 5 get considerably reduced for increasing T , while this happens much more mildly for their approximate counterparts.

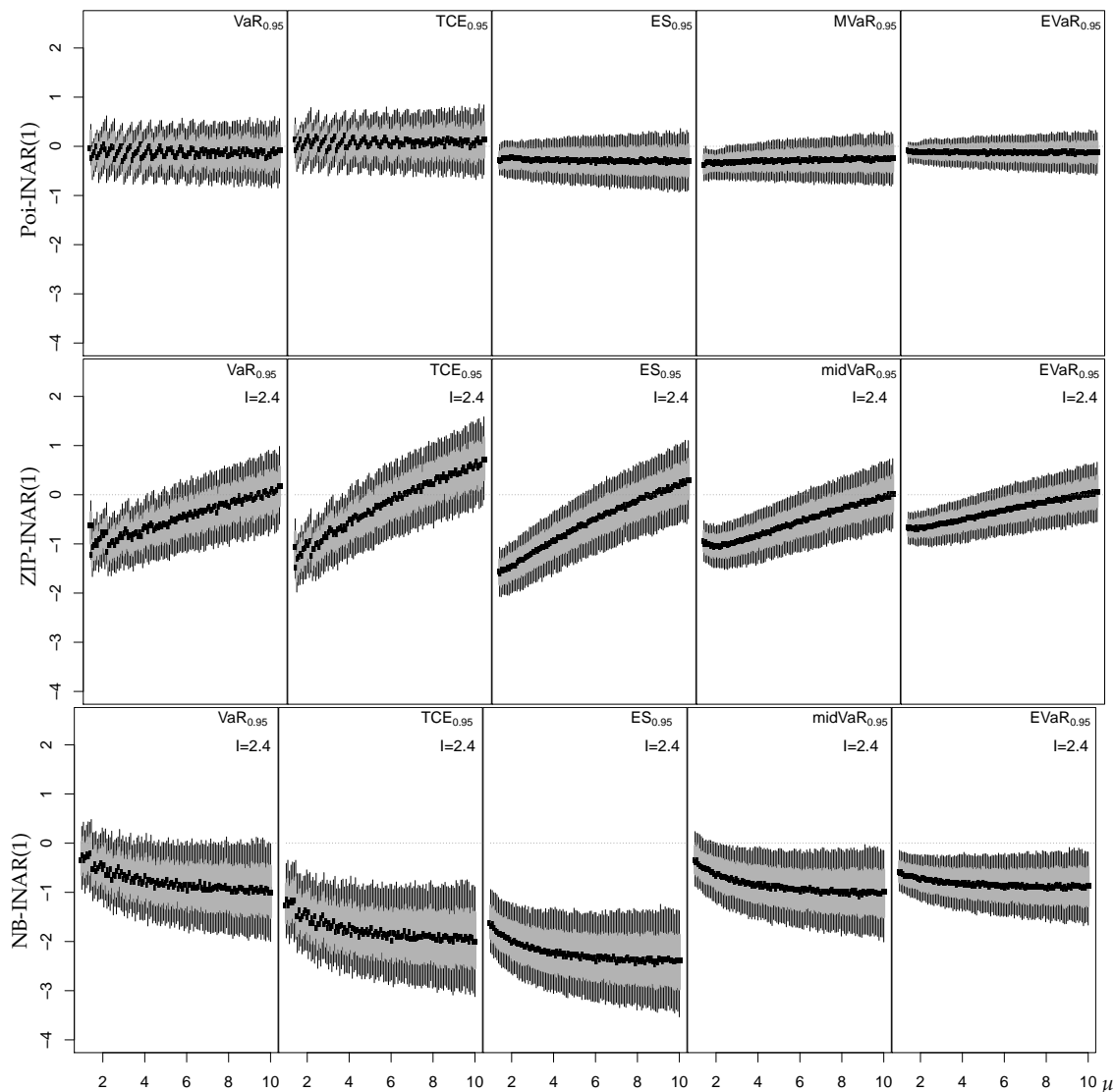


Figure 5. Boxplots of deviations for $VaR_{0.95}$, $TCE_{0.95}$, $ES_{0.95}$, $MVaR_{0.95}$, and $EVaR_{0.95}$. One-step-ahead forecasts based on Gaussian approximation for Poi-, ZIP-, and NB-INAR(1) DGP with $\alpha = 0.55$ and sample size $T = 250$.

In summary, EVaR and MVar appear to be less affected by estimation and approximation error than TCE and ES. The boxplots show that the coherent risk forecasts are generally close to the true risk forecast, and the estimation error creates a rather balanced under- and overrating of the true risk. The performance of these measures is difficult to compare to that of the VaR, because the VaR is the only discrete risk measure. But also the TCE shows strong discreteness effects in the Poi- and ZIP-case. The Gaussian AR(1) approximation leads to a much more frequent and severe underrating of the actual risk, and it is not able to mimic the integer-valued VaR. In particular, additional overdispersion leads to a severe deterioration of the approximation quality, with different patterns in the NB- and the ZIP-case. Since the coherent VaR forecasts and, to some degree, also the TCE forecasts are clearly affected by discreteness, a Gaussian approximation of these measures has an inherent significant disadvantage. The coherent VaR and TCE forecasts also show a rather different performance than the coherent ES, MVar, and EVaR forecasts. Therefore, we shall concentrate the following discussion on the latter three risk measures, while the supplemental material provides full results on all risk measures.

Table 4. Relative frequency of risk underrating for Poi-, ZIP, and NB-INAR(1) DGPs, coherent (“Coh”) vs. approximate (“Gau”) forecasts, for different (α, T), $I = 2.4$ in case of overdispersion and forecast horizon $h = 1$, computed across all simulation runs and all levels of μ .

	Poi-INAR(1) DGP						ZIP-INAR(1) DGP						NB-INAR(1) DGP						
	$\alpha = 0.33$		$\alpha = 0.55$		$\alpha = 0.80$		$\alpha = 0.33$		$\alpha = 0.55$		$\alpha = 0.80$		$\alpha = 0.33$		$\alpha = 0.55$		$\alpha = 0.80$		
	Coh	Gau	Coh	Gau	Coh	Gau	Coh	Gau	Coh	Gau	Coh	Gau	Coh	Gau	Coh	Gau	Coh	Gau	
VaR_{0.95}																			
T = 75	0.171	0.631	0.142	0.587	0.076	0.485	0.269	0.572	0.293	0.708	0.383	0.850	0.388	0.797	0.397	0.775	0.365	0.674	
T = 250	0.099	0.663	0.086	0.608	0.061	0.498	0.158	0.574	0.176	0.769	0.240	0.947	0.261	0.897	0.266	0.872	0.249	0.747	
T = 2500	0.034	0.685	0.030	0.619	0.026	0.505	0.051	0.575	0.057	0.813	0.072	0.990	0.083	0.988	0.086	0.968	0.078	0.828	
TCE_{0.95}																			
T = 75	0.475	0.527	0.408	0.458	0.284	0.328	0.524	0.423	0.542	0.589	0.656	0.841	0.579	0.916	0.598	0.920	0.624	0.894	
T = 250	0.482	0.525	0.443	0.435	0.373	0.299	0.513	0.388	0.527	0.604	0.600	0.936	0.548	0.987	0.567	0.988	0.603	0.970	
T = 2500	0.495	0.526	0.480	0.418	0.461	0.281	0.503	0.368	0.509	0.604	0.534	0.985	0.513	1.000	0.524	1.000	0.541	0.992	
ES_{0.95}																			
T = 75	0.495	0.758	0.440	0.697	0.308	0.603	0.523	0.558	0.537	0.716	0.628	0.919	0.572	0.959	0.592	0.961	0.630	0.952	
T = 250	0.497	0.861	0.464	0.779	0.388	0.682	0.514	0.549	0.522	0.768	0.580	0.987	0.541	0.998	0.557	0.998	0.591	0.997	
T = 2500	0.499	0.964	0.489	0.863	0.466	0.763	0.504	0.533	0.506	0.801	0.525	1.000	0.511	1.000	0.518	1.000	0.533	1.000	
MVaR_{0.95}																			
T = 75	0.500	0.752	0.456	0.727	0.333	0.719	0.524	0.618	0.537	0.756	0.601	0.888	0.562	0.826	0.572	0.807	0.559	0.723	
T = 250	0.499	0.854	0.474	0.820	0.404	0.829	0.515	0.640	0.523	0.837	0.565	0.978	0.534	0.939	0.547	0.919	0.551	0.826	
T = 2500	0.500	0.962	0.492	0.910	0.470	0.928	0.505	0.639	0.505	0.900	0.522	0.997	0.509	1.000	0.514	0.999	0.516	0.931	
EVaR_{0.95}																			
T = 75	0.501	0.661	0.467	0.626	0.362	0.556	0.518	0.580	0.527	0.700	0.587	0.847	0.557	0.841	0.572	0.846	0.599	0.836	
T = 250	0.500	0.748	0.480	0.690	0.420	0.610	0.512	0.602	0.516	0.782	0.555	0.962	0.532	0.951	0.545	0.955	0.573	0.952	
T = 2500	0.500	0.891	0.494	0.788	0.477	0.682	0.504	0.597	0.502	0.862	0.518	1.000	0.509	1.000	0.514	1.000	0.526	1.000	

Table 5. Mean severity of risk underrating for Poi-, ZIP, and NB-INAR(1) DGPs, coherent (“Coh”) vs. approximate (“Gau”) forecasts, for different (α, T), $I = 2.4$ in case of overdispersion and forecast horizon $h = 1$, computed across all simulation runs and all levels of μ .

	Poi-INAR(1) DGP						ZIP-INAR(1) DGP						NB-INAR(1) DGP						
	$\alpha = 0.33$		$\alpha = 0.55$		$\alpha = 0.80$		$\alpha = 0.33$		$\alpha = 0.55$		$\alpha = 0.80$		$\alpha = 0.33$		$\alpha = 0.55$		$\alpha = 0.80$		
	Coh	Gau	Coh	Gau	Coh	Gau	Coh	Gau	Coh	Gau	Coh	Gau	Coh	Gau	Coh	Gau	Coh	Gau	
VaR_{0.95}																			
T = 75	-1.034	-0.582	-1.020	-0.538	-1.005	-0.424	-1.133	-0.777	-1.148	-0.926	-1.244	-1.066	-1.384	-1.283	-1.365	-1.240	-1.247	-0.985	
T = 250	-1.002	-0.444	-1.001	-0.416	-1.000	-0.332	-1.016	-0.599	-1.017	-0.734	-1.068	-0.915	-1.084	-0.984	-1.074	-0.939	-1.042	-0.737	
T = 2500	-1.000	-0.378	-1.000	-0.357	-1.000	-0.292	-1.000	-0.520	-1.000	-0.641	-1.004	-0.865	-1.000	-0.844	-1.000	-0.787	-1.000	-0.578	
TCE_{0.95}																			
T = 75	-0.402	-0.553	-0.386	-0.499	-0.304	-0.371	-0.622	-0.776	-0.681	-0.948	-0.934	-1.216	-1.339	-1.959	-1.427	-2.003	-1.431	-1.773	
T = 250	-0.224	-0.384	-0.219	-0.352	-0.190	-0.256	-0.336	-0.607	-0.376	-0.750	-0.555	-1.043	-0.757	-1.721	-0.810	-1.777	-0.868	-1.597	
T = 2500	-0.074	-0.294	-0.071	-0.272	-0.067	-0.195	-0.109	-0.542	-0.123	-0.663	-0.178	-0.979	-0.242	-1.674	-0.266	-1.731	-0.299	-1.539	
ES_{0.95}																			
T = 75	-0.381	-0.676	-0.347	-0.597	-0.289	-0.426	-0.619	-0.899	-0.676	-1.100	-0.894	-1.443	-1.349	-2.317	-1.433	-2.371	-1.427	-2.150	
T = 250	-0.212	-0.506	-0.203	-0.437	-0.190	-0.295	-0.335	-0.726	-0.369	-0.899	-0.521	-1.330	-0.755	-2.157	-0.804	-2.222	-0.861	-2.052	
T = 2500	-0.068	-0.418	-0.067	-0.350	-0.070	-0.226	-0.105	-0.663	-0.116	-0.801	-0.167	-1.315	-0.239	-2.134	-0.256	-2.197	-0.272	-2.029	
MVaR_{0.95}																			
T = 75	-0.350	-0.589	-0.307	-0.547	-0.229	-0.449	-0.576	-0.794	-0.614	-0.941	-0.738	-1.090	-0.945	-1.301	-0.925	-1.254	-0.772	-0.995	
T = 250	-0.193	-0.436	-0.177	-0.405	-0.146	-0.342	-0.312	-0.609	-0.332	-0.746	-0.426	-0.962	-0.525	-1.006	-0.510	-0.953	-0.452	-0.738	
T = 2500	-0.061	-0.355	-0.058	-0.331	-0.054	-0.289	-0.098	-0.531	-0.104	-0.650	-0.134	-0.942	-0.165	-0.911	-0.161	-0.836	-0.145	-0.587	
EVaR_{0.95}																			
T = 75	-0.323	-0.441	-0.284	-0.402	-0.202	-0.309	-0.505	-0.617	-0.514	-0.713	-0.561	-0.807	-0.809	-1.100	-0.816	-1.096	-0.745	-0.962	
T = 250	-0.178	-0.292	-0.160	-0.264	-0.124	-0.193	-0.275	-0.429	-0.277	-0.520	-0.311	-0.662	-0.449	-0.866	-0.451	-0.870	-0.437	-0.778	
T = 2500	-0.057	-0.200	-0.052	-0.181	-0.044	-0.128	-0.086	-0.340	-0.086	-0.416	-0.096	-0.630	-0.141	-0.800	-0.143	-0.806	-0.138	-0.717	

4.2. Further Autoregressive Count DGPs

In this section, we do not further stress the overdispersion phenomena considered in Section 4.1, but take a closer look at the autocorrelation structure. It shall become clear that for DGPs with higher-order dependence, the discrepancies between the coherent and approximate forecasts’ performance are even intensified compared to the INAR(1) case. Note that the INAR(1) model of Section 4.1 is not the only AR-type model for unbounded counts, also see Appendix A.1. Thus, it is important to check to what extent the performance results for coherent and approximate risk forecasting can be transferred to other types of AR-like count DGPs. More precisely, we consider higher-order Poi-INAR models,

alternative AR-like models from the Poi-INARCH family, and non-stationary extensions (with trend or seasonality) of the II-Poi-AR(1) model, see Appendix A.1 for details.

We start by extending the Poi-INAR(1) model to a second-order model, the Poi-INAR(2) model, to see whether an increased AR-order has an effect on the risk measures' performance. With exceptions for high autocorrelation and small sample sizes ($\alpha = 0.8$, $T = 75$), where the Poi-INAR(2) model's frequency for risk underrating is more close to 0.5 than in the first-order case (see Tables S5, S7 and S9 in the supplement, first block of columns), there is no noticeable difference in the performance of the coherent risk measures. In contrast, the approximate risk forecasts are clearly affected by the additional dependence parameter. We notice a further deterioration in performance, especially for a moderate to strong autocorrelation ($\alpha \geq 0.55$) and for the mean severity of underrating (see Tables S6, S8 and S10). More generally, for given μ value, the variation among the approximate forecast values clearly increases for the second-order INAR model, as illustrated by Figure 6 (second row compared to first row of boxplots).

The INARCH family constitutes a common alternative to the INAR family. The corresponding summarizing statistics are provided in Tables S5–S10 of the supplemental material (second and third block of columns), while boxplots of deviations are presented in the third and fourth row of Figure 6. For the coherent risk forecasts, the variation among the risk forecasts and, along with this, the severity of risk underrating slightly increase for the Poi-INARCH DGPs (somewhat stronger for the second-order than for the first-order case). At the same time, the frequencies of risk under- and overrating are rather balanced for both INARCH models. In contrast to their coherent counterparts, the approximate risk forecasts of the INARCH models show a strong increase in variation and, thus, also in the severity of risk underrating. We recognize a more heterogeneous pattern of variation among the deviations for the Poi-INARCH(1) DGP (both coherent and approximate) in the third row of Figure 6. As an interim conclusion, the findings of Section 4.1, regarding a much more frequent and severe risk underrating if using a Gaussian approximation, also hold true for the other types of stationary and AR-like count DGPs.

Finally, we investigate the performance of risk forecasting for another AR-like regression model, the II-Poi-AR(1) model, but now in the additional presence of deterministic trend or seasonality, i.e., we are concerned with non-stationary count DGPs. For approximation, we use a Gaussian regression model with ARMA innovations, including a linear trend and harmonic oscillation like the II-Poi-AR(1) DGP (an approximation with a SARIMA model having stochastic trend or seasonality did considerably worse). For different parameter values of γ_0 (intercept), γ_1 (trend), (γ_2, γ_3) (seasonality), α , and T , we simulated 1000 time series and computed the corresponding risk forecasts, see Tables S11 and S12 in the supplement for summarizing statistics. Despite having additional trend or seasonality, the performance of the coherent risk forecasts is not considerably worse if compared to the stationary DGPs: Although trend and seasonality (going along with additional parameters to be estimated) as well as an increase of the intercept cause more variation among the coherent forecast values, the underrating frequencies are rather balanced throughout (like for the INARCH models). The approximate risk forecasts, in contrast, underrate the true risk far more frequently, with ES and MVaR performing worst in this matter. While risk underrating happens more often and more severely with increasing trend γ_1 , an increasing seasonality (γ_2, γ_3) leads to more severe but less frequent underrating. Altogether, the mean severity of risk underrating is 2–6 times higher for the approximate risk forecasts than for the coherent ones.

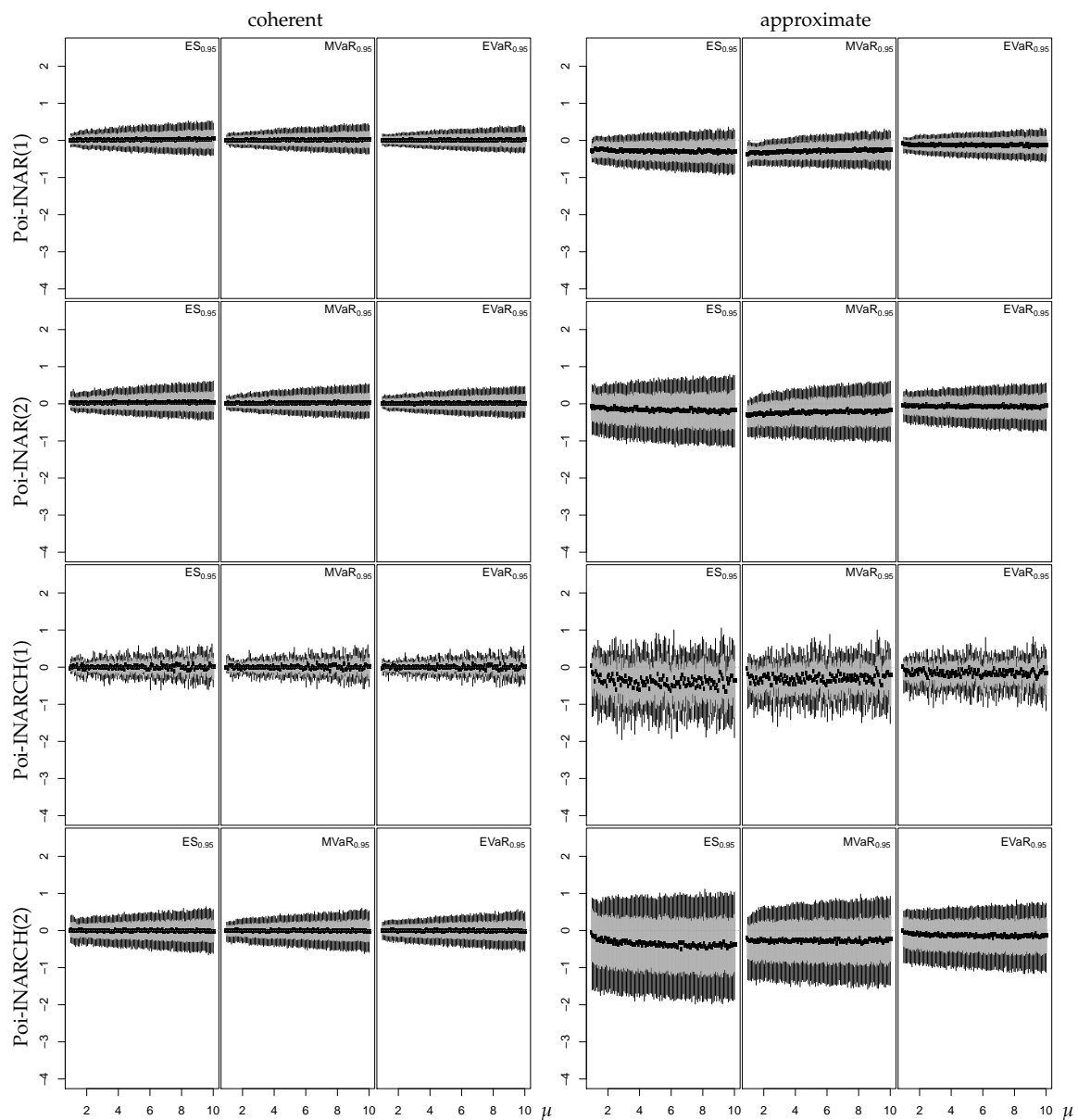


Figure 6. Boxplots of deviations for $ES_{0.95}$, $MVaR_{0.95}$, and $EVaR_{0.95}$. One-step-ahead forecasts based on the coherent models of the Poisson INAR(1), INAR(2), INARCH(1) and INARCH(2) DGP (left column) and their Gaussian approximations (right column) with $\alpha = 0.55$, $\alpha_2 = 0.45$ in case of of the second-order models, and sample size $T = 250$.

In summary, for all the alternative types of AR-like count DGP discussed in this section, the approximate risk forecasts perform considerably worse than the coherent ones, with even larger discrepancy than for the Poi-INAR(1) DGP of Section 4.1. While an increasing sample size T leads to more balanced (and less severe) deviations in the coherent case, such an effect cannot be generally observed if using a Gaussian approximation: In case of the Poi-INARCH models, for example, the underrating frequencies of the approximate risk forecasts often increase far beyond 0.5 with increasing T . As already observed in Section 4.1, the ES is most affected by the approximation error, while the approximate EVaR generally exhibits the lowest variation, though still clearly more than in the coherent case.

4.3. DGPs for Bounded Counts

In the previous sections, we discussed DGPs of unbounded counts, i.e., where the range consists of the full set of non-negative integers, \mathbb{N}_0 . In this section, we turn our attention to bounded counts having the range $\{0, \dots, n\}$, where the upper bound $n \in \mathbb{N}$

cannot be exceeded. As a consequence, also the corresponding risk measures can never exceed the upper bound $n \in \mathbb{N}$. We consider two types of DGP for bounded counts: the thinning-based BinAR(1) model as well as the regression-type BinARCH(1) model, both exhibiting an AR(1)-like ACF (see Appendix A.1 for details). Besides the illustrative graphs with boxplots shown in Figure 7 below, the full set of summarizing statistics is provided by Tables S13–S16 in the supplemental material.

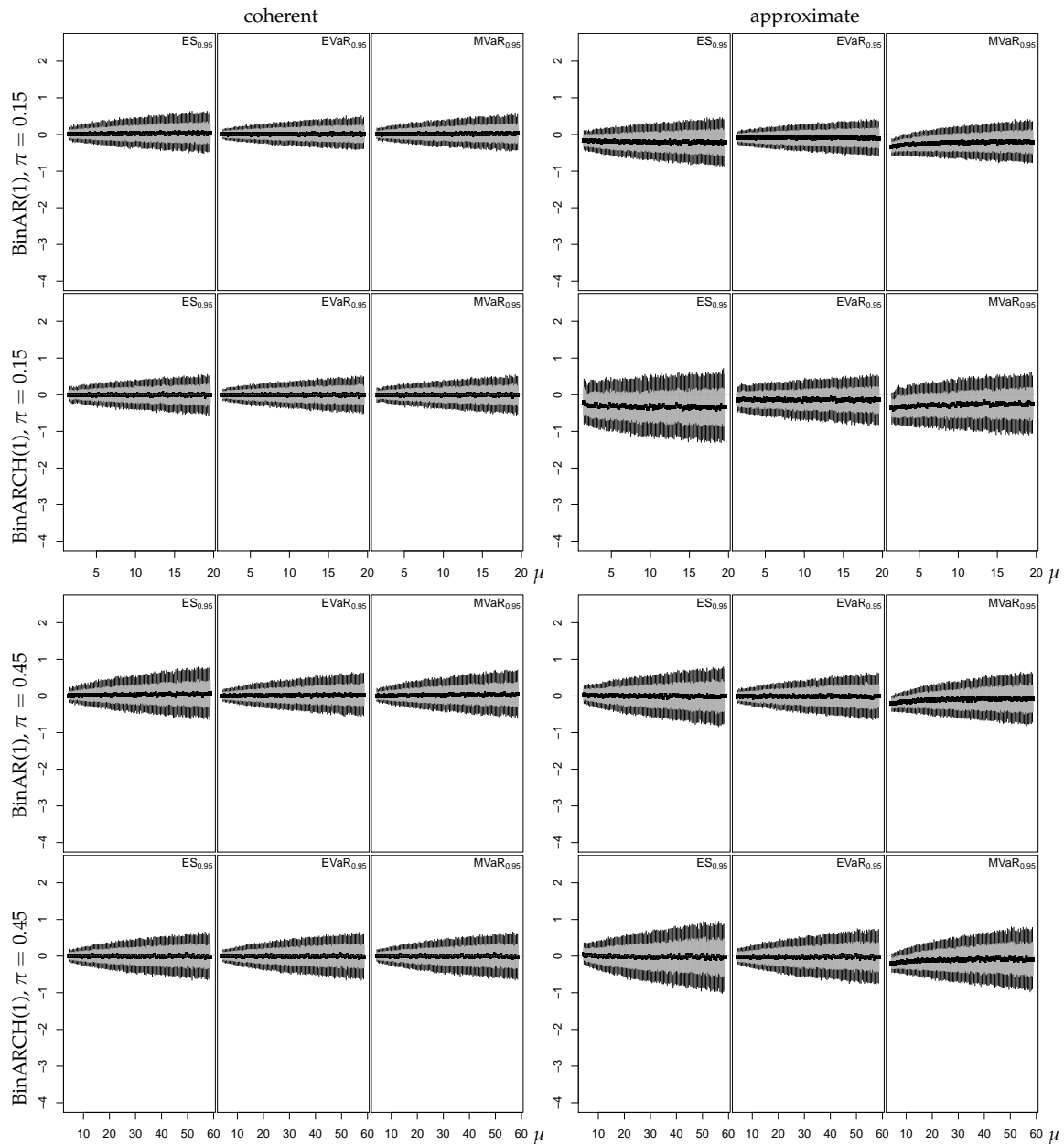


Figure 7. Boxplots of deviations for $ES_{0.95}$, $MVaR_{0.95}$, and $EVaR_{0.95}$. One-step-ahead forecasts based on the coherent models of the BinAR(1) and BinARCH(1) DGP (left column) and their Gaussian approximations (right column) with $\alpha = 0.55$, $\pi \in \{0.15, 0.45\}$, and sample size $T = 250$.

Let us start with the BinAR(1) DGP (see the first block of columns in Tables S13–S16). For a low “success probability” ($\pi = 0.15$, see Tables S13–S14), the coherent and approximate risk forecasts perform similar to the Poi-INAR(1) case, which is not surprising in view of the Poisson limit theorem. For $\pi = 0.45$ (Tables S15–S16), in contrast, the BinAR(1) process has a nearly symmetric and bell-shaped marginal distribution. While the proportion of underrating for the coherent risk forecasts is not affected by the rise in π , the underrating

becomes more severe. This differs from the case of approximate risk forecasts, where the deviations of ES and EVaR from their true risk values become more balanced (and less severe for large T) if $\pi = 0.45$, which might be explained by the de Moivre–Laplace theorem. It can also be recognized from the third row in Figure 7 that for $\pi = 0.45$, there is not much difference between the approximate and the coherent forecast approach. For $\pi = 0.15$ (first row), in contrast, the approximate forecasts' boxplots look clearly misplaced, with a particularly severe extent of underrating for the MVaR with small n .

Next, we turn to the BinARCH(1) model (second block of columns in Tables S13–S16). In analogy to the Poi-INARCH DGPs of Section 4.2, the coherent risk forecasts of the BinARCH(1) DGP show very balanced proportions of risk over- and underrating for all parametrizations (more balanced than in the BinAR(1) case). The approximate forecasts of the BinARCH(1) DGP, in contrast, behave quite similar to the BinAR(1) case: they tend to frequently underrate the true risk, especially for increasing T , and MVaR performs much worse than ES and EVaR. The mean severities are even considerably worse than in the BinAR(1) case (especially for $\pi = 0.15$), which can also be seen from the second and fourth row in Figure 7, showing much more variation among the approximate risk forecasts than in any other scenario of Figure 7. Altogether, we conclude that for both bounded-counts DGPs, the coherent risk forecasts always outperform the approximate ones, although the difference is less pronounced in the nearly symmetric situation with $\pi = 0.45$.

5. Application: Risk Forecasts for Transaction Counts Data

Risk analysis plays an important role at the financial markets. Measuring the value at risk or calculating the expected risk-adjusted rate of return is a typical example, but there are many other financial applications (see, e.g., Hull 2018; McNeil et al. 2015). Several financial examples have also been reported in articles on count time series analysis, such as Brännäs and Quoreshi (2010), who analyze time series of intra-day counts of transactions on the Stockholm stock exchange, or Jung and Tremayne (2011), who consider daily counts of so-called “iceberg orders” concerning shares being traded on the Frankfurt stock exchange. Our data examples to be discussed below are provided by the Deutsche Börse Group, which is an international exchange organisation, a market infrastructure provider, and the operator of the Frankfurt stock exchange. As part of its services, Deutsche Börse offers real-time risk alerts to inform traders “about unusual price and order book behaviour,” which are “triggered, when certain thresholds for changes in traded prices, bid-ask spreads and order book resilience are exceeded. Thresholds are calibrated on a daily basis so that messages are only generated when exceptional order book situations could occur.” (Retrieved 15 December 2020, from <https://www.deutsche-boerse.com/dbg-en/media/press-releases/Deutsche-B-rse-supports-traders-with-real-time-risk-alerts-for-most-liquid-Eurex-futures-683428>).

Inspired by such risk analyses, we applied our approaches for coherent risk forecasting to time series of daily counts of transactions (number of trades by settlement date) of structured products from on-market (e.g., Xetra) and off-market trading (over the counter, OTC), as offered by the Cascade-Turnoverdata of Deutsche Börse (Retrieved 15 December 2020, from <https://datashop.deutsche-boerse.com/reference-data>). The different types of one-step-ahead risk forecasts (again, we use the risk level $\rho = 0.95$ for convenience) provide threshold values on a daily basis, which refer to the maximal trading activity in 95% of all cases. For illustration, let us consider the two example time series shown in Figure 8. Both refer to factor long certificates with leverage, i.e., the daily return of the underlying asset is multiplied by a constant positive factor (a “bet on rising prices” in a sense); see Anic (2020) for detailed background on leverage certificates. The first data set comprises the $T_1 = 381$ counts for the trading days between February 2017 and July 2018 for model fitting, see Figure 8a, whereas the 23 counts from August 2018 are later used for out-of-sample forecasting. The second data set includes one additional year of data, i.e., the $T_2 = 636$ counts from February 2017 to July 2019 from Figure 8c are used for model fitting, and the 22 counts from August 2019 for forecasting.

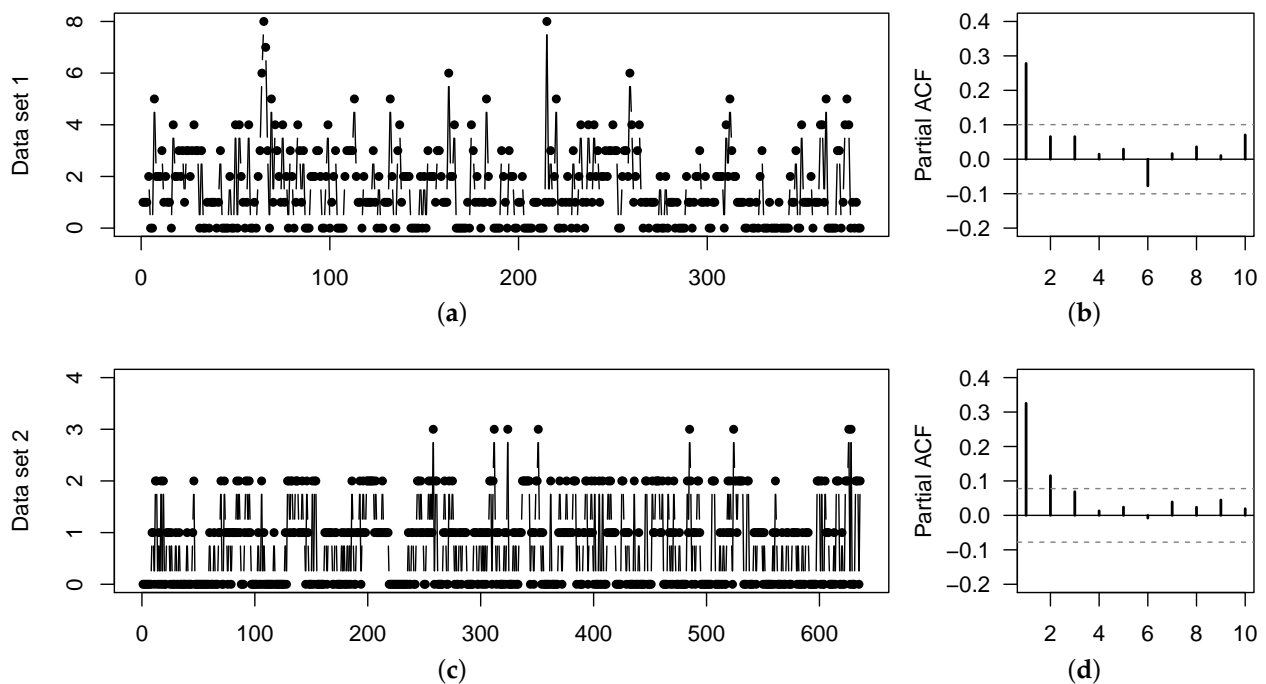


Figure 8. Two data sets of transaction counts: their time series plots in (a) and (c), respectively, and their sample PACFs in (b) and (d), respectively.

The two data sets do not only differ in their sizes, but also in their properties. The second data set consists of much lower counts than the first one (samples means $\hat{\mu}_1 \approx 1.493$ vs. $\hat{\mu}_2 \approx 0.719$), and also exhibits less dispersion (samples dispersion indexes $\hat{I}_1 \approx 1.518$ vs. $\hat{I}_2 \approx 0.913$). And while the first data set has a clearly AR(1)-like partial ACF (PACF), the second PACF indicates that an AR(2)-like model might be a better solution, see Figure 8b,d. Using the dispersion test in (2.14) of Weiß (2018) with level 5%, we conclude on significant overdispersion for data set 1, but we do not observe a significant violation of equidispersion for data set 2. Altogether, an NB-INAR(1) model appears to be a reasonable choice for data set 1, and Poi-INAR(1) or Poi-INAR(2) for data set 2. To check for possible effects of model misspecification on risk forecasting, we try out both candidate models for data set 2, and we additionally consider a Poi-INAR(1) model also for data set 1.

In Figure 9, we applied both types of candidate models to compute the one-step-ahead risk forecasts (plotted as lines) for the respective hold-out sample (shown by dots). Especially for data set 1, a rather strong effect of model choice can be recognized. This is plausible as the models differ in their dispersion behavior, and as the risk measures are computed from the upper tail of the forecast distribution. For data set 2, the models differ in the length of their memory. It can be seen that the risk forecasts of the second-order model sometimes decay later than their first-order counterparts. Certainly, we do not know the true DGP behind the data, but the aforementioned sample properties indicate that for data set 1 (2), the NB-INAR(1) (Poi-INAR(2)) model should be preferred. The corresponding risk forecasts are represented by the darker curves in Figure 9.

Before using these risk forecasts for decision making, it would be wise to account for estimation uncertainty; recall our analyses of Section 4. For this purpose, after having fitted the respective INAR model to the data sets, a corresponding parametric INAR bootstrap (see Jentsch and Weiß 2019) with $B = 10^4$ replications is done to account for the estimation uncertainty. That is, for each bootstrap replication, the model parameters are estimated anew and used to compute the risk forecasts. Thus, each point forecast value in Figure 9 is accompanied by B bootstrap replicates. These can be used, for example, to approximate the quantiles of the forecast’s sample distribution. In analogy to the lean boxplots used in Section 4, we computed the 10%-, 25%-, 50%-, 75%-, and 90%-quantiles from the bootstrap

replicates and plotted them as lines in Figure 10. The width of these quantile bands expresses the strength of estimation uncertainty, which is more pronounced for data set 1 where only $T_1 = 381$ instead of $T_2 = 636$ counts were available for model fitting. If, for example, the risk measures are applied as thresholds for generating risk alerts, then especially the observations at times $t = 12, 17$ for data set 2 appear to be exceptionally large, see Figure 10. This could give rise to inform traders about an unusual order book behavior.

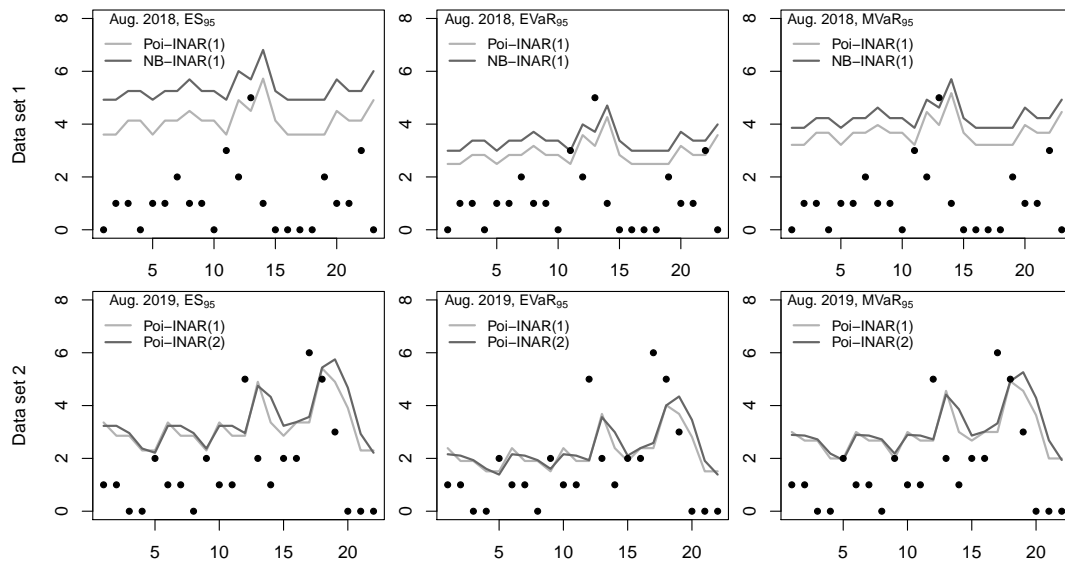


Figure 9. Two data sets of transaction counts: out-of-sample risk forecasts for different candidate models.

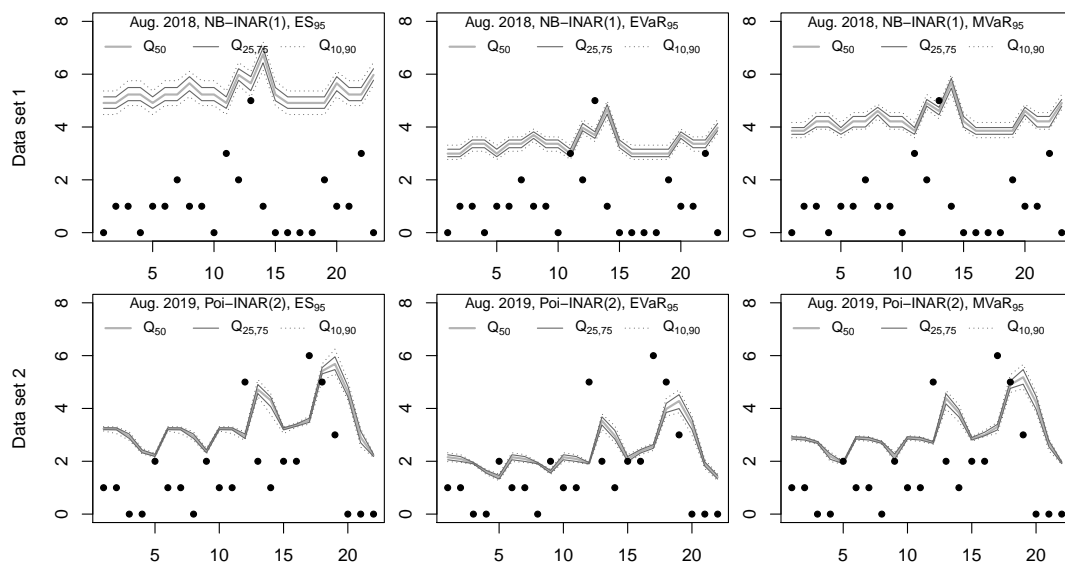


Figure 10. Two data sets of transaction counts: bootstrapped quantiles (“ Q_γ ” refers to γ -quantile) of out-of-sample risk forecasts for given model.

6. Conclusions

We considered five types of risk measures for evaluating a possible prospective “loss” related to a count process. Besides the widely used VaR, TCE, and ES, the more recent EVaR based on expectiles was also included in our investigations, as well as the newly proposed MVaR based on mid-quantiles. The real-valued ES, EVaR, and MVaR are less affected by discreteness, and are thus advantageous for estimation and forecasting. For our comprehensive simulation study, we considered a broad variety of count DGPs, covering

the most important features of count time series as they may occur in real applications. It turned out that risk forecasting relying on a Gaussian approximation to the count DGP causes rather misleading risk assessments, with a strong tendency to risk underrating. By contrast, while coherent risk forecasting suffers from estimation uncertainty, too, and is also affected by count phenomena such as overdispersion or zero inflation, we observed a rather balanced risk performance throughout. In our real applications, we accounted for the effect of estimation uncertainty by using parametric bootstrap approaches for count time series.

Supplementary Materials: The following are available online at <https://www.hsu-hh.de/mathstat/en/research/projects/forecastingrisk> (accessed on 26 March 2021), S1: Figures with Boxplots for Section 4; S2: Tables with Summarizing Statistics for Section 4.

Author Contributions: Conceptualization (all authors); Data curation (A.H., C.H.W.); Funding acquisition (A.H., C.H.W.); Methodology (all authors); Software (A.H.); Supervision (C.H.W., G.F.); Writing—original draft preparation (A.H., C.H.W., G.F.); Writing—review and editing (C.W., G.F., L.C.A., R.G.). All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Projektnummer 394832307.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to confidentiality reasons.

Acknowledgments: The authors thank all referees for their useful comments on an earlier draft of this article. The authors are grateful to Oliver Krüger (Helmholtz-Zentrum Geesthacht) for his support regarding the storm surges data discussed in Section 2.3. The transaction counts data of Section 5 were kindly made available to the authors by the Deutsche Börse AG.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ACF	autocorrelation function
AR	autoregressive
ARMA	autoregressive moving-average
Bin	binomial
BinAR	binomial autoregressive
BinARCH	binomial autoregressive conditional heteroskedastic
CDF	cumulative distribution function
DGP	data-generating process
ES	expected shortfall
EVaR	expectile-value at risk
INAR	integer-valued autoregressive
INARCH	integer-valued autoregressive conditional heteroskedastic
ll-Poi-AR	log-linear Poisson autoregressive
MVaR	mid-value at risk
NB	negative binomial
PACF	partial autocorrelation function
PMF	probability mass function
Poi	Poisson
SARIMA	seasonal autoregressive integrated moving-average
TCE	tail conditional expectation
VaR	value at risk
ZIP	zero-inflated Poisson

Appendix A. Summary of Considered Count DGPs

In what follows, we summarize those models that were used as a DGP in our numerical and simulation studies; see Table 3 for the considered parametrizations. 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 (2018).

Table A1. Relevant count distributions and their PMF.

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

Appendix A.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 (Bin) thinning: For $\alpha \in (0, 1)$ and a count r. v. X , it is defined by requiring $\alpha \circ X | X \sim \text{Bin}(X, \alpha)$, see Table A1. The following models assume that all thinnings are executed independently of each other, of the i. i. d. count innovations (ϵ_t), and of past observations.

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

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

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

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

It is referred to as Poi-, NB-, or ZIP-INAR(1) model, respectively, if ϵ_t follows a Poisson, negative binomial, or zero-inflated Poisson distribution (see Table A1).

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

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

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

BinAR(1) model for bounded range $\{0, \dots, n\}$ 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 $\gamma := \beta + \alpha$. The BinAR(1) model recursion is

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

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

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

Table A2. Forecast distributions $p(x|x_T, x_{T-1}) = p(X_{T+1} = x | X_T = x_T, X_{T-1} = x_{T-1}, \dots)$ of relevant count time series models.

INAR(1) model	$p(x x_T) = \sum_{s=0}^{\min\{x, x_T\}} \binom{x_T}{s} \alpha^s (1-\alpha)^{x_T-s} \cdot P(\epsilon_t = x-s)$	for $x \in \mathbb{N}_0$.
INAR(2) model	$p(x x_T, x_{T-1}) = \sum_{j_1=0}^{\min\{x, x_T\}} \sum_{j_2=0}^{\min\{x-j_1, x_{T-1}\}} \binom{x_T}{j_1} \alpha_1^{j_1} (1-\alpha_1)^{x_T-j_1} \cdot \binom{x_{T-1}}{j_2} \alpha_2^{j_2} (1-\alpha_2)^{x_{T-1}-j_2} \cdot P(\epsilon_t = x-j_1-j_2)$	for $x \in \mathbb{N}_0$.
BinAR(1) model	$p(x x_T) = \sum_{m=\max\{0, x+x_T-n\}}^{\min\{x, x_T\}} \binom{x_T}{m} \binom{n-x_T}{x-m} \gamma^m (1-\gamma)^{x_T-m} \beta^{x-m} (1-\beta)^{n-x_T+m-x}$	for $0 \leq x \leq n$.
Poi-regression	$p(x m_{T+1}) = \exp(-m_{T+1}) \frac{m_{T+1}^x}{x!}$ for $x \in \mathbb{N}_0$, where $m_{T+1} = E[X_{T+1} x_T, x_{T-1}, \dots]$ satisfies	<ul style="list-style-type: none"> • $m_{T+1} = \beta + \alpha x_T$ for Poi-INARCH(1) model; • $m_{T+1} = \beta + \alpha_1 x_T + \alpha_2 x_{T-1}$ for Poi-INARCH(2) model; • $\ln m_{T+1} = \ln \mu_{T+1} + \alpha_1 (\ln(x_T + 1) - \ln(\mu_T + 1))$ for ll-Poi-AR(1) model, with $\ln \mu_t = \gamma_0 + \gamma_1 t + \gamma_2 \cos(\omega t) + \gamma_3 \sin(\omega t)$.
Bin-regression	$p(x p_{T+1}) = \binom{n}{x} p_{T+1}^x (1-p_{T+1})^{n-x}$ for $0 \leq x \leq n$, where $p_{T+1} = \frac{1}{n} E[X_{T+1} x_T, x_{T-1}, \dots]$ satisfies $P_{T+1} = \beta + \alpha \frac{x_T}{n}$	for BinARCH(1) model.

Appendix A.2. Regression Models

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

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

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

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 $X_t | X_{t-1}, \dots \sim \text{Poi}(\beta + \alpha_1 X_{t-1} + \alpha_2 X_{t-2})$ 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 $X_t | X_{t-1}, \dots \sim \text{Bin}(n, \beta + \alpha \frac{X_{t-1}}{n})$ with $\beta, \beta + \alpha \in (0, 1)$ and transition probabilities

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

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

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

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

References

- Acerbi, Carlo, and Dirk Tasche. 2002. On the coherence of expected shortfall. *Journal of Banking & Finance* 26: 1487–503.
- Alwan, Layth C., and Christian H. Weiß. 2017. INAR implementation of newsvendor model for serially dependent demand counts. *International Journal of Production Research* 55: 1085–99. [\[CrossRef\]](#)
- Anic, Vladimir. 2020. Constant leverage certificates: Dynamics, performance, and risk-return characteristics. *Quantitative Finance and Economics* 4: 693–724. [\[CrossRef\]](#)
- Artzner, Philippe, Freddy Delbaen, Jean-Marc Eber, and David Heath. 1999. Coherent measures of risk. *Mathematical Finance* 9: 203–28. [\[CrossRef\]](#)
- Bellini, Fabio, and Elena di Bernardino. 2017. Risk management with expectiles. *European Journal of Finance* 23: 487–506. [\[CrossRef\]](#)
- Brännäs, Kurt, and A. M. M. Shahiduzzaman Quoreshi. 2010. Integer-valued moving average modelling of the number of transactions in stocks. *Applied Financial Economics* 20: 1429–40. [\[CrossRef\]](#)
- Chan, Stephen, and Saralees Nadarajah. 2019. Risk: An R package for financial risk measures. *Computational Economics* 53: 1337–51. [\[CrossRef\]](#)
- Cont, Rama, and Peter Tankov. 2004. *Financial Modeling with Jump Processes*. Boca Raton: Chapman & Hall/CRC.
- Emmer, Susanne, Marie Kratz, and Dirk Tasche. 2015. What is the best risk measure in practice? A comparison of standard measures. *Journal of Risk* 18: 31–60. [\[CrossRef\]](#)
- Freeland, R. Keith, and Brendan P. M. McCabe. 2004. Forecasting discrete valued low count time series. *International Journal of Forecasting* 20: 427–34. [\[CrossRef\]](#)
- Göb, Rainer. 2011. Estimating value at risk and conditional value at risk for count variables. *Quality and Reliability Engineering International* 27: 659–72. [\[CrossRef\]](#)
- Homburg, Annika. 2020. Criteria for evaluating approximations of count distributions. *Communications in Statistics—Simulation and Computation* 49: 3152–70. [\[CrossRef\]](#)
- Homburg, Annika, Christian H. Weiß, Layth C. Alwan, Gabriel Frahm, and Rainer Göb. 2019. Evaluating approximate point forecasting of count processes. *Econometrics* 7: 30. [\[CrossRef\]](#)
- Homburg, Annika, Christian H. Weiß, Layth C. Alwan, Gabriel Frahm, and Rainer Göb. 2020. A performance analysis of prediction intervals for count time series. *Journal of Forecasting*. [\[CrossRef\]](#)
- Hull, John C. 2018. *Risk Management and Financial Institutions*, 5th ed. Hoboken: John Wiley & Sons, Inc.
- Jensen, Jürgen, and Sylvain H. Müller-Navarra. 2008. Storm surges on the German coast. *Die Küste* 74: 92–124.
- Jentsch, Carsten, and Anne Leucht. 2016. Bootstrapping sample quantiles of discrete data. *Annals of the Institute of Statistical Mathematics* 68: 491–539. [\[CrossRef\]](#)
- Jentsch, Carsten, and Christian H. Weiß. 2019. Bootstrapping INAR models. *Bernoulli* 25: 2359–408. [\[CrossRef\]](#)
- Jung, Robert C., and Andrew R. Tremayne. 2011. Useful models for time series of counts or simply wrong ones? *ASTA Advances in Statistical Analysis* 95: 59–91. [\[CrossRef\]](#)
- Kerkhof, Jeroen, Bertrand Melenberg, and Hans Schumacher. 2010. Model risk and capital reserves. *Journal of Banking & Finance* 34: 267–79.
- Klüppelberg, Claudia, Daniel Straub, and Isabell M. Welpel, eds. 2014. *Risk—A Multidisciplinary Introduction*. Cham: Springer International Publishing Switzerland.
- Ma, Yanyuan, Marc G. Genton, and Emanuel Parzen. 2011. Asymptotic properties of sample quantiles of discrete distributions. *Annals of the Institute of Statistical Mathematics* 63: 227–43. [\[CrossRef\]](#)
- McNeil, Alexander J., Rüdiger Frey, and Paul Embrechts. 2015. *Quantitative Risk Management: Concepts, Techniques, and Tools*, 2nd ed.; Princeton: Princeton University Press.
- Newey, Whitney K., and James L. Powell. 1987. Asymmetric least squares estimation and testing. *Econometrica* 55: 819–47. [\[CrossRef\]](#)
- Parzen, Emanuel. 1997. Concrete statistics. In *Statistics in Quality*. Edited by Subir Ghosh, William R. Schucany and William B. Smith. New York: Marcel Dekker, pp. 309–32.
- Rahardja, Dewi. 2020. Statistical methodological review for time-series data. *Journal of Statistics and Management Systems* 23: 1445–61. [\[CrossRef\]](#)

- Rockafellar, R. Tyrrell, and Stanislav Uryasev. 2002. Conditional value-at-risk for general loss distributions. *Journal of Banking & Finance* 26: 1443–71.
- Ruszczynski, Andrzej. 2010. Risk-averse dynamic programming for Markov decision processes. *Mathematical Programming* 125: 235–61. [[CrossRef](#)]
- Stone, Bernell K. 1973. A general class of three-parameter risk measures. *Journal of Finance* 28: 675–85. [[CrossRef](#)]
- Weiß, Christian H. 2018. *An Introduction to Discrete-Valued Time Series*. Chichester: John Wiley & Sons, Inc.
- Weiß, Christian H., and Murat C. Testik. 2019. Risk-based metrics for performance evaluation of control charts. *Quality and Reliability Engineering International* 35: 280–91. [[CrossRef](#)]