To test the null hypothesis of a Poisson marginal distribution, test statistics based on the Stein–Chen identity are proposed. For a wide class of Poisson count time series, the asymptotic distribution of different types of Stein–Chen statistics is derived, also if multiple statistics are jointly applied. The performance of the tests is analyzed with simulations, as well as the question which Stein–Chen functions should be used for which alternative. Illustrative data examples are presented, and possible extensions of the novel Stein–Chen approach are discussed as well.

**KEYWORDS**

bivariate Poisson distribution, bootstrap, count time series, diagnostic tests, INARMA models, Stein–Chen identity

**1 | INTRODUCTION**

Stein (1972) introduced an approach for bounding the error if approximating distributions, nowadays referred to as the “Stein’s method.” In the special case of the Poisson approximation of a count distribution (i.e., a distribution on the set $\mathbb{N}_0 = \{0, 1, \ldots \}$), which was developed by Chen (1975), it is referred to as the “Stein–Chen method” (Erhardsson, 2005; Stein, 1986). The Stein–Chen characterization of a Poisson-distributed count random variable $X$ with mean $\mu > 0$, $X \sim \text{Poi}(\mu)$, uses expectations of functions $f : \mathbb{N}_0 \to \mathbb{R}$ defined on the nonnegative integers. More precisely, the (univariate) Stein–Chen identity states that

$$E[X \cdot f(X)] = \mu \cdot E[f(X + 1)] \quad (1)$$
holds for any bounded \( f \), if and only if \( X \sim \text{Poi}(\mu) \). Note that the necessity part of the Stein–Chen identity holds for any function \( f : \mathbb{N}_0 \rightarrow \mathbb{R} \), provided that the involved moments exist.

Besides its original application within Poisson approximations, Weiß and Aleksandrov (2020) demonstrated that Equation (1) also constitutes a useful tool for computing diverse Poisson moments. In the present work, however, our aim is different: we use the Stein–Chen identity (1) to derive novel goodness-of-fit (GoF) tests for the Poisson distribution; see Beltrán-Beltrán and O’Reilly (2019) and Gürtler and Henze (2000) for references on Poi-GoF tests. Generally, there are two strategies for defining a GoF test statistic: one can try to capture (almost) the whole distribution, or one can focus on specific characteristic properties thereof. The first group comprises, for example, the famous Pearson statistic, while the second group includes the widely used Fisher’s index of dispersion (Beltrán-Beltrán & O’Reilly, 2019; Gürtler & Henze, 2000; Kyriakoussis, Li, & Papadopoulos, 1998). Tests from the first group offer the potential of being powerful to nearly any violation of the null hypothesis, but they are often difficult to implement in practice, especially if not being concerned with independent and identically distributed (i.i.d.) but with time series data (Weiß, 2018b). The tests from the second group, by contrast, often have feasible asymptotic distributions also for the time series case. They are sensitive only regarding selected alternative scenarios (e.g., the dispersion test can detect alternatives going along with changes in the dispersion structure), but then, they often exhibit an excellent power.

The Stein–Chen characterization (1) of the Pois-distribution might be used to construct a GoF test belonging to the first group, as it was done by Betsch, Ebner, and Nestmann (2020) for the i.i.d. case, but this test requires a bootstrap implementation. In this article, by contrast, we use the Stein–Chen identity to develop GoF tests from the second group, but with great flexibility regarding the covered alternatives, and with a unique approach for computing the test statistics’ asymptotic distribution (also comprising the time series case). The outstanding feature of our proposed Stein–Chen tests is the fact that the function \( f \) in (1), that is, the weighting scheme used for the expectations, is chosen flexibly with respect to the relevant alternative scenario. Furthermore, we apply our novel GoF tests not only to i.i.d. counts, but also to a rather wide class of time series models having a Poisson marginal distribution. The relevant types of data generating process (DGP) have the property that any pair \((X_t, X_{t-k})\) with time lag \( k \in \mathbb{N} = \{1, 2, \ldots \} \) follows a bivariate Poisson (BPoi) distribution (Johnson, Kotz, & Balakrishnan, 1997, chapter 37), and corresponding models can be found within the Poi-INARMA family (Poisson INTeger-valued AutoRegressive Moving Average), see Section 2 for details. In Section 3, these pairwise BPoi-distributions, together with the bivariate Stein–Chen identities of Weiß and Aleksandrov (2020), are used to derive the asymptotic distribution of our novel GoF test statistics. The performance of the resulting Stein–Chen GoF tests is investigated with simulations for diverse types of DGPs and alternative scenarios, see Section 5. In Section 6, we consider the joint application of different Stein–Chen tests, which might be used for further improving the power performance. The proposed approaches are also illustrated by several real-data examples in Sections 3 and 6. Finally, Section 7 concludes the article and outlines issues for future research.

2 | ON POISSON INARMA PROCESSES

INARMA models have been proposed as integer-valued counterparts to the ordinary ARMA models for real-valued time series. Their basic idea is to replace the multiplication in the ARMA’s model recursion by an appropriate type of thinning operation such that the integer nature of the counts is preserved (Weiß, 2018a). For the Poi-INARMA models considered here, one uses the
operation of binomial thinning “o” (Steutel & van Harn, 1979): \(a \circ X|X \sim \text{Bin}(X, a)\) with thinning parameter \(a \in [0; 1)\). Then, the INAR(p) model with \(p \in \mathbb{N}\) (Alzaid & Al-Osh, 1990) is defined by the recursion

\[
X_t = \alpha_1 \circ_t X_{t-1} + \ldots + \alpha_p \circ_p X_{t-p} + \epsilon_t \quad \text{with} \quad \alpha_* := \sum_{j=1}^{p} \alpha_j < 1,
\]

where the innovations \((\epsilon_t)_{t \in \mathbb{Z}}\) are i.i.d. count random variables. We added the subscript “t” to the \(p\) thinnings in (2) to express that these are applied at time \(t\) for generating the \(t\)th observation, \(X_t\). The thinnings \(\alpha_1 \circ_t, \ldots, \alpha_p \circ_t\) at time \(t\) are executed independently of each other, and independent of the innovations \((\epsilon_t)_{t \in \mathbb{Z}}\). But thinnings corresponding to different times might be mutually dependent: in the model proposed by Alzaid and Al-Osh (1990), the conditional distribution of \((\alpha_1 \circ_{t+1} X_t, \ldots, \alpha_p \circ_{t+p} X_t)\) given \(X_t\) (these are the thinnings of \(X_t\) used for the observations at times \(t + 1, \ldots, t + p\) is a multinomial distribution. Then, Poi\((\mu (1 - \alpha_*))\)-innovations imply Poi\((\mu)\)-observations, where \(\mu > 0\). This specific choice is referred to as the Poi-INAR(p) model. For the Poi-INAR(1) and Poi-INAR(2) process, it is known that we do not only have a Poisson marginal distribution, but also bivariate Poisson distributions for pairs \((X_t, X_{t-k})\) with any time lag \(k \in \mathbb{N}\) (Weiβ, 2018b):

\[
(X_t, X_{t-k}) \sim \text{BPoi}_{\mu, \rho(k)} := \text{BPoi}(\rho(k) \mu; (1 - \rho(k)) \mu, (1 - \rho(k)) \mu).
\]

Here, \(\text{BPoi}(\lambda_0; \lambda_1, \lambda_2)\) refers to the bivariate Poisson distribution with parameters \(\lambda_0 \geq 0\) and \(\lambda_1, \lambda_2 > 0\), which is defined as the distribution of \((Z_0 + Z_1, Z_0 + Z_2)\) with independent Poisson variates \(Z_i \sim \text{Poi}(\lambda_i)\) for \(i = 0, 1, 2\) (Johnson et al., 1997, chapter 37). The autocorrelation function (ACF) involved in (3) equals \(\rho(k) = \alpha_1^k\) if \(p = 1\), and if \(p = 2\), we get (Alzaid & Al-Osh, 1988)

\[
\rho(1) = \alpha_1, \quad \rho(k) = \alpha_1 \rho(k - 1) + \alpha_2 \rho(k - 2) \quad \text{for} \quad k \geq 2.
\]

Note that the type of Poi-INAR(p) model proposed by Du and Li (1991), where \((\alpha_1 \circ_{t+1} X_t, \ldots, \alpha_p \circ_{t+p} X_t)\) given \(X_t\) are conditionally independent (and not multinomial as in the model by Alzaid & Al-Osh, 1990), does not satisfy the “BPoi\(_{\mu, \rho(k)}\)"-property” (3) for \(p \geq 2\); in fact, it does not even have a Poisson marginal distribution.

A comprehensive analysis of Poi-INMA(q) models is provided by Weiβ (2008). They rely on the model recursion

\[
X_t = \beta_0 \circ_t \epsilon_t + \beta_1 \circ_t \epsilon_{t-1} + \ldots + \beta_q \circ_t \epsilon_{t-q}, \quad q \in \mathbb{N},
\]

where the \(q + 1\) thinnings at time \(t\) are performed independently of each other (commonly, one sets \(\beta_0 = 1\) by using the convention \(1 \circ X = X\)). With the notation \(\beta_* := \sum_{j=0}^{q} \beta_j\), it holds that Poi\((\mu/\beta_*)\)-innovations lead to Poi\((\mu)\)-observations. The pairwise bivariate distributions and the ACF further depend on the actual choice for the conditional distribution of \((\beta_0 \circ_t \epsilon_t, \beta_1 \circ_{t+1} \epsilon_t, \ldots, \beta_q \circ_{t+q} \epsilon_t)\) given \(\epsilon_t\), see Weiβ (2008) for details. But in any case, the relation (3) still holds true.

Hence, several types of Poi-INARMA processes have the property that pairs \((X_t, X_{t-k})\) are bivariately Poisson distributed. Their distribution \(\text{BPoi}_{\mu, \rho(k)}\) is even symmetric, because the \(\lambda_1\)- and \(\lambda_2\)-parameter agree with each other due to stationarity. This symmetry is later used to simplify
the asymptotic derivations. In particular, it allows to simplify the bivariate Stein–Chen identities in theorem 4.1 of Weiß and Aleksandrov (2020), leading to the following necessary conditions for \((X_t, X_{t+k})\) being \(\text{BPoi}_{\mu, \rho(k)}\) distributed.

**Proposition 1.** If \((X_t, X_{t+k}) \sim \text{BPoi}_{\mu, \rho(k)}\), then for any functions \(f, g : \mathbb{N}_0 \rightarrow \mathbb{R}\), we have the identities

\[
E[X_t \cdot f(X_t) g(X_{t+k})] = (1 - \rho(k)) \mu E[f(X_t) g(X_{t+k} + 1)] + \rho(k) \mu E[f(X_t + 1) g(X_{t+k} + 1)]
\]

and

\[
E[X_t f(X_t) g(X_{t+k})] = \rho(k)^2 \mu^2 E[f(X_t + 2) g(X_{t+k} + 2)] + ((1 - \rho(k))^2 \mu^2 + \rho(k) \mu) E[f(X_t + 1) g(X_{t+k} + 1)] + \rho(k) (1 - \rho(k)) \mu^2 E[f(X_t + 2) g(X_{t+k} + 1)] + \rho(k) (1 - \rho(k)) \mu^2 E[f(X_t + 1) g(X_{t+k} + 2)],
\]

provided that the involved expectations exist.

### 3 | **STEIN–CHEN GOF TESTS**

One of the most well-known characterizations of the Poisson distribution is its equidispersion property, that is, its variance is equal to the mean, whereas common non-Poisson distributions often have a variance being larger (overdispersion) or smaller (underdispersion) than the mean. Thus, to test the null hypothesis of a Poisson distribution, one may focus on testing for equidispersion. This is the idea behind the widely used Fisher’s index of dispersion (Beltrán-Beltrán & O’Reilly, 2019; Gürtler & Henze, 2000; Kyriakoussis et al., 1998), which has also been used for some types of Poi-INARMA processes, see Aleksandrov and Weiß (2020a) and the references therein. The dispersion index \(\hat{I}\) is the quotient of the empirical variance to the mean, which can be rewritten as

\[
\hat{I} = \frac{1}{n} \sum_{t=1}^{n} X_t^2 - \frac{1}{n} \sum_{t=1}^{n} X_t = 1 + \frac{1}{n} \sum_{t=1}^{n} X_t f(X_t) - \frac{1}{n} \sum_{t=1}^{n} X_t f(X_t + 1)
\]

with \(f(x) = x - 1\). The equidispersion property itself can be stated as

\[
E[X \cdot f(X)] = \mu \cdot E[f(X + 1)] \quad \text{with} \quad f(x) = x - 1.
\]

Equations (8) and (9) show that \(\hat{I}\) and the equidispersion property correspond to the Stein–Chen identity (1) for the specific choice \(f(x) = x - 1\). While the dispersion index (and thus the choice \(f(x) = x - 1\) for (1)) is known to perform very well for uncovering overdispersion (this is confirmed later in Section 5), there might be better solutions for different types of alternative (i.e., other choices of \(f\) that lead to an improved power).
3.1 Stein–Chen statistics

In what follows, we utilize the general Stein–Chen identity (1) to derive GoF tests for detecting deviations from a Poisson distribution, given a count time series $X_1, \ldots, X_n$ of length $n \in \mathbb{N}$. Therefore, we replace the theoretical moments involved in (1) by the following sample moments:

$$
\overline{X} f(X) = \frac{1}{n} \sum_{t=1}^{n} X_t f(X_t), \quad \overline{X} = \frac{1}{n} \sum_{t=1}^{n} X_t, \quad \overline{f(X+1)} = \frac{1}{n} \sum_{t=1}^{n} f(X_t + 1).
$$

At this point, the main idea behind the proposed Stein–Chen approach becomes clear: the function $f$ serves as some kind of weighting function, which gives a large weight to, for example, low counts or a certain segment of counts. Several example choices for $f$ are later discussed in Section 5; the dispersion index (8) uses the linear function $f(x) = x - 1$.

There are many different ways of defining test statistics via (1). For each considered statistic, we shall derive a normal approximation based on asymptotic considerations (under the Poisson null hypothesis). But since it is not clear in advance for which type of statistic such a normal approximation will perform best, we include the following candidates into our investigations:

$$
\hat{T}_1 = \frac{\overline{X} \cdot f(X+1)}{\overline{X} f(X)}, \quad \hat{T}_2 = \frac{\overline{X f(X)}}{\overline{X} \cdot f(X+1)} = \frac{1}{\hat{T}_1},
\hat{T}_3 = \frac{\overline{X f(X)} - \overline{X} \cdot f(X+1)}{f(X+1) - \overline{X}}.
$$

(10)

If the value of the test statistics $\hat{T}_1$ and $\hat{T}_2$ deviates significantly from 1, then the DGP is supposed not to be Poisson distributed. In the case of $\hat{T}_3$ and $\hat{T}_4$, we look for departures from 0. For the ratio statistics $\hat{T}_1, \hat{T}_2, \hat{T}_4$, possible issues regarding a division by zero have to be kept in mind. Recall that the test statistic $\hat{T}_4$ is equivalent (except a constant) to the ordinary dispersion index (8) if choosing $f(x) = x - 1$ (linear weighting scheme), whereas $\hat{T}_2$ then corresponds to the modified dispersion statistic in Kyriakoussis et al. (1998). Since the dispersion test is one of the most widely used Poi-GoF tests in practice (also for Poi-INARMA processes, see Aleksandrov & Weiß, 2020a), it constitutes the natural benchmark test in this work.

3.2 Univariate asymptotics

We start our asymptotic investigations with a central limit theorem (CLT) for the three-dimensional vector-valued process $(Y_t)_{t \in \mathbb{N}}$ given by

$$
Y_t = \begin{pmatrix} Y_{t,1} \\ Y_{t,2} \\ Y_{t,3} \end{pmatrix} := \begin{pmatrix} X_t - \mu \\ X_t f(X_t) - \mu \phi_1 \\ f(X_t + 1) - \phi_1 \end{pmatrix},
$$

(11)

where $\mu = E[X_t]$ and $\phi_i := E[f(X_t + i)]$. Under the null $X_t \sim \text{Poi}(\mu)$, because of the Stein–Chen identity (1), we have $E[Y_t] = 0$. Furthermore, we assume that the DGP (under the null) is such that the CLT of Ibragimov (1962) can be applied to $n^{-1/2} \sum_{t=1}^{n} Y_t$. This is the case if the process
is $\alpha$-mixing with exponentially decreasing weights as satisfied by the above Poi-INAR(p) process (Doukhan, Fokianos, & Li, 2012) (the Poi-INMA(q) process is even q-dependent), but the requirements in Ibragimov (1962) are even slightly weaker. The following lemma summarizes the resulting asymptotics.

**Lemma 1.** Let the DGP $(X_t)_{t \in \mathbb{Z}}$ satisfy $(X_t, X_{t-k}) \sim \text{BPoi}_{\mu, \rho(k)}$ as well as the aforementioned mixing conditions. Denote $\phi_{ij}(k) := E[f(X_t + i) f(X_{t+k} + j)]$. Then, $n^{-1/2} \sum_{i=1}^n Y_i$ according to (11) is asymptotically normally distributed with mean $0$ and covariance matrix $\Sigma = (\sigma_{ij})$, where

\[
\begin{align*}
\sigma_{11} &= \mu \left( 1 + 2 \sum_{k=1}^{\infty} \rho(k) \right), \\
\sigma_{12} &= (\phi_1 + \mu (\phi_2 - \phi_1)) \sigma_{11}, \quad \sigma_{13} = (\phi_2 - \phi_1) \sigma_{11}, \\
\sigma_{22} &= \mu \phi_{11}(0) + \mu^2 \phi_{22}(0) - \mu^2 \phi_1^2 + 2 \mu^2 \sum_{k=1}^{\infty} \left( (1 - \rho(k))^2 + \rho(k)/\mu \right) \phi_{11}(k) \\
&\quad + 2 \rho(k) (1 - \rho(k)) \phi_{12}(k) + \rho(k)^2 \phi_{22}(k) - \phi_1^2), \\
\sigma_{23} &= \mu \phi_{12}(0) - \mu \phi_1^2 + 2 \mu \sum_{k=1}^{\infty} \left( \phi_{11}(k) + \rho(k) (\phi_{12}(k) - \phi_{11}(k)) - \phi_1^2 \right), \\
\sigma_{33} &= \phi_{11}(0) - \phi_1^2 + 2 \sum_{k=1}^{\infty} \left( \phi_{11}(k) - \phi_1^2 \right).
\end{align*}
\]

The proof of Lemma 1 can be found in Appendix A.1. In the following Theorem 1, we derive the asymptotics of the statistics $\hat{T}_1, \ldots, \hat{T}_4$ from (10). To allow for improved size properties of the corresponding GoF tests (relevant for small $n$), we offer bias-corrected formulas for the means of $\hat{T}_1, \ldots, \hat{T}_4$.

**Theorem 1.** Let the DGP $(X_t)_{t \in \mathbb{Z}}$ satisfy the assumptions in Lemma 1. Then, the distribution of the statistic $\hat{T}_i$, $i = 1, \ldots, 4$, according to (10) can be approximated by the normal distribution $N(\mu_{T_i}, \sigma_{T_i}^2/n)$, where the bias-corrected means are given by

\[
\begin{align*}
\mu_{T_1} &= 1 + \frac{1}{n} \frac{\sigma_{22} - \mu \sigma_{23} - \phi_1^2 \sigma_{11}}{\mu^2 \phi_1^2}, \\
\mu_{T_2} &= 1 + \frac{1}{n} \frac{\mu \sigma_{33} - \sigma_{23}}{\mu \phi_1^2}, \\
\mu_{T_3} &= \frac{1}{n} \sigma_{13}, \\
\mu_{T_4} &= \frac{1}{n} \frac{\sigma_{33} - \sigma_{23}}{\phi_1^2},
\end{align*}
\]

and the variances by

\[
\begin{align*}
\sigma_{T_1}^2 &= \frac{-\sigma_{11}}{\mu^2} + \frac{1}{\mu^2 \phi_1^2} (\sigma_{22} - 2 \mu \sigma_{23} + \mu^2 \sigma_{33}), \\
\sigma_{T_2}^2 &= \sigma_{T_1}^2, \\
\sigma_{T_1}^2 &= \phi_1^2 \mu^2 \sigma_{T_1}^2, \\
\sigma_{T_4}^2 &= \mu^2 \sigma_{T_1}^2.
\end{align*}
\]

Here, the expressions for $\sigma_{ij}$ are provided by Lemma 1.

Theorem 1 follows from Lemma 1 by utilizing Taylor expansions of the statistics (“Delta method”); the proof can be found in Appendix A.2. The required values of $\phi_i$ and $\phi_{ij}(k)$ can be computed numerically by truncating the involved infinite sums, for example, $\phi_i = E[f(X + i)] \approx \ldots$
\[ \sum_{x=0}^{M} f(x+i) \cdot P(X = x) \] with \( M \) sufficiently large. Further implementation details are provided in Section 4.

### 3.3 Multivariate asymptotics

For a joint application of multiple Stein–Chen GoF statistics, their joint (asymptotic) distribution is required for computing the resulting size. Therefore, the asymptotics of Section 3 have to be extended in the following way. Let us focus on the case of two Stein–Chen functions \( f, g \) : \( \mathbb{N}_0 \to \mathbb{R} \); a further extension to more than two functions is straightforward. First, we substitute the three-dimensional definition of \( (Y_t)_t \) according to (11) by a five-dimensional one, namely,

\[
Y_t = \begin{pmatrix}
Y_{t,1} \\
Y_{t,2} \\
Y_{t,3} \\
Y_{t,4} \\
Y_{t,5}
\end{pmatrix}
= \begin{pmatrix}
X_t - \mu \\
X_t f(X_t) - \mu \phi_1 \\
f(X_t + 1) - \phi_1 \\
X_t g(X_t) - \mu \varphi_1 \\
g(X_t + 1) - \varphi_1
\end{pmatrix}
\] (12)

with \( \phi_k := E[f(X_t + k)] \) and \( \varphi_k := E[g(X_t + k)] \). Furthermore, in analogy to Lemma 1, we denote \( \phi_{ij}(k) := E[f(X_t + i) f(X_{t+k} + j)] \), \( \varphi_{ij}(k) := E[g(X_t + i) g(X_{t+k} + j)] \), and \( \theta_{ij}(k) := E[f(X_t + i) g(X_{t+k} + j)] \). If the DGP (under the null) satisfies the mixing conditions of Lemma 1, then \( n^{-1/2} \sum_{t=1}^{n} Y_t \) is asymptotically normally distributed.

**Lemma 2.** Let the DGP \( (X_t)_t \) satisfy the assumptions in Lemma 1. Then, \( n^{-1/2} \sum_{t=1}^{n} Y_t \) according to (12) is asymptotically normally distributed with mean \( \mathbf{0} \) and covariance matrix \( \Sigma = (\sigma_{ij}) \), where

- \( \sigma_{11}, \sigma_{22}, \sigma_{33}, \sigma_{12}, \sigma_{13}, \sigma_{23} \) are the same as in Lemma 1,
- \( \sigma_{44}, \sigma_{55}, \sigma_{14}, \sigma_{15}, \sigma_{45} \) agree with \( \sigma_{22}, \sigma_{33}, \sigma_{12}, \sigma_{13}, \sigma_{23} \), respectively, after having replaced \( \phi \) by \( \varphi \), and where

\[
\sigma_{24} = \mu \theta_{11}(0) + \mu^2 \theta_{22}(0) - \mu^2 \phi_1 \varphi_1 + 2 \mu^2 \sum_{k=1}^{\infty} \left( \frac{(1 - \rho(k))^2 + \rho(k)/\mu}{\rho(k)} \theta_{11}(k) \right) \\
+ \rho(k) (1 - \rho(k)) (\theta_{21}(k) + \theta_{12}(k)) + \rho(k)^2 \theta_{22}(k - \varphi_1 \phi_1),
\]

\[
\sigma_{25} = \mu \theta_{12}(0) - \mu \phi_1 \varphi_1 + 2 \mu \sum_{k=1}^{\infty} (\theta_{11}(k) + \rho(k) (\theta_{12}(k - \theta_{11}(k)) - \varphi_1 \phi_1),
\]

\[
\sigma_{34} = \mu \theta_{21}(0) - \mu \phi_1 \varphi_1 + 2 \mu \sum_{k=1}^{\infty} (\theta_{11}(k) + \rho(k) (\theta_{21}(k - \theta_{11}(k)) - \varphi_1 \phi_1),
\]

\[
\sigma_{35} = \theta_{11}(0) - \varphi_1 \phi_1 + 2 \sum_{k=1}^{\infty} (\theta_{11}(k) - \varphi_1 \phi_1).
\]

The proof of Lemma 2 is provided by Appendix A.3. In the second step, we apply the statistics \( T_1, \ldots, T_4 \) from (10) to the Stein–Chen functions \( f, g \) and derive their joint asymptotic distributions in analogy to our derivation of Theorem 1. To keep it simple, we use a unique type of statistic for both \( f, g \), but the derivations are easily adapted to the case of different statistics as well. Since
the bias corrections and asymptotic variances are already provided by Theorem 1, the subsequent
Theorem 2 concentrates on the asymptotic cross-covariance of both statistics.

Theorem 2. Let the DGP \((X_t)\) satisfy the assumptions in Lemma 1, and let \(T_{f,i}\) express that statistic
\(T_i\) is combined with function \(f\). Then, the joint distribution of the statistics \((\hat{T}_{f,i}, \hat{T}_{g,i})\) can be approxi-
mated by the multivariate normal distribution with the bias-corrected means \((\mu_{T_{f,i}}, \mu_{T_{g,i}})\) according to
Theorem 1, and where

\[
\begin{align*}
\text{Cov}[T_{f,1}, T_{g,1}] &= \text{Cov}[T_{f,2}, T_{g,2}] = -\frac{\sigma_{11}}{\mu^2} + \frac{\sigma_{24} - \mu (\sigma_{25} + \sigma_{34}) + \mu^2 \sigma_{35}}{\phi_1 \phi_1 \mu^2}, \\
\text{Cov}[T_{f,3}, T_{g,3}] &= \phi_1^2 \mu^2 \cdot \text{Cov}[T_{f,1}, T_{g,1}], \quad \text{Cov}[T_{f,4}, T_{g,4}] = \mu^2 \cdot \text{Cov}[T_{f,1}, T_{g,1}].
\end{align*}
\]

For the proof of Theorem 2, see Appendix A.4. As an example, in the special case of i.i.d.
Poisson counts, we explicitly get

\[
\text{Cov}[T_{f,1}, T_{g,1}] = \frac{\theta_{11}(0) + \mu (\theta_{11}(0) - \theta_{12}(0) - \theta_{21}(0) + \theta_{22}(0)) - 1}{\mu \phi_1 \phi_1}. \tag{13}
\]

Joint tests based on the multivariate asymptotics derived before are further discussed in Section 6.

4 | IMPLEMENTATION OF STEIN–CHEN GOF TESTS

Before studying the performance of the proposed Stein–Chen GoF tests by simulations, let us
briefly discuss their implementation in practice. This is done step-by-step by illustrative examples,
starting with the most simple case of i.i.d. counts and for processes with finitely many nonva-
nishing ACF values such as, for example, INMA(1) counts, and ending up with INAR(1) counts
where all pairs \((X_t, X_{t-k})\) are dependent for any lag \(k \in \mathbb{N}\).

4.1 | Application to i.i.d. counts

In the case of an i.i.d. DGP, the formulas derived in Lemma 1 and Theorem 1 simplify consider-
ably, because the infinite sums “\(\sum_{k=1}^{\infty} \ldots\)” completely vanish. Thus, the asymptotic covariances
in Lemma 1 reduce to

\[
\begin{align*}
\sigma_{11} &= \mu, \quad \sigma_{12} = (\phi_1 + \mu (\phi_2 - \phi_1)) \sigma_{11}, \\
\sigma_{13} &= (\phi_2 - \phi_1) \sigma_{11}, \quad \sigma_{22} = \mu \phi_1(0) + \mu^2 \phi_2(0) - \mu^2 \phi_1^2, \\
\sigma_{23} &= \mu \phi_1(0) - \mu \phi_1^2, \quad \sigma_{33} = \phi_1(0) - \phi_1^2.
\end{align*}
\]  

(14)

As stated before, the involved moments \(\phi_1\) and \(\phi_{ij}(0)\) have to be approximated by truncating
the summation after the \(M\)th summand, with \(M\) being sufficiently large. We decided to set \(M\) equal
to the \((1 - 10^{-7})\)-quantile of the hypothetical Poisson distribution. Let us illustrate this approach
and the whole application of the Stein–Chen tests with a first data example.

We analyze a time series of \(n = 156\) strike counts (for 4-week periods in 1948–1959) in the
U.K. coal mining industry, which have been published as a frequency table in Kendall (1961, table
Application to INMA(1) counts

The asymptotics of Lemma 1 and Theorem 1 do not only simplify in the i.i.d. case. For the Poi-INMA(1) DGP \(X_t = \epsilon_t + \beta \circ \epsilon_{t-1}\), which also satisfies the conditions of Lemma 1, we have independence between \(X_t\) and \(X_{t-k}\) if \(k \geq 2\). Therefore, only the summand \(k = 1\) in the sums \(\sum_{k=1}^{\infty} \cdots\) in Lemma 1 is nonzero. Compared with the i.i.d. case, we additionally have to compute the \(\phi_i(1)\), while \(\rho(1) = \beta / (1 + \beta)\) (Aleksandrov & Weiß, 2020a).

As an example, let us look at the time series “1*” from Freeland (1998), where claimants from the logging industry with burn-related injuries are counted on a monthly basis (1985–1994, so \(n = 120\)). In Aleksandrov and Weiß (2020a), it was shown that these data are well described by a Poi-INMA(1) model, although the sample variance \(\approx 0.162\) is slightly smaller than the mean 0.175. To cross-check this result, we apply the Stein–Chen test statistic \(\hat{T}_2\) with \(f(x) = \exp(-x)\), which takes the value \(\approx 1.053\). Using again moment estimates, the two-sided critical values are computed from an approximate normal distribution, whose mean and variance have to account for the apparent INMA(1) autocorrelation (\(\hat{\rho}(1) \approx 0.222\)). We get \(\mu_{\hat{T}_2} \approx 1.008\) and \(\sigma_{\hat{T}_2} / \sqrt{n} \approx 0.088\) (compared with \(\approx 1.006\) and \(\approx 0.084\), respectively, under an i.i.d. assumption) such that we cannot reject the Poisson null hypothesis, in accordance with the findings of Aleksandrov and Weiß (2020a).

Application to INAR(1) counts

For Poi-INAR processes, the pairs \((X_t, X_{t-k})\) are dependent for any lag \(k \in \mathbb{N}\), that is, the sums \(\sum_{k=1}^{\infty} \cdots\) in Lemma 1 are truly infinite. Nevertheless, also for Poi-INAR processes, some simplifications are possible. If, for example, the DGP \((X_t)_{t\geq 1}\) is a Poi-INAR(1) process according to Section 2, so \(X_t = \alpha \circ X_{t-1} + \epsilon_t\), then the conditions of Lemma 1 and Theorem 1 are satisfied. Furthermore, the expression for \(\sigma_{11}\) (and thus also for \(\sigma_{12}, \sigma_{13}\)) can be simplified by using the geometric sum formula:

\[1 + 2 \sum_{k=1}^{\infty} \rho(k) = 1 + 2 \sum_{k=1}^{\infty} \alpha^k = \frac{1 + \alpha}{1 - \alpha}.
\]

For the remaining \(\sigma_{i,j}\), however, a closed-form solution is not possible, that is, we additionally have to truncate the sums \(\sum_{k=1}^{\infty} \cdots\) after \(N\) summands with \(N\) sufficiently large. We implemented a dynamic choice of \(N\), where \(N\) is not further increased if the Euclidean distance between
(μT, σT/√n) computed with N summands and (μT, σT/√n) computed with N − 1 summands felt below 10−7.

Let us discuss the emergency counts in section 3.1 of Weiβ (2013) for illustration. The length of this time series is n = 96, its sample mean ≈ 2.563, while the sample variance ≈ 1.870 is considerably smaller. The ACF is of AR(1)-type with rather large lag-1 ACF, ̂ρ(1) ≈ 0.664, such that Weiβ (2013) decided to use an INAR(1) model for these data. To test the null hypothesis of a Poi-INAR(1) model, we use the dispersion test as well as the Stein–Chen test statistic ̂T 2 with f(x) = exp(−x). The asymptotic implementation of the tests is done by using the Poi-INAR(1) moment estimates. Although the dispersion index ̂I ≈ 0.722 is quite low, we cannot reject the null hypothesis. This is due to the large standard error σI/√n ≈ 0.232 (while μI ≈ 0.948), which is caused by the large ACF in combination with the low sample size. This decision is confirmed by the ̂T 2-test, which leads to ̂T 2 ≈ 1.237 while μT 2 ≈ 1.058, σT 2/√n ≈ 0.197. To achieve the latter values, the upper summation limit was chosen as N = 31. Thus, altogether 2 × 3 × 32 = 98 ϕ-moments had to be computed, causing a lot of programming work.

Therefore, for DGPs with a nonvanishing ACF, one may think of an alternative to an asymptotic implementation, namely, a parametric bootstrap solution. Detailed investigations of INAR bootstrap schemes can be found in Jentsch and Weiβ (2019) and Weiβ and Jentsch (2019). The bootstrap solution requires less programming work but at the cost of increased computing times. Given the moment estimates for the emergency counts time series, B = 500 bootstrap time series of length n = 96 are generated from the fitted Poi-INAR(1) model and used to compute replicates of the statistic ̂T 2 with f(x) = exp(−x). Then, we compute the 2.5%-quantile from these replicated statistics as the lower critical value, leading to ≈ 0.702 (asymptotic: ≈ 0.672), and the 97.5%-quantile as the upper critical value ≈ 1.511 (asymptotic: ≈ 1.445). Thus, ̂T 2 ≈ 1.237 again does not lead to a rejection. This type of INAR(1) bootstrap implementation is further investigated in our simulations for Section 5.3, while we consider the asymptotic implementations for the i.i.d. and INMA(1) DGPs in Section 5.

5 RESULTS FROM A SIMULATION STUDY

To analyze the performance of the Stein–Chen GoF tests proposed in Section 3, we did a comprehensive simulation study. Relevant excerpts thereof are presented in the sequel, while further simulation results are available from the authors upon request. In Sections 5.1 and 5.2, when the GoF tests are implemented based on the asymptotic results of Theorem 1, we use 104 replications per scenario. The simulations are done for the means μ ∈ {2, 5} and for the sample sizes n ∈ {100, 250, 500, 1000}. Later in Section 5.3, when considering the computer-intensive bootstrap implementations, we had to reduce the number of replications to 5000 and considered less scenarios, details are given below. The length of the prerun (if required) was chosen as 100, the nominal level for the GoF tests is 5% throughout.

We compared the performance of the statistics ̂T 1, … , ̂T 4 with two-sided critical values. For the asymptotic implementations based on Theorem 1, we used the plug-in approach, that is, the true parameter values needed for the asymptotic approximations are replaced by the corresponding moment estimates. Doing this, it turned out that the statistic ̂T 4 has the same performance as the statistic ̂T 2, as long as we plug-in the sample mean instead of the true mean in Theorem 1. Therefore, ̂T 4 is not further considered in the sequel.
One of the major aims of our simulation study was to find out which function $f$ (weighting scheme) in combination with which statistic $\hat{T}_i$ leads to the best size and power values with respect to different alternative scenarios and in the presence of serial dependence. To keep the results manageable, we present them only for those functions $f$ and statistics $\hat{T}_i$, where, provided a decent size, the power was among the best in almost all scenarios. We also explain why which functions $f$ work best for which scenario such that guidance for the choice of $f$ is provided. As alternative scenarios, we considered the cases of equidispersion but not Poisson, underdispersion, overdispersion, and zero inflation. We considered the negative binomial (NB) distribution in the overdispersion case, the Good distribution for equidispersion and underdispersion, and the zero-inflated Poisson (ZIP) distribution for zero inflation. Background and references on the aforementioned distributions can be found in appendix A.1 of Weiß (2018a). Example plots of their probability mass function (PMF) are shown in Figure 1, where the hypothetical Poisson distribution is always plotted by gray bars.

5.1 | GoF tests for i.i.d. counts

If applying the Stein–Chen tests to i.i.d. counts, the asymptotic results of Lemma 1 and Theorem 1 simplify considerably, recall Section 4.1. For i.i.d. counts in the case of an equidispersed alternative, the Good distribution (see Table 1), we did not only consider the dispersion index $\hat{I}$ (corresponding to $f(x) = x - 1$) as a competitor, but also the function $f(x) = (x - 1)(x - 2)$ together with the statistic $\hat{T}_2$, because this leads to a statistic similar to the skewness index “$\hat{\theta}_Y$” used by Schweer and Weiß (2016). If, in the sequel, we refer to the skewness index, we actually refer to $\hat{T}_2$ with $f(x) = (x - 1)(x - 2)$. The results in Table 1 show that we achieve the best power in the equidispersion case if using the function $\exp(-x)$ (this function was already used for the data examples in Sections 4.1–4.3). This is reasonable as $\exp(-x)$ puts a large weight on small outcomes (where the difference between the Poisson and the equidispersed Good distribution is most apparent, recall Figure 1), whereas both $x - 1$ and $(x - 1)(x - 2)$ put their weight on large outcomes (which are hardly observed for an equidispersed distribution with low $\mu$). However, even if using $\exp(-x)$ as the weighting scheme, the power values are not particularly large for the lowest mean, $\mu = 2$, because the Poisson and the equidispersed Good distribution are quite similar in this case (Schweer & Weiß, 2016), also see Figure 1. But as we shall see in
Section 6, an appropriate type of joint Stein–Chen test allows to achieve a good power even for $\mu = 2$.

In Table 2, results are presented for an underdispersed alternative, where the dispersion index and the skewness index again serve as the benchmark tests. For underdispersion even more than for equidispersion, it is crucial to focus on the lower tail of the count distribution, recall Figure 1. Thus, it is plausible that the best power is again achieved for the exponential weighting scheme, $f(x) = \exp(-x)$, in combination with statistic $\hat{T}_2$ (clearly better than dispersion and skewness index). But also the inverse weighting scheme $1/(x + 0.1)$ with its large weight on low counts does quite well, whereas the remaining choices for $f$ in Table 2 lead to less power because of too much weight in the upper tail. Hence, if aiming at uncovering a certain type of alternative, it is important to identify the region of strongest deviation from the Poisson null, and to choose a function $f$ with sufficient weight in that region. For the equidispersed and underdispersed alternatives analyzed so far, the function $\exp(-x)$ with its weight on low counts can be recommended for achieving the best power.

In Table 3, we investigate the alternative of i.i.d. ZIP counts. Since the ZIP distribution approaches the Poisson null for $\sigma^2/\mu \to 1$, we also present a local power analysis this time (due to the use of potentially highly nonlinear functions $f$, a theory-based local power analysis of the proposed GoF tests, for several choices of $f$ and for different local alternatives, is nontrivial and not within the scope of this article). While in the equidispersion and underdispersion scenarios, the statistic $\hat{T}_2$ was superior, we now observed an advantage for $\hat{T}_1$. But except using a different construction for the test statistic, it is again $f(x) = \exp(-x)$ and $f(x) = 1/(x + 0.1)$ that perform best in terms of (local) power, both putting their maximal weight into zero where the zero inflation becomes most apparent, recall Figure 1. Note that the dispersion index does considerably worse than these two tests, which might be surprising as zero inflation goes along with overdispersion, and the latter is well detected by the dispersion index (also see the next paragraph). But because of the isolated point mass in zero (recall Figure 1), zero inflation causes a rather irregular type of overdispersion. This confirms our guidance to choose the weighting scheme $f$ in accordance with the region of strongest deviation from the Poisson null that one would like to detect.
Table 2  Simulated size and power (nominal level 0.05) of different two-sided tests in the situation of underdispersion with mean $\mu = 2$

| $\sigma^2/\mu$ | n   | $\hat{I}$ | $\hat{T}_2$ (x−1)(x−2) | $\hat{T}_2$ exp(−x) | $\hat{T}_2$ 1/(x + 0.1) | $\hat{T}_1$ $\sqrt{|x−1|}$ | $\hat{T}_1$ log(x + 1) |
|---------------|-----|-----------|--------------------------|---------------------|--------------------------|--------------------------|--------------------------|
| Size 1 100 0.046 0.037 0.047 0.046 0.044 0.046 0.046 |
| 250 0.051 0.043 0.052 0.051 0.051 0.050 |
| 500 0.050 0.047 0.051 0.053 0.052 0.052 |
| 1000 0.046 0.048 0.048 0.047 0.049 0.047 |
| Power 0.9 100 0.094 0.060 0.300 0.250 0.216 0.166 |
| 250 0.210 0.079 0.594 0.549 0.466 0.367 |
| 500 0.369 0.089 0.866 0.836 0.747 0.621 |
| 1000 0.623 0.101 0.991 0.987 0.958 0.893 |
| 0.8 100 0.284 0.069 0.611 0.562 0.524 0.444 |
| 250 0.638 0.174 0.936 0.922 0.893 0.828 |
| 500 0.912 0.305 0.998 0.997 0.994 0.985 |
| 1000 0.996 0.534 1.000 1.000 1.000 1.000 |

Note: Null: i.i.d. Poisson DGP. Alternative: i.i.d. Good distribution with underdispersion.

Finally, let us look at the regular type of overdispersion caused by an NB-alternative. In this case, we did not find any function $f$ from the above proposals that lead to a better power than the dispersion index $\hat{I}$ (i.e., $f(x) = x − 1$). Some functions perform almost as good as the dispersion index, but none does better. In Table 4, we show the NB-results of the functions $|x−1|^a$ with $a \in \{2, \ldots , 0.6\}$ for $\mu = 2$, where $a = 1$ corresponds to the dispersion index (its rejection rates are printed in italic font). These functions have increasing weight for increasing $x$, which is reasonable for uncovering overdispersion. We see that increasing deviations from $a = 1$ (especially decreases) lead to a deterioration of the power, showing that the dispersion index is the best choice for detecting such regular overdispersion.

To sum up, if constructing a Stein–Chen statistic aiming to detect a particular type of deviation from the Poisson null hypothesis (such as, e.g., underdispersion or zero inflation), one should identify the region in $\mathbb{N}_0$ of strongest deviation from the Poisson null, and then choose the weighting scheme $f$ such that it puts sufficient weight into this region. This guidance is also followed for the subsequent power analyses, where the counts exhibit additional serial dependence.

5.2 GoF tests for INMA(1) counts

After having analyzed how to choose the weighting scheme $f$ for achieving a good power, we now study the effect of additional serial dependence on the performance of the Stein–Chen GoF tests. As the first step, in this section, we investigate INMA(1) dependence, recall Section 4.2 for implementation details. More precisely, we consider the null hypothesis of a Poi-INMA(1) DGP, and INMA(1) DGPs with differently distributed innovations as the alternatives. For $n \leq 250$, a few simulation runs lead to the estimate $\hat{\rho}(1)$ being outside the interval $[0, 0.5]$, contradicting the INMA(1) model, and were thus excluded from our performance analyses. Generally, the obtained
| \( \sigma^2 / \mu \) | \( n \) | \( \hat{I} \) | \( \hat{T}_1 \) \( \exp(-x) \) | \( \hat{T}_1 \) \( 1/(x + 0.1) \) | \( \hat{I} \) | \( \hat{T}_1 \) \( \exp(-x) \) | \( \hat{T}_1 \) \( 1/(x + 0.1) \) |
|---|---|---|---|---|---|---|---|
| **Size** | | | | | | | |
| 1 | 100 | 0.049 | 0.050 | 0.048 | 0.049 | 0.048 | 0.047 |
| | 250 | 0.050 | 0.048 | 0.048 | 0.049 | 0.046 | 0.047 |
| | 500 | 0.050 | 0.051 | 0.051 | 0.049 | 0.048 | 0.049 |
| | 1000 | 0.047 | 0.049 | 0.050 | 0.050 | 0.050 | 0.050 |
| 2 | 100 | 0.118 | 0.226 | 0.207 | 0.138 | 0.536 | 0.473 |
| | 250 | 0.204 | 0.395 | 0.374 | 0.222 | 0.806 | 0.746 |
| | 500 | 0.355 | 0.633 | 0.618 | 0.360 | 0.959 | 0.932 |
| | 1000 | 0.596 | 0.891 | 0.883 | 0.588 | 0.998 | 0.996 |
| 1.25 | 100 | 0.416 | 0.712 | 0.687 | 0.421 | 0.933 | 0.905 |
| | 250 | 0.755 | 0.965 | 0.962 | 0.733 | 0.999 | 0.997 |
| | 500 | 0.956 | 0.999 | 0.999 | 0.940 | 1.000 | 1.000 |
| | 1000 | 1.000 | 1.000 | 1.000 | 0.999 | 1.000 | 1.000 |
| 1.5 | 100 | 0.883 | 0.993 | 0.992 | 0.850 | 0.999 | 0.998 |
| | 250 | 0.998 | 1.000 | 1.000 | 0.993 | 1.000 | 1.000 |
| | 500 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | 1000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| **Local power** | | | | | | | |
| 1.100 | 100 | 0.124 | 0.220 | 0.202 | 0.131 | 0.521 | 0.463 |
| | 1.063 | 250 | 0.122 | 0.203 | 0.192 | 0.132 | 0.554 | 0.483 |
| | 1.045 | 500 | 0.115 | 0.200 | 0.189 | 0.121 | 0.559 | 0.489 |
| | 1.032 | 1000 | 0.118 | 0.199 | 0.191 | 0.119 | 0.562 | 0.488 |
| | 1.250 | 100 | 0.405 | 0.708 | 0.684 | 0.421 | 0.924 | 0.899 |
| | 1.158 | 250 | 0.417 | 0.724 | 0.706 | 0.425 | 0.964 | 0.942 |
| | 1.112 | 500 | 0.416 | 0.726 | 0.711 | 0.429 | 0.979 | 0.962 |
| | 1.079 | 1000 | 0.416 | 0.729 | 0.718 | 0.425 | 0.987 | 0.974 |
| | 1.500 | 100 | 0.888 | 0.994 | 0.991 | 0.851 | 0.998 | 0.997 |
| | 1.316 | 250 | 0.905 | 0.997 | 0.996 | 0.877 | 1.000 | 1.000 |
| | 1.224 | 500 | 0.911 | 0.998 | 0.997 | 0.894 | 1.000 | 1.000 |
| | 1.158 | 1000 | 0.926 | 0.998 | 0.998 | 0.913 | 1.000 | 1.000 |

**Note:** Null: i.i.d. Poisson DGP. Alternative: i.i.d. ZIP distribution. For the local power analysis, \( \sigma^2 / \mu - 1 \) decreases with \( n^{-1/2} \).
TABLE 4 Simulated size and power (nominal level 0.05) of different two-sided tests of the test statistic $\hat{T}_3$ in the case of overdispersion with mean $\mu = 2$

| $\sigma^2 / \mu$ | $n$ | 2   | 1.8 | 1.4 | 1.2 | 1.1 | 1   | 0.9 | 0.8 | 0.6 |
|------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Size             | 1   | 100 | 0.040 | 0.042 | 0.045 | 0.047 | 0.046 | 0.046 | 0.047 | 0.051 | 0.050 |
|                  | 250 | 0.042 | 0.044 | 0.044 | 0.045 | 0.045 | 0.045 | 0.044 | 0.050 | 0.050 |
|                  | 500 | 0.044 | 0.046 | 0.046 | 0.046 | 0.046 | 0.046 | 0.047 | 0.051 | 0.053 |
|                  | 1000| 0.048 | 0.048 | 0.048 | 0.046 | 0.045 | 0.045 | 0.046 | 0.048 | 0.050 |
| Power            | 1.1 | 100 | 0.136 | 0.138 | 0.139 | 0.137 | 0.137 | 0.136 | 0.135 | 0.067 | 0.062 |
|                  | 250 | 0.208 | 0.214 | 0.223 | 0.221 | 0.221 | 0.220 | 0.217 | 0.066 | 0.055 |
|                  | 500 | 0.319 | 0.330 | 0.348 | 0.356 | 0.356 | 0.356 | 0.354 | 0.352 | 0.072 | 0.057 |
|                  | 1000| 0.515 | 0.538 | 0.573 | 0.585 | 0.589 | 0.590 | 0.587 | 0.085 | 0.067 |
|                  | 1.25| 100 | 0.356 | 0.369 | 0.393 | 0.401 | 0.403 | 0.403 | 0.400 | 0.096 | 0.065 |
|                  | 250 | 0.648 | 0.670 | 0.704 | 0.716 | 0.720 | 0.725 | 0.726 | 0.131 | 0.084 |
|                  | 500 | 0.887 | 0.902 | 0.927 | 0.936 | 0.937 | 0.939 | 0.940 | 0.180 | 0.113 |
|                  | 1000| 0.992 | 0.994 | 0.997 | 0.998 | 0.998 | 0.999 | 0.999 | 0.269 | 0.153 |
|                  | 1.5 | 100 | 0.743 | 0.764 | 0.798 | 0.809 | 0.813 | 0.815 | 0.815 | 0.209 | 0.127 |
|                  | 250 | 0.974 | 0.979 | 0.986 | 0.989 | 0.991 | 0.991 | 0.991 | 0.338 | 0.202 |
|                  | 500 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.514 | 0.311 |
|                  | 1000| 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.766 | 0.515 |

Note: Null: i.i.d. Poisson DGP. Alternative: i.i.d. NB distribution.

results for the INMA(1) case are quite similar to those of the i.i.d. case. Therefore, we reduced the number of presented scenarios to save some space.

Table 5 shows size and power results for the alternative of an INMA(1) DGP with equidispersed but Good-distributed innovations. Compared with the i.i.d. case in Table 1, we recognize a clearly worse power, that is, the additional serial dependence complicates the detection of this type of deviation from a Poisson distribution. Especially for the low mean $\mu = 2$, there is virtually no power left. Nevertheless, for $\mu = 5$, the Stein–Chen statistic with weight function $\exp(-x)$ emphasizing low counts clearly outperforms the dispersion and the skewness index.

Also in the case of an underdispersed alternative, the power deteriorates in the presence of serial dependence, see Table 6 in comparison with Table 2. But by contrast with the equidispersion case, we still have a reasonable power for detecting underdispersion in the case of an INMA(1) DGP. In particular, any of the considered Stein–Chen tests is still superior to the dispersion and skewness index. Altogether, the favorite solution appears to be the statistic $\hat{T}_2$ together with the exponential weighting scheme.

Finally, let us turn to the overdispersion scenarios. For the regular type of overdispersion caused by the NB-INMA(1) alternative, we still did not find any other function $f$ that improves over the dispersion index, recall the discussion of Table 4. We omitted plotting the results, because they are very similar to the i.i.d. case in Table 4 except having less power with increasing dependence. For overdispersion caused by zero inflation, that is, the alternative is an INMA(1) DGP
with ZIP-distributed innovations, we show some simulation results in Table 7. More precisely, we have to compare the upper part of Table 7 with the results in Table 3, because this upper part still relies on an implementation of the asymptotic results provided by Lemma 1 and Theorem 1. We see that both functions \( \exp(-x) \) and \( 1/(x+0.1) \) are superior to the dispersion index, although the actual power values are again decreased because of the apparent serial dependence. Thus, to sum up, it became clear that the guidance for choosing \( f \) provided in Section 5.1 still remains valid, but we have to be aware of a deterioration of power caused by the additional serial dependence.

By contrast with our previous analyses, this time, we did a further simulation experiment, which becomes more relevant in the next Section 5.3. Besides the asymptotic implementation of the Stein–Chen tests, we also did a parametric bootstrap implementation in analogy to Section 4.3, see the lower part of Table 7. So given the moment estimates for the available Monte–Carlo time series, \( B = 500 \) bootstrap time series of length \( n \) were generated from the fitted Poi-INMA(1) model and used to compute the considered statistics. The 2.5%- and 97.5%-quantile from these replicated statistics were used as the critical values for deciding about the null hypothesis. Table 7 compares the performance resulting from these different implementations. The size of the bootstrap implementation tends to exceed the nominal 5%-level, so it is slightly worse than the asymptotic implementation. Regarding the power, the asymptotic implementation is clearly better for the small sample size \( n = 100 \), while both implementations perform similarly well for \( n \geq 500 \). Thus, for INMA-type data, there is a preference for using the asymptotic implementation, also in view of the lower computational effort. But one should note that the bootstrap implementation generally performs similarly well, which is important to know for such situations where an asymptotic implementation is difficult—such as for the INAR-type DGP considered in Section 5.3.

### 5.3 GoF tests for INAR(1) counts

In the INMA(1) case discussed in Section 5.2, the motivation for the bootstrap implementation was solely academic, because an asymptotic implementation is easily possible and
TABLE 6 Simulated size and power (nominal level 0.05) of different two-sided tests in the situation of underdispersion with mean $\mu = 2$ and thinning parameter $\beta = 0.5$

| $\sigma^2/\mu$ | $n$ | $\hat{I}$ | $\hat{T}_2$ | $\hat{T}_2$ | $\hat{T}_2$ | $\hat{T}_1$ | $\hat{T}_1$ |
|---------------|-----|-----------|-------------|-------------|-------------|-------------|-------------|
| Size          |     |           |             |             |             |             |             |
| 1             | 100 | 0.052     | 0.034       | 0.055       | 0.052       | 0.050       | 0.052       |
|               | 250 | 0.047     | 0.043       | 0.051       | 0.049       | 0.050       | 0.049       |
|               | 500 | 0.049     | 0.043       | 0.047       | 0.047       | 0.049       | 0.051       |
|               | 1000| 0.048     | 0.050       | 0.048       | 0.047       | 0.048       | 0.047       |
| Power         | 0.9 | 100       | 0.077       | 0.039       | 0.158       | 0.130       | 0.128       | 0.115       |
|               |     |           |             |             |             |             |             |
|               | 250 | 0.171     | 0.073       | 0.269       | 0.245       | 0.238       | 0.221       |
|               | 500 | 0.303     | 0.108       | 0.475       | 0.452       | 0.432       | 0.406       |
|               | 1000| 0.545     | 0.195       | 0.749       | 0.733       | 0.712       | 0.677       |
| 0.8           | 100 | 0.211     | 0.071       | 0.362       | 0.318       | 0.324       | 0.309       |
|               |     |           |             |             |             |             |             |
|               | 250 | 0.560     | 0.248       | 0.687       | 0.664       | 0.677       | 0.667       |
|               | 500 | 0.857     | 0.495       | 0.927       | 0.924       | 0.932       | 0.924       |
|               | 1000| 0.992     | 0.800       | 0.998       | 0.998       | 0.999       | 0.999       |

Note: Null: Poi-INMA(1) DGP. Alternative: INMA(1) with innovations from Good distribution with underdispersion.

computationally more efficient. Nevertheless, it was important to recognize that at least for larger sample sizes, the performance of bootstrap and asymptotic implementation was nearly the same. The situation is different for the INAR(1) case. Then, also the summands for $k > 1$ have to be computed in Lemma 1 such that there is no computational advantage of an asymptotic implementation anymore, recall the discussion in Section 4.3. Therefore, we now consider the parametric bootstrap implementation of the INAR(1) case (with 5000 Monte-Carlo and $B = 500$ bootstrap replications, where the latter again use the Poi-INAR(1) fit relying on moment estimates). In our first simulation experiment, we compare the parametric bootstrap’s performance between the ZIP-INMA(1) and ZIP-INAR(1) case, see Table 8. There is no big difference between both cases. The sizes are a bit worse in the INMA(1) case, whereas the power is considerably better. The latter is easily explained: although both DGPs have the same value for their thinning parameter, the extent of serial dependence is much stronger for the INAR(1) DGP. In accordance with our previous observations, we conclude that the deterioration of the power is caused by this increase of dependence.

Next, let us discuss the alternative of Good-distributed innovations having equidispersion. Comparing Table 5 with Table 9, we see that the power in the INAR(1) case becomes again worse than in the INMA(1) case for increasing values of the dependence parameter $\alpha$. In particular, for $\alpha \geq 0.5$, there is virtually no power left despite the large sample size $n = 1000$. Comparing Table 6 with Table 10 for the case of Good-distributed innovations with underdispersion, we see that the power of the INAR(1) bootstrap implementation is comparable with the one of the asymptotic INMA(1) implementation for small $\alpha$, while the power gets again worse if the dependence increases. Another pattern is more notable: the ranking among the statistics changes with increasing $\alpha$. For $\alpha = 0.75$, the best power is observed for the dispersion index and the statistic $\hat{T}_1$ with $f(x) = \log(x + 1)$. 

\[ f(x) = \log(x + 1) \]
TABLE 7 Simulated size and power (nominal level 0.05) of different two-sided tests in the situation of zero inflation with mean $\mu = 2$ and thinning parameter $\beta = 0.5$.

| $\sigma^2 / \mu$ | $n$ | $i$ | $\hat{T}_1$ (exp($-x$)) | $1/(x + 0.1)$ | $\hat{T}_1$ (exp($-x$)) | $1/(x + 0.1)$ |
|------------------|-----|-----|--------------------------|----------------|--------------------------|----------------|
| Asym, 1          | 100 | 0.046 | 0.049                    | 0.049         | 0.049                    | 0.047         |
| size             | 500 | 0.049 | 0.047                    | 0.047         | 0.055                    | 0.050         |
|                  | 1000| 0.047 | 0.048                    | 0.050         | 0.049                    | 0.050         |
| Asym, 1.1        | 100 | 0.107 | 0.140                    | 0.128         | 0.114                    | 0.203         |
| power            | 500 | 0.298 | 0.386                    | 0.383         | 0.310                    | 0.506         |
|                  | 1000| 0.516 | 0.625                    | 0.627         | 0.515                    | 0.759         |
|                  | 1.25| 0.345 | 0.461                    | 0.446         | 0.360                    | 0.545         |
|                  | 500 | 0.916 | 0.967                    | 0.967         | 0.908                    | 0.978         |
|                  | 1000| 0.997 | 1.000                    | 1.000         | 0.996                    | 1.000         |
| Boot, 1          | 100 | 0.049 | 0.049                    | 0.049         | 0.051                    | 0.055         |
| size             | 500 | 0.053 | 0.052                    | 0.052         | 0.053                    | 0.053         |
|                  | 1000| 0.059 | 0.057                    | 0.055         | 0.054                    | 0.056         |
| Boot, 1.1        | 100 | 0.090 | 0.112                    | 0.114         | 0.091                    | 0.137         |
| power            | 500 | 0.277 | 0.347                    | 0.358         | 0.280                    | 0.451         |
|                  | 1000| 0.492 | 0.607                    | 0.615         | 0.503                    | 0.729         |
|                  | 1.25| 0.304 | 0.395                    | 0.404         | 0.302                    | 0.428         |
|                  | 500 | 0.902 | 0.962                    | 0.965         | 0.887                    | 0.972         |
|                  | 1000| 0.998 | 0.999                    | 1.000         | 0.994                    | 1.000         |

Note: Null: Poi-INMA(1) DGP. Alternative: INMA(1) with ZIP innovations. Upper part: Tests implemented based on asymptotic results (“asym”). Lower part: Tests implemented based on parametric bootstrap (“boot”).

In our final simulation experiment, we apply the block bootstrap to the INAR(1) DGP (with 5000 replications and $B = 500$). The motivation for this experiment is as follows: if we know that we are concerned with an INAR(1) DGP and just want to test for a Poisson distribution within this setup (as assumed before), there is no reason why one should prefer the block bootstrap over the parametric one. But if we do not know the actual type of DGP, then it seems plausible to use the block bootstrap to account for the (unknown) dependence structure while testing the Poisson null hypothesis. The performance of this approach is investigated for the INAR(1) DGP and compared with the parametric bootstrap’s performance. The idea is to take the 2.5%- and 97.5%-quantiles from the bootstrapped deviations $\hat{T}_{i,b} - \hat{T}_i$, $b = 1, \ldots, B$, as the critical values for the deviation of the test statistic $\hat{T}_i$ to its hypothetical value 0 or 1, respectively. For implementation, we used the stationary bootstrap (SB) method described in Politis and Romano (1994). The beginning of the first block was chosen according to a discrete uniform distribution, the length of the blocks themselves according to a geometric distribution with mean $b$. Here, the mean block length $b$ was either chosen as the fixed value 24 (which appears reasonable for sample size $n = 1000$), or in a data-driven way by using the function $b.star$ from R’s “np-package” (which implements...
TABLE 8  Parametric bootstrap, simulated size, and power (nominal level 0.05) of different two-sided tests in the case of zero inflation for the sample size $n = 1000$ with thinning parameter $\alpha = 0.5$ or $\beta = 0.5$, respectively

| $\sigma^2 / \mu$ | $\hat{T}_1 \exp(-x)$ | $\hat{T}_1 1/(x + 0.1)$ | $\hat{T}_1 \exp(-x)$ | $\hat{T}_1 1/(x + 0.1)$ |
|-----------------|-----------------------|--------------------------|-----------------------|--------------------------|
| ZIP- 1          | 1                     | 0.048                    | 0.055                 | 0.054                    |
| INAR(1) 1.1     | 0.390                 | 0.425                    | 0.437                 | 0.379                    |
|                 | 1.25                  | 0.980                    | 0.978                 | 0.982                    |
| ZIP- 1          | 1                     | 0.059                    | 0.057                 | 0.055                    |
| INMA(1) 1.1     | 0.492                 | 0.607                    | 0.615                 | 0.503                    |
|                 | 1.25                  | 0.998                    | 0.999                 | 1.000                    |

Note: Upper part: Poi-INAR(1) DGP, ZIP-INAR(1) alternative. Lower part: Poi-INMA(1) DGP, ZIP-INMA(1) alternative.

TABLE 9  Parametric bootstrap, simulated size, and power (nominal level 0.05) of different two-sided tests in the case of equidispersion for $n = 1000$

| $\mu$ = 2 | $\hat{T}_2 (x - 1)(x - 2) \exp(-x)$ | $\hat{T}_2 (x - 1)(x - 2) \exp(-x)$ |
|-----------|-------------------------------------|-------------------------------------|
| $\alpha$ | $\hat{T}_2$                          | $\hat{T}_2$                          |
| 0.25      | 0.058                               | 0.057                               |
| 0.50      | 0.055                               | 0.054                               |
| 0.75      | 0.061                               | 0.054                               |
| 0.25      | 0.071                               | 0.126                               |
| 0.50      | 0.063                               | 0.070                               |
| 0.75      | 0.059                               | 0.062                               |

Note: Null: Poi-INAR(1) DGP. Alternative: INAR(1) with innovations from Good distribution with equidispersion.

the approach of Patton, Politis, & White, 2009). For both SB implementations, we “wrapped” the data around a circle such that $X_n$ is followed by $X_1$. So if a block extends beyond $X_n$, we continue with $X_1$, … for building the block, that is, we take $X_{(t-1)(\text{mod} n)+1}$ at time $t$. The results obtained for the SB implementations, in comparison with those of the parametric bootstrap, are summarized in Table 11. While the SB implementations lead to roughly the same power, we observe a rather strong oversizing. In addition, for sample sizes smaller than $n = 1000$, we did not observe a reasonable performance at all. Hence, the SB implementations have to be used with caution.

6 | JOINT STEIN–CHEN GOF TESTING

If testing the Poisson null hypothesis against a specific type of alternative, we provided guidance in Section 5 of how to choose an appropriate weighting scheme $f$ for the Stein–Chen statistic.
**Table 10**  Parametric bootstrap, simulated size and power (nominal level 0.05) of different two-sided tests in the situation of underdispersion for \( n = 1000 \) and \( \mu = 2 \)

| \( \alpha \) | \( \sigma^2/\mu \) | \( \hat{l} \) | \( \hat{T}_2 \) | \( \hat{T}_2 \) | \( \hat{T}_2 \) | \( \hat{T}_1 \) | \( \hat{T}_1 \) |
|---|---|---|---|---|---|---|---|
| 0.25 | 1 | 0.056 | 0.055 | 0.062 | 0.061 | 0.059 | 0.058 |
| 0.9 | 0.591 | 0.210 | 0.844 | 0.836 | 0.799 | 0.750 |
| 0.8 | 0.995 | 0.799 | 1.000 | 1.000 | 1.000 | 1.000 |
| 0.50 | 1 | 0.061 | 0.062 | 0.055 | 0.055 | 0.053 | 0.057 |
| 0.9 | 0.475 | 0.292 | 0.490 | 0.502 | 0.522 | 0.520 |
| 0.8 | 0.967 | 0.833 | 0.969 | 0.972 | 0.975 | 0.978 |
| 0.75 | 1 | 0.061 | 0.063 | 0.058 | 0.059 | 0.056 | 0.061 |
| 0.9 | 0.272 | 0.237 | 0.214 | 0.224 | 0.246 | 0.265 |
| 0.8 | 0.755 | 0.673 | 0.630 | 0.654 | 0.705 | 0.741 |

Note: Null: Poi-INAR(1) DGP. Alternative: INAR(1) with innovations distributed as Good distribution with underdispersion.

**Table 11**  First block corresponds to parametric bootstrap, second block corresponds to SB with fixed \( b = 24 \) (“SB24”), and third block corresponds to SB with \( b \) selected by \( b \).star (“SB*”)

| \( \mu \) | \( \sigma^2/\mu \) | \( \hat{l} \) | \( \hat{T}_1 \) | \( \hat{T}_1 \) | \( \hat{l} \) | \( \hat{T}_1 \) | \( \hat{T}_1 \) |
|---|---|---|---|---|---|---|---|
| Par. | 1 | 0.048 | 0.055 | 0.054 | 0.053 | 0.049 | 0.049 |
| boot | 1.1 | 0.390 | 0.425 | 0.437 | 0.379 | 0.404 | 0.500 |
| 1.25 | 0.980 | 0.978 | 0.982 | 0.976 | 0.960 | 0.992 |
| SB24 | 1 | 0.078 | 0.075 | 0.073 | 0.071 | 0.103 | 0.086 |
| 1.1 | 0.431 | 0.339 | 0.382 | 0.385 | 0.266 | 0.382 |
| 1.25 | 0.981 | 0.962 | 0.974 | 0.976 | 0.921 | 0.981 |
| SB* | 1 | 0.070 | 0.067 | 0.066 | 0.077 | 0.096 | 0.088 |
| 1.1 | 0.422 | 0.312 | 0.359 | 0.380 | 0.251 | 0.367 |
| 1.25 | 0.982 | 0.960 | 0.972 | 0.976 | 0.915 | 0.980 |

Note: Simulated size and power (nominal level 0.05) of different two-sided tests in the case of zero inflation for the sample size \( n = 1000 \) with thinning parameter \( \alpha = 0.5 \). Null: Poi-INAR(1) DGP. Alternative: ZIP-INAR(1).

To achieve a good power for multiple types of alternatives, different choices of the Stein–Chen function \( f \) could be necessary. Thus, it suggests itself to apply multiple functions and the corresponding test statistics jointly for getting some kind of omnibus test. The required asymptotics have already been derived in Section 3.3 before.

As an application of the proposed joint Stein–Chen tests, let us return to the i.i.d. simulations discussed in Section 5.1. In Table 1, we analyzed the performance of single Stein–Chen statistics to test against an equidispersed alternative. Using the function \( \exp(-x) \), we obtained a better power...
than for the dispersion or skewness test, but the power was still not satisfactory for $\mu = 2$. Therefore, let us now analyze the joint test statistic $(\hat{T}_{f,2}, \hat{T}_{g,2})$ with functions $g(x) = \exp(-x)$ and either $f(x) = x - 1$ (dispersion test) or $f(x) = (x - 1)(x - 2)$ (skewness test), see Table 12. Using (13), the test was implemented by utilizing that the quadratic form $n (\hat{T}_{f,2}, \hat{T}_{g,2})\Sigma^{-1}(\hat{T}_{f,2}, \hat{T}_{g,2})^\top$ is asymptotically $\chi^2_2$-distributed. Comparing Table 1 with Table 12, it turns out that both combined tests outperform any of the one-dimensional statistics, with a good power not only for $\mu = 5$ but also for $\mu = 2$. Table 12 also indicates a slight advantage for the skewness–exp combination. Overall, we recognize that the combination of multiple Stein–Chen statistics offers the potential to achieve an improved performance than if testing based on a single statistic.

Let us apply this joint test in a final data example, where we analyze the pollonium data of Rutherford, Geiger, and Bateman (1910). For $n = 2608$ time intervals of length 1/8-min, the number of scintillations caused by the radioactive decay of polonium was determined. Since successive intervals were interrupted by breaks of several minutes, it is reasonable to assume an i.i.d. count DGP for these data. Rutherford et al. (1910) argued that a Poisson model is suitable for the data; in fact, we observe nearly equidispersed counts (sample mean $\approx 3.872$ and variance $\approx 3.696$). The Poisson null hypothesis was checked by Gürtler and Henze (2000) with altogether 18 GoF tests, where two of them lead to a rejection at the 5%-level. Thus, they conjectured that the data might originate from an equidispersed non-Poisson distribution. To further investigate this conjecture, we apply both aforementioned joint Stein–Chen tests to the data. The $\chi^2$-statistic of the dispersion–exp (skewness–exp) test takes the value $\approx 4.638$ ($\approx 2.594$). Since the 95%-quantile of the $\chi^2_2$-distribution equals $\approx 5.991$, we do not reject the Poisson null at the 5%-level. So in contrast to Gürtler and Henze (2000), we do not find evidence against a Poisson distribution for the pollonium data.

### TABLE 12

Simulated size and power (nominal level 0.05) of joint statistic with $(\hat{T}_{f,2}, \hat{T}_{g,2})$ in the case of equidispersion

| $\mu = 2$ | $\mu = 5$ |
|------------|------------|
| $f$: $(x - 1)$ | $(x - 1)(x - 2)$ | $f$: $(x - 1)$ | $(x - 1)(x - 2)$ |
| $n$ | $g$: $\exp(-x)$ | $\exp(-x)$ | $g$: $\exp(-x)$ | $\exp(-x)$ |
| Size | | | |
| 100 | 0.051 | 0.048 | 0.039 | 0.042 |
| 250 | 0.051 | 0.049 | 0.046 | 0.044 |
| 500 | 0.051 | 0.045 | 0.049 | 0.049 |
| 1000 | 0.050 | 0.049 | 0.047 | 0.047 |
| Power | | | |
| 100 | 0.175 | 0.171 | 0.132 | 0.166 |
| 250 | 0.291 | 0.300 | 0.402 | 0.448 |
| 500 | 0.504 | 0.498 | 0.765 | 0.793 |
| 1000 | 0.792 | 0.772 | 0.981 | 0.987 |

Note: Null: i.i.d. Poisson DGP. Alternative: i.i.d. Good distribution with equidispersion.

7 | CONCLUSIONS AND FUTURE RESEARCH

The newly proposed Stein–Chen tests allow to choose different weighting schemes $f$ for different types of alternative scenarios, thus being a flexible tool for uncovering diverse deviations from a
Poisson model. For example, the function \( f(x) = \exp(-x) \) turned out to be a good choice if testing against equidispersion, underdispersion, or zero inflation, where null model and alternative model mainly differ for low counts. By contrast, the linear function \( x - 1 \) (dispersion index) is recommended for NB-like overdispersion, where deviations to the Poisson null mainly manifest themselves in the upper tail of the distribution. We also showed how different Stein–Chen statistics can be applied jointly to achieve a better power performance. The asymptotic derivations used for the computation of critical values rely on a bivariate Poisson distribution for lagged pairs \((X_t, X_{t-k})\) under the null hypothesis and are thus widely applicable, for example, to DGP from the Poi-INARMA family.

There are several directions for future research. If choosing \( f(x) = x - 1 \), our Stein–Chen approach covers the test statistics relying on second-order factorial moments that were proposed by Kyriakoussis et al. (1998). In fact, we also considered the third-order case \( f(x) = (x - 1)(x - 2) \), but both cases are yet limited to a Poisson null hypothesis. So besides an analysis of higher-order statistics, it would be relevant to extend our Stein–Chen approach to other types of null hypothesis, such as the (negative) binomial distribution \( \text{Bin}(n, \pi) \) (NB\((n, \pi)\)) considered by Kyriakoussis et al. (1998). Test statistics could be defined by considering adequate types of Stein identity, such as \( E[X \cdot f(X)] = \mu \cdot E[\frac{1 - X/n}{1 - \mu/n} \cdot f(X + 1)] \) in the Bin case or \( E[X \cdot f(X)] = \mu \cdot E[\frac{1 + X/n}{1 + \mu/n} \cdot f(X + 1)] \) in the NB case, see Sudheesh and Tibiletti (2012) for these and further identities. The resulting tests could then be applied to, for example, the geometric INAR-type DGP of Ristić, Bakouch, and Nastić (2009), which uses the NB-thinning operation instead of the binomial one. Related to this, also an asymptotic analysis of the local power of the proposed Stein–Chen tests, with respect to different local alternatives, could be relevant to achieve a theory-based guidance for choosing \( f \).

Another direction for future research is trying to apply the Stein–Chen approach to DGP from the Poi-INGARCH (INteger-valued Generalized AutoRegressive Conditional Heteroscedasticity) family. For these conditional Poisson regression models, the conditional distribution of \( X_t \) given \( X_{t-1}, \ldots \) is \( \text{Poi}(M_t) \), where the conditional mean \( M_t = E[X_t | X_{t-1}, \ldots] \) depends linearly on past observations and possibly on past conditional means (Weiß, 2018a). Then, the Stein–Chen identity (1) could be applied to the conditional \( \text{Poi}(M_t) \)-distribution, leading to

\[
E[X_t \cdot f(X_t) | X_{t-1}, \ldots] = M_t \cdot E[f(X_t + 1) | X_{t-1}, \ldots].
\]

This is used to define standardized Stein–Chen residuals as

\[
R_t := \frac{X_t \cdot f(X_t) - M_t \cdot E[f(X_t + 1) | X_{t-1}, \ldots]}{\sqrt{V[X_t \cdot f(X_t) | X_{t-1}, \ldots]}}, \tag{15}
\]

where the required conditional variance follows from Equation (A3) as

\[
V[X_t \cdot f(X_t) | X_{t-1}, \ldots] = M_t^2 \cdot E[f(X_t + 2)^2 | X_{t-1}, \ldots]
+ M_t \cdot E[f(X_t + 1)^2 | X_{t-1}, \ldots] - M_t^2 \cdot E[f(X_t + 1) | X_{t-1}, \ldots]^2.
\]

These residuals might be used for diagnostic model checking in analogy to Zhu and Wang (2010). The choice \( f(x) \equiv 1 \) in (15) results in the ordinary (standardized) Pearson residuals, see Aleksandrov and Weiß (2020b) for a recent investigation, but we conjecture that nonconstant Stein–Chen functions will lead to an improved performance for some types of alternatives.
ACKNOWLEDGMENTS
The authors thank the associate editor and 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 437270842.

ORCID
Christian H. Weiß https://orcid.org/0000-0001-8739-6631
Carsten Jentsch https://orcid.org/0000-0001-7824-1697

REFERENCES
Aleksandrov, B., & Weiß, C. H. (2020a). Parameter estimation and diagnostic tests for INMA(1) processes. TEST, 29(1), 196–232.
Aleksandrov, B., & Weiß, C. H. (2020b). Testing the dispersion structure of count time series using Pearson residuals. AStA Advances in Statistical Analysis, 104(3), 325–361.
Alzaid, A. A., & Al-Osh, M. A. (1988). First-order integer-valued autoregressive process: Distributional and regression properties. Statistica Neerlandica, 42(1), 53–61.
Alzaid, A. A., & Al-Osh, M. A. (1990). An integer-valued pth-order autoregressive structure (INAR(p)) process. Journal of Applied Probability, 27(2), 314–324.
Beltrán-Beltrán, N., & O’Reilly, F. J. (2019). On goodness of fit tests for the Poisson, negative binomial and binomial distributions. Statistical Papers, 60(1), 1–18.
Betsch, S., Ebner, B., & Nestmann, F. (2020). Characterizations of non-normalized discrete probability distributions and their application in statistics. Preprint, arXiv:2011.04369v1.
Chen, L. H. Y. (1975). Poisson approximation for dependent trials. Annals of Probability, 3(3), 534–545.
Doukhan, P., Fokianos, K., & Li, X. (2012). On weak dependence conditions: The case of discrete valued processes. Statistics and Probability Letters, 82(11), 1941–1948 with a correction in Vol. 83(2), 674–675.
Du, J.-G., & Li, Y. (1991). The integer-valued autoregressive (INAR(p)) model. Journal of Time Series Analysis, 12(2), 129–142.
Erhardsson, T. (2005). Stein’s method for Poisson and compound Poisson approximation. In A. H. Barbour & L. H. Y. Chen (Eds.), An introduction to Stein’s method (pp. 61–113). Singapore, Asia: Singapore University Press and World Scientific Publishing Co. Pte. Ltd.
Freeland, R.K. (1998). Statistical analysis of discrete time series with applications to the analysis of workers compensation claims data (PhD thesis). University of British Columbia, Canada. https://open.library.ubc.ca/cIRcle/collections/ubctheses/831/items/1.0088709
Gürtler, N., & Henze, N. (2000). Recent and classical goodness-of-fit tests for the Poisson distribution. Journal of Statistical Planning and Inference, 90(2), 207–225.
Ibragimov, I. (1962). Some limit theorems for stationary processes. Theory of Probability & Its Applications, 7(4), 349–382.
Jentsch, C., & Weiß, C. H. (2019). Bootstrapping INAR models. Bernoulli, 25(3), 2359–2408.
Johnson, N. L., Kotz, S., & Balakrishnan, N. (1997). Discrete multivariate distributions. Hoboken, NJ: John Wiley & Sons, Inc.
Kendall, M. G. (1961). Presidential address: Natural law in the social science. Journal of the Royal Statistical Society, Series A, 124(1), 1–16.
Kyriakoussis, A., Li, G., & Papadopoulos, A. (1998). On characterization and goodness-of-fit test of some discrete distribution families. Journal of Statistical Planning and Inference, 74(2), 215–228.
Patton, A. J., Politis, D. N., & White, H. (2009). Correction: Automatic block-length selection for the dependent bootstrap. Econometric Reviews, 28(4), 372–375.
Politis, D. N., & Romano, J. P. (1994). The stationary bootstrap. Journal of the American Statistical Association, 89(428), 1303–1313.
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.
Rutherford, E., Geiger, H., & Bateman, H. (1910). LXXVI. The probability variations in the distribution of α particles. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 20(118), 698–707.

Schweer, S., & Weiß, C. H. (2016). Testing for Poisson arrivals in INAR(1) processes. *TEST*, 25(3), 503–524.

Stein, C. (1972). *A bound for the error in the normal approximation to the distribution of a sum of dependent random variables*. Paper presented at: Proceedings of the 6th Berkeley Symposium on Mathematical Statistics and Probability (vol. 2, pp. 583–602). University of California Press, Berkeley, CA.

Stein, C. (1986). *Approximate computation of expectations IMS Lecture Notes* (Vol. 7). Hayward, CA: Hayward Industries, Inc.

Steutel, F. W., & van Harn, K. (1979). Discrete analogues of self-decomposability and stability. *Annals of Probability*, 7(5), 893–899.

Sudheesh, K. K., & Tibiletti, L. (2012). Moment identity for discrete random variable and its applications. *Statistics*, 46(6), 767–775.

Weiß, C. H. (2008). Serial dependence and regression of Poisson INARMA models. *Journal of Statistical Planning and Inference*, 138(10), 2975–2990.

Weiß, C. H. (2013). Integer-valued autoregressive models for counts showing underdispersion. *Journal of Applied Statistics*, 40(9), 1931–1948.

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

Weiß, C. H. (2018b). Goodness-of-fit testing of a count time series’ marginal distribution. *Metrika*, 81(6), 619–651.

Weiß, C. H., & Aleksandrov, B. (2020). Computing (bivariate) Poisson moments using Stein–Chen identities. *The American Statistician* (forthcoming).

How to cite this article: Aleksandrov, B., Weiß, C. H., & Jentsch, C. (2021). Goodness-of-fit tests for Poisson count time series based on the Stein–Chen identity. *Statistica Neerlandica*, 76(1), 35–64. https://doi.org/10.1111/stan.12252

### APPENDIX A

#### A.1 Proof of Lemma 1

The CLT by Ibragimov (1962) implies to calculate all covariances as

\[
\sigma_{ij} = E[Y_{t,i} \cdot Y_{t,j}] + \sum_{k=1}^{\infty} (E[Y_{t+k,i} \cdot Y_{t,j}] + E[Y_{t,i} \cdot Y_{t+k,j}]).
\]

Because of the symmetry of the BPOi_{\mu,\rho(k)}-distribution, both infinite sums are equal to each other for all components of Y_t. Hence, the above expression reduces to

\[
\sigma_{ij} = E[Y_{t,i} \cdot Y_{t,j}] + 2 \sum_{k=1}^{\infty} E[Y_{t,i} \cdot Y_{t+k,j}]. \tag{A1}
\]

To derive closed-form formulas for the \(\sigma_{ij}\) from (A1), we make use of the equality (Weiß & Aleksandrov, 2020)

\[
E[X_n \cdot f(X)] = \mu^n \cdot E[f(X + n)] \quad \text{for all } n \in \mathbb{N}, \tag{A2}
\]
where \((x)_{(n)} = x \cdots (x - n + 1)\) denotes the \(n\)th falling factorial. Equation (A2) allows to complement the Stein–Chen identity (1) by a corresponding variance formula, namely,

\[
V[X f(X)] = E[(X)(2) f(X)^2] + E[X f(X)^2] - E[X f(X)]^2 \\
= \mu^2 E[f(X + 2)^2] + \mu E[f(X + 1)^2] - \mu^2 E[f(X + 1)^2]. \quad (A3)
\]

The expression for \(\sigma_{11}\) follows from (A1) by considering the Poisson’s equidispersion property \(V[X_i] = E[X_i] = \mu\):

\[
\sigma_{11} = E[X_i^2] - \mu^2 + 2 \sum_{k=1}^{\infty} (E[X_{i+k}X_i] - \mu^2) \\
= V[X_i] + 2 \sum_{k=1}^{\infty} \text{Cov}[X_i, X_{i+k}] = \mu \left( 1 + 2 \sum_{k=1}^{\infty} \rho(k) \right).
\]

Next,

\[
\sigma_{12} = E[X_i^2 f(X_i)] - \mu^2 \phi_1 + 2 \sum_{k=1}^{\infty} (E[X_i f(X_i)X_{i+k}] - \mu^2 \phi_1) \\
\overset{(A2)}{=} \mu^2 E[f(X_i + 2)] + \mu E[f(X_i + 1)] - \mu^2 \phi_1 + 2 \sum_{k=1}^{\infty} (E[X_i f(X_i)X_{i+k}] - \mu^2 \phi_1) \\
\overset{(7)}{=} \mu^2 \phi_2 + \mu \phi_1 - \mu^2 \phi_1 \\
+ 2 \sum_{k=1}^{\infty} ((1 - \rho(k))\mu^2 + \rho(k)\mu) E[f(X_i + 1)] + \rho(k)(1 - \rho(k))\mu^2 E[f(X_i + 2)] \\
+ \rho(k)(1 - \rho(k))\mu^2 E[f(X_i + 1)] + \rho(k)^2 \mu^2 E[f(X_i + 2)] - \mu^2 \phi_1 \\
= \mu \phi_1 + \mu^2 (\phi_2 - \phi_1) + 2 \sum_{k=1}^{\infty} (\rho(k)(\mu - \mu^2) \phi_1 + \rho(k)\mu^2 \phi_2) \\
= (\mu \phi_1 + \mu^2 (\phi_2 - \phi_1)) \left( 1 + 2 \sum_{k=1}^{\infty} \rho(k) \right),
\]

where we used (7) with \(g \equiv 1\). This simplifies to the expression given in Lemma 1. Next,

\[
\sigma_{13} = E[X_i f(X_i + 1)] - \mu \phi_1 + 2 \sum_{k=1}^{\infty} (E[X_{i+k}f(X_i + 1)] - \mu \phi_1) \\
\overset{(A2)}{=} \mu \phi_2 - \mu \phi_1 + 2 \sum_{k=1}^{\infty} (E[X_{i+k}f(X_i + 1)] - \mu \phi_1) \\
\overset{(6)}{=} \mu(\phi_2 - \phi_1) + 2 \sum_{k=1}^{\infty} ((1 - \rho(k))\mu \phi_1 + \rho(k)\mu \phi_2 - \mu \phi_1) \\
= \mu(\phi_2 - \phi_1) \left( 1 + 2 \sum_{k=1}^{\infty} \rho(k) \right),
\]

where we used (6) with \(g \equiv 1\). Next,
\[
\sigma_{22} = E \left[ X_t^2 f(X_t)^2 \right] - \mu^2 \phi_1^2 + 2 \sum_{k=1}^{\infty} (E \left[ X_{t+k} f(X_{t+k}) X_t f(X_t) \right] - \mu^2 \phi_1^2)
\]

\[
\quad \overset{(A3)}{=} \mu E \left[ f(X_t + 1)^2 \right] + \mu^2 E \left[ f(X_t + 2)^2 \right] - \mu^2 \phi_1^2
\]

\[
\quad + 2 \sum_{k=1}^{\infty} (E \left[ X_{t+k} f(X_{t+k}) X_t f(X_t) \right] - \mu^2 \phi_1^2)
\]

\[
\quad \overset{(7)}{=} \mu E \left[ f(X_t + 1)^2 \right] + \mu^2 E \left[ f(X_t + 2)^2 \right] - \mu^2 \phi_1^2
\]

\[
\quad + 2 \sum_{k=1}^{\infty} (\mu^2 (1 - \rho(k))^2 + \rho(k) \mu) E \left[ f(X_{t+k} + 1) f(X_{t+1}) \right]
\]

\[
\quad + \rho(k) \mu^2 (1 - \rho(k)) E \left[ f(X_{t+k} + 1) f(X_{t+2}) \right]
\]

\[
\quad + \rho(k) \mu^2 (1 - \rho(k)) E \left[ f(X_{t+k} + 2) f(X_{t+1}) \right]
\]

\[
\quad + \rho(k) \mu^2 E \left[ f(X_{t+k} + 2) f(X_{t+2}) \right] - \mu^2 \phi_1^2).
\]

Note that we have the equality \( E \left[ f(X_{t+k} + 1) f(X_{t+1}) \right] = E \left[ f(X_{t+k} + 2) f(X_{t+1}) \right] \) because of the symmetry of \( \text{BPoI}_{\mu, \rho(k)} \). So \( \sigma_{22} \) simplifies to the expression given in Lemma 1.

Next, applying (A2) to \( f(x) f(x + 1) \), we get

\[
\sigma_{23} = E \left[ X_t f(X_t) f(X_t + 1) \right] - \mu \phi_1^2 + 2 \sum_{k=1}^{\infty} (E \left[ X_{t+k} f(X_{t+k}) f(X_{t+1}) \right] - \mu \phi_1^2)
\]

\[
\quad \overset{(A2)}{=} \mu E \left[ f(X_t + 1) f(X_t + 2) \right] - \mu \phi_1^2 + 2 \sum_{k=1}^{\infty} (E \left[ X_{t+k} f(X_{t+k}) f(X_{t+1}) \right] - \mu \phi_1^2)
\]

\[
\quad \overset{(6)}{=} \mu E \left[ f(X_t + 1) f(X_t + 2) \right] - \mu \phi_1^2 + 2 \sum_{k=1}^{\infty} ((1 - \rho(k)) \mu E \left[ f(X_{t+k} + 1) f(X_{t+1}) \right]
\]

\[
\quad + \rho(k) \mu E \left[ f(X_{t+k} + 1) f(X_{t+2}) \right] - \mu \phi_1^2).
\]

Finally,

\[
\sigma_{33} = E \left[ f(X_t + 1)^2 \right] - \phi_1^2 + 2 \sum_{k=1}^{\infty} (E \left[ f(X_{t+k} + 1) f(X_{t+1}) \right] - \phi_1^2)
\]

does not further simplify. So the proof of Lemma 1 is complete.

### A.2 Proof of Theorem 1

In Lemma 1, we established the asymptotic normality of

\[
\sqrt{n} \frac{\overline{Y}}{\overline{X}} = \sqrt{n} \left( \overline{X} - \mu, \overline{Xf(X)} - \mu \phi_1, \overline{f(X + 1)} - \phi_1 \right)^	op.
\]

The statistics \( T_i, i = 1, \ldots, 4 \), according to (10) are expressed as functions \( g_i : \mathbb{R}^3 \rightarrow \mathbb{R} \) applied to \( (\overline{X}, \overline{Xf(X)}, \overline{f(X + 1)}) \) in the following way:

\[
g_1(y) = \frac{y_1 \cdot y_3}{y_2}, \quad \text{where } g_1(\mu, \phi_1 \mu, \phi_1) = 1;
\]

\[
g_2(y) = \frac{y_2}{y_1 \cdot y_3}, \quad \text{where } g_2(\mu, \phi_1 \mu, \phi_1) = 1;
\]
\[ g_3(y) = y_2 - y_1 \cdot y_3, \quad \text{where } g_3(\mu, \phi_1 \mu, \phi_1) = 0; \]
\[ g_4(y) = \frac{y_2}{y_3} - y_1, \quad \text{where } g_4(\mu, \phi_1 \mu, \phi_1) = 0. \]

Then, the Delta method implies the asymptotic normality of \( \sqrt{n} (T_i - g_i(\mu, \phi_1 \mu, \phi_1)) \). To derive the asymptotic variance, we need the gradients of \( g_i \), which are given by

\[ \nabla g_1(y) = \left( \begin{array}{c} \frac{y_3}{y_2}, -\frac{y_1 y_3}{y_2^2}, \frac{y_1}{y_2} \end{array} \right), \quad \nabla g_3(y) = (-y_3, 1, -y_1), \]
\[ \nabla g_2(y) = \left( -\frac{y_2}{y_1 y_3}, \frac{1}{y_1 y_3}, -\frac{y_2}{y_1 y_3} \right), \quad \nabla g_4(y) = \left( -1, \frac{1}{y_3}, -\frac{y_2}{y_3^2} \right). \]

These are evaluated in \((\mu, \phi_1 \mu, \phi_1)\), leading to \( D_i := \nabla g_i(\mu, \phi_1 \mu, \phi_1) \) as

\[ D_1 = \left( \frac{1}{\mu}, -\frac{1}{\phi_1 \mu}, -\frac{1}{\phi_1} \right), \quad D_3 = (-\phi_1, 1, -\mu) = -\phi_1 \mu D_1, \]
\[ D_2 = \left( -\frac{1}{\mu}, \frac{1}{\phi_1 \mu}, -\frac{1}{\phi_1} \right) = -D_1, \quad D_4 = \left( -1, \frac{1}{\phi_1}, -\frac{\mu}{\phi_1} \right) = -\mu D_1. \quad (A4) \]

Thus, we calculate the variance of the statistic \( \sqrt{n} T_i \) as \( \sigma^2_{T_i} = D_i \Sigma D_i^\top \), with \( \Sigma \) taken from Lemma 1. We start with

\[ \sigma^2_{T_i} = \frac{1}{\mu^2} \sigma_{11} + \frac{1}{\phi_1 \mu^2} \sigma_{22} + \frac{1}{\phi_1} \sigma_{33} - \frac{2}{\phi_1 \mu^2} \sigma_{12} + \frac{2}{\phi_1 \mu} \sigma_{13} - \frac{2}{\phi_1} \sigma_{23} \]
\[ = \frac{1}{\phi_1 \mu^2} (\phi_1^2 \sigma_{11} - 2 \phi_1 \sigma_{12} + 2 \phi_1 \mu \sigma_{13} + \sigma_{22} - 2 \mu \sigma_{23} + \mu^2 \sigma_{33}). \]

If we insert that \( \sigma_{12} = (\phi_1 + \mu (\phi_2 - \phi_1)) \sigma_{11} \) and \( \sigma_{13} = (\phi_2 - \phi_1) \sigma_{11} \) according to Lemma 1, it follows that

\[ \sigma^2_{T_i} = \frac{\sigma_{11}}{\phi_1 \mu^2} (\phi_1 - 2 (\phi_1 + \mu (\phi_2 - \phi_1)) + 2 \mu (\phi_2 - \phi_1)) \]
\[ + \frac{1}{\phi_1 \mu^2} (\sigma_{22} - 2 \mu \sigma_{23} + \mu^2 \sigma_{33}) \]
\[ = \frac{\sigma_{11}}{\mu^2} + \frac{1}{\phi_1 \mu^2} (\sigma_{22} - 2 \mu \sigma_{23} + \mu^2 \sigma_{33}). \]

The last terms is further simplified by using Lemma 1:

\[ \sigma_{22} - 2 \mu \sigma_{23} + \mu^2 \sigma_{33} = \mu(1 + \mu)\phi_{11}(0) + \mu^2 \phi_{22}(0) - 2 \mu^2 \phi_{12}(0) \]
\[ + 2 \mu^2 \sum_{k=1}^{\infty} (\rho(k)(\rho(k) + \frac{1}{\mu})\phi_{11}(k) + \rho(k)^2(\phi_{22}(k) - 2\phi_{12}(k))). \]

The remaining \( \sigma^2_{T_i} \) immediately follow as

\[ \sigma^2_{T_2} = \sigma^2_{T_1}, \quad \sigma^2_{T_3} = \phi_1^2 \mu^2 \sigma^2_{T_1}, \quad \sigma^2_{T_4} = \mu^2 \sigma^2_{T_1}. \]
Finally, we compute a bias correction based on a second-order Taylor expansion. For this purpose, we derive the Hessians $H_{g_i}(y_1, y_2, y_3)$ for $i = 1, \ldots, 4$, and evaluate them as $H_i := H_{g_i}(\mu, \phi_1 \mu, \phi_1)$. Then, the approximate bias correction is obtained as

$$\frac{1}{n} \left( \frac{1}{2} h_{i;11} \sigma_{11} + \frac{1}{2} h_{i;22} \sigma_{22} + \frac{1}{2} h_{i;33} \sigma_{33} + h_{i;12} \sigma_{12} + h_{i;13} \sigma_{13} + h_{i;23} \sigma_{23} \right).$$

For $T_1$, we get

$$H_{g_1} = \begin{pmatrix}
\frac{-y_3}{y_2^2} & -\frac{y_2}{y_3} & \frac{1}{y_2} \\
\frac{1}{y_3} & -\frac{y_3}{y_2^2} & \frac{1}{y_3} \\
\frac{1}{y_2} & \frac{1}{y_3} & 0
\end{pmatrix}, \quad H_1 = \begin{pmatrix}
\frac{-1}{\phi_1 \mu} & \frac{1}{\phi_1 \mu} \\
\frac{2}{\phi_1^2 \mu^2} & \frac{1}{\phi_1 \mu} \\
\frac{1}{\phi_1 \mu} & \frac{-1}{\phi_1^2 \mu} & 0
\end{pmatrix}.$$

Then, the approximate mean formula becomes

$$\mu_{T_1} = 1 + \frac{1}{n} \left( \frac{\mu}{\phi_1 \mu} - \frac{\sigma_{22} - \mu \sigma_{23}}{\phi_1^2 \mu^2} \right) \frac{1}{1 + \frac{1}{n} \left( \frac{-\sigma_{11}}{\mu^2} + \frac{\sigma_{22} - \mu \sigma_{23}}{\phi_1^2 \mu^2} \right)}.$$

By using the Lemma 1, we further simplify

$$\sigma_{22} - \mu \sigma_{23} = \mu \phi_{11}(0) + \mu^2 \phi_{22}(0) - \mu^2 \phi_{12}(0) + 2 \mu^2 \sum_{k=1}^{\infty} \left( \rho(k)^2 - \rho(k) + \frac{\rho(k)}{\mu} \right) \phi_{11}(k) + \left( \rho(k) - 2 \rho(k)^2 \right) \phi_{12}(k) + \rho(k)^2 \phi_{22}(k).$$

Next, we compute the Hessian

$$H_{g_2} = \begin{pmatrix}
\frac{2 y_2}{y_1 y_3} & -\frac{1}{y_1 y_3} & \frac{y_2}{y_1 y_3} \\
\frac{1}{y_2 y_3} & -\frac{1}{y_1 y_3} & -\frac{1}{y_2 y_3} \\
\frac{1}{y_2 y_3} & \frac{1}{y_1 y_3} & -\frac{2 y_3}{y_1 y_3}
\end{pmatrix}, \quad H_2 = \begin{pmatrix}
\frac{1}{\mu^2} & -\frac{1}{\phi_1 \mu^2} & \frac{1}{\phi_1 \mu} \\
\frac{-1}{\phi_1 \mu^2} & 0 & \frac{-1}{\phi_1 \mu} \\
\frac{1}{\phi_1 \mu} & \frac{1}{\phi_1 \mu} & \frac{2}{\phi_1^2}
\end{pmatrix}.$$

The approximate mean formula equals

$$\mu_{T_2} = 1 + \frac{1}{n} \frac{\phi_1^2 \sigma_{11} - \phi_1 \sigma_{12} + \phi_1 \mu \sigma_{13} - \mu \sigma_{23} + \mu^2 \sigma_{33}}{\mu^2 \phi_1^2} = 1 + \frac{1}{n} \frac{\mu \sigma_{33} - \sigma_{23}}{\mu \phi_1^2},$$

because $\mu \sigma_{13} - \sigma_{12} = -\phi_1 \sigma_{11}$. By using the Lemma 1, we further simplify

$$\mu \sigma_{33} - \sigma_{23} = -\mu \phi_{12}(0) + \mu \phi_{11}(0) + 2 \mu \sum_{k=1}^{\infty} \rho(k) \left( \phi_{11}(k) - \phi_{12}(k) \right).$$

The bias correction for $T_3$ is particularly simple, because the only nonzero entry of $H_{g_3}$ is the value $-1$ at positions $(1, 3)$ and $(3, 1)$. Thus,

$$\mu_{T_3} = -\frac{1}{n} \sigma_{13}.$$
Finally, we compute the Hessian

\[
\mathbf{H}_4 = \begin{pmatrix}
0 & 0 & 0 \\
0 & 0 & -\frac{1}{\phi_1^2} \\
0 & -\frac{1}{\phi_1^2} & \frac{2\nu_2}{\phi_1^2}
\end{pmatrix}, \quad \mathbf{H}_4 = \begin{pmatrix}
0 & 0 & 0 \\
0 & 0 & -\frac{1}{\phi_2^2} \\
0 & -\frac{1}{\phi_2^2} & \frac{2\nu_2}{\phi_2^2}
\end{pmatrix}.
\]

The approximate mean formula becomes

\[
\mu_{T_4} = \frac{1}{n} \frac{\mu \sigma_{23} - \sigma_{25}}{\phi_1^2}.
\]

This completes the proof of Theorem 1.

**A.3 Proof of Lemma 2**

We proceed in the same way as in the proof of Lemma 1 in Appendix A.1.

\[
\sigma_{24} = E[X_t^2f(X_t)g(X_t)] - \mu^2 \phi_1 \varphi_1 + 2 \sum_{k=1}^{\infty} (E[X_tX_{t+k}f(X_t)g(X_{t+k})] - \mu^2 \phi_1 \varphi_1)
\]

\[
\overset{(A2)}{=} \mu E[f(X_t + 1)g(X_t + 1)] + \mu^2 E[f(X_t + 2)g(X_t + 2)] - \mu^2 \phi_1 \varphi_1
\]

\[
+ 2 \sum_{k=1}^{\infty} (E[X_tX_{t+k}f(X_t)g(X_{t+k})] - \mu^2 \phi_1 \varphi_1)
\]

\[
\overset{(7)}{=} \mu E[f(X_t + 1)g(X_t + 1)] + \mu^2 E[f(X_t + 2)g(X_t + 2)] - \mu^2 \phi_1 \varphi_1
\]

\[
+ 2 \sum_{k=1}^{\infty} ((1 - \rho(k))^2 \mu^2 + \rho(k) \mu) E[f(X_t + 1)g(X_{t+k} + 1)]
\]

\[
+ \rho(k)(1 - \rho(k)) \mu^2 E[f(X_t + 2)g(X_{t+k} + 1)]
\]

\[
+ \rho(k)(1 - \rho(k)) \mu^2 E[f(X_t + 1)g(X_{t+k} + 2)]
\]

\[
+ (\rho(k) \mu)^2 E[f(X_t + 2)g(X_{t+k} + 2)] - \mu^2 \phi_1 \varphi_1,
\]

which leads to the expression in Lemma 2. Next,

\[
\sigma_{34} = E[X_tf(X_t + 1)g(X_t)] - \mu \phi_1 \varphi_1 + 2 \sum_{k=1}^{\infty} (E[X_{t+k}f(X_t + 1)g(X_{t+k})] - \mu \phi_1 \varphi_1)
\]

\[
\overset{(A2)}{=} \mu E[f(X_t + 2)g(X_t + 1)] - \mu \phi_1 \varphi_1 + 2 \sum_{k=1}^{\infty} (E[X_{t+k}f(X_t + 1)g(X_{t+k})] - \mu \phi_1 \varphi_1)
\]

\[
\overset{(6)}{=} \mu E[f(X_t + 2)g(X_t + 1)] - \mu \phi_1 \varphi_1 + 2 \sum_{k=1}^{\infty} ((1 - \rho(k)) \mu E[f(X_t + 1)g(X_{t+k} + 1)]
\]

\[
+ \rho(k) \mu E[f(X_t + 2)g(X_{t+k} + 1)] - \mu \phi_1 \varphi_1.
\]

\[
\sigma_{25} \text{ is derived in the same way by exchanging the roles of } f \text{ and } g. \text{ Finally,}
\]

\[
\sigma_{35} = E[f(X_t + 1)g(X_t + 1)] - \phi_1 \varphi_1 + 2 \sum_{k=1}^{\infty} (E[f(X_{t+k} + 1)g(X_t + 1)] - \phi_1 \varphi_1).
\]
This completes the proof of Lemma 2.

A.4 Proof of Theorem 2 (sketch)

We proceed in analogy to Appendix A.2. Define the function \( h : \mathbb{R}^5 \to \mathbb{R}^2 \) such that \( h(\bar{Y}) \) is equal to the vector \( (T_{f,1}, T_{g,1}) \), where \( \bar{Y} \) denotes the mean across the vectors \( Y_t \) from (12). Then, we compute the Jacobian of \( h \) and evaluate it in the means \( (\mu, \mu \phi_1, \mu \varphi_1, \phi_1, \varphi_1) ^\top \), leading to the \( 2 \times 5 \)-matrix \( D \). With the Delta method, we conclude on asymptotic normality, where the asymptotic covariance matrix computes as \( D \Sigma D ^\top \) with \( \Sigma \) from Lemma 2. Since we are only interested in the cross-covariance, we build a quadratic form of the first row of \( D \), of \( \Sigma \), and of the last column of the transpose matrix \( D ^\top \). For \( i = 1 \), we thus get

\[
\text{Cov}[T_{f,1}, T_{g,1}] = \begin{pmatrix}
\phi_1 \varphi_1 \sigma_{11} + \phi_1 (\mu \sigma_{13} - \sigma_{12}) + \phi_1 (\mu \sigma_{15} - \sigma_{14}) + \sigma_{24} - \mu (\sigma_{25} + \sigma_{34}) + \mu^2 \sigma_{35} \\
-\phi_1 \varphi_1 \sigma_{11} + \sigma_{24} - \mu (\sigma_{25} + \sigma_{34}) + \mu^2 \sigma_{35}
\end{pmatrix}
\]

where we used that \( \mu \sigma_{13} - \sigma_{12} = -\phi_1 \sigma_{11} \) and \( \mu \sigma_{15} - \sigma_{14} = -\varphi_1 \sigma_{11} \). Using the relations (A4) between the different \( D_i \) derived in Appendix A.2, we conclude that

\[
\begin{align*}
\text{Cov}[T_{f,2}, T_{g,2}] &= \text{Cov}[T_{f,1}, T_{g,1}], \\
\text{Cov}[T_{f,3}, T_{g,3}] &= \phi_1^2 \mu^2 \cdot \text{Cov}[T_{f,1}, T_{g,1}], \\
\text{Cov}[T_{f,4}, T_{g,4}] &= \mu^2 \cdot \text{Cov}[T_{f,1}, T_{g,1}].
\end{align*}
\]

So it remains to compute the expression for \( \text{Cov}[T_{f,1}, T_{g,1}] \), which is done by plugging-in the \( \sigma \)'s from Lemma 2. For illustration, we consider the i.i.d. case, where all infinite sums \( \sum_{k=1}^{\infty} \ldots \) vanish. After some calculations, we end up with

\[
\text{Cov}[T_{f,1}, T_{g,1}] = \frac{(1 + \mu) \theta_{11} (0) - \mu \theta_{12} (0) - \mu \theta_{21} (0) + \mu \theta_{22} (0) - \phi_1 \varphi_1}{\mu \phi_1 \varphi_1},
\]

which proves (13). Note that we could easily get all other covariances \( \text{Cov}[T_{f,i}, T_{g,j}] \) for \( i \neq j \). We just have to redefine \( h \), which leads to different rows of the matrix \( D \). \( D \) has the structure

\[
D = \begin{pmatrix}
d_{1,1} & d_{1,2} & d_{1,3} & 0 & 0 \\
d_{2,1} & 0 & 0 & d_{2,4} & d_{2,5}
\end{pmatrix},
\]

where the required elements \( d_{r,s} \) can be taken from the matrices \( D_i \) in Appendix A.2.