Article

A Semiparametric Approach to Test for the Presence of INAR: Simulations and Empirical Applications

Lucio Palazzo [1] and Riccardo Ievoli [2,*]

1 Department of Political Sciences, University of Naples Federico II, 80138 Naples, Italy; lucio.palazzo@unina.it
2 Department of Chemical, Pharmaceutical and Agricultural Sciences, University of Ferrara, 44121 Ferrara, Italy
* Correspondence: riccardo.ievoli@unife.it

Abstract: The present paper explores the application of bootstrap methods in testing for serial dependence in observed driven Integer-AutoRegressive (models) considering Poisson arrivals (P-INAR). To this end, a new semiparametric and restricted bootstrap algorithm is developed to ameliorate the performance of the score-based test statistic, especially when the time series present small or moderately small lengths. The performance of the proposed bootstrap test, in terms of empirical size and power, is investigated through a simulation study even considering deviation from Poisson assumptions for innovations, i.e., overdispersion and underdispersion. Under non-Poisson innovations, the semiparametric bootstrap seems to "restore" inference, while the asymptotic test usually fails. Finally, the usefulness of this approach is shown via three empirical applications.

Keywords: semiparametric bootstrap; score test; low counts; overdispersion; underdispersion

MSC: 62F40

1. Introduction

INteger AutoRegressive models (INAR) are widely used in the case of non-negative integer-valued time series. One of the well-known specifications of INAR models involves the employment of equidispersed Poisson (hereafter P-INAR) for the model disturbances. Different contributions have been focused on testing for the presence of a (possibly unknown) serial dependence in stable P-INAR models, especially because conventional methods for continuous time series may fail.

A first contribution can be found in [1], where a test statistic, based on the score function, was proposed for the P-INAR(1) model. Then, score-based statistics were compared to other proposals (e.g., the runs test and Portmentau-type statistics) in terms of empirical size and power [2], while [3] developed generalized score-based statistics to take into account under- and equidispersion in INAR models. Recently [4], found that the (conditional) maximum likelihood ratio may be an efficient alternative compared to the score test in the P-INAR(1) models. Under the null hypothesis of non-serial dependence, i.e., $\alpha = 0$ in the P-INAR(1), these statistics can be generally approximated in large samples by well-known free parameter distributions, such as standard normal or Chi-squared.

In practice, asymptotic approximation issues can lead to poor performance in such tests. The simulations in [2,4] confirmed that the score-type tests are undersized in case of series of small or moderately small sample sizes (i.e., when $T = 50, 100$ and T is the sample size), also showing bias from the nominal level with moderately large samples (e.g., $T = 500, 1000$).

In terms of empirical power, the Monte Carlo results of [3] evidenced a considerable gap between performance in small samples (e.g., $T = 100$) and moderately large samples (e.g., $T = 500$) under a set of local alternatives ($\alpha \in \{0.01, \ldots, 0.15\}$). The presence of underdispersion and overdispersion may play a relevant role in this context, and the tests can exhibit more or less sensitivity with respect to the possible deviations from the

Poisson assumption. To overcome these issues, researchers have employed response surface regression to adjust the critical values of the tests. Moreover, this method is strongly based on arbitrary choices for both the Monte Carlo setup and the model specification.

In this paper we investigate, through both a simulation study and three empirical applications, if and how bootstrap methods can improve inference in testing serial dependence for P-INAR models. Bootstrap methods in INAR were introduced by [5,6] in the context of confidence bounds of forecasting. More recently, refs. [7,8] developed either parametric and semiparametric bootstrap methods to obtain more reliable inference in point estimation (via bootstrap-based bias correction) and confidence bounds, while [9] extended resampling methods in a more model-based forecasting perspective. From another point of view, ref. [10] proposed a parametric bootstrap procedure to test distributional assumption for INAR innovations. All authors pointed out the inconsistency of the conventional time series bootstrap, proposing methods that take into account the nature of integer-valued data in the resampling scheme.

Starting from [7], we propose a straightforward semiparametric bootstrap imposing the null hypothesis (of non-serial dependence, i.e., $\alpha = 0$) in the bootstrap data-generating process (DGP). The usage of "restricted" methods, which appears quite novel in the INAR context, can be found in [11,12] in bootstrapping linear regressions, also considering endogenous regressors. To the best of our knowledge, this is the first work proposing a semiparametric bootstrap algorithm to test for the presence of the INAR effect, especially considering time series of small or moderately small length. Bootstrap algorithms for score-based statistics have been proposed to solve other econometric issues. For instance, ref. [13] considered bootstrap methods for score-based statistics in the case of instrumental variables with possible weak instruments.

The remainder of this paper is structured as follows: the P-INAR(1) model and the score-based test statistic based on Poisson assumption are presented in Section 2. Section 3 introduces the new semiparametric bootstrap algorithm, while the results of Monte Carlo simulations are shown in Section 4. Applications of these methods on real datasets are presented in Section 5, and a general discussion is provided in Section 6. Finally, Section 7 contains some conclusions and further possible advances.

2. The Model and Score Test Statistic

Consider the following stable INAR(p) model, introduced in [14,15], defined as:

$$X_t = \sum_{i=1}^{p} \alpha_i \odot X_{t-i} + \varepsilon_t = \alpha_1 \odot X_{t-1} + \alpha_p \odot X_{t-p} \ldots + \varepsilon_t, \tag{1}$$

where $\{\epsilon_t\}$ is an i.i.d. nonnegative integer-valued having a finite mean μ_ε and variance $\sigma_\varepsilon^2 < \infty$. The processes $\alpha_i \odot X_{t-i}$ denote p mutually independent binomial thinning operators, representing a stochastic sum of i.i.d. stochastic processes (see [16] for further details).

This work focuses on testing serial dependence in one-lagged INAR models. Thus, without loss of generality, we consider throughout the rest of the paper the following stable INAR(1) [15]:

$$X_t = \alpha \odot X_{t-1} + \varepsilon_t, \tag{2}$$

where $\alpha \in (0,1)$ is the parameter of interest, also denoted as the thinning parameter. In (2), the symbol $\odot$ is the binomial thinning operator, i.e., a random sum of i.i.d. random variables $\{Y_i\}$, with $Y_i \sim \text{Ber}(\alpha)$, independent of X_t, such that $\text{E}(Y_i) = \alpha$ and $\text{Var}(Y_i) = \alpha(1-\alpha)$.

The DGP of marginal process varies according to the distribution of the innovations $\{\varepsilon_t\}$. In the case of i.i.d. $\varepsilon_t \sim Po(\lambda)$, the model is called P-INAR(1), also assuming equidispersion, i.e., $E(\varepsilon_t) = V(\varepsilon_t) = \lambda$.

Under such assumptions, parameter estimation can be conventionally carried out through Yule–Walker equations, conditional least squares and conditional maximum likelihood; see, e.g., [14,17]. In what follows, we consider the score-based test statistic for serial

dependence in P-INAR model, introduced in [1,2]. To test for the presence of the INAR(1) effect, the following system of hypothesis is considered:

$$H_0 : \alpha = 0 \qquad \text{vs.} \qquad H_1 : \alpha > 0. \tag{3}$$

where α, i.e., the parameter of interest, comes from Equation (2).

Score statistic for testing P-INAR(1) model, with parameters (α, λ), takes the following specification [2,3]:

$$S^P(\hat{\lambda}) = T^{-1/2} \frac{\sum_{t=1}^{T} (x_{t-1} - \hat{\lambda})(x_t - \hat{\lambda})}{\hat{\lambda}} \tag{4}$$

where $\hat{\lambda} = \bar{x} = T^{-1} \sum_{t=1}^{T} x_t$. The statistic in (4) converges in distribution to a standard normal [1,3].

3. Bootstrap Algorithm for Testing INAR

In this Section, a new semiparametric bootstrap method for the test statistic in Equation (4) is presented. We remark that conventional non-parametric approaches for continuous time series, e.g., block bootstrap, ref. [18] and the semiparametric autoregressive bootstrap [19] should not be applied because they do not take into account the true characteristics of integer-valued time series, leading to inconsistent results. In addition, the infeasibility of conventional methods for time series has been shown in [7].

We consider a semiparametric bootstrap for its suitability, employing a "restricted" algorithm, i.e., imposing $\alpha = 0$ in the bootstrap DGP and obtaining $\hat{\varepsilon}_t = x_t$. This restriction ensures that residuals have the same support of the innovations' DGP. In practice, the pseudo residuals are sampled from the empirical distribution function (EDF) of the restricted residuals (under the null hypothesis of $\alpha = 0$).

The following algorithm summarizes the proposed semiparametric method.

Semiparametric Bootstrap Algorithm

Given a random sample $x_1, \ldots, x_T$ of size T,

Step 1. Estimate the parameters $(\hat{\alpha}, \hat{\lambda})$ and the test statistic $\hat{S}$. Residuals can be obtained imposing $\alpha = 0$, i.e., $\hat{\varepsilon}_t = x_t$;

Step 2. Use $\hat{\varepsilon}_t$ to obtain bootstrap pseudo-residuals $\varepsilon_1^*, \ldots, \varepsilon_T^*$, i.e., $\varepsilon_t^* \sim \text{EDF}(\hat{\varepsilon}_t)$;

Step 3. Create $x_1^*, \ldots, x_T^*$, plugging the pseudo-residuals in the bootstrap DGP;

Step 4. Compute the bootstrapped score statistic

$$\hat{S}^* = S^*(\hat{\lambda}^*) = T^{-1/2} \frac{\sum_{t=1}^{T} (x_{t-1}^* - \hat{\lambda}^*)(x_t^* - \hat{\lambda}^*)}{\hat{\lambda}^*} \tag{5}$$

where $\hat{\lambda}^* = T^{-1} \sum_{t=1}^{T} x_t^*$;

Step 5. Repeat B times steps 1-4, producing $\hat{S}_1^*, \ldots, \hat{S}_b^*, \ldots, \hat{S}_B^*$;

Step 6. Obtain the bootstrap *p*-value as:

$$p^* = B^{-1} \sum_{b=1}^{B} I(|\hat{S}_b^*| > |\hat{S}|)$$

Moreover, the pseudo-residuals of Step 2 can be also obtained by using a parametric method, where the bootstrap DGP is constructed based on more specific assumptions. Specifically, for the P-INAR(1) model, the restricted residual $\{\varepsilon_t^*\}_{t=1}^{T}$ is sampled from a Poisson distribution with parameter equal to the estimate of λ. To summarize, it is assumed that $\varepsilon_t^* \sim Po(\hat{\lambda})$. A possible drawback of the parametric method in the P-INAR case is

the sensitivity with respect to deviations from Poisson assumption, especially for what concerns the degree of dispersion.

4. Simulation Study

In this Section, a simulation study is performed to assess the proposed methodology via the comparison between the semiparametric bootstrap and the parametric bootstrap, using the asymptotic test as a benchmark.

4.1. Setup

The finite sample behaviour of the bootstrap-based score test, illustrated in the previous Section, was analysed by generating M = 10,000 samples according to the following DGP:

$$x_t = \alpha \odot x_{t-1} + \varepsilon_t,$$

where $\varepsilon_t \sim Po(\lambda)$ considering the following alternative parameter settings: $\lambda = \{2, 5, 10\}$. Different sample sizes were used for the simulations, such that $T = \{50, 75, 100, 250, 500\}$, while the considered nominal level for the test was 0.05. To generate Monte Carlo samples, a pre-run of 500 observations was carried out. The empirical size of bootstrap-based statistic $\hat{S}^*$ was evaluated under $\alpha = 0$, and an increasing sequence of α by 0.05 (starting from $\alpha = 0$) was considered for the empirical power, stopping at $\alpha = 0.8$ to avoid the near-unit root situation [20]. The number of replications used to compute the bootstrap p-values was set equal to $B = 999$. We also computed empirical rejection frequencies both for the parametric bootstrap illustrated in Section 3 and asymptotic rejection frequencies for the score statistic. Performance was evaluated through both the empirical size and the empirical power.

Finally, computational times were evaluated to show the straightforward applicability of the proposed bootstrap test. Time series of lengths ranging from 50 to 500, increasing by 50, were considered.

Deviations from Poisson Assumptions

We firstly evaluated the presence of overdispersion in the innovation process. In this regard, the simulated DGP follows a negative binomial distribution, i.e., $\varepsilon_t \sim NB(r, p)$. Given the (Fisher) index of dispersion, defined as the ratio between variance and mean of the series, $I_d = \sigma_\varepsilon^2 / \mu_\varepsilon$, we considered three following cases inspired by the design of [2]:

- **Small overdispersion,** considering $\{r = 10, p = 5/6\}$ such that $I_d = 1.2$;
- **Moderate overdispersion,** with $\{r = 4, p = 2/3\}$ resulting in $I_d = 1.5$;
- **High overdispersion,** with $\{r = 1, p = 1/3\}$ and $I_d = 3.0$.

In both cases, expected values of ε_t are equal to 2.

Then, we consider three cases of under-dispersion using a binomial distribution, i.e., $\varepsilon_t \sim Bin(n, p)$, with the three following parametrisations:

- **Small underdispersion,** considering $\{n = 2, p = 0.1\}$ such that $I_d = 0.9$;
- **Moderate underdispersion,** with $\{n = 2, p = 0.5\}$ and $I_d = 0.5$;
- **High underdispersion,** with $\{n = 2, p = 0.7\}$ and $I_d = 0.3$.

The number of Monte Carlo simulations and bootstrap iterations were equal to those considered for the Poisson-based DGP.

4.2. Main Results

We start from the Poisson case (equidispersion). Table 1 summarizes the main results in terms of empirical size.

Even in the equidispersed case, the asymptotic rejection frequencies can be quite below the nominal level, especially with series of moderately small length (i.e., $T \leq 100$ and $\lambda = 2, 5, 10$). Nevertheless, the distribution of rejection frequencies obtained through semiparametric bootstrap (hereafter SPB) shows the successfulness of proposed method even with series presenting moderately small length. Indeed, the good performance of parametric bootstrap (hereafter PB), which outperforms the SPB in some simulation

scenarios, can be due to the combination of (a) the imposition of the true DGP in the simulation setup and (b) the usage of a score statistic which is specifically suited for equidispersed Poisson arrivals.

Table 1. Empirical size of the bootstrap-based score test under Poisson DGP. ASY: asymptotic, SPB: semiparametric bootstrap, PB: parametric bootstrap.

| DGP	Mean	Variance	I_d	T	ASY	SPB	PB
Po (2)	2.00	2.00	1.00	50	0.0350	0.0469	0.0501
				75	0.0356	0.0461	0.0491
				100	0.0385	0.0463	0.0555
				250	0.0445	0.0473	0.0486
				500	0.0449	0.0495	0.0483
Po (5)	5.00	5.00	1.00	50	0.0316	0.0472	0.0515
				75	0.0350	0.0444	0.0494
				100	0.0382	0.0490	0.0522
				250	0.0464	0.0509	0.0508
				500	0.0451	0.0524	0.0437
Po (10)	10.00	10.00	1.00	50	0.0329	0.0462	0.0462
				75	0.0367	0.0503	0.0496
				100	0.0361	0.0476	0.0530
				250	0.0424	0.0501	0.0477
				500	0.0458	0.0498	0.0491

Figure 1 shows the performance of the bootstrap test in terms of empirical power, considering 15 different scenarios. The overall performances of SPB and PB are comparable with respect to the asymptotic test. Although the two bootstraps exhibit a conservative trend, especially with $\alpha \leq 0.4$ when $T = 50, 75$, and $\alpha \leq 0.2$ when $T = 250$, the SPB outperforms the PB in all considered scenarios, especially when $\alpha > 0.2$ and with moderately small T ($T = 50, 75, 100$). In addition, the PB outperforms the asymptotic test in the case of moderately small series ($T = 50, 75$) and for a reasonably large α (i.e., $\alpha \geq 0.4$).
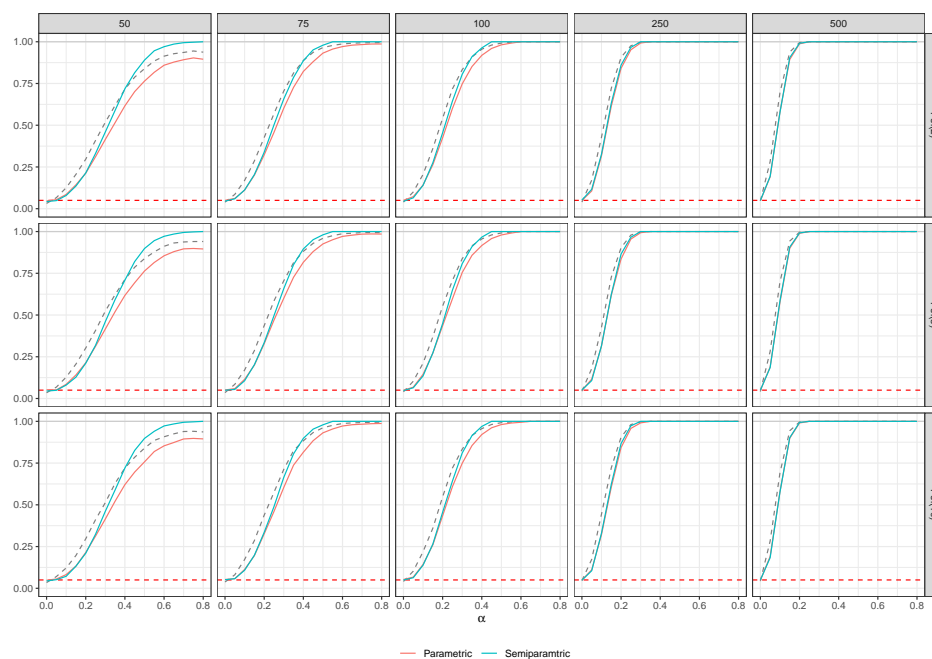
Figure 1. Empirical power of the bootstrap-based score test under Poisson DGP. The black dashed line is the empirical power of the asymptotic test, while the red dashed line represents the nominal level of 0.05.

The considered tests do not appear to be particularly sensitive with respect to different choices of the λ parameter. To conclude, in the case of Poisson innovations, both SPB and PB are reasonable choices to improve inference in testing for the presence of INAR(1).

Figure 2 illustrates the results of computational costs in terms of the median computed through Monte Carlo replications and considering the 95% quantile intervals. To summarize, the computational cost appears very satisfactory, ranging from 5 and 30 ms, while the semiparametric bootstrap outperforms the parametric one. The gap between the two methods grows as the sample size increases.
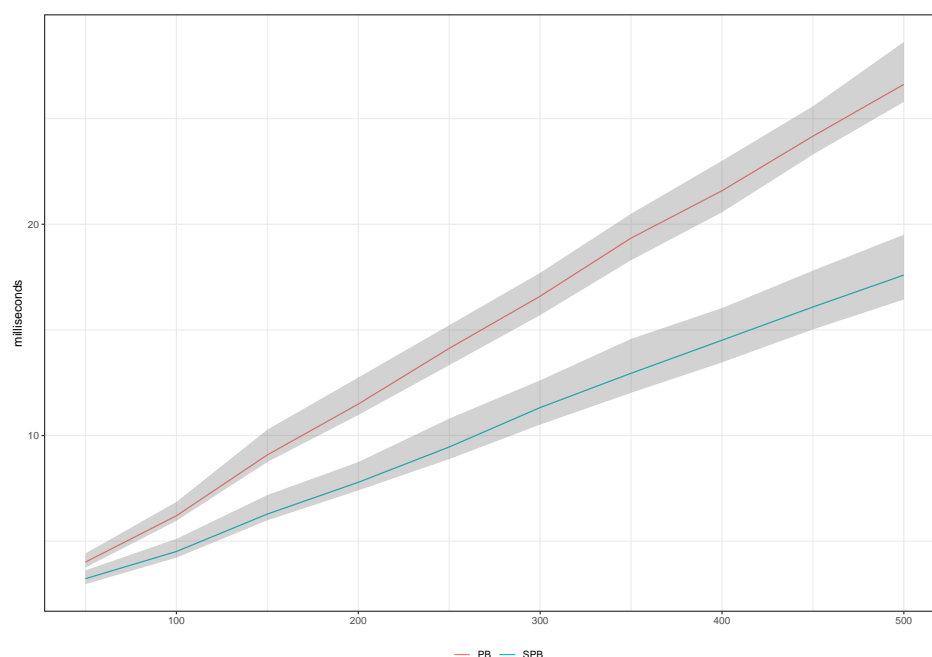
Figure 2. Plots of the median computational cost of the semiparametric (blue line) and parametric (red line) bootstrap procedures ($B = 999$), computed through Monte Carlo replications, with increasing sample size. Upper and lower bands of the grey area represent the 2.5 and the 97.5 quantiles of the distribution of the computational costs.

In the cases of DGPs deviating from Poisson assumption, the results of the tests show substantial differences. Empirical sizes of SPB and PB in the case of overdispersion are depicted in Table 2 along with the asymptotic size. Even with a low value of overdispersion ($I_d = 1.2$), the PB shows worse performance than the asymptotic test, exhibiting rejection frequencies that doubled the considered nominal value. Furthermore, in the case of either a moderate or high degree of overdispersion ($I_d = 1.5, 3.0$), both the PB and the asymptotic test are severely oversized, appearing totally unreliable. Indeed, they exhibit an increasing trend of empirical sizes as the sample length of the series increases. Surprisingly, the SPB performs well throughout the three considered scenarios, especially when T is sufficiently large.

Regarding the empirical power, illustrated in Figure 3, when the overdispersion is low ($I_d = 1.2$), the three tests show similar behaviour as the INAR parameter α increases. When T is quite large (e.g., $T = 250, 500$), the SPB, the PB, and the asymptotic test rapidly reach the unity for $\alpha \geq 0.2$. However, severe overdispersion ($I_d = 3.0$) leads to a similar behaviour for the PB and the asymptotic test, producing unreliable over-rejections even for small values of α. Conversely, the SPB, which is generally dominated by both the PB and the asymptotic test, presents a behaviour that is compared to the case of equidispersed Poisson innovations.

Table 2. Empirical size of the bootstrap-based score test under negative binomial DGP. ASY: asymptotic, SPB: semiparametric bootstrap, PB: parametric bootstrap.

X	E(X)	Var(X)	I_d	T	ASY	SPB	PB
NB (10, 5/6)	2.00	2.40	1.20	50	0.0567	0.0467	0.0958
				75	0.0642	0.0466	0.0924
				100	0.0638	0.0485	0.1004
				250	0.0719	0.0485	0.1024
				500	0.0722	0.0492	0.0971
NB (4, 2/3)	2.00	3.00	1.50	50	0.0934	0.0440	0.1776
				75	0.1023	0.0474	0.1810
				100	0.1069	0.0486	0.1825
				250	0.1177	0.0504	0.1899
				500	0.1252	0.0505	0.1911
NB (1, 1/3)	2.00	6.00	3.00	50	0.2029	0.0435	0.4541
				75	0.2228	0.0438	0.4698
				100	0.2306	0.0483	0.4881
				250	0.2571	0.0511	0.5021
				500	0.2620	0.0518	0.5080

Considering underdispersed innovations (e.g., when they follow a binomial distribution), the PB and the asymptotic tests appear useless once again. Table 3 illustrates how the PB and the asymptotic test are both quite undersized even with slight underdispersion ($I_d = 0.9$). Therefore, when $I_d = 0.5$ and $I_d = 0.3$, the rejection frequencies are practically equal to 0 for each considered T. As in the case of overdispersed innovations, rejection frequencies of SPB are distributed around the nominal level of 0.05.
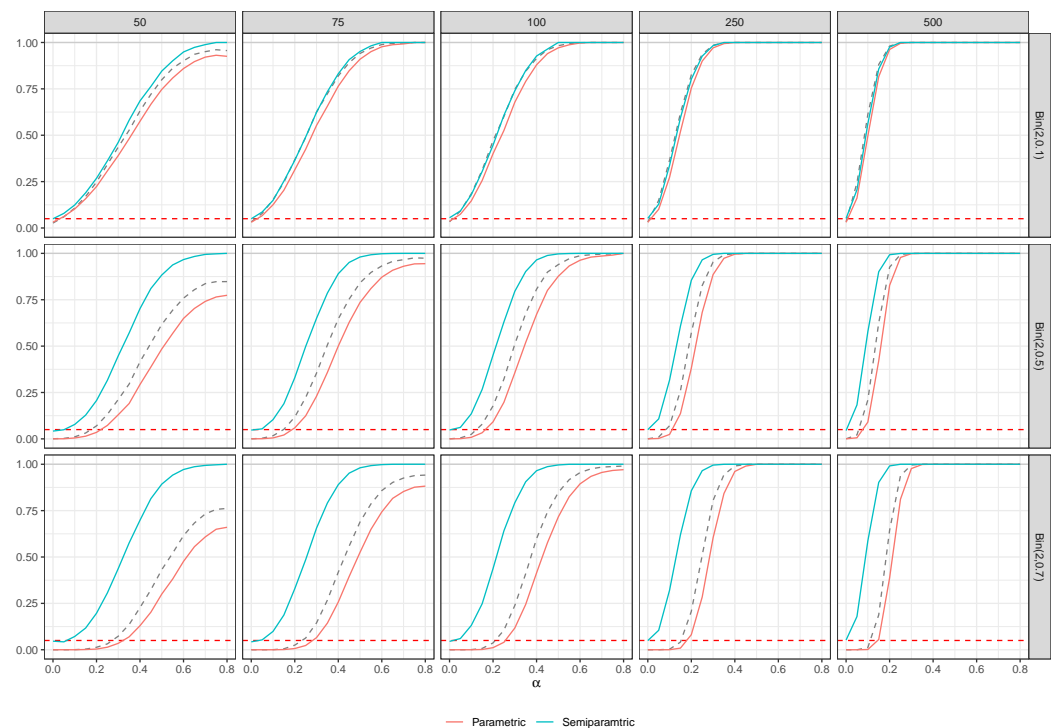
Table 3. Empirical size of the bootstrap-based score test under binomial DGP. ASY: asymptotic, SPB: semiparametric bootstrap, PB: parametric bootstrap.

X	E(X)	Var(X)	I_d	T	ASY	SPB	PB
Bin (2, 0.1)	0.20	0.18	0.90	50	0.0251	0.0523	0.0323
				75	0.0314	0.0499	0.0341
				100	0.0308	0.0489	0.0323
				250	0.0353	0.0524	0.0312
				500	0.0356	0.0500	0.0320
Bin (2, 0.5)	1.00	0.50	0.50	50	0.0004	0.0472	0.0005
				75	0.0005	0.0465	0.0003
				100	0.0006	0.0427	0.0001
				250	0.0006	0.0493	0.0002
				500	0.0003	0.0512	0.0000
Bin (2, 0.7)	1.40	0.42	0.30	50	0.0000	0.0421	0.0000
				75	0.0000	0.0478	0.0000
				100	0.0000	0.0494	0.0000
				250	0.0000	0.0473	0.0000
				500	0.0000	0.0471	0.0000

The empirical power in the case of binomial distribution of the innovations is summarized in Figure 4. Considering slight underdispersion ($I_d = 0.9$), the SPB, the PB, and the asymptotic test share a similar behaviour: when $T \geq 100$, the rejection frequencies are practically stackable. Moreover, when the underdispersion is moderate or severe ($I_d = 0.5, 0.3$), the PB and the asymptotic test suffer from the under-rejection, as already seen in the empirical size. Thus, the PB seems to perform worse than the asymptotic test, while the SPB confirms its apparent insensitivity with respect to the deviations from the equidispersion. In addition, the SPB is more powerful with respect to both the PB and the asymptotic test.
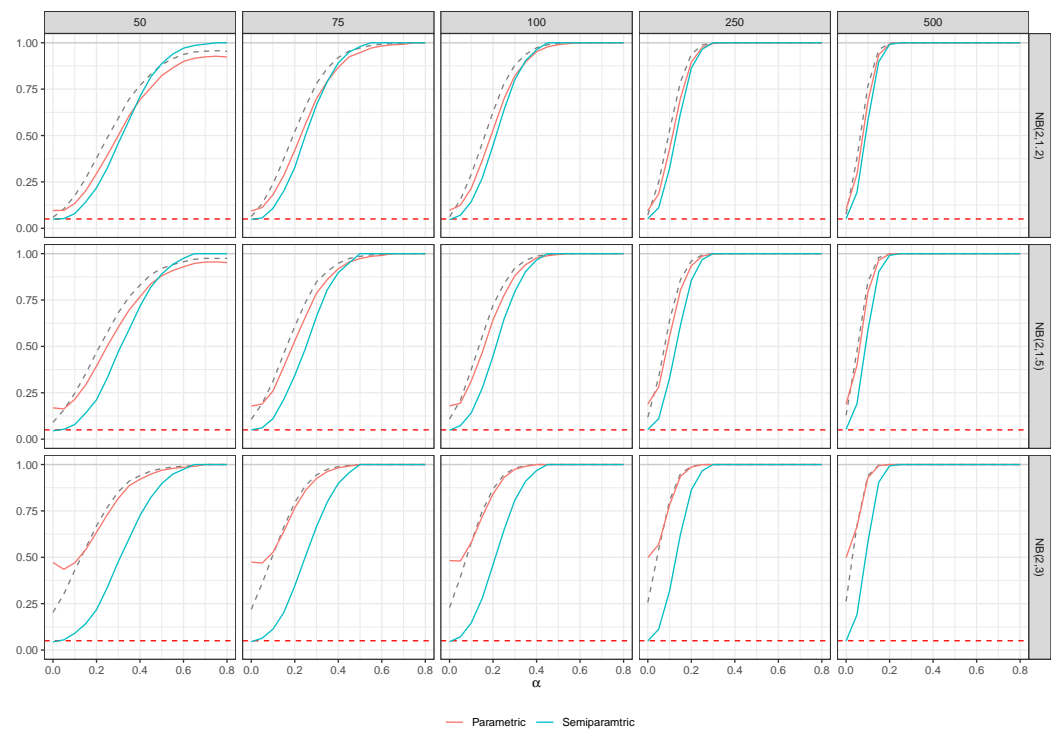
Figure 3. Empirical power of the bootstrap-based score test under negative binomial DGP. The black dashed line is the empirical power of the asymptotic test, while the red dashed line represents the nominal level of 0.05.


Figure 4. Empirical power of the bootstrap-based score test under binomial DGP. The black dashed line is the empirical power of the asymptotic test, while the red dashed line represents the nominal level of 0.05.

5. Empirical Applications

Here, the proposed SPB is applied to three case studies.

5.1. Independent Counts: Scored Goals by a Football Team

The first example concerns the series of scored goals of a football team, representing a reasonable case of Poisson time series without persistence in time. Scored goals have been previously used as a data example in the estimation of bivariate INAR [21–23], modelling scored goals in the first and in the second half. Our data include scored goals by the Arsenal Football team in the English Premier League between Season 2009-2010 and Season 2018-2019 (10 Seasons), for a total of $T = 380$ matches (Figure 5).

The plot of the series can be found in Figure 5. Descriptive statistics show that the average number of goals is $\bar{x} = 1.92$, and the estimated dispersion index is $\hat{I}_d = 1.02$.
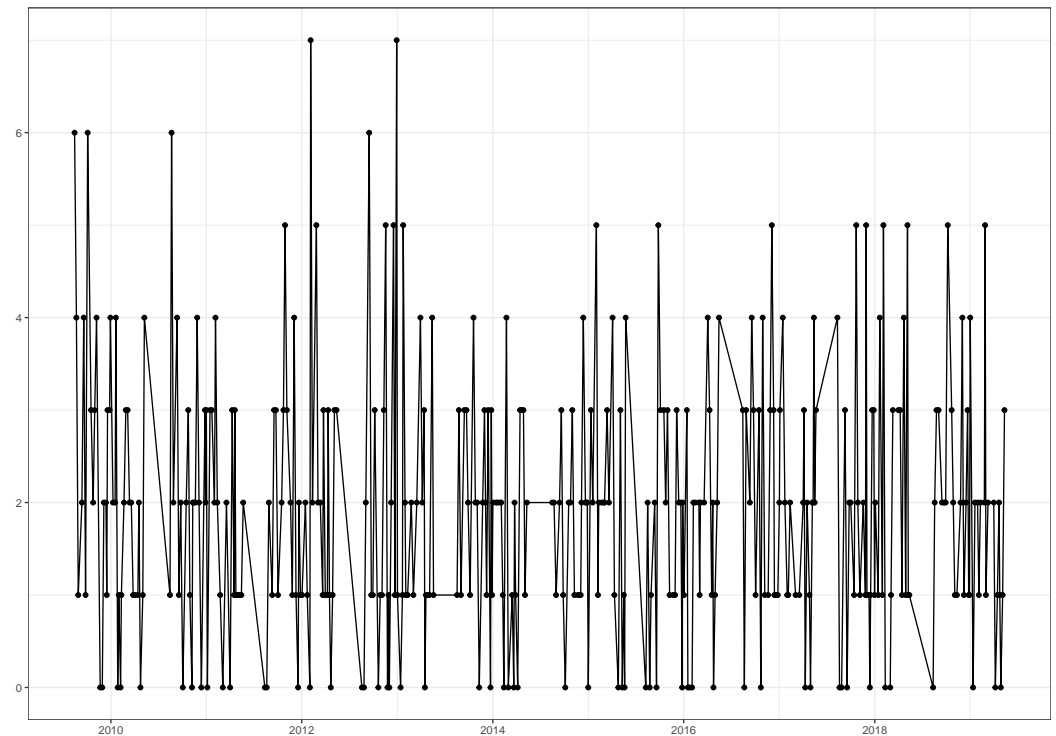
Figure 5. Time series of scored goals by Arsenal in English Premier League between Season 2009–2010 and Season 2018–2019.

5.2. INAR (1) with Equidispersed Poisson Innovations: IPs Data

To introduce Poisson INAR-based control charts, Weiss [24] presented a count of different IP addresses registered by the Department of Statistics of the University of Wurzburg. The data were collected in eight hours on 29 November 2005 and are available from [25], see Figure 6. The time unit is equal to 2 min, and the length of the series is $T = 241$.

Descriptive statistics show that the average number of IP counts is $\bar{x} = 1.32$, and the estimated dispersion index is again close to the unity, $\hat{I}_d = 1.06$. According to the Yule–Walker estimation, the estimated thinning parameter is $\hat{\alpha} = 0.22$.

5.3. INAR (1) with Possible Overdispersion: Strikes Data

Finally, we also consider a dataset of 108 monthly work major stoppages in U.S between 1994 and 2002. This dataset has been considered in many contributions regarding integer-valued time series models [25–28]. The data are illustrated in Figure 7. The mean is close to 5 strikes per month, while the estimated dispersion index is $\hat{I}_d = 1.60$, suggesting the presence of overdispersion. The estimated thinning parameter is $\hat{\alpha} = 0.57$.
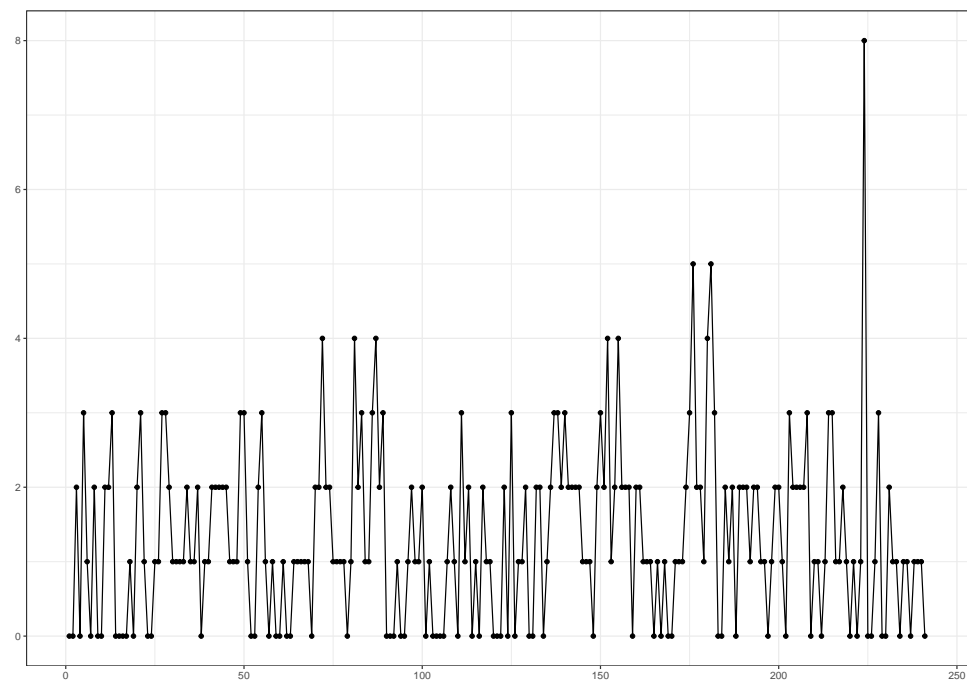
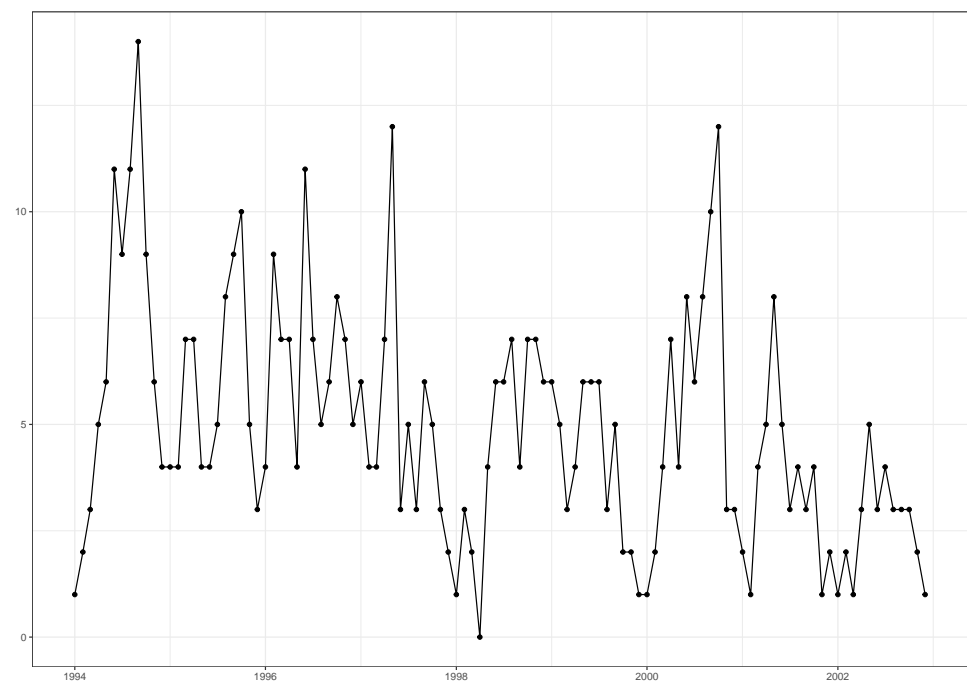**Figure 6.** Time series of IP addresses.

Figure 7. Montly time series of major strikes in U.S. between 1994 and 2002.

### *5.4. Main Results of Applications*

Firstly, the estimated autocorrelation and partial autocorrelation functions (ACF and PACF, respectively) of the mentioned dataset are depicted in Figures 8–10. Table 4 summarizes descriptive statistics and results of SPB, compared with PB and asymptotic test (in terms of *p*-values) where the number of iterations is set equal to $B = 99{,}999$.

Considering `scored goals`, SPB confirms the inability to reject the null hypothesis. Moreover, it is possible to appreciate lower bootstrap-based *p*-values (SPB and BB) when compared with the asymptotic one. In the case of the IP dataset, the SPB suggests to reject the null hypothesis considering a nominal level of 0.05 (but even lower), but the bootstrap-

based *p*-value is greater than the one obtained with the PB and than the asymptotic *p*-value. This result is in line with the empirical power observed in the simulation section in the case of INAR(1) with Poisson innovations. For the last dataset (Strikes), all the *p*-values are practically equal to zero since the estimated thinning operator appears very different from zero according to the Yule-Walker estimation. However, simulations of Section 4 raise further doubts on the reliability of the asymptotic method and the PB.
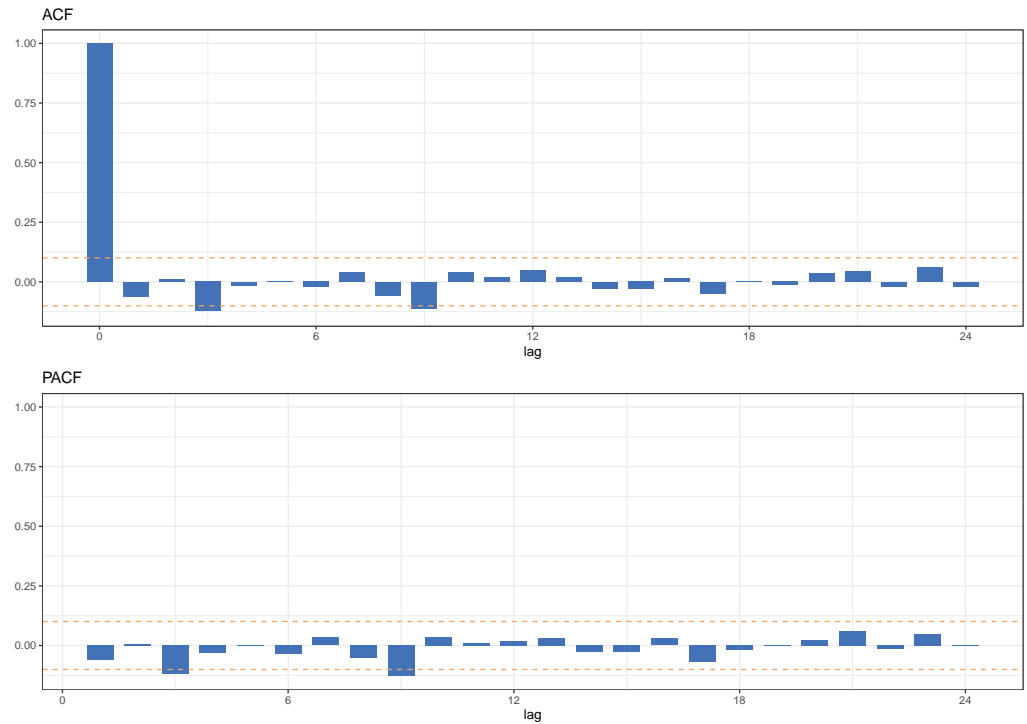
Figure 8. Estimated auto-correlations of the dataset Scored goals: (**top panel**) autocorrelation function (ACF); (**bottom panel**) partial autocorrelation function (PACF).
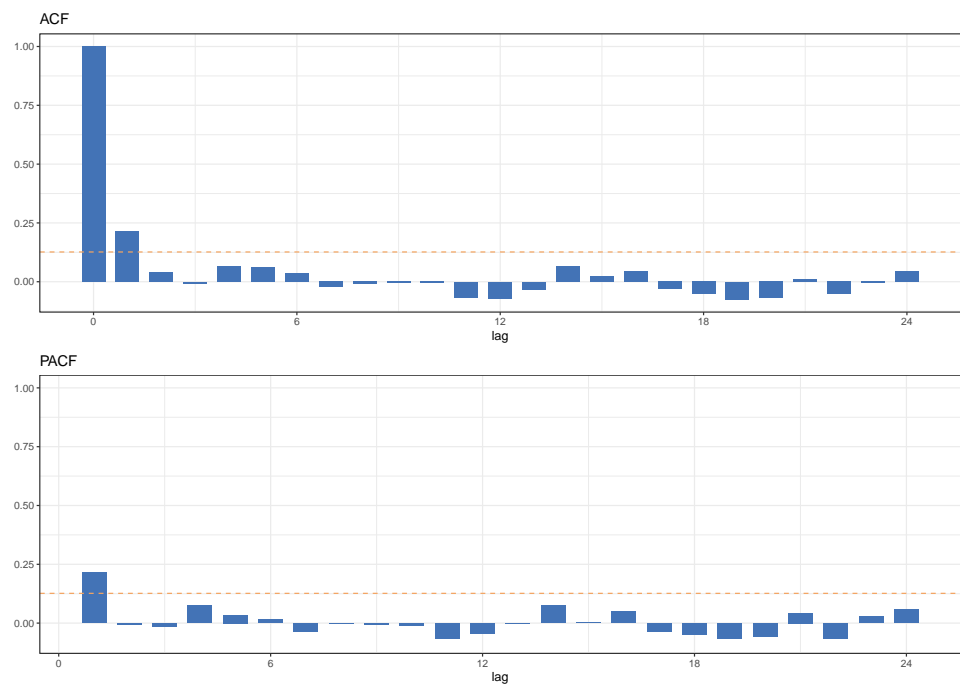
Figure 9. Estimated auto-correlations of the dataset IPs: (**top panel**) autocorrelation function (ACF); (**bottom panel**) partial autocorrelation function (PACF).
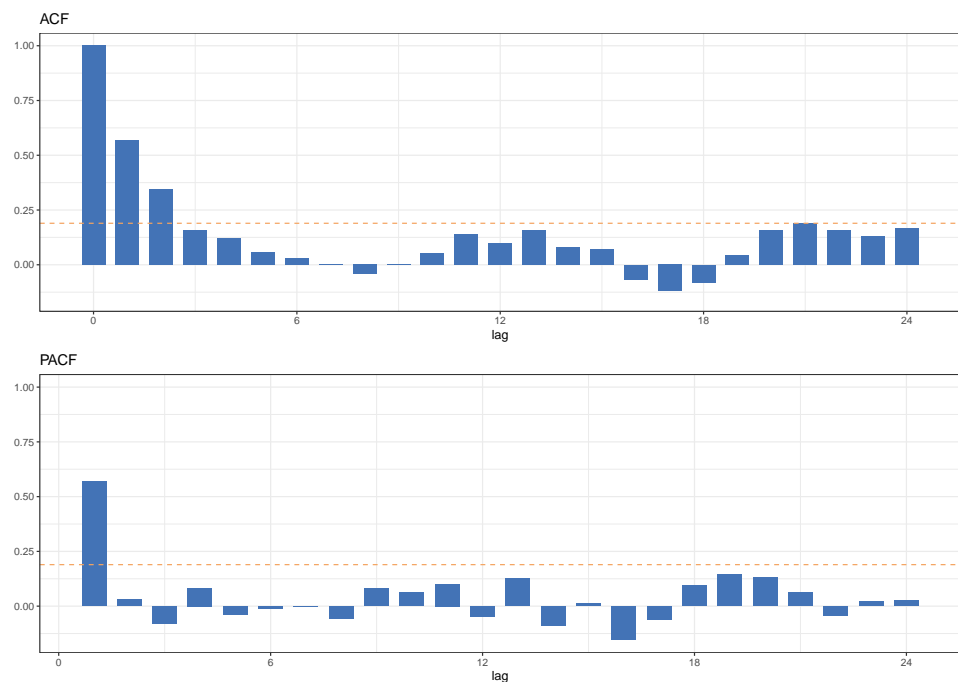
Figure 10. Estimated auto-correlations of the dataset `Strikes`: (**top panel**) autocorrelation function (ACF); (**bottom panel**) partial autocorrelation function (PACF).

Table 4. Results of empirical applications.

Dataset	$\bar{x}$	$\hat{\sigma}_x^2$	$\hat{I}_d$	$\hat{\alpha}$	ASY	SPB	PB
Scored Goals	1.92	1.96	1.02	-0.06	0.8853	0.2402	0.2284
IPs	1.32	1.39	1.06	0.22	0.0002	0.0013	0.0005
Strikes	4.94	7.92	1.60	0.57	0.0000	0.0000	0.0000

6. Discussion

The proposed semiparametric bootstrap helps to improve the performance of the score-based statistic in the case of the P-INAR model in terms of empirical size, also considering series of moderately small length. Under the i.i.d Poisson assumption for the innovations, the parametric bootstrap also exhibits excellent results due to the specific features of the simulation setup, while the satisfying performance of the semiparametric method suggests its usefulness, especially in a more generalized context (e.g., under several possible distributions for the innovations). In terms of empirical power, the semiparametric bootstrap generally dominates the parametric one.

In this regard, an analysis on the asymptotic theory will be carried out in further studies. Therefore, under i.i.d. Poisson disturbances, numerical exercises suggest that S^* may converge to a $N(0,1)$ in the bootstrap sense (i.e., conditionally on the data), which is also the limit distribution of the score-based statistic S^P [1–3]. Table 5 shows the averaged estimated moments of S^* computed using a $B = 999$ bootstrap iterations and 10,000 Monte Carlo replications in the case of $\lambda = 2$ and $\alpha = 0$, with a series of length $T = 1000$. The Jarque–Bera test is also used to check normality of S^*. The presented exercise shows how the averaged estimated moments of S^* are reasonably close to the moments of a standard Gaussian distribution, while the rejection frequencies of the Jarque–Bera test on the two statistics S^* slightly exceed the nominal value used for the normality test (0.05).

Table 5. Numerical exercise: averaged estimated moments of S^* computed with two bootstrap algorithms (PB and SPB). The last column contains the rejection frequencies of the Jarque–Bera test considering a nominal level of 0.05.

| Method	Mean	Variance	Skewness	Kurtosis	JB Test
PB	−0.031	0.998	0.014	3.011	0.063
SPB	−0.031	0.997	0.015	3.013	0.059

Moreover, previous simulation studies show that the S^P statistic can fail in case of different parametric arrivals [2,3]. This is confirmed by the simulations carried out in Section 4, while the Figure depicted in the Appendix A (Figure A1) shows how S^P is sensitive to both the degree and the type of dispersion. For instance, the (simulated) distribution of S^P under the null hypothesis is flatter under moderate overdispersion ($I_d = 1.5$), and then it less rejects $H_0 : \alpha = 0$. Under these situations, the parametric bootstrap fails since the degree of dispersion is not included in the bootstrap DGP. Thus, simulations suggest that the distribution of S^* in the case of the parametric bootstrap converges (conditional to the data and under the null hypothesis) to a standard Gaussian distribution, even when $I_d << 1$ or $I_d >> 1$. Conversely, the semiparametric algorithm is able to include the level of dispersion in the bootstrap DGP. Thus, numerical exercises employing two-sample Kolmogorov–Smirnov test show that S^* reasonably mimics the asymptotic distribution of S^P under the null hypothesis for any (finite) value of $I_d > 0$.

These arguments can be strengthened by looking at the distribution of the bootstrap p-values. Indeed, conventional bootstrap validity can be also checked when the bootstrap p-values are (asymptotically) uniformly distributed between 0 and 1 (see e.g., [29]). Figure 11 presents a comparison between (simulated) asymptotic and bootstrap p-values in the case of i.i.d. Poisson innovations and $\alpha = 0$. For both algorithms, the bootstrap p-values are close to the 45-degree line, suggesting that they are (asymptotically) uniformly distributed. Moreover, the other two subsequent figures illustrate the simulated distributions of bootstrap p-values in case of deviations from Poisson assumptions under $\alpha = 0$. In the case of moderate overdispersion (Figure 12), i.e., $\varepsilon_t \sim NB(4, 2/3)$ and $I_d = 1.5$, the parametric bootstrap p-values are systematically lower than the asymptotic ones. In addition, numerical evidence shows that they are not uniformly distributed, e.g., the mean of the p-values is not close to the expected value (i.e., 0.5), and the one sample Kolmogorov–Smirnov test rejects the null hypothesis of uniform distribution between 0 and 1. On the other hand, p-values obtained through semiparametric bootstrap are distributed around the 45 degree line, and numerical evidence shows that they are uniformly distributed (estimated mean is close to 0.5, and the Kolmogorov–Smirnov test does not reject the null hypothesis). In the case of underdispersed innovations, i.e., $\varepsilon_t \sim Bin(2, 0.5)$, an opposing behaviour can be observed (Figure 13). The parametric bootstrap p-values are always greater than the asymptotic ones and are not uniformly distributed, while the semiparametric p-values are, again, uniformly distributed and close to the 45 degree line.

A last consideration may regard the generation of underdispersed innovations. We remark that results of INAR(1) with binomial innovations (both in terms of empirical size and power) may be partially influenced by the intrinsic characteristics of the series, which involves counts that are constrained to assume few modalities, especially for small values of the thinning parameter α. Indeed, the performance of the semiparametric bootstrap is also checked using the Good distribution (see, e.g., [25,30]), also denoted as the polylogarithmic distribution, which is more appropriate to model underdispersed counts. Details of the used DGP and the results of the simulation study are presented in the Appendix B.
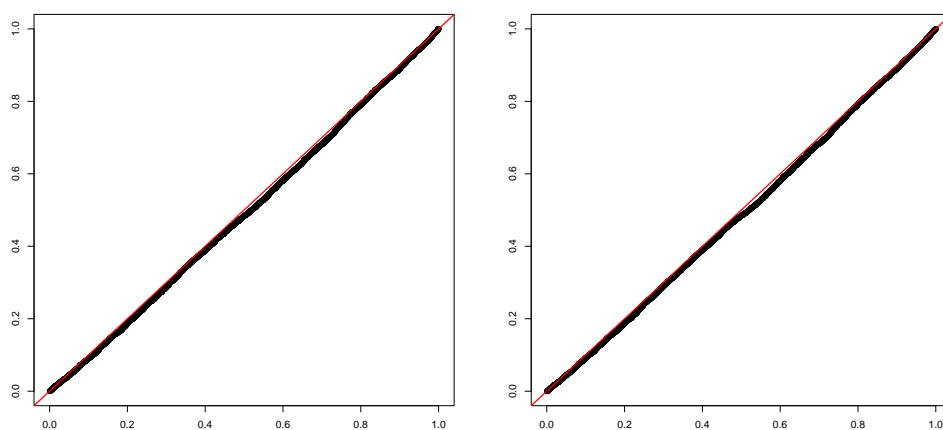
Figure 11. Comparison of asymptotic and bootstrap *p*-values under the null hypothesis and Poisson innovations: (**left panel**) semiparametric bootstrap; (**right panel**) parametric bootstrap. The red dashed line represents the 45-degree line.
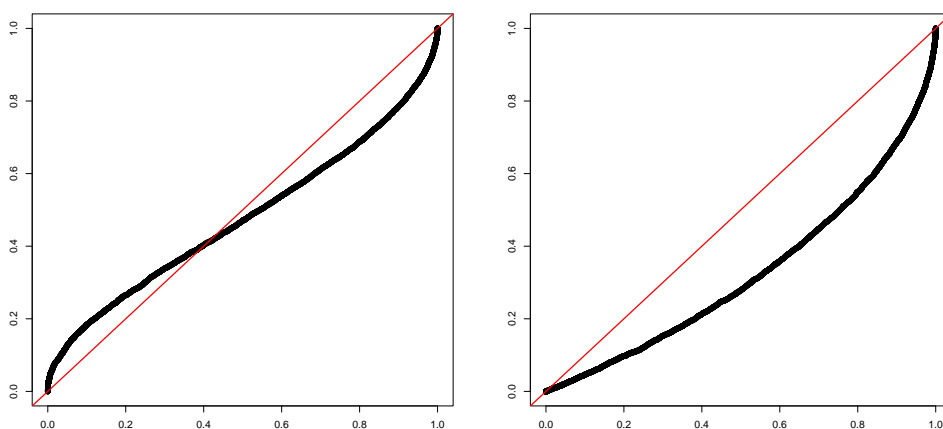
Figure 12. Comparison of asymptotic and bootstrap *p*-values under the null hypothesis and Negative Binomial innovations: (**left panel**) semiparametric bootstrap; (**right panel**) parametric bootstrap. The red dashed line represents the 45-degree line.
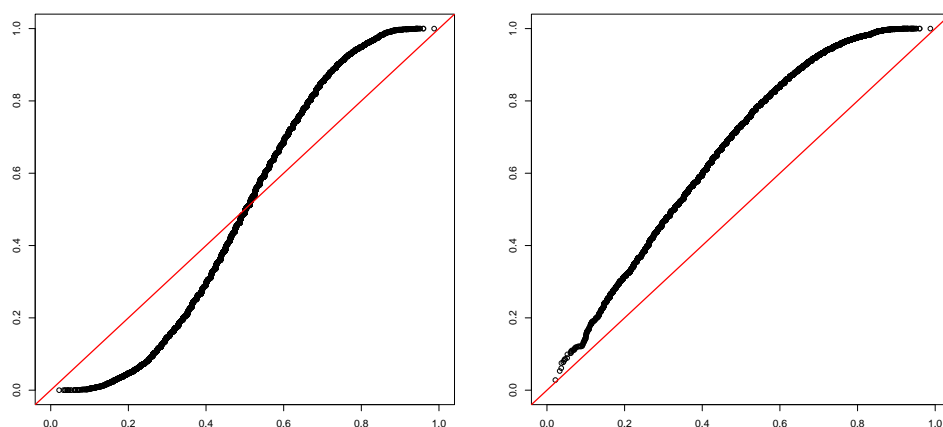
Figure 13. Comparison of asymptotic and bootstrap *p*-values under the null hypothesis and binomial innovations: (**left panel**) semiparametric bootstrap; (**right panel**) parametric bootstrap. The red dashed line represents the 45-degree line.

7. Concluding Remarks

The score-based statistic, formalized in [2,3], is a reasonable way to test for the presence of serial dependence in integer-valued time series. In the case of Poisson innovations

(P-INAR model), a semiparametric bootstrap algorithm can represent a straightforward solution to improve the performance of the test in terms of empirical size, especially with series of short (or moderately short) length. The method also shows a good performance in terms of empirical power, especially for a combination of reasonably large values of time persistence parameter and sample size. Furthermore, the parametric bootstrap represents also a possible competitor.

Considering not-equidispersed innovations, both the asymptotic test and the parametric bootstrap appear practically useless. Conversely, simulations and numerical exercises suggest that the semiparametric algorithm may be able to "restore" inference either in the case of overdispersion or underdispersion.

Further research will regard asymptotic theory to investigate the theoretical behaviour of the bootstrap-based score statistic (S^*) under both parametric and semiparametric approaches. The validity of semiparametric bootstrap in the case of dispersed innovations will be proved through a broader concept of validity occurring in the case of randomness of limit bootstrap measures [29]. In addition, the proposed bootstrap algorithm can be extended to more generalized versions of the score statistic [3], even considering possible other sources of misspecifications (e.g., zero inflation) arising in discrete time series. The applicability of score-based bootstrap test should be also investigated through the analysis of real integer-valued time series in many fields, such as finance, healthcare, and environment.

Appendix A

Figure A1. Probability density function of S^P under the null hypothesis ($\alpha = 0$, solid black line) against the $N(0, 1)$ (red dashed line): (**left panel**) moderate overdispersion ($\varepsilon_t \sim NB(10, 0.5)$); (**right panel**) moderate underdispersion ($\varepsilon_t \sim Bin(2, 0.5)$).

Appendix B

Here, we present the results of empirical size and power of the semiparametric bootstrap considering the Good distribution for the arrivals. The Good distribution is a non-negative integer-valued distribution with parameters z and s allowing for the presence of underdispersion, presenting the following probability mass function:

$$P(X = x) = \frac{z^{x+1}(x+1)^{-s}}{F(z,s)}, \tag{A1}$$

for $0 < z < 1$ and $s \in \mathbb{R}$, where $F(z,s)$ represents the following polylogarithm function:

$$F(z,s) = \sum_{n=1}^{\infty} \frac{z^n}{n^s}.$$

The Good distribution is a particular case of the Lerch three-parameter distribution [31], when $s \in \mathbb{R}$, $0 < z < 1$ and $v = 1$. Both mean and variance, as well as the moment generating function, depend on the parameters s, z and the polylogarithm function.

The results concerning the SPB-based test computed on series generated by using the Good distribution show a similar behaviour as in the case of binomial innovations. In particular, the empirical size in Table A1 of SPB is more powerful than its asymptotic counterpart. Moreover, the empirical power plots in Figure A2 again confirm the presence of an insensitivity to deviations from equidispersion, as already mentioned in the previous cases. The SPB rapidly reaches the unity, even with moderate sample size and, more in general, the SPB power curve is more powerful with respect to the asymptotic test.
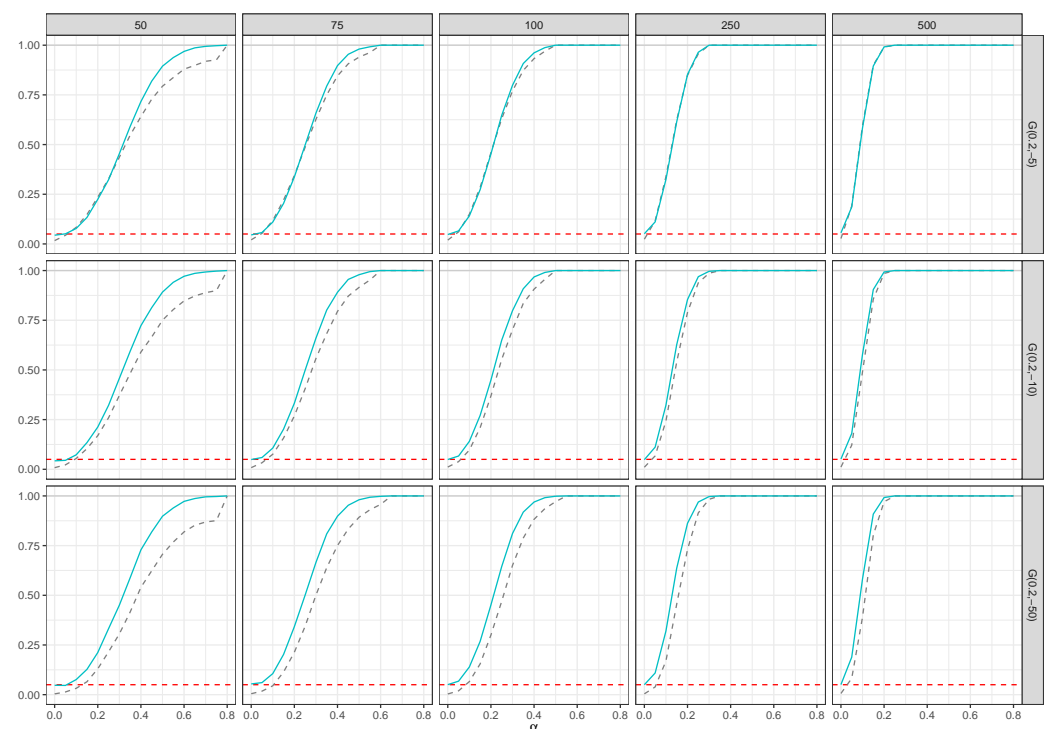
Figure A2. Empirical power of the bootstrap-based score test under Good arrivals (lightblue line) . The black dashed line is the empirical power of asymptotic test while the red dashed line represents the nominal level of 0.05.

Table A1. Empirical size of the bootstrap-based score test under Good DGP. ASY: asymptotic, SPB: semiparametric bootstrap.

X	E(X)	Var(X)	I_d	T	ASY	SPB
G (0.2;−5)	2.73	2.32	0.85	50	0.0178	0.0447
				75	0.0215	0.0481
				100	0.0226	0.0490
				250	0.0211	0.0461
				500	0.0267	0.0518
G (0.2;−10)	5.84	4.24	0.73	50	0.0089	0.0458
				75	0.0102	0.0466
				100	0.0106	0.0492
				250	0.0099	0.0492
				500	0.0117	0.0502
G (0.2;−50)	30.69	19.67	0.64	50	0.0050	0.0470
				75	0.0044	0.0477
				100	0.0037	0.0419
				250	0.0052	0.0518
				500	0.0056	0.0509

References

1. Freeland, R. Statistical Analysis of Discrete-Time Series with Applications to the Analysis of Workers Compensation Claims Data. Ph.D. Thesis, University of British Columbia, Vancouver, BC, Canada, 1998.
2. Jung, R.C.; Tremayne, A.R. Testing for serial dependence in time series models of counts. *J. Time Ser. Anal.* **2003**, *24*, 65–84. [CrossRef]
3. Sun, J.; McCabe, B.P. Score statistics for testing serial dependence in count data. *J. Time Ser. Anal.* **2013**, *34*, 315–329. [CrossRef]
4. Larsson, R. Testing for INAR effects. *Commun. Stat.-Simul. Comput.* **2019**, *49*, 1–20. [CrossRef]
5. Cardinal, M.; Roy, R.; Lambert, J. On the application of integer-valued time series models for the analysis of disease incidence. *Stat. Med.* **1999**, *18*, 2025–2039. [CrossRef]
6. Kim, H.Y.; Park, Y.S. Forecasting interval for the INAR (p) process using sieve bootstrap. In *Proceedings of the Korean Statistical Society Conference*; The Korea Institute of Science and Technology Information: Daejeon, Korea, 2005; pp. 159–165.
7. Jentsch, C.; Weiß, C.H. Bootstrapping INAR models. *Bernoulli* **2019**, *25*, 2359–2408. [CrossRef]
8. Weiß, C.; Jentsch, C. Bootstrap-based bias corrections for INAR count time series. *J. Stat. Comput. Simul.* **2019**, *89*, 1248–1264. [CrossRef]
9. Bisaglia, L.; Gerolimetto, M. Model-based INAR bootstrap for forecasting INAR (p) models. *Comput. Stat.* **2019**, *34*, 1815–1848. [CrossRef]
10. Meintanis, S.G.; Karlis, D. Validation tests for the innovation distribution in INAR time series models. *Comput. Stat.* **2014**, *29*, 1221–1241. [CrossRef]
11. Godfrey, L. *Bootstrap Tests for Regression Models*; Springer: Berlin/Heidelberg, Germany, 2009.
12. Davidson, R.; MacKinnon, J.G. Wild bootstrap tests for IV regression. *J. Bus. Econ. Stat.* **2010**, *28*, 128–144. [CrossRef]
13. Moreira, M.J.; Porter, J.R.; Suarez, G.A. Bootstrap validity for the score test when instruments may be weak. *J. Econom.* **2009**, *149*, 52–64. [CrossRef]
14. Du, J.G.; Li, Y. The Integer Valued Autoregressive (INAR(p)) model. *J. Time Ser. Anal.* **1991**, *12*, 129–142.
15. Al-Osh, M.A.; Alzaid, A.A. First–Order Integer–Valued Autoregressive (INAR (1)) Process. *J. Time Ser. Anal.* **1987**, *8*, 261–275. [CrossRef]
16. Steutel, F.W.; van Harn, K. Discrete Analogues of Self-Decomposability and Stability. *Ann. Probab.* **1979**, *7*, 893–899. [CrossRef]
17. McKenzie, E. Discrete variate time series. In *Stochastic Processes, Handbook of Statistics*; Elsevier: Amsterdam, The Netherlands, 2003; pp. 573–606. [CrossRef]
18. Politis, D.N.; Romano, J.P. A circular block-resampling procedure for stationary data. In *Exploring the Limits of Bootstrap*; John Wiley & Sons: Hoboken, NJ, USA, 1992; p. 2635270.
19. Kreiss, J.P.; Paparoditis, E.; Politis, D.N. On the range of validity of the autoregressive sieve bootstrap. *Ann. Stat.* **2011**, *39*, 2103–2130. [CrossRef]
20. Drost, F.C.; Van Den Akker, R.; Werker, B.J. The asymptotic structure of nearly unstable non-negative integer-valued AR (1) models. *Bernoulli* **2009**, *15*, 297–324. [CrossRef]
21. Jowaheer, V.; Sunecher, Y.; Khan, N.M. A non-stationary BINAR (1) process with negative binomial innovations for modeling the number of goals in the first and second half: The case of Arsenal Football Club. *Commun. Stat. Case Stud. Data Anal. Appl.* **2016**, *2*, 21–33. [CrossRef]
22. Mamode Khan, N.; Sunecher, Y.; Jowaheer, V. Modelling Football Data Using a GQL Algorithm based on Higher Ordered Covariances. *Electron. J. Appl. Stat. Anal.* **2017**, *10*, 654–665.

23. Sunecher, Y.; Khan, N.M.; Jowaheer, V.; Bourguignon, M.; Arashi, M. A Primer on a Flexible Bivariate Time Series Model for Analyzing First and Second Half Football Goal Scores: The Case of the Big 3 London Rivals in the EPL. *Ann. Data Sci.* **2019**, *6*, 531–548. [CrossRef]
24. Weiß, C.H. Controlling correlated processes of Poisson counts. *Qual. Reliab. Eng. Int.* **2007**, *23*, 741–754. [CrossRef]
25. Weiss, C.H. *An Introduction to Discrete-Valued Time Series*; John Wiley & Sons: Hoboken, NJ, USA, 2018.
26. Jung, R.C.; Ronning, G.; Tremayne, A.R. Estimation in conditional first order autoregression with discrete support. *Stat. Pap.* **2005**, *46*, 195–224. [CrossRef]
27. Weiß, C.H. The INARCH (1) model for overdispersed time series of counts. *Commun. Stat.-Simul. Comput.* **2010**, *39*, 1269–1291. [CrossRef]
28. Weiß, C.; Scherer, L.; Aleksandrov, B.; Feld, M. Checking model adequacy for count time series by using Pearson residuals. *J. Time Ser. Econom.* **2020**, *12*, 20180018. [CrossRef]
29. Cavaliere, G.; Georgiev, I. Inference under random limit bootstrap measures. *Econometrica* **2020**, *88*, 2547–2574. [CrossRef]
30. Weiß, C.H. Integer-valued autoregressive models for counts showing underdispersion. *J. Appl. Stat.* **2013**, *40*, 1931–1948. [CrossRef]
31. Zörnig, P.; Altmann, G. Unified representation of Zipf distributions. *Comput. Stat. Data Anal.* **1995**, *19*, 461–473. [CrossRef]