Communications in Statistics - Simulation and Computation

ISSN: 0361-0918 (Print) 1532-4141 (Online) Journal homepage: https://www.tandfonline.com/loi/lssp20

Testing for INAR effects

Rolf Larsson

To cite this article: Rolf Larsson (2020) Testing for INAR effects, Communications in Statistics - Simulation and Computation, 49:10, 2745-2764, DOI: 10.1080/03610918.2018.1530784

To link to this article: https://doi.org/10.1080/03610918.2018.1530784

Published online: 21 Jan 2019.

COMMUNICATIONS IN STATISTICS - SIMULATION AND COMPUTATION®
2020, VOL. 49, NO. 10, 2745–2764
https://doi.org/10.1080/03610918.2018.1530784

Testing for INAR effects

Rolf Larsson

Department of Mathematics, Uppsala University, Uppsala, Sweden

ABSTRACT
In this article, we focus on the integer valued autoregressive model, INAR (1), with Poisson innovations. We test the null of serial independence, where the INAR parameter is zero, versus the alternative of a positive INAR parameter. To this end, we propose different explicit approximations of the likelihood ratio (LR) statistic. We derive the limiting distributions of our statistics under the null. In a simulation study, we compare size and power of our tests with the score test, proposed by Sun and McCabe [2013. Score statistics for testing serial dependence in count data. Journal of Time Series Analysis 34 (3):315–29]. The size is either asymptotic or derived via response surface regressions of critical values. We find that our statistics are superior to score in terms of power and work just as well in terms of size. Another finding is that the powers of our approximate LR statistics compare well with the power of the numerical LR statistic. Power simulations are also performed under an INAR(2) framework, with similar outcome.

ARTICLE HISTORY
Received 10 April 2018
Accepted 25 September 2018

KEYWORDS
INAR model; Likelihood ratio test

1. Introduction

In recent years, integer valued autoregressive (INAR) models have gained a lot of interest. For overviews, see e.g. Weiss (2008) and Scotto, Weiss and Gouveia (2015).

The simplest integer valued autoregressive model, INAR(1), is described by the equation

$$X_t = \alpha \circ X_{t-1} + R_t, \tag{1}$$

where $0 \leq \alpha < 1, X_0 = 0$ and the error terms R_t are integer-valued and *iid*. The R_t may e.g. be assumed to follow the Poisson distribution or, to allow for over dispersion, the negative binomial distribution. The operator is called the binomial thinning operator. It is defined through

$$\alpha \circ X_t \stackrel{\text{def}}{=} \sum_{i=1}^{X_t} Y_{it},$$

where, conditional on X_t, $\{Y_{it}\}_{i=1}^{X_t}$ is a sequence of *iid* Bernoulli variables such that

$$P(Y_{it} = 1) = \alpha = 1 - P(Y_{it} = 0).$$

Conditional on all X_t, the sequences $\{Y_{it}\}_{i=1}^{X_t}$ are independent for different t.

CONTACT Rolf Larsson rolf.larsson@math.uu.se Uppsala University, Department of Mathematics, P.O. Box 480, SE-751 06 Uppsala, Sweden.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/lssp.

The parameters may be estimated by maximum likelihood, least squares or moment based methods (Yule-Walker). See Al-Osh and Alzaid (1987) for further details.

Putting $\alpha = 0$ in (1), we get a simple Poisson (or e.g. negative binomial) model. Hence, it is of interest to test the hypothesis $H_0 : \alpha = 0$ vs the alternative $H_1 : \alpha > 0$. For this purpose, Sun and McCabe (2013) derived explicit formulae for the score test statistic under different forms of innovation distributions. They also performed simulation studies to examine size and power.

Simulations regarding the score test, as well as alternative non parametric tests, were performed already by Jung and Tremayne (2003). A modified version of the score test was seen to compete very well with the non parametric tests in terms of size. Moreover, under an INAR(1) alternative, it was seen to be superior in terms of power. They also examined the power under the INAR(2) framework by Alzaid and Al-Osh (1990). Here, they found a loss of power in the case of oscillatory behavior of the ACF.

In this paper, we assume Poisson innovations and discuss the corresponding likelihood ratio (LR) test statistic. Unlike score, it does not have an explicit form. However, in the style of Larsson (2014), we derive explicit approximations of the LR statistic. We derive their asymptotic properties, and find that they need to be adjusted to become asymptotically similar (not depending on the parameters of the innovation distributions). Then, in a simulation study, we compare the new tests with the score test in terms of size and power. For size, we compare using asymptotic and response surface regression based critical values. We find that our statistics perform better than the score test in terms of power and work just as well in terms of size. Their powers also turn out well in comparison with the numerical likelihood ratio test. We also simulate power under the INAR(2) model, with similar conclusions.

The rest of the paper is as follows. In Sec. 2, we review the asymptotic properties of the score test and give a new result under the test alternative. Moreover, we present the new tests and derive the corresponding limit properties. Sec. 3 contains the simulation study, while Sec. 4 concludes. Proofs are collected in the Appendix.

2. Theoretical results

2.1. The score test

Assume that we have observations $x_1, \ldots, x_n$, and that the R_t are Poisson distributed with unknown parameter λ.

We begin by reviewing some results about the score test statistic. The statistic is given by (Freeland 1998; Sun and McCabe 2013)

$$S_n \overset{\text{def}}{=} \frac{1}{\bar{x}} \sum_{t=1}^{n} x_{t-1}(x_t - \bar{x}). \tag{2}$$

Moreover, Theorem 1 of Sun and McCabe (2013) (see also Freeland 1998, p.116) states that under $H_0 : \alpha = 0$, as $n \to \infty$,

$$n^{-1/2} S_n \xrightarrow{\mathcal{L}} U \tag{3}$$

where U is standard normal and $\xrightarrow{\mathcal{L}}$ denotes convergence in distribution.

To give some intuition for the asymptotics of the test under the alternative where α is not fixed to zero, we have the following result ($\xrightarrow{p}$ denotes convergence in probability):

Proposition 1. *For S_n as in (2), as $n \to \infty$,*

$$n^{-1} S_n \xrightarrow{p} \alpha.$$

Proof. See the Appendix. □

2.2. The likelihood ratio test

Introduce

$$p(u|v) \stackrel{\text{def}}{=} P(X_t = u | X_{t-1} = v).$$

Following Sun and McCabe (2013), the log likelihood, conditional on $X_0 = x_0$, is

$$l(\alpha, \lambda) \stackrel{\text{def}}{=} \sum_{t=1}^{n} \log p(x_t | x_{t-1}), \tag{4}$$

where, writing

$$P(R_t = k) \stackrel{\text{def}}{=} r(k) = \frac{\lambda^k}{k!} e^{-\lambda} \tag{5}$$

for $k = 0, 1, 2, \ldots, \Delta x_t \stackrel{\text{def}}{=} x_t - x_{t-1}$ and defining $a \vee b = \max(a, b)$,

$$\begin{aligned} p(x_t | x_{t-1}) &= \sum_{k=0 \vee \Delta x_t}^{x_t} P(\alpha \circ X_{t-1} = x_t - k | X_{t-1} = x_{t-1}) r(k) \\ &= \sum_{k=0 \vee \Delta x_t}^{x_t} q_{k,t}(\alpha) r(k), \end{aligned} \tag{6}$$

with

$$q_{k,t}(\alpha) \stackrel{\text{def}}{=} \binom{x_{t-1}}{x_t - k} \alpha^{x_t - k} (1 - \alpha)^{x_{t-1} - (x_t - k)}.$$

Since the log likelihood is a rather complicated function of α, it seems hard to derive the likelihood ratio (LR) test, Q_n say, explicitly. (We define Q_n as the maximum likelihood under the null divided by the maximum likelihood under the alternative hypothesis.) However, it is possible to find its limiting distribution. This was done by Freeland (1998), p.121, who stated that under H_0: $\alpha = 0$, for positive x,

$$P(-2 \log Q_n \le x) \to \frac{1}{2} + \frac{1}{2} P(Y \le x),$$

as $n \to \infty$, where Y is χ^2 distributed with one degree of freedom. (For a more general setting, see also Silvapulle and Sen 2005, chap. 4.8.)

We may reexpress this is terms of the standard normal variate U as

$$-2\log Q_n \xrightarrow{\mathcal{L}} U^2 I\{U \geq 0\}, \tag{7}$$

under H_0: $\alpha = 0$, as $n \to \infty$, where $I\{U \geq 0\} = 1$ if $U \geq 0$ and 0 otherwise.

2.3. Approximate likelihood ratio tests

In this section, we will derive approximations of $-2\log Q_n$ based on a second order Taylor expansion of the log likelihood with respect to α. We have the following result.

Proposition 2. *For the log likelihood $l(\alpha, \lambda)$ as in (4),*

$$l(\alpha, \lambda) = V_0(\lambda) + V_1(\lambda)\alpha + \frac{1}{2}V_2(\lambda)\alpha^2 + O(\alpha^3), \tag{8}$$

where

$$V_0(\lambda) \stackrel{\text{def}}{=} \sum_{t=1}^{n} \log\{r(x_t)\} = \log\lambda \sum_{t=1}^{n} x_t - \lambda n - \sum_{t=1}^{n} \log(x_t!), \tag{9}$$

$$V_1(\lambda) \stackrel{\text{def}}{=} -\sum_{t=1}^{n} x_{t-1} + \lambda^{-1}\sum_{t=1}^{n} x_t x_{t-1}, \tag{10}$$

$$\begin{aligned} V_2(\lambda) \stackrel{\text{def}}{=} & -\sum_{t=1}^{n} x_{t-1} + 2\lambda^{-1}\sum_{t=1}^{n} x_t x_{t-1} \\ & -\lambda^{-2}\left\{\sum_{t=1}^{n} x_t^2 x_{t-1}^2 - \sum_{t=2}^{n} x_t(x_t-1)x_{t-1}(x_{t-1}-1)\right\}. \end{aligned} \tag{11}$$

Proof. See the Appendix. □

In particular, observe that $V_1(\bar{x})$ is the score test given in (2).

We may use the approximation in (8) to derive an approximation of the LR test. To this end, assume for a moment that λ is known. Differentiating (8) yields

$$l'(\alpha, \lambda) = V_1(\lambda) + V_2(\lambda)\alpha + O(\alpha^2),$$

and so, the first order approximation to the solution of $l'(\alpha, \lambda) = 0$ is

$$\hat{\alpha} \stackrel{\text{def}}{=} -\frac{V_1(\lambda)}{V_2(\lambda)}. \tag{12}$$

In the following, we will refer to (12) as the approximative MLE of α.

To obtain an approximation of $-2\log Q_n$, we have that

$$-2\log Q_n = -2\{l(0, \lambda) - l(\hat{\alpha}, \lambda)\},$$

where (8) yields

$$\begin{aligned} l(0, \lambda) &= V_0(\lambda), \\ l(\hat{\alpha}, \lambda) &= V_0(\lambda) + V_1(\lambda)\hat{\alpha} + O_P(\hat{\alpha}^2), \end{aligned}$$

implying

$$-2\log Q_n = 2Z_n(\lambda) + O_P(\hat{\alpha}^2),$$

where we have the approximative LR statistic, avoiding negative estimates $\hat{\alpha}$,

$$Z_n(\lambda) \stackrel{\text{def}}{=} (\hat{\alpha}\vee 0)V_1(\lambda) = \left\{-\frac{V_1(\lambda)}{V_2(\lambda)}\vee 0\right\}V_1(\lambda), \tag{13}$$

using (12).

Treating λ as unknown, we need to plug in an estimator in (13). Preferrably, the exact maximum likelihood estimator (MLE) under H_1 should be inserted, but then we loose the advantage with having an explicit expression. So instead, our idea is to insert an approximate MLE.

We suggest two alternative ways to do this. The first, and simplest, is to replace λ by its MLE under H_0, which is $\bar{x}$. From (13), we then get the statistic

$$Z_n(\bar{x}) = \left\{-\frac{V_1(\bar{x})}{V_2(\bar{x})}\vee 0\right\}V_1(\bar{x}), \tag{14}$$

where $V_1(\bar{x})$ and $V_2(\bar{x})$ are found by inserting $\bar{x}$ for λ in Proposition 2.

The second alternative is to maximize the quadratic approximation of the log likelihood given in (8) of Proposition 2 with respect to λ, and use this as an approximation of the MLE. Unfortunately, when putting the first derivative equal to zero, this results in solving a non linear equation. Hence, a further approximation is needed. The idea here is to write $\lambda = \bar{x} + \delta$, Taylor expand around $\delta = 0$ and then solve for δ. For these details, we refer the reader to the Appendix. Here, we just give the resulting approximative MLE as

$$\hat{\lambda} = \bar{x} + \frac{B_1}{B_2}, \tag{15}$$

where

$$B_1 \stackrel{\text{def}}{=} \bar{x}\bar{x}_1\left(A_1^2 - A_2\right)(A_1 - \bar{x}), \tag{16}$$

$$B_2 \stackrel{\text{def}}{=} \left(A_2 - 2A_1\bar{x} + \bar{x}^2\right)^2 + \bar{x}\bar{x}_1\left(A_1^2 - A_2\right) - \bar{x}_1\left(A_1^2 - A_2\right)(A_1 - \bar{x}), \tag{17}$$

with $\bar{x}_1 =^{\text{def}} n^{-1}\sum_{t=1}^n x_{t-1}$,

$$A_1 \stackrel{\text{def}}{=} \frac{\sum_{t=1}^n x_t x_{t-1}}{\sum_{t=1}^n x_{t-1}}, \tag{18}$$

$$A_2 \stackrel{\text{def}}{=} \frac{\sum_{t=1}^n x_t^2 x_{t-1}^2 - \sum_{t=2}^n x_t(x_t - 1)x_{t-1}(x_{t-1} - 1)}{\sum_{t=1}^n x_{t-1}}. \tag{19}$$

We may then insert $\lambda = \hat{\lambda}$ in (13) to obtain the statistic

$$Z_n(\hat{\lambda}) \stackrel{\text{def}}{=} \left\{-\frac{V_1(\hat{\lambda})}{V_2(\hat{\lambda})}\vee 0\right\}V_1(\hat{\lambda}). \tag{20}$$

As can be seen from the following proposition, $\bar{x}$ and $\hat{\lambda}$ are not consistent for λ under H_1.

Proposition 3. *As $n \to \infty$, denoting by $\bar{X}$ the mean of $X_1, \ldots, X_n$ generated by (1),*

$$\bar{X} \xrightarrow{p} \frac{\lambda}{1-\alpha} = \lambda + \lambda\alpha + O(\alpha^2), \tag{21}$$

$$\hat{\lambda} \xrightarrow{p} \lambda + (1+3\lambda)\alpha^2 + O(\alpha^3). \tag{22}$$

Proof. See the Appendix. □

Note that for small α, $\hat{\lambda}$ has smaller asymptotic bias than $\bar{x}$.

The asymptotic distributions of $Z_n(\bar{x})$ and $Z_n(\hat{\lambda})$ are given in the following proposition.

Proposition 4. *As $n \to \infty$, under $H_0 : \alpha = 0$, with $\bar{X}$ as in proposition 3,*

$$Z_n(\bar{X}) \xrightarrow{\mathcal{L}} (\lambda+1)^{-1} U^2 I\{U \geq 0\}, \tag{23}$$

$$Z_n(\hat{\lambda}) \xrightarrow{\mathcal{L}} (\lambda+1) U^2 I\{U \geq 0\}, \tag{24}$$

where U is standard normal.

Proof. See the Appendix. □

Observe that our limit distributions are of the same form as the limit distribution of the "exact" LR test given in (7). Unfortunately however, they depend on λ. This seems to go against some general intuition. However, note that the tests are based on an expansion (cf (8)) which is not asymptotic in the sense that the higher order terms are of smaller order as n tends to infinity. They are just of smaller order in terms of the "deviation" α from the null hypothesis.

To get rid of the asymptotic dependency on λ in (23) and (24), we propose the asymptotically similar alternative statistics

$$\tilde{Z}_n \stackrel{\text{def}}{=} (\bar{x}+1) Z_n(\bar{x}) \tag{25}$$

and

$$Z_n^* \stackrel{\text{def}}{=} (\hat{\lambda}+1)^{-1} Z_n(\hat{\lambda}), \tag{26}$$

which by the Slutsky theorem both converge in distribution to $U^2 I\{U \geq 0\}$ as $n \to \infty$. In the remainder of the paper, we will concentrate on these statistics.

We may also derive asymptotic expectations of the test statistics, in the same style as in Proposition 1.

Proposition 5. *For the statistics defined in (25) and (26), as $n \to \infty$,*

$$n^{-1} \tilde{Z}_n \xrightarrow{p} \alpha^2 - \frac{\alpha^3}{\lambda} + O(\alpha^4), \tag{27}$$

$$n^{-1} Z_n^* \xrightarrow{p} \alpha^2 - \frac{1+4\lambda+5\lambda^2}{\lambda(1+\lambda)} \alpha^3 + O(\alpha^4). \tag{28}$$

Proof. See the Appendix. □

Table 1. Coefficients for response surface regression, significance level 0.01. For b coefficients with a †, we have fitted the model (30).

Statistic	a_0	a_1	a_2	b_1	b_2	b_3
Score	2.312	−5.089	0	0	0	0
$\tilde{Z}_n$	4.417	−83.17	844.4	0.1761†	−0.006558†	5.794†
$Z_n^{*1/2}$	1.960	−60.17	1700	0	0	0

Observe that to first order, the asymptotic expectations of the modified statistics (multiplied by n) are both α^2, hence no functions of λ. This should be expected, given their asymptotic similarity. By Proposition 1, the same is true for the score test statistic, where the asymptotic expectation is α. This also gives a hint that for very small α, the power might be marginally higher for the score test than for the other two. However, to compare the powers for larger α, we need to resort to simulations.

3. Finite sample simulation

3.1. Empirical size and response surface regression

For practical use, it is of course important to have reliable and easily accessed critical values. Because of proposition 4, for large n, regarding $\tilde{Z}_n$ and Z_n^* we may use the $\chi^2(1)$ distribution for this purpose. However, it turns out that the convergence to the asymptotic distribution is relatively slow, so there is a need for refinement. To this end, we propose to use critical values obtained from response surface regression. (Cf Jung and Tremayne 2003.)

All simulations are performed in Matlab R2014b.

As a basis of the response surface regression, we have run 10 000 000 replications each to find empirical critical values for the tests using
$n \in \{25, 50, 100, 200, 400, 800\}$ and $\lambda \in \{0.5, 1, 2, 5, 10, 20\}$. Then, we have regressed the so obtained critical values on various combinations of n^{-1} and λ^{-1} or λ. Based on these, by trial and error, we estimated regressions of the type

$$k_\delta = a_0 + a_1 n^{-1} + a_2 n^{-2} + b_1 \lambda^{-1} + b_2 \lambda^{-2} + b_3 n^{-1} \lambda^{-1} \tag{29}$$

or

$$k_\delta = a_0 + a_1 n^{-1} + a_2 n^{-2} + b_1 \lambda + b_2 \lambda^2 + b_3 n^{-1} \lambda, \tag{30}$$

choosing the one with the highest coefficient of determination, where k_δ is the critical value of a level δ test. We incorporated either λ^{-1} terms or λ terms, depending on which fit best. For the five different test statistics and $\delta \in \{0.01, 0.05\}$, we give the estimated coefficients of these regressions in Tables 1 and 2. Observe that we have used the square root of the Z_n^* statistic, since we got better fits for this one than for the non transformed statistic. Also observe that the response surface regressions for score and $Z_n^{*1/2}$ do not depend on the nuisance parameter λ, which is advantegeous. A third observation is that, for the asymptotically similar tests, the estimated intercepts a_0 should be close to the corresponding asymptotic values. (For the score test and $Z^{*1/2}$, these are 2.33 and 1.64, respectively, whereas for $\tilde{Z}_n$ they are the squares of these values, i.e. 5.41 and 2.71, respectively.)

Table 2. Coefficients for response surface regression, significance level 0.05. For b coefficients with a †, we have fitted the model (30).

Statistic	a_0	a_1	a_2	b_1	b_2	b_3
Score	1.619	−9.592	57.83	0	0	0
$\tilde{Z}_n$	2.298	−43.45	450.0	0.05901†	−0.002126†	2.507†
$Z_n^{*1/2}$	1.417	−28.38	485.8	0	0	0

In Tables 3–6, we give estimated sizes for the tests when using asymptotic critical values as well as critical values obtained from response surface regressions. We obtained these from new simulations with 10 000 000 replications, $n \in \{50, 200, 800\}$ and $\lambda \in \{2, 10\}$. The asymptotic critical value for the score test at level δ is given by the normal percentile, u_δ say, so that $P(U > u_\delta) = \delta$. Observe that, to mimic the practical situation, $\bar{x}$ (for the score test and $\tilde{Z}_n$) and $\hat{\lambda}$ (for Z_n^*) are inserted replicate wise for λ in the response surface equations.

We find that, except for the score test in very large samples, tests based on asymptotic critical values are undersized. The size converges most rapidly for the score test, and the convergence is decent for $\tilde{Z}_n$. For $Z_n^{*1/2}$, the asymptotic critical values converge very slowly with n. Our finding that the score test is undersized corroborates with the simulation study of Jung and Tremayne (2003).

With only a few exceptions, the response surface based critical values work well.

3.2. Size adjusted power under an INAR(1) alternative

In this section, we compare the size adjusted powers of the tests by means of simulation. We also compare to the numerical LR test.

We have chosen to study size adjusted power, and not raw power, for the following reasons: Comparisons of raw power often end up by saying that the most oversized test has the highest power, and this is a very non surprising and non informative conclusion. Also, one may imagine that size distortions already have been taken care of one way or another, for example by response surface regression (as in the previous subsection) or by bootstrap.

We simulated the size adjusted power of the three tests that we have discussed. The sample sizes are $n \in \{50, 200\}$ and the Poisson parameter $\lambda \in \{2, 10\}$. The number of replications is 5 000. The critical value comes from a simulation under the null hypothesis with the same random seed as for the simulations under all entertained alternatives.

The results are given in Figures 1–4. Except for very close to the null hypothesis, where all tests perform about equally well, we find that $\tilde{Z}_n$ works best closer to the null, but further out it is outperformed by Z_n^*. Moreover, as expected from propositions 1 and 5, close to the null hypothesis the score test is slightly better than the other tests (although this is hardly visible from the graphs). However, further away it is overall comparatively worse than $\tilde{Z}_n$ and Z_n^*. The performance difference is more pronounced for small n and large λ.

Also, note that in terms of power, the approximate LR tests perform very similar to the numerical LR test, and in fact, clearly better for small n and large λ.

3.3. Size adjusted power under an INAR(2) alternative

To see how our tests perform under higher order INAR models, we also simulated size-adjusted power under an INAR(2) assumption. The INAR(2) model may be formulated as

Table 3. Estimated sizes in per cent, nominal size 0.01, $\lambda = 2$.

Statistic	n	Asymptotic	Response surface
Score	50	0.8	1.0
	200	0.9	1.0
	800	0.9	1.0
$\tilde{Z}_n$	50	0.3	1.2
	200	0.6	1.1
	800	0.8	1.2
$Z_n^{*1/2}$	50	0.5	1.6
	200	0.0	1.0
	800	0.2	1.2

Table 4. Estimated sizes in per cent, nominal size 0.05, $\lambda = 2$.

Statistic	n	Asymptotic	Response surface
Score	50	3.4	5.0
	200	4.3	5.0
	800	4.7	5.0
$\tilde{Z}_n$	50	2.6	5.4
	200	3.7	5.2
	800	4.3	5.4
$Z_n^{*1/2}$	50	0.8	6.6
	200	1.2	4.8
	800	2.7	5.6

Table 5. Estimated sizes in per cent, nominal size 0.01, $\lambda = 10$.

Statistic	n	Asymptotic	Response surface
Score	50	0.7	1.0
	200	0.8	0.9
	800	0.9	1.0
$\tilde{Z}_n$	50	1.0	1.0
	200	0.9	0.9
	800	0.9	0.9
$Z_n^{*1/2}$	50	0.1	0.6
	200	0.0	0.7
	800	0.2	1.1

Table 6. Estimated sizes in per cent, nominal size 0.05, $\lambda = 10$.

Statistic	n	Asymptotic	Response surface
Score	50	3.3	5.0
	200	4.2	4.9
	800	4.6	5.0
$\tilde{Z}_n$	50	4.4	5.1
	200	4.4	4.8
	800	4.6	4.8
$Z_n^{*1/2}$	50	0.3	4.6
	200	1.0	4.3
	800	2.5	5.4

$$X_t = \alpha_1 \circ X_{t-1} + \alpha_2 \circ X_{t-2} + R_t, \tag{31}$$

where the $\circ$ operation and R_t is as before. As in Jung and Tremayne (2003), we use the INAR(2) specification of Alzaid and Al-Osh (1990), where$(\alpha_1 \circ X_n, \alpha_2 \circ X_n)$ given $X_n = x_n$ is trinomial with parameters $(\alpha_1, \alpha_2, x_n)$. The trinomial assumption introduces a

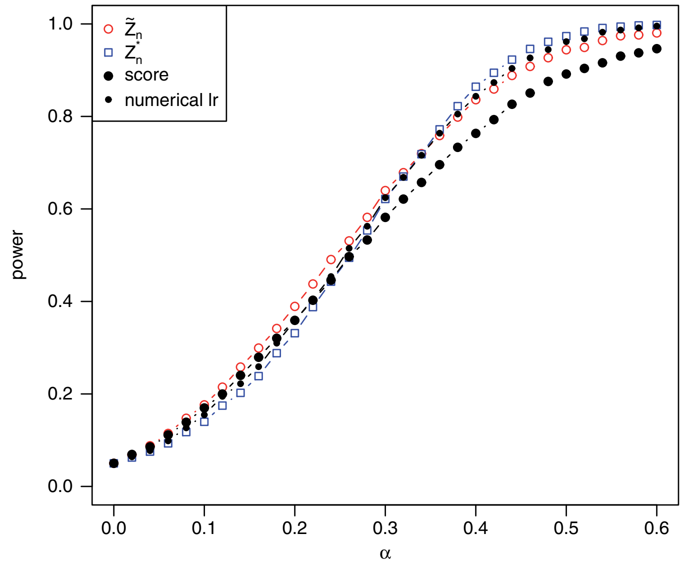

Figure 1. Simulated power, 5000 replicates, $\lambda = 2$, $n = 50$.

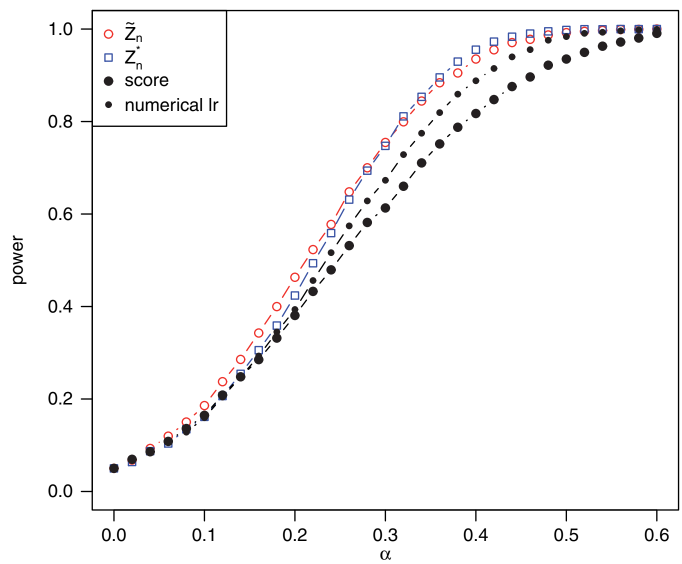

Figure 2. Simulated power, 5000 replicates, $\lambda = 10$, $n = 50$.

moving average type of dependency that makes the partial autocorrelation function (PACF) behave like for a standard ARMA process. Without this restriction, the PACF cuts off after lag two like that of an AR(2) process, see further Du and Li (1991) and Alzaid and Al-Osh (1990).

Jung and Tremayne (2003) plot the power vs $\alpha_1 + \alpha_2$. They distinguish between two cases. The first case is when $\alpha_2 < \alpha_1 - \alpha_1^2$, corresponding to an autocorrelation function (ACF) that decays exponentially to zero with increasing lag order. In the second case, where $\alpha_2 > \alpha_1 - \alpha_1^2$, the ACF damps out in an oscillatory manner. For a modified version

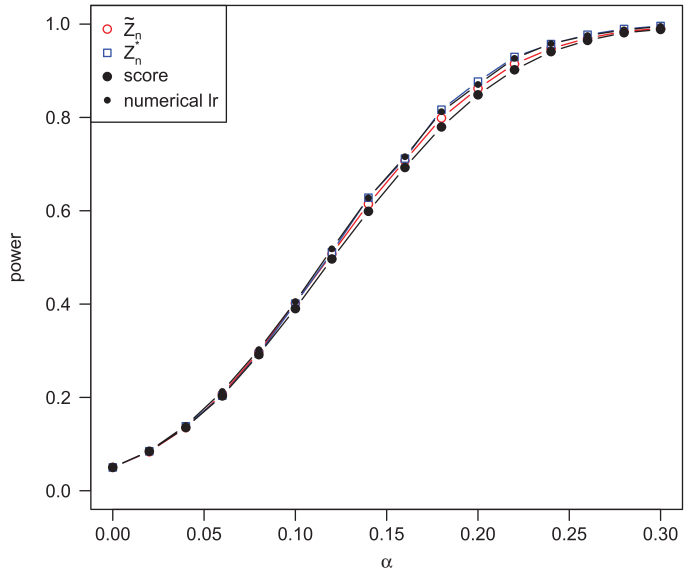

Figure 3. Simulated power, 5000 replicates, $\lambda = 2$, $n = 200$.

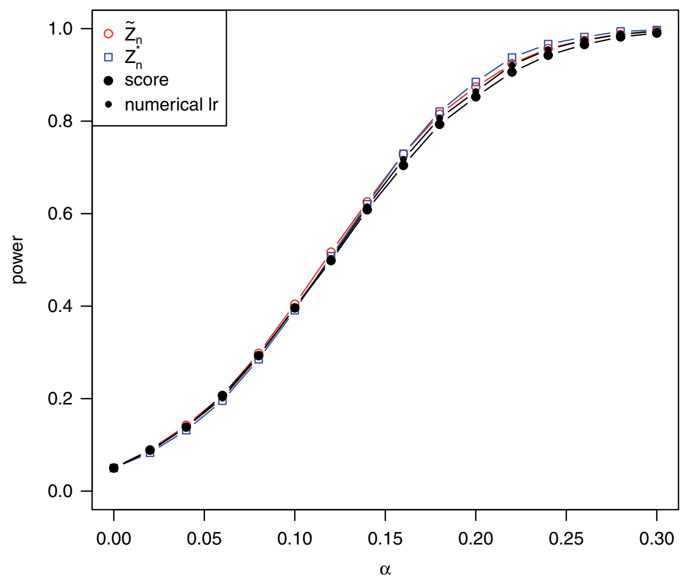

Figure 4. Simulated power, 5000 replicates, $\lambda = 10$, $n = 200$.

of the score test, they find that the power is good in the first case. In the second case, the power is lower, and for a large range of parameter values, some of the non parametric tests have better power.

However, it is not clear how α_1 and α_2 were choosen to get a specific value of $\alpha_1 + \alpha_2$. In our study, we introduce a parameter

$$\gamma = \frac{\alpha_1 - \alpha_2 - \alpha_1^2}{\alpha_1 + \alpha_2}.$$

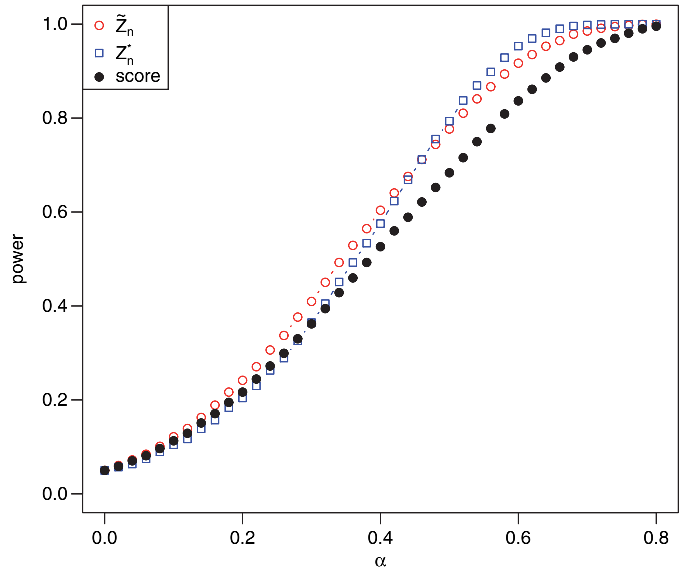

Figure 5. Simulated power, INAR(2), 20000 replicates, $\lambda = 2$, $n = 50$, $\gamma = 0.2$.

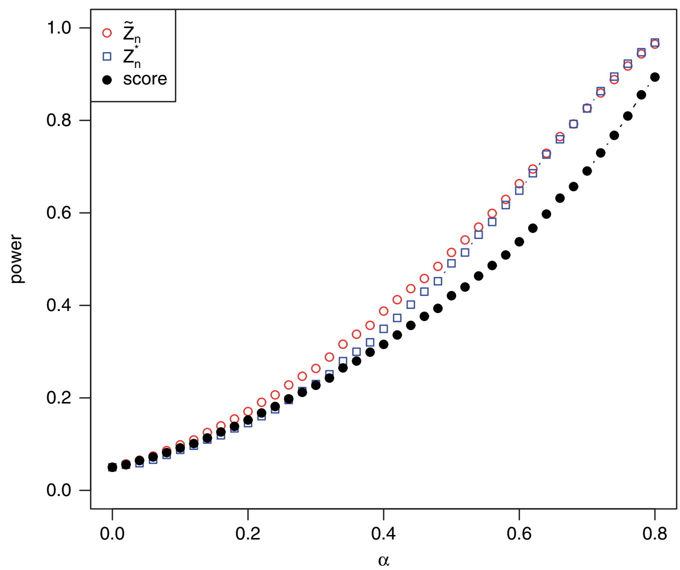

Figure 6. Simulated power, INAR(2), 20000 replicates, $\lambda = 2$, n = 50, $\gamma = -0.2$.

This means that $\gamma > 0$ corresponds to the first case above and $\gamma < 0$ corresponds to the second case. Letting $\alpha = \alpha_1 + \alpha_2$, we need to solve a non linear system to get α_1 and α_2. One solution is given by

$$\alpha_1 = 1 - \sqrt{1 - \alpha(1+\gamma)}, \quad \alpha_2 = \alpha - 1 + \sqrt{1 - \alpha(1+\gamma)}.$$

In our simulations, we have coosen $\gamma = \pm 0.2$ and $0 < \alpha_1 + \alpha_2 \leq 0.8$. This corresponds to $0 < \alpha_1 < 0.8$ for $\gamma = 0.2$ and $0 < \alpha_1 < 0.4$ for $\gamma = -0.2$. Hence, in the latter case, $\alpha_2 = \alpha - \alpha_1$ is larger in general and powers for tests desinged to be optimal for INAR(1)

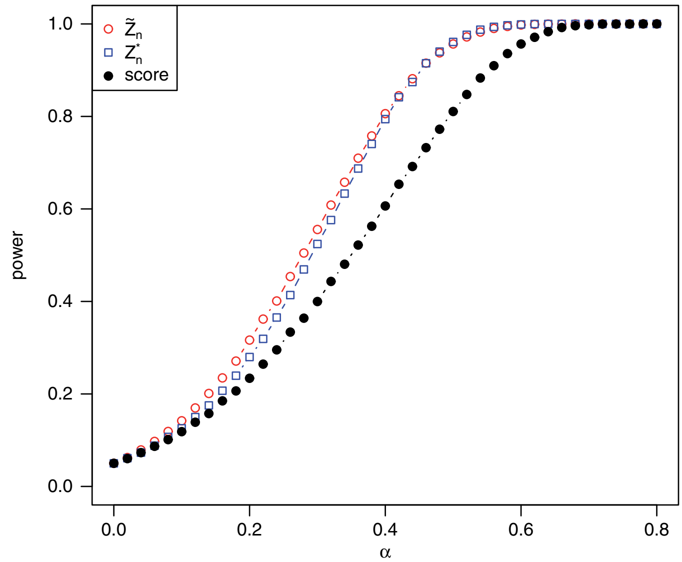

Figure 7. Simulated power, INAR(2), 20000 replicates, $\lambda = 10$, $n = 50$, $\gamma = 0.2$.

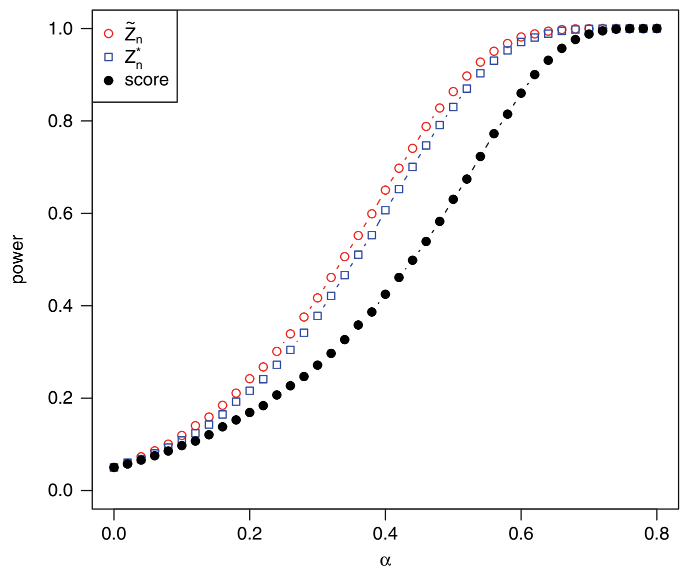

Figure 8. Simulated power, INAR(2), 20000 replicates, $\lambda = 10$, $n = 50$, $\gamma = -0.2$.

alternatives are expected to be lower. This was also what was found by Jung and Tremayne (2003) in their simulations.

The reported results, given in Figures 5–8, are from simulations with $n = 50$ and $\lambda \in \{2, 10\}$. The number of replicates is 20 000. We also ran simulations for $n = 200$, giving similar results. As expected, the power is lower for negative γ. Apart from this, much the same pattern as in the INAR(1) case is seen.

4. Concluding remarks

In this paper, we have proposed likelihood based alternatives to the score test by Sun and McCabe (2013) to test for no serial dependence in the INAR(1) model with Poisson innovations. In our simulation study, we find that when using likelihood ratio based statistics, we may gain power compared to score.

It should not be too difficult to extend our study to other types of innovation distributions that allow for over dispersion, such as negative binomial, binomial or more general distribution families like the Katz system. See further Sun and McCabe (2013) for the score test. Extensions to higher order INAR models or multivariate models would also be interesting.

Acknowledgments

I would like to thank the two anonymous referees for their careful reading and for their constructive suggestions to improve the paper.

Appendix: Proofs and derivations

According to Weiss (2008), INAR(1) processes $\{X_t\}$ may be given initial distributions such that they are strictly stationary, implying that all X_t are distributed as some random variable X, say. Define the moments of this stationary distribution by $m_k = E(X^k)$. For subsequent proofs and derivations, the following lemma will be useful. Observe that the lemma may be applied to general forms of innovation distributions, not just Poisson.

Lemma 1. *Let X_t be as in (1), where the R_t are integer-valued and iid.*

Let $\mu = E(R_t)$ and $\sigma^2 = V(R_t)$. Moreover, let $m_k = E(X^k)$ where X is distributed according to the stationary distribution of $\{X_t\}$, with a suitable choice of initial distribution. As $n \to \infty$,

$$n^{-1}\sum_{t=1}^{n} X_t \xrightarrow{p} m_1, \tag{32}$$

$$n^{-1}\sum_{t=1}^{n} X_{t-1} \xrightarrow{p} m_1, \tag{33}$$

$$n^{-1}\sum_{t=1}^{n} X_t X_{t-1} \xrightarrow{p} \mu m_1 + \alpha m_2, \tag{34}$$

$$n^{-1}\sum_{t=1}^{n} X_t X_{t-1}^2 \xrightarrow{p} \mu m_2 + \alpha m_3, \tag{35}$$

$$n^{-1}\sum_{t=1}^{n} X_t^2 X_{t-1} \xrightarrow{p} \left(\mu^2 + \sigma^2\right) m_1 + \alpha(1-\alpha + 2\mu) m_2 + \alpha^2 m_3. \tag{36}$$

Proof. Corollary 1 of Elton (1987) implies that the mean of any continuous function of a stationary Markov process with arbitrary initial distribution converges almost surely to the corresponding expectation of the stationary distribution. This implies that (32) holds. It is a trivial fact that this extends to (33).

To prove (34), at first write

$$Y_t \stackrel{\text{def}}{=} X_t - \alpha X_{t-1} - \mu. \tag{37}$$

Consequently,

$$n^{-1}\sum_{t=1}^{n} X_t X_{t-1} = n^{-1}\sum_{t=1}^{n} Y_t X_{t-1} + \alpha n^{-1}\sum_{t=1}^{n} X_{t-1}^2 + \mu n^{-1}\sum_{t=1}^{n} X_{t-1}. \tag{38}$$

Here, since $E(X_t|X_{t-1}) = \alpha X_{t-1} + \mu$, we have

$$E(Y_t X_{t-1}) = E\{E(Y_t X_{t-1}|X_{t-1})\} = E\{X_{t-1}E(Y_t|X_{t-1})\} = 0.$$

Since $Y_t X_{t-1}$ is a Markov process, it follows as above from corollary 1 of Elton (1987) that $n^{-1}\sum_{t=1}^{n} Y_t X_{t-1} \xrightarrow{p} 0$. Similarly, $n^{-1}\sum_{t=1}^{n} X_{t-1}^2 \xrightarrow{p} m_2$, and so, via (33), (38) implies that (34) holds. The proof of (35) is similar.

Finally, from (37),

$$\begin{aligned} & n^{-1}\sum_{t=1}^{n} X_t^2 X_{t-1} \\ & n^{-1}\sum_{t=1}^{n} Y_t^2 X_{t-1} + 2\alpha n^{-1}\sum_{t=1}^{n} Y_t X_{t-1}^2 + 2\mu n^{-1}\sum_{t=1}^{n} Y_t X_{t-1} \\ & +\alpha^2 n^{-1}\sum_{t=1}^{n} X_{t-1}^3 + 2\alpha\mu n^{-1}\sum_{t=1}^{n} X_{t-1}^2 + \lambda^2 n^{-1}\sum_{t=1}^{n} X_{t-1}. \end{aligned} \tag{39}$$

Here, since

$$E(Y_t^2|X_{t-1}) = V(Y_t|X_{t-1}) = \alpha(1-\alpha)X_{t-1} + \sigma^2,$$

it follows as above that

$$E(Y_t^2 X_{t-1}) = \alpha(1-\alpha)m_2 + \sigma^2 m_1,$$

implying

$$n^{-1}\sum_{t=1}^{n} Y_t^2 X_{t-1} \xrightarrow{p} \alpha(1-\alpha)m_2 + \sigma^2 m_1.$$

It is then analogous to above to derive (36) from (39). □

The rest of the appendix only concerns the Poisson case.

Lemma 2. *For a stationary INAR(1) process with Poisson innovations defined as in (1), the first three moments are*

$$\begin{aligned} m_1 &= \frac{\lambda}{1-\alpha}, \\ m_2 &= \frac{\lambda}{1-\alpha}\left(1 + \frac{\lambda}{1-\alpha}\right), \\ m_3 &= \frac{\lambda}{1-\alpha}\left\{1 + 3\frac{\lambda}{1-\alpha} + \frac{\lambda^2}{(1-\alpha)^2}\right\}. \end{aligned}$$

Proof. Weiss (2008) (example 3.3) states that if the innovations R_t are $\mathrm{Po}(\lambda)$ and if X_0 follows a $\mathrm{Po}\{\lambda/(1-\alpha)\}$ distribution, then this is also the stationary distribution. The lemma then follows from simple moment formulae for the Poisson distribution. □

Proof of proposition 1. Rewrite (2) as

$$S_n = -\sum_{t=1}^{n} X_{t-1} + \bar{X}^{-1}\sum_{t=1}^{n} X_t X_{t-1}.$$

Then, from lemma 2 and 3 and simplifications,

$$n^{-1}S_n \xrightarrow{p} -m_1 + m_1^{-1}(\lambda m_1 + \alpha m_2) = \alpha.$$

□

Proof of proposition 2. Introduce the notation (observe that $x_t - k \geq 0 \vee \Delta x_t$ is equivalent to $x_t \geq k$ and $x_{t-1} \geq k$ for all $k \geq 0$)

$$s_{0t} \stackrel{\text{def}}{=} r(x_t) I\{x_t \geq 0 \vee \Delta x_t\} = r(x_t), \tag{40}$$

$$\begin{aligned} s_{1t} &\stackrel{\text{def}}{=} r(x_t - 1) I\{x_t - 1 \geq 0 \vee \Delta x_t\} \\ &= r(x_t - 1) I\{x_t \geq 1\} I\{x_{t-1} \geq 1\}, \end{aligned} \tag{41}$$

$$\begin{aligned} s_{2t} &\stackrel{\text{def}}{=} r(x_t - 2) I\{x_t - 2 \geq 0 \vee \Delta x_t\} \\ &= r(x_t - 2) I\{x_t \geq 2\} I\{x_{t-1} \geq 2\}, \end{aligned} \tag{42}$$

where $I\{A\}$ is the indicator function of the event A. Now, spelling out the terms of the sum in (6) "backwards" and using binomial expansion,

$$\begin{aligned} &p(x_t | x_{t-1}) \\ &= (1-\alpha)^{x_{t-1}} s_{0t} + x_{t-1}\alpha(1-\alpha)^{x_{t-1}-1} s_{1t} \\ &+ \binom{x_{t-1}}{2}\alpha^2(1-\alpha)^{x_{t-1}-2} s_{2t} + O(\alpha^3) \\ &= \left\{1 - x_{t-1}\alpha + \binom{x_{t-1}}{2}\alpha^2\right\} s_{0t} + x_{t-1}\alpha\{1 - (x_{t-1} - 1)\alpha\} s_{1t} \\ &+ \binom{x_{t-1}}{2}\alpha^2 s_{2t} + O(\alpha^3) \\ &= p_{0t} + p_{1t}\alpha + p_{2t}\alpha^2 + O(\alpha^3), \end{aligned}$$

where

$$\begin{aligned} p_{0t} &\stackrel{\text{def}}{=} s_{0t}, \\ p_{1t} &\stackrel{\text{def}}{=} -x_{t-1}(s_{0t} - s_{1t}), \\ p_{2t} &\stackrel{\text{def}}{=} \binom{x_{t-1}}{2}(s_{0t} - 2s_{1t} + s_{2t}). \end{aligned}$$

Hence, inserting into (4), we find

$$\begin{aligned} l(\alpha) &= \sum_{t=1}^{n} \log\left\{p_{0t} + p_{1t}\alpha + p_{2t}\alpha^2 + O(\alpha^3)\right\} \\ &= \sum_{t=1}^{n} \log p_{0t} + \sum_{t=1}^{n} \log\left\{1 + u_{1t}\alpha + u_{2t}\alpha^2 + O(\alpha^3)\right\}, \end{aligned}$$

where $u_{it} \stackrel{\text{def}}{=} p_{it}/p_{0t}$ for $i = 1,\ 2$. Next, Taylor expanding according to $\log(1+x) = x - x^2/2 + O(x^3)$, and defining $w_{it} \stackrel{\text{def}}{=} s_{it}/s_{0t}$ for $i = 0,\ 1,\ 2$,

$$v_{1t} \stackrel{\text{def}}{=} u_{1t} = -x_{t-1} + w_{1t}x_{t-1}, \tag{43}$$

$$v_{2t} \stackrel{\text{def}}{=} 2u_{2t} - u_{1t}^2 = -x_{t-1} + 2w_{1t}x_{t-1} - w_{1t}^2 x_{t-1}^2 + w_{2t}x_{t-1}(x_{t-1}-1), \tag{44}$$

we get

$$l(\alpha) = V_0(\lambda) + V_1(\lambda)\alpha + \frac{1}{2}V_2(\lambda)\alpha^2 + O(\alpha^3), \tag{45}$$

where $V_0(\lambda) \stackrel{\text{def}}{=} \sum_{t=1}^{n} \log p_{0t} = \sum_{t=1}^{n} \log s_{0t}$ and $V_i(\lambda) \stackrel{\text{def}}{=} \sum_{t=1}^{n} v_{it}$ for $i = 1,\ 2$.

Using (5), it is easy to see that (9) follows. Furthermore, we have via (40), (41) and (5) that

$$w_{1t} = \frac{s_{1t}}{s_{0t}} = \frac{r(x_t - 1)I\{x_t \geq 1\}I\{x_{t-1} \geq 1\}}{r(x_t)} = \frac{x_t}{\lambda}I\{x_{t-1} \geq 1\}. \tag{46}$$

Thus, via (43), we obtain (10). Similarly, from (40), (42) and (5),

$$w_{2t} = \frac{x_t(x_t - 1)}{\lambda^2}I\{x_t \geq 2\}I\{x_{t-1} \geq 2\},$$

implying via (44) and (46) that (11) holds. □

Derivation of (15)-(19). Via (12), write

$$g(\lambda) \stackrel{\text{def}}{=} V_0(\lambda) + V_1(\lambda)\hat{\alpha} + \frac{1}{2}V_2(\lambda)\hat{\alpha}^2 = V_0(\lambda) - \frac{1}{2}\frac{V_1(\lambda)^2}{V_2(\lambda)}. \tag{47}$$

It follows from (10) and (11) that

$$\frac{V_1(\lambda)^2}{V_2(\lambda)} = -\sum_{t=1}^{n} x_{t-1}\frac{\left(1 - A_1\lambda^{-1}\right)^2}{1 - 2A_1\lambda^{-1} + A_2\lambda^{-2}},$$

where A_1 and A_2 are as in (18) and (19), respectively. Hence, inserting (9) and differentiating,

$$g'(\lambda) = n\left\{\bar{x}\lambda^{-1} - 1 + \bar{x}_1\frac{\left(A_1^2 - A_2\right)(A_1 - \lambda)}{\left(A_2 - 2A_1\lambda + \lambda^2\right)^2}\right\}.$$

Hence, the equation $g'(\lambda) = 0$ implies

$$0 = \left(A_2 - 2A_1\lambda + \lambda^2\right)^2(\bar{x} - \lambda) + \bar{x}_1\left(A_1^2 - A_2\right)\lambda(A_1 - \lambda). \tag{48}$$

This equation does not seem to have simple explicit solutions. However, since λ is estimated by $\bar{x}$ under the null hypothesis, it seems natural to put $\lambda = \bar{x} + \delta$, expand the right hand side of (48) to first order in δ and then solve for δ. This results in the equation

$$0 = B_1 - B_2\delta + O(\delta^2),$$

where B_1 and B_2 are given by (16) and (17), respectively. Thus, the approximative solution $\delta = B_1/B_2$ follows. □

Proof of proposition 3. Lemma 1 and 2 immedeately give (21). Equation (22) follows from (15)–(19), rewriting (19) as

$$A_2 = \frac{-\sum_{t=1}^{n} x_t x_{t-1} + \sum_{t=1}^{n} x_t x_{t-1}^2 + \sum_{t=1}^{n} x_t^2 x_{t-1}}{\sum_{t=1}^{n} x_{t-1}},$$

Lemma 1 and 2 and some tedious algebra. □

Proof of proposition 4. To find the asymptotic distribution of $Z_n(\bar{X})$, we already know the asymptotic properties of $V_1(\bar{X}) = S_n$. Moreover, inserting $\lambda = \bar{X}$ into (11), we get after some simplification

$$\begin{aligned} V_2(\bar{X}) = &-\sum_{t=1}^{n} X_{t-1} + 2\bar{X}^{-1}\sum_{t=1}^{n} X_t X_{t-1} \\ &-\bar{X}^{-2}\left\{\sum_{t=1}^{n} X_t X_{t-1}^2 + \sum_{t=1}^{n} X_t^2 X_{t-1} - \sum_{t=1}^{n} X_t X_{t-1}\right\}. \end{aligned} \tag{49}$$

Lemma 1 and 2 with $\alpha = 0$ the Slutsky theorem and simplifications yield

$$n^{-1} V_2(\bar{X}) \xrightarrow{p} -(\lambda + 1). \tag{50}$$

Thus, by the fact that $V_1(\bar{X})$ equals the score statistic S_n, (3), (14) and the Slutsky theorem,

$$Z_n(\bar{X}) \xrightarrow{\mathcal{L}} (\lambda + 1)^{-1} U^2 I\{U \geq 0\},$$

which proves (23).

Our next task is to find the asymptotic distribution of $Z_n(\hat{\lambda})$. To this end, it follows from (18) and lemma 1 and 2 that

$$A_1 = \frac{n^{-1}\sum_{t=1}^{n} X_t X_{t-1}}{n^{-1}\sum_{t=1}^{n} X_{t-1}} \xrightarrow{p} \frac{\lambda^2}{\lambda} = \lambda$$

and similarly, (19) implies

$$A_2 \xrightarrow{p} 2\lambda^2 + \lambda.$$

This, in turn, yields

$$\begin{aligned} A_1 - \bar{X} &\xrightarrow{p} 0, \\ A_1^2 - A_2 &\xrightarrow{p} -(\lambda^2 + \lambda), \\ A_2 - 2A_1\bar{X} + \bar{X}^2 &\xrightarrow{p} \lambda^2 + \lambda, \end{aligned} \tag{51}$$

and inserting into (16) and (17) and simplifying, we find

$$B_1 \xrightarrow{p} 0, \tag{52}$$

$$B_2 \xrightarrow{p} (\lambda+1)\lambda^2. \tag{53}$$

Via (15), this proves that

$$\hat{\lambda} \xrightarrow{p} \lambda, \tag{54}$$

and in particular we note that under H_0, $\hat{\lambda}$ and $\bar{x}$ are both consistent. (This may also be seen directly from proposition 3.)

To go further, we need to focus on the limit of $n^{1/2}B_1$. To this end, via (18) and (2), we at first note the simplification

$$A_1 - \bar{X} = \frac{\sum_{t=1}^n X_t X_{t-1} - n^{-1}\sum_{t=1}^n X_t \sum_{t=1}^n X_{t-1}}{\sum_{t=1}^n X_{t-1}} = \frac{\bar{X}}{n\bar{X}_1} S_n. \tag{55}$$

Now, because of (54), the limit of $V_2(\hat{\lambda})$ is as the limit of $V_2(\bar{x})$ in (50). Moreover, observe that from (16), (15) and (55),

$$\bar{X} - \hat{\lambda} = -\frac{B_1}{B_2} = -\frac{\bar{X}\bar{X}_1\left(A_1^2 - A_2\right)}{B_2}(A_1 - \bar{X}) = -\frac{\bar{X}^2\left(A_1^2 - A_2\right)}{nB_2} S_n.$$

Hence, since

$$\begin{aligned} V_1(\hat{\lambda}) &= -\sum_{t=1}^n X_{t-1} + \hat{\lambda}^{-1}\sum_{t=1}^n X_t X_{t-1} \\ &= \hat{\lambda}^{-1}\left\{\sum_{t=1}^n X_t X_{t-1} - \bar{X}\sum_{t=1}^n X_{t-1} + (\bar{X} - \hat{\lambda})\sum_{t=1}^n X_{t-1}\right\} \\ &= \hat{\lambda}^{-1}\left\{1 - \frac{\bar{X}\left(A_1^2 - A_2\right)}{nB_2}\sum_{t=1}^n X_{t-1}\right\}\bar{x}S_n, \end{aligned}$$

it follows via (51) and (53) that

$$n^{-1/2}V_1(\hat{\lambda}) \xrightarrow{\mathcal{L}} \lambda^{-1}(1+\lambda)\lambda U = (\lambda+1)U. \tag{56}$$

Hence, via (20) and (50), we finally have

$$Z_n(\hat{\lambda}) \xrightarrow{\mathcal{L}} (\lambda+1)(U \vee 0)U = (\lambda+1)U^2 I\{U \geq 0\},$$

which proves (24). □

Proof of proposition 5. Inserting $\lambda = \bar{X}$ in (11), applying lemma 1 and 2 and simplifying, we get

$$n^{-1}V_2(\bar{X}) \xrightarrow{p} -\frac{1}{(1-\alpha)\lambda}\left\{\alpha(1-\alpha)^2 + (1-\alpha)(1+2\alpha)\lambda + \lambda^2\right\}.$$

Now, observing that the right hand side is non positive for all $0 \leq \alpha < 1$, using $V_1(\bar{X}) = S_n$, proposition 1, (14) and the Slutsky theorem, we get

$$n^{-1}Z_n(\bar{X}) \xrightarrow{p} \frac{\alpha^2(1-\alpha)\lambda}{\alpha(1-\alpha)^2 + (1-\alpha)(1+2\alpha)\lambda + \lambda^2} = \frac{\alpha^2}{1+\lambda} - \frac{\alpha^3}{\lambda(1+\lambda)} + O(\alpha^4),$$

and (27) follows by the Slutsky theorem, applying (25).

Eq. (28) follows similarly, in combination with arguments from the proof of proposition 4. □

References

Al-Osh, M. A., and A. A. Alzaid. 1987. First order integer valued autoregressive (INAR(1)) process. *Journal of Time Series Analysis* 8 (3):261–75.
Alzaid, A. A., and M. A. Al-Osh. 1990. An integer-valued pth-order autoregressive structure (INAR(p)) process. *Journal of Applied Probability* 27(2):314–24.
Du, J.-G., and Y. Li. 1991. The integer-valued autoregressive (INAR(p)) model. *Journal of Time Series Analysis* 12(2):129–42.
Elton, J. H. 1987. An ergodic theorem for iterated maps. *Ergodic Theory and Dynamic Systems* 7 (4):481–8.
Freeland, R. K. 1998. Statistical Analysis of Discrete Time Series with Applications to the Analysis of Workers Compensation Claims Data. PhD thesis, The University of British Columbia, Canada.
Jung, R. C., and A. R. Tremayne. 2003. Testing for serial dependence in time series models of counts. *Journal of Time Series Analysis* 24 (1):65–84.
Larsson, R. 2014. A likelihood ratio type test for invertibility in moving average processes. *Computational Statistics and Data Analysis* 76:489–501.
Scotto, M. G., C. H. Weiss, and S. Gouveia. 2015. Thinning-based models in the analysis of integer-values time series: a review. *Statistical Modeling* 15 (6):590–618.
Silvapulle, M. J., and P. K. Sen. 2005. *Constrained statistical inference: Inequality, order, and shape restrictions*. Hoboken, NJ: Wiley.
Sun, J., and B. P. McCabe. 2013. Score statistics for testing serial dependence in count data. *Journal of Time Series Analysis* 34 (3):315–29.
Weiss, C. H. 2008. Thinning operations for modeling time series of counts - a survey. *AStA Advances in Statistical Analysis* 92:319–41.