



Semiparametric estimation of INAR models using roughness penalization

Maxime Faymonville¹ · Carsten Jentsch¹ · Christian H. Weiß² · Boris Aleksandrov²

Accepted: 17 August 2022 / Published online: 21 September 2022
© The Author(s) 2022

Abstract

Popular models for time series of count data are integer-valued autoregressive (INAR) models, for which the literature mainly deals with parametric estimation. In this regard, a semiparametric estimation approach is a remarkable exception which allows for estimation of the INAR models without any parametric assumption on the innovation distribution. However, for small sample sizes, the estimation performance of this semiparametric estimation approach may be inferior. Therefore, to improve the estimation accuracy, we propose a penalized version of the semiparametric estimation approach, which exploits the fact that the innovation distribution is often considered to be smooth, i.e. two consecutive entries of the PMF differ only slightly from each other. This is the case, for example, in the frequently used INAR models with Poisson, negative binomially or geometrically distributed innovations. For the data-driven selection of the penalization parameter, we propose two algorithms and evaluate their performance. In Monte Carlo simulations, we illustrate the superiority of the proposed penalized estimation approach and argue that a combination of penalized and unpenalized estimation approaches results in overall best INAR model fits.

Keywords Count data · Penalized estimation · Integer-valued autoregressions · Innovation distribution · Validation

Carsten Jentsch, Christian Weiß and Boris Aleksandrov contributed equally to this work.

Extended author information available on the last page of the article

1 Introduction

According to Du and Li (1991), the INAR(p) model is defined by the recursion

$$X_t = \alpha_1 \circ X_{t-1} + \alpha_2 \circ X_{t-2} + \dots + \alpha_p \circ X_{t-p} + \varepsilon_t, \quad t \in \mathbb{Z}, \tag{1}$$

with innovation process $\varepsilon_t \stackrel{\text{i.i.d.}}{\sim} G$, where the distribution G has range $\mathbb{N}_0 = \{0, 1, 2, \dots\}$. Furthermore, let $\alpha = (\alpha_1, \dots, \alpha_p)' \in (0, 1)^p$ denote the vector of model coefficients with $\sum_{i=1}^p \alpha_i < 1$ and

$$\alpha_i \circ X_{t-i} = \sum_{j=1}^{X_{t-i}} Z_j^{(t,i)},$$

where “ \circ ” is the binomial thinning operator first introduced by Steutel and Van Harn (1979). Here, $(Z_j^{(t,i)}, j \in \mathbb{N}, t \in \mathbb{Z}), i = 1, \dots, p$, are mutually independent Bernoulli-distributed random variables $Z_j^{(t,i)} \sim \text{Bin}(1, \alpha_i)$ with $P(Z_j^{(t,i)} = 1) = \alpha_i$ independent of $(\varepsilon_t, t \in \mathbb{Z})$. The special case $p = 1$ results in the INAR(1) model introduced by McKenzie (1985) and Al-Osh and Alzaid (1987). All the thinning operations “ \circ ” are independent of each other and of $\varepsilon_t, t \in \mathbb{Z}$. Furthermore, the thinning operation at time t and ε_t are independent of $X_s, s < t$.

Most researchers deal with parametric estimation of INAR models (see for example Franke and Seligmann (1993), Freeland and McCabe (2005), Brännäs and Hellström (2001) and Jung et al. (2005)), i.e. they assume G to lie in some parametric class of distributions $(G_\theta \mid \theta \in \Theta \subset \mathbb{R}^q)$ for some finite $q \in \mathbb{N}$. In contrast, Drost et al. (2009) introduced a semiparametric estimator, which on the one hand keeps the parametric assumption of the binomial thinning operation, but on the other hand allows to estimate the innovation distribution nonparametrically. Using empirical process theory, they derive asymptotic theory in terms of consistency and asymptotic normality results and proved efficiency. Consequently, their estimation approach does not require any parametric assumption regarding the innovation distribution, and avoids the risk of a falsely specified parametric assumption and its undesirable consequences. The approach estimates the coefficients of INAR models and the innovation distribution simultaneously. The resulting semiparametric maximum likelihood estimator

$$(\hat{\alpha}_{sp}, \hat{G}_{sp}) = (\hat{\alpha}_{sp,1}, \dots, \hat{\alpha}_{sp,p}, \hat{G}_{sp}(0), \hat{G}_{sp}(1), \hat{G}_{sp}(2), \dots),$$

where $\hat{\alpha}_{sp} = (\hat{\alpha}_{sp,1}, \dots, \hat{\alpha}_{sp,p})$ denotes the vector of the estimated INAR coefficients and $\{\hat{G}_{sp}(k), k \in \mathbb{N}_0\}$ are the estimated entries of the probability mass function (PMF) of G , maximizes the conditional log-likelihood function $\log(\mathcal{L}(\alpha, G))$, i.e.

$$\forall n \in \mathbb{Z}_+ : (\hat{\alpha}_{sp}, \hat{G}_{sp}) \in \arg \max_{(\alpha, G) \in [0, 1]^p \times \tilde{\mathcal{G}}} \left(\prod_{t=0}^n P_{(X_{t-1}, \dots, X_{t-p}), X_t}^{\alpha, G} \right). \tag{2}$$

Here, $\tilde{\mathcal{G}}$ is the set of all probability measures on \mathbb{Z}_+ and $P_{(X_{t-1}, \dots, X_{t-p}), X_t}^{\alpha, G}$ are the transition probabilities under the true model parameters α and G , i.e.

$$\begin{aligned} P_{(X_{t-1}, \dots, X_{t-p}), X_t}^{\alpha, G} &= \mathbb{P}_{\alpha, G} \left(\sum_{i=1}^p \alpha_i \circ X_{t-i} + \varepsilon_t = x_t \mid X_{t-1} = x_{t-1}, \dots, X_{t-p} = x_{t-p} \right) \\ &= (\text{Bin}(x_{t-1}, \alpha_1) * \dots * \text{Bin}(x_{t-p}, \alpha_p) * G)\{x_t\}, \end{aligned}$$

with \mathbb{P} the underlying probability measure and “ $*$ ” denoting the convolution of distributions. In the special case of an INAR(1) model the transition probabilities are given by

$$\mathbb{P}_{\alpha, G}(X_t = x_t \mid X_{t-1} = x_{t-1}) = \sum_{j=0}^{\min(x_t, x_{t-1})} \binom{x_{t-1}}{j} \alpha^j (1 - \alpha)^{x_{t-1}-j} \mathbb{P}_{\alpha, G}(\varepsilon_t = x_t - j),$$

where α is the coefficient of the INAR(1) model (McKenzie 1985; Al-Osh and Alzaid 1987). For $k < \min\{X_t - \sum_{i=1}^p X_{t-i} \mid t = p + 1, \dots, n\}$ or $k > \max\{X_t \mid t = 1, \dots, n\}$,

the values $\hat{G}_{sp}(k)$, $k \in \mathbb{N}_0$, are equal to 0. For further details, see Drost et al. (2009).

In practice, discrete probability distributions such as the Poisson, the negative binomial or the geometric distribution are often used as innovation distribution G , see Weiß (2018), Yang (2019), Al-Osh and Alzaid (1987), Al-Osh and Alzaid (1990). The common feature of all these distributions is their smoothness in the sense that consecutive entries of their PMFs differ only slightly from each other. However, for a small sample size n , the semiparametric estimation approach of Drost et al. (2009) may lead to rather non-smooth innovation distributions with unnatural gaps in their PMF. For illustration, we consider a time series containing counts of transactions of structured products (factor long certificates with leverage) from on-market and off-market trading per trading day between February 1, 2017 and July 31, 2018 (thus $n = 381$). These data, which are plotted in Fig. 1, have first been presented by Homburg et al. (2021), who derived them from the Cascade-Turnoverdata of the Deutsche Börse Group. In the upper right corner, we see the estimated innovation distribution using the semiparametric procedure of Drost et al. (2009) which turns out to be smooth. In the second row, we consider only the first 100 observations of the time series, where the first plot shows indeed a bimodal estimated innovation distribution. In the third row, we only considered the first 20 observations. The lower-left plot shows the resulting estimated PMF, which contains an unnatural gap with $\hat{G}_{sp}(3)$ being estimated exactly equal to zero while its neighbors $\hat{G}_{sp}(2)$ and $\hat{G}_{sp}(4)$ are estimated positive. Hence, the resulting estimation is not smooth contrary to the estimated innovation distribution on the whole time series. In general, such non-smooth innovation distributions are not common in practice and instead, smoothly estimated innovation distributions are often desired. In this paper, we want to use this

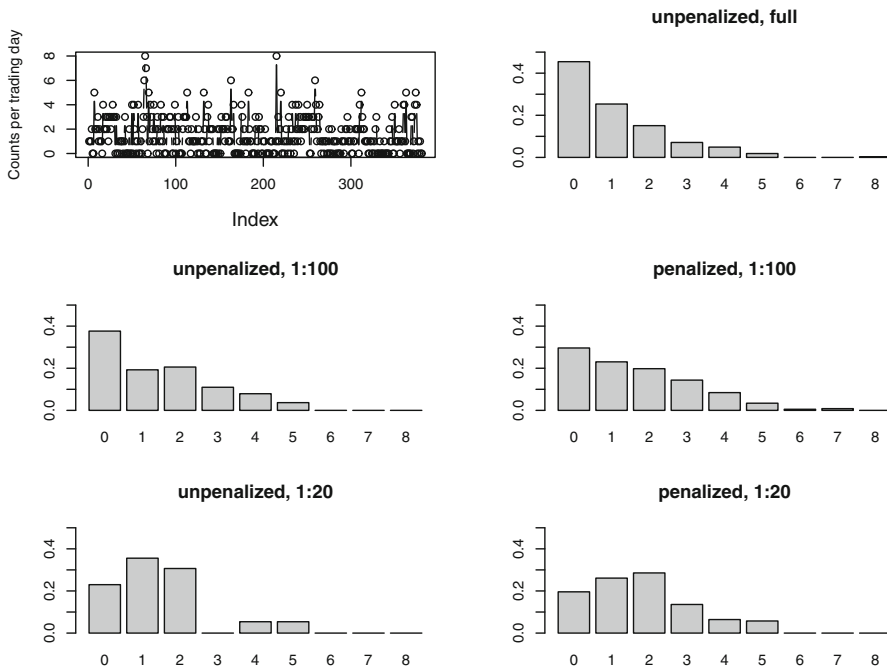


Fig. 1 From left to right and top to bottom: Plot of time series of counts of transactions of structured products per trading day, the unpenalized estimation of the corresponding innovation distribution based on the full data and the (un)penalized estimated innovation distribution for the first 100 and 20 observations, respectively

prior knowledge and take advantage of a natural qualitative smoothness assumption on the innovation distribution by proposing a version of the semiparametric estimation approach, which penalizes the roughness of the innovation distribution. The resulting estimated PMFs of this approach are contained in the right plots in the second and third row, respectively. In comparison, the penalized estimation now leads to a smoother estimation of the PMF without any gaps. We will have a closer look at additional real data examples in Sect. 4. For long time series, the smoothing caused by penalization is not of such great importance, because the distribution estimated without penalization will be sufficiently smooth by itself. But for short time series, estimation without smoothing will commonly lead to jagged estimated innovation distributions although the true distribution behind the data might be smooth. So the need for smoothing is of particular importance for short time series.

The paper is organized as follows. In Sect. 2, we introduce a penalized estimation approach using roughness penalization and propose two algorithms for the data-driven selection of the penalization parameter. Section 3 examines our estimation approach in a comprehensive simulation study, where we compare the estimation performance of the penalized and the unpenalized approach for different settings. In a real data application in Sect. 4, we analyze the monthly demand of car spare parts to illustrate our method and its practical relevance. In the conclusion in Sect. 5, we summarize the results and give an outlook on further research questions.

2 Penalized approach of fitting INAR models

Penalized estimation of count data is a modern topic in current statistical research. Bui et al. (2021) consider parameter estimation in count data models using penalized likelihood methods. In a time series context, Nardi and Rinaldo (2011) studied LASSO penalization for fitting autoregressive time series models to get sparse solutions, i.e. where some autoregressive coefficients are estimated exactly as zero. Fokianos (2010) proposed an alternative estimation scheme for the estimation of INAR models based on minimizing the least square criterion under ridge type of constraints. Wang (2020) proposed a variable selection procedure for INAR(1) models with Poisson distributed innovations including covariables by using penalized estimation and Wang et al. (2021) introduced an order selection procedure for INAR(p) and INARCH(p) models also by using penalized estimation. By contrast, in this paper, we propose a penalized estimation approach for INAR models which does not rely on a penalization of the INAR coefficients (towards zero), but on a penalization of the roughness of the innovation distribution (towards smoothness).

2.1 Penalized estimation approach using roughness penalty

The idea of our approach is to penalize the log-likelihood used in the semiparametric estimation of the INAR model according to Drost et al. (2009). Thus, we still do not assume a parametric class of distributions, we only use the assumed qualitative (i.e. nonparametric) property of smoothness. More precisely, this refers to a roughness penalization as introduced by Scott et al. (1980), which is e.g. used by Adam et al. (2019) for developing a nonparametric approach to fit hidden Markov models to time series of counts. We design the penalty term based on the idea of Tibshirani et al. (2005), where differences of successive parameters are penalized. In this regard, we allow for differences of order $m \in \mathbb{N}$. Applied to our setting, the estimation approach based on Drost et al. (2009) now maximizes the penalized log-likelihood (compare (2))

$$\log(\mathcal{L}_{\text{pen}}(\alpha, G)) = \log(\mathcal{L}(\alpha, G)) - \eta \cdot d_{G,m},$$

where $\eta > 0$ is the so-called smoothing or penalization parameter, $d_{G,m}$ denotes a suitable measure to quantify the roughness of G and m corresponds to the order of difference. According to Tibshirani et al. (2005), a first possible roughness measure for the penalization term is based on the L_1 distance (LASSO penalization), i.e.

$$d_{G,m,1} = \sum_{i=m}^{\max(x_1, \dots, x_n)} |\Delta^m G(i)|, \quad (3)$$

where $\Delta^m G(i) = \Delta^{m-1}(\Delta G(i))$ and $\Delta G(i) = G(i) - G(i-1)$. In addition, we consider the squared L_2 distance (Ridge penalization) as second roughness measure, i.e.

$$d_{G,m,2} = \sum_{i=m}^{\max(x_1, \dots, x_n)} (\Delta^m G(i))^2. \tag{4}$$

The idea behind choosing this second roughness measure is that it does not shrink the differences of the successive entries of the PMF exactly to 0 (contrary to the first roughness measure), but the differences become close to 0, which is more in line with the idea of a smooth distribution (note the analogy of penalized regression, where the L_1 penalization is used for variable selection because of this property, see Fahrmeir et al. (2013)). The order of the differences m is a tuning parameter. For $m = 1$, we penalize only the distance between two directly consecutive entries, for $m = 2$ the smoothness is extended to a triple of values, etc.

Remark 1 A possible extension would be to allow for different penalization weights (η_i) for the individual (higher-order) differences of the entries of the PMF. For instance, in the case of L_1 penalization, the goal could be to maximize

$$\log(\mathcal{L}(\alpha, G)) - \sum_{i=m}^{\max(x_1, \dots, x_n)} \eta_i | \Delta^m G(i) |,$$

analogously for the case of L_2 penalization.

Figure 2 shows a first exemplary result on a sample of an INAR(1) process with $n = 25$ observations, order of difference $m = 1$ and smoothing parameter $\eta = 1$ roughly chosen by eye. In this example, the benefit of penalization already becomes clear. The penalized estimated innovation distributions are much closer to the true Poi(1) innovation distribution (which was truncated at value six for clarity) than the unpenalized estimated innovation distribution. Also, the difference between the L_1 and the L_2 penalization becomes visible. When using the L_2 penalization, the distances between the values of the PMF become small, when using the L_1 penalization they are shrunked to zero.

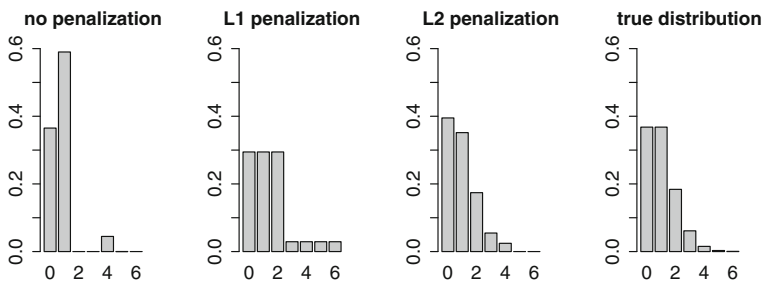


Fig. 2 Barplots of the (estimated) innovation distributions for one realization in the four cases (no penalization, L_1 penalization, L_2 penalization, true distribution)

2.2 Selection of the penalization parameter

Now, we propose two approaches to determine for a fixed roughness measure the optimal smoothing/penalization parameter η , which is a trade-off between fit to the data and the smoothness assumption. For this purpose, we adapt as a first approach the cross-validation procedure described in Adam et al. (2019) to our setting. Therefore, we split the data set into s blocks F_i , $i = 1, \dots, s$, of roughly equal size. In each fold i , $F_{(-i)}$ denotes the in-sample data (data without F_i) and F_i the out-of-sample data. This replicates the correct dependence structure except for the “glue points”, which only has a minor effect in practice when the data originate from an INAR model of small order. The greedy search algorithm is structured as follows:

Algorithm 1

- (1) Choose an initial $\eta^{(0)} > 0$ and set $z = 0$.
- (2) For each fold i and for each value on a specified grid

$$\{ \dots, \eta^{(z)} - 2c, \eta^{(z)} - c, \eta^{(z)}, \eta^{(z)} + c, \eta^{(z)} + 2c, \dots \}$$

where $c \in \mathbb{R}$ is a small constant, estimate the model with penalization on $F_{(-i)}$ and compute the penalized log-likelihood on F_i .

- (3) Average the resulting log-likelihood values across all folds i and choose $\eta^{(z+1)}$ as the penalization parameter on the grid that yields the maximum value.
- (4) Repeat steps 2) and 3) until $\eta^{(z+1)} = \eta^{(z)}$ and define $\eta^{\text{opt}} := \eta^{(z+1)}$.

Furthermore, to avoid a potentially non-optimal selection of the penalization parameter η caused by an inappropriate choice of the initial value $\eta^{(0)}$, we propose a second optimization algorithm. How we split the data in each fold j , $j = 1, \dots, \tilde{s}$, in in- and out-of-sample data is specified later in Sect. 3.

Algorithm 2

- (1) For each fold j and each value η on a specified grid $\{0, \tilde{c}, 2\tilde{c}, 3\tilde{c}, \dots, u\}$ on the interval $[0, u]$ for an appropriate upper bound u , estimate the model with penalization on the in-sample data and compute the penalized log-likelihood on the out-of-sample data.
- (2) Average the resulting log-likelihood values across all folds j .
- (3) Fit a polynomial of order r to the curve resulting from plotting the average out-of-sample log-likelihood against the grid.
- (4) Choose η^{opt} as the value on the grid, where the curve takes its maximum value.

3 Simulation study

We investigate the performance of the proposed procedure in a simulation study with $K = 500$ Monte Carlo samples of size $n \in \{20, 50, 100, 250, 500, 1000\}$ generated from an INAR(1) process according to (1) for $p = 1$ with different coefficients

$\alpha \in \{0.2, 0.5, 0.8\}$ and innovation distributions $G \in \{\text{Poi}(1), \text{NB}(2, \frac{2}{3}), \text{Geo}(\frac{1}{2}), \text{ZIP}(\frac{1}{2}, 2)\}$, where ZIP denotes a zero-inflated Poisson distribution as in Jazi et al. (2012). The parameters of the negative binomial, geometric and zero-inflated Poisson distribution are hereby chosen to have the same expected value as the $\text{Poi}(1)$ distribution. But contrary to the $\text{Poi}(1)$ distribution which is equidispersed, i.e. the variance of the distribution equals its mean, they are overdispersed, i.e. their variances are larger than their mean values. Another difference between the considered innovation distributions is their (non-) smoothness, see also Fig. 12 in the appendix. The $\text{Poi}(1)$, $\text{NB}(2, \frac{2}{3})$ and $\text{Geo}(\frac{1}{2})$ distributions are rather smooth, but the $\text{ZIP}(\frac{1}{2}, 2)$ distribution, which shows a pronounced zero probability, is not. The effect of this property on the roughness penalization is investigated in Subsect. 3.5. Moreover, in Subsect. 3.2, we also provide a small simulation setting for higher-order INAR processes and consider the case of an INAR(2) model. The implementation is straightforward but is a lot more demanding such that we restrict the considered setting to a rather small extent. To ensure the stationarity of the time series, we actually generate $n + 100$ observations and remove the first 100 observations. We consider first ($m = 1$) and second ($m = 2$) order differences in the penalization term (see Subsect. 3.4). As initialization for the smoothing parameter $\eta^{(0)}$, we set $\eta^{(0)} = 1$ as in the example in Fig. 2 for the sample sizes $n \in \{20, 50, 100, 250\}$ and for computing time reasons $\eta^{(0)} = 0.5$ for $n \in \{500, 1000\}$.¹ For the considered grid around the smoothing parameter (see Algorithm 1) we choose $c = 0.05$ resulting in $\{\eta^{(z)} - 0.1, \eta^{(z)} - 0.05, \eta^{(z)}, \eta^{(z)} + 0.05, \eta^{(z)} + 0.1\}$. Unless stated otherwise, we use $\alpha = 0.5$ as true INAR(1) coefficient and Algorithm 1 with 10-fold cross validation ($s = 10$) as optimization algorithm. For the realization of the simulation study, we use the statistical programming language R 4.1.2 (R Core Team 2021).

3.1 Roughness penalty for smooth innovations distributions and first order differences

Figure 3 shows the L_2 distances of the estimated innovation distributions to the true $\text{Poi}(1)$ innovation distribution,

$$d(\hat{G}, G) = \sum_{i=0}^M (\hat{G}(i) - G(i))^2,$$

for the different sample sizes and the respective estimation methods (unpenalized (up), L_1 penalization and L_2 penalization) for some large enough M . We use $M = 70$ as upper bound for the observations x_1, \dots, x_n since after this value the corresponding probabilities of occurrence are negligibly small. When the sample size n is small, the penalized estimation of the innovation distribution provides a large benefit compared to the unpenalized estimation: The L_2 distances of the penalized estimated to the true innovation distribution are much smaller than those of the unpenalized estimated to the true innovation distribution. Furthermore, the L_2 penalization performs better

¹ For large sample sizes n , η^{opt} will be close to zero, so a lower initial value decreases the number of iterations needed for Algorithm 1, which saves computing time.

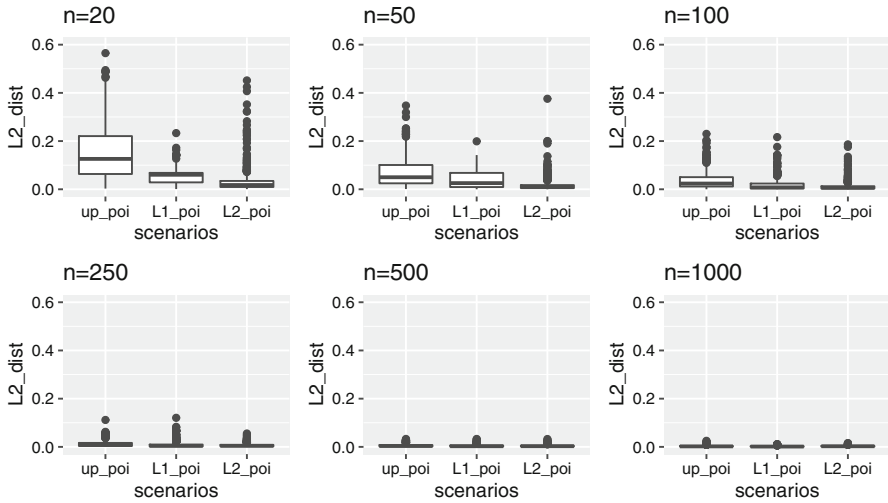


Fig. 3 Boxplots of the L_2 distances of the estimated innovation distribution to the true Poi(1) innovation distribution of an INAR(1) process for different sample sizes n . We report results for unpenalized (up), L_1 and L_2 penalized estimation

than the L_1 penalization. In Table 3 in the appendix, we also report the variance, the bias and the MSE of the first five estimated entries of the PMF resulting from the different procedures for the different sample sizes n . We see that the penalized estimation reduces both the variance, the absolute bias and consequently also the MSE of the estimated innovation distribution, especially for small n . Figures 17 and 18 and Tables 6 and 7 in the appendix show the analog results for a true $NB(2, \frac{2}{3})$ and $Geo(\frac{1}{2})$ distribution, respectively. In general, regardless of the distribution and up to a sample size of $n = 100$, we see a clear improvement concerning the estimation performance when using penalization. From a sample size of $n = 250$ on, this improvement can only be seen marginally with the different methods essentially coinciding for large n . In Fig. 13 and Table 4 in the appendix, we show the results for INAR coefficient $\alpha = 0.2$ and Poi(1) innovation distribution and, correspondingly, in Fig. 15 and Table 5 for $\alpha = 0.8$. In the latter case, the benefit of the penalized estimation compared to the unpenalized estimation is even larger than in the case $\alpha = 0.5$. This is plausible because it is in general more difficult to estimate the innovation distribution for a larger value of α as this also leads to a larger observations mean with innovations mean remaining constant. Therefore, more entries of the PMF have to be estimated with the same amount of data. Contrary, for $\alpha = 0.2$, we have (with analog arguments) less entries of the PMF which have to be estimated with the same amount of data, which simplifies the estimation of the PMF in general and the benefit of penalization decreases. Altogether, we can conclude that the benefit of penalization is more pronounced with larger α , that is, with larger serial dependency.

We get confirming conclusions, when we consider the values of the optimal smoothing parameter η , which approaches zero with increasing n , see Fig. 4 for the

case of a true $\text{Poi}(1)$ innovation distribution, Figs. 19 and 20 in the appendix for the cases of a true $\text{NB}(2, \frac{2}{3})$ and $\text{Geo}(\frac{1}{2})$ innovation distribution and Figs. 14 and 16 in the appendix in case of a true $\text{Poi}(1)$ innovation distribution with $\alpha = 0.2$ and $\alpha = 0.8$, respectively. Thus, for increasing n , the penalized and the unpenalized estimation coincide as intuitively expected: For large n , there are enough observations to learn the smoothness of the innovation distribution from the data even without imposing smoothness through penalization.

3.2 Higher-order INAR processes

To show that our proposed procedure is also applicable for higher-order INAR processes, we consider the case of a true $\text{INAR}(2)$ process according to (1) for $p = 2$ with coefficients $\alpha_1 = 0.3$, $\alpha_2 = 0.2$ and $G = \text{Poi}(1)$. Due to the high computing time for the semiparametric estimation, we only consider a small simulation setup with $n = 50$ observations and $K = 100$ Monte Carlo samples. We consider L_1 and L_2 penalization with first order differences and compare the performance with the case of estimation without penalization. In Fig. 21 in the appendix, we see that also for higher-order INAR models, penalized estimation of the innovation distribution provides a clear benefit compared to unpenalized estimation. With penalization we are closer to the true innovation distribution than without and we are able to reduce the variance, the absolute bias and consequently the MSE of our estimation, see Table 1. Again, L_2 penalization works best.

3.3 Alternative selection of the penalization parameter

To investigate whether the results depend on the chosen initial parameter, we now determine the optimal penalization parameter alternatively using Algorithm 2 with $u = 5$, $\tilde{c} = 0.1$ and $r = 5$. In this context, we want to address a potential practical issue of Algorithm 1: the generation of the in- and out-of-sample data. For each of

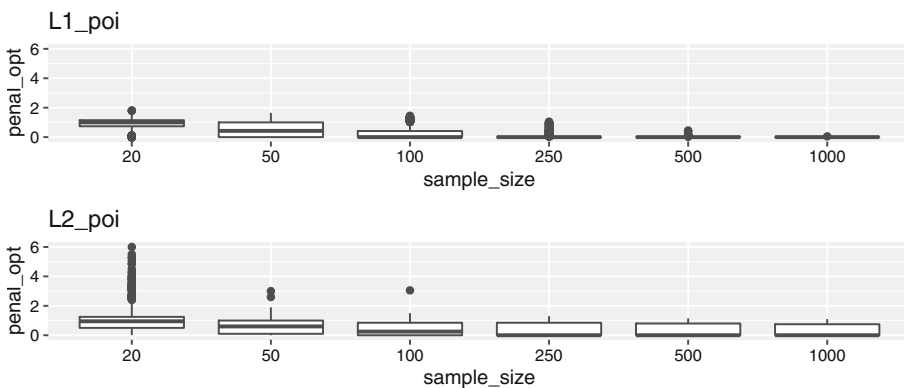


Fig. 4 Boxplots of the penalization parameter η selected by L_1 penalization (upper panel) and L_2 penalization (lower panel) for the different sample sizes n in the case of a true $\text{Poi}(1)$ innovation distribution of an $\text{INAR}(1)$ process

Table 1 Variance, bias and MSE of the first five estimated entries of the PMF for $n = 50$ in case of a true Poi(1) innovation distribution of an INAR(2) process. We report results for unpenalized (up), L_1 and L_2 penalized estimation

n		g0_up	g0_L1	g0_L2	g1_up	g1_L1	g1_L2
50	Variance	0.0269	0.0080	0.0070	0.0244	0.0037	0.0028
	Bias	-0.0499	-0.0571	-0.0016	-0.0317	-0.0786	-0.0393
	MSE	0.0294	0.0113	0.0070	0.0254	0.0099	0.0043
n		g2_up	g2_L1	g2_L2	g3_up	g3_L1	g3_L2
50	Variance	0.0193	0.0054	0.0045	0.0087	0.0073	0.0030
	Bias	0.0369	0.0266	0.0193	0.0227	0.0719	0.0202
	MSE	0.0207	0.0061	0.0048	0.0092	0.0124	0.0034
n		g4_up	g4_L1	g4_L2			
50	Variance	0.0025	0.0043	0.0009			
	Bias	0.0127	0.0511	0.0070			
	MSE	0.0026	0.0069	0.0009			

the 10 folds, 90% of the data becomes the in-sample data and the remaining 10% the out-sample data. For small n , 10% of the data is small. To avoid this, we now use an n -fold cross-validation ($\tilde{s} = n$) for sample sizes $n \in \{20, 50\}$ with Algorithm 2, where starting from each observation the following 50% of the data is in- and the other 50% is out-of-sample. When reaching the end of the time series, we start again from its beginning.

In Fig. 5, we see the results of this alternative procedure compared to the previous (iterative) procedure in Algorithm 1. It gives slightly better results than the iterative method, but overall the distances are very similar. The same can be concluded when considering Table 8. The alternative procedure leads to slightly lower MSE values, but altogether the values resemble each other. The 10-fold cross-validation also seems to be suitable and the resulting optimal parameters of the two procedures are close to each other (see Fig. 6). In conclusion, if we determine the optimal parameter from a sequence on a grid as in Algorithm 2, we tend to get slightly better results. However, the price to pay is a much higher computing time than with the iterative procedure. The iterative method needs a reasonably chosen starting value, but then it gives similarly good results in considerably less computing time. In addition, when using the alternative method, the question arises how to choose the upper limit of the interval adequately. In the following, we will continue to use the iterative method from Algorithm 1 but one should keep in mind that Algorithm 2 is also a practically useful procedure.

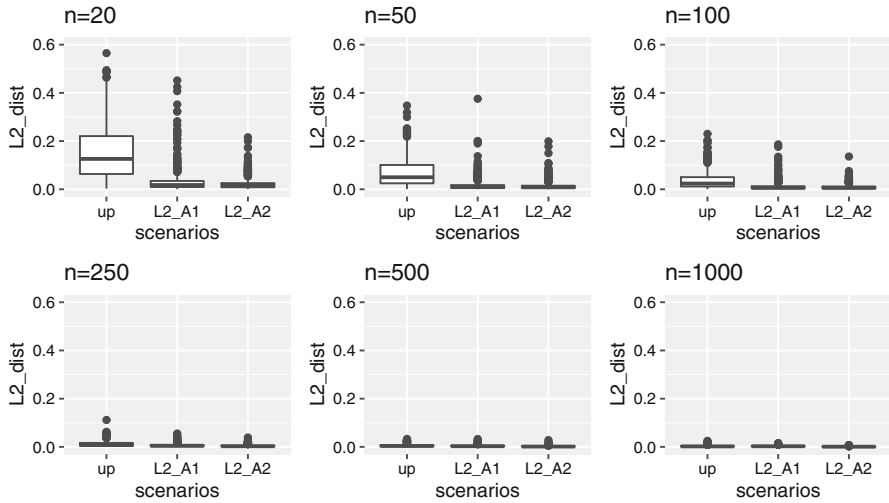


Fig. 5 Boxplots of the L_2 distances of the estimated innovation distribution to the true Poi(1) innovation distribution of an INAR(1) process for the different sample sizes n . We report results for unpenalized (up) and L_2 penalized estimation using either the iterated Algorithm 1 (A1) or the alternative Algorithm 2 (A2)

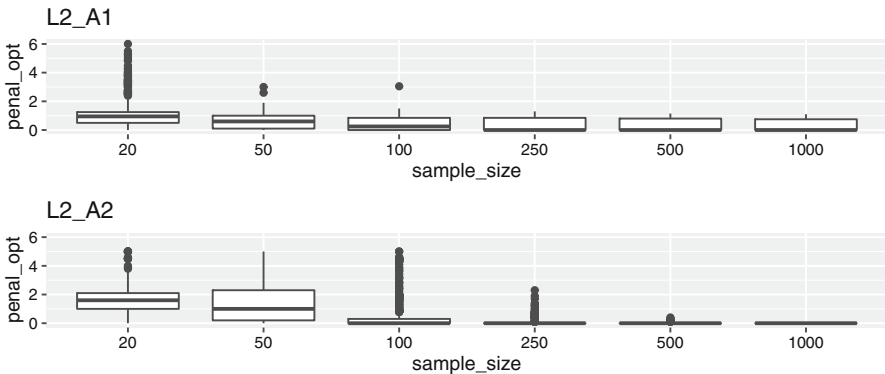


Fig. 6 Boxplots of the penalization parameter η selected by L_2 penalization using Algorithm 1 (A1, upper panel) and Algorithm 2 (A2, lower panel) for the different sample sizes n in the case of a true Poi(1) innovation distribution of an INAR(1) process

3.4 Higher-order differences in penalization term

So far we only considered first order differences ($m = 1$). Now we want to see if penalizing higher-order differences (e.g. $m = 2$) is able to improve the performance of our penalized estimation method. In Fig. 7 and Table 10 in the appendix it is visible for the case of a true Poi(1) innovation distribution and L_2 penalization that also the penalization of differences of higher order performs better than the unpenalized estimation in the cases of small sample sizes, and that it comes close to the penalization of first order differences. Similar results are in Fig. 22 and Table 9,

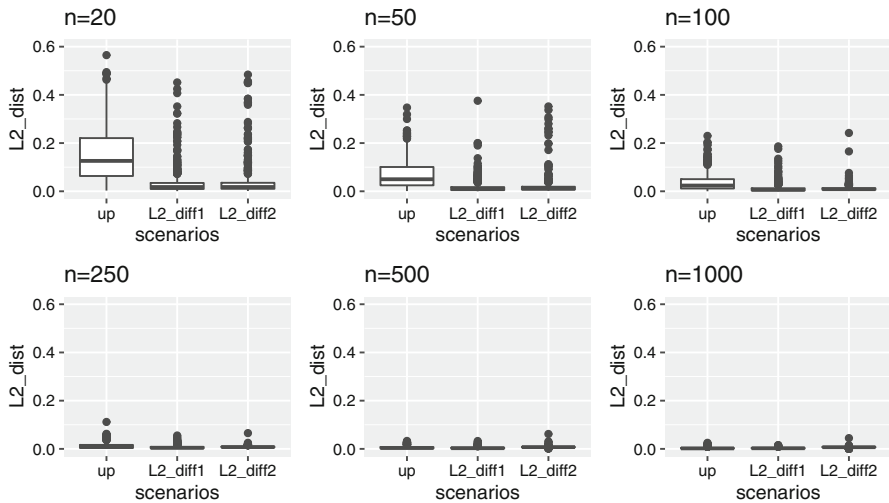


Fig. 7 Boxplots of the L_2 distances of the estimated innovation distribution to the true $Poi(1)$ innovation distribution of an $INAR(1)$ process for the different sample sizes n . We report results for unpenalized (up) and L_2 penalized estimation using either first order (diff1) or second order (diff2) differences

both in the appendix, where we see the results of first and second order differences for the L_1 penalization. In case of L_1 penalization, we would prefer second order differences for small sample sizes. Overall, however, the L_2 penalization of first-order differences performs best.

3.5 Non-smooth innovation distribution

Finally, let us consider the case of $ZIP(\frac{1}{2}, 2)$ distributed innovations and consider Fig. 8. The results are as expected. Since the ZIP distribution is not smooth (see Fig. 12 in the appendix), the smoothness assumption and hence the penalization is not suitable. The boxplots reflect this: Except for sample size $n = 20$, the penalized estimation procedure provides no benefit and for some n even leads to slightly higher L_2 distances from the true $ZIP(\frac{1}{2}, 2)$ distribution than the unpenalized procedure. As we can see in Table 11, the penalized estimation leads to a higher absolute bias when estimating the first (non-smooth) entry, $G(0)$, of the PMF. As sample size n increases, the penalization has less impact, as there is enough data to detect the incorrect assumption such that the unpenalized and the penalized procedures coincide.

For comparison, let's take a look at the results with a true $ZIP(\frac{1}{2}, 2)$ distribution when we exclude $G(0)$ from the penalization displayed in Fig. 9, i.e. when we consider

$$\tilde{d}_{G,m,1} = \sum_{i=m+1}^{\max(x_1, \dots, x_n)} |\Delta^m G(i)| \quad \text{and} \quad \tilde{d}_{G,m,2} = \sum_{i=m+1}^{\max(x_1, \dots, x_n)} (\Delta^m G(i))^2,$$

instead of $d_{G,m,1}$ and $d_{G,m,2}$ defined in (3) and (4). It becomes clear what we would

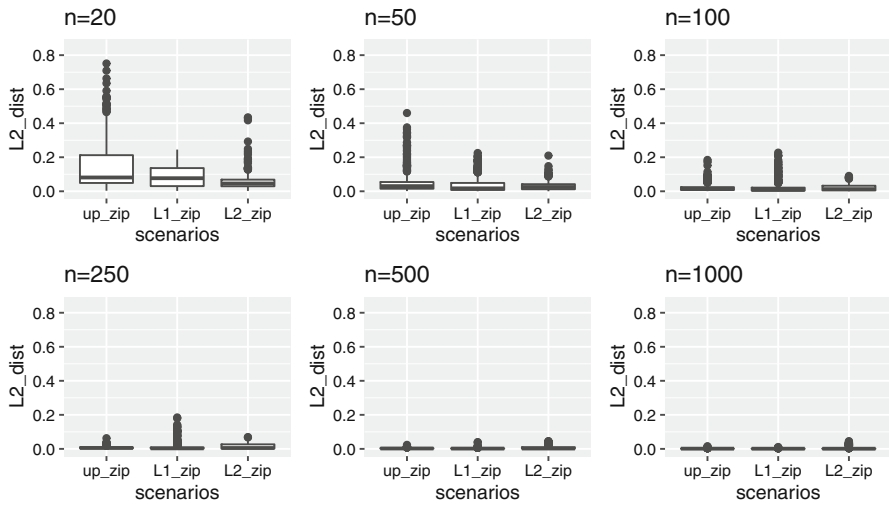


Fig. 8 Boxplots of the L_2 distances of the estimated innovation distribution to the true $\text{ZIP}(\frac{1}{2}, 2)$ innovation distribution of an INAR(1) process for the different sample sizes n . We report results for unpenalized (up), L_1 and L_2 penalized estimation

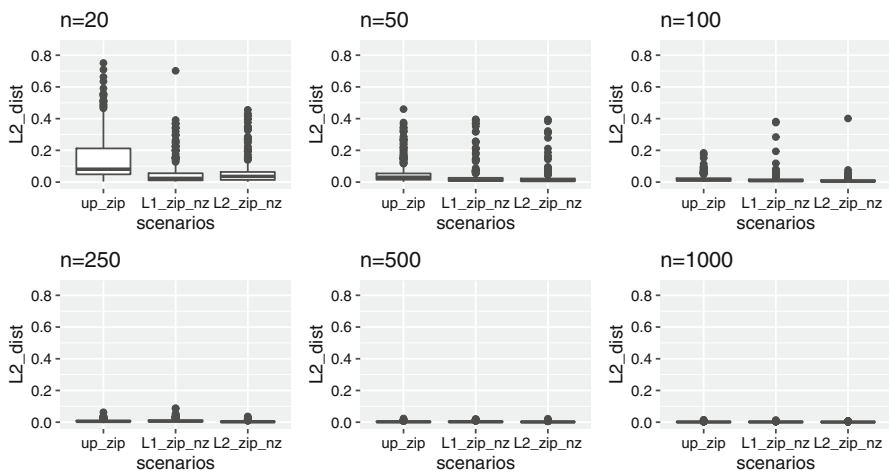


Fig. 9 Boxplots of the L_2 distances of the estimated innovation distribution to the true $\text{ZIP}(\frac{1}{2}, 2)$ innovation distribution of an INAR(1) process for the different sample sizes n . We report results for unpenalized (up), L_1 and L_2 penalized estimation without smoothing of $G(0)$ (nz)

expect: By excluding the “non-smooth entry” $G(0)$ of the PMF of the innovation distribution from penalization, the penalized estimation works well again and provides a benefit for small n . In this case, the penalized estimation now results in a lower absolute bias of the estimated PMF’s first entry compared to the unpenalized estimation (compare Table 12). However, this benefit is not as pronounced as in the cases of a true $\text{Poi}(1)$, $\text{NB}(2, \frac{2}{3})$ and $\text{Geo}(\frac{1}{2})$ innovation distribution. This can probably be explained by the fact that the $\text{ZIP}(\frac{1}{2}, 2)$ distribution has most of its mass

in zero and the corresponding entry of the PMF, $G(0)$, remains unaffected by the penalization. Consequently, the results from penalized and unpenalized estimation do not differ substantially from each other.

In summary, if the smoothness assumption of the innovation distribution is correctly imposed, it provides a large benefit for small sample size n . This holds whether the true underlying distribution is equidispersed or overdispersed. The best results are obtained for L_2 penalization and first-order differences.

3.6 Estimation of the INAR coefficient

A drawback of the penalized estimation is that the estimation of the INAR coefficient α no longer works well for small sample size n , see Fig. 23 in the appendix. A strength of the semiparametric estimation approach of Drost et al. (2009) is the accurate joint estimation of the INAR coefficient and the innovation distribution. This joint estimation accuracy is not maintained when penalization is used for small n . The L_2 distances of the penalized estimated INAR coefficient α to the true value are higher than for the unpenalized estimated coefficient. For increasing n , the estimation of α improves, but since the benefit of the penalized estimation lies in the cases where n is small, this is no comfort.

Instead, we can solve this problem by taking only the estimator for the innovation distribution from the penalized approach and estimating the INAR coefficient with the unpenalized (efficient) estimation approach of Drost et al. (2009). In Fig. 23, we see that it is indeed preferable to combine the unpenalized estimation of the INAR coefficient α and the penalized estimation of the innovation distribution G . Also when looking at the MSE, it is clear that this combination outperforms all other estimation approaches under consideration.

4 Real data example

For modeling intermittent demand, Syntetos and Boylan (2021) consider the equidispersed Poisson distribution on the one hand, and, as the demand variability may be severe when demand is intermittent, overdispersed distributions from the Compound-Poisson family (such as the negative binomial distribution) on the other hand. All these parametric distributions are smooth. With our novel penalized semiparametric estimation approach, we get smooth distributions without parametric assumptions, and as we saw in our simulations, our penalization procedure works well for both equi- and overdispersed distributions. By contrast, if using an unpenalized non-parametric estimation approach such as the empirical distribution function (EDF), Syntetos and Boylan (2021) criticize that demand values not observed in the past are automatically assigned zero probabilities for the future. Furthermore, they state that an EDF provides a perfect fit to the historical data, but it does not ensure the goodness of fit to the demand over the forecast horizon, especially with respect to higher percentiles. Again, these drawbacks are omitted with our penalized estimation approach. Finally, historical demand time series are often rather short, see the demand count time series provided by Snyder (2002) as

an example, such that smoothing approaches would be particularly welcome. For these reason, the forecasting of intermittent demand appears to be a promising application area for our proposed penalized semiparametric estimation procedure.

Therefore, we consider time series ($n = 51$) of the monthly demand of different car spare parts offered by an Australian subsidiary of a Japanese car company from January 1998 to March 2002 (Snyder 2002). Figure 10 contains an exemplary time series of car part 2404. The observations vary between 0 and 5 and the up and down movements indicate a moderate autocorrelation level. After inspecting the corresponding (P)ACF also included in Fig. 10, we conclude that an AR(1)-like model might be appropriate for describing the serial dependence of the time series. Moreover, L_2 penalization with first order differences leads to an estimated innovation distribution without any unnatural gaps, i.e. zero values, in the PMF.

Now consider the 1-step median prediction and the 90% quantile of the 1-step prediction of the demand for car spare part 2404. The latter serves here as a worst-case scenario for spare parts requirements. Therefore, we determine the median and the 90% quantile of the predictive distribution $P(\dots | y)$, where $y \in 0, \dots, 10$. Based on the results of the simulation study in Subsect. 3.6, we use the penalized estimated innovation distribution and the unpenalized estimated INAR coefficient to determine the conditional predictive distribution. Table 2 shows that the penalized estimation tends to lead to higher predicted values (more conservative prediction). Consequently, without penalizing the innovation distribution, the predictions for the demand for spare parts may be too low, which can lead to a lack of spare parts. Moreover, the penalization of the innovation distribution (especially for such short

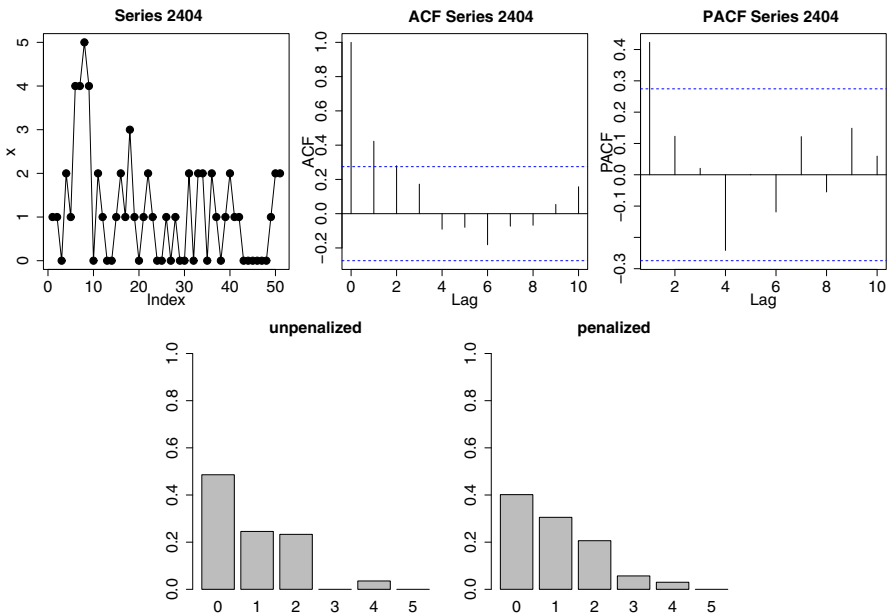


Fig. 10 From left to right and top to bottom: Plot of time series of monthly demand for car spare part 2404, its corresponding ACF and PACF and the unpenalized and the penalized estimated innovation distribution

Table 2 Unpenalized and penalized 1-step median prediction and 90% quantile of the 1-step prediction of the demand for car spare part 2404 when observing demand y

y	0	1	2	3	4	5	6	7	8	9	10
Median (unpenalized)	1	1	1	1	2	2	2	3	3	3	3
Median (penalized)	1	1	1	2	2	2	2	3	3	3	3
90% quantile (unpenalized)	2	2	3	3	4	4	4	5	5	5	6
90% quantile (penalized)	2	3	3	4	4	4	5	5	5	6	6

time series) can serve as a robustness analysis to identify possible uncertainties in the forecast at an early stage.

In addition, we consider car spare part 1971. Figure 11 again suggests an AR(1)-like model and a moderate autocorrelation level. The observations vary between 0 and 4 and there may be zero inflation in this time series. Therefore, in addition to the unpenalized and penalized estimates, we also consider the penalized estimate of the innovation distribution, where $G(0)$ is not smoothed (see Subsect. 3.5). It becomes clear that this last estimation procedure yields more plausible results than when $G(0)$ is smoothed. Again, the penalized estimation procedure yields a slightly smoother innovation distribution than the unpenalized estimation. In summary, if there is a reasonable suspicion of zero inflation, $G(0)$ should not be smoothed.

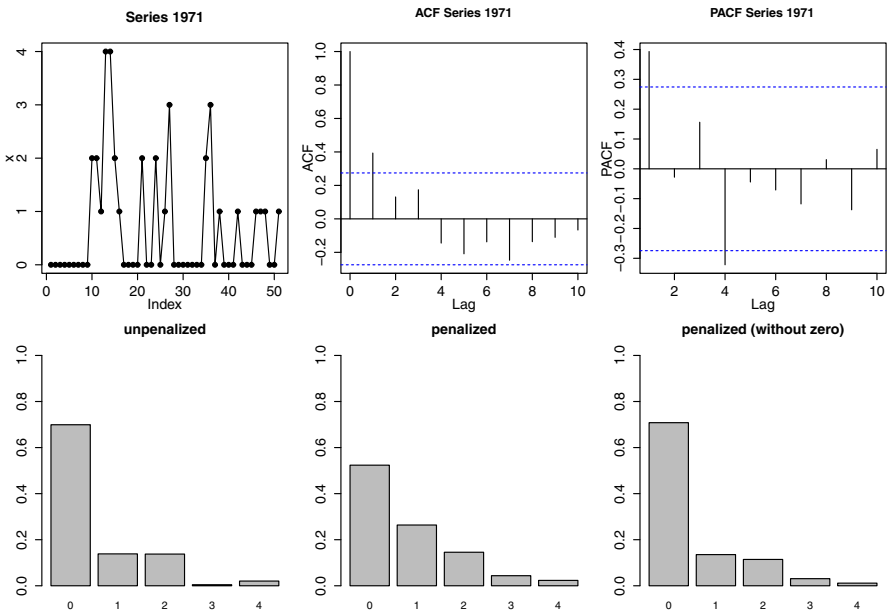


Fig. 11 Plot of time series of monthly demand for car spare part 1971, its corresponding ACF and PACF, the unpenalized and the penalized estimated innovation distribution and the penalized estimated innovation distribution excluding the first entry of the PMF (from left to right and from top to bottom)

5 Conclusion

Although semiparametric estimation yields a decent fit in INAR models, its performance is often not convincing for small sample sizes. Therefore, we proposed a penalization approach that exploits a qualitative smoothness assumption fulfilled by commonly used innovation distributions. A simulation study showed that our penalization approach provides a large benefit in estimating the innovation distribution, especially for small sample sizes. Additionally, we showed that the combination of unpenalized estimation of INAR coefficients and penalized estimation of the innovation distribution provided the best performance. Future research should investigate whether additional penalization of the INAR coefficients may result in further benefit. Furthermore, as the penalization approach proved to be beneficial for forecasting, one may also think of an application in statistical process control, i.e. for the design of control charts relying on a fitted INAR(1) model. Another interesting issue for future research is the application of our proposed method on integer-valued autoregressive models on \mathbb{Z} , such as those proposed by Kim and Park (2008) or Liu et al. (2021).

Appendix

See Figures 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22 and 23.

See Tables 3, 4, 5, 6, 7, 8, 9, 10, 11 and 12.

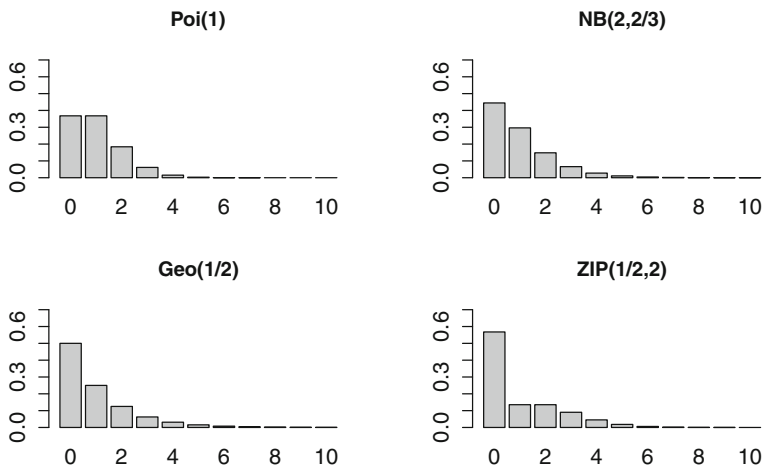


Fig. 12 Probability density functions of the $\text{Poi}(1)$, $\text{NB}(2, \frac{2}{3})$, $\text{Geo}(\frac{1}{2})$ and $\text{ZIP}(\frac{1}{2}, 2)$ distributions (truncated at value 10 for clarity)

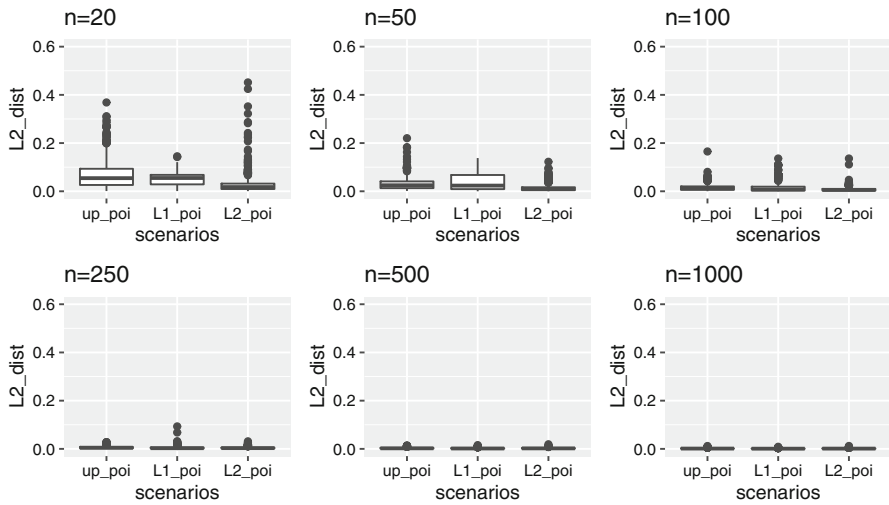


Fig. 13 Boxplots of the L_2 distances of the estimated innovation distribution to the true $Poi(1)$ innovation distribution of an INAR(1) process for the different sample sizes n and $\alpha = 0.2$. We report results for unpenalized (up), L_1 and L_2 penalized estimation

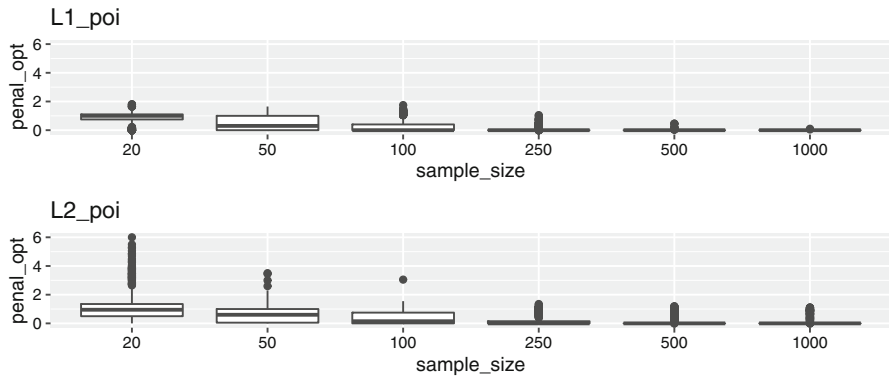


Fig. 14 Boxplots of the penalization parameter η selected by L_1 penalization (upper panel) and L_2 penalization (lower panel) for the different sample sizes n in the case of a true $Poi(1)$ innovation distribution of an INAR(1) process and $\alpha = 0.2$

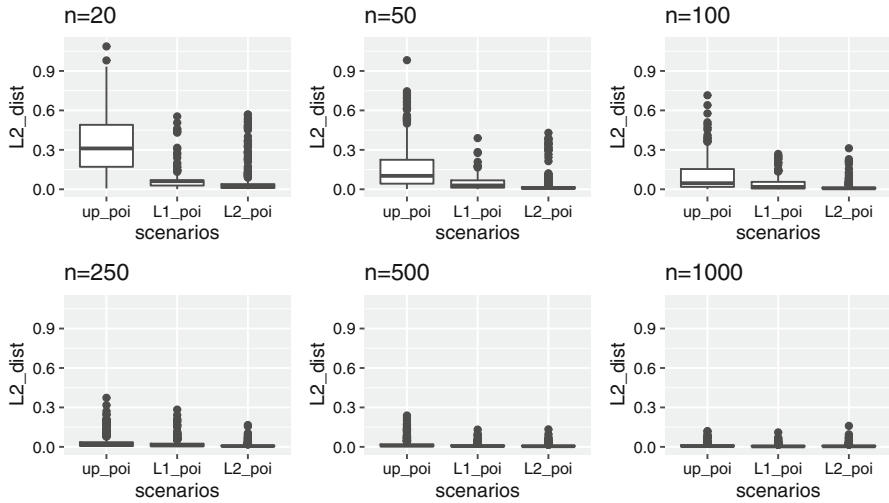


Fig. 15 Boxplots of the L_2 distances of the estimated innovation distribution to the true Poi(1) innovation distribution of an INAR(1) process for the different sample sizes n and $\alpha = 0.8$. We report results for unpenalized (up), L_1 and L_2 penalized estimation

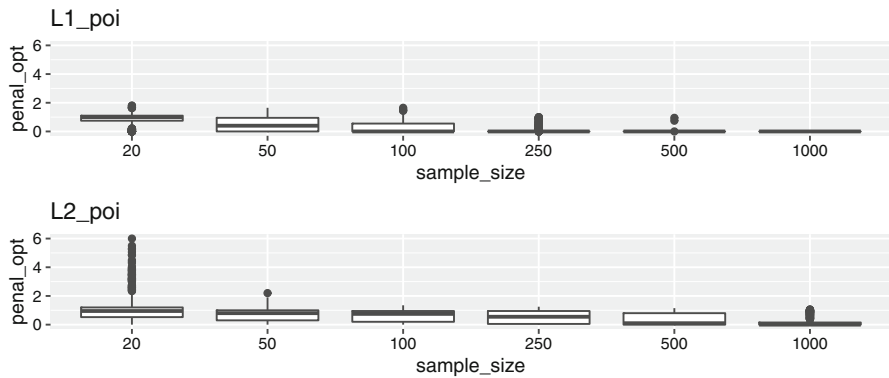


Fig. 16 Boxplots of the penalization parameter η selected by L_1 penalization (upper panel) and L_2 penalization (lower panel) for the different sample sizes n in the case of a true Poi(1) innovation distribution of an INAR(1) process and $\alpha = 0.8$

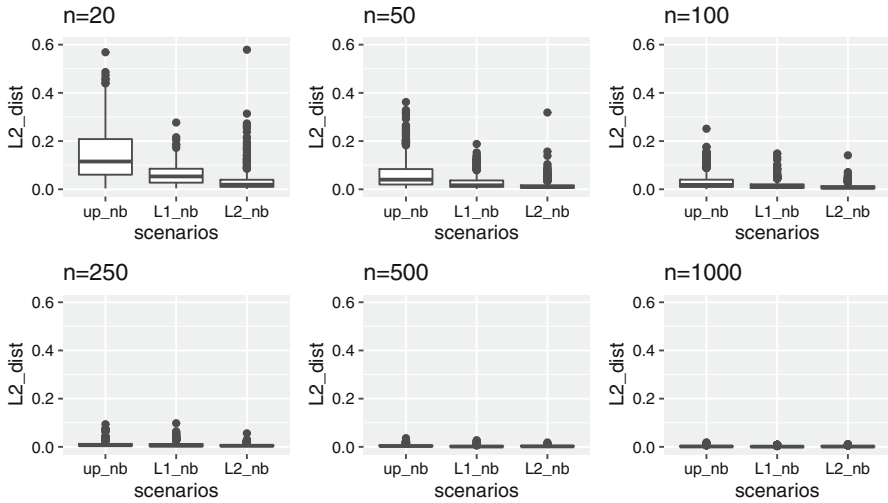


Fig. 17 Boxplots of the L_2 distances of the estimated innovation distribution to the true $NB(2, \frac{2}{3})$ innovation distribution of an INAR(1) process for the different sample sizes n . We report results for unpenalized (up), L_1 and L_2 penalized estimation

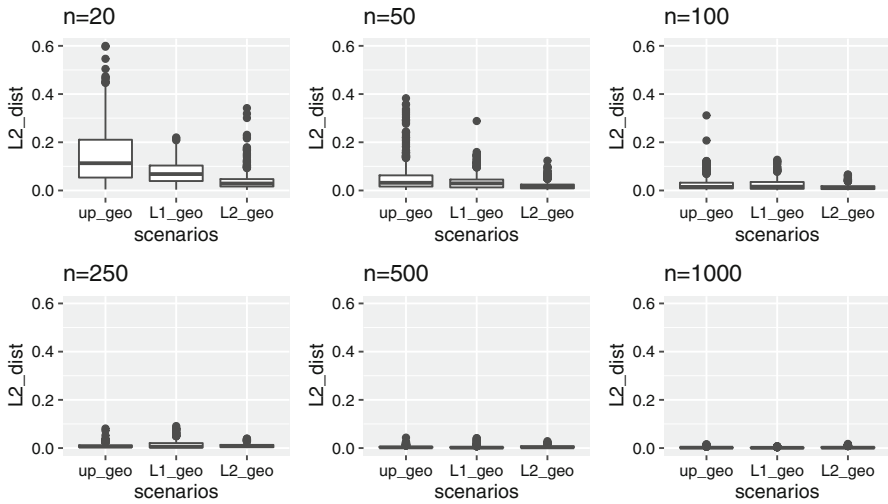


Fig. 18 Boxplots of the L_2 distances of the estimated innovation distribution to the true $Geo(\frac{1}{2})$ innovation distribution of an INAR(1) process for the different sample sizes n . We report results for unpenalized (up), L_1 and L_2 penalized estimation

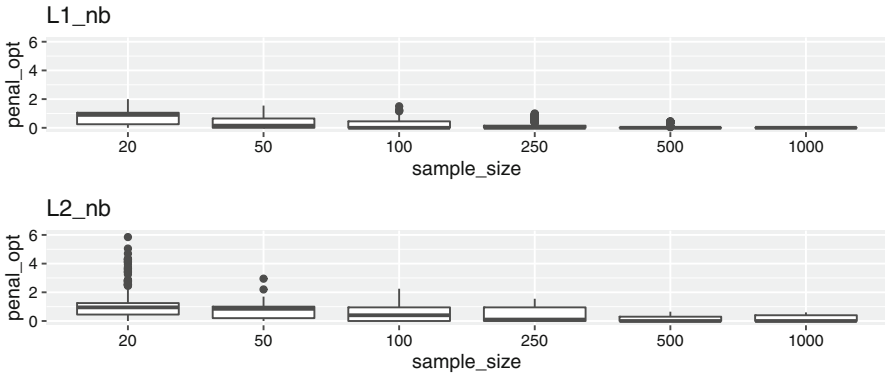


Fig. 19 Boxplots of the penalization parameter η selected by L_1 penalization (upper panel) and L_2 penalization (lower panel) for the different sample sizes n in the case of a true $NB(2, \frac{2}{3})$ innovation distribution of an INAR(1) process

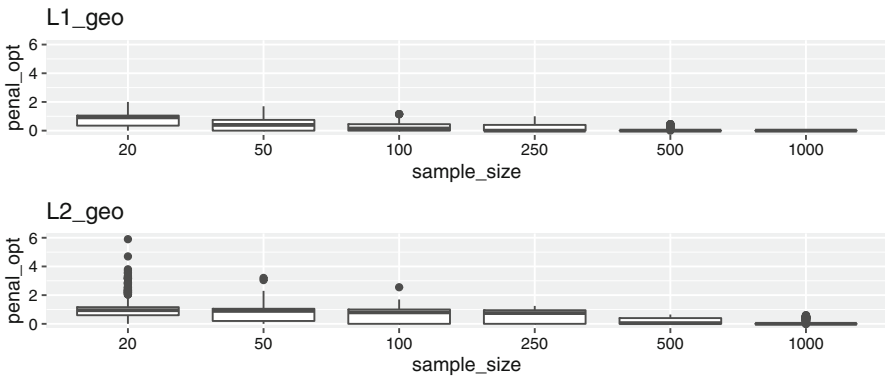


Fig. 20 Boxplots of the penalization parameter η selected by L_1 penalization (upper panel) and L_2 penalization (lower panel) for the different sample sizes n in the case of a true $Geo(\frac{1}{2})$ innovation distribution of an INAR(1) process

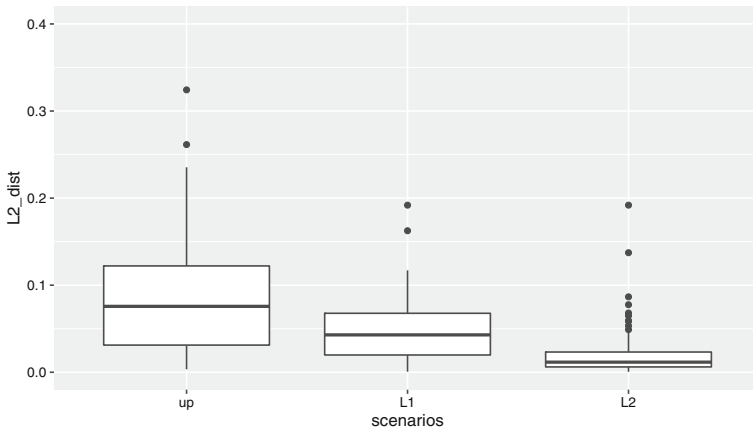


Fig. 21 Boxplots of the L_2 distances of the estimated innovations distribution to the true $Poi(1)$ innovation distribution of an INAR(2) process and $n = 50$. We report results for unpenalized (up), L_1 and L_2 penalized estimation

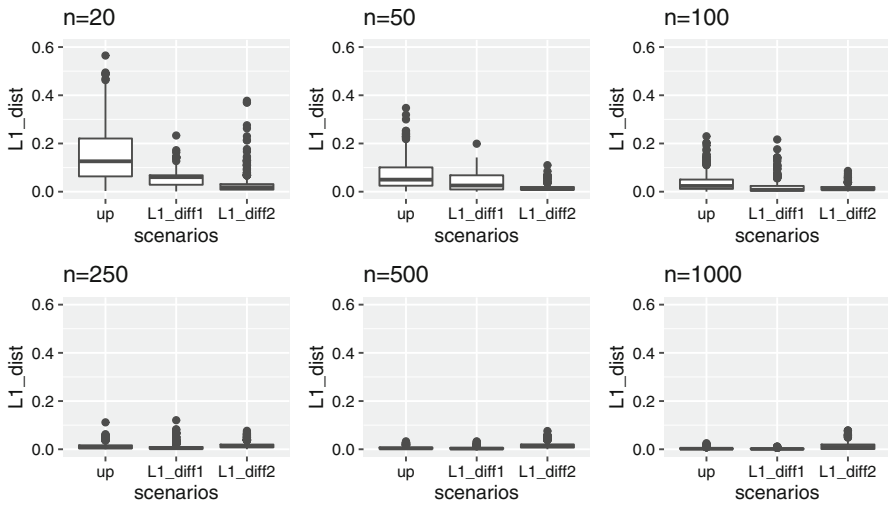


Fig. 22 Boxplots of the L_2 distances of the estimated innovation distribution to the true $Poi(1)$ innovation distribution of an INAR(1) process for the different sample sizes n . We report results for unpenalized (up) and L_1 penalized estimation using either first order (diff1) or second order (diff2) differences

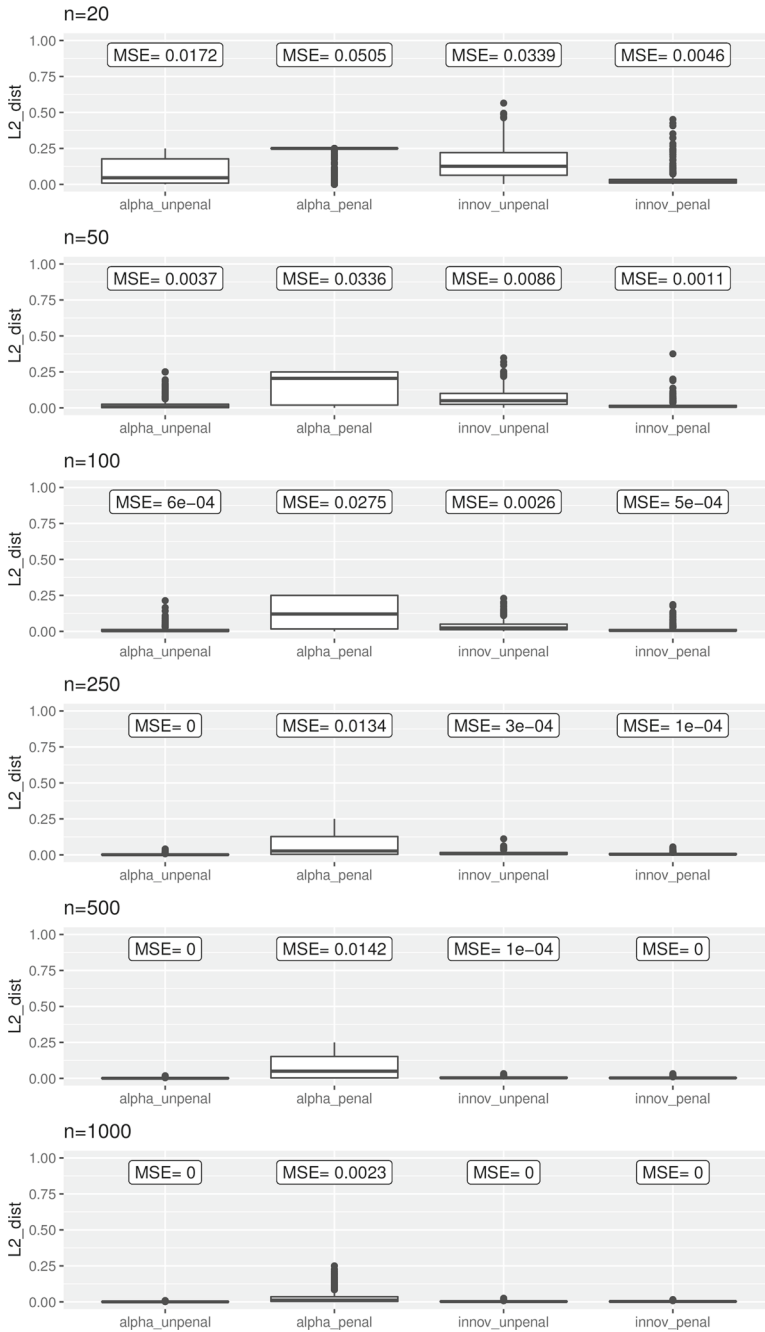


Fig. 23 Boxplots of the L_2 distances between (1) unpenalized estimated INAR coefficient α and the true INAR coefficient α , (2) L_2 penalized and true α , (3,4) their related estimated innovation distribution and the true Poi(1) innovation distribution of an INAR(1) process for different sample sizes n and their corresponding MSE (rounded to four digits)

Table 3 Variance, bias and MSE of the first five estimated entries of the PMF for the different sample sizes n in case of a true $Poi(1)$ innovation distribution of an INAR(1) process. We report results for unpenalized (up), L_1 and L_2 penalized estimation

n	g_0_up	$g_0_L_1$	$g_0_L_2$	g_1_up	$g_1_L_1$	$g_1_L_2$	g_2_up	$g_2_L_1$	$g_2_L_2$	g_3_up	$g_3_L_1$	$g_3_L_2$	g_4_up	$g_4_L_1$	$g_4_L_2$	
20	Variance	0.0458	0.0069	0.0126	0.0382	0.0066	0.0321	0.0059	0.0063	0.0156	0.0078	0.0053	0.0048	0.0042	0.0029	
	Bias	-0.0921	-0.0700	-0.0114	-0.0217	-0.0371	0.0513	0.0456	0.0220	0.0376	0.1064	0.0330	0.0250	0.0890	0.0400	
	MSE	0.0543	0.0118	0.0127	0.0387	0.0110	0.0080	0.0348	0.0080	0.0068	0.0170	0.0191	0.0064	0.0054	0.0122	0.0045
50	Variance	0.0246	0.0077	0.0060	0.0198	0.0037	0.0155	0.0045	0.0034	0.0073	0.0079	0.0025	0.0012	0.0028	0.0006	
	Bias	-0.0312	-0.0414	0.0003	-0.0017	-0.0468	-0.0251	0.0102	0.0149	0.0116	0.0165	0.0550	0.0118	0.0037	0.0307	0.0063
	MSE	0.0256	0.0094	0.0060	0.0198	0.0089	0.0043	0.0156	0.0047	0.0036	0.0076	0.0110	0.0026	0.0012	0.0037	0.0006
100	Variance	0.0125	0.0052	0.0035	0.0114	0.0052	0.0080	0.0031	0.0027	0.0031	0.0042	0.0014	0.0006	0.0010	0.0003	
	Bias	-0.0225	-0.0206	-0.0030	0.0032	-0.0276	-0.0216	0.0103	0.0120	0.0105	0.0067	0.0234	0.0136	0.0019	0.0096	0.0006
	MSE	0.0130	0.0056	0.0035	0.0114	0.0059	0.0035	0.0081	0.0032	0.0028	0.0032	0.0047	0.0016	0.0006	0.0011	0.0003
250	Variance	0.0043	0.0027	0.0019	0.0039	0.0023	0.0020	0.0025	0.0012	0.0010	0.0012	0.0007	0.0002	0.0002	0.0002	
	Bias	-0.0075	-0.0113	-0.0053	0.0006	-0.0086	-0.0164	0.0054	0.0106	0.0120	0.0026	0.0063	0.0084	-0.0014	0.0011	
	MSE	0.0044	0.0029	0.0019	0.0039	0.0024	0.0022	0.0025	0.0021	0.0014	0.0010	0.0012	0.0008	0.0002	0.0002	0.0002
500	Variance	0.0021	0.0015	0.0010	0.0019	0.0012	0.0013	0.0011	0.0009	0.0006	0.0004	0.0004	0.0001	0.0001	0.0001	
	Bias	-0.0044	-0.0029	-0.0050	0.0030	0.0007	-0.0142	0.0004	0.0017	0.0096	0.0009	0.0005	0.0001	-0.0000	0.0011	
	MSE	0.0022	0.0015	0.0010	0.0019	0.0012	0.0015	0.0011	0.0009	0.0007	0.0004	0.0004	0.0005	0.0001	0.0001	0.0001
1000	Variance	0.0012	0.0007	0.0006	0.0011	0.0006	0.0010	0.0006	0.0004	0.0004	0.0002	0.0003	0.0000	0.0000	0.0000	
	Bias	-0.0034	-0.0018	-0.0044	0.0014	0.0002	-0.0165	0.0007	0.0013	0.0108	0.0013	0.0002	0.0000	0.0000	0.0000	
	MSE	0.0012	0.0007	0.0006	0.0011	0.0006	0.0013	0.0006	0.0004	0.0005	0.0002	0.0002	0.0000	0.0000	0.0000	0.0000

Table 4 Variance, bias and MSE of the first five estimated entries of the PMF for the different sample sizes n in case of a true $\text{Poi}(1)$ innovation distribution of an INAR(1) process and $\alpha = 0.2$. We report results for unpenalized (up), L_1 and L_2 penalized estimation

n	g_0 _up	g_0 _L1	g_0 _L2	g_1 _up	g_1 _L1	g_1 _L2	g_2 _up	g_2 _L1	g_2 _L2	g_3 _up	g_3 _L1	g_3 _L2	g_4 _up	g_4 _L1	g_4 _L2
20	Variance	0.0244	0.0063	0.0103	0.0225	0.0047	0.0066	0.0169	0.0059	0.0061	0.0079	0.0045	0.0018	0.0041	0.0024
	Bias	-0.0126	-0.0662	-0.0073	-0.0064	-0.0737	-0.0317	0.0132	0.0473	0.0208	0.1064	0.0284	0.0223	0.0851	0.0419
50	MSE	0.0245	0.0107	0.0103	0.0225	0.0102	0.0076	0.0170	0.0081	0.0065	0.0192	0.0053	0.0023	0.0114	0.0041
	Variance	0.0118	0.0072	0.0046	0.0091	0.0060	0.0033	0.0075	0.0044	0.0029	0.0078	0.0019	0.0006	0.0022	0.0005
100	Bias	-0.0082	-0.0355	0.0027	-0.0021	-0.0457	-0.0210	0.0043	0.0160	0.0081	0.0516	0.0084	0.0067	0.0263	0.0093
	MSE	0.0119	0.0085	0.0046	0.0091	0.0081	0.0038	0.0075	0.0047	0.0030	0.0029	0.0105	0.0007	0.0029	0.0006
250	Variance	0.0058	0.0042	0.0027	0.0040	0.0039	0.0023	0.0034	0.0025	0.0017	0.0043	0.0009	0.0004	0.0010	0.0003
	Bias	-0.0074	-0.0201	-0.0001	-0.0039	-0.0236	-0.0164	0.0054	0.0131	0.0075	0.0036	0.0219	0.0074	0.0013	0.0027
500	MSE	0.0059	0.0046	0.0027	0.0041	0.0045	0.0026	0.0035	0.0027	0.0018	0.0015	0.0048	0.0004	0.0011	0.0003
	Variance	0.0024	0.0016	0.0013	0.0016	0.0013	0.0013	0.0014	0.0012	0.0009	0.0005	0.0004	0.0001	0.0001	0.0001
1000	Bias	-0.0057	-0.0037	-0.0022	-0.0009	-0.0060	-0.0079	0.0041	0.0072	0.0069	0.0014	0.0009	0.0025	0.0007	0.0008
	MSE	0.0024	0.0016	0.0013	0.0016	0.0013	0.0014	0.0014	0.0013	0.0009	0.0005	0.0004	0.0001	0.0001	0.0001
2000	Variance	0.0011	0.0008	0.0007	0.0009	0.0006	0.0008	0.0006	0.0005	0.0005	0.0003	0.0002	0.0003	0.0001	0.0001
	Bias	-0.0014	0.0005	-0.0012	0.0009	-0.0003	-0.0080	0.0005	0.0013	0.0048	0.0001	-0.0013	0.0038	-0.0003	0.0004
5000	MSE	0.0011	0.0008	0.0007	0.0009	0.0006	0.0009	0.0006	0.0006	0.0005	0.0003	0.0002	0.0003	0.0001	0.0001
	Variance	0.0005	0.0004	0.0004	0.0005	0.0003	0.0005	0.0004	0.0003	0.0003	0.0001	0.0001	0.0001	0.0000	0.0000
10000	Bias	-0.0020	0.0005	-0.0009	0.0006	0.0006	-0.0056	0.0013	-0.0006	0.0029	0.0001	-0.0003	0.0029	0.0002	0.0001
	MSE	0.0006	0.0004	0.0004	0.0005	0.0003	0.0005	0.0004	0.0003	0.0003	0.0001	0.0001	0.0002	0.0000	0.0000

Table 5 Variance, bias and MSE of the first five estimated entries of the PMF for the different sample sizes n in case of a true $\text{Poi}(1)$ innovation distribution of an INAR(1) process and $\alpha = 0.8$. We report results for unpenalized (up), L_1 and L_2 penalized estimation

n	g_0 _up	g_0 _L1	g_0 _L2	g_1 _up	g_1 _L1	g_1 _L2	g_2 _up	g_2 _L1	g_2 _L2	g_3 _up	g_3 _L1	g_3 _L2	g_4 _up	g_4 _L1	g_4 _L2
20	Variance	0.0575	0.0084	0.0162	0.0639	0.0063	0.0096	0.0491	0.0062	0.0071	0.0393	0.0065	0.0237	0.0050	0.0049
	Bias	-0.2006	-0.0784	-0.0249	-0.1267	-0.0837	-0.0475	0.0533	0.0438	0.0172	0.1200	0.1056	0.0385	0.0842	0.0442
	MSE	0.0977	0.0145	0.0168	0.0800	0.0133	0.0119	0.0519	0.0081	0.0074	0.0537	0.0191	0.0080	0.0308	0.0122
50	Variance	0.0462	0.0096	0.0086	0.0377	0.0084	0.0042	0.0312	0.0052	0.0024	0.0174	0.0077	0.0036	0.0058	0.0021
	Bias	-0.1184	-0.0450	-0.0112	-0.0111	-0.0533	-0.0424	0.0519	0.0202	0.0183	0.0559	0.0556	0.0235	0.0155	0.0261
	MSE	0.0602	0.0117	0.0087	0.0378	0.0113	0.0060	0.0339	0.0056	0.0028	0.0205	0.0108	0.0041	0.0061	0.0038
100	Variance	0.0351	0.0096	0.0044	0.0230	0.0086	0.0032	0.0194	0.0052	0.0024	0.0074	0.0054	0.0016	0.0013	0.0004
	Bias	-0.0793	-0.0396	-0.0085	0.0020	-0.0324	-0.0374	0.0478	0.0223	0.0222	0.0237	0.0351	0.0230	0.0052	0.0119
	MSE	0.0414	0.0112	0.0045	0.0230	0.0097	0.0046	0.0216	0.0057	0.0029	0.0080	0.0066	0.0021	0.0013	0.0004
250	Variance	0.0150	0.0087	0.0031	0.0103	0.0066	0.0027	0.0045	0.0033	0.0015	0.0017	0.0019	0.0008	0.0002	0.0002
	Bias	-0.0361	-0.0207	-0.0051	0.0217	-0.0053	-0.0323	0.0078	0.0147	0.0194	0.0059	0.0076	0.0165	0.0008	0.0021
	MSE	0.0163	0.0091	0.0031	0.0108	0.0066	0.0037	0.0046	0.0035	0.0019	0.0017	0.0020	0.0011	0.0002	0.0002
500	Variance	0.0087	0.0041	0.0027	0.0058	0.0034	0.0028	0.0023	0.0017	0.0010	0.0006	0.0006	0.0001	0.0001	0.0001
	Bias	-0.0136	-0.0018	-0.0041	0.0083	0.0013	-0.0203	0.0049	0.0000	0.0125	-0.0003	0.0006	0.0107	0.0005	-0.0003
	MSE	0.0089	0.0041	0.0027	0.0059	0.0034	0.0032	0.0024	0.0017	0.0012	0.0006	0.0006	0.0007	0.0001	0.0001
1000	Variance	0.0053	0.0029	0.0034	0.0035	0.0022	0.0029	0.0009	0.0008	0.0009	0.0003	0.0003	0.0004	0.0001	0.0001
	Bias	0.0013	-0.0031	-0.0074	-0.0003	0.0029	-0.0061	0.0002	-0.0001	0.0084	-0.0009	0.0001	0.0045	-0.0004	0.0008
	MSE	0.0053	0.0029	0.0034	0.0035	0.0022	0.0030	0.0009	0.0008	0.0009	0.0003	0.0003	0.0004	0.0001	0.0001

Table 6 Variance, bias and MSE of the first five estimated entries of the PMF for the different sample sizes n in case of a true $NB(2, \frac{2}{3})$ innovation distribution of an INAR(1) process. We report results for unpenalized (up), L_1 and L_2 penalized estimation

n	g0_up	g0_L1	g0_L2	g1_up	g1_L1	g1_L2	g2_up	g2_L1	g2_L2	g3_up	g3_L1	g3_L2	g4_up	g4_L1	g4_L2	
20	Variance	0.0537	0.0153	0.0137	0.0338	0.0062	0.0054	0.0265	0.0068	0.0053	0.0135	0.0063	0.0041	0.0053	0.0037	0.0022
	Bias	-0.0720	-0.1058	-0.0340	-0.0186	-0.0201	-0.0048	0.0366	0.0377	0.0184	0.0277	0.0567	0.0152	0.0178	0.0580	0.0201
	MSE	0.0588	0.0265	0.0148	0.0341	0.0066	0.0054	0.0279	0.0083	0.0056	0.0143	0.0095	0.0044	0.0056	0.0070	0.0026
50	Variance	0.0236	0.0115	0.0057	0.0184	0.0043	0.0025	0.0115	0.0037	0.0024	0.0058	0.0035	0.0017	0.0019	0.0019	0.0009
	Bias	-0.0412	-0.0554	-0.0363	0.0129	-0.0069	0.0004	0.0092	0.0162	0.0168	0.0112	0.0222	0.0119	0.0038	0.0167	0.0073
	MSE	0.0253	0.0145	0.0070	0.0185	0.0044	0.0025	0.0116	0.0040	0.0026	0.0060	0.0040	0.0018	0.0019	0.0021	0.0009
100	Variance	0.0117	0.0068	0.0043	0.0093	0.0026	0.0016	0.0054	0.0026	0.0015	0.0025	0.0014	0.0009	0.0010	0.0005	0.0004
	Bias	-0.0279	-0.0408	-0.0343	0.0164	0.0053	0.0025	0.0050	0.0166	0.0183	0.0016	0.0087	0.0088	0.0037	0.0050	0.0034
	MSE	0.0125	0.0085	0.0054	0.0096	0.0026	0.0016	0.0054	0.0029	0.0019	0.0025	0.0015	0.0010	0.0010	0.0005	0.0004
250	Variance	0.0037	0.0038	0.0023	0.0033	0.0014	0.0009	0.0016	0.0014	0.0009	0.0008	0.0006	0.0004	0.0004	0.0003	0.0002
	Bias	-0.0069	-0.0290	-0.0308	0.0040	0.0081	0.0030	0.0011	0.0113	0.0160	0.0009	0.0054	0.0084	0.0006	0.0025	0.0030
	MSE	0.0038	0.0046	0.0032	0.0033	0.0015	0.0009	0.0016	0.0015	0.0011	0.0008	0.0006	0.0006	0.0004	0.0003	0.0002
500	Variance	0.0017	0.0013	0.0011	0.0016	0.0007	0.0007	0.0008	0.0004	0.0005	0.0004	0.0001	0.0001	0.0002	0.0001	0.0001
	Bias	-0.0029	-0.0077	-0.0167	0.0011	0.0046	0.0031	0.0007	-0.0027	0.0060	0.0006	0.0045	0.0068	0.0005	-0.0001	-0.0011
	MSE	0.0017	0.0013	0.0014	0.0016	0.0007	0.0007	0.0008	0.0004	0.0005	0.0004	0.0002	0.0002	0.0002	0.0001	0.0001
1000	Variance	0.0008	0.0004	0.0009	0.0007	0.0005	0.0003	0.0004	0.0002	0.0003	0.0002	0.0001	0.0001	0.0001	0.0001	0.0001
	Bias	-0.0003	0.0031	-0.0170	0.0003	-0.0026	0.0026	-0.0008	-0.0002	0.0080	0.0002	-0.0002	0.0050	0.0006	-0.0015	-0.0007
	MSE	0.0008	0.0005	0.0012	0.0007	0.0005	0.0003	0.0004	0.0002	0.0002	0.0004	0.0001	0.0001	0.0001	0.0001	0.0001

Table 7 Variance, bias and MSE of the first five estimated entries of the PMF for the different sample sizes n in case of a true $\text{Geo}(\frac{1}{2})$ innovation distribution of an INAR(1) process. We report results for unpenalized (up), L_1 and L_2 penalized estimation

n		g0_up	g0_L1	g0_L2	g1_up	g1_L1	g1_L2	g2_up	g2_L1	g2_L2	g3_up	g3_L1	g3_L2	g4_up	g4_L1	g4_L2
20	Variance	0.0551	0.0189	0.0137	0.0364	0.0060	0.0050	0.0217	0.0074	0.0055	0.0118	0.0054	0.0033	0.0056	0.0030	0.0022
	Bias	-0.1071	-0.1490	-0.0743	0.0251	0.0128	0.0318	0.0279	0.0429	0.0230	0.0308	0.0553	0.0159	0.0181	0.0497	0.0205
	MSE	0.0666	0.0411	0.0192	0.0371	0.0062	0.0060	0.0225	0.0093	0.0093	0.0060	0.0128	0.0084	0.0036	0.0060	0.0055
50	Variance	0.0214	0.0143	0.0072	0.0140	0.0044	0.0025	0.0095	0.0037	0.0023	0.0046	0.0022	0.0014	0.0020	0.0013	0.0009
	Bias	-0.0296	-0.0868	-0.0618	0.0129	0.0188	0.0246	0.0062	0.0241	0.0219	0.0041	0.0152	0.0084	0.0044	0.0162	0.0064
	MSE	0.0223	0.0218	0.0110	0.0142	0.0048	0.0031	0.0096	0.0043	0.0043	0.0028	0.0046	0.0024	0.0015	0.0020	0.0009
100	Variance	0.0090	0.0094	0.0046	0.0078	0.0032	0.0016	0.0037	0.0023	0.0013	0.0019	0.0010	0.0007	0.0009	0.0005	0.0004
	Bias	-0.0162	-0.0737	-0.0561	0.0157	0.0263	0.0225	-0.0040	0.0214	0.0196	0.0023	0.0095	0.0080	0.0004	0.0071	0.0034
	MSE	0.0093	0.0148	0.0077	0.0080	0.0039	0.0021	0.0037	0.0027	0.0027	0.0017	0.0019	0.0011	0.0007	0.0009	0.0005
250	Variance	0.0030	0.0065	0.0034	0.0025	0.0020	0.0009	0.0014	0.0013	0.0006	0.0007	0.0005	0.0003	0.0004	0.0002	0.0002
	Bias	-0.0045	-0.0432	-0.0531	0.0060	0.0210	0.0215	-0.0006	0.0114	0.0199	-0.0024	0.0037	0.0062	0.0020	0.0035	0.0034
	MSE	0.0030	0.0084	0.0062	0.0026	0.0025	0.0014	0.0014	0.0014	0.0010	0.0007	0.0005	0.0004	0.0004	0.0002	0.0002
500	Variance	0.0014	0.0015	0.0020	0.0013	0.0009	0.0007	0.0007	0.0005	0.0004	0.0003	0.0002	0.0002	0.0002	0.0001	0.0001
	Bias	-0.0022	-0.0044	-0.0316	0.0043	0.0036	0.0164	-0.0015	0.0004	0.0097	-0.0007	-0.0007	0.0023	0.0002	0.0009	0.0019
	MSE	0.0014	0.0015	0.0030	0.0013	0.0009	0.0010	0.0007	0.0005	0.0005	0.0005	0.0003	0.0002	0.0002	0.0001	0.0001
1000	Variance	0.0007	0.0005	0.0013	0.0007	0.0004	0.0005	0.0003	0.0002	0.0003	0.0002	0.0001	0.0001	0.0001	0.0001	0.0001
	Bias	-0.0007	0.0001	-0.0171	0.0018	0.0008	0.0094	-0.0013	-0.0017	0.0039	0.0000	0.0001	0.0018	0.0003	0.0007	0.0012
	MSE	0.0007	0.0005	0.0016	0.0007	0.0004	0.0006	0.0003	0.0002	0.0003	0.0002	0.0001	0.0001	0.0001	0.0001	0.0001

Table 8 Variance, bias and MSE of the first five estimated entries of the PMF for the different sample sizes n in case of a true $\text{Poi}(1)$ innovation distribution of an INAR(1) process. We report results for unpenalized (up), L_2 penalized estimation using either the iterated Algorithm 1 (A1) or the alternative Algorithm 2 (A2)

n	$g0_up$	$g0_A1$	$g0_A2$	$g1_up$	$g1_A1$	$g1_A2$	$g2_up$	$g2_A1$	$g2_A2$	$g3_up$	$g3_A1$	$g3_A2$	$g4_up$	$g4_A1$	$g4_A2$	
20	Variance	0.0458	0.0126	0.0063	0.0382	0.0066	0.0041	0.0321	0.0063	0.0039	0.0156	0.0053	0.0037	0.0048	0.0029	0.0017
	Bias	-0.0921	-0.0114	-0.0127	-0.0217	-0.0371	-0.0433	0.0513	0.0220	0.0260	0.0376	0.0330	0.0428	0.0250	0.0400	0.0414
	MSE	0.0543	0.0127	0.0065	0.0387	0.0080	0.0060	0.0348	0.0068	0.0046	0.0170	0.0064	0.0055	0.0054	0.0045	0.0034
50	Variance	0.0246	0.0060	0.0043	0.0198	0.0037	0.0029	0.0155	0.0034	0.0020	0.0073	0.0025	0.0019	0.0012	0.0006	0.0007
	Bias	-0.0312	0.0003	-0.0007	-0.0017	-0.0251	-0.0345	0.0102	0.0116	0.0134	0.0165	0.0118	0.0172	0.0037	0.0063	0.0148
	MSE	0.0256	0.0060	0.0043	0.0198	0.0043	0.0041	0.0156	0.0036	0.0022	0.0076	0.0026	0.0022	0.0012	0.0006	0.0009
100	Variance	0.0125	0.0035	0.0030	0.0114	0.0031	0.0030	0.0080	0.0027	0.0020	0.0031	0.0014	0.0011	0.0006	0.0003	0.0003
	Bias	-0.0225	-0.0030	0.0044	0.0032	-0.0216	-0.0159	0.0103	0.0105	0.0056	0.0067	0.0136	0.0044	0.0019	0.0006	0.0033
	MSE	0.0130	0.0035	0.0030	0.0114	0.0035	0.0032	0.0081	0.0028	0.0020	0.0032	0.0016	0.0011	0.0006	0.0003	0.0003
250	Variance	0.0043	0.0019	0.0014	0.0039	0.0020	0.0013	0.0025	0.0012	0.0010	0.0010	0.0007	0.0004	0.0002	0.0002	0.0001
	Bias	-0.0075	-0.0053	0.0032	0.0006	-0.0164	-0.0013	0.0054	0.0120	-0.0002	0.0026	0.0084	-0.0012	-0.0014	0.0011	-0.0002
	MSE	0.0044	0.0019	0.0014	0.0039	0.0022	0.0013	0.0025	0.0014	0.0010	0.0010	0.0008	0.0004	0.0002	0.0002	0.0001
500	Variance	0.0021	0.0010	0.0007	0.0019	0.0013	0.0007	0.0011	0.0006	0.0004	0.0004	0.0004	0.0002	0.0001	0.0001	0.0000
	Bias	-0.0044	-0.0050	0.0035	0.0030	-0.0142	0.0006	0.0004	0.0096	-0.0017	0.0009	0.0081	-0.0015	0.0001	0.0011	-0.0006
	MSE	0.0022	0.0010	0.0008	0.0019	0.0015	0.0007	0.0011	0.0007	0.0004	0.0004	0.0005	0.0002	0.0001	0.0001	0.0000
1000	Variance	0.0012	0.0007	0.0006	0.0011	0.0006	0.0010	0.0006	0.0004	0.0004	0.0002	0.0002	0.0003	0.0000	0.0000	0.0000
	Bias	-0.0034	-0.0018	-0.0044	0.0014	0.0002	-0.0165	0.0007	0.0013	0.0108	0.0013	0.0002	0.0088	0.0002	0.0002	0.0015
	MSE	0.0012	0.0007	0.0006	0.0011	0.0006	0.0013	0.0006	0.0004	0.0005	0.0002	0.0002	0.0004	0.0000	0.0000	0.0000

Table 9 Variance, bias and MSE of the first five estimated entries of the PMF for the different sample sizes n in case of a true $\text{Poi}(1)$ innovation distribution of an INAR(1) process. We report results for unpenalized (up) and L_1 penalized estimation using either first order (diff1) or second order (diff2) differences

n		g0_up	g0_d1	g0_d2	g1_up	g1_d1	g1_d2	g2_up	g2_d1	g2_d2	g3_up	g3_d1	g3_d2	g4_up	g4_d1	g4_d2
20	Variance	0.0458	0.0069	0.0105	0.0382	0.0054	0.0021	0.0321	0.0059	0.0023	0.0156	0.0078	0.0048	0.0048	0.0042	0.0032
	Bias	-0.0921	-0.0700	0.0154	-0.0217	-0.0751	-0.0760	0.0513	0.0456	0.0112	0.0376	0.1064	0.0467	0.0250	0.0890	0.0408
	MSE	0.0543	0.0118	0.0108	0.0387	0.0110	0.0079	0.0348	0.0080	0.0024	0.0170	0.0191	0.0069	0.0054	0.0122	0.0049
50	Variance	0.0246	0.0077	0.0029	0.0198	0.0068	0.0007	0.0155	0.0045	0.0002	0.0073	0.0079	0.0014	0.0012	0.0028	0.0012
	Bias	-0.0312	-0.0414	0.0074	-0.0017	-0.0468	-0.0813	0.0102	0.0149	0.0120	0.0165	0.0550	0.0447	0.0037	0.0307	0.0234
	MSE	0.0256	0.0094	0.0030	0.0198	0.0089	0.0073	0.0156	0.0047	0.0004	0.0076	0.0110	0.0034	0.0012	0.0037	0.0017
100	Variance	0.0125	0.0052	0.0020	0.0114	0.0052	0.0006	0.0080	0.0031	0.0001	0.0031	0.0042	0.0007	0.0006	0.0010	0.0010
	Bias	-0.0225	-0.0206	0.0008	0.0032	-0.0276	-0.0841	0.0103	0.0120	0.0124	0.0067	0.0234	0.0488	0.0019	0.0096	0.0208
	MSE	0.0130	0.0056	0.0020	0.0114	0.0059	0.0076	0.0081	0.0032	0.0002	0.0032	0.0047	0.0030	0.0006	0.0011	0.0015
250	Variance	0.0043	0.0027	0.0016	0.0039	0.0023	0.0011	0.0025	0.0020	0.0002	0.0010	0.0012	0.0006	0.0002	0.0002	0.0010
	Bias	-0.0075	-0.0113	-0.0117	0.0006	-0.0086	-0.0846	0.0054	0.0106	0.0121	0.0026	0.0063	0.0514	-0.0014	0.0013	0.0265
	MSE	0.0044	0.0029	0.0017	0.0039	0.0024	0.0083	0.0025	0.0021	0.0003	0.0010	0.0012	0.0033	0.0002	0.0002	0.0017
500	Variance	0.0021	0.0015	0.0013	0.0019	0.0012	0.0022	0.0011	0.0009	0.0002	0.0004	0.0004	0.0010	0.0001	0.0001	0.0009
	Bias	-0.0044	-0.0029	-0.0171	0.0030	0.0007	-0.0738	0.0004	0.0017	0.0083	0.0009	0.0005	0.0460	0.0001	-0.0000	0.0298
	MSE	0.0022	0.0015	0.0016	0.0019	0.0012	0.0077	0.0011	0.0009	0.0003	0.0004	0.0004	0.0032	0.0001	0.0001	0.0018
1000	Variance	0.0012	0.0007	0.0012	0.0011	0.0006	0.0031	0.0006	0.0004	0.0002	0.0002	0.0002	0.0012	0.0000	0.0000	0.0009
	Bias	-0.0034	-0.0018	-0.0194	0.0014	0.0002	-0.0586	0.0007	0.0013	0.0052	0.0013	0.0002	0.0369	0.0002	0.0002	0.0270
	MSE	0.0012	0.0007	0.0016	0.0011	0.0006	0.0065	0.0006	0.0004	0.0002	0.0002	0.0002	0.0026	0.0000	0.0000	0.0016

Table 10 Variance, bias and MSE of the first five estimated entries of the PMF for the different sample sizes n in case of a true $\text{Poi}(1)$ innovation distribution of an INAR(1) process. We report results for unpenalized (up) and L_2 penalized estimation using either first order (diff1) or second order (diff2) differences

n	g_0_up	g_0_d1	g_0_d2	g_1_up	g_1_d1	g_1_d2	g_2_up	g_2_d1	g_2_d2	g_3_up	g_3_d1	g_3_d2	g_4_up	g_4_d1	g_4_d2	
20	Variance	0.0458	0.0126	0.0181	0.0382	0.0066	0.0053	0.0321	0.0063	0.0057	0.0156	0.0053	0.0058	0.0029	0.0042	
	Bias	-0.0921	-0.0114	0.0238	-0.0217	-0.0371	-0.0573	0.0513	0.0220	0.0089	0.0376	0.0330	0.0212	0.0250	0.0294	
	MSE	0.0543	0.0127	0.0186	0.0387	0.0080	0.0086	0.0348	0.0068	0.0058	0.0170	0.0064	0.0063	0.0054	0.0045	0.0050
50	Variance	0.0246	0.0060	0.0092	0.0198	0.0037	0.0011	0.0155	0.0034	0.0017	0.0073	0.0025	0.0028	0.0012	0.0006	0.0011
	Bias	-0.0312	0.0003	0.0433	-0.0017	-0.0251	-0.0595	0.0102	0.0116	0.0014	0.0165	0.0118	0.0122	0.0037	0.0063	0.0035
	MSE	0.0256	0.0060	0.0111	0.0198	0.0043	0.0046	0.0156	0.0036	0.0017	0.0076	0.0026	0.0029	0.0012	0.0006	0.0011
100	Variance	0.0125	0.0035	0.0025	0.0114	0.0031	0.0006	0.0080	0.0027	0.0008	0.0031	0.0014	0.0008	0.0006	0.0003	0.0002
	Bias	-0.0225	-0.0030	0.0517	0.0032	-0.0216	-0.0606	0.0103	0.0105	-0.0006	0.0067	0.0136	0.0118	0.0019	0.0006	-0.0017
	MSE	0.0130	0.0035	0.0052	0.0114	0.0035	0.0042	0.0081	0.0028	0.0008	0.0032	0.0016	0.0009	0.0006	0.0003	0.0002
250	variance	0.0043	0.0019	0.0010	0.0039	0.0020	0.0004	0.0025	0.0012	0.0003	0.0010	0.0007	0.0003	0.0002	0.0002	0.0001
	Bias	-0.0075	-0.0053	0.0503	0.0006	-0.0164	-0.0591	0.0054	0.0120	-0.0001	0.0026	0.0084	0.0110	-0.0014	0.0011	-0.0011
	MSE	0.0044	0.0019	0.0035	0.0039	0.0022	0.0038	0.0025	0.0014	0.0003	0.0010	0.0008	0.0004	0.0002	0.0002	0.0001
500	Variance	0.0021	0.0010	0.0008	0.0019	0.0013	0.0005	0.0011	0.0006	0.0002	0.0004	0.0004	0.0002	0.0001	0.0001	0.0001
	Bias	-0.0044	-0.0050	0.0495	0.0030	-0.0142	-0.0577	0.0004	0.0096	-0.0013	0.0009	0.0081	0.0110	0.0001	0.0011	-0.0005
	MSE	0.0022	0.0010	0.0032	0.0019	0.0015	0.0039	0.0011	0.0007	0.0002	0.0004	0.0005	0.0003	0.0001	0.0001	0.0001
1000	Variance	0.0012	0.0006	0.0009	0.0011	0.0010	0.0009	0.0006	0.0004	0.0002	0.0002	0.0003	0.0001	0.0000	0.0000	0.0000
	Bias	-0.0006	0.0010	0.0036	0.0003	-0.0005	0.0001	-0.0001	-0.0016	-0.0060	0.0005	0.0012	0.0012	-0.0001	0.0003	0.0014
	MSE	0.0005	0.0005	0.0005	0.0004	0.0004	0.0002	0.0003	0.0003	0.0002	0.0002	0.0002	0.0001	0.0001	0.0001	0.0001

Table 11 Variance, bias and MSE of the first five estimated entries of the PMF for the different sample sizes n in case of a true ZIP($\lambda, 2$) innovation distribution of an INAR (1) process. We report results for unpenalized (up), L_1 and L_2 penalized estimation

n	g_0_up	g_0_L1	g_0_L2	g_1_up	g_1_L1	g_1_L2	g_2_up	g_2_L1	g_2_L2	g_3_up	g_3_L1	g_3_L2	g_4_up	g_4_L1	g_4_L2
20	Variance	0.0581	0.0287	0.0174	0.0293	0.0066	0.0174	0.0038	0.0045	0.0138	0.0042	0.0040	0.0072	0.0035	0.0026
	Bias	-0.1037	-0.1670	-0.1142	0.0424	0.0508	0.0752	0.0153	0.0289	0.0210	0.0400	0.0130	0.0165	0.0556	0.0179
	MSE	0.0689	0.0566	0.0305	0.0310	0.0086	0.0123	0.0176	0.0046	0.0050	0.0142	0.0058	0.0042	0.0075	0.0066
50	Variance	0.0205	0.0185	0.0104	0.0121	0.0044	0.0087	0.0025	0.0022	0.0052	0.0020	0.0016	0.0026	0.0017	0.0010
	Bias	-0.0242	-0.0785	-0.0810	0.0118	0.0321	0.0565	0.0066	0.0098	0.0119	0.0043	0.0128	0.0077	-0.0022	0.0111
	MSE	0.0211	0.0247	0.0169	0.0123	0.0054	0.0077	0.0087	0.0026	0.0023	0.0053	0.0022	0.0017	0.0026	0.0018
100	Variance	0.0067	0.0119	0.0086	0.0053	0.0027	0.0039	0.0038	0.0016	0.0013	0.0025	0.0013	0.0009	0.0007	0.0006
	Bias	-0.0132	-0.0496	-0.0598	0.0101	0.0194	0.0408	-0.0002	0.0090	0.0089	0.0028	0.0092	0.0059	0.0007	0.0045
	MSE	0.0069	0.0144	0.0122	0.0054	0.0031	0.0055	0.0038	0.0017	0.0014	0.0025	0.0014	0.0010	0.0013	0.0007
250	Variance	0.0023	0.0065	0.0063	0.0017	0.0019	0.0029	0.0013	0.0009	0.0006	0.0009	0.0006	0.0004	0.0005	0.0002
	Bias	-0.0032	-0.0274	-0.0528	0.0013	0.0121	0.0353	0.0004	0.0063	0.0082	0.0018	0.0038	0.0045	-0.0002	0.0029
	MSE	0.0023	0.0073	0.0091	0.0017	0.0021	0.0042	0.0013	0.0009	0.0007	0.0009	0.0006	0.0004	0.0005	0.0002
500	Variance	0.0009	0.0010	0.0035	0.0007	0.0006	0.0019	0.0006	0.0004	0.0003	0.0004	0.0003	0.0002	0.0002	0.0001
	Bias	0.0006	0.0010	-0.0298	0.0004	-0.0007	0.0230	-0.0012	-0.0008	0.0020	0.0008	0.0015	0.0030	0.0002	-0.0007
	MSE	0.0009	0.0010	0.0044	0.0007	0.0006	0.0025	0.0006	0.0004	0.0003	0.0004	0.0003	0.0002	0.0002	0.0001
1000	Variance	0.0005	0.0004	0.0016	0.0004	0.0003	0.0009	0.0003	0.0002	0.0002	0.0002	0.0002	0.0001	0.0001	0.0001
	Bias	-0.0006	-0.0011	-0.0100	0.0003	0.0007	0.0073	-0.0001	-0.0004	0.0008	0.0005	0.0009	0.0011	-0.0001	0.0001
	MSE	0.0005	0.0004	0.0017	0.0004	0.0003	0.0009	0.0003	0.0002	0.0002	0.0002	0.0002	0.0002	0.0001	0.0001

Table 12 Variance, bias and MSE of the first five estimated entries of the PMF for the different sample sizes n in case of a true ZIP $_{(\frac{1}{2}, 2)}$ innovation distribution of an INAR (1) process. We report results for unpenalized (up), L_1 and L_2 penalized estimation without smoothing of $G(0)$

n	$g0_up$	$g0_L1$	$g0_L2$	$g1_up$	$g1_L1$	$g1_L2$	$g2_up$	$g2_L1$	$g2_L2$	$g3_up$	$g3_L1$	$g3_L2$	$g4_up$	$g4_L1$	$g4_L2$	
20	Variance	0.0581	0.0320	0.0338	0.0293	0.0031	0.0090	0.0174	0.0025	0.0055	0.0138	0.0022	0.0038	0.0072	0.0021	0.0022
	Bias	-0.1037	-0.0027	0.0034	0.0424	-0.0208	-0.0077	0.0153	-0.0217	-0.0021	0.0211	0.0077	0.0048	0.0165	0.0352	0.0100
	MSE	0.0689	0.0320	0.0338	0.0310	0.0036	0.0090	0.0176	0.0030	0.0055	0.0142	0.0022	0.0039	0.0075	0.0033	0.0023
50	Variance	0.0205	0.0183	0.0132	0.0121	0.0021	0.0033	0.0087	0.0017	0.0020	0.0052	0.0014	0.0015	0.0026	0.0016	0.0011
	Bias	-0.0242	0.0178	0.0117	0.0118	-0.0237	-0.0007	0.0066	-0.0282	-0.0101	0.0043	0.0002	-0.0022	-0.0022	0.0150	0.0004
	MSE	0.0211	0.0186	0.0134	0.0123	0.0027	0.0033	0.0087	0.0025	0.0021	0.0053	0.0014	0.0015	0.0026	0.0018	0.0011
100	Variance	0.0067	0.0088	0.0054	0.0053	0.0016	0.0017	0.0038	0.0013	0.0010	0.0025	0.0010	0.0007	0.0013	0.0010	0.0006
	Bias	-0.0132	0.0261	0.0080	0.0101	-0.0217	0.0002	-0.0002	-0.0229	-0.0081	0.0028	0.0024	0.0012	0.0007	0.0073	0.0002
	MSE	0.0069	0.0095	0.0055	0.0054	0.0021	0.0017	0.0038	0.0018	0.0011	0.0025	0.0010	0.0007	0.0013	0.0010	0.0006
250	Variance	0.0023	0.0042	0.0018	0.0017	0.0014	0.0006	0.0013	0.0012	0.0004	0.0009	0.0006	0.0003	0.0005	0.0005	0.0002
	Bias	-0.0032	0.0249	0.0092	0.0013	-0.0148	-0.0011	0.0004	-0.0166	-0.0096	0.0018	0.0030	-0.0001	-0.0002	0.0035	0.0024
	MSE	0.0023	0.0048	0.0019	0.0017	0.0017	0.0006	0.0013	0.0015	0.0005	0.0009	0.0006	0.0003	0.0005	0.0005	0.0002
500	Variance	0.0009	0.0013	0.0009	0.0007	0.0007	0.0004	0.0006	0.0007	0.0003	0.0004	0.0005	0.0002	0.0002	0.0003	0.0002
	Bias	0.0006	0.0090	0.0057	0.0004	-0.0043	-0.0002	-0.0012	-0.0061	-0.0065	0.0008	0.0020	0.0004	0.0002	-0.0001	0.0011
	MSE	0.0009	0.0013	0.0009	0.0007	0.0007	0.0004	0.0006	0.0007	0.0003	0.0004	0.0005	0.0002	0.0002	0.0003	0.0002
1000	Variance	0.0005	0.0005	0.0005	0.0004	0.0004	0.0002	0.0003	0.0003	0.0001	0.0002	0.0002	0.0001	0.0001	0.0001	0.0001
	Bias	-0.0006	0.0010	0.0036	0.0003	-0.0005	0.0001	-0.0001	-0.0016	-0.0060	0.0005	0.0012	0.0012	-0.0001	0.0003	0.0014
	MSE	0.0005	0.0005	0.0005	0.0004	0.0004	0.0002	0.0003	0.0003	0.0002	0.0002	0.0002	0.0001	0.0001	0.0001	0.0001

Acknowledgements The authors thank the two referees for their useful comments on an earlier draft of this article. This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - Projektnummer 437270842.

Funding Open Access funding enabled and organized by Projekt DEAL.

Declarations

Conflict of interests The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Adam T, Langrock R, Weiß C (2019) Penalized estimation of flexible hidden Markov models for time series of counts. *METRON* 77:87–104
- Al-Osh MA, Alzaid AA (1987) First-order integer-valued autoregressive (INAR(1)) process. *J Time Ser Anal* 8(3):261–275
- Al-Osh MA, Alzaid AA (1990) An integer-valued p th order autoregressive structure (INAR(p)) process. *J Appl Probab* 27(2):314–324
- Brännäs K, Hellström J (2001) Generalized integer-valued autoregression. *Economet Rev* 20:425–443
- Bui MT, Potgieter CJ, Kamata A (2021) Penalized likelihood methods for modeling count data. *arXiv preprint arXiv:2109.14010*
- Core Team R (2021): A language and environment for statistical computing. URL <https://www.R-project.org/>
- Drost F, Van den Akker R, Werker B (2009) Efficient estimation of auto-regression parameters and innovation distributions for semiparametric integer-valued AR(p) models. *J Royal Stat Soc Ser B* 71:467–485
- Du JG, Li Y (1991) The integer valued autoregressive (INAR(p)) model. *J Time Ser Anal* 12(2):129–142
- Fahrmeir L, Kneib T, Lang S (2013) *Regression: models, methods and applications*. Springer
- Fokianos K (2010) Penalized estimation for integer autoregressive models. In: *Statistical modelling and regression structures*. Springer, pp 337–352
- Franke J, Seligmann T (1993) Conditional maximum-likelihood estimates for INAR(1) processes and their applications to modelling epileptic seizure counts. *Developments in Time Series*, pp 310–330
- Freeland R, McCabe B (2005) Asymptotic properties of cls estimators in the poisson AR(1) model. *Stat Probab Lett* 73:147–153
- Homburg A, Weiß C, Frahm G et al (2021) Analysis and forecasting of risk in count processes. *J Risk Finan Manag* 14(4):182
- Jazi M, Jones G, Lai C (2012) First-order integer valued AR processes with zero inflated Poisson innovations. *J Time Ser Anal* 33:954–963
- Jung R, Ronning G, Tremayne A (2005) Estimation in conditional first order autoregression with discrete support. *Stat Pap* 46:195–224
- Kim HY, Park Y (2008) A non-stationary integer-valued autoregressive model. *Stat Pap* 49(3):485–502
- Liu Z, Li Q, Zhu F (2021) Semiparametric integer-valued autoregressive models on \mathbb{Z} . *Can J Stat* 49:1317–1337

- McKenzie E (1985) Some simple models for discrete variate time series. *Water Resour Bull* 21(4):645–650
- Nardi Y, Rinaldo A (2011) Autoregressive process modeling via the lasso procedure. *J Multivar Anal* 102:528–549
- Scott DW, Tapia RA, Thompson JR (1980) Nonparametric probability density estimation by discrete maximum penalized-likelihood criteria. *Ann Stat* 8(4):820–832
- Snyder R (2002) Forecasting sales of slow and fast moving inventories. *Eur J Oper Res* 140:684–699
- Steutel FW, Van Harn K (1979) Discrete analogues of self-decomposability and stability. *Ann Probab* 7(5):893–899
- Syntetos AA, Boylan JE (2021) Intermittent demand forecasting: Context, methods and applications. John Wiley & Sons, UK
- Tibshirani R, Saunders M, Rosset S et al (2005) Sparsity and smoothness via the fused lasso. *J Roy Stat Soc B* 67(1):91–108
- Wang X (2020) Variable selection for first-order poisson integer-valued autoregressive model with covariables. *Aust N Z J Stat* 62(2):278–295
- Wang X, Wang D, Yang K (2021) Integer-valued time series model order shrinkage and selection via penalized quasi-likelihood approach. *Metrika* 84:713–750
- Weiß C (2018) *An Introduction to Discrete-Valued Time Series*, 1st edn. Wiley, UK
- Yang L (2019) The predictive distributions of thinning-based count processes. *Scand J Stat* 48(1):42–67

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Maxime Faymonville¹ · Carsten Jentsch¹ · Christian H. Weiß² · Boris Aleksandrov²

✉ Maxime Faymonville
faymonville@statistik.tu-dortmund.de

Carsten Jentsch
jentsch@statistik.tu-dortmund.de

Christian H. Weiß
weissc@hsu-hh.de

Boris Aleksandrov
aleksanb@hsu-hh.de

¹ Department of Statistics, TU Dortmund University, D-44221 Dortmund, Germany

² Department of Mathematics and Statistics, Helmut-Schmidt-University, D-22008 Hamburg, Germany