



The INARCH(1) Model for Overdispersed Time Series of Counts

Christian Weiss

► To cite this version:

Christian Weiss. The INARCH(1) Model for Overdispersed Time Series of Counts. Communications in Statistics - Simulation and Computation, 2010, 39 (06), pp.1269-1291. 10.1080/03610918.2010.490317 . hal-00602988

HAL Id: hal-00602988

<https://hal.science/hal-00602988>

Submitted on 24 Jun 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



The INARCH(1) Model for Overdispersed Time Series of Counts

| | |
|---|--|
| Journal: | <i>Communications in Statistics - Simulation and Computation</i> |
| Manuscript ID: | LSSP-2010-0003.R1 |
| Manuscript Type: | Original Paper |
| Date Submitted by the Author: | 28-Apr-2010 |
| Complete List of Authors: | Weiß, Christian; Darmstadt University of Technology, Department of Mathematics |
| Keywords: | ACP model, c control chart, INGARCH model, overdispersion, Poisson-Charlier expansion, simultaneous confidence regions |
| Abstract: | The INARCH(1) model for overdispersed time series of counts has a simple structure, a parsimonious parametrization, and a great potential for applications in practice. We analyze two approaches to approximate the marginal process distribution: a Markov chain approach and the Poisson-Charlier expansion. Then approaches for estimating the two model parameters are discussed. We derive explicit expressions for the asymptotic distribution of the maximum likelihood and conditional least squares estimators. They are used for constructing simultaneous confidence regions, the finite-sample performance of which is analyzed in a simulation study. A real-data example from economics illustrates the application of the INARCH(1) model. |
| Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online. | |
| INARCH1_CommStatSimComp2.tex | |

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



For Peer Review Only

The INARCH(1) Model for Overdispersed Time Series of Counts

Christian H. Weiß

Full address:

Christian H. Weiß

Darmstadt University of Technology

Department of Mathematics

Schlossgartenstraße 7

D-64289 Darmstadt

Germany

Email: weiss@mathematik.tu-darmstadt.de

Tel.: +49-6151-16-3787

Fax: +49-6151-16-6535

The INARCH(1) Model for Overdispersed
Time Series of Counts

Christian H. Weiß*

April 28, 2010

Abstract

The INARCH(1) model for overdispersed time series of counts has a simple structure, a parsimonious parametrization, and a great potential for applications in practice. We analyze two approaches to approximate the marginal process distribution: a Markov chain approach and the Poisson-Charlier expansion. Then approaches for estimating the two model parameters are discussed. We derive explicit expressions for the asymptotic distribution of the maximum likelihood

*Department of Mathematics, Darmstadt University of Technology, Schlossgartenstraße 7, D-64289 Darmstadt, Germany.

Corresponding author. Email: weiss@mathematik.tu-darmstadt.de

and conditional least squares estimators. They are used for constructing simultaneous confidence regions, the finite-sample performance of which is analyzed in a simulation study. A real-data example from economics illustrates the application of the INARCH(1) model.

KEY WORDS: ACP model; c control chart; INGARCH model; overdispersion; Poisson-Charlier expansion; simultaneous confidence regions.

1 Introduction

Count data processes $(X_t)_{\mathbb{Z}}$, where the X_t have state space $\mathbb{N}_0 = \{0, 1, \dots\}$ and where the time indices t are from $\mathbb{Z} = \{\dots, -1, 0, 1, \dots\}$, are commonly observed in real-world applications, e. g., in insurance (e. g., time series of claim counts), economics (e. g., counts of price changes), statistical process control (e. g., counts of defects), traffic (e. g., counts of accidents), network monitoring (e. g., as a part of an intrusion detection system), epidemiology (e. g., counts of cases of a certain disease) and others; for references, see (Heinen, 2003; Weiß, 2008). In line with this practical relevance, diverse models for count data time series have been proposed in literature, designed for different types of marginal distribution and autocorrelation structure, see, e. g., the recent reviews by (Kedem & Fokianos, 2002; Jung et al., 2006; Weiß, 2008). Particularly popular are ARMA-type models based on an ap-

appropriate type of thinning operation, which are referred to as integer-valued ARMA (INARMA) models. While models based on binomial thinning are commonly used for processes with Poisson marginals, alternative approaches like random coefficient thinning, iterated thinning or quasi-binomial thinning allow to consider marginal distributions with different types of *overdispersion* (Weiß, 2008). Overdispersion (i. e., a variance greater than the expectation) is commonly observed in practice. Typical reasons for overdispersion are the presence of positive correlation between the monitored events (Friedman, 1993; Poortema, 1999; Paroli et al., 2000) or a variation in the probability of the monitored events (Heimann, 1996; Poortema, 1999; Christensen et al., 2003); further potential causes of overdispersion are discussed by (Jackson, 1972).

An alternative approach for modeling time series of overdispersed counts are the rather novel *INGARCH models*, the *integer-valued* counterpart to the usual *generalized autoregressive conditional heteroskedasticity* models (for the latter, see (Bera & Higgins, 1993)). The INGARCH models were introduced by (Heinen, 2003)¹ and further investigated by (Ferland et al., 2006; Fokianos et al., 2009; Weiß, 2009; Zhu & Wang, 2009, 2010). The

¹(Heinen, 2003) referred to these models as the *autoregressive conditional Poisson (ACP)* models. But since these models are closely related to the cardinal GARCH models, (Ferland et al., 2006) suggested to refer to these models as the INGARCH models.

INGARCH models are defined by an ARMA-like recursion, and if the range of the model parameters is restricted in an appropriate manner, then a strictly stationary solution of this recursion exists, which also has finite first and second order moments (Ferland et al., 2006). In fact, these models even have an ARMA-like autocorrelation structure, which can be determined from a system of equations being similar to the usual Yule-Walker equations (Weiß, 2009).

In the following, we refer to the INGARCH($p, 0$) models as the *INARCH(p) models*, in analogy to the relation of the usual ARCH and GARCH models, also see (Zhu & Wang, 2009, 2010). These models can also be understood as special generalized linear models (GLM): Using the terminology of Section 1 in (Kedem & Fokianos, 2002), the INARCH(p) model is a GLM with Poisson distribution as random component and the identity link as systematic component. These purely autoregressive INARCH(p) models appear to be attractive especially from a practical point of view, since they allow to perform tasks like model identification, model estimation and forecasting relatively easily, also see (Zhu & Wang, 2009, 2010). Also compared to different model families, the INARCH models exhibit several advantages. In contrast to the overdispersed p^{th} order autoregressive models based on a thinning operation, for instance, they only require $p + 1$ instead of $p + 2$ parameters,

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

and they have simple transition probabilities as well as a less complex model structure if $p > 1$. An application of the INARCH models was reported by (Weiß, 2009), who analyzed a time series of monthly claims counts of workers in the heavy manufacturing industry (burn related injuries); this time series was originally presented by (Freeland, 1998). In fact, (Weiß, 2009) showed that already the simple two-parametric INARCH(1) model is able to describe both the observed serial dependence and overdispersion satisfactorily. Also (Zhu & Wang, 2010) proposed to use an INARCH(1) model for a time series of download counts. A further real-data example, where again an INARCH(1) model proves to be appropriate, is presented in Section 5 below: monthly strike data published by the U.S. Bureau of Labor Statistics. So in spite of its simple structure and its parsimonious parametrization, the INARCH(1) model seems to be of great practical relevance. It can be considered as a counterpart to the very popular Poisson INAR(1) model, but being able to describe overdispersion. For these reasons, we shall analyze the INARCH(1) model in more [details](#) in this article.

In Section 2, we review the definition and basic properties of the INARCH(1) model. Section 3 shows how to approximate the marginal process distribution with the help of the Poisson-Charlier expansion. The goodness of this approximation is investigated, also in view of approximating the average run

lengths (ARLs) of a c control chart. Section 4 discusses approaches for the estimation of the two model parameters. In particular, an explicit expression for the asymptotic distribution of the conditional least squares (CLS) estimators is derived. It can be used for constructing simultaneous confidence regions, the finite-sample performance of which is investigated in a simulation study. After having illustrated the application of these methods with the real-data example of Section 5, we conclude in Section 6.

2 The INARCH(1) Model: Definition and Basic Properties

In this section, we shall briefly review the definition and known basic properties of the INARCH(1) model.

2.1 Definition (INARCH(1) Model) *Let $(X_t)_{\mathbb{Z}}$ be a process with state space \mathbb{N}_0 , let $\beta > 0$ and $0 < \alpha < 1$. The process $(X_t)_{\mathbb{Z}}$ is said to follow an INARCH(1) model if X_t , conditioned on X_{t-1}, X_{t-2}, \dots , is Poisson distributed according to $Po(\beta + \alpha \cdot X_{t-1})$.*

The INARCH(1) model of Definition 2.1 is a stationary Markov chain (Ferland et al., 2006). Since the transition probabilities

$$p_{i|j} := P(X_t = i \mid X_{t-1} = j) = \exp(-\beta - \alpha \cdot j) \cdot \frac{(\beta + \alpha \cdot j)^i}{i!} > 0, \quad (1)$$

it is also irreducible and aperiodic and hence ergodic (Zhu & Wang, 2009), also see (Fokianos et al., 2009), who showed ergodicity for the more general INGARCH(1,1) model. All moments exist (Ferland et al., 2006), and its autocorrelation function $\rho_X(n) := Corr[X_t, X_{t-n}]$ simply equals α^n , like in the standard AR(1) case (Weiß, 2009). It has to be mentioned, however, that the value of α and therefore also of $\rho_X(n)$ is restricted to positive values, like in the case of the popular INAR(1) models. The marginal distribution of the INARCH(1) model can be expressed in terms of its cumulants: If $\mu_X(z) := E[\exp(z \cdot X_t)]$ denotes the moment generating function (mgf), the *cumulant generating function (cgf)* is defined as $\kappa_X(z) := \ln(\mu_X(z))$. The coefficients κ_r of the series expansion $\kappa_X(z) = \sum_{r=1}^{\infty} \frac{\kappa_r}{r!} \cdot z^r$ are referred to as the *cumulants* of X , with $\kappa_r = \kappa_X^{(r)}(0)$. The following Proposition 2.2, which was proven by (Weiß, 2009), summarizes essential results about the marginal cumulants of an INARCH(1) process.

2.2 Proposition (Marginal Cumulants) *Let $(X_t)_{\mathbb{Z}}$ be an INARCH(1) process according to Definition 2.1. Then the cumulants are determined recursively from*

$$\kappa_1 = \frac{\beta}{1-\alpha}, \quad \kappa_n = -(1-\alpha^n)^{-1} \cdot \sum_{j=1}^{n-1} s_{n,j} \cdot \kappa_j \quad \text{for } n \geq 2,$$

where the coefficients $s_{n,j}$ are the Stirling numbers of the first kind (Douglas,

1980; Appendix 9.1), given by

$$s_{n,0} = 0 \quad \text{and} \quad s_{n,n} = 1 \quad \text{for } n \geq 1,$$

$$s_{n+1,j} = s_{n,j-1} - n \cdot s_{n,j} \quad \text{for } j = 1, \dots, n \text{ and } n \geq 1.$$

In particular,

$$\kappa_1 = \frac{\beta}{1-\alpha}, \quad \kappa_2 = \frac{\beta}{(1-\alpha)(1-\alpha^2)}, \quad \kappa_3 = \frac{1+2\alpha^2}{1-\alpha^3} \cdot \kappa_2, \quad \kappa_4 = \frac{1+6\alpha^2+5\alpha^3+6\alpha^5}{(1-\alpha^3)(1-\alpha^4)} \cdot \kappa_2.$$

As a consequence of Proposition 2.2, mean μ_X , variance σ_X^2 , skewness and excess of X_t are given by $\frac{\beta}{1-\alpha}$, $\frac{\beta}{(1-\alpha)(1-\alpha^2)}$, $\frac{1+2\alpha^2}{1+\alpha+\alpha^2} \cdot \sqrt{\frac{1+\alpha}{\beta}}$ and $\frac{1+6\alpha^2+5\alpha^3+6\alpha^5}{\beta(1+\alpha+\alpha^2)(1+\alpha^2)}$, respectively. Since $\sigma_X^2 > \mu_X$, the INARCH(1) model allows to describe overdispersion.

3 The Marginal Distribution of INARCH(1) Processes

An explicit expression for the marginal distribution of an INARCH(1) process $(X_t)_{\mathbb{Z}}$ is not known. But since $(X_t)_{\mathbb{Z}}$ is an ergodic Markov chain, it follows for the marginal probabilities $p_i := P(X_t = i)$ that

$$p_i = \lim_{n \rightarrow \infty} p_{i|j}(n) \quad \text{for all } i, j \in \mathbb{N}_0, \quad (2)$$

where the n -step transition probabilities $p_{i|j}(n) := P(X_t = i \mid X_{t-n} = j)$ with $n \in \mathbb{N}$ are determined recursively via

$$p_{i|j}(n) = \sum_{r=0}^{\infty} p_{i|r} \cdot p_{r|j}(n-1). \quad (3)$$

Relations (2) and (3) allow to determine the marginal probabilities numerically up to an arbitrary precision: Choosing $M, N \in \mathbb{N}$ sufficiently large, we approximate

$$p_i \approx p_{i|j}(N), \quad \text{where } p_{i|j}(n) \approx \sum_{r=0}^M p_{i|r} \cdot p_{r|j}(n-1) \quad (4)$$

for arbitrary $i, j \in \mathbb{N}_0$. As an example, one may choose $j := \lceil \mu_X \rceil$ (smallest integer not less than μ_X).

This Markov chain approach has the disadvantage that it is computationally difficult and that it requires an appropriate choice of M, N , also see Section 3.1 below. Therefore, it would be desirable to find a more simple approximation of the marginal distribution. Since the marginal distribution is defined through assuming a conditional Poisson distribution, it appears to be a quite natural approach to try to approximate the marginal distribution with the help of the *Poisson-Charlier expansion* of (Barbour, 1987). Details and notations of this approach are outlined in Appendix A. The Poisson-Charlier approximations of the true marginal distribution of X_t incorporate the information about this distribution provided by the factorial cumulants $\kappa_{(n)}$, see formula (A.1).

3.1 Proposition (Marginal Factorial Cumulants) *Let $(X_t)_{\mathbb{Z}}$ be an IN-ARCH(1) process according to Definition 2.1. Then the factorial cumulants are determined from the usual cumulants, for the latter see Proposition 2.2,*

according to

$$\kappa_{(1)} = \kappa_1, \quad \kappa_{(n)} = \alpha^n \cdot \kappa_n \quad \text{for } n \geq 2.$$

The proof of Proposition 3.1 is provided by Appendix B.1. In Sections 3.1 and 3.2, we shall investigate the performance of the Poisson-Charlier approximations by comparing the marginal distributions *overally and by only* considering the effect on the ARL of a c control chart, respectively. We consider approximations up to order 4, see Table A.1, since higher order approximations become too complex for practice and one may apply the Markov chain approach instead.

3.1 Study: Approximate Marginal Distribution

To analyze the performance of the first four Poisson-Charlier approximations to the marginal distribution of the INARCH(1) process as given by Table A.1, we compared them to the true marginal distribution in terms of relative errors for a large variety of model parametrizations (β, α) . *For this purpose, we computed the true marginal distribution according to the Markov chain approach described by formula (4). We chose M such that a Poisson variate Y with mean $\beta + \alpha \cdot \lceil \mu_X \rceil$ would satisfy $P(Y = y) < 10^{-60}$ for $y > M$. Then we computed the corresponding transition matrix $\mathbf{P} = (p_{i|j})_{i,j=0,\dots,M}$ as well as the initial approximation $\mathbf{p}^{(0)} = (p_{0|\lceil \mu_X \rceil}, \dots, p_{M|\lceil \mu_X \rceil})^\top$. We repeated the*

iteration $\mathbf{p}^{(k+1)} = \mathbf{P}\mathbf{p}^{(k)}$ as long as the Chebycheff distance between $\mathbf{p}^{(k+1)}$ and $\mathbf{p}^{(k)}$ was larger than 10^{-15} , which determines N . In the following, the main results are reported and illustrated by selected graphs.

Independent of the concrete parametrization, it turned out that the approximations can only be used in the central region of the marginal distribution; for $k \rightarrow \infty$, the relative error of the approximations to the probability p_k tends to -100% , since the approximations decrease faster to 0 than the true probabilities. Otherwise, the performance of the approximations is only moderately influenced by the choice of β , but depends heavily on the choice of α . This is exemplified by the graphs of Figure 1, which express the performance for several values of α , and where β is always chosen such that the marginal mean equals 5 (a choice being motivated by the real-data example discussed later in Section 5). It becomes clear, see (a), that a pure Poisson approximation already leads to visible deviations for small values of α such as 0.1. The second order Poisson-Charlier approximation can be used at least for $\alpha \leq 0.25$, the third order one for $\alpha \leq 0.4$, see (b) and (c). The fourth order approximation still works well for $\alpha = 0.5$, see (d), but for $\alpha = 0.6$, we observe deviations between -11% and $+6\%$ even in the central region of the marginal distribution, see (e). In this situation, higher-order Poisson-Charlier approximations would be necessary to reach a satisfactory

performance, see Figure 2. But then also this type of approximation [requires complex computations](#), so the Markov chain approximation appears to be preferable. The problems for large α seem to be due to the fact that the marginal (factorial) cumulants (and therefore the whole marginal distribution), see Propositions 2.2 and 3.1, of the INARCH(1) model are seriously influenced by α .

So in a nutshell, satisfactory (and sufficiently simple at the same time) Poisson-Charlier approximations for the marginal probabilities can only be found for moderate values of α such as $\alpha \leq 0.5$, otherwise the Markov chain approximation is preferable. For such moderate values of α , the higher-order Poisson-Charlier approximations clearly outperform the simple Poisson approximation (and, by the way, also a negative binomial or generalized Poisson approximation). The Poisson-Charlier approximation might show a better performance also for large degrees of autocorrelation if applied to models, where the marginal distribution is separated more clearly from the serial dependence structure. This would be an interesting issue for future research, also see Section 6.

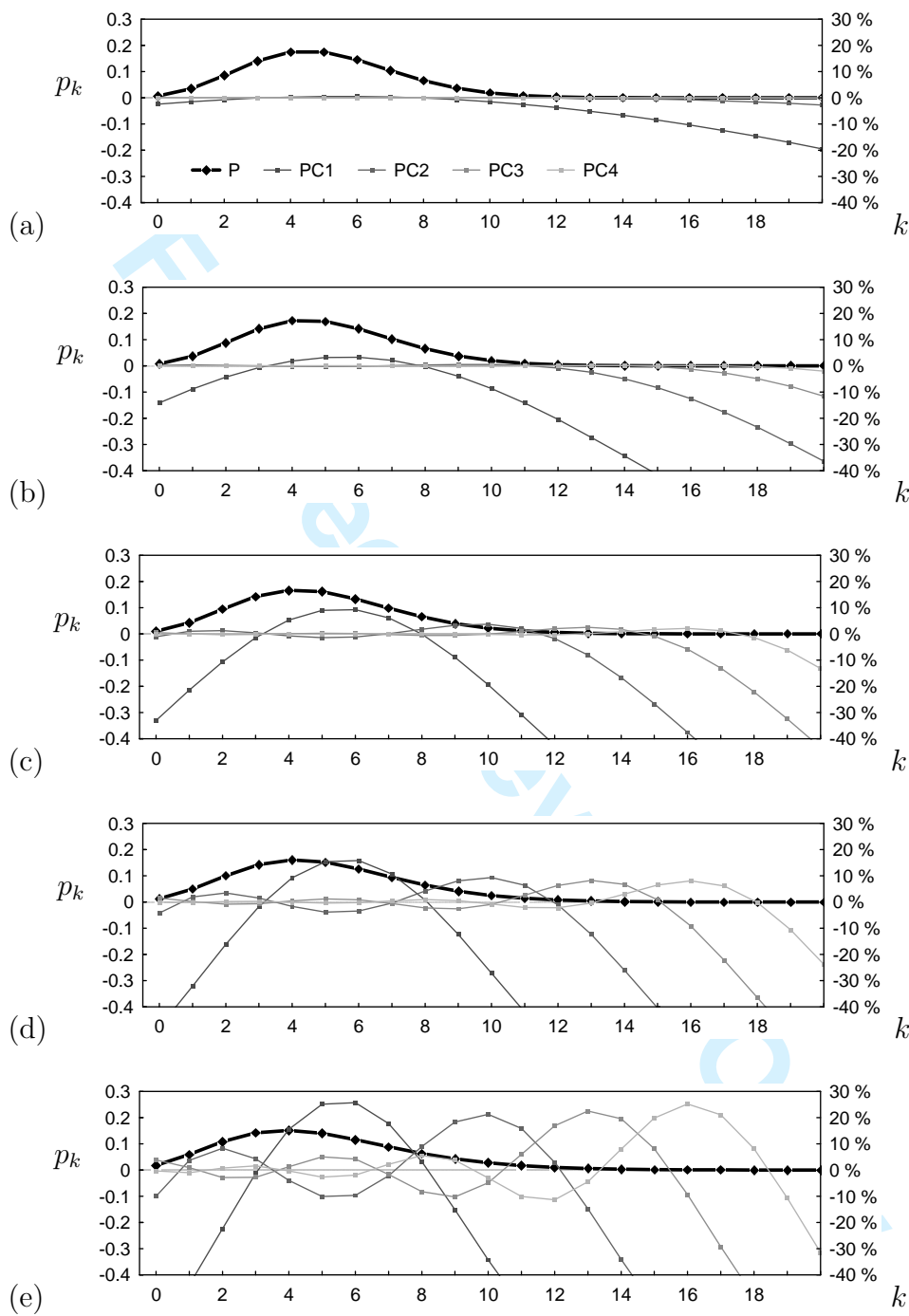


Figure 1: Marginal distribution (P) and relative errors of Poisson-Charlier approximations (PCn) for (a) $(\beta, \alpha) = (4.5, 0.1)$, (b) $(\beta, \alpha) = (3.75, 0.25)$, (c) $(\beta, \alpha) = (3, 0.4)$, (d) $(\beta, \alpha) = (2.5, 0.5)$, (e) $(\beta, \alpha) = (2, 0.6)$.

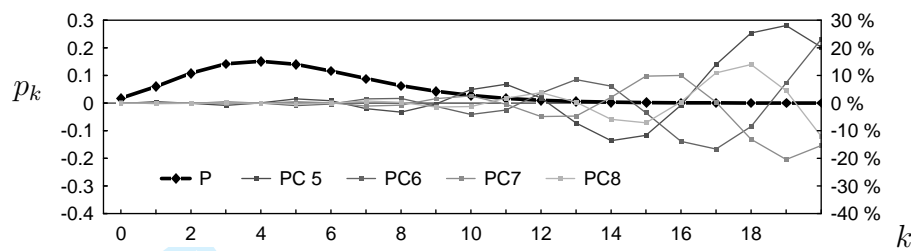


Figure 2: Marginal distribution (P) and relative errors of Poisson-Charlier approximations (PCn) for $(\beta, \alpha) = (2, 0.6)$, also see Figure 1 (e).

3.2 Study: Approximate ARLs of c Chart

A possible application, where knowledge about the central marginal probabilities of an INARCH(1) process is required, is the computation of average run lengths (ARLs) of control charts. Since such ARLs are computed by averaging about marginal probabilities, it seems possible that the Poisson-Charlier approximations lead to satisfactory results also for larger values of α .

For this purpose, c control charts for different in-control parametrizations and different types of shift in α and β were investigated. A c chart with lower and upper control limits LCL and UCL monitors the observed counts X_t directly. Since an INARCH(1) process is a Markov chain, ARLs can be computed with the well-known Markov chain approach of (Brook & Evans, 1972), which was done using either the exact marginal distribution or a Poisson-Charlier approximation. Some illustrative graphs are shown in Figure 3, where shifts in β compared to the respective in-control value β_0 are considered, while

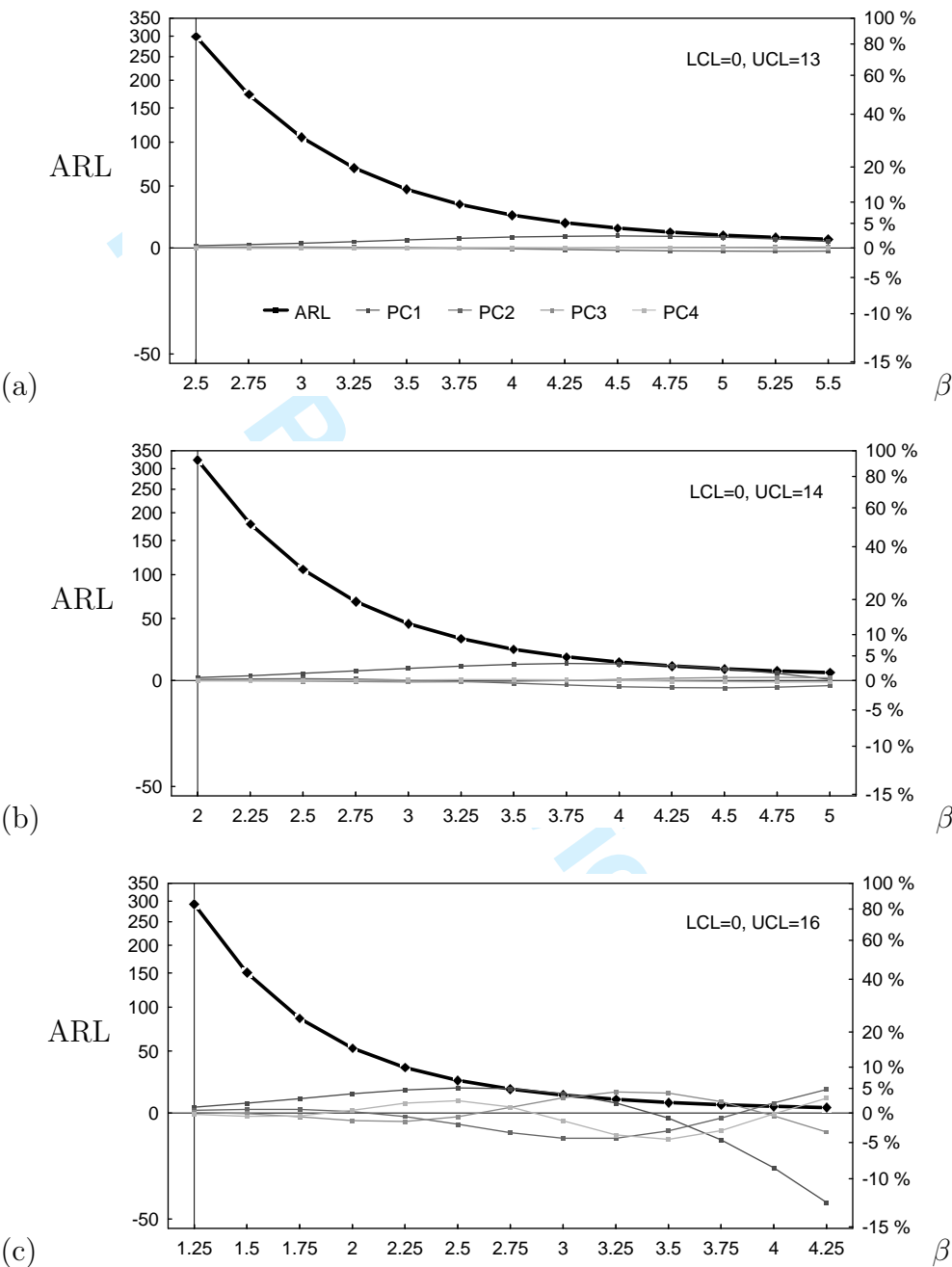


Figure 3: $ARL(\beta)$ of c chart and relative errors of Poisson-Charlier approximations (PC n) for (a) $(\beta_0, \alpha_0) = (2.5, 0.5)$, (b) $(\beta_0, \alpha_0) = (2, 0.6)$, (c) $(\beta_0, \alpha_0) = (1.25, 0.75)$.

$\alpha = \alpha_0$. It becomes clear that for $\alpha_0 \leq 0.5$, any Poisson-Charlier approximation of order ≥ 2 leads to a satisfactory approximation of the ARLs, also see (a). This result appears to be reasonable in view of the results of Section 3.1. For $\alpha_0 = 0.6$, at least the approximations of order ≥ 3 can be used, while these approximations lead to errors between -5% and $+5\%$ for $\alpha_0 = 0.75$.

4 Estimation of Parameters

Let x_1, \dots, x_T be a time series from an INARCH(1) process with parameters α, β according to Definition 2.1. In the following, we discuss and compare approaches for estimating these two model parameters, also see (Zhu & Wang, 2009). A related study for the case of an INGARCH(1,1) model was done by (Ferland et al., 2006; Fokianos et al., 2009).

4.1 Maximum Likelihood Approach

As already proposed by (Ferland et al., 2006), α and β can be estimated with a *conditional maximum likelihood (ML)* approach. This type of likelihood function is easy to compute, since the process has a conditional Poisson

distribution:

$$\begin{aligned}
 L(\beta, \alpha) &:= P(X_T = x_T, \dots, X_2 = x_2 \mid X_1 = x_1) \\
 &= \prod_{t=2}^T e^{-\beta - \alpha x_{t-1}} \cdot (\beta + \alpha x_{t-1})^{x_t} / x_t! \\
 &= e^{-(T-1)\beta} \cdot e^{-\alpha \sum_{t=2}^T x_{t-1}} \cdot \prod_{t=2}^T (\beta + \alpha x_{t-1})^{x_t} / \left(\prod_{t=2}^T x_t! \right).
 \end{aligned} \tag{5}$$

The ML estimates are either obtained by numerically maximizing the logarithmic likelihood function $\ell(\beta, \alpha) := \ln(L(\beta, \alpha))$ directly, or by solving the following system of equations:

$$\begin{aligned}
 \sum_{t=2}^T x_t / (\hat{\beta}_{\text{ML}} + \hat{\alpha}_{\text{ML}} x_{t-1}) &= T - 1, \\
 \hat{\beta}_{\text{ML}} &= \frac{1}{T-1} \cdot \left(\sum_{t=2}^T x_t - \hat{\alpha}_{\text{ML}} \cdot \sum_{t=2}^T x_{t-1} \right).
 \end{aligned} \tag{6}$$

(Zhu & Wang, 2009) analyzed the mean square error of the ML estimator in a simulation study.

The observed Fisher information $\mathbf{J}(\beta, \alpha)$, i. e., the negative Hessian of $\ell(\beta, \alpha)$, is given by $\mathbf{J}(\beta, \alpha) = \sum_{t=2}^T \mathbf{J}_t(\beta, \alpha)$ with

$$\mathbf{J}_t(\beta, \alpha) = \begin{pmatrix} \frac{x_t}{(\beta + \alpha x_{t-1})^2} & \frac{x_t x_{t-1}}{(\beta + \alpha x_{t-1})^2} \\ \frac{x_t x_{t-1}}{(\beta + \alpha x_{t-1})^2} & \frac{x_t x_{t-1}^2}{(\beta + \alpha x_{t-1})^2} \end{pmatrix}. \tag{7}$$

The proof of formulae (6) and (7) is provided by Appendix B.2. An explicit expression for the expectation of $\mathbf{J}_t(\beta, \alpha)$, the expected Fisher information $\mathbf{I}(\beta, \alpha)$ per observation, is not yet available. (Zhu & Wang, 2009; Theorem 3)

showed that the ML estimators are asymptotically normally distributed:

$$\sqrt{T-1} \cdot (\hat{\beta}_{\text{ML}} - \beta, \hat{\alpha}_{\text{ML}} - \alpha)^\top \underset{a}{\sim} N(\mathbf{0}, \mathbf{I}^{-1}(\beta, \alpha)), \quad (8)$$

where $\underset{a}{\sim}$ is an abbreviation for asymptotic distribution. In practice, the observed Fisher information $\mathbf{J}(\beta, \alpha)$ can be used to approximate $(T-1) \cdot \mathbf{I}(\beta, \alpha)$, and plugging in the obtained estimates allows, for instance, to approximate the asymptotic standard errors of the ML estimators, also see (Ferland et al., 2006). Related investigations (together with finite-sample results from a simulation study), but concerning the INGARCH(1,1) model, were done by (Fokianos et al., 2009).

4.2 Conditional Least Squares Approach

The *conditional least squares (CLS)* estimators are obtained by minimizing

$$CSS(\beta, \alpha) := \sum_{t=2}^T (x_t - E[X_t | X_{t-1} = x_{t-1}])^2 = \sum_{t=2}^T (x_t - \beta - \alpha x_{t-1})^2. \quad (9)$$

Evaluating the derivatives of (9), one obtains the explicit expression

$$\hat{\alpha}_{\text{CLS}} = \frac{\sum_{t=2}^T X_t X_{t-1} - \frac{1}{T-1} \cdot \sum_{t=2}^T X_t \cdot \sum_{s=2}^T X_{s-1}}{\sum_{t=2}^T X_{t-1}^2 - \frac{1}{T-1} \cdot \left(\sum_{t=2}^T X_{t-1} \right)^2}, \quad (10)$$

$$\hat{\beta}_{\text{CLS}} = \frac{1}{T-1} \left(\sum_{t=2}^T X_t - \hat{\alpha}_{\text{CLS}} \cdot \sum_{t=2}^T X_{t-1} \right)$$

for the CLS estimators, also see (Klimko & Nelson, 1978). The asymptotic distribution of the CLS estimators (10) is given by

$$\sqrt{T-1}(\hat{\beta}_{\text{CLS}} - \beta, \hat{\alpha}_{\text{CLS}} - \alpha)^\top \xrightarrow{D} N(\mathbf{0}, \Sigma_{\beta, \alpha}) \quad \text{with} \quad (11)$$

$$\Sigma_{\beta, \alpha} = \begin{pmatrix} \frac{\beta}{1-\alpha}(\beta(1+\alpha) + \frac{1+2\alpha^4}{1+\alpha+\alpha^2}) & -\beta(1+\alpha) - \frac{(1+2\alpha)\alpha^3}{1+\alpha+\alpha^2} \\ -\beta(1+\alpha) - \frac{(1+2\alpha)\alpha^3}{1+\alpha+\alpha^2} & (1-\alpha^2)(1 + \frac{\alpha(1+2\alpha^2)}{\beta(1+\alpha+\alpha^2)}) \end{pmatrix},$$

where \xrightarrow{D} abbreviates convergence in distribution. The proof of formula (11) is provided by Appendix B.3. Among others, formula (11) allows to approximate the asymptotic standard errors of the CLS estimators by plugging in the obtained estimates. Compared to the ML estimators of the previous Section 4.1, the CLS estimators have the advantage of being computed more easily (explicit formula instead of numerical maximization) and of having an explicit, data-independent expression for the asymptotic distribution, while the ML estimators usually have smaller standard errors within a class of estimating functions, see (Godambe & Heyde, 1987), as also indicated by the results of (Fokianos et al., 2009) concerning the INGARCH(1,1) model and by Table 5 below. A type of weighted CLS estimators, with the same asymptotic properties as the ML estimators of Section 4.1, was proposed by (Zhu & Wang, 2009), who also investigated its mean square error in a simulation study.

4.3 Method of Moments

Since the INARCH(1) process $(X_t)_{\mathbb{Z}}$ satisfies $E[X_t] = \beta/(1 - \alpha)$ and $\rho_X(1) = \alpha$, the method of moments implies to estimate α and β as

$$\hat{\alpha}_{\text{MM}} = \frac{\sum_{t=2}^T (X_t - \bar{X}_T)(X_{t-1} - \bar{X}_T)}{\sum_{t=1}^T (X_t - \bar{X}_T)^2}, \quad \hat{\beta}_{\text{MM}} = \bar{X}_T \cdot (1 - \hat{\alpha}_{\text{MM}}), \quad (12)$$

where $\bar{X}_T = \frac{1}{T} \sum_{t=1}^T X_t$ denotes the empirical mean of the segment X_1, \dots, X_T .

So both estimators (10) and (12) are nearly equal to each other. In fact, the arguments of Theorem 3 in (Freeland & McCabe, 2005) carry over, i. e., both estimators have the same asymptotic properties.

4.4 Simultaneous Confidence Regions

In the sequel, we shall propose different types of confidence region based either on the CLS estimators (10), the asymptotic distribution of which is given by formula (11), or based on the ML estimator of Section 4.1, where the inverse of the observed Fisher information (7) is used to approximate the true covariance.² We present results from a simulation study, where we investigated the performance of the confidence regions for time series of finite length. Explicit recommendations for the use of the confidence regions are derived.

²Note that the CLS and ML estimates can be computed using the available R functions `lm(data)` and `glm(data, family=poisson(link="identity"))`, respectively.

The first confidence region is based on the fact that $(\mathbf{Y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \boldsymbol{\mu})$ is χ_p^2 -distributed for a p -dimensional normal vector $\mathbf{Y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. If z denotes the γ -quantile of the χ_2^2 -distribution, then we obtain from the asymptotic distribution (11) of the CLS estimators (10)

$$\{(\beta, \alpha) \mid (\hat{\beta}_{\text{CLS}} - \beta, \hat{\alpha}_{\text{CLS}} - \alpha) \boldsymbol{\Sigma}_{\beta, \alpha}^{-1} (\hat{\beta}_{\text{CLS}} - \beta, \hat{\alpha}_{\text{CLS}} - \alpha)^\top < \frac{z}{T-1}\}, \quad (13)$$

which defines an asymptotically exact simultaneous confidence region for (β, α) on level γ , the CLS_1 confidence region. In complete analogy, we define the ML_1 confidence region based on the ML estimators as

$$\{(\beta, \alpha) \mid (\hat{\beta}_{\text{ML}} - \beta, \hat{\alpha}_{\text{ML}} - \alpha) \mathbf{J}(\beta, \alpha) (\hat{\beta}_{\text{ML}} - \beta, \hat{\alpha}_{\text{ML}} - \alpha)^\top < z\}. \quad (14)$$

The shape of these regions can be recognized from Figures 5 (a) and (c), respectively, which are computed from the data discussed in Section 5.

Because of computational reasons and the non-rectangular shape of regions (13) and (14), one might prefer an only approximate but rectangular confidence region for the parameters of the INARCH(1) model. Let z denote the $(3 + \gamma)/4$ -quantile of the $N(0, 1)$ -distribution, then the region obtained by replacing $\boldsymbol{\Sigma}_{\beta, \alpha}$ by $\boldsymbol{\Sigma}_{\hat{\beta}_{\text{CLS}}, \hat{\alpha}_{\text{CLS}}}$ and by applying the Bonferroni inequality equals

$$\begin{aligned} \{(\beta, \alpha) \mid (\hat{\beta}_{\text{CLS}} - \beta)^2 &< \frac{z^2}{T-1} \cdot \frac{\hat{\beta}_{\text{CLS}}(1+2\hat{\alpha}_{\text{CLS}}^4 + \hat{\beta}_{\text{CLS}}(1+\hat{\alpha}_{\text{CLS}})(1+\hat{\alpha}_{\text{CLS}}+\hat{\alpha}_{\text{CLS}}^2))}{1-\hat{\alpha}_{\text{CLS}}^3}, \\ (\hat{\alpha}_{\text{CLS}} - \alpha)^2 &< \frac{z^2}{T-1} \cdot \left(1 - \hat{\alpha}_{\text{CLS}}^2 + \frac{\hat{\alpha}_{\text{CLS}} + \hat{\alpha}_{\text{CLS}}^3 - 2\hat{\alpha}_{\text{CLS}}^5}{\hat{\beta}_{\text{CLS}}(1+\hat{\alpha}_{\text{CLS}}+\hat{\alpha}_{\text{CLS}}^2)}\right)\}. \end{aligned} \quad (15)$$

It is referred to as the CLS_2 confidence region, while we refer to

$$\begin{aligned} \{(\beta, \alpha) \mid (\hat{\beta}_{ML} - \beta)^2 < z^2 \cdot (\mathbf{J}(\hat{\beta}_{ML}, \hat{\alpha}_{ML})^{-1})_{11}, \\ (\hat{\alpha}_{ML} - \alpha)^2 < z^2 \cdot (\mathbf{J}(\hat{\beta}_{ML}, \hat{\alpha}_{ML})^{-1})_{22}\} \end{aligned} \quad (16)$$

as the ML_2 confidence region. The shape of these regions can be recognized from Figures 5 (b) and (d), respectively.

The practitioner is interested in the performance of these confidence regions for finite values of T , say $T = 100, 200, 500$ or 1000 . To estimate the true coverage probability $P(I_\gamma \ni (\beta, \alpha))$ of the confidence region I_γ , where (β, α) are the true process parameters, a simulation study was done. For each combination of the true parameters (β, α) , for each γ , T and each type of confidence region, 50,000 time series were generated. From each time series, the respective estimates and confidence region were computed. Then it was checked if this region contained the true parameter tuple. The number of ‘successes’ divided by 50,000 is an estimate of the true coverage probability. The results are tabulated in Tables 1 to 4.

The results of Table 1 show that the performance of the asymptotically exact CLS_1 region (13) depends heavily on the length T of the available time series and the true autocorrelation parameter α , but only slightly on the particular choice of γ . In general, the difference between γ and the estimated coverage probability (usually $< \gamma$) decreases for increasing T , as expected. But the

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table 1: Estimated coverage probabilities of the confidence region CLS_1 : (13).

| β | α | $\gamma = 0.90, T =$ | | | | $\gamma = 0.95, T =$ | | | |
|---------|----------|----------------------|-------|-------|-------|----------------------|-------|-------|-------|
| | | 100 | 200 | 500 | 1000 | 100 | 200 | 500 | 1000 |
| 2 | 0.2 | 0.895 | 0.898 | 0.898 | 0.901 | 0.944 | 0.947 | 0.947 | 0.950 |
| | 0.4 | 0.887 | 0.893 | 0.898 | 0.899 | 0.935 | 0.942 | 0.948 | 0.949 |
| | 0.6 | 0.867 | 0.884 | 0.893 | 0.896 | 0.917 | 0.933 | 0.942 | 0.948 |
| | 0.8 | 0.811 | 0.852 | 0.881 | 0.891 | 0.866 | 0.904 | 0.931 | 0.940 |
| 4 | 0.2 | 0.894 | 0.895 | 0.897 | 0.900 | 0.942 | 0.945 | 0.950 | 0.950 |
| | 0.4 | 0.884 | 0.891 | 0.898 | 0.899 | 0.934 | 0.940 | 0.947 | 0.949 |
| | 0.6 | 0.863 | 0.882 | 0.892 | 0.895 | 0.914 | 0.932 | 0.942 | 0.946 |
| | 0.8 | 0.808 | 0.850 | 0.882 | 0.889 | 0.864 | 0.902 | 0.932 | 0.940 |
| 6 | 0.2 | 0.893 | 0.898 | 0.900 | 0.899 | 0.944 | 0.947 | 0.949 | 0.949 |
| | 0.4 | 0.882 | 0.892 | 0.897 | 0.899 | 0.931 | 0.942 | 0.947 | 0.949 |
| | 0.6 | 0.864 | 0.883 | 0.894 | 0.897 | 0.913 | 0.932 | 0.943 | 0.947 |
| | 0.8 | 0.809 | 0.849 | 0.882 | 0.892 | 0.863 | 0.900 | 0.931 | 0.942 |

| β | α | $\gamma = 0.975, T =$ | | | | $\gamma = 0.99, T =$ | | | |
|---------|----------|-----------------------|-------|-------|-------|----------------------|-------|-------|-------|
| | | 100 | 200 | 500 | 1000 | 100 | 200 | 500 | 1000 |
| 2 | 0.2 | 0.968 | 0.972 | 0.974 | 0.975 | 0.984 | 0.988 | 0.989 | 0.989 |
| | 0.4 | 0.961 | 0.967 | 0.973 | 0.973 | 0.979 | 0.984 | 0.988 | 0.989 |
| | 0.6 | 0.945 | 0.959 | 0.968 | 0.973 | 0.967 | 0.978 | 0.986 | 0.989 |
| | 0.8 | 0.903 | 0.935 | 0.958 | 0.966 | 0.936 | 0.961 | 0.977 | 0.983 |
| 4 | 0.2 | 0.967 | 0.970 | 0.975 | 0.975 | 0.984 | 0.987 | 0.989 | 0.989 |
| | 0.4 | 0.961 | 0.967 | 0.972 | 0.975 | 0.980 | 0.985 | 0.988 | 0.989 |
| | 0.6 | 0.943 | 0.960 | 0.969 | 0.971 | 0.966 | 0.978 | 0.986 | 0.988 |
| | 0.8 | 0.900 | 0.934 | 0.959 | 0.965 | 0.931 | 0.959 | 0.978 | 0.983 |
| 6 | 0.2 | 0.969 | 0.972 | 0.973 | 0.974 | 0.985 | 0.987 | 0.989 | 0.989 |
| | 0.4 | 0.959 | 0.967 | 0.972 | 0.974 | 0.977 | 0.984 | 0.988 | 0.989 |
| | 0.6 | 0.943 | 0.958 | 0.969 | 0.972 | 0.967 | 0.977 | 0.985 | 0.988 |
| | 0.8 | 0.899 | 0.932 | 0.958 | 0.968 | 0.931 | 0.958 | 0.977 | 0.985 |

Table 2: Estimated coverage probabilities of the confidence region CLS_2 : (15).

| β | α | $\gamma = 0.90, T =$ | | | | $\gamma = 0.95, T =$ | | | |
|---------|----------|----------------------|-------|-------|-------|----------------------|-------|-------|-------|
| | | 100 | 200 | 500 | 1000 | 100 | 200 | 500 | 1000 |
| 2 | 0.2 | 0.920 | 0.922 | 0.924 | 0.922 | 0.957 | 0.959 | 0.962 | 0.961 |
| | 0.4 | 0.922 | 0.925 | 0.924 | 0.925 | 0.959 | 0.961 | 0.961 | 0.961 |
| | 0.6 | 0.925 | 0.927 | 0.927 | 0.927 | 0.960 | 0.961 | 0.962 | 0.962 |
| | 0.8 | 0.919 | 0.928 | 0.931 | 0.932 | 0.956 | 0.964 | 0.965 | 0.965 |
| 4 | 0.2 | 0.927 | 0.931 | 0.930 | 0.929 | 0.960 | 0.963 | 0.964 | 0.964 |
| | 0.4 | 0.928 | 0.929 | 0.931 | 0.933 | 0.962 | 0.963 | 0.966 | 0.966 |
| | 0.6 | 0.928 | 0.930 | 0.933 | 0.931 | 0.962 | 0.963 | 0.965 | 0.965 |
| | 0.8 | 0.920 | 0.930 | 0.932 | 0.934 | 0.956 | 0.962 | 0.965 | 0.965 |
| 6 | 0.2 | 0.928 | 0.931 | 0.932 | 0.932 | 0.962 | 0.964 | 0.965 | 0.966 |
| | 0.4 | 0.931 | 0.932 | 0.933 | 0.935 | 0.963 | 0.965 | 0.966 | 0.966 |
| | 0.6 | 0.930 | 0.931 | 0.935 | 0.935 | 0.963 | 0.963 | 0.966 | 0.966 |
| | 0.8 | 0.924 | 0.930 | 0.935 | 0.936 | 0.958 | 0.962 | 0.967 | 0.968 |

| β | α | $\gamma = 0.975, T =$ | | | | $\gamma = 0.99, T =$ | | | |
|---------|----------|-----------------------|-------|-------|-------|----------------------|-------|-------|-------|
| | | 100 | 200 | 500 | 1000 | 100 | 200 | 500 | 1000 |
| 2 | 0.2 | 0.976 | 0.978 | 0.980 | 0.980 | 0.990 | 0.990 | 0.992 | 0.992 |
| | 0.4 | 0.978 | 0.980 | 0.979 | 0.980 | 0.989 | 0.992 | 0.992 | 0.992 |
| | 0.6 | 0.979 | 0.980 | 0.980 | 0.980 | 0.991 | 0.991 | 0.992 | 0.992 |
| | 0.8 | 0.976 | 0.981 | 0.983 | 0.982 | 0.990 | 0.992 | 0.993 | 0.993 |
| 4 | 0.2 | 0.979 | 0.981 | 0.981 | 0.982 | 0.990 | 0.992 | 0.992 | 0.992 |
| | 0.4 | 0.979 | 0.980 | 0.982 | 0.982 | 0.990 | 0.991 | 0.993 | 0.993 |
| | 0.6 | 0.980 | 0.980 | 0.982 | 0.983 | 0.991 | 0.992 | 0.993 | 0.993 |
| | 0.8 | 0.975 | 0.981 | 0.982 | 0.982 | 0.989 | 0.991 | 0.993 | 0.992 |
| 6 | 0.2 | 0.979 | 0.981 | 0.982 | 0.983 | 0.991 | 0.992 | 0.992 | 0.993 |
| | 0.4 | 0.980 | 0.981 | 0.982 | 0.982 | 0.991 | 0.993 | 0.992 | 0.993 |
| | 0.6 | 0.980 | 0.981 | 0.982 | 0.983 | 0.991 | 0.992 | 0.993 | 0.993 |
| | 0.8 | 0.977 | 0.980 | 0.983 | 0.984 | 0.990 | 0.991 | 0.993 | 0.993 |

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table 3: Estimated coverage probabilities of the confidence region ML_1 : (14).

| β | α | $\gamma = 0.90, \quad T =$ | | | | $\gamma = 0.95, \quad T =$ | | | |
|---------|----------|----------------------------|-------|-------|-------|----------------------------|-------|-------|-------|
| | | 100 | 200 | 500 | 1000 | 100 | 200 | 500 | 1000 |
| 2 | 0.2 | 0.903 | 0.902 | 0.898 | 0.901 | 0.948 | 0.950 | 0.948 | 0.950 |
| | 0.4 | 0.900 | 0.900 | 0.900 | 0.901 | 0.948 | 0.948 | 0.949 | 0.951 |
| | 0.6 | 0.895 | 0.896 | 0.899 | 0.898 | 0.945 | 0.946 | 0.949 | 0.949 |
| | 0.8 | 0.885 | 0.888 | 0.897 | 0.898 | 0.937 | 0.941 | 0.947 | 0.947 |
| 4 | 0.2 | 0.901 | 0.902 | 0.899 | 0.904 | 0.950 | 0.950 | 0.950 | 0.953 |
| | 0.4 | 0.900 | 0.899 | 0.899 | 0.898 | 0.950 | 0.949 | 0.947 | 0.948 |
| | 0.6 | 0.897 | 0.898 | 0.897 | 0.901 | 0.946 | 0.948 | 0.949 | 0.950 |
| | 0.8 | 0.884 | 0.891 | 0.898 | 0.896 | 0.939 | 0.943 | 0.948 | 0.948 |
| 6 | 0.2 | 0.902 | 0.900 | 0.898 | 0.899 | 0.949 | 0.950 | 0.949 | 0.950 |
| | 0.4 | 0.898 | 0.899 | 0.897 | 0.899 | 0.949 | 0.948 | 0.949 | 0.949 |
| | 0.6 | 0.893 | 0.898 | 0.899 | 0.901 | 0.944 | 0.947 | 0.949 | 0.950 |
| | 0.8 | 0.882 | 0.891 | 0.898 | 0.897 | 0.938 | 0.944 | 0.949 | 0.949 |

| β | α | $\gamma = 0.975, \quad T =$ | | | | $\gamma = 0.99, \quad T =$ | | | |
|---------|----------|-----------------------------|-------|-------|-------|----------------------------|-------|-------|-------|
| | | 100 | 200 | 500 | 1000 | 100 | 200 | 500 | 1000 |
| 2 | 0.2 | 0.972 | 0.974 | 0.974 | 0.975 | 0.987 | 0.989 | 0.989 | 0.989 |
| | 0.4 | 0.971 | 0.973 | 0.974 | 0.975 | 0.987 | 0.988 | 0.989 | 0.990 |
| | 0.6 | 0.969 | 0.972 | 0.974 | 0.974 | 0.985 | 0.988 | 0.989 | 0.989 |
| | 0.8 | 0.964 | 0.968 | 0.972 | 0.973 | 0.983 | 0.985 | 0.989 | 0.990 |
| 4 | 0.2 | 0.975 | 0.975 | 0.975 | 0.976 | 0.989 | 0.990 | 0.989 | 0.990 |
| | 0.4 | 0.973 | 0.973 | 0.973 | 0.974 | 0.988 | 0.989 | 0.988 | 0.990 |
| | 0.6 | 0.971 | 0.973 | 0.975 | 0.975 | 0.988 | 0.989 | 0.990 | 0.990 |
| | 0.8 | 0.967 | 0.970 | 0.974 | 0.974 | 0.986 | 0.987 | 0.989 | 0.989 |
| 6 | 0.2 | 0.974 | 0.974 | 0.974 | 0.974 | 0.989 | 0.989 | 0.989 | 0.990 |
| | 0.4 | 0.974 | 0.974 | 0.975 | 0.975 | 0.989 | 0.989 | 0.990 | 0.989 |
| | 0.6 | 0.970 | 0.974 | 0.974 | 0.975 | 0.987 | 0.990 | 0.990 | 0.990 |
| | 0.8 | 0.968 | 0.971 | 0.974 | 0.976 | 0.986 | 0.988 | 0.989 | 0.990 |

Table 4: Estimated coverage probabilities of the confidence region ML_2 : (16).

| β | α | $\gamma = 0.90, T =$ | | | | $\gamma = 0.95, T =$ | | | |
|---------|----------|----------------------|-------|-------|-------|----------------------|-------|-------|-------|
| | | 100 | 200 | 500 | 1000 | 100 | 200 | 500 | 1000 |
| 2 | 0.2 | 0.921 | 0.921 | 0.923 | 0.926 | 0.958 | 0.960 | 0.960 | 0.962 |
| | 0.4 | 0.923 | 0.922 | 0.922 | 0.927 | 0.960 | 0.959 | 0.959 | 0.963 |
| | 0.6 | 0.920 | 0.926 | 0.927 | 0.925 | 0.959 | 0.960 | 0.963 | 0.962 |
| | 0.8 | 0.915 | 0.921 | 0.923 | 0.926 | 0.956 | 0.959 | 0.960 | 0.962 |
| 4 | 0.2 | 0.930 | 0.929 | 0.931 | 0.930 | 0.964 | 0.963 | 0.964 | 0.965 |
| | 0.4 | 0.931 | 0.932 | 0.931 | 0.931 | 0.964 | 0.964 | 0.964 | 0.965 |
| | 0.6 | 0.929 | 0.929 | 0.930 | 0.930 | 0.963 | 0.964 | 0.965 | 0.963 |
| | 0.8 | 0.920 | 0.926 | 0.929 | 0.932 | 0.958 | 0.961 | 0.964 | 0.966 |
| 6 | 0.2 | 0.934 | 0.936 | 0.934 | 0.934 | 0.966 | 0.967 | 0.965 | 0.966 |
| | 0.4 | 0.933 | 0.935 | 0.935 | 0.934 | 0.965 | 0.967 | 0.967 | 0.965 |
| | 0.6 | 0.932 | 0.934 | 0.934 | 0.936 | 0.964 | 0.967 | 0.967 | 0.968 |
| | 0.8 | 0.924 | 0.931 | 0.932 | 0.934 | 0.959 | 0.964 | 0.965 | 0.967 |

| β | α | $\gamma = 0.975, T =$ | | | | $\gamma = 0.99, T =$ | | | |
|---------|----------|-----------------------|-------|-------|-------|----------------------|-------|-------|-------|
| | | 100 | 200 | 500 | 1000 | 100 | 200 | 500 | 1000 |
| 2 | 0.2 | 0.977 | 0.979 | 0.980 | 0.980 | 0.990 | 0.991 | 0.992 | 0.992 |
| | 0.4 | 0.979 | 0.979 | 0.980 | 0.981 | 0.991 | 0.991 | 0.992 | 0.992 |
| | 0.6 | 0.978 | 0.980 | 0.980 | 0.981 | 0.990 | 0.991 | 0.991 | 0.992 |
| | 0.8 | 0.977 | 0.979 | 0.980 | 0.981 | 0.990 | 0.991 | 0.992 | 0.992 |
| 4 | 0.2 | 0.982 | 0.981 | 0.982 | 0.982 | 0.992 | 0.993 | 0.993 | 0.993 |
| | 0.4 | 0.982 | 0.981 | 0.982 | 0.982 | 0.992 | 0.992 | 0.993 | 0.993 |
| | 0.6 | 0.980 | 0.982 | 0.982 | 0.981 | 0.992 | 0.993 | 0.993 | 0.992 |
| | 0.8 | 0.978 | 0.979 | 0.982 | 0.982 | 0.990 | 0.992 | 0.993 | 0.992 |
| 6 | 0.2 | 0.983 | 0.984 | 0.982 | 0.983 | 0.993 | 0.994 | 0.993 | 0.993 |
| | 0.4 | 0.982 | 0.983 | 0.983 | 0.984 | 0.993 | 0.993 | 0.993 | 0.994 |
| | 0.6 | 0.981 | 0.983 | 0.982 | 0.984 | 0.993 | 0.993 | 0.993 | 0.993 |
| | 0.8 | 0.979 | 0.982 | 0.982 | 0.982 | 0.991 | 0.993 | 0.993 | 0.993 |

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

difference increases for increasing α . In contrast, the ML_1 region (14), see Table 3, appears to be much more robust against T and α , and is therefore clearly preferable especially for small T .

The approximate confidence regions (15) and (16), see Tables 2 and 4, derived from the Bonferroni inequality are always conservative, i. e., their true coverage probability is larger than the nominal level, although this difference decreases for increasing γ . Compared to the asymptotically exact region, the Bonferroni regions lead to an equal or even better performance for large values of γ .

From a practical point of view, the following recommendations can be derived from the Tables 1 and 4: The ML_1 region (14) appears to be a good choice for any α, γ, T . In contrast, the CLS_1 region (13) leads to a reliable result for a small degree of autocorrelation, say $\alpha \leq 0.2$, or if the sample size satisfies $T \geq 500$ ($0.4 \leq \alpha \leq 0.6$) or even $T \geq 1000$ ($\alpha \approx 0.8$); it is interesting to note that (Fokianos et al., 2009) found similar restrictions for the sample size concerning the INGARCH(1,1) model. In those situations, where the true coverage probability of the CLS_1 region is too small, i. e., if T is small and α large, the approximate and rectangular Bonferroni regions (15) or (16) can be used as a further alternative, at least if $\gamma \geq 0.95$.

5 Real-Data Example: Counts of Strikes

In the following, we analyze monthly strike data published by the U.S. Bureau of Labor Statistics (<http://www.bls.gov/wsp/>). The considered time series of counts describes the number of work stoppages leading to 1000 workers or more being idle in effect in the period. We restrict to the period from January 1994 to December 2002 (108 observations), which was already investigated by (Jung et al., 2005). A line plot of the data is shown in Figure 4 (a), and from the partial autocorrelation function in (c), it becomes clear that a first order autoregressive model seems to be reasonable. (Jung et al., 2005) fitted a Poisson INAR(1) model to the data, but since the estimates obtained with different methods deviated heavily from each other, they concluded that such a model is not appropriate. And in fact, see (Jung et al., 2005), the data exhibits overdispersion (the empirical variance equals 7.92, being much larger than the mean 4.94), making the Poisson marginal distribution an unreasonable choice, also see Figure 4 (d).

Therefore, we fit an INARCH(1) model to the data, using the approaches discussed in Section 4 before. The estimates obtained with different methods, see Table 5, are close to each other; the numbers shown in parentheses are the respective asymptotic standard errors computed as described in Section 4.

The variance of the ML estimated INARCH(1) model equals 8.37, being

close to the empirical variance 7.92. Also an analysis of the corresponding empirical residuals

$$\hat{\epsilon}_t := x_t - \hat{E}[X_t | X_{t-1} = x_{t-1}] = x_t - \hat{\beta} - \hat{\alpha} \cdot x_{t-1} \quad \text{for } t \geq 2$$

shows that the fitted INARCH(1) model seems to be a good choice. Finally, we also compare the INARCH(1) model to different types of AR(1)-like model for time series of counts: the two-parametric equidispersed Poisson INAR(1) model and the three-parametric overdispersed negative binomial random coefficient thinning INAR(1) model (NB RCINAR(1)), negative binomial iterated thinning INAR(1) model (NB IINAR(1)) and generalized Poisson quasi-binomial thinning INAR(1) model (GP QINAR(1)). For a description of these models as well as for their parametrization, see (Weiß, 2008). The corresponding ML estimates and asymptotic standard errors (again shown in parentheses and computed from the respective observed Fisher information) are given in Table 5. The performance of these models is measured in terms of the information criteria AIC and BIC, see Table 5, and it becomes clear that the INARCH(1) model clearly performs best.

Finally, we computed the simultaneous confidence regions according to Section 4.4 for different values of γ ; the results are shown in Figure 5. Obviously, the estimated value of α deviates significantly from 0 in all cases. The performance of these confidence regions can be recognized from the results of

Table 5: Estimated parameters of several models.

| Model | Method | Par. 1 | Par. 2 | Par. 3 | AIC | BIC |
|--|--------|--------------------|---------------------|--------------------|-------|-------|
| INARCH(1) (α, β) | MM | 0.5735 (0.0881) | 2.1090 (0.4277) | | 464.3 | 469.7 |
| | CLS | 0.5841 (0.0876) | 2.0716 (0.4273) | | | |
| | ML | 0.6364 (0.0807) | 1.8114 (0.3860) | | | |
| Po INAR(1) (α, λ) | ML | 0.5061 (0.0561) | 2.4603 (0.2989) | | 473.1 | 478.5 |
| NB RCINAR(1) (ρ, n, p) | ML | 0.5918 (0.0632) | 10.1779 (4.8583) | 0.6701 (0.1076) | 467.1 | 475.2 |
| NB IINAR(1) (ρ, n, α) | ML | 0.6308 (0.0797) | 7.7564 (3.8682) | 4.2181 (1.5811) | 466.8 | 474.8 |
| GP QINAR(1) (ρ, λ, θ) | ML | 0.5928 (0.0627) | 4.0912 (0.4826) | 0.1841 (0.0667) | 466.9 | 475.0 |

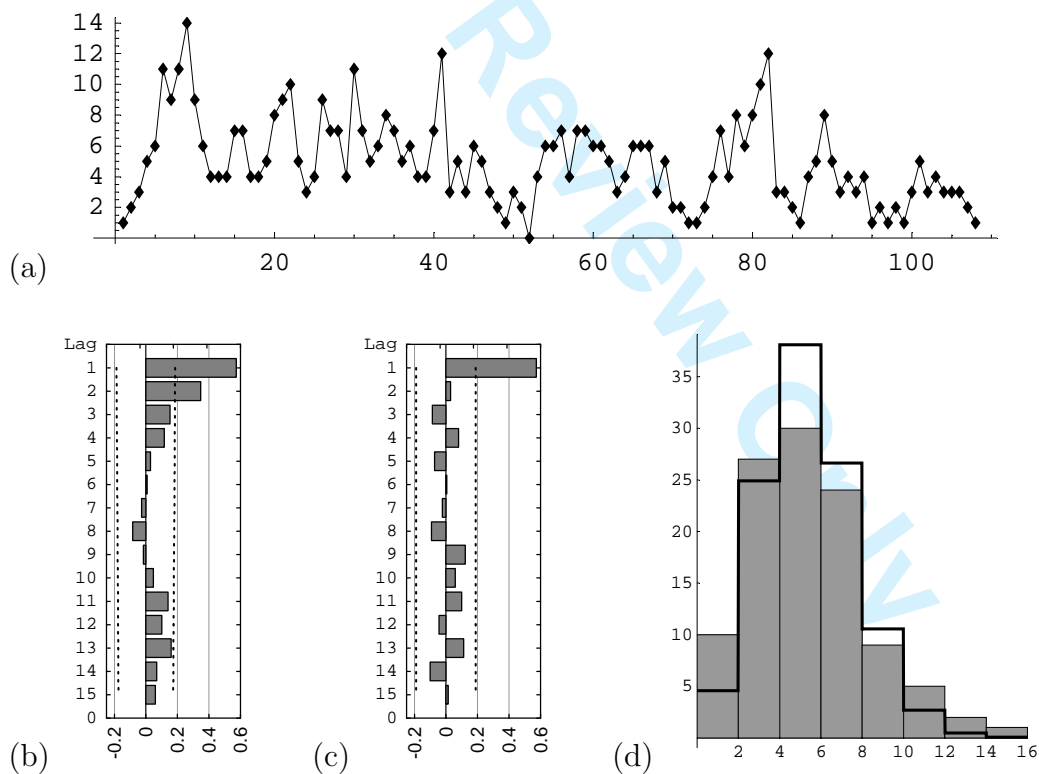


Figure 4: Monthly strike data: (a) Line plot, (b) empirical autocorrelation and (c) partial autocorrelation function, (d) histogram with $Po(4.94)$ fit.

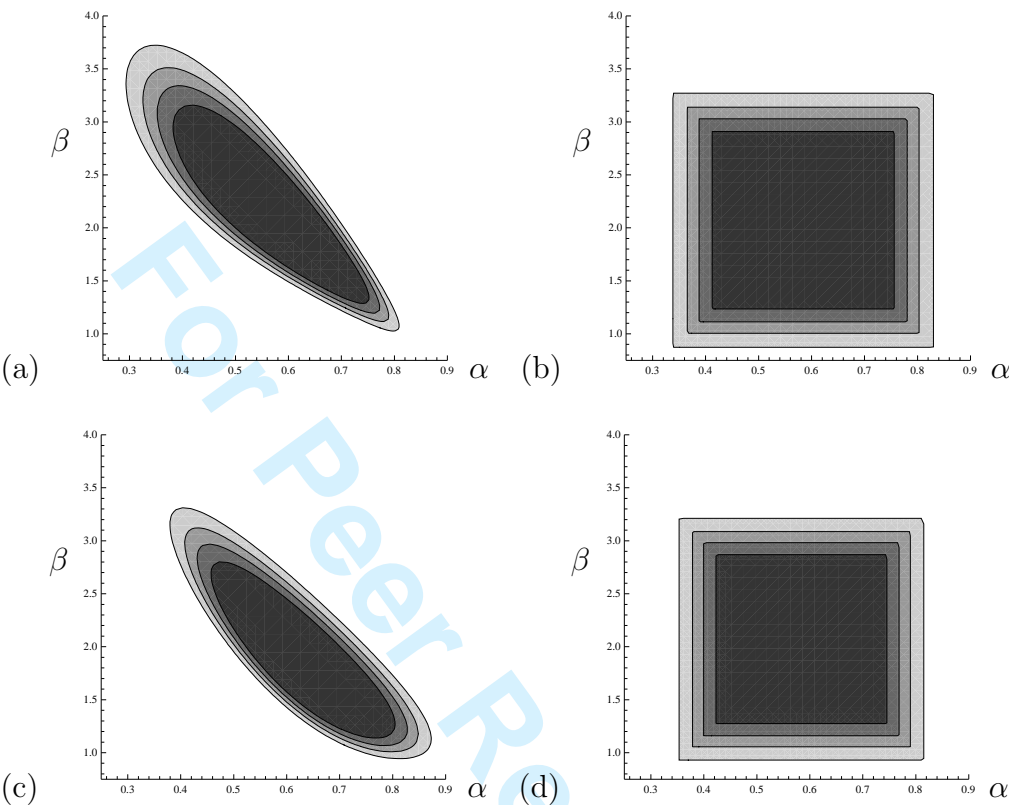


Figure 5: Simultaneous confidence regions according to Section 4.4: (a) and (b) based on CLS estimates, (c) and (d) based on ML estimates. Confidence levels $\gamma = 0.9$ (darkest gray), 0.95, 0.975 and 0.99 (lightest gray).

Tables 1 to 4, where the cells highlighted in gray give the estimated coverages for an INARCH(1) process with true parameters $\beta = 2$ and $\alpha = 0.6$, being close to the situation considered here. It becomes clear that the confidence region CLS_1 has a lower coverage, while ML_1 reaches its nominal level quite well. Also the approximate regions CLS_2 and ML_2 perform quite well for $\gamma \geq 0.975$.

6 Conclusion and Future Work

In this article, we recommended the INARCH(1) model as a simple and parsimoniously parametrized model for time series of overdispersed counts. After having reviewed the definition and basic properties of the INARCH(1) model, we showed how to approximate the marginal process distribution with the help of the Poisson-Charlier expansion. It became clear that this approximation always leads to an improvement compared to a Poisson approximation, but can be applied really satisfactorily only for INARCH(1) processes with an at most moderate degree of autocorrelation; otherwise, the more complex Markov chain method is preferable. Then we considered the problem of parameter estimation and derived explicit expressions for the asymptotic distribution of the ML and CLS estimators. Furthermore, we constructed several simultaneous confidence regions for the two model parameters, the performance of which is influenced by the length of the available time series and the degree of autocorrelation. A real-data example showed that the INARCH(1) model has a great potential for applications in practice.

A particularly important issue for future research is to investigate if the Poisson-Charlier expansion might be applicable also to different types of count data time series. For none of the INGARCH models, the marginal distribution is explicitly known (comparable to the case of the usual GARCH

models), but which is also true for different model types like the INAR(p) models of (Du & Li, 1991).

Acknowledgements. The author would like to thank the three referees for many useful comments on an earlier draft of this article.

References

Audenaert, K. (2009). Inverse moments of univariate discrete distributions via the Poisson expansion. arXiv:0809.4155v1. Submitted to *Sankhya A*.

Barbour, A.D. (1987). Asymptotic expansions in the Poisson limit theorem. *Ann. Prob.* 15(2):748-766.

Bera, A.K., Higgins, M.L. (1993). ARCH models: Properties, estimation and testing. *Jour. Economic Surveys* 7(4):305-366.

Brook, D., Evans, D.A. (1972). An approach to the probability distribution of cusum run length. *Biometrika* 59(3):539-549.

Christensen, A., Melgaard, H., Iwersen, J., Thyregod, P. (2003). Environmental monitoring based on a hierarchical Poisson-Gamma model. *Jour. Qual. Tech.* 35(3):275-285.

Douglas, J.B. (1980). *Analysis with standard contagious distributions*. Int. Co-operative Publishing House, Fairland, Maryland USA.

- 1
2
3
4
5
6
7
8 Du, J.-G., Li, Y. (1991). The integer-valued autoregressive (INAR(p)) model. *Jour.*
9
10 *Time Series Analysis* 12(2):129-142.
11
12
13
14 Ferland, R., Latour, A., Oraichi, D. (2006). Integer-valued GARCH processes.
15
16 *Jour. Time Series Analysis* 27(6):923-942.
17
18
19 Fokianos, K., Rahbek, A., Tjøstheim, D. (2009). *Poisson autoregression. Jour.*
20
21 *Amer. Stat. Assoc.* 10(4):1430-1439.
22
23
24
25 Freeland, R.K. (1998). *Statistical analysis of discrete time series with applications*
26
27 *to the analysis of workers compensation claims data.* PhD thesis, University of
28
29 British Columbia, Canada.
30
31
32 Freeland, R.K., McCabe, B. (2005). Asymptotic properties of CLS estimators in
33
34 the Poisson AR(1) model. *Stat. Prob. Letters* 73(2):147-153.
35
36
37
38 Friedman, D.J. (1993). Some considerations in the use of quality control techniques
39
40 in integrated circuit fabrication. *Int. Stat. Review* 61(1):97-107.
41
42
43
44 Godambe, V.P., Heyde, C.C. (1987). Quasi-likelihood and optimal estimation. *Int.*
45
46 *Stat. Review* 55(3):231-244.
47
48
49 Heimann, P.A. (1996). Attributes control charts with large sample sizes. *Jour.*
50
51 *Qual. Tech.* 28(4):451-459.
52
53
54
55 Heinen, A. (2003). Modelling Time Series Count Data: An Autoregressive Condi-
56
57
58
59
60

- 1
2
3
4
5
6
7
8 tional Poisson Model. *CORE Discussion Paper*, 2003-63, University of Louvain,
9 Belgium.
10
11
12
13
14 Jackson, J.E. (1972). All count distributions are not alike. *Jour. Qual. Tech.*
15
16 4(2):86-92.
17
18
19 Jung, R.C, Ronning, G., Tremayne, A.R. (2005). Estimation in Conditional First
20
21 Order Autoregression with Discrete Support. *Stat. Papers* 46:195-224.
22
23
24 Jung, R.C., Kukuk, M., Liesenfeld, R. (2006). Time series of count data: Mod-
25
26 elling, estimation and diagnostics. *Comp. Stat. Data Analysis* 51:2350-2364.
27
28
29
30 Kedem, B., Fokianos, K. (2002). *Regression models for time series analysis*. J.
31
32 Wiley & Sons, Inc., Hoboken, New Jersey.
33
34
35 Klimko, L.A., Nelson, P.I. (1978). On conditional least squares estimation for
36
37 stochastic processes. *Ann. Stat.* 6(3):629-642.
38
39
40
41 Paroli, R., Redaelli, G., Spezia, L. (2000). Poisson hidden Markov models for
42
43 time series of overdispersed insurance counts. *ASTIN Colloquium International*
44
45 *Actuarial Association*, Brussels, Belgium, 461-474.
46
47
48
49 Poortema, K. (1999). On modelling overdispersion of counts. *Statistica Neerlandica*
50
51 53(1):5-20.
52
53
54
55 Weiß, C.H. (2008). Thinning operations for modelling time series of counts – A
56
57 survey. *Adv. Stat. Analysis* 92(3):319-341.
58
59
60

Wei, C.H. (2009). Modelling time series of counts with overdispersion. *Stat. Meth.*

Appl. 18(4):507-519.

Zhu, F., Wang, D. (2009). Estimation and testing for a Poisson autoregressive model. To appear in *Metrika*.

Zhu, F., Wang, D. (2010). Diagnostic checking integer-valued ARCH(p) models using conditional residual autocorrelations. *Comp. Stat. Data Analysis* 54:496-508.

A The Poisson-Charlier Approximation

The Poisson-Charlier expansion was introduced by (Barbour, 1987), a good description is provided in Section 4 of (Audenaert, 2009). The Poisson-Charlier expansion is essentially based on the following observation:

If $p_X(z) := E[z^X]$ denotes the probability generating function (pgf) of the random count variable X with marginal probabilities p_i for $i \in \mathbb{N}_0$ and mean μ_X , then the *factorial cumulant generating function (fcgf)* of X is defined as

$$k_X(z) := \ln(p_X(1+z)) = \ln E[(1+z)^X] =: \sum_{r=1}^{\infty} \frac{\kappa_{(r)}}{r!} \cdot z^r, \quad (\text{A.1})$$

where the coefficients $\kappa_{(r)}$ of the series expansion of $k_X(z)$ are referred to as the *factorial cumulants* of X . The factorial cumulants are related to the usual cumulants through the relation (Douglas, 1980; p. 470)

$$\kappa_{(n)} = \sum_{j=1}^n s_{n,j} \cdot \kappa_j, \quad (\text{A.2})$$

where the $s_{n,j}$ denote again the Stirling numbers of the first kind, see Proposition 2.2. In the case of the Poisson distribution $Po(\lambda)$, one simply obtains $k_X(z) = \lambda z$, i. e., $\kappa_{(1)} = \kappa_1 = \lambda$ and $\kappa_{(r)} = 0$ for $r \geq 2$.

If now, in turn, the fcgf of X is known, then its pgf is obtained as

$$p_X(z) = \exp(k_X(z-1)) = \exp\left(\sum_{r=1}^{\infty} \frac{\kappa_{(r)}}{r!} \cdot (z-1)^r\right). \quad (\text{A.3})$$

This leads to the idea to approximate the true pgf of X by the m^{th} order approximation

$$p_X(z) \approx \exp\left(\sum_{r=1}^m \frac{\kappa_{(r)}}{r!} \cdot (z-1)^r\right), \quad (\text{A.4})$$

where in the case of the Poisson distribution, the first order approximation already gives the exact pgf.

The Poisson-Charlier expansion of (Barbour, 1987) is a further refinement of this approach. Let $\pi_i := e^{-\kappa_1} \cdot \kappa_1^i / i!$ for $i \in \mathbb{N}_0$ denote the Poisson probabilities of the $Po(\kappa_1)$ -distribution, where $\kappa_1 = \mu_X = \kappa_{(1)}$, and let ∇ denote the (backward) difference operator, i. e., $\nabla \pi_i = \pi_i - \pi_{i-1}$. Then the m^{th} order Poisson-Charlier approximation of the true probability p_i is given by $f_m(\nabla) \cdot \pi_i$, where f_m is defined as the $(m-1)^{\text{th}}$ order Taylor polynomial in z around $z = 0$ and evaluated in $z = 1$ of the function

$$f(z, \nabla) := \exp\left(\frac{1}{z} \cdot \sum_{r=2}^{\infty} \frac{\kappa_{(r)}}{r!} \cdot (-z\nabla)^r\right), \quad (\text{A.5})$$

see (Audenaert, 2009). The first four approximations are summarized in Table A.1.

It becomes clear that the first order approximation is just the Poisson approximation, i. e., the true probabilities p_i are approximated by the Poisson probabilities π_i

Table A.1: The first four Poisson-Charlier approximations.

$$\begin{aligned}
f_1(\nabla) &= 1, \\
f_2(\nabla) &= 1 + \frac{1}{2}\kappa_{(2)}\nabla^2, \\
f_3(\nabla) &= 1 + \frac{1}{2}\kappa_{(2)}\nabla^2 - \frac{1}{6}\kappa_{(3)}\nabla^3 + \frac{1}{8}\kappa_{(2)}^2\nabla^4, \\
f_4(\nabla) &= 1 + \frac{1}{2}\kappa_{(2)}\nabla^2 - \frac{1}{6}\kappa_{(3)}\nabla^3 + \left(\frac{\kappa_{(2)}^2}{8} + \frac{\kappa_{(4)}}{24}\right)\nabla^4 - \frac{1}{12}\kappa_{(2)}\kappa_{(3)}\nabla^5 + \frac{1}{48}\kappa_{(2)}^3\nabla^6.
\end{aligned}$$

belonging to the $Po(\mu_X)$ -distribution. The higher order approximations account for the deviation of the true distribution of X from the Poisson distribution, where the deviation is measured in terms of the factorial cumulants.

B Proofs

B.1 Proof of Proposition 3.1

One can prove Proposition 3.1 by applying formula (A.2) to Proposition 2.2, but a direct proof appears to be more easy. According to Definition 2.1, the marginal pgf $p_X(z)$ of the INARCH(1) model is characterized by the equality

$$p_X(z) = \exp(\beta(z-1)) \cdot p_X(\exp(\alpha(z-1))),$$

also see (Weiß, 2009). Hence, we obtain for the fcgf $k_X(z)$ the relation

$$k_X(z) = \beta z + \kappa_X(\alpha z).$$

Differentiating this relation, it follows that

$$k'_X(z) = \beta + \alpha \cdot \kappa'_X(\alpha z), \quad \text{and} \quad k_X^{(n)}(z) = \alpha^n \cdot \kappa_X^{(n)}(\alpha z) \quad \text{for } n \geq 2.$$

Using $\kappa_{(n)} = k_X^{(n)}(0)$ and $\kappa_n = \kappa_X^{(n)}(0)$ completes the proof.

B.2 Proof of Formulae (6) and (7)

Maximizing $\ell(\beta, \alpha)$ is equivalent to maximizing

$$\tilde{\ell}(\beta, \alpha) = -(T-1)\beta - \alpha \sum_{t=2}^T x_{t-1} + \sum_{t=2}^T x_t \cdot \ln(\beta + \alpha x_{t-1}), \quad (\text{B.1})$$

see formula (5). Taking partial derivatives, one obtains

$$\frac{\partial}{\partial \beta} \ell(\beta, \alpha) = -(T-1) + \sum_{t=2}^T x_t / (\beta + \alpha x_{t-1}) \stackrel{!}{=} 0 \quad (\text{a}), \quad (\text{B.2})$$

$$\frac{\partial}{\partial \alpha} \ell(\beta, \alpha) = -\sum_{t=2}^T x_{t-1} + \sum_{t=2}^T x_t x_{t-1} / (\beta + \alpha x_{t-1}) \stackrel{!}{=} 0 \quad (\text{b}).$$

Taking $\beta \cdot (\text{a}) + \alpha \cdot (\text{b})$, it follows that

$$\beta(T-1) + \alpha \sum_{t=2}^T x_{t-1} = \sum_{t=2}^T x_t (\beta + \alpha x_{t-1}) / (\beta + \alpha x_{t-1}) = \sum_{t=2}^T x_t. \quad (\text{B.3})$$

Combining formulae (B.2) and (B.3), the proof of formula (6) is complete. Furthermore, formula (7) follows from formula (B.2) by taking again partial derivatives.

B.3 Proof of Formula (11)

The INARCH(1) process according to Definition 2.1 is an ergodic stationary Markov chain, having the moments

$$\begin{aligned} \mu_X &= \kappa_1 = \frac{\beta}{1-\alpha}, \quad \mu_{X,2} := E[X_t^2] = \kappa_2 + \mu_X^2 = \frac{\beta(1+\beta(1+\alpha))}{(1-\alpha)(1-\alpha^2)}, \\ \mu_{X,3} &:= E[X_t^3] = \kappa_3 + 3\kappa_2\mu_X + \mu_X^3 = \frac{\beta}{(1-\alpha)^3} \cdot \left(\frac{1+2\alpha^2}{(1+\alpha)(1+\alpha+\alpha^2)} + \frac{3\beta}{1+\alpha} + \beta^2 \right), \end{aligned} \quad (\text{B.4})$$

see Proposition 2.2 and formula [1,5] on page 452 in (Douglas, 1980). Therefore, the asymptotic distribution of the CLS estimators (10) can be determined by applying the results of Section 3 in (Klimko & Nelson, 1978) (alternatively, Theorem 4 in (Zhu & Wang, 2009) might be applied). Adapting their notations, we define

$$\begin{aligned} g_{t-1}(\beta, \alpha) &:= E[X_t | X_{t-1}, \dots] = \beta + \alpha \cdot X_{t-1}, \\ u_t(\beta, \alpha) &:= X_t - g_{t-1}(\beta, \alpha) = X_t - \beta - \alpha \cdot X_{t-1}. \end{aligned} \quad (\text{B.5})$$

The function $g_{t-1}(\beta, \alpha)$ is easily shown to satisfy the regularity conditions (i) to (iii) defined on page 634 of (Klimko & Nelson, 1978). Especially, we obtain that

$$\begin{aligned} \mathbf{V} &:= \begin{pmatrix} E[(\frac{\partial}{\partial \beta} g_{t-1}) \cdot (\frac{\partial}{\partial \beta} g_{t-1})] & E[(\frac{\partial}{\partial \beta} g_{t-1}) \cdot (\frac{\partial}{\partial \alpha} g_{t-1})] \\ E[(\frac{\partial}{\partial \alpha} g_{t-1}) \cdot (\frac{\partial}{\partial \beta} g_{t-1})] & E[(\frac{\partial}{\partial \alpha} g_{t-1}) \cdot (\frac{\partial}{\partial \alpha} g_{t-1})] \end{pmatrix} \\ &= \begin{pmatrix} E[1] & E[X_{t-1}] \\ E[X_{t-1}] & E[X_{t-1}^2] \end{pmatrix} = \begin{pmatrix} 1 & \frac{\beta}{1-\alpha} \\ \frac{\beta}{1-\alpha} & \frac{\beta(1+\beta(1+\alpha))}{(1-\alpha)(1-\alpha^2)} \end{pmatrix}, \end{aligned} \quad (\text{B.6})$$

see formula (B.4). The inverse of \mathbf{V} equals

$$\mathbf{V}^{-1} = \begin{pmatrix} 1 + \beta(1 + \alpha) & -(1 - \alpha^2) \\ -(1 - \alpha^2) & \frac{(1-\alpha)(1-\alpha^2)}{\beta} \end{pmatrix}. \quad (\text{B.7})$$

Furthermore, also the following expectations exist, see (B.4):

$$\begin{aligned}
 \mathbf{W} &:= \begin{pmatrix} E\left[\left(\frac{\partial}{\partial\beta} g_{t-1}\right) \cdot \left(\frac{\partial}{\partial\beta} g_{t-1}\right) \cdot u_t^2\right] & E\left[\left(\frac{\partial}{\partial\beta} g_{t-1}\right) \cdot \left(\frac{\partial}{\partial\alpha} g_{t-1}\right) \cdot u_t^2\right] \\ E\left[\left(\frac{\partial}{\partial\alpha} g_{t-1}\right) \cdot \left(\frac{\partial}{\partial\beta} g_{t-1}\right) \cdot u_t^2\right] & E\left[\left(\frac{\partial}{\partial\alpha} g_{t-1}\right) \cdot \left(\frac{\partial}{\partial\alpha} g_{t-1}\right) \cdot u_t^2\right] \end{pmatrix} \\
 &= \begin{pmatrix} E[1 \cdot (\beta + \alpha \cdot X_{t-1})] & E[X_{t-1} \cdot (\beta + \alpha \cdot X_{t-1})] \\ E[X_{t-1} \cdot (\beta + \alpha \cdot X_{t-1})] & E[X_{t-1}^2 \cdot (\beta + \alpha \cdot X_{t-1})] \end{pmatrix} \\
 &= \frac{\beta}{1-\alpha} \cdot \begin{pmatrix} 1 & \frac{\alpha+\beta(1+\alpha)}{1-\alpha^2} \\ \frac{\alpha+\beta(1+\alpha)}{1-\alpha^2} & \frac{\beta(1+2\alpha)}{(1-\alpha)(1-\alpha^2)} + \frac{\beta^2}{(1-\alpha)^2} + \frac{\alpha(1+2\alpha^2)}{(1-\alpha^2)(1-\alpha^3)} \end{pmatrix}.
 \end{aligned} \tag{B.8}$$

Formula (B.8) follows, since

$$\begin{aligned}
 E[f(X_{t-1}) \cdot u_t^2(\beta, \alpha)] &= E[f(X_{t-1}) \cdot E[(X_t - \beta - \alpha \cdot X_{t-1})^2 \mid X_{t-1}]] \\
 &= E[f(X_{t-1}) \cdot V[X_t - \beta - \alpha \cdot X_{t-1} \mid X_{t-1}]] \\
 &= E[f(X_{t-1}) \cdot (\beta + \alpha \cdot X_{t-1})]
 \end{aligned}$$

because of the conditional Poisson distribution. So we can apply Theorem 3.2 of (Klimko & Nelson, 1978) to obtain the asymptotic distribution of the CLS estimators (10):

$$\sqrt{T-1}(\hat{\beta}_{\text{CLS}} - \beta, \hat{\alpha}_{\text{CLS}} - \alpha)^\top \xrightarrow{D} N(\mathbf{0}, \mathbf{V}^{-1} \mathbf{W} \mathbf{V}^{-1}), \quad \text{where}$$

$$\mathbf{V}^{-1} \mathbf{W} \mathbf{V}^{-1} = \begin{pmatrix} \frac{\beta}{1-\alpha}(\beta(1+\alpha) + \frac{1+2\alpha^4}{1+\alpha+\alpha^2}) & -\beta(1+\alpha) - \frac{(1+2\alpha)\alpha^3}{1+\alpha+\alpha^2} \\ -\beta(1+\alpha) - \frac{(1+2\alpha)\alpha^3}{1+\alpha+\alpha^2} & (1-\alpha^2)(1 + \frac{\alpha(1+2\alpha^2)}{\beta(1+\alpha+\alpha^2)}) \end{pmatrix}. \tag{B.9}$$