Chapter 2
Detection of Changes in INAR Models

Šárka Hudecová, Marie Hušková, and Simos Meintanis

Abstract In the present paper we develop on-line procedures for detecting changes in the parameters of integer valued autoregressive models of order one. Tests statistics based on probability generating functions are constructed and studied. The asymptotic behavior of the tests under the null hypothesis as well as under certain alternatives is derived.

2.1 Introduction

Studying the stability in time series is one of the important tasks of data analysis. In many cases such tasks are formulated in terms of hypothesis testing (stability of the system versus system instability) or as an estimation problem whereby certain unknown quantities defining the system are estimated in order to detect a possible change in the values of these quantities. This area is known as change point analysis or structural break problem. Corresponding procedures come in two basic variants: off-line (with all data being available at the beginning of the analysis) or on-line procedures whereby observations arrive sequentially (one at a time) and statistical analysis is performed with each incoming observation.

Š. Hudecová · M. Hušková (✉)
Department of Statistics, Charles University of Prague, Sokolovská 83, 186 75 Praha 8, Czech Republic
e-mail: huskova@karlin.mff.cuni.cz

Š. Hudecová
e-mail: hudecova@karlin.mff.cuni.cz

S. Meintanis
Department of Economics, National and Kapodistrian University of Athens, 1 Sophocleous Str., and Aristidou Str., 105 59 Athens, Greece
e-mail: simosmei@econ.uoa.gr

S. Meintanis
Unit for Business Mathematics and Informatics, North–West University, Potchefstroom, South Africa

A. Steland et al. (eds.), *Stochastic Models, Statistics and Their Applications*,
Springer Proceedings in Mathematics & Statistics 122,
DOI 10.1007/978-3-319-13881-7_2

So far the change-point problem has been studied mostly in time series with continuous observations and consequently there is a huge literature on the problem, either in classical ARMA-type time series or more recently in the popular GARCH model; see for instance a recent survey paper [3].

There is a current interest however in studying the same problem with time series of counts. This interest has been developed along with the introduction of several corresponding models for such time series, which in turn is due to the fact that count time series can prove useful in the analysis of data occurring in many applications, such as finance (occurrence of events in a time period), climatology [16], medicine, etc.

There are only a few papers dealing with detection of changes in integer valued time series, a review of recent results can be found in [7] and [13]. In *the off-line* setting, [10] derived results on likelihood ratio type statistics for detection of a change in binary autoregressive time series, while [8] published results on CUSUM type test statistics. Papers [9, 14] and [6] proposed and studied procedures in Poisson autoregressive models.

The on-line procedures for detection of changes were studied in [19, 20] and by [17] and [18] in connection of control charts, while in [14] the authors developed and studied sequential CUSUM type procedures in various integer valued time series.

Here we focus on detection of changes in integer-valued autoregressive (INAR) time series. The INAR model of order one (INAR(1) for short) (see [1, 2, 15]) is specified by Eq. (2.1) below, and it incorporates a Bernoulli probability parameter as well as another parameter indexing the family of the so-called innovations. In what follows we develop detector statistics for detecting changes in these parameters in the context of INAR(1) models. As already mentioned we will work with monitoring schemes and hence propose sequential-type detector statistics. In the remainder of the paper we introduce the INAR process and the test statistics in Sect. 2.2, while in Sect. 2.3 we derive the limit properties of the procedures under the null as well as under a certain class of alternatives. The proofs are postponed in Sect. 2.4.

2.2 Model and Procedures

The INAR(1) process $\{Y_t\}$ is defined by the equation

$$Y_t = p \circ Y_{t-1} + e_t, \tag{2.1}$$

where $p \circ Y_{t-1}$ denotes a sum of Y_{t-1} independent Bernoulli variables all of which are independent of Y_{t-1}, the parameter $p \in (0, 1)$ denotes the probability of the aforementioned Bernoulli variables and $\{e_t\}$ is a sequence (often referred to as 'innovations') of independent and identically distributed (i.i.d.) nonnegative integer valued random variables with finite second moment and probability generating function (PGF) denoted by $g_e(u)$that is assumed to belong to a given family, i.e., $g_e(\cdot) \in \mathcal{G}_\Theta = \{g_e(u; \theta);\ u \in [0, 1],\ \theta \in \Theta\}$ with Θ being an open subset of R.

Under the above conditions the sequence $\{Y_t\}$ is stationary and ergodic. Given the family $\mathcal{G}_\Theta$ of possible PGF for $\{e_t\}$ the model depends on two parameters $(p, \theta) \in$

$(0, 1) \times \Theta$. In this connection we note that while the Poisson family has been by far the most studied case, alternative families for $\{e_t\}$ such as the zero-inflated Poisson of [12], and the Poisson mixture of [16], have also been considered.

The proposed sequential test procedures for detecting changes in INAR(1) processes will be based on properties of probability generating function (PGF) of the observed variables. In this connection recall that the PGF of a random variable Y is defined as

$$g_Y(u) = Eu^Y, \quad u \in [0, 1],$$

and that under very mild conditions this PGF uniquely determines the underlying distribution function of Y. The empirical version of the PGF is defined by

$$\widehat{g}_{Y,n}(u) = \frac{1}{n} \sum_{i=1}^{n} u^{Y_i}, \quad u \in [0, 1],$$

and was employed by [11] in the context of goodness-of-fit testing with certain integer valued time series. This empirical PGF can be further used as the main tool for the construction of detector statistics in count time series of a more general nature. This is in fact a subject of a research project which is already in progress. Here, however, and in order to stay within a relatively simple context, we focus on procedures for detecting changes in the parameters of INAR(1) processes.

We are interested in investigating whether or not the parameters (p, θ) are the same during the observational period. Toward this we introduce a slightly more general model

$$Y_t = p_t \circ Y_{t-1} + e_t, \tag{2.2}$$

where $p_t \circ Y_{t-1}$ denotes a sum of Y_{t-1} independent Bernoulli variables all of which are independent of Y_{t-1} and all have a success probability $p_t \in (0, 1)$, and $\{e_t\}$ is a sequence of independent nonnegative integer valued random variables with finite second moments such that PGF of e_t is $g_e(u; \theta_t)$, $\theta_t \in \Theta$.

We consider a sequential setup where the observations arrive one after the other and, additionally, assume that a historical data set (or training data) $Y_1, \ldots, Y_m$ following the INAR(1) model specified in Eq. (2.1) are given. Then we wish to test the null hypothesis:

$$H_0: \quad (p_t, \theta_t) = (p_0, \theta_0), \quad 1 \le t < \infty,$$

against the alternative

$$\begin{aligned} H_1: \quad & \text{there exist } t_0 \text{ such that } (p_t, \theta_t) = (p_0, \theta_0), \quad 1 \le t \le m + t_0 \\ & \text{but } (p_t, \theta_t) = \left(p^0, \theta^0\right) \quad m + t_0 < t < \infty, \quad (p_0, \theta_0) \ne \left(p^0, \theta^0\right), \end{aligned}$$

where the parameters $p_0, p^0 \in (0, 1)$ and $\theta_0, \theta^0 \in \Theta$ are unknown, and where $m + t_0$ is an unknown change point. Clearly we are interested in testing the null hypothesis that the parameters (p_0, θ_0) do not change, which means that model (2.1) holds true with $(p, \theta) = (p_0, \theta_0)$ while under the alternative the first $m + t_0$ observations follow model (2.1) with parameter (p_0, θ_0) and afterwards it changes to another INAR(1) model with parameter values (p^0, θ^0).

For detection of changes in the above model we apply the method developed in [5] which was first applied in the context of linear regression and later on extended to various other setups. In principle, we estimate the unknown parameters from the historical data, then, having $m + t$ observations, we calculate the test statistic $Q(m, t)$ that is sensitive w.r.t. a change in either of the parameters and according to value of this statistic, we decide whether a change in either of the parameters is indicated or not. In case of no indication of a change we continue with the next observation. We note in this context of change-detection for the parameters of a certain model, CUSUM type procedures are often used. Another possibility is to use some functionals of estimators of unknown parameters based on historical data $Y_1, \dots, Y_m$ and on $Y_{m+1}, \dots, Y_{m+t}$, $t = 1, 2, \dots$.

Here we deal with procedures based on probability generating function utilizing the following property of the PGF of $\{Y_t\}$ under model (2.2)

$$E\big(u^{Y_t}|Y_{t-1}\big) = \big(1 + p_t(u-1)\big)^{Y_{t-1}} g_e(u;\theta_t), \quad t \ge 1,\ u \in [0,1],$$
$$E\big(u^{Y_t}\big) = E\big(1 + p_t(u-1)\big)^{Y_{t-1}} g_e(u;\theta_t), \quad t \ge 1,\ u \in [0,1].$$

Then under model (2.2), the quantities

$$\sum_{s=1}^{t}\Big(u^{Y_s} - \big(1 + p_s(u-1)\big)^{Y_{s-1}} g_e(u;\theta_s)\Big), \quad t \ge 2,$$

are partial sums of martingale differences for fixed $u \in [0, 1]$ which prompts the idea of utilizing these quantities for constructing test procedures.

We suggest to test the null hypothesis H_0 by means of the test statistics based on the first $m + t$ observations

$$S_m(t) = \int_0^1 \left(Q_{m,m+t}(u, \widehat{p}_m, \widehat{\theta}_m) - \frac{t}{m} Q_{0,m}(u, \widehat{p}_m, \widehat{\theta}_m)\right)^2 w(u)du, \quad t \ge 1, \quad (2.3)$$

where $w(u)$ is a nonnegative weight function and

$$Q_{\ell,j}(u, \widehat{p}_m, \widehat{\theta}_m) = \frac{1}{\sqrt{m}} \sum_{s=\ell+1}^{j} \Big(u^{Y_s} - \big(1 + \widehat{p}_m(u-1)\big)^{Y_{s-1}} g_e(u;\widehat{\theta}_m)\Big),$$
$$\ell, j = 0, \dots$$

with $(\widehat{p}_m, \widehat{\theta}_m)$ being estimators of (p, θ) based on the historical data $Y_1, \dots, Y_m$.

The null hypothesis is rejected as soon as for some t

$$S_m(t)/q_\gamma^2(t/m) \ge c,$$

for an appropriately chosen c, where

$$q_\gamma(s) = (1+s)\left(\frac{s}{s+1}\right)^\gamma, \quad s \in (0,\infty),\ \gamma \in [0, 1/2),$$

is a boundary function. (Possible choices of boundary functions $q_\gamma(s)$ are discussed, e.g., in [4].) In this case, we usually stop and confirm a change, otherwise we continue monitoring. The related stopping rule is defined as

$$\tau_m(\gamma, T) = \inf\{1 \le t \le mT\colon\ S_m(t)/q_\gamma^2(t/m) \ge c\},$$
$$\tau_m(\gamma, T) = \infty \quad \text{if } S_m(t)/q_\gamma^2(t/m) < c \text{ for all } 1 \le t \le Tm,$$

for some fixed integer $T > 0$. It is required that under H_0

$$\lim_{m\to\infty} P\big(\tau_m(\gamma, T) < \infty\big) = \alpha$$

for prechosen $\alpha \in (0, 1)$ and under alternatives

$$\lim_{m\to\infty} P\big(\tau_m(\gamma, T) < \infty\big) = 1.$$

The former requirement guarantees asymptotic level α, while the later one ensures consistency. Hence in order to get an approximation for $c = c_\alpha$, the limit behavior ($m \to \infty$) of

$$\max_{1\le t\le mT} S_m(t)/q_\gamma^2(t/m) \tag{2.4}$$

under H_0 has to be studied, while for consistency one has to investigate its limit behavior under alternatives. Both tasks are taken up in the next section.

The question of the optimal choices of the weight function w and the boundary function q_γ in order the detection lag is as small as possible remains open. Some practical recommendations are in the next section.

2.3 Asymptotic Results

Consider the INAR(1) process in Eq. (2.2) and denote the true value of $\vartheta = (p, \theta)$ under the null hypothesis H_0 by $\vartheta_0 = (p_0, \theta_0)$. To study the limit distribution under the null hypothesis H_0 we assume the following:

(A.1) $\{Y_t\}_{t\in\mathcal{N}}$ is a sequence of random variables satisfying (2.1) with $\{e_t\}_{t\in\mathcal{N}}$ being a sequence of i.i.d. discrete nonnegative random variables with finite second moment and PGF $g_e(\cdot;\theta)$, $\theta \in \Theta$, where Θ is an open subset of $\mathcal{R}$.

(A.2) $g_e(u;\theta)$ has the first partial derivative w.r.t. θ for all $u \in [0, 1]$ fulfilling Lipschitz condition:

$$\left|\frac{\partial g_e(u;\theta)}{\partial\theta} - \frac{\partial g_e(u;\theta)}{\partial\theta}\bigg|_{\theta=\theta_0}\right| \le D_1|\theta_0 - \theta|v(u),$$
$$u \in [0, 1],\ \ |\theta - \theta_0| \le D_2,$$

and

$$\left|\frac{\partial g_e(u;\theta)}{\partial\theta}\right| \le D_3 v(u), \qquad u \in [0, 1],\ \ |\theta - \theta_0| \le D_2$$

for some $D_j > 0$, $j = 1, 2, 3$, and some measurable function $v(\cdot)$.

(A.3) $0 < \int_0^1 w(u)du < \infty$, $\int_0^1 w(u)v^2(u)du < \infty$.

(A.4) $\widehat{\vartheta}_m = (\widehat{p}_m, \widehat{\theta}_m)'$ is estimator of $\vartheta_0 = (p_0, \theta_0)'$ satisfying

$$\sqrt{m}(\widehat{\vartheta}_m - \vartheta_0) = O_P(1).$$

Assumption (A.3) is satisfied by rather wide class of weight function w. Simple practical choices are $w(u) = u^a$, $u \in [0, 1]$, $a \geq 0$.

In the following theorem we formulate the main assertion on limit behavior of our test statistic defined in Eq. (2.4) under the null hypothesis H_0.

Theorem 2.1 *Let assumptions* (*A*.1)–(*A*.4) *be satisfied in model* (2.1). *Then under the null hypothesis* H_0 *the limit distribution* $(m \to \infty)$ *of* $\max_{1 \leq t \leq mT} S_m(t)/q_\gamma^2(t/m)$ *with* $T > 0$ *fixed is the same as that of*

$$\sup_{s \in (0, T/(T+1))} \int_0^1 \frac{1}{s^{2\gamma}} Z^2(s, u; p_0, \theta_0) w(u) du,$$

where $\{Z(s, u; p, \theta);\, s \in (0, T/(T+1)),\, u \in [0, 1]\}$ *is a Gaussian process with zero mean and covariance structure described by*

$$\begin{aligned} &cov\big(Z(s_1, u_1; p, \theta), Z(s_2, u_2; p, \theta)\big) \\ &\quad = \min(s_1, s_2) E\big(u_1^{Y_2} - E\big(u_1^{Y_2}|Y_1\big)\big)\big(u_2^{Y_2} - E\big(u_2^{Y_2}|Y_1\big)\big) \end{aligned}$$

where

$$E\big(u^{Y_2}|Y_1\big) = \big(1 + p(u-1)\big)^{Y_1} g_e(u; \theta).$$

The explicit form of the limit distribution is not known. In order to approximate this distribution one can replace the unknown parameters and covariance structure by the respective estimators based on historical data and simulate the resulting process. Another possibility is to use parametric bootstrap by estimating (p, θ) from the historical data and then generate bootstrap observations along Eq. (2.1) with (p, θ) replaced by their estimators. This possibility also leads to an asymptotically correct approximation of the limit null distribution of the test statistic.

Next we shortly discuss the limit behavior of our test statistic under the following class of alternatives:

$\widetilde{H}_1$: there exists $0 < \nu_0 < T$ such that for $t_0 = \lfloor m\nu_0 \rfloor$ variables $\{Y_t\}_{t \leq m+t_0}$ follow (2.2) with (p_0, θ_0) and $\{Y_{m+t_0+t}\}_{t \geq 1} =^d \{Y_t^0\}_{t \geq 1}$, where $\{Y_t^0\}_{t \geq 1}$ follow (2.2) with $(p^0, \theta^0) \neq (p_0, \theta_0)$.

Notice that $\widetilde{H}_1$ slightly differs from the alternative H_1. In particular, $\widetilde{H}_1$ assumes that the process $\{Y_t\}$ changes from one INAR(1) process to another one, both possibly strictly stationary. This simplifies the formulation of the succeeding theorem and the corresponding proof.

Theorem 2.2 *Let* $\{Y_t\}_{t \leq m+t_0}$ *and* $\{Y_t^0\}_{t \geq 1}$ *from* $\widetilde{H}_1$ *satisfy assumptions* (*A*.1) *with parameters* (p_0, θ_0) *and* (p^0, θ^0), *respectively, and let also* (A.2)–(A.4) *be satisfied.*

Then under the alternative hypothesis $\widetilde{H}_1$ for any $\nu_0 < s < T$

$$\frac{1}{m} S_m\big(\lfloor ms \rfloor\big) \to (s - \nu_0) \int_0^1 \Big(E\big[\big(1 + p^0(u-1)\big)^{Y_1^0} g_e\big(u, \theta^0\big) - \big(1 + p_0(u-1)\big)^{Y_1} g_e(u, \theta_0)\big]\Big)^2 w(u) du$$

in probability as $m \to \infty$.

Studying carefully the proofs one realizes that the proposed test procedures are sensitive not only w.r.t. changes in the parameters p and/or θ but also w.r.t. changes that leave these parameters invariant but involve a change in the distribution (and hence the PGF) of the innovations e_t.

2.4 Proofs

Due to the space restriction and due to a certain similarity to the proof of Theorem 4.1 in [11] we present only main steps of the proof of our Theorem 1.

By the Taylor expansion of $Q_{m,m+t}(u, \widehat{p}_m, \widehat{\theta}_m) - \frac{t}{m} Q_{0,m}(u, \widehat{p}_m, \widehat{\theta}_m)$ at p_0, θ_0 and by convergence properties of stationary sequences we realize that under H_0 the limit behavior of $\max_{1 \le t \le mT} S_m(t)/q_\gamma^2(t/m)$ does not change if the estimators $\widehat{p}_m, \widehat{\theta}_m$ are replaced by their true values p_0, θ_0.

Since $Q_{m,m+t}(u, p_0, \theta_0)$ and $Q_{0,m}(u, p_0, \theta_0)$ are partial sums of bounded martingale differences we can apply theorems on their limit behavior. The proof can be finished combining the arguments in the last part of the proof of Theorem 4.1 in [11] and the proof of Theorem 1 in [4].

Acknowledgements The research of Simos Meintanis was partially supported by grant number 11699 of the Special Account for Research Grants of the National and Kapodistrian University of Athens (ELKE). The research of Marie Hušková was partially supported by grant GAČR P201/12/1277 and by AP research network grant Nr. P7/06 of the Belgian government (Belgian Science Policy). The research of Šárka Hudecová was partially supported by the Czech Science Foundation project DYME Dynamic Models in Economics No. P402/12/G097.

References

1. Al-Osh MA, Alzaid AA (1987) First-order integer-valued autoregressive (INAR(1)) process. J Time Ser Anal 8:261–275
2. Alzaid AA, Al-Osh MA (1988) First-order integer-valued autoregressive (INAR(1)) process: distributional and regression properties. Stat Neerl 42:53–61
3. Aue A, Horváth L (2013) Structural breaks in time series. J Time Ser Anal 34:1–16
4. Aue A, Horváth L, Hušková M, Kokoszka P (2006) Change-point monitoring in linear models with conditionally heteroskedastic errors. Econom J 9:373–403
5. Chu C-SJ, Stinchcombe M, White H (1996) Monitoring structural change. Econometrica 64:1045–1065

6. Doukhan P, Kengne W (2013) Inference and testing for structural change in time series of counts model. Available via ArXiv http://arxiv.org/pdf/1305.1751v1.pdf. Cited 13 Sep 2014
7. Fokianos K (2012) Count time series models. In: Subba Rao T, Subba Rao S, Rao CR (eds) Handbook of statistics. Time series — methods and applications, vol 30. Elsevier, Amsterdam, pp 315–348
8. Fokianos K, Gombay E, Hussein A (2014) Retrospective change detection for binary time series models. J Stat Plan Inference 145:102–112
9. Franke J, Kirch C, Tadjuidje Kamgaing J (2012) Changepoints in time series of counts. J Time Ser Anal 33(5):757–770
10. Hudecová Š (2013) Structural changes in autoregressive models for binary time series. J Stat Plan Inference 143(10):1744–1752
11. Hudecová Š, Hušková M, Meintanis S (2014) Tests for time series of counts based on the probability generating function. Statistics, doi:10.1080/02331888.2014.979826 (in press)
12. Jadi MA, Jones G, Lai CD (2012) First-order integer valued AR processes with zero inflated Poisson innovations. J Time Ser Anal 33(6):954–963
13. Kirch C, Tadjuidje Kamgaing J (2015) Detection of change points in discrete valued time series. In: Davis RA, Holan SA, Lund RB, Ravishanker N (eds) Handbook of discrete valued time series. Chapman & Hall, London
14. Kirch C, Tadjuidje Kamgaing J (2014) Monitoring time series based on estimating functions. Preprint. Available at http://www.math.kit.edu/stoc/~ckirch/seite/publications. Cited 13 Sep 2014
15. McKenzie E (1985) Some simple models for discrete variate time series. Water Resour Bull 21:645–650
16. Pavlopoulos H, Karlis D (2008) INAR(1) modeling of overdispersed count series with an environmental application. Environmetrics 19(4):369–393
17. Weiss CH (2009) Modelling time series of counts with overdispersion. Stat Methods Appl 18:507–519
18. Weiss CH (2011) Detecting mean increases in Poisson INAR(1) processes with EWMA control charts. J Appl Stat 38(2):338–398
19. Weiss CH, Testik MC (2011) The Poisson INAR(1) CUSUM chart under overdispersion and estimation error. IIE Trans 43:805–818
20. Yontay P, Weiss CH, Testik MC (2013) A two-sided cumulative sum chart for first-order integer-valued autoregressive processes of Poisson counts. Qual Reliab Eng Int 29(1):33–42

Springer
http://www.springer.com/978-3-319-13880-0

Stochastic Models, Statistics and Their Applications
Wrocław, Poland, February 2015
Steland, A.; Rafajłowicz, E.; Szajowski, K. (Eds.)
2015, XX, 492 p. 65 illus., 37 illus. in color., Hardcover
ISBN: 978-3-319-13880-0