

Test Statistics for Parameter Changes in INAR(p) Models and a Simulation Study

Tamás T. Szabó
University of Szeged, Hungary

Abstract: In the paper a change detection procedure is introduced for integer-valued autoregressive processes which are inspired by the ordinary autoregressive model. The power of the test is investigated in a simulation study to determine how effectively it can detect changes in the key parameters of the process.

Zusammenfassung: In diesem Artikel schlagen wir eine Methode vor, die für das Erkennen von Veränderungen in einem ganzzahligen autoregressiven Prozess gebraucht werden kann. Dann wird eine Simulationsstudie ausgeführt, um zu bestimmen, wie oft der Test eine Veränderung in den verschiedenen Parametern ermitteln kann.

Keywords: Change Detection, Integer-valued Autoregression, Computer Simulation.

1 Introduction

Detecting a change of a parameter in a stochastic model is an important task, and one that has been investigated since decades. In econometrics, medicine or population dynamics, failing to notice a change in a parameter will lead to unreliable estimates of the parameters, and, ultimately, false predictions. Several statistics have been constructed for a wide range of stochastic models to test for parameter change. For a review of these techniques we refer to Csörgő and Horváth (1997).

In this paper we will consider the integer-valued autoregressive process of order p , abbreviated INAR(p) and test for a change in several crucial parameters. The INAR(p) model, introduced by Alzaid and Al-Osh (1987) and Alzaid and Al-Osh (1990) is an integer analogue of the AR(p) model that is extensively used in time series analysis. Its main advantage is that it is nonnegative and integer-valued, therefore more applicable in situations where we have data of this type – most frequently count data.

The approach we will use to test for a parameter change is similar to the CUSUM approach described in Kang and Lee (2009) and Lee and Na (2005). After introducing the test statistics we will describe an empirical simulation study which investigates the power of the test as the function of several parameters of the process.

In this first section we will introduce the INAR(p) model and describe the statistical problem that has motivated this paper. The second section contains the details of the calculation of the test statistics. The fourth section is dedicated to the simulation study.

1.1 The INAR(p) Model

Definition. An INAR(p) time series model is a stochastic process $(X_k)_{k \geq -p+1}$, $k \in \mathbb{Z}$, given by

$$X_k = \sum_{j=1}^{X_{k-1}} \xi_{1,k}(j) + \sum_{j=1}^{X_{k-2}} \xi_{2,k}(j) + \cdots + \sum_{j=1}^{X_{k-p}} \xi_{p,k}(j) + \varepsilon_k, \quad k \in \mathbb{N}, \quad (1)$$

where for all $i \in \mathbb{N}$, $\{\xi_{i,k}(j)\}_{j \in \mathbb{N}}$ is a sequence of i.i.d. Bernoulli random variables with mean $\alpha_{i,k}$ and $(\varepsilon_k)_{k \in \mathbb{N}}$ is a sequence of independent nonnegative integer-valued random variables with mean μ_k and variance σ_k^2 , such that all these sequences are independent of each other. For technical reasons, we will suppose that $X_0, X_{-1}, \dots, X_{-p+1}$ are deterministic throughout the paper. The α 's are called the coefficients, the distribution of the ε_k 's is called the innovation distribution. We will also call the distributions of the $\xi_{i,k}(j)$'s the offspring distribution.

Remark. Using the binomial thinning operator due to Steutel and Harn (1979) we can also write the process in the following form:

$$X_k = \alpha_{1,k} \circ X_{k-1} + \alpha_{2,k} \circ X_{k-2} + \cdots + \alpha_{p,k} \circ X_{k-p} + \varepsilon_k.$$

The meaning of \circ is the following: if X is a nonnegative integer-valued random variable and $0 \leq \alpha \leq 1$, then

$$\alpha \circ X := \begin{cases} 0, & \text{if } X = 0, \\ \sum_{j=1}^X \xi_j, & \text{otherwise,} \end{cases}$$

where $(\xi_j)_{j \in \mathbb{N}}$ is a sequence of i.i.d. Bernoulli random variables with mean α .

Remark. The \circ operator emphasizes the similarity to the AR(p) model – indeed, many of our results were inspired by analogous results for AR(p) processes. There are, however, two very important differences. The first is the innovation. In the AR(p) model this is assumed to have zero mean. Here it is nonnegative. The second is that the deterministic regression (not counting the innovation) in AR(p) is replaced here by a stochastic regression, which makes it more fruitful to consider the INAR(p) model as a special multitype branching process with immigration. We will also use this approach later on.

Remark. The motivation of the process is easy to understand: let us take a population in which every member has 0 or 1 offspring at p consequent time steps and then deceases. The number of offspring is independent in the different time steps and the members act independently of each other. If we augment this process with a stochastic immigration we obtain the INAR(p) model.

1.2 The Goal of the Analysis

The parameters of the process are contained in the parameter vector θ_k defined by

$$\theta_k := [\alpha_{1,k}, \dots, \alpha_{p,k}, \mu_{\varepsilon_k}]^\top.$$

The variance of the innovation, denoted by σ_k^2 , is not included in θ_k because we will not test for change in that quantity. The reason that a notation was introduced for it is that its estimation is necessary to calculate the test statistics.

Let us split $\theta_k = (\phi_k, \eta_k)$ into parameter vectors of interest ϕ_k and nuisance vectors η_k . We want to test the following hypothesis:

$$H_0 : \theta_1 = \theta_2 = \dots = \theta_n \text{ and } \varepsilon_1, \varepsilon_2, \dots, \varepsilon_n \text{ are i.i.d.}$$

against

$$H_A : \exists k^* \in \{1, \dots, n-1\} : \phi_1 = \phi_2 = \dots = \phi_{k^*} \neq \phi_{k^*+1} = \dots = \phi_n \text{ but} \\ \eta_1 = \eta_2 = \dots = \eta_n, \varepsilon_1, \dots, \varepsilon_{k^*} \text{ are i.i.d and } \varepsilon_{k^*+1}, \dots, \varepsilon_n \text{ are i.i.d.}$$

2 Calculation of the Test Statistics

In this section we will work under the null hypothesis and derive test statistics whose limiting distribution is a well-known process. Under the null hypothesis, the process takes the following form:

$$X_k = \sum_{i=1}^p \sum_{j=1}^{X_{k-i}} \xi_i(k, j) + \varepsilon(k), \quad k \in \mathbb{N}, \tag{2}$$

where for all $i \in \mathbb{N}$, $\{\xi_i(k, j)\}_{j \in \mathbb{N}}$ is a sequence of i.i.d. Bernoulli random variables with mean α_i and $(\varepsilon(\ell))_{\ell \in \mathbb{N}}$ is a sequence of i.i.d. nonnegative integer-valued random variables with mean μ_ε and variance σ_ε^2 , such that all these sequences are independent of each other.

Remark. Note that the index k has moved from subscript to within parentheses for each variable. This emphasizes the fact that their distribution is no longer dependent of k .

We can distinguish three types of INAR(p) processes:

- stable: $\alpha_1 + \dots + \alpha_p < 1$
- unstable: $\alpha_1 + \dots + \alpha_p = 1$
- explosive: $\alpha_1 + \dots + \alpha_p > 1$.

In the paper we will only consider the stable case, i.e., when under the null hypothesis $\alpha_1 + \dots + \alpha_p < 1$.

First we calculate the estimates of the parameters using the conditional least squares (CLS) method. This method, first proposed by Klimko and Nelson (1978) relies on the sequence

$$M_n = X_n - E(X_n | \mathcal{F}_{n-1}) = X_n - \alpha_1 X_{n-1} - \dots - \alpha_p X_{n-p} - \mu_\varepsilon,$$

where \mathcal{F}_n is the σ -algebra generated by X_1, \dots, X_n . Given an observed trajectory X_1, \dots, X_n , the sequence $(M_k)_{k=1}^n$ depends only on the parameters. The conditional least

squares estimates of the parameters are the values for which the sum of squares $\sum_{k=1}^n M_k^2$ is minimal. After simple calculations the estimates are

$$\widehat{\theta}^{(n)} := \begin{bmatrix} \widehat{\alpha}_1^{(n)} \\ \vdots \\ \widehat{\alpha}_p^{(n)} \\ \widehat{\mu}_\varepsilon^{(n)} \end{bmatrix} = \left[\begin{array}{c} \sum_{k=1}^n \mathbf{X}_{k-1} \mathbf{X}_{k-1}^\top \\ \sum_{k=1}^n \mathbf{X}_{k-1}^\top \\ \sum_{k=1}^n \mathbf{X}_{k-1} \\ n \end{array} \right]^{-1} \sum_{k=1}^n X_k \begin{bmatrix} \mathbf{X}_{k-1} \\ 1 \end{bmatrix} \quad (3)$$

with

$$\mathbf{X}_k = \begin{bmatrix} X_k \\ X_{k-1} \\ \vdots \\ X_{k-p+1} \end{bmatrix}.$$

Replacing the parameters in the M_k with their estimates we obtain the sequence $\widehat{M}_k^{(n)}$. The test statistics will be based on the stochastic process

$$\widehat{\mathcal{M}}_n(t) := \widehat{I}_n(\widehat{\theta}^{(n)})^{-1/2} \sum_{k=1}^{\lfloor nt \rfloor} \begin{bmatrix} \widehat{M}_k^{(n)} X_{k-1} \\ \vdots \\ \widehat{M}_k^{(n)} X_{k-p} \\ \widehat{M}_k^{(n)} \end{bmatrix}, \quad n \in \mathbb{N}, t \in [0, 1],$$

where $\lfloor x \rfloor$ denotes the integer part of a real number $x \in \mathbb{R}$, $A^{-1/2}$ denotes the unique symmetric positive definite square root of the inverse of a symmetric positive definite matrix A , and $\widehat{I}_n(\widehat{\theta}^{(n)})$ is a strongly consistent estimator of an analogue of the Fisher information matrix.

Using techniques similar to those of Lee and Na (2005) we can prove the following theorem, essentially similar to Theorem 2.1 of that paper:

Theorem 2.1. *Suppose that $(X_k)_{k \geq -p+1}$ is an INAR(p) process, H_0 holds, $E(\varepsilon_1^4) < \infty$, $\alpha_1 + \dots + \alpha_p < 1$, and $\text{Var}(\varepsilon_1) > 0$. Then there is a standard $p+1$ -dimensional Brownian bridge $(B(t))_{0 \leq t \leq 1}$ such that*

$$\widehat{\mathcal{M}}_n \xrightarrow{\mathcal{D}} \mathcal{B} \quad \text{as } n \rightarrow \infty.$$

Using Theorem 2.1 we can prove the following result:

Theorem 2.2. *Let us denote the i -th component of $(\widehat{\mathcal{M}}_n(t))_{0 \leq t \leq 1}$ by $(\widehat{\mathcal{M}}_n^{(i)}(t))_{0 \leq t \leq 1}$ ($i = 1, \dots, p+1$), and let $(B(t))_{0 \leq t \leq 1}$ a standard one-dimensional Brownian bridge. Then under the conditions of Theorem 2.1*

$$\begin{aligned} \sup_{0 \leq t \leq 1} \widehat{\mathcal{M}}_n^{(i)}(t) &\xrightarrow{\mathcal{D}} \sup_{0 \leq t \leq 1} B(t) \\ \sup_{0 \leq t \leq 1} |\widehat{\mathcal{M}}_n^{(i)}(t)| &\xrightarrow{\mathcal{D}} \sup_{0 \leq t \leq 1} |B(t)| \\ \sup_{0 \leq t \leq 1} \widehat{\mathcal{M}}_n^{(i)}(t) - \inf_{0 \leq t \leq 1} \widehat{\mathcal{M}}_n^{(i)}(t) &\xrightarrow{\mathcal{D}} \sup_{0 \leq t \leq 1} B(t) - \inf_{0 \leq t \leq 1} B(t) \end{aligned}$$

as $n \rightarrow \infty$.

Using Theorem 2.2 we can construct three tests for a change in any parameter, similarly to Berkes, Gombay, and Horváth (2009). The components of the Brownian bridge $(\mathcal{B}(t))_{0 \leq t \leq 1}$ in Theorem 2.1 are independent, therefore we can test for change in several parameters simultaneously. If we perform d tests together, all of which have a significance level of $1 - \alpha^*$, then under H_0 each test accepts H_0 with probability $1 - \alpha^*$. We accept H_0 if and only if all tests accept it, and the tests are independent, therefore the probability of the simultaneous test accepting H_0 is $(1 - \alpha^*)^d$. Hence, for the simultaneous test to have a significance level of α , we have to put

$$\alpha^* = 1 - (1 - \alpha)^{1/d}.$$

Test 1. (one-sided): if

$$\sup_{0 \leq t \leq 1} \widehat{\mathcal{M}}_n^{(i)}(t) > Q_1(1 - \alpha^*),$$

then we reject H_0 and conclude that there was a downward change in the i -th parameter. Similarly, if

$$\inf_{0 \leq t \leq 1} \widehat{\mathcal{M}}_n^{(i)}(t) < -Q_1(1 - \alpha^*),$$

then we reject H_0 and conclude that there was an upward change in the i -th parameter. Here $Q_1(1 - \alpha^*)$ is the $(1 - \alpha^*)$ -quantile of $\sup_{0 \leq t \leq 1} B(t)$.

Test 2. (two-sided): if

$$\sup_{0 \leq t \leq 1} |\widehat{\mathcal{M}}_n^{(i)}(t)| > Q_2(1 - \alpha^*),$$

then we reject H_0 and conclude that there was a change in the i -th parameter. Here $Q_2(1 - \alpha^*)$ is the $(1 - \alpha^*)$ -quantile of $\sup_{0 \leq t \leq 1} |B(t)|$.

Test 3. (temporary change): if

$$\sup_{0 \leq t \leq 1} \widehat{\mathcal{M}}_n^{(i)}(t) - \inf_{0 \leq t \leq 1} \widehat{\mathcal{M}}_n^{(i)}(t) > Q_3(1 - \alpha^*),$$

then we reject H_0 and conclude that there was a temporary change (also called epidemic alternative) in the i -th parameter, as defined in Csörgő and Horváth (1997), 1.7.4. Here $Q_3(1 - \alpha^*)$ is the $(1 - \alpha^*)$ -quantile of $\sup_{0 \leq t \leq 1} B(t) - \inf_{0 \leq t \leq 1} B(t)$.

The cumulative distribution functions to compute the appropriate quantiles are the following:

$$\begin{aligned} \mathbb{P} \left(\sup_{0 \leq t \leq 1} B(t) > x \right) &= e^{-2x^2}, \quad x \geq 0 \\ \mathbb{P} \left(\sup_{0 \leq t \leq 1} |B(t)| > x \right) &= 2 \sum_{k=1}^{\infty} (-1)^{k+1} e^{-2k^2x^2}, \quad x \geq 0 \\ \mathbb{P} \left(\sup_{0 \leq t \leq 1} B(t) - \inf_{0 \leq t \leq 1} B(t) > x \right) &= 2 \sum_{k=1}^{\infty} (4k^2x^2 - 1) e^{-2k^2x^2}, \quad x \geq 0. \end{aligned}$$

The first two results are well-known, the third is due to Kuiper (1960).

3 Calculation of the Information Matrix

The heuristic reasoning behind the definition of $(\widehat{M}_n(t))_{0 \leq t \leq 1}$ is the following: due to the martingale central limit theorem,

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{\lceil nt \rceil} M_k \xrightarrow{\mathcal{D}} \sqrt{c} \mathcal{W}_t, \quad t \in [0, 1],$$

where c is a constant dependent on θ and σ_ε^2 , and $(\mathcal{W}_t)_{0 \leq t \leq 1}$ is a standard Brownian motion. Therefore, by a rough approximation

$$\frac{1}{\sqrt{n}} (M_1, \dots, M_n) \sim N(0, cI_n),$$

where I_n is the $n \times n$ identity matrix. The likelihood function is

$$-\frac{1}{(2\pi c)^{n/2}} \exp \left\{ -\frac{1}{2c^2 n} \sum_{k=1}^n M_k^2 \right\}.$$

We will take the derivative of the loglikelihood function and work with that quantity. The first term will be regarded as constant. This is a simplification because c actually depends on the parameters but taking this into account leads to calculations that are difficult to handle. Also, we will not take into account the constant factor before the sum of the M_k but will rather work with the analogue of the information matrix. Therefore, the analogue of the loglikelihood function that we will consider will be

$$Q_n(\alpha_1, \dots, \alpha_p, \mu_\varepsilon) = -2^{-1} \sum_{k=1}^n M_k^2.$$

Using this function we are developing similar results to those of the likelihood ratio tests described in Csörgő and Horváth (1997). The proof that this heuristic reasoning leads to valid results is Theorem 2.1. Based on Q_n , the analogue of the efficient score vector is

$$\nabla Q_k(\widehat{\theta}^{(n)}) = \begin{bmatrix} \sum_{j=1}^k \widehat{M}_j^{(n)} X_{j-1} \\ \vdots \\ \sum_{j=1}^k \widehat{M}_j^{(n)} X_{j-p} \\ \sum_{j=1}^k \widehat{M}_j^{(n)} \end{bmatrix}.$$

The gradient of $Q_k(\theta)$ is

$$\nabla Q_k(\theta) = \begin{bmatrix} \sum_{j=1}^k M_j X_{j-1} \\ \vdots \\ \sum_{j=1}^k M_j X_{j-p} \\ \sum_{j=1}^k M_j \end{bmatrix}.$$

The appropriate analogue of the Fisher information matrix is

$$I_n(\theta) := \sum_{k=1}^n \mathbb{E} \left[\{ \nabla Q_k(\theta) - \nabla Q_{k-1}(\theta) \} \{ \nabla Q_k(\theta) - \nabla Q_{k-1}(\theta) \}^\top \mid \mathcal{F}_{k-1} \right]. \quad (4)$$

We can write $I_n(\theta)$ in the form

$$I_n(\theta) = \sum_{k=1}^n \mathbb{E} \left(\left[\begin{array}{cc} M_k^2 \mathbf{X}_{k-1} \mathbf{X}_{k-1}^\top & M_k^2 \mathbf{X}_{k-1} \\ M_k^2 \mathbf{X}_{k-1}^\top & M_k^2 \end{array} \right] \mid \mathcal{F}_{k-1} \right).$$

The vector \mathbf{X}_{k-1} is measurable with respect to \mathcal{F}_{k-1} , therefore

$$I_n(\theta) = \sum_{k=1}^n \mathbb{E}(M_k^2 \mid \mathcal{F}_{k-1}) \begin{bmatrix} \mathbf{X}_{k-1} \mathbf{X}_{k-1}^\top & \mathbf{X}_{k-1} \\ \mathbf{X}_{k-1}^\top & 1 \end{bmatrix}.$$

By simple calculation

$$I_n(\theta) = \sum_{k=1}^n (\alpha_1(1 - \alpha_1)X_{k-1} + \dots + \alpha_p(1 - \alpha_p)X_{k-p} + \sigma_\varepsilon^2) \begin{bmatrix} \mathbf{X}_{k-1} \mathbf{X}_{k-1}^\top & \mathbf{X}_{k-1} \\ \mathbf{X}_{k-1}^\top & 1 \end{bmatrix}.$$

Since σ_ε^2 appears in $I_n(\theta)$ we have to provide an estimator for it. To do this, we introduce the sequence of martingale differences

$$N_k = M_k^2 - \mathbb{E}(M_k^2 \mid \mathcal{F}_{k-1}) = M_k^2 - \alpha_1(1 - \alpha_1)X_{k-1} - \dots - \alpha_p(1 - \alpha_p)X_{k-p} - \sigma_\varepsilon^2, \quad k \geq 0.$$

Replacing the α coefficients and μ_ε both in the formula and in M_k^2 by their estimates we have

$$\tilde{N}_k := \left(\widehat{M}_k^{(n)} \right)^2 - \widehat{\alpha}_1^{(n)} \left(1 - \widehat{\alpha}_1^{(n)} \right) X_{k-1} - \widehat{\alpha}_p^{(n)} \left(1 - \widehat{\alpha}_p^{(n)} \right) X_{k-1} - \sigma_\varepsilon^2, \quad k \geq 0.$$

For the conditional least squares estimation we need to minimize the sum of squares

$$\sum_{k=1}^n \tilde{N}_k^2.$$

Taking the derivative with respect to σ_ε^2 we obtain the CLS estimate

$$\widehat{(\sigma_\varepsilon^2)}^{(n)} = \frac{1}{n} \sum_{k=1}^n \left(\left(\widehat{M}_k^{(n)} \right)^2 - \widehat{\alpha}_1^{(n)} \left(1 - \widehat{\alpha}_1^{(n)} \right) X_{k-1} - \widehat{\alpha}_p^{(n)} \left(1 - \widehat{\alpha}_p^{(n)} \right) X_{k-1} \right).$$

The matrix $\widehat{I}_n(\theta)$ is defined by replacing in $I_n(\theta)$ the variance σ_ε^2 by its CLS estimate.

Remark. Using the ergodic theorem for Markov chains it can be proved that

$$\frac{\widehat{I}_n(\widehat{\theta}^{(n)})}{n} \rightarrow I(\theta) \quad \text{a.s.}$$

for some matrix $I(\theta)$. The entries of $I(\theta)$ are increasing functions of the variances of the offspring and innovation distributions. This will explain some of the phenomena arising later.

4 Simulation Study

The second part of the paper deals with a simulation study that was carried out using Octave 2.3.4. Both the recursive generation of the trajectories and the calculation of the test statistics are given explicitly in the previous sections, therefore the implementation was straightforward and no further details are needed here.

4.1 Technical Details

The simulations were run on dual-core Intel processors under Windows XP. The generation of a trajectory with length 400 and the analysis of that case lasted approximately 0.4 seconds—therefore the typical experiment of 1000 simulations took approximately 400 seconds. The running time increased approximately linearly with trajectory length. Higher order cases also required significantly more time due to the higher dimension matrix calculations.

4.2 Rate of Convergence and the Optimal Trajectory Length

Theorem 2.2 provides the asymptotic distribution of the test statistics but says nothing about the rate of convergence. First we investigated this issue by performing Test 2 both individually and simultaneously for INAR(1) processes with no change. For a trajectory length of 400, an asymptotic significance level of 95 % resulted in typical Type I error rates of 3.4–3.8 %. This effect disappeared slowly, the Type I error being between 4.7 % and 5.1 % for a trajectory length of 4000. A natural conclusion would be to perform the simulation study for a trajectory length of 4000. Apart from a tenfold increase in running time, there were two major problems with this. Firstly, the average researcher is quite unlikely to have a time series with 4000 data points, and therefore a study like this would say little to prospective users of the method. Secondly, for a trajectory length of 4000, the empirical power is larger than 0.9 for almost any change and therefore the phenomena described below cannot be observed. Based on these reasons, we decided to investigate trajectories with 400 data points, with the caveat that in these cases for significance level $1 - \alpha$ the Type I error rate may differ from α by a few percents.

4.3 The First-order Case – General Observations

The first investigated case was an INAR(1) process where the innovation was Poisson distributed. We performed Test 2 for both parameters simultaneously with overall significance 0.95. The length of a trajectory was $n = 400$, change occurred at $k = 200$. Before the change, the coefficient α_1 was 0.5 and the innovation was Poisson distributed with mean $\mu_\varepsilon = 1$. After the change, both parameters took different values: α_1 was selected from the set $\{0, 0.1, 0.2, \dots, 0.9\}$, and μ_ε was selected from $\{0, 0.2, 0.4, \dots, 2\}$. For all such possible pairs of after-change parameters, a Monte-Carlo experiment with 1000 repetitions was run to determine the empirical power of the simultaneous two-sided test for both parameters. The result can be seen in Figure 1. Tables 1 and 2 are also provided for the cases when there was only a change in one of the two parameters.

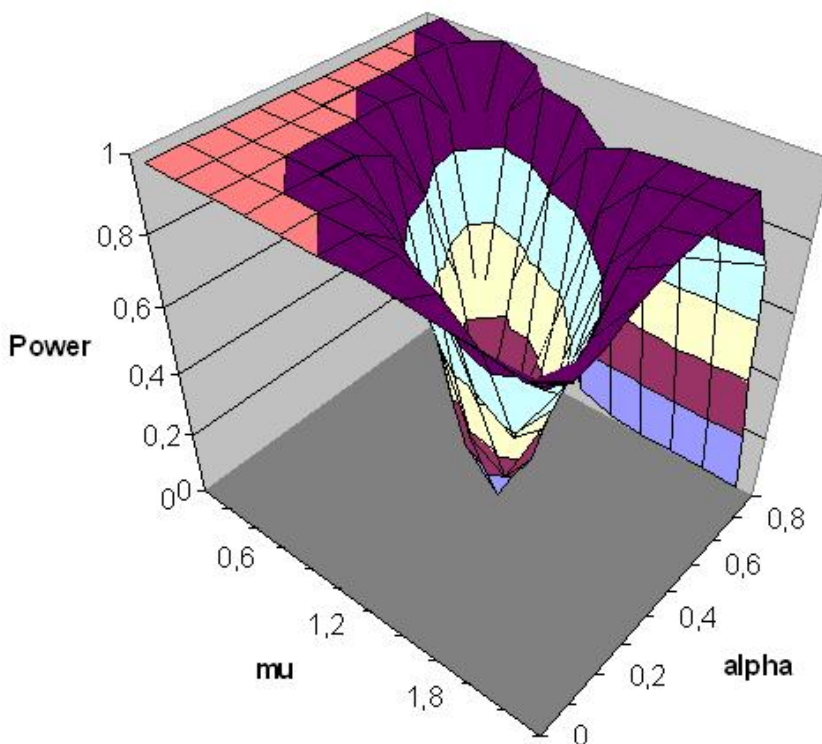


Figure 1: Empirical power of the test for INAR(1)

It is important to note that the power is significantly lower along the line

$$\frac{\mu_\varepsilon}{1 - \alpha_1} = 2.$$

The explanation is that this formula provides the mean of the unique stationary distribution, which initially is equal to 2. Apparently, the test is very sensitive to a change in this quantity. Figure 2 illustrates this effect with two typical trajectories. In the first one, the mean of the stationary distribution remains the same, in the second one, it decreases. The change can be detected by eye. Another shortcoming of the test is that the power drops steeply along $\alpha_1 = 0.9$ when there is an increase in μ_ε . This is due to the fact that in this case α_1 is slightly overestimated (if α_1 changes to 0.9 and μ_ε to 1.2 then the mean of the CLS estimates for α_1 was 0.975) and for α_1 values close to 1 the behavior of the process changes drastically. This is the nearly unstable case, and in the unstable case the process loses its stationarity, $E(X_k) \rightarrow \infty$ linearly and Theorem 2.2 is no longer valid—see e.g. Barczy, Ispány, and Pap (2011) This case is illustrated by Figure 3, where we can see that a small change in the coefficient can be accompanied by a large change in the innovation without any visible change occurring. In Figure 3 the coefficient and the innovation mean are constant throughout the trajectory.

Tables 1 and 2 are provided for the cases where there was a change in only on parameter.

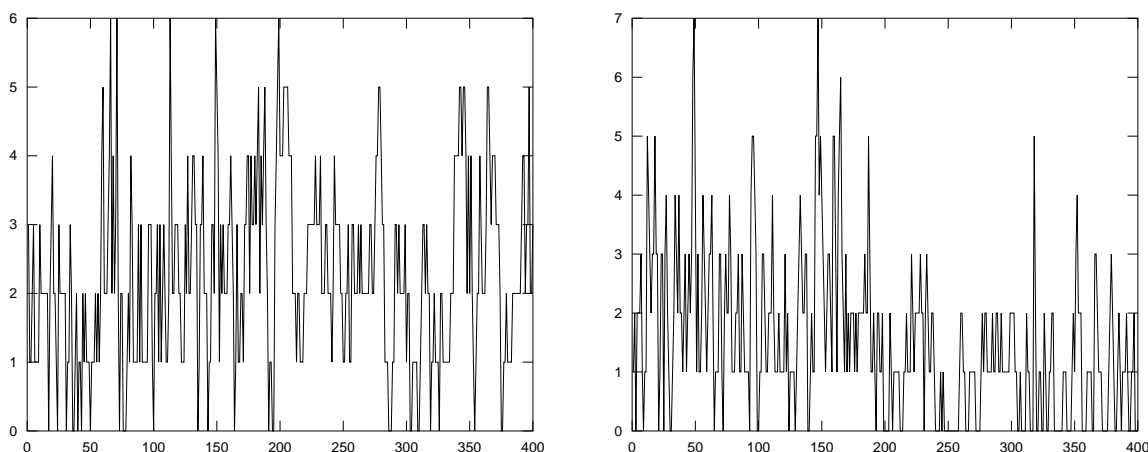


Figure 2: Typical trajectories: α_1 changes to 0.8 and 0.6, respectively, μ_ε changes to 0.4

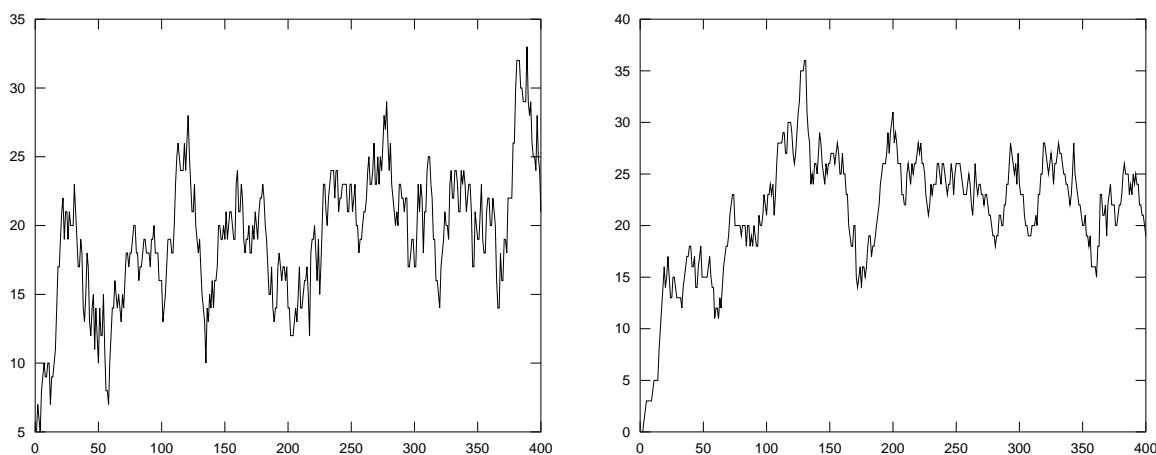


Figure 3: Typical trajectories: α_1 is 0.92 and 0.95, respectively, μ_ε is 1.6 and 1, respectively

4.4 The Effect of Different Trajectory Lengths and Change Points

The above anomalies can be avoided by investigating longer trajectories. We did this for three different cases. The first one was the nearly unstable case of $\alpha_1 = 0.9$ and $\mu_\varepsilon = 2$ after the change. Here the power increased rapidly and reached 0.985 at the trajectory length of 1400. The second case was $\alpha_1 = 0.4$ and $\mu_\varepsilon = 1.2$ after the change. Here both changes are very small and in opposite directions, partly cancelling each other out. It is no surprise that the test performed very poorly for a trajectory length of 400, and longer trajectories repaired this problem only slowly, with the power reaching 0.9 only at the trajectory length of 6600. However, the third case demonstrated that this is mostly due to the small change in the values and not the fact that the mean of the stationary distribution remained constant. Here, the mean of the stationary distribution remained constant again but the changes were greater in value – α_1 changed to 0.8 and μ_ε to 0.4. The power reached 0.9 for a trajectory length of only 600. Tables 3, 4 and 5 summarize these results.

It can also be expected that the test would perform poorly when the change occurs near the beginning or the end of the trajectory. The simulation confirmed this conjecture.

Table 1: First-order process, change in the coefficient

Empirical power for constant $\mu_\varepsilon = 1$, α_1 changes from 0.5

α_1 to	0	0.1	0.2	0.3	0.4	0.6	0.7	0.8	0.9
Power	1	0.996	0.928	0.635	0.204	0.190	0.806	0.969	0.055

Table 2: First-order process, change in the innovation

Empirical power for constant $\alpha_1 = 0.5$, μ_ε changes from 1

μ to	0	0.2	0.4	0.6	0.8	1.2	1.4	1.6	1.8	2
Power	1	1	0.995	0.802	0.198	0.175	0.534	0.839	0.978	0.998

Table 3: Effect of trajectory length for the nearly unstable case

Empirical power, μ_ε from 1 to 2, α_1 changes from 0.5 to 0.9

n	800	900	1000	1100	1200	1300	1400
Power	0.037	0.141	0.378	0.632	0.843	0.931	0.985

Table 4: Effect of trajectory length for small change

Empirical power, μ_ε from 1 to 1.2, α_1 changes from 0.5 to 0.4

n	400	1000	1800	2600	3400	4200	5000	5800	6600
Power	0.092	0.229	0.349	0.552	0.631	0.762	0.806	0.874	0.914

Table 5: Effect of trajectory length when the mean of the stationary distribution is constant

Empirical power, μ_ε from 1 to 0.4, α_1 changes from 0.5 to 0.8

n	200	300	400	500	600	700	800	900	1000
Power	0.379	0.595	0.75	0.863	0.914	0.961	0.988	0.993	0.996

However, it also showed that the test performed well in a broader range near the middle of the trajectory than would have been expected. The results can be found in Tables 6 and 7. Trajectory length was again 400 for comparison with the previous subsection, α_1 changed to 0.3 and 0.7 and μ_ε to 0.8 and 1.2 for a downward and upward change, respectively.

The simulation also revealed an interesting phenomenon: the power was significantly greater if the change occurred at the beginning of the trajectory for a downward change. However, an upward change was detected with greater probability near the end of the trajectory.

One heuristic reasoning for this is the following: large values of the test process appear because the parameters are misestimated. The effect is that in every step \widehat{M}_k differs from M_k largely. These deviations sum up and result in large absolute values of the test process—and, ultimately, the rejection of the null hypothesis. However, this effect is partly canceled out by the smaller values of $(\widehat{I}_n(\widehat{\theta}^{(n)}))^{-1/2}$. The values of $\widehat{I}_n(\widehat{\theta}^{(n)})$ increase with the increasing estimated variance of the innovation and offspring distributions. In this case, all these variances increase with the respective means increasing. For a downward change near the beginning of the trajectory, the smaller mean is dominant and the estimate will be close to it, therefore $\widehat{I}_n(\widehat{\theta}^{(n)})$ will have smaller values and will decrease extreme values of the process to a smaller extent. If the change occurs near the end of

Table 6: Effect of the change point for downward change

Empirical power, μ_ε from 1 to 0.8, α_1 changes from 0.5 to 0.3

Change point	40	80	120	160	200	240	280	320	360
Power	0.389	0.695	0.846	0.899	0.875	0.780	0.585	0.335	0.086

Table 7: Effect of the change point for upward change

Empirical power, μ_ε from 1 to 1.2, α_1 changes from 0.5 to 0.7

Change point	40	80	120	160	200	240	280	320	360
Power	0.151	0.468	0.668	0.807	0.935	0.980	0.980	0.931	0.652

the trajectory, then the larger mean is dominant and larger values of $\hat{I}_n(\hat{\theta}^{(n)})$ will decrease the extreme values to a greater extent, therefore the null hypothesis will be accepted with a greater probability. For an upward change, these phenomena occur similarly, only for opposite ends of the trajectory.

4.5 Individual Tests

The next area of investigation was the performance of the individual tests. First we checked whether a change in one of the parameters caused a false detection for the other one. For this we applied Test 2 for each parameter and changed only the other one. The range of after-change parameters and every other parameter was the same as in 4.3. The results are contained in Tables 8 and 9.

Table 8: First-order process, two-sided test for the unchanged coefficient

Accept rates of Test 2 for α_1 for constant $\alpha_1 = 0.5$, μ_ε changes from 1

μ to	0	0.2	0.4	0.6	0.8	1.2	1.4	1.6	1.8	2
Rate	0.929	0.976	0.899	0.864	0.923	0.940	0.853	0.720	0.620	0.510

Table 9: First-order process, two-sided test for the unchanged innovation

Accept rates of Test 2 for μ_ε for constant $\mu_\varepsilon = 1$, α_1 changes from 0.5

α to	0	0.1	0.2	0.3	0.4	0.6	0.7	0.8	0.9
Power	0.405	0.533	0.754	0.854	0.939	0.921	0.810	0.725	0.993

The findings are the following: for the α_1 coefficient the test performed poorly only if there was a great upward change in μ . This is due to a fundamental shortcoming of the application of CLS estimation to these processes. The CLS estimates are designed for an INAR(p) process with no change and perform very well under those conditions. If, however, the parameters change over time, the behavior of the CLS estimates cannot be simply described. Especially if a drastic change occurs in the parameters and the variance of the process is relatively large, the best-fitting model may well be one with relatively small innovation but α_1 close to 1. It was indeed found that for large changes of μ_ε the CLS estimate of α_1 differed greatly from the real value even if there was no change in

α_1 . This phenomenon was even more pronounced for larger innovations (see 4.6) and also affected μ_ε . It is important to note, however, that while the estimates for μ_ε were incorrect, the null hypothesis was accepted for μ_ε with more probability than for α_1 .

Another point of interest can be the the value of the last two entries in Table 9. When α changes to 0.8, a change is falsely detected in μ_ε with 0.275 probability. However, when α changes to 0.9, a larger value, the rate of false detections drops to 0.007. This is again due to the nearly unstable phenomenon. The estimation of the parameters differs from the real values, but this effect is countered by the increased variance, and therefore the much smaller values of $(\hat{I}_n(\hat{\theta}^{(n)}))^{-1/2}$. Therefore, the absolute value of the process remains small and the null hypothesis is accepted.

Next, we investigated the power of the one-sided tests. For this, we used Test 1, which was performed individually for both parameters with no change occurring in the other parameter. As can be seen in Tables 10 and 11 the one-sided test did not perform notably better than the simultaneous two-sided test.

Table 10: First-order process, one-sided test for the innovation

		Empirical power for constant $\alpha_1 = 0.5$, μ_ε changes from 1									
μ to		0	0.2	0.4	0.6	0.8	1.2	1.4	1.6	1.8	2
Power		1	1	0.999	0.846	0.245	0.222	0.582	0.873	0.974	0.993

Table 11: First-order process, one-sided test for the coefficient

		Empirical power for constant $\mu_\varepsilon = 1$, α_1 changes from 0.5									
α to		0	0.1	0.2	0.3	0.4	0.6	0.7	0.8	0.9	
Power		1	0.997	0.961	0.707	0.215	0.198	0.872	0.984	0.129	

Here a little explanation may be needed on the design of Test 1. As noted in 4.4, extreme values of the test statistics arise due to a misestimation in the parameters. An upward change in one of the parameters means that the parameter will be overestimated before the change and underestimated after it (this is not always the case as will be seen in 4.6 but is generally true). Therefore the test statistics takes negative values before the change and then begins to increase after it. Hence, if it takes a value outside the critical interval, then that will be a negative value. Similarly, a downward change will result in large positive values of the test statistics.

4.6 The Magnitude of the Innovation

In the next case, the innovation distribution was again Poisson, but this time with a mean of $\mu_\varepsilon = 100$. The question here was whether the power was dependent on the absolute or the relative value of change. The coefficient did not change, all other parameters were the same as in 4.3 and the simultaneous two-sided test was carried out for both parameters. The results can be found in Table 12.

There were two important findings: first, the Type II error rate seems to depend on the absolute rather than the relative amount of change in the innovation mean, because here the empirical power was 1 for almost every case. There were two important exceptions.

Table 12: First-order process, large innovation
 Empirical power for constant $\alpha_1 = 0.5$, μ_ε changes from 100
 Simultaneous two-sided test

μ to	0	20	40	60	80	120	140	160	180	200
Power	0	0.182	1	1	1	1	1	1	0.648	0

A large upward change in μ_ε resulted in smaller power—this was similar to the nearly unstable case in 4.3. However, a large downward change also caused an empirical power of 0. The explanation is again connected with the CLS estimates and $\hat{I}_n(\hat{\theta}^{(n)})$. As explained in 4.5, the CLS estimates differ greatly from the real values and the fitted process has a very high variance, which results in large entries in $\hat{I}_n(\hat{\theta}^{(n)})$. As a consequence, the test process has small values and stays almost always within the critical band.

4.7 The Second-Order Case

In the next phase we investigated the INAR(2) model and especially the question of how much a change in one parameter can be separated from the others.

To investigate false detections, the coefficients were 0.3–0.3 before the change. After the change, one of the coefficients changed to a value from $\{0, 0.1, \dots, 0.6\}$ with no change occurring in the other coefficient. Test 2 was then performed for the unchanged parameter. The innovation mean did not change. All other parameters were the same as in 4.3.

The results were satisfactory: a change in either of the coefficients did not lead to false detection in the other in more than 10 % of the cases. This result is in contrast with 4.5, which shows that a large change in either the coefficients or the innovation mean can lead to false detection in the other one. Tables 13 and 14 summarize these results.

On the other hand, the higher order of the process decreases the power of the individual two-sided test. We tested with the following settings: the coefficients before the change were 0.25–0.25 and the innovation mean 1. After the change, either one of the coefficients changed to a value from $\{0.05, 0.15, \dots, 0.65\}$ or the innovation mean to a value from $\{0, 0.2, \dots, 1.8\}$. Test 2 was then performed for the changed parameter. All the other settings were the same as in 4.3.

The results were not satisfactory: the empirical power was notably smaller for the coefficients than in the first-order case. The innovation mean did not seem to be affected by this phenomenon. The results for the coefficients are contained in Table 15. The empirical powers increased with longer trajectories in the same manner as described in 4.4.

A final test was performed to see how accurately the test can detect a shift from α_1 to α_2 . Here, initially we had $\alpha_1 = \alpha_2 = 0.3$. After the change α_1 was chosen from $\{0, 0.1, \dots, 0.6\}$ and $\alpha_2 = 0.6 - \alpha_1$. The innovation mean was $\mu_\varepsilon = 1$. All other parameters were the same as in 4.3. The finding here was that the shift in both directions was detected with roughly equal probability and the power was smaller than in the case where only one of the coefficients changed. It is, however, not very surprising that a smaller

Table 13: Second-order process, change in the coefficient

Empirical power of Test 2 for all three parameters
 $\alpha_2 = 0.25, \mu_\varepsilon = 1, \alpha_1$ changes from 0.25

α_1 to	0.05	0.15	0.2	0.3	0.35	0.45	0.55	0.65
Power	0.593	0.171	0.066	0.05	0.17	0.501	0.519	0.194

Table 14: Second-order process, change in the innovation

Empirical power for constant $\alpha_1 = \alpha_2 = 0.25, \mu_\varepsilon$ changes from 1

μ to	0	0.2	0.4	0.6	0.8	1.2	1.4	1.6	1.8
Power	1	1	0.982	0.718	0.19	0.159	0.469	0.736	0.896

Table 15: Second-order process, two-sided test for the unchanged coefficient

Accept rates of Test 2 for α_2
 α_1 changes from 0.3, $\alpha_2 = 0.3, \mu_\varepsilon = 1$

α_1 to	0	0.1	0.2	0.4	0.5	0.6
Rate	0.900	0.924	0.946	0.958	0.960	0.988

Table 16: Second-order process, change in the structure

Empirical power for constant $\mu = 1$
 α_1 and α_2 both change from 0.3, $\alpha_1 + \alpha_2 = 0.6$ throughout

α_1 to	0	0.05	0.1	0.15	0.2	0.25	0.35	0.4	0.45	0.5	0.55
Power	0.585	0.421	0.312	0.164	0.073	0.043	0.048	0.090	0.182	0.344	0.626

change in the structure of the process, like this, is detected with smaller probability. The exact numbers can be found in Table 16.

5 Conclusions

From the described simulation we can draw the following conclusions: the test is able to detect change in any of the described parameters. However, the power of the test decreases for changes which occur very close to the beginning and the end of the trajectory. Larger changes are detected with greater probability and a change in the mean of the stationary distribution adds to the power of the test greatly. The absolute, rather than the relative magnitude of the change in the innovation influences the power of the test. One-sided tests are more powerful than their two-sided alternatives.

Unfortunately, the test also has several shortcomings. The individual test falsely detects change if there is a large change in one of the other parameters. If the coefficient is close to 1 or the change in the parameters is large enough, the CLS fitted process has large variance and the test fails in almost every case. Therefore, the test is best applied if the change in the parameters is moderate and the coefficient is not too close to 1.

These observations are valid for a trajectory length of 400 but longer trajectories increase the power greatly, and for a trajectory with several thousand data points, changes are almost always detected. For long trajectories the Type I error rate is approximately the

same as can be expected from significance levels, while for shorter ones it may be a few percents lower. With these limitations, however, the test can be used to detect parameter change.

Acknowledgements

First of all I would like to thank my consultant Gyula Pap for all the invaluable comments that he has made to this manuscript, and for introducing me to this field of statistics. I am also grateful to the kind staff at the Bolyai Institute at the University of Szeged for being there whenever I had to make use of the Institute's computer facilities. The reviewer at the Austrian Journal of Statistics has made many insightful comments which have helped me clarify several crucial points in the paper.

References

- Alzaid, A. A., and Al-Osh, M. A. (1987). First order integer-valued autoregressive (INAR(1)) processes. *Journal of Time Series Analysis*, 8, 261-275.
- Alzaid, A. A., and Al-Osh, M. A. (1990). An integer-valued p th order autoregressive structure (INAR(p)) process. *Journal of Applied Probability*, 27, 314-324.
- Barczy, M., Ispány, M., and Pap, G. (2011). Asymptotic behavior of unstable INAR(p) processes. *Stochastic Processes and Applications*, 121, 583-608.
- Berkes, I., Gombay, E., and Horváth, L. (2009). Testing for changes in the covariance structure of linear processes. *Journal of Statistical Planning and Inference*, 139, 2044-2063.
- Csörgő, M., and Horváth, L. (1997). *Limit Theorems in Change-point Analysis*. New York: Wiley.
- Kang, J., and Lee, S. (2009). Parameter change test for random coefficient integer-valued autoregressive processes with application to polio data analysis. *Journal of Time Series Analysis*, 30, 239-258.
- Klimko, L. A., and Nelson, P. I. (1978). On conditional least squares estimation for stochastic processes. *The Annals of Statistics*, 6, 629-642.
- Kuiper, N. H. (1960). Tests concerning random points on a circle. *Nederl. Akad. Wetensch. Proc. Ser. A*, 63, 38-47.
- Lee, S., and Na, O. (2005). Test for parameter change in stochastic processes based on conditional least-squares estimator. *Journal of Multivariate Analysis*, 93, 375-393.
- Steutel, F. W., and Harn, K. van. (1979). Discrete analogues of self-decomposability and stability. *The Annals of Probability*, 8, 893-899.

Author's address:

Tamás T. Szabó
Szegedi Tudományegyetem Bolyai Intézet
Szeged, Aradi vértanúk tere 1.
H-6725 Hungary
E-mail: szabo.tamas.10@gmail.com