Change detection problems in branching processes

Outline of Ph.D. thesis

by

Tamás T. Szabó

Thesis advisor:

Professor Gyula Pap

Doctoral School of Mathematics and Computer Science
Bolyai Institute, University of Szeged
Szeged, 2017

1 Introduction

Change detection is a naturally occurring question in statistics, and time series analysis in particular. One of the most widely used assumptions is that the dynamics of the process do not change over time, which allows us to collect a large enough samples for analysis. Obtaining a test for that assumption is therefore a natural desire.

There have been numerous attempts to solve this problem, mainly in the domain of independent observations and classical time series – for an overview of these, see the excellent monograph of Csörgő and Horváth (1997). The standard methods in this space are either likelihood-based, or the cumulative sums (CUSUM) method, first proposed by Page (1954, 1955)

Change detection in branching processes has received relatively little attention from researchers. The extra randomness introduced by the branching mechanism makes likelihood-based methods largely unworkable, so alternative approaches have to be used. Our test process will draw on the so-called quasi-likelihood approach (see Gombay, 2008), which offers suggestions and conjectures based on earlier work, and can actually be written in CUSUM form. The proofs, however, rely on classical martingale theory – especially as set forth in Karatzas and Shreve (1991) and Jacod and Shiryaev (2003) – and a critical Hájek-Rényi type inequality by Kokoszka and Leipus (1998).

We will use a few standard notations: $\otimes$ will denote the Kronecker product, and $\mathbf{1}_i$ the i-th unit vector. Filtrations will always be the natural ones of the process. We will use Landau notation for rates of convergence: for a stochastic process X_t, the notation $X_t = \mathrm{O}_{\mathbb{P}}(g(t))$ means that the collection of measures $\big(\mathcal{L}\big(\frac{X_t}{g(t)}\big)\big)_{t>t_0}$ is tight for some $t_0 \geqslant 0$ ($\mathcal{L}$ stands for distribution). Also, $X_t = \mathrm{o}_{\mathbb{P}}(g(t))$ means simply that $\frac{X_t}{g(t)} \xrightarrow{\mathbb{P}} 0$.

2 Change detection: basic approach

The basic setup will be the following:

1. We will take a vector-valued process $\boldsymbol{X}_t$, indexed either by the natural numbers or the nonnegative real numbers and take a sample of it on the interval $0 \leqslant t \leqslant T$.

2. We will choose a parameter $\boldsymbol{\theta}_t$ governing the dynamics of the process. The main question will be whether this parameter is constant in t, or, formally, we would like to test

$$\mathrm{H}_0 : \ \exists \boldsymbol{\theta} : \ \boldsymbol{\theta}_t = \boldsymbol{\theta}, \quad t \in [0, T]$$

against the alternative hypothesis

$$\mathrm{H}_\mathrm{A} : \exists \rho \in (0,1) : \ \boldsymbol{\theta}_t = \boldsymbol{\theta}', \ t \in [0, \rho T) \ \text{ and } \ \boldsymbol{\theta}_t = \boldsymbol{\theta}'', \ t \in [\rho T, T]$$

for some $\boldsymbol{\theta}' \neq \boldsymbol{\theta}''$. An important additional condition will be for stability: $\boldsymbol{\theta}$, $\boldsymbol{\theta}'$, $\boldsymbol{\theta}''$ have to be such that $\boldsymbol{X}$ have a unique stationary distribution under H_0, and both parts of the process (before and after the change) have a unique stationary distribution under H_A.

3. We will find an appropriate vector-valued function f such that

$$\boldsymbol{M}_t := \boldsymbol{X}_t - \boldsymbol{X}_0 - \int_0^t f(\boldsymbol{\theta}_s; \boldsymbol{X}_{s-}) \, \mathrm{d}s$$

will be a martingale. Here $\boldsymbol{X}_{s-}$, a slightly informal notation, means $\boldsymbol{X}_s$ for continuous s and $\boldsymbol{X}_{s-1}$ for discrete s. Similarly, the integral is simply a sum for discrete s.

4. Assuming $\boldsymbol{\theta}_t = \boldsymbol{\theta}$ for all t, we will estimate $\boldsymbol{\theta}$ with $\widehat{\boldsymbol{\theta}}_T$ based on the conditional least squares (CLS) method of Klimko and Nelson (1978).

5. We will replace $\boldsymbol{\theta}_t$ with $\widehat{\boldsymbol{\theta}}_T$ in the definition of $\boldsymbol{M}_t$ to obtain $\widehat{\boldsymbol{M}}_t^{(T)}$.

6. We will prove that if $\boldsymbol{\theta}_t$ is constant in t, then

$$\widehat{\boldsymbol{\mathcal{M}}}_u^{(T)} := \widehat{\boldsymbol{I}}_T^{-1/2} \widehat{\boldsymbol{M}}_{uT}^{(T)}, \quad u \in [0,1]$$

converges in distribution to a Brownian bridge on $[0,1]$, for some random normalizing matrix $\widehat{\boldsymbol{I}}_T$, which is calculable from the sample.

7. Consequently, we will construct tests for the change in $\boldsymbol{\theta}$, using the supremum or infimum of $\widehat{\boldsymbol{\mathcal{M}}}_u^{(T)}$ as a test statistic (based on the direction of change).

8. We will prove that if there is a single change in $\boldsymbol{\theta}_t$ on $[0, T]$, then the test statistic will tend to infinity stochastically as $T \to \infty$.

9. We will prove that the *arg max*, or *arg min*, of $\widehat{\boldsymbol{\mathcal{M}}}_u^{(T)}$ is a good estimator of the change point in $\boldsymbol{\theta}_t$.

3 Change detection: the INAR(p) process

First we give an overview of the results achieved in Pap and Szabó (2013). The integer valued autoregressive process of order p (abbreviated INAR(p)) was first proposed by Alzaid and Al-Osh (1987) for $p = 1$ and Du and Li (1991) for higher p values, and is defined by the following equation:

$$X_k = \alpha_1 \circ X_{k-1} + \cdots + \alpha_p \circ X_{k-p} + \varepsilon_k, \ k = 1, 2, \ldots, \tag{3.1}$$

where the ε_k are i.i.d nonnegative integer-valued random variables with mean μ, and for a random nonnegative integer-valued random variable Y and $\alpha \in (0, 1)$, $\alpha \circ Y$ denotes the sum of Y i.i.d Bernoulli random variables with mean α, also independent of Y. We extend the state space to create a Markov chain, and define our parameter vector:

$$\boldsymbol{X}_k := \begin{bmatrix} X_k \\ X_{k-1} \\ \vdots \\ X_{k-p} \end{bmatrix}, \quad \boldsymbol{\theta} := \begin{bmatrix} \boldsymbol{\alpha} \\ \mu \end{bmatrix} := \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_p \\ \mu \end{bmatrix}.$$

Change detection methods for INAR(p) processes in general (i.e., with no prespecified innovation distribution) have only been proposed in a few papers – we refer to Kang and Lee (2009) especially, where the authors give a test statistics similar to ours for a more general model. However, no result is available there under the alternative hypothesis and the asymptotics of the change-point estimator are not given – we will give some answers to both of these questions, which strengthens the theoretical foundations of the method considerably.

Our martingale will be defined by its differences, which will be given by subtracting the conditional expectation from the process values:

$$M_k = X_k - \mathbb{E}(X_k|\mathcal{F}_{k-1}) = X_k - \boldsymbol{\alpha}^\top \boldsymbol{X}_{k-1} - \mu, \qquad k = 1, 2, \ldots. \tag{3.2}$$

Minimizing the sum of squares for these quantities, we obtain the following CLS estimates:

(3.3)

$$\widehat{\boldsymbol{\theta}}_n := \begin{bmatrix} \widehat{\boldsymbol{\alpha}}_n \\ \widehat{\mu}_n \end{bmatrix} := \boldsymbol{Q}_n^{-1} \sum_{k=1}^n X_k \begin{bmatrix} \boldsymbol{X}_{k-1} \\ 1 \end{bmatrix}, \; \boldsymbol{Q}_n := \sum_{k=1}^n \begin{bmatrix} \boldsymbol{X}_{k-1} \\ 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_{k-1} \\ 1 \end{bmatrix}^\top.$$

Now we define our test process as

$$\widehat{\boldsymbol{\mathcal{M}}}_n(t) := \widehat{\boldsymbol{I}}_n^{-1/2} \sum_{k=1}^{\lfloor nt \rfloor} \widehat{M}_k^{(n)} \begin{bmatrix} \boldsymbol{X}_{k-1} \\ 1 \end{bmatrix}, \tag{3.4}$$

where $\widehat{\boldsymbol{I}}_n$ is an estimate of the information matrix (obtainable by a "score vector" analogy):

$$\widehat{\boldsymbol{I}}_n := \sum_{k=1}^n ((\widehat{\boldsymbol{\alpha}}_n^*)^\top \boldsymbol{X}_{k-1} + \widehat{\sigma}_n^2) \begin{bmatrix} \boldsymbol{X}_{k-1} \\ 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_{k-1} \\ 1 \end{bmatrix}^\top,$$

with

$$\widehat{\boldsymbol{\alpha}}_n^* := \begin{bmatrix} \widehat{\alpha}_n^{(1)}(1 - \widehat{\alpha}_n^{(1)}) \\ \vdots \\ \widehat{\alpha}_n^{(p)}(1 - \widehat{\alpha}_n^{(p)}) \end{bmatrix}, \quad \widehat{\sigma}_n^2 := -\frac{1}{n} \sum_{k=1}^n \left(\left(\widehat{M}_k^{(n)} \right)^2 - (\widehat{\boldsymbol{\alpha}}_n^*)^\top \boldsymbol{X}_{k-1} \right).$$

We will impose some regularity conditions on the process, most notably that the sum of the α parameters be less than 1 – that is, that our process be *stable.* These conditions guarantee the ergodicity of the process and the existence of a unique stationary distribution.

3.1 Definition. *An INAR(p) process* $(X_k)_{k\geqslant -p+1}$ *is said to satisfy condition* $\mathbf{C}_0$, *if* $\mathbb{E}(X_0^6) < \infty, \ldots, \mathbb{E}(X_{-p+1}^6) < \infty$, $\mathbb{E}(\varepsilon_1^6) < \infty$, $\alpha_1 + \cdots + \alpha_p < 1$, $\mu > 0$ *all hold for it, and if, furthermore,* $\alpha_p > 0$ *and the greatest common denominator of the numbers* i *such that* $\alpha_i > 0$ *is 1.*

Under the null hypothesis of no change, our main result is the following:

3.2 Theorem. *If* $(X_k)_{k\geqslant -p+1}$ *satisfies condition* $\mathbf{C}_0$ *and* H_0 *holds, then*

$$\widehat{\boldsymbol{\mathcal{M}}}_n \xrightarrow{\mathcal{D}} \boldsymbol{\mathcal{B}} \qquad \text{as} \quad n \to \infty,$$

where $(\boldsymbol{\mathcal{B}}(t))_{0\leqslant t\leqslant 1}$ *is a* $(p+1)$*-dimensional standard Brownian bridge, and* $\xrightarrow{\mathcal{D}}$ *denotes convergence in distribution in the Skorokhod space* $D([0,1])$.

Based on this result it is easy to construct tests for parameter change, by comparing some functional (e.g., supremum or infimum) of the test process's components with the known distribution of the respective functional for a Brownian bridge.

Under the alternative hypothesis we will work under the following assumptions:

3.3 Definition. *We will say that an INAR(p) process* $(X_k)_{k\geqslant -p+1}$ *satisfies* $\mathbf{C}_{\mathrm{A}}$ *if* $\tau = \lfloor n\rho \rfloor$ *for some* $\rho \in (0,1)$, *both* $(X_k)_{-p+1\leqslant k\leqslant \tau}$ *and* $(X_k)_{k\geqslant \tau+1}$ *satisfy condition* $\mathbf{C}_0$, *and the parameter vectors for the processes* $(X_k)_{-p+1\leqslant k\leqslant \tau}$ *and* $(X_k)_{k\geqslant \tau+1}$ *are*

$$\boldsymbol{\theta}' := \begin{bmatrix} \boldsymbol{\alpha}' \\ \mu' \end{bmatrix}, \quad \text{and} \quad \boldsymbol{\theta}'' := \begin{bmatrix} \boldsymbol{\alpha}'' \\ \mu'' \end{bmatrix},$$

respectively. In this case, $\widetilde{\boldsymbol{X}}'$ *and* $\widetilde{\boldsymbol{X}}''$ *will denote variables with the unique stationary distributions of the two halves of the process, respectively. We* *will use the following notations:*

$$\boldsymbol{Q}' := \mathbb{E}\left(\begin{bmatrix} \widetilde{\boldsymbol{X}}' \\ 1 \end{bmatrix} \begin{bmatrix} \widetilde{\boldsymbol{X}}' \\ 1 \end{bmatrix}^\top\right), \; \boldsymbol{Q}'' := \mathbb{E}\left(\begin{bmatrix} \widetilde{\boldsymbol{X}}'' \\ 1 \end{bmatrix} \begin{bmatrix} \widetilde{\boldsymbol{X}}'' \\ 1 \end{bmatrix}^\top\right), \; \widetilde{\boldsymbol{I}} := \lim_{n\to\infty} \frac{\widehat{\boldsymbol{I}}_n}{n}.$$

The test's weak consistence is proved by the following theorem.

3.4 Theorem. *Suppose that $\boldsymbol{C}_A$ holds. For $i = 1, 2, \ldots, p+1$, let us define*

$$\psi_i := \mathbf{1}_i^\top \widetilde{\boldsymbol{I}}^{-1/2}((\rho\boldsymbol{Q}')^{-1} + ((1-\rho)\boldsymbol{Q}'')^{-1})^{-1}(\boldsymbol{\theta}' - \boldsymbol{\theta}'').$$

If $\psi_i > 0$ then for the i-th component of the test process,

$$\sup_{0\leqslant t\leqslant 1} \widehat{\boldsymbol{\mathcal{M}}}_n(t)^{(i)} = n^{1/2}\psi_i + \mathrm{o}_{\mathbb{P}}(n^{1/2}),$$

and conversely, if $\psi_i < 0$ then

$$\inf_{0\leqslant t\leqslant 1} \widehat{\boldsymbol{\mathcal{M}}}_n(t)^{(i)} = n^{1/2}\psi_i + \mathrm{o}_{\mathbb{P}}(n^{1/2}).$$

As mentioned earlier, the estimate of the change point, $\widehat{\tau}_n$, will be the arg min or arg max of the test process. For this, we have the following result.

3.5 Theorem. *If $\mathbf{C}_A$ holds, then we have*

$$\widehat{\tau}_n - \lfloor n\rho \rfloor = \mathrm{O}_{\mathbb{P}}(1) \qquad \text{as } \; n \to \infty.$$

Consequently, if we define $\widehat{\rho}_n := \frac{\widehat{\tau}_n}{n}$, then $\widehat{\rho}_n - \rho = \mathrm{O}_{\mathbb{P}}(n^{-1})$.

4 Change detection: the Cox–Ingersoll–Ross process

The approach for the previous procedure can be extended to continuous time as well. We will now give a brief account of this based on Pap and Szabó (2016). The Cox–Ingersoll–Ross (CIR) process is defined by

$$\mathrm{d}Y_t = (a - bY_t)\,\mathrm{d}t + \sigma\sqrt{Y_t}\,\mathrm{d}W_t, \quad t \geqslant 0, \tag{4.1}$$

where $a > 0$, $b > 0$, $\sigma > 0$ and $(W_t)_{t \geqslant 0}$ is a standard Wiener process. These constraints ensure that our process has a unique stationary distribution and is ergodic, and also that any solution starting from a nonnegative value stays nonnegative almost surely.

The CIR process (also known as Feller diffusion) was first investigated by Feller (1951), proposed as a short-term interest-rate model by Cox et al. (1985), and became one of the most widespread "short rate" models in financial mathematics. Inevitably, therefore, describing its statistical properties is of great importance and has received considerable interest.

There are a handful of change detection tests for the CIR process in the literature: Schmid and Tzotchev (2004) used control charts and a sequential method i.e., an online procedure, while Guo and Härdle (2017) used the local parameter approach based on approximate maximum likelihood estimates – in essence, they wanted to find the largest interval for which the sample fits the model.

Our parameter vector in this case will be

$$\boldsymbol{\theta} := \begin{bmatrix} a \\ b \end{bmatrix}.$$

Change detection in σ is not necessary, since we can establish almost surely whether σ is constant across our sample, as it can be calculated almost surely from an arbitrarily small part of a continuous sample.

We will use the following martingale:

$$M_s := Y_s - Y_0 - \int_0^s (a - bY_u)\,\mathrm{d}u = \sigma \int_0^s \sqrt{Y_u}\,\mathrm{d}W_u, \qquad s \geqslant 0. \tag{4.2}$$

Our estimates will be based on a formal analogy, but they can also be obtained as the limit of the discrete time CLS estimates (Overbeck and Rydén, 1997, Theorems 3.1 and 3.3). Nevertheless, the exact method by which we arrive at them is less important than their structure, and the consequent results we can deduce with their help.

$$\widehat{\boldsymbol{\theta}}_T := \begin{bmatrix} \widehat{a}_T \\ \widehat{b}_T \end{bmatrix} := \left(\int_0^T \begin{bmatrix} 1 \\ -Y_s \end{bmatrix} \begin{bmatrix} 1 \\ -Y_s \end{bmatrix}^\top \mathrm{d}s \right)^{-1} \int_0^T \begin{bmatrix} 1 \\ -Y_s \end{bmatrix} \mathrm{d}Y_s. \tag{4.3}$$

Based on this, our test process will be

$$\widehat{\boldsymbol{\mathcal{M}}}_t^{(T)} := \boldsymbol{I}_T^{-1/2} \int_0^{tT} \begin{bmatrix} 1 \\ -Y_s \end{bmatrix} \mathrm{d}\widehat{M}_s^{(T)}, \qquad t \in [0,1]. \tag{4.4}$$

In this case the information matrix has a relatively simple form:

$$\boldsymbol{I}_{tT} := \sigma^2 \int_0^{tT} \begin{bmatrix} Y_s & -Y_s^2 \\ -Y_s^2 & Y_s^3 \end{bmatrix} \mathrm{d}s, \qquad t \in [0,1].$$

Under the null hypothesis we have the following result, which will again enable us to construct change detection tests for a change in either direction. The extra moment conditions in $\mathbf{C}_0$ are not necessary here, since all moments of a CIR process are always finite.

4.1 Theorem. *Let $(Y_t)_{t\in\mathbb{R}_+}$ be CIR process such that $\mathbb{P}(Y_0 \in \mathbb{R}_+) = 1$. Then*

$$\widehat{\boldsymbol{\mathcal{M}}}^{(T)} \xrightarrow{\mathcal{D}} \boldsymbol{\mathcal{B}} \qquad \text{as } T \to \infty,$$

where $(\boldsymbol{\mathcal{B}}(t))_{0\leqslant t\leqslant 1}$ is a 2-dimensional standard Brownian bridge, and the convergence happens in distribution on the Skorokhod space $D([0,1])$.

Under the alternative hypothesis we have the following result, which closely matches Theorem 3.4.

4.2 Theorem. *Let us suppose that $\boldsymbol{\theta}$ changes from $\boldsymbol{\theta}'$ to $\boldsymbol{\theta}''$ at time ρT, where $\rho \in (0,1)$, and both $\boldsymbol{\theta}' > \mathbf{0}$ and $\boldsymbol{\theta}'' > \mathbf{0}$ componentwise. Let us take $i \in \{1,2\}$, and then define*

$$\psi_i := \mathbf{1}_i^\top \widetilde{\boldsymbol{I}}^{-1/2}((\rho\boldsymbol{Q}')^{-1} + ((1-\rho)\boldsymbol{Q}'')^{-1})^{-1}(\boldsymbol{\theta}' - \boldsymbol{\theta}''), \ \widetilde{\boldsymbol{I}} := \lim_{T\to\infty} \frac{\boldsymbol{I}_T}{T}.$$

If $\psi_i > 0$, then we have

$$\sup_{t\in[0,T]} \widehat{\boldsymbol{\mathcal{M}}}_{t,i}^{(T)} = T^{1/2}\psi_i + \mathrm{o}_{\mathbb{P}}(T^{1/2}).$$

On the other hand, if $\psi_i < 0$, we have

$$\inf_{t\in[0,T]} \widehat{\boldsymbol{\mathcal{M}}}_{t,i}^{(T)} = T^{1/2}\psi_i + \mathrm{o}_{\mathbb{P}}(T^{1/2}).$$

For the CIR process we estimate the point of change with $\widehat{\tau}_T$ and investigate the properties of $\widehat{\rho}_T := \frac{\widehat{\tau}_T}{T}$. The following theorem is easily obtained from the counterpart of Theorem 3.5 for the CIR process.

4.3 Theorem. *Under the assumptions of Theorem 4.2, if there is a change only in a or only in b, then, for the appropriate change-point estimate we have $\widehat{\rho}_T - \rho = \mathrm{O}_{\mathbb{P}}(T^{-1})$.*

5 CLS-like parameter estimation for the Heston model

Lastly, based on Barczy et al. (2016), we propose conditional least squares estimators for the Heston model, which is a solution, for $t \geqslant 0$, of the two-dimensional stochastic differential equation

$$\begin{cases} \mathrm{d}Y_t = (a - bY_t)\,\mathrm{d}t + \sigma_1\sqrt{Y_t}\,\mathrm{d}W_t, \\ \mathrm{d}X_t = (\alpha - \beta Y_t)\,\mathrm{d}t + \sigma_2\sqrt{Y_t}\big(\varrho\,\mathrm{d}W_t + \sqrt{1-\varrho^2}\,\mathrm{d}B_t\big), \end{cases} \tag{5.1}$$

where $a \in \mathbb{R}_{++}$, $b, \alpha, \beta \in \mathbb{R}$, $\sigma_1 \in \mathbb{R}_{++}$, $\sigma_2 \in \mathbb{R}_{++}$, $\varrho \in (-1, 1)$, and $(W_t, B_t)_{t\geqslant 0}$ is a 2-dimensional standard Wiener process, see Heston (1993). It is immediately apparent that Y is just the Cox–Ingersoll–Ross process introduced in (4.1). For various interpretations of Y and X in financial mathematics, see, e.g., Hurn et al. (2013, Section 4).

Historically, most efforts have concentrated on parameter estimation for the CIR model only, and not the higher dimension Heston model. For the CIR model and discrete time observations, Theorems 3.1 and 3.3 in Overbeck and Rydén (1997) correspond to our Theorem 5.2, but they estimate the volatility coefficient σ_1 as well, which we assume to be known. Our estimates also offer a plausible way towards change detection, because their form is quite similar to the ones obtained before.

As with the CIR process, we will focus on the subcritical case exclusively, i.e., when $b > 0$. We also do not estimate the parameters σ_1, σ_2 and ϱ, since these parameters could be determined (rather than estimated) using an arbitrarily short continuous time observation X. In any case, it turns out that for the calculation of the estimator of (a, b, α, β), one does not need to know the values of the parameters σ_1, σ_2 and ϱ.

We would like to introduce CLS estimators (CLSE's) for (a, b, α, β) based on discrete time observations. However, this is highly impractical, as the resulting partial derivatives depend on the parameters in a nonlinear manner, so we cannot calculate the estimates analytically. Hence we transform the parameter space, and derive CLSE's for the transformed parameter vector:

$$\begin{bmatrix} c \\ d \\ \gamma \\ \delta \end{bmatrix} = \begin{bmatrix} a \int_0^1 \mathrm{e}^{-bu} \, \mathrm{d}u \\ \mathrm{e}^{-b} \\ \alpha - a\beta \int_0^1 \left(\int_0^u \mathrm{e}^{-bv} \, \mathrm{d}v \right) \mathrm{d}u \\ -\beta \int_0^1 \mathrm{e}^{-bu} \, \mathrm{d}u \end{bmatrix} \tag{5.2}$$

For these parameters the partial derivatives are linear, and our estimates can be established by the standard CLS method and are thus remarkably similar to the estimates obtained in the previous two cases:

(5.3)
$$\begin{bmatrix} \widehat{c}_n \\ \widehat{d}_n \\ \widehat{\gamma}_n \\ \widehat{\delta}_n \end{bmatrix} := \left(\boldsymbol{E}_2 \otimes \left(\sum_{i=1}^n \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix} \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix}^\top \right)^{-1} \right) \sum_{i=1}^n \begin{bmatrix} Y_i \\ X_i - X_{i-1} \end{bmatrix} \otimes \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix}.$$

These estimates possess two very desirable qualities per the following theorem:

5.1 Theorem. *If* $a, b > 0$, $\alpha, \beta \geqslant 0$, $\sigma_1, \sigma_2 > 0$, $\varrho \in (-1, 1)$ *and* $(Y_0, X_0) =$

(y_0, x_0) *with* $y_0 > 0$, $x_0 \in \mathbb{R}$, *then* $(\widehat{c}_n, \widehat{d}_n, \widehat{\gamma}_n, \widehat{\delta}_n)$ *is strongly consistent and asymptotically normal, i.e., as* $n \to \infty$,

$$(\widehat{c}_n, \widehat{d}_n, \widehat{\gamma}_n, \widehat{\delta}_n) \xrightarrow{\text{a.s.}} (c, d, \gamma, \delta) \quad \textit{and} \quad \sqrt{n} \begin{bmatrix} \widehat{c}_n - c \\ \widehat{d}_n - d \\ \widehat{\gamma}_n - \gamma \\ \widehat{\delta}_n - \delta \end{bmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}_4(\mathbf{0}, \boldsymbol{G})$$

with some strictly positive definite matrix $\boldsymbol{G}$.

We apply the inverse of the transformation in (5.2) to estimate the original parameters – this can be done with an asymptotic probability of 1.

$$\begin{bmatrix} \widehat{a}_n \\ \widehat{b}_n \\ \widehat{\alpha}_n \\ \widehat{\beta}_n \end{bmatrix} := \begin{bmatrix} -\widehat{c}_n \frac{\log \widehat{d}_n}{1-\widehat{d}_n} \\ -\log \widehat{d}_n \\ \widehat{\gamma}_n - \widehat{c}_n \widehat{\delta}_n \frac{\widehat{d}_n - 1 - \log \widehat{d}_n}{(1-\widehat{d}_n)^2} \\ \widehat{\delta}_n \frac{\log \widehat{d}_n}{1-\widehat{d}_n} \end{bmatrix}. \tag{5.4}$$

After this transformation, with the help of the so-called delta method, we arrive at the following result.

5.2 Theorem. *Under the conditions of Theorem 5.1,* $(\widehat{a}_n, \widehat{b}_n, \widehat{\alpha}_n, \widehat{\beta}_n)$ *is strongly consistent and asymptotically normal, i.e., as* $n \to \infty$,

$$(\widehat{a}_n, \widehat{b}_n, \widehat{\alpha}_n, \widehat{\beta}_n) \xrightarrow{\text{a.s.}} (a, b, \alpha, \beta) \quad \textit{and} \quad \sqrt{n} \begin{bmatrix} \widehat{a}_n - a \\ \widehat{b}_n - b \\ \widehat{\alpha}_n - \alpha \\ \widehat{\beta}_n - \beta \end{bmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}_4\left(0, \boldsymbol{J}\boldsymbol{G}\boldsymbol{J}^\top\right),$$

where

$$\boldsymbol{J} := \begin{bmatrix} -\frac{\log d}{1-d} & -c\frac{\log d - 1 + d^{-1}}{(1-d)^2} & 0 & 0 \\ 0 & -\frac{1}{d} & 0 & 0 \\ \delta\frac{\log d + 1 - d}{(1-d)^2} & c\delta\frac{2\log d - d + d^{-1}}{(1-d)^3} & 1 & c\frac{\log d + 1 - d}{(1-d)^2} \\ 0 & \delta\frac{\log d - 1 + d^{-1}}{(1-d)^2} & 0 & \frac{\log d}{1-d} \end{bmatrix},$$

the Jacobian of the transformation in (5.4).

References

Alzaid, A. A. and Al-Osh, M. A. (1987). First order integer-valued autoregressive (INAR(1)) processes. *J. Time Series Anal.*, 8:261–275.

Barczy, M., Pap, G., and T. Szabó, T. (2016). Parameter estimation for the subcritical Heston model based on discrete time observations. *Acta Sci. Math.*, 82(1-2):313–338.

Cox, J. C., Ingersoll, J. E., and Ross, S. A. (1985). A theory of the term structure of interest rates. *Econometrica*, 53(2):385–407.

Csörgő, M. and Horváth, L. (1997). *Limit Theorems in Change-point Analysis.* Wiley.

Du, J.-G. and Li, Y. (1991). The integer valued autoregressive (INAR(p)) model. *J. Time Series Anal.*, 12:129–142.

Feller, W. (1951). Two singular diffusion problems. *Ann. of Math.*, 54:173–182.

Gombay, E. (2008). Change detection in autoregressive time series. *J. Multivariate Anal.*, 99:451–464.

Guo, M. and Härdle, W. K. (2017). Adaptive interest rate modelling. *J. Forecast.*, 36(3):241–256.

Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, 6:327–343.

Hurn, A., Lindsay, K., and McClelland, A. (2013). A quasi-maximum likelihood method for estimating the parameters of multivariate diffusions. *J. Econometrics*, 172(1):106–126.

Jacod, J. and Shiryaev, A. N. (2003). *Limit Theorems for Stochastic Processes.* Springer-Verlag.

Kang, J. and Lee, S. (2009). Parameter change test for random coefficient integer-valued autoregressive processes with application to polio data analysis. *J. Time Series Anal.*, 30:239–258.

Karatzas, I. and Shreve, S. E. (1991). *Brownian Motion and Stochastic Calculus*. Springer-Verlag.

Klimko, L. A. and Nelson, P. I. (1978). On conditional least squares estimation for stochastic processes. *Ann. Statist.*, 6:629–642.

Kokoszka, P. and Leipus, R. (1998). Change-point in the mean of dependent observations. *Statist. Probab. Lett.*, 40:385–393.

Overbeck, L. and Rydén, T. (1997). Estimation in the Cox–Ingersoll–Ross model. *Econometric Theory*, 13(3):430–461.

Page, E. (1954). Continuous inspection schemes. *Biometrika*, 41(1-2):100–115.

Page, E. (1955). A test for a change in a parameter occurring at an unknown point. *Biometrika*, 42(3-4):523–527.

Pap, G. and Szabó, T. T. (2013). Change detection in INAR(p) processes against various alternative hypotheses. *Comm. Statist. Theory Methods*, 42(7):1386–1405.

Pap, G. and Szabó, T. T. (2016). Change detection in the Cox–Ingersoll–Ross model. *Statistics & Risk Modeling*, 33(1-2):21–40.

Schmid, W. and Tzotchev, D. (2004). Statistical surveillance of the parameters of a one-factor Cox–Ingersoll–Ross model. *Sequential Anal.*, 23(3):379–412.