CentER

Discussion Paper

No. 2006–44

AN ASYMPTOTIC ANALYSIS OF NEARLY UNSTABLE INAR (1) MODELS

By Feike C. Drost, Ramon van den Akker, Bas J.M. Werker

April 2006

ISSN 0924-7815

Tilburg University

An asymptotic analysis of nearly unstable INAR(1) models

Feike C. Drost*, Ramon van den Akker*, and Bas J.M. Werker*

April 28, 2006

Abstract

This paper considers integer-valued autoregressive processes where the autoregression parameter is close to unity. We consider the asymptotics of this 'near unit root' situation. The local asymptotic structure of the likelihood ratios of the model is obtained, showing that the limit experiment is Poissonian. This Poisson limit experiment is used to construct efficient estimators and tests.

Key words: integer-valued time series, Poisson limit experiment, local-to-unity asymptotics

JEL: C12, C13, C19

1 Introduction

Integer-valued autoregressive processes of the order 1 (INAR(1)) were introduced by Al-Osh and Alzaid (1987) as a nonnegative integer-valued analogue of the familiar AR(1) processes. An INAR(1) process (starting at 0) is defined by the recursion, $X_0 = 0$, and,

$$X_t = \vartheta \circ X_{t-1} + \varepsilon_t, \quad t \in \mathbb{N}, \tag{1}$$

where,

$$\vartheta \circ X_{t-1} = \sum_{j=1}^{X_{t-1}} Z_j^{(t)}.$$

Here $(Z_j^{(t)})_{j\in\mathbb{N},t\in\mathbb{N}}$ is a collection of i.i.d. Bernoulli variables with success probability $\theta \in [0,1]$, independent of the i.i.d. innovation sequence $(\varepsilon_t)_{t\in\mathbb{N}}$ with distribution G on $\mathbb{Z}_+ = \mathbb{N} \cup \{0\}$. All these variables are defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P}_{\theta,G})$. If we work with fixed G, we usually drop the subscript G. The random variable $\vartheta \circ X_{t-1}$ is called the Binomial thinning of X_{t-1} (this operator was introduced by

*Econometrics group, CentER, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands. The authors are grateful to Marc Hallin and Johan Segers for useful discussions.

Steutel and Van Harn (1979)) and, conditionally on X_{t-1}, it follows a Binomial distribution with success probability θ and number of trials equal to X_{t-1}. Equation (1) can be interpreted as a branching process with immigration. The outcome X_t is composed of $\vartheta \circ X_{t-1}$, the elements of X_{t-1} that survive during $(t-1, t]$, and ε_t, the number of immigrants during $(t-1, t]$. Here the number of immigrants is independent of the survival of elements of X_{t-1}. Moreover, each element of X_{t-1} survives with probability θ and its survival has no effect on the survival of the other elements. From a statistical point of view, the difference between the literature on INAR processes and the literature on branching processes with immigration is that in the latter setting one observes both the X process and the ε process, whereas one only observes the X process in the INAR setting, which complicates inference drastically. Compared to the familiar AR(1) processes inference for INAR(1) processes is also more complicated, since, even if θ is known, observing X does not imply observing ε. From the definition of an INAR process it immediately follows that $\mathbb{E}_{\theta,G}\left[X_t \mid X_{t-1}, \dots, X_0\right] = \mathbb{E}_G\varepsilon_1 + \theta X_{t-1}$, which (partially) explains the 'AR' in 'INAR'. It is well-known that, if $\theta \in [0, 1)$ and $\mathbb{E}_G\varepsilon_1 < \infty$, the 'stable' case, there exists an initial distribution, $\nu_{\theta,G}$, such that X is stationary if $\mathcal{L}(X_0) = \nu_{\theta,G}$. Of course, the INAR(1) process is non-stationary if $\theta = 1$: under $\mathbb{P}_1$ the process X is nothing but a standard random walk with drift on $\mathbb{Z}_+$ (but note that X is nondecreasing under $\mathbb{P}_1$). We call this situation 'unstable' or say that the process has a 'unit root'. Although the unit root is on the boundary of the parameter space, it is an important parameter value since in many applications the estimates for θ are close to 1.

Denote the law of $(X_0, \dots, X_n)$ under $\mathbb{P}_{\theta,G}$ on the measurable space $(\mathcal{X}_n, \mathcal{A}_n) = (\mathbb{Z}_+^{n+1}, 2^{\mathbb{Z}_+^{n+1}})$ by $\mathbb{P}_{\theta,G}^{(n)}$. In our applications we mainly consider two sets of assumptions on G: (i) G is known or (ii) G is completely unknown (apart from some regularity conditions). For expository reasons, let us, for the moment, focus on the case that G is completely known and that the goal is to estimate θ. We use 'local-to-unity' asymptotics to take the 'increasing statistical difficulty' in the neighborhood of the unit root into account, i.e. we consider local alternatives to the unit root in such a way that the increasing degree of difficulty to discriminate between these alternatives and the unit root compensates the increase of information contained in the sample as the number of observations grows. This approach is well-known; it originated by the work of Chan and Wei (1987) and Philips (1987), who studied the behavior of a given estimator (OLS) in a nearly unstable AR(1) setting, and Jeganathan (1995), whose results yield the asymptotic structure of nearly unstable AR models. Following this approach, we introduce the sequence of nearly unstable INAR experiments $\mathcal{E}_n(G) = (\mathcal{X}_n, \mathcal{A}_n, (\mathbb{P}_{1-h/n^2}^{(n)} \mid h \geq 0))$, $n \in \mathbb{N}$. The 'localizing rate' n^2 (for the nearly unstable AR(1) model one has rate $n\sqrt{n}$ (non-zero intercept) or n (no intercept)) will become apparent later on. Suppose that we have found an estimator $\widehat{h}_n$ with 'nice properties', then this yields the estimate $1 - \widehat{h}_n/n^2$ of θ in the experiment of interest. To our knowledge Ispány et al. (2003) were the first to study estimation in a nearly unstable INAR(1) model. These authors study the behavior of the OLS estimator and they use a localizing rate n instead of n^2. However, as we will see shortly, n^2 is indeed the proper localizing rate and in Proposition 4.4 we show that the OLS estimator is an exploding estimator in $(\mathcal{E}_n(G))_{n\in\mathbb{N}}$, i.e. it has not even the 'right' rate of convergence. The question then arises how we should estimate h. Instead of analyzing the asymptotic behavior of a given estimator, we derive the asymptotic structure of the experiments themselves by determining the limit experiment (in the Le Cam sense) of $(\mathcal{E}_n(G))_{n\in\mathbb{N}}$. This limit experiment gives bounds to the accuracy of inference procedures and suggests how to construct efficient ones.

The main goal of this paper is to determine the limit experiment of $(\mathcal{E}_n(G))_{n\in\mathbb{N}}$. Remember that (see, for example, Chapter 9 in Van der Vaart (2000) or Van der Vaart (1991)), the sequence of experiments $(\mathcal{E}_n(G))_{n\in\mathbb{N}}$ is said to converge to a limit experiment (in Le Cam's weak topology) $\mathcal{E} = (\mathcal{X}, \mathcal{A}, (\mathbb{Q}_h \mid h \geq 0))$ if, for every finite subset $I \subset \mathbb{R}_+$ and every $h_0 \in \mathbb{R}_+$, we have

$$\left(\frac{\mathrm{d}\mathbb{P}^{(n)}_{1-\frac{h}{n^2}}}{\mathrm{d}\mathbb{P}^{(n)}_{1-\frac{h_0}{n^2}}} \right)_{h\in I} \xrightarrow{d} \left(\frac{\mathrm{d}\mathbb{Q}_h}{\mathrm{d}\mathbb{Q}_{h_0}} \right)_{h\in I}, \text{ under } \mathbb{P}^{(n)}_{1-\frac{h_0}{n^2}}.$$

To see that it is indeed reasonable to expect n^2 as the proper localizing rate, we briefly discuss the case of Geometrically distributed innovations (in the remainder we treat general G). In case $G = \text{Geometric}(1/2)$, i.e., G puts mass $(1/2)^{k+1}$ at $k \in \mathbb{Z}_+$, it is an easy exercise to verify for $h > 0$ (the Geometric distribution allows us, using Newton's Binomial formula, to obtain explicit expressions for the transition-probabilities from X_{t-1} to X_t if $X_t \geq X_{t-1}$),

$$\frac{\mathrm{d}\mathbb{P}^{(n)}_{1-\frac{h}{r_n}}}{\mathrm{d}\mathbb{P}^{(n)}_1} \xrightarrow{p} \begin{cases} 0 & \text{if } \frac{r_n}{n^2} \to 0, \\ \exp\left(-\frac{hG(0)\mathbb{E}_G\varepsilon_1}{2}\right) & \text{if } \frac{r_n}{n^2} \to 1, \\ 1 & \text{if } \frac{r_n}{n^2} \to \infty, \end{cases} \quad \text{under } \mathbb{P}_1.$$

This has two important implications. First, it indicates that n^2 is indeed the proper localizing rate. Intuitively, if we go faster than n^2 we cannot distinguish $\mathbb{P}^{(n)}_{1-h/r_n}$ from $\mathbb{P}^{(n)}_1$, and if we go slower we can distinguish $\mathbb{P}^{(n)}_{1-h/r_n}$ perfectly from $\mathbb{P}^{(n)}_1$. Secondly, since $\exp(-hG(0)\mathbb{E}_G\varepsilon_1/2) < 1$ we cannot, by Le Cam's first lemma, hope, in general, for contiguity of $\mathbb{P}^{(n)}_{1-h/n^2}$ with respect to $\mathbb{P}^{(n)}_1$ (Remark 2 after Theorem 2.1 gives an example of sets that yield this non-contiguity). This lack of contiguity is unfortunate for several reasons. Most importantly, if we would have contiguity the limiting behavior of $(\mathrm{d}\mathbb{P}^{(n)}_{1-h/n^2}/\,\mathrm{d}\mathbb{P}^{(n)}_1)_{h\in I}$ determines the limit experiment, whereas we now need to consider the behavior of $(\mathrm{d}\mathbb{P}^{(n)}_{1-h/n^2}/\,\mathrm{d}\mathbb{P}^{(n)}_{1-h_0/n^2})_{h\in I}$ for all $h_0 \geq 0$ (so to be clear: the preceding display does not yet yield the limit experiment for the Geometric(1/2) case). And it implies that the global sequence of experiments has not the common LAQ structure (see, for example, Definition 1 in Jeganathan (1995)) at $\theta = 1$. The traditional AR(1) process $Y_0 = 0$, $Y_t = \mu + \theta Y_{t-1} + u_t$, u_t i.i.d. $\mathrm{N}(0, \sigma^2)$, with $\mu \neq 0$ and σ^2 known, does enjoy this LAQ property at $\theta = 1$, the limit experiment at $\theta = 1$ is the usual normal location experiment (i.e., the model has the

LAN property) and the localizing rate is $n^{3/2}$. The limit experiment at $\theta = 1$ for $Y_0 = 0$, $Y_t = \theta Y_{t-1} + u_t$, u_t i.i.d. $\mathrm{N}(0, \sigma^2)$, with σ^2 known, does not have the LAN-structure; the limit experiment is of the LABF type and the localizing rate is n. Thus although the INAR(1) process behaves the same as the traditional AR(1) process at $\theta = 1$, their statistical properties 'near $\theta = 1$' are very different. In Section 3 we prove that the limit-experiment of $(\mathcal{E}_n(G))_{n\in\mathbb{N}}$ corresponds to one draw from a Poisson distribution with mean $hG(0)\mathbb{E}_G\varepsilon_1/2$, i.e.

$$\mathcal{E}(G) = \left(\mathbb{Z}_+, 2^{\mathbb{Z}_+}, \left(\text{Poisson}\left(\frac{hG(0)\mathbb{E}_G\varepsilon_1}{2}\right) \mid h \geq 0\right)\right).$$

We indeed recognize $\exp\left(-hG(0)\mathbb{E}_G\varepsilon_1/2\right)$ as the likelihood ratio at h relative to $h_0 = 0$ in the experiment $\mathcal{E}(G)$. Due to the lack of enough smoothness of the likelihood ratios around the unit root, this convergence of experiments is not obtained by the usual (general applicable) techniques, but by a direct approach. Since the transition probability is the convolution of a Binomial distribution with G and the fact that certain Binomial experiments converge to a Poisson limit experiment (see Remark 4 after Theorem 3.1 for the precise statement), one might be tempted to think that the convergence $\mathcal{E}_n(G) \to \mathcal{E}(G)$ follows, in some way, from this convergence. Remark 4 after Theorem 3.1 shows that this reasoning is not valid.

The rest of the paper is organized as follows. In Section 2 we discuss some preliminary properties which provide insight in the behavior of a nearly unstable INAR(1) process and are needed in the rest of the paper. The main result is stated and proved in Section 3. Section 4 uses our main result to analyze some estimation and testing problems. In Section 4.1 we consider efficient estimation of h, the deviation from a unit root, in the nearly unstable case for two settings. The first setting, discussed in Section 4.1.1, treats the case that the immigration distribution G is completely known. And the second setting, analyzed in 4.1.2, considers a semiparametric model, where hardly any conditions on G are imposed. Furthermore, we show in Section 4.1.1 that the OLS-estimator is explosive. In Section 4.2 we provide an efficient estimator of θ in the 'global' model. Finally, we discuss testing for a unit root in Section 4.3. We show that the traditional Dickey-Fuller test has no (local) power, but that an intuitively obvious test is efficient.

2 Preliminaries

This section discusses some basic properties of nearly unstable INAR(1) processes. Besides giving insight in the behavior of a nearly unstable INAR(1) process, these properties are a key input in the next sections.

First, we introduce the following notation. The mean of ε_t is denoted by μ_G and its variance by σ_G^2. The probability mass function corresponding to G, the distribution of the innovations ε_t, is denoted by g. The probability mass function of the Binomial distribution with parameters $\theta \in [0, 1]$ and $n \in \mathbb{Z}_+$ is denoted by $\mathrm{b}_{n,\theta}$.

Given $X_{t-1} = x_{t-1}$, the random variables ε_t and $\vartheta \circ X_{t-1}$ are independent and $X_{t-1} - \vartheta \circ X_{t-1}$, 'the number of deaths during $(t-1, t]$', follows a Binomial$(X_{t-1}, 1-\theta)$ distribution. This interpretation yields the following representation of the transition probabilities, for $x_{t-1}, x_t \in \mathbb{Z}_+$,

$$\begin{aligned} P^{\theta}_{x_{t-1},x_t} &= \mathbb{P}_\theta \left\{ X_t = x_t \mid X_{t-1} = x_{t-1} \right\} = \sum_{k=0}^{x_{t-1}} \mathbb{P}_\theta \left\{ \varepsilon_t = x_t - x_{t-1} + k, X_{t-1} - \vartheta \circ X_{t-1} = k \mid X_{t-1} = x_{t-1} \right\} \\ &= \sum_{k=0}^{x_{t-1}} \mathrm{b}_{x_{t-1},1-\theta}(k) g(\Delta x_t + k), \end{aligned}$$

where $\Delta x_t = x_t - x_{t-1}$, and $g(i) = 0$ for $i < 0$. Under $\mathbb{P}_1$ we have $X_t = \mu_G t + \sum_{i=1}^{t} (\varepsilon_i - \mu_G)$, and $P^1_{x_{t-1},x_t} = g(\Delta x_t)$, $x_{t-1}, x_t \in \mathbb{Z}_+$. Hence, under $\mathbb{P}_1$, an INAR(1) process is nothing but a random walk with drift.

The next proposition is basic, but often applied in the sequel.

Proposition 2.1 *If $\sigma_G^2 < \infty$, we have, for $h \geq 0$,*

$$\lim_{n\to\infty} \mathbb{E}_{1-\frac{h}{n^2}} \left[\frac{1}{n^2} \sum_{t=1}^{n} X_t - \frac{\mu_G}{2} \right]^2 = 0. \tag{2}$$

If $\sigma_G^2 < \infty$, then we have, for $\alpha > 0$ and every sequence $(\theta_n)_{n\in\mathbb{N}}$ in $[0,1]$,

$$\lim_{n\to\infty} \frac{1}{n^{3+\alpha}} \sum_{t=1}^{n} \mathbb{E}_{\theta_n} X_t^2 = 0. \tag{3}$$

PROOF:
We obviously have, $\mathrm{Var}_1 \left(\sum_{t=1}^n X_t \right) = O(n^3)$ and $\lim_{n\to\infty} n^{-2} \sum_{t=1}^n \mathbb{E}_1 X_t = \mu_G/2$, which yields (2) for $h = 0$. Next, we prove (2) for $h > 0$. Straightforward calculations show, for $\theta < 1$,

$$\mathbb{E}_\theta \sum_{t=1}^{n} X_t = \mu_G \sum_{t=1}^{n} \frac{1-\theta^t}{1-\theta} = \mu_G \left[\frac{n}{1-\theta} - \frac{\theta - \theta^{n+1}}{(1-\theta)^2} \right],$$

which yields

$$\lim_{n\to\infty} \frac{1}{n^2} \mathbb{E}_{1-\frac{h}{n^2}} \sum_{t=1}^{n} X_t = \lim_{n\to\infty} \frac{\mu_G}{n^2} \left[\frac{n}{h/n^2} - \frac{1 - \frac{h}{n^2} - \left[1 - (n+1)\frac{h}{n^2} + \frac{(n+1)n}{2}\frac{h^2}{n^4} + o\left(\frac{1}{n^2}\right) \right]}{h^2/n^4} \right] = \frac{\mu_G}{2}. \tag{4}$$

To treat the variance of $n^{-2}\sum_{t=1}^n X_t$, we use the following simple relations, see also Ispány et al. (2003), for $0<\theta<1$, $s,t\geq 1$,

$$\begin{aligned}\operatorname{Cov}_\theta(X_t,X_s) &= \theta^{|t-s|}\operatorname{Var}_\theta X_{t\wedge s},\\ \operatorname{Var}_\theta X_t &= \frac{1-\theta^{2t}}{1-\theta^2}\sigma_G^2+\frac{(\theta-\theta^t)(1-\theta^t)}{1-\theta^2}\mu_G\leq(\sigma_G^2+\mu_G)\frac{1-\theta^{2t}}{1-\theta^2}.\end{aligned} \tag{5}$$

From this we obtain

$$\begin{aligned}\operatorname{Var}_{1-\frac{h}{n^2}}\left(\frac{1}{n^2}\sum_{t=1}^n X_t\right) &= \frac{1}{n^4}\sum_{t=1}^n\left(1+2\sum_{s=t+1}^n\left(1-\frac{h}{n^2}\right)^{s-t}\right)\operatorname{Var}_{1-\frac{h}{n^2}}X_t\\ &\leq \frac{1}{n}2n(\sigma_G^2+\mu_G)\frac{1}{n^2}\frac{1}{1-\left(1-\frac{h}{n^2}\right)^2}\frac{1}{n}\sum_{t=1}^n\left(1-\left(1-\frac{h}{n^2}\right)^{2t}\right)\to 0,\end{aligned}$$

as $n\to\infty$. Together with (4) this completes the proof of (2) for $h>0$.
To prove (3), note that $X_t\leq\sum_{i=1}^t\varepsilon_i$. Hence $\mathbb{E}_{\theta_n}X_t^2\leq\mathbb{E}_1X_t^2=\sigma_G^2t+\mu_G^2t^2$, which yields the desired conclusion. □

Remark 1 *Convergence in probability for the case $h>0$ in (2) cannot be concluded from the convergence in probability in (2) for $h=0$ by contiguity arguments. The reason is (see Remark 2 after the proof of Theorem 2.1) that $\mathbb{P}_{1-h/n^2}^{(n)}$ is not contiguous with respect to $\mathbb{P}_1^{(n)}$.*

Next, we consider the thinning process $(\vartheta\circ X_{t-1})_{t\geq 1}$. Under $\mathbb{P}_{1-h/n^2}$, $X_{t-1}-\vartheta\circ X_{t-1}$, conditional on X_{t-1}, follows a Binomial$(X_{t-1},h/n^2)$ distribution. So we expect that there do not occur many 'deaths' in any time-interval $(t-1,t]$. The following proposition gives a precise statement, where we use the notation, for $h\geq 0$ and $n\in\mathbb{N}$,

$$A_n^h=\left\{z\in\mathbb{Z}_+\;\middle|\;\frac{h(z+1)}{n^2}<\frac{1}{2}\right\},\quad \mathcal{A}_n^h=\{(X_0,\ldots,X_{n-1})\in A_n^h\times\cdots\times A_n^h\}. \tag{6}$$

The reasons for the introduction of these sets are the following. By Proposition A.1 we have for $x\in A_n^h$ $\sum_{k=r}^x \mathrm{b}_{x,h/n^2}(k)\leq 2\,\mathrm{b}_{x,h/n^2}(r)$ for $r=2,3$ and terms of the form $(1-\frac{h}{n^2})^{-2}$ can be bounded neatly, without having to make statements of the form 'for n large enough', or having to refer to 'upto a constant depending on h'. Furthermore, recall the notation $\Delta X_t=X_t-X_{t-1}$.

Proposition 2.2 *Assume G satisfies $\sigma_G^2<\infty$. Then we have for all sequences $(\theta_n)_{n\in\mathbb{N}}$ in $[0,1]$, $h\geq 0$,*

$$\lim_{n\to\infty}\mathbb{P}_{\theta_n}(\mathcal{A}_n^h)=1. \tag{7}$$

For $h \geq 0$ we have,

$$\lim_{n\to\infty} \mathbb{P}_{1-\frac{h}{n^2}} \{\exists t \in \{1,\ldots,n\} : X_{t-1} - \vartheta \circ X_{t-1} \geq 2\} = 0. \tag{8}$$

PROOF:
For a sequence $(\theta_n)_{n\in\mathbb{N}}$ in $[0,1]$, (3) implies

$$\mathbb{P}_{\theta_n} \left\{\exists\, 0 \leq t \leq n : X_t > n^{7/4}\right\} \leq \frac{1}{n^{7/2}} \sum_{t=1}^{n} \mathbb{E}_{\theta_n} X_t^2 \to 0 \text{ as } n \to \infty. \tag{9}$$

From this we easily obtain (7).
To obtain (8) note that, for $X_{t-1} \in A_n^h$ we have, using the bound (43),

$$\mathbb{P}_{1-\frac{h}{n^2}} \{X_{t-1} - \vartheta \circ X_{t-1} \geq 2 \mid X_{t-1}\} = \sum_{k=2}^{X_{t-1}} \mathrm{b}_{X_{t-1},\frac{h}{n^2}}(k) \leq 2\,\mathrm{b}_{X_{t-1},\frac{h}{n^2}}(2) \leq \frac{h^2 X_{t-1}^2}{n^4}.$$

By (3) this yields,

$$\lim_{n\to\infty} \mathbb{P}_{1-\frac{h}{n^2}} \left(\{\exists t \in \{1,\ldots,n\} : X_{t-1} - \vartheta \circ X_{t-1} \geq 2\} \cap \mathcal{A}_n^h\right) \leq \lim_{n\to\infty} \frac{h^2}{n^4} \sum_{t=1}^{n} \mathbb{E}_{1-\frac{h}{n^2}} X_{t-1}^2 = 0.$$

Since we already showed $\lim_{n\to\infty} \mathbb{P}_{1-h/n^2}(\mathcal{A}_n^h) = 1$, this yields (8). □

Finally, we derive the limit distribution of the number of downward movements of X during $[0,n]$. The probability that the Bernoulli variable $1\{\Delta X_t < 0\}$ equals one is small. Intuitively the dependence over time of this indicator-process is not too strong, so it is not unreasonable to expect that a 'Poisson law of small numbers' holds. As the following theorem shows, this is indeed the case.

Theorem 2.1 *Assume that G satisfies $\sigma_G^2 < \infty$. Then we have, for $h \geq 0$,*

$$\sum_{t=1}^{n} 1\{\Delta X_t < 0\} \xrightarrow{d} \text{Poisson}\left(\frac{hg(0)\mu_G}{2}\right), \text{ under } \mathbb{P}_{1-\frac{h}{n^2}}. \tag{10}$$

PROOF:
If $g(0) = 0$ then $\Delta X_t < 0$ implies $X_{t-1} - \vartheta \circ X_{t-1} \geq 2$. Hence, (8) implies $\sum_{t=1}^{n} 1\{\Delta X_t < 0\} \xrightarrow{p} 0$ under $\mathbb{P}_{1-h/n^2}$. Since the Poisson distribution with mean 0 concentrates all its mass at 0, this yields the result. The cases $h = 0$ or $g(0) = 1$ (since $X_0 = 0$) are also trivial.
So we consider the case $h > 0$ and $0 < g(0) < 1$. For notational convenience, abbreviate $\mathbb{P}_{1-h/n^2}$ by $\mathbb{P}_n$

and $\mathbb{E}_{1-h/n^2}$ by $\mathbb{E}_n$. Put $Z_t = 1\{\Delta X_t = -1, \varepsilon_t = 0\}$. From (8) it follows that

$$\sum_{t=1}^n 1\{\Delta X_t < 0\} - \sum_{t=1}^n Z_t = \sum_{t=1}^n (1\{\Delta X_t \leq -2\} + 1\{\Delta X_t = -1, \varepsilon_t \geq 1\}) \xrightarrow{p} 0, \text{ under } \mathbb{P}_n.$$

Thus it suffices to prove that $\sum_{t=1}^n Z_t \xrightarrow{d} \text{Poisson}(hg(0)\mu_G/2)$ under $\mathbb{P}_n$. We do this by applying Lemma A.1. Introduce random variables Y_n, where Y_n follows a Poisson distribution with mean $\lambda_n = \sum_{t=1}^n \mathbb{E}_n Z_t$. And let Z follow a Poisson distribution with mean $hg(0)\mu_G/2$. From Lemma A.1 we obtain the bound

$$\sup_{A \subset \mathbb{Z}_+} \left| \mathbb{P}_n \left\{ \sum_{t=1}^n Z_t \in A \right\} - \Pr\{Y_n \in A\} \right| \leq \sum_{t=1}^n (\mathbb{E}_n Z_t)^2 + \sum_{t=1}^n \mathbb{E}_n \left| \mathbb{E}_n \left[Z_t - \mathbb{E}_n Z_t \mid Z_1, \ldots, Z_{t-1} \right] \right|.$$

If we prove that

$$\text{(i) } \sum_{t=1}^n (\mathbb{E}_n Z_t)^2 \to 0, \text{ (ii) } \sum_{t=1}^n \mathbb{E}_n Z_t \to \frac{hg(0)\mu_G}{2}, \text{ (iii) } \sum_{t=1}^n \mathbb{E}_n \left| \mathbb{E}_n \left[Z_t - \mathbb{E}_n Z_t \mid Z_1, \ldots, Z_{t-1} \right] \right| \to 0,$$

all hold as $n \to \infty$, then the result follows since we then have, for all $k \in \mathbb{Z}_+$,

$$\left| \mathbb{P}_n \left\{ \sum_{t=1}^n Z_t = k \right\} - \Pr(Z = k) \right| \leq \left| \mathbb{P}_n \left\{ \sum_{t=1}^n Z_t = k \right\} - \Pr\{Y_n = k\} \right| + \left| \Pr\{Y_n = k\} - \Pr(Z = k) \right| \to 0.$$

First we tackle (i). Notice that, condition on X_{t-1},

$$\mathbb{E}_n Z_t = \frac{hg(0)}{n^2} \mathbb{E}_n X_{t-1} \left(1 - \frac{h}{n^2}\right)^{X_{t-1}-1} \leq \frac{hg(0)}{n^2} \mathbb{E}_n X_{t-1}.$$

Then (i) is easily obtained using (3),

$$\lim_{n\to\infty} \sum_{t=1}^n (\mathbb{E}_n Z_t)^2 \leq \lim_{n\to\infty} \frac{h^2 g^2(0)}{n^4} \sum_{t=1}^n \mathbb{E}_n X_{t-1}^2 = 0.$$

Next we consider (ii). If we prove the relation,

$$\lim_{n\to\infty} \left| \frac{1}{n^2} \sum_{t=1}^n \mathbb{E}_n X_{t-1} - \frac{1}{n^2} \sum_{t=1}^n \mathbb{E}_n X_{t-1} \left(1 - \frac{h}{n^2}\right)^{X_{t-1}-1} \right| = 0,$$

it is immediate that (ii) follows from (2). To prove the previous display, we introduce $B_n = \{\forall t \in \{1, \ldots, n\} : X_t \leq n^{7/4}\}$ with $\lim_{n\to\infty} \mathbb{P}_n(B_n) = 1$ (see (9)). On the event B_n we have $n^{-2} X_t \leq n^{-1/4}$ for

$t = 1, \dots, n$. This yields

$$0 \le \mathbb{E}_n X_{t-1}\left(1-\left(1-\frac{h}{n^2}\right)^{X_{t-1}-1}\right) \le \mathbb{E}_n X_{t-1}\left(1-\left(1-\frac{h}{n^2}\right)^{X_{t-1}}\right)1_{B_n} + \mathbb{E}_n X_{t-1}1_{B_n^c}$$
$$\le \mathbb{E}_n\left[1_{B_n}X_{t-1}\sum_{j=1}^{X_{t-1}}\binom{X_{t-1}}{j}\left(\frac{h}{n^2}\right)^j\right] + \mathbb{E}_n X_{t-1}1_{B_n^c} \le \frac{1}{n^{1/4}}\exp(h)\mathbb{E}_n X_{t-1} + \mathbb{E}_n X_{t-1}1_{B_n^c}.$$

Using $\mathbb{P}_n(B_n) \to 1$ and (2) we obtain,

$$\lim_{n\to\infty}\frac{1}{n^2}\sum_{t=1}^{n}\mathbb{E}_n X_{t-1}1_{B_n^c} \le \lim_{n\to\infty}\sqrt{\mathbb{E}_n\left(\frac{1}{n^2}\sum_{t=1}^{n}X_{t-1}\right)^2\mathbb{P}_n(B_n^c)} = \sqrt{\left(\frac{\mu_G}{2}\right)^2\cdot 0} = 0.$$

By (3) we have $\lim_{n\to\infty} n^{-9/4}\sum_{t=1}^{n}\mathbb{E}_n X_{t-1} = 0$. Combination with the previous two displays yields the result.

Finally, we prove (iii). Let $\mathcal{F}^\varepsilon = (\mathcal{F}_t^\varepsilon)_{t\ge1}$ and $\mathcal{F}^X = (\mathcal{F}_t^X)_{t\ge0}$ be the filtrations generated by $(\varepsilon_t)_{t\ge1}$ and $(X_t)_{t\ge0}$ respectively. Note that we have, for $t \ge 2$,

$$\mathbb{E}_n\left|\mathbb{E}_n\left[Z_t - \mathbb{E}_n Z_t \mid Z_1,\dots,Z_{t-1}\right]\right| \le \mathbb{E}_n\left|\mathbb{E}_n\left[Z_t - \mathbb{E}_n Z_t \mid \mathcal{F}_{t-1}^\varepsilon, \mathcal{F}_{t-1}^X\right]\right|$$
$$= \frac{hg(0)}{n^2}\mathbb{E}_n\left|X_{t-1}\left(1-\frac{h}{n^2}\right)^{X_{t-1}-1} - \mathbb{E}_n X_{t-1}\left(1-\frac{h}{n^2}\right)^{X_{t-1}-1}\right|. \quad (11)$$

Using the reverse triangle-inequality we obtain

$$\left|\mathbb{E}_n\left|X_{t-1}\left(1-\frac{h}{n^2}\right)^{X_{t-1}-1} - \mathbb{E}_n X_{t-1}\left(1-\frac{h}{n^2}\right)^{X_{t-1}-1}\right| - \mathbb{E}_n\left|X_{t-1}-\mathbb{E}_n X_{t-1}\right|\right|$$
$$\le \mathbb{E}_n\left|X_{t-1}\left(1-\left(1-\frac{h}{n^2}\right)^{X_{t-1}-1}\right) - \mathbb{E}_n X_{t-1}\left(1-\left(1-\frac{h}{n^2}\right)^{X_{t-1}-1}\right)\right|$$
$$\le 2\mathbb{E}_n X_{t-1}\left(1-\left(1-\frac{h}{n^2}\right)^{X_{t-1}-1}\right).$$

We have already seen in the proof of (ii) that

$$\lim_{n\to\infty}\frac{1}{n^2}\sum_{t=1}^{n}\mathbb{E}_n X_{t-1}\left(1-\left(1-\frac{h}{n^2}\right)^{X_{t-1}-1}\right) = 0$$

holds. A combination of the previous two displays with (11) now easily yields the bound

$$\sum_{t=1}^{n} \mathbb{E}_n \left| \mathbb{E}_n \left[Z_t - \mathbb{E}_n Z_t \mid Z_1, \ldots, Z_{t-1} \right] \right| \leq o(1) + \frac{hg(0)}{n^2} \sum_{t=1}^{n} \sqrt{\mathrm{Var}_n X_{t-1}}. \tag{12}$$

From (5) we obtain, for $\theta < 1$, $\mathrm{Var}_\theta X_t \leq \left(\sigma_G^2 + \mu_G\right)(1-\theta^{2t})(1-\theta^2)^{-1}$. And for $1 \leq t \leq n$ we have $0 \leq 1 - \left(1 - h/n^2\right)^{2t} \leq n^{-1}\exp(2h)$. Now we easily obtain

$$\frac{1}{n^2}\sum_{t=1}^{n} \sqrt{\mathrm{Var}_n X_{t-1}} \leq \sqrt{\sigma_G^2 + \mu_G} \sqrt{\frac{1}{n^2} \frac{1}{1 - \left(1 - \frac{h}{n^2}\right)^2} \frac{1}{n} n \sqrt{\frac{\exp(2h)}{n}}} \to 0 \text{ as } n \to \infty.$$

A combination with (12) yields (iii). This concludes the proof. □

Remark 2 *Since* $\sum_{t=1}^{n} 1\{\Delta X_t < 0\}$ *equals zero under* $\mathbb{P}_1^{(n)}$ *and converges in distribution to a non-degenerated limit under* $\mathbb{P}_{1-h/n^2}^{(n)}$ *(*$h > 0$, $0 < g(0) < 1$*), we see that* $\mathbb{P}_{1-h/n^2}^{(n)}$ *is not contiguous with respect to* $\mathbb{P}_1^{(n)}$ *for* $h > 0$.

3 The limit experiment: one observation from a Poisson distribution

For easy reference, we introduce the following assumption.

Assumption 3.1 *A probability distribution G on $\mathbb{Z}_+$ is said to satisfy Assumption 3.1 if one of the following two condition holds.*

(1) $\mathrm{support}(G) = \{0, \ldots, M\}$ *for some* $M \in \mathbb{N}$;

(2) $\mathrm{support}(G) = \mathbb{Z}_+$, $\sigma_G^2 < \infty$ *and g is eventually decreasing, i.e. there exists $M \in \mathbb{N}$ such that* $g(k+1) \leq g(k)$ *for* $k \geq M$.

The rest of this section is devoted to the following theorem.

Theorem 3.1 *Suppose G satisfies Assumption 3.1. Then the limit experiment of* $(\mathcal{E}_n(G))_{n\in\mathbb{N}}$ *is given by*

$$\mathcal{E}(G) = \left(\mathbb{Z}_+, 2^{\mathbb{Z}_+}, (\mathbb{Q}_h \mid h \geq 0)\right),$$

with $\mathbb{Q}_h = \mathrm{Poisson}\left(hg(0)\mu_G/2\right)$.

Notice that the likelihood-ratios for this Poisson limit experiment are given by,

$$\frac{\mathrm{d}\mathbb{Q}_h}{\mathrm{d}\mathbb{Q}_{h_0}}(Z) = \exp\left(-\frac{(h-h_0)g(0)\mu_G}{2}\right)\left(\frac{h}{h_0}\right)^Z, \tag{13}$$

for $h \geq 0$, $h_0 > 0$ and,

$$\frac{\mathrm{d}\mathbb{Q}_h}{\mathrm{d}\mathbb{Q}_0}(Z) = \exp\left(-\frac{hg(0)\mu_G}{2}\right)1_{\{0\}}(Z), \tag{14}$$

for $h \geq 0$.

PROOF:
To determine the limit-experiment we need to determine the limit-distribution of the log-likelihood ratios, $h, h_0 \geq 0$,

$$\mathcal{L}_n(h,h_0) = \log \frac{\mathrm{d}\mathbb{P}^{(n)}_{1-\frac{h}{n^2}}}{\mathrm{d}\mathbb{P}^{(n)}_{1-\frac{h_0}{n^2}}} = \sum_{t=1}^{n} \log \frac{P^{1-\frac{h}{n^2}}_{X_{t-1},X_t}}{P^{1-\frac{h_0}{n^2}}_{X_{t-1},X_t}},$$

under $\mathbb{P}_{1-h_0/n^2}$. Notice that for $h_0 > 0$ $\mathcal{L}_n(0,h_0) = -\infty$, and thus $\mathrm{d}\mathbb{P}^{(n)}_0 / \mathrm{d}\mathbb{P}^{(n)}_{1-h_0/n^2} = 0$, if $\sum_{t=1}^n 1\{\Delta X_t < 0\} > 0$. Because $\mathcal{L}_n(h,h_0)$ is complicated to analyze, we make an approximation of this object. Split the transition-probability $P^{1-h/n^2}_{x_{t-1},x_t}$ into a leading term,

$$L_n(x_{t-1},x_t,h) = \begin{cases} \sum_{k=-\Delta x_t}^{-\Delta x_t+1} \mathrm{b}_{x_{t-1},\frac{h}{n^2}}(k)g(\Delta x_t + k) & \text{if } \Delta x_t < 0, \\ \sum_{k=0}^{1} \mathrm{b}_{x_{t-1},\frac{h}{n^2}}(k)g(\Delta x_t + k) & \text{if } \Delta x_t \geq 0, \end{cases}$$

and a remainder term,

$$R_n(x_{t-1},x_t,h) = \begin{cases} \sum_{k=-\Delta x_t+2}^{x_{t-1}} \mathrm{b}_{x_{t-1},\frac{h}{n^2}}(k)g(\Delta x_t + k) & \text{if } \Delta x_t < 0, \\ \sum_{k=2}^{x_{t-1}} \mathrm{b}_{x_{t-1},\frac{h}{n^2}}(k)g(\Delta x_t + k) & \text{if } \Delta x_t \geq 0. \end{cases}$$

We introduce a simpler version of $\mathcal{L}_n(h,h_0)$ in which the remainder terms are removed,

$$\tilde{\mathcal{L}}_n(h,h_0) = \sum_{t=1}^{n} \log \frac{L_n(X_{t-1},X_t,h)}{L_n(X_{t-1},X_t,h_0)}.$$

The difference between $\tilde{\mathcal{L}}_n(h,h_0)$ and $\mathcal{L}_n(h,h_0)$ is negligible. To enhance readability we organize this result and its proof in a lemma.

Lemma 3.1 *We have, for $h, h_0 \geq 0$,*

$$\tilde{\mathcal{L}}_n(h,h_0) = \mathcal{L}_n(h,h_0) + o(\mathbb{P}_{1-\frac{h_0}{n^2}};1). \tag{15}$$

Proof:
We obtain, for $h > 0, h_0 \geq 0$, using the inequality $|\log((a+b)/(c+d)) - \log(a/c)| \leq b/a + d/c$ for $a, c > 0$, $b, d \geq 0$, the bound

$$\left|\mathcal{L}_n(h, h_0) - \tilde{\mathcal{L}}_n(h, h_0)\right| \leq \sum_{t=1}^{n} \frac{R_n(X_{t-1}, X_t, h)}{L_n(X_{t-1}, X_t, h)} + \sum_{t=1}^{n} \frac{R_n(X_{t-1}, X_t, h_0)}{L_n(X_{t-1}, X_t, h_0)} \quad \mathbb{P}_{1-\frac{h_0}{n^2}} - \text{a.s.} \tag{16}$$

It is easy to see that, for $h_0 > 0$, $\mathcal{L}_n(0, h_0)$ and $\tilde{\mathcal{L}}_n(0, h_0)$ both equal minus infinity if $\sum_{t=1}^{n} 1\{\Delta X_t < 0\} \geq 1$, and for $\sum_{t=1}^{n} 1\{\Delta X_t < 0\} = 0$ we have

$$\left|\mathcal{L}_n(0, h_0) - \tilde{\mathcal{L}}_n(0, h_0)\right| \leq \sum_{t=1}^{n} \frac{R_n(X_{t-1}, X_t, h_0)}{L_n(X_{t-1}, X_t, h_0)} \quad \mathbb{P}_{1-\frac{h_0}{n^2}} - \text{a.s.}$$

Thus if we show that

$$\sum_{t=1}^{n} \frac{R_n(X_{t-1}, X_t, h')}{L_n(X_{t-1}, X_t, h')} \xrightarrow{p} 0, \text{ under } \mathbb{P}_{1-\frac{h_0}{n^2}},$$

holds for $h' = h$ and $h' = h_0$ the lemma is proved (exclude the case $h' = 0$ and $h_0 > 0$, which need not be considered). We split the expression in the previous display into four nonnegative parts

$$\begin{aligned}\sum_{t=1}^{n} \frac{R_n(X_{t-1}, X_t, h')}{L_n(X_{t-1}, X_t, h')} &= \sum_{t:\, \Delta X_t \leq -2} \frac{R_n(X_{t-1}, X_t, h')}{L_n(X_{t-1}, X_t, h')} + \sum_{t:\, \Delta X_t = -1} \frac{R_n(X_{t-1}, X_t, h')}{L_n(X_{t-1}, X_t, h')} \\ &+ \sum_{t:\, 0 \leq \Delta X_t \leq M} \frac{R_n(X_{t-1}, X_t, h')}{L_n(X_{t-1}, X_t, h')} + \sum_{t:\, \Delta X_t > M} \frac{R_n(X_{t-1}, X_t, h')}{L_n(X_{t-1}, X_t, h')}.\end{aligned}$$

Since $\Delta X_t \leq -2$ implies $X_{t-1} - \vartheta \circ X_{t-1} \geq 2$ (8) implies

$$\sum_{t:\, \Delta X_t \leq -2} \frac{R_n(X_{t-1}, X_t, h')}{L_n(X_{t-1}, X_t, h')} \xrightarrow{p} 0, \text{ under } \mathbb{P}_{1-\frac{h_0}{n^2}}.$$

Next we treat the terms for which $\Delta X_t = -1$. If $h_0 = 0$ we do not have such terms (under $\mathbb{P}_{1-h_0/n^2}$), and remember that the case $h' = 0$ and $h_0 > 0$ need not be considered. So we only need to consider this term for $h', h_0 > 0$. On the event $\mathcal{A}_n^{h'}$ (see (6) for the definition of this event), an application of (43) yields,

$$\begin{aligned}\sum_{t:\Delta X_t = -1} \frac{R_n(X_{t-1}, X_t, h')}{L_n(X_{t-1}, X_t, h')} &\leq \sum_{t:\Delta X_t = -1} \frac{\sum_{k=3}^{X_{t-1}} \mathrm{b}_{X_{t-1}, \frac{h'}{n^2}}(k)}{g(0)\, \mathrm{b}_{X_{t-1}, \frac{h'}{n^2}}(1)} \leq 2 \sum_{t=1}^{n} \frac{\frac{X_{t-1}^3}{3!} \frac{h'^3}{n^6} \left(1 - \frac{h'}{n^2}\right)^{X_{t-1}-3}}{g(0) X_{t-1} \frac{h'}{n^2} \left(1 - \frac{h'}{n^2}\right)^{X_{t-1}-1}} \\ &\leq \frac{4h'^2}{3g(0)n^4} \sum_{t=1}^{n} X_{t-1}^2,\end{aligned}$$

since $(1-h'/n^2)^{-2} \leq 4$ by definition of $A_n^{h'}$ (see (6) for the definition of this set). From (3) and (7) it now easily follows that we have

$$\sum_{t:\Delta X_t=-1} \frac{R_n(X_{t-1},X_t,h')}{L_n(X_{t-1},X_t,h')} \xrightarrow{p} 0, \text{ under } \mathbb{P}_{1-\frac{h_0}{n^2}}.$$

Next, we analyze the terms for which $0 \leq \Delta X_t \leq M$. We have, by (43), on the event $\mathcal{A}_n^{h'}$,

$$\sum_{t:\,0\leq \Delta X_t\leq M} \frac{R_n(X_{t-1},X_t,h')}{L_n(X_{t-1},X_t,h')} \leq \sum_{t:\,0\leq \Delta X_t\leq M} \frac{\sum_{k=2}^{X_{t-1}} \mathrm{b}_{X_{t-1},\frac{h'}{n^2}}(k) g(\Delta X_t+k)}{g(\Delta X_t)\, \mathrm{b}_{X_{t-1},\frac{h'}{n^2}}(0)} \leq \frac{2}{m^*} \sum_{t:0\leq \Delta X_t\leq M} \frac{\mathrm{b}_{X_{t-1},\frac{h'}{n^2}}(2)}{\mathrm{b}_{X_{t-1},\frac{h'}{n^2}}(0)}$$
$$\leq \frac{4h'^2}{m^* n^4} \sum_{t=1}^{n} X_{t-1}^2,$$

where $m^* = \min\{g(k)|0\leq k\leq M\} > 0$. Now (3), and (7) yield the desired convergence,

$$\sum_{t:\,0\leq \Delta X_t\leq M} \frac{R_n(X_{t-1},X_t,h')}{L_n(X_{t-1},X_t,h')} \xrightarrow{p} 0, \text{ under } \mathbb{P}_{1-\frac{h_0}{n^2}}.$$

Finally, we discuss the terms for which $\Delta X_t > M$. If the support of G was given by $\{0,\dots,M\}$ there are no such terms. So we only need to consider the case, where the support of G is $\mathbb{Z}_+$. Since g is non-increasing on $\{M, M+1, \dots\}$, we have, by (43),

$$R_n(X_{t-1},X_t,h') \leq 2g(\Delta X_t)\, \mathrm{b}_{X_{t-1},\frac{h'}{n^2}}(2), \quad X_{t-1}\in A_n^{h'},$$

which yields,

$$0 \leq \frac{R_n(X_{t-1},X_t,h')}{L_n(X_{t-1},X_t,h')} \leq \frac{2g(\Delta X_t)\frac{X_{t-1}^2}{2}\frac{h'^2}{n^4}\left(1-\frac{h'}{n^2}\right)^{X_{t-1}-2}}{g(\Delta X_t)\left(1-\frac{h'}{n^2}\right)^{X_{t-1}}} \leq \frac{4h'^2}{n^4} X_{t-1}^2, \quad X_{t-1}\in A_n^{h'}.$$

From (3), and (7) it now easily follows that we have

$$\sum_{t:\Delta X_t\geq M} \frac{R_n(X_{t-1},X_t,h')}{L_n(X_{t-1},X_t,h')} \xrightarrow{p} 0, \text{ under } \mathbb{P}_{1-\frac{h_0}{n^2}}.$$

This concludes the proof of the lemma. □

Hence, the limit-distribution of the random vector $(\mathcal{L}_n(h,h_0))_{h\in I}$, for a finite subset $I\subset\mathbb{R}_+$, is the same as the limit-distribution of $(\tilde{\mathcal{L}}_n(h,h_0))_{h\in I}$. It easily follows, using (8), that $\tilde{\mathcal{L}}_n(h,h_0)$ can be decomposed as

$$\tilde{\mathcal{L}}_n(h,h_0) = \sum_{t=1}^{n} \frac{X_{t-1}-2}{n^2} \log \left(\frac{1-\frac{h}{n^2}}{1-\frac{h_0}{n^2}} \right)^{n^2} + S_n^+(h,h_0) + S_n^-(h,h_0) + o(\mathbb{P}_{1-h_0/n^2};1), \tag{17}$$

where $S_n^+(h,h_0) = \sum_{t:\,\Delta X_t \geq 0} W_{tn}^+$ and $S_n^-(h,h_0) = \sum_{t:\,\Delta X_t = -1} W_{tn}^-$, are defined by (here $\sum_{t:\,\Delta X_t=-1}$ is shorthand for $\sum_{1\leq t\leq n:\,\Delta X_t=-1}$, and for $\sum_{t:\,\Delta X_t\geq 0}$ the same convention is used),

$$W_{tn}^+ = \log \left[\frac{g(\Delta X_t)\left(1-\frac{h}{n^2}\right)^2 + X_{t-1}\frac{h}{n^2}\left(1-\frac{h}{n^2}\right) g(\Delta X_t+1)}{g(\Delta X_t)\left(1-\frac{h_0}{n^2}\right)^2 + X_{t-1}\frac{h_0}{n^2}\left(1-\frac{h_0}{n^2}\right) g(\Delta X_t+1)} \right],$$

and

$$W_{tn}^- = \log \left[\frac{X_{t-1}\frac{h}{n^2}\left(1-\frac{h}{n^2}\right) g(0) + \frac{X_{t-1}(X_{t-1}-1)}{2}\frac{h^2}{n^4} g(1)}{X_{t-1}\frac{h_0}{n^2}\left(1-\frac{h_0}{n^2}\right) g(0) + \frac{X_{t-1}(X_{t-1}-1)}{2}\frac{h_0^2}{n^4} g(1)} \right].$$

First, we treat the first term in (17). By (2) we have,

$$\log \left[\left(\frac{1-\frac{h}{n^2}}{1-\frac{h_0}{n^2}} \right)^{n^2} \right] \frac{1}{n^2} \sum_{t=1}^{n} (X_{t-1}-2) \xrightarrow{p} -\frac{(h-h_0)\mu_G}{2}, \text{ under } \mathbb{P}_{1-\frac{h_0}{n^2}}. \tag{18}$$

Next, we discuss the behavior of $S_n^+(h,h_0)$, the second term of (17). This is the content of the next lemma.

Lemma 3.2 *We have, for $h, h_0 \geq 0$,*

$$S_n^+(h,h_0) \xrightarrow{p} \frac{(h-h_0)\,(1-g(0))\mu_G}{2}, \text{ under } \mathbb{P}_{1-\frac{h_0}{n^2}}. \tag{19}$$

PROOF:
We write,

$$S_n^+(h,h_0) = \sum_{t:\,\Delta X_t \geq 0} \log\left[1+U_{tn}^+\right],$$

where

$$U_{tn}^+ = \frac{g(\Delta X_t)\left[\frac{h^2-h_0^2}{n^4} - 2\frac{h-h_0}{n^2}\right] + X_{t-1} g(\Delta X_t+1)\left[\frac{h-h_0}{n^2} - \frac{h^2-h_0^2}{n^4}\right]}{g(\Delta X_t)\left(1-\frac{h_0}{n^2}\right)^2 + X_{t-1} g(\Delta X_t+1)\frac{h_0}{n^2}\left(1-\frac{h_0}{n^2}\right)}.$$

Notice that, for n large enough,

$$U_{tn}^{+2} \le \frac{2\left(g^2(\Delta X_t)\left[\frac{h^2-h_0^2}{n^4} - 2\frac{h-h_0}{n^2}\right]^2 + X_{t-1}^2 g^2(\Delta X_t+1)\left[\frac{h-h_0}{n^2} - \frac{h^2-h_0^2}{n^4}\right]^2\right)}{g^2(\Delta X_t)\left(1-\frac{h_0}{n^2}\right)^4} \le \frac{C}{n^4}\left(X_{t-1}^2+1\right),$$

for some constant C, where we used that $e \mapsto g(e+1)/g(e)$ is bounded. From (3) we obtain,

$$\lim_{n\to\infty} \mathbb{E}_{1-\frac{h_0}{n^2}} \sum_{t:\,\Delta X_t \ge 0} U_{tn}^{+2} \le 0 + \lim_{n\to\infty} \mathbb{E}_{1-\frac{h_0}{n^2}} \frac{C}{n^4}\sum_{t=1}^n X_{t-1}^2 = 0.$$

Hence

$$\sum_{t:\,\Delta X_t \ge 0} U_{tn}^{+2} \xrightarrow{p} 0, \text{ under } \mathbb{P}_{1-h_0/n^2}, \tag{20}$$

and

$$\lim_{n\to\infty} \mathbb{P}_{1-\frac{h_0}{n^2}}\left\{\max_{t:\,\Delta X_t\ge 0} |U_{tn}^+| \in [-1/2, 1/2]\right\} = 1. \tag{21}$$

Using the expansion $\log(1+x) = x + r(x)$, where the remainder term r satisfies $|r(x)| \le 2x^2$ for $x \ge 2^{-1}$, we obtain from (20) and (21),

$$S_n^+(h,h_0) = \sum_{t:\,\Delta X_t\ge 0} \log\left[1+U_{tn}^+\right] = \sum_{t:\,\Delta X_t \ge 0} U_{tn}^+ + o(\mathbb{P}_{1-h_0/n^2};1).$$

Thus the problem reduces to determining the asymptotic behavior of $\sum_{t:\,\Delta X_t\ge 0} U_{tn}^+$. Note that,

$$\sum_{t:\,\Delta X_t\ge 0} U_{tn}^+ = \sum_{t:\,\Delta X_t \ge 0} \frac{X_{t-1} g(\Delta X_t+1)\left[\frac{h-h_0}{n^2} - \frac{h^2-h_0^2}{n^4}\right]}{g(\Delta X_t)\left(1-\frac{h_0}{n^2}\right)^2 + X_{t-1} g(\Delta X_t+1)\frac{h_0}{n^2}\left(1-\frac{h_0}{n^2}\right)} + o(\mathbb{P}_{1-h_0/n^2};1).$$

Using that $e \mapsto g(e+1)/g(e)$ is bounded and (3), we obtain

$$\begin{aligned}\sum_{t:\,\Delta X_t\ge 0} &\left| \frac{X_{t-1} g(\Delta X_t+1)\left[\frac{h-h_0}{n^2} - \frac{h^2-h_0^2}{n^4}\right]}{g(\Delta X_t)\left(1-\frac{h_0}{n^2}\right)^2 + X_{t-1} g(\Delta X_t+1)\frac{h_0}{n^2}\left(1-\frac{h_0}{n^2}\right)} - \frac{(h-h_0)}{n^2}\frac{X_{t-1}g(\Delta X_t+1)}{g(\Delta X_t)}\right| \\ &\le \frac{C}{n^4}\sum_{t=1}^n X_{t-1}^2 \xrightarrow{p} 0, \text{ under } \mathbb{P}_{1-\frac{h_0}{n^2}}.\end{aligned}$$

Thus the previous three displays and (8) yield

$$S_n^+(h,h_0) = \frac{h-h_0}{n^2}\sum_{t=1}^n X_{t-1}\frac{g(\Delta X_t+1)}{g(\Delta X_t)}1\{\Delta X_t \geq 0,\, X_{t-1}-\vartheta\circ X_{t-1}\leq 1\} + o(\mathbb{P}_{1-h_0/n^2};1).$$

Finally, we will show that

$$\frac{1}{n^2}\sum_{t=1}^n X_{t-1}\frac{g(\Delta X_t+1)}{g(\Delta X_t)}1\{\Delta X_t \geq 0,\, X_{t-1}-\vartheta\circ X_{t-1}\leq 1\} \xrightarrow{p} \frac{(1-g(0))\mu_G}{2}, \text{ under } \mathbb{P}_{1-\frac{h_0}{n^2}}, \tag{22}$$

which will conclude the proof. For notational convenience we introduce

$$\begin{aligned} Z_t &= \frac{g(\Delta X_t+1)}{g(\Delta X_t)}1\{\Delta X_t \geq 0,\, X_{t-1}-\vartheta\circ X_{t-1}\leq 1\} \\ &= \frac{g(\varepsilon_t+1)}{g(\varepsilon_t)}1\{X_{t-1}-\vartheta\circ X_{t-1}=0\} + \frac{g(\varepsilon_t)}{g(\varepsilon_t-1)}1\{\varepsilon_t\geq 1,\, X_{t-1}-\vartheta\circ X_{t-1}=1\}. \end{aligned}$$

Using that ε_t is independent of $X_{t-1}-\vartheta\circ X_{t-1}$ we obtain

$$\begin{aligned} \mathbb{E}_{1-\frac{h_0}{n^2}}\left[Z_t \mid X_{t-1}-\vartheta\circ X_{t-1}\right] &= (1-g(0))1\{X_{t-1}-\vartheta\circ X_{t-1}=0\} \\ &\quad + 1\{X_{t-1}-\vartheta\circ X_{t-1}=1\}\mathbb{E}\frac{g(\varepsilon_t)}{g(\varepsilon_t-1)}1\{\varepsilon_t\geq 1\}, \end{aligned}$$

where we used that $\mathbb{E}g(\varepsilon_1+1)/g(\varepsilon_1) = 1-g(0)$ and $\mathbb{E}1\{\varepsilon_1\geq 1\}g(\varepsilon_1)/g(\varepsilon_1-1) < \infty$, since we assumed that g is eventually decreasing. So we have

$$\begin{aligned} Z_t - \mathbb{E}_{1-\frac{h_0}{n^2}}\left[Z_t \mid X_{t-1}-\vartheta\circ X_{t-1}\right] &= \left(\frac{g(\varepsilon_t+1)}{g(\varepsilon_t)} - \mathbb{E}\frac{g(\varepsilon_t+1)}{g(\varepsilon_t)}\right)1\{X_{t-1}-\vartheta\circ X_{t-1}=0\} \\ &\quad + \left(\frac{g(\varepsilon_t)}{g(\varepsilon_t-1)}1\{\varepsilon_t\geq 1\} - \mathbb{E}\frac{g(\varepsilon_t)}{g(\varepsilon_t-1)}1\{\varepsilon_t\geq 1\}\right)1\{X_{t-1}-\vartheta\circ X_{t-1}=1\}. \end{aligned}$$

From this it is not hard to see that we have,

$$\mathbb{E}_{1-\frac{h_0}{n^2}}X_{t-1}\left(Z_t - \mathbb{E}_{1-\frac{h_0}{n^2}}\left[Z_t \mid X_{t-1}-\vartheta\circ X_{t-1}\right]\right) = 0,$$

for $s<t$,

$$\mathbb{E}_{1-\frac{h_0}{n^2}}X_{t-1}\left(Z_t - \mathbb{E}_{1-\frac{h_0}{n^2}}\left[Z_t \mid X_{t-1}-\vartheta\circ X_{t-1}\right]\right)X_{s-1}\left(Z_s - \mathbb{E}_{1-\frac{h_0}{n^2}}\left[Z_s \mid X_{s-1}-\vartheta\circ X_{s-1}\right]\right) = 0.$$

and,

$$\mathbb{E}_{1-\frac{h_0}{n^2}}\left(Z_t - \mathbb{E}_{1-\frac{h_0}{n^2}}\left[Z_t \mid X_{t-1}-\vartheta\circ X_{t-1}\right]\right)^2 \leq C, \tag{23}$$

for $C = 2(\mathrm{Var}\,(g(\varepsilon_1+1)/g(\varepsilon_1)) + \mathrm{Var}\,\big(1_{\{\varepsilon_t \geq 1\}} g(\varepsilon_1)/g(\varepsilon_1 - 1)\big)$. Thus, by (3), it follows that

$$\begin{aligned}
&\mathbb{E}_{1-\frac{h_0}{n^2}} \left(\frac{1}{n^2} \sum_{t=1}^n X_{t-1} \left(Z_t - \mathbb{E}_{1-\frac{h_0}{n^2}} \left[Z_t \mid X_{t-1} - \vartheta \circ X_{t-1} \right] \right) \right)^2 \\
&\quad = \frac{1}{n^4} \sum_{t=1}^n \mathbb{E}_{1-\frac{h_0}{n^2}} X_{t-1}^2 \left(Z_t - \mathbb{E}_{1-\frac{h_0}{n^2}} \left[Z_t \mid X_{t-1} - \vartheta \circ X_{t-1} \right] \right)^2 \leq \frac{C}{n^4} \sum_{t=1}^n \mathbb{E}_{1-\frac{h_0}{n^2}} X_{t-1}^2 \to 0.
\end{aligned}$$

Hence (22) is equivalent to,

$$\frac{1}{n^2} \sum_{t=1}^n X_{t-1} \mathbb{E}_{1-\frac{h_0}{n^2}} \left[Z_t \mid X_{t-1} - \vartheta \circ X_{t-1} \right] \xrightarrow{p} \frac{(1-g(0))\mu_G}{2}, \text{ under } \mathbb{P}_{1-\frac{h_0}{n^2}}. \tag{24}$$

Since, by (3),

$$\frac{1}{n^2} \sum_{t=1}^n \mathbb{E}_{1-\frac{h_0}{n^2}} X_{t-1} 1\{X_{t-1} - \vartheta \circ X_{t-1} = 1\} = \frac{h_0}{n^4} \sum_{t=1}^n \mathbb{E}_{1-\frac{h_0}{n^2}} X_{t-1}^2 \left(1 - \frac{h_0}{n^2}\right)^{X_{t-1}-1} \leq \frac{h_0}{n^4} \sum_{t=1}^n \mathbb{E}_{1-\frac{h_0}{n^2}} X_{t-1}^2 \to 0,$$

we have, using (8),

$$\begin{aligned}
&\left| \frac{1}{n^2} \sum_{t=1}^n X_{t-1} \mathbb{E}_{1-\frac{h_0}{n^2}} \left[Z_t \mid X_{t-1} - \vartheta \circ X_{t-1} \right] - \frac{1-g(0)}{n^2} \sum_{t=1}^n X_{t-1} \right| \\
&\quad \leq \left| \mathbb{E} \frac{g(\varepsilon_t)}{g(\varepsilon_t - 1)} 1\{\varepsilon_t \geq 1\} - (1-g(0)) \right| \frac{1}{n^2} \sum_{t=1}^n X_{t-1} 1\{X_{t-1} - \vartheta \circ X_{t-1} = 1\} \\
&\quad + \frac{1-g(0)}{n^2} \sum_{t=1}^n X_{t-1} 1\{X_{t-1} - \vartheta \circ X_{t-1} \geq 2\} \xrightarrow{p} 0, \text{ under } \mathbb{P}_{1-\frac{h_0}{n^2}},
\end{aligned}$$

we conclude (24), which finally concludes the proof of the lemma. □

Finally, we discuss the term $S_n^-(h, h_0)$ in (17). Under $\mathbb{P}_1$ this term is not present, so we only need to consider $h_0 > 0$. We organize the result and its proof in the following lemma.

Lemma 3.3 *We have, for $h_0 > 0$, $h \geq 0$,*

$$S_n^-(h, h_0) = \log \left[\frac{h}{h_0} \right] \sum_{t=1}^n 1\{\Delta X_t < 0\} + o(\mathbb{P}_{1-h_0/n^2}; 1), \tag{25}$$

where we set $\log(0) = -\infty$ and $-\infty \cdot 0 = 0$.

Proof:
First we consider $h=0$. From the definition of $S_n^-(0,h_0)$ we see that $S_n^-(0,h_0)=0$ if $\sum_{t=1}^n 1\{\Delta X_t<0\}=0$ (since an empty sum equals zero by definition). And if $\sum_{t=1}^n 1\{\Delta X_t<0\}\geq 1$ we have $S_n^-(0,h_0)=-\infty$ (since $W_{tn}^-=-\infty$ for $h=0$). This concludes the proof for $h=0$.
So we now consider $h>0$. We rewrite

$$W_{tn}^- = \log\left[\frac{\frac{h}{h_0}\left(\frac{1-\frac{h}{n^2}}{1-\frac{h_0}{n^2}}\right)+\frac{X_{t-1}-1}{2n^2}\frac{h^2 g(1)}{g(0)h_0\left(1-\frac{h_0}{n^2}\right)}}{1+\frac{X_{t-1}-1}{2n^2}\frac{h_0 g(1)}{g(0)\left(1-\frac{h_0}{n^2}\right)}}\right].$$

By (8), the proof is finished, if we show that

$$\sum_{t:\,\Delta X_t=-1}\left|W_{tn}^- - \log\left[\frac{h}{h_0}\right]\right| \xrightarrow{p} 0, \text{ under } \mathbb{P}_{1-\frac{h_0}{n^2}}.$$

Using the inequality $|\log((a+b)/(c+d))-\log(a/c)|\leq b/a+d/c$ for $a,c>0$, $b,d\geq 0$, we obtain

$$\begin{aligned}\left|W_{tn}^- - \log\left[\frac{h}{h_0}\right]\right| &\leq \left|W_{tn}^- - \log\left[\frac{h}{h_0}\left(\frac{1-\frac{h}{n^2}}{1-\frac{h_0}{n^2}}\right)\right]\right| + O(n^{-2})\\ &\leq \frac{X_{t-1}-1}{2n^2}\left[\frac{h^2 g(1)}{g(0)h_0\left(1-\frac{h_0}{n^2}\right)}\left(\frac{h}{h_0}\left(\frac{1-\frac{h}{n^2}}{1-\frac{h_0}{n^2}}\right)\right)^{-1}+\frac{h_0 g(1)}{g(0)\left(1-\frac{h_0}{n^2}\right)}\right]+O(n^{-2}).\end{aligned}$$

Hence, it suffices to show that

$$\sum_{t:\,\Delta X_t=-1}\frac{X_{t-1}}{n^2} \xrightarrow{p} 0, \text{ under } \mathbb{P}_{1-\frac{h_0}{n^2}}.$$

Note first that we have, by (8),

$$0\leq \frac{1}{n^2}\sum_{t=1}^n X_{t-1}1\{\Delta X_t=-1\} = \frac{1}{n^2}\sum_{t=1}^n X_{t-1}1\{\Delta X_t=-1,\varepsilon_t=0\}+o(\mathbb{P}_{1-h_0/n^2};1).$$

We show that the expectation of the first term on the right-hand side in the previous display converges to zero, which will conclude the proof. We have, by (3),

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n^2}\sum_{t=1}^n \mathbb{E}_{1-\frac{h_0}{n^2}}X_{t-1}1\{\Delta X_t=-1,\varepsilon_t=0\} &= \lim_{n\to\infty}\frac{h_0}{n^4}\sum_{t=1}^n \mathbb{E}_{1-\frac{h_0}{n^2}}g(0)X_{t-1}^2\left(1-\frac{h_0}{n^2}\right)^{X_{t-1}-1}\\ &\leq \lim_{n\to\infty}\frac{h_0 g(0)}{n^4}\sum_{t=1}^n \mathbb{E}_{1-\frac{h_0}{n^2}}X_{t-1}^2=0,\end{aligned}$$

which concludes the proof of the lemma. □

To complete the proof of the theorem, note that we obtain from Lemma 3.1, (17), (18), Lemma 3.2 and Lemma 3.3,

$$\mathcal{L}_n(h,h_0) = \tilde{\mathcal{L}}_n(h,h_0) + o(\mathbb{P}_{1-h_0/n^2};1) = -\frac{(h-h_0)\,g(0)\mu_G}{2} + \log\left[\frac{h}{h_0}\right]\sum_{t=1}^n 1\{\Delta X_t < 0\} + o(\mathbb{P}_{1-h_0/n^2};1),$$

where we interpret $\log(0) = -\infty$, $\log(0)\cdot 0 = 0$ and $\log(h/0)\sum_{t=1}^n 1\{\Delta X_t < 0\} = 0$ when $h_0 = 0$, $h > 0$. Hence, Theorem 2.1 implies that, for a finite subset $I \subset \mathbb{R}_+$,

$$(\mathcal{L}_n(h,h_0))_{h\in I} \xrightarrow{d} \frac{\mathrm{d}\mathbb{Q}_h}{\mathrm{d}\mathbb{Q}_{h_0}}(Z), \text{ under } \mathbb{P}_{1-\frac{h_0}{n^2}},$$

which concludes the proof. □

Remark 3 *In the proof we have seen that,*

$$\log\frac{\mathrm{d}\mathbb{P}^{(n)}_{1-\frac{h}{n^2}}}{\mathrm{d}\mathbb{P}^{(n)}_{1-\frac{h_0}{n^2}}} = -\frac{(h-h_0)\,g(0)\mu_G}{2} + \log\left[\frac{h}{h_0}\right]\sum_{t=1}^n 1\{\Delta X_t < 0\} + o(\mathbb{P}_{1-h_0/n^2};1).$$

So, heuristically, we can interpret $\sum_{t=1}^n 1\{\Delta X_t < 0\}$ *as an 'approximately sufficient statistic'.*

Remark 4 *It is straightforward to see that the experiments*

$$\mathcal{B}_n^0 = \left(\mathbb{Z}_+, 2^{\mathbb{Z}_+}, \left(\mathit{Binomial}\left(n, \frac{h}{n}\right) \mid h \geq 0\right)\right), \text{ and } \mathcal{B}_n^1 = \left(\mathbb{Z}_+, 2^{\mathbb{Z}_+}, \left(\mathit{Binomial}\left(n, 1-\frac{h}{n}\right) \mid h \geq 0\right)\right),$$

$n \in \mathbb{N}$, *both converge to the Poisson experiment* $\mathcal{P} = \left(\mathbb{Z}_+, 2^{\mathbb{Z}_+}, (\mathit{Poisson}(h) \mid h \geq 0)\right)$. *Since the law of* X_t *given* X_{t-1} *is the convolution of a Binomial*(X_{t-1}, θ) *distribution with* G, *one might be tempted to think that the convergence of experiments* $\mathcal{E}_n(G) \to \mathcal{E}(G)$ *somehow follows from the convergence* $\mathcal{B}_n^1 \to \mathcal{P}$. *However, a similar reasoning would yield that the sequence of experiments*

$$\mathcal{E}_n^0(G) = \left(\mathbb{Z}_+^{n+1}, 2^{\mathbb{Z}_+^{n+1}}, \left(\mathbb{P}^{(n)}_{\frac{h}{\sqrt{n}}} \mid h \geq 0\right)\right), \quad n \in \mathbb{N},$$

converges to some Poisson experiment. This is not the case. As Proposition 4.6 shows, the sequence $\left(\mathcal{E}_n^0(G)\right)_{n\in\mathbb{N}}$ *converges to the normal location experiment* $(\mathbb{R}, \mathcal{B}(\mathbb{R}), (\mathrm{N}(h,\tau) \mid h \geq 0))$, *for some* $\tau > 0$.

4 Applications

This section addresses the following applications. In Section 4.1 we discuss efficient estimation of h, the deviation from a unit root, in the nearly unstable case for two settings. The first setting, discussed in Section 4.1.1, treats the case that the immigration distribution G is completely known. And the second setting, analyzed in 4.1.2, considers a semiparametric model, where hardly any conditions on G are imposed. In Section 4.2 we provide an efficient estimator of θ in the 'global' INAR model. Finally, we discuss testing for a unit root in Section 4.3.

4.1 Efficient estimation of h in nearly unstable INAR models

4.1.1 G known

In this section G is assumed to be known. So we consider the sequence of experiments $(\mathcal{E}_n(G))_{n\in\mathbb{N}}$. As before, we denote the observation from the limit experiment $\mathcal{E}(G)$ by Z, and $\mathbb{Q}_h = \text{Poisson}(hg(0)\mu_G/2)$.

Since we have established convergence of $(\mathcal{E}_n(G))_{n\in\mathbb{N}}$ to $\mathcal{E}(G)$, an application of the Le Cam-Van der Vaart Asymptotic Representation Theorem yields the following proposition.

Proposition 4.1 *Suppose G satisfies Assumption 3.1. If $(T_n)_{n\in\mathbb{N}}$ is a sequence of estimators of h in the sequence of experiments $(\mathcal{E}_n(G))_{n\in\mathbb{N}}$ such that $\mathcal{L}(T_n \mid \mathbb{P}_{1-h/n^2}) \to Z_h$ for all $h \geq 0$, then there exists a map $t : \mathbb{Z}_+ \times [0,1] \to \mathbb{R}$ such that $Z_h = \mathcal{L}(t(Z,U) \mid \mathbb{Q}_h \times \text{Uniform}[0,1])$ (i.e. U is distributed uniformly on $[0,1]$ and independent of the observation Z from the limit experiment $\mathcal{E}(G)$).*

PROOF:
Under the stated conditions the sequence of experiments $(\mathcal{E}_n(G))_{n\in\mathbb{N}}$ converges to the Poisson limit experiment $\mathcal{E}(G)$ (by Theorem 3.1). Since this experiment is dominated by counting measure on $\mathbb{Z}_+$, the result follows by applying the Le Cam-Van der Vaart Asymptotic Representation Theorem (see, for instance, Theorem 3.1 in Van der Vaart (1991) or Theorem 9.3 in Van der Vaart (2000)). □

Thus, to any set of limit-laws of an estimator there is a randomized estimator in the limit experiment which has the same set of laws. If the asymptotic performance of an estimator is considered to be determined by its sets of limit laws, the limit experiment thus gives a lower bound to what is possible: along the sequence of experiments you cannot do better than the best procedure in the limit experiment.

To discuss efficient estimation we need to prescribe what we judge to be optimal in the Poisson limit experiment. Often a normal location experiment is the limit experiment. For such a normal location experiment, i.e. estimate h on basis of one observation Y from $\text{N}(h,\tau)$ (τ known), it is natural to restrict to location-equivariant estimators. For this class one has a convolution-property (see, e.g., Proposition 8.4

in Van der Vaart (2000)): the law of every location-equivariant estimator T of h can be decomposed as $T \stackrel{d}{=} Y + V$, where V is independent of Y. This yields, by Anderson's lemma (see, e.g., Lemma 8.5 in Van der Vaart (2000)), efficiency of Y (within the class of location-equivariant estimators) for all bowl-shaped loss functions. More general, there are convolution-results for shift-experiments. However, the Poisson limit experiment $\mathcal{E}(G)$ has not a natural shift structure. In such a Poisson setting it seems reasonable to minimize variance amongst the unbiased estimators.

Proposition 4.2 *Suppose G is such that $0 < g(0) < 1$ and $\mu_G < \infty$. In the experiment,*

$$\mathcal{E}(G) = \left(\mathbb{Z}_+, 2^{\mathbb{Z}_+}, (\mathbb{Q}_h = \textit{Poisson}(hg(0)\mu_G/2) \mid h \geq 0)\right),$$

the unbiased estimator $2Z/g(0)\mu_G$ minimizes the variance amongst all randomized estimators $t(Z,U)$ for which $\mathbb{E}_h t(Z,U) = h$ for all $h \geq 0$, i.e.

$$\mathrm{Var}_h\, t(Z,U) \geq \mathrm{Var}_h \left(\frac{2Z}{g(0)\mu_G}\right) = \frac{2h}{g(0)\mu_G} \text{ for all } h \geq 0.$$

Proof:
This is an immediate consequence of the Lehmann-Scheffé theorem. □

A combination of this proposition with Proposition 4.1 yields a variance lower-bound to asymptotically unbiased estimators in the sequence of experiments $(\mathcal{E}_n(G))_{n\in\mathbb{N}}$.

Proposition 4.3 *Suppose G satisfies Assumption 3.1. If $(T_n)_{n\in\mathbb{N}}$ is an estimator of h in the sequence of experiments $(\mathcal{E}_n(G))_{n\in\mathbb{N}}$ such that $\mathcal{L}(T_n|\mathbb{P}_{1-h/n^2}) \to Z_h$ with $\int z\, \mathrm{d}Z_h(z) = h$ for all $h \geq 0$, then we have*

$$\int (z-h)^2\, \mathrm{d}Z_h(z) \geq \frac{2h}{g(0)\mu_G}, \text{ for all } h \geq 0. \tag{26}$$

Proof:
By Proposition 4.1 there exists a randomized estimator $t(Z,U)$ in the limit experiment such that $Z_h = \mathcal{L}(t(Z,U) \mid \mathbb{Q}_h \times \mathrm{Uniform}[0,1])$. Hence $\mathbb{E}_h t(Z,U) = h$ and $\mathrm{Var}_h\, t(Z,U) = \int (z-h)^2\, \mathrm{d}Z_h(z)$. Now the result follows from Proposition 4.2. □

It is not unnatural to restrict to estimators that satisfy $\mathcal{L}(T_n \mid \mathbb{P}_{1-h/n^2}) \to Z_h$. We make the additional restriction that $\int z\, \mathrm{d}Z_h(Z) = h$, i.e. the limit-distribution is unbiased. Now, based on the previous proposition, it is natural to call an estimator in this class efficient if it attains the variance-bound (26). To demonstrate efficiency of a given estimator, one only needs to show that it belongs to the class of asymptotically unbiased estimators, and that it attains the bound.

First we discuss the OLS estimator. Let $\theta_n = 1 - h/n^2$. Rewriting $X_t = \vartheta \circ X_{t-1} + \varepsilon_t = \mu_G + \theta_n X_t + u_t$ for $u_t = \varepsilon_t - \mu_G + \vartheta \circ X_{t-1} - \theta_n X_{t-1}$, we obtain the regression-equation $X_t - \mu_G = \theta_n X_{t-1} + u_t$, which can also be written as $n^2(X_t - X_{t-1} - \mu_G) = h(-X_{t-1}) + n^2 u_t$ (note that indeed $\mathbb{E}_{\theta_n} u_t = \mathbb{E}_{\theta_n} X_{t-1} u_t = 0$). So the OLS estimator of θ_n is given by,

$$\widehat{\theta}_n^{\mathrm{OLS}} = \frac{\sum_{t=1}^n X_{t-1}(X_t - \mu_G)}{\sum_{t=1}^n X_{t-1}^2}, \tag{27}$$

and the OLS estimator of h is given by,

$$\widehat{h}_n^{\mathrm{OLS}} = -\frac{n^2 \sum_{t=1}^n X_{t-1}(X_t - X_{t-1} - \mu_G)}{\sum_{t=1}^n X_{t-1}^2} = n^2 \left(1 - \widehat{\theta}_n^{\mathrm{OLS}}\right).$$

Ispány et al. (2003) analyzed the asymptotic behavior of the OLS estimator under localizing rate n. However, since the convergence of experiments takes place at rate n^2, we analyze the behavior of the OLS estimator also under localizing rate n^2. The next proposition gives this behavior.

Proposition 4.4 *If $\mathbb{E}_G \varepsilon_1^4 < \infty$, then we have, for all $h \geq 0$,*

$$\left|\widehat{h}_n^{OLS}\right| \xrightarrow{p} \infty, \text{ under } \mathbb{P}_{1-\frac{h}{n^2}}.$$

PROOF:
Let $h \geq 0$ and set $\theta_n = 1 - h/n^2$, $\mathbb{P}_n = \mathbb{P}_{\theta_n}$, and $\mathbb{E}_n(\cdot) = \mathbb{E}_{\theta_n}(\cdot)$. We have

$$n^{3/2}\left(\widehat{\theta}_n^{\mathrm{OLS}} - \theta_n\right) = \frac{n^{-3/2} \sum_{t=1}^n X_{t-1}\left(\varepsilon_t - \mu_G + \vartheta \circ X_{t-1} - \theta_n X_{t-1}\right)}{n^{-3} \sum_{t=1}^n X_{t-1}^2}.$$

We prove that,

$$n^{-3/2} \sum_{t=1}^n X_{t-1}\left(\vartheta \circ X_{t-1} - \theta_n X_{t-1}\right) \xrightarrow{p} 0, \text{ under } \mathbb{P}_n, \tag{28}$$

$$n^{-3} \sum_{t=1}^n X_{t-1}^2 \xrightarrow{p} \frac{\mu_G^2}{3}, \text{ under } \mathbb{P}_n, \tag{29}$$

$$n^{-3/2} \sum_{t=1}^n X_{t-1}\left(\varepsilon_t - \mu_G\right) \xrightarrow{d} \mathrm{N}\left(0, \frac{\sigma_G^2 \mu_G^2}{3}\right), \text{ under } \mathbb{P}_n, \tag{30}$$

all hold, which yields,

$$n^{3/2}\left(\widehat{\theta}_n^{\mathrm{OLS}} - \theta_n\right) \xrightarrow{d} \mathrm{N}\left(0, \frac{3\sigma_G^2}{\mu_G^2}\right), \text{ under } \mathbb{P}_n, \tag{31}$$

which in turn will yield the result, since,

$$\left|\widehat{h}_n^{\mathrm{OLS}}\right| = \sqrt{n}\left|-n^{3/2}\left(\widehat{\theta}_n^{\mathrm{OLS}} - \theta_n\right) + \frac{h}{\sqrt{n}}\right|.$$

First, we treat (28). Using the martingale structure and the fact that, conditional on X_{t-1}, $\vartheta \circ X_{t-1}$ has a Binomial(X_{t-1}, θ_n) distribution, we obtain,

$$\mathbb{E}_n\left(\frac{1}{n^{3/2}}\sum_{t=1}^n X_{t-1}\left(\vartheta \circ X_{t-1} - \theta_n X_{t-1}\right)\right)^2 = \frac{1}{n^3}\sum_{t=1}^n \mathbb{E}_n X_{t-1}^2\left(\vartheta \circ X_{t-1} - \theta_n X_{t-1}\right)^2 \leq \frac{h}{n^5}\sum_{t=1}^n \mathbb{E}_n X_{t-1}^3 \to 0,$$

where the last step follows from $\mathbb{E}_n X_t^3 \leq \mathbb{E}_G(\sum_{i=1}^t \varepsilon_i)^3 = O(t^3)$.
Next, we discuss (29). Introduce $S_t = \sum_{i=1}^t \varepsilon_i$ and $Y_t = S_t - X_t$. Notice that Y_t is nonnegative, $Y_s = Y_{s-1} + (X_{s-1} - \vartheta \circ X_{s-1})$ for $s \geq 1$, $Y_0 = 0$, and thus $Y_t = \sum_{i=1}^t (X_{i-1} - \vartheta \circ X_{i-1})$. Decompose $X_t^2 = Y_t^2 + S_t^2 - 2S_tY_t$. It is straightforward to check that $n^{-3}\sum_{t=1}^n S_t^2 \xrightarrow{p} \mu_G^2/3$, under $\mathbb{P}_n$. To obtain (29), it thus suffices to prove that $n^{-3}\sum_{t=1}^n Y_t^2$ and $n^{-3}\sum_{t=1}^n S_tY_t$ both converge to zero in probability under $\mathbb{P}_n$. We have, use that conditional on X_t, $X_t - \vartheta \circ X_t$ has a Binomial$(X_t, h/n^2)$ distribution,

$$\begin{aligned}
\mathbb{E}_n Y_t^2 &= \sum_{i=1}^t \sum_{j=1}^t \mathbb{E}_n(X_{i-1} - \vartheta \circ X_{i-1})(X_{j-1} - \vartheta \circ X_{j-1}) \\
&= \sum_{i=1}^t \mathbb{E}_n(X_{i-1} - \vartheta \circ X_{i-1})^2 + 2\sum_{i=1}^t \sum_{j=1}^{i-1} \frac{h}{n^2}\mathbb{E}_n(X_{j-1} - \vartheta \circ X_{j-1})X_{i-1} \\
&\leq \sum_{i=1}^t \left(\frac{h}{n^2}\mathbb{E}_n X_{i-1} + \frac{h^2}{n^4}\mathbb{E}_n X_{i-1}^2\right) + \frac{2ch}{n^2}\sum_{i=1}^t \sum_{j=1}^{i-1} i^2,
\end{aligned}$$

where we used the (very crude) bounds $\mathbb{E}_n(X_{s-1} - \vartheta \circ X_{s-1})X_{v-1} \leq \mathbb{E}_G S_{s-1}S_{v-1} \leq \mathbb{E}_G S_v^2$ for $s < v$ and $\mathbb{E}_G S_v^2 \leq cv^2$ for some constant $c > 0$ (not depending on v). Since $n^{-4}\sum_{t=1}^n \sum_{s=1}^n \mathbb{E}_n X_s X_t$ converges by (2), we now easily obtain $n^{-3}\sum_{t=1}^n Y_t^2 \xrightarrow{p} 0$, under $\mathbb{P}_n$. Furthermore, we have,

$$\begin{aligned}
\frac{1}{n^3}\sum_{t=1}^n \mathbb{E}_n S_t Y_t &\leq \frac{1}{n^3}\sum_{t=1}^n \sqrt{\mathbb{E}_n S_t^2 \mathbb{E}_n Y_t^2} \leq \frac{\sqrt{\mu^2 + \sigma^2}}{n^3}\sum_{t=1}^n t\sqrt{\mathbb{E}_n Y_t^2} \\
&\leq \frac{\sqrt{\mu^2 + \sigma^2}}{n^3}\sqrt{\frac{n(2n+1)(n+1)}{6}}\sqrt{\sum_{t=1}^n \mathbb{E}_n Y_t^2} \to 0,
\end{aligned}$$

which concludes the proof of (29).
Finally, we treat (30). By a martingale central limit theorem for arrays (see Theorem 3.2, Corollary 3.1 and the remark after that corollary in Hall and Heyde (1980)) we have (30), if the following two conditions

are satisfied,

$$\frac{1}{n^3}\sum_{t=1}^{n} X_{t-1}^2 \mathbb{E}_n\left[(\varepsilon_t-\mu_G)^2 \mid X_{t-1}\right] \xrightarrow{p} \frac{\sigma_G^2\mu_G^2}{3}, \text{ under } \mathbb{P}_n, \tag{32}$$

and for all $\epsilon > 0$,

$$\frac{1}{n^3}\sum_{t=1}^{n} X_{t-1}^2 \mathbb{E}_n\left[(\varepsilon_t-\mu_G)^2 1\{X_{t-1}|\varepsilon_t-\mu_G| > \epsilon n^{3/2}\} \mid X_{t-1}\right] \xrightarrow{p} 0, \text{ under } \mathbb{P}_n. \tag{33}$$

Since ε_t is independent of X_{t-1} (32) immediately follows from (29). To see that the Lindeberg condition (33) is satisfied, notice that, using the independence of ε_t and X_{t-1}, Cauchy-Schwarz, and Markov's inequality, we have

$$\begin{aligned}\mathbb{E}_n\left[(\varepsilon_t-\mu_G)^2 1\{X_{t-1}|\varepsilon_t-\mu_G| > \epsilon n^{3/2}\} \mid X_{t-1}\right] &\le \sqrt{\mathbb{E}_G(\varepsilon_1-\mu_G)^4 \mathbb{P}_n\left[|\varepsilon_t-\mu_G| > \frac{\epsilon n^{3/2}}{X_{t-1}} \mid X_{t-1}\right]} \\ &\le \frac{\sigma_G X_{t-1}}{\epsilon n^{3/2}}\sqrt{\mathbb{E}_G\left(\varepsilon_1-\mu_G\right)^4},\end{aligned}$$

which yields,

$$\frac{1}{n^3}\sum_{t=1}^{n} X_{t-1}^2 \mathbb{E}_n\left[(\varepsilon_t-\mu_G)^2 1\{X_{t-1}|\varepsilon_t-\mu_G| > \epsilon n^{3/2}\} \mid X_{t-1}\right] \le \frac{\sigma_G\sqrt{\mathbb{E}_G(\varepsilon_1-\mu_G)^4}}{\epsilon n^{9/2}}\sum_{t=1}^{n} X_{t-1}^3 \xrightarrow{p} 0 \text{ under } \mathbb{P}_n,$$

since we noticed in the proof of (28) that $n^{-(4+\alpha)}\sum_{t=1}^n X_{t-1}^3 \xrightarrow{p} 0$, under $\mathbb{P}_n$, for $\alpha > 0$. This concludes the proof. □

Thus the OLS estimator explodes. How should we estimate h then? Recall, that we interpreted $\sum_{t=1}^n 1\{\Delta X_t < 0\}$ as an approximately sufficient statistic for h. Hence, it is natural to try to construct an efficient estimator based on this statistic. Using Theorem 2.1 we see that this is indeed possible.

Corollary 4.1 *Let G satisfy Assumption 3.1. The estimator,*

$$\widehat{h}_n = \frac{2\sum_{t=1}^n 1\{\Delta X_t < 0\}}{g(0)\mu_G}, \tag{34}$$

is an efficient estimator of h in the sequence $(\mathcal{E}_n(G))_{n\in\mathbb{N}}$.

4.1.2 A semiparametric model

So far we assumed that G is known. In this section, where we instead consider a semiparametric model, we hardly impose conditions on G (see, for example, Wefelmeyer (1999) for an introduction to semiparametric stationary Markov models). The dependence of $\mathbb{P}_\theta$ upon G is made explicit by adding a subscript: $\mathbb{P}_{\theta,G}$. Formally, we consider the sequence of experiments,

$$\mathcal{E}_n = \left(\mathbb{Z}_+^{n+1}, 2^{\mathbb{Z}_+^{n+1}}, \left(\mathbb{P}^{(n)}_{1-\frac{h}{n^2},G} \mid (h,G) \in [0,\infty) \times \mathcal{G}\right)\right), \quad n \in \mathbb{N},$$

where $\mathcal{G}$ is the set of all distributions on $\mathbb{Z}_+$ that satisfy Assumption 3.1.

The goal is to estimate h efficiently. Here efficient, just as in the previous section, means asymptotically unbiased with minimal variance. Since the semiparametric model is more realistic, the estimation of h becomes more difficult. As we will see, the situation for our semiparametric model is quite fortunate: we can estimate h with the same asymptotic precision as in the case that G is known. In semiparametric statistics this is called adaptive estimation.

The efficient estimator for the case that G is known cannot be used anymore, since it depends on $g(0)$ and μ_G. The obvious idea is to replace these objects by estimates. The next proposition provides consistent estimators.

Proposition 4.5 *Let $h \geq 0$ and G satisfy $\sigma_G^2 < \infty$. Then we have,*

$$\widehat{g}_n(0) = \frac{1}{n}\sum_{t=1}^n 1\{X_t = X_{t-1}\} \xrightarrow{p} g(0) \text{ and } \widehat{\mu}_{G,n} = \frac{X_n}{n} \xrightarrow{p} \mu_G \text{ under } \mathbb{P}_{1-\frac{h}{n^2},G}.$$

Proof:
Notice first that we have,

$$\frac{1}{n}\sum_{t=1}^n (X_{t-1} - \vartheta \circ X_{t-1}) \xrightarrow{p} 0 \text{ under } \mathbb{P}_{1-\frac{h}{n^2},G}, \tag{35}$$

since, condition on X_{t-1} and use (2),

$$0 \leq \frac{1}{n}\sum_{t=1}^n \mathbb{E}_{1-\frac{h}{n^2},G}(X_{t-1} - \vartheta \circ X_{t-1}) = \frac{h}{n^3}\sum_{t=1}^n \mathbb{E}_{1-\frac{h}{n^2},G}X_{t-1} \to 0.$$

Using that $|1\{X_t = X_{t-1}\} - 1\{\varepsilon_t = 0\}| = 1$ only if $X_{t-1} - \vartheta \circ X_{t-1} \geq 1$, we easily obtain, by using (35),

$$\left| \widehat{g}_n(0) - \frac{1}{n} \sum_{t=1}^{n} 1\{\varepsilon_t = 0\} \right| \leq \frac{1}{n} \sum_{t=1}^{n} 1\{X_{t-1} - \vartheta \circ X_{t-1} \geq 1\} \leq \frac{1}{n} \sum_{t=1}^{n} (X_{t-1} - \vartheta \circ X_{t-1}) \xrightarrow{p} 0.$$

Now the result for $\hat{g}_n(0)$ follows by applying the weak law of large numbers to $n^{-1} \sum_{t=1}^{n} 1\{\varepsilon_t = 0\}$. Next, consider $\widehat{\mu}_{G,n}$. We have, use (35) and the weak law of large numbers for $n^{-1} \sum_{t=1}^{n} \varepsilon_t$,

$$\widehat{\mu}_{G,n} = \frac{X_n}{n} = \frac{1}{n} \sum_{t=1}^{n} (X_t - X_{t-1}) = \frac{1}{n} \sum_{t=1}^{n} \varepsilon_t - \frac{1}{n} \sum_{t=1}^{n} (X_{t-1} - \vartheta \circ X_{t-1}) \xrightarrow{p} \mu_G \text{ under } \mathbb{P}_{1-\frac{h}{n^2},G},$$

which concludes the proof. □

From the previous proposition it is obvious that the asymptotic behavior of $\widehat{h}_n$, in (34), is identical to the asymptotic behavior of,

$$\tilde{h}_n = \frac{2 \sum_{t=1}^{n} 1\{\Delta X_t < 0\}}{\widehat{g}_n(0) \widehat{\mu}_{G,n}}.$$

This implies that estimation of h in the semiparametric experiments $(\mathcal{E}_n)_{n \in \mathbb{N}}$ is not harder than estimation of h in $(\mathcal{E}_n(G))_{n \in \mathbb{N}}$. In semiparametric parlor: the semiparametric problem is adaptive to $\mathcal{G}$. The precise statement is given in the following corollary; the proof is trivial.

Corollary 4.2 *If $(T_n)_{n \in \mathbb{N}}$ is a sequence of estimators in the semiparametric sequence of experiments $(\mathcal{E}_n)_{n \in \mathbb{N}}$ such that $\mathcal{L}\left(T_n \mid \mathbb{P}_{1-h/n^2,G}\right) \to Z_{h,G}$ with $\int z \, \mathrm{d}Z_{h,G}(z) = h$ for all $(h, G) \in [0, \infty) \times \mathcal{G}$, then we have*

$$\int (z-h)^2 \, \mathrm{d}Z_{h,G}(z) \geq \frac{2h}{g(0)\mu_G} \text{ for all } (h, G) \in [0, \infty) \times \mathcal{G}.$$

The estimator $\tilde{h}_n$ satisfies the conditions and achieves the variance bound.

4.2 Efficient estimation in the global model in case G is known

For convenience we introduce $\mathcal{X}_n = \mathbb{Z}_+^{n+1}$ and $\mathcal{A}_n = 2^{\mathbb{Z}_+^{n+1}}$, and the following assumption.

Assumption 4.1 *A probability distribution G on $\mathbb{Z}_+$ is said to satisfy Assumption 4.1 if $g(k) > 0$ for all $k \in \mathbb{Z}_+$, g is eventually decreasing, i.e. there exists $M \in \mathbb{N}$ such that $g(k+1) \leq g(k)$ for $k \geq M$ and $\mathbb{E}_G \varepsilon_1^4 < \infty$.*

So far we considered nearly unstable INAR experiments. This section considers global experiments for the case G known, i.e. $\mathcal{D}_n(G) = \left(\mathcal{X}_n, \mathcal{A}_n, \left(\mathbb{P}_\theta^{(n)} \mid \theta \in [0,1]\right)\right)$, $n \in \mathbb{N}$. The goal is to estimate the autoregression parameter θ efficiently.

The 'stable' sequence of experiments $\mathcal{D}_n^{(0,1)}(G) = \left(\mathcal{X}_n, \mathcal{A}_n, \left(\mathbb{P}_\theta^{(n)} \mid \theta \in (0,1)\right)\right)$, $n \in \mathbb{N}$, is analyzed by Drost et al. (2006). Under Assumption 4.1 it follows from their results that these experiments are of the Local Asymptotic Normal form (at $\sqrt{n}$-rate). Recall that an estimator T_n of θ is regular if for all $\theta \in (0,1)$ there exists a law L_θ such that for all $h \in \mathbb{R}$,

$$\mathcal{L}\left(\sqrt{n}\left(T_n - \left(\theta + \frac{h}{\sqrt{n}}\right)\right) \mid \mathbb{P}_{\theta+h/\sqrt{n}}\right) \to L_\theta,$$

i.e. vanishing perturbations do not influence the limiting distribution (or more accurately: the associated estimators in the local limit experiment are location-equivariant). For LAN experiments, the Hájek-Le Cam convolution theorem tells us that for every regular estimator T_n of θ we have: $L_\theta = \mathrm{N}(0, I_\theta^{-1}) \oplus \Delta_{\theta,(T_n)}$, where $I_\theta > 0$ (which does not depend on the estimator, and thus is unavoidable noise) is the Fisher-information (see Drost et al. (2006) for the formula). Since $\Delta_{\theta,(T_n)}$ is additional noise, one calls a regular estimator efficient if $\Delta_{\theta,(T_n)}$ is degenerated at $\{0\}$. Drost et al. (2006) provide an (computationally attractive) efficient estimator of θ by updating the OLS estimator into an efficient estimator. Let us recall this estimator. Let $\hat{\theta}_n^*$ be a discretized version of $\hat{\theta}_n^{\mathrm{OLS}}$ (for $n \in \mathbb{N}$ make a grid of intervals with lengths $1/\sqrt{n}$, over $\mathbb{R}$ and, given $\hat{\theta}_n^{\mathrm{OLS}}$, define $\hat{\theta}_n^*$ to be the midpoint of the interval into which $\hat{\theta}_n^{\mathrm{OLS}}$ falls). Then,

$$\theta_n^{(0,1)} = \hat{\theta}_n^* + \frac{1}{n}\sum_{t=1}^n \hat{I}_{\theta,n}^{-1}\dot{\ell}_\theta(X_{t-1}, X_t; \hat{\theta}_n^*, G), \tag{36}$$

where, for $\theta \in (0,1)$,

$$\dot{\ell}_\theta(x_{t-1}, x_t; \theta, G) = \mathbb{E}_{\theta,G}\left[\frac{\vartheta \circ X_{t-1} - \theta X_{t-1}}{\theta(1-\theta)} \mid X_t = x_t, X_{t-1} = x_{t-1}\right] = \frac{\sum_{k=0}^{x_{t-1}}(k - \theta x_{t-1})\,\mathrm{b}_{x_{t-1},\theta}(k)g(x_t - k)}{\theta(1-\theta)P^\theta_{x_{t-1},x_t}},$$

and,

$$\hat{I}_{n,\theta} = \frac{1}{n}\sum_{t=1}^n \dot{\ell}_\theta^2(X_{t-1}, X_t; \hat{\theta}_n^*, G),$$

is an efficient estimator of θ in the sequence of experiments $\mathcal{D}_n^{(0,1)}(G)$, $n \in \mathbb{N}$.

The difference between $\mathcal{D}_n^{(0,1)}(G)$ and $\mathcal{D}_n(G)$ is that in $\mathcal{D}_n(G)$ the full parameter space is used. To consider estimation in the full model, we also need to consider the local asymptotic structure of $\mathcal{D}_n(G)$ at $\theta = 0$ and $\theta = 1$. For $\theta = 1$ we have already done this by determining the limit experiment of $(\mathcal{E}_n(G))_{n\in\mathbb{N}}$. The next proposition shows that for $\theta = 0$ the situation is standard: we have the LAN-property.

Proposition 4.6 *Suppose G satisfies Assumption 4.1. Then $(\mathcal{D}_n(G))_{n\in\mathbb{N}}$ has the LAN-property at $\theta = 0$,* *i.e. for $h \geq 0$ we have,*

$$\sum_{t=1}^{n} \log \frac{P^{h/\sqrt{n}}_{X_{t-1},X_t}}{P^{0}_{X_{t-1},X_t}} = \sum_{t=1}^{n} \log \frac{P^{h/\sqrt{n}}_{X_{t-1},X_t}}{g(X_t)} = hS_n^0 - \frac{h^2}{2} I_0 + o(\mathbb{P}_0; 1), \tag{37}$$

where,

$$I_0 = (\sigma_G^2 + \mu_G^2)\mathbb{E}_G \left(\frac{g(\varepsilon_1) - g(\varepsilon_1 - 1)}{g(\varepsilon_1)} \right)^2,$$
$$S_n^0 = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} -X_{t-1} \left(\frac{g(X_t) - g(X_t - 1)}{g(X_t)} \right) \xrightarrow{d} \mathrm{N}(0, I_0) \text{ under } \mathbb{P}_0.$$

PROOF:
Note first that under $\mathbb{P}_0$ we have $X_t = \varepsilon_t$. Since we are localizing at $\theta = 0$, the following representation of the transition probabilities is convenient, $P^{\theta}_{x_{t-1},x_t} = \sum_{k=0}^{x_{t-1}} \mathrm{b}_{x_{t-1},\theta}(k) g(x_t - k)$. Using the inequality $\log((a+b)/c) - \log(a/c) \leq b/a$ for $a, c > 0$, $b \geq 0$ we obtain, for $h > 0$,

$$\log \frac{\mathrm{d}\mathbb{P}^{(n)}_{\frac{h}{\sqrt{n}}}}{\mathrm{d}\mathbb{P}^{(n)}_0} = \log \frac{P^{h/\sqrt{n}}_{X_{t-1},X_t}}{g(X_t)} - \log \frac{\sum_{k=0}^{2} \mathrm{b}_{X_{t-1},\frac{h}{\sqrt{n}}}(k) g(X_t - k)}{g(X_t)} \leq R_t = \frac{\sum_{k=3}^{X_{t-1}} \mathrm{b}_{X_{t-1},\frac{h}{\sqrt{n}}}(k) g(X_t - k)}{\sum_{k=0}^{2} \mathrm{b}_{X_{t-1},\frac{h}{\sqrt{n}}}(k) g(X_t - k)}.$$

On the event $A_n = \{\forall t \in \{1, \ldots, n\} : h\varepsilon_t < \sqrt{n}\}$ we have for some constant $K \geq 0$, using (43) and the assumption that G is eventually decreasing,

$$R_t \leq \frac{2K\, \mathrm{b}_{X_{t-1},\frac{h}{\sqrt{n}}}(3)}{\left(1 - \frac{h}{\sqrt{n}}\right)^{X_{t-1}}} \leq \frac{Kh^3 X_{t-1}^3}{3n\sqrt{n}\left(1 - \frac{h}{\sqrt{n}}\right)^3}.$$

Using $\mathbb{E}_G \varepsilon_1^3 < \infty$ and Markov's inequality, it is easy to see that $\lim_{n\to\infty} \mathbb{P}_0(A_n^c) = 0$. From this it easily follows that $\sum_{t=1}^{n} R_t \xrightarrow{p} 0$ under $\mathbb{P}_0$. We decompose,

$$L_{tn} = \log \frac{\sum_{k=0}^{2} \mathrm{b}_{X_{t-1},\frac{h}{\sqrt{n}}}(k) g(X_t - k)}{g(X_t)} = (X_{t-1} - 2) \log \left(1 - \frac{h}{\sqrt{n}}\right) + \log\left(1 + A_n + B_{tn} + C_{tn}\right),$$

where,

$$A_n = -\frac{2h}{\sqrt{n}} + \frac{h^2}{n}, \quad B_{tn} = \frac{h}{\sqrt{n}} X_{t-1} \left(1 - \frac{h}{\sqrt{n}}\right) \frac{g(X_t - 1)}{g(X_t)}, \quad C_{tn} = \frac{X_{t-1}(X_{t-1} - 1)}{2} \frac{h^2}{n} \frac{g(X_t - 2)}{g(X_t)}.$$

From here on the proof continues in the classical way. Using the Taylor expansion $\log(1+x) = x - x^2/2 + x^2 r(x)$, where r satisfies $r(x) \to 0$ as $x \to 0$, we make the decomposition,

$$\log(1 + A_n + B_{tn} + C_{tn}) = A_n + B_{tn} + C_{tn} - \frac{1}{2}\left(A_n^2 + B_{tn}^2 + C_{tn}^2 + 2A_nB_{tn} + 2A_nC_{tn} + 2B_{tn}C_{tn}\right) + R_{tn},$$

where $R_{tn} = (A_n + B_{tn} + C_{tn})^2\, r(A_n+B_{tn}+C_{tn})$. It is easy to see that the terms $\sum_{t=1}^n C_{tn}^2$, $\sum_{t=1}^n B_{tn}C_{tn}$ and $\sum_{t=1}^n A_nC_{tn}$ are all $o(\mathbb{P}_0; 1)$. Furthermore, we have,

$$\sum_{t=1}^n \left\{(X_{t-1} - 2)\log\left(1 - \frac{h}{\sqrt{n}}\right) + A_n - \frac{1}{2}A_n^2\right\} = -\frac{h}{\sqrt{n}}\sum_{t=1}^n X_{t-1} - \frac{h^2}{2n}\sum_{t=1}^n X_{t-1} + o(\mathbb{P}_0; 1),$$

and,

$$-\frac{h}{\sqrt{n}}\sum_{t=1}^n X_{t-1} + \sum_{t=1}^n B_{tn} = hS_n^0 - \frac{h^2}{n}\sum_{t=1}^n X_{t-1}\frac{g(X_t-1)}{g(X_t)}.$$

Combining the previous displays we obtain,

$$L_{tn} = hS_n^0 + \sum_{t=1}^n \left\{C_{tn} - \frac{1}{2}B_{tn}^2 - A_nB_{tn} - \frac{h^2}{2n}X_{t-1} - \frac{h^2}{n}X_{t-1}\frac{g(X_t-1)}{g(X_t)}\right\} + R_{tn} + o(\mathbb{P}_0; 1).$$

By the law of large numbers we have (note that $\mathbb{E}_0 g(X_t - i)/g(X_t) = 1$, $i = 1, 2$), under $\mathbb{P}_0$, $\sum_{t=1}^n C_{tn} \xrightarrow{p} h^2\left(\sigma_G^2 + \mu_G^2 - \mu_G\right)/2$, $\sum_{t=1}^n A_nB_{tn} \xrightarrow{p} -2h^2\mu_G$, $(1/n)\sum_{t=1}^n X_{t-1} \xrightarrow{p} \mu_G$, $(1/n)\sum_{t=1}^n X_{t-1}g(X_t - 1)/g(X_t) \xrightarrow{p} \mu_G$, and $\sum_{t=1}^n B_{tn}^2 \xrightarrow{p} h^2(\sigma_G^2 + \mu_G^2)\mathbb{E}_G(g^2(\varepsilon_1 - 1)/g^2(\varepsilon_1))$. Thus, once we show that $\sum_{t=1}^n R_{tn} = o(\mathbb{P}_0; 1)$ the proposition is proved. Using the inequality $(x+y+z)^2 \leq 9(x^2+y^2+z^2)$ we easily obtain $\sum_{t=1}^n (A_n + B_{tn} + C_{tn})^2 = O(\mathbb{P}_0; 1)$. And using Markov's inequality it is easy to see that, for $\epsilon > 0$, $\mathbb{P}_0\{\max_{1\leq t\leq n}|A_n + B_{tn} + C_{tn}| > \epsilon\} \leq \sum_{t=1}^n \mathbb{P}_0\{|A_n + B_{tn} + C_{tn}| > \epsilon\} \to 0$. Thus $\sum_{t=1}^n (A_n + B_{tn} + C_{tn})^2 r(A_n + B_{tn} + C_{tn}) \xrightarrow{p} 0$ under $\mathbb{P}_0$, which concludes the proof. □

Remark 5 *(i) The meaning of this LAN-result is that the sequence,* $\left(\mathcal{X}_n, \mathcal{A}_n, \left(\mathbb{P}_{h/\sqrt{n}}^{(n)} \mid h \geq 0\right)\right)$, $n \in \mathbb{N}$, *of local experiments, converges to the experiment* $\left((\mathbb{R}, \mathcal{B}(\mathbb{R}), \left(\mathrm{N}\left(h, I_0^{-1}\right) \mid h \geq 0\right)\right)$. *(ii) Note that we are dealing here with a 'one-sided' LAN-result, i.e. we only consider h positive. As a consequence, it is not possible to apply the standard results for experiments with the LAN-structure directly (this, since these are formulated for interior points of the parameter space). Since we do not want to discuss this issue further, we consider asymptotically centered estimators with minimal asymptotic variance as a best estimator at $\theta = 0$ (see below). (iii) The 'information-loss principle', which is used in Drost et al. (2006) to establish the LAN-property for $\theta \in (0,1)$, cannot be used here since the score of a Binomial distribution does not exist (in the usual sense) at $\theta = 0$.*

Now we completed the picture of the local asymptotic structures of $(\mathcal{D}_n(G))_{n\in\mathbb{N}}$ we can discuss efficient estimation. First, we describe the class of estimators in which we are interested. We consider estimators T_n that satisfy,

(i) $(\theta = 0)$ for all $h \geq 0$,

$$\mathcal{L}\left(\sqrt{n}\left(T_n - \frac{h}{\sqrt{n}}\right) \mid \mathbb{P}_{h/\sqrt{n}}\right) \to L_h, \text{ with } \int z \,\mathrm{d}L_h(z) = 0, \tag{38}$$

(ii) $(0 < \theta < 1)$ T_n is regular, i.e. for all $h \in \mathbb{R}$,

$$\mathcal{L}\left(\sqrt{n}\left(T_n - \left(\theta + \frac{h}{\sqrt{n}}\right)\right) \mid \mathbb{P}_{\theta + h/\sqrt{n}}\right) \to L_\theta, \tag{39}$$

(ii) $(\theta = 1)$ for all $h \geq 0$,

$$\mathcal{L}\left(n^2\left(T_n - \left(1 - \frac{h}{n^2}\right)\right) \mid \mathbb{P}_{1-h/n^2}\right) \to R_h \text{ with } \int z \,\mathrm{d}R_h(z) = 0. \tag{40}$$

So for $\theta \in (0,1)$ we ask for regularity which we discussed earlier. For $\theta = 0$ and $\theta = 1$ we only ask for a limiting distribution with mean zero. For any such estimator we have (the first inequality follows by arguments completely analogue to the derivation of the third inequality, we already discussed the second statement, and the third follows from Proposition 4.3 by taking $\widehat{h}_n = n^2(1 - T_n)$ as estimator of h),

$$\int z^2 \,\mathrm{d}L_h(z) \geq I_0^{-1}, \quad L_\theta = \mathrm{N}(0, I_\theta^{-1}) \oplus \Delta_{\theta,(T_n)}, \quad \int z^2 \,\mathrm{d}R_h(z) \geq \frac{2h}{g(0)\mu_G} \text{ for all } h \geq 0,\, \theta \in (0,1). \tag{41}$$

Hence it is natural to call an estimator in the global model efficient if it satisfies (38)-(40) with $L_\theta = \mathrm{N}(0, I_\theta^{-1})$, $\int z^2 \,\mathrm{d}L_h(z) = I_0^{-1}$, and $\int z^2 \,\mathrm{d}R_h(z) = 2h/g(0)\mu_G$ for all $h \geq 0$, $\theta \in (0,1)$.

Proposition 4.7 *Suppose G satisfies Assumption 4.1. Let $\alpha, \beta \in (0, 1/2)$, and $c_\alpha, c_\beta > 0$. The estimator,*

$$\widehat{\theta}_n = \theta_n^0 1\left\{\left|\widehat{\theta}_n^{OLS}\right| \leq c_\alpha n^{-\alpha}\right\} + \theta_n^{(0,1)} 1\left\{\left|\widehat{\theta}_n^{OLS}\right| > c_\alpha n^{-\alpha},\, |\widehat{\theta}_n^{OLS} - 1| > c_\beta n^{-\beta}\right\} + \theta_n^1 1\left\{|\widehat{\theta}_n^{OLS} - 1| \leq c_\beta n^{-\beta}\right\},$$

where $\theta_n^{(0,1)}$ is defined in (36) and,

$$\theta_n^0 = \frac{1}{\sqrt{n}} I_0^{-1} S_n^0, \quad \theta_n^1 = 1 - \frac{2\sum_{t=1}^n 1\{\Delta X_t < 0\}}{n^2 g(0)\mu_G},$$

is an efficient estimator of θ in the sequence of experiments $(\mathcal{D}_n(G))_{n\in\mathbb{N}}$.

PROOF:

From Le Cam's third lemma and Proposition 4.6 it easily follows that θ_n^0 satisfies (38) and attains its variance lower-bound in (41). Since $\theta_n^{(0,1)}$ is an efficient estimator in the 'stable experiments' $(\mathcal{D}_n^{(0,1)}(G))_{n\in\mathbb{N}}$ it follows, by definition, that $\theta_n^{(0,1)}$ satisfies (39) and attains the convolution lower-bound in (41). And it is also clear (from Proposition 4.1) that θ_n^1 satisfies (40) and attains its variance lower-bound in (41). Thus it suffices to show that $\sqrt{n}\left(\widehat{\theta}_n - \theta_n^0\right) \xrightarrow{p} 0$ under $\mathbb{P}_{h/\sqrt{n}}$ for all $h \geq 0$, $\sqrt{n}\left(\widehat{\theta}_n - \theta_n^{(0,1)}\right) \xrightarrow{p} 0$ under $\mathbb{P}_{\theta+h/\sqrt{n}}$ for all $\theta \in (0,1)$, $h \in \mathbb{R}$, and $n^2\left(\widehat{\theta}_n - \theta_n^1\right) \xrightarrow{p} 0$ under $\mathbb{P}_{1-h/n^2}$ for all $h \geq 0$. It is an easy exercise, using a martingale central limit theorem, to show that $\sqrt{n}\left(\widehat{\theta}_n^{\mathrm{OLS}} - (\theta + h/\sqrt{n})\right)$ converges to a normal distribution under $\mathbb{P}_{\theta+h/\sqrt{n}}$ for all $\theta \in [0,1)$ and $h \in \mathbb{R}$ (for $\theta = 0$ we only consider $h \geq 0$). And from (31) we have that $n^{3/2}(\widehat{\theta}_n^{\mathrm{OLS}} - (1 - h/n^2))$ converges to a normal distribution under $\mathbb{P}_{1-h/n^2}$ for $h \geq 0$. This implies that $n^\alpha \widehat{\theta}_n^{\mathrm{OLS}} \xrightarrow{p} 0$ under $\mathbb{P}_{h/\sqrt{n}}$ and $\mathbb{P}_{h/\sqrt{n}}\left\{\left|\widehat{\theta}_n^{\mathrm{OLS}} - 1\right| \leq c_\beta n^{-\beta}\right\} \to 0$ for $h \geq 0$, $n^\alpha \widehat{\theta}_n^{\mathrm{OLS}} \xrightarrow{p} \infty$ and $n^\beta \left|\widehat{\theta}_n^{\mathrm{OLS}} - 1\right| \xrightarrow{p} \infty$ under $\mathbb{P}_{\theta+h/\sqrt{n}}$, for $\theta \in (0,1)$, $h \in \mathbb{R}$, and we have $n^\beta\left(\widehat{\theta}_n^{\mathrm{OLS}} - 1\right) \xrightarrow{p} 0$ under $\mathbb{P}_{1-h/n^2}$ and $\mathbb{P}_{1-h/n^2}\left\{\left|\widehat{\theta}_n^{\mathrm{OLS}}\right| \leq c_\alpha n^{-\alpha}\right\} \to 0$ for $h \geq 0$. This concludes the proof. □

4.3 Testing for a unit root

This section discusses testing for a unit root in an INAR(1) model. We consider the case that G is known and satisfies Assumption 3.1.

In the global experiments $\mathcal{D}_n(G) = (\mathcal{X}_n, \mathcal{A}_n, (\mathbb{P}_\theta^{(n)} \mid \theta \in [0,1]))$, $n \in \mathbb{N}$, we want to test the hypothesis $\mathrm{H}_0 : \theta = 1$ versus $\mathrm{H}_1 : \theta < 1$. In other words, we want to test the null hypothesis of a unit root. Hellström (2001) considered this problem, from the perspective that one wants to use standard (that is, OLS) software routines in the testing. He derives, by Monte Carlo simulations, the finite sample null-distributions for a Dickey-Fuller test of a random walk with Poisson distributed errors. This (standard) Dickey-Fuller test statistic is given by the *usual* (i.e. non-corrected) t-test that the slope parameter equals 1, i.e.

$$\tau_n = \frac{\widehat{\theta}_n^{\mathrm{OLS}} - 1}{\sqrt{\sigma_G^2 \left(\sum_{t=1}^n X_{t-1}^2\right)^{-1}}},$$

where $\widehat{\theta}_n^{\mathrm{OLS}}$ is given by (27). Under H_0, i.e. under $\mathbb{P}_1$, we have (we are now dealing with a random walk with drift), $\tau_n \xrightarrow{d} \mathrm{N}(0,1)$. Hence, the size $\alpha \in (0,1)$ Dickey-Fuller test rejects H_0 if and only if $\tau_n < \Phi^{-1}(\alpha)$. To analyze the performance of a test, one needs to consider the local asymptotic behavior of the test. Since $\mathcal{E}_n(G) \to \mathcal{E}(G)$ we should consider the performance of τ_n along the sequence $\mathcal{E}_n(G)$. The following proposition shows, however, that the asymptotic probability that the null hypothesis is rejected equals α for all alternatives. Hence, the standard Dickey-Fuller test has no power.

Proposition 4.8 *If $\mathbb{E}_G\varepsilon_1^4 < \infty$ we have for all $h \geq 0$,*

$$\tau_n \xrightarrow{d} \mathrm{N}(0,1), \text{ under } \mathbb{P}_{1-\frac{h}{n^2}},$$

which yields

$$\lim_{n\to\infty} \mathbb{P}_{1-\frac{h}{n^2}}(\textit{reject } \mathrm{H}_0) = \alpha.$$

PROOF:
From (29) and (31) the result easily follows. □

So the standard Dickey-Fuller test for a unit root does not behave well in the nearly unstable INAR(1) setting. In our sequence of experiments $\mathcal{E}_n(G)$, $n \in \mathbb{N}$, we propose the intuitively obvious tests

$$\psi_n(X_0,\ldots,X_n) = \begin{cases} \alpha, & \text{if } \sum_{t=1}^n 1\{\Delta X_t < 0\} = 0, \\ 1, & \text{if } \sum_{t=1}^n 1\{\Delta X_t < 0\} \geq 1, \end{cases}$$

i.e. reject H_0 if the process ever moves down and reject H_0 with probability α if there are no downward movements. We will see that this obvious test is in fact efficient.

To discuss efficiency of tests, we recall the implication of the Le Cam-Van der Vaart asymptotic representation theorem to testing (see Theorem 7.2 in Van der Vaart (1991)). Let $\alpha \in (0,1)$ and ϕ_n be a sequence of tests in $(\mathcal{E}_n(G))_{n\in\mathbb{N}}$ such that $\limsup_{n\to\infty} \mathbb{E}_1\phi_n(X_0,\ldots,X_n) \leq \alpha$. Then we have

$$\limsup_{n\to\infty} \mathbb{E}_{1-\frac{h}{n^2}}\phi_n(X_0,\ldots,X_n) \leq \sup_{\phi\in\Phi_\alpha} \mathbb{E}_h\phi(Z) \text{ for all } h > 0,$$

where Φ_α is the collection of all level α tests for testing $\mathrm{H}_0 : h = 0$ versus $\mathrm{H}_1 : h > 0$ in the Poisson limit experiment $\mathcal{E}(G)$. If we have equality in the previous display, it is natural to call a test ϕ_n efficient. It is obvious that the uniform most powerful test in the Poisson limit experiment is given by

$$\phi(Z) = \begin{cases} \alpha, & \text{if } Z = 0, \\ 1, & \text{if } Z \geq 1. \end{cases}$$

Its power function is given by $\mathbb{E}_0\phi(Z) = \alpha$ and $\mathbb{E}_h\phi(Z) = 1-(1-\alpha)\exp(-hg(0)\mu_G/2)$. Using Theorem 2.1 we find

$$\lim_{n\to\infty} \mathbb{E}_1\psi_n(X_0,\ldots,X_n) = \alpha, \text{ and } \lim_{n\to\infty} \mathbb{E}_{1-\frac{h}{n^2}}\psi_n(X_0,\ldots,X_n) = 1-(1-\alpha)\exp\left(-\frac{hg(0)\mu_G}{2}\right) \text{ for } h > 0.$$

We conclude that the test ψ_n is efficient.

A Auxiliaries

The following result is basic (see, for instance, Feller (1968) pages 150-151), but since it is heavily applied, we state it here for convenience.

Proposition A.1 *Let $m \in \mathbb{N}$, $p \in (0,1)$. If $r > mp$, we have*

$$\sum_{k=r}^{m} \mathrm{b}_{m,p}(k) \leq \mathrm{b}_{m,p}(r) \frac{r(1-p)}{r-mp}. \tag{42}$$

So, if $1 > mp$, we have for $r = 2, 3$,

$$\sum_{k=r}^{m} \mathrm{b}_{m,p}(k) \leq 2\, \mathrm{b}_{m,p}(r). \tag{43}$$

For convenience we recall Theorem 1 in Serfling (1975).

Lemma A.1 *Let $Z_1, \ldots, Z_n$ (possibly dependent) 0-1 valued random variables and set $S_n = \sum_{t=1}^{n} Z_t$. Let Y be Poisson distributed with mean $\sum_{t=1}^{n} \mathbb{E} Z_t$. Then we have*

$$\sup_{A \subset \mathbb{Z}_+} |\mathbb{P}\{S_n \in A\} - \mathbb{P}\{Y \in A\}| \leq \sum_{t=1}^{n} (\mathbb{E} Z_t)^2 + \sum_{t=1}^{n} \mathbb{E} \left| \mathbb{E}[Z_t \mid Z_1, \ldots, Z_{t-1}] - \mathbb{E} Z_t \right|.$$

References

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

Chan, N. and C. Wei (1987). Asymptotic inference for nearly nonstationary AR(1) processes. *Ann. Statist.*, **15**, 1050–1063.

Drost, F., R. van den Akker, and B. Werker (2006). Local asymptotic normality and efficient estimation in parametric INAR(p) models. *Working paper*.

Feller, W. (1968). *An introduction to probability theory and its applications: volume I* (3rd ed.). John Wiley & Sons, New York.

Hall, P. and C. Heyde (1980). *Martingale limit theory and its applications.* Academic Press, New York.

Hellström, J. (2001). Unit root testing in integer-valued AR(1) models. *Economics Letters*, **70**, 9–14.

Ispány, M., G. Pap, and M. van Zuijlen (2003). Asymptotic inference for nearly unstable INAR(1) models. *J. Appl. Probab.*, **40**, 750–765.

Jeganathan, P. (1995). Some aspects of asymptotic theory with applications to time series models. *Econometric Theory*, **11**, 818–887.

Philips, P. (1987). Towards a unified asymptotic theory for autoregression. *Biometrika*, **74**, 535–547.

Serfling, R. (1975). A general Poisson approximation theorem. *Ann. Probab.*, **3**, 726–731.

Steutel, F. and K. van Harn (1979). Discrete analogues of self-decomposability and stability. *Ann. Probab.*, **7**, 893–899.

Van der Vaart, A. (1991). An asymptotic representation theorem. *Internat. Statist. Rev.*, **59**, 97–121.

Van der Vaart, A. (2000). *Asymptotic Statistics* (1 ed.). Cambridge University Press, Cambridge.

Wefelmeyer, W. (1999). Efficient estimation in Markov chain models: an introduction. In: S. Ghosh (Ed.), *Asymptotics, Nonparametrics, and Time Series*, pp. 427–459. Statistics: Textbooks and Monographs 158, Dekker, New York 1999.