CentER

Discussion Paper

No. 2006–45

LOCAL ASYMPTOTIC NORMALITY AND EFFICIENT ESTIMATION FOR INAR (P) MODELS

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

April 2006

ISSN 0924-7815

TILBURG UNIVERSITY

Local Asymptotic Normality and efficient estimation for INAR(p) models

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

April 28, 2006

Abstract

Integer-valued autoregressive (INAR) processes have been introduced to model nonnegative integer-valued phenomena that evolve in time. The distribution of an INAR(p) process is determined by two parameters: a vector of survival probabilities and a probability distribution on the nonnegative integers, called an immigration or innovation distribution. This paper provides an efficient estimator of the parameters, and in particular, shows that the INAR(p) model has the Local Asymptotic Normality property.

Key words: count data, integer-valued time series, information loss structure

JEL: C12, C13, C19

1 Introduction

The INAR(1) process has been introduced by Al-Osh and Alzaid (1987) to model nonnegative integer-valued phenomena that evolve in time. The INAR(1) process is defined by the recursion,

$$X_t = \vartheta \circ X_{t-1} + \varepsilon_t, \quad t \in \mathbb{Z}_+ = \mathbb{N} \cup \{0\}, \tag{1}$$

where,

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

Here $\left(Z_j^{(t)}\right)_{j\in\mathbb{N},t\in\mathbb{Z}_+}$ is a collection of i.i.d. Bernoulli variables with success probability $\theta \in (0,1)$, independent of the i.i.d. innovation sequence $(\varepsilon_t)_{t\in\mathbb{Z}_+}$ with distribution G on $\mathbb{Z}_+$. Finally, the starting value X_{-1},

[*]Econometrics group, CentER, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands.

with distribution ν on $\mathbb{Z}_+$, is independent of $(\varepsilon_t)_{t\in\mathbb{Z}_+}$ and $(Z_j^{(t)})_{j\in\mathbb{N},t\in\mathbb{Z}_+}$. Equation (1) can be interpreted as a branching process with immigration. The outcome X_t is composed of the surviving elements of X_{t-1} during the period $(t-1,t]$, $\vartheta \circ X_{t-1}$, and the number of immigrants during this period, ε_t. Each element of X_{t-1} survives with probability θ and its survival has no effect on the survival of the other elements, nor on the number of immigrants. The more general INAR(p) processes were first introduced by Al-Osh and Alzaid (1990) but Du and Li (1991) proposed a different setup. In the setup by Du and Li (1991) the autocorrelation structure of an INAR(p) process is the same as that of an AR(p) process, whereas it corresponds to the one of an ARMA($p,p-1$) process in the setup by Al-Osh and Alzaid (1990). The setup by Du and Li (1991) has been followed by most authors, and we use their setup as well. The INAR(p) process is an analogue of (1) with p lags. An INAR(p) process is recursively defined by,

$$X_t = \vartheta_1 \circ X_{t-1} + \vartheta_2 \circ X_{t-2} + \cdots + \vartheta_p \circ X_{t-p} + \varepsilon_t, \quad t \in \mathbb{Z}_+, \tag{2}$$

where, for $i = 1,\ldots,p$,

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

Here $Z_j^{(t,i)}$, $i \in \{1,\ldots,p\}$, $j \in \mathbb{N}$, $t \in \mathbb{Z}_+$, are independent Bernoulli distributed variables, where $Z_j^{(t,i)}$ has success probability $\theta_i \in (0,1)$, independent of the $\mathbb{Z}_+$-valued i.i.d. G-distributed innovations $(\varepsilon_t)_{t\in\mathbb{Z}_+}$. The starting value $(X_{-1},\ldots,X_{-p})'$ is independent of $(\varepsilon_t)_{t\in\mathbb{Z}_+}$ and $(Z_j^{(t,i)})_{i\in\{1,\ldots,p\},j\in\mathbb{N},t\in\mathbb{Z}_+}$, and has distribution ν on $\mathbb{Z}_+^p$. The corresponding probability space is denoted by $(\Omega,\mathcal{F},\mathbb{P}_{\nu,\theta,G})$.

Applications of INAR(p) processes in the medical sciences can be found in, for example, Franke and Seligmann (1993) and Cardinal et al. (1999); applications to economics in, for example, Böckenholt (1999), Berglund and Brännäs (2001), Brännäs and Hellström (2001), Rudholm (2001), Böckenholt (2003), Brännäs and Shahiduzzaman (2004), Freeland and McCabe (2004), and Gourieroux and Jasiak (2004).

In this paper we allow for parametric INAR(p) models, i.e., G belongs to a parametric class of distributions, say $(G_\alpha | \alpha \in A \subset \mathbb{R}^q)$. Estimators of the parameters are provided by several authors. For $p = 1$ and $G_\alpha = \text{Poisson}(\alpha)$, Franke and Seligmann (1993) analyzed maximum likelihood. Du and Li (1991) derived the limit-distribution of the OLS-estimator of θ. Brännäs and Hellström (2001) considered GMM estimation, and Silva and Oliveira (2005) proposed a frequency domain based estimator of θ. In this paper we are interested in asymptotic efficient estimation of the parameters in an INAR(p) model. In Appendix C, we review the modern notion of asymptotic efficiency: the Hájek-Le Cam convolution theorem gives a lower-bound to the accuracy of regular estimators for experiments which have the Local Asymptotic Normality (LAN) structure. An estimator is called efficient if it is regular and attains this lower-bound. Hence, once we have established the LAN-property, the next goal is to construct an estimator which attains the lower-bound. Maximum likelihood is, in general, not feasible for INAR(p) models, since it is not clear which optimization routines could be used to determine (a point close to) a maximum location of the likelihood. This is caused by the complicated nature of the transition-probabilities. Therefore, we will not try to establish efficiency of the maximum likelihood estimator. Instead, we provide a one-step method, which updates an initial $\sqrt{n}$-consistent estimator into an efficient one. This yields a computationally attractive and statistically efficient estimator.

The setup of the paper is as follows. In Section 2 we discuss some preliminary properties of INAR(p) processes: the existence of a stationary distribution and existence of moments. Section 3 introduces two models. In the first model the immigration distribution G is completely known and the vector of survival-probabilities, $\theta = (\theta_1, \ldots, \theta_p)$, is the only unknown parameter. The goal is to estimate θ efficiently. In the second model, G belongs to a parametric class of distributions, say $(G_\alpha | \alpha \in A \subset \mathbb{R}^q)$. For this model, the goal is to estimate the parameter (θ, α) efficiently. In Section 3.1 we establish the LAN-property, that yields, by the Hájek-Le Cam convolution theorem, a lower-bound to the accuracy of regular estimators. The proof of the LAN-property is facilitated by a certain representation of the transition-scores, which is explained by an information-loss interpretation of the model. Section 3.2 provides efficient estimators. A one-step update estimator is proposed, that, besides being efficient, is also computationally attractive (compared to maximum likelihood). The proofs can be found in Appendix B. Appendix A contains some auxiliary results. As already mentioned, Appendix C gives a short review of the modern notion of asymptotic efficiency for parametric models.

2 Preliminary results

This section discusses some preliminary properties of INAR(p) processes. Throughout the paper the number of lags, $p \in \mathbb{N}$, is fixed. The following notation is used: $\mathcal{G}$ denotes the set of all probability measures on $\mathbb{Z}_+ = \mathbb{N} \cup \{0\}$. The Binomial distribution with parameters $\theta \in [0,1]$ and $n \in \mathbb{Z}_+$ is denoted by $\text{Bin}_{n,\theta}$ ($\text{Bin}_{0,\theta}$ is the Dirac-measure concentrated in 0), $\text{b}_{n,\theta}$ denotes the corresponding point mass function, and δ_x denotes the Dirac measure concentrated in x. In general, we denote a probability measure on $\mathbb{Z}_+$ by a capital, and denote the associated probability mass function by the corresponding lower case. For $G \in \mathcal{G}$, μ_G denotes the mean of G, and σ_G^2 denotes its variance. As usual $\mathbb{E}_{\nu,\theta,G}\left(\cdot\right)$ is shorthand for $\int (\cdot)\, \mathrm{d}\mathbb{P}_{\nu,\theta,G}$. For (probability) measures F and G, $F * G$ denotes the convolution of F and G. Finally, $\mathbb{F} = (\mathcal{F}_t)_{t \geq -p}$ is the natural filtration generated by X, i.e. $\mathcal{F}_t = \sigma\left(X_{-p}, \ldots, X_t\right)$. Note that, contrary to classical AR(p) processes, $\mathcal{F}_t \neq \sigma\left(X_{-p}, \ldots, X_{-1}, \varepsilon_0, \ldots, \varepsilon_t\right)$.

We compute the first two conditional moments of an INAR(p) process to gain some insight in its dependence structure. It immediately follows from (2) that, for $t \in \mathbb{Z}_+$,

$$\mathbb{E}_{\theta,G}\left[X_t \mid \mathcal{F}_{t-1}\right] = \mathbb{E}_{\theta,G}\left[X_t \mid X_{t-1},\dots,X_{t-p}\right] = \mu_G + \sum_{i=1}^{p}\theta_i X_{t-i} \in [0,\infty],$$

$$\mathrm{Var}_{\theta,G}\left[X_t \mid \mathcal{F}_{t-1}\right] = \mathrm{Var}_{\theta,G}\left[X_t \mid X_{t-1},\dots,X_{t-p}\right] = \sigma_G^2 + \sum_{i=1}^{p}\theta_i(1-\theta_i) X_{t-i} \in [0,\infty].$$

Hence an INAR(p) process has the same auto-regression function as an AR(p) process. However, an INAR(p) process has conditional heteroskedasticity of autoregressive form, whereas the conditional variance is constant for AR(p) processes. By straightforward calculations one can see that an INAR(p) process has the same autocorrelation structure as an AR(p) process.

Next we determine the conditional distribution of X_t given $\mathcal{F}_{t-1}$. From (2) it follows, for $t \in \mathbb{Z}_+$,

$$\mathbb{P}_{\nu,\theta,G}\left\{X_t = x_t \mid \mathcal{F}_{t-1}\right\} = \mathbb{P}_{\nu,\theta,G}\left\{X_t = x_t \mid X_{t-1},\dots,X_{t-p}\right\} = P^{\theta,G}_{(X_{t-1},\dots,X_{t-p}),x_t},$$

where, for $x_{t-p},\dots,x_t \in \mathbb{Z}_+$, the transition-probability $P^{\theta,G}_{(x_{t-1},\dots,x_{t-p}),x_t}$ is given by,

$$\begin{aligned} P^{\theta,G}_{(x_{t-1},\dots,x_{t-p}),x_t} &= \mathbb{P}_{\nu,\theta,G}\left\{\sum_{i=1}^{p}\vartheta_i \circ X_{t-i} + \varepsilon_t = x_t \mid X_{t-1} = x_{t-1},\dots,X_{t-p} = x_{t-p}\right\} \\ &= \left(\mathrm{Bin}_{x_{t-1},\theta_1} * \cdots * \mathrm{Bin}_{x_{t-p},\theta_p} * G\right)\{x_t\}. \end{aligned}$$

Du and Li (1991) and Latour (1998) prove the existence of a (second order) stationary INAR(p) process in case $\mathbb{E}_G\varepsilon_0^2 < \infty$ and $\sum_{i=1}^{p}\theta_i < 1$. Franke and Seligmann (1993) give conditions for the existence of a (strictly) stationary INAR(1) process using generating functions. Dion et al. (1995) give conditions for the existence of (strictly) stationary INAR(p) using multitype branching processes with immigration. Using elementary Markov chain techniques, we give an alternative (shorter) proof under the same conditions.

Theorem 2.1 *For all $G \in \mathcal{G}$ with $g(0) \in [0,1)$, $\mu_G < \infty$, and $\theta \in (0,1)^p$ with $\sum_{i=1}^{p}\theta_i < 1$, there exists a probability measure $\nu_{\theta,G}$ on $\mathbb{Z}_+^p$ such that X is a strictly stationary process under $\mathbb{P}_{\nu_{\theta,G},\theta,G}$. The support of $\nu_{\theta,G}$ is given by $\{\alpha,\alpha+1,\dots\}^p$, where $\alpha = \min\{k \in \mathbb{Z}_+ \mid g(k) > 0\}$.*

Clearly, in case $g(0) = 1$, a strictly stationary solution is given by $X_t = 0$ for all t, i.e. $\nu_{\theta,G} = \delta_0$. The next lemma gives sufficient conditions for the existence of the first three moments of the stationary distribution.

Lemma 2.1 *Let $G \in \mathcal{G}$ with $g(0) \in [0,1)$, $\mu_G < \infty$, and $\theta \in (0,1)^p$ with $\sum_{i=1}^{p}\theta_i < 1$. Then, for $k \in \{1,2,3\}$, $\mathbb{E}_G\varepsilon_0^k < \infty$ if and only if $\mathbb{E}_{\nu_{\theta,G},\theta,G}X_0^k < \infty$.*

3 Efficient estimation

In this section we consider INAR(p) models, where we always restrict ourselves to the stationary parameter regime, i.e., $\theta \in (0,1)^p$ with $\sum_{i=1}^p \theta_i < 1$ (see Drost et al. (2006) for the asymptotic structure of an INAR(1) model at the boundary of the parameter space). In a first model, the immigration-distribution G and the initial distribution ν are completely known, which leads to the following sequence of statistical experiments induced by observing $(X_{-p}, \dots, X_n)$,

$$\mathcal{E}_1^{(n)}(\nu, G) = \left(\mathbb{Z}_+^{n+1+p}, 2^{\mathbb{Z}_+^{n+1+p}}, \left(\mathbb{P}_{\nu,\theta,G}^{(n)} \mid \theta \in \Theta \right) \right), \quad n \in \mathbb{Z}_+,$$

where the initial distribution ν and the immigration distribution $G \in \mathcal{G}$ are fixed, $\Theta = \{\theta \in (0,1)^p \mid \sum_{i=1}^p \theta_i < 1\}$, and $\mathbb{P}_{\nu,\theta,G}^{(n)}$ denotes the law of $(X_{-p}, \dots, X_n)$ on the measurable space $\left(\mathbb{Z}_+^{n+1+p}, 2^{\mathbb{Z}_+^{n+1+p}} \right)$ under $\mathbb{P}_{\nu,\theta,G}$. In this model G is completely known, but we also want to consider the case where G belongs to a parametric model, for example, $G = \text{Poisson}(\alpha)$. So let $A \subset \mathbb{R}^q$ and $\mathcal{G}_A = (G_\alpha)_{\alpha \in A}$ be a family of elements in $\mathcal{G}$, such that $\alpha \mapsto G_\alpha$ is sufficiently smooth (this will be made precise later). We then consider the sequence of experiments, induced by observing $(X_{-p}, \dots, X_n)$,

$$\mathcal{E}_2^{(n)}(\nu, \mathcal{G}_A) = \left(\mathbb{Z}_+^{n+1+p}, 2^{\mathbb{Z}_+^{n+1+p}}, \left(\mathbb{P}_{\nu,\theta,\alpha}^{(n)} \mid \theta \in \Theta, \alpha \in A \right) \right), \quad n \in \mathbb{Z}_+,$$

where, for notational convenience, we abbreviate G_α in sub- and superscripts by α.

The goal of this section is to estimate the parameters in both models efficiently. As mentioned in the introduction, a short review of the modern notion of asymptotic efficiency is given in Appendix C.

3.1 The LAN-property

In this subsection we prove the LAN-property for the sequence of experiments $\mathcal{E}_2^{(n)}(\nu, \mathcal{G}_A)$, $n \in \mathbb{Z}_+$, immediately implying the LAN-property for the sequence of experiments $\mathcal{E}_1^{(n)}(\nu, G)$, $n \in \mathbb{Z}_+$.

Let $\mathcal{G}_A = (G_\alpha | \alpha \in A)$ be a parametric family of immigration-distributions, where A is an open, convex subset of $\mathbb{R}^q$ such that,

(A1) the support of G_α does not depend on α and we have $0 < g_\alpha(0) < 1$;

(A2) for all $e \in \mathbb{Z}_+$ and $\alpha \in A$, the expressions,

$$h_\alpha(e) = \frac{\partial}{\partial \alpha} \log \left(g_\alpha(e) \right) 1_{(0,1]} \left(g_\alpha(e) \right) \in \mathbb{R}^q,$$
$$\dot{h}_\alpha(e) = \frac{\partial^2}{\partial \alpha^T \partial \alpha} \log \left(g_\alpha(e) \right) 1_{(0,1]} \left(g_\alpha(e) \right) \in \mathbb{R}^{q \times q},$$

are defined and, for all $e \in \mathbb{Z}_+$, they are continuous in α;

(A3) for every $(\theta, \alpha) \in \Theta \times A$, there exists $\delta > 0$ and random variables $M_1^{\theta,\alpha}$ and $M_2^{\theta,\alpha}$ such that,

$$\sup_{(\tilde{\theta},\tilde{\alpha}):|(\tilde{\theta},\tilde{\alpha})-(\theta,\alpha)|<\delta} \mathbb{E}_{\tilde{\theta},\tilde{\alpha}} \left[|h_{\tilde{\alpha}}(\varepsilon_0)|^2 \mid X_0, \ldots, X_{-p} \right] \leq M_1^{\theta,\alpha}, \quad \mathbb{E}_{\nu_{\theta,\alpha},\theta,\alpha} M_1^{\theta,\alpha} < \infty, \tag{3}$$

and, for $i, j = 1, \ldots, q$,

$$\sup_{(\tilde{\theta},\tilde{\alpha}):|(\tilde{\theta},\tilde{\alpha})-(\theta,\alpha)|<\delta} \mathbb{E}_{\tilde{\theta},\tilde{\alpha}} \left[\left| \dot{h}_{\tilde{\alpha},ij}(\varepsilon_0) \right| \mid X_0, \ldots, X_{-p} \right] \leq M_2^{\theta,\alpha}, \quad \mathbb{E}_{\nu_{\theta,\alpha},\theta,\alpha} M_2^{\theta,\alpha} < \infty, \tag{4}$$

where $\dot{h}_{\alpha,ij}(e)$ is the (i,j)-entry of the matrix $\dot{h}_\alpha(e)$;

(A4) the information-equality $\mathbb{E}_\alpha h_\alpha h_\alpha^T(\varepsilon_0) = -\mathbb{E}_\alpha \dot{h}_\alpha(\varepsilon_0)$ is satisfied, and $\mathbb{E}_\alpha h_\alpha h_\alpha^T(\varepsilon_0)$ is non-singular and continuous in α;

(A5) $\mathbb{E}_\alpha \varepsilon_0^2 < \infty$;

(A6) $G_\alpha = G_{\alpha'}$ implies $\alpha = \alpha'$.

Remark 1 *It is well-known that Assumptions (A2) and (A4) ensure that $\mathcal{G}_A$ is differentiable in quadratic mean with score $h_\alpha(\varepsilon_0)$ (see, for example, Lemma 7.6 in Van der Vaart (2000)) and consequently $\mathbb{E}_\alpha h(\varepsilon_0) = 0$, see the proof of Theorem 7.2 in Van der Vaart (2000).*

Remark 2 *Assumptions (A1)-(A6) are of the Cramér-type. Conditions (3) and (4) in Assumption A3 are rather awkward. A simple sufficient condition is given by $|h_{\alpha,i}(e)| \leq a_\alpha + c_\alpha e$ and $|\dot{h}_{\alpha,ij}| \leq b_\alpha + d_\alpha e^2$ for $a_\alpha, b_\alpha, c_\alpha$ and d_α that are (locally) bounded in α. Now it is easy to see that the (in the literature often-used) example $A = (0, \infty)$ and $G_\alpha = \text{Poisson}(\alpha)$ satisfies the conditions above.*

To see that the sequence of experiments $(\mathcal{E}_2^{(n)}(\nu, \mathcal{G}_A))_{n \in \mathbb{Z}_+}$ has the LAN-property, we need to determine the asymptotic behavior of a localized log-likelihood ratio. To that end we first write down the likelihood. By the p-th order Markov-structure, the likelihood is given by,

$$L_n(\theta, \alpha \mid X_{-p}, \ldots, X_n) = \nu\{X_{-1}, \ldots, X_{-p}\} \prod_{t=0}^{n} P^{\theta,\alpha}_{(X_{t-1},\ldots,X_{t-p}),X_t}.$$

Since the likelihood is extremely smooth in (θ, α), it seems to be appropriate to establish the LAN-property directly, using a Taylor-expansion. This is the path we take. To get a useful expression of the transition-scores for θ and α, we briefly discuss how we can view upon the model as an information-loss model. Suppose that, instead of just observing $X_{-p}, \ldots, X_n$, we would also be able to observe $\vartheta_i \circ X_{t-i}$, $i = 1, \ldots, p$, $t = 0, \ldots, n$. Then $\varepsilon_t = X_t - \sum_{i=1}^{p} \vartheta_i \circ X_{t-i}$ also belongs to the information set at time

t, just as in the classical AR(p) model. In our model, with only observations on $X_{-p}, \ldots, X_n$, this does not hold true; there is loss of information. The 'information-loss principle', see for example Proposition A.5.5 in Bickel et al. (1998), suggests that the transition-score for θ_i in the model where we only observe $X_{t-p}, \ldots, X_t$, equals the conditional expectation, given $X_{t-p}, \ldots, X_t$, of the transition-score for θ_i in the model with also observations on $\vartheta_i \circ X_{t-i}$, $i = 1, \ldots, p$. It is not difficult to see that the transition-score for θ_i in the model with the additional observations $\vartheta_i \circ X_{t-i}$ is nothing but the score of a $\mathrm{Bin}_{X_{t-i},\theta_i}$ distribution. Recall that the score of a $\mathrm{Bin}_{x,\theta}$ distribution is given by, for $\theta \in (0,1)$,

$$\dot{s}_{x,\theta}(k) = \left(\frac{\partial}{\partial\theta} \log \mathrm{b}_{x,\theta}(k)\right) 1_{(0,1]}(\mathrm{b}_{x,\theta}(k)) = \frac{k - \theta x}{\theta(1-\theta)} 1_{\{0,\ldots,x\}}(k), \quad x \in \mathbb{Z}_+. \tag{5}$$

Hence, the information-loss structure suggests that the transition-score for θ_i in our model equals,

$$\mathbb{E}_{\theta,\alpha}\left[\dot{s}_{X_{t-i},\theta_i}(\vartheta_i \circ X_{t-i}) \mid X_t, \ldots, X_{t-p}\right].$$

Similarly, the transition-score for α is conjectured to be equal to,

$$\mathbb{E}_{\theta,\alpha}\left[h_\alpha(\varepsilon_t) \mid X_t, \ldots, X_{t-p}\right].$$

One way to make this reasoning precise, is to show that the model is differentiable in quadratic mean with respect to (θ, α). Instead, since the model is extremely smooth, we may derive the transition-scores directly by calculating the partial derivatives of $\log P^{\theta,\alpha}_{(x_{t-1},\ldots,x_{t-p}),x_t}$ with respect to both θ and α. It is easy to see that, for $x_{t-p}, \ldots, x_t \in \mathbb{Z}_+$, $i = 1, \ldots, p$, $\theta \in (0,1)^p$, we have,

$$\begin{aligned}
\dot{\ell}_{\theta,i}(x_{t-p}, \ldots, x_{t-1}, x_t; \theta, \alpha) &= \frac{\partial}{\partial\theta_i} \log\left(P^{\theta,\alpha}_{(x_{t-1},\ldots,x_{t-p}),x_t}\right) 1_{(0,1]}\left(P^{\theta,\alpha}_{(x_{t-1},\ldots,x_{t-p}),x_t}\right) \\
&= \frac{\sum_k \dot{s}_{x_{t-i},\theta_i}(k)\, \mathrm{b}_{x_{t-i},\theta_i}(k) \left(G_\alpha * \underset{j=1,\ldots,p}{\overset{j\neq i}{*}} \mathrm{Bin}_{x_{t-j},\theta_j}\right)\{x_t - k\}}{P^{\theta,\alpha}_{(x_{t-1},\ldots,x_{t-p}),x_t}} 1_{(0,1]}\left(P^{\theta,\alpha}_{(x_{t-1},\ldots,x_{t-p}),x_t}\right) \\
&= \mathbb{E}_{\theta,\alpha}\left[\dot{s}_{X_{t-i},\theta_i}(\vartheta_i \circ X_{t-i}) \mid X_t = x_t, \ldots, X_{t-p} = x_{t-p}\right],
\end{aligned} \tag{6}$$

where we put $\mathbb{E}_{\theta,\alpha}[\cdot \mid X_t = x_t, \ldots, X_{t-p} = x_{t-p}] = 0$ if $\mathbb{P}_{\nu,\theta,\alpha}\{X_{t-p} = x_{t-p}, \ldots, X_t = x_t\} = 0$. Similarly we find, for $x_{t-p}, \ldots, x_t \in \mathbb{Z}_+$, and $i = 1, \ldots, q$,

$$\begin{aligned}
\dot{\ell}_{\alpha,i}(x_{t-p}, \ldots, x_{t-1}, x_t; \theta, \alpha) &= \frac{\partial}{\partial\alpha_i} \log\left(P^{\theta,\alpha}_{(x_{t-1},\ldots,x_{t-p}),x_t}\right) 1_{(0,1]}\left(P^{\theta,\alpha}_{(x_{t-1},\ldots,x_{t-p}),x_t}\right) \\
&= \frac{\sum_e h_{\alpha,i}(e) g_\alpha(e) \left(*_{j=1,\ldots,p} \mathrm{Bin}_{x_{t-j},\theta_j}\right)\{x_t - e\}}{P^{\theta,\alpha}_{(x_{t-1},\ldots,x_{t-p}),x_t}} 1_{(0,1]}\left(P^{\theta,\alpha}_{(x_{t-1},\ldots,x_{t-p}),x_t}\right) \\
&= \mathbb{E}_{\theta,\alpha}\left[h_{\alpha,i}(\varepsilon_t) \mid X_t = x_t, \ldots, X_{t-p} = x_{t-p}\right].
\end{aligned} \tag{7}$$

For the case $p = 1$ and $G_\alpha = \text{Poisson}(\alpha)$, representation (6) was also found by Freeland and McCabe (2004). Although we established (6) and (7) by direct calculations, we stress that the structure is due to the information-loss interpretation of the model. It follows that the score is a martingale. A Taylor-expansion of the localized log-likelihood ratio, a martingale central limit theorem, and a law of large numbers now suggest that the sequence of experiments $(\mathcal{E}_2^{(n)}(\nu, \mathcal{G}_A))_{n\in\mathbb{Z}_+}$ has the LAN-property. The following theorem gives the precise result.

Theorem 3.1 *Let $\mathcal{G}_A \subset \mathcal{G}$ satisfy Assumptions (A1)-(A5) and let ν a probability measure on $\mathbb{Z}_+^p$ with finite support. Let $\theta \in \Theta$ and $\alpha \in A$. Then the sequence of experiments $(\mathcal{E}_2^{(n)}(\nu, \mathcal{G}_A))_{n\in\mathbb{Z}_+}$ has the LAN-property in (θ, α), i.e. for every $u = (u_1, u_2) \in \mathbb{R}^p \times \mathbb{R}^q$ the following expansion holds,*

$$\begin{aligned}\log \frac{\mathrm{d}\mathbb{P}^{(n)}_{\nu,\theta+u_1/\sqrt{n},\alpha+u_2/\sqrt{n}}}{\mathrm{d}\mathbb{P}^{(n)}_{\nu,\theta,\alpha}}(X_{-p},\dots,X_n) &= \log \frac{L_n\left(\theta+\frac{u_1}{\sqrt{n}}, \alpha+\frac{u_2}{\sqrt{n}} \mid X_{-p},\dots,X_n\right)}{L_n\left(\theta,\alpha \mid X_{-p},\dots,X_n\right)} \\ &= u^T S_n - \frac{1}{2}u^T J u + R_n,\end{aligned}$$

where the score (also called central sequence),

$$S_n = S_n(\theta,\alpha) = \frac{1}{\sqrt{n}}\sum_{t=0}^{n}\begin{pmatrix}\dot{\ell}_\theta(X_{t-p},\dots,X_t;\theta,\alpha)\\ \dot{\ell}_\alpha(X_{t-p},\dots,X_t;\theta,\alpha)\end{pmatrix}, \tag{8}$$

satisfies

$$S_n \xrightarrow{d} \mathrm{N}(0, J), \text{ under } \mathbb{P}_{\nu,\theta,\alpha}. \tag{9}$$

The Fisher-information defined by,

$$J = J(\theta,\alpha) = \begin{pmatrix} J_\theta & J_{\theta,\alpha} \\ J_{\alpha,\theta} & J_\alpha\end{pmatrix} = \begin{pmatrix}\mathbb{E}_{\nu_{\theta,\alpha},\theta,\alpha}\dot{\ell}_\theta\dot{\ell}_\theta^T(X_{-p},\dots,X_0;\theta,\alpha) & \mathbb{E}_{\nu_{\theta,\alpha},\theta,\alpha}\dot{\ell}_\theta\dot{\ell}_\alpha^T(X_{-p},\dots,X_0;\theta,\alpha)\\ \mathbb{E}_{\nu_{\theta,\alpha},\theta,\alpha}\dot{\ell}_\alpha\dot{\ell}_\theta^T(X_{-p},\dots,X_0;\theta,\alpha) & \mathbb{E}_{\nu_{\theta,\alpha},\theta,\alpha}\dot{\ell}_\alpha\dot{\ell}_\alpha^T(X_{-p},\dots,X_0;\theta,\alpha)\end{pmatrix},$$

is non-singular, and $R_n = R_n(u,\theta,\alpha) \xrightarrow{p} 0$ under $\mathbb{P}_{\nu,\theta,G}$.

Remark 3 *If one wants to draw the initial value, $(X_{-1},\dots,X_{-p})'$, according to the stationary distribution, one considers the sequence of experiments $\tilde{\mathcal{E}}_2^{(n)}(\mathcal{G}_A) = (\mathbb{Z}_+^{n+1+p}, 2^{\mathbb{Z}_+^{n+1+p}}, (\mathbb{P}^{(n)}_{\nu_{\theta,\alpha},\theta,\alpha} \mid \theta\in\Theta, \alpha \in A))$, $n \in \mathbb{Z}_+$. If the conditions in Theorem 3.1 are satisfied and if the initial value is negligible, i.e. for all $u = (u_1, u_2) \in \mathbb{R}^p \times \mathbb{R}^q$ we have $\nu_{\theta+u_1/\sqrt{n},\alpha+u_2/\sqrt{n}}\{X_{-1},\dots,X_{-p}\} - \nu_{\theta,\alpha}\{X_{-1},\dots,X_{-p}\} = o(\mathbb{P}_{\nu_{\theta,\alpha},\theta,\alpha}; 1)$, then we can obtain the LAN-property for $(\tilde{\mathcal{E}}_2^{(n)}(\mathcal{G}_A))_{n\in\mathbb{Z}_+}$ analogous to the proof of Theorem 3.1. For this 'stationary case' the LAN-property can alternatively be established using results in Roussas (1972). In case $p = 1$, $A = (0,\infty)$, and $G_\alpha = \text{Poisson}(\alpha)$, it is easy to see, using generating functions, that*

$\nu_{\theta,\alpha} = Poisson\,(\alpha/(1-\theta))$. *For this case the negligibility of the initial value readily follows.*

Remark 4 *For the case* $p = 1$ *and* $G_\alpha = \text{Poisson}(\alpha)$, *the non-singularity of* J *is obtained via direct calculation by Franke and Seligmann (1993).*

If we want to consider the sequence of experiments $\mathcal{E}_1^{(n)}(\nu, G)$, $n \in \mathbb{Z}_+$, we can always embed G in a parametric model $\mathcal{G}_A$ which satisfies Assumptions (A1)-(A5). Then an application of the preceding theorem with $u_2 = 0$ immediately yields the following corollary.

Corollary 3.1 *Let* $\theta \in \Theta$, *let* $G \in \mathcal{G}$ *with* $\mathbb{E}_G\varepsilon_0^2 < \infty$, *and* $g(0) \in (0,1)$, *and let* ν *be a probability measure on* $\mathbb{Z}_+^p$ *with finite support. Then the sequence of experiments* $(\mathcal{E}_1^{(n)}(\nu, G))_{n\in\mathbb{Z}_+}$ *has the LAN-property in* θ, *i.e. for every* $u \in \mathbb{R}^p$ *the following expansion holds,*

$$\log \frac{\mathrm{d}\mathbb{P}^{(n)}_{\nu,\theta+\frac{u}{\sqrt{n}},G}}{\mathrm{d}\mathbb{P}^{(n)}_{\nu,\theta,G}}(X_{-p}, \dots, X_n) = u^T S_n^\theta - \frac{1}{2} u^T J_\theta u + R_n,$$

where $S_n^\theta = n^{-1/2} \sum_{t=0}^n \dot{\ell}_\theta(X_{t-p}, \dots, X_t; \theta, G) \xrightarrow{d} \mathrm{N}(0, J_\theta)$ *under* $\mathbb{P}_{\nu,\theta,G}$, $J_\theta = J_\theta(\theta, G)$ *is invertible, and* $R_n = R_n(u, \theta, G) \xrightarrow{p} 0$ *under* $\mathbb{P}_{\nu,\theta,G}$.

3.2 Efficient estimators

This section provides efficient estimators of the parameters in an INAR(p) model based on the ubiquitous one-step update method.

3.2.1 G is known

In case $\mu_G < \infty$, an initial estimator of θ is the OLS-estimator,

$$\hat{\theta}_n^G = \begin{pmatrix} \sum_{t=0}^n X_{t-1}^2 & \dots & \sum_{t=0}^n X_{t-1}X_{t-p} \\ \vdots & \ddots & \vdots \\ \sum_{t=0}^n X_{t-p}X_{t-1} & \dots & \sum_{t=0}^n X_{t-p}^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{t=0}^n X_{t-1}(X_t - \mu_G) \\ \vdots \\ \sum_{t=0}^n X_{t-p}(X_t - \mu_G) \end{pmatrix}.$$

If we assume the existence of a third moment of X_0 under the stationary distribution (which is, by Lemma 2.1, equivalent to imposing $\mathbb{E}_G\varepsilon_0^3 < \infty$), $\hat{\theta}_n^G$ yields a $\sqrt{n}$-consistent estimator of θ. The following proposition is well-known (see Du and Li (1991)).

Proposition 3.1 *Let* $\theta \in \Theta$, ν *a probability measure on* $\mathbb{Z}_+^p$ *with finite support,* $G \in \mathcal{G}$ *with* $g(0) \in (0,1)$ *and* $\mathbb{E}_G\varepsilon_0^3 < \infty$. *Then* $\sqrt{n}\left(\hat{\theta}_n^G - \theta\right)$ *converges in distribution under* $\mathbb{P}_{\nu,\theta,G}$.

Next, we apply the one-step-Newton-Raphson-method to update this initial $\sqrt{n}$-consistent estimator into an efficient estimator. To state this theorem, we need the concept of a discretized estimator. For $n \in \mathbb{N}$ make a grid of cubes, with sides of length $1/\sqrt{n}$, over $\mathbb{R}^p$ and, given $\hat{\theta}_n^G$, define $\hat{\theta}_n^{G,*}$ to be the midpoint of the cube into which $\hat{\theta}_n$ has fallen (for ties take one of the possibilities). Then $\hat{\theta}_n^{G,*}$ is also $\sqrt{n}$-consistent and is called a discretized version of $\hat{\theta}_n$.

Theorem 3.2 *Let ν a probability measure on $\mathbb{Z}_+^p$ with finite support, $G \in \mathcal{G}$ with $g(0) \in (0,1)$ and $\mathbb{E}_G \varepsilon_0^3 < \infty$. Let $\hat{\theta}_n^*$ be a discretized version of $\hat{\theta}_n$. Then*

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

where,

$$\hat{J}_{n,\theta} = \frac{1}{n} \sum_{t=0}^{n} \dot{\ell}_\theta \dot{\ell}_\theta^T (X_{t-p}, \ldots, X_t; \hat{\theta}_n^*, G),$$

is an efficient estimator of θ in the sequence of experiments $(\mathcal{E}_1^{(n)}(\nu, G))_{n \in \mathbb{Z}_+}$. Moreover, $\hat{J}_{n,\theta}^{-1}$ is a consistent estimator of the asymptotic covariance matrix of $\hat{\theta}_n^G$, i.e.

$$\hat{J}_{n,\theta}^{-1} \xrightarrow{p} J_\theta^{-1}, \text{ under } \mathbb{P}_{\nu,\theta,G}.$$

Remark 5 *Instead of $\hat{\theta}_n^G$, any other $\sqrt{n}$-consistent estimator of θ can be used.*

Remark 6 *If one can find a $\sqrt{n}$-consistent initial estimator of θ under the weaker assumption $\mathbb{E}_G \varepsilon_0^2 < \infty$, the condition $\mathbb{E}_G \varepsilon_0^3 < \infty$ may be replaced by $\mathbb{E}_G \varepsilon_0^2 < \infty$.*

The proof of this theorem runs along the same lines as the proof of Theorem 3.3.

3.2.2 G belongs to a parametric model

To use the OLS-estimator as an initial estimator of θ we need the existence of a third moment of X_t under the stationary distribution. Therefore we replace, in this section, Assumption (A5) on $\mathcal{G}_A$ by,

(A5′) for all $\alpha \in A$: $\mathbb{E}_\alpha \varepsilon_0^3 < \infty$.

This yields, by Lemma 2.1, the existence of a third moment of X_0 under the stationary distribution. Just as for the case G known, OLS yields a $\sqrt{n}$-consistent estimator of (θ, μ_G) (see, for example, Du and Li (1991)).

Proposition 3.2 *Let $\theta \in \Theta$, ν a probability measure on $\mathbb{Z}_+^p$ with finite support, $G \in \mathcal{G}$ with $\mathbb{E}_G \varepsilon_0^3 < \infty$*

and $g(0) \in (0,1)$*. Then* $\left(\sqrt{n}\left(\hat{\theta}_n - \theta\right), \sqrt{n}\left(\hat{\mu}_{G,n} - \mu_G\right)\right)$ *converges in distribution under* $\mathbb{P}_{\nu,\theta,G}$*, where,*

$$\begin{pmatrix} \hat{\mu}_{G,n} \\ \hat{\theta}_n \end{pmatrix} = \begin{pmatrix} n & \sum_{t=0}^n X_{t-1} & \dots & \sum_{t=0}^n X_{t-p} \\ \sum_{t=0}^n X_{t-1} & \sum_{t=0}^n X_{t-1}^2 & \dots & \sum_{t=0}^n X_{t-1}X_{t-p} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{t=0}^n X_{t-p} & \sum_{t=0}^n X_{t-p}X_{t-1} & \dots & \sum_{t=0}^n X_{t-p}^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{t=0}^n X_t \\ \sum_{t=0}^n X_{t-1}X_t \\ \vdots \\ \sum_{t=0}^n X_{t-p}X_t \end{pmatrix}.$$

The OLS-estimator also yields a $\sqrt{n}$-consistent estimator of α in case $\mathcal{G}_A = (\text{Poisson}(\alpha) \mid \alpha > 0)$, since then $\mu_{G_\alpha} = \alpha$.

For other specific choices of G_α, it might by easy to find a (moment-based) estimator of α. This is the approach, we recommend. However, it would be reassuring to know that a $\sqrt{n}$-consistent estimator of α always exists. The following observation is the key to the general existence of a $\sqrt{n}$-consistent estimator of α. Although we do not observe the innovation process $(\varepsilon_t)_{t\in\mathbb{Z}_+}$, we have observations on some innovations (if $g(0) > 0$), since

$$X_t 1\{X_{t-1} = 0, \dots, X_{t-p} = 0\} = \varepsilon_t. \tag{10}$$

By Assumptions (A1)-(A6) $\mathcal{G}_A$ is an identified regular parametric model (see Definition 2.1.1 and Proposition 2.1.1 in Bickel et al. (1998)). By a theorem by Le Cam (see, e.g., Theorem 2.5.1 in Bickel et al. (1998)) there exists an 'estimator' $T_n = t_n(\varepsilon_1, \dots, \varepsilon_n)$ of α such that $\sqrt{n}(T_n - \alpha)$ is tight under $\mathbb{P}_{\nu,\theta,\alpha}$ for all $\alpha \in A$. Using display (10) we could use such an 'estimator' to construct a $\sqrt{n}$-consistent estimator of α.

Proposition 3.3 *Let* $\theta \in \Theta$*,* ν *a probability measure on* $\mathbb{Z}_+^p$ *and* $G \in \mathcal{G}$ *with* $g(0) \in (0,1)$ *and* $\sigma_G^2 < \infty$*. Let*

$$\tau_0 = 0, \;\; \tau_k = \inf\{t > \tau_{k-1} \mid X_{t-p} = \dots = X_{t-1} = 0\}, \;\; k \in \mathbb{N}, \;\; \textit{and } N_n = \max\{j \in \mathbb{Z}_+ \mid \tau_j \leq n\}.$$

Then $\hat{\alpha}_n = t_{N_n}\left(X_{\tau_1}, \dots, X_{\tau_{N_n}}\right)$*, defines a* $\sqrt{n}$*-consistent estimator of* α*. In particular, if for some* $\sigma^2 > 0$*,*

$$\sqrt{n}\left(t_n(\varepsilon_1, \dots, \varepsilon_n) - \alpha\right) \xrightarrow{d} \mathrm{N}(0, \sigma^2), \;\; \textit{under } \mathbb{P}_{\nu,\theta,\alpha},$$

then we have,

$$\sqrt{n}\left(\hat{\alpha}_n - \alpha\right) \xrightarrow{d} \mathrm{N}\left(0, \frac{\sigma^2}{\nu_{\theta,\alpha}\{0, \dots, 0\}}\right), \;\; \textit{under } \mathbb{P}_{\nu,\theta,\alpha}.$$

Since we have a $\sqrt{n}$-consistent estimator of (θ, α), we can update this estimator into an efficient estimator.

Theorem 3.3 *Let ν a probability measure on $\mathbb{Z}_+^p$ with finite support, and $\mathcal{G}_A \subset \mathcal{G}$ satisfying Assumptions (A1)-(A6) with (A5$'$) instead of (A5). Let $(\hat{\theta}_n, \hat{\alpha}_n)$ be a $\sqrt{n}$-consistent estimator of (θ, α) and $(\hat{\theta}_n^*, \hat{\alpha}_n^*)$ a discretized version of it. Then,*

$$\begin{pmatrix} \hat{\theta}_n^{**} \\ \hat{\alpha}_n^{**} \end{pmatrix} = \begin{pmatrix} \hat{\theta}_n^{*} \\ \hat{\alpha}_n^{*} \end{pmatrix} + \frac{1}{n} \sum_{t=0}^{n} \hat{J}_n^{-1} \begin{pmatrix} \dot{\ell}_\theta(X_{t-p}, \ldots, X_t; \hat{\theta}_n^*, \hat{\alpha}_n^*) \\ \dot{\ell}_\alpha(X_{t-p}, \ldots, X_t; \hat{\theta}_n^*, \hat{\alpha}_n^*) \end{pmatrix},$$

with,

$$\hat{J}_n = \begin{pmatrix} \frac{1}{n} \sum_{t=0}^{n} \dot{\ell}_\theta \dot{\ell}_\theta^T(X_{t-p}, \ldots, X_t; \hat{\theta}_n^*, \hat{\alpha}_n^*) & \frac{1}{n} \sum_{t=0}^{n} \dot{\ell}_\theta \dot{\ell}_\alpha^T(X_{t-p} \ldots, X_t; \hat{\theta}_n^*, \hat{\alpha}_n^*) \\ \frac{1}{n} \sum_{t=0}^{n} \dot{\ell}_\alpha \dot{\ell}_\theta^T(X_{t-p}, \ldots, X_t; \hat{\theta}_n^*, \hat{\alpha}_n^*) & \frac{1}{n} \sum_{t=0}^{n} \dot{\ell}_\alpha \dot{\ell}_\alpha^T(X_{t-p}, \ldots, X_t; \hat{\theta}_n^*, \hat{\alpha}_n^*) \end{pmatrix},$$

is an efficient estimator of (θ, α) in the sequence of experiments $(\mathcal{E}_2^{(n)}(\nu, \mathcal{G}_A))_{n \in \mathbb{Z}_+}$. Moreover, $\hat{J}_n^{-1}$ yields a consistent estimator of the asymptotic covariance matrix of $(\hat{\theta}_n^{**}, \hat{\alpha}_n^{**})$, i.e.,

$$\hat{J}_n^{-1} \xrightarrow{p} J^{-1}, \text{ under } \mathbb{P}_{\nu,\theta,\alpha}.$$

Remark 7 *The same remarks as after Theorem 3.2 apply.*

A Auxiliary results

The setup is as described in Sections 1 and 2 of the main text.

Note that $X = (X_t)_{t \geq -p}$ is a p-th order Markov chain. To exploit this Markovian structure we introduce a process $Y = (Y_t)_{t \geq 0}$ defined by $Y_t = (X_{t-1}, X_{t-2}, \ldots, X_{t-p})'$. Under $\mathbb{P}_{\nu,\theta,G}$ the process Y is a Markov chain in $\mathbb{Z}_+^p$ with initial distribution ν and transition-probabilities

$$Q_{y_0,y_1}^{\theta,G} = \mathbb{P}_{\nu,\theta,G}\{Y_{t+1} = y_1 \mid Y_t = y_0\} = \begin{cases} P_{y_0,x_0}^{\theta,G} & , \text{ if } y_1 = (x_0, x_{-1}, \ldots, x_{-p+1})'; \\ 0 & , \text{ otherwise,} \end{cases}$$

where $y_0 = (x_{-1}, \ldots, x_{-p})'$, $y_1 \in \mathbb{Z}_+^p$ and $x_0 \in \mathbb{Z}_+$. It is easy to see that, in case $g(0) < 1$ and $\theta \in (0,1)^p$, the Markov chain Y is irreducible on $\mathcal{S} = \mathcal{S}(G) = \{\alpha, \alpha+1, \ldots\}^p$, where $\alpha = \min\{k \in \mathbb{Z}_+ \mid g(k) > 0\} \in \mathbb{Z}_+$. It is also not hard to see that, under these conditions, the chain is also aperiodic.

Next we recall some (adaptions of) results from the literature.

Lemma A.1 *Let $G \in \mathcal{G}$ with $g(0) \in [0,1)$ and $\theta \in (0,1)^p$. Assume that we have a unique distribution $\nu_{\theta,G}$ such that Y is stationary under $\mathbb{P}_{\nu_{\theta,G},\theta,G}$. If there exists a finite set $A \subset \mathbb{Z}_+^p$, $\delta > 0$, and a nonnegative function $f : \mathbb{Z}_+^p \to \mathbb{R}$ such that $\mathbb{E}_{\theta,G}[f(Y_t) \mid Y_{t-1} = y] \leq (1-\delta)f(y)$, for $y \in A^c$, and $f > 0$ on A, then f is $\nu_{\theta,G}$-integrable, i.e. $\mathbb{E}_{\nu_{\theta,G},\theta,G} f(Y_t) = \sum_{y \in \mathcal{S}(G)} f(y)\nu_{\theta,G}\{y\} < \infty$.*

Proof:
Since Y lives in a countable state space and is recurrent and irreducible, the chain Y is Harris recurrent. We already noticed that Y is also aperiodic under the imposed conditions. Now we can apply Theorem 3 in Tweedie (1983) (display (1.3) in this paper is trivially satisfied since A is finite). $\square$

Lemma A.2 *Let ν be a probability measure on $\mathbb{Z}_+^p$, $G \in \mathcal{G}$ with $g(0) \in [0,1)$ and $\mu_G < \infty$, $\theta \in (0,1)^p$ with $\sum_{i=1}^p \theta_i < 1$. Then for every function $h : \mathbb{Z}_+^{p+1} \to \mathbb{R}$ satisfying $\mathbb{E}_{\nu_{\theta,G},\theta,G}|h(X_{-p},\ldots,X_0)| < \infty$, the following strong law of large numbers holds*

$$\mathbb{P}_{\nu,\theta,G}\left\{\lim_{n\to\infty}\frac{1}{n}\sum_{t=0}^{n} h(X_{t-p},\ldots,X_t) = \mathbb{E}_{\nu_{\theta,G},\theta,G} h(X_{-p},\ldots,X_0)\right\} = 1.$$

Proof:
Since Y is an irreducible Markov chain with stationary distribution $\nu_{\theta,G}$, $Z_t = (Y_t, Y_{t-1})'$, yields an irreducible Markov chain with stationary distribution $\nu_{\theta,G} \otimes Q^{\theta,G}$. Now the result follows from a law of large numbers for Markov chains (see, e.g., Theorem 4.3.15 in Dacunha-Castelle and Duflo (1986)). $\square$

Lemma A.3 *Let ν be a probability measure on $\mathbb{Z}_+^p$, $G \in \mathcal{G}$ with $g(0) < 1$, $\mu_G < \infty$, and $\theta \in (0,1)^p$ with $\sum_{i=1}^p \theta_i < 1$. If $h : \mathbb{Z}_+^{p+1} \to \mathbb{R}$ satisfies $\mathbb{E}_{\nu_{\theta,G},\theta,G} h^2(X_{-p},\ldots,X_{-1},X_0) < \infty$, then,*

$$\frac{1}{\sqrt{n}}\sum_{t=0}^{n}\left(h(X_{t-p},\ldots,X_t) - \mathbb{E}_{\theta,G}\left[h(X_{t-p},\ldots,X_t) \mid X_{t-1},\ldots,X_{t-p}\right]\right) \xrightarrow{d} \mathrm{N}(0,\sigma^2), \text{ under } \mathbb{P}_{\nu,\theta,G},$$

where σ^2 is given by,

$$\sigma^2 = \mathbb{E}_{\nu_{\theta,G},\theta,G} h^2(X_{-p},\ldots,X_{-1},X_0) - \mathbb{E}_{\nu_{\theta,G}}\left(\mathbb{E}_{\theta,G}\left[h(X_{-p},\ldots,X_0) \mid X_{-1},\ldots,X_{-p}\right]\right)^2.$$

Proof:
Since Y is a positive recurrent Markov chain on $\mathcal{S}$, this follows from Theorem 4.3.16 in Dacunha-Castelle and Duflo (1986) in case $\nu = \delta_y$. From this, the result extends to arbitrary ν by looking at pointwise convergence of characteristic functions, conditioning on the initial value and using dominated convergence. $\square$

Lemma A.4 *Let $G \in \mathcal{G}$ with $g(0) \in [0,1)$, $\mu_G < \infty$, $\theta \in (0,1)^p$ with $\sum_{i=1}^p \theta_i < 1$, and ν a probability measure on $\mathbb{Z}_+^p$ whose support is a finite subset of $\mathcal{S}$. If $g : \mathbb{Z}_+^{p+1} \to [0,\infty)$ satisfies $\mathbb{E}_{\nu_{\theta,G},\theta,G} g(X_{-p},\ldots,X_0) < \infty$*

then,

$$\lim_{t\to\infty} \sup_{0\leq h\leq g} \left|\mathbb{E}_{\nu,\theta,G} h(X_{t-p},\ldots,X_t) - \mathbb{E}_{\nu_{\theta,G},\theta,G} h(X_{-p},\ldots,X_0)\right| = 0. \tag{11}$$

PROOF:
This follows by Theorem 2 and the remark after this theorem in Tweedie (1983), since $W_t = (Y_{t+1}, Y_t)'$ is an aperiodic Harris recurrent Markov chain with stationary distribution $\nu_{\theta,G} \otimes P^{\theta,G}$. □

Lemma A.5 *Let $G \in \mathcal{G}$ with $g(0) \in [0,1)$, $\mu_G < \infty$, $\theta \in (0,1)^p$ with $\sum_{i=1}^p \theta_i < 1$, and ν a probability measure on $\mathbb{Z}_+^p$ whose support is a finite subset of $\mathcal{S}$. Let K be a compact subset of $\mathbb{R}^k$. Let, for every $\kappa \in K$, $f(\cdot;\kappa) : \mathbb{Z}_+^{p+1} \to \mathbb{R}$ such that the following conditions hold,*

(C1)

$$\mathbb{E}_{\nu_{\theta,G},\theta,G} \sup_{\kappa\in K} |f(X_{-p},\ldots,X_0;\kappa)| < \infty;$$

(C2) for every $x_{-p},\ldots,x_0 \in \mathbb{Z}_+$ the map $\kappa \mapsto f(x_{-p},\ldots,x_0;\kappa)$ is continuous.

Then,

$$\sup_{\kappa\in K} \left| \frac{1}{n}\sum_{t=0}^n f(X_{t-p},\ldots,X_t;\kappa) - \mathbb{E}_{\nu_{\theta,G},\theta,G} f(X_{-p},\ldots,X_0;\kappa) \right| \xrightarrow{p} 0, \text{ under } \mathbb{P}_{\nu,\theta,G}. \tag{12}$$

And, for $K \ni \kappa_n \to \kappa_0$,

$$\frac{1}{n}\sum_{t=0}^n f(X_{t-p},\ldots,X_t;\kappa_n) \xrightarrow{p} \mathbb{E}_{\nu_{\theta,G},\theta,G} f(X_{-p},\ldots,X_0;\kappa_0), \text{ under } \mathbb{P}_{\nu,\theta,G}. \tag{13}$$

PROOF:
By Assumption (C1) and Lemma A.4 we have, for $M > 0$,

$$\begin{aligned}\lim_{t\to\infty} \mathbb{E}_{\nu,\theta,G} \sup_{\kappa\in K} |f(X_{t-p},\ldots,X_t;\kappa)|\, 1_{[M,\infty)}\left(\sup_{\kappa\in K} |f(X_{t-p},\ldots,X_t;\kappa)|\right) \\ = \mathbb{E}_{\nu_{\theta,G},\theta,G} \sup_{\kappa\in K} |f(X_{-p},\ldots,X_0;\kappa)|\, 1_{\sup_{\kappa\in K}|f(X_{-p},\ldots,X_0;\kappa)|>M}.\end{aligned}$$

Hence we obtain, for $M > 0$,

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{n}\sum_{t=0}^n \mathbb{E}_{\nu,\theta,G} \sup_{\kappa\in K} |f(X_{t-p},\ldots,X_t;\kappa)|\, 1_{[M,\infty)}\left(\sup_{\kappa\in K} |f(X_{t-p},\ldots,X_t;\kappa)|\right) \\ = \mathbb{E}_{\nu_{\theta,G},\theta,G} \sup_{\kappa\in K} |f(X_{-p},\ldots,X_0;\kappa)|\, 1_{[M,\infty)}\left(\sup_{\kappa\in K} |f(X_{-p},\ldots,X_0;\kappa)|\right).\end{aligned}$$

Using (C1) we now obtain,

$$\lim_{M\to\infty}\lim_{n\to\infty}\frac{1}{n}\sum_{t=0}^{n}\mathbb{E}_{\nu,\theta,G}\sup_{\kappa\in K}|f(X_{t-p},\ldots,X_t;\kappa)|\,1_{[M,\infty)}\left(\sup_{\kappa\in K}|f(X_{t-p},\ldots,X_t;\kappa)|\right)=0.$$

Hence Assumption DM in Andrews (1992) is satisfied. Assumption TSE-1B in that paper is also satisfied, by the compactness of K, by (C2) and by $\lim_{n\to\infty}(1/n)\sum_{t=0}^{n}\mathbb{P}_{\nu,\theta,G}\left\{(X_{t-p},\ldots,X_t)\in A\right\}=\nu_{\theta,G}\otimes P^{\theta,G}(A)$ for $A\subset\mathbb{Z}_+^{p+1}$ (Lemma A.4). Since we have Lemma A.2, an application of Theorem 4 in Andrews (1992) yields,

$$\sup_{\kappa\in K}\left|\frac{1}{n}\sum_{t=0}^{n}f(X_{t-p},\ldots,X_t;\kappa)-\mathbb{E}_{\nu,\theta,G}f(X_{-p},\ldots,X_0;\kappa)\right|\xrightarrow{p}0,\ \text{under}\ \mathbb{P}_{\nu,\theta,G}.$$

This yields (12), since

$$\sup_{\kappa\in K}\left|\mathbb{E}_{\nu,\theta,G}f(X_{t-p},\ldots,X_t;\kappa)-\mathbb{E}_{\nu_{\theta,G},\theta,G}f(X_{-p},\ldots,X_0;\kappa)\right|\to 0\text{ as }t\to\infty,$$

which follows by applying Lemma A.4 to f^+ and f^-. Display (13) follows by applying dominated convergence. □

B Proofs

Proof of Theorem 2.1:
First note that it suffices to prove that the Markov chain Y, introduced in Appendix A, has a stationary distribution. We prove this for the case $g(0)>0$. The case $g(0)=0$ runs along the same lines.
By $Q_{i,j}^n$ we denote the n-step probability of moving from state i to j of the process Y, i.e., $Q_{i,j}^n=\mathbb{P}_{\delta_i,\theta,G}\{Y_n=j\}$. Since, under the imposed conditions, Y is aperiodic and irreducible on $\mathbb{Z}_+^p$, it suffices, by, for example, Theorem 8.8 in Billingsley (1995), to prove that there exist states $i,j\in\mathbb{Z}_+^p$ for which $Q_{i,j}^n$ does not converge to 0 as $n\to\infty$.
It is easy to see that, for all $t\in\mathbb{Z}_+$, $\mathbb{E}_{\delta_0,\theta,G}X_t<\infty$ when $\mathbb{E}_G\varepsilon_0<\infty$. We first show that we even have $\sup_{t\in\mathbb{Z}_+}\mathbb{E}_{\delta_0,\theta,G}X_t<\infty$. Note that this statement indeed holds if we can show that $\mathbb{E}_{\delta_0,\theta,G}X_t\le\mu_G\sum_{j=0}^{t}\theta_*^j$, where $\theta_*=\sum_{i=1}^{p}\theta_i$ which is less than 1 by assumption. Obviously we have $\mathbb{E}_{\delta_0,\theta,G}X_{-1}=\cdots=\mathbb{E}_{\delta_0,\theta,G}X_{-p}=0$ and $\mathbb{E}_{\delta_0,\theta,G}X_0=\mu_G$. Hence the statement holds for $t\in\{-p,\ldots,0\}$. Let $N\in\mathbb{Z}_+$. Assuming that $\mathbb{E}_{\delta_0,\theta,G}X_t\le\mu_G\sum_{j=0}^{t}\theta_*^j$ is valid for all $t\in\{-p,\ldots,N\}$ we obtain

$$\mathbb{E}_{\delta_0,\theta,G}X_{N+1}=\mu_G+\sum_{i=1}^{p}\theta_i\mathbb{E}_{\delta_0,\theta,G}X_{N+1-i}\le\mu_G+\sum_{i=1}^{p}\theta_i\mu_G\sum_{j=0}^{N}\theta_*^j=\mu_G\sum_{j=0}^{N+1}\theta_*^j,$$

which concludes the induction argument.
Using $\sup_{t\in\mathbb{Z}_+}\mathbb{E}_{\delta_0,\theta,G}X_t<\infty$, Markov's inequality yields, for $M>0$,

$$\sup_{t\in\mathbb{Z}_+}\mathbb{P}_{\delta_0,\theta,G}\left\{\max_{i=1,\ldots,p}X_{t-i}>M\right\}\le\frac{p}{M}\sup_{t\in\mathbb{Z}_+}\mathbb{E}_{\delta_0,\theta,G}X_t.$$

Hence there exists $M\in\mathbb{N}$ such that, for all $t\in\mathbb{Z}_+$, $\mathbb{P}_{\delta_0,\theta,G}\left\{\max_{i=1,\ldots,p}X_{t-i}\le M\right\}\ge 1/2$.
Define

$$B_M=\{(x_1,\ldots,x_p)\in\mathbb{Z}_+^p\mid x_i\le M,\,\forall i\in\{1,\ldots,p\}\},$$

then, for $n\ge 1$,

$$Q_{0,0}^{n+p}=\mathbb{P}_{\delta_0,\theta,G}\{Y_{n+p}=0\}\ge\mathbb{P}_{\delta_0,\theta,G}\{Y_n\in B_M,Y_{n+p}=0\}=\sum_{\substack{(i_1,\ldots,i_{n-1})\in\mathbb{Z}_+^{(n-1)p}\\ i_n\in B_M\\ (i_{n+1},\ldots,i_{n+p-1})\in\mathbb{Z}_+^{(p-1)p}}}Q_{0,i_1}\cdots Q_{i_{n+p-1},0}$$

$$=\sum_{\substack{(i_1,\ldots,i_{n-1})\in\mathbb{Z}_+^{(n-1)p}\\ i_n\in B_M}}Q_{0,i_1}\cdots Q_{i_{n-1},i_n}Q_{i_n,0}^p.$$

Using $i_n\in B_M$ we obtain the (very crude) bound

$$Q_{i_n,0}^p\ge\left[g(0)(1-\theta_*)^{pM}\right]^p,$$

where $\theta_*=\max_{i=1,\ldots,p}\theta_i$. Since,

$$\sum_{(i_1,\ldots,i_{n-1})\in\mathbb{Z}_+^{(n-1)p},i_n\in B_M}Q_{0,i_1}\cdots Q_{i_{n-1},i_n}=\mathbb{P}_{\delta_0,\theta,G}\left\{X_{n-p}\le M,\ldots,X_{n-1}\le M\right\}\ge\frac{1}{2},$$

we obtain,

$$Q_{0,0}^{n+p}\ge\left[g(0)(1-\theta_*)^{pM}\right]^p\sum_{\substack{(i_1,\ldots,i_{n-1})\in\mathbb{Z}_+^{(n-1)p}\\ i_n\in B_M}}Q_{0,i_1}\cdots Q_{i_{n-1},i_n}\ge\frac{1}{2}\left[g(0)(1-\theta_*)^{pM}\right]^p>0.$$

This concludes the proof.

Remark 8 *It is also possible to prove this theorem using the, quite mechanical, proof of Lemma 2.1 by applying Theorem 1 in Feigin and Tweedie (1985). We prefer the present proof since it is far more intuitive.*

□

Proof of Lemma 2.1:

We prove the lemma by applying Lemma A.1. The difficulty is to pick the proper function f in Lemma A.1. Note that Theorem 2.1 yields the stationary distribution of Y. Checking that the chosen function indeed satisfies the conditions in Lemma A.1 is easy, but boring and tedious, calculus. Remember that $\mathcal{S} = \{\alpha, \alpha+1, \dots\}^p$, where $\alpha = \min\{k \in \mathbb{Z}_+ \mid g(k) > 0\}$.

The case $k = 1$ For $k = 1$ we take $g : \mathcal{S} \to \mathbb{R}_+$ given by $g(y) = 1 + \sum_{i=1}^p a_i y_i$ with $a_i = (\theta_i + \cdots + \theta_p)$ for $i = 1, \dots, p$, and $0 < \delta < (1 - a_1) \min_{j=1,\dots,p} \theta_j < 1$. We prove that the conditions in Lemma A.1 are satisfied, from which we can conclude that $\int g \, \mathrm{d}\nu_{\theta,G} < \infty$, which implies that $\mathbb{E}_{\nu_{\theta,G},\theta,G} X_0 < \infty$. Note first that $g \geq 1$ on $\mathcal{S}$ and,

$$(1-\delta)g(y) - \mathbb{E}_{\theta,G}\left[g(Y_{t+1}) \mid Y_t = y\right] = a + \sum_{j=1}^p \left[(1-\delta)a_i - c_i\right] y_i,$$

for some constant a and c_i given by $c_i = a_1\theta_i + a_{i+1}$, where we set $a_{p+1} = 0$. We show that $(1-\delta)a_i - c_i > 0$ which implies that,

$$(1-\delta)g(y) - \mathbb{E}_{\theta,G}\left[g(Y_{t+1}) \mid Y_t = y\right], > 0,$$

outside a finite set. We have (use $a_i = \theta_i + a_{i+1}$),

$$(1-\delta)a_i - a_1\theta_i - a_{i+1} = -\delta a_i + (1-a_1)\theta_i > -\delta + (1-a_1) \min_{j=1,\dots,p} \theta_j > 0,$$

which concludes the proof for the case $k = 1$.

The case $k = 2$ For $k = 2$ we take $g : \mathcal{S} \to \mathbb{R}_+$ given by $g(y) = 1 + \sum_{i=1}^p \sum_{j=1}^p a_{ij} y_i y_j$, where $a_{ij} = (\theta_i + \cdots + \theta_p)(\theta_j + \cdots + \theta_p) < 1$ and $a_{ij} = 0$ if $\max\{i, j\} > p$, and $0 < \delta < (1 - a_{11})(\min_j \theta_j)^2$. After some calculus we find,

$$\mathbb{E}_{\theta,G}\left[g(Y_t) \mid Y_{t-1} = y\right] = a + \sum_{i=1}^p \alpha_i y_i + \sum_{i=1}^p \sum_{j=1}^p \beta_{ij} y_i y_j,$$

where $\beta_{ij} = a_{11}\theta_i\theta_j + \theta_i a_{1,j+1} + \theta_j a_{i+1,1} + a_{i+1,j+1}$. Analogous to the previous case it suffices to prove that $(1-\delta)a_{ij} - \beta_{ij} > 0$ for all $i, j = 1, \dots, p$. Using that,

$$a_{ij} = \theta_i\theta_j + \theta_i(\theta_{j+1} + \cdots + \theta_p) + \theta_j(\theta_{i+1} + \cdots + \theta_p) + a_{i+1,j+1},$$

and $a_{ij} < (\theta_i + \cdots + \theta_p)$, we obtain,

$$(1-\delta)a_{ij} - \beta_{ij} > (1-a_{11})\theta_i\theta_j - \delta a_{ij} \geq (1-a_{11})(\min_j \theta_j)^2 - \delta > 0,$$

which concludes the proof for $k = 2$.

The case $k = 3$ Finally we consider the case $k = 3$. Define $g : \mathcal{S} \to \mathbb{R}_+$ by $g(y) = \sum_{i=1}^p \sum_{j=1}^p \sum_{k=1}^p a_{ijk} y_i y_j y_k$, where $a_{ijk} = (\theta_i + \cdots + \theta_p)(\theta_j + \cdots + \theta_p)(\theta_k + \cdots + \theta_p) < 1$ for $i, j, k = 1, \ldots, p$, and $a_{ijk} = 0$ if $\max\{i, j, k\} > p$. And we take $0 < \delta < (1 - a_{111})(\min_j \theta_j)^3$. Using that the third moment of a Binomial(n, p) distribution is given by $np(1 - 3p + 3np + 2p^2 - 3np^2 + n^2p^2)$, it follows, after some nasty calculus,

$$\mathbb{E}_{\theta,G}\left[g(Y_t) \mid Y_{t-1} = y\right] = a + \sum_{i=1}^p b_i y_i + \sum_{i=1}^p \sum_{j=1}^p c_{ij} y_i y_j + \sum_{i=1}^p \sum_{j=1}^p \sum_{k=1}^p \beta_{ijk} y_i y_j y_k,$$

where,

$$\begin{aligned}\beta_{ijk} &= a_{111}\theta_i\theta_j\theta_k + a_{i+1,j+1,k+1} + \theta_i a_{1,j+1,k+1} + \theta_j a_{i+1,1,k+1} + \theta_k a_{i+1,j+1,1} \\ &\quad + \theta_i\theta_j a_{1,1,k+1} + \theta_j\theta_k a_{i+1,1,1} + \theta_i\theta_k a_{1,j+1,1}.\end{aligned}$$

Note that,

$$\begin{aligned}a_{ijk} &= \theta_i\theta_j\theta_k + \theta_i\theta_k(\theta_{j+1} + \cdots + \theta_p) + \theta_j\theta_k(\theta_{i+1} + \cdots + \theta_p) + \theta_i\theta_j(\theta_{k+1} + \cdots + \theta_p) \\ &\quad + \theta_i(\theta_{j+1} + \cdots + \theta_p)(\theta_{k+1} + \cdots + \theta_p) + \theta_j(\theta_{i+1} + \cdots + \theta_p)(\theta_{k+1} + \cdots + \theta_p) \\ &\quad + \theta_k(\theta_{i+1} + \cdots + \theta_p)(\theta_{j+1} + \cdots + \theta_p) + a_{i+1,j+1,k+1}.\end{aligned}$$

Now the rest of the proof proceeds as in the case '$k = 2$'. □

PROOF OF THEOREM 3.1:

Using Assumption (A5) on $\mathcal{G}_A$ and Lemma 2.1 we obtain $\mathbb{E}_{\nu_0,\theta,\alpha} X_0^2 < \infty$, where, for notational convenience, we denote $\nu_0 = \nu_{\theta,\alpha}$.

Expansion of log-likelihood ratio:

Let $u = (u_1, u_2) \in \mathbb{R}^p \times \mathbb{R}^q$, $u \neq 0$ (the case $u = 0$ is trivial). Since $\Theta \times A$ is open and convex we obtain, by Taylor's theorem,

$$\log \frac{L_n\left(\theta + \frac{u_1}{\sqrt{n}}, \alpha + \frac{u_2}{\sqrt{n}} \mid X_{-p}, \ldots, X_n\right)}{L_n\left(\theta, \alpha \mid X_{-p}, \ldots, X_n\right)} = u^T S_n(\theta, \alpha) - \frac{1}{2} u^T J_n(\tilde{\theta}_n, \tilde{\alpha}_n) u, \tag{14}$$

where $(\tilde{\theta}_n, \tilde{\alpha}_n)$ is a random point on the line-segment between (θ, α) and $(\theta + u_1/\sqrt{n}, \alpha + u_2/\sqrt{n})$ and

$$J_n(\theta, \alpha) = -\frac{1}{\sqrt{n}} \frac{\partial}{\partial(\theta, \alpha)^T} S_n(\theta, \alpha). \tag{15}$$

First, we give some auxiliary calculations in Part 0. In Part 1 we show that $S_n(\theta,\alpha) \xrightarrow{d} \mathrm{N}(0,J)$ under $\mathbb{P}_{\nu,\theta,\alpha}$, in Part 2 we prove that $J_n(\tilde{\theta}_n,\tilde{\alpha}_n) \xrightarrow{p} J$ under $\mathbb{P}_{\nu,\theta,\alpha}$, and, finally, in Part 3 we prove the non-singularity of J.

Part 0: auxiliary calculations

In this part we show that certain expressions are integrable, which is needed in Step 1 and Step 2.

It is easy to see that, for $\theta \in (0,1)$, $\ell \in \mathbb{N}$, we have

$$\frac{\partial^\ell}{\partial\theta^\ell}\log \mathrm{b}_{x,\theta}(k) = (-1)^{\ell+1}(\ell-1)!\frac{k}{\theta^\ell} - (\ell-1)!\frac{x-k}{(1-\theta)^\ell},$$

and hence

$$\left|\frac{\partial^\ell}{\partial\theta^\ell}\log \mathrm{b}_{x,\theta}(k)\right| \le (\ell-1)!x\left(\frac{1}{(1-\theta)^\ell}\vee\frac{1}{\theta^\ell}\right) \le (\ell-1)!\frac{x}{(1-\theta)^\ell\theta^\ell}. \tag{16}$$

For notational convenience we denote $\dot{s}_{x,\theta}(k) = (\partial/\partial\theta)\log \mathrm{b}_{x,\theta}(k)$ and $\ddot{s}_{x,\theta}(k) = (\partial^2/\partial\theta^2)\log \mathrm{b}_{x,\theta}(k)$. From (6) and (16) we obtain the bound

$$\left|\dot{\ell}_{\theta,i}(X_{-p},\dots,X_0;\theta,\alpha)\right| \le \frac{1}{\theta_i(1-\theta_i)}X_{-i}.$$

From Assumption (A3) on $\mathcal{G}_A$ we obtain $\delta > 0$. If necessary, decrease δ such that the ball round θ with radius δ is a subset of Θ. Of course, this has no influence on the validity of (3) and (4).

Using the previous display and Cauchy-Schwarz, we obtain, for $i,j = 1,\dots,p$,

$$\begin{aligned}\mathbb{E}_{\nu_0,\theta,\alpha}\sup_{(\tilde{\theta},\tilde{\alpha}):|(\tilde{\theta},\tilde{\alpha})-(\theta,\alpha)|<\delta}\left|\dot{\ell}_{\theta,i}\dot{\ell}_{\theta,j}(X_{-p},\dots,X_0;\tilde{\theta},\tilde{\alpha})\right| &\le M_\theta\sqrt{\mathbb{E}_{\nu_0,\theta,\alpha}X_{-i}^2\mathbb{E}_{\nu_0,\theta,\alpha}X_{-j}^2}\\ &= M_\theta\mathbb{E}_{\nu_0,\theta,\alpha}X_0^2 < \infty,\end{aligned} \tag{17}$$

where $M_\theta \in \mathbb{R}_+$. Using (3) from Assumption (A3) on $\mathcal{G}_A$ we obtain, for $i,j = 1,\dots,q$,

$$\begin{aligned}&\mathbb{E}_{\nu_0,\theta,\alpha}\sup_{(\tilde{\theta},\tilde{\alpha}):|(\tilde{\theta},\tilde{\alpha})-(\theta,\alpha)|<\delta}|\dot{\ell}_{\alpha,i}\dot{\ell}_{\alpha,j}(X_{-p},\dots,X_0;\tilde{\theta},\tilde{\alpha})|\\ &\quad\le \mathbb{E}_{\nu_0,\theta,\alpha}\sup_{(\tilde{\theta},\tilde{\alpha}):|(\tilde{\theta},\tilde{\alpha})-(\theta,\alpha)|<\delta}\sqrt{\mathbb{E}_{\tilde{\theta},\tilde{\alpha}}\left[|h_{\tilde{\alpha},i}(\varepsilon_0)|^2 \mid X_0,\dots,X_{-p}\right]\mathbb{E}_{\tilde{\theta},\tilde{\alpha}}\left[|h_{\tilde{\alpha},j}(\varepsilon_0)|^2 \mid X_0,\dots,X_{-p}\right]}\\ &\quad\le \mathbb{E}_{\nu_0,\theta,\alpha}M_1^{\theta,\alpha} < \infty.\end{aligned} \tag{18}$$

Using Cauchy-Schwarz, (16) and (3) from Assumption (A3) on $\mathcal{G}_A$ we also have, for $i = 1,\dots,p$, $j = 1,\dots,q$,

$$\mathbb{E}_{\nu_0,\theta,\alpha}\sup_{(\tilde{\theta},\tilde{\alpha}):|(\tilde{\theta},\tilde{\alpha})-(\theta,\alpha)|<\delta}\left|\dot{\ell}_{\theta,i}\dot{\ell}_{\alpha,j}(X_{-p},\dots,X_0;\tilde{\theta},\tilde{\alpha})\right| \le M_\theta\sqrt{\mathbb{E}_{\nu_0,\theta,\alpha}M_1^{\theta,\alpha}\mathbb{E}_{\nu_0,\theta,\alpha}X_0^2} < \infty. \tag{19}$$

In the same way as we derived (6) we obtain the representations,

$$\frac{\frac{\partial^2}{\partial \theta_i^2} P^{\theta,\alpha}_{(X_{-1},\ldots,X_{-p}),X_0}}{P^{\theta,\alpha}_{(X_{-1},\ldots,X_{-p}),X_0}} = \mathbb{E}_{\theta,\alpha}\left[\ddot{s}_{X_{-i},\theta_i}(\vartheta_i \circ X_{-i}) + \dot{s}^2_{X_{-i},\theta_i}(\vartheta_i \circ X_{-i}) \mid X_0,\ldots,X_{-p}\right],$$

and for $i \neq j$,

$$\frac{\frac{\partial^2}{\partial \theta_j \partial \theta_i} P^{\theta,\alpha}_{(X_{-1},\ldots,X_{-p}),X_0}}{P^{\theta,\alpha}_{(X_{-1},\ldots,X_{-p}),X_0}} = \mathbb{E}_{\theta,\alpha}\left[\dot{s}_{X_{-i},\theta_i}(\vartheta_i \circ X_{-i})\dot{s}_{X_{-j},\theta_j}(\vartheta_j \circ X_{-j}) \mid X_0,\ldots,X_{-p}\right].$$

Using (16) we obtain the bound,

$$\left|\frac{\frac{\partial^2}{\partial \theta_i \partial \theta_j} P^{\theta,\alpha}_{(X_{-1},\ldots,X_{-p}),X_0}}{P^{\theta,\alpha}_{(X_{-1},\ldots,X_{-p}),X_0}}\right| \leq \frac{1}{\theta_i(1-\theta_i)}\frac{1}{\theta_j(1-\theta_j)}\left(X_{-i}^2 + X_{-j}^2\right),$$

which implies,

$$\mathbb{E}_{\nu_0,\theta,\alpha} \sup_{(\tilde\theta,\tilde\alpha):|(\tilde\theta,\tilde\alpha)-(\theta,\alpha)|<\delta} \left|\frac{\frac{\partial^2}{\partial \theta_i \partial \theta_j} P^{\tilde\theta,\tilde\alpha}_{(X_{-1},\ldots,X_{-p}),X_0}}{P^{\tilde\theta,\tilde\alpha}_{(X_{-1},\ldots,X_{-p}),X_0}}\right| < \infty. \tag{20}$$

In the same way as we derived (7) we obtain the representation

$$\frac{\frac{\partial^2}{\partial \alpha_i \partial \alpha_j} P^{\theta,\alpha}_{(X_{-1},\ldots,X_{-p}),X_0}}{P^{\theta,\alpha}_{(X_{-1},\ldots,X_{-p}),X_0}} = \mathbb{E}_{\theta,\alpha}\left[\dot{h}_{\alpha,ji}(\varepsilon_0) + h_{\alpha,j}(\varepsilon_0)h_{\alpha,i}(\varepsilon_0) \mid X_0,\ldots,X_{-p}\right].$$

Using (3) from Assumption (A3) on $\mathcal{G}_A$ we obtain,

$$\begin{aligned}
&\mathbb{E}_{\nu_0,\theta,\alpha} \sup_{(\tilde\theta,\tilde\alpha):|(\tilde\theta,\tilde\alpha)-(\theta,\alpha)|<\delta} \left|\mathbb{E}_{\tilde\theta,\tilde\alpha}\left[h_{\tilde\alpha,j}(\varepsilon_0)h_{\tilde\alpha,i}(\varepsilon_0) \mid X_0,\ldots,X_{-p}\right]\right| \\
&\leq \mathbb{E}_{\nu_0,\theta,\alpha} \sup_{(\tilde\theta,\tilde\alpha):|(\tilde\theta,\tilde\alpha)-(\theta,\alpha)|<\delta} \sqrt{\mathbb{E}_{\tilde\theta,\tilde\alpha}\left[|h_{\tilde\alpha,j}(\varepsilon_0)|^2 \mid X_0,\ldots,X_{-p}\right]\mathbb{E}_{\tilde\theta,\tilde\alpha}\left[|h_{\tilde\alpha,i}(\varepsilon_0)|^2 \mid X_0,\ldots,X_{-p}\right]} \\
&\leq \mathbb{E}_{\nu_0,\theta,\alpha} M_1^{\theta,\alpha} < \infty.
\end{aligned}$$

Hence, an combination with (4) from Assumption (A3), yields,

$$\mathbb{E}_{\nu_0,\theta,\alpha} \sup_{(\tilde\theta,\tilde\alpha):|(\tilde\theta,\tilde\alpha)-(\theta,\alpha)|<\delta} \left|\frac{\frac{\partial^2}{\partial \alpha_i \partial \alpha_j} P^{\tilde\theta,\tilde\alpha}_{(X_{-1},\ldots,X_{-p}),X_0}}{P^{\tilde\theta,\tilde\alpha}_{(X_{-1},\ldots,X_{-p}),X_0}}\right| < \infty. \tag{21}$$

Next we compute for $i = 1, \dots, p$, $j = 1, \dots, q$, the representation,

$$\frac{\frac{\partial^2}{\partial \alpha_j \partial \theta_i} P^{\theta,\alpha}_{(X_{-1},\dots,X_{-p}),X_0}}{P^{\theta,\alpha}_{(X_{-1},\dots,X_{-p}),X_0}} = \mathbb{E}_{\theta,\alpha}\left[h_{\alpha,j}(\varepsilon_0)\dot{s}_{X_{-i},\theta_i}(\vartheta_i \circ X_{-i}) \mid X_0, \dots, X_{-p}\right],$$

which, using (16) and (3), yields,

$$\begin{aligned}\mathbb{E}_{\nu_0,\theta,\alpha} & \sup_{(\tilde{\theta},\tilde{\alpha}):|(\tilde{\theta},\tilde{\alpha})-(\theta,\alpha)|<\delta} \left|\mathbb{E}_{\tilde{\theta},\tilde{\alpha}}\left[h_{\tilde{\alpha},j}(\varepsilon_0)\dot{s}_{X_{-i},\tilde{\theta}_i}(\vartheta_i \circ X_{-i}) \mid X_0, \dots, X_{-p}\right]\right| \\ & \leq M_\theta \sqrt{\mathbb{E}_{\nu_0,\theta,\alpha} X_0^2} \sqrt{\mathbb{E}_{\nu_0,\theta,\alpha} M_1^{\theta,\alpha}} < \infty. \end{aligned} \tag{22}$$

<u>Part 1: the score</u>

From (6) it follows that,

$$\mathbb{E}_{\theta,\alpha}\left[\dot{\ell}_{\theta,i}(X_{t-p},\dots,X_t;\theta,\alpha) \mid X_{t-1},\dots,X_{t-p}\right] = \mathbb{E}_{\theta,\alpha}\left[\dot{s}_{X_{t-i},\theta_i}(\vartheta_i \circ X_{t-i}) \mid X_{t-1},\dots,X_{t-p}\right] = 0, \tag{23}$$

since $\vartheta_i \circ X_{t-i}$, conditional on $X_{t-p}, \dots, X_{t-1}$, has expectation $\theta_i X_{t-i}$. From (7) it follows that,

$$\mathbb{E}_{\theta,\alpha}\left[\dot{\ell}_{\alpha,j}(X_{t-p},\dots,X_t;\theta,\alpha) \mid X_{t-1},\dots,X_{t-p}\right] = \mathbb{E}_{\theta,\alpha}\left[h_{\alpha,j}(\varepsilon_t) \mid X_{t-1},\dots,X_{t-p}\right] = 0, \tag{24}$$

since ε_t is independent of $X_{t-p}, \dots, X_{t-1}$ and $\mathbb{E}_\alpha h_{\alpha,j}(\varepsilon_0) = 0$. Let $w = (w_1, w_2) \in \mathbb{R}^p \times \mathbb{R}^q$. From (23) and (24) it follows that,

$$\mathbb{E}_{\theta,\alpha}\left[w_1^T \dot{\ell}_\theta(X_{t-p},\dots,X_t;\theta,\alpha) + w_2^T \dot{\ell}_\alpha(X_{t-p},\dots,X_t;\theta,\alpha) \mid X_{t-1},\dots,X_{t-p}\right] = 0,$$

and, by (17) and (18),

$$\mathbb{E}_{\nu_0,\theta,\alpha}\left[w_1^T \dot{\ell}_\theta(X_{-p},\dots,X_0;\theta,\alpha) + w_2^T \dot{\ell}_\alpha(X_{-p},\dots,X_0;\theta,\alpha)\right]^2 = w^T J w < \infty.$$

Hence we have, by Lemma A.3,

$$\frac{1}{\sqrt{n}} \sum_{t=0}^{n} \left[w_1^T \dot{\ell}_\theta(X_{t-p},\dots,X_t;\theta,\alpha) + w_2^T \dot{\ell}_\alpha(X_{t-p},\dots,X_t;\theta,\alpha)\right] \xrightarrow{d} w^T \mathrm{N}(0,J), \text{ under } \mathbb{P}_{\nu,\theta,\alpha}.$$

Display (9) now follows by applying the Cramér-Wold device, which concludes Part 1.

<u>Part 2: the Fisher information</u>

In this part we prove that $J_n(\tilde{\theta}_n, \tilde{\alpha}_n) \xrightarrow{p} J$ under $\mathbb{P}_{\nu,\theta,\alpha}$, where

$$J_n(\theta,\alpha) = \begin{pmatrix} J_n^{\theta} & J_n^{\theta,\alpha} \\ J_n^{\alpha,\theta} & J_n^{\alpha} \end{pmatrix} = -\frac{1}{n}\sum_{t=0}^{n} \begin{pmatrix} \frac{\partial}{\partial\theta^T}\dot{\ell}_\theta(X_{t-p},\ldots,X_t;\theta,\alpha) & \frac{\partial}{\partial\alpha^T}\dot{\ell}_\theta(X_{t-p},\ldots,X_t;\theta,\alpha) \\ \frac{\partial}{\partial\theta^T}\dot{\ell}_\alpha(X_{t-p},\ldots,X_t;\theta,\alpha) & \frac{\partial}{\partial\alpha^T}\dot{\ell}_\alpha(X_{t-p},\ldots,X_t;\theta,\alpha) \end{pmatrix}. \tag{25}$$

Using Assumption (A2) on $\mathcal{G}_A$ it is easy to see that for fixed $x_{-p},\ldots,x_0 \in \mathbb{Z}_+$ the maps $(\theta,\alpha) \mapsto (\partial/\partial\theta)\log\dot{\ell}_\theta(x_{-p},\ldots,x_0;\theta,\alpha)$, $(\theta,\alpha) \mapsto (\partial/\partial\theta)\log\dot{\ell}_\alpha(x_{-p},\ldots,x_0;\theta,\alpha)$, $(\theta,\alpha) \mapsto (\partial/\partial\alpha)\log\dot{\ell}_\theta(x_{-p},\ldots,x_0;\theta,\alpha)$ and $(\theta,\alpha) \mapsto (\partial/\partial\alpha)\log\dot{\ell}_\alpha(x_{-p},\ldots,x_0;\theta,\alpha)$ all are continuous. Since we already proved (17), (18), (19), (20), (21), and (22), it is sufficient, by Lemma A.5, to prove that we have $J_n(\theta,\alpha) \xrightarrow{p} J$.

First we consider the diagonal of J_n^{θ}. For $i \in \{1,\ldots,p\}$, the calculations in Part 0 and Lemma A.2, yield,

$$\begin{aligned} J_{n,ii}^{\theta} &\xrightarrow{p} -\mathbb{E}_{\nu_0,\theta,\alpha}\Big[\mathbb{E}_{\theta,\alpha}\Big[\ddot{s}_{X_{-i},\theta_i}(\vartheta_i \circ X_{-i}) + \dot{s}^2_{X_{-i},\theta_i}(\vartheta_i \circ X_{-i}) \mid X_0,\ldots,X_{-p}\Big] - \dot{\ell}^2_{\theta,i}(X_{-p},\ldots,X_0;\theta,\alpha)\Big] \\ &= J_{ii}^{\theta}, \end{aligned}$$

where the last equality follows from,

$$\begin{aligned} &\mathbb{E}_{\nu_0,\theta,\alpha}\Big[\ddot{s}_{X_{-i},\theta_i}(\vartheta_i \circ X_{-i}) + \dot{s}^2_{X_{-i},\theta_i}(\vartheta_i \circ X_{-i})\Big] \\ &\quad = \mathbb{E}_{\nu_0,\theta,\alpha}\mathbb{E}_{\theta,\alpha}\Big[\ddot{s}_{X_{-i},\theta_i}(\vartheta_i \circ X_{-i}) + \dot{s}^2_{X_{-i},\theta_i}(\vartheta_i \circ X_{-i}) \mid X_{-1},\ldots,X_{-p}\Big] = 0, \end{aligned}$$

which is standard once one realizes that $\vartheta_i \circ X_{-i}$ given $X_{-p},\ldots,X_{-1}$ is $\mathrm{Bin}_{X_{-i},\theta_i}$ distributed.

Next we consider the off-diagonal elements of J^{θ}. Let $i \neq j$. Applying the representations in Part 0 and Lemma A.2 gives,

$$\begin{aligned} J_{n,ij}^{\theta} &\xrightarrow{p} -\mathbb{E}_{\nu_0,\theta,\alpha}\Big[\mathbb{E}_{\theta,\alpha}\big[\dot{s}_{X_{-i},\theta_i}(\vartheta_i \circ X_{-i})\dot{s}_{X_{-j},\theta_j}(\vartheta_j \circ X_{-j}) \mid X_0,\ldots,X_{-p}\big] - \dot{\ell}_{\theta,i}\dot{\ell}_{\theta,j}(X_{-p},\ldots,X_0;\theta,\alpha)\Big] \\ &= \mathbb{E}_{\nu_0,\theta,\alpha}\dot{\ell}_{\theta,i}\dot{\ell}_{\theta,j}(X_{-p},\ldots,X_0;\theta,\alpha) = J_{ij}^{\theta}, \end{aligned}$$

since,

$$\begin{aligned} &\mathbb{E}_{\nu_0,\theta,\alpha}\dot{s}_{X_{-i},\theta_i}(\theta_i \circ X_{-i})\dot{s}_{X_{-j},\theta_j}(\vartheta_j \circ X_{-j}) \\ &\quad = \mathbb{E}_{\nu_0,\theta,\alpha}\mathbb{E}_{\theta,\alpha}\big[\dot{s}_{X_{-i},\theta_i}(\vartheta_i \circ X_{-i})\dot{s}_{X_{-j},\theta_j}(\vartheta_j \circ X_{-j}) \mid X_{-1},\ldots,X_{-p}\big] = 0, \end{aligned}$$

because $\vartheta_i \circ X_{-i}$ and $\vartheta_j \circ X_{-j}$ given $X_{-p},\ldots,X_{-1}$ are mean-zero and independent.

Next we consider the block $J_n^{\theta,\alpha}$ (by symmetry this also yields the result for the block $J_n^{\alpha,\theta}$). Using the representations derived in Part 0 and Lemma A.2 we obtain,

$$\begin{aligned} J_{n,ij}^{\theta,\alpha} &\xrightarrow{p} -\mathbb{E}_{\nu_0,\theta,\alpha}\Big[\mathbb{E}_{\theta,\alpha}\big[h_{\alpha,j}(\varepsilon_0)\dot{s}_{X_{-i},\theta_i}(\vartheta_i \circ X_{-i}) \mid X_0,\ldots,X_{-p}\big] - \dot{\ell}_{\alpha,j}\dot{\ell}_{\theta,i}(X_{-p},\ldots,X_0;\theta,\alpha)\Big] \\ &= \mathbb{E}_{\nu_0,\theta,\alpha}\dot{\ell}_{\alpha,j}\dot{\ell}_{\theta,i}(X_{-p},\ldots,X_0;\theta,\alpha) = I_{ij}^{\theta,\alpha}, \end{aligned}$$

since,

$$\mathbb{E}_{\nu_0,\theta,\alpha}\left[\mathbb{E}_{\theta,\alpha}\left[h_{\alpha,j}(\varepsilon_0)\dot{s}_{X_{-i},\theta_i}(\vartheta_i \circ X_{-i}) \mid X_0,\ldots,X_{-p}\right]\right] = \mathbb{E}_{\nu_0,\theta,\alpha} h_{\alpha,j}(\varepsilon_0)\dot{s}_{X_{-i},\theta_i}(\vartheta_i \circ X_{-i}) = 0,$$

because $h_{\alpha,j}(\varepsilon_0)$ and $\dot{s}_{X_{-i},\theta_i}(\vartheta_i \circ X_{-i})$ are independent and have mean zero.
Finally we treat J_n^α. Using the representations in Part 0 and Lemma A.2 again, we obtain

$$\begin{aligned} J_n^\alpha &\xrightarrow{p} -\mathbb{E}_{\nu_0,\theta,\alpha}\left[\mathbb{E}_{\theta,G}\left[\dot{h}_\alpha(\varepsilon_0) + h_\alpha h_\alpha^T(\varepsilon_0) \mid X_0,\ldots,X_{-p}\right] - \dot{\ell}_\alpha \dot{\ell}_\alpha^T(X_{-p},\ldots,X_0;\theta,\alpha)\right] \\ &= J^\alpha, \end{aligned}$$

since, by Assumption (A4) on $\mathcal{G}_A$,

$$\mathbb{E}_{\nu_0,\theta,\alpha}\mathbb{E}_{\theta,\alpha}\left[\dot{h}_\alpha(\varepsilon_0) + h_\alpha h_\alpha^T(\varepsilon_0) \mid X_0,\ldots,X_{-p}\right] = \mathbb{E}_\alpha \dot{h}_\alpha(\varepsilon_0) + \mathbb{E}_\alpha h_\alpha h_\alpha^T(\varepsilon_0) = 0.$$

Part 3: non-singularity of I

Finally we show that J is non-singular. First we prove that J^α is non-singular. If J^α would be singular we would have,

$$a_2^T \dot{\ell}_\alpha\left(X_{-p},\ldots,X_0;\theta,\alpha\right) = 0 \quad \mathbb{P}_{\nu_0,\theta,\alpha}\text{-a.s. for certain } a_2 \in \mathbb{R}^q \backslash \{0\}.$$

Note that we have, for all $k \in \text{support}(G_\alpha)$, $\mathbb{P}_{\nu_0,\theta,\alpha}\left\{X_{-p} = \cdots = X_{-1} = 0, X_0 = k\right\} > 0$, and on the event $E_k = \{X_{-p} = \cdots = X_{-1} = 0, X_0 = k\}$ we have $\varepsilon_0 = k$. Hence we obtain, on the event E_k,

$$0 = a_2^T \dot{\ell}_\alpha\left(X_{-p},\ldots,X_0;\theta,\alpha\right) = a_2^T \mathbb{E}_{\theta,\alpha}\left[h(\varepsilon_0) \mid X_0,\ldots,X_{-p}\right] = a_2^T h_\alpha(k) \text{ for all } k \in \text{support}(G_\alpha),$$

which contradicts Assumption (A4) on $\mathcal{G}_A$ that $\mathbb{E}_\alpha h_\alpha h_\alpha^T(\varepsilon_0)$ is non-singular. Hence J^α is indeed non-singular.
Suppose that

$$a_1^T \dot{\ell}_\theta\left(X_{-p},\ldots,X_0;\theta,\alpha\right) + a_2^T \dot{\ell}_\alpha\left(X_{-p},\ldots,X_0;\theta,\alpha\right) = 0 \quad \mathbb{P}_{\nu_0,\theta,\alpha} - \text{a.s.} \tag{26}$$

Let $i \in \{1,\ldots,p\}$ and note that for $k \in \mathbb{Z}_+$ the event $\{X_j = 0 \text{ for } j \in \{-p,\ldots,0\}\backslash\{-i\},\ X_{-i} = k\}$ has positive probability under $\mathbb{P}_{\nu_0,\theta,\alpha}$ and that on this event we have,

$$\dot{\ell}_{\theta,i}\left(X_{-p},\ldots,X_0;\theta,\alpha\right) = \mathbb{E}_{\theta,\alpha}\left[\dot{s}_{X_{-i},\theta_i}(\vartheta_i \circ X_{-i}) \mid X_0,\ldots,X_{-p}\right] = -\frac{\theta_i k}{\theta_i(1-\theta_i)},$$

for $j \neq i$ we have,

$$\dot{\ell}_{\theta,j}\left(X_{-p},\ldots,X_0;\theta,\alpha\right)=\mathbb{E}_{\theta,\alpha}\left[\dot{s}_{X_{-j},\theta_j}(\vartheta_j\circ X_{-j})\mid X_0,\ldots,X_{-p}\right]=0,$$

and,

$$\dot{\ell}_{\alpha,m}(X_{-p},\ldots,X_{-0};\theta,\alpha)=\mathbb{E}_{\theta,\alpha}\left[h_{\alpha,m}(\varepsilon_0)\mid X_0,\ldots,X_{-p}\right]=h_{\alpha,m}(0).$$

Hence we obtain from (26), for $k\in\mathbb{Z}_+, i=1,\ldots,p$, the equality,

$$\frac{-a_{1,i}\theta_i k}{\theta_i(1-\theta_i)}+a_2^T h_\alpha(0)=0,$$

which is only possible if $a_1=0$. Hence $a_1=0$, so from (26) we get $a_2^T\dot{\ell}_\alpha\left(X_{-p},\ldots,X_0;\theta,\alpha\right)=0$ $\mathbb{P}_{\nu_0,\theta,\alpha}$-a.s. This is only possible if $a_2=0$, since we already proved that J^α is non-singular. We conclude that J is non-singular. □

Proof of Proposition 3.3:
By Prohorov's theorem it suffices to prove that there exists a subsequence n_k such that $\sqrt{n_k}(\hat{\alpha}_{n_k}-\alpha)$ converges in distribution.
Note first that, by Lemma A.2,

$$\frac{N_n}{n}\to\nu_{\theta,\alpha}\{0,\ldots,0\}>0,\quad\mathbb{P}_{\nu,\theta,\alpha}-\text{a.s.}$$

Let, for $u\in\mathbb{R}^q$,

$$\phi_n(u)=\mathbb{E}_\alpha\exp\left(iu^T(\sqrt{n}(t_n(\varepsilon_1,\ldots,\varepsilon_n)-\alpha))\right).$$

Since $t_n(\varepsilon_1,\ldots,\varepsilon_n)$ is a $\sqrt{n}$-consistent estimator of α, there exists, by Prohorov's theorem, a subsequence n_k such that $\sqrt{n_k}(t_{n_k}(\varepsilon_1,\ldots,\varepsilon_{n_k})-\alpha)$ converges in distribution under $\mathbb{P}_{\nu,\theta,\alpha}$. Hence for all $u\in\mathbb{R}^q$,

$$\lim_{k\to\infty}\phi_{n_k}(u)=\phi(u),$$

where ϕ is a characteristic function of an $\mathbb{R}^q$-valued random variable, which we denote by Z. Using the strong Markov property, it is not very hard to see that $(X_{\tau_k})_{k\in\mathbb{N}}$ are i.i.d. G-distributed independent of N_n. Hence (use dominated convergence),

$$\lim_{k\to\infty}\mathbb{E}_{\nu,\theta,\alpha}\exp\left(iu^T(\sqrt{N_{n_k}}(\hat{\alpha}_{n_k}-\alpha))\right)=\lim_{k\to\infty}\mathbb{E}_{\nu,\theta,\alpha}\phi_{N_{n_k}}(u)=\mathbb{E}_{\nu,\theta,\alpha}\phi(u)=\phi(u),$$

which yields,

$$\sqrt{N_{n_k}}(\hat{\alpha}_{n_k} - \alpha) \xrightarrow{d} Z, \text{ under } \mathbb{P}_{\nu,\theta,\alpha} \text{ as } k \to \infty.$$

Now,

$$\sqrt{n_k}(\hat{\alpha}_{n_k} - \alpha) = \sqrt{\frac{n_k}{N_{n_k}}}\sqrt{N_{n_k}}(\hat{\alpha}_{n_k} - \alpha) \xrightarrow{d} \frac{1}{\sqrt{\nu_{\theta,\alpha}\{0,\ldots,0\}}} Z, \text{ under } \mathbb{P}_{\nu,\theta,\alpha} \text{ as } k \to \infty,$$

which concludes the proof. □

Proof of Theorem 3.3:

Let $(\theta, \alpha) \in \Theta \times A$. To prove that $(\hat{\theta}_n^{**}, \hat{\alpha}_n^{**})$ is efficient at (θ, α) it suffices (see, for example, Theorem 2.3.1 in Bickel et al. (1998)) to prove that it is asymptotically linear in the efficient influence function at (θ, α), i.e.

$$\sqrt{n}\begin{pmatrix} \hat{\theta}_n^{**} - \theta \\ \hat{\alpha}_n^{**} - \alpha \end{pmatrix} = J^{-1}(\theta, \alpha) S_n(\theta, \alpha) + o(\mathbb{P}_{\nu,\theta,\alpha}; 1).$$

If we can show that the following conditions hold,

C1 $S_n(\theta, \alpha)$ converges in distribution under $\mathbb{P}_{\nu,\theta,\alpha}$;

C2 for every deterministic sequence $(\theta_n, \alpha_n) = (\theta_0, \alpha_0) + O(1/\sqrt{n})$ we have,

$$S_n(\theta_n, \alpha_n) - S_n(\theta, \alpha) + J(\theta, \alpha)\sqrt{n}\begin{pmatrix} \theta_n - \theta \\ \alpha_n - \alpha \end{pmatrix} \xrightarrow{p} 0, \text{ under } \mathbb{P}_{\nu,\theta,\alpha};$$

C3 $\hat{J}_n \xrightarrow{p} J(\theta, \alpha)$ under $\mathbb{P}_{\nu,\theta,\alpha}$,

then we obtain, from Theorem 5.48 in Van der Vaart (2000) ($(\hat{\theta}_n^{*}, \hat{\alpha}_n^{*})$ is consistent and discretized) the desired result.

Condition 1 has already been proved in Part 1 of the proof of Theorem 3.1; Condition 3 is proved in Part 0 and Part 2 of the proof of Theorem 3.1. Let $(\theta_n, \alpha_n) = (\theta, \alpha) + O(n^{-1/2})$ be a deterministic sequence. From the proof of Theorem 3.1 we have,

$$S_n(\theta_n, \alpha_n) = S_n(\theta, \alpha) - J_n(\tilde{\theta}_n, \tilde{\alpha}_n)\sqrt{n}\begin{pmatrix} \theta_n - \theta \\ \alpha_n - \alpha \end{pmatrix},$$

where $(\tilde{\theta}_n, \tilde{\alpha}_n)$ is a point between (θ, α) and (θ_n, α_n). Using Part 0 in the proof of Theorem 3.1 and Lemma A.5 we obtain $J_n(\tilde{\theta}_n, \tilde{\alpha}_n) \xrightarrow{p} J(\theta, \alpha)$ under $\mathbb{P}_{\nu,\theta,\alpha}$. This yields Condition 2, which concludes the proof. □

C Review of modern notion of efficiency for parametric models

This appendix briefly reviews the modern notion of asymptotic efficiency for parametric models. For details and proofs see, for example, Le Cam and Yang (1990), Bickel et al. (1998), and Van der Vaart (2000). Given a sequence of statistical experiments $\mathcal{E}_n = (\mathcal{X}_n, \mathcal{A}_n, (P_{n,h} \mid h \in H))$, $n \in \mathbb{N}$, suppose that the parameter space H is an open subset of $\mathbb{R}^k$. The sequence of experiments $(\mathcal{E}_n)_{n\in\mathbb{N}}$ has the Local Asymptotic Normality (LAN) property in $h_0 \in H$ if there exist a $k \times k$ information-matrix I_{h_0}, a score Δ_{n,h_0} with $\Delta_{n,h_0} \xrightarrow{d} \mathrm{N}_k(0, I_{h_0})$ under $(P_{n,h_0})_{n\in\mathbb{N}}$ such that for all $a \in \mathbb{R}^k$

$$\log \frac{\mathrm{d}P_{n,h_0+a/\sqrt{n}}}{\mathrm{d}P_{n,h_0}} = a^T \Delta_{n,h_0} - \frac{1}{2} a^T I_{h_0} a + R_n, \tag{27}$$

where $R_n = R_n(a, h_0) \xrightarrow{p} 0$ under $(P_{n,h_0})_{n\in\mathbb{N}}$. $(\mathcal{E}_n)_{n\in\mathbb{N}}$ has the Local Asymptotic Normality (LAN) property if it has the LAN-property for all $h_0 \in H$. A method to prove efficiency of an estimator is to provide a lower-bound to the precision of an estimator and subsequently device an estimator that attains this bound. For sequences of experiments that have the LAN-property the Hájek-Le Cam convolution theorem gives a bound to the accuracy of regular estimators. Before we state this important theorem, we recall the concept of a regular estimator. An estimator $(T_n)_{n\in\mathbb{N}}$ of h is regular in h_0, if for all $a \in \mathbb{R}^k$, we have $\sqrt{n}\,(T_n - (h_0 + a/\sqrt{n})) \xrightarrow{d} L_{h_0}$, under $(P_{n,h_0+a/\sqrt{n}})_{n\in\mathbb{N}}$, where the limit-distribution L_{h_0} does not depend on a. If this holds for all $h_0 \in H$ then we say that T_n is a regular estimator of h. An interpretation is that for regular estimators the convergence to the limiting-distribution is, in some sense, uniform. Now we can state the Hájek-Le Cam convolution theorem.

Theorem C.1 *Assume that $\mathcal{E}_n = (\mathcal{X}_n, \mathcal{A}_n, (P_{n,h} \mid h \in H))$, $n \in \mathbb{N}$ has the LAN-property in h_0 with non-singular information-matrix I_{h_0} and score Δ_{n,h_0}. If T_n is a regular estimator of h in h_0, then we have*

$$\begin{pmatrix} \sqrt{n}\left(T_n - h_0 - \frac{1}{\sqrt{n}} I_{h_0}^{-1} \Delta_{n,h_0}\right) \\ I_{h_0}^{-1} \Delta_{n,h_0} \end{pmatrix} \xrightarrow{d} \begin{pmatrix} Y_{h_0} \\ Z_{h_0} \end{pmatrix}, \text{ under } (P_{n,h_0})_{n\in\mathbb{N}}, \tag{28}$$

where Y_{h_0} and $Z_{h_0} \sim \mathrm{N}(0, I_{h_0}^{-1})$ are independent.

So the limiting distribution of a regular estimator T_n is the convolution of a $\mathrm{N}(0, I_{h_0}^{-1})$ distribution, which only depends on the model, and a distribution which will also depend on the estimator. Hence the scaled estimation error $\sqrt{n}(T_n - h_0)$ consists of an unavoidable part Z_{h_0} and additional noise Y_{h_0}. Thus the following definition is indeed natural: an estimator T_n of h is efficient in h_0 if it is regular, and if $\sqrt{n}(T_n - h_0) \xrightarrow{d} \mathrm{N}(0, I_{h_0}^{-1})$ under $(P_{n,h_0})_{n\in\mathbb{N}}$. If T_n is efficient for all $h \in H$, then we call T_n an efficient estimator of h.

References

Al-Osh, M. and A. Alzaid (1987). First-order integer-valued autoregressive (INAR(1)) processes. *J. Time Ser. Anal.*, **8**, 261–75.

Al-Osh, M. and A. Alzaid (1990). An integer-valued pth-order autoregressive structure (INAR(p)) process. *J. Appl. Prob.*, **27**, 314–324.

Andrews, D. (1992). Generic uniform convergence. *Econometric Theory*, **8**, 241–257.

Berglund, E. and K. Brännäs (2001). Plants' entry and exit in Swedish municipalities. *Ann. Reg. Sci.*, **35**, 431–448.

Bickel, P., C. Klaassen, Y. Ritov, and J. Wellner (1998). *Efficient and adaptive estimation for semiparametric models* (2nd ed.). Springer.

Billingsley, P. (1995). *Probability and Measure* (2nd ed.). Wiley series in probability and mathematical statistics.

Böckenholt, U. (1999). Mixed INAR(1) Poisson regression models: analyzing heterogeneity and serial dependencies in longitudinal count data. *J. Econometrics*, **89**, 317–338.

Böckenholt, U. (2003). Analysing state dependences in emotional experiences by dynamic count data models. *J. Roy. Statist. Soc. Ser. C*, **52**, 213–226.

Brännäs, K. and J. Hellström (2001). Generalized integer-valued autoregression. *Econometric Rev.*, **20**, 425–443.

Brännäs, K. and Q. Shahiduzzaman (2004). Integer-valued moving average modelling of the number of transactions in stocks. Working paper Umeå Economic Studies 637.

Cardinal, M., R. Roy, and J. Lambert (1999). On the application of integer-valued time series models for the analysis of disease incidence. *Stat. Med.*, **18**, 2025–2039.

Dacunha-Castelle, D. and M. Duflo (1986). *Probability and Statistics* (1st ed.). Springer-Verlag.

Dion, J.-P., G. Gauthier, and A. Latour (1995). Branching processes with immigration and integer-valued time series. *Serdica Math. J.*, **21**, 123–136.

Drost, F., R. van den Akker, and B. Werker (2006). An asymptotic analysis of nearly unstable INAR(1) models. *Working paper*.

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

Feigin, P. and R. Tweedie (1985). Random coefficient autoregressive processes: a Markov chain analysis of stationarity and finiteness of moments. *J. Time Ser. Anal.*, **6**, 1–14.

Franke, J. and T. Seligmann (1993). Conditional maximum-likelihood estimates for INAR(1) processes and their applications to modelling epileptic seizure counts. In: T. Subba Rao (Ed.), *Developments in time series*, pp. 310–330. London: Chapman & Hall.

Freeland, R. and B. McCabe (2004). Analysis of low count time series data by Poisson autoregression. *J. Time Ser. Anal.*, **25**, 701–722.

Gourieroux, C. and J. Jasiak (2004). Heterogeneous INAR(1) model with application to car insurance. *Ins.: Mathematics Econ.*, **34**, 177–192.

Latour, A. (1998). Existence and stochastic structure of a nonnegative integer-valued autoregressive process. *J. Time Ser. Anal.*, **19**, 439–455.

Le Cam, L. and G. Yang (1990). *Asymptotics in statistics* (1 ed.). New-York: Springer-Verlag.

Roussas, G. (1972). *Contiguity of probability measures* (1st ed.). Cambridge university press.

Rudholm, N. (2001). Entry and the number of firms in the Swedish pharmaceuticals market. *Rev. Ind. Organ.*, **19**, 351–364.

Silva, M. and V. Oliveira (2005). Difference equations for the higher order moments and cumulants of the INAR(p) model. *J. Time Ser. Anal.*, **26**, 17–36.

Tweedie, R. (1983). The existence of moments for stationary Markov chains. *J. Appl. Prob.*, **20**, 191–196.

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