Electronic Journal of Statistics
Vol. 7 (2013) 793–819
ISSN: 1935-7524
DOI: 10.1214/13-EJS790

A goodness-of-fit test for Poisson count processes

Konstantinos Fokianos*

University of Cyprus
Department of Mathematics and Statistics
P.O. Box 20537
CY – 1678 Nicosia
Cyprus
e-mail: fokianos@ucy.ac.cy

and

Michael H. Neumann†

Friedrich-Schiller-Universität Jena
Institut für Stochastik
Ernst-Abbe-Platz 2
D – 07743 Jena
Germany
e-mail: michael.neumann@uni-jena.de

Abstract: We are studying a novel class of goodness-of-fit tests for parametric count time series regression models. These test statistics are formed by considering smoothed versions of the empirical process of the Pearson residuals. Our construction yields test statistics which are consistent against Pitman's local alternatives and they converge weakly at the usual parametric rate. To approximate the asymptotic null distribution of the test statistics, we propose a parametric bootstrap method and we study its properties. The methodology is applied to simulated and real data.

AMS 2000 subject classifications: Primary 60G10, 62M07; secondary 62F40.
Keywords and phrases: Autoregression, conditional mean, ergodicity, goodness-of-fit test, integer-valued processes, local alternatives.

Received July 2012.

Contents

*Supported by a Leventis Foundation Grant.
†Supported by the German Research Foundation DFG, project NE 606/2-2.
‡We would like to thank an anonymous reviewer for several comments and remarks that improved our work. The authors acknowledge the use of resources provided by the European Grid Infrastructure. For more information, please reference the EGI-InSPIRE paper (http://go.egi.eu/pdnon). In particular, K. Fokianos is acknowledging computing support by Prof. M. Dikaiakos and Ms. A. Bella (CyGrid).

1. Introduction

Recently, several authors have been developing theory for linear and nonlinear parametric models for regression analysis of count time series; see for instance Fokianos et al. [9], Neumann [19], Woodard et al. [27], Fokianos and Tjøstheim [11], Doukhan et al. [2] and Fokianos and Tjøstheim [11]. The main issues addressed in those contributions, are the study of conditions under which the observed process is ergodic and the associated problem of maximum likelihood estimation. However, a missing part of those studies is the advancement of goodness-of-fit methods for assessing the adequacy of the proposed models. In this contribution, we fill this void, by developing methodology for testing goodness of fit of parametric linear and nonlinear count time series models. The problem of goodness of fit has been studied recently by Neumann [19] but the proposed test statistic cannot be employed to detect local alternatives, in general. In our approach, this obstacle is removed because the proposed test statistics are based on a smooth approximation of the empirical process of Pearson residuals; see McCullagh and Nelder [18].

In what follows, we assume the following setup. Consider a stationary vector valued process $\{(Y_t, \lambda_t)\}$, for $t \in \mathbb{N}_0 = \mathbb{N} \cup \{0\}$. In general, the process $\{Y_t\}$ denotes the response count process which is assumed to follow the Poisson distribution conditionally on the past and $\{\lambda_t\}$ denotes the mean process. In other words, if we denote by $\mathcal{F}_t = \sigma(\lambda_0, Y_s \colon s \le t)$, we assume that the model is given by

$$Y_t \mid \mathcal{F}_{t-1} \sim \text{Poisson}(\lambda_t), \quad \lambda_t = f(\lambda_{t-1}, Y_{t-1}), \tag{1.1}$$

for some function $f \colon [0, \infty) \times \mathbb{N}_0 \to [\lambda_{\min}, \infty)$ with $\lambda_{\min} > 0$.

To understand the usefulness of model (1.1), suppose that a count time series is available and its autocorrelation function assumes relatively high values for large lags. Then, it is natural to model such data by including a large number of lagged regressor variables into a model. However such an approach can be avoided when employing model (1.1); it simply provides a parsimonious way to model this type of data; see Fokianos et al. [9] and Fokianos [8] for more. Such a modeling approach follows the spirit of GARCH models (see Bollerslev [1]) whereby the volatility is regressed to its past and lagged regressors. For the case of Poisson distribution, the mean equals the variance and therefore model (1.1) can be viewed as an integer GARCH type of model.

A special case of model (1.1) corresponds to the linear specification which is given by $\lambda_t = \theta_1 + \theta_2\lambda_{t-1} + \theta_3 Y_{t-1}$ and $\theta_1, \theta_2, \theta_3$ are nonnegative constants. Below we will assume that $\theta_1 > 0$ and $\theta_2 + \theta_3 < 1$. Rydberg and Shephard [22] proposed such a model for describing the number of trades on the New York Stock Exchange in certain time intervals and called it BIN(1,1) model. Stationarity and mixing properties for this model were derived by Streett [23], Ferland et al. [7] who referred to it as INGARCH(1,1) model, Fokianos et al. [9] and Neumann [19].

Recalling now the general model (1.1) we note that although estimation and inference is quite well developed, goodness-of-fit methods have not attracted a lot of attention. In this article we will be considering testing goodness of fit of model (1.1) by focusing on the following two forms of hypotheses. The first one refers to the simple hypothesis

$$H_0^{(s)}: \quad f = f_0 \qquad \text{against} \qquad H_1^{(s)}: \quad f \neq f_0, \tag{1.2}$$

for some completely specified function f_0 which satisfies the contractivity assumption **(C)** below. However, in applications, the most interesting testing problem is given by the following composite hypotheses

$$H_0: \quad f \in \{f_\theta : \theta \in \Theta\} \qquad \text{against} \qquad H_1: \quad f \notin \{f_\theta : \theta \in \Theta\}, \tag{1.3}$$

where $\Theta \subseteq \mathbb{R}^d$ and the function f_θ is known up to a parameter θ and satisfies assumption **(C)** as before.

Those testing problems have been considered in a recent contribution by Neumann [19]. Assuming that observations $Y_0, \ldots, Y_n$ are available he proposed a simple procedure for testing (1.2) based on the following test statistic

$$T_{n,0} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left\{ (Y_t - f_0(\widetilde{\lambda}_{t-1}, Y_{t-1}))^2 - Y_t \right\}.$$

Here, the initial value $\widetilde{\lambda}_0$ is arbitrarily chosen and, for $t = 1, \ldots, n$, $\widetilde{\lambda}_t = f_0(\widetilde{\lambda}_{t-1}, Y_{t-1})$. It is clear that the test statistic $T_{n,0}$ is based on the observation that the conditional variance of Y_t is equal to its mean because of the Poisson assumption. In the case of the composite hypotheses (1.3), the above test statistic is suitably modified as

$$T_n = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left\{ (Y_t - f_{\widehat{\theta}_n}(\widehat{\lambda}_{t-1}, Y_{t-1}))^2 - Y_t \right\},$$

where $\widehat{\theta}_n$ is any $\sqrt{n}$-consistent estimator of θ, $\widehat{\lambda}_0$ is chosen arbitrarily, and, as before, $\widehat{\lambda}_t = f_{\widehat{\theta}_n}(\widehat{\lambda}_{t-1}, Y_{t-1})$, for $t = 1, \ldots, n$. It turns out that both of these test statistics are asymptotically normally distributed and offer easily implemented specification tests for the intensity function f. However, they are not suitable to detect local alternatives in general. *In this contribution, we propose alternative test statistics which are able to detect local alternatives with* *a nontrivial asymptotic power.* Those test statistics are formed on the basis of Pearson residuals and a smoothing function as we show in Section 2. We point out that the smoothing function does not need to depend upon any bandwidth; in other words the proposed test statistic's power is not influenced by any suboptimal choice of a bandwidth. In developing these test statistics, we extend the theory of specification testing to integer valued time series. In fact, we study the asymptotic distribution of the proposed test statistics and we show that they are consistent against Pitman's local alternatives, at the usual rate of $n^{-1/2}$, where n denotes the sample size. In some sense, we extend the theory of nonparametric goodness-of-fit tests for autoregressions (Koul and Stute [16]) to include *count time series with a feedback mechanism*; recall (1.1). A major difference though between our work and that of Koul and Stute [16] is that the proposed methodology includes models with feedback mechanism and employs the empirical process of the Pearson residuals. Based on our approach, we relax the number of moments needed to obtain asymptotic results following a recent approach outlined by Escanciano [5]. Furthermore, the goodness-of-fit test statistics are easily computed and they are valid under a fairly large class of problems. To implement the proposed test statistic for testing (1.3) we take full advantage of model (1.1) by resorting to parametric bootstrap; see Section 3. It is shown theoretically that the bootstrap replications successfully approach the null distribution of the test statistic. Furthermore, Section 4 demonstrates empirically that this method works well in practice and does not depend on the particular smoothing function heavily.

2. Main results

Throughout this paper we assume that observations $Y_0, \ldots, Y_n$ are available and they are generated from a stationary process $((Y_t, \lambda_t))_{t \in T}$ which satisfies model (1.1). For simplicity of presentation we chose the index set $T = \mathbb{Z}$ which is always possible due to Kolmogorov's theorem. Before stating our main results, we will impose the following contractive condition on the function f:

Assumption (C).

$$|f(\lambda, y) \,-\, f(\lambda', y')| \,\leq\, \kappa_1 |\lambda - \lambda'| \,+\, \kappa_2 |y - y'| \qquad \forall \lambda, \lambda' \geq 0, \forall y, y' \in \mathbb{N}_0,$$

where κ_1 and κ_2 are nonnegative constants with $\kappa := \kappa_1 + \kappa_2 < 1$.

Assumption (C) is actually a key condition for proving absolute regularity of the count process $(Y_t)_{t \in \mathbb{Z}}$ and ergodicity of the bivariate process $((Y_t, \lambda_t))_{t \in \mathbb{Z}}$; see Neumann [19, Thm 3.1] but also Fokianos et al. [9, Prop. 2.3], where regularity of a perturbed version of model (1.1) is shown. The generality of assumption (C) is chosen to include also nonlinear specifications such as the exponential AR model proposed in Doukhan et al. [2] and Fokianos and Tjøstheim [11]. In this case, the intensity function is specified as $f(\lambda, y) = \theta_1 + (\theta_2 + \theta_3 \exp(-\theta_4 \lambda^2))\lambda + \theta_5 y$, where $\theta_1, \ldots, \theta_5 > 0$. It follows from $\partial f(\lambda, y)/\partial y = \theta_5$ and $|\partial f(\lambda, y)/\partial \lambda| = |\theta_2 + \theta_3 (1 - 2\theta_4 \lambda^2) \exp(-\theta_4 \lambda^2)| \leq \theta_2 + \theta_3$, that assumption (C) is fulfilled if $\theta_2 + \theta_3 + \theta_5 < 1$.

Adapting an idea from Gao et al. [14], Fokianos and Tjøstheim [11] proposed another nonlinear model given by the specification $f(\lambda, y) = \theta_1(1+\lambda)^{-\theta_2} + \theta_3\lambda + \theta_4 y$, $\theta_1, \ldots, \theta_4 > 0$. Then $\partial f(\lambda, y)/\partial y = \theta_4$ and $\partial f(\lambda, y)/\partial\lambda = \theta_3 - \theta_1\theta_2/(1+\lambda)^{1+\theta_2}$ which implies that assumption (C) is fulfilled if $\max\{\theta_3, \theta_1\theta_2 - \theta_3\} + \theta_4 < 1$. Several other non-linear specifications for count time series models can be studied within the framework of assumption (C), including smooth transition autoregressions; for an excellent survey see Teräsvirta et al. [25].

2.1. Test statistic for testing simple hypothesis

It is instructive first to consider the case of the simple hypothesis (1.2). We first give some notation. Denote by $I_t = (\lambda_t, Y_t)'$ the vector consisting of the values of the intensity process and the count process at time t. Recall that if $\widetilde{\lambda}_0$ is some initial value, then the process $\widetilde{\lambda}_t = f_0(\widetilde{\lambda}_{t-1}, Y_{t-1})$ can be recursively defined. Put $\widetilde{I}_t = (\widetilde{\lambda}_t, Y_t)'$. Define further $\xi_t = (Y_t - \lambda_t)/\sqrt{\lambda_t}$ and $\widetilde{\xi}_t = (Y_t - \widetilde{\lambda}_t)/\sqrt{\widetilde{\lambda}_t}$. Note that the sequence $\{\xi_t\}$ consists of the so called Pearson residuals; see Kedem and Fokianos [15, Ch.1]. For an $x \in \Pi := [0, \infty)^2$, define the following supremum type of test statistic by

$$\begin{aligned} G_n(x) &= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \widetilde{\xi}_t \, w(x - \widetilde{I}_{t-1}), \\ T_n &= \sup_{x\in\Pi} |G_n(x)|, \end{aligned} \tag{2.1}$$

where $w(\cdot)$ is some suitably defined weight function. In the applications, we consider the weight function to be of the form $w(x) = w(x_1, x_2) = K(x_1)K(x_2)$ where $K(\cdot)$ is a univariate kernel and $x = (x_1, x_2) \in \Pi$. We employ the uniform, Gaussian and the Epanechnikov kernels. For instance, when the uniform kernel is employed, compute the test statistic (2.1), by using the weights

$$w(x - I_{t-1}) = K(x_1 - \lambda_{t-1})K(x_2 - Y_{t-1}) = \frac{1}{4}\mathbb{1}(|x_1 - \lambda_{t-1}| \le 1)\mathbb{1}(|x_2 - Y_{t-1}| \le 1),$$

where $\mathbb{1}(A)$ is the indicator function of a set A. Then, the test statistic (2.1) becomes

$$T_n = \sup_{x\in\Pi} |G_n(x)|,$$

where

$$G_n(x_1, x_2) = \frac{1}{4\sqrt{n}} \sum_{t=1}^{n} \widetilde{\xi}_t \mathbb{1}(|x_1 - \widetilde{\lambda}_{t-1}| \le 1)\mathbb{1}(|x_2 - Y_{t-1}| \le 1),$$

Obvious formulas hold for the other kernel functions that we are using and for the case of the composite hypotheses that we consider in Section 2.2; see equation (2.2). Then Assumption (A1)(ii) is trivially satisfied for the Epanechnikov kernel. We include also in simulations test statistics formed after employing the uniform and Gaussian kernel even though they do not satisfy (A1)(ii); see below.

It is not clear whether there exists an optimal choice of the weight function $w(\cdot)$. However, given the theory developed in this paper, we suggest users to employ weight functions that satisfy the following assumption.

Assumption (A1).

(i) The true data generating process is given by (1.1) such that the contraction assumption (C) is fulfilled.
(ii) The weight function $w(\cdot)$ is not identical to zero and has bounded support contained in $[-C, C]^2$, for some constant C, and satisfies $|w(x) - w(y)| \le L_w \max\{|x_1 - y_1|, |x_2 - y_2|\}$ for all $x, y \in \Pi$ and some $L_w < \infty$.

Proposition 2.1. *Suppose that $H_0^{(s)}$ is true and assumption (A1) is fulfilled.*

(i) Let $G_{n,0}(x) = n^{-1/2} \sum_{t=1}^n \xi_t w(x - I_{t-1})$. Then

$$\sup_{x \in \Pi} |G_n(x) \,-\, G_{n,0}(x)| \;=\; o_P(1).$$

(ii) It holds that

$$G_n \;\xrightarrow{d}\; G$$

where $G = (G(x))_{x \in \Pi}$ is a centered Gaussian process with covariance function K,

$$K(x, y) \;=\; E\left[w(x - I_0)\, w(y - I_0)\right],$$

and convergence holds true w.r.t. the supremum metric.

Corollary 2.1. *Suppose that the assumptions of Proposition 2.1 are satisfied. Then*

$$T_n \;\xrightarrow{d}\; \sup_{x \in \Pi} |G(x)|.$$

The distribution of $\sup_{x \in \Pi} |G(x)|$ is absolutely continuous with respect to the Lebesgue measure and

$$\sup_{u \in \mathbb{R}} \left| P(T_n \le u) \,-\, P(\sup_{x \in \Pi} |G(x)| \le u) \right| \;\underset{n \to \infty}{\longrightarrow}\; 0.$$

2.2. Test statistic for testing composite hypotheses

By employing a similar notation as above, put $\widehat{I}_t = (\widehat{\lambda}_t, Y_t)'$, where $\widehat{\lambda}_t = f_{\widehat{\theta}_n}(\widehat{\lambda}_{t-1}, Y_{t-1})$ is the estimated unobserved process. Similar to the case of simple hypothesis, we define the estimated Pearson residuals, by $\widehat{\xi}_t = (Y_t - \widehat{\lambda}_t)/\sqrt{\widehat{\lambda}_t}$.

It is obvious that test statistic (2.1) can be adapted for testing (1.3) as follows:

$$\begin{aligned} \widehat{G}_n(x) &= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \widehat{\xi}_t \; w(x - \widehat{I}_{t-1}), \\ \widehat{T}_n &= \sup_{x \in \Pi} |\widehat{G}_n(x)|. \end{aligned} \tag{2.2}$$

Note that the processes $(\widetilde{\lambda}_t)_{t\in\mathbb{N}_0}$ and $(\widehat{\lambda}_t)_{t\in\mathbb{N}_0}$ are not stationary in general. It can be shown by backward iterations that, for given $(Y_t)_{t\in\mathbb{Z}}$ and $\theta \in \Theta$, the system of equations $\lambda_t = f_\theta(\lambda_{t-1}, Y_{t-1})$ $(t \in \mathbb{Z})$ has a unique stationary solution $(\lambda_t(\theta))_{t\in\mathbb{Z}}$, where $\lambda_t(\theta) = g_\theta(Y_{t-1}, Y_{t-2}, \ldots)$ for some measurable function g_θ; see also the proof of Theorem 3.1 in Neumann [19]. In the technical part below, we will use $\lambda_t(\widehat{\theta}_n)$ as an approximation for $\widehat{\lambda}_t$. Under H_0, we have $\lambda_t = \lambda_t(\theta_0)$, where θ_0 denotes the true parameter value. The assumptions below are essential to study the asymptotic behavior of test statistic (2.2). These are mild assumptions and are satisfied for many useful models; see Fokianos and Tjøstheim [11] and Doukhan et al. [2].

Assumption (A2).

(i) Assume that $\widehat{\theta}_n$ is an estimator of $\theta \in \Theta \subseteq \mathbb{R}^d$ which satisfies

$$\widehat{\theta}_n - \theta_0 = \frac{1}{n}\sum_{t=1}^{n} l_t + o_P(n^{-1/2}), \tag{2.3}$$

where

- $l_t = h(Y_t, Y_{t-1}, \ldots)$,
- $E(l_t \mid \mathcal{F}_{t-1}) = 0$ a.s.,
- $E[\|l_t\|^2/\lambda_t] < \infty$.

(ii) $\lambda_t(\theta)$ is a continuously differentiable function with respect to θ such that,

- $E\|\dot{\lambda}_t(\theta_0)\|_{l_1} < \infty$, where $\dot{\lambda}_t = \partial\lambda_t(\theta)/\partial\theta = (\partial\lambda_t(\theta)/\partial\theta_1, \ldots, \partial\lambda_t(\theta)/\partial\theta_d)$.
- $E_{\theta_0}\left[\sup_{\theta:\ \|\theta-\theta_0\|\le\delta} \|\dot{\lambda}_t(\theta) - \dot{\lambda}_t(\theta_0)\|^2\right] \underset{\delta\to 0}{\longrightarrow} 0.$

Assumption (A3). We assume that there exist $C < \infty$, $\kappa_1, \kappa_2 \ge 0$ with $\kappa := \kappa_1 + \kappa_2 < 1$ such that

(i) $\quad |f_{\theta'}(\lambda, y) - f_{\theta_0}(\lambda, y)| \le C\,\|\theta' - \theta_0\|\,(\lambda + y + 1) \qquad \forall \lambda, y$

(ii) $\quad |f_{\theta'}(\lambda, y) - f_{\theta'}(\widetilde{\lambda}, \widetilde{y})| \le \kappa_1|\lambda - \widetilde{\lambda}| + \kappa_2|y - \widetilde{y}| \quad \forall \lambda, y, \widetilde{\lambda}, \widetilde{y}.$

hold for all $\theta' \in \Theta$ with $\|\theta' - \theta_0\| \le \delta$, for some $\delta > 0$.

Assumption (A3)(i) will be used when terms such as $|f_{\widehat{\theta}_n}(\lambda_{t-1}, Y_{t-1}) - f_{\theta_0}(\lambda_{t-1}, Y_{t-1})|$ have to be estimated; e.g. for the derivation of (A-11) below. It is obvious that this condition is satisfied for the linear specification, $f_\theta(\lambda, y) = \theta_1 + \theta_2\lambda + \theta_3 y$. Furthermore, by considering $\partial f_\theta(\lambda, y)/\partial\theta_i$, it can be easily seen that this condition is also fulfilled for the specifications $f_\theta(\lambda, y) = \theta_1 + (\theta_2 + \theta_3\exp(-\theta_4\lambda^2))\lambda + \theta_5 y$ and $f_\theta(\lambda, y) = \theta_1(1+\lambda)^{-\theta_2} + \theta_3\lambda + \theta_4 y$ mentioned in the discussion right after Assumption (C).

Proposition 2.2. *Suppose that H_0 is true and assume further that (A1) to (A3) are satisfied. Then we have the following:*

(i) $\quad \sup_{x\in\Pi} |\widehat{G}_n(x) - \widehat{G}_{n,0}(x)| = o_P(1),$

where $\widehat{G}_{n,0}(x) = n^{-1/2} \sum_{t=1}^n \{ \xi_t w(x - I_{t-1}) - E_{\theta_0}[\dot{\lambda}_1(\theta_0) w(x - I_0)/\sqrt{\lambda_1(\theta_0)}]\, l_t \}$.

(ii) $\qquad \widehat{G}_n \xrightarrow{d} \widehat{G},$

where $\widehat{G} = (\widehat{G}(x))_{x \in \Pi}$ *is a centered Gaussian process with covariance function* $\widehat{K}$,

$$\widehat{K}(x,y) = E_{\theta_0}\Big[\Big(\xi_1\, w(x - I_0) - E_{\theta_0}[\dot{\lambda}_1(\theta_0)\, w(x - I_0)/\sqrt{\lambda_1(\theta_0)}]\, l_1\Big) \times \Big(\xi_1\, w(y - I_0) - E_{\theta_0}[\dot{\lambda}_1(\theta_0)\, w(y - I_0)/\sqrt{\lambda_1(\theta_0)}]\, l_1\Big)\Big].$$

Corollary 2.2. *Suppose that the assumptions of Proposition 2.2 are satisfied. Then*

$$\widehat{T}_n \xrightarrow{d} \sup_{x \in \Pi} |\widehat{G}(x)|.$$

If, additionally, there exists an $x \in \Pi$ *with* $\operatorname{var}(\widehat{G}(x)) > 0$, *then the distribution of* $\sup_{x \in \Pi} |\widehat{G}(x)|$ *is absolutely continuous with respect to the Lebesgue measure and*

$$\sup_{u \in \mathbb{R}} \Big| P(\widehat{T}_n \le u) - P(\sup_{x \in \Pi} |\widehat{G}(x)| \le u) \Big| \underset{n \to \infty}{\longrightarrow} 0.$$

2.3. Behavior of the test statistic under fixed and local alternatives

In this part of the article we consider the behavior of the test statistic (2.2) under fixed and local alternatives. In particular, for studying convergence under a sequence of local alternatives, we impose assumptions analogous to A6 and A3 of Escanciano [4] and Escanciano [5], respectively.

Proposition 2.3. *Suppose that assumptions (A1)-(A3) are satisfied.*

(i) Consider the fixed alternative

$$H_1\colon \quad f \notin \{f_\theta : \theta \in \Theta\}$$

for the testing problem (1.3)*. Then, if* $\widehat{\theta}_n \xrightarrow{P} \bar{\theta} \in \Theta$ *where* Θ *is a set where the contraction assumption (C) is satisfied, we have that*

$$\sup_{x \in \Pi} \Big| n^{-1/2} \widehat{G}_n(x) - E\Big[(f(I_0) - f_{\bar{\theta}}(I_0))\; w(x - I_0)/\sqrt{f_{\bar{\theta}}(I_0)}\Big]\Big| \xrightarrow{P} 0.$$

(ii) Consider the sequence of local alternatives

$$H_{1,n}\colon \quad f(\lambda, y) = f_{\theta_0}(\lambda, y) + \frac{a(\lambda, y)}{\sqrt{n}} + o(n^{-1/2}),$$

where $a\colon [0,\infty)\times\mathbb{N}_0 \to \mathbb{R}$ *is a Lipschitz continuous function, for the testing problem* (1.3)*. Suppose further that under* $H_{1,n}$*:*

$$\widehat{\theta}_n \;=\; \theta_0 \,+\, \frac{1}{\sqrt{n}}\theta_a \,+\, \frac{1}{n}\sum_{t=1}^{n} l_t \,+\, o_{P_n}(n^{-1/2}),$$

where $(P_n)_{n\in\mathbb{N}}$ *are the distributions according to* $H_{1,n}$*. Then*

$$\widehat{G}_n \;\xrightarrow{d}\; G \,+\, D,$$

where

$$D(x) \;=\; E_{\theta_0}\left[\left\{a(I_0) \,-\, \theta_a\,\dot{\lambda}_1(\theta_0)\right\}\; w(x-I_0)/\sqrt{\lambda_1(\theta_0)}\right].$$

The above results in conjunction with Proposition 3.1 indicate that the test is expected to have nontrivial power.

Remark 1. Our initial attempt towards the problem of goodness-of-fit testing for count time series was based on supremum-type tests of the following form (Koul and Stute [16])

$$H_n(x) = n^{-1/2}\sum_{t=1}^{n}\widehat{\xi}_t \mathbb{1}(\widehat{I}_{t-1}\le x). \tag{2.4}$$

In fact, Escanciano [5] introduced a method to prove stochastic equicontinuity for processes such as $(H_n)_{n\in\mathbb{N}}$ under ergodicity and minimal moment conditions. Unfortunately, we were not able to verify conditions similar to those in that paper. In particular, for the condition W3 of Escanciano [5] to hold true, the author gives as a sufficient condition that some conditional densities are dominated by the unconditional density. For count time series models, the conditional distribution of $(\lambda_t, Y_t)'$ given $\mathcal{F}_{t-2}$ is clearly discrete while this property is not guaranteed for the unconditional distribution in general. Furthermore, Escanciano and Mayoral [6] gave a set of alternative sufficient conditions for proving stochastic equicontinuity. Because of the discrete distributions we consider, we were not able to verify the counterpart of their condition A1(c). Nevertheless, the asymptotic distribution of supremum-type test statistics based on (2.4) is possible to be studied following the arguments of Koul and Stute [16] and utilizing the recent results on weak dependence properties obtained by Doukhan et al. [2].

Remark 2. It is some sort of folklore, mainly in the context of i.i.d. observations, that alternatives with a difference of order less than $n^{-1/2}$ from the null model cannot be detected with an asymptotically nontrivial power. We believe that this is also true in the case of Poisson count processes. As a simple example, consider the simplified case where we have i.i.d. observations $Y_1,\ldots,Y_n$, either with $Y_t\sim P_0 = \text{Poisson}(\lambda_0)$, $\lambda>0$, or with $Y_t\sim P_n = \text{Poisson}(\lambda_n)$, where $\lambda_n=\lambda_0+c_n$, $n^{1/2}c_n \longrightarrow_{n\to\infty} 0$. Then the squared Hellinger distance between

P_0 and P_n fulfills

$$\begin{aligned} H^2(P_0, P_n) &= \frac{1}{2} \sum_{k=0}^{\infty} \left(\sqrt{e^{-\lambda_0} \lambda_0^k / k!} - \sqrt{e^{-\lambda_n} \lambda_n^k / k!} \right)^2 \\ &= 1 - \exp\left\{ -(\sqrt{\lambda_0} - \sqrt{\lambda_n})^2 / 2 \right\} \leq (\sqrt{\lambda_0} - \sqrt{\lambda_n})^2 / 2 \leq |c_n| / 2. \end{aligned}$$

Denote by $P_0^{\star n}$ and $P_n^{\star n}$ the n-fold products of the distributions P_0 and P_n, respectively. Then we obtain for the Hellinger affinity of these distributions that

$$\begin{aligned} \rho(P_0^{\star n}, P_n^{\star n}) &= \sum_{k_1, \ldots, k_n = 0}^{\infty} \sqrt{P_0^{\star n}(\{(k_1, \ldots, k_n)\})} \sqrt{P_n^{\star n}(\{(k_1, \ldots, k_n)\})} \\ &= \left(\sum_{k=0}^{\infty} \sqrt{P_0(\{k\})} \sqrt{P_n(\{k\})} \right)^n \\ &= \big(1 - H^2(P_0, P_n)\big)^n \\ &\geq (1 - |c_n|/2)^n \underset{n \to \infty}{\longrightarrow} 1. \end{aligned}$$

Therefore, we obtain for the Hellinger distance between the product measures that

$$H^2(P_0^{\star n}, P_n^{\star n}) = 1 - \rho(P_0^{\star n}, P_n^{\star n}) \underset{n \to \infty}{\longrightarrow} 0,$$

which implies by the Cauchy-Schwartz inequality that

$$\begin{aligned} & \sum_{k_1, \ldots, k_n} |P_0^{\star n}(\{(k_1, \ldots, k_n)\}) - P_n^{\star n}(\{(k_1, \ldots, k_n)\})| \\ & \quad \leq \sqrt{8}\, H(P_0^{\star n}, P_n^{\star n}) \underset{n \to \infty}{\longrightarrow} 0. \end{aligned}$$

This proves that no test of the hypothesis $H_0\colon Y_t \sim P_0$ versus $H_1^{(n)}\colon Y_t \sim P_n$ can have an asymptotically nontrivial power.

3. Bootstrap

Resampling methods have been used in different contexts by several authors to approximate the null distribution of test statistics; see for instance Stute et al. [24] and Franke et al. [12] among other authors. In the case of testing the composite hypotheses (1.3) for count time series, we will take advantage of the Poisson assumption and employ the following version of parametric bootstrap to compute p-values for testing (1.3).

Parametric Bootstrap Algorithm

Step 1 Estimate the parameter θ by $\widehat{\theta}_n$ which satisfies assumption (A2)(i); e.g. the MLE.

Step 2 Take any starting value λ_0^* and, conditioned on λ_0^*, generate $Y_0^* \sim$ Poisson(λ_0^*).

Step 3 Given $Y^*_{t-1}, \lambda^*_{t-1}, \ldots, Y^*_0, \lambda^*_0$, compute $\lambda^*_t = f_{\widehat{\theta}_n}(\lambda^*_{t-1}, Y^*_{t-1})$ and generate $Y^*_t \sim \text{Poisson}(\lambda^*_t)$.

Step 4 Based on the above sample, compute the bootstrap counterparts $\widehat{G}^*_n(x)$ and $\widehat{T}^*_n$ of $\widehat{G}_n(x)$ and $\widehat{T}_n$, respectively.

The choice of the starting value λ^*_0 will be discussed in the next section. The next theorem shows the appropriateness of the bootstrap approximation.

Proposition 3.1. *Suppose that either the conditions of Proposition 2.2 or of Proposition 2.3 are fulfilled.*

(i) Under H_0 and $H_{1,n}$:

$$\widehat{G}^*_n \xrightarrow{d} \widehat{G} \quad \textit{in probability}$$

(ii) Under H_1:

$$\widehat{G}^*_n \xrightarrow{d} \bar{G} \quad \textit{in probability,}$$

where $\bar{G}$ is a zero mean Gaussian process.

Proposition 2.3 and part (ii) of Proposition 3.1 yield under H_1 that

$$P\left(\widehat{T}_n > t^*_n\right) \underset{n\to\infty}{\longrightarrow} 0$$

if $E\left[(f(I_0) - f_{\bar{\theta}}(I_0))\ w(x-I_0)/\sqrt{f_{\bar{\theta}}(I_0)}\right] \neq 0$ for some $x \in \Pi$, i.e., the proposed test is consistent against fixed alternatives under this condition. The following corollary states that our bootstrap-based test has asymptotically the prescribed size.

Corollary 3.1. *Suppose that the conditions of Proposition 2.2 are fulfilled. Then*

(i) $\widehat{T}^*_n \xrightarrow{d} \widehat{T}$ in probability,

(ii) If, additionally, $\text{var}(\widehat{G}(x)) > 0$ for some $x \in \Pi$, then

(a)

$$\sup_{x\in[0,\infty)} \left| P\left(\widehat{T}^*_n \leq x \mid Y_1, \ldots, Y_n\right) - P(\widehat{T}_n \leq x)\right| \xrightarrow{P} 0.$$

(b) $P(\widehat{T}_n > t^*_n) \underset{n\to\infty}{\longrightarrow} \alpha$.

4. Examples

4.1. An empirical study

We illustrate the performance of the proposed goodness-of-fit methodology by presenting a limited simulation study. More precisely, we study the test statistic (2.2). To compute the value of the test statistic, we use weight functions $w(\cdot)$ of the form $w(x) = w(x_1, x_2) = K(x_1)K(x_2)$ where $K(\cdot)$ is a univariate kernel and $x = (x_1, x_2) \in \Pi$ as discussed before. Note again that we employ three types of kernels; the Gaussian, uniform and Epanechnikov kernels. The first two do not satisfy assumption (A1)(ii). They are included for completeness of the presentation and for demonstrating that assumption (A1)(ii) might be possible to be relaxed. We also study empirically the asymptotic distribution of the supremum type test statistic based on $H_n(x)$ for the purpose of comparison; recall (2.4).

We use the maximum likelihood estimator to calculate the Pearson residuals. We set $\lambda_0 = 0$ and $\partial\lambda_0/\partial\theta = 0$ for initialization of the recursions in the case of the linear model. To compute the p-value of the test statistic we use the proposed bootstrap methodology. Accordingly, the test statistic is recalculated for B parametric bootstrap replications of the data set. Then, if $\widehat{T}_n$ denotes the observed value of the test statistic and $\widehat{T}^{\star}_{i;n}$ denotes the value of the test statistic in i'th bootstrap run, the corresponding p-value used to determine acceptance/rejection is given by

$$\text{p-value} = \frac{\#\{i : \widehat{T}^{\star}_{i;n} \geq \widehat{T}_n\}}{B+1}.$$

Throughout the simulations (and for the data analysis) we use $B = 499$. All results are based on 500 runs.

We first study the empirical size of the test. Table 1 shows the achieved significance levels for testing the linear model given by $\lambda_t = \theta_1 + \theta_2\lambda_{t-1} + \theta_3 Y_{t-1}$. It can be shown, that for the linear model, the stationary mean is given by $E(Y_t) = \theta_1/(1 - \theta_2 - \theta_3)$. Table 1 reports results when the stationary mean is equal to 2 (relatively low value) for two different configurations of the parameters and for different sample sizes. In both cases, we see that all test statistics achieve the nominal significance levels quite satisfactorily, especially for larger sample sizes. Furthermore, significance levels obtained by the test statistic based on (2.4) tend to underestimate the significance levels computed by the proposed goodness of fit tests.

Table 2 shows the power of the various test statistics for the model $\lambda_t = \theta_1 + \theta_2\lambda_{t-1} + (\theta_3 + \theta_4 \exp(-\theta_5 Y^2_{t-1}))Y_{t-1}$ with $\theta_4, \theta_5 > 0$ such that $\theta_2 + \theta_3 + \theta_4 < 1$. This is a case of a model which belongs to the local alternative specification studied in Proposition (2.3). In both cases, the power of all test statistics increase with the sample size. Note that in the first case the power is relatively low for the test statistic based on the Epanechnikov kernel but this behavior changes when considering the other model. For the case of second model and for large sample sizes the test statistics achieve relatively high power.

We study now the power of the test for testing the linear model against a log-linear model specification as suggested by Fokianos and Tjøstheim [10]. More specifically, if we denote by $\nu_t = \log \lambda_t$ the log-mean process, then the log-linear model studied by Fokianos and Tjøstheim [10] is given by $\nu_t = \theta_1 + \theta_2\nu_{t-1} + \theta_3 \log(1 + Y_{t-1})$, for $t \geq 1$. Note that for this model, the parameters are allowed to be either positive or negative valued whereas for the linear model all the parameters have to be positive. It is rather challenging to compute the

TABLE 1
Achieved significance levels for testing goodness of fit for the linear model. Results are based on 499 bootstrap replications and 500 simulations

	$\theta_1 = 0.4,\ \theta_2 = 0.2,\ \theta_3 = 0.6$				
n	Level	H_n	Uniform kernel	Gaussian kernel	Epanechnikov kernel
500	10	9.2	9.2	10.6	10.4
	5	4.8	5.4	5.4	5.8
	1	1.6	1.0	0.6	0.6
1000	10	8.6	8.6	9.4	8.8
	5	4.6	5.0	3.8	4.2
	1	1.0	1.2	1.0	0.6
	$\theta_1 = 1,\ \theta_2 = 0.1,\ \theta_3 = 0.4$				
500	10	7.5	12.1	10.3	13.0
	5	4.3	6.1	4.9	6.3
	1	0.6	0.6	1.4	1.6
1000	10	8.0	9.9	9.7	10.7
	5	4.2	4.6	5.0	5.0
	1	0.8	0.6	0.6	0.8

TABLE 2
Power for testing goodness of fit for the linear model. Results are based on 499 bootstrap replications and 500 simulations

	$\theta_1 = 0.4,\ \theta_2 = 0.2,\ \theta_3 = 0.6, \theta_4 = 0.4,\ \theta_5 = 0.1$				
n	Level	H_n	Uniform kernel	Gaussian kernel	Epanechnikov kernel
500	10	34.6	35.6	30.0	22.2
	5	23.8	23.4	19.2	12.4
1000	10	62.2	60.0	51.4	36.4
	5	48.4	44.8	35.8	23.8
	$\theta_1 = 1,\ \theta_2 = 0.1,\ \theta_3 = 0.4,\ \theta_4 = 0.4,\ \theta_5 = 0.1$				
500	10	38.8	67.6	66.0	46.7
	5	26.6	55.7	52.9	31.0
1000	10	81.1	97.6	96.2	86.7
	5	68.5	95.2	93.6	78.9

mean and the autocovariance function of the process $\{Y_t\}$ with this particular log-mean specification. However, Fokianos and Tjøstheim [10] have shown that when θ_2 and θ_3 have the same sign such that $|\theta_2 + \theta_3| < 1$ then the maximum likelihood estimator is consistent and asymptotically normally distributed. The same conclusion is true when θ_2 and θ_3 have opposite signs but in this case the required condition is $\theta_2^2 + \theta_3^2 < 1$. For the log-linear model, we use as starting values $\nu_0 = 1$ and $\partial\nu_0/\partial\theta = 0$.

Table 3 shows values of the empirical power of the test statistic $\widehat{T}_n$ when the data are generated by a log-linear model, as described before, for different sample sizes. We note that when the coefficient of the feedback process (that is θ_2) is positive, then the proposed test statistic achieves good power even though the sample size is relatively small. The performance of the test statistic based on $\widehat{H}_n$ compares favorably with the performance of all other test statistics in this case. In particular, we note that the power of the test statistic $\widehat{T}_n$ formed after

Table 3
Power for testing goodness of fit for the linear model when the data are generated by a log linear model. Results are based on 499 bootstrap replications and 500 simulations

	$\theta_1 = 0.50$, $\theta_2 = 0.65$, $\theta_3 = -0.35$				
n	Level	H_n	Uniform kernel	Gaussian kernel	Epanechnikov kernel
200	10	82.1	72.6	83.0	68.1
	5	72.4	63.6	73.8	51.8
500	10	99.8	98.7	99.4	96.4
	5	99.6	96.8	99.2	93.2
	$\theta_1 = 0.50$, $\theta_2 = -0.55$, $\theta_3 = 0.45$				
200	10	64.2	47.9	30.1	19.6
	5	52.9	34.1	18.6	10.7
500	10	88.5	78.3	55.9	34.6
	5	81.6	67.6	38.7	22.1

employing the Gaussian kernel is superior when compared with all other test statistics. When the signs of coefficients change then we have a significant decrease of power for all test statistics. This significance decrease can be explained by considering the autocorrelation function of the log–linear model. Although the autocorrelation function cannot be computed analytically, it can be studied by simulation. In the first case, in which the data are generated by a log-linear model with $\theta_1 = 0.5$, $\theta_2 = 0.65$, $\theta_3 = -0.35$, the autocorrelation function takes on negative values; hence the linear model will be rejected often since its autocorrelation function is always positive (see Fokianos and Tjøstheim [10], and Fokianos [8]). For the second case where data are generated by a log-linear model with $\theta_1 = 0.5$, $\theta_2 = -0.55$, $\theta_3 = 0.45$, the autocorrelation function of the observed process behaves like an alternating sequence with lag one value positive. Therefore, the linear model might accommodate such dependence structure and therefore it is rejected less often. We note that the test statistic based on $\widehat{H}_n$ performs better than all other test statistics.

4.2. Some data examples

The theory is illustrated empirically by two real data examples discussed in detail by Fokianos [8]. We first examine the number of transactions for the stock Ericsson B, for one day period; these data were reported by Fokianos et al. [9] where a linear model was fitted to the data. This series is an example of a count processes where the autocorrelation function decays slowly to zero; therefore it is expected that a linear feedback model would fit the data adequately; for further information see Fokianos [8]. Table 4 shows the test statistics accompanied with their p-values which have been obtained after $B = 499$ bootstrap replications. We note that the test statistic $\widehat{T}_n$ points to the adequacy of the linear model regardless of the chosen kernel function that is employed. However, the test statistic based on H_n raises some doubt about the linearity assumption though not overwhelming. Given the simplicity of fitting the linear model to the data, we conclude that it can describe the data adequately.

TABLE 4
Test statistics and p-values for real data examples when testing for the linear model. Results are based on $B = 499$ bootstrap replications

Transactions Data			
H_n	Uniform kernel	Gaussian kernel	Epanechnikov kernel
0.028	0.356	0.340	0.272
Series C3			
0.580	0.238	0.530	0.672

Another data example that we consider is a time series of claims, referred to as Series C3, of short-term disability benefits made by cut-injured workers in the logging industry; see Zhu and Joe [28] and Fokianos [8]. In contrast to the previous example, this is an example of count time series whose autocorrelation function decays rapidly towards zero; hence the feedback mechanism might not provide any further insight into the modeling of this series; for more see Fokianos [8]. Table 4 shows that all test statistics point to the adequacy of linear model for fitting those data. In this case, we note that the results are similar regardless of the chosen test statistic.

Appendix

In what follows, we denote by $\|w\|_\infty = \sup_{x\in\Pi} \|w(x)\|$.

Proof of Proposition 2.1. To prove the first part of the proposition we use the following decomposition:

$$\begin{aligned} |G_n(x) - G_{n,0}(x)| &= \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left(\widetilde{\xi}_t \; w(x - \widetilde{I}_{t-1}) - \xi_t \; w(x - I_{t-1}) \right) \right| \\ &\le \frac{1}{\sqrt{n}} \sum_{t=1}^{n} |\widetilde{\xi}_t - \xi_t| \; \|w\|_\infty + \frac{1}{\sqrt{n}} \sum_{t=1}^{n} |\xi_t| \; L_w \; |\widetilde{\lambda}_{t-1} - \lambda_{t-1}|. \end{aligned}$$

From

$$\begin{aligned} |\widetilde{\xi}_t - \xi_t| &\le \frac{|\widetilde{\lambda}_t - \lambda_t|}{\sqrt{\widetilde{\lambda}_t}} + |Y_t - \lambda_t| \left| \frac{1}{\sqrt{\widetilde{\lambda}_t}} - \frac{1}{\sqrt{\lambda_t}} \right| \\ &\le |\widetilde{\lambda}_t - \lambda_t| \left(\frac{1}{\sqrt{\lambda_{\min}}} + \frac{|Y_t - \lambda_t|}{2\, \lambda_{\min}^{3/2}} \right) \end{aligned} \tag{A-1}$$

and the fact that $|\widetilde{\lambda}_t - \lambda_t| \le \kappa_1^t |\widetilde{\lambda}_0 - \lambda_0|$ which follows directly from assumption (C), we obtain that assertion (i) holds true. In view of this result, it suffices to prove that

$$G_{n,0} \xrightarrow{d} G. \tag{A-2}$$

Towards this goal, we need to prove convergence of the finite dimensional distributions and stochastic equicontinuity. Weak convergence of the finite-dimensional distributions of $G_{n,0}$ to those of the process G follows from the Cramér-Wold device and the CLT for martingale difference arrays; see Pollard [21, p. 171].

To study stochastic equicontinuity, it suffices to show that, for each $\epsilon, \eta > 0$, there exist a finite partition $\{B_j\colon j = 0, \ldots, N\}$ of the set Π and points $x_j \in B_j$ such that

$$\limsup_{n\to\infty} P\left(\max_{0\le j\le N} \sup_{x\in B_j} |G_{n,0}(x) \,-\, G_{n,0}(x_j)| \,>\, \epsilon \right) \,\le\, \eta. \tag{A-3}$$

Inspired by Escanciano [5], we base our proof on a Bernstein-type inequality rather than on upper estimates of fourth moments of increments of the processes; compare with Koul and Stute [16]. To this end, we truncate the ξ_t and define

$$\xi_{n,t} \,=\, \xi_t \mathbb{1}(|\xi_t| \le \sqrt{n}) \qquad \text{and} \qquad \bar{\xi}_{n,t} \,=\, \xi_{n,t} \,-\, E(\xi_{n,t} \mid \mathcal{F}_{t-1}).$$

Then,

$$\begin{aligned} G_{n,0}(x) \quad &= \quad \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \bar{\xi}_{n,t}\, w(x - I_{t-1}) \,+\, \frac{1}{\sqrt{n}} \sum_{t=1}^{n} E(\xi_{n,t} \mid \mathcal{F}_{t-1})\, w(x - I_{t-1}) \\ &\qquad + \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \xi_t \mathbb{1}(|\xi_t| > \sqrt{n})\, w(x - I_{t-1}) \\ &=: \quad T_{n,1}(x) \,+\, T_{n,2}(x) \,+\, T_{n,3}(x), \end{aligned} \tag{A-4}$$

say. But

$$\begin{aligned} & P\left(|\xi_t| \,>\, \sqrt{n} \quad \text{for at least one } t \in \{1, \ldots, n\}\right) \\ &\le \quad nP(|\varepsilon_1| \,>\, \sqrt{n}\,\sqrt{\lambda_{\min}}) \\ &\le \quad E\left[\frac{\varepsilon_1^2}{\lambda_{\min}}\, \mathbb{1}(|\varepsilon_1| \,>\, \sqrt{n}\,\sqrt{\lambda_{\min}}) \right] \underset{n\to\infty}{\longrightarrow} 0, \end{aligned} \tag{A-5}$$

where $\varepsilon_1 = Y_1 - \lambda_1$. Therefore

$$P\left(\sup_{x\in\Pi} |T_{n,3}(x)| \,\ne\, 0 \right) \underset{n\to\infty}{\longrightarrow} 0, \tag{A-6}$$

i.e., the effect of the truncation is asymptotically negligible. Furthermore, since $E(\xi_{n,t} \mid \mathcal{F}_{t-1}) = -E(\xi_t \mathbb{1}(|\xi_t| > \sqrt{n}) \mid \mathcal{F}_{t-1})$ we obtain

$$\begin{aligned} E\left[\sup_{x\in\Pi} |T_{n,2}(x)| \right] \quad &\le \quad E\left[\|w\|_\infty \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left| E(\xi_t \mathbb{1}(|\xi_t| > \sqrt{n}) \mid \mathcal{F}_{t-1}) \right| \right] \\ &\le \quad \|w\|_\infty \frac{1}{\sqrt{n}} \sum_{t=1}^{n} E\left[\frac{E(\xi_t^2\, \mathbb{1}(|\xi_t| > \sqrt{n}) \mid \mathcal{F}_{t-1})}{\sqrt{n}} \right] \\ &= \quad \|w\|_\infty\, E\left[\frac{\varepsilon_1^2}{\lambda_{\min}}\, \mathbb{1}(|\varepsilon_1| \,>\, \sqrt{n}\sqrt{\lambda_{\min}}) \right] \underset{n\to\infty}{\longrightarrow} 0, \end{aligned}$$

which implies that

$$P\left(\sup_{x\in\Pi}|T_{n,2}(x)| > \epsilon/2\right) \underset{n\to\infty}{\longrightarrow} 0. \tag{A-7}$$

Hence, it remains to show that the processes $(T_{n,1}(x))_{x\in\Pi}$ can be approximated by their values on a finite grid. We define, for $j\in\mathbb{N}$ and $k_1,k_2\in\{1,2,\ldots,2^{3j}\}$,

$$B_{j;k_1,k_2} = [(k_1-1)2^{-j}, k_1 2^{-j}) \times [(k_2-1)2^{-j}, k_2 2^{-j}).$$

Moreover, let $\Pi_j = \bigcup_{k_1,k_2=1}^{2^{3j}} B_{j;k_1,k_2} = [0,2^{2j})^2$ and $B_{j;0,0}=\Pi\backslash\Pi_j$. From these sets, we choose the points $x_{j;k_1,k_2}=((k_1-1)2^{-j},(k_2-1)2^{-j})$ (for $k_1,k_2\geq 1$) and $x_{j;0,0}=(2^{2j},2^{2j})$. Finally, we define functions $\pi_j\colon \Pi\longrightarrow\Pi$ as $\pi_j(x)=\sum_{k_1,k_2} x_{j;k_1,k_2}\mathbb{1}(x\in B_{j;k_1,k_2})$.

For any fixed $\alpha\in(0,1)$, we define thresholds $\lambda_j=2^{-\alpha j}$ and "favorable events" as

$$\Omega_j = \left\{|T_{n,1}(x_{j;k_1,k_2}) - T_{n,1}(\pi_{j-1}(x_{j;k_1,k_2}))| \leq \lambda_j \quad \text{for all } k_1,k_2\right\}.$$

Since $T_{n,1}$ has continuous sample paths and $\pi_J(x)\underset{J\to\infty}{\longrightarrow} x\ \forall x\in\Pi$ we obtain that

$$T_{n,1}(\pi_J(x)) \underset{J\to\infty}{\longrightarrow} T_{n,1}(x) \qquad \forall x\in\Pi,$$

i.e., $T_{n,1}(x)$ is the pointwise limit of the approximations $T_{n,1}(\pi_J(x))$. Note that Ω_j can be rewritten as

$$\Omega_j = \left\{|T_{n,1}(\pi_j(x)) - T_{n,1}(\pi_{j-1}(x))| \leq \lambda_j \quad \text{for all } x\in\Pi\right\}.$$

Hence, if the event $\Omega_{j_0}^\infty=\bigcap_{j=j_0+1}^\infty \Omega_j$ occurs, then we obtain from $|T_{n,1}(x)-T_{n,1}(\pi_{j_0}(x))|\leq |T_{n,1}(x)-T_{n,1}(\pi_J(x))| + \sum_{j=j_0+1}^J \lambda_j\ \forall J\geq j_0$ that

$$\sup_{x\in\Pi}\left\{|T_{n,1}(x) - T_{n,1}(\pi_{j_0}x)|\right\} \leq \sum_{j=j_0+1}^{\infty}\lambda_j \leq \epsilon/2, \tag{A-8}$$

provided that j_0 is sufficiently large.

It remains to show that the probability of the unfavorable event $\Omega\setminus\Omega_{j_0}^\infty$ is smaller than or equal to η for sufficiently large j_0. Since

$$E\left(\bar{\xi}_{n,t}^2\middle|\,\mathcal{F}_{t-1}\right) \leq E(\xi_{n,t}^2\mid\mathcal{F}_{t-1}) \leq E(\xi_t^2\mid\mathcal{F}_{t-1}) = 1$$

we have, for $1\leq k_1,k_2\leq 2^{3j}$, that

$$\begin{aligned}
v_{j;k_1,k_2} &:= \frac{1}{n}\sum_{t=1}^n E\Big(\left[\bar{\xi}_{n,t}\left\{w(x_{j;k_1,k_2}-I_{t-1}) - w(\pi_{j-1}(x_{j;k_1,k_2})-I_{t-1})\right\}\right]^2\Big|\mathcal{F}_{t-1}\Big)\\
&\leq \frac{1}{n}\sum_{t=1}^n \left(w(x_{j;k_1,k_2}-I_{t-1}) - w(\pi_{j-1}(x_{j;k_1,k_2})-I_{t-1})\right)^2\\
&\leq L_w^2\, 2^{-2j}
\end{aligned}$$

and

$$\begin{aligned} a_{j;k_1,k_2} &:= \sup_{x\in\Pi} \left\{ n^{-1/2}\, |\bar{\xi}_{n,t}|\, |w(x_{j;k_1,k_2} - I_{t-1}) \,-\, w(\pi_{j-1}(x_{j;k_1,k_2}) - I_{t-1})| \right\} \\ &\leq L_w\, 2^{-j+1}. \end{aligned}$$

Furthermore, we have

$$\begin{aligned} & E\left(T_{n,1}(x_{j;0,0}) \,-\, T_{n,1}(\pi_{j-1}(x_{j;0,0}))\right)^2 \\ &\quad = E\left[\frac{1}{n}\sum_{t=1}^{n}\left(w(x_{j;0,0} - I_{t-1}) \,-\, w(\pi_{j-1}(x_{j;0,0}) - I_{t-1})\right)^2\right] \\ &\quad \leq 4\, \|w\|_\infty\, P\left(w(x_{j;0,0} - I_0) \,-\, w(\pi_{j-1}(x_{j;0,0}) - I_0) \,\neq\, 0\right) \\ &\quad = O\left(E[\|I_0\|^2]\, 2^{-2j}\right). \end{aligned}$$

The last equality follows from the boundedness of the support of w.

Therefore, we obtain by the Bernstein-Freedman inequality (see Freedman [13, Proposition 2.1]) that

$$\begin{aligned} P(\Omega_j^c) &\leq \sum_{k_1,k_2\leq 2^{3j}} P\left(|T_{n,1}(x_{j;k_1,k_2}) \,-\, T_{n,1}(\pi_{j-1}(x_{j;k_1,k_2}))| \,>\, \lambda_j\right) \\ &\leq \sum_{1\leq k_1,k_2\leq 2^{3j}} 2\, \exp\left(-\frac{\lambda_j^2}{2\, v_{j;k_1,k_2} \,+\, 2\, a_{j;k_1,k_2}\, \lambda_j}\right) \\ &\qquad + \frac{E\left(T_{n,1}(x_{j;0,0}) \,-\, T_{n,1}(\pi_{j-1}x_{j;0,0})\right)^2}{\lambda_j^2} \\ &= O\left(2^{6j}\, \exp\left(-\frac{2^{-2\alpha j}}{C_1\, 2^{-2j} \,+\, C_2\, 2^{-(1+\alpha)j}}\right)\right) \,+\, O\left(2^{2(\alpha-1)j}\right). \end{aligned}$$

Hence, we obtain, for j_0 sufficiently large, that

$$P\left(\Omega \setminus \bigcap_{j=j_0+1}^{\infty} \Omega_j\right) \,\leq\, \sum_{j=j_0+1}^{\infty} P(\Omega_j^c) \,\leq\, \eta. \tag{A-9}$$

Stochastic equicontinuity of $G_{n,0}$, i.e. (A-3), follows now from displays (A-4), (A-6) to (A-9). □

Proof of Corollary 2.1. The first assertion follows from Proposition 2.1 and the continuous mapping theorem.

Absolute continuity of the distribution of $\sup_{x\in\Pi} |G(x)|$ will be derived from a result from Lifshits [17]. Since $G(\lambda, y) \to 0$ as either $\lambda \to \infty$ or $y \to \infty$, this supremum has the same distribution as $\sup_{\lambda,y\in[0,1]} |\widetilde{G}(\lambda, y)|$, where

$$\widetilde{G}(\lambda, y) \,=\, \begin{cases} G\left(\frac{\lambda}{1-\lambda}, \frac{y}{1-y}\right), & \text{if } \lambda, y \in [0,1), \\ 0, & \text{if } \lambda = 1 \text{ or } y = 1 \end{cases}$$

The process $(\widetilde{G}(\lambda,y))_{\lambda,y\in[0,1]}$ is a centered Gaussian process defined on a compact set and with continuous sample paths. Hence, Proposition 3 of Lifshits [17] can be applied and it follows that $\sup_{\lambda,y}\widetilde{G}(\lambda,y)$ is absolutely continuous on $(0,\infty)$ with respect to the Lebesgue measure. For the same reason, the distribution of $\sup_{\lambda,y}(-\widetilde{G}(\lambda,y))$ is also absolutely continuous on $(0,\infty)$. Hence, the distribution of $\sup_{\lambda,y}|G(\lambda,y)|$ has no atom unequal to 0. However, since $P(\sup_{\lambda,y}|G(\lambda,y)|\neq 0)=1$, we obtain the second assertion. Convergence of the distribution functions in the uniform norm can be deduced from the distributional convergence in conjunction with the continuity of the limiting distribution function; see e.g. van der Vaart [26, Lemma 2.11]. □

Proof of Proposition 2.2. (i) To simplify calculations, we approximate the $\widehat{\lambda}_t$ by their stationary counterparts, $\lambda_t(\widehat{\theta}_n)=g_{\widehat{\theta}_n}(Y_{t-1},Y_{t-2},\ldots)$. It follows from the contractive condition (A3)(ii) that

$$\left|\widehat{\lambda}_t\,-\,\lambda_t(\widehat{\theta}_n)\right|\;\leq\;\kappa_1^t\left|\widehat{\lambda}_0\,-\,\lambda_0(\widehat{\theta}_n)\right|;$$

see e.g. inequality (5.13) in Neumann [19]. Therefore, we obtain that

$$\sup_{x\in\Pi}\left|\widehat{\widehat{G}}_n(x)\,-\,\widehat{G}_n(x)\right|\;=\;o_P(1),\tag{A-10}$$

where $\widehat{\widehat{G}}_n(x)\;=\;n^{-1/2}\sum_{t=1}^n\xi_t(\widehat{\theta}_n)\,w(x-I_{t-1}(\widehat{\theta}_n))$ and $I_{t-1}(\theta)=(\lambda_{t-1}(\theta),Y_{t-1})'$.
Consider the following decomposition

$$\begin{aligned}
\widehat{\widehat{G}}_n(x)=\widehat{G}_{n,0}(x)\,&+\,\frac{1}{\sqrt{n}}\sum_{t=1}^n\xi_t\left[w(x-I_{t-1}(\widehat{\theta}_n))\,-\,w(x-I_{t-1}(\theta_0))\right]\\
&+\,\frac{1}{\sqrt{n}}\sum_{t=1}^n(\xi_t-\xi_t(\widehat{\theta}_n))\left[w(x-I_{t-1}(\widehat{\theta}_n))\,-\,w(x-I_{t-1}(\theta_0))\right]\\
&-\,\frac{1}{\sqrt{n}}\sum_{t=1}^n\left[\xi_t(\widehat{\theta}_n)-\xi_t\,-\,\dot{\xi}_t(\theta_0)\left(\frac{1}{n}\sum_{s=1}^n l_s\right)\right]w(x-I_{t-1}(\theta_0))\\
&-\,\left(\frac{1}{\sqrt{n}}\sum_{t=1}^n l_t\right)\,\times\,\frac{1}{n}\sum_{s=1}^n\left[\dot{\xi}_s(\theta_0)w(x-I_{s-1})\,-\,E_{\theta_0}[\dot{\xi}_1(\theta_0)w(x-I_0)]\right]\\
=:\widehat{G}_{n,0}(x)\,&+\,R_{n,1}(x,\widehat{\theta}_n)\,+\,R_{n,2}(x,\widehat{\theta}_n)\,-\,R_{n,3}(x,\widehat{\theta}_n)\,-\,R_{n,4}(x),
\end{aligned}$$

say. According to Neumann [19, Eq. (5.13)], if $\|\widehat{\theta}_n-\theta_0\|\leq\delta$, we obtain that

$$\begin{aligned}
&|\lambda_t(\widehat{\theta}_n)\,-\,\lambda_t(\theta_0)|\\
&\quad\leq\;C\,\|\widehat{\theta}_n-\theta_0\|\,\big\{(\lambda_{t-1}+Y_{t-1}+1)\,+\,\kappa_1(\lambda_{t-2}+Y_{t-2}+1)\,+\,\cdots\\
&\qquad+\,\kappa_1^{t-1}\,(\lambda_0+Y_0+1)\big\}\,+\,\kappa_1^t\,|\lambda_0(\widehat{\theta}_n)\,-\,\lambda_0(\theta_0)|.
\end{aligned}\tag{A-11}$$

Since $|w(x-I_{t-1}(\widehat{\theta}_n))-w(x-I_{t-1}(\theta_0))|\leq L_w\,|\lambda_{t-1}(\widehat{\theta}_n)\,-\,\lambda_{t-1}(\theta_0)|$ we see that the summands in $R_{n,1}(x,\widehat{\theta}_n)$ are essentially of order $O_P(n^{-1/2})$ which suggests that the $\sup_{x\in\Pi}|R_{n,1}(x,\widehat{\theta}_n)|$ should be of negligible order. Apart from this heuristic, a sound proof of this fact is however more delicate since we have a supremum over x and the $\lambda_t(\widehat{\theta}_n)$ depend via $\widehat{\theta}_n$ on the whole sample. To proceed, we will first approximate $\widehat{\theta}_n$ by values from a sufficiently fine grid, apply an exponential inequality to the corresponding sums, and use finally continuity arguments to conclude. To invoke the Bernstein-Freedman inequality, we will replace as in the proof of Proposition 2.1 the ξ_t by $\bar{\xi}_{n,t} = \xi_{n,t} - E(\xi_{n,t} \mid \mathcal{F}_{t-1})$, where $\xi_{n,t} = \xi_t \mathbb{1}(|\xi_t| \leq \sqrt{n})$. We split up

$$\begin{aligned} R_{n,1}(x,\widehat{\theta}_n) &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n} \bar{\xi}_{n,t}\ \left[w(x-I_{t-1}(\widehat{\theta}_n))\ -\ w(x-I_{t-1}(\theta_0))\right] \\ &\quad +\ \frac{1}{\sqrt{n}}\sum_{t=1}^{n} E(\xi_{n,t}\mid\mathcal{F}_{t-1})\ \left[w(x-I_{t-1}(\widehat{\theta}_n))\ -\ w(x-I_{t-1}(\theta_0))\right] \\ &\quad +\ \frac{1}{\sqrt{n}}\sum_{t=1}^{n} \xi_t\ \mathbb{1}(|\xi_t|>\sqrt{n})\ \left[w(x-I_{t-1}(\widehat{\theta}_n))\ -\ w(x-I_{t-1}(\theta_0))\right] \\ &=: R_{n,11}(x,\widehat{\theta}_n)\ +\ R_{n,12}(x,\widehat{\theta}_n)\ +\ R_{n,13}(x,\widehat{\theta}_n). \end{aligned}$$

We conclude from (A-5) that

$$\sup_{x\in\Pi}\left|R_{n,13}(x,\widehat{\theta}_n)\right|\ =\ o_P(1). \tag{A-12}$$

Furthermore, it follows from $E(\xi_{n,t}\mid\mathcal{F}_{t-1}) = -E(\xi_t\mathbb{1}(|\xi_t|>\sqrt{n})\mid\mathcal{F}_{t-1})$ that

$$\begin{aligned} E_{\theta_0}\left[\sup_{x\in\Pi}|R_{n,12}(x,\widehat{\theta}_n)|\right] \quad &\leq \quad 2\,\|w\|_\infty\ \frac{1}{\sqrt{n}}\sum_{t=1}^{n} E\left[|\xi_t|\mathbb{1}(|\xi_t|>\sqrt{n})\right] \\ &\leq \quad 2\,\|w\|_\infty\ E\left[(\varepsilon_1^2/\lambda_{\min})\mathbb{1}(|\xi_1|>\sqrt{n}\sqrt{\lambda_{\min}})\right] \underset{n\to\infty}{\longrightarrow} 0, \end{aligned}$$

which shows that

$$\sup_{x\in\Pi}\left|R_{n,12}(x,\widehat{\theta}_n)\right|\ =\ o_P(1). \tag{A-13}$$

It remains to show that $R_{n,11}(x,\widehat{\theta}_n)$ is asymptotically negligible uniformly in $x\in\Pi$. Let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence of positive reals tending to infinity as $n\to\infty$. We choose two sequences of grids, $\mathcal{G}_n \subseteq \{\theta\in\Theta\colon \|\theta-\theta_0\|\leq\gamma_n n^{-1/2}\}$ and $\mathcal{X}_n\subseteq\{x\in\Pi\colon\|x\|\leq\gamma_n\}$, with mesh sizes γ_n^{-1} and of cardinality of order $O(n^\gamma)$, for some $\gamma<\infty$. We conclude from the Bernstein-Freedman inequality that

$$\max_{x\in\mathcal{X}_n,\,\theta\in\mathcal{G}_n}|R_{n,11}(x,\theta)|\ \stackrel{P}{\longrightarrow}\ 0.$$

Using the Lipschitz continuity of w and the fact that $P(\|\widehat{\theta}_n-\theta_0\|>\gamma_n n^{-1/2})\longrightarrow_{n\to\infty} 0$ we obtain that

$$\sup_{x\in\mathcal{X}_n}\inf_{\theta\in\mathcal{G}_n}\|R_{n,11}(x,\widehat{\theta}_n)\ -\ R_{n,11}(x,\theta)\|\ =\ o_P(1),$$

which implies

$$\sup_{x:\,\|x\|\le\gamma_n} |R_{n,11}(x,\widehat{\theta}_n)| \;=\; o_P(1).$$

Finally, since the weight function w has compact support $[-C,C]^2$ we obtain

$$\begin{aligned}\sup_{x:\,\|x\|>\gamma_n} |R_{n,11}(x,\widehat{\theta}_n)| \quad &\le \quad \frac{1}{\sqrt{n}}\sum_{t=1}^{n} |\bar{\xi}_{n,t}|\; L_w\; |\lambda_{t-1}(\widehat{\theta}_n) - \lambda_{t-1}(\theta_0)| \\ &\qquad \times \mathbb{1}(\max\{\|I_{t-1}(\theta_0)\|, \|I_{t-1}(\widehat{\theta}_n)\|\} > \gamma_n - C\sqrt{2})\end{aligned}$$

and, therefore we obtain that

$$\sup_{x:\,\|x\|>\gamma_n} |R_{n,11}(x,\widehat{\theta}_n)| \;=\; o_P(1).$$

This implies, in conjunction with (A-12) and (A-13) that

$$\sup_{x\in\Pi} |R_{n,1}(x,\widehat{\theta}_n)| \;=\; o_P(1). \tag{A-14}$$

Using the Lipschitz continuity of w we obtain the estimate

$$\left|R_{n,2}(x,\widehat{\theta}_n)\right| \;\le\; \frac{L_w}{\sqrt{n}}\,\sum_{t=1}^{n} |\xi_t(\widehat{\theta}_n) - \xi_t|\; |\lambda_{t-1}(\widehat{\theta}_n) - \lambda_{t-1}|.$$

Therefore, together with

$$\begin{aligned}|\xi_t(\widehat{\theta}_n) - \xi_t| \quad &\le \quad \frac{|\lambda_t(\widehat{\theta}_n) \,-\, \lambda_t|}{\sqrt{\lambda_t(\widehat{\theta}_n)}} \;+\; |Y_t - \lambda_t| \left|\frac{1}{\sqrt{\lambda_t(\widehat{\theta}_n)}} - \frac{1}{\sqrt{\lambda_t}}\right| \\ &\le \quad |\lambda_t(\widehat{\theta}_n) - \lambda_t| \left(\frac{1}{\sqrt{\lambda_{\min}}} \;+\; \frac{|Y_t - \lambda_t|}{2\,\lambda_{\min}^{3/2}}\right)\end{aligned}$$

and the fact that $\max_{1\le t\le n} |Y_t - \lambda_t| = o_P(\sqrt{n})$, we obtain that

$$\sup_{x\in\Pi} |R_{n,2}(x,\widehat{\theta}_n)| \;=\; o_P(1). \tag{A-15}$$

To estimate $R_{n,3}(x,\widehat{\theta}_n)$, we split up

$$\begin{aligned}|R_{n,3}(x,\widehat{\theta}_n)| \quad &\le \quad \frac{\|w\|_\infty}{\sqrt{n}}\sum_{t=1}^{n} \left|\xi_t(\widehat{\theta}_n) \,-\, \xi_t(\theta_0) \,-\, \dot{\xi}_t(\theta_0)(\widehat{\theta}_n \,-\, \theta_0)\right| \\ &\qquad +\; \frac{\|w\|_\infty}{\sqrt{n}}\sum_{t=1}^{n} \|\dot{\xi}_t(\theta_0)\|\; \left\|(\widehat{\theta}_n \,-\, \theta_0) \,-\, \frac{1}{n}\sum_{s=1}^{n} l_s\right\|. \end{aligned} \tag{A-16}$$

Note that

$$\dot{\xi}_t(\theta) \;=\; -\dot{\lambda}_t(\theta)/\lambda_t(\theta)^{1/2} - (Y_t - \lambda_t(\theta))\dot{\lambda}_t(\theta)/(2\lambda_t(\theta)^{3/2}). \tag{A-17}$$

Hence, it follows from (A2)(ii) and (A-1) that

$$E_{\theta_0}\left[\sup_{\theta:\,\|\theta-\theta_0\|\le\delta}\|\dot{\xi}_1(\theta)\,-\,\dot{\xi}_1(\theta_0)\|\right]\underset{\delta\to 0}{\longrightarrow}0.$$

Since $|\xi_t(\widehat{\theta}_n)-\xi_t(\theta_0)\,-\,\dot{\xi}_t(\theta_0)(\widehat{\theta}_n\,-\,\theta_0)|\,\le\,\|\dot{\xi}_t(\widetilde{\theta}_n)-\dot{\xi}_t(\theta_0)\|\|\widehat{\theta}_n\,-\,\theta_0\|$, for some $\widetilde{\theta}_n$ between θ_0 and $\widehat{\theta}_n$ we see that the first term on the right-hand side of (A-16) is $o_P(1)$. We obtain from (A-17) and the ergodic theorem that $n^{-1}\sum_{t=1}^n\|\dot{\xi}_t(\theta_0)\|\overset{a.s.}{\longrightarrow}E_{\theta_0}\|\dot{\xi}_0(\theta_0)\|$. This shows that, in conjunction with assumption (A2)(i) that the second term is also negligible. Hence, we obtain that

$$\sup_{x\in\Pi}|R_{n,3}(x,\widehat{\theta}_n)|\,=\,o_P(1). \tag{A-18}$$

To estimate $\sup_x|R_{n,4}(x)|$, we use truncation by letting $\bar{\dot{\xi}}_s(\theta_0)\,=\,\dot{\xi}_s(\theta_0)\times\mathbb{1}(|\dot{\xi}_s(\theta_0)|\le M)$, for some $M<\infty$. Note that the random functions $\bar{\dot{\xi}}_s(\theta_0)w(x-I_{s-1})$ are bounded and equicontinuous in x. Moreover, it follows from the ergodicity of $((I_{t-1},\bar{\dot{\xi}}_t(\theta_0)))_t$ that

$$\frac{1}{n}\sum_{t=1}^n\delta_{(I_{t-1},\bar{\dot{\xi}}_t(\theta_0))}\;\Longrightarrow\;P_{\theta_0}^{I_0,\bar{\dot{\xi}}_1(\theta_0)}.$$

Hence, we conclude from Corollary 11.3.4 in Dudley [3, p. 311] that

$$\sup_{x\in\Pi}\left|n^{-1}\sum_{s=1}^n\bar{\dot{\xi}}_s(\theta_0)\;w(x-I_{s-1})\;-\;E_{\theta_0}[\bar{\dot{\xi}}_1(\theta_0)\;w(x-I_0)]\right|\;\overset{a.s.}{\longrightarrow}\;0.$$

Furthermore, it follows from majorized convergence that

$$E_{\theta_0}\left[\sup_{x\in\Pi}\left|n^{-1}\sum_{s=1}^n(\dot{\xi}_s(\theta_0)-\bar{\dot{\xi}}_s(\theta_0))w(x-I_{s-1})-E_{\theta_0}[(\dot{\xi}_1(\theta_0)-\bar{\dot{\xi}}_1(\theta_0))w(x-I_0)]\right|\right]$$
$$\le 2\,\|w\|_\infty\;E_{\theta_0}[|\bar{\xi}_1|\mathbb{1}(|\bar{\xi}_1|>M)]\;\underset{M\to\infty}{\longrightarrow}\;0.$$

These two estimates yield that

$$\sup_{x\in\Pi}|R_{n,4}(x)|\,=\,o_P(1). \tag{A-19}$$

The first assertion of the proposition follows now from (A-14) to (A-19).

To show the second part of the result, we need to establish stochastic equicontinuity and convergence of finite dimensional distributions. Stochastic equicontinuity of $(G_{n,0})_{n\in\mathbb{N}}$ was already shown in the second part of the proof of Proposition 2.1. Hence, it remains to show that the process $(n^{-1/2}\sum_{t=1}^n E_{\theta_0}[\dot{\lambda}_1(\theta_0)w(x-I_0)/\sqrt{\lambda_1(\theta_0)}]l_t)_{x\in\Pi}$ also possesses this property. We obtain from the Lipschitz continuity of the weight function w that

$$\left\|E_{\theta_0}[\dot{\lambda}_1(\theta_0)\;w(x-I_0)/\sqrt{\lambda_1(\theta_0)}]\;-\;E_{\theta_0}[\dot{\lambda}_1(\theta_0)\;w(y-I_0)/\sqrt{\lambda_1(\theta_0)}]\right\|$$
$$\le\frac{L_w}{\sqrt{\lambda_{\min}}}\,\|x-y\|\;E_{\theta_0}\|\dot{\lambda}_1(\theta_0)\|.$$

Furthermore, we have that

$$\begin{aligned}
&\sup_{x:\,\|x\|\geq c} \left\| E_{\theta_0}[\dot{\lambda}_1(\theta_0)\; w(x-I_0)/\sqrt{\lambda_1(\theta_0)}] \right\| \\
&\quad \leq \quad \frac{\|w\|_\infty}{\sqrt{\lambda_{\min}}} \sup_{x:\,\|x\|\geq c} E_{\theta_0}[\|\dot{\lambda}_1(\theta_0)\|\; \mathbb{1}(I_0 \in \operatorname{supp}(w(x-\cdot)))] \underset{c\to\infty}{\longrightarrow} 0.
\end{aligned}$$

These two facts yield, together with $n^{-1/2}\sum_{t=1}^n l_t = O_P(1)$ the desired stochastic equicontinuity.

In addition, weak convergence of the finite-dimensional distributions of $\widehat{G}_{n,0}$ to those of the process $\widehat{G}$ follows from the Cramér-Wold device and the CLT for martingale difference arrays given on page 171 in Pollard [21]. $\square$

Proof of Proposition 2.3. To prove the first part, note that from $\widehat{\theta}_n \xrightarrow{P} \bar{\theta}$, we obtain

$$\sup_{x\in\Pi} \left| n^{-1/2}\widehat{G}_n(x) \;-\; \frac{1}{n}\sum_{t=1}^n \frac{Y_t - f_{\bar{\theta}}(I_{t-1})}{\sqrt{f_{\bar{\theta}}(I_{t-1})}}\; w(x \;-\; I_{t-1}(\bar{\theta})) \right| \;=\; o_P(1).$$

We split up

$$\begin{aligned}
&\frac{1}{n}\sum_{t=1}^n \frac{Y_t - f_{\bar{\theta}}(I_{t-1})}{\sqrt{f_{\bar{\theta}}(I_{t-1})}}\; w(x \;-\; I_{t-1}(\bar{\theta})) \\
&\quad = \quad \frac{1}{n}\sum_{t=1}^n \frac{Y_t - f(I_{t-1})}{\sqrt{f_{\bar{\theta}}(I_{t-1})}}\; w(x \;-\; I_{t-1}(\bar{\theta})) \\
&\qquad\quad + \frac{1}{n}\sum_{t=1}^n \frac{f(I_{t-1}) - f_{\bar{\theta}}(I_{t-1})}{\sqrt{f_{\bar{\theta}}(I_{t-1})}}\; w(x \;-\; I_{t-1}(\bar{\theta})) \\
&\quad =: \quad R_{n,1}(x) \;+\; R_{n,2}(x).
\end{aligned}$$

We can see by calculations analogous to that for $G_{n,0}$ that $(n^{1/2}R_{n,1}(x))_{x\in\Pi}$ converges to a certain Gaussian process that yields

$$\sup_{x\in\Pi} |R_{n,1}(x)| \;=\; o_P(1).$$

Furthermore, it follows from Corollary 11.3.4 of Dudley [3] and an appropriate truncation argument that

$$\sup_{x\in\Pi} \left| R_{n,2}(x) \;-\; E[(f(I_0) - f_{\bar{\theta}}(I_0))\; w(x - I_0)/\sqrt{f_{\bar{\theta}}(I_0)}] \right| \;=\; o_P(1),$$

which completes the proof of (i).

To prove the second part, note that

$$\widehat{G}_n(x) \quad = \quad \frac{1}{\sqrt{n}}\sum_{t=1}^n \frac{Y_t \;-\; f_{\widehat{\theta}_n}(I_{t-1})}{\sqrt{f_{\widehat{\theta}_n}(I_{t-1})}}\; w(x \;-\; \widehat{I}_{t-1})$$

$$= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{Y_t - f(I_{t-1})}{\sqrt{f(I_{t-1})}} \, w(x - \widehat{I}_{t-1}) + \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{f(I_{t-1}) - f_{\widehat{\theta}_n}(I_{t-1})}{\sqrt{f(I_{t-1})}} \, w(x - \widehat{I}_{t-1}) + R_{n,3}(x), \tag{A-20}$$

say. It can be shown analogously to the proof of Proposition 2.2 that

$$\sup_{x \in \Pi} |R_{n,3}(x)| = o_{P_n}(1).$$

Furthermore, we can show analogously to part (ii) of the proof of Proposition 2.1 that $(n^{-1/2} \sum_{t=1}^{n} (Y_t - f(I_{t-1}))/\sqrt{f(I_{t-1})} w(x - I_{t-1}))_{x \in \Pi}$ is stochastically equicontinuous.

To analyze the second term on the right-hand side of (A-20), we decompose further:

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{f(I_{t-1}) - f_{\widehat{\theta}_n}(I_{t-1})}{\sqrt{f(I_{t-1})}} \, w(x - I_{t-1})$$
$$= \frac{1}{n} \sum_{t=1}^{n} a(I_{t-1}) \, w(x - I_{t-1}) / \sqrt{f_{\theta_0}(I_{t-1})} - \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{f_{\theta_0}(I_{t-1}) - f_{\widehat{\theta}_n}(I_{t-1})}{\sqrt{f(I_{t-1})}} \, w(x - I_{t-1}) + R_{n,4}(x),$$

where

$$\sup_{x \in \Pi} |R_{n,4}(x)| = o_{P_n}(1).$$

But

$$\sup_{x \in \Pi} \left| \frac{1}{n} \sum_{t=1}^{n} a(I_{t-1}) w(x - I_{t-1}) / \sqrt{f_{\theta_0}(I_{t-1})} - E_{\theta_0}[a(I_0) w(x - I_0) / \sqrt{f_{\theta_0}(I_0)}] \right| \xrightarrow{P_n} 0.$$

Furthermore, we obtain from

$$\widehat{\theta}_n = \theta_0 + \frac{1}{\sqrt{n}} \theta_a + \frac{1}{n} \sum_{t=1}^{n} l_t + o_{P_n}(n^{-1/2})$$

that

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{f_{\theta_0}(I_{t-1}) - f_{\widehat{\theta}_n}(I_{t-1})}{\sqrt{f(I_{t-1})}} \, w(x - I_{t-1})$$
$$= -\left(\theta_a + \frac{1}{\sqrt{n}} \sum_{t=1}^{n} l_t \right) E_{\theta_0}[\dot{\lambda}_1(\theta_0) w(x - I_0) / \sqrt{\lambda_1(\theta_0)}] + R_{n,5}(x),$$

where

$$\sup_{x\in\Pi} |R_{n,5}(x)| \;=\; o_{P_n}(1).$$

Summarizing the above calculations we obtain

$$\widehat{G}_n(x) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left\{ \frac{Y_t - f(I_{t-1})}{\sqrt{f(I_{t-1})}} w(x - I_{t-1}) - l_t E_{\theta_0}[\dot{\lambda}_1(\theta_0) w(x - I_0)/\sqrt{\lambda_1(\theta_0)}] \right\}$$
$$+ \; D(x) \; + \; R_n(x) \tag{A-21}$$

with

$$\sup_{x\in\Pi} |R_n(x)| \;=\; o_{P_n}(1).$$

Moreover, the first term on the right-hand side of (A-21) is stochastically equicontinuous. The assertion follows now from the Cramér-Wold device and the CLT for martingale difference arrays given on page 171 in Pollard [21]. □

Proof of Proposition 3.1. In all cases, stochastic equicontinuity can be proved as above. And we obtain from the contractive property that $E^*(\lambda_0^*)^2 = O_P(1)$. To obtain part (i) of the proposition, note that under H_0 and $H_{1,n}$, the conditional distributions of the bootstrap variables converge weakly to those of the original random variables,

$$P^{I_t^*|I_{t-1}^*=x} \;\Longrightarrow\; P_{\theta_0}^{I_t|I_{t-1}=x}, \qquad \text{in probability.} \tag{A-22}$$

Therefore, we obtain by Lemma 4.2 of Neumann and Paparoditis [20] that also the marginal distributions converge,

$$P^{I_t^*} \;\Longrightarrow\; P_{\theta_0}^{I_t}, \qquad \text{in probability.} \tag{A-23}$$

Equations (A-22) and (A-23) eventually lead to

$$\widehat{G}_n^* \xrightarrow{d} \widehat{G},$$

again in probability.

To prove the second part of the results, note that under H_1, we obtain

$$P^{I_t^*|I_{t-1}^*=x} \;\Longrightarrow\; P_{\bar{\theta}}^{I_t|I_{t-1}=x}, \qquad \text{in probability.} \tag{A-24}$$

This implies, again by Lemma 4.2 of Neumann and Paparoditis [20], that

$$P^{I_t^*} \;\Longrightarrow\; P_{\bar{\theta}}^{I_t}, \qquad \text{in probability.} \tag{A-25}$$

Finally, (A-24) and (A-25) yield

$$\widehat{G}_n^* \xrightarrow{d} \bar{G},$$

again in probability. □

References

[1] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics* **31**: 307–327. MR0853051

[2] Doukhan, P., Fokianos, K., and Tjøstheim, D. (2012). On weak dependence conditions for Poisson autoregressions, *Statistics & Probability Letters* **82**: 942–948. MR2910041

[3] Dudley, R. M. (1989). *Real Analysis and Probability*, Wadsworth & Brooks/Cole, Pacific Grove, CA. MR0982264

[4] Escanciano, J. C. (2006). Goodness-of-fit tests for linear and nonlinear time series models, *Journal of the American Statistical Association* **101**: 531–541. MR2256173

[5] Escanciano, J. C. (2007). Model checks using residual marked empirical processes, *Statistica Sinica* **17**: 115–138. MR2352505

[6] Escanciano, J. C. and Mayoral, S. (2010). Data-driven smooth tests for the martingale difference hypothesis, *Computational Statistics & Data Analysis* **54**: 1983–1998. MR2640302

[7] Ferland, R., Latour, A. and Oraichi, D. (2006). Integer–valued GARCH processes, *Journal of Time Series Analysis* **27**: 923–942. MR2328548

[8] Fokianos, K. (2012). Count time series, *in* T. Subba Rao, S. Subba Rao and C. R. Rao (eds). *Handbook of Statistics: Time Series - Methods and Applications*, Vol. 30, Elsevier B. V., Amsterdam, 315–347

[9] Fokianos, K., Rahbek, A., and Tjøstheim, D. (2009). Poisson autoregression, *Journal of the American Statistical Association* **104**: 1430–1439. MR2596998

[10] Fokianos, K. and Tjøstheim, D. (2011). Log–linear Poisson autoregression, *Journal of Multivariate Analysis* **102**: 563–578. MR2755016

[11] Fokianos, K. and Tjøstheim, D. (2012). Nonlinear Poisson autoregression, *Annals of the Institute for Statistical Mathematics* **64**: 1205–1225. MR2981620

[12] Franke, J., Kreiss, J.-P., and Mammen, E. (2002). Bootstrap of kernel smoothing in nonlinear time series, *Bernoulli* **8**: 1–37. MR1884156

[13] Freedman, D. A. (1975). On tail probabilities for martingales, *Annals of Probability* **3**: 100–118. MR0380971

[14] Gao, J., King, M., Lu, Z., and Tjøstheim, D. (2009). Specification testing in nonlinear and nonstationary time series regression, *Annals of Statistics* **37**: 3893–3928. MR2572447

[15] Kedem, B. and Fokianos, K. (2002). *Regression Models for Time Series Analysis*, Wiley, Hoboken, NJ. MR1933755

[16] Koul, H. L. and Stute, W. (1999). Nonparametric model checks for time series, *The Annals of Statistics* **27**: 204–236. MR1701108

[17] Lifshits, M. A. (1984). Absolute continuity of functionals of “supremum” type for gaussian processes, *J. Sov. Math.* **27**: 3103–3112.

[18] McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*, 2nd edn, Chapman & Hall, London. MR0727836

[19] Neumann, M. H. (2011). Absolute regularity and ergodicity of Poisson count processes, *Bernoulli* **17**: 1268–1284. MR2854772

[20] Neumann, M. H. and Paparoditis, E. (2008). Goodness-of-fit tests for Markovian time series models: central limit theory and bootstrap approximations, *Bernoulli* **14**: 14–46. MR2401652

[21] Pollard, D. (1984). *Convergence of Stochastic Processes*, Springer-Verlag, New York. MR0762984

[22] Rydberg, T. H. and Shephard, N. (2000). A modeling framework for the prices and times of trades on the New York stock exchange, *in* W. J. Fitzgerlad, R. L. Smith, A. T. Walden and P. C. Young (eds), *Nonlinear and Nonstationary Signal Processing*, Isaac Newton Institute and Cambridge University Press, Cambridge, pp. 217–246. MR1831319

[23] Streett, S. (2000). *Some observation driven models for time series of counts*, PhD thesis, Colorado State University, Department of Statistics. MR2701787

[24] Stute, W., González Manteiga, W., and P. Quindimil, M. (1998). Bootstrap approximations in model checks for regression, *Journal of the American Statistical Association* **93**: 141–149. MR1614600

[25] Teräsvirta, T., Tjøstheim, D., and Granger, C. W. J. (2010). *Modelling Nonlinear Economic Time Series*, Oxford University Press, Oxford.

[26] van der Vaart, A. W. (1998). *Asymptotic Statistics*, Cambridge University Press, Cambridge. MR1652247

[27] Woodard, D. W., Matteson, D. S., and Henderson, S. G. (2011). Stationarity of count-valued and nonlinear time series models, *Electronic Journal of Statistics* **5**: 800–828. MR2824817

[28] Zhu, R. and Joe, H. (2006). Modelling count data time series with Markov processes based on binomial thinning, *Journal of Time Series Analysis* **27**: 725–738. MR2291211