Electronic Journal of Statistics
Vol. 9 (2015) 1357–1377
ISSN: 1935-7524
DOI: 10.1214/15-EJS1044

Estimation and testing linearity for non-linear mixed poisson autoregressions*

Vasiliki Christou†

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

and

Konstantinos Fokianos†

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

Abstract: Non-linear mixed Poisson autoregressive models are studied for the analysis of count time series. Given a correct mean specification of the model, we discuss quasi maximum likelihood estimation based on Poisson log-likelihood function. A score testing procedure for checking linearity of the mean process is developed. We consider the cases of identifiable and non identifiable parameters under the null hypothesis. When the parameters are identifiable then a chi-square approximation to the distribution of the score test is obtained. In the case of non identifiable parameters, a supremum score type test statistic is employed for checking linearity of the mean process. The methodology is applied to simulated and real data.

MSC 2010 subject classifications: Primary 62M09; secondary 62M10.

Keywords and phrases: Bootstrap, chi-square, contraction, identifiability, quasi maximum likelihood, score test, threshold model.

Received January 2015.

*The authors would like to acknowledge the project eMammoth - Compute and Store on Grids and Clouds infrastructure (ANABATHMISI/06609/09), which is co-funded by the Republic of Cyprus and the European Regional Development Fund of the EU. In particular, we are acknowledging computing support by Prof. M. Dikaiakos and Ms. A. Balla (CyGrid). Special thanks to P. Doukhan, A. C. Davison, D. Tjøstheim, the Editor, the Associate Editor and three anonymous reviewers for several constructive comments.

†Supported by Cyprus Research Promotion Foundation TEXNOLOGIA/THEPIS/0609(BE)/02.

Contents

1. Introduction

The aim of this contribution is to study estimation and testing for non-linear mixed Poisson autoregressions. In fact, we enlarge the existing framework of Poisson and negative binomial based autoregressive models, as discussed by [35, Ch.4], [12] and [22], among others. It turns out that the class of mixed Poisson models contains numerous examples of integer-valued models including Poisson-stopped-sum distributions ([32, Sec. 3]), Tweedie-Poisson models ([33, 37]) and other. Mixed Poisson models do not necessarily belong to the exponential family models as discussed by [35] and more recently by [13]; a case in point is the negative binomial distribution with an unknown dispersion parameter.

We discuss ergodicity and stationarity conditions of those models by employing the notion of weak dependence (see [15]). Furthermore, assuming that the mean of the process has been correctly specified, we develop quasi maximum likelihood inference by employing the Poisson log-likelihood function. This approach avoids complicated likelihood functions–as in the case of mixed Poisson models–yet it produces consistent estimates under the correct mean specification. We omit details regarding estimation theory because these can be found in [9] for the case of negative binomial process.

We are particularly interested in testing linearity of the assumed model for the mean process. This question can be attacked by using the likelihood ratio, Wald or score (or Lagrange Multiplier) tests. The score test is often a very convenient tool because it does not require estimation of the model under the alternative. However, its application in the context of mixed Poisson time series modes, needs to be done with care because the test statistic is calculated by employing a quasi–likelihood function. All aforementioned types of test statistics are asymptotically equivalent (cf. [26, Ch. 8]).

Special attention is paid to two classes of nonlinear models specifying the mean process of a count time series under the alternative hypotheses. The first class consists of identifiable models. In this case and under the null hypothesis of linearity, the score test statistic possesses an asymptotic chi-square distribution. The second class consists of models in which a nonnegative nuisance parameter exists only under the alternative hypothesis. Then the testing problem is non-standard and the classical asymptotic theory does not apply; see [11, 2, 44, 1, 40] among others. A notable example of a time series model which is not identifiable under the null, is the threshold model ([45]). We propose a supremum score-type test statistic for dealing with the problem of non-identifiability. Our contribution is summarized by the following:

- An introduction of a general framework for modeling count time series. Indeed, mixed Poisson models include several parametric models which are quite useful to integer valued data analysis.
- Inference and testing linearity based on a quasi-likelihood function. In fact, testing hypotheses by employing a score test which is based on a quasi-likelihood function requires suitable adjustment, as we show in Proposition 3.1.
- The proposed score test gives an additional tool to the data analyst for checking linearity of a given model. Further diagnostics for identification of a suitable model can be found in [34, 10, 13].

The paper is organized as follows. Section 2 discusses mixed Poisson autoregression and gives conditions for proving ergodicity, stationarity and existence of moments for the proposed models. In addition, we develop quasi-likelihood inference to obtain consistent regression parameters. This section generalizes previous results (see [9]); the interested reader is referred to this work for further details. This section lays out basic ideas and notation used throughout the paper. In Section 3, we discuss the score test for testing linearity of the assumed model for the mean process. We consider the cases of identifiable and non identifiable parameters. Section 4 reports empirical results for the performance of the proposed test statistics. In Section 5 the proposed testing methodology is illustrated to a real count time series.

2. Mixed poisson autoregression and inference

Assume that $\{Y_t, t \in \mathbb{Z}\}$ denotes a count time series and let $\{\lambda_t, t \in \mathbb{Z}\}$ be a sequence of mean processes. Denote by $\mathcal{F}_t^{Y,\lambda}$ the past of the process up to and including time t, that is $\mathcal{F}_t^{Y,\lambda} = \sigma(Y_s, s \leq t, \lambda_0)$, where λ_0 denotes some starting value. We will study the following class of count time series models defined by

$$Y_t = \tilde{N}_t(0, Z_t\lambda_t], \quad \lambda_t = f(Y_{t-1}, \lambda_{t-1}), \quad t \geq 1. \tag{2.1}$$

In the above, $\tilde{N}_t$ is a standard homogeneous Poisson process (that is a Poisson process with rate equal to 1) and $\{Z_t\}$ denotes a sequence of independent and identically distributed positive random variables with mean 1, such that $E|Z_t|^r < \infty$, $r \in \mathbb{N}$, which are independent of $\tilde{N}_t$. In addition, we assume that λ_t is measurable with respect to $\{Y_s, s < t\}$ and Z_t is independent of $\{Y_s, s < t\}$.

The family of processes belonging to (2.1) is called mixed Poisson process (see [41], for instance). Two important distributional assumptions are implied by (2.1) and they are routinely employed for the analysis of count time series. Namely, the Poisson distribution given by

$$\mathrm{P}[Y_t = y \mid \mathcal{F}_{t-1}^{Y,\lambda}] = \frac{\exp(-\lambda_t)\lambda_t^y}{y!}, \quad y = 0, 1, 2, \dots \tag{2.2}$$

and the negative binomial distribution given by

$$\mathrm{P}[Y_t = y \mid \mathcal{F}_{t-1}^{Y,\lambda}] = \frac{\Gamma(\nu + y)}{\Gamma(y+1)\Gamma(\nu)} \Big(\frac{\nu}{\nu + \lambda_t}\Big)^{\nu} \Big(\frac{\lambda_t}{\nu + \lambda_t}\Big)^{y}, \quad y = 0, 1, 2, \dots, \tag{2.3}$$

where $\nu > 0$. Note that (2.2) is a special case of (2.1) when $\{Z_t\}$ is a sequence of degenerate random variables with mean 1. Furthermore, (2.3) is a special case of (2.1) when $\{Z_t\}$ are iid Gamma with mean 1 and variance $1/\nu$. Related work for the case of the negative binomial is that of [49] but with a different parametrization and when the function $f(\cdot, \cdot)$ of (2.1) is linear (see (2.5)). Regardless of the choice of $Z's$, the conditional mean of $\{Y_t\}$ as given by (2.1) is always equal to λ_t. Furthermore, the variance of (2.1) is given by $\lambda_t + \sigma_Z^2 \lambda_t^2$, with $\sigma_Z^2 = \mathrm{Var}(Z_t)$. The conditional variance of the Poisson distribution is equal to λ_t, whereas the conditional variance of (2.3) is equal to $\lambda_t + \lambda_t^2/\nu$.

We show that modeling based on (2.1) generalizes several existing results reported in the literature. Consider, for instance, the work by [50] who suggests the generalized Poisson distribution for modeling count time series data; see [32, pp. 396] who prove that the generalized Poisson distribution is proper distribution for a certain range of parameter values. In fact, the generalized Poisson distribution is a Poisson–stopped–sum distribution; hence it is an infinitely divisible distribution. [31] use the property of infinite divisibility to show that a proper generalized Poisson distribution is a mixed Poisson distribution. Hence, (2.1) covers the case of generalized Poisson distribution after suitable model reparametrization. Moreover, likelihood based inference for the generalized Poisson distribution is implemented by constraining its parameters; see [32, pp. 399] for more. In general, mixed Poisson distributions with an infinitely divisible mixing random variable Z are Poisson–stopped–sum distribution; [32, pp. 324].

Another count time series model, which was suggested by [51], is that of zero-inflated Poisson (negative binomial) distribution. These models are included in the framework introduced by (2.1). When the mixing variables Z_t are binary, then we obtain a zero-inflated Poisson model. A similar argument yields a zero-inflated negative binomial model. In this contribution, we will employ quasi-maximum likelihood for estimation. In the context of zero-inflated Poisson and negative binomial models, [51] employs the E-M algorithm. We point out that [50, 51] considers linear autoregressions but we consider nonlinear models, in general. As a final remark, the works of [33] and [37] show that suitable choice of the mixing variables Z_t in (2.1) yields to Tweedie–Poisson exponential dispersion models which include the Neyman type A and the Poisson-inverse Gaussian distributions.

2.1. Modeling

Assume that the mean process of (2.1) is given by

$$\lambda_t = f(Y_{t-1}, \lambda_{t-1}), \quad t \geq 1, \tag{2.4}$$

where $f(\cdot,\cdot)$ is a parametric function defined on $\mathbb{N}_0 \times \mathbb{R}_+$ and taking values on $(0,\infty)$, where $\mathbb{N}_0 = \{0,1,2,\dots\}$. For instance, consider the linear model

$$\lambda_t = d + a_1\lambda_{t-1} + b_1 Y_{t-1}, \tag{2.5}$$

where $d > 0, a_1 > 0, b_1 > 0$, such that $a_1 + b_1 < 1$. Model (2.5) has been shown to be stationary with any moments (see [22, 43] for the Poisson case and [9] for the negative binomial case). Some examples of (2.4) include

$$\lambda_t = \frac{d}{(1+Y_{t-1})^{\gamma}} + a_1\lambda_{t-1} + b_1 Y_{t-1}, \tag{2.6}$$

$$\lambda_t = d + a_1\lambda_{t-1} + (b_1 + c_1\exp(-\gamma Y_{t-1}^2))Y_{t-1}, \tag{2.7}$$

and

$$\lambda_t = d + a_1\lambda_{t-1} + b_1 Y_{t-1} + (d_2 + a_2\lambda_{t-1} + b_2 Y_{t-1}) I(Y_{t-1} \leq r). \tag{2.8}$$

Models (2.6) and (2.7) are modifications of analogous models studied by [43, 23, 19, 9]. Model (2.6) introduces a deviation of (2.5) in the sense that small values of γ make (2.6) to approach (2.5). [9] show that when $\max\{b_1, d\gamma - b_1\} + a_1 < 1$, model (2.6) is ergodic and stationary whose moments are finite. The same holds for the mixed Poisson process specification given by (2.1). Similarly model (2.7) can be viewed as a Smooth Transition Autoreggressive (STAR) model (see [44]). It turns out that when $0 < a_1 + b_1 + c_1 < 1$, model (2.7) is ergodic, stationary and it has moments of all orders. Model (2.8) is a threshold model and has been studied recently by [47, 18, 46]. In particular, [46] show that (2.8) posses a stationary and ergodic solution, with any moments if $0 < a_1 + b_1 < 1$ and $a_1 + a_2 < 1$, under the Poisson assumption (provided that $\min(d_1, a_1, b_1, d+d_2, a_1+a_2, b_1+b_2) > 0$). Furthermore, to obtain the asymptotic distribution of the MLE, they stipulate the condition $b_1 + b_2 < 1$. If all the coefficients are positive, then the condition $a_1 + a_2 + b_1 + b_2 < 1$ guarantees that (2.8) has a unique stationary solution which possess any moments; in addition the MLE of the parameter vector $(a_1, a_2, b_1, b_2)'$ is consistent and asymptotically normally distributed (see also [47, 18] for stationarity and consistency of the MLE for model (2.8)). We are not aware of any study regarding the probabilistic properties of the threshold model under any other distributional assumption. The following proposition generalizes [9, Thm.1] and it applies to (2.6) and (2.7); see also [13] in the context of exponential family.

Proposition 2.1. *Suppose that $\{Y_t, t \in \mathbb{Z}\}$ is a count time series specified by (2.1) with a mean process $\{\lambda_t, t \in \mathbb{Z}\}$ given by (2.4). Let $\{Z_t\}$ be a sequence of* *independent and identically distributed positive random variables with mean 1, such that* $E|Z_t|^r < \infty$, $r \in \mathbb{N}$, *which are independent of* $\tilde{N}_t$ *and independent of* $\{Y_s, s < t\}$. *Suppose that there exist constants* α_1, α_2 *of non-negative real numbers such that*

$$|f(y, \lambda) - f(y', \lambda')| \leq \alpha_1 |\lambda - \lambda'| + \alpha_2 |y - y'|.$$

Assume that $\alpha = \alpha_1 + \alpha_2 < 1$. *Then there exists a unique causal solution* $\{(Y_t, \lambda_t), t \in \mathbb{Z}\}$ *to model (2.4) which is stationary, ergodic and for any* $r \in \mathbb{N}$ *satisfies* $E\|(Y_0, \lambda_0)\|^r < \infty$.

2.2. Inference

For the case of mixed Poisson models (2.1), it is rather challenging, in general, to have readily available a likelihood function, because the distribution of the mixing variable Z_t is generally unknown. Hence, we resort to a quasi maximum likelihood (QMLE) methodology. This method is quite analogous to quasi-likelihood inference developed for estimation and fitting of ordinary GARCH models. For instance [4, 25, 42, 3] among others, study the Gaussian likelihood function irrespectively of the assumed error distribution. It turns out that QMLE are consistent estimators of regression parameters under a correct mean process specification (see also [48, 28, 30], for instance). To define properly the QMLE, consider the Poisson log-likelihood function, conditional on some starting value λ_0,

$$l_n(\boldsymbol{\theta}) = \sum_{t=1}^{n} l_t(\boldsymbol{\theta}) = \sum_{t=1}^{n} \left(Y_t \log \lambda_t(\boldsymbol{\theta}) - \lambda_t(\boldsymbol{\theta})\right), \tag{2.9}$$

where $\boldsymbol{\theta}$ denotes the unknown parameter vector. The quasi-score function is defined by

$$\boldsymbol{S}_n(\boldsymbol{\theta}) \quad = \quad \frac{\partial l_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{t=1}^{n} \frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{t=1}^{n} \left(\frac{Y_t}{\lambda_t(\boldsymbol{\theta})} - 1\right) \frac{\partial \lambda_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}. \tag{2.10}$$

The solution of the system of nonlinear equations $\boldsymbol{S}_n(\boldsymbol{\theta}) = \mathbf{0}$, if it exists, yields the QMLE of $\boldsymbol{\theta}$ which we denote by $\hat{\boldsymbol{\theta}}$. The conditional information matrix is defined by

$$\boldsymbol{G}_n(\boldsymbol{\theta}) \quad = \quad \sum_{t=1}^{n} \mathrm{Var}\Big[\frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \Big| \mathcal{F}_{t-1}^{Y,\lambda}\Big] = \sum_{t=1}^{n} \Big(\frac{1}{\lambda_t(\boldsymbol{\theta})} + \sigma_Z^2\Big) \left(\frac{\partial \lambda_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right) \left(\frac{\partial \lambda_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)'.$$

It can be shown, under the assumptions of [9, Thm. 2] that $\hat{\boldsymbol{\theta}}$ is consistent and asymptotically normally distributed; that is

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta_0}) \xrightarrow{D} \mathcal{N}(\mathbf{0}, \boldsymbol{G}^{-1}(\boldsymbol{\theta_0}) \boldsymbol{G}_1(\boldsymbol{\theta_0}) \boldsymbol{G}^{-1}(\boldsymbol{\theta_0})), \tag{2.11}$$

where the matrices $\boldsymbol{G}$ and $\boldsymbol{G}_1$ are given by

$$\boldsymbol{G}(\boldsymbol{\theta}) = \mathrm{E}\left(\frac{1}{\lambda_t(\boldsymbol{\theta})}\left(\frac{\partial\lambda_t}{\partial\boldsymbol{\theta}}\right)\left(\frac{\partial\lambda_t}{\partial\boldsymbol{\theta}}\right)'\right), \tag{2.12}$$

and

$$\boldsymbol{G}_1(\boldsymbol{\theta}) = \mathrm{E}\left(\left(\frac{1}{\lambda_t(\boldsymbol{\theta})} + \sigma_Z^2\right)\left(\frac{\partial\lambda_t}{\partial\boldsymbol{\theta}}\right)\left(\frac{\partial\lambda_t}{\partial\boldsymbol{\theta}}\right)'\right). \tag{2.13}$$

If $\sigma_Z^2 > 0$, then an estimator, say $\hat{\sigma}_Z^2$, is given as the solution of

$$\sum_{t=1}^{n} \frac{(Y_t - \hat{\lambda}_t)^2}{\hat{\lambda}_t(1+\hat{\lambda}_t\sigma_Z^2)} = n - m, \tag{2.14}$$

where m denotes the dimension of $\boldsymbol{\theta}$ and $\hat{\lambda}_t = \lambda_t(\hat{\boldsymbol{\theta}})$. There exists at most one solution of (2.14). If it exists, then it is consistent and can be calculated by existing software (cf. [14, 6]). For the case $\sigma_Z^2 = 1/\nu$, see [7, Ch. 3] and [38].

3. Testing linearity

Testing linearity, within a parametric framework, is a problem which is attacked by computing the likelihood ratio, Wald and score tests. The likelihood ratio and Wald tests require estimation for the full model which can be computationally challenging. Consider (2.7), for instance. To obtain the maximum likelihood estimation of γ, we need rather large sample sizes. The problem's complexity increases especially for small values of γ, as empirical experience has shown; see also [52, Ch.18, p. 684] who consider smooth transition models. For (2.8), it is well known, from the linear model estimation theory, that the threshold parameter r is not estimated at the rate $n^{-1/2}$ (see [8]). Hence, testing and inference in the context of (2.8) rise challenging computational and theoretical issues.

The computational advantage of the score test is that it is calculated after estimating the constrained model under the null. In other words, testing for linearity requires estimation of the simple linear model (2.5). In addition, the asymptotic distribution of the score statistic is not affected when parameters lie at the boundary of the hypothesis, (see [26, Ch. 8]). Denote by $\boldsymbol{\theta} = (\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)})$ the unknown parameter, where $\boldsymbol{\theta}^{(1)}$ and $\boldsymbol{\theta}^{(2)}$ are vectors of dimension m_1 and m_2, respectively, such that $m_1 + m_2 = m$. The hypotheses of interest are

$$H_0 : \boldsymbol{\theta}^{(2)} = \boldsymbol{\theta}_0^{(2)} \quad \text{vs.} \quad H_1 : \boldsymbol{\theta}^{(2)} > \boldsymbol{\theta}_0^{(2)}, \quad \text{componentwise.}$$

Let $\tilde{\boldsymbol{\theta}}_n = (\tilde{\boldsymbol{\theta}}_n^{(1)}, \tilde{\boldsymbol{\theta}}_n^{(2)})$ be the constrained quasi-likelihood estimator of $\boldsymbol{\theta} = (\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)})$. Put $\boldsymbol{S}_n = (\boldsymbol{S}_n^{(1)}, \boldsymbol{S}_n^{(2)})$ for the corresponding partition of the score function. The general form of the score statistic is given by (cf. [6] and [29, Ch. 5]),

$$LM_n = \boldsymbol{S}_n^{(2)'}(\tilde{\boldsymbol{\theta}}_n)\widetilde{\boldsymbol{\Sigma}}^{-1}(\tilde{\boldsymbol{\theta}}_n)\boldsymbol{S}_n^{(2)}(\tilde{\boldsymbol{\theta}}_n). \tag{3.1}$$

In the above, $\widetilde{\boldsymbol{\Sigma}}$ is an appropriate estimator for the covariance matrix $\boldsymbol{\Sigma} = \mathrm{Var}(\boldsymbol{S}_n^{(2)}(\tilde{\boldsymbol{\theta}}_n)/\sqrt{n})$.
Because this approach is based on quasi–score instead of the true score, certain adjustments should be made for obtaining its asymptotic distribution. Recall (2.12) and consider the following partition

$$\boldsymbol{G} = \begin{pmatrix} \boldsymbol{G}_{11} & \boldsymbol{G}_{12} \\ \boldsymbol{G}_{21} & \boldsymbol{G}_{22} \end{pmatrix},$$

and similarly for $\boldsymbol{G}_1$ (see (2.13)). Then it can be shown that

$$\boldsymbol{\Sigma} \equiv \boldsymbol{\Sigma}_{MP} = \boldsymbol{G}_{1,22} - \boldsymbol{G}_{21}\boldsymbol{G}_{11}^{-1}\boldsymbol{G}_{1,12} - \boldsymbol{G}_{1,21}\boldsymbol{G}_{11}^{-1}\boldsymbol{G}_{12} + \boldsymbol{G}_{21}\boldsymbol{G}_{11}^{-1}\boldsymbol{G}_{1,11}\boldsymbol{G}_{11}^{-1}\boldsymbol{G}_{12}. \tag{3.2}$$

If the true distribution is Poisson, then the matrices $\boldsymbol{G}$ and $\boldsymbol{G}_1$ coincide and

$$\boldsymbol{\Sigma} \equiv \boldsymbol{\Sigma}_P = \boldsymbol{G}_{22} - \boldsymbol{G}_{21}\boldsymbol{G}_{11}^{-1}\boldsymbol{G}_{12}.$$

3.1. Standard implementation

If all the parameters are identified under the null hypothesis, then the standard asymptotic theory holds as the following result shows.

Proposition 3.1. *Suppose that $\{Y_t, t = 1, \ldots, n\}$ is a count time series specified by (2.1) with mean process $\{\lambda_t\}$ defined by (2.4) and suppose that the assumptions of Proposition 2.1 and [9, Thm.2] hold true. Suppose that the function $f(\cdot)$ depends upon a vector $(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)})$, where $\boldsymbol{\theta}^{(i)}$ is of dimension $m_i, i = 1, 2$. Consider the problem*

$$H_0 : \boldsymbol{\theta}^{(2)} = \boldsymbol{\theta}_0^{(2)} \quad vs. \quad H_1 : \boldsymbol{\theta}^{(2)} > \boldsymbol{\theta}_0^{(2)}, \quad componentwise. \tag{3.3}$$

Then the score test statistic converges to a chi-square random variable, i.e.

$$LM_n = \boldsymbol{S}_n^{(2)'}(\tilde{\boldsymbol{\theta}}_n)\widetilde{\boldsymbol{\Sigma}}^{-1}(\tilde{\boldsymbol{\theta}}_n)\boldsymbol{S}_n^{(2)}(\tilde{\boldsymbol{\theta}}_n) \xrightarrow{D} \mathcal{X}_{m_2}^2,$$

as $n \to \infty$, when H_0 is true. In addition, consider testing

$$H_0 : \boldsymbol{\theta}^{(2)} = \boldsymbol{\theta}_0^{(2)} \quad vs. \quad H_1 : \boldsymbol{\theta}^{(2)} = \boldsymbol{\theta}_0^{(2)} + n^{-1/2}\boldsymbol{\delta},$$

where $\boldsymbol{\delta}$ is a fixed vector in $\mathbb{R}_+^{m_2}$. Then

$$LM_n = \boldsymbol{S}_n^{(2)'}(\tilde{\boldsymbol{\theta}}_n)\widetilde{\boldsymbol{\Sigma}}^{-1}(\tilde{\boldsymbol{\theta}}_n)\boldsymbol{S}_n^{(2)}(\tilde{\boldsymbol{\theta}}_n) \xrightarrow{D} \mathcal{X}_{m_2}^2(\boldsymbol{\delta}'\boldsymbol{\Delta}\boldsymbol{\delta}),$$

with $\boldsymbol{\Delta} = \widetilde{\boldsymbol{\Sigma}}_P\widetilde{\boldsymbol{\Sigma}}_{MP}^{-1}\widetilde{\boldsymbol{\Sigma}}_P$.

The proof is postponed to the appendix.

Recall (2.6) and let $\boldsymbol{\theta}^{(1)} = (d, a_1, b_1)'$ and $\boldsymbol{\theta}^{(2)} = \gamma$. The hypothesis $H_0 : \gamma = 0$ implies that the parameter γ lies at the boundary of the parameter space. Following [26, pp. 196], Proposition 3.1 assures that the quasi-likelihood based score test statistic follows asymptotically the chi-square distribution with one degree of freedom for testing $\gamma = 0$.

3.2. Non-standard implementation

We deviate from the notation used so far in this section so that we can bring across the main ideas. We are interested on testing linearity of the mean process when a non-linear model contains nuisance parameters that are not identified under the null; recall (2.7) and (2.8). Then the quasi maximum likelihood estimators have nonstandard behavior. The lack of identification affects also the score test and the classical asymptotic theory does not apply.
Consider (2.7). When $c_1 = 0$, the parameter γ is not identified under the null. Consequently, testing of

$$H_0 : c_1 = 0 \quad \text{vs.} \quad H_1 : c_1 > 0, \tag{3.4}$$

cannot be implemented in a standard way and the traditional large sample theory does not apply. To deal with this problem, we consider a fixed arbitrary value of γ. Then the model is still linear in the parameters and the score statistic LM_n is asymptotically distributed as a chi–square random variable with one degree of freedom, under the null. Despite the fact that this test is in general consistent, even for alternatives where $\gamma_0 \neq \gamma$, it may lack power for alternatives where γ_0 is far from γ. To avoid low values of power obtained by the above approach, we propose a supremum type of test statistic. Consider Γ, a grid of values for the parameter γ. Then the sup-score test statistic is given by

$$LM_n = \sup_{\gamma \in \Gamma} LM_n(\gamma), \tag{3.5}$$

where $LM_n(\gamma)$ is given by (3.1). We reject hypothesis (3.4) for large values of LM_n. Critical values are calculated by bootstrapping the test statistic.

Recall now (2.8) and consider testing the following hypotheses

$$H_0 : d_2 = a_2 = b_2 = 0 \quad \text{vs.} \quad H_1 : d_2 > 0 \;\text{ or }\; a_2 > 0 \;\text{ or }\; b_2 > 0. \tag{3.6}$$

Similarly, consider a grid of values for the threshold parameter r, say Γ, and calculate

$$LM_n = \sup_{r \in \Gamma} LM_n(r).$$

We reject (3.6) for large values of the test statistic.

Remark 3.1. It is of interest to develop the asymptotic distribution of the supremum type score test statistic LM_n similarly to the works by [24] and [39]. Note that Proposition 2.1 shows that the joint process (Y_t, λ_t) is τ-weakly dependent ([15, 20]). However a uniform central limit theorem for multivariate τ-weakly dependent process is not available in the literature, to the best of our knowledge (but see [16] who prove a central limit theorem for the case of univariate empirical distribution function). For univariate functions of mixing process, see additionally the recent work by [17]. The main challenge is to extend the work of [16] to multivariate empirical processes indexed by classes of functions. We conjecture that under the null hypothesis and using the assumptions of Proposition 2.1, the process $\sup_\gamma LM_n(\gamma)$, as defined by (3.5), will converge to the supremum of a chi-square process (see also [11]). Our work focuses explicitly on the computational aspects of LM_n and the problem of obtaining its asymptotic distribution will be examined elsewhere.

In the following, we examine the finite sample behavior of the score test under both cases of identifiable and non identifiable parameters.

4. Simulation study

We present a limited simulation study to demonstrate empirically the theoretical results. For this work, we employ parametric bootstrap based either on the Poisson distribution or the negative binomial distribution. More specifically, given the data, we estimate by QMLE the parameters of model (2.5) and calculate the score test statistic LM_n given by (3.1). Then, we generate B time series of length n using the estimated model under H_0. For each time series, $\mathrm{b} = 1, \ldots, \mathrm{B}$, we compute the value of the score statistic denoted by LM_{b}^* and compute the p-values by the formula

$$\text{p-value} = \frac{\#\{\mathrm{b} : LM_{\mathrm{b}}^* \geq LM\} + 1}{\mathrm{B} + 1}.$$

In the case of a non identifiable parameter, the score test is computed by choosing a grid for the values of the non identifiable parameter and then take the supremum over this grid. We use $\mathrm{B} = 499$ bootstrap replicates and 200 simulations for various sample sizes.

4.1. Results for identifiable models

We consider the size of the proposed score test statistic. Recall model (2.5) and choose $(d, a_1, b_1) = (1.5, 0.05, 0.6)$ and $n = 250, 500$ and 1000. We maximize the quasi log-likelihood function (2.9) by a quasi-Newton method using the `constrOptim()` function of `R` to obtain the QMLE. For selected nominal significance level $\alpha = 1\%, \alpha = 5\%$ or $\alpha = 10\%$ we obtain the results reported by Table 1 for the size of the test statistic (3.1). For comparison, the last three columns list the achieved significance levels of the test derived from the asymptotic chi-square distribution with one degree of freedom. The empirical results show that the bootstrap approximation performs well; in fact it yields achieved significance levels which are closer to nominal significance levels, especially for smaller sample size.

To study the power of the test statistic we work analogously under the model

$$\lambda_t = \frac{1.5}{(1 + Y_{t-1})^\gamma} + 0.05\lambda_{t-1} + 0.6Y_{t-1},$$

where $\gamma = 0.3, 0.5, 1$. Table 2–which is obtained in the same manner as Table 1–shows that as γ assumes larger values, the power of the test statistic (3.1) approaches unity; see also Figure 1.

TABLE 1

Empirical size of the test statistic (3.1) for testing $H_0 : \gamma = 0$ of model (2.6) and for sample sizes $n = 250, 500$ and 1000. Data are generated from the linear model (2.5) with true values $(d, a_1, b_1) = (1.5, 0.05, 0.6)$

	Bootstrap test for $n = 250$			Approximation test for $n = 250$		
Nominal significance level	Poisson	NegBin ($\nu = 2$)	NegBin ($\nu = 4$)	Poisson	NegBin ($\nu = 2$)	NegBin ($\nu = 4$)
$\alpha = 1\%$	0.000	0.015	0.000	0.005	0.015	0.010
$\alpha = 5\%$	0.045	0.055	0.020	0.055	0.065	0.060
$\alpha = 10\%$	0.090	0.111	0.050	0.145	0.115	0.115
	Bootstrap test for $n = 500$			Approximation test for $n = 500$		
$\alpha = 1\%$	0.010	0.005	0.005	0.020	0.010	0.015
$\alpha = 5\%$	0.041	0.068	0.037	0.075	0.030	0.055
$\alpha = 10\%$	0.096	0.094	0.084	0.150	0.075	0.110
	Bootstrap test for $n = 1000$			Approximation test for $n = 1000$		
$\alpha = 1\%$	0.010	0.010	0.026	0.010	0.005	0.010
$\alpha = 5\%$	0.057	0.037	0.051	0.060	0.020	0.035
$\alpha = 10\%$	0.114	0.115	0.097	0.115	0.055	0.090

TABLE 2

Empirical power of the test statistic (3.1) for testing $H_0 : \gamma = 0$ of model (2.6) for sample sizes $n = 250, 500$ and 1000. Data are generated from (2.6) with true values $(d, a_1, b_1) = (1.5, 0.05, 0.6)$ and $\gamma \in \{0.3, 0.5, 1\}$. The nominal significance level is set to $\alpha = 5\%$

Nonlinear model (2.6)	Bootstrap test for $n = 250$			Approximation test for $n = 250$		
γ	Poisson	NegBin ($\nu = 2$)	NegBin ($\nu = 4$)	Poisson	NegBin ($\nu = 2$)	NegBin ($\nu = 4$)
$\gamma = 0.3$	0.115	0.050	0.117	0.110	0.025	0.065
$\gamma = 0.5$	0.255	0.082	0.184	0.230	0.065	0.165
$\gamma = 1$	0.652	0.508	0.579	0.650	0.370	0.545
	Bootstrap test for $n = 500$			Approximation test for $n = 500$		
$\gamma = 0.3$	0.207	0.061	0.157	0.210	0.045	0.130
$\gamma = 0.5$	0.450	0.251	0.424	0.485	0.240	0.300
$\gamma = 1$	0.924	0.756	0.837	0.920	0.710	0.870
	Bootstrap test for $n = 1000$			Approximation test for $n = 1000$		
$\gamma = 0.3$	0.271	0.144	0.212	0.360	0.185	0.250
$\gamma = 0.5$	0.740	0.548	0.688	0.745	0.455	0.635
$\gamma = 1$	1.000	0.974	0.995	0.995	0.955	0.990

4.2. Results for non identifiable models

To examine the performance of the sup-score test statistic (3.5) for testing model (2.7), put

$$\lambda_t = 0.5 + 0.3\lambda_{t-1} + (0.2 + c_1 \exp(-\gamma Y_{t-1}^2))Y_{t-1},$$

where $c_1 \in \{0.2, 0.4\}$ and $\gamma \in \{0.05, 0.2, 0.5, 1\}$. Consider testing the hypotheses (3.4). The sup-score test statistic (3.5) is calculated by choosing a grid which consists of a sequence of thirty equidistant values in the interval $[0.01, 2]$. Table 3 shows that the test statistic, generally, achieves its nominal significance level. However, larger sample sizes yield to more accurate approximation of the nominal level. Table 4 shows that the power of the test statistic is relatively

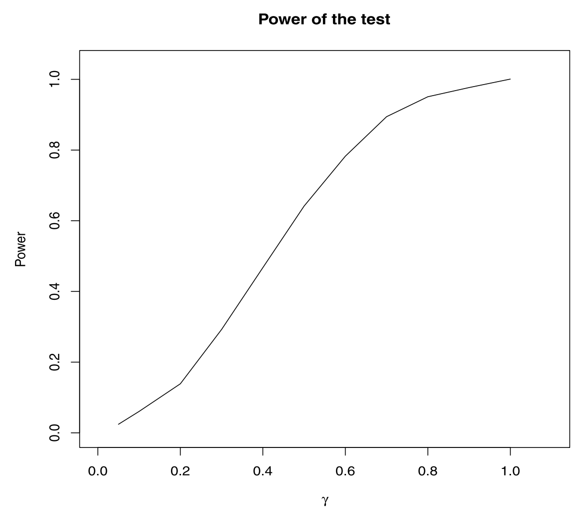

Fig 1. *Plot for the power of the test statistic* (3.1) *for testing* $H_0 : \gamma = 0$ *for model* (2.6). *Data are generated from the nonlinear model (2.6) with true values* $(d, a_1, b_1) = (1.5, 0.05, 0.6)$, $\nu = 4$ *and for different values of* γ. *Results are based on* 1000 *data points,* 500 *simulations and on the asymptotic approximation based on the chi-square distribution with one degree of freedom.*

Table 3

Empirical size of the test statistic (3.5) *for testing* (3.3) *of model* (2.7) *and for sample sizes* $n = 250, 500$ *and* 1000. *Data are generated from the linear model* (2.5) *with true values* $(d, a_1, b_1) = (0.5, 0.3, 0.2)$

	Bootstrap test for $n = 250$		
Nominal significance level	Poisson	NegBin ($\nu = 2$)	NegBin ($\nu = 4$)
$\alpha = 1\%$	0.000	0.000	0.009
$\alpha = 5\%$	0.028	0.004	0.062
$\alpha = 10\%$	0.112	0.065	0.106
	Bootstrap test for $n = 500$		
$\alpha = 1\%$	0.000	0.000	0.010
$\alpha = 5\%$	0.048	0.043	0.037
$\alpha = 10\%$	0.122	0.102	0.099
	Bootstrap test for $n = 1000$		
$\alpha = 1\%$	0.000	0.000	0.005
$\alpha = 5\%$	0.020	0.045	0.060
$\alpha = 10\%$	0.046	0.075	0.105

large for increasing sample sizes and for values of γ close to zero provided that the parameter c_1 is also of large magnitude.

Consider next the threshold model which specifies the time series mean process by model (2.8). The empirical results for testing (3.6) are based on the Poisson assumption. The sup-score test statistic is computed by defining a grid of values for the threshold parameter r. Following the suggestion of [46], let the grid defined by a sequence of ten equidistant values in the interval $[q_1, q_2]$,

TABLE 4
Empirical power for test statistic (3.5) *for testing* (3.3) *of model* (2.7) *and for sample sizes* $n = 250, 500$ *and* 1000. *Data are generated from the exponential model* (2.7) *with true values* $(d, a_1, b_1) = (0.5, 0.3, 0.2)$, $c_1 \in \{0.2, 0.4\}$ *and* $\gamma \in \{0.05, 0.2, 0.5, 1\}$. *The nominal significance level is* $\alpha = 5\%$

Nonlinear model (2.7)		Bootstrap test for $n = 250$		
c_1	γ	Poisson	NegBin ($\nu = 2$)	NegBin ($\nu = 4$)
$c_1 = 0.2$	$\gamma = 0.05$	0.105	0.132	0.069
$c_1 = 0.2$	$\gamma = 0.2$	0.110	0.120	0.090
$c_1 = 0.2$	$\gamma = 0.5$	0.044	0.062	0.037
$c_1 = 0.2$	$\gamma = 1$	0.000	0.045	0.055
$c_1 = 0.4$	$\gamma = 0.05$	0.356	0.302	0.404
$c_1 = 0.4$	$\gamma = 0.2$	0.349	0.341	0.319
$c_1 = 0.4$	$\gamma = 0.5$	0.202	0.212	0.263
$c_1 = 0.4$	$\gamma = 1$	0.091	0.102	0.067
		Bootstrap test for $n = 500$		
$c_1 = 0.2$	$\gamma = 0.05$	0.140	0.178	0.220
$c_1 = 0.2$	$\gamma = 0.2$	0.199	0.175	0.231
$c_1 = 0.2$	$\gamma = 0.5$	0.111	0.156	0.122
$c_1 = 0.2$	$\gamma = 1$	0.077	0.070	0.084
$c_1 = 0.4$	$\gamma = 0.05$	0.755	0.690	0.739
$c_1 = 0.4$	$\gamma = 0.2$	0.746	0.609	0.628
$c_1 = 0.4$	$\gamma = 0.5$	0.420	0.358	0.469
$c_1 = 0.4$	$\gamma = 1$	0.174	0.154	0.235
		Bootstrap test for $n = 1000$		
$c_1 = 0.2$	$\gamma = 0.05$	0.315	0.390	0.355
$c_1 = 0.2$	$\gamma = 0.2$	0.485	0.335	0.372
$c_1 = 0.2$	$\gamma = 0.5$	0.312	0.190	0.265
$c_1 = 0.2$	$\gamma = 1$	0.105	0.122	0.135
$c_1 = 0.4$	$\gamma = 0.05$	0.985	0.950	0.985
$c_1 = 0.4$	$\gamma = 0.2$	0.985	0.905	0.925
$c_1 = 0.4$	$\gamma = 0.5$	0.855	0.650	0.775
$c_1 = 0.4$	$\gamma = 1$	0.465	0.255	0.340

where q_1 and q_2 are set to be respectively the empirical 0.2 and 0.8 quantile of each time series replication. Table 5 (respectively, Table 6) reports simulation results on the size (respectively, power) of the test statistic. The approximation of nominal significance levels improves for large sample sizes. Furthermore, the power of the test statistic increases for values of parameters a_2 and b_2 of large magnitude.

5. Example

We consider a time series of weekly number of measles at Sheffield for the period between September 8th, 1978 and April 17th, 1987. The total number of observations is 450. The sample mean (sample variance, respectively) of these data is 17.151 (265.781, respectively), showing strong overdispersion. Employing the quasi-likelihood methodology as outlined in Section 2.2, we obtain the QMLE of the regression parameters for each of models (2.5)–(2.8). Table 7 summarizes the results. We also report standard errors (in parentheses) where the first row shows

Table 5

Empirical size of test statistic (3.5) for testing (3.6) of model (2.8) and for sample sizes $n = 250, 500$ and 1000. Data are generated from the linear model (2.5) with true values $(d, a_1, b_1) = (0.5, 0.05, 0.1)$

Nominal significance level	$n = 250$	$n = 500$	$n = 1000$
$\alpha = 1\%$	0.000	0.010	0.005
$\alpha = 5\%$	0.000	0.031	0.050
$\alpha = 10\%$	0.063	0.061	0.076

Table 6

Empirical power of test statistic (3.5) for testing (3.6) of model (2.8) and for sample sizes $n = 250, 500$ and 1000. Data are generated from the threshold model (2.8) with true values $(d_1, a_1, b_1, d_2) = (0.5, 0.05, 0.1, 0.7)$, $a_2 \in \{0.1, 0.2\}$, $b_2 \in \{0.1, 0.3\}$ and $r = 5$. The nominal significance level is $\alpha = 5\%$

a_2	b_2	$n = 250$	$n = 500$	$n = 1000$
$a_2 = 0.1$	$b_2 = 0.1$	0.030	0.032	0.075
$a_2 = 0.2$	$b_2 = 0.1$	0.032	0.071	0.232
$a_2 = 0.2$	$b_2 = 0.3$	0.326	0.813	0.960

Table 7

Quasi maximum likelihood estimators and their standard errors (in parentheses) for the linear model (2.5) and the nonlinear models (2.6), (2.7) and (2.8), for the monthly number of measles at Sheffield for the period between September 8th, 1978 and April 17th, 1987. The total number of observations is 450

Model	Quasi Maximum Likelihood Estimators								
	$\hat{d}$	$\hat{a}_1$	$\hat{b}_1$	$\hat{\gamma}$	$\hat{c}_1$	$\hat{d}_2$	$\hat{a}_2$	$\hat{b}_2$	$\hat{r}$
(2.5)	0.720	0.490	0.469	–	–	–	–	–	–
	(0.122)	(0.024)	(0.023)						
	(0.235)	(0.057)	(0.055)						
(2.6)	1.274	0.490	0.486	0.468	–	–	–	–	–
	(0.507)	(0.024)	(0.024)	(0.305)					
	(0.858)	(0.058)	(0.062)	(0.591)					
(2.7)	0.712	0.492	0.433	0.596	0.041	–	–	–	–
	(0.270)	(0.043)	(0.043)	(15.771)	(1.300)				
	(0.504)	(0.090)	(0.099)	(21.751)	(1.785)				
(2.8)	0.720	0.490	0.469	–	–	0.700	0.009	0.010	3
	(0.172)	(0.026)	(0.025)			(0.640)	(0.093)	(0.269)	
	(0.302)	(0.050)	(0.050)			(0.915)	(0.140)	(0.381)	

the standard errors under the Poisson assumption and the second row shows the standard errors obtained by the sandwich matrix $\boldsymbol{G}^{-1}(\hat{\boldsymbol{\theta}})\boldsymbol{G}_1(\hat{\boldsymbol{\theta}})\boldsymbol{G}^{-1}(\hat{\boldsymbol{\theta}})$; recall (2.11). For model (2.5) and under the negative binomial working assumption the estimator of ν is $\hat{\nu} = 5.309$. Similarly, for model (2.6) is given by $\hat{\nu} = 5.303$, for model (2.7) is $\hat{\nu} = 5.283$ and for (2.8) is $\hat{\nu} = 5.309$. The estimator of the threshold parameter r is obtained by a profiling procedure whereby we calculate the log-likelihood function (2.9) for a grid of values of r and then we choose r as the value that maximizes the log-likelihood function. As a general comment, note that all results obtained indicate that there exists some non stationarity in these data.

TABLE 8

Bootstrap p-values for testing linearity for the mean process of the monthly number of measles at Sheffield for the period between September 8*th,* 1978 *and April* 17*th,* 1987*. The total number of observations is 450*

Model	Poisson	NegBin
(2.6)	0.362	0.600
(2.7)	0.622	0.618
(2.8)	0.104	0.368

To test for linearity for each of these models, we use B = 499 bootstrap replications and calculate the p-values for each model and for each distributional assumption. Table 8 reports the results and indicates that the linear model (2.5) is always accepted. This is in accordance with the results obtained by Table 7 when we compare the estimators with their standard errors. To support further this conclusion we have calculated the scoring rules considered by [34, 10, 13]. All results obtained by fitting different models under different response distributions are close to the results obtained by [10] and they are not reported again. However, all these methods point out to the suitability of the linear mixed Poisson model even though some non-stationarity is quite evident.

Appendix

Proof of Proposition 3.1. Recall that $\hat{\sigma}_Z^2$ is defined by the solution of (2.14) and is consistent. Consider first the case where the data are generated from the Poisson distribution and recall again that $\tilde{\boldsymbol{\theta}}_n = (\tilde{\boldsymbol{\theta}}_n^{(1)}, \tilde{\boldsymbol{\theta}}_n^{(2)})$ is the consistent estimator of $\boldsymbol{\theta} = (\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)})$ under the null hypothesis. Since $\tilde{\boldsymbol{\theta}}_n^{(1)}$ is a consistent estimator of $\boldsymbol{\theta}^{(1)}$ and $\boldsymbol{\theta}^{(1)} > \mathbf{0}$, then for n large enough we have that $\tilde{\boldsymbol{\theta}}_n^{(1)} > \mathbf{0}$ and $S_{n,i}(\tilde{\boldsymbol{\theta}}_n) := \partial l_n(\tilde{\boldsymbol{\theta}}_n)/\partial\theta_i = 0, \quad \forall\ i = 1, \ldots, m_1$. That is, $\tilde{\boldsymbol{\theta}}_n^{(1)} > \mathbf{0}$ and $\boldsymbol{S}_n^{(1)}(\tilde{\boldsymbol{\theta}}_n) = \mathbf{0}$.

If we define the matrices $\boldsymbol{K} = (\boldsymbol{O}_{m_2 \times m_1}, \boldsymbol{I}_{m_2})$ and $\widetilde{\boldsymbol{K}} = (\boldsymbol{I}_{m_1}, \boldsymbol{O}_{m_1 \times m_2})$, we have that

$$\boldsymbol{S}_n^{(1)}(\tilde{\boldsymbol{\theta}}_n) = \widetilde{\boldsymbol{K}}\boldsymbol{S}_n(\tilde{\boldsymbol{\theta}}_n) = \mathbf{0} \quad \text{and} \quad \boldsymbol{S}_n(\tilde{\boldsymbol{\theta}}_n) = \boldsymbol{K}'\boldsymbol{S}_n^{(2)}(\tilde{\boldsymbol{\theta}}_n). \tag{A.1}$$

Since

$$\frac{1}{n}\boldsymbol{H}_n(\boldsymbol{\theta}_0) = -\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} \xrightarrow{p} \boldsymbol{G}(\boldsymbol{\theta}_0),$$

a Taylor expansion shows that

$$\boldsymbol{S}_n(\tilde{\boldsymbol{\theta}}_n) \overset{o_p(1)}{=} \boldsymbol{S}_n(\boldsymbol{\theta}_0) - \boldsymbol{G}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0).$$

Therefore, the last m_2 components of the above relation give

$$\boldsymbol{S}_n^{(2)}(\tilde{\boldsymbol{\theta}}_n) \overset{o_p(1)}{=} \boldsymbol{S}_n^{(2)}(\boldsymbol{\theta}_0) - \boldsymbol{K}\boldsymbol{G}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0), \tag{A.2}$$

and the first m_1 components yield

$$\mathbf{0} = \boldsymbol{S}_n^{(1)}(\tilde{\boldsymbol{\theta}}_n) \overset{o_p(1)}{=} \boldsymbol{S}_n^{(1)}(\boldsymbol{\theta}_0) - \widetilde{\boldsymbol{K}}\boldsymbol{G}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0).$$

In addition,

$$(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = \widetilde{\boldsymbol{K}}'(\tilde{\boldsymbol{\theta}}_n^{(1)} - \boldsymbol{\theta}_0^{(1)}). \tag{A.3}$$

Thus, we have that

$$\mathbf{0} \overset{o_p(1)}{=} \boldsymbol{S}_n^{(1)}(\boldsymbol{\theta}_0) - \widetilde{\boldsymbol{K}}\boldsymbol{G}\widetilde{\boldsymbol{K}}'(\tilde{\boldsymbol{\theta}}_n^{(1)} - \boldsymbol{\theta}_0^{(1)}),$$

or

$$(\tilde{\boldsymbol{\theta}}_n^{(1)} - \boldsymbol{\theta}_0^{(1)}) \overset{o_p(1)}{=} (\widetilde{\boldsymbol{K}}\boldsymbol{G}\widetilde{\boldsymbol{K}}')^{-1}\boldsymbol{S}_n^{(1)}(\boldsymbol{\theta}_0). \tag{A.4}$$

Substituting (A.1), (A.2) and (A.3) in the general expression of the score test, we have that

$$\begin{aligned}
LM_n &= \boldsymbol{S}_n'(\tilde{\boldsymbol{\theta}}_n)\boldsymbol{I}^{-1}(\tilde{\boldsymbol{\theta}}_n)\boldsymbol{S}_n(\tilde{\boldsymbol{\theta}}_n)\\
&= \boldsymbol{S}_n'(\tilde{\boldsymbol{\theta}}_n)\boldsymbol{G}^{-1}(\tilde{\boldsymbol{\theta}}_n)\boldsymbol{S}_n(\tilde{\boldsymbol{\theta}}_n)\\
&= \boldsymbol{S}_n^{(2)'}(\tilde{\boldsymbol{\theta}}_n)\boldsymbol{K}\boldsymbol{G}^{-1}\boldsymbol{K}'\boldsymbol{S}_n^{(2)}(\tilde{\boldsymbol{\theta}}_n)\\
&\overset{o_p(1)}{=} (\boldsymbol{S}_n^{(2)}(\boldsymbol{\theta}_0) - \boldsymbol{K}\boldsymbol{G}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0))'\boldsymbol{K}\boldsymbol{G}^{-1}\boldsymbol{K}'(\boldsymbol{S}_n^{(2)}(\boldsymbol{\theta}_0) - \boldsymbol{K}\boldsymbol{G}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0))\\
&= (\boldsymbol{S}_n^{(2)}(\boldsymbol{\theta}_0) - \boldsymbol{K}\boldsymbol{G}\widetilde{\boldsymbol{K}}'(\tilde{\boldsymbol{\theta}}_n^{(1)} - \boldsymbol{\theta}_0^{(1)}))'\boldsymbol{K}\boldsymbol{G}^{-1}\boldsymbol{K}'(\boldsymbol{S}_n^{(2)}(\boldsymbol{\theta}_0) - \boldsymbol{K}\boldsymbol{G}\widetilde{\boldsymbol{K}}'(\tilde{\boldsymbol{\theta}}_n^{(1)} - \boldsymbol{\theta}_0^{(1)})).
\end{aligned}$$

Let

$$\boldsymbol{W} = \begin{pmatrix}\boldsymbol{W}_1\\ \boldsymbol{W}_2\end{pmatrix} = \frac{1}{\sqrt{n}}\begin{pmatrix}\boldsymbol{S}_n^{(1)}(\boldsymbol{\theta}_0)\\ \boldsymbol{S}_n^{(2)}(\boldsymbol{\theta}_0)\end{pmatrix} \xrightarrow{D} \mathcal{N}(\mathbf{0}, \boldsymbol{G}) \equiv \mathcal{N}\Big(\mathbf{0}, \begin{pmatrix}\boldsymbol{G}_{11} & \boldsymbol{G}_{12}\\ \boldsymbol{G}_{21} & \boldsymbol{G}_{22}\end{pmatrix}\Big).$$

It holds that $\boldsymbol{K}\boldsymbol{G}\widetilde{\boldsymbol{K}}' = \boldsymbol{G}_{21}$, $\widetilde{\boldsymbol{K}}\boldsymbol{G}\boldsymbol{K}' = \boldsymbol{G}_{12}$, $\widetilde{\boldsymbol{K}}\boldsymbol{G}\widetilde{\boldsymbol{K}}' = \boldsymbol{G}_{11}$ and $\boldsymbol{K}\boldsymbol{G}^{-1}\boldsymbol{K}' = (\boldsymbol{G}_{22} - \boldsymbol{G}_{21}\boldsymbol{G}_{11}^{-1}\boldsymbol{G}_{12})^{-1}$, where the last equality comes from the inversion of the block matrix $\boldsymbol{G}$ and denotes the matrix $\widetilde{\boldsymbol{\Sigma}}_P^{-1}$.

Thus, the score statistic is

$$LM_n = (\boldsymbol{W}_2 - \boldsymbol{G}_{21}(\tilde{\boldsymbol{\theta}}_n^{(1)} - \boldsymbol{\theta}_0^{(1)}))'(\boldsymbol{G}_{22} - \boldsymbol{G}_{21}\boldsymbol{G}_{11}^{-1}\boldsymbol{G}_{12})^{-1}(\boldsymbol{W}_2 - \boldsymbol{G}_{21}(\tilde{\boldsymbol{\theta}}_n^{(1)} - \boldsymbol{\theta}_0^{(1)})),$$

and using (A.4) we finally have that

$$LM_n = (\boldsymbol{W}_2 - \boldsymbol{G}_{21}\boldsymbol{G}_{11}^{-1}\boldsymbol{W}_1)'\widetilde{\boldsymbol{\Sigma}}_P^{-1}(\boldsymbol{W}_2 - \boldsymbol{G}_{21}\boldsymbol{G}_{11}^{-1}\boldsymbol{W}_1). \tag{A.5}$$

Then $LM_n \xrightarrow{D} LM \overset{H_0}{\sim} \mathcal{X}_{m_2}^2$, because of the fact that

$$\boldsymbol{W}_2 - \boldsymbol{G}_{21}\boldsymbol{G}_{11}^{-1}\boldsymbol{W}_1 \overset{H_0}{\sim} \mathcal{N}(\mathbf{0}, \widetilde{\boldsymbol{\Sigma}}_P).$$

Consider now the misspecified model where the data are generated from the mixed Poisson model. Following [26] and [27, pp. 126], (see also [5, 36, 21]), the score statistic is given by

$$LM_n = \boldsymbol{S}_n'(\tilde{\boldsymbol{\theta}}_n)\boldsymbol{G}^{-1}\boldsymbol{K}'(\boldsymbol{K}\boldsymbol{G}^{-1}\boldsymbol{G}_1\boldsymbol{G}^{-1}\boldsymbol{K}')^{-1}\boldsymbol{K}\boldsymbol{G}^{-1}\boldsymbol{S}_n(\tilde{\boldsymbol{\theta}}_n)$$

$$= \quad \boldsymbol{S}_n^{(2)'}(\tilde{\boldsymbol{\theta}}_n)\boldsymbol{K}\boldsymbol{G}^{-1}\boldsymbol{K}'(\boldsymbol{K}\boldsymbol{G}^{-1}\boldsymbol{G}_1\boldsymbol{G}^{-1}\boldsymbol{K}')^{-1}\boldsymbol{K}\boldsymbol{G}^{-1}\boldsymbol{K}'\boldsymbol{S}_n^{(2)}(\tilde{\boldsymbol{\theta}}_n).$$

Some calculations yield that $\boldsymbol{K}\boldsymbol{G}^{-1}\boldsymbol{K}' = \boldsymbol{G}^{22} = (\boldsymbol{G}_{22} - \boldsymbol{G}_{21}\boldsymbol{G}_{11}^{-1}\boldsymbol{G}_{12})^{-1}$ and $(\boldsymbol{K}\boldsymbol{G}^{-1}\boldsymbol{G}_1\boldsymbol{G}^{-1}\boldsymbol{K}')^{-1} = (\boldsymbol{G}^{22})^{-1}\widetilde{\boldsymbol{\Sigma}}_{MP}^{-1}(\boldsymbol{G}^{22})^{-1}$, where $\widetilde{\boldsymbol{\Sigma}}_{MP}$ is given by (3.2), that is

$$\widetilde{\boldsymbol{\Sigma}}_{MP} = \boldsymbol{G}_{1,22} - \boldsymbol{G}_{21}\boldsymbol{G}_{11}^{-1}\boldsymbol{G}_{1,12} - \boldsymbol{G}_{1,21}\boldsymbol{G}_{11}^{-1}\boldsymbol{G}_{12} + \boldsymbol{G}_{21}\boldsymbol{G}_{11}^{-1}\boldsymbol{G}_{1,11}\boldsymbol{G}_{11}^{-1}\boldsymbol{G}_{12}$$

and thus, the score test statistic is given by

$$LM_n = \boldsymbol{S}_n^{(2)'}(\tilde{\boldsymbol{\theta}}_n)\widetilde{\boldsymbol{\Sigma}}_{MP}^{-1}\boldsymbol{S}_n^{(2)}(\tilde{\boldsymbol{\theta}}_n). \tag{A.6}$$

It can be shown that

$$\boldsymbol{S}_n^{(2)}(\tilde{\boldsymbol{\theta}}_n) \overset{o_p(1)}{=} \boldsymbol{W}_2 - \boldsymbol{G}_{21}\boldsymbol{G}_{11}^{-1}\boldsymbol{W}_1 \overset{H_0}{\sim} \mathcal{N}(\boldsymbol{0}, \widetilde{\boldsymbol{\Sigma}}_{MP}).$$

and therefore, $LM_n \xrightarrow{D} LM \overset{H_0}{\sim} \mathcal{X}_{m_2}^2$.

Assume now that the data are generated again from a Poisson model under the local Pitman-type alternatives $H_1 : \boldsymbol{\theta}^{(2)} = \boldsymbol{\theta}_0^{(2)} + n^{-1/2}\boldsymbol{\delta}$. In order to obtain the limiting distribution of the score test statistic (3.1) under the alternative, we only need the asymptotic distribution of the score function under H_1. By a Taylor expansion of $\boldsymbol{S}_n(\boldsymbol{\theta}_0)$ about $\boldsymbol{\theta}_n = \boldsymbol{\theta}_0 + n^{-1/2}\boldsymbol{\delta}^*$, where now $\boldsymbol{\delta}^*$ is a fixed vector in $\mathbb{R}_+^m$ of the form $\boldsymbol{\delta}^* = (\boldsymbol{\delta_1}, \boldsymbol{\delta})$, we have that

$$n^{-1/2}\boldsymbol{S}_n(\boldsymbol{\theta}_0) \overset{o_p(1)}{=} n^{-1/2}\boldsymbol{S}_n(\boldsymbol{\theta}_n) - n^{-1}\boldsymbol{H}_n(\boldsymbol{\theta}_n)\boldsymbol{\delta}^*.$$

Since $n^{-1/2}\boldsymbol{S}_n(\boldsymbol{\theta}_n) \xrightarrow{D} \mathcal{N}(\boldsymbol{0}, \boldsymbol{G}(\boldsymbol{\theta}_0))$ and $n^{-1}\boldsymbol{H}_n(\boldsymbol{\theta}_n)\boldsymbol{\delta}^* \xrightarrow{p} \boldsymbol{G}(\boldsymbol{\theta}_0)\boldsymbol{\delta}^*$ under the alternative, then $\boldsymbol{W} = n^{-1/2}\boldsymbol{S}_n(\boldsymbol{\theta}_0) \xrightarrow{D} \mathcal{N}(-\boldsymbol{G}(\boldsymbol{\theta}_0)\boldsymbol{\delta}^*, \boldsymbol{G}(\boldsymbol{\theta}_0))$. Therefore,

$$\boldsymbol{S}_n^{(2)}(\tilde{\boldsymbol{\theta}}_n) \overset{o_p(1)}{=} \boldsymbol{W}_2 - \boldsymbol{G}_{21}\boldsymbol{G}_{11}^{-1}\boldsymbol{W}_1 \overset{H_1}{\sim} \mathcal{N}(-\widetilde{\boldsymbol{\Sigma}}_P\boldsymbol{\delta}, \widetilde{\boldsymbol{\Sigma}}_P),$$

and considering again the expression of the score statistic given by (A.5), we have that $LM_n \xrightarrow{D} LM \overset{H_1}{\sim} \mathcal{X}_{m_2}^2(\boldsymbol{\delta}'\boldsymbol{\Delta}\boldsymbol{\delta})$, where for the case of the Poisson assumption $\boldsymbol{\Delta} = \widetilde{\boldsymbol{\Sigma}}_P = \boldsymbol{G}_{22} - \boldsymbol{G}_{21}\boldsymbol{G}_{11}^{-1}\boldsymbol{G}_{12}$ evaluated at $\tilde{\boldsymbol{\theta}}_n$.

If the data are generated from the mixed Poisson model, then following the same steps as for the Poisson and since now it holds that $n^{-1/2}\boldsymbol{S}_n(\boldsymbol{\theta}_n) \xrightarrow{D} \mathcal{N}(\boldsymbol{0}, \boldsymbol{G}_1(\boldsymbol{\theta}_0))$ and $n^{-1}\boldsymbol{H}_n(\boldsymbol{\theta}_n)\boldsymbol{\delta}^* \xrightarrow{p} \boldsymbol{G}(\boldsymbol{\theta}_0)\boldsymbol{\delta}^*$ under the alternative, we have that

$$\boldsymbol{W} = n^{-1/2}\boldsymbol{S}_n(\boldsymbol{\theta}_0) \xrightarrow{D} \mathcal{N}(-\boldsymbol{G}(\boldsymbol{\theta}_0)\boldsymbol{\delta}^*, \boldsymbol{G}_1(\boldsymbol{\theta}_0)).$$

Thus,

$$\boldsymbol{S}_n^{(2)}(\tilde{\boldsymbol{\theta}}_n) \overset{o_p(1)}{=} \boldsymbol{W}_2 - \boldsymbol{G}_{21}\boldsymbol{G}_{11}^{-1}\boldsymbol{W}_1 \overset{H_1}{\sim} \mathcal{N}(-\widetilde{\boldsymbol{\Sigma}}_P\boldsymbol{\delta}, \widetilde{\boldsymbol{\Sigma}}_{MP})$$

and considering again the expression of the score statistic given by (A.6), we have that $LM_n \xrightarrow{D} LM \overset{H_1}{\sim} \mathcal{X}_{m_2}^2(\boldsymbol{\delta}'\boldsymbol{\Delta}\boldsymbol{\delta})$, where for this case $\boldsymbol{\Delta} = \widetilde{\boldsymbol{\Sigma}}_P'\widetilde{\boldsymbol{\Sigma}}_{MP}^{-1}\widetilde{\boldsymbol{\Sigma}}_P$ evaluated at $\tilde{\boldsymbol{\theta}}_n$. □

References

[1] Andrews, D. W. K. (2001). Testing when a parameter is on the boundary of the maintained hypothesis, *Econometrica* **69**, 683–734. MR1828540

[2] Andrews, D. W. K. and Ploberger, W. (1994). Optimal tests when a nuisance parameter is present only under the alternative, *Econometrica* **62**, 1383–1414. MR1303238

[3] Bardet, J.-M. and Wintenberger, O. (2009). Asymptotic normality of the quasi-maximum likelihood estimator for multidimensional causal processes, *The Annals of Statistics* **37**, 2730–2759. MR2541445

[4] Berkes, I., Horváth, L. and Kokoszka, P. (2003). GARCH processes: Structure and estimation, *Bernoulli* **9**, 201–227. MR1997027

[5] Boos, D. D. (1992). On generalized score tests, *The American Statistician* **46**, 327–333.

[6] Breslow, N. (1990). Tests of hypotheses in overdispersed Poisson regression and other quasi-likelihood models, *Journal of the American Statistical Association* **85**, 565–571.

[7] Cameron, A. C. and Trivedi, P. K. (1998). *Regression Analysis of Count Data*, 1st ed, Cambridge University Press, Cambridge. MR1648274

[8] Chan, K. S. (1993). Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model, *The Annals of Statistics* **21**, 520–533. MR1212191

[9] Christou, V. and Fokianos, K. (2014). Quasi–likelihood inference for negative binomial time series models, *Journal of Time Series Analysis* **35**, 55–78. MR3148248

[10] Christou, V. and Fokianos, K. (2015). On count time series prediction, *Journal of Statistical Computation and Simulation* **85**, 357–373. MR3270681

[11] Davies, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative, *Biometrika* **74**, 33–43. MR0885917

[12] Davis, R. A., Dunsmuir, W. T. M. and Streett, S. B. (2003). Observation-driven models for Poisson counts, *Biometrika* **90**, 777–790. MR2024757

[13] Davis, R. A. and Liu, H. (2015). Theory and inference for a class of observation-driven models with application to time series of counts, *Statistica Sinica*. to appear.

[14] Dean, C. B., Eaves, D. M. and Martinez, C. J. (1995). A comment on the use of empirical covariance matrices in the analysis of count data, *Journal of Statistical Planning and Inference* **48**, 197–205. MR1366790

[15] Dedecker, J., Doukhan, P., Lang, G., León, J. R., Louhichi, S. and Prieur, C. (2007). *Weak Dependence: With Examples and Applications*, Vol. 190 of *Lecture Notes in Statistics*, Springer, New York. MR2338725

[16] Dedecker, J. and Prieur, C. (2007). An empirical central limit theorem for dependent sequences, *Stochastic Process. Appl.* **117**, 121–142. MR2287106

[17] Dehling, H., Durieu, O. and Tusche, M. (2014). Approximating class approach for empirical processes of dependent sequences indexed by functions, *Bernoulli* **20**, 1372–1403. MR3217447

[18] Douc, R., Doukhan, P. and Moulines, E. (2013). Ergodicity of observation-driven time series models and consistency of the maximum likelihood estimator, *Stochastic Processes and their Applications* **123**, 2620–2647. MR3054539

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

[20] Doukhan, P. and Wintenberger, O. (2008). Weakly dependent chains with infinite memory, *Stochastic Processes and Their Applications* **118**, 1997–2013. MR2462284

[21] Engle, R. F. (1984). *Wald, Likelihood Ratio, and Lagrange Multiplier Tests in Econometrics*, Vol. 2 of *Handbook of Econometrics*, Elsevier, Amsterdam: North Holland, 775–826.

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

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

[24] Francq, C., Horvath, L. and Zakoïan, J.-M. (2010). Sup-tests for linearity in a general nonlinear AR(1) model, *Econometric Theory* **26**, 965–993. MR2660290

[25] Francq, C. and Zakoïan, J.-M. (2004). Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes, *Bernoulli* **10**, 605–637. MR2076065

[26] Francq, C. and Zakoïan, J.-M. (2010). *GARCH Models: Structure, Statistical Inference and Financial Applications*, John Wiley, United Kingdom. MR3186556

[27] Gallant, A. R. (1987). *Nonlinear statistical models*, Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics, John Wiley, New York. MR0921029

[28] Godambe, V. P. and Heyde, C. C. (1987). Quasi–likelihood and optimal estimation, *International Statistical Review* **55**, 231–244. MR0963141

[29] Harvey, A. C. (1990). *The Econometric Analysis of Time Series*, 2nd ed, MIT Press, Cambridge, MA. MR1085719

[30] Heyde, C. C. (1997). *Quasi-Likelihood and its Applications: A General Approach to Optimal Parameter Estimation*, Springer, New York. MR1461808

[31] Joe, H. and Zhu, R. (2005). Generalized Poisson distribution: The property of mixture of Poisson and comparison with negative binomial distribution, *Biometrical Journal* **47**, 219–229. MR2137236

[32] Johnson, N. L., Kotz, S. and Kemp, A. W. (1992). *Univariate Discrete Distributions*, 2nd ed, Wiley, New York. MR1224449

[33] Jørgensen, B. (1997). *The Theory of Dispersion Models*, Chapman & Hall, London. MR1462891
[34] Jung, R. and Tremayne, A. (2011). Useful models for time series of counts or simply wrong ones?, *AStA Advances in Statistical Analysis* **95**, 59–91. MR2775582
[35] Kedem, B. and Fokianos, K. (2002). *Regression Models for Time Series Analysis*, John Wiley, Hoboken, NJ. MR1933755
[36] Kent, J. T. (1982). Robust properties of likelihood ratio tests, *Biometrika* **69**, 19–27. MR0655667
[37] Kokonendji, C. C., Dossou-Gbété, S. and Demétrio, C. G. B. (2004). Some discrete exponential dispersion models: Poisson-Tweedie and Hinde-Demétrio classes, *Statistics and Operations Research Transactions* **28**, 201–213. MR2116192
[38] Lawless, J. F. (1987). Negative binomial and mixed Poisson regression, *The Canadian Journal of Statistics* **15**, 209–225. MR0926553
[39] Li, G. and Li, W. K. (2011). Testing a linear time series model against its threshold extension, *Biometrika* **98**, 243–250. MR2804225
[40] Luukkonen, R., Saikkonen, P. and Teräsvirta, T. (1988). Testing linearity against smooth transition autoregressive models, *Biometrika* **75**, 491–499. MR0967588
[41] Mikosch, T. (2009). *Non-life Insurance Mathematics, An Introduction with the Poisson Process*, 2nd ed, Springer-Verlag, Berlin. MR2503328
[42] Mikosch, T. and Straumann, D. (2006). Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: A stochastic recurrence equations approach, *The Annals of Statistics* **34**, 2449–2495. MR2291507
[43] Neumann, M. (2011). Absolute regularity and ergodicity of Poisson count processes, *Bernoulli* **17**, 1268–1284. MR2854772
[44] Teräsvirta, T. (1994). Specification, estimation, and evaluation of smooth transition autoregressive models, *Journal of the American Statistical Association* **89**, 208–218.
[45] Tong, H. (1990). *Nonlinear Time Series: A Dynamical System Approach*, Oxford University Press, New York. MR1079320
[46] Wang, C., Liu, H., Yao, J.-F., Davis, R. A. and Li, W. K. (2014). Self-excited threshold Poisson autoregression, *Journal of the American Statistical Association* **109**, 777–787. MR3223749
[47] 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
[48] Zeger, S. L. and Qaqish, B. (1988). Markov regression models for time series: A quasi-likelihood approach, *Biometrics* **44**, 1019–1031. MR0980997
[49] Zhu, F. (2011). A negative binomial integer-valued GARCH model, *Journal of Time Series Analysis* **32**, 54–67. MR2790672
[50] Zhu, F. (2012*a*). Modeling overdispersed or underdispersed count data with generalized Poisson integer-valued GARCH models, *Journal of Mathematical Analysis and Applications* **389**, 58–71. MR2876481

[51] Zhu, F. (2012*b*). Zero-inflated Poisson and negative binomial integer-valued GARCH models, *J. Statist. Plann. Inference* **142**, 826–839. MR2863870

[52] Zivot, E. and Wang, J. (2006). *Modeling Financial Time Series with S-Plus®*, 2nd ed, Springer-Verlag, New York.