Bootstrapping Network Autoregressive Models for Testing Linearity

Mirko Armillotta, Konstantinos Fokianos, and Ioannis Krikidis

Abstract We develop methodology for network data with special attention to epidemic network spatio-temporal structures. We provide estimation methodology for linear network autoregressive models for both continuous and count multivariate time series. A study of non-linear models for inference under the assumption of known network structure is provided. We propose a family of test statistics for testing linearity of the imposed model. In particular, we compare empirically two bootstrap versions of a supremum-type quasi-score test. Synthetic data are employed to demonstrate the validity of the methodological results. Finally, an epidemic application of the proposed methodology to daily COVID-19 cases detected on province-level geographical network in Italy complements the work.

1 Modelling Network Time Series

New sources of data like social networks, GPS data, or epidemic counting processes, usually recorded over a timespan and a specific geographical area, has motivated a lot of interest in network data modelling. In particular, understanding the effect of a network to a multivariate time series is of essential importance for many applications and has attracted considerable recent attention. The methodology outlined in this work has potential application in several network science related fields.

Knight et al. [30] defined such multivariate streaming data as network time series and proposed a methodology for modelling them. This approach has been originally

M. Armillotta (✉) · K. Fokianos
Department of Mathematics and Statistics, University of Cyprus, P.O. Box 20537, 1678 Nicosia, Cyprus
e-mail: armillotta.mirko@ucy.ac.cy

K. Fokianos
e-mail: fokianos@ucy.ac.cy

I. Krikidis
Department of Electrical and Computer Engineering, University of Cyprus,
P.O. Box 20537, 1678 Nicosia, Cyprus

G. Dzemyda et al. (eds.), *Data Science in Applications*,
Studies in Computational Intelligence 1084,
https://doi.org/10.1007/978-3-031-24453-7_6

proposed in the context of spatio-temporal data analysis and is referred to Space-Time Autoregressive Moving Average (STARMA) models. See Cliff and Ord [9] and Martin and Oeppen [36], among many others. Indeed, a wide variety of available spatial streaming data related to physical phenomena fits this framework. In general, any stream of data for a sample of units whose relations can be modelled through an adjacency matrix (neighborhood structure), adhere to statistical techniques reviewed in this work.

We review some recent literature for network time series. Zhu et al. [54] developed inferential theory for Network Autoregressive models (NAR) when the network dimension N is increasing ($N \to \infty$), under the Independent Identic Distributed (IID) assumption on the innovation error sequence, where a continuous response random variable is observed for each node of a network. Technically speaking, in this approach the observed variable Y, for the node i at time t, is denoted by $Y_{i,t}$. To understand its behavior, as it evolves in time, it is assumed to depend on the past value of the variable for the node itself, say $Y_{i,t-1}$, and of the past values of the average variable between its neighbors, i.e. the mean of the variable Y, at time $t-1$ observed among the nodes connected to the node i. These authors develop Ordinary Least Squares (OLS) inference and study the asymptotic behaviour of the related estimator. Further extensions of network autoregressive models consider quantile autoregression [55], grouped least squares estimation [53], as well as a network extension for GARCH models [52]. The latter has been considered only for the case of fixed network dimension. Finally, Knight et al. [31] studied the more elaborate neighbourhood structures of STARMA models in the context of network analysis, named as Generalized NAR (GNAR), which considers the effect of several layers of connections between the nodes of the network and provide R software for fitting such models, for continuous variables only.

1.1 The Case of Discrete Responses

Interesting datasets collected from social network analysis have integer-valued nature, e.g. number of characters contained in users posts, number of likes, etc. However, the literature on models for multivariate count time series is sparse; see Fokianos [17] for a recent review. To fill this gap, Armillotta and Fokianos [3] proposed a linear and a log-linear Poisson network autoregression model (PNAR) for Poisson distributed data, under the assumption of α-mixing innovations. For details about and weak dependence related literature see Rosenblatt [43] and Doukhan [15]. This model generalizes the linear NAR model, by linking it with the context of Generalized Linear Models [37], since the observations are marginally Poisson distributed, conditionally to their past history. The joint dependence among different variables is specified by a copula construction, see Fokianos et al. [21, Sect. 2]. Armillotta and Fokianos [3] have further established parametric estimation under the framework of quasi maximum likelihood inference [26, 50] and associated asymptotic theory when

the network dimension increases. Bracher and Held [6] study the related problem from a Bayesian point of view.

1.2 Nonlinear Models

All previous contributions assume linearity of the model, which is restrictive assumption in practice. Literature for univariate nonlinear time series models is well established; this is especially true for continuous-valued variables. The interested reader can see Tong [46], Fan and Yao [16], Gao [23] and Teräsvirta et al. [45], among many others, for more details. For integer-valued data there exists a more recent stream of works, although still under development. Suitable smoothing conditions for inference on nonlinear models are provided by Fokianos et al. [20], Neumann [38] with Poisson data, Christou and Fokianos [7] for the Negative Binomial case, and Gorgi [25] for the Beta Negative Binomial distribution. See also Wang et al. [47] for a threshold autoregressive model with Poisson distribution. In a more general framework, related works are by Ahmad and Francq [1], Davis and Liu [12] and Douc et al. [14], among others. For a recent review see Davis et al. [13]. Despite this flourishing literature related to nonlinear models, the previous works are not directly applicable to network autoregressive models, because of their multivariate structure. Multivariate models for discrete observations include the work by Pedeli and Karlis [40–42] and Fokianos et al. [21], among others, who consider linear models. Armillotta and Fokianos [4] specified a general nonlinear network autoregressive model for both continuous and discrete-valued processes, establishing also the related stationarity results and asymptotic theory of suitable quasi maximum likelihood estimators.

1.3 Testing for Linearity

Testing the linearity of a given model is a classical subject of study in time series analysis and econometrics. For continuous-valued random variables, general results have been reported when the parameters are identifiable or non-identifiable under the null hypothesis; see Boos [5] for the former and Francq et al. [22] for the latter case. Other linearity tests for specific nonlinear models and with non identifiable parameters, have been specified in Luukkonen et al. [35], for the Smooth Transition Autoregression (STAR) case, Li and Li [33], for the Threshold Autoregression (TAR) model, among others. For discrete-valued time series, Christou and Fokianos [8] suggest a score type test for univariate (mixed) Poisson random variables, in the case of correctly identifiable parameters. Finally, Andrews and Ploberger [2] and Hansen [28] proposed general methods for testing linearity under non-identifiability for univariate models. Non parametric tests have been also proposed; see, for example, Gao et al. [24] and Fokianos and Neumann [18], for continuous and count data, respectively. However, these latter test become computationally intensive when considering

multivariate time series models. Armillotta and Fokianos [4] proposed testing procedures for examining linearity (or nonlinearity) of NAR models, for both continuous and count data, with and without the presence of non identifiable parameters under the null hypothesis.

1.4 Outline

The main aim of the work is to compare different bootstrap methods for testing linearity of NAR models. Such comparison will be conducted with the use of simulated synthetic data as well as by an application to real world data.

The paper is organized as follows: Sect. 2 introduces general nonlinear frameworks for network time series autoregressive models, for continuous and count processes and also discusses specific models of interest. Details about the inference to unknown parameters of the model are also provided. Then, in Sect. 3, results concerning the quasi-score test for testing linearity in network autoregressive models are discussed. The testing methodology is analyzed with and without non identifiable parameters under the null assumption. Practical computational aspects are taken into account, by describing different ways to compute the p-values of the proposed test statistics, by feasible bounds and bootstrap methodologies. Section 4 presents the results obtained on simulated data regarding the comparison between different computations of the linearity test. Finally, the proposed methodology is also applied to a real data analysis on epidemic networks to daily new COVID-19 cases observed on province-level geographical network in Italy.

Notation

For a $q \times p$ matrix $M = (m_{ij})$, $i = 1, \ldots, q$, $j = 1, \ldots, p$, denotes the generalized matrix norm $|||M|||_r = \max_{|x|_r=1} |Mx|_r$. If r=1, then $|||M|||_1 = \max_{1\le j\le p} \sum_{i=1}^{q} |m_{ij}|$. If $r = 2$, $|||M|||_2 = \rho^{1/2}(M'M)$, where $\rho(\cdot)$ is the spectral radius. If $r = \infty$, $|||M|||_\infty = \max_{1\le i\le q} \sum_{j=1}^{p} |m_{ij}|$. If $q = p$, these norms are matrix norms. The symbol I denotes an identity matrix, 1 a vector of ones, 0 a vector of zeros, whose dimensions depend on the context in which they are applied.

2 Network Autoregressive Models

When a network with N nodes, indexed by $i = 1, \ldots N$ is a priori known to the researcher, the neighbourhood structure of such a network is completely described by using its adjacency matrix $A = (a_{ij}) \in \mathbb{R}^{N\times N}$. The single element of such matrix would be $a_{ij} = 1$, if there is a directed edge from i to j (e.g. user i follows j on Twitter,

a flight take off from airport i landing to airport j), and $a_{ij} = 0$ otherwise. Undirected graphs are allowed ($A = A'$), which means that the edge between two nodes, i and j, has no specific direction. Typically, self-relationships are excluded i.e. $a_{ii} = 0$ for any $i = 1, \ldots, N$. This is a restriction for many applications, such as social networks; see Wasserman et al. [49] and Kolaczyk and Csárdi [32], for more details on network definitions. Since the information on the network is assumed to be known in advance, the network structure is treated as a known component of the analysis. The row-normalised adjacency matrix is defined by $W = \text{diag}\{n_1, \ldots, n_N\}^{-1} A$ where $n_i = \sum_{j=1}^{N} a_{ij}$ is the total number of connections starting from the node i, such that $i \to j$; it is called out-degree. Then, W is constructed with the property $|||W|||_\infty = 1$. Moreover, define e_i the N-dimensional unit vector with 1 in the ith position and 0 everywhere else, such that $w_i = a_{ij}/n_i = (e_i' W)' = (w_{i1} \ldots, w_{iN})'$ is the vector containing the ith row of W.

Define a N-dimensional vector of time series $\{Y_t,\ t = 1, 2 \ldots, T\}$, where $Y_t = (Y_{1,t}, \ldots, Y_{i,t}, \ldots Y_{N,t})'$, which is observed on the given network; in this way, a univariate time series is detected for each node, say $Y_{i,t}$, with corresponding conditional expectation $\lambda_{i,t}$, denoted by $\{\lambda_t \equiv \mathrm{E}(Y_t|\mathcal{F}_{t-1}),\ t = 1, 2 \ldots, T\}$, with $\lambda_t = (\lambda_{1,t}, \ldots, \lambda_{i,t}, \ldots, \lambda_{N,t})'$ being the conditional expectation vector, and denote the history of the process by $\mathcal{F}_t = \sigma(Y_s : s \le t)$. When the stochastic process $\{Y_t : t \in \mathbb{Z}\}$ is integer-valued, the first lag order nonlinear Poisson Network Autoregression (PNAR) is generally specified as follow [4]

$$Y_t = N_t(\lambda_t), \qquad \lambda_t = f(Y_{t-1}, W, \theta) \tag{1}$$

where $f(\cdot)$ is a function depending on the past lags of the count random vector, the known network structure W, and an m-dimensional parameter vector θ. The process $\{N_t\}$ is a sequence of N-variate copula-Poisson processes describing the joint dependence structure of the time series vector Y_t, where the marginal probability distribution of the count variables is $Y_{i,t}|\mathcal{F}_{t-1} \sim Poisson(\lambda_{i,t})$, for $i = 1, \ldots, N$. The joint distribution between univariate variables is generated by a copula structure, say $C(\cdot, \rho)$, on waiting times of a Poisson process, defined by Armillotta and Fokianos [3, Sect. 2.1]. An extension of (1) for a general lag order $p > 1$ models is given by, see Armillotta and Fokianos [4]

$$\lambda_t = f(Y_{t-1}, \ldots, Y_{t-p}, W, \theta)\,.$$

When the time series are continuous-valued, the nonlinear Network Autoregression (NAR) is defined by Armillotta and Fokianos [4] such that

$$Y_t = \lambda_t + \xi_t, \qquad \lambda_t = f(Y_{t-1}, W, \theta) \tag{2}$$

where $\xi_{i,t} \sim IID(0, \sigma^2)$, for $1 \le i \le N$ and $1 \le t \le T$. Obviously, we can extend (2) by incorporating a larger number of lags.

Models (1)–(2) have been proved to be stationary under suitable smoothness conditions on the function $f(\cdot)$ which are easily verifiable. See Armillotta and Fokianos [4, Sects. 2.2–2.3] for details about stability conditions.

Denote by $X_{i,t} = n_i^{-1} \sum_{j=1}^{N} a_{ij} Y_{j,t}$ the so called network effect; it represents the average impact of node i's connections. Recall models (1)–(2). The parameter vector can be split in two parts $\theta = (\theta^{(1)\prime}, \theta^{(2)\prime})'$, where the vectors $\theta^{(1)}$ and $\theta^{(2)}$ are of dimension m_1 and m_2, respectively, such that $m_1 + m_2 = m$. In general, $\theta^{(1)}$ denotes parameters associated with the linear part the model, whereas $\theta^{(2)}$ denotes the vector of nonlinear parameters. For $t = 1 \ldots, T$, both (1)–(2) have element-wise components

$$\lambda_{i,t} = f_i(X_{i,t-1}, Y_{i,t-1}; \theta^{(1)}, \theta^{(2)}), \quad i = 1, \ldots, N, \tag{3}$$

where $f_i(\cdot)$ is defined as the ith component of the function $f(\cdot)$, and it ultimately depends on the specific nonlinear model of interest which is taken into account.

2.1 Examples of Specific Models of Interest

We give some illustrative examples of specific nonlinear models of (3). We first introduce the linear model as a special case.

Linear Model

Recall that $X_{i,t} = n_i^{-1} \sum_{j=1}^{N} a_{ij} Y_{j,t}$ is the neighbourhood mean. The first order linear NAR(1) model,

$$\lambda_{i,t} = \beta_0 + \beta_1 X_{i,t-1} + \beta_2 Y_{i,t-1}, \tag{4}$$

is a special case of (3), with $\theta^{(1)} = (\beta_0, \beta_1, \beta_2)'$, but without nonlinear parameters $\theta^{(2)}$. For each single node i, model (4) allows the conditional mean of the process to depend on the past of the variable itself, for the same node i, and the average of the other nodes $j \neq i$ by which the focal node i is connected. Implicitly, only the nodes connected with the node i can affect its conditional mean $\lambda_{i,t}$. The parameter β_1 measures the impact of the network effect $X_{i,t-1}$. The coefficient β_2 determines the impact of the lagged variable $Y_{i,t-1}$. Model (4) was originally introduced by Knight et al. [30] and Zhu et al. [54] for the case of continuous random variables Y_t, with $Y_{i,t} = \lambda_{i,t} + \xi_{i,t}$. Armillotta and Fokianos [3] extended (4) to count random variables. In this case, (4) is the linear PNAR(1) model with $Y_{i,t}|\mathcal{F}_{t-1} \sim Poisson(\lambda_{i,t})$ for $i = 1, \ldots, N$ and a copula structure for joint distribution.

Intercept Drift (ID)

When Y_t is integer-valued, a drift in the intercept term of (4) introduces the nonlinear model

$$\lambda_{i,t} = \frac{\beta_0}{(1 + X_{i,t-1})^{\gamma}} + \beta_1 X_{i,t-1} + \beta_2 Y_{i,t-1}\,, \tag{5}$$

where $\gamma \geq 0$. Model (5) behaves like a linear model for small values of γ, and $\gamma = 0$ reduces (5) to (4) exactly. Instead, when γ takes values far from zero, model (5) introduce a perturbation, deviating from the linear model (4). Hence, (5) is a special case of (3), with $\theta^{(1)} = (\beta_0, \beta_1, \beta_2)'$ and $\theta_0^{(2)} = \gamma$. A slightly modified version of (5) allows to treat the case where $Y_t \in \mathbb{R}^N$, by taking the absolute value of $X_{i,t-1}$ defined at the denominator of the intercept term.

Smooth Transition (STNAR)

A Smooth Transition version of the NAR model, say STNAR(1), is specified as

$$\lambda_{i,t} = \beta_0 + (\beta_1 + \alpha \exp(-\gamma X_{i,t-1}^2)) X_{i,t-1} + \beta_2 Y_{i,t-1}\,, \tag{6}$$

where $\gamma \geq 0$. This models introduces a smooth regime switching behaviour on the network effect, by mimicking the smooth transition time series models suggested by Haggan and Ozaki [27], Teräsvirta [44] and Fokianos and Tjøstheim [19]. When $\alpha = 0$ in (6), the linear NAR model (4) is recovered. Moreover, (6) is a special case of (3), with $\theta^{(1)} = (\beta_0, \beta_1, \beta_2)'$ and $\theta_0^{(2)} = (\alpha, \gamma)'$.

Threshold Effect (TNAR)

Another regime switching nonlinear time series model of particular interest is Threshold NAR model, TNAR(1), defined by

$$\lambda_{i,t} = \beta_0 + \beta_1 X_{i,t-1} + \beta_2 Y_{i,t-1} + (\alpha_0 + \alpha_1 X_{i,t-1} + \alpha_2 Y_{i,t-1}) I(X_{i,t-1} \leq \gamma)\,, \tag{7}$$

where $I(\cdot)$ is the indicator function and γ is the threshold parameter. Unlike the STNAR model, (7) induces an abrupt shift in the parameters of the models. For details about threshold-type models, the reader is referred to Lim and Tong [34], Wang et al. [47] and Christou and Fokianos [8], among others. When $\alpha_0 = \alpha_1 = \alpha_2 = 0$, model (7) reduces to the linear counterpart (4). Clearly, $\theta^{(1)} = (\beta_0, \beta_1, \beta_2)'$ and $\theta^{(2)} = (\alpha_0, \alpha_1, \alpha_2, \gamma)'$ show that (7) is a special case of (3).

2.2 Inference

Estimation for the true unknown parameter vector θ_0 in models (1) is developed by means of quasi-maximum likelihood methodology, see Wedderburn [50], Gourieroux et al. [26] and Heyde [29], for example. The Quasi Maximum Likelihood Estimator (QMLE) is the vector of parameters $\hat{\theta}$ maximizing the function

$$l_T(\theta) = \sum_{t=1}^{T} \sum_{i=1}^{N} \left(Y_{i,t} \log \lambda_{i,t}(\theta) - \lambda_{i,t}(\theta) \right), \tag{8}$$

which is not necessarily the *true* log-likelihood of the process but it serves as an approximation. In particular, following Armillotta and Fokianos [4], (8) is the log-likelihood that it would have been obtained if all time series were contemporaneously independent. Note that although the joint copula structure $C(\ldots, \rho)$ and the corresponding set of parameters ρ are not included in the maximization of (8), the QMLE is still computed under the assumption of dependence as it is implicitly taken into account in the past values of multivariate counts Y_t. Maximizing (8) simplifies computations of the estimation and guarantees consistency and asymptotic normality of the resulting estimator. The derivative of (8) yields the score function

$$S_T(\theta) = \sum_{t=1}^{T} \sum_{i=1}^{N} \left(\frac{Y_{i,t}}{\lambda_{i,t}(\theta)} - 1 \right) \frac{\partial \lambda_{i,t}(\theta)}{\partial \theta} \equiv \sum_{t=1}^{T} s_t(\theta) \,. \tag{9}$$

Define $\partial \lambda_t(\theta)/\partial \theta'$ the $N \times m$ matrix of derivatives, $D_t(\theta)$ the $N \times N$ diagonal matrix with elements equal to $\lambda_{i,t}(\theta)$, for $i = 1, \ldots, N$ and $\xi_t(\theta) = Y_t - \lambda_t(\theta)$ is a Martingale Difference Sequence (MDS). Then, the empirical Hessian and conditional information matrices are given, respectively, by

$$H_T(\theta) = \sum_{t=1}^{T} \sum_{i=1}^{N} \frac{Y_{i,t}}{\lambda_{i,t}^2(\theta)} \frac{\partial \lambda_{i,t}(\theta)}{\partial \theta} \frac{\partial \lambda_{i,t}(\theta)}{\partial \theta'} - \sum_{t=1}^{T} \sum_{i=1}^{N} \left(\frac{Y_{i,t}}{\lambda_{i,t}(\theta)} - 1 \right) \frac{\partial^2 \lambda_{i,t}(\theta)}{\partial \theta \partial \theta'} \,,$$

$$B_T(\theta) = \sum_{t=1}^{T} \frac{\partial \lambda_t'(\theta)}{\partial \theta} D_t^{-1}(\theta) \Sigma_t(\theta) D_t^{-1}(\theta) \frac{\partial \lambda_t(\theta)}{\partial \theta'} \,,$$

where $\Sigma_t(\theta) = \mathrm{E}\left(\xi_t(\theta)\xi_t'(\theta) \mid \mathcal{F}_{t-1}\right)$ is the conditional covariance matrix evaluated at θ. Under suitable network assumptions and smoothness conditions on the nonlinear function $f(\cdot)$, Armillotta and Fokianos [4] proved the consistency and asymptotic normality of the estimator, that is $\sqrt{NT}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, H^{-1} B H^{-1})$, when $N \to \infty$ and $T_N \to \infty$, where H and B are the theoretical Hessian and information matrices, respectively, evaluated at the true value of the parameters $\theta = \theta_0$.

Analogous inferential result are obtained for model (2), by maximizing the quasi log-likelihood $l_{NT}(\theta) = -\sum_{t=1}^{T} (Y_t - \lambda_t(\theta))' (Y_t - \lambda_t(\theta))$, being equivalent to perform a nonlinear Least Squares (LS) estimation of the unknown parameters.

3 Linearity Test

In this section we introduce the linearity test for nonlinear networks autoregressive models (1)–(2), discussed in Sect. 2. Recall model (3) and consider the following hypothesis testing problems

$$H_0 : \theta^{(2)} = \theta_0^{(2)} \quad \text{versus} \quad H_1 : \theta^{(2)} \neq \theta_0^{(2)}, \quad \text{componentwise}\,, \tag{10}$$

where, under the null hypothesis H_0, the nonlinear parameters take a value $\theta_0^{(2)}$, which yields the linear model (4). For example, when Y_t is integer-valued following (1) and the mean process λ_t is defined as in (5), then $\theta^{(2)} = \gamma$ and $\theta_0^{(2)} = 0$. Indeed the problem $H_0 : \gamma = 0$ versus $H_1 : \gamma > 0$ becomes an hypothesis test between a linear null assumption versus ID alternative model.

To develop a test statistic for (10), we employ a quasi-score test based on the quasi-log-likelihood (8). This is a convenient choice, since such type of test requires only the estimation of model under the null hypothesis, which will be the linear model (4), say $\tilde{\theta} = (\tilde{\beta}_0, \tilde{\beta}_1, \tilde{\beta}_2)'$; this is usually a simpler task compared to the estimation of the nonlinear alternative model. Recall the partition of the parameters θ in (3), then $S_T(\theta) = (S_T^{(1)'}(\theta), S_T^{(2)'}(\theta))'$ denotes the corresponding partition of the quasi-score function (9). The quasi-score test statistic is given by

$$LM_T = S_T^{(2)'}(\tilde{\theta}) \Sigma_T^{-1}(\tilde{\theta}) S_T^{(2)}(\tilde{\theta})\,, \tag{11}$$

with $\Sigma_T(\tilde{\theta}) = J H_T^{-1}(\tilde{\theta}) J' \left(J H_T^{-1}(\tilde{\theta}) B_T(\tilde{\theta}) H_T^{-1}(\tilde{\theta}) J' \right)^{-1} J H_T^{-1}(\tilde{\theta}) J'$, where $J = (O_{m_2 \times m_1}, I_{m_2})$, I_s is a $s \times s$ identity matrix and $O_{a \times b}$ is a $a \times b$ matrix of zeros. $\Sigma_T(\tilde{\theta})$ is a the estimator for the unknown covariance matrix $\Sigma = \text{Var}[S_T^{(2)}(\tilde{\theta})]$. It can be proved that the quasi-score test (11) converges, asymptotically, to a $\chi^2_{m_2}$ distribution [4, Theorem 7]. Then, we reject H_0, if the value of LM_T computed in the available sample is greater than the critical values of the $\chi^2_{m_2}$ distribution, computed at ordinary significance levels. Analogous results hold for the continuous-valued model (2).

3.1 The Case of Non Identifiable Parameters

For model (3), consider the case where $f_i(\cdot)$ is defined as

$$\lambda_{i,t} = \beta_0 + \beta_1 X_{i,t-1} + \beta_2 Y_{i,t-1} + h_i(Y_{t-1}, \gamma)\alpha\,, \tag{12}$$

where $h_i(Y_{t-1}, \gamma)$ is a B-dimensional vector of nonlinear functions, say $h_i^b(Y_{t-1}, \gamma)$, with $b = 1, \ldots, B$, and α is the associated B-dimensional vector of nonlinear parameters. In practice, model (12) assumes that the nonlinear part of the network autoregressive models is of the form of an additive component. Note that the function $h_i(\cdot)$ depends on the lags of the variable and on k-dimensional vector of parameters γ. Several nonlinear models are included in (12). For example, the STNAR model (6), where $B = 1$ and $h_i(Y_{t-1}, \gamma) = \exp(-\gamma X_{i,t-1}^2)X_{i,t-1}$, for $i = 1, \ldots, N$, and the TNAR model (7), where $B = 3$ and $h_i^1(Y_{t-1}, \gamma) = I(X_{i,t-1} \leq \gamma)$, $h_i^2(Y_{t-1}, \gamma)) = X_{i,t-1} I(X_{i,t-1} \leq \gamma)$ and $h_i^3(Y_{t-1}, \gamma) = Y_{i,t-1} I(X_{i,t-1} \leq \gamma)$. Testing linearity on model (12) is equivalent to testing

$$H_0 : \alpha = 0\,, \quad \text{versus} \quad H_1 : \alpha \neq 0\,, \quad \text{elementwise}, \tag{13}$$

However, in this particular case, it is not possible to estimate the value of the parameter γ, because it is not identifiable under the null hypothesis H_0. Note that the parameter γ exists in the score partition function (9) because it is related to the nonlinear parameter $\theta^{(2)} = \alpha$. We conclude that the relevant quantities for inference and testing-see (11)—depend on γ, that is $S_T^{(2)}(\tilde{\theta}, \gamma)$, $\Sigma_T(\tilde{\theta}, \gamma)$ and $LM_T(\gamma)$. The model is then subject to non identifiable parameters γ under the null assumption. When this problem appears the standard theory does not apply and a chi-square type test is not suitable any more; see Davies [11] and Hansen [28], among several other references. Clearly, the value of the test changes over different values of $\gamma \in \Gamma$, where Γ is the domain of γ. A summary function of the test computed under different values of γ is then required; a typical choice is $g_T = \sup_{\gamma \in \Gamma} LM_T(\gamma)$. In practice, the space Γ is replaced by $\Gamma_F = (\gamma_L, \gamma_1, \ldots, \gamma_l, \gamma_U)$, a grid of values for the non identifiable parameters γ, and the maximum of the tests computed over such grid would be the test statistics employed for the evaluation of the test (13). Armillotta and Fokianos [4] established the convergence of the test g_T to g, when $T \to \infty$, being a function of a chi-square process, $LM(\gamma)$, in symbol $g = \sup_{\gamma \in \Gamma} LM(\gamma)$. The values of the latter asymptotic distribution cannot be tabulated, as this depends on unknown values of γ. We describe next methodology for computing the p-values of the sup-type test statistic.

3.2 Bootstrapping Test Statistics

Based on the previous arguments, we suggest to approximate the p-values of the test statistic by employing the following bootstrap algorithm

Algorithm 1 Score bootstrap

1: **for** $j = 1, \ldots, J$ **do**
2: Generate $\left\{ v_{t,j} : t = 1, \ldots, T \right\} \sim IIDN(0, 1)$.
3: Compute $S_T^{v_j}(\tilde{\theta}, \gamma) = \sum_{t=1}^{T} s_t(\tilde{\theta}, \gamma) v_{t,j}$.
4: Compute the test $LM_T^{v_j}(\gamma)$, for $\gamma \in \Gamma_F$, and $g_T^j = \sup_{\gamma \in \Gamma_F} LM_T^{v_j}(\gamma)$.
5: **end for**
6: Compute $p_T^J = J^{-1} \sum_{j=1}^{J} I(g_T^j \geq g_T)$.

An approximation of the p-values is obtained from step 6 of Algorithm 1, where g_T is the value of the test statistic computed on the available sample. When the bootstrap replication J is big enough p_T^J is a good approximation of the unknown p-values of the test. Then, the null hypothesis H_0 is rejected if p_T^J is smaller than a given significance level. In order to test the robustness and performances of Algorithm 1, we propose here a comparison with an alternative parametric bootstrap procedure.

Algorithm 2 Parametric bootstrap

Estimate parameters of the linear model (4), $\tilde{\theta}$.
for $j = 1, \ldots, J$ **do**
By using $\tilde{\theta}$ at step 1, generate from (1), with $f(\cdot)$ defined as in (4), a bootstrap sample $\bar{Y}_t^j$, with $t = 1, \ldots, T$.
Compute $\bar{S}_T^j(\tilde{\theta}, \gamma)$ from (9), by using the observations generated at step 3.
Compute the test $\overline{LM}_T^j(\gamma)$, for $\gamma \in \Gamma_F$, and $\bar{g}_T^j = \sup_{\gamma \in \Gamma_F} \overline{LM}_T^j(\gamma)$.
end for
Compute $\bar{p}_T^J = J^{-1} \sum_{j=1}^{J} I(\bar{g}_T^j \geq g_T)$.

The bootstrap p-values are obtained from step 7 of Algorithm 2. The parametric bootstrap method differs from the former approach because the source of randomness in the bootstrap iterations is not a multiplicative Gaussian noise $v_{t,j}$ but a resampling process which generates new pseudo-observations from the estimated model. The same methods apply unaffected to the continuous-valued model (2). We omit the details. In the following section we compare the performances of the testing methods proposed so far.

4 Applications

In this part of the chapter we illustrate the described methodologies for testing linearity for network autoregressive models on a set of synthetic and real data.

4.1 Simulation Results

Synthetic data obtained by Monte Carlo simulation are considered in this section. A network structure is required in the application of NAR models. Moreover, recall that the structure of the network is completely described by its adjacency matrix $A = (a_{ij}) \in \mathbb{R}^{N \times N}$ with a_{ij} such that $a_{ij} = 1$, if there is a directed edge from i to j and 0 otherwise. In this simulation study such network is generated following one of the most popular network structure models, the Stochastic Block Model (SBM), see Nowicki and Snijders [39], Wang and Wong [48] and Zhao et al. [51]. A block label ($l = 1, \ldots, K$) is assigned for each node with equal probability and K is the total number of blocks. Then, set $\mathrm{P}(a_{ij} = 1) = N^{-0.3}$ if i and j belong to the same block, and $\mathrm{P}(a_{ij} = 1) = N^{-1}$ otherwise. Practically, the model assumes that nodes within the same block are more likely to be connected with respect to nodes from different blocks. Throughout we assume the existence of two blocks ($K = 2$) and $N = 8$. The network is practically generated by using the `igraph` package of `R` software [10].

The observed count time series $\{Y_t : t = 1, \ldots, T = 1000\}$ is generated recursively as in (1), with λ_t coming from the linear model (4), using the copula-based data generating process of Armillotta and Fokianos [3, Sect. 2.1]. A choice of the copula function $C(\cdot)$ and the starting N-dimensional vector of the process λ_0 are required. The selected copula structure is Gaussian, $C_R^{Ga}(\ldots)$, with correlation matrix $R = \rho \bar{I}$, where $\bar{I}$ is a $N \times N$ matrix of ones; $\rho = 0.5$ is the copula parameter. Then $C_R^{Ga}(\ldots) = C^{Ga}(\ldots, \rho)$. We set $\lambda_0 = 1$ and use a burnout sample, by discarding the 300 first temporal observations to reduce the impact of the starting value of the process. The time series observations are obtained by setting the value of the linear parameters equal to $\theta^{(1)} = (\beta_0, \beta_1, \beta_2)' = (0.5, 0.2, 0.1)'$. This procedure is replicated $S = 200$ times. Then, the linear QMLE estimation $\tilde{\theta}$ optimising (8) is computed for each replication.

To generate the process Y_t in the continuous-valued case, the random errors $\xi_{i,t}$ are simulated from standard normal distribution $N(0, 1)$. For the data generating process of the vector Y_t, the initial value Y_0 is randomly simulated according to its stationary distribution [54, Proposition 1]. This is Gaussian with mean $\mu = \beta_0(1 - \beta_1 - \beta_2)^{-1}1$ and covariance matrix $\mathrm{vec}[\mathrm{Var}(Y_t)] = (I_{N^2} - G \otimes G)^{-1}\mathrm{vec}(I)$, where $1 = (1 \ldots 1)' \in \mathbb{R}^N$, I is the $N \times N$ identity matrix, $G = \beta_1 W + \beta_2 I$,$\otimes$ denotes the Kronecker product and $\mathrm{vec}(\cdot)$ the vec operator. Once the starting value Y_0 is given, the process $\{Y_t : t = 1, \ldots, T\}$ is generated recursively according to (4) and $Y_t = \lambda_t + \xi_t$, coming from (2). Then, the LS estimation of the linear parameters is computed for each replication. In this case, the resulting estimator is the ordinary least squares, which has closed form solution [54, Eq. 2.9].

We give here an example of a non standard case, by testing the linearity of model (4) versus the STNAR model; this is done by setting the hypothesis test $H_0 : \alpha = 0$ versus $H_1 : \alpha > 0$ in (6), inducing lack of identifiability on the parameter γ. According to Sect. 3.1, for each of the S replications, we can approximate the p-values of the sup-type test, $\sup_{\gamma \in \Gamma_F} LM_T(\gamma)$, where Γ_F is a grid of 10 equidistant values picked on [0.01, 3], by the two bootstrap approximation procedures described in Sect. 3.2,

Table 1 Empirical size at nominal significance levels $\alpha_{H_0} = \{0.1, 0.05, 0.01\}$ of the test statistics (11) for testing $H_0 : \alpha = 0$ in $S = 200$ simulations of model (6), for $N = 8$, $T = 1000$. Data are integer-valued and generated from (1), with the linear model (4). The empirical power is also reported for data generated from model (6) with $\alpha = \{0.3, 0.4\}$ and $\gamma = \{0.1, 0.2\}$. The network is derived from the SBM. The approximated p-values are computed by score bootstrap (p_T^J), in the first row, and parametric bootstrap ($\bar{p}_T^J$), second row

Method	Size			Power								
				$\gamma = 0.2, \alpha = 0.3$			$\gamma = 0.1, \alpha = 0.4$			$\gamma = 0.2, \alpha = 0.4$		
	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
p_T^J	0.020	0.015	0.000	0.260	0.155	0.075	0.445	0.300	0.075	0.590	0.475	0.270
$\bar{p}_T^J$	0.020	0.010	0.000	0.265	0.180	0.060	0.510	0.300	0.085	0.590	0.510	0.275

Table 2 Empirical size at nominal significance levels $\alpha_{H_0} = \{0.1, 0.05, 0.01\}$ of the test statistics (11) for testing $H_0 : \alpha = 0$ in $S = 200$ simulations of model (6), for $N = 8$, $T = 1000$. Data are continuous-valued and generated from (2), with the linear model (4). The empirical power is also reported for data generated from model (6) with $\alpha = \{0.3, 0.4\}$ and $\gamma = \{0.1, 0.2\}$. The network is derived from the SBM. The approximated p-values are computed by score bootstrap (p_T^J), in the first row, and parametric bootstrap ($\bar{p}_T^J$), second row

Method	Size			Power								
				$\gamma = 0.2, \alpha = 0.3$			$\gamma = 0.1, \alpha = 0.4$			$\gamma = 0.2, \alpha = 0.4$		
	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
p_T^J	0.070	0.025	0.000	0.970	0.940	0.755	0.370	0.140	0.020	0.995	0.990	0.950
$\bar{p}_T^J$	0.070	0.020	0.000	0.970	0.905	0.720	0.275	0.105	0.005	0.990	0.985	0.915

with $J = 299$ bootstrap replications. The fraction of cases over S simulations in which the p-value approximations is smaller than the usual significance levels 0.1, 0.05 and 0.01 is the frequency of cases where H_0 is rejected and constitutes the empirical size of the test. The empirical power of the test is again the frequency of cases where H_0 is rejected but obtained when data were generated by the model (6) instead. This is accomplished by using the same generating mechanism described for the linear model, by setting various combinations of values of nonlinear parameters $\alpha = \{0.3, 0.4\}$ and $\gamma = \{0.1, 0.2\}$.

The results of the simulation study for the count data case are reported in Table 1. We note that the empirical size is smaller than or close to the expected nominal levels; the empirical power is low when α is small and tends to grow for larger values of α far from the value of the null assumption. The two bootstrap methods show similar behavior, but the parametric bootstrap yields slightly better when compared to the score based bootstrap. Such results show that both tests works satisfactorily with a slight preference given to the parametric bootstrap methodology.

Table 2 considers results regarding the continuous case. Firstly, we see an overall improvement of the performances compared with the integer-valued case. This is expected since here the errors ξ_t are generated from Normal random variables and also the stationary distribution of the process Y_t is Gaussian. Hence, the χ^2 (process)

distribution of the test is approached more quickly. Instead in the integer-valued case such distribution will be reached only asymptotically, with $N \to \infty$, $T_N \to \infty$. The results of the two bootstrap procedures are again similar, but we note that the score bootstrap slightly outperforms the parametric one.

4.2 New COVID-19 Cases on Italian Provinces

We study a dataset which consists of daily new cases of COVID-19 virus detected for each province of Italy, according to the Nomenclature of Territorial Units for Statistics, Level 3 (NUTS-3) classification, as established on Regulation (EC) No 1059/2003 of the European Parliament and of the Council. Data is provided by the Presidenza del Consiglio dei Ministri—Dipartimento della Protezione Civile.[1] The total number of provinces is $N = 107$. The time series starts at 25/02/2020 and is updated daily until 07/02/2022 ($T = 714$). For the considered regions and time window, we observed two instances of negative numbers of new cases. These values are replaced by zero counts.

An undirected network structure can be derived by exploiting available data on geographical coordinates. The geodesic distance between the centroids of pairs of provinces $\{i, j\}$ are computed, say d_{ij}. Then, two provinces $\{i, j\}$ are connected with an undirected edge if $d_{ij} \leq 200$ km. We consider such cut-off reasonable by considering that a smaller distance would results in few connections for most remote regions, like the islands, whereas a bigger distance will result in a fully connected network, i.e. a network which connects each node to all the others, which is not of interest in the current analysis. The density of the network is 21.58%. The histogram of the number of connections is shown in Fig. 1. The maximum number of connections is 45. The median number of connections is 22.

We see from Fig. 2 a typical time series for each province. The data show that it is possible to detect at least two regimes of variation; one during pandemic seasonal waves, with high numbers of daily new cases and one where the virus cases are relatively stable for several months. We address the question that a linear model is suitable for fitting such data. The partial autocorrelation function (PACF) of the time series indicates a significant effect of the past counts so an autoregressive model may be adequate to model the dataset. The median number of daily new cases is 27.

Estimation of the linear PNAR model (4) is performed by QMLE. For testing linearity, the quasi-score linearity test is computed according to (11). For the identifiable case, the asymptotic chi-square test is employed, for the nonlinear model (5), testing $H_0 : \gamma = 0$ versus $H_1 : \gamma > 0$. For non identifiable case, we test linearity against the presence of smooth transition effects, as in (6), with $H_0 : \alpha = 0$ versus $H_1 : \alpha > 0$. A grid of 10 equidistant values in the interval $\Gamma_F \equiv [0.001, 3]$ is chosen for values of the nuisance parameter γ. The p-values are computed for the test

[1] Dataset available at https://github.com/pcm-dpc/COVID-19/blob/master/dati-province/dpc-covid19-ita-province.csv.

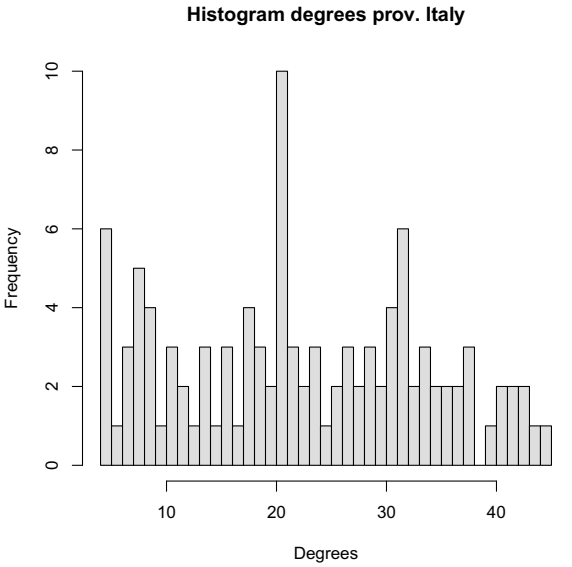

Fig. 1 Histograms of number of connections (degrees) between provinces of Italy

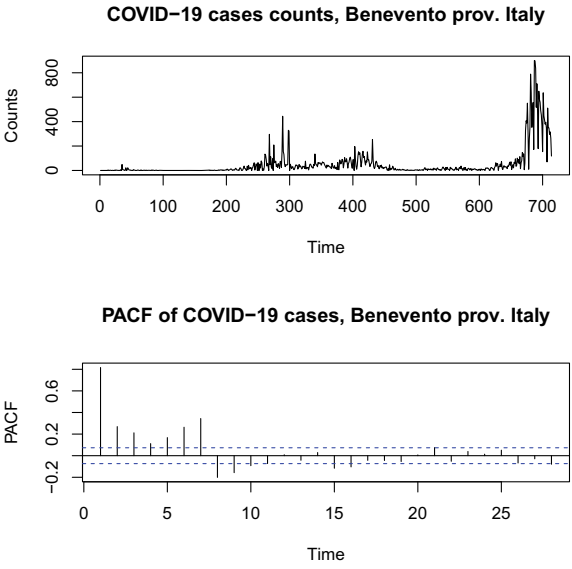

Fig. 2 Time series of counts and partial autocorrelation function for the number of daily new COVID-19 cases in Benevento province, Italy. Dashed blue line: 5% confidence bands

Table 3 QMLE estimates of the linear model (4) for daily COVID-19 new cases in Italy. Standard errors in brackets. Linearity is tested against the ID nonlinear model (5), with χ^2_1 asymptotic test (11); against the STNAR model (6), with approximated p-values computed by score bootstrap (p^J_T), parametric bootstrap ($\bar{p}^J_T$); and versus TNAR model (7)

Models	$\tilde{\beta}_0$	$\tilde{\beta}_1$	$\tilde{\beta}_2$
Linear	1.665 (0.462)	0.149 (0.016)	0.842 (0.025)
Models	χ^2_1	p^J_T	$\bar{p}^J_T$
ID	3.585	–	–
STNAR	–	<0.001	<0.001
TNAR	–	0.600	0.800

$\sup LM_T = \sup_{\gamma \in \Gamma_F} LM_T(\gamma)$ through the two bootstrap approximation procedures described in this work. The number of bootstrap replication is set to $J = 299$. For the parametric bootstrap, the generation of pseudo-observations requires the choice of a copula and related parameter. We chose the Gaussian copula with correlation matrix $R = \rho \bar{I}$, and $\rho = 0.5$. Finally, a linearity test against threshold effects, as in (7), is also performed, which leads to the test $H_0 : \alpha_0 = \alpha_1 = \alpha_2 = 0$ versus $H_1 : \alpha_l > 0$, for some $l = 0, 1, 2$. In order to determine a feasible range of values for the non identifiable threshold parameter, we compute the quantiles at 10% and 90% of the empirical distribution for the process $\{X_{i,t} : t = 1, \ldots, T\}$, at each $i = 1, \ldots, N$. Then, we take the minimum of 10% quantiles and the maximum of 90% quantiles as the extremes of Γ_F, from which a grid of 10 equidistant values is picked.

The results are summarized in Table 3. The estimated parameters for the linear model (4) are highly significant. The magnitude of the network effect β_1 appears to agree with intuition, as an increasing number of cases in a province can lead to a growth in cases found in a close geographic area. The effect of the lagged variable has a upwards impact on the number of cases, as expected by the observed temporal dependence. The linearity test against the nonlinear model (5) is rejected at 0.1 significance level, since the value of the test statistics is greater than the critical values of the χ^2_1 distribution, but not at 0.05 and 0.01 levels. This gives a mild evidence for possible nonlinear drifts in the intercept. The linearity is strongly rejected when tested against the STNAR model, by both bootstrap tests at all levels 0.1, 0.05 and 0.01. Nevertheless, bootstrap sup-type tests do not show evidence of threshold effects in the model. Then, we conclude that there is a clear evidence in accordance to regime switching effects with smooth switching rather than abrupt shifts. These findings are in line with the values of the time series, as shown in Fig. 2.

Acknowledgements This work has been co-financed by the European Regional Development Fund and the Republic of Cyprus through the Research and Innovation Foundation, under the project INFRASTRUCTURES/1216/0017 (IRIDA).

References

1. Ahmad, A., Francq, C.: Poisson QMLE of count time series models. J. Time Ser. Anal. **37**, 291–314 (2016)
2. Andrews, D.W.K., Ploberger, W.: Optimal tests when a nuisance parameter is present only under the alternative. Econometrica **62**, 1383–1414 (1994)
3. Armillotta, M., Fokianos, K.: Poisson network autoregression (2021). arXiv:2104.06296
4. Armillotta, M., Fokianos, K.: Testing linearity for network autoregressive models (2022). arXiv:2202.03852
5. Boos, D.D.: On generalized score tests. Amer. Stat. **46**, 327–333 (1992)
6. Bracher, J., Held, L.: Endemic-epidemic models with discrete-time serial interval distributions for infectious disease prediction. Inter. J. Forecast. in press (2020)
7. Christou, V., Fokianos, K.: Quasi-likelihood inference for negative binomial time series models. J. Time Ser. Anal. **35**, 55–78 (2014)
8. Christou, V., Fokianos, K.: Estimation and testing linearity for non-linear mixed Poisson autoregressions. Electron. J. Stat. **9**, 1357–1377 (2015)
9. Cliff, A., Ord, J.K.: Space-time modelling with an application to regional forecasting. Trans. Inst. Brit. Geograph. 119–128 (1975)
10. Csardi, G., Nepusz, T.: The igraph software package for complex network research. Inter. J. Complex Syst. 1695 (2006). https://igraph.org
11. Davies, R.B.: Hypothesis testing when a nuisance parameter is present only under the alternative. Biometrika **74**, 33–43 (1987)
12. Davis, R.A., Liu, H.: Theory and inference for a class of nonlinear models with application to time series of counts. Stat. Sinica **26**, 1673–1707 (2016)
13. Davis, R.A., Fokianos, K., Holan, S.H., Joe, H., Livsey, J., Lund, R., Pipiras, V., Ravishanker, N.: Count time series: a methodological review. J. Amer. Stat. Assoc. **116**, 1533–1547 (2021)
14. Douc, R., Fokianos, K., Moulines, E.: Asymptotic properties of quasi-maximum likelihood estimators in observation-driven time series models. Electr. J. Stat. **11**, 2707–2740 (2017)
15. Doukhan, P.: Mixing. Lecture Notes in Statistics, vol. 85. Springer, New York (1994)
16. Fan, J., Yao, Q.: Nonlinear Time Series: Nonparametric and Parametric Methods. Springer, New York (2003)
17. Fokianos, K.: Multivariate count time series modelling. To appear in Econometrics and Statistics (2022)
18. Fokianos, K., Neumann, M.H.: A goodness-of-fit test for Poisson count processes. Electr. J. Stat. **7**, 793–819 (2013)
19. Fokianos, K., Tjøstheim, D.: Nonlinear Poisson autoregression. Ann. Inst. Stat. Math. **64**, 1205–1225 (2012)
20. Fokianos, K., Rahbek, A., Tjøstheim, D.: Poisson autoregression. J. Amer. Stat. Assoc. **104**, 1430–1439 (2009)
21. Fokianos, K., Støve, B., Tjøstheim, D., Doukhan, P.: Multivariate count autoregression. Bernoulli **26**, 471–499 (2020)
22. Francq, C., Horvath, L., Zakoïan, J.M.: Sup-tests for linearity in a general nonlinear AR(1) model. Econom. Theory **26**, 965–993 (2010)
23. Gao, J.: Nonlinear Time Series: Semiparametric and Nonparametric Methods. CRC Press, Boca Raton (2007)
24. Gao, J., King, M., Lu, Z., Tjøstheim, D.: Specification testing in nonlinear and nonstationary time series autoregression. Ann. Stat. **37**, 3893–3928 (2009)
25. Gorgi, P.: Beta-negative binomial auto-regressions for modelling integer-valued time series with extreme observations. J. R. Stat. Soc.: Ser. B **82**, 1325–1347 (2020)
26. Gourieroux, C., Monfort, A., Trognon, A.: Pseudo maximum likelihood methods: theory. Econometrica 681–700 (1984)
27. Haggan, V., Ozaki, T.: Modelling nonlinear random vibrations using an amplitude-dependent autoregressive time series model. Biometrika **68**(1), 189–196 (1981)

28. Hansen, B.E.: Inference when a nuisance parameter is not identified under the null hypothesis. Econometrica **64**, 413–430 (1996)
29. Heyde, C.C.: Quasi-likelihood and its Application. A General Approach to Optimal Parameter Estimation. Springer Series in Statistics. Springer, New York (1997)
30. Knight, M., Nunes, M., Nason, G.: Modelling, detrending and decorrelation of network time series (2016). arXiv:1603.03221
31. Knight, M., Leeming, K., Nason, G., Nunes, M.: Generalized network autoregressive processes and the GNAR package. J. Stat. Softw. **96**, 1–36 (2020). https://www.jstatsoft.org/v096/i05
32. Kolaczyk, E.D., Csárdi, G.: Statistical Analysis of Network Data with R, vol. 65. Springer, Berlin (2014)
33. Li, G., Li, W.K.: Testing a linear time series model against its threshold extension. Biometrika **98**, 243–250 (2011)
34. Lim, K., Tong, H.: Threshold autoregressions, limit cycles, and data. J. R. Stat. Soc. Ser B **42**, 245–92 (1980)
35. Luukkonen, R., Saikkonen, P., Teräsvirta, T.: Testing linearity against smooth transition autoregressive models. Biometrika **75**, 491–499 (1988)
36. Martin, R.L., Oeppen, J.: The identification of regional forecasting models using space: time correlation functions. Trans. Inst. Brit. Geograph. 95–118 (1975)
37. McCullagh, P., Nelder, J.A.: Generalized Linear Models, 2nd edn. Chapman & Hall, London (1989)
38. Neumann, M.: Absolute regularity and ergodicity of Poisson count processes. Bernoulli **17**, 1268–1284 (2011)
39. Nowicki, K., Snijders, T.A.B.: Estimation and prediction for stochastic blockstructures. J. Amer. Stat. Assoc. **96**, 1077–1087 (2001)
40. Pedeli, X., Karlis, D.: A bivariate INAR (1) process with application. Stat. Modell. **11**, 325–349 (2011)
41. Pedeli, X., Karlis, D.: On composite likelihood estimation of a multivariate INAR (1) model. J. Time Ser. Anal. **34**, 206–220 (2013)
42. Pedeli, X., Karlis, D.: Some properties of multivariate INAR (1) processes. Comput. Stat. & Data Anal. **67**, 213–225 (2013)
43. Rosenblatt, M.: A central limit theorem and a strong mixing condition. Proc. Natl. Acad. Sci. U. S. A. **42**, 43–47 (1956)
44. Teräsvirta, T.: Specification, estimation, and evaluation of smooth transition autoregressive models. J. Amer. Stat. Assoc. **89**, 208–218 (1994)
45. Teräsvirta, T., Tjøstheim, D., Granger, C.W.J.: Modelling Nonlinear Economic Time Series. Oxford University Press, Oxford (2010)
46. Tong, H.: Non-linear Time Series: A Dynamical System Approach. Oxford University Press, Oxford (1990)
47. Wang, C., Liu, H., Yao, J.F., Davis, R.A., Li, W.K.: Self-excited threshold Poisson autoregression. J. Amer. Stat. Assoc. **109**, 777–787 (2014)
48. Wang, Y.J., Wong, G.Y.: Stochastic blockmodels for directed graphs. J. Amer. Stat. Assoc. **82**, 8–19 (1987)
49. Wasserman, S., Faust, K., et al.: Social Network Analysis: Methods and Applications, vol. 8. Cambridge University Press, Cambridge (1994)
50. Wedderburn, R.W.: Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. Biometrika **61**(3), 439–447 (1974)
51. Zhao, Y., Levina, E., Zhu, J., et al.: Consistency of community detection in networks under degree-corrected stochastic block models. Ann. Stat. **40**(4), 2266–2292 (2012)
52. Zhou, J., Li, D., Pan, R., Wang, H.: Network GARCH model. Stat. Sin. **30**, 1–18 (2020)
53. Zhu, X., Pan, R.: Grouped network vector autoregression. Stat. Sin. **30**, 1437–1462 (2020)
54. Zhu, X., Pan, R., Li, G., Liu, Y., Wang, H.: Network vector autoregression. Ann. Stat. **45**, 1096–1123 (2017)
55. Zhu, X., Wang, W., Wang, H., Härdle, W.K.: Network quantile autoregression. J. Econometr. **212**, 345–358 (2019)