The Annals of Statistics
2023, Vol. 51, No. 6, 2526–2552
https://doi.org/10.1214/23-AOS2345

NONLINEAR NETWORK AUTOREGRESSION

BY MIRKO ARMILLOTTA[1,a] AND KONSTANTINOS FOKIANOS[2,b]

[1]*Department of Econometrics and Data Science, Vrije Universiteit Amsterdam,* [a]*m.armillotta@vu.nl*
[2]*Department of Mathematics and Statistics, University of Cyprus,* [b]*fokianos.konstantinos@ucy.ac.cy*

We study general nonlinear models for time series networks of integer and continuous-valued data. The vector of high-dimensional responses, measured on the nodes of a known network, is regressed nonlinearly on its lagged value and on lagged values of the neighboring nodes by employing a smooth link function. We study stability conditions for such multivariate process and develop quasi-maximum likelihood inference when the network dimension is increasing. In addition, we study linearity score tests by treating separately the cases of identifiable and nonidentifiable parameters. In the case of identifiability, the test statistic converges to a chi-square distribution. When the parameters are not identifiable, we develop a supremum-type test whose p-values are approximated adequately by employing a feasible bound and bootstrap methodology. Simulations and data examples support further our findings.

1. Introduction. The availability of network data recorded over a timespan in several applications (social networks, GPS data, epidemics, etc.) requires assessing the effect of a network structure to a multivariate time series process. This problem has attracted considerable attention. Recently, [66] proposed a Network Autoregressive model (NAR), under Independent Identically Distributed (*IID*) innovation sequence, where a continuous response variable is observed for each node of a network. The high-dimensional vector of such responses is modelled linearly on the past values of the response, measured on the node itself and the average lagged response of the neighbours connected to the node. Motivated by the fact that real data networks are usually of large dimension, the authors study least squares inference under two large-sample regimes (a) with increasing time sample size, that is, $T \to \infty$, and fixed network dimension, say N, and (b) with $N \to \infty$ and $T_N \to \infty$, where the temporal size depends on N. In this contribution, we extend this work in various directions by proposing appropriate inference and testing methodology, which is applicable to continuous and integer valued data.

1.1. *Related works*. The NAR model has been the focus of recent research, for example, logistic network models [63], network quantile model [67], grouped least squares estimation [65], network GARCH models [64] and applications [8]. In addition, [42] consider the Generalized NAR model, for the continuous case, which incorporates the effect of several layers of connections between nodes of the network. This study is accompanied by appropriate R software. As pointed out by [66], discrete response variables are frequently encountered in applications and are strongly related to network data. For example, in the social network analysis data are counts, for example, number of characters contained in posts of single users, number of posts shared, etc. Models for count time series have been studied by [6] who introduced linear and log-linear Poisson network autoregression models (PNAR). Such extensions show that the NAR model is a member of the broad class of Generalized Linear Models

Received November 2022; revised July 2023.
MSC2020 subject classifications. Primary 62M10; secondary 62J02.
Key words and phrases. Contraction, hypothesis testing, increasing dimension, multivariate count time series.

(GLM) [45]; for the count case, the observations are in fact marginally Poisson distributed, conditionally to their past history. The joint dependence is imposed by employing a copula construction, as introduced by [30] and is outlined in [7], Section S-1.1. In addition, [6] have studied thoroughly the two related types of asymptotic inference (a)-(b) discussed above, in the context of quasi-maximum likelihood inference (QMLE) [41].

1.2. *Nonlinear models and testing linearity.* Theory related to NAR model relies on the assumption of linearity. However, there are many real world examples where a nonlinear model might be more appropriate. For instance, in economics, the theory supports occasionally nonlinear behaviour; see [55], Chapter 2, for several examples. In modelling economic/financial time series, it seems natural to allow for the existence of different states, or regimes (e.g., expansion/crisis), such that dynamics depend upon the specific regime; [68], Chapter 18. Government agencies, research institutes and central banks may typically employ nonlinear models [55], p. 16. As far as social network analysis concerned, nonlinear phenomena are frequently observed, for example, "superstars" with huge number of followers having an exponentially higher impact on other users' behaviour when compared to the "standard" user [66]. So from both theoretical and applied point of view, there exists a necessity to develop nonlinear autoregression model theory for the case of network time series. Literature in univariate nonlinear time series is vast and well developed, in particular for continuous random variables. The interested reader is referred to [11, 56] and [27] among many others. For integer-valued data, the literature is more recent, although still flourishing; see [12, 16, 18, 29, 31, 58] and [37]. General results are given by [1, 17, 21, 22, 61] and [9].

Estimation for nonlinear models has been traditionally accompanied by tests who examine the assumption of linearity. Such tests are routinely used because they provide a sound framework where evidence of linearity can be examined thoroughly. In addition, they offer guidance about the specific nonlinear model to be fitted; see [54], Section 3, who suggests that "The first step of a specification strategy for any type of nonlinear model should therefore consist of testing linearity". Furthermore, proper inference is developed especially when the linear model is nested within a nonlinear model and as such the resulting estimators obtained after fitting a nonlinear model may be inconsistent; see [55], Section 5.1,5.5. Finally, it is always important to have additional tools for both practical usefulness and for explanatory data analysis—detecting latent variables, change point testing, etc. These points motivated the growth of a large literature on linearity tests, especially for continuous-valued random variables. A survey of general type results for test statistics whose application depends on identifiable/nonidentifiable parameters are given by [32] and [40]. Finally, [5] and [38] established a general framework for testing linearity when some parameters are nonidentifiable under the hypothesis of linearity. Nonlinear models for count time series and the associated testing linearity problem has been studied by [13] who suggest a quasi-score test for (mixed) Poisson random variables. All the above works are concerned with univariate time series. Related literature on multivariate observation-driven models for discrete-valued data considers only linear cases; see [47–49] and [30] among others.

1.3. *Main challenges.* Existing theory does not cover the case of NAR models, which are multivariate and their properties depend on both N (size) and T (time). Therefore, asymptotically, both indices, N and T, tend to infinity and it is a great challenge to address the properties of such multivariate processes. Moreover, QMLE inference breaks down when estimating network models because N is large. Consequently, nonstandard proofs are required for establishing stationarity of infinite-dimensional processes and to obtain sound inference within the double asymptotic regime (b). Note that even a simple weak law does not hold under regime (b). In particular, quantities related to the inference are of the order $\mathcal{O}(N)$. Consider, for example, the sample information matrix, which depends on the network structure and diverges with $N \to \infty$. Then the covariance of estimators explodes and proving existence of limiting Hessian and Fisher matrices is a challenging problem. These issues become more persistent when testing linearity, especially in the case of nonidentifiable parameters. Then a double indexed asymptotic theory with infinite-dimensional vectors over a uniform metric space for the score and related matrices is relevant and asks for development.

1.4. *Our contribution.* The main results of our contribution are the following: (i) Specification of a novel general nonlinear network autoregressive model for both continuous and discrete valued multivariate network observations (Section 2); (ii) Under mild conditions, stationarity results (Sections 2.2–2.3) and asymptotic theory of QMLE are established, when both time and network dimensions increases (Section 3). These are new results because nonlinear NAR models have not been treated in the literature; (iii) Development of testing procedures for examining linearity of the model by employing a quasi-score based test under both asymptotic regimes (a)–(b); see Section 4. We focus on score tests, as they require fitting NAR models under the null hypothesis. This is a computationally simpler task. Their asymptotic distribution is (noncentral) chi-square even when the parameters to be tested lie on the boundary of the parameter space; (iv) Finally, we consider the situation where nonidentifiable parameters, say γ, are present under the null. In such case, the results of Sections 2–4 are extended. This is done by proving stochastic equi-continuity of the score with diverging number of nodes and double-indexed convergence of Hessian/information matrices uniformly over γ. Then, as $N \to \infty$ and $T_N \to \infty$, we show that the quasi-score linearity test asymptotically approximates a (noncentral) chi-square process (Section 5). We discuss two ways to approximate the p-values of the tests: by the upper bound developed in [15], and by bootstrap approximation relying on stochastic permutations [38]. The double asymptotic convergence of bootstrap p-values to their theoretical counterpart is also established. We are not aware of other contributions, to the best of our knowledge, attacking the problem of general asymptotic inference with increasing dimension network time series models for both count and continuous data. The last sections discuss results of a simulation study (Section 6) and real data examples (Section 7). All the methodology is implemented in the new released `R` package `PNAR` [10, 57]. The paper concludes with an Appendix, containing the proofs for Sections 2 and 5. The Supplementary Material [7] contains the proofs for Sections 3 and 4 and additional results for the threshold network autoregressive model, under asymptotic regime (a).

1.5. *Notation.* We denote $|x|_r = (\sum_{j=1}^{p} |x_j|^r)^{1/r}$ the l^r-norm of a $p \times 1$ vector x. If $r = \infty$, $|x|_\infty = \max_{1 \le j \le p} |x_j|$. Let $\|X\|_r = (\sum_{j=1}^{p} \mathrm{E}(|X_j|^r))^{1/r}$ the L^r-norm for a random vector X. For a $q \times p$ matrix $M = (m_{ij})$, $i = 1, \dots, q$, $j = 1, \dots, p$, denote 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. Define the entrywise norms $|M|_r = (\sum_{i=1}^{q} \sum_{j=1}^{p} |m_{ij}|^r)^{1/r}$. If $q = p$ and $1 \le r \le 2$, these are matrix norms. Define $|x|_{\mathrm{vec}}^r = (|x_1|^r, \dots, |x_p|^r)'$, $\|X\|_{r,\mathrm{vec}} = (\mathrm{E}^{1/r}|X_1|^r, \dots, \mathrm{E}^{1/r}|X_p|^r)'$, $|M|_{\mathrm{vec}} = (|m_{ij}|)_{(i,j)}$ and $\preceq$ a partial order relation on $x, y \in \mathbb{R}^p$ such that $x \preceq y$ means $x_i \le y_i$ for $i = 1, \dots, p$. The same notation holds for random vectors X, Y such that $X \preceq Y$ means $X_i \le Y_i$ almost surely (a.s.) for $i = 1, \dots, p$. Set the compact notation $\max_{1 \le i < \infty} x_i = \max_{i \ge 1} x_i$. The notation C_r denote a constant, which depend on r, where $r \in \mathbb{N}$, and C is a generic constant. The symbol I denotes an identity matrix, 1 (0) a vector of ones (zeros), whose dimension depends on context. Let $\Rightarrow$ denote weak convergence with respect to the uniform metric. Finally, the notation $\{N, T_N\} \to \infty$ will be used as a shorthand for $N \to \infty$ and $T_N \to \infty$.

2. Nonlinear NAR model specification. Consider a network with N nodes (network size) indexed by $i = 1, \dots N$. The neighbourhood structure of the network is explicitly described by its adjacency matrix $A = (a_{ij}) \in \mathbb{R}^{N \times N}$ where $a_{ij} = 1$, 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 ($A = A'$) but self-relationships are excluded, that is, $a_{ii} = 0$ for any $i = 1, \dots, N$. This is a typical and realistic assumption, for example, social networks; see [60] and [43] among others. The network structure, equivalently the matrix A, is assumed to be nonrandom. A row-normalized adjacency matrix is defined by $W = \operatorname{diag}\{n_1, \dots, n_N\}^{-1} A$ where $n_i = \sum_{j=1}^{N} a_{ij}$ is the so-called out-degree, the total number of edges starting from the node i. Then W satisfies $\|\|W\|\|_\infty = 1$ and $W1 = 1$. Moreover, define e_i the N-dimensional unit vector with 1 in the ith position and 0 everywhere else, such that $w_i' = e_i' W = (w_{i1} \dots, w_{iN})$ the ith row of the matrix W, with $w_{ij} = a_{ij}/n_i$.

Define a N-dimensional vector of time series $\{Y_t = (Y_{1,t} \dots Y_{i,t} \dots Y_{N,t})', t = 1, 2 \dots, T\}$ which is observed on a given network; that is, a univariate time series is measured for each node, with rate $\lambda_{i,t}$. Denote by $\{\lambda_t \equiv \mathrm{E}(Y_t | \mathcal{F}_{t-1}) = (\lambda_{1,t} \dots \lambda_{i,t} \dots \lambda_{N,t})', t = 1, 2 \dots, T\}$, the corresponding conditional expectation vector, and denote the history of the process by $\mathcal{F}_t = \sigma(Y_s : s \le t)$. Assume that $\{Y_t : t \in \mathbb{Z}\}$ is integer-valued and consider the following first-order nonlinear Poisson Network Autoregression (PNAR):

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

where $\{N_t\}$ is a sequence of *IID* N-variate copula-Poisson process with intensity 1, counting the number of events in $[0, \lambda_{1,t}] \times \cdots \times [0, \lambda_{N,t}]$ and $f(\cdot)$ is a deterministic function depending on the past lag values of the count process, the known network structure W and an m-dimensional parameter vector θ. Examples will be given below. More precisely, the conditional marginal probability distribution of the count variables is $Y_{i,t} | \mathcal{F}_{t-1} \sim \text{Poisson}(\lambda_{i,t})$, for $i = 1, \dots, N$, and the joint distribution is generated by a copula, which depends on a parameter ρ, say $C(\cdot, \rho)$, and it is imposed on waiting times of a Poisson process specified as in [6], Section 2.1; see [7], Section S-1.1. Several alternative models resembling multivariate Poisson distributions have been proposed in the literature; see [28], Section 2, for a discussion about the issues of available multivariate count distributions. A copula-based approach for the data generating process (henceforth DGP) is used throughout this paper. Results for higher-order models are derived straightforwardly; see Remark 5.

Similar to the case of integer-valued time series, we define the nonlinear Network Autoregression (NAR) for continuous-valued time series by

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

where $\xi_{i,t} \sim IID(0, \sigma^2)$, for $1 \le i \le N$ and $1 \le t \le T$ and $\lambda_t = \mathrm{E}(Y_t | \mathcal{F}_{t-1})$.

Denote by $X_{i,t} = n_i^{-1} \sum_{j=1}^{N} a_{ij} Y_{j,t}$ the network effect, that is, the average impact of node i's connections. Consider the partition of the parameter vector $\theta = (\theta^{(1)\prime}, \theta^{(2)\prime})'$, where the vectors $\theta^{(1)}$ and $\theta^{(2)}$ are of dimension m_1 and m_2, respectively, such that $m_1 + m_2 = m$. For $t = 1 \dots, T$, both (1)–(2) have elementwise components

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

where $f_i(\cdot)$ is the ith component of the function $f(\cdot)$ depending on the specific model of interest, which can contain linear and nonlinear effects. In general, $\theta^{(1)}$ will denote an $m_1 \times 1$ vector associated with linear model parameters, whereas $\theta^{(2)}$ will denote the $m_2 \times 1$ vector of nonlinear parameters. Some examples are given below.

2.1. *Examples.*

EXAMPLE 1. Consider (2) and the first-order linear NAR(1),

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

which is a special case of (3), with $\theta^{(1)} = (\beta_0, \beta_1, \beta_2)'$. Model (4) was originally introduced by [66] for the case of continuous random variables Y_t, such that $Y_{i,t} = \lambda_{i,t} + \xi_{i,t}$. 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 directly connected with the focal node i can impact on the conditional mean process $\lambda_{i,t}$. This is reasonable assumption in many applications; for example, in the social network analysis, if the focal node i does not follow a node l, so $a_{il} = 0$, the effect of the activity related to the latter does not affect the former. The parameter β_1 is called network effect, as it measures the average impact of the i'th node connections. The coefficient β_2 is called autoregressive effect because it determines the impact of the lagged variable $Y_{i,t-1}$. Model (4) has been extended to the case of count time series by [6]; it is called the linear PNAR(1) with $Y_{i,t}|\mathcal{F}_{t-1} \sim \text{Poisson}(\lambda_{i,t})$ for $i = 1, \dots, N$ and the copula-based DGP, as described earlier.

EXAMPLE 2. A nonlinear deviation of (4), when Y_t takes integer values is given by

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

where $\gamma \geq 0$. Clearly, (5) approaches a linear model for small values of γ, and $\gamma = 0$ reduces to the linear model (4). Instead, when γ is larger than zero, (5) introduces 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 $\theta^{(2)} = \gamma$. Model (5) introduces a nonlinear drift in the intercept so that the baseline effect varies over time as a function of the network. If $Y_{i,t}$ counts activities of users in a social network (likes, reactions, etc.) and the community becomes more active, then the average magnitude of $X_{i,t-1}$ grows, and thus the baseline for each node i varies. When $Y_t \in \mathbb{R}^N$, the following model

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

is analogous to (5) but for continuous-valued time series. To the best of our knowledge, we are not aware of any stability or inferential results for models (5)–(6) when $\{N, T_N\} \to \infty$. When $N = 1$ and $T \to \infty$, such nonlinear models have been studied by by [31, 34] among others.

EXAMPLE 3. Another example of (3) is given by the smooth transition version of the NAR model, say STNAR(1),

$$\lambda_{i,t} = \beta_0 + (\beta_1 + \alpha \exp(-\gamma X_{i,t-1}^2)) X_{i,t-1} + \beta_2 Y_{i,t-1}, \tag{7}$$

where $\gamma \geq 0$; see [53] for an introduction to STAR models. This models introduces a smooth regime switching behaviour of the network effect making it possible to vary smoothly from β_1 to $\beta_1 + \alpha$, as γ varies from large to small values. When $\alpha = 0$ in (7), the linear NAR model (4) is obtained. Moreover, (7) is a special case of (3), with $\theta^{(1)} = (\beta_0, \beta_1, \beta_2)'$ and $\theta^{(2)} = (\alpha, \gamma)'$. In the case of univariate count time series, see [29, 31] for more.

EXAMPLE 4. Define the threshold NAR model ([44]), say TNAR(1), by

$$\lambda_{i,t} = \beta_0 + \beta_1 X_{i,t-1} + \beta_2 Y_{i,t-1} + (\alpha_0 + \alpha_1 X_{i,t-1} + \alpha_2 Y_{i,t-1}) I(X_{i,t-1} \leq \gamma), \tag{8}$$

where $I(\cdot)$ is the indicator function and γ is the threshold parameter. When $\alpha_0 = \alpha_1 = \alpha_2 = 0$, model (8) reduces to the linear model (4). In this case, $\theta^{(1)} = (\beta_0, \beta_1, \beta_2)'$ and $\theta^{(2)} = (\alpha_0, \alpha_1, \alpha_2, \gamma)'$ show that (8) is a special case of (3). In the case of univariate count time series, see [13, 21, 58, 61] for more.

Nonlinear functions, such as (7)–(8), provide examples of switching models accounting for regime specific dynamics of the observed process. The switching mechanism depends on the network effect. For example, consider A as the network matrix connecting regional districts, which share at least a border. Let $Y_{i,t}$ denote the numbers of reported cases for some disease in each of these districts. Then, for each district i, the historical average of neighbours ($X_{i,t-1}$) determines a switching effect, say from exponentially expanding pandemic to dying out pandemic (and vice versa). Note that for (7), the network effect is regime-dependent but this can be modified suitably as in (8). In conclusion, the dichotomy between STNAR and TNAR models is that the former accounts for smooth transitions while the latter models sudden changes; see [55] for more on nonlinear modelling of time series.

2.2. *Stability conditions for fixed network size.* Set $f(\cdot, W, \theta) = f(\cdot)$.

THEOREM 2.1. *Consider model (1), with fixed N. Define $G = \mu_1 W + \mu_2 I$, where μ_1, μ_2 are nonnegative constants such that $\rho(G) < 1$ and assume that for $y, y^* \in \mathbb{N}^N$,*

$$\left|f(y) - f(y^*)\right|_{\text{vec}} \preceq G\left|y - y^*\right|_{\text{vec}}. \tag{9}$$

Then the process $\{Y_t,\ t \in \mathbb{Z}\}$ is stationary, ergodic and $\mathrm{E}|Y_t|_a^a < \infty$ for any $a \geq 1$.

The parallel result for continuous variables is also established.

THEOREM 2.2. *Consider model (2), with fixed N. Define $G = |\mu_1|W + |\mu_2|I$, where μ_1, μ_2 are real constants such that $\rho(G) < 1$ and the contraction condition (9) holds. Then the process $\{Y_t,\ t \in \mathbb{Z}\}$ is stationary ergodic with $\mathrm{E}|Y_t|_1 < \infty$. Moreover, if $\mathrm{E}|\xi_t|_a^a < \infty$ for some $a \geq 8$, then $\mathrm{E}|Y_t|_a^a < \infty$.*

The proof of Theorems 2.1–2.2 is given in Appendix A.1. Theorem 2.1 extends [6], Prop. 1, which was established for the linear PNAR model (4). Theorem 2.2 similarly extends [66], Theorem 1. In particular, existence of some moments for $\{\xi_t : t \in \mathbb{Z}\}$ guarantees the conclusions of Theorem 2.2. Such assumption is not necessary in the linear case considered by [66] (see their equation (2.1) because of the assumed normality.

For each $i = 1 \dots, N$, the contraction condition (9) follows by assuming that for $x_i, x_i^* \in \mathbb{R}_+$ and $y_i, y_i^* \in \mathbb{N}$,

$$\left|f_i(x_i, y_i) - f_i(x_i^*, y_i^*)\right| \leq \mu_1|x_i - x_i^*| + \mu_2|y_i - y_i^*|, \tag{10}$$

because the left-hand side of (10) is bounded by $\mu_1|\sum_{j=1}^N w_{ij}(y_j - y_j^*)| + \mu_2|y_i - y_i^*| \leq (\mu_1 w_i' + \mu_2 e_i')|y - y^*|_{\text{vec}}$, where $\mu_1 w_i' + \mu_2 e_i' = e_i' G$ is the ith row of the matrix G. Condition (10) is verified elementwise. When the nonlinear functions $f_i(\cdot)$ cannot be expressed in a vector form, for example, $f = (f_1, \dots, f_N)'$, verification of (10) is helpful; see (5)–(7). Moreover, the condition $\rho(G) < 1$ of Theorem 2.1 is implied by (10) when $\mu_1 + \mu_2 < 1$, because $\rho(G) \leq \|G\|_\infty \leq \mu_1 \|W\|_\infty + \mu_2 \leq \mu_1 + \mu_2$, since $\|W\|_\infty = 1$, by construction.

Theorem 2.2 follows again by (10) but with $|\mu_s|$, for $s = 1, 2$ and assuming that $|\mu_1| + |\mu_2| < 1$. Some illustrative examples are given below.

EXAMPLE 1 (continued). For model (4), $\lambda_t = \beta_0 1 + G Y_{t-1}$, with $G = \beta_1 W + \beta_2 I$. In this case, the sharp condition $\rho(G) < 1$ is easily verifiable, and under (9), it implies the results of Theorem 2.1. However, the assumptions of the theorem are also satisfied by the set of sufficient conditions (10) with $\mu_1 = \beta_1$, $\mu_2 = \beta_2$ and $\beta_1 + \beta_2 < 1$, for integer-valued processes. For the continuous-valued case, a similar argument shows that $|\beta_1| + |\beta_2| < 1$.

EXAMPLE 2 (continued). Consider model (5). By the mean value theorem (MVT),

$$\begin{aligned}\left|f(x_i, y_i) - f\left(x_i^*, y_i^*\right)\right| &\leq \max_{x_i \in \mathbb{R}_+}\left|\frac{\partial f(x_i, y_i)}{\partial x_i}\right|\left|x_i - x_i^*\right| + \max_{y_i \in \mathbb{N}}\left|\frac{\partial f(x_i, y_i)}{\partial y_i}\right|\left|y_i - y_i^*\right| \\ &\leq \beta_1^*\left|x_i - x_i^*\right| + \beta_2\left|y_i - y_i^*\right|,\end{aligned}$$

where $\beta_1^* = \max\{\beta_1, \beta_0\gamma - \beta_1\}$. Theorem 2.1 holds with $G = \beta_1^* W + \beta_2 I$ and $\beta_1^* + \beta_2 < 1$. Similar to model (5), by considering all the possible combinations of signs of x, β_0 and β_1 in model (6), we have $|\partial f(x_i, y_i)/\partial x_i| = |\beta_1 - \beta_0\gamma/(1 + x_i)^{\gamma+1} x_i/|x_i|| \leq \bar{\beta}_1 \equiv \max\{|\beta_1|, |\beta_0\gamma - \beta_1|, |\beta_1 - \beta_0\gamma|\}$. Theorem 2.2 holds with $G = \bar{\beta}_1 W + |\beta_2| I$ and $\bar{\beta}_1 + |\beta_2| < 1$.

EXAMPLE 3 (continued). In the integer-valued case, Theorem 2.1 applies to model (7) with $G = (\beta_1 + \alpha)W + \beta_2 I$ and $\beta_1 + \alpha + \beta_2 < 1$, which coincides with the stationarity condition developed for the standard STAR model [53]. By considering all the possible combinations of signs for β_1 and α, it is not difficult to show that Theorem 2.2 is verified, for model (7), under the similar sufficient condition $\beta_1^\star + |\beta_2| < 1$, where $\beta_1^\star = \max\{|\beta_1|, |\beta_1 + \alpha|\}$.

EXAMPLE 4 (continued). The threshold model (8) does not satisfy the contraction conditions (9)–(10). For the case of count data and N fixed, we develop a different proof to show that $\{Y_t\}$ is stationary and ergodic provided that it has a positive conditional probability mass function and $\|\|G\|\|_1 < 1$, where $G = (\beta_1 + \alpha_1)W + (\beta_2 + \alpha_2)I$. Analogous result holds also for continuous data; see [7], Section S-4.

2.3. *Stability conditions for increasing network size.* In this section, following the works by [66] and [6], we investigate the stability conditions of the process $\{Y_t \in E^N\}$, with $E = \mathbb{R}$ or $E = \mathbb{N}$, respectively, when the network size diverges ($N \to \infty$). We use a working definition of stationarity for increasing dimensional processes following [66], Definition 1; see [7], Section S-1.2.

THEOREM 2.3. *Consider model* (1) *and* $N \to \infty$. *Define* $G = \mu_1 W + \mu_2 I$, *where* μ_1, $\mu_2 \geq 0$ *are constants such that* $\mu_1 + \mu_2 < 1$ *and the contraction condition* (9) *holds*, *with* $\max_{i\geq 1} f_i(0, 0) < \infty$. *Then there exists a unique strictly stationary solution* $\{Y_t \in \mathbb{N}^N,\ t \in \mathbb{Z}\}$ *to the nonlinear PNAR model*, *with* $\max_{i\geq 1} \mathrm{E}|Y_{i,t}|^a \leq C_a < \infty$, *for any* $a \geq 1$.

THEOREM 2.4. *Consider model* (2) *and* $N \to \infty$. *Define* $G = |\mu_1|W + |\mu_2|I$, *where* μ_1, μ_2 *are real constants such that* $|\mu_1| + |\mu_2| < 1$ *and the contraction condition* (9) *holds*, *with* $\max_{i\geq 1}|f_i(0, 0)| < \infty$. *Then there exists a unique strictly stationary solution* $\{Y_t \in \mathbb{R}^N,\ t \in \mathbb{Z}\}$ *to the nonlinear NAR model. In addition*, *if* $\max_{i\geq 1} \mathrm{E}|\xi_{i,t}|^a \leq C_{\xi,a} < \infty$ *for some* $a \geq 8$, *then* $\max_{i\geq 1} \mathrm{E}|Y_{i,t}|^a \leq C_a < \infty$.

Theorems 2.3–2.4 (whose proof is given in Appendices A.2 and A.3) extend the increasing network-type results of [6], Theorem 1, and [66], Theorem 2, to nonlinear versions of the PNAR and NAR models, respectively. For models (4)–(8), $\max_{i\geq 1}|f_i(0, 0)| = \beta_0$. Moreover, the contraction condition (10), with $\mu_1 + \mu_2 < 1$ ($|\mu_1| + |\mu_2| < 1$), fulfils the conditions of Theorem 2.3 (Theorem 2.4), that is, we obtain identical sufficient conditions, which guarantee stationarity with fixed and diverging N. We emphasize again that (8) does not satisfy (9)–(10). This fact makes difficult to show stationarity, when N increases. More importantly, all stability results do not depend on the network structure, as specified by the matrix W, and on the data generating process describing the joint dependence.

3. Quasi-maximum likelihood inference. Consider model (3). Estimation for the unknown parameter vector θ is developed by means of QMLE. Define the quasi-log-likelihood function for θ by

$$l_{NT}(\theta) = \sum_{t=1}^{T}\sum_{i=1}^{N} l_{i,t}(\theta), \tag{11}$$

where $l_{i,t}(\theta)$ is the log-likelihood contribution of a single network node whose form depends on the type of data (discrete or continuous). Observe that (11) is not necessarily the *true* log-likelihood. The QMLE is denoted by $\hat{\theta}$ and maximizes (11). It is obtained by solving the system of equations $S_{NT}(\theta) = 0$, where

$$S_{NT}(\theta) = \frac{\partial l_{NT}(\theta)}{\partial \theta} = \sum_{t=1}^{T} s_{Nt}(\theta) \tag{12}$$

is the quasi-score function. Moreover, define the following matrices:

$$H_{NT}(\theta) = -\frac{\partial^2 l_{NT}(\theta)}{\partial \theta \partial \theta'}, \qquad B_{NT}(\theta) = \sum_{t=1}^{T} \mathrm{E}\big(s_{Nt}(\theta)s'_{Nt}(\theta)|\mathcal{F}_{t-1}\big), \tag{13}$$

as the sample Hessian matrix and the conditional information matrix, respectively. Henceforth, we drop the dependence on θ when a quantity is evaluated at the true value θ_0.

3.1. *Inference for PNAR models.* Consider model (1). In this case, the QMLE estimator, $\hat{\theta}$, maximizes

$$l_{NT}(\theta) = \sum_{t=1}^{T}\sum_{i=1}^{N} \big(Y_{i,t} \log \lambda_{i,t}(\theta) - \lambda_{i,t}(\theta)\big), \tag{14}$$

which is the log-likelihood obtained if all time series were contemporaneously independent. This simplifies computations allowing to establish consistency and asymptotic normality of the resulting estimator. It is worth noting that the joint copula structure, say $C(\cdot, \rho)$, with set of parameters ρ, do not enter into the maximization problem of the working log-likelihood (14). However, this does not imply that inference does not take into account dependence among observations. The corresponding score function is given by

$$S_{NT}(\theta) = \sum_{t=1}^{T}\sum_{i=1}^{N} \left(\frac{Y_{i,t}}{\lambda_{i,t}(\theta)} - 1\right)\frac{\partial \lambda_{i,t}(\theta)}{\partial \theta} = \sum_{t=1}^{T} s_{Nt}(\theta). \tag{15}$$

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 $i = 1, \dots, N$ and $\xi_t(\theta) = Y_t - \lambda_t(\theta)$ is a Martingale Difference Sequence (*MDS*) at $\theta = \theta_0$. Then the empirical Hessian and conditional information matrices are given, respectively, by

$$H_{NT}(\theta) = \sum_{t=1}^{T}\sum_{i=1}^{N} \frac{Y_{i,t}}{\lambda_{i,t}^2(\theta)} \frac{\partial \lambda_{i,t}(\theta)}{\partial \theta}\frac{\partial \lambda_{i,t}(\theta)}{\partial \theta'} - \sum_{t=1}^{T}\sum_{i=1}^{N} \left(\frac{Y_{i,t}}{\lambda_{i,t}(\theta)} - 1\right)\frac{\partial^2 \lambda_{i,t}(\theta)}{\partial \theta \partial \theta'}, \tag{16}$$

$$B_{NT}(\theta) = \sum_{t=1}^{T} \frac{\partial \lambda'_t(\theta)}{\partial \theta} D_t^{-1}(\theta)\Sigma_t(\theta)D_t^{-1}(\theta)\frac{\partial \lambda_t(\theta)}{\partial \theta'}, \tag{17}$$

where $\Sigma_t(\theta) = \mathrm{E}(\xi_t(\theta)\xi_t'(\theta)|\mathcal{F}_{t-1})$ is the conditional covariance matrix evaluated at θ. We impose the following standard assumptions:

A The parameter space Θ is compact and the true value θ_0 belongs to its interior.

B For $i = 1, \dots, N$, the function $f_i(\cdot)$ is three times differentiable with respect to θ and satisfies, for $x_i, x_i^* \in \mathbb{R}_+$ and $y_i, y_i^* \in \mathbb{N}$,

$$\left|\frac{\partial f_i(x_i, y_i, \theta)}{\partial \theta_g} - \frac{\partial f_i(x_i^*, y_i^*, \theta)}{\partial \theta_g}\right| \le c_{1g}|x_i - x_i^*| + c_{2g}|y_i - y_i^*|, \quad g = 1, \dots, m,$$

$$\left|\frac{\partial^2 f_i(x_i, y_i, \theta)}{\partial \theta_g \partial \theta_l} - \frac{\partial^2 f_i(x_i^*, y_i^*, \theta)}{\partial \theta_g \partial \theta_l}\right| \le c_{1gl}|x_i - x_i^*| + c_{2gl}|y_i - y_i^*|, \quad g, l = 1, \dots, m,$$

$$\left|\frac{\partial^3 f_i(x_i, y_i, \theta)}{\partial \theta_g \partial \theta_l \partial \theta_s} - \frac{\partial^3 f_i(x_i^*, y_i^*, \theta)}{\partial \theta_g \partial \theta_l \partial \theta_s}\right| \le c_{1gls}|x_i - x_i^*| + c_{2gls}|y_i - y_i^*|, \quad g, l, s = 1, \dots, m.$$

Furthermore, $\forall g, l, s$, $\max_{i\ge 1}|\partial f_i(0,0,\theta)/\partial\theta_g| < \infty$, $\max_{i\ge1}|\partial^2 f_i(0,0,\theta)/\partial\theta_g\partial\theta_l| < \infty$, $\max_{i\ge1}|\partial^3 f_i(0,0,\theta)/\partial\theta_g\partial\theta_l\partial\theta_s| < \infty$ and $\sum_g(c_{1g} + c_{2g}) < \infty$, $\sum_{g,l}(c_{1gl} + c_{2gl}) < \infty$, $\sum_{g,l,s}(c_{1gls} + c_{2gls}) < \infty$. In addition, the components of $\partial f_i/\partial\theta$ are linearly independent.

C For $i = 1, \dots, N$, $f_i(x_i, y_i, \theta) \ge C > 0$, where C is a generic constant.

Such regularity conditions have been employed in the literature to guarantee consistency and asymptotic normality of the QMLE in the context of nonlinear time series models; see [52], Chapter 3, among others. We now give additional assumptions employed for developing inference when $\{N, T_N\} \to \infty$ and the necessary network properties. Define

$$H_N(\theta) = \mathrm{E}\left(\frac{\partial\lambda_t'(\theta)}{\partial\theta} D_t^{-1}(\theta)\frac{\partial\lambda_t(\theta)}{\partial\theta'}\right), \tag{18}$$

$$B_N(\theta) = \mathrm{E}\left(\frac{\partial\lambda_t'(\theta)}{\partial\theta} D_t^{-1}(\theta)\Sigma_t(\theta)D_t^{-1}(\theta)\frac{\partial\lambda_t(\theta)}{\partial\theta'}\right), \tag{19}$$

as, respectively, (minus) the expected Hessian matrix and the information matrix. Consider the following assumptions.

H1 The process $\{\xi_t, \mathcal{F}_t : N \in \mathbb{N}, t \in \mathbb{Z}\}$ is α-mixing with mixing coefficients $\{\alpha(J)\}$.

H2 Define the standardized random process $\dot{Y}_t = D_t^{-1/2}(Y_t - \lambda_t)$. There exists a nonnegative, nonincreasing sequence $\{\varphi_h\}_{h=1,\dots,\infty}$ such that $\sum_{h=1}^{\infty} h\varphi_h = \Phi < \infty$, and for $i < j < k < l$, a.s.

$$\left|\mathrm{Cov}(\dot{Y}_{i,t}, \dot{Y}_{j,t}\dot{Y}_{k,t}\dot{Y}_{l,t}|\mathcal{F}_{t-1})\right| \le \varphi_{j-i}, \qquad \left|\mathrm{Cov}(\dot{Y}_{i,t}\dot{Y}_{j,t}\dot{Y}_{k,t}, \dot{Y}_{l,t}|\mathcal{F}_{t-1})\right| \le \varphi_{l-k},$$

$$\left|\mathrm{Cov}(\dot{Y}_{i,t}\dot{Y}_{j,t}, \dot{Y}_{k,t}\dot{Y}_{l,t}|\mathcal{F}_{t-1})\right| \le \varphi_{k-j}, \qquad \left|\mathrm{Cov}(\dot{Y}_{i,t}, \dot{Y}_{j,t}|\mathcal{F}_{t-1})\right| \le \varphi_{j-i}.$$

H3 For model (1) with network W, the following limits exist, at $\theta = \theta_0$:

H3.1 $\lim_{N\to\infty} N^{-1}H_N = H$, with H a $m \times m$ positive definite matrix.

H3.2 $\lim_{N\to\infty} N^{-1}B_N = B$.

H3.3 The third derivative of the quasi-log-likelihood (14) is bounded by functions $m_{i,t}$, which satisfy $\lim_{N\to\infty} N^{-1}\sum_{i=1}^{N}\mathrm{E}(m_{i,t}) = M$, where M is a finite constant.

Assumption **H1** is useful for studying processes with dependent errors [23]. When N is fixed, a combination of Theorems 1–2 in [24] and Remark 2.1 in [25] shows that the process $\{\xi_t : t \in \mathbb{Z}\}$ is α-mixing, with exponentially decaying coefficients, provided that $|||G|||_1 < 1$. Analogous conclusion follow by [30], Propositions 3.1–3.4. Condition **H2** represents a contemporaneous weak dependence assumption. Indeed, even in the simple case of the independence model, i.e. $\lambda_{i,t} = \beta_0$, for all $i = 1, \dots, N$, the reader can easily verify that, without any further constraints, $N^{-1}B_N = \mathcal{O}(N)$, so the limiting variance of the QMLE diverges. Note that **H2** does not guarantee finiteness of the Hessian and information matrices, as $N \to \infty$. Such requirement is imposed by Assumption **H3**. Obviously, such properties depend on the structure of W and on the functional form of $f(\cdot)$ in (1); without the knowledge of these components it cannot be simplified any further. We present a detailed example involving the nonlinear PNAR model (5) in Section 3.2 to offer further insight about **H3**. Proofs for all the following results are given in the Supplementary Material [7], Section S-2.

LEMMA 3.1. *Consider model* (1) *with* S_{NT}, H_{NT} *and* B_{NT} *defined as in* (15), (16) *and* (17), *respectively. Let* $\theta \in \Theta \subset \mathbb{R}^m_+$. *Suppose the conditions of Theorem* 2.3, *Assumption* ***B***–***C*** *and* ***H*1**–***H*3** *hold. Then, as* $\{N, T_N\} \to \infty$:

1. $(NT_N)^{-1} H_{NT_N} \xrightarrow{p} H$,
2. $(NT_N)^{-1} B_{NT_N} \xrightarrow{p} B$,
3. $(NT_N)^{-\frac{1}{2}} S_{NT_N} \xrightarrow{d} N(0, B)$,
4. $\max_{g,l,s} \sup_{\theta \in \mathcal{O}(\theta_0)} |\frac{1}{NT_N} \sum_{t=1}^{T_N} \sum_{i=1}^{N} \frac{\partial^3 l_{i,t}(\theta)}{\partial\theta_g \partial\theta_l \partial\theta_s}| \leq M_{NT_N} \xrightarrow{p} M$,

where $M_{NT_N} := (NT_N)^{-1} \sum_{t=1}^{T_N} \sum_{i=1}^{N} m_{i,t}$ *and* $\mathcal{O}(\theta_0) = \{\theta : |\theta - \theta_0|_2 < \delta\}$ *is a neighbourhood of* θ_0.

THEOREM 3.2. *For model* (1), *suppose that Assumption* ***A*** *and conditions of Lemma* 3.1 *hold. Then there exists a fixed open neighbourhood* $\mathcal{O}(\theta_0) = \{\theta : |\theta - \theta_0|_2 < \delta\}$ *of* θ_0 *such that with probability tending to* 1, *as* $\{N, T_N\} \to \infty$, *the equation* $S_{NT_N}(\theta) = 0$ *has a unique solution, denoted by* $\hat{\theta}$, *such that* $\hat{\theta} \xrightarrow{p} \theta_0$ *and* $\sqrt{NT_N}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, H^{-1}BH^{-1})$.

Theorem 3.2 follows by Lemma 3.1 as proved by [6], Section S-3.3. In addition, it extends the results of [6], Theorem 3, to nonlinear Poisson NAR models. The novelty of Theorem 3.2 is that both N and T tend to infinity as opposed to the standard case (when N is fixed). Additional conditions guarantee strong consistency of the estimators, that is, we have the following.

THEOREM 3.3. *If* $T_N = \lambda N$, *for some* $\lambda > 0$ *and Assumption* ***H*1** *is such that the mixing coefficients satisfy* $\alpha(J)^{1-1/r} = \mathcal{O}(J^{-3-\epsilon})$, *for some* $r > 2$ *and some* $\epsilon > 0$, *then as* $\{N, T_N\} \to \infty$, *all the convergences* "$\xrightarrow{p}$" *in Lemma* 3.1 *are replaced by* "$\xrightarrow{a.s.}$" *and Theorem* 3.2 *holds with* $\hat{\theta} \xrightarrow{a.s.} \theta_0$.

For instance, exponential decay of the mixing coefficients $\alpha(J)$ satisfies the assumption for any r and ϵ. Theorem 3.3 is a new result to the best of our knowledge as strong laws of large numbers for generally dependent double-indexed processes are scarce in the literature (for an exception, see [19]); see the discussion in [2], Comment 6 and [3], p. 256. It is pointed out again that the proof of Theorem 3.2 does not depend on the specification of the data generating process for the joint dependence of $\{Y_t\}$.

3.2. *A detailed example.* We give a detailed discussion for proving Theorem 3.2 for the nonlinear PNAR model (5) case. Let $\Sigma_\xi = \mathrm{E}|\xi_t \xi_t'|_{\mathrm{vec}}$ and $\lambda_{\max}(X)$ the largest absolute eigenvalue of an arbitrary symmetric matrix X. Consider the vector form of model (5): $\lambda_t = \beta_0 C_{t-1} + GY_{t-1}$, where $G = \beta_1 W + \beta_2 I$ and $C_{t-1} = (1 + X_{t-1})^{-\gamma}$. Under the conditions of Theorem 2.3, and by using a infinite backward substitution argument on Y_{t-1} we can rewrite the model as $Y_t = \mu_{t-1}^{\infty} + \tilde{Y}_t$, where $\mu_{t-1}^{\infty} = \beta_0 \sum_{j=0}^{\infty} G^j C_{t-1-j}$

and $\tilde{Y}_t = \sum_{j=0}^{\infty} G^j \xi_{t-j}$. The proof of such representation is given in the Supplementary Material [7], Section S-2.3. Define the following quantities: $L_{t-1} = \log(1 + X_{t-1})$, $E_{t-1} = C_{t-1} \odot L_{t-1}$, $F_{t-1} = \log^2(1 + X_{t-1}) \odot C_{t-1}$, $J_{t-1} = \log^3(1 + X_{t-1}) \odot C_{t-1}$, where $\odot$ is the Hadamard product [51], Section 11.7; $I_{1,t-1} = C_{t-1}$, $I_{2,t-1} = W\mu_{t-1}^{\infty}$, $I_{3,t-1} = \mu_{t-1}^{\infty}$, $\Lambda_t = \Sigma_t^{1/2} D_t^{-1}$, $\Gamma(0) = \mathrm{E}[\Lambda_t(Y_{t-1} - \mu_{t-1}^{\infty})(Y_{t-1} - \mu_{t-1}^{\infty})'\Lambda_t']$ and $\Delta(0) = \mathrm{E}[\Lambda_t W(Y_{t-1} - \mu_{t-1}^{\infty})(Y_{t-1} - \mu_{t-1}^{\infty})'W'\Lambda_t']$, $\Gamma_{1,t-1} = I_{1,t-1}$, $\Gamma_{2,t-1} = X_{t-1}$, $\Gamma_{3,t-1} = Y_{t-1}$, $\Gamma_{4,t-1} = E_{t-1}$; moreover, let $(j^*, l^*, k^*) = \arg\max_{j,l,k} |N^{-1} \sum_{i=1}^N \partial^3 l_{i,t}(\theta)/\partial\theta_j \partial\theta_l \partial\theta_k|$, $\Pi_{jlk} = N^{-1} \sum_{i=1}^N \mathrm{E}(\Gamma_{i,j,t-1}\Gamma_{i,l,t-1}\Gamma_{i,k,t-1}/\lambda_{i,t})$, $\Pi_{F,k} = N^{-1}\mathrm{E}(F_{t-1}' D_t^{-1} \Gamma_{k,t-1})$ and set $\Pi_J = N^{-1}\mathrm{E}(J_{t-1}' D_t^{-1} |\xi_t|_{\mathrm{vec}})$, for $j, l, k = 1, \dots, 4$. Consider the following assumptions.

Q1 Let W be a sequence of matrices with nonstochastic entries indexed by N.

Q1.1 Consider W as a transition probability matrix of a Markov chain, whose state space is defined as the set of all the nodes in the network (i.e., $\{1, \dots, N\}$). The Markov chain is assumed to be irreducible and aperiodic. Further, define $\pi = (\pi_1, \dots, \pi_N)' \in \mathbb{R}^N$ as the stationary distribution of the Markov chain, where $\pi \geq 0$, $\sum_{i=1}^N \pi_i = 1$ and $\pi = W'\pi$. Furthermore, assume that $\lambda_{\max}(\Sigma_\xi) \sum_{i=1}^N \pi_i^2 \to 0$ as $N \to \infty$.

Q1.2 Define $W^* = W + W'$ and assume $\lambda_{\max}(W^*) = \mathcal{O}(\log N)$ and $\lambda_{\max}(\Sigma_\xi) = \mathcal{O}((\log N)^\delta)$, for some $\delta \geq 1$.

Q2 Assume that the following limits exist: $l_{kj}^B = \lim_{N\to\infty} N^{-1}\mathrm{E}(I_{k,t-1}' \Lambda_t' \Lambda_t I_{j,t-1})$, for $k, j = 1, \dots, 4$, $u_1^B = \lim_{N\to\infty} N^{-1} \mathrm{tr}[\Delta(0)]$, $u_2^B = \lim_{N\to\infty} N^{-1} \mathrm{tr}[W\Gamma(0)]$, $u_3^B = \lim_{N\to\infty} N^{-1} \mathrm{tr}[\Gamma(0)]$, $v_{k4}^B = \lim_{N\to\infty} N^{-1}\mathrm{E}(\Gamma_{k,t-1}' \Lambda_t' \Lambda_t E_{t-1})$, $d^* = \lim_{N\to\infty} \Pi_{j^*,l^*,k^*}$. If at least two indices among (j^*, l^*, k^*) equal 4, $d_F^* = \lim_{N\to\infty} \Pi_{F,s^*}$, where $s^* = j^*, l^*, k^*$. Moreover, if all three (j^*, l^*, k^*) equal 4, $d_J^* = \lim_{N\to\infty} \Pi_J$.

THEOREM 3.4. *Consider* (5) *and suppose the conditions of Theorem* 2.3, *Assumptions* ***A*–*C***, ***H*1–*H*2** *and* ***Q*1–*Q*2** *hold. Then the conclusions of Theorem* 3.2 *hold true for model* (5), *with corresponding limiting matrices*:

$$H = \begin{pmatrix} l_{11}^H & l_{12}^H & l_{13}^H & -\beta_0 v_{14}^H \\ & l_{22}^H + u_1^H & l_{23}^H + u_2^H & -\beta_0 v_{24}^H \\ & & l_{33}^H + u_3^H & -\beta_0 v_{34}^H \\ & & & \beta_0^2 v_{44}^H \end{pmatrix},$$

$$B = \begin{pmatrix} l_{11}^B & l_{12}^B & l_{13}^B & -\beta_0 v_{14}^B \\ & l_{22}^B + u_1^B & l_{23}^B + u_2^B & -\beta_0 v_{24}^B \\ & & l_{33}^B + u_3^B & -\beta_0 v_{34}^B \\ & & & \beta_0^2 v_{44}^B \end{pmatrix}, \tag{20}$$

where the elements of the Hessian matrix are obtained by the elements of the information matrix with $\Sigma_t = D_t$.

REMARK 1. Clearly, the network structure influences the results of Theorem 3.4. Indeed, Assumption **Q1** requires a well-behaved underlying network: (i) there should exist a non-zero probability to connect each pair of nodes; this allows the network to converge to its stationary distribution, that is, $\lim_{N\to\infty} W^N = 1\pi'$; (ii) The growth of the network should be such that certain regularity properties hold. For instance, the covariances of the errors do not diverge fast, as $N \to \infty$. The proof in [7], Section S-2.3, shows that the leading terms of Hessian and information matrices depend on the error component ξ_{t-j} and the pseudo covariance matrix Σ_ξ and are asymptotically negligible (compare also with [6], Lemma S-1). In this way, the remaining terms appearing in Assumption **Q2** show existence of the limiting Hessian matrix H and (together with Assumption **H2**) of the limiting information B. Without any assumptions for the network, the structure of matrices H and B is unknown and conditions of finiteness of the limiting matrices could not be specified explicitly.

3.3. *Inference for NAR models.* In this case, define $\hat{\theta}$, as the maximizer of the least squares criterion

$$l_{NT}(\theta) = -\sum_{t=1}^{T} (Y_t - \lambda_t(\theta))'(Y_t - \lambda_t(\theta)). \tag{21}$$

It follows that

$$S_{NT}(\theta) = \sum_{t=1}^{T} \frac{\partial \lambda_t'(\theta)}{\partial \theta} (Y_t - \lambda_t(\theta)) = \sum_{t=1}^{T} s_{Nt}(\theta). \tag{22}$$

The empirical Hessian and information matrices are respectively

$$H_{NT}(\theta) = \sum_{t=1}^{T}\sum_{i=1}^{N} \frac{\partial \lambda_{i,t}(\theta)}{\partial \theta}\frac{\partial \lambda_{i,t}(\theta)}{\partial \theta'} - \sum_{t=1}^{T}\sum_{i=1}^{N} (Y_{i,t} - \lambda_{i,t}(\theta)) \frac{\partial^2 \lambda_{i,t}(\theta)}{\partial \theta \partial \theta'}, \tag{23}$$

$$B_{NT}(\theta) = \sum_{t=1}^{T} \frac{\partial \lambda_t'(\theta)}{\partial \theta} \Sigma_t(\theta) \frac{\partial \lambda_t(\theta)}{\partial \theta'}, \tag{24}$$

where notation is as in Section 3.1. In addition,

$$H_N(\theta) = \mathrm{E}\left(\frac{\partial \lambda_t'(\theta)}{\partial \theta}\frac{\partial \lambda_t(\theta)}{\partial \theta'}\right), \tag{25}$$

$$B_N = \mathrm{E}\left(\frac{\partial \lambda_t'(\theta)}{\partial \theta} \xi_t(\theta)\xi_t'(\theta) \frac{\partial \lambda_t(\theta)}{\partial \theta'}\right), \tag{26}$$

and the latter equals $\sigma^2 H_N$, when $\theta = \theta_0$, because ξ_t an is *IID*$(0, \sigma^2)$ process. For the same reasons Assumption **H1**–**H2** hold trivially. Assumption **H3** is modified as the following:

H3′ For model (2) with network W, the following limits exist, at $\theta = \theta_0$:

H3′.1 $\lim_{N\to\infty} N^{-1} H_N = H$, with H a $m \times m$ positive definite matrix.

H3′.2 The third derivative of the quasi-log-likelihood (21) is bounded by functions $m_{i,t}$, which satisfy $\lim_{N\to\infty} N^{-1}\sum_{i=1}^{N} \mathrm{E}(m_{i,t}) = M$, where M is a finite constant.

THEOREM 3.5. *Consider model* (2) *with* S_{NT}, H_{NT} *and* B_{NT} *defined as in* (22), (23) *and* (24), *respectively. Let* $\theta \in \Theta \subset \mathbb{R}^m$. *Suppose that the conditions of Theorem* 2.4, *Assumptions* ***A***–***B*** *and* ***H3′*** *hold. Then there exists a fixed open neighbourhood* $\mathcal{O}(\theta_0) = \{\theta : |\theta - \theta_0|_2 < \delta\}$ *of* θ_0 *such that with probability tending to* 1 *as* $\{N, T_N\} \to \infty$, *the equation* $S_{NT_N}(\theta) = 0$ *has a unique solution, denoted by* $\hat{\theta}$, *such that* $\hat{\theta} \xrightarrow{p} \theta_0$ *and* $\sqrt{NT_N}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, B^{-1})$, *where* $B = \sigma^2 H$ *and* H *is defined as in* (20), *with* $\Sigma_t = D_t = I$.

The proof is omitted since it is analogous to the proof of Theorem 3.2. Theorem 3.5 generalises the results of [66], Theorem 3, to nonlinear NAR models, and it can be proved to entail results analogous to Proposition 3.4, by considering [66], Assumption C2, instead of **Q1**, and **Q2** holding, with $D_t = \Lambda_t = I$ and $\Pi_{jlk} = 0$, for $j, k, l = 1, \dots, 4$; see [66], Theorem 3, for a detailed proof concerning the case of model (4). A result similar to Theorem 3.3 for model (2) is also established by setting $T_N = \lambda N$.

REMARK 2. Reiterating the discussion following Theorems 2.3–2.4 and noting that Assumption **B** does not hold for (8), the double asymptotic based inference derived in this section and the associated testing theory (see Section 5) do not hold for threshold models when N is increasing. However, the Supplementary Material ([7], Section S-4) provides all these results for the threshold model if N is fixed.

REMARK 3. The asymptotic theory of this section applies for parameter values satisfying the conditions of Theorems 2.3–2.4. In practical applications, the QMLE is obtained using constrained optimization where the constraints satisfy such conditions. In the integer-valued case, additional constraints should be introduced so that the mean process is positive.

4. Hypothesis testing on network autoregressive models. With the same notation as in Sections 2 and 3, recall (3) and consider the following testing problems:

$$H_0 : \theta^{(2)} = \theta_0^{(2)} \quad \text{vs.} \quad H_1 : \theta^{(2)} \neq \theta_0^{(2)}, \quad \text{componentwise,} \tag{27}$$

against the Pitman's local alternatives

$$H_0 : \theta^{(2)} = \theta_0^{(2)} \quad \text{vs.} \quad H_1 : \theta^{(2)} = \theta_0^{(2)} + \frac{\delta_2}{\sqrt{NT}}, \quad \delta_2 \in \mathbb{R}^{m_2}. \tag{28}$$

To develop a test statistic for testing (27)–(28), we employ a quasi-score test based on (11). An appealing property of the score test is that it is computed under the null, which is computationally simpler. Moreover, the asymptotic distribution of the test is not affected when $\theta^{(2)}$ belongs to the boundary of the parameter space. Define $\tilde{\theta} = (\tilde{\theta}^{(1)\prime}, \tilde{\theta}^{(2)\prime})'$ the constrained quasi-likelihood estimator of $\theta = (\theta^{(1)\prime}, \theta^{(2)\prime})'$, under the null hypothesis, and $S_{NT}(\theta) = (S_{NT}^{(1)\prime}(\theta), S_{NT}^{(2)\prime}(\theta))'$ denote the corresponding partition of the quasi-score function. Because we study a quasi-score test, we correct the test statistic to obtain thoroughly its limiting distribution; see [33], among others. Accordingly, the test statistic is given by [7], Section S-3,

$$LM_{NT} = S_{NT}^{(2)\prime}(\tilde{\theta}) \Sigma_{NT}^{-1}(\tilde{\theta}) S_{NT}^{(2)}(\tilde{\theta}). \tag{29}$$

Here, $(NT)^{-1}\Sigma_{NT}(\tilde{\theta})$ is a suitable estimator for the covariance matrix defined as $\Sigma = \mathrm{Var}[(NT)^{-1/2} S_{NT}^{(2)}(\tilde{\theta})]$, where

$$\Sigma = B_{22} - H_{21}H_{11}^{-1}B_{12} - B_{21}H_{11}^{-1}H_{12} + H_{21}H_{11}^{-1}B_{11}H_{11}^{-1}H_{12}, \tag{30}$$

with $B_{gl}, H_{gl}, g, l = 1, 2$ with dimension $m_g \times m_l$, are blocks of the matrices H, B such that

$$H = \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix}, \qquad B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}.$$

If (11) is the *true* likelihood, then LM_{NT} reduces to the standard score test with $B \equiv H$ and $\Sigma = B_{22} - B_{21}B_{11}^{-1}B_{12} =: \Sigma_B$.

REMARK 4. Following [6], the estimator $\Sigma_{NT}(\tilde{\theta})$ of (29) is computed as the sample counterpart of (30), obtained by replacing the partitioned matrices H and B, respectively, by $H_{NT}(\tilde{\theta})$ and $B_{NT}(\tilde{\theta})$, where $H_{NT}(\theta)$ is defined in (13) and $B_{NT}(\theta)$ is the sample information matrix.

A′ The parameter space Θ is compact. Define the partition of the parameter space $\Theta^{(1)}$ such that $\theta^{(1)} \in \Theta^{(1)}$ and the true value $\theta_0^{(1)}$ belongs to its interior.

THEOREM 4.1. *Suppose that model* (3) *admits a stationary solution, for* $N \to \infty$. *Consider* l_{NT}, S_{NT}, H_{NT} *and* B_{NT} *defined by* (11), (12) *and* (13), *respectively. Assume that, under* H_0, $\boldsymbol{A}'$ *is satisfied such that, as* $\{N, T_N\} \to \infty$, *Lemma* 3.1 *and Theorem* 3.2 *holds. Recall the testing problem* (27). *Then, as* $\{N, T_N\} \to \infty$, *the quasi-score test statistic* (29) *converges to a chi-square random variable,*

$$LM_{NT_N} \xrightarrow{d} \chi^2_{m_2},$$

under H_0. *Moreover, under the alternative* (28), (29) *converges to a noncentral chi-square random variable,*

$$LM_{NT_N} \xrightarrow{d} \chi^2_{m_2}(\delta_2' \tilde{\Delta} \delta_2),$$

where $\tilde{\Delta} = \tilde{\Sigma}_H \tilde{\Sigma}^{-1} \tilde{\Sigma}_H$ *and* $\Sigma_H := H_{22} - H_{21} H_{11}^{-1} H_{12}$; $\tilde{\Sigma}$ *and* $\tilde{\Sigma}_H$ *are sample counterparts of* Σ *and* Σ_H, *respectively, evaluated at* $\tilde{\theta}$.

Theorem 4.1 extends the results of [13] for the case of multivariate discrete and continuous network autoregressive models with infinite-dimensional data. In addition, it implies that even though $\theta^{(2)}$ belongs to the boundary of the parameter space, the asymptotic χ^2 distribution remains unaffected. Instead, the asymptotic distribution of the Wald and likelihood ratio tests depends on the null hypothesis and do not converge to χ^2 distributed when N is fixed; see [33], Section 8.3.2 and [1]. We illustrate some applications of Theorem 4.1 to the network models (1)–(2) but we emphasize that its conclusion applies to more general settings.

PROPOSITION 4.2. *Assume* Y_t *follows* (1) *and the process* λ_t *is defined as in* (3). *Consider the test* $H_0 : \theta^{(2)} = \theta^{(2)}_0$ *versus* $H_1 : \theta^{(2)} \neq \theta^{(2)}_0$. *Then, under* H_0, $\boldsymbol{A}'$ *and the conditions of Lemma* 3.1, *Theorem* 4.1 *is true.*

Proposition 4.2 follows by Lemma 3.1, Theorems 2.3 and 4.1.

4.1. *A detailed example* (*Continued*). For model (5), the linearity test (27) is equivalent to testing $H_0 : \gamma = 0$ versus $H_1 : \gamma > 0$. Convergence for all necessary asymptotic quantities is required only under the null. Recall the notation of Section 3.2. Then, under H_0, $C_{t-1} = 1$, the decomposition of the count process simplifies to $Y_t = \mu + \sum_{j=0}^{\infty} G^j \xi_{t-j}$, since $\mu_{t-1} = \mu = \beta_0/(1 - \beta_1 - \beta_2)^{-1}$. This entails that $\Gamma_{1,t-1} = 1$, $\Gamma_{4,t-1} = L_{t-1}$, $F_{t-1} = \log^2(1 + X_{t-1})$ and $J_{t-1} = \log^2(1 + X_{t-1})$. Moreover, set $\Lambda = \mathrm{E}(\Lambda_t' \Lambda_t)$, $\Gamma(0) = \mathrm{E}[\Lambda_t (Y_{t-1} - \mu)(Y_{t-1} - \mu)' \Lambda_t']$ and $\Delta(0) = \mathrm{E}[\Lambda_t W (Y_{t-1} - \mu)(Y_{t-1} - \mu)' W' \Lambda_t']$. So, condition **Q2** simplifies as follows:

Q2$'$ Assume that the following limits exist: $f_1 = \lim_{N\to\infty} N^{-1}(1'\Lambda 1)$, $f_2 = \lim_{N\to\infty} N^{-1} \mathrm{tr}[\Gamma(0)]$, $f_3 = \lim_{N\to\infty} N^{-1} \mathrm{tr}[W\Gamma(0)]$, $f_4 = \lim_{N\to\infty} N^{-1} \mathrm{tr}[\Delta(0)]$, $\dot{v}^B_{k4} = \lim_{N\to\infty} N^{-1} \mathrm{E}(\Gamma'_{k,t-1} \Lambda_t' \Lambda_t L_{t-1})$, $d^* = \lim_{N\to\infty} \Pi_{j^*,l^*,k^*}$. If at least two indices among (j^*, l^*, k^*) equal 4, $d^*_F = \lim_{N\to\infty} \Pi_{F,s^*}$, where $s^* = j^*, l^*, k^*$. Moreover, if all three indices (j^*, l^*, k^*) equal 4, $d^*_J = \lim_{N\to\infty} \Pi_J$.

In this case, the limiting Hessian and information matrices in (20) are equal to the respective matrices obtained by the linear model fitting [6], equation (22), plus the addition of the fourth row and column whose elements are given by $(-\beta_0)^\nu \dot{v}^B_{k4}$, for $k = 1, \dots, 4$ and $\nu = 2$, when $k = 4$ and $\nu = 1$, otherwise.

PROPOSITION 4.3. *Assume* Y_t *follows* (1) *and the process* λ_t *is defined as in* (5). *Suppose the conditions of Theorem* 2.3, *Assumptions* $\boldsymbol{A}'$, ***B***–***C***, ***H*1**–***H*2** *and* ***Q*1**–***Q*2**$'$ *hold. Consider the test* $H_0 : \gamma = 0$ *versus* $H_1 : \gamma > 0$. *Then Theorem* 4.1 *holds true.*

Denoting the constrained QMLE by $\tilde{\theta} = (\tilde{\theta}^{(1)}, 0)'$, where $\tilde{\theta}^{(1)}$ is the QMLE of the linear model (4), the partial quasi-score (29) is given by $S_{NT}^{(2)}(\tilde{\theta}) = \sum_{t=1}^{T}\sum_{i=1}^{N}(Y_{i,t}/\lambda_{i,t}(\tilde{\theta}) - 1)\partial\lambda_{i,t}(\tilde{\theta})/\partial\gamma$, with $\partial\lambda_{i,t}(\tilde{\theta})/\partial\gamma = -\tilde{\beta}_0\log(1 + X_{i,t-1})$, where $\tilde{\beta}_0$ is the QMLE of the intercept β_0 in the linear model (4). Furthermore, the covariance estimator $\Sigma_{NT}(\tilde{\theta})$ for the test statistic (29) is defined as in Remark 4. Note that in Proposition 4.3, the nonlinear perturbation is due to the network structure. Moreover, since the asymptotic distribution of the score test (29) depends on the convergence of sample Hessian and information matrices to (20), the approximation to the chi-square distribution depends by the convergence of the network according to the regularity properties given by **Q1**–**Q2**$'$ (see Remark 1). The analogous result and conclusions are obtained for (2), by using Theorems 2.4, 3.5 and 4.1 and, therefore, it is omitted. Consider the following condition:

Q2$''$ For Y_t defined as in (2) and λ_t following (6), Assumption **Q2**$'$ holds, with $D_t = \Lambda_t = I$ and $\Pi_{jlk} = 0$, for $j, k, l = 1, \dots, 4$.

PROPOSITION 4.4. *Assume Y_t follows* (2) *and the process λ_t is defined as in* (6). *Suppose the conditions of Theorem* 2.4, *Assumptions **A**$'$–**B**, **H1**–**H2**,* [66], *Assumption C2, and **Q2**$''$ hold. Consider the test $H_0 : \gamma = 0$ versus $H_1 : \gamma > 0$. Then Theorem* 4.1 *holds true.*

5. Testing under nonidentifiable parameters. We develop testing theory when the parameters are not identifiable under the linearity hypothesis. A case in point is model (7), with $\theta^{(1)} = (\beta_0, \beta_1, \beta_2)'$ and $\theta^{(2)} = (\alpha, \gamma)'$. Then testing $H_0 : \alpha = 0$, makes γ nonidentifiable but the score partition (15)—and consequently the test statistic—still depends on the value of γ. Hence, the theory of Section 4 does not apply any more. Similar remarks hold for the threshold parameter γ of model (8), when testing $H_0 : \alpha_0 = \alpha_1 = \alpha_2 = 0$. Assigning a fixed arbitrary value for γ resolves such issues but this approach might lack power as the test is sensitive to the choice of γ, especially when γ is far from its true value. It is well known (see [55], Section 5.1,5.5) that testing linearity is an important issue because nonidentifiable parameters have tremendous impact on properties of estimators. Usually a sup-type test, say $g_{NT} = \sup_{\gamma\in\Gamma} LM_{NT}(\gamma)$, is employed in applications, where $\Gamma = [\gamma_L, \gamma_U]$ is a compact domain for γ; for example, [15, 32] and [13], paragraph 3.2.

Define Z a random variable, and suppose that the function $g(\cdot) : \Gamma \to \mathbb{R}$ is continuous with respect to the uniform metric, monotonic for each γ, and such that, as $Z \to \infty$, then $g(Z) \to \infty$ in a subset of Γ with a nonzero probability $\mathbb{P}$. For the standard asymptotics, that is, $T \to \infty$, such functions have been employed in applications. Examples include $g_T = g(LM_T)$ [38] and [5] who considered $g(LM_T) = \int_\Gamma LM_T(\gamma)d\mathbb{P}(\gamma)$ and $g(LM_T) = \log(\int_\Gamma \exp(1/2LM_T(\gamma))d\mathbb{P}(\gamma))$. We extend this theory to the case of both $T, N \to \infty$.

5.1. *Specification.* In this section, we use a more convenient notation. Accordingly, consider the nonlinear PNAR model defined in (1) as

$$Y_t = N_t(\lambda_t(\gamma)), \quad \lambda_t(\gamma) = Z_{1t}(W)\beta + h(Y_{t-1}, W, \gamma)\alpha, \tag{31}$$

where β is a $k_1 \times 1$ vector of identifiable parameters associated with the linear component of the model, α is a $k_2 \times 1$ vector of identifiable nonlinear parameters and γ denote nuisance parameters. We set $\theta = (\phi', \gamma')'$, $\phi = (\beta', \alpha')'$. With this notation, the dimension of θ is $m = k + m^*$, where m^* is the dimension of γ and $k = k_1 + k_2$. In addition, $Z_{1t}(W) = (1, WY_{t-1}, Y_{t-1})$ is a $N \times k_1$ matrix associated to the linear part of the network autoregressive model (for the order 1 model $k_1 = 3$), and $h(Y_{t-1}, W, \gamma)$ is a $N \times k_2$ matrix describing the nonlinear part of the model. Set $Z_{1t} = Z_{1t}(W)$, $h_t(\gamma) \equiv h(Y_{t-1}, W, \gamma)$ and $h_t(\gamma) = (h_t^1(\gamma) \dots h_t^b(\gamma) \dots h_t^{k_2}(\gamma))$, where each column indicates a nonlinear regressor $h_t^b(\gamma)$, for $b = 1, \dots, k_2$, being a $N \times 1$ vector whose elements are $h_{i,t}^b(\gamma)$, where

$i = 1, \ldots, N$. Then the conditional expectation of (31) is $\lambda_t(\gamma) = Z_t(\gamma)\phi$ where $Z_t(\gamma) = (Z_{1t}, h_t(\gamma))$ is the $N \times k$ matrix of regressors. Analogously, for continuous-valued time series and $\xi_t \sim IID(0, \sigma^2)$, equation (2) becomes

$$Y_t = \lambda_t(\gamma) + \xi_t, \quad \lambda_t(\gamma) = Z_{1t}(W)\beta + h(Y_{t-1}, W, \gamma)\alpha. \tag{32}$$

Many nonlinear models are included in this general frameworks provided by (31)–(32); for example, the STNAR model (7), where $k_2 = 1$ and $h_{i,t}(\gamma) = \exp(-\gamma X_{i,t-1}^2)X_{i,t-1}$, for $i = 1, \ldots, N$, and the TNAR model (8), where $k_2 = 3$ and $h_{i,t}^1(\gamma) = I(X_{i,t-1} \leq \gamma)$, $h_{i,t}^2(\gamma) = X_{i,t-1}I(X_{i,t-1} \leq \gamma)$ and $h_{i,t}^3(\gamma) = Y_{i,t-1}I(X_{i,t-1} \leq \gamma)$; see [38], p. 414.

5.2. *Testing linearity.* For models (31)–(32), consider testing linearity in the presence of a non identifiable parameters γ,

$$H_0 : \alpha = 0, \quad \text{vs.} \quad H_1 : \alpha \neq 0, \quad \text{elementwise.} \tag{33}$$

Consider first the case of count time series, that is, equation (31). In this case, the score (15), Hessian (16) and the sample information matrix (17), for the quasi-log-likelihood (14) are $S_{NT}(\gamma) = \sum_{t=1}^T s_{Nt}(\gamma)$, $H_{NT}(\gamma_1, \gamma_2) = \sum_{t=1}^T \sum_{i=1}^N Y_{i,t}\bar{Z}_{i,t}(\gamma_1)\bar{Z}_{i,t}'(\gamma_2)$ and $B_{NT}(\gamma_1, \gamma_2) = \sum_{t=1}^T \mathrm{E}[s_t(\gamma_1)s_t'(\gamma_2)|\mathcal{F}_{t-1}]$, where $s_{Nt}(\gamma) = Z_t'(\gamma)D_t^{-1}(\gamma)(Y_t - Z_t(\gamma)\phi)$ and $\bar{Z}_{i,t}(\gamma) = Z_{i,t}(\gamma)/\lambda_{i,t}(\gamma)$. The theoretical counterpart of such quantities are then denoted by $H_N(\gamma_1, \gamma_2) = \sum_{i=1}^N \mathrm{E}(Y_{i,t}\bar{Z}_{i,t}(\gamma_1)\bar{Z}_{i,t}'(\gamma_2))$, $H(\gamma_1, \gamma_2) = \lim_{N\to\infty} N^{-1}H_N(\gamma_1, \gamma_2)$ and $B_N(\gamma_1, \gamma_2) = \mathrm{E}(s_t(\gamma_1)s_t'(\gamma_2))$, $B(\gamma_1, \gamma_2) = \lim_{N\to\infty} N^{-1}B_N(\gamma_1, \gamma_2)$. Following the discussion of Section 4, the quasi-score function is partitioned again in two components: the part concerning linear parameters and the component associated with the nonlinear part of the model. We denote this by $S_{NT}(\gamma) = (S_{NT}^{(1)\prime}, S_{NT}^{(2)\prime}(\gamma))'$. Moreover, consider $S(\gamma) = (S^{(1)\prime}, S^{(2)\prime}(\gamma))'$ a mean zero Gaussian process with covariance kernel $B(\gamma_1, \gamma_2)$. Define the matrix $\Sigma(\gamma_1, \gamma_2)$ as in (30), with partitioned matrices B_{gl}, H_{gl}, for $g, l = 1, 2$, of dimension $k_g \times k_l$, being blocks of the matrices $B(\gamma_1, \gamma_2)$, $H(\gamma_1, \gamma_2)$, with obvious rearrangement of the notation. Then $S^{(2)}(\gamma)$ is a Gaussian process with covariance kernel $\Sigma(\gamma_1, \gamma_2)$. Define $\tilde{\phi} = (\tilde{\beta}', 0')'$ the constrained estimator under the null hypothesis and use the tilde notation for all quantities, which correspond to constrained QMLE. Then, for testing (33), we consider the test statistic

$$LM_{NT}(\gamma) = \tilde{S}_{NT}^{(2)\prime}(\gamma)\tilde{\Sigma}_{NT}^{-1}(\gamma, \gamma)\tilde{S}_{NT}^{(2)}(\gamma), \tag{34}$$

where, according to Remark 4, $\tilde{\Sigma}_{NT}(\gamma, \gamma)$ is the estimator of $\Sigma(\gamma, \gamma)$, obtained by substituting $H(\gamma, \gamma)$, $B(\gamma, \gamma)$ with $\tilde{H}_{NT}(\gamma, \gamma)$, $\tilde{B}_{NT}(\gamma, \gamma)$, respectively.

Define $Z_{1,i,t} = (1, X_{i,t-1}, Y_{i,t-1})'$ and $\eta_{Nt} = N^{-1/2}\sum_{i=1}^N Y_{i,t}(Z_{1,i,t}'Z_{1,i,t} - 1) + X_{1,i,t} + Y_{1,i,t}$. An extra condition is required.

B$'$ Assumption **B** holds with all constants not depending on $\gamma \in \Gamma$, where Γ is compact, and $\|\eta_{Nt}\|_q < \infty$, for some $q > \max\{1 + \delta, m^*\}$, with $0 < \delta < 1$.

Assumption **B**$'$ is similar to assumption **B** for the particular case we consider. An extra moment assumption is required to guarantee stochastic equi-continuity of the score. It can be easily shown that a sufficient condition for obtaining $\|\eta_{Nt}\|_q < \infty$ would be, for example, the weak dependence condition $|\mathrm{E}(Y_{i,t}^r Y_{j,t}^r|\mathcal{F}_{t-1})| \leq \phi_{j-i}$, such that $\sum_{h=1}^\infty \phi_h^{1/r} < \infty$, where $r = q/2$, if q is even, and $r = (q+1)/2$, if q is odd. For instance, in the STNAR model (7), $m^* = 1$, $q = 2$ and $r = 1$, so the condition simplifies to a special case of Assumption **H2**. From **B**$'$, Assumption **C** holds trivially for (31), because a.s. $\lambda_{i,t}(\gamma) \geq \beta_0 + h_i'(0, \gamma)\alpha = C > 0$, for $i = 1, \ldots, N$. Define $\delta_2 \in \mathbb{R}_+^{k_2}$ and $J_2 = (O_{k_2\times k_1}, I_{k_2})$, where I_s is a $s \times s$ identity matrix and $O_{a\times b}$ is a $a \times b$ matrix of zeros.

THEOREM 5.1. *Assume Y_t is integer-valued, following* (31) *and suppose the conditions of Theorem* 2.3, *Assumption* $\boldsymbol{A}'$–$\boldsymbol{B}'$ *and* $\boldsymbol{H1}$–$\boldsymbol{H3}$ *hold. Consider the test* $H_0 : \alpha = 0$ *versus* $H_1 : \alpha > 0$, *componentwise. Then, under* H_0, *as* $\{N, T_N\} \to \infty$, $S_{NT_N}(\gamma) \Rightarrow S(\gamma)$, $LM_{NT_N}(\gamma) \Rightarrow LM(\gamma)$ *and* $g_{NT_N} \Rightarrow g = g(LM(\gamma))$ *where*

$$LM(\gamma) = S^{(2)\prime}(\gamma)\Sigma^{-1}(\gamma,\gamma)S^{(2)}(\gamma).$$

Moreover, the same result holds under local alternatives $H_1 : \alpha = (NT_N)^{-1/2}\delta_2$, *with* $S^{(2)}(\gamma)$ *having mean* $J_2H^{-1}(\gamma,\gamma)J_2'\delta_2$.

Theorem 5.1 (the proof is given in Appendix A.4) extends [38], Theorem 1, in three directions: (i) develops testing for NAR models; (ii) proves convergence to asymptotic process, where both time and network dimension diverge together; (iii) the results hold for both continuous-valued data (see below) and integer-valued multivariate random variables. In line with Section 4.1, for each single model encompassed in (31) one can substitute **H3** with network conditions **Q1** and suitable limits existence as in **Q2**$'$. An analogous result holds for continuous-valued time series, as in (32). Its proof is omitted. Consider $s_t(\gamma) = Z_t'(\gamma)\xi_t$ and $H_T(\gamma_1,\gamma_2) = \sum_{t=1}^{T}\sum_{i=1}^{N} Z_{i,t}(\gamma_1)Z_{i,t}'(\gamma_2)$. In this case, no additional weak dependence assumption is required since the error sequence is independent.

B$''$ Assumption **B** holds with all constants not depending on $\gamma \in \Gamma$, where Γ is compact.

THEOREM 5.2. *Assume Y_t is continuous-valued, following* (32) *and suppose the conditions of Theorem* 2.4, *Assumptions* $\boldsymbol{A}'$–$\boldsymbol{B}''$ *and* $\boldsymbol{H3}'$ *hold. Consider the test* $H_0 : \alpha = 0$ *versus* $H_1 : \alpha \neq 0$, *componentwise. Then the results of Theorem* 5.1 *hold true.*

REMARK 5. The results of this paper extend straightforwardly to the case of model order $p > 1$, that is, $\lambda_t = f(Y_{t-1}, \dots, Y_{t-p}, W, \theta)$. Indeed, the proof of stability conditions of Theorems 2.1–2.2 are based on the fact that the process $\{Y_t : t \in \mathbb{Z}\}$ is a first-order Markov chain. All proofs adapt directly to a Markov chain of generic order p, by suitable adjustment of the contraction property (9). A similar remark holds for asymptotic properties of the QMLE and Theorems 3.2–5.2 by a suitable extension.

5.3. *Computations of p-values.* The null distribution of the process $g(\cdot)$ cannot be tabulated in general, apart from special cases; see [4]. To overcome this obstacle, we consider two different approaches. Consider the sup-type test, $g = \sup_{\gamma\in\Gamma}(LM(\gamma))$. By Theorems 5.1–5.2, under H_0, $LM(\gamma)$ is a chi-square process with k_2 degrees of freedom. If the nuisance parameter γ is scalar, [15] proves that the p-value of the sup-test is approximately bounded by

$$\mathrm{P}\Big[\sup_{\gamma\in\Gamma_F}\big(LM(\gamma)\big) \geq M\Big] \leq \mathrm{P}\big(\chi^2_{k_2} \geq M\big) + VM^{\frac{1}{2}(k_2-1)}\frac{\exp(-\frac{M}{2})2^{-\frac{k_2}{2}}}{\Gamma(\frac{k_2}{2})}, \tag{35}$$

where M is the maximum of the test statistic $LM_{NT}(\gamma)$, with $\gamma \in \Gamma_F$ and $\Gamma_F = (\gamma_L, \gamma_1, \dots, \gamma_l, \gamma_U)$ is a grid of values for Γ. The quantity V is the approximated total variation, defined by

$$V = \Big|LM_{NT}^{\frac{1}{2}}(\gamma_1) - LM_{NT}^{\frac{1}{2}}(\gamma_L)\Big| + \cdots + \Big|LM_{NT}^{\frac{1}{2}}(\gamma_U) - LM_{NT}^{\frac{1}{2}}(\gamma_l)\Big|.$$

Such method is attractive because of its simplicity and its computational speed. This last point is of great importance in network models, especially when the dimension N is large. However, the method suffers from three main drawbacks. First, (35) leads to a conservative test, because usually the p-values are smaller than their bound. Second, the results of [15] hold only for scalar nuisance parameters. Though this observation applies to several models discussed so far, like the STNAR model (7), more complex models may require inclusion of more than one nuisance parameter. Finally, (35) cannot be applied to the TNAR model (8), because $LM(\gamma)$, under the null hypothesis, has to be differentiable [15], p. 36, [38], Section 4. Following [38], we develop a bootstrap method based on stochastic permutations.

Define $F(\cdot)$, the distribution function of the process g with $p_{NT} = 1 - F(g_{NT})$. From Theorem 5.1 and the Continuous Mapping Theorem (CMT), $p_{NT} \Rightarrow p$, where $p = 1 - F(g)$ and $p \sim U(0, 1)$, under the null. Hence, the test rejects H_0 if $p_{NT} \leq a_{H_0}$, where a_{H_0} is the asymptotic size of the test. Define $\{\nu_t : t = 1, \dots, T\} \sim \text{IIDN}(0, 1)$, such that $\tilde{S}^{\nu}_{NT}(\gamma) = \sum_{t=1}^{T} \tilde{s}^{\nu}_{Nt}(\gamma)$, with $\tilde{s}^{\nu}_{Nt}(\gamma) = \tilde{s}_{Nt}(\gamma)\nu_t$, is the version of the estimated score perturbed by a Gaussian noise. Similarly, the perturbed score test is defined by $LM^{\nu}_{NT} = \tilde{S}^{\nu(2)\prime}_{NT}(\gamma)\tilde{\Sigma}^{-1}_{NT}(\gamma, \gamma)\tilde{S}^{\nu(2)}_{NT}(\gamma)$ and $\tilde{g}_{NT} = g(LM^{\nu}_{NT})$. Finally, $\tilde{p}_{NT} = 1 - \tilde{F}_{NT}(g_{NT})$ is the approximation of p-values obtained by stochastic permutations, where $\tilde{F}_{NT}(\cdot)$ denotes the distribution function of $\tilde{g}_{NT}$, conditional to the available sample. The following result shows that such a bootstrap approximation provides adequate approximation to the null distribution.

THEOREM 5.3. *Assume the conditions of Theorems 3.3, 5.1 hold. Then $\tilde{p}_{NT_N} - p_{NT_N} = o_p(1)$ and $\tilde{p}_{NT_N} \Rightarrow p$. Moreover, under H_0, $\tilde{p}_{NT_N} \xrightarrow{d} U(0, 1)$.*

The proof of this theorem is given in Appendix A.5. An analogous result is obtained for continuous-valued network models and it is omitted. Although $\tilde{p}_{NT}$ is close to p_{NT} asymptotically, the conditional distribution $\tilde{F}_{NT}(\cdot)$ is not observed. We can approximate this by Monte Carlo simulations following (i)–(iv) of [38], p. 419. A Gaussian sequence $\{\nu_{t,j} : t = 1, \dots, T\} \sim \text{IIDN}(0, 1)$ is generated, and at each iteration compute the quantities $\tilde{S}^{\nu_j}_{NT}(\gamma)$, $LM^{\nu_j}_{NT}(\gamma)$ and $\tilde{g}^{j}_{NT} = g(LM^{\nu_j}_{NT}(\gamma))$, for $j = 1, \dots, J$. Hence, an approximation of the p-values is obtained by $\tilde{p}^{J}_{NT} = J^{-1}\sum_{j=1}^{J} I(\tilde{g}^{j}_{NT} \geq g_{NT})$. The Glivenko–Cantelli theorem implies that $\tilde{p}^{J}_{NT} \xrightarrow{p} \tilde{p}_{NT}$, as $J \to \infty$, and choosing J large enough allows to make $\tilde{p}^{J}_{NT}$ arbitrary close to $\tilde{p}_{NT}$.

The proposed bootstrap methodology provides a direct approximation of the p-values instead of an approximate bound, given by (35). Furthermore, it is suitable even when testing linearity in the presence of more than one nuisance parameter. As a final remark, the stochastic permutation bootstrap method has been preferred instead of parametric bootstrap as it requires only the generation of standard univariate normal sequences at each step. This considerably reduces the computational burden of generating a $N \times 1$ vector of observations at each step of the procedure. This is especially relevant in the case of count data, since the simulation of copula can be time consuming.

REMARK 6. Following up on Remark 2, note that the previous results do not apply to TNAR model (8), if $N \to \infty$. The stochastic equi-continuity and uniform convergence assumptions require $h_t(\gamma)$ to be continuous with respect to γ, which is not satisfied for (8). For instance, when trying to establish stochastic equi-continuity of the score, it can be proved that for (8), the Lipschitz property (36) can be obtained in expectation but with magnitude $\lambda = 1/(2q)$. However, to establish the result of Theorem 5.1 we need $\lambda > m^*/q$ [39], p. 357. This can happen only when $m^* < 1/2$ but for the TNAR model $m^* = 1$, so the condition is not satisfied. [7], Section S-4, provides properties, estimation and testing for TNAR models when N is fixed.

6. Simulations. We provide two different cases for the network generating mechanism to verify empirically the above results. Additional results are reported in [7], Section S-5.

TABLE 1
Empirical size and power of the test statistics (29) for testing $H_0 : \gamma = 0$ versus $H_1 : \gamma > 0$, in model (6), with $S = 1000$ simulations, for various values of N and T. Data are continuous-valued and generated from the linear model (4)

				Size			Power ($\gamma = 0.5$)			Power ($\gamma = 1$)		
Model	K	N	T	10%	5%	1%	10%	5%	1%	10%	5%	1%
SBM	2	4	500	0.093	0.043	0.009	1.000	1.000	0.999	1.000	1.000	1.000
		500	10	0.019	0.004	0.000	0.158	0.063	0.002	0.164	0.067	0.001
		200	300	0.110	0.044	0.006	0.495	0.337	0.125	0.994	0.990	0.933
		500	300	0.101	0.048	0.009	0.716	0.583	0.288	1.000	1.000	0.995
		500	400	0.105	0.050	0.006	0.751	0.619	0.311	1.000	1.000	0.999
SBM	5	10	500	0.119	0.062	0.015	1.000	1.000	1.000	1.000	1.000	1.000
		200	300	0.091	0.051	0.006	0.667	0.542	0.268	1.000	1.000	1.000
		500	300	0.098	0.047	0.006	0.847	0.748	0.448	1.000	1.000	1.000
		500	400	0.086	0.039	0.006	0.885	0.807	0.541	1.000	1.000	1.000
ER	-	30	500	0.066	0.029	0.004	0.272	0.156	0.048	0.888	0.802	0.565
		500	30	0.026	0.005	0.000	0.392	0.235	0.044	0.935	0.847	0.523
		200	300	0.085	0.031	0.004	0.411	0.272	0.080	0.974	0.949	0.798
		500	300	0.082	0.042	0.004	0.649	0.476	0.192	0.999	0.998	0.974
		500	400	0.089	0.051	0.008	0.666	0.519	0.206	1.000	1.000	0.992

EXAMPLE N-1 (Stochastic Block Model (SBM)). First, consider the stochastic block model; see [59] and [46] among others. A block label ($k = 1, \dots, K$) 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. We assume $K \in \{2, 5\}$.

EXAMPLE N-2 (Erdős–Rényi (ER) model). Introduced by [26] and [35], this graph model is simple. The network is constructed by connecting N nodes randomly. Each edge is included in the graph with probability p, independently from every other edge. In this example, we set $p = \mathrm{P}(a_{ij} = 1) = N^{-0.3}$.

Consider testing $H_0 : \gamma = 0$ versus $H_1 : \gamma > 0$ for models (5) and (6). Under H_0, the model reduces to (4). For the continuous-valued case, we test linearity of the NAR against the nonlinear version in (6). The random errors $\xi_{i,t}$ are simulated from $N(0, 1)$. For the data generating process of the vector Y_t, the initial value Y_0 is randomly simulated according to its stationary distribution [66], Proposition 1, which 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 vec(.) denotes the vec operator and $\otimes$ denotes the Kronecker product. We set $\theta^{(1)} = (\beta_0, \beta_1, \beta_2)' = (1.5, 0.4, 0.5)'$. This procedure is replicated $S = 1000$ times. Then $\tilde{\theta}^{(1)}$ is computed for each replication. By Proposition 4.4, the quasi-score statistic (29) is evaluated and compared with the critical values of a χ^2_1 distribution. Results of this simulation study are reported in Table 1. The empirical size of the test does not exceed the nominal significance level in all cases considered. When N is small and T is large enough, the power of the test statistics tends to 1. In the case of small temporal size T and large network dimension N, the approximation suffers. This is expected and is explained by (i) the double asymptotic results of Section 4 hold when $T_N \to \infty$ as $N \to \infty$; see the proof of Lemma 3.1; (ii) the temporal dependence induced by the error term requires a sufficiently large T for successful model identification; (iii) the quasi-likelihood might not approximate the true likelihood (see also [6], Section 4.1). When both N, T are large enough, the test approximates adequately its asymptotic distribution. As expected, when $\gamma = 1$, the test statistic's power improves, because γ is far from 0. Improved performance of the test statistic is observed when either $K = 5$ or when the Erdős–Rényi model is employed. Histograms and Q-Q plots of the simulated score test against the χ_1^2 distribution are plotted in [7], Figure S-2. For all the network models, the histogram is positively skewed and approximates satisfactory the χ_1^2 distribution. The Q-Q plots lie into the confidence bands quite satisfactorily and the empirical mean and variance of the simulated score tests are close to 1 and 2, respectively. Further simulations results for the integer-valued case can be found in [7], Section S-5, together with simulation results related to the nonidentifiable case.

7. Empirical example. We discuss an example of the testing methods for integer data. For an example concerning continuous data, see [7], Section S-5.3. The data set consists of monthly number of burglaries on the south side of Chicago from 2010–2015, that is, $T = 72$ and $N = 552$ census block groups of Chicago; see [14], https://github.com/nick3703/Chicago-Data. To predict the future number of burglaries, the ordinary Vector Autoregressive (VAR) model can be applied but we should take into account that data are counts and dimensionality issues because the number of VAR parameters is large compared to the sample size. A simple method, like fitting AR(1) models separately to each individual census blocks, is applicable but still requires $2N$ parameters to be fitted. More crucially, the relationship across different time series is not taken into account. To overcome such issues, we appeal to geographic network information between blocks to fit a PNAR model, which takes into account dependence among count valued data. An undirected network structure is defined by geographical proximity: two blocks are connected if they share at least a border. The density of this network is 1.74%. The median number of connections is 5. The QMLE is employed for fitting linear PNAR model (4). The results are summarized in Table 2. The magnitude of the network effect β_1 shows that an increasing number of burglaries in a block can lead to a growth in the same type of crime committed in a neighbourhood area. The effect of the lagged variable has a upwards impact on the number of burglaries as well. We evaluate the out-of-sample forecasting performance of the linear PNAR(1) model versus a baseline AR(1) model fitted separately to each individual census block. We evaluate the one-step ahead forecast by

TABLE 2

QMLE estimates of the linear model (4) for Chicago burglary counts. Standard errors in brackets. Linearity is tested against the nonlinear model (5), with χ_1^2 asymptotic test (29); against the STNAR model (7), with p-values computed by (DV) Davies bound (35), bootstrap p-values of the sup-type test and versus the TNAR model (8)

Models	β_0	β_1	β_2
(4)	0.455	0.322	0.284
SE	(0.022)	(0.013)	(0.008)

Models	Chi-sq.	DV	Bootstrap
(5)	8.999	-	-
(7)	-	0.038	0.515
(8)	-	-	0.498

computing its Root Mean Square Error (RMSE). The RMSE for the PNAR model is 0.038. This is considerably smaller than the RMSE obtained by the AR(1) models (which is 0.167). In conclusion, the PNAR model gives significant accuracy improvement of the one-step prediction and at the same time achieves parsimony. We apply now the proposed linearity tests. A quasi-score linearity test is computed according to (29), by using the asymptotic chi-square test, for the nonlinear model (5), testing $H_0 : \gamma = 0$ versus $H_1 : \gamma > 0$. We also test linearity against the presence of smooth transition effects, as in (7), with $H_0 : \alpha = 0$ versus $H_1 : \alpha > 0$. A grid of 100 equi-distant values in an interval of values $\Gamma_F = [\gamma_L, \gamma_U]$ is selected for the nuisance parameter γ, where the extremes are defined as in [57], p. 9. According to the results of Theorem 5.1, the p-values are computed with the Davies bound approximation (35) for the test $\sup LM_T = \sup_{\gamma \in \Gamma_F} LM_T(\gamma)$ as well as through the bootstrap approximation procedure. The number of bootstrap replications is $J = 299$. Finally, a linearity test against threshold effects, as in (8), 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 $l = 0, 1, 2$. A feasible range values for the nonidentifiable threshold parameter has been considered as in [57], p. 11. From Table 2, the linearity test against (5) is rejected at standard levels. This gives an intuition for possible nonlinear drifts in the intercept. Davies bound gives evidence in favour of STNAR effects at the 5% level. Conversely, bootstrap sup-tests reject nonlinearity coming from both smooth (7) and abrupt transition (8) models. We conclude that there is no clear evidence of the regime switching effect.

APPENDIX

A.1. Proof of Theorem 2.1. Consider the $N \times 1$ Markov chain $Y_t = F(Y_{t-1}, N_t)$ where $\{N_t, t \in \mathbb{Z}\}$ defined in (1) is a sequence of *IID* N-dimensional count processes such that $N_{i,t}$, for $i = 1 \dots, N$, are Poisson processes with intensity 1. $F(\cdot)$ is a measurable function such that $F(y, N_t) = (N_t[f(y)])$ and $f(\cdot)$ is defined in (1) for $y \in \mathbb{N}^N$. By (9), $f(y) \preceq C + Gy$, where $C = f(0)$, we have

$$\mathrm{E}\left|F(y, N_1)\right|_1 = 1' f(y) \le 1'[C + Gy] < \infty$$

since the expectation of the Poisson process is $\mathrm{E}N_1(\lambda) = \lambda$. Moreover, for $y, y^* \in \mathbb{N}^N$,

$$\mathrm{E}\left|F(y, N_1) - F\left(y^*, N_1\right)\right|_{\mathrm{vec}} \preceq G\left|y - y^*\right|_{\mathrm{vec}}$$

as $\mathrm{E}|N_1(\lambda_1) - N_1(\lambda_2)|_{\mathrm{vec}} = |\lambda_1 - \lambda_2|_{\mathrm{vec}}$. Note that $\rho(G) < 1$, Therefore, by [20], Theorem 1, $\{Y_t,\ t \in \mathbb{Z}\}$ is a stationary and ergodic process with $\mathrm{E}|Y_t|_1 < \infty$. Now set $\delta > 0$ such that $\rho(G_\delta) < 1$, if $G_\delta = (1 + \delta)G$. From [20], Lemma 2, we have that

$$\left\|N_t\left[f(y)\right]\right\|_{a,\mathrm{vec}} \preceq (1 + \delta)\left|f(y)\right|_{\mathrm{vec}} + b1 \preceq C_{\delta b} + G_\delta |y|_{\mathrm{vec}}$$

by recalling that $|f(y)|_{\mathrm{vec}} \preceq C + G|y|_{\mathrm{vec}}$ and $\rho(G_\delta) < 1$, where $b > 0$ and $C_{\delta b} = (1 + \delta)C + b$. Then, by [20], Theorem 1, we get $\mathrm{E}|Y_t|_a^a < \infty$, $\forall a \ge 1$. Theorem 2.2 follows analogously.

A.2. Proof of Theorem 2.3. For any arbitrary N, $\mathrm{E}(Y_t) = \mathrm{E}(\lambda_t) = \mathrm{E}[f(Y_{t-1})] \preceq c1 + G\mathrm{E}(Y_{t-1})$, by (9), where $\max_{i \ge 1} f_i(0, 0) = c > 0$. Define $\mu = c(1 - \mu_1 - \mu_2)^{-1}$. Note that $\rho(G) \le |||G|||_\infty \le \mu_1 |||W|||_\infty + \mu_2 \le \mu_1 + \mu_2$. This is so because $|||W|||_\infty = 1$, by construction. Since $\mu_1 + \mu_2 < 1$, we have $|||G|||_\infty < 1$ and by [51], 19.16(a), $(I - G)^{-1}$ exists. Moreover, $(I - G)^{-1}1 = (1 - \mu_1 - \mu_2)^{-1}1$, implying that $\mathrm{E}(Y_t) \preceq \mu 1$ and $\max_{i \ge 1} E(Y_{i,t}) \le \mu$. It holds that $\xi_t = Y_t - \lambda_t$, $\mathrm{E}|\xi_{i,t}| \le 2\mathrm{E}(Y_{i,t}) \le 2\mu < \infty$. Furthermore, by using backward substitution and (9), we have $Y_t \preceq \mu 1 + \sum_{j=0}^{\infty} G^j \xi_{t-j} = \sum_{j=0}^{\infty} G^j (c1 + \xi_{t-j})$.

From the definition in [7], Section S-1.2, we have that $\mathcal{W} = \{\omega \in \mathbb{R}^\infty : \omega_\infty = \sum|\omega_i| < \infty\}$, where $\omega = (\omega_i \in \mathbb{R} : 1 \le i \le \infty)' \in \mathbb{R}^\infty$. For each $\omega \in \mathcal{W}$, let $\omega_N = (\omega_1, \dots, \omega_N)' \in \mathbb{R}^N$ be its truncated N-dimensional version. For any $\omega \in \mathcal{W}$, $\mathrm{E}|c1 + \xi_t|_{\mathrm{vec}} \preceq (c + 2\mu)1 = C1 < \infty$,

$G^j 1 = (\mu_1 + \mu_2)^j 1$ and $\mathrm{E}|\omega_N' Y_t| \le \mathrm{E}(|\omega_N|_{\mathrm{vec}}' \sum_{j=0}^{\infty} G^j (c1 + \xi_{t-j})) \le C\omega_\infty \sum_{j=0}^{\infty}(\mu_1 + \mu_2)^j = C_*$, since $\mu_1 + \mu_2 < 1$, which implies that $Y_t^\omega = \lim_{N\to\infty} \omega_N' Y_t < \infty$ with probability 1. Moreover, Y_t^ω is strictly stationary and, therefore, $\{Y_t\}$ is strictly stationary, following [7], Section S-1.2. To verify the uniqueness of the solution, take another stationary solution Y_t^* to the PNAR model. Then $\mathrm{E}|\omega_N' Y_t - \omega_N' Y_t^*| \le |\omega_N|_{\mathrm{vec}}' \mathrm{E}|N_t(\lambda_t) - N_t(\lambda_t^*)|_{\mathrm{vec}} \le |\omega_N|_{\mathrm{vec}}' G \mathrm{E}|Y_{t-1} - Y_{t-1}^*|_{\mathrm{vec}} = 0$, by infinite backward substitution, for any N and weight ω. So, $Y_t^\omega = Y_t^{*,\omega}$ with probability one. In addition, $\mu_1 w_i' + \mu_2 e_i' = e_i' G$ and condition (9) is equivalent to require, for $i = 1, \dots, N$, a.s.

$$
\begin{aligned}
|\lambda_{i,t} - \lambda_{i,t}^*| &= e_i' |f(Y_{t-1}) - f(Y_{t-1}^*)|_{\mathrm{vec}} \le (\mu_1 w_i' + \mu_2 e_i') |Y_{t-1} - Y_{t-1}^*|_{\mathrm{vec}} \\
&= \mu_1 \sum_{j=1}^{N} w_{ij} |Y_{j,t-1} - Y_{j,t-1}^*| + \mu_2 |Y_{i,t-1} - Y_{i,t-1}^*|
\end{aligned}
$$

which leads a.s. to $\lambda_{i,t} = f_i(X_{i,t-1}, Y_{i,t-1}) \le c + \mu_1 X_{i,t-1} + \mu_2 Y_{i,t-1}$. Then, when N is increasing, [6], Proposition 2, applies directly by a recursion argument [6], Section S-1.1, and all moments of the process $\{Y_t\}$ are uniformly bounded.

A.3. Proof of Theorem 2.4. Similar to A.1, by assuming that $\max_{i\ge 1} \mathrm{E}|\xi_{i,t}|^a \le C_{\xi,a} < \infty$, the first ath moments of Y_t are uniformly bounded. By (9), $|Y_t|_{\mathrm{vec}} \preceq \sum_{j=0}^{\infty} G^j (c1 + |\xi_{t-j}|_{\mathrm{vec}})$, where $c = \max_{i\ge 1} |f_i(0,0)|$. Analogously to A.2, since $G^j 1 = (|\mu_1| + |\mu_2|)^j 1$ and $|\mu_1| + |\mu_2| < 1$, $\{Y_t\}$ defined as in (2), is strictly stationary, following [66], Definition 1. The uniqueness of the solution follows by $|Y_t - Y_t^*|_{\mathrm{vec}} = |\lambda_t - \lambda_t^*|_{\mathrm{vec}}$ and the infinite backward substitution argument.

A.4. Proof of Theorem 5.1. First, we show the weak convergence of $(NT)^{-1/2} S_{NT}(\gamma)$ to a Gaussian process with kernel $B(\gamma_1, \gamma_2)$. For all nonnull $\eta \in \mathbb{R}^k$, consider the triangular array $s_{Nt}^*(\gamma) = \eta'(N^{-1/2} \sum_{i=1}^{N} s_{i,t}(\gamma))$. By Assumption $\mathbf{B}'$, Γ is compact, and by the continuity of the score $s_{Nt}^*(\gamma)$ is compact. Note that $s_{Nt}^*(\gamma)$ is a martingale difference array. So, by the results of Lemma 3.1, the multivariate pointwise central limit theorem and $(NT_N)^{-1} B_{NT_N}(\gamma_1, \gamma_2) \xrightarrow{p} B(\gamma_1, \gamma_2)$ establish the finite-dimensional convergence. It remains to show the stochastic equi-continuity, that is, ([39], Theorem 2),

$$
|s_{Nt}^*(\gamma) - s_{Nt}^*(\gamma^*)| \le \delta_{Nt} |\gamma - \gamma^*|_1^{\lambda}, \tag{36}
$$

a.s. with $\|\delta_{Nt}\|_q < \infty$ and $\|s_{Nt}^*(\gamma)\|_q < \infty$, where $q \ge 2$ and λ such that $q > m^*/\lambda$. By [6], Section S-6, and Assumption **H1**–**H3**, $\|s_{Nt}^*(\gamma)\|_4 < \infty$. For $q > 4$, a similar result can be obtained following the arguments of [62], Remark 2.3, by requiring higher-order covariances in Assumption **H2**. To prove (36), we recall the following uniform bounds. By Assumption $\mathbf{B}'$, for $i = 1, \dots, N$, $|\partial f_i(x_i, y_i, \theta)/\partial \alpha_b| = h_{i,t}^b(\gamma) \le c_b + c_{1b} x_i + c_{2b} y_i$ for $b = 1, \dots, k_2$ and $l = 1, \dots, m^*$, where $c_b = h_i^b(0, \gamma)$ $\forall \gamma \in \Gamma$. Let $C, C_0, C_1, C_2 > 0$ be generic constants varying from place to place, which do not depend on γ. Then a.s. $|h_{i,t}(\gamma)|_1 \le C_0 + C_1 X_{i,t-1} + C_2 Y_{i,t-1}$. Similar bounds hold for $\lambda_{i,t}(\gamma)$ and $|Z_{i,t}(\gamma)|_1$. By Theorem 2.3, all the moments of the Poisson process Y_t exist as well as those associated to the error $\xi_t(\gamma) = Y_t - \lambda_t(\gamma)$. This fact and the multinomial theorem imply that every moment of all the previously defined random variables is uniformly bounded. Define $h_{i,t}(0, \gamma) = c$, $\forall \gamma \in \Gamma$ and $h_{i,t}^*(\gamma) = h_{i,t}(\gamma) - h_{i,t}(0, \gamma)$. For $i = 1, \dots, N$ and MVT $|h_{i,t}(\gamma) - h_{i,t}(\gamma^*)|_1 = |h_{i,t}^*(\gamma) - h_{i,t}^*(\gamma^*)|_1 = |\partial h_{i,t}^*(\tilde{\gamma})/\partial \gamma|_1 |\gamma - \gamma^*|_1 \le A_{i,t-1} |\gamma - \gamma^*|_1$ a.s. where $A_{i,t-1} = C_1 X_{i,t-1} + C_2 Y_{i,t-1}$, $\tilde{\gamma}_l$ are intermediate points between γ_l and γ_l^*, for $l = 1, \dots, m^*$ and the last inequality holds by Assumption $\mathbf{B}'$ since $\partial h_{i,t}^*(\tilde{\gamma})/\partial \gamma = \partial^2 f_i(x_i, y_i, \theta)/\partial \alpha \partial \gamma - \partial^2 f_i(0, 0, \theta)/\partial \alpha \partial \gamma$. For all $\gamma, \gamma^* \in \Gamma$, standard algebra and previous bounds show that a.s. $|s^*_{Nt}(\gamma) - s^*_{Nt}(\gamma^*)| \leq \delta_{Nt}|\gamma - \gamma^*|_1$ where $\delta_{Nt} = C/\sqrt{N}\sum_{i=1}^N A_{i,t-1}(1 + C_0Y_{i,t} + C_1Y_{i,t}X_{i,t-1} + C_2Y_{i,t}Y_{i,t-1})$ proving (36) with $\lambda = 1$. By Assumption **B**′ $\|\delta_{Nt}\|_q < \infty$ since $\delta_{Nt} \leq C\eta_{Nt}$, then $(NT_N)^{-1/2}S_{NT_N}(\gamma) = T_N^{-1/2}\sum_{t=1}^{T_N} N^{-1/2}s_{Nt}(\gamma)$ is stochastically equi-continuous and as $\{N, T_N\} \to \infty$, $(NT_N)^{-1/2}S_{NT_N}(\gamma) \Rightarrow S(\gamma)$.

We now prove uniform convergence of $\tilde{\Sigma}_{NT}(\gamma_1, \gamma_2)$ by showing stochastic equi-continuity for Hessian and information matrices. For all $\eta \in \mathbb{R}^k$, $\eta \neq 0$, consider the triangular array $b_{Nt}(\gamma_1, \gamma_2) = \eta'(N^{-1}B_{Nt}(\gamma_1, \gamma_2))\eta$ where $B_{Nt}(\gamma_1, \gamma_2)$ is the single summand of $B_{NT}(\gamma_1, \gamma_2)$. Define $\rho_{ijt}(\gamma_1, \gamma_2) = \mathrm{E}[\xi_{i,t}(\gamma_1)\xi_{j,t}(\gamma_2)|\mathcal{F}_{t-1}]/\sqrt{\lambda_{i,t}(\gamma_1)\lambda_{j,t}(\gamma_2)}$ the conditional correlation. Then a.s.

$$\begin{aligned}
&\left|b_{Nt}(\gamma_1, \gamma_2) - b_{Nt}(\gamma_1^*, \gamma_2^*)\right| \\
&\quad \leq \eta'\left(\frac{1}{N}\sum_{i,j=1}^N \frac{Z_{i,t}(\gamma_1)\rho_{ijt}(\gamma_1, \gamma_2)Z'_{j,t}(\gamma_2)}{\sqrt{\lambda_{i,t}(\gamma_1)}\sqrt{\lambda_{j,t}(\gamma_2)}} - \frac{Z_{i,t}(\gamma_1^*)\rho_{ijt}(\gamma_1^*, \gamma_2^*)Z'_{j,t}(\gamma_2^*)}{\sqrt{\lambda_{i,t}(\gamma_1^*)}\sqrt{\lambda_{j,t}(\gamma_2^*)}}\right)\eta \\
&\quad \leq C\sum_{r=1}^5 D_r,
\end{aligned}$$

and the inequality follows since for a matrix M, $\eta'M\eta \leq |\eta\eta'|_1|M|_1$ and by $\lambda_{i,t}(\gamma) \geq C$ $\forall\gamma \in \Gamma$. The elements D_r are obtained by consecutive addition and subtraction. We focus on one element (say D_1), and the other terms are treated analogously. Some tedious algebra shows that a.s.

$$\begin{aligned}
&\left|\rho_{ijt}(\gamma_1, \gamma_2) - \rho_{ijt}(\gamma_1^*, \gamma_2^*)\right| \\
&\quad \leq \left|\lambda_{i,t}^{\frac{1}{2}}(\gamma_1) - \lambda_{i,t}^{\frac{1}{2}}(\gamma_1^*)\right|\left|\rho_{ijt}(\gamma_1^*, \gamma_2^*)\right|\lambda_{j,t}^{\frac{1}{2}}(\gamma_2) + \left|\lambda_{j,t}^{\frac{1}{2}}(\gamma_2) - \lambda_{j,t}^{\frac{1}{2}}(\gamma_2^*)\right|\left|\rho_{ijt}(\gamma_1^*, \gamma_2^*)\right|\lambda_{i,t}^{\frac{1}{2}}(\gamma_1^*) \\
&\quad \leq C_1\left|\lambda_{i,t}(\gamma_1) - \lambda_{i,t}(\gamma_1^*)\right|\varphi_{j-i}A^*_{j,t-1} + C_2\left|\lambda_{j,t}(\gamma_2) - \lambda_{j,t}(\gamma_2^*)\right|\varphi_{j-i}A^*_{i,t-1} \\
&\quad \leq C_1^*\varphi_{j-i}\tilde{A}_{ij,t-1}\left|\gamma_1 - \gamma_1^*\right|_1 + C_2^*\varphi_{j-i}\tilde{A}_{ji,t-1}\left|\gamma_2 - \gamma_2^*\right|_1,
\end{aligned}$$

where $A^*_{i,t-1} = A_{i,t-1} + C_0$ and $\tilde{A}_{ij,t-1} = A_{i,t-1}A^*_{j,t-1}$. The first inequality follows by addition and subtraction. The second inequality is a consequence of Assumption **H2** and $|\sqrt{x} - \sqrt{y}| = |x - y|/(\sqrt{x} + \sqrt{y})$; the third is due to Lipschitz continuity of $h_{i,t}(\gamma)$. Set $\pi_{ijt}(\gamma, \gamma^*) = |\sqrt{\lambda_{i,t}(\gamma_1^*)\lambda_{j,t}(\gamma_2^*)}Z_{i,t}(\gamma_1)Z'_{j,t}(\gamma_2)|_1 \leq \pi_{ijt}$ a.s. with the inequality coming from previous uniform bounds where π_{ijt} is a linear combination of $X_{i,t-1}$ and $Y_{i,t-1}$ not depending on γ. Then

$$\begin{aligned}
D_1 &= \frac{1}{N}\sum_{i,j=1}^N \left|\rho_{ijt}(\gamma_1, \gamma_2) - \rho_{ijt}(\gamma_1^*, \gamma_2^*)\right|\pi_{ijt}(\gamma, \gamma^*) \\
&\leq \frac{C_1^*}{N}\sum_{i,j=1}^N \varphi_{j-i}\tilde{A}_{ij,t-1}\pi_{ijt}\left|\gamma_1 - \gamma_1^*\right|_1 + \frac{C_2^*}{N}\sum_{i,j=1}^N \varphi_{j-i}\tilde{A}_{ji,t-1}\pi_{ijt}\left|\gamma_2 - \gamma_2^*\right|_1.
\end{aligned}$$

This shows that $|b_{Nt}(\gamma_1, \gamma_2) - b_{Nt}(\gamma_1^*, \gamma_2^*)| \leq b^*_{1,Nt}|\gamma_1 - \gamma_1^*|_1 + b^*_{2,Nt}|\gamma_2 - \gamma_2^*|_1$ a.s. with $b^*_{s,Nt}$ defined by obvious notation, not depending on γ and such that $\mathrm{E}(b^*_{s,Nt}) < \infty$, for $s = \{1, 2\}$. Rewriting in matrix form, we have $b^*_{s,Nt} = \eta'(N^{-1}B^*_{s,Nt}(\gamma_1, \gamma_2))\eta$. According to [3], Lemma 1, this is a sufficient condition for the information matrix to be stochastic equi-continuous and by [3], Theorem 1, $(NT_N)^{-1}B_{NT_N}(\gamma_1, \gamma_2) \xrightarrow{p} B(\gamma_1, \gamma_2)$ uniformly over $\gamma_1, \gamma_2 \in \Gamma$, as $\{N, T_N\} \to \infty$. An analogous result for the Hessian follows by MVT with respect to γ_1, γ_2 and the uniform boundedness of the third derivative (Lemma 3.1). By standard

Taylor expansion arguments, and CMT, $(NT_N)^{-1}\tilde{\Sigma}_{NT_N}(\gamma_1,\gamma_2) \xrightarrow{p} \Sigma(\gamma_1,\gamma_2)$ uniformly over $\gamma_1,\gamma_2 \in \Gamma$. Following analogous steps of [7], Section S-3.1, for the identifiable parameters ϕ, [7], equation (S-2), leads to

$$\text{(37)} \qquad \frac{\tilde{S}^{(2)}_{NT_N}(\gamma)}{\sqrt{NT_N}} \doteq P(\gamma,\gamma)\frac{S_{NT_N}(\gamma)}{\sqrt{NT_N}} \quad \Rightarrow \quad P(\gamma,\gamma)S(\gamma) := S^{(2)}(\gamma) \equiv N\bigl(0, \Sigma(\gamma,\gamma)\bigr),$$

with $P(\gamma,\gamma) = [-J_2H(\gamma,\gamma)J_1'(J_1H(\gamma,\gamma)J_1')^{-1}, I_{k_2}]$, $\Sigma(\gamma,\gamma) = P(\gamma,\gamma)B(\gamma,\gamma)P(\gamma,\gamma)'$. Finally, the CMT shows that $LM_{NT_N}(\gamma) \Rightarrow LM(\gamma)$ and $g_{NT_N} \Rightarrow g$. A similar conclusion is obtained for the local alternatives $\alpha = (NT)^{-1/2}\delta_2$, where $\delta_2 \in \mathbb{R}^{k_2}$, by [7], equation (S-4), with $S^{(2)}(\gamma) \equiv N(J_2H^{-1}(\gamma,\gamma)J_2'\delta_2, \Sigma(\gamma,\gamma))$ in (37). This completes the proof.

A.5. Proof of Theorem 5.3. Following the results of Section A.4, the information matrix $B_{Nt}(\gamma_1,\gamma_2)$ is Lipschitz for γ_1, γ_2 with constants $B^*_{1,Nt}$, $B^*_{2,Nt}$ having finite absolute moments. Moreover, by [7], Section S-2.2, $B_{NT_N}(\gamma_1,\gamma_2) \xrightarrow{a.s.} B(\gamma_1,\gamma_2)$ $\forall \gamma_1,\gamma_2 \in \Gamma$. Define $B^*_{s,NT} = T^{-1}\sum_{t=1}^T B^*_{s,Nt}$, for $s = \{1,2\}$. Following the same arguments of [7], Section S-2.1, S-2.2, it can be proved that $B^*_{s,NT_N} - B^*_{s,N} \xrightarrow{a.s.} 0$ where $B^*_{s,N} = \mathrm{E}(B^*_{s,Nt})$. Assumptions **H1**–**H3** imply that $B^*_s = \lim_{N\to\infty} B^*_N$ is finite. Then $B^*_{s,NT_N} \xrightarrow{a.s.} B^*_s$. This is a sufficient condition for B_{NT} to be strongly stochastically equi-continuous [3], Lemma 1, and together with pointwise almost sure convergence, [3], Theorem 2, shows that $B_{NT_N}(\gamma_1,\gamma_2) \xrightarrow{a.s.} B(\gamma_1,\gamma_2)$ uniformly over γ_1,γ_2.

Consider $\omega \in \Omega$, where Ω denotes a set of samples. We operate conditionally on the sample ω, so randomness is through the IID standard normal process ν_t. Set $S^\nu_{NT}(\gamma) = \sum_{t=1}^T s^\nu_{Nt}(\gamma)$, with $s^\nu_{Nt}(\gamma) = s_{Nt}(\gamma)\nu_t$. Then $\tilde{S}^\nu_{NT}(\gamma) = S^\nu_{NT}(\gamma) + \bar{S}_{NT}(\gamma)$, $\bar{S}_{NT}(\gamma) = \sum_{t=1}^T(\tilde{s}^\nu_{Nt}(\gamma) - s^\nu_{Nt}(\gamma)) = \sum_{t=1}^T\sum_{i=1}^N(\tilde{s}^\nu_{i,t}(\gamma) - s^\nu_{i,t}(\gamma))$, and a.s.

$$\begin{aligned}
\bar{S}_{NT}(\gamma) &= \sum_{t=1}^T\sum_{i=1}^N\biggl(\frac{Z_{i,t}(\gamma)\tilde{\xi}_{i,t}\nu_t}{\tilde{\lambda}_{i,t}} - \frac{Z_{i,t}(\gamma)\xi_{i,t}(\gamma)\nu_t}{\lambda_{i,t}(\gamma)}\biggr) \\
&= \sum_{t=1}^T\sum_{i=1}^N Z_{i,t}(\gamma)\biggl(\frac{\lambda_{i,t}(\gamma)\tilde{\xi}_{i,t} - \tilde{\lambda}_{i,t}\xi_{i,t}(\gamma)}{\tilde{\lambda}_{i,t}\lambda_{i,t}(\gamma)}\biggr)\nu_t \\
&\le \beta_0^{-2}\sum_{t=1}^T\sum_{i=1}^N Z_{i,t}(\gamma)Z'_{i,t}(\gamma)(\phi - \tilde{\phi})Y_{i,t}\nu_t.
\end{aligned}$$

Set $G^\nu_{NT}(\gamma) := \beta_0^{-2}\sum_{t=1}^T\sum_{i=1}^N Y_{i,t}Z_{i,t}(\gamma)Z'_{i,t}(\gamma)\nu_t$, so

$$\sup_{\gamma\in\Gamma}\biggl|\frac{\bar{S}_{NT}(\gamma)}{\sqrt{NT}}\biggr|_1 \le \sup_{\gamma\in\Gamma}\biggl\|\frac{G^\nu_{NT}(\gamma)}{NT}\biggr\|_1 \bigl|\sqrt{NT}(\tilde{\phi} - \phi)\bigr|_1.$$

By Section A.4, $s_t(\gamma)$ is L^2 integrable. Then, from the assumptions of Theorems 3.3, 5.1 and Pollard's central limit theorem for triangular empirical processes [50], Theorem 10.6, the arguments in [38], pp. 426–427, prove that $(NT_N)^{-1/2}S^\nu_{NT_N}(\gamma) \Rightarrow_p S(\gamma)$, where $\Rightarrow_p$ denotes the weak convergence in probability, as defined in [36]. Furthermore, we have $(NT_N)^{-1}G^\nu_{NT_N}(\gamma) \xrightarrow{a.s.} O_{k\times k}$. Then $(NT_N)^{-1/2}\bar{S}_{NT_N}(\gamma) \Rightarrow_p 0$, $(NT_N)^{-1/2}\tilde{S}^\nu_{NT_N}(\gamma) \Rightarrow_p S(\gamma)$, $LM^\nu_{NT_N}(\gamma) \Rightarrow_p LM(\gamma)$, $\tilde{g}_{NT_N} \Rightarrow_p g$, $\tilde{F}_{NT_N}(x) \xrightarrow{p} F(x)$, uniformly over x and $\tilde{p}_{NT_N} = 1 - \tilde{F}_{NT_N}(g_{NT_N}) = 1 - F(g_{NT_N}) + o_p(1) = p_{NT_N} + o_p(1)$.

Acknowledgements. This work was completed when M. Armillotta was with the Department of Mathematics and Statistics at the University of Cyprus. We greatly appreciate comments made by the Editor, Assistant Editor and two reviewers on an earlier version of the manuscript.

Funding. The research was supported by the European Regional Development Fund and the Republic of Cyprus through the Research and Innovation Foundation, under the project INFRASTRUCTURES/1216/0017 (IRIDA).

SUPPLEMENTARY MATERIAL

Supplement to "Nonlinear network autoregression" (DOI: 10.1214/23-AOS2345SUPP; .pdf). The supplementary material contains the proofs for Sections 3 and 4, additional simulations and empirical results, some important concepts and theoretical analysis of the TNAR model (8) with fixed network dimension.

REFERENCES

[1] AHMAD, A. and FRANCQ, C. (2016). Poisson QMLE of count time series models. *J. Time Series Anal.* **37** 291–314. MR3512959 https://doi.org/10.1111/jtsa.12167

[2] ANDREWS, D. W. K. (1987). Consistency in nonlinear econometric models: A generic uniform law of large numbers. *Econometrica* **55** 1465–1471. MR0923471 https://doi.org/10.2307/1913568

[3] ANDREWS, D. W. K. (1992). Generic uniform convergence. *Econometric Theory* **8** 241–257. MR1179512 https://doi.org/10.1017/S0266466600012780

[4] ANDREWS, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica* **61** 821–856. MR1231678 https://doi.org/10.2307/2951764

[5] 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 https://doi.org/10.2307/2951753

[6] ARMILLOTTA, M. and FOKIANOS, K. Count network autoregression. *J. Time Series Anal.* To appear. Available at https://arxiv.org/abs/2104.06296.

[7] ARMILLOTTA, M. and FOKIANOS, K. (2023). Supplement to "Nonlinear Network Autoregression." https://doi.org/10.1214/23-AOS2345SUPP

[8] ARMILLOTTA, M., FOKIANOS, K. and GUIZZARDI, A. (2023). Unveiling Venice's hotels competition networks from dynamic pricing digital market. *J. Roy. Statist. Soc. Ser. A.* qnad085. https://doi.org/10.1093/jrsssa/qnad085

[9] ARMILLOTTA, M., LUATI, A. and LUPPARELLI, M. (2022). Observation-driven models for discrete-valued time series. *Electron. J. Stat.* **16** 1393–1433. MR4387846 https://doi.org/10.1214/22-ejs1989

[10] ARMILLOTTA, M., TSAGRIS, M. and FOKIANOS, K. Inference for network count time series with the R package PNAR. *R J.*. To appear. Available at https://arxiv.org/abs/2211.02582.

[11] CHAN, K.-S. and TONG, H. (2001). *Chaos: A Statistical Perspective. Springer Series in Statistics*. Springer, New York. MR1851668 https://doi.org/10.1007/978-1-4757-3464-5

[12] CHRISTOU, V. and FOKIANOS, K. (2014). Quasi-likelihood inference for negative binomial time series models. *J. Time Series Anal.* **35** 55–78. MR3148248 https://doi.org/10.1111/jtsa.12050

[13] CHRISTOU, V. and FOKIANOS, K. (2015). Estimation and testing linearity for non-linear mixed Poisson autoregressions. *Electron. J. Stat.* **9** 1357–1377. MR3360730 https://doi.org/10.1214/15-EJS1044

[14] CLARK, N. J. and DIXON, P. M. (2021). A class of spatially correlated self-exciting statistical models. *Spat. Stat.* **43** Paper No. 100493, 18. MR4219886 https://doi.org/10.1016/j.spasta.2021.100493

[15] DAVIES, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* **74** 33–43. MR0885917 https://doi.org/10.1093/biomet/74.1.33

[16] DAVIS, R. A., DUNSMUIR, W. T. M. and STREETT, S. B. (2003). Observation-driven models for Poisson counts. *Biometrika* **90** 777–790. MR2024757 https://doi.org/10.1093/biomet/90.4.777

[17] DAVIS, R. A. and LIU, H. (2016). Theory and inference for a class of nonlinear models with application to time series of counts. *Statist. Sinica* **26** 1673–1707. MR3586234

[18] DAVIS, R. A. and WU, R. (2009). A negative binomial model for time series of counts. *Biometrika* **96** 735–749. MR2538769 https://doi.org/10.1093/biomet/asp029

[19] DE JONG, R. M. (1996). A strong law of large numbers for triangular mixingale arrays. *Statist. Probab. Lett.* **27** 1–9. MR1394169 https://doi.org/10.1016/0167-7152(95)00036-4

[20] DEBALY, Z. M. and TRUQUET, L. (2021). A note on the stability of multivariate non-linear time series with an application to time series of counts. *Statist. Probab. Lett.* **179** Paper No. 109196, 7. MR4293765 https://doi.org/10.1016/j.spl.2021.109196

[21] DOUC, R., DOUKHAN, P. and MOULINES, E. (2013). Ergodicity of observation-driven time series models and consistency of the maximum likelihood estimator. *Stochastic Process. Appl.* **123** 2620–2647. MR3054539 https://doi.org/10.1016/j.spa.2013.04.010

[22] DOUC, R., FOKIANOS, K. and MOULINES, E. (2017). Asymptotic properties of quasi-maximum likelihood estimators in observation-driven time series models. *Electron. J. Stat.* **11** 2707–2740. MR3679907 https://doi.org/10.1214/17-EJS1299

[23] DOUKHAN, P. (1994). *Mixing: Properties and Examples. Lecture Notes in Statistics* **85**. Springer, New York. MR1312160 https://doi.org/10.1007/978-1-4612-2642-0

[24] DOUKHAN, P., FOKIANOS, K. and LI, X. (2012). On weak dependence conditions: The case of discrete valued processes. *Statist. Probab. Lett.* **82** 1941–1948. MR2970296 https://doi.org/10.1016/j.spl.2012.06.020

[25] DOUKHAN, P., FOKIANOS, K. and TJØSTHEIM, D. (2012). On weak dependence conditions for Poisson autoregressions. *Statist. Probab. Lett.* **82** 942–948. MR2910041 https://doi.org/10.1016/j.spl.2012.01.015

[26] ERDŐS, P. and RÉNYI, A. (1959). On random graphs. I. *Publ. Math. Debrecen* **6** 290–297. MR0120167 https://doi.org/10.5486/pmd.1959.6.3-4.12

[27] FAN, J. and YAO, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods. Springer Series in Statistics*. Springer, New York. MR1964455 https://doi.org/10.1007/b97702

[28] FOKIANOS, K. (2024). Multivariate count time series modelling. *Econometrics and Statistics*. To appear. Available at https://doi.org/10.1016/j.ecosta.2021.11.006.

[29] FOKIANOS, K., RAHBEK, A. and TJØSTHEIM, D. (2009). Poisson autoregression. *J. Amer. Statist. Assoc.* **104** 1430–1439. With electronic supplementary materials available online. MR2596998 https://doi.org/10.1198/jasa.2009.tm08270

[30] FOKIANOS, K., STØVE, B., TJØSTHEIM, D. and DOUKHAN, P. (2020). Multivariate count autoregression. *Bernoulli* **26** 471–499. MR4036041 https://doi.org/10.3150/19-BEJ1132

[31] FOKIANOS, K. and TJØSTHEIM, D. (2012). Nonlinear Poisson autoregression. *Ann. Inst. Statist. Math.* **64** 1205–1225. MR2981620 https://doi.org/10.1007/s10463-012-0351-3

[32] 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 https://doi.org/10.1017/S0266466609990430

[33] FRANCQ, C. and ZAKOÏAN, J.-M. (2010). *GARCH Models: Structure, Statistical Inference and Financial Applications*. Wiley, Chichester. MR3186556 https://doi.org/10.1002/9780470670057

[34] GAO, J., KING, M., LU, Z. and TJØSTHEIM, D. (2009). Specification testing in nonlinear and nonstationary time series autoregression. *Ann. Statist.* **37** 3893–3928. MR2572447 https://doi.org/10.1214/09-AOS698

[35] GILBERT, E. N. (1959). Random graphs. *Ann. Math. Stat.* **30** 1141–1144. MR0108839 https://doi.org/10.1214/aoms/1177706098

[36] GINÉ, E. and ZINN, J. (1990). Bootstrapping general empirical measures. *Ann. Probab.* **18** 851–869. MR1055437

[37] GORGI, P. (2020). Beta-negative binomial auto-regressions for modelling integer-valued time series with extreme observations. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** 1325–1347. MR4176345

[38] HANSEN, B. E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica* **64** 413–430. MR1375740 https://doi.org/10.2307/2171789

[39] HANSEN, B. E. (1996). Stochastic equicontinuity for unbounded dependent heterogeneous arrays. *Econometric Theory* **12** 347–359. MR1395036 https://doi.org/10.1017/S0266466600006629

[40] HANSEN, B. E. (1999). Testing for linearity. *J. Econ. Surv.* **13** 551–576.

[41] HEYDE, C. C. (1997). *Quasi-Likelihood and Its Application: A General Approach to Optimal Parameter Estimation. Springer Series in Statistics*. Springer, New York. MR1461808 https://doi.org/10.1007/b98823

[42] KNIGHT, M., LEEMING, K., NASON, G. and NUNES, M. (2020). Generalized network autoregressive processes and the GNAR package. *J. Stat. Softw.* **96** 1–36.

[43] KOLACZYK, E. D. and CSÁRDI, G. (2014). *Statistical Analysis of Network Data with R. Use R!* Springer, New York. MR3288852 https://doi.org/10.1007/978-1-4939-0983-4

[44] LIM, K. S. and TONG, H. (1980). Threshold autoregressions, limit cycles, and data. *J. Roy. Statist. Soc. Ser. B* **42** 245–92.

[45] McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. *Monographs on Statistics and Applied Probability*. CRC Press, London. Second edition [of MR0727836]. MR3223057 https://doi.org/10.1007/978-1-4899-3242-6

[46] Nowicki, K. and Snijders, T. A. B. (2001). Estimation and prediction for stochastic blockstructures. *J. Amer. Statist. Assoc.* **96** 1077–1087. MR1947255 https://doi.org/10.1198/016214501753208735

[47] Pedeli, X. and Karlis, D. (2011). A bivariate INAR(1) process with application. *Stat. Model.* **11** 325–349. MR2906704 https://doi.org/10.1177/1471082X1001100403

[48] Pedeli, X. and Karlis, D. (2013). On composite likelihood estimation of a multivariate INAR(1) model. *J. Time Series Anal.* **34** 206–220. MR3028367 https://doi.org/10.1111/jtsa.12003

[49] Pedeli, X. and Karlis, D. (2013). Some properties of multivariate INAR(1) processes. *Comput. Statist. Data Anal.* **67** 213–225. MR3079598 https://doi.org/10.1016/j.csda.2013.05.019

[50] Pollard, D. (1990). *Empirical Processes: Theory and Applications*. *NSF-CBMS Regional Conference Series in Probability and Statistics* **2**. IMS. MR1089429

[51] Seber, G. A. F. (2008). *A Matrix Handbook for Statisticians*. *Wiley Series in Probability and Statistics*. Wiley-Interscience, Hoboken. MR2365265

[52] Taniguchi, M. and Kakizawa, Y. (2000). *Asymptotic Theory of Statistical Inference for Time Series*. *Springer Series in Statistics*. Springer, New York. MR1785484 https://doi.org/10.1007/978-1-4612-1162-4

[53] Teräsvirta, T. (1994). Specification, estimation, and evaluation of smooth transition autoregressive models. *J. Amer. Statist. Assoc.* **89** 208–218.

[54] Teräsvirta, T., Tjøstheim, D. and Granger, C. W. J. (1994). Aspects of modelling nonlinear time series. In *Handbook of Econometrics*, *Vol. IV*. *Handbooks in Econom.* **2** 2917–2957. North-Holland, Amsterdam. MR1315983

[55] Teräsvirta, T., Tjøstheim, D. and Granger, C. W. J. (2010). *Modelling Nonlinear Economic Time Series*. *Advanced Texts in Econometrics*. Oxford Univ. Press, Oxford. MR3185399 https://doi.org/10.1093/acprof:oso/9780199587148.001.0001

[56] Tong, H. (1990). *Nonlinear Time Series*. *Oxford Statistical Science Series* **6**. Clarendon, New York. A dynamical system approach, With an appendix by K. S. Chan, Oxford Science Publications. MR1079320

[57] Tsagris, M., Armillotta, M. and Fokianos, K. (2023). PNAR: Poisson network autoregressive models. R package version 1.6. Available at https://CRAN.R-project.org/package=PNAR.

[58] Wang, C., Liu, H., Yao, J., Davis, R. A. and Li, W. K. (2014). Self-excited threshold Poisson autoregression. *J. Amer. Statist. Assoc.* **109** 777–787. MR3223749 https://doi.org/10.1080/01621459.2013.872994

[59] Wang, Y. J. and Wong, G. Y. (1987). Stochastic blockmodels for directed graphs. *J. Amer. Statist. Assoc.* **82** 8–19. MR0883333

[60] Wasserman, S., Faust, K. et al. (1994). *Social Network Analysis: Methods and Applications* 8. Cambridge Univ. Press, Cambridge.

[61] Woodard, D. B., Matteson, D. S. and Henderson, S. G. (2011). Stationarity of generalized autoregressive moving average models. *Electron. J. Stat.* **5** 800–828. MR2824817 https://doi.org/10.1214/11-EJS627

[62] Yaskov, P. (2015). Variance inequalities for quadratic forms with applications. *Math. Methods Statist.* **24** 309–319. MR3437388 https://doi.org/10.3103/S1066530715040055

[63] Zhou, J., Huang, D. and Wang, H. (2017). A dynamic logistic regression for network link prediction. *Sci. China Math.* **60** 165–176. MR3585056 https://doi.org/10.1007/s11425-015-0807-8

[64] Zhou, J., Li, D., Pan, R. and Wang, H. (2020). Network GARCH model. *Statist. Sinica* **30** 1723–1740. MR4260742 https://doi.org/10.5705/ss.202018.0234

[65] Zhu, X. and Pan, R. (2020). Grouped network vector autoregression. *Statist. Sinica* **30** 1437–1462. MR4257540 https://doi.org/10.5705/ss.202017.0533

[66] Zhu, X., Pan, R., Li, G., Liu, Y. and Wang, H. (2017). Network vector autoregression. *Ann. Statist.* **45** 1096–1123. MR3662449 https://doi.org/10.1214/16-AOS1476

[67] Zhu, X., Wang, W., Wang, H. and Härdle, W. K. (2019). Network quantile autoregression. *J. Econometrics* **212** 345–358. MR3994021 https://doi.org/10.1016/j.jeconom.2019.04.034

[68] Zivot, E. and Wang, J. (2003). *Modeling Financial Time Series with S-Plus®*. Springer, New York. MR2000944 https://doi.org/10.1007/978-0-387-21763-5