Generalized Linear Models Network Autoregression

Mirko Amillotta(✉), Konstantinos Fokianos, and Ioannis Krikidis

University of Cyprus, PO BOX 20537, Nicosia, Cyprus
{armillotta.mirko,fokianos,krikidis}@ucy.ac.cy

Abstract. We discuss a unified framework for the statistical analysis of streaming data obtained by networks with a known neighborhood structure. In particular, we deal with autoregressive models that make explicit the dependence of current observations to their past values and the values of their respective neighborhoods. We consider the case of both continuous and count responses measured over time for each node of a known network. We discuss least squares and quasi maximum likelihood inference. Both methods provide estimators with good properties. In particular, we show that consistent and asymptotically normal estimators of the model parameters, under this high-dimensional data generating process, are obtained after optimizing a criterion function. The methodology is illustrated by applying it to wind speed observed over different weather stations of England and Wales.

Keywords: Adjacency matrix · autocorrelation · least squares estimation · link function · multivariate time series · network analysis · quasi-likelihood estimation

1 Introduction

Measuring the impact of a network structure to a multivariate time series process has attracted considerable attention over the last years, mainly due to the growing availability of streaming network data (social networks, GPS data, epidemics, air pollution monitoring systems and more generally environmental wireless sensor networks, among many other applications). The methodology outlined in this work has potential application in several network science fields. In general, any stream of data for a sample of units whose relations can be modeled as an adjacency matrix (neighborhood structure) the statistical techniques reviewed in this work are directly applicable. Indeed, a wide variety of available spatial streaming data related to physical phenomena can fit this framework. As an illustrative example, we analyze wind speed data observed over different weather stations

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

© Springer Nature Switzerland AG 2022
P. Ribeiro et al. (Eds.): NetSci-X 2022, LNCS 13197, pp. 112–125, 2022.
https://doi.org/10.1007/978-3-030-97240-0_9

of England and Wales. Network autoregressions allows meaningful analysis of the actual wind speed, for each node, based on the effect of past speeds and the velocity measured on its neighbor stations; see Sect. 4. This methodology is potentially useful to model sensor networks for environmental monitoring. See [6,8,22,25], among others, who discuss application of wireless sensor network for environmental, agricultural and intelligent home automation systems. See also [41] for an application to social network analysis. We discuss a statistical framework which encompasses the case of both continuous and count responses measured over time for each node of a known network.

1.1 The Case of Continuous Responses

When a response random variable, say $Y_{i,t}$, is measured for each node i of a known network, with N nodes, at time t, a $N \times 1$-dimensional random vector is obtained, say $\mathbf{Y}_t \in \mathbb{R}^N = (Y_{1,t} \dots Y_{i,t} \dots Y_{N,t})'$, for each measured time $t = 1, \dots, T$. The Vector Autoregressive (VAR) model, is a standard tool for continuous time series analysis and it has been widely applied to model multivariate processes. However, if the size of the network is N, then the number of unknown parameters to be estimated is of the order $\mathcal{O}(N^2)$ which is much larger than the temporal sample size T. The VAR model cannot then be applied for modeling such data.

Other modelling strategies have been proposed to describe the dynamics of such processes. One method is based on sparsity, see for example [21], among other. Accordingly, the parameters of the model which have less impact to the response are automatically set to zero, allowing to estimate the remaining ones. Alternatively, a dimension reduction method which accounts for network impact has been recently developed by [41], who introduced the Network vector Autoregressive model (NAR). In this methodology, for each node $i = 1, \dots, N$ the current response, $Y_{i,t}$, for the node i, at time t, is assumed to depend only on the lagged value of the response itself, say $Y_{i,t-1}$, and the mean of the past responses computed only over the nodes connected to the node i; this can be broadly thought as a factor which accounts for the impact of the network structure to node i. The NAR representation allows considerable simplification for the final model fitted to the data as it depends only on a few parameters. In addition, such representation still includes all essential information, i.e. the impact of the past values of the response and the influence of the network neighbors on each node.

NAR models are tailored to continuous response data. The parameters of the model are estimated via ordinary least squares (OLS), under two asymptotic regimes (a) with increasing time sample size $T \to \infty$ and fixed network dimension N (which is standard assumption for multivariate time series analysis) and (b) with both N, T increasing, i.e. $\min\{N, T\} \to \infty$. The latter is important in network science, since the asymptotic behavior of the network when its dimension grows ($N \to \infty$) is a crucial interest in network analysis. In practice, when only a sample of the network is available, the results obtained under (b) guarantee that the estimates of unknown parameters of the model have good statistical properties, even if N is big and, ultimately, bigger than T.

More recently, an extension to network quantile autoregressive models has been studied by [42]. Further works in this line of research includes the grouped least squares estimation, [40], and a Network GARCH model, see [39] under the standard asymptotic regime (a). Related work was developed by [23] who specified a Generalized Network Autoregressive model (GNAR) for continuous random variables, by taking into account different layers of relationships within neighbors of the network. All network time series models discussed so far are defined in terms of Independent Identically Distributed (IID) error random innovations; such an assumption is crucial for most of theoretical analysis.

1.2 The Case of Discrete Responses

Increasing availability of discrete-valued data, from diverse applications, has advanced the growth of a rich literature on modelling and inference for count time series processes. In this contribution, we consider the generalized linear model (GLM) framework, see [27], which includes both continuous-valued time series and integer-valued processes. Likelihood inference and testing can be developed in the GLM framework. Some examples of GLM models for count processes include the works by [9,15] and [14], among others. In [17] and [19], stability conditions and inference for linear and log-linear count time series models are developed. Further related contributions can be found in [5] for inference of negative binomial time series, [1,7,10,11] and [12], among others, for further generalizations. Even though a vast literature on the univariate case is available, results on multivariate count time series models for network data are still missing; see [26,30–32] for some exceptions. Recently [18], introduced multivariate linear and log-linear Poisson autoregression models. These authors described the joint distribution of the counts by means of a copula construction. Copulas are useful because of Sklar's theorem which shows that marginal distributions are combined to give a joint distribution when applying a copula, i.e. a N-dimensional distribution function all of whose marginals are standard uniforms. Further details are also available in the review of [16]. Recent work by [2] studied linear and log-linear multivariate count-valued extensions of the NAR model, called Poisson Network Autoregression (PNAR). These authors developed associated theory for the two types of asymptotic inference (a)–(b) discussed earlier, under the α-mixing property of the innovation term, see [13,33]. Intuitively, this assumption requires only *asymptotic independence* over time. The marginal distribution of the resulting count process is Poisson (but other marginals are possible including the Negative Binomial distribution) whereas the dependence among them is captured by the copula construction described in [18]. Inference relies on the Quasi Maximum Likelihood Estimation (QMLE), see [20], among others.

1.3 Outline

This paper summarizes some of the work by [41] and [2] and provides a unified framework for both continuous and integer-valued data. In addition it reviews the recent developments in this research area and illustrates the potential usefulness of this methodology. The paper is divided into three parts: Sect. 2 discusses the linear and log-linear NAR and PNAR model specifications. In Sect. 3, the quasi likelihood inference is described, for the two types of asymptotics (a)–(b). Finally, Sect. 4 reports the results of an application on a wind speed network in England and Wales, and gives a model selection procedure for the lag order of the NAR model.

Notation

For a $q \times p$-dimensional matrix $\mathbf{A}$ whose elements are a_{ij}, for $i = 1, \dots, q$, $j = 1, \dots, p$, denotes generalized matrix norm, defined as $|||\mathbf{A}|||_r = \max_{|\mathbf{x}|_r=1} |\mathbf{A}\mathbf{x}|_r$. If $r = 1$, $|||\mathbf{A}|||_1 = \max_{1\le j\le p} \sum_{i=1}^{q} |a_{ij}|$. $|||\mathbf{A}|||_2 = \rho^{1/2}(\mathbf{A}'\mathbf{A})$, where $\rho(\cdot)$ is the spectral radius, if $r = 2$. $|||\mathbf{A}|||_\infty = \max_{1\le i\le q} \sum_{j=1}^{p} |a_{ij}|$, if $r = \infty$. If $q = p$, then these norms are matrix norms.

2 Models

We study a network of size N (number of nodes), indexed by $i = 1, \dots N$, and adjacency matrix $\mathbf{A} = (a_{ij}) \in \mathbb{R}^{N\times N}$ where $a_{ij} = 1$, if there is a directed edge from i to j, $i \to j$ (e.g. user i follows user j on Twitter), and $a_{ij} = 0$ otherwise. Undirected graphs are also allowed ($i \leftrightarrow j$). The neighborhood structure is assumed to be known but self-relationships are not allowed, i.e. $a_{ii} = 0$ for any $i = 1, \dots, N$ (this is reasonable because e.g. user i cannot follow himself). For more on networks see [24,36]. Define a variable $Y_{i,t} \in \mathbb{R}$ for the node i at time t. The interest in on assessing the effect of the network structure on the stochastic process $\{\mathbf{Y}_t = (Y_{i,t},\, i = 1, 2 \dots N,\, t = 0, 1, 2 \dots, T)\}$, with the corresponding N-dimensional conditional mean process defined in the following way $\{\boldsymbol{\lambda}_t = (\lambda_{i,t},\, i = 1, 2 \dots N,\, t = 1, 2 \dots, T)\}$, where $\boldsymbol{\lambda}_t = \mathrm{E}(\mathbf{Y}_t|\mathcal{F}_{t-1})$ and $\mathcal{F}_{t-1} = \sigma(\mathbf{Y}_s : s \le t-1)$ is the σ-algebra generated by the past of the process.

2.1 NAR Model

For $i = 1, \dots, N$, the Network Autoregressive model of order 1, NAR(1), is given by

$$\lambda_{i,t} = \beta_0 + \beta_1 n_i^{-1} \sum_{j=1}^{N} a_{ij} Y_{j,t-1} + \beta_2 Y_{i,t-1}\,, \tag{1}$$

where $n_i = \sum_{j\ne i} a_{ij}$ is the out-degree, i.e. the total number of nodes which i has an edge with. The NAR(1) model implies that, for every single node i, the conditional mean of the process is regressed on the past of the variable itself for node i and the weighted average over the other nodes $j \neq i$ which have a connection with i. Hence only the nodes which are directly followed by the focal node i (neighborhoods) may have an impact on the mean process of the focal node i. It is a reasonable assumption in many applications; for example, in a social network the activity of node k, which satisfies $a_{ik} = 0$, does not affect node i. However, extensions to several layers of neighborhoods are also possible, see [23] and [2, Rem. 2]. The parameter β_1 is called network effect and it measures the average impact of node i's connections $n_i^{-1} \sum_{j=1}^{N} a_{ij} Y_{j,t-1}$. The coefficient β_2 is called autoregressive (or lagged) effect because it provides a weight for the impact of past process $Y_{i,t-1}$.

For a continuous-valued time series Y_t, [41] defined $Y_{i,t} = \lambda_{i,t} + \xi_{i,t}$, where $\lambda_{i,t}$ is specified in (1) and $\xi_{i,t} \sim IID(0, \sigma^2)$ across both $1 \leq i \leq N$ and $0 \leq t \leq T$ and with finite fourth moment. Then first two moments of the process $\mathbf{Y}_t$ modelled by (1) are given by [41, Prop. 1]

$$\begin{aligned}
\mathrm{E}(\mathbf{Y}_t) &= \beta_0 (1 - \beta_1 - \beta_2)^{-1} \mathbf{1}_N \,, \\
\mathrm{vec}[\mathrm{Var}(\mathbf{Y}_t)] &= \sigma^2 (\mathbf{I}_{N^2} - \mathbf{G} \otimes \mathbf{G})^{-1} \mathrm{vec}(\mathbf{I}_N) \,,
\end{aligned}$$

where $\mathbf{1}_N = (1, 1, \ldots, 1)' \in \mathbb{R}^N$ and $\mathbf{I}_N$ is the identity matrix $N \times N$ and $\mathbf{G} = \beta_1 \mathbf{W} + \beta_2 \mathbf{I}_N$, with $\mathbf{W} = \mathrm{diag}\left\{ n_1^{-1}, \ldots, n_N^{-1} \right\} \mathbf{A}$ being the row-normalized adjacency matrix. Note that the matrix $\mathbf{W}$ is a stochastic matrix, as $|||\mathbf{W}|||_\infty = 1$, [34, Def. 9.16].

More generally, the NAR(p) model is defined by

$$\lambda_{i,t} = \beta_0 + \sum_{h=1}^{p} \beta_{1h} \left(n_i^{-1} \sum_{j=1}^{N} a_{ij} Y_{j,t-h} \right) + \sum_{h=1}^{p} \beta_{2h} Y_{i,t-h} \,, \tag{2}$$

allowing dependence on the last p values of the response node. Obviously, when $p = 1$, $\beta_{11} = \beta_1$, $\beta_{22} = \beta_2$ and we obtain (1). Without loss of generality, coefficients can be set equal to zero if the parameter order is different for the summands of (2).

2.2 PNAR Model

Consider the process $Y_{i,t}$, for $i = 1, \ldots, N$, is integer-valued (that is $\mathbf{Y}_t \in \mathbb{N}^N$) and it is assumed to be marginally Poisson, such as $Y_{i,t} | \mathcal{F}_{t-1} \sim Poisson(\lambda_{i,t})$. Other models can be developed, including the Negative Binomial distribution, but the marginal mean has to parameterized as in (1). The univariate conditional mean of the count process is still specified as (1), more generally (2), above. The interpretation of all coefficients is identical to the case of continuous-valued case. The innovation term is given by $\boldsymbol{\xi}_t = \mathbf{Y}_t - \boldsymbol{\lambda}_t$ and forms a martingale difference sequence by construction but, in general, it is not an IID sequence. This adds a level of complexity in the model because a joint count distribution is required for modelling and inference. Several alternatives of multivariate Poisson-type probability mass function (p.m.f) have been proposed in the literature, see the review in [16, Sect. 2]. However, they usually have a complicated closed form, the associated inference is theoretically cumbersome, and numerically difficult; moreover, the resulting model is largely constrained. Then, a copula approach has been preferred as in [2], where the joint distribution of the vector $\{\mathbf{Y}_t\}$ is constructed imposing a copula structure on waiting times of a Poisson process, see [18, p. 474]. More precisely, consider a set of values $(\beta_0, \beta_1, \beta_2)'$ and a starting vector $\boldsymbol{\lambda}_0 = (\lambda_{1,0}, \ldots, \lambda_{N,0})'$,

1. Let $\mathbf{U}_l = (U_{1,l}, \ldots, U_{N,l})$, for $l = 1, \ldots, L$ a sample from a N-dimensional copula $C(u_1, \ldots, u_N)$, where $U_{i,l}$ follows a Uniform(0,1) distribution, for $i = 1, \ldots, N$.
2. The transformation $X_{i,l} = -\log U_{i,l}/\lambda_{i,0}$ is exponential with parameter $\lambda_{i,0}$, for $i = 1, \ldots, N$.
3. If $X_{i,1} > 1$, then $Y_{i,0} = 0$, otherwise $Y_{i,0} = \max\left\{k \in [1, K] : \sum_{l=1}^{k} X_{i,l} \le 1\right\}$, by taking K large enough. Then, $Y_{i,0} \sim Poisson(\lambda_{i,0})$, for $i = 1, \ldots, N$. So, $\mathbf{Y}_0 = (Y_{1,0}, \ldots, Y_{N,0})$ is a set of marginal Poisson processes with mean $\boldsymbol{\lambda}_0$.
4. By using the model (1), $\boldsymbol{\lambda}_1$ is obtained.
5. Return back to step 1 to obtain $\mathbf{Y}_1$, and so on.

This constitutes an innovative data generating process with desired Poisson marginal distributions and flexible correlation. With the distribution structure presented above, the resulting model for the count process $\mathbf{Y}_t$, with conditional mean specified as in (1) for all i, has been introduced by [2], called linear Poisson Network Autoregression of order 1, PNAR(1), written in matrix notation:

$$\mathbf{Y}_t = \mathbf{N}_t(\boldsymbol{\lambda}_t), \quad \boldsymbol{\lambda}_t = \boldsymbol{\beta}_0 + \mathbf{G}\mathbf{Y}_{t-1}\,, \tag{3}$$

where $\{\mathbf{N}_t\}$ is a sequence of independent N-variate copula-Poisson process (see above), which counts the number of events in the time intervals $[0, \lambda_{1,t}] \times \cdots \times [0, \lambda_{N,t}]$. Moreover, $\boldsymbol{\beta}_0 = \beta_0 \mathbf{1}_N \in \mathbb{R}^N$. By considering the conditional mean specified as in (2) for all i, it is immediate to define the PNAR(p) model:

$$\mathbf{Y}_t = \mathbf{N}_t(\boldsymbol{\lambda}_t), \quad \boldsymbol{\lambda}_t = \boldsymbol{\beta}_0 + \sum_{h=1}^{p} \mathbf{G}_h \mathbf{Y}_{t-h}\,, \tag{4}$$

where $\mathbf{G}_h = \beta_{1h}\mathbf{W} + \beta_{2h}\mathbf{I}_N$ for $h = 1, \ldots, p$. Clearly, $\lambda_{i,t} > 0$ so $\beta_0, \beta_{1h}, \beta_{2h} \ge 0$ for all $h = 1 \ldots, p$. Although the network effect β_1 of model (1) is typically expected to be positive, see [4], in order to allow a connection to the wider GLM theory, [27], and allow coefficients which take values on the entire real line the following log-linear version of the PNAR(p) is proposed in [2]:

$$\nu_{i,t} = \beta_0 + \sum_{h=1}^{p} \beta_{1h} \left(n_i^{-1} \sum_{j=1}^{N} a_{ij} \log(1 + Y_{j,t-h}) \right) + \sum_{h=1}^{p} \beta_{2h} \log(1 + Y_{i,t-h})\,, \tag{5}$$

where $\nu_{i,t} = \log(\lambda_{i,t})$ for every $i = 1, \ldots, N$. The model (5) do not require any constraints on the parameters, since $\nu_{i,t} \in \mathbb{R}$. The interpretation of coefficients and the summands of (5) is similar to that of linear model but in the log scale.

The condition $\sum_{h=1}^{p}(|\beta_{1h}| + |\beta_{2h}|) < 1$ is sufficient to obtain the process $\{\mathbf{Y}_t,\ t \in \mathbb{Z}\}$ to be stationary and ergodic for every Network Autoregressive model of order p. See [41, Thm. 4] and [2, Thm. 1–2]. For model (3), such stationary distribution has the first two moments

$$\mathrm{E}(\mathbf{Y}_t) = (\mathbf{I}_N - \mathbf{G})^{-1}\boldsymbol{\beta}_0 = \beta_0(1-\beta_1-\beta_2)^{-1}\mathbf{1}_N\,,$$
$$\mathrm{vec}[\mathrm{Var}(\mathbf{Y}_t)] = (\mathbf{I}_{N^2} - \mathbf{G}\otimes\mathbf{G})^{-1}\mathrm{vec}[\mathrm{E}(\boldsymbol{\Sigma}_t)]\,,$$

where $\boldsymbol{\Sigma}_t = \mathrm{E}(\boldsymbol{\xi}_t\boldsymbol{\xi}_t'|\mathcal{F}_{t-1})$ denotes the *true* conditional covariance matrix of the vector $\mathbf{Y}_t$.

3 Inference

We approach the estimation problem by using the theory of estimating functions; see [3,37] and [20], among others. Consider the vector of unknown parameters $\boldsymbol{\theta} = (\beta_0, \beta_{11}, \ldots, \beta_{1p}, \beta_{21}, \ldots, \beta_{2p})' \in \mathbb{R}^m$, satisfying the stationarity condition, where $m = 2p+1$. Define the quasi-log-likelihood function for $\boldsymbol{\theta}$ as $l_{NT}(\boldsymbol{\theta}) = \sum_{t=1}^{T}\sum_{i=1}^{N} l_{i,t}(\boldsymbol{\theta})$, which is not constrained to be the *true* log-likelihood of the process. The quasi maximum likelihood estimator (QMLE) is the vector of parameters $\hat{\boldsymbol{\theta}}$ which maximize the quasi-log-likelihood $l_{NT}(\boldsymbol{\theta})$. Such maximization is performed by solving the system of equations $S_{NT}(\boldsymbol{\theta}) = \mathbf{0}_m$, with respect to $\boldsymbol{\theta}$, where $\mathbf{S}_{NT}(\boldsymbol{\theta}) = \partial l_{NT}(\boldsymbol{\theta})/\partial\boldsymbol{\theta} = \sum_{t=1}^{T}\mathbf{s}_{Nt}(\boldsymbol{\theta})$ is the quasi-score function, and $\mathbf{0}_m$ is a $m\times 1$-dimensional vector of 0's. Moreover define the matrices

$$\mathbf{H}_{NT}(\boldsymbol{\theta}) = -\frac{\partial^2 l_{NT}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}, \quad B_{NT}(\theta) = \mathrm{E}\left(\sum_{t=1}^{T}\mathbf{s}_{Nt}(\boldsymbol{\theta})\mathbf{s}_{Nt}(\boldsymbol{\theta})'\middle|\mathcal{F}_{t-1}\right)\,, \tag{6}$$

as the sample Hessian matrix and the sample conditional information matrix, respectively. We drop the dependence on $\boldsymbol{\theta}$ when a quantity is evaluated at the true value $\boldsymbol{\theta}_0$.

Define $X_{i,t} = n_i^{-1}\sum_{j=1}^{N} a_{ij}Y_{j,t-1}$ and $\mathbf{Z}_{i,t-1} = (1, X_{i,t-1}, Y_{i,t-1})'$. For continuous variables, the QMLE estimator for the NAR(1) model defined in (1) maximizes the quasi-log-likelihood

$$l_{NT}(\boldsymbol{\theta}) = -\sum_{t=1}^{T}(\mathbf{Y}_t - \mathbf{Z}_{t-1}\boldsymbol{\theta})'(\mathbf{Y}_t - \mathbf{Z}_{t-1}\boldsymbol{\theta})\,, \tag{7}$$

where $\mathbf{Z}_{t-1} = (\mathbf{Z}_{1,t-1}, \ldots, \mathbf{Z}_{N,t-1})' \in \mathbb{R}^{N\times m}$, with associated score function

$$\mathbf{S}_{NT}(\boldsymbol{\theta}) = \sum_{t=1}^{T}\mathbf{Z}_{t-1}'(\mathbf{Y}_t - \mathbf{Z}_{t-1}\boldsymbol{\theta})\,. \tag{8}$$

The maximization problem (8) has a closed form solution,

$$\hat{\boldsymbol{\theta}} = \left(\sum_{t=1}^{T}\mathbf{Z}_{t-1}'\mathbf{Z}_{t-1}\right)^{-1}\sum_{t=1}^{T}\mathbf{Z}_{t-1}'\mathbf{Y}_t \tag{9}$$

which is equivalent to perform an OLS estimation of the model $\mathbf{Y}_t = \mathbf{Z}_{t-1}\boldsymbol{\theta} + \boldsymbol{\xi}_t$. The extension to the NAR(p) model is straightforward, by defining $\mathbf{Z}_{i,t-1} = (1, X_{i,t-1}, \ldots, X_{i,t-p}, Y_{i,t-1}, \ldots, Y_{i,t-p})' \in \mathbb{R}^m$, see [41, Eq. 2.13]. Under regularity assumptions on the matrix $\mathbf{W}$ and $\xi_{i,t} \sim IID(0, \sigma^2)$, the OLS estimator (9) is consistent and $\sqrt{NT}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}_m, \sigma^2\boldsymbol{\Sigma})$, as $\min\{N, T\} \to \infty$, where $\boldsymbol{\Sigma}$ is defined in [41, Eq. 2.10]. For details see [41, Thm. 3, 5]. The limiting covariance matrix $\boldsymbol{\Sigma}$ is consistently estimated by the Hessian matrix in (6), which takes the form $(NT)^{-1}\mathbf{H}_{NT} = (NT)^{-1}\sum_{t=1}^{T} \mathbf{Z}'_{t-1}\mathbf{Z}_{t-1}$. The error variance σ^2 is substituted by the sample variance $\hat{\sigma}^2 = (NT)^{-1}\sum_{i,t}(Y_{i,t} - \mathbf{Z}'_{i,t-1}\hat{\boldsymbol{\theta}})$.

For count variables, the QMLE defined in [2] maximizes the following quasi-log-likelihood

$$l_{NT}(\boldsymbol{\theta}) = \sum_{t=1}^{T}\sum_{i=1}^{N}\Big(Y_{i,t}\log\lambda_{i,t}(\boldsymbol{\theta}) - \lambda_{i,t}(\boldsymbol{\theta})\Big)\,, \tag{10}$$

which is the independence log-likelihood, such as the likelihood obtained if processes $Y_{i,t}$ defined in (4), for $i = 1, \ldots, N$ were independent. This simplifies computations but guarantees consistency and asymptotic normality of the estimator. Note that, although for this choice the joint copula structure $C(\ldots)$ does not appear in the maximization of the "working" log-likelihood (10), this does not imply that inference is carried out under the assumption of independence of the observed process; dependence is taken into account because of the dependence of the likelihood function on the past values of the process through the regression coefficients.

With the same notation, the score function is

$$\mathbf{S}_{NT}(\boldsymbol{\theta}) = \sum_{i=1}^{T}\frac{\partial\boldsymbol{\lambda}'_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\mathbf{D}_t^{-1}(\boldsymbol{\theta})\Big(\mathbf{Y}_t - \boldsymbol{\lambda}_t(\boldsymbol{\theta})\Big)\,, \tag{11}$$

where

$$\frac{\partial\boldsymbol{\lambda}_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} = (\mathbf{1}_N, \mathbf{W}\mathbf{Y}_{t-1}, \ldots, \mathbf{W}\mathbf{Y}_{t-p}, \mathbf{Y}_{t-1}, \ldots, \mathbf{Y}_{t-p})$$

is a $N \times m$ matrix and $\mathbf{D}_t(\boldsymbol{\theta})$ is the $N \times N$ diagonal matrix with diagonal elements equal to $\lambda_{i,t}(\boldsymbol{\theta})$ for $i = 1, \ldots, N$. It should be noted that (11) equals the score (8), up to a scaling matrix $\mathbf{D}_t^{-1}(\boldsymbol{\theta})$, as $\mathbf{Z}_{t-1} = \partial\boldsymbol{\lambda}_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}'$ and $\boldsymbol{\lambda}_t(\boldsymbol{\theta}) = \mathbf{Z}_{t-1}\boldsymbol{\theta}$. The Hessian matrix has the form

$$\mathbf{H}_{NT}(\boldsymbol{\theta}) = \sum_{t=1}^{T}\frac{\partial\boldsymbol{\lambda}'_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\mathbf{C}_t(\boldsymbol{\theta})\frac{\partial\boldsymbol{\lambda}_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\,, \tag{12}$$

with $\mathbf{C}_t(\boldsymbol{\theta}) = \operatorname{diag}\left\{Y_{1,t}/\lambda^2_{1,t}(\boldsymbol{\theta}) \ldots Y_{N,t}/\lambda^2_{N,t}(\boldsymbol{\theta})\right\}$ and the conditional information matrix is

$$\mathbf{B}_{NT}(\boldsymbol{\theta}) = \sum_{t=1}^{T}\frac{\partial\boldsymbol{\lambda}'_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\mathbf{D}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\mathbf{D}_t^{-1}(\boldsymbol{\theta})\frac{\partial\boldsymbol{\lambda}_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\,, \tag{13}$$

where $\mathbf{\Sigma}_t(\boldsymbol{\theta}) = \boldsymbol{\xi}_t(\boldsymbol{\theta})\boldsymbol{\xi}'_t(\boldsymbol{\theta})$ and $\boldsymbol{\xi}_t(\boldsymbol{\theta}) = \mathbf{Y}_t - \boldsymbol{\lambda}_t(\boldsymbol{\theta})$. Consider the linear PNAR(p) model (4). By [2, Thm. 3–4], under regularity assumptions on the matrix $\mathbf{W}$ and the α-mixing property of the errors $\{\xi_{i,t}, t \in \mathbb{Z}, i \in \mathbb{N}\}$, the system of equations $\mathbf{S}_{NT}(\boldsymbol{\theta}) = \mathbf{0}_m$ has a unique solution, say $\hat{\boldsymbol{\theta}}$ (QMLE), which is consistent and $\sqrt{NT}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}_m, \mathbf{H}^{-1}\mathbf{B}\mathbf{H}^{-1})$, as $\min\{N, T\} \to \infty$, where

$$\mathbf{H} = \lim_{N\to\infty} N^{-1}\mathrm{E}\left[\frac{\partial \boldsymbol{\lambda}'_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_0}\mathbf{D}_t^{-1}(\boldsymbol{\theta}_0)\frac{\partial \boldsymbol{\lambda}_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'_0}\right],$$

$$\mathbf{B} = \lim_{N\to\infty} N^{-1}\mathrm{E}\left[\frac{\partial \boldsymbol{\lambda}'_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_0}\mathbf{D}_t^{-1}(\boldsymbol{\theta}_0)\mathbf{\Sigma}_t(\boldsymbol{\theta}_0)\mathbf{D}_t^{-1}(\boldsymbol{\theta}_0)\frac{\partial \boldsymbol{\lambda}_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'}\right].$$

Both $\mathbf{H}$ and $\mathbf{B}$ are consistently estimated by (12)–(13), respectively after divided by NT and evaluated at $\hat{\boldsymbol{\theta}}$ [2, Thm. 6]. Similar results are developed for the log-linear PNAR(p) model [2, Thm. 5].

All the results of this section work immediately for the classical time series inference, with N fixed and $T \to \infty$, as a particular case.

4 Applications

4.1 Simulated Example

In this section a limited simulation example regarding the estimation of the linear PNAR model is provided. First, a network structure is generated following one of the most popular network model, the stochastic block model (SBM), [28,35] and [38] which assigns a block label $k = 1, \ldots, K$ for each node with equal probability and K is the total number of blocks. Define $\mathrm{P}(a_{ij} = 1) = \alpha N^{-0.3}$ the probability of an edge between nodes i and j, if they belong to the same block, and $\mathrm{P}(a_{ij} = 1) = \alpha N^{-1}$ otherwise. In this way, the model implicitly assumes that nodes within the same block are more likely to be connected with respect to nodes from different blocks. Here we set $K = 5$, $\alpha = 1$ and $N = 30$. This allow to obtain the weighted adjacency matrix $\mathbf{W}$. Now a vector of count variables $\mathbf{Y}_t$ is simulated according to the data generating mechanism (DGM) described in Sect. 2.2, for $t = 1, \ldots, T$, with $T = 400$ and starting value $\boldsymbol{\lambda}_0 = \mathbf{1}_N$. The PNAR(1) model is employed in the simulation with $(\beta_0, \beta_1, \beta_2) = (1, 0.3, 0.4)$. The Gaussian copula is selected in the DGM, with copula parameter $\rho = 0.5$, that is $C(u_1, \ldots, u_N) = \Phi_R\left(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_N)\right)$, where Φ^{-1} is the inverse cumulative distribution function of a standard normal and Φ_R is the joint cumulative distribution function of a multivariate normal distribution with mean vector zero and covariance matrix equal to the correlation matrix $R = \rho^{N\times N}$, i.e. an $N \times N$ matrix whose all elements are equal to ρ. Results are based on 100 simulations.

Then, a PNAR model with one and two lags is estimated for the generated data by optimizing the quasi log-likelihood (10) with the `nloptr` `R` package. Results of the estimation are presented in Table 1. The standard errors (SE) are estimated as the square root from the main diagonal of the sandwich estimator matrix $\mathbf{H}_{NT}^{-1}(\hat{\boldsymbol{\theta}})\mathbf{B}_{NT}(\hat{\boldsymbol{\theta}})\mathbf{H}_{NT}^{-1}(\hat{\boldsymbol{\theta}})$, coming from (12) and (13). The t-statistic column is given by the ratio $Estimate/SE$. The first-order estimated coefficients are significant and close to the real values while the others are not significantly different from zero, as expected.

Table 1. QML estimation results for different PNAR models.

	PNAR(1)		
	Estimate	***SE***	***t-statistic***
β_0	1.0456	0.0732	14.29
β_1	0.2999	0.0161	18.64
β_2	0.3763	0.0135	27.87
	PNAR(2)		
β_0	1.0356	0.0810	12.79
β_{11}	0.2954	0.0209	14.16
β_{12}	0.0082	0.0203	0.40
β_{21}	0.3741	0.0157	23.80
β_{22}	0.0019	0.0133	0.14

4.2 Data Example

Here an application of the network autoregressive models on real data is provided, regarding 721 wind speeds taken at each of 102 weather stations in England and Wales. By considering weather stations as nodes of the potential network, if two weather stations share a border, an edge between them will be drawn. Then, an undirected network of such stations is drawn on geographic proximity. See Fig. 1. The dataset is available in the `GNAR` `R` package [23] incorporating the time series data `vswindts` and the associated network `vswindnet`. Moreover, a character vector of the weather station location names, `vswindnames`, and coordinates of the stations in two column matrix, `vswindcoords`, are reported. Full details can be found in the help file of the `GNAR` package.

As the wind speed is continuous-valued, the NAR(p) model is estimated with $p = 1, 2, 3$ by OLS (9). The results are summarised in Table 2. Standard errors are computed as the elements on the main diagonal of the matrix $\sqrt{\hat{\sigma}^2 \sum_{t=1}^{T} \mathbf{Z}'_{t-1}\mathbf{Z}_{t-1}}$. The estimated error variance is about $\hat{\sigma}^2 \approx 0.15$ for NAR models of every order analysed. All the coefficients are significant at 5% level.

The intercept and the coefficients of the lagged effect (β_{2h}, $h = 1, 2, 3$) are always positive. In particular, the lagged effect seems to have a predominant magnitude, especially at the first lag. Some network effects are also detected but their impact tends to become small after the first lag.

The OLS estimators is the maximizer of the quasi log-likelihood (7). This allows to compare the goodness of fit performances of competing models through information criteria. We compute usual Akaike information criterion (AIC) and the Bayesian information criterion (BIC) together with the Quasi information criterion (QIC) introduced by [29]. Such information criterion is a version of the AIC which takes into account the fact that a QMLE is performed instead of the standard MLE. In fact the QIC coincides with the AIC when the quasi likelihood equals the true likelihood of the model. In Table 3, all the information criteria select the NAR(1) as the best. This means that the expected wind speed for a weather station is mainly determined by its past speed and the past wind speeds detected on close stations, which gives a reasonable interpretation in practice.

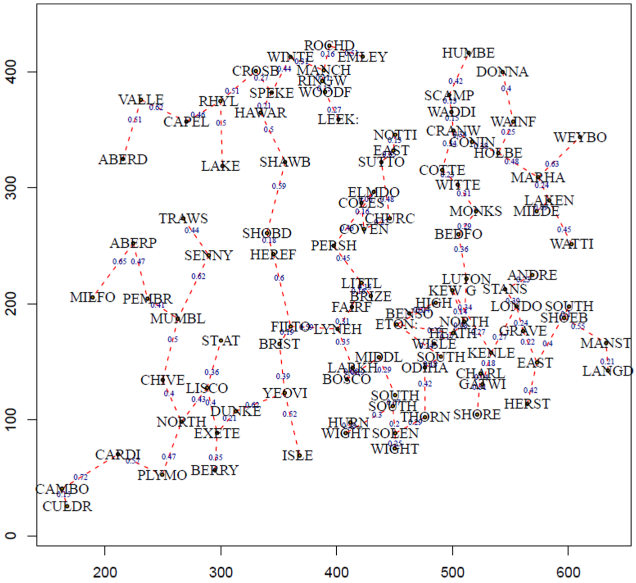

Fig. 1. Plot of the wind speed network. Geographic coordinates on the axis; numbers are relative distances between sites; labels are the site name. See [23].

Table 2. QML estimation results for wind speed data after fitting NAR(p) models for $p = 1, 2, 3$

	NAR(1)		
	Estimate	***SE***$(\times 10^2)$	***t-statistic***
β_0	0.1540	0.4616	33.37
β_1	0.1568	0.2717	57.48
β_2	0.7682	0.2429	316.26
	NAR(2)		
β_0	0.1202	0.4553	26.40
β_{11}	0.1409	0.4811	29.28
β_{12}	–0.0263	0.4806	–5.48
β_{21}	0.5828	0.3620	160.99
β_{22}	0.2442	0.3618	67.52
	NAR(3)		
β_0	0.1161	0.5297	21.91
β_{11}	0.1457	0.4927	29.56
β_{12}	–0.0116	0.5799	–2.00
β_{13}	–0.0222	0.4855	–4.56
β_{21}	0.5815	0.3623	160.53
β_{22}	0.2467	0.3637	67.84
β_{23}	0.0046	0.1763	2.63

Table 3. Information criteria for wind speed data model assessment

Model	**AIC** $(\times 10^{-3})$	**BIC** $(\times 10^{-3})$	**QIC** $(\times 10^{-3})$
NAR(1)	–22.91	–22.89	–22.91
NAR(2)	–21.49	–21.47	–21.50
NAR(3)	–21.44	–21.41	–21.45

References

1. Ahmad, A., Francq, C.: Poisson QMLE of count time series models. J. Time Ser. Anal. **37**, 291–314 (2016)
2. Armillotta, M., Fokianos, K.: Poisson network autoregression. arXiv preprint arXiv:2104.06296 (2021)
3. Basawa, I.V., Prakasa Rao, B.L.S.: Statistical Inference for Stochastic Processes. Academic Press Inc, London (1980)
4. Chen, X., Chen, Y., Xiao, P.: The impact of sampling and network topology on the estimation of social intercorrelations. J. Mark. Res. **50**, 95–110 (2013)
5. Christou, V., Fokianos, K.: Quasi-likelihood inference for negative binomial time series models. J. Time Ser. Anal. **35**, 55–78 (2014)

6. Corke, P., Wark, T., Jurdak, R., Hu, W., Valencia, P., Moore, D.: Environmental wireless sensor networks. Proc. IEEE **98**(11), 1903–1917 (2010)
7. Cui, Y., Zheng, Q.: Conditional maximum likelihood estimation for a class of observation-driven time series models for count data. Stat. Probab. Lett. **123**, 193–201 (2017)
8. Dardari, D., Conti, A., Buratti, C., Verdone, R.: Mathematical evaluation of environmental monitoring estimation error through energy-efficient wireless sensor networks. IEEE Trans. Mob. Comput. **6**(7), 790–802 (2007)
9. Davis, R.A., Dunsmuir, W.T.M., Streett, S.B.: Observation-driven models for Poisson counts. Biometrika **90**, 777–790 (2003)
10. Davis, R.A., Liu, H.: Theory and inference for a class of nonlinear models with application to time series of counts. Stat. Sin. **26**, 1673–1707 (2016)
11. Douc, R., Doukhan, P., Moulines, E.: Ergodicity of observation-driven time series models and consistency of the maximum likelihood estimator. Stochast. Process. Appl. **123**, 2620–2647 (2013)
12. Douc, R., Fokianos, K., Moulines, E.: Asymptotic properties of quasi-maximum likelihood estimators in observation-driven time series models. Electron. J. Stat. **11**, 2707–2740 (2017)
13. Doukhan, P.: Mixing: Properties and Examples. Lecture Notes in Statistics, vol. 85. Springer, New York (1994). https://doi.org/10.1007/978-1-4612-2642-0
14. Ferland, R., Latour, A., Oraichi, D.: Integer-valued GARCH process. J. Time Ser. Anal. **27**, 923–942 (2006)
15. Fokianos, K., Kedem, B.: Partial likelihood inference for time series following generalized linear models. J. Time Ser. Anal. **25**, 173–197 (2004)
16. Fokianos, K.: Multivariate count time series modelling. arXiv preprint arXiv:2103.08028 (2021)
17. Fokianos, K., Rahbek, A., Tjøstheim, D.: Poisson auto regression. J. Am. Stat. Assoc. **104**, 1430–1439 (2009)
18. Fokianos, K., Støve, B., Tjøstheim, D., Doukhan, P.: Multivariate count autoregression. Bernoulli **26**, 471–499 (2020)
19. Fokianos, K., Tjøstheim, D.: Log-linear Poisson autoregression. J. Multivar. Anal. **102**, 563–578 (2011)
20. Heyde, C.C.: Quasi-Likelihood and its Application. A General Approach to Optimal Parameter Estimation. Springer Series in Statistics. Springer, New York (1997). https://doi.org/10.1007/b98823
21. Hsu, N.J., Hung, H.L., Chang, Y.M.: Subset selection for vector autoregressive processes using Lasso. Comput. Stat. Data Anal. **52**, 3645–3657 (2008)
22. Kelly, S.D.T., Suryadevara, N.K., Mukhopadhyay, S.C.: Towards the implementation of IoT for environmental condition monitoring in homes. IEEE Sens. J. **13**(10), 3846–3853 (2013)
23. Knight, M., Leeming, K., Nason, G., Nunes, M.: Generalized network autoregressive processes and the GNAR package. J. Stat. Softw. **96**, 1–36 (2020)
24. Kolaczyk, E.D., Csárdi, G.: Statistical Analysis of Network Data with R, vol. 65. Springer, Cham (2014). https://doi.org/10.1007/978-1-4939-0983-4
25. Kularatna, N., Sudantha, B.: An environmental air pollution monitoring system based on the IEEE 1451 standard for low cost requirements. IEEE Sens. J. **8**(4), 415–422 (2008)
26. Latour, A.: The multivariate GINAR(p) process. Adv. Appl. Probab. **29**, 228–248 (1997)
27. McCullagh, P., Nelder, J.A.: Generalized Linear Models, 2nd edn. Chapman & Hall, London (1989)

28. Nowicki, K., Snijders, T.A.B.: Estimation and prediction for stochastic blockstructures. J. Am. Stat. Assoc. **96**, 1077–1087 (2001)
29. Pan, W.: Akaike's information criterion in generalized estimating equations. Biometrics **57**, 120–125 (2001)
30. Pedeli, X., Karlis, D.: A bivariate INAR(1) process with application. Stat. Model. **11**, 325–349 (2011)
31. Pedeli, X., Karlis, D.: On composite likelihood estimation of a multivariate INAR(1) model. J. Time Ser. Anal. **34**, 206–220 (2013)
32. Pedeli, X., Karlis, D.: Some properties of multivariate INAR(1) processes. Comput. Stat. Data Anal. **67**, 213–225 (2013)
33. Rosenblatt, M.: A central limit theorem and a strong mixing condition. Proc. Natl. Acad. Sci. U.S.A. **42**, 43–47 (1956)
34. Seber, G.A.F.: A Matrix Handbook for Statisticians. Wiley Series in Probability and Statistics, Wiley-Interscience. Wiley, Hoboken (2008)
35. Wang, Y.J., Wong, G.Y.: Stochastic blockmodels for directed graphs. J. Am. Stat. Assoc. **82**, 8–19 (1987)
36. Wasserman, S., Faust, K., et al.: Social Network Analysis: Methods and Applications, vol. 8. Cambridge University Press, Cambridge (1994)
37. Zeger, S.L., Liang, K.Y.: Longitudinal data analysis for discrete and continuous outcomes. Biometrics **42**, 121–130 (1986)
38. Zhao, Y., Levina, E., Zhu, J., et al.: Consistency of community detection in networks under degree-corrected stochastic block models. Ann. Stat. **40**(4), 2266–2292 (2012)
39. Zhou, J., Li, D., Pan, R., Wang, H.: Network GARCH model. Stat. Sin. **30**, 1–18 (2020)
40. Zhu, X., Pan, R.: Grouped network vector autoregression. Stat. Sin. **30**, 1437–1462 (2020)
41. Zhu, X., Pan, R., Li, G., Liu, Y., Wang, H.: Network vector autoregression. Ann. Stat. **45**, 1096–1123 (2017)
42. Zhu, X., Wang, W., Wang, H., Härdle, W.K.: Network quantile autoregression. J. Econometrics **212**, 345–358 (2019)