Count Network Autoregression

ECODEP Conference

Mirko Armillotta [1] Konstantinos Fokianos [2]

[1] Vrije Universiteit Amsterdam

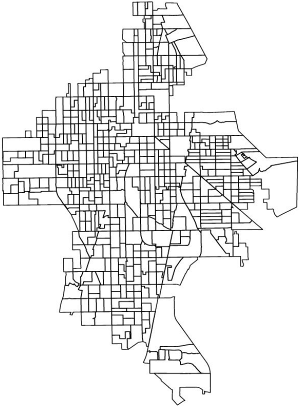
[2] University of Cyprus

February 12th, 2024

Table of contents

Monthly number of burglaries

Monthly number of burglaries on the south side of Chicago from 2010-2015. Counts registered for $N = 552$ blocks; (Clark and Dixon, 2021)



Census block groups in South Chicago. Undirected network, edge between block i and j is set if locations share a border.

Multivariate Count Autoregressions

For a recent survey, see Fokianos (2022).

Multivariate Integer AR models:

$$\boldsymbol{Y}_t = \sum_{j=1}^{p} \boldsymbol{A}_j \circ \boldsymbol{Y}_{t-j} + \boldsymbol{\epsilon}_t,$$

where $\circ$ denotes the thinning operation. Introduced by Latour (1997) (but see also Franke and Rao (1995)). Some properties of this model have been recently discussed by Pedeli and Karlis (2013a,b) and Karlis (2016).

Estimation by LSE or MLE (but computationally demanding).

Parameter Driven Models

- The observed process is driven by an unobserved process.
- A state space model for multivariate longitudinal count data has been suggested by Jørgensen et al. (1999).
- Jung et al. (2011) suggested a factor model for multivariate count time series.
- More recent contributions include Aktekin et al. (2018) (see also Gamerman et al. (2013)) Berry and West (2020), Serhiyenko (2015), Ravishanker et al. (2014), Ravishanker et al. (2015). The previous articles and the recent work of Davis et al. (2021) give further references and list other approaches.

Observation Driven Models

Fokianos et al. (2020a) studied a broad class of observation-driven models whose dynamics are driven by past observations plus noise. In particular their contribution is the following:

- Study a class of linear and log-linear models for multivariate count time series
- Prove ergodicity and stationarity by employing Markov chain theory and weak dependence approaches
- Suggest a class of estimating functions for QMLE inference and study the properties of the estimators.
- Apply these results to real data.

Assume that $\{\mathbf{Y}_t = (Y_{i,t})\}$ denotes a N–dimensional count time series and suppose further that $\{\boldsymbol{\lambda}_t = (\lambda_{i,t})\}$ is a corresponding p-dimensional intensity process, for $t = 1, 2, \ldots, T$.

Questions

- How can we describe the joint distribution of $\boldsymbol{Y}_t$ given the past?

Assume that $\{\mathbf{Y}_t = (Y_{i,t})\}$ denotes a N–dimensional count time series and suppose further that $\{\boldsymbol{\lambda}_t = (\lambda_{i,t})\}$ is a corresponding p-dimensional intensity process, for $t = 1, 2, \ldots, T$.

Questions

- How can we describe the joint distribution of $\boldsymbol{Y}_t$ given the past?
- Can we develop autoregressive models for count time series?

Multivariate Modeling 1

Assume that $\{\mathbf{Y}_t = (Y_{i,t})\}$ denotes a N–dimensional count time series and suppose further that $\{\boldsymbol{\lambda}_t = (\lambda_{i,t})\}$ is a corresponding p-dimensional intensity process, for $t = 1, 2, \ldots, T$.

Questions

- How can we describe the joint distribution of $\boldsymbol{Y}_t$ given the past?
- Can we develop autoregressive models for count time series?
- Estimation

Assume that $\{\mathbf{Y}_t = (Y_{i,t})\}$ denotes a N–dimensional count time series and suppose further that $\{\boldsymbol{\lambda}_t = (\lambda_{i,t})\}$ is a corresponding p-dimensional intensity process, for $t = 1, 2, \ldots, T$.

Questions

- How can we describe the joint distribution of $\boldsymbol{Y}_t$ given the past?
- Can we develop autoregressive models for count time series?
- Estimation
- Prediction

Multivariate Modeling 2

The multivariate linear model is given by (see also Heinen and Rengifo (2007), Jung et al. (2011), Liu (2012))

$$\begin{aligned} Y_{i,t} \mid \mathcal{F}_{t-1}^{\boldsymbol{Y},\boldsymbol{\lambda}} &\sim \text{independent} \;\; \text{Poisson}(\lambda_{i,t}), i = 1, 2, \ldots, N, \\ \boldsymbol{\lambda}_t &= \boldsymbol{d} + \boldsymbol{A}\boldsymbol{\lambda}_{t-1} + \boldsymbol{B}\boldsymbol{Y}_{t-1}, \end{aligned} \tag{1}$$

where $\boldsymbol{d}$, $\mathbf{A}$ and $\boldsymbol{B}$ are matrices with non-negative elements.

Multivariate Modeling 2

The multivariate linear model is given by (see also Heinen and Rengifo (2007), Jung et al. (2011), Liu (2012))

$$\begin{aligned} Y_{i,t} \mid \mathcal{F}_{t-1}^{\boldsymbol{Y},\boldsymbol{\lambda}} &\sim \text{independent Poisson}(\lambda_{i,t}), i = 1, 2, \ldots, N, \\ \boldsymbol{\lambda}_t &= \boldsymbol{d} + \boldsymbol{A}\boldsymbol{\lambda}_{t-1} + \boldsymbol{B}\boldsymbol{Y}_{t-1}, \end{aligned} \tag{1}$$

where $\boldsymbol{d}$, $\mathbf{A}$ and $\boldsymbol{B}$ are matrices with non-negative elements.

- The above specification implies that $Y_{i,t}$ are marginally Poisson processes with parameter $\lambda_{i,t}$, $i = 1, 2, \ldots, N$.

Multivariate Modeling 2

The multivariate linear model is given by (see also Heinen and Rengifo (2007), Jung et al. (2011), Liu (2012))

$$\begin{aligned} Y_{i,t} \mid \mathcal{F}_{t-1}^{\boldsymbol{Y},\boldsymbol{\lambda}} &\sim \text{independent Poisson}(\lambda_{i,t}), i = 1, 2, \ldots, N, \\ \boldsymbol{\lambda}_t &= \boldsymbol{d} + \boldsymbol{A}\boldsymbol{\lambda}_{t-1} + \boldsymbol{B}\boldsymbol{Y}_{t-1}, \end{aligned} \quad (1)$$

where $\boldsymbol{d}$, $\mathbf{A}$ and $\boldsymbol{B}$ are matrices with non-negative elements.

- The above specification implies that $Y_{i,t}$ are marginally Poisson processes with parameter $\lambda_{i,t}$, $i = 1, 2, \ldots, N$.
- However, their joint distribution is not multivariate Poisson, as we explain next.

Multivariate Modeling 2

The multivariate linear model is given by (see also Heinen and Rengifo (2007), Jung et al. (2011), Liu (2012))

$$\begin{aligned} Y_{i,t} \mid \mathcal{F}_{t-1}^{\boldsymbol{Y},\boldsymbol{\lambda}} &\sim \text{independent Poisson}(\lambda_{i,t}), i = 1, 2, \ldots, N, \\ \boldsymbol{\lambda}_t &= \boldsymbol{d} + \boldsymbol{A}\boldsymbol{\lambda}_{t-1} + \boldsymbol{B}\boldsymbol{Y}_{t-1}, \end{aligned} \tag{1}$$

where $\boldsymbol{d}$, $\mathbf{A}$ and $\boldsymbol{B}$ are matrices with non-negative elements.

- The above specification implies that $Y_{i,t}$ are marginally Poisson processes with parameter $\lambda_{i,t}$, $i = 1, 2, \ldots, N$.
- However, their joint distribution is not multivariate Poisson, as we explain next.
- In fact, our construction allows for dependence between $Y_{i,t}$ and $Y_{j,t}$, for $i \neq j$.

Suppose that $\boldsymbol{\lambda}_0 = (\lambda_{1,0}, \ldots, \lambda_{N,0})$ is some starting value. Then:

- Generate $\boldsymbol{U}_l = (U_{1;l}, \ldots, U_{N;l})$ for $l = 1, 2, \ldots, K$, from a copula $C(u_1, \ldots, u_N)$. Then $U_{i;l}$, $l = 1, 2, \ldots, K$ follow marginally the uniform distribution on $(0, 1)$, $i = 1, 2, \ldots, N$.

Multivariate Modeling 3

Suppose that $\boldsymbol{\lambda}_0 = (\lambda_{1,0}, \ldots, \lambda_{N,0})$ is some starting value. Then:

- Generate $\boldsymbol{U}_l = (U_{1;l}, \ldots, U_{N;l})$ for $l = 1, 2, \ldots, K$, from a copula $C(u_1, \ldots, u_N)$. Then $U_{i;l}$, $l = 1, 2, \ldots, K$ follow marginally the uniform distribution on $(0, 1)$, $i = 1, 2, \ldots, N$.
- Introduce the transformation

$$X_{i,l} = -\frac{\log U_{i,l}}{\lambda_{i,0}}, \quad i = 1, 2, \ldots, N.$$

The marginal distribution of $X_{i,l}$, $l = 1, 2, \ldots, K$ is exponential with parameter $\lambda_{i,0}$, $i = 1, 2, \ldots, N$.

Suppose that $\boldsymbol{\lambda}_0 = (\lambda_{1,0}, \ldots, \lambda_{N,0})$ is some starting value. Then:

- Generate $\boldsymbol{U}_l = (U_{1;l}, \ldots, U_{N;l})$ for $l = 1, 2, \ldots, K$, from a copula $C(u_1, \ldots, u_N)$. Then $U_{i;l}$, $l = 1, 2, \ldots, K$ follow marginally the uniform distribution on $(0, 1)$, $i = 1, 2, \ldots, N$.
- Introduce the transformation

$$X_{i,l} = -\frac{\log U_{i,l}}{\lambda_{i,0}}, \quad i = 1, 2, \ldots, N.$$

  The marginal distribution of $X_{i,l}$, $l = 1, 2, \ldots, K$ is exponential with parameter $\lambda_{i,0}$, $i = 1, 2, \ldots, N$.
- If $X_{i,1} > 1$, set $Y_{i,0} = 0$, otherwise

$$Y_{i,0} = \max\left\{K : \sum_{l=1}^{K} X_{i,l} \leq 1\right\}, \; i = 1, 2, \ldots, N.$$

  Then $\mathbf{Y}_0 = (Y_{1,0}, \ldots, Y_{N,0})$ is marginally a Poisson process with parameter $\boldsymbol{\lambda}_0$.

Multivariate Modeling 3

Suppose that $\boldsymbol{\lambda}_0 = (\lambda_{1,0}, \ldots, \lambda_{N,0})$ is some starting value. Then:

- Generate $\boldsymbol{U}_l = (U_{1;l}, \ldots, U_{N;l})$ for $l = 1, 2, \ldots, K$, from a copula $C(u_1, \ldots, u_N)$. Then $U_{i;l}$, $l = 1, 2, \ldots, K$ follow marginally the uniform distribution on $(0, 1)$, $i = 1, 2, \ldots, N$.
- Introduce the transformation

$$X_{i,l} = -\frac{\log U_{i,l}}{\lambda_{i,0}}, \quad i = 1, 2, \ldots, N.$$

  The marginal distribution of $X_{i,l}$, $l = 1, 2, \ldots, K$ is exponential with parameter $\lambda_{i,0}$, $i = 1, 2, \ldots, N$.
- If $X_{i,1} > 1$, set $Y_{i,0} = 0$, otherwise

$$Y_{i,0} = \max\left\{K : \sum_{l=1}^{K} X_{i,l} \leq 1\right\}, \; i = 1, 2, \ldots, N.$$

  Then $\mathbf{Y}_0 = (Y_{1,0}, \ldots, Y_{N,0})$ is marginally a Poisson process with parameter $\boldsymbol{\lambda}_0$.
- Use model (1) to obtain $\boldsymbol{\lambda}_1$.

Suppose that $\boldsymbol{\lambda}_0 = (\lambda_{1,0}, \ldots, \lambda_{N,0})$ is some starting value. Then:

- Generate $\boldsymbol{U}_l = (U_{1;l}, \ldots, U_{N;l})$ for $l = 1, 2, \ldots, K$, from a copula $C(u_1, \ldots, u_N)$. Then $U_{i;l}$, $l = 1, 2, \ldots, K$ follow marginally the uniform distribution on $(0, 1)$, $i = 1, 2, \ldots, N$.
- Introduce the transformation

$$X_{i,l} = -\frac{\log U_{i,l}}{\lambda_{i,0}}, \quad i = 1, 2, \ldots, N.$$

  The marginal distribution of $X_{i,l}$, $l = 1, 2, \ldots, K$ is exponential with parameter $\lambda_{i,0}$, $i = 1, 2, \ldots, N$.
- If $X_{i,1} > 1$, set $Y_{i,0} = 0$, otherwise

$$Y_{i,0} = \max\left\{K : \sum_{l=1}^{K} X_{i,l} \leq 1\right\}, \; i = 1, 2, \ldots, N.$$

  Then $\mathbf{Y}_0 = (Y_{1,0}, \ldots, Y_{N,0})$ is marginally a Poisson process with parameter $\boldsymbol{\lambda}_0$.
- Use model (1) to obtain $\boldsymbol{\lambda}_1$.
- Back to step 1 to obtain $\mathbf{Y}_1$, and so on

Multivariate Modeling 4

- Easy conceptual construction.

Multivariate Modeling 4

- Easy conceptual construction.
- Multivariate Poisson distribution available in the literature are hard to work with.

Multivariate Modeling 4

- Easy conceptual construction.
- Multivariate Poisson distribution available in the literature are hard to work with.
- Keeping the Poisson process property marginally.

Multivariate Modeling 4

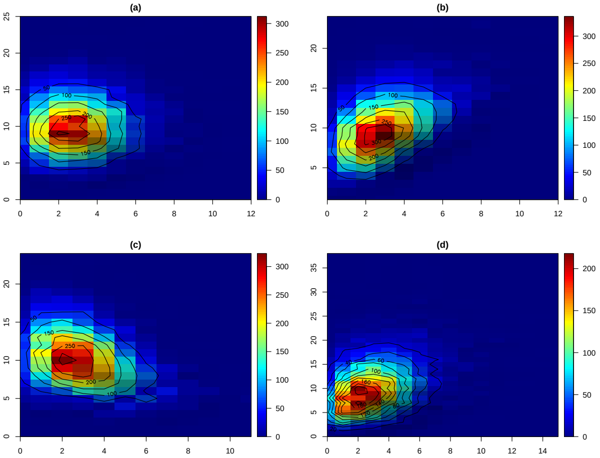
- Easy conceptual construction.
- Multivariate Poisson distribution available in the literature are hard to work with.
- Keeping the Poisson process property marginally.
- Copula is imposed on continuous random variables.

Multivariate Modeling 4

- Easy conceptual construction.
- Multivariate Poisson distribution available in the literature are hard to work with.
- Keeping the Poisson process property marginally.
- Copula is imposed on continuous random variables.
- Can be extended to other marginal count processes if they can be generated by continuous inter arrival times (mixed Poisson processes).

An example

Joint p.m.f of a bivariate count distribution using a Gaussian copula with correlation coefficient ρ. (a) $\rho = 0$ (independence) (b) $\rho = 0.8$ (positive correlation) (c) $\rho = -0.8$ (negative correlation). Plots are based on 10000 independent observations where the marginals are Poisson with $\lambda_1 = 3$ and $\lambda_2 = 10$. (d) Joint p.m.f of negative multinomial distribution.



Consider the case of $p = 2$. Then the second equation of (1) becomes

$$\begin{aligned}\lambda_{1,t} &= d_1 + a_{11}\lambda_{1,t-1} + a_{12}\lambda_{2,t-1} + b_{11}Y_{1,t-1} + b_{12}Y_{2,t-1},\\ \lambda_{2,t} &= d_2 + a_{21}\lambda_{1,t-1} + a_{22}\lambda_{2,t-1} + b_{21}Y_{1,t-1} + b_{22}Y_{2,t-1},\end{aligned}$$

where d_i is the ith element of $\mathbf{d}$ and a_{ij} (b_{ij}, respectively) is the (i,j)th element of $\mathbf{A}$ (**B**, respectively).

Consider the case of $p = 2$. Then the second equation of (1) becomes

$$\begin{aligned} \lambda_{1,t} &= d_1 + a_{11}\lambda_{1,t-1} + a_{12}\lambda_{2,t-1} + b_{11}Y_{1,t-1} + b_{12}Y_{2,t-1}, \\ \lambda_{2,t} &= d_2 + a_{21}\lambda_{1,t-1} + a_{22}\lambda_{2,t-1} + b_{21}Y_{1,t-1} + b_{22}Y_{2,t-1}, \end{aligned}$$

where d_i is the ith element of $\mathbf{d}$ and a_{ij} (b_{ij}, respectively) is the (i,j)th element of $\mathbf{A}$ (**B**, respectively).

1. When $a_{12} = b_{12} = 0$, then λ_{1t} depends only on its own past. If this is not true, then the parameters denote the linear dependence of λ_{1t} on $\lambda_{2,t-1}$ and $Y_{2,t-1}$ in the presence of $\lambda_{1,t-1}$ and $Y_{1,t-1}$.

Consider the case of $p = 2$. Then the second equation of (1) becomes

$$\begin{aligned} \lambda_{1,t} &= d_1 + a_{11}\lambda_{1,t-1} + a_{12}\lambda_{2,t-1} + b_{11}Y_{1,t-1} + b_{12}Y_{2,t-1}, \\ \lambda_{2,t} &= d_2 + a_{21}\lambda_{1,t-1} + a_{22}\lambda_{2,t-1} + b_{21}Y_{1,t-1} + b_{22}Y_{2,t-1}, \end{aligned}$$

where d_i is the ith element of $\mathbf{d}$ and a_{ij} (b_{ij}, respectively) is the (i,j)th element of $\mathbf{A}$ (**B**, respectively).

1. When $a_{12} = b_{12} = 0$, then λ_{1t} depends only on its own past. If this is not true, then the parameters denote the linear dependence of λ_{1t} on $\lambda_{2,t-1}$ and $Y_{2,t-1}$ in the presence of $\lambda_{1,t-1}$ and $Y_{1,t-1}$.
2. Similar results hold when $a_{21} = b_{21} = 0$.

Similarly, we can define a log-linear model (Fokianos and Tjøstheim (2011)) for multivariate count time series:

$$Y_{i,t} \mid \mathcal{F}_t^{\mathbf{Y},\boldsymbol{\lambda}} \sim \text{marginally Poisson}(\lambda_{i,t}), \quad \boldsymbol{\nu}_t = \mathbf{d} + \mathbf{A}\boldsymbol{\nu}_{t-1} + \mathbf{B}\log(\mathbf{Y}_{t-1} + \mathbf{1}_p), \qquad (2)$$

where $\boldsymbol{\nu}_t \equiv \log \boldsymbol{\lambda}_t$ is defined componentwise (i.e. $\nu_{i,t} = \log \lambda_{i,t}$) and $\mathbf{1}_p$ denotes the p–dimensional vector which consists of ones.
A log-linear model enjoys the following properties:

1. Modeling is done on the logarithmic scale (more suitable for count data).

Similarly, we can define a log-linear model (Fokianos and Tjøstheim (2011)) for multivariate count time series:

$$Y_{i,t} \mid \mathcal{F}_t^{\mathbf{Y},\boldsymbol{\lambda}} \sim \text{marginally} \;\; \text{Poisson}(\lambda_{i,t}), \quad \boldsymbol{\nu}_t = \mathbf{d} + \mathbf{A}\boldsymbol{\nu}_{t-1} + \mathbf{B}\log(\mathbf{Y}_{t-1} + \mathbf{1}_p), \qquad (2)$$

where $\boldsymbol{\nu}_t \equiv \log \boldsymbol{\lambda}_t$ is defined componentwise (i.e. $\nu_{i,t} = \log \lambda_{i,t}$) and $\mathbf{1}_p$ denotes the p–dimensional vector which consists of ones.

A log-linear model enjoys the following properties:

1. Modeling is done on the logarithmic scale (more suitable for count data).
2. Parameters are allowed to get negative values.

Similarly, we can define a log-linear model (Fokianos and Tjøstheim (2011)) for multivariate count time series:

$$Y_{i,t} \mid \mathcal{F}_t^{\mathbf{Y},\boldsymbol{\lambda}} \sim \text{marginally} \;\; \text{Poisson}(\lambda_{i,t}), \quad \boldsymbol{\nu}_t = \mathbf{d} + \mathbf{A}\boldsymbol{\nu}_{t-1} + \mathbf{B}\log(\mathbf{Y}_{t-1} + \mathbf{1}_p), \qquad (2)$$

where $\boldsymbol{\nu}_t \equiv \log \boldsymbol{\lambda}_t$ is defined componentwise (i.e. $\nu_{i,t} = \log \lambda_{i,t}$) and $\mathbf{1}_p$ denotes the p–dimensional vector which consists of ones.
A log-linear model enjoys the following properties:

1. Modeling is done on the logarithmic scale (more suitable for count data).
2. Parameters are allowed to get negative values.
3. Encompasses both positive and negative correlation.

Similarly, we can define a log-linear model (Fokianos and Tjøstheim (2011)) for multivariate count time series:

$$Y_{i,t} \mid \mathcal{F}_t^{\mathbf{Y},\boldsymbol{\lambda}} \sim \text{marginally Poisson}(\lambda_{i,t}), \quad \boldsymbol{\nu}_t = \mathbf{d} + \mathbf{A}\boldsymbol{\nu}_{t-1} + \mathbf{B}\log(\mathbf{Y}_{t-1} + \mathbf{1}_p), \tag{2}$$

where $\boldsymbol{\nu}_t \equiv \log \boldsymbol{\lambda}_t$ is defined componentwise (i.e. $\nu_{i,t} = \log \lambda_{i,t}$) and $\mathbf{1}_p$ denotes the p–dimensional vector which consists of ones.
A log-linear model enjoys the following properties:

1. Modeling is done on the logarithmic scale (more suitable for count data).
2. Parameters are allowed to get negative values.
3. Encompasses both positive and negative correlation.
4. Covariates can be included.

Similarly, we can define a log-linear model (Fokianos and Tjøstheim (2011)) for multivariate count time series:

$$Y_{i,t} \mid \mathcal{F}_t^{\mathbf{Y},\boldsymbol{\lambda}} \sim \text{marginally} \ \ \text{Poisson}(\lambda_{i,t}), \quad \boldsymbol{\nu}_t = \mathbf{d} + \mathbf{A}\boldsymbol{\nu}_{t-1} + \mathbf{B}\log(\mathbf{Y}_{t-1} + \mathbf{1}_p), \qquad (2)$$

where $\boldsymbol{\nu}_t \equiv \log \boldsymbol{\lambda}_t$ is defined componentwise (i.e. $\nu_{i,t} = \log \lambda_{i,t}$) and $\mathbf{1}_p$ denotes the p–dimensional vector which consists of ones.

A log-linear model enjoys the following properties:

1. Modeling is done on the logarithmic scale (more suitable for count data).
2. Parameters are allowed to get negative values.
3. Encompasses both positive and negative correlation.
4. Covariates can be included.
5. Interpretation of the parameters as in the case of linear model.

Network Autoregression

What is a network time series?

Network N nodes, index $i = 1, \dots N \iff$ adjacency matrix $\mathbf{A} = (a_{ij}) \in \mathbb{R}^{N \times N}$

$a_{ij} = 1$, if $i \to j$ (e.g. user i follows j),
$a_{ij} = 0$, otherwise

Undirected graphs are allowed ($i \leftrightarrow j$), $\mathbf{A} = \mathbf{A}^T$.

$\mathbf{A}$ nonrandom : reasonable for various applications (e.g. social networks, space points, transportation).

Let $\mathbf{Y}_t = (Y_{i,t}, i = 1, 2 \dots N, t = 1, 2 \dots, T) \in \mathbb{R}^N$. High-dimensional

Network time series: Mult. t.s. + Network structure

Target: Assess the network effect on $\mathbf{Y}_t$ over time.

Model $\mathbf{Y}_t$ by vector autoregressive model (VAR) $\Rightarrow$ parameters $\mathcal{O}(N^2) \gg T$.

Network autoregression, **NAR(1)**, (Zhu et al., 2017):

$$Y_{i,t} = \beta_0 + \beta_1 n_i^{-1} \sum_{j=1}^{N} a_{ij} Y_{j,t-1} + \beta_2 Y_{i,t-1} + \varepsilon_{i,t}\,, \quad \varepsilon_{i,t} \sim IID(0,\sigma)\ \forall i,t$$

$n_i = \sum_{j=1}^{N} a_{ij}$ out-degree.

β_1 network effect: average impact of node i's connections $\sum_{j=1}^{N} w_{ij} Y_{j,t-1}$
β_2 autoregressive effect: impact of past $Y_{i,t-1}$

$w_{ij} = a_{ij}/n_i$ for $j = 1, \ldots, N$ weights

$\sum_{j=1}^{N} w_{ij} = 1$, for $i = 1, \ldots, N$.

Main limits:

- Only for continuous variables.
- Relies on IID assumption
- OLS

$\{\mathbf{Y}_t\}$ multiv. count time series, $\boldsymbol{\lambda}_t = \mathrm{E}(\mathbf{Y}_t|\mathcal{F}_{t-1}) \in \mathbb{R}^N_+$, $\mathcal{F}_t = \sigma(\mathbf{Y}_s, s \le t)$.

Poisson Network Autoregression, PNAR(1):

$$Y_{i,t}|\mathcal{F}_{t-1} \sim Pois(\lambda_{i,t}), \quad \lambda_{i,t} = \beta_0 + \beta_1 n_i^{-1} \sum_{j=1}^{N} a_{ij} Y_{j,t-1} + \beta_2 Y_{i,t-1} \tag{3}$$

Non IID errors, $\xi_{i,t} = Y_{i,t} - \lambda_{i,t}$, Martingale diff. (MDS)

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

$$\mathbf{G} = \beta_1 \mathbf{W} + \beta_2 \mathbf{I}_N, \qquad \mathbf{W} = \mathrm{diag}\left\{n_1^{-1}, \dots, n_N^{-1}\right\} \mathbf{A}$$

$\mathbf{W}$ nonrandom matrix carrying network information.

$\{\mathbf{N}_t\}$ is a sequence of N-variate copula-Poisson processes.

Stability Results

PNAR(p):

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

where $\beta_0, \beta_{1h}, \beta_{2h} \geq 0$ for all $h = 1 \ldots, p$. If $p = 1$, $\beta_{11} = \beta_1$, $\beta_{22} = \beta_2$ to obtain (3).

$$\mathbf{Y}_t = \mathbf{N}_t(\boldsymbol{\lambda}_t), \quad \boldsymbol{\lambda}_t = \boldsymbol{\beta}_0 + \sum_{h=1}^{p} \mathbf{G}_h \mathbf{Y}_{t-h} \,, \tag{5}$$

where $\mathbf{G}_h = \beta_{1h}\mathbf{W} + \beta_{2h}\mathbf{I}_N$, for $h = 1, \ldots, p$.

Proposition 1

Consider model (5). Suppose that $\sum_{h=1}^{p}(\beta_{1h} + \beta_{2h}) < 1$. Then the process $\{\mathbf{Y}_t,\ t \in \mathbb{Z}\}$ is stationary, ergodic and $\max_{1 \leq i \leq N} \mathrm{E}\left|Y_{i,t}\right|^r < C_r < \infty$, $\forall r \geq 1$. (even when $N \to \infty$)

Note: similarly to Multiv. ARMA models, stability conditions independent of the correlations in the innovation.

$\{\mathbf{Y}_t\}$ multiv. count time series, $\boldsymbol{\lambda}_t = \mathrm{E}(\mathbf{Y}_t|\mathcal{F}_{t-1}) \in \mathbb{R}_+^N$, $\mathcal{F}_t = \sigma(\mathbf{Y}_s, s \leq t)$.

Nonlinear Poisson Network Autoregression

$$\mathbf{Y}_t = \mathbf{N}_t(\boldsymbol{\lambda}_t), \qquad \boldsymbol{\lambda}_t = f(\mathbf{Y}_{t-1}, \mathbf{W}, \boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}) \tag{6}$$

$\mathbf{W} = \mathrm{diag}\left\{n_1^{-1}, \ldots, n_N^{-1}\right\}\mathbf{A}$ carrying network information.

$n_i = \sum_{j=1}^{N} a_{ij}$ out-degree

$f(\cdot)$ satisfies suitable smoothness conditions

- $\boldsymbol{\theta}^{(1)}$ $m_1 \times 1$ vector of linear model parameters.
- $\boldsymbol{\theta}^{(2)}$ $m_2 \times 1$ vector of nonlinear parameters.

$\{\mathbf{N}_t\}$ is a sequence of N-variate copula-Poisson processes. (Fokianos et al., 2020b)

Why linear models?

- Evidence of significant usefulness of nonlinear model (e.g. modelling economic/financial time series, existence of different states of the world or regimes (Zivot and Wang, 2006, Ch. 18))
- Government agencies, research institutes and central banks may typically employ nonlinear models (Teräsvirta et al., 2010, p. 16).
- In social network analysis nonlinear behaviors are often encountered; e.g. “superstars" with huge number of followers having an exponentially higher impact on other users' behavior with respect to the “standard" user (Zhu et al., 2017).

Nonlinear model examples

- **Intercept drift NAR (ID-NAR)**, $\gamma \geq 0$, linearity $\gamma = 0$

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

- **Smooth Transition NAR (ST-NAR)**, $\gamma \geq 0$ smoothing par., lin. $\alpha = 0$

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

- **Threshold NAR (T-NAR)**, lin. $\alpha_0 = \alpha_1 = \alpha_2 = 0$

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

$I(\cdot)$ indicator function, γ is the threshold par.

Many other models fall within this framework; see Teräsvirta et al. (2010).

Stability condition

Define $f(\cdot, \mathbf{W}, \boldsymbol{\theta}) = f(\cdot)$.

(I) Set $\mathbf{F} = \mu_1 \mathbf{W} + \mu_2 \mathbf{I}_N$, $\mu_1, \mu_2 \geq 0$ and

$$|f(y) - f(y^*)|_{vec} \preceq \mathbf{F} \, |y - y^*|_{vec} \; ,$$

Theorem 1

Consider model (6). Suppose (I) holds with $\mu_1 + \mu_2 < 1$. Then, when $N \to \infty$, there exists a unique strictly stationary solution $\{\mathbf{Y}_t \in \mathbb{N}^N, \; t \in \mathbb{Z}\}$ to the Nonlinear Poisson NAR model. Moreover, $\max_{1 \leq i < \infty} \mathrm{E} \left| Y_{i,t} \right|^r \leq C_r < \infty$, $\forall r \geq 1$.

Def. stationarity with increasing dimension (Zhu et al., 2017).

- **NAR:** $\beta_1 + \beta_2 < 1$
- **ID-NAR:** $\max \{\beta_1, \beta_0 \gamma - \beta_1\} + \beta_2 < 1$
- **ST-NAR:** $\beta_1 + \beta_2 + \alpha < 1$
- ...

Log-linear model

Log-linear PNAR(p):

$$Y_{i,t}|\mathcal{F}_{t-1} \sim Poisson(\exp(\nu_{i,t})),$$

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

where $\nu_{i,t} = \log(\lambda_{i,t})$ for every $i = 1, \ldots, N$.

- Better link to the GLM theory (McCullagh and Nelder, 1989).
- Allows covariates and coefficients in $\mathbb{R}$.

Analogous results established.

Quasi maximum likelihood inference

For parameters $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}_+^m$, quasi log-likelihood:

$$l_{NT}(\boldsymbol{\theta}) = \sum_{t=1}^{T} \sum_{i=1}^{N} \left(Y_{i,t} \log \lambda_{i,t}(\boldsymbol{\theta}) - \lambda_{i,t}(\boldsymbol{\theta}) \right) \tag{7}$$

Copula structure $C(\ldots, \rho)$ not included. (7) allows inference.

$$\mathbf{S}_{NT}(\boldsymbol{\theta}) = \frac{\partial l_{NT}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} = \sum_{t=1}^{T} \mathbf{s}_{Nt}(\boldsymbol{\theta}),$$

$$\mathbf{H}_N = \mathrm{E}\left[-\frac{\partial^2 l_{NT}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right], \quad \mathbf{B}_N = \mathrm{E}\left[\mathbf{s}_{Nt}(\boldsymbol{\theta}_0) \mathbf{s}'_{Nt}(\boldsymbol{\theta}_0) \right]$$

- N can be large in applications $\Longrightarrow$ Interest in the asymptotics with $N \to \infty$.

Main result

Theorem 2

Under mild assumptions as $\{N, T_N\} \to \infty$, the equation $\mathbf{S}_{NT}(\boldsymbol{\theta}) = \mathbf{0}_m$ has a unique solution, $\hat{\boldsymbol{\theta}}$, s.t. $\hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}_0$ and $\sqrt{NT}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(0, \mathbf{H}^{-1}\mathbf{B}\mathbf{H}^{-1})$.

where $\{N, T_N\} \to \infty$ is shorthand for $N \to \infty$ and $T_N \to \infty$.

- Result holds for all models
- Assumptions depend on network structure
- Assumption guarantee existence of Hessian and information matrices.

Why testing for linearity?

1. (*Evidence*) Provide evidence to the researcher.

2. (*Model selection*) Theory might give indication of nonlinearity, but no clue on the type of nonlinearity. Linearity tests give guidance.

3. (*Consistent inference*) Nonlinear models nesting the linear model suffer from identifiability issues, when the "true" model is linear but instead a nonlinear model is estimated. Inference will be inconsistent. (link)

4. (*Practical usefulness*) In practice, testing linearity convenient before attempting estimation of complex nonlinear models.

5. (*General inspection*) Not only to provide alternative specifications but can be used as a general tool; e.g. for detecting latent variables, change point testing, checking adequacy of Box-Cox transformations, etc.

"Thus linearity testing has to precede any nonlinear modelling and estimation" (Teräsvirta et al., 2010, Sec. 5.1,5.5).

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

where under H_0, the linear NAR model is restored. $\mathbf{S}_{NT}(\boldsymbol{\theta}) = \left(\mathbf{S}_{NT}^{(1)}(\boldsymbol{\theta}), \mathbf{S}_{NT}^{(2)}(\boldsymbol{\theta})\right)'$

Quasi-score test statistic:

$$LM_{NT} = \mathbf{S}_{NT}^{(2)\prime}(\hat{\boldsymbol{\theta}})\boldsymbol{\Sigma}_{NT}(\hat{\boldsymbol{\theta}})^{-1}\mathbf{S}_{NT}^{(2)}(\hat{\boldsymbol{\theta}}),$$

where $\boldsymbol{\Sigma}_{NT}(\hat{\boldsymbol{\theta}})$ suitable estimator for covariance matrix $\boldsymbol{\Sigma} = \mathrm{Var}[\mathbf{S}_{NT}^{(2)}(\hat{\boldsymbol{\theta}})]$.

Two cases

- Identifiable parameters:

$$LM_{NT} \xrightarrow{d} \chi^2_k$$

- Non-identifiable parameters
  - $\mathbf{S}_{NT}(\gamma)$, $LM_{NT}(\gamma)$ depend on $\gamma \implies$ Standard theory not applicable. (Davies, 1987)

- $\mathbf{S}_{NT}(\gamma) \Rightarrow \mathbf{S}(\gamma)$ and $LM_{NT}(\gamma) \Rightarrow LM(\gamma)$ where

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

  is a chi-square process.

- In general, asymptotic distribution of $g(LM(\gamma))$ cannot be tabulated.

Implementation of p-values

Bound for p-values (Davies, 1987)

$$P\left[\sup_{\gamma\in\Gamma_F}(LM(\gamma)) \geq M\right] \leq P(\chi^2_k \geq M) + VM^{\frac{1}{2}(k-1)}\frac{\exp(-\frac{M}{2})2^{-\frac{k}{2}}}{\Gamma(\frac{k}{2})}, \tag{8}$$

where M is the maximum of the test statistic $LM_{NT}(\gamma)$, computed by the available sample and $\Gamma_F = (\gamma_L, \gamma_1, \ldots, \gamma_l, \gamma_U)$ is a grid of values for $\Gamma = [\gamma_L, \gamma_U]$. V is the approximated total variation

$$V = \left|LM^{\frac{1}{2}}_{NT}(\gamma_1) - LM^{\frac{1}{2}}_{NT}(\gamma_L)\right| + \cdots + \left|LM^{\frac{1}{2}}_{NT}(\gamma_U) - LM^{\frac{1}{2}}_{NT}(\gamma_l)\right|$$

1. Simple and fast.
2. Only a bound $\Longrightarrow$ conservative test.
3. Only for scalar γ.
4. Requires differentiability of $LM(\gamma)$ w.r.t. γ (Threshold NAR)

Bootstrap on stochastic permutations (Hansen, 1996)

- $\{v_{t,b} : t = 1, \ldots, T\} \sim N(0,1)$ for $b = 1, \ldots, B$
- $\mathbf{S}^b_{NT}(\gamma) = \sum_{t=1}^{T} \mathbf{s}_{Nt}(\hat{\boldsymbol{\theta}}, \gamma) \times v_{t,b}$
- $LM^b_{NT}(\gamma)$ and $g^b_{NT} = \sup_{\gamma \in \Gamma} LM^b_{NT}(\gamma)$
- $p^B_{NT} = B^{-1} \sum_{b=1}^{B} I(g^b_{NT} \geq g_{NT})$

Does not suffer from 2-4 but time consuming when N is large.

Application

Monthly number of burglaries on the south side of Chicago from 2010-2015. Counts registered for $N = 552$ blocks. (Clark and Dixon, 2021)

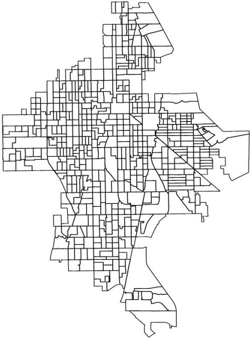


Figure 1: Census block groups in South Chicago.

Undirected network, edge between block i and j is set if locations share (at least) a border.

Table 1: Estimation results for Chicago crime data.

	Linear PNAR(1)			Log-linear PNAR(1)		
	Estimate	SE ($\times 10^2$)	p-value	Estimate	SE ($\times 10^2$)	p-value
β_0	0.4551	2.1607	$<$0.01	-0.5158	3.8461	$<$0.01
β_1	0.3215	1.2544	$<$0.01	0.4963	2.8952	$<$0.01
β_2	0.2836	0.8224	$<$0.01	0.5027	1.2105	$<$0.01
	Linear PNAR(2)			Log-linear PNAR(2)		
	Estimate	SE ($\times 10^2$)	p-value	Estimate	SE ($\times 10^2$)	p-value
β_0	0.3209	1.8931	$<$0.01	-0.5059	4.7605	$<$0.01
β_{11}	0.2076	1.1742	$<$0.01	0.2384	3.4711	$<$0.01
β_{21}	0.2287	0.7408	$<$0.01	0.3906	1.2892	$<$0.01
β_{12}	0.1191	1.4712	$<$0.01	0.0969	3.3404	$<$0.01
β_{22}	0.1626	0.7654	$<$0.01	0.2731	1.2465	$<$0.01

Table 2: Information criteria for Chicago crime data. Smaller values in bold.

	AIC$\times 10^{-3}$		BIC$\times 10^{-3}$		QIC$\times 10^{-3}$	
	linear	log-linear	linear	log-linear	linear	log-linear
PNAR(1)	115.06	115.37	115.07	115.38	115.11	115.44
PNAR(2)	**111.70**	**112.58**	**111.72**	**112.60**	**111.76**	**112.68**

Table 3: Chicago burglaries counts. Linearity is tested against:
ID-NAR model, with χ^2_1 asymptotic test;
ST-NAR model, p-values computed by (DV) Davies bound (8), bootstrap sup test (p^B_{NT});
T-NAR model (only bootstrap). Boot. replications $J = 499$.

Models	χ^2_1	DV	p^B_{NT}
ID-NAR	0.005	-	-
ST-NAR	-	0.01	0.90
T-NAR	-	-	0.77

Conclude for nonlinear shift in intercept but no clear evidence of regime switching.

Conclusion

- New useful models allowing to measure impact of networks on multivariate time series of counts.
- Non IID errors ξ_t.
- Minimal stationarity conditions.
- QMLE with standard and double asymptotics $N \to \infty$, $T \to \infty$.

Further developments

- Problem of unknown network $\Longrightarrow$ Challenging extension adjacency matrix $\mathbf{W}$ stochastic.
- Overdispersion, heavy tails, zero inflation.
- More suitable estimation tools (GEE).
- Time-varying networks
- ...
- Suggestions are welcome!

- M. Armillotta and K. Fokianos: “Poisson Network Autoregression", 2024, to appear in *Journal of Time Series Analysis*

  Available at https://arxiv.org/pdf/2104.06296.pdf

- M. Armillotta and K. Fokianos: “Nonlinear Network Autoregression", 2023, *Annals of Statistics*

  Available at https://arxiv.org/pdf/2202.03852.pdf

- M. Tsagris, M. Armillotta, K. Fokianos. **R Package ‘PNAR’**, 2024 to appear in *R-Journal* ,

  https://cran.r-project.org/web/packages/PNAR/index.html

Retirement may be an ending, a closing, but it is also a new beginning!!

Figure 2: RATS 2012–Protaras, Cyprus

References

Aktekin, T., N. Polson, and R. Soyer (2018). Sequential bayesian analysis of multivariate count data. *Bayesian Analysis 13*, 385 – 409.

Berry, L. R. and M. West (2020). Bayesian forecasting of many count-valued time series. *Journal of Business & Economic Statistics 38*, 872–887.

Clark, N. J. and P. M. Dixon (2021). A class of spatially correlated self-exciting statistical models. *Spatial Statistics 43*, 1–18.

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

Davis, R. A., K. Fokianos, S. H. Holan, H. Joe, J. Livse, R. Lund, V. Pipiras, and N. Ravishanker (2021). Count time series: A methodological review. *Journal of the American Statistical Association 116*, 1533–1547.

Fokianos, K. (2022). Multivariate count time series modelling. *Econometrics and Statistics*. to appear.

Fokianos, K., B. Støve, D. Tjøstheim, and P. Doukhan (2020a). Multivariate count autoregression. *Bernoulli 26*, 471–499.

Fokianos, K., B. Støve, D. Tjøstheim, and P. Doukhan (2020b). Multivariate count autoregression. *Bernoulli 26*, 471–499.

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

Franke, J. and T. S. Rao (1995). Multivariate first-order integer values autoregressions. Technical report, Department of Mathematics, UMIST.

Gamerman, D., T. R. dos Santos, and G. C. Franco (2013). A non-Gaussian family of state-space models with exact marginal likelihood. *Journal of Time Series Analysis 34*, 625–645.

Hansen, B. E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica 64*, 413–430.

Heinen, A. and E. Rengifo (2007). Multivariate autoregressive modeling of time series count data using copulas. *Journal of Empirical Finance 14*, 564 – 583.

Jørgensen, B., S. Lundbye-Christensen, P. X.-K. Song, and L. Sun (1999). A state space model for multivariate longitudinal count data. *Biometrika 86*, 169–181.

References (cont.)

Jung, R., R. Liesenfeld, and J.-F. Richard (2011). Dynamic factor models for multivariate count data: an application to stock–market trading activity. *Journal of Business & Economic Statistics 29*, 73–85.

Karlis, D. (2016). Modelling multivariate times series for counts. In R. Davis, S. Holan, R. Lund, and N. Ravishanker (Eds.), *Handbook of Discrete-Valued Time Series*, Handbooks of Modern Statistical Methods, pp. 407–424. London: CRC Press, Boca Raton, FL.

Latour, A. (1997). The multivariate GINAR(p) process. *Advances in Applied Probability 29*, 228–248.

Liu, H. (2012). *Some models for time series of counts*. Ph. D. thesis, Columbia University, USA.

McCullagh, P. and J. A. Nelder (1989). *Generalized Linear Models* (2nd ed.). London: Chapman & Hall.

Pedeli, X. and D. Karlis (2013a). On composite likelihood estimation of a multivariate INAR(1) model. *Journal of Time Series Analysis 34*, 206–220.

Pedeli, X. and D. Karlis (2013b). Some properties of multivariate INAR(1) processes. *Computational Statistics & Data Analysis 67*, 213 – 225.

Ravishanker, N., V. Serhiyenko, and M. R. Willig (2014). Hierarchical dynamic models for multivariate times series of counts. *Statistics and its Interface 7*, 559–570.

Ravishanker, N., R. Venkatesan, and S. Hu (2015). Dynamic models for time series of counts with a marketing application. In R. Davis, S. Holan, R. Lund, and N. Ravishanker (Eds.), *Handbook of Discrete-Valued Time Series*, Handbooks of Modern Statistical Methods, pp. 425–446. London: CRC Press, Boca Raton, FL.

Serhiyenko, V. (2015). *Dynamic Modeling of Multivariate Counts - Fitting, Diagnostics, and Applications*. Ph. D. thesis, University of Connecticut, USA.

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

Zhu, X., R. Pan, G. Li, Y. Liu, and H. Wang (2017). Network vector autoregression. *The Annals of Statistics 45*, 1096–1123.

Zivot, E. and J. Wang (2006). *Modelling Financial Time Series with S-PLUS*. Springer-Verlag.