Nonlinear Network Autoregressive Models

Mirko Armillotta[1] Konstantinos Fokianos[2]

[1]Vrije Universiteit Amsterdam

[2]University of Cyprus

3th May 2023, University of Crete, Greece

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).

Table of contents

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 ($i \leftrightarrow j$), $\mathbf{A} = \mathbf{A}'$.

$\mathbf{A}$ nonrandom (e.g. social networks, space points, transportation).

Let $\mathbf{Y}_t = (Y_{1,t} \dots Y_{i,t} \dots Y_{N,t})' \in \mathbb{R}^N$ for $t = 1, 2 \dots, T$.
High-dimensional (continuous or count)

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 $\mathcal{O}(N^2) \gg T$.

Nonlinear Network Autoregression

$\{\mathbf{Y}_t\}$ 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

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

$\mathbf{W} = \mathrm{diag}\left\{n_1^{-1}, \ldots, n_N^{-1}\right\} \mathbf{A}$ 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., 2020)

Start. value $\boldsymbol{\lambda}_0 = (\lambda_{1,0}, \ldots, \lambda_{N,0})$,

1. From copula $C(u_1, \ldots, u_N; \rho)$ generate $\mathbf{U}_l = (U_{1;l}, \ldots, U_{N;l})'$ for $l = 1, 2, \ldots, K$. $U_{i;l} \sim Unif(0,1)$.

2. Introduce the transformation

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

   where $Z_{i,l} \sim Exp(\lambda_{i,0})$, $l = 1, 2, \ldots, K$.

3. If $Z_{i,1} > 1$, set $Y_{i,0} = 0$, otherwise

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

   Then $\mathbf{Y}_0 = (Y_{1,0}, \ldots, Y_{N,0})'$ is (cond.) marginal Poisson: $Y_{i,0}|\boldsymbol{\lambda}_0 \sim Pois(\lambda_{i,0})$.

4. Use model (1), $\boldsymbol{\lambda}_1 = f(\mathbf{Y}_0, \mathbf{W}, \boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)})$

5. Back to step 1 to obtain $\mathbf{Y}_1$, and so on.

- Poisson-type joint distribution $\mathbf{Y}_t|\mathcal{F}_{t-1}$ problematic,
    - Complicated closed form $\rightarrow$ inference theoretically cumbersome.
    - Numerically challenging.
    - Implies strong constraints (e.g. covariances positive, constant correlations).
- Avoid complex Poisson-type joint distribution
- Easy conceptual construction.
- Keeping the Poisson process property marginally.
- Avoid identifiability problem (Sklar, 1959)
- Copula is imposed on continuous random variables.

For further details see Fokianos (2022).

For continuous r.v. set $\mathbf{Y}_t = \boldsymbol{\lambda}_t + \boldsymbol{\xi}_t$, where $\xi_{i,t} \sim IID(0, \sigma^2)$, $\forall i, t$.
(Analogous results established)

Element-wise components of (1):

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

where $f_i(\cdot)$ is the i^{th} component of $f(\cdot)$.

Lagged network mean: $X_{i,t-1} = n_i^{-1} \sum_{j=1}^{N} a_{ij} Y_{j,t-1}$.

- **Linear Network Autoregression (NAR)**, Zhu et al. (2017) (continuous r.v.) and Armillotta and Fokianos (2021) (counts)

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

β_1 network effect: average impact of node i's connections $X_{i,t-1}$
β_2 autoregressive effect: impact of past ($Y_{i,t-1}$)

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 behaviours are often encountered; e.g. "superstars" with huge number of followers having an exponentially higher impact on other users' behaviour with respect to the "standard" user (Zhu et al., 2017).

- **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 others... (go back)

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

$$\left| f(\mathbf{Y}_{t-1}) - f(\mathbf{Y}^*_{t-1}) \right|_{vec} \preceq \mathbf{F} \left| \mathbf{Y}_{t-1} - \mathbf{Y}^*_{t-1} \right|_{vec} ,$$

Theorem 1

Consider model (1). 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:** $\beta_1 + \beta_2 + \alpha < 1$
- ...

Quasi maximum likelihood inference

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

$$l_{NT}(\boldsymbol{\theta}) = \sum_{t=1}^{T} \sum_{i=1}^{N} \left(Y_{i,t} \log \lambda_{i,t}(\boldsymbol{\theta}) - \lambda_{i,t}(\boldsymbol{\theta}) \right) \tag{2}$$

Copula structure $C(\ldots, \rho)$ not included. (2) 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 $N \to \infty$.

Assumptions

Define $\mathbf{W}^* = \mathbf{W} + \mathbf{W}'$, $\boldsymbol{\xi}_t = \mathbf{Y}_t - \boldsymbol{\lambda}_t$ and $\boldsymbol{\Sigma}_{\boldsymbol{\xi}} = \mathrm{E}\,|\boldsymbol{\xi}_t \boldsymbol{\xi}_t'|_{vec}$.

(A) $\boldsymbol{\Theta}$ is compact and $\boldsymbol{\theta}_0 \in (\text{Int.}\boldsymbol{\Theta})$. At $\boldsymbol{\theta}_0$, the conditions of Thm. 1 hold.

(B) For $i = 1, \ldots, N$, $f_i(x_i, y_i, \theta) \geq C > 0$. For $g = 1, \ldots, m$

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

with $\sum_g (c_{1g} + c_{2g}) < \infty$. Analogous conditions for second and third order. (II)

(C) Consider $\{1, \ldots, N\}$ are states of an irreducible and aperiodic Markov chain, with $\mathbf{W}$ be its transition probability matrix, and $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_N)' \in \mathbb{R}^N$ the stationary distribution. Moreover:

- $\lambda_{\max}(\boldsymbol{\Sigma}_{\boldsymbol{\xi}}) \sum_{i=1}^N \pi_i^2 \to 0$ as $N \to \infty$.
- $\lambda_{\max}(\mathbf{W}^*) = \mathcal{O}(\log N)$ and $\lambda_{\max}(\boldsymbol{\Sigma}_{\boldsymbol{\xi}}) = \mathcal{O}((\log N)^\delta)$, $\delta \geq 1$.

(D) Some regularity conditions allowing $\mathbf{H} = \lim_{N\to\infty} N^{-1}\mathbf{H}_N < \infty$.

(E) $\left\{\boldsymbol{\xi}_t \in \mathbb{N}^N,\ t \in \mathbb{Z}, N \in \mathbb{N}\right\}$ is α-mixing; i.e. when $J \to \infty$

$$\alpha(J) = \sup_{t\in\mathbb{Z}, N\in\mathbb{N}} \ \sup_{A\in\mathcal{F}^N_{-\infty,t}, B\in\mathcal{F}^N_{t+J,\infty}} |\mathrm{P}(A\cap B) - \mathrm{P}(A)\mathrm{P}(B)| \to 0$$

$$\mathcal{F}^N_{-\infty,t} = \sigma\left(\xi_{i,s} : 1 \le i \le N, s \le t\right),\ \mathcal{F}^N_{t+J,\infty} = \sigma\left(\xi_{i,s} : 1 \le i \le N, s \ge t+J\right).$$

(F) (Weak dependence) There exists a non negative, non increasing sequence $\{\varphi_h\}_{h=1,\dots,\infty}$ s.t. $\sum_{h=1}^{\infty} \varphi_h = \Phi < \infty$ and, for $i < j$,

$$|\mathrm{Corr}(Y_{i,t}, Y_{j,t} \mid \mathcal{F}_{t-1})| \le \varphi_{j-i}\,.$$

Analogous conditions for second and third corr.

(Not unique) $N^{-1}\sum_{i,j=1}^{N} |\mathrm{Corr}(Y_{i,t}, Y_{j,t} \mid \mathcal{F}_{t-1})| \le \varphi_c$

Double asymptotic regime

Assumption (C)-(F) is needed, e.g. $\lambda_{i,t} = \beta_0$, for all $i = 1, \ldots, N$, no assumptions $\Rightarrow N^{-1}\mathbf{B}_N = \mathcal{O}(N)$.

Theorem 2

Consider model (1). Assume (A)-(F) hold. Then, 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 $T_N \to \infty$.

Theorem 3

If $T_N = \lambda N$, for some $\lambda > 0$ and Assumption (E) 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 $\{N, T_N\} \to \infty$, Theorem 2 holds with strong consistency, i.e. $\hat{\boldsymbol{\theta}} \xrightarrow{a.s.} \boldsymbol{\theta}_0$.

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).

Testing linearity

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

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 $\boldsymbol{\Sigma} = \mathrm{Var}[\mathbf{S}_{NT}^{(2)}(\hat{\boldsymbol{\theta}})]$.

Theorem 4

Suppose conditions of Theorem 2 hold. Then, under H_0,

$$LM_{NT} \xrightarrow{d} \chi_k^2, \quad \{N, T_N\} \to \infty.$$

Non identifiable parameters

Suppose the nonlinear function $f(\cdot)$ in (1) is

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

$\mathbf{G} = \beta_1 \mathbf{W} + \beta_2 \mathbf{I}_N$. Testing linearity

$$H_0 : \boldsymbol{\alpha} = 0 \quad \text{vs.} \quad H_1 : \boldsymbol{\alpha} \neq 0, \quad \text{componentwise}.$$

Parameters $\boldsymbol{\gamma}$ non identifiable under the null H_0.

$\mathbf{S}_{NT}(\boldsymbol{\gamma})$, $LM_{NT}(\boldsymbol{\gamma})$ depend on $\boldsymbol{\gamma} \Longrightarrow$ Standard theory not applicable. (Davies, 1987)

(II) Assumption (B) holds with all constants not depending on $\boldsymbol{\gamma} \in \Gamma$, where Γ compact. Additional moment conditions.

Testing linearity: non identifiable parameters

Theorem 5

Consider model (3) and the test $H_0 : \boldsymbol{\alpha} = 0$ vs. $H_1 : \boldsymbol{\alpha} \neq 0$. Suppose conditions of Theorem 2 and (II) hold. Then, under H_0, as $\{N, T_N\} \to \infty$, $\mathbf{S}_{NT}(\gamma) \Rightarrow \mathbf{S}(\gamma)$ and $LM_{NT}(\gamma) \Rightarrow LM(\gamma)$ where

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

is a chi-square process.

Define $g_{NT} = g(LM_{NT}(\boldsymbol{\gamma}))$, e.g. $g_{NT} = \sup_{\boldsymbol{\gamma} \in \Gamma} LM_{NT}(\boldsymbol{\gamma})$.

$$g_{NT} \Rightarrow g = g(LM(\boldsymbol{\gamma})) , \quad \{N, T_N\} \to \infty .$$

- In general, asymp. distribution of $g(LM(\boldsymbol{\gamma}))$ cannot be tabulated.

Bound for p-values (Davies, 1987)

$$\mathrm{P}\left[\sup_{\gamma\in\Gamma_F}(LM(\gamma)) \geq M\right] \leq \mathrm{P}(\chi^2_k \geq M) + VM^{\frac{1}{2}(k-1)}\frac{\exp(-\frac{M}{2})2^{-\frac{k}{2}}}{\Gamma(\frac{k}{2})}, \quad (4)$$

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)

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

Theorem 6

Assume the conditions of Theorems 3 and 5 hold. Then, as $\{N, T_N\} \to \infty$ and $B \to \infty$, $p_{NT}^{B} \Rightarrow p$.

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

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

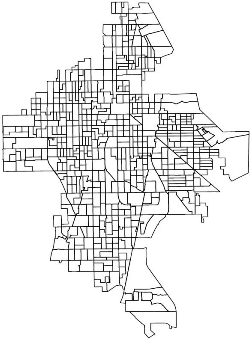

Figure 1: Census block groups in South Chicago.

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

Table 1: 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 (4), bootstrap sup test (p^B_{NT});
T-NAR model (only bootstrap). Boot. replications $J = 299$.

Models	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$
NAR	0.455	0.322	0.284
Std.	(0.022)	(0.013)	(0.008)
Models	χ^2_1	DV	p^B_{NT}
ID-NAR	8.999	-	-
ST-NAR	-	0.038	0.515
T-NAR	-	-	0.498

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

Summary

- New useful nonlinear models allowing to measure impact of networks on multivariate time series (counts and cont.).
- Very general, for $f(\cdot)$ smooth.
- Minimal stationarity conditions ($N \to \infty$).
- QMLE nonlinear NAR models with double asymptotics $N \to \infty$, $T_N \to \infty$.
- Testing linearity of NAR model parameters, standard and non identifiable case (double asymp.)
- Provide tools to compute p-values.

Further developments

- Overdispersion, zero inflation. $\Longrightarrow$ Beyond the Poisson: Negative Binomial, etc.
- Improve efficiency of estimators.
- Other ways to compute p-values.
- ...
- Suggestions are welcome!

- M. Armillotta and K. Fokianos: *Poisson Network Autoregression*, 2021+.
- M. Armillotta and K. Fokianos. *Testing Linearity for Network Autoregressive Models*, 2022.
- M. Armillotta, M. Tsagris and K. Fokianos. *The R-package PNAR for modelling count network time series*, 2022.
- M. Tsagris, M. Armillotta, K. Fokianos. **R Package 'PNAR'**, 2022.

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

- Email: `m.armillotta@vu.nl`

Many thanks for your invitation and attention!

Armillotta, M. and Fokianos, K. (2021). Poisson network autoregression. *arXiv preprint arXiv:2104.06296*.

Clark, N. J. and Dixon, P. M. (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.

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

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

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

Sklar, M. (1959). Fonctions de repartition an dimensions et leurs marges. *Publ. inst. statist. univ. Paris*, 8:229–231.

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

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

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