Mixtures of Nonlinear Poisson Autoregressions

J. Rynkiewicz[1], Joint work with P. Doukhan, K. Fokianos

1-Université de Paris I, Paris, France

Online Opening EcoDep Conference,
September 2020

Introduction

We consider models for count time series, which allow for the mean process to change according to the values of an unobservable discrete random variable.

- We introduce Markov switching non linear autoregressive Poisson models and study their properties by considering two cases :
  1. For the mixture setup, we show that infinite order models are weakly dependent.
  2. For the Markov switching case, we prove that finite order models are geometric β-mixing.
- We study the statistical properties of the maximum likelihood estimator for the case of finite order autoregression of Markov switching models.
  1. When the number of regimes is known.
  2. When the number of regimes is unknown, we show that a marginal likelihood ratio test for testing the number of hidden regimes converges to a Gaussian process.
- We apply the theoretical results to the weekly number of E.coli cases and compare mixture models to models with interventions.

- We assume that $(Z_t)_{t\in\mathbb{N}}$ and $(Y_t)_{t\in\mathbb{N}}$ are two random sequences defined on a probability space $(\Omega, \mathcal{G}, P)$ and taking values in a finite set $\mathcal{K} = \{1, \cdots, K\}$ and in $\mathbb{N}$, respectively.
- Suppose that $(Z_t)_{t\in\mathbb{N}}$ is the unobserved "state sequence" which is a time-homogeneous Markov chain with transition probability matrix $Q = (q_{i,j})_{1\leq i,j\leq K}$, i.e., for any integer $t \in \mathbb{N}$, and for any $i, j \in \mathcal{K}$,

$$P(Z_{t+1} = j \,|\, Z_t = i) = q_{i,j}. \tag{1}$$

  Moreover, $p = (p_i)_{i\in\mathcal{K}}$ will denote the initial probability distribution of the process $(Z_t)_{t\in\mathbb{N}}$.
- The observations $(Y_t)_{t\in\mathbb{N}}$, conditionally on $(Z_t)_{t\in\mathbb{N}}$ and on the σ-field $\mathcal{F}_t$, generated by $\{Y_s, s \leq t\}$, are distributed according to a Poisson law.

$$\mathrm{P}[Y_t = y | Z_t = k, \mathcal{F}_{t-1}] = \frac{\exp(-\lambda_{k,t})\lambda_{k,t}^{y}}{y!}, \quad y = 0, 1, 2, \ldots, \tag{2}$$

A Simple Mixture Model

- When (Z_t) is an iid sequence of random variables with $\mathrm{P}[Z_t = k] = p_k$ for $k = 1, \ldots, K$, the marginal probability mass function (pmf) of (Y_t), given the past, is given by

$$\mathrm{P}[Y_t = y \mid \mathcal{F}_{t-1}] = \sum_{k=1}^{K} p_k \frac{\exp(-\lambda_{k,t})\lambda_{k,t}^{y}}{y!}, \quad y = 0, 1, 2, \ldots,$$

  We say that the conditional distribution of Y_t, given the past, is a K-class mixture of Poisson distributions.

- Elementary calculations show that

$$\mathrm{E}[Y_t \mid \mathcal{F}_{t-1}] = \textstyle\sum_{k=1}^{K} p_k \lambda_{k,t},$$
$$\mathrm{Var}[Y_t \mid \mathcal{F}_{t-1}] = \textstyle\sum_{k=1}^{K} p_k \lambda_{k,t}(1 + \lambda_{k,t}) - \left(\sum_{k=1}^{K} p_k \lambda_{k,t}\right)^2.$$

  Since $\sum_{k=1}^{K} p_k \lambda_{k,t}^2 - \left(\sum_{k=1}^{K} p_k \lambda_{k,t}\right)^2 = \sum_{\substack{k=1 \\ k \neq l}}^{K} \sum_{l=1}^{K} p_k p_l (\lambda_{k,t} - \lambda_{l,t})^2 > 0$
  we have shown that simple a Poisson mixture models takes into account overdispersion, which usually observed in count time series analysis.

Theorem

Consider the MINARCH(K, ∞) model. Assume that for any $\boldsymbol{x}$ and $\boldsymbol{x}'$ in $\mathbb{N}^\infty \times \mathbb{N}^\infty$ (with $\mathbb{N} = \{0, 1, 2, \ldots\}$), there exist sequences $(\alpha_{kl})_{l\geq 1}$, $k = 1, \ldots, K$ of non-negative real numbers such that

$$|f_k(\boldsymbol{x}) - f_k(\boldsymbol{x}')| \leq \sum_{l=1}^{\infty} \alpha_{kl}|x_l - x_l'| \quad k = 1, 2, \ldots, K.$$

Denote by $A_k = \sum_l \alpha_{kl}$ and let $B_s = \sum_{k=1}^K p_k A_k^s < 1$. Then (i) If $s = 1$ there exists a $\tau-$weakly dependent strictly stationary process $\{Y_t, t \in \mathbb{Z}\}$ which belongs to $\mathcal{L}^1$ and (ii) for $s > 1$ this solution belongs to $\mathcal{L}^s$. Moreover

$$\tau(r) \leq \frac{2}{1 - B_1} \max_{1\leq k\leq K} f_k(\boldsymbol{0}) \inf_{1\leq p\leq r} \left\{ B_1^{\frac{r}{p}} + \frac{1}{1 - B_1} \sum_{q=p+1}^{\infty} \sum_{k=1}^{K} p_k \alpha_{kq} \right\},$$

where $(\tau(r))_{r\in\mathbb{N}}$ denotes the sequence of τ-dependence coefficients.

Mixing for MS-INARCH(K, L) models

We can show that the MS-INARCH(K, ∞) is geometrically β-mixing under stricter conditions.

Proposition

For $L < \infty$, consider the MS-INARCH(K, L) model. Assume that for any x and x' in $\mathbb{N}^L \times \mathbb{N}^L$ (with $\mathbb{N} = \{0, 1, 2, \dots\}$), there exist sequences $(\alpha_{kl})_{l \geq 1}$, $k = 1, \dots, K$ of non-negative real numbers such that

$$|f_k(x) - f_k(x')| \leq \sum_{l=1}^{L} \alpha_{kl} |x_l - x'_l| \quad k = 1, 2, \dots, K.$$

Suppose further that $\alpha = \sum_{l=1}^{L} \max_{1 \leq k \leq K} \alpha_{kl} < 1$. Then the Markov chain $W_t = (Y_t, \dots, Y_{t-L+1}, Z_t) \in \mathbb{N}^L \times \mathcal{K}$ is geometrically $\beta-$mixing.

Likelihood for MS-INARCH(K, L) models

- We investigate the behavior of maximum likelihood estimator (MLE) for MS-INARCH(K,L) models.
- Assume that the observations $\{Y_t, t = 1, \dots, n\}$ is a realization of a strictly stationary process $\{Y_t, t \in \mathbb{Z}\}$.
- Denote the vector of unknown parameters by $\psi = (\mathrm{vec}(Q), \psi_1^T, \dots, \psi_K^T)^T$ where Q is the unknown transition probability matrix $Q = (q_{i,j})_{1 \le i,j \le K}$.
- We introduce the function

$$g_k(y, y_1, \dots, y_L; \psi_k) = \frac{\exp(-f(y_1, \dots, y_L; \psi_k)) f^y(y_1, \dots, y_L; \psi_k)}{y!},$$

Suppose that we have observations $y_{-L+1}, \cdots, y_n$ from the $(Y_t)_{t \in \mathbb{N}}$ sequence. Then, the conditional likelihood, given $y_{-L+1}, \cdots, y_0$, is equal to

$$L_n(y_1, \cdots, y_n | y_{-L+1}, \cdots, y_0 ; \psi) =$$
$$\sum_{z_1=1}^K L_n(y_1, \cdots, y_n | y_{-L+1}, \cdots, y_0, Z_1 = z_1 ; \psi) P_\psi(Z_1 = z_1 | y_{-L+1}, \cdots, y_0)$$
$$= \sum_{z_1=1}^K \cdots \sum_{z_n=1}^K \prod_{t=2}^n q_{z_{t-1}, z_t}$$
$$\prod_{t=1}^n g_k(y_t, y_{t-1}, \dots, y_{t-L}; \psi_k) P_\psi(Z_1 = z_1 | y_{-L+1}, \cdots, y_0).$$

Modified conditional likelihood

- Multiplying the preceding equation by the density $L(y_{-L+1},\cdots,y_0;\psi)$ of $(Y_{-L+1},\cdots,Y_0)$ yields the likelihood. In general, it is not feasible to compute the stationary distributions of $(Z_1, Y_{-L+1},\cdots,Y_0)$; thus we will work with the modified conditional likelihood :

$$\tilde{L}_n(y_1,\cdots,y_n\,|y_{-L+1},\cdots,y_0\,;\psi) = \sum_{z_1=1}^{K}\cdots\sum_{z_n=1}^{K}\pi_{z_1}\prod_{t=2}^{n}q_{z_{t-1},z_t}\prod_{t=1}^{n}g_k(y_t,y_{t-1},\ldots,y_{t-L};\psi_k).$$

- The vector π_{z_1} of initial state probabilities may be chosen as an arbitrary stochastic vector with positive entries.
- With a slight abuse of language, we call Maximum Likelihood Estimator (MLE) a maximum point of the modified likelihood, say $\widehat{\psi}_n$ defined by

$$\widehat{\psi}_n = \arg\max_{\psi\in\Psi}\tilde{L}_n(y_1,\cdots,y_n\,|y_{-L+1},\cdots,y_0\,;\psi).$$

It turns out that the modified likelihood can be calculated in linear time by employing a recursive state filter.

Assumptions for the consistency and asymptotic normality of the conditional MLE

H-1 : The Markov chain $(Y_t, Z_t)_{t\in\mathbb{N}}$ admits a stationary solution.

H-2 : The parameter vector $\boldsymbol{\psi}$ belongs to a compact set $\boldsymbol{\Psi} \subset \mathbb{R}^d$, and the true parameter $\boldsymbol{\psi}^0$ belongs to the interior of $\boldsymbol{\Psi}$.

H-3 The regression functions $f_k(\cdot)$ are continuous and they satisfy that $f_k(\cdot) \geq C_1$ for some constant $C_1 > 0$ and for all $k = 1, \dots, K$.

H-4 : Assume that $P_{\boldsymbol{\psi}} = P_{\boldsymbol{\psi}^0}$ if and only if $\boldsymbol{\psi} = \boldsymbol{\psi}^0$.

H-5 : For all $(y_1, \dots, y_L) \in \mathbb{N}^L$, the functions $\boldsymbol{\psi}_k \mapsto f_k(\cdot)$ are twice continuously differentiable on $\boldsymbol{\Psi}_\eta$.

H-6 : $\mathrm{E}\left(\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}_\eta} \sup_{k\in\mathcal{K}} \|\nabla_{\boldsymbol{\psi}} f_k(\cdot)\|^2\right) < \infty$ and $\mathrm{E}\left(\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}_\eta} \sup_{k\in\mathcal{K}} \|\nabla^2_{\boldsymbol{\psi}} f_k(\cdot)\|\right) < \infty$.

Consistency

Lemma

Consider the MS-INARCH(K, L) model. Then under assumptions **(H-1)–(H-3)** it holds that, as $n \to \infty$,

$$\frac{1}{n}\tilde{L}_n(y_1, \dots, y_n \,|\, y_{-L+1}, \cdots, y_0 \,; \boldsymbol{\psi}) \xrightarrow{a.s.} H(\boldsymbol{\psi}),$$

where $H(.)$ is the entropy function. We define the Kullback-Leibler divergence by $KL(\boldsymbol{\psi}) = H(\boldsymbol{\psi}^0) - H(\boldsymbol{\psi})$.

Theorem

Consider the MS-INARCH(K, L) model. Then under assumptions **(H-1)–(H-3)** it holds that, as $n \to \infty$,

$$\rho(\widehat{\boldsymbol{\psi}}_n, \boldsymbol{\Psi}^0) \xrightarrow{a.s.} 0.$$

Inference for known number of regimes

- If the regressions functions $f_k(.)$ are identifiable, and the number of regimes is known then we can easily show that the model MS-INARCH(K, L) is identifiable and the MLE is consistent.

Theorem

Consider the MS-INARCH(K, L) model and assume that **(H-1)–(H-4)** hold true. Then, as $n \to \infty$, $\lim_{n\to\infty} \widehat{\psi}_n = \psi^0$, a.s.

- Asymptotic normality of the MLE is implied by a central limit theorem for the score function $\nabla_\psi \log \tilde{L}_n / \sqrt{n}$, and a locally uniform law of large numbers for the observed Fisher information $-\nabla^2_\psi \log \tilde{L}_n / n$ in a neighborhood of ψ^0.

Theorem

Suppose that **(H-1)–(H-6)** hold true and $I(\psi^0)$ is positive definite. Then, as $n \to \infty$

1. $$\frac{1}{\sqrt{n}} \nabla_\psi \tilde{L}_n(y_1, \ldots, y_n | y_{-L+1}, \ldots, y_0 ; \psi^0) \longrightarrow \mathcal{N}\left(0, I(\psi^0)\right),$$

2. $$\frac{1}{n} \nabla^2_\psi \tilde{L}_n(y_1, \ldots, y_n | y_{-L+1}, \ldots, y_0 ; \psi^0) \longrightarrow I(\psi^0),$$

3. $$\sqrt{n}\left(\widehat{\psi}_n - \psi^0\right) \longrightarrow \mathcal{N}\left(0, I^{-1}(\psi^0)\right).$$

Unknown number of regimes

- If the number of hidden regimes is unknown, then the MS-INARCH model is not identifiable and therefore the likelihood ratio test statistic (LRT) has a non-standard behavior.
- The following simple example illustrates this issue. Assume that the true data generating process has only $K^0 = 1$ regime and the autoregressive function is linear with $L = 1$, i.e.

$$P(Y_t = y \,|\, Y_{t-1}) = \exp(-(\psi_0^0 + \psi_1^0 Y_{t-1}) \frac{(\psi_0^0 + \psi_1^0 Y_{t-1})^y}{y!}.$$

We fit a linear model with $K = 2$ regimes and transition matrix with elements $q_{i,1} = p_1$ and $q_{i,2} = 1 - p_1$, for $i = 1, 2$,

$$P(Y_t = y \,|\, Y_{t-1}) = p_1 \exp(-(\psi_{1,0} + \psi_{1,1} Y_{t-1})) \frac{(\psi_{1,0}+\psi_{1,1} Y_{t-1})^y}{y!} + (1 - p_1) \exp(-(\psi_{2,0} + \psi_{2,1} Y_{t-1})) \frac{(\psi_{2,0}+\psi_{2,1} Y_{t-1})^y}{y!}.$$

Any parameter vector $\boldsymbol{\psi} = (p_1, \psi_{1,0}, \psi_{1,1}, \psi_{2,0}, \psi_{2,1})$ with
$\psi_{1,0} = \psi_{2,0} = \psi_0^0, \psi_{1,1} = \psi_{2,1} = \psi_1^0,\ p_1 \in [0; 1]$ or
$\psi_{2,0} = \psi_0^0, \psi_{2,1} = \psi_1^0, (\psi_{1,0}, \psi_{1,1}) \in \mathbb{R}^2,\ p_1 = 0$ or
$\psi_{1,0} = \psi_0^0, \psi_{1,1} = \psi_1^0, (\psi_{2,0}, \psi_{2,1}) \in \mathbb{R}^2,\ p_1 = 1$ satisfies the true conditional pmf. This simple observation shows that for any parametric model, whose number of regimes is overestimated, has a non-invertible Fisher information matrix and it explains why the LRT has a non-standard asymptotic behavior.

The marginal-conditional density function

- we study the asymptotic distribution of the marginal LRT (MLRT), when the number of mixture regimes is overestimated.
- Denote by $g(y, y_1, \cdots, y_L; \psi)$ the marginal-conditional density function of y knowing $y_1, \cdots, y_L$, i.e.

$$g(y, y_1, \cdots, y_L; \psi) = \sum_{k=1}^{K} p_k(y_1, \ldots, y_L; \psi) g_k(y, y_1, \ldots, y_L; \psi_k), \qquad (3)$$

  where $p_k(y_1, \ldots, y_L; \psi)$ is the stationary law of conditional expectation of the prediction filter $p_{k,t}$ knowing the immediate past $Y_{t-1} = y_1, \cdots, Y_{t-L} = y_L$.
- Moreover, we will denote by g^0 the true marginal-conditional density :

$$g^0(y, y_1, \cdots, y_L) = \sum_{k=1}^{K^0} p_k(y_1, \cdots, y_l; \psi^0) g_k(y, y_1, \ldots, y_L; \psi_k^0),$$

  for any parameter vector $\psi^0 \in \Psi^0$, realising the true density function of the observations.

Some notations

- for any parameter vector $\boldsymbol{\psi}^0 \in \boldsymbol{\Psi}^0$, realising the true density function of the observations.
- For $\eta > 0$, denote by $\mathcal{G}_\eta = \left\{ g \in \mathcal{G}, \|g - g^0\|_{L^2(\mu)} \leq \eta \right\}$ the set of conditional densities in a η-neighborhood of the true density where $\mathcal{G} = \{g(.;\boldsymbol{\psi}), \boldsymbol{\psi} \in \boldsymbol{\Psi}\}$.
- In a η-neighborhood of the true density, the extended set of score-functions $\mathcal{S}_\eta$ is defined by

$$\mathcal{S}_\eta = \left\{ s_g = \frac{\frac{g}{g^0} - 1}{\left\| \frac{g}{g^0} - 1 \right\|_{L^2(\mu)}}, g \in \mathcal{G}_\eta \right\}.$$

- We define the limit-set of score functions $\mathcal{D}$

$$\mathcal{D} = \left\{ d \in \mathbb{L}^2(\mu) \mid \exists\, (g_n) \in \mathcal{G},\ \left\| \frac{g_n - g^0}{g^0} \right\|_{\mathbb{L}^2(\mu)} \xrightarrow[n\to\infty]{} 0, \right.$$
$$\left. \|d - s_{g_n}\|_{\mathbb{L}^2(\mu)} \xrightarrow[n\to\infty]{} 0 \right\}.$$

- Let

$$\ln(\boldsymbol{\psi}) = \sum_{k=1}^{n} \ln g(y_k, y_{k-1}, \cdots, y_{k-L})$$

be the marginal likelihood of $y_1, \cdots, y_n$ knowing, $y_0, \cdots, y_{1-L}$.

Assumptions

We assume the following :

T-1 The process $(Y_t)_{t\in\mathbb{N}}$ is geometrically β-mixing.

T-2 The parameter space Ψ is a compact set of $\mathbb{R}^d$, and the parameter space Ψ contains a neighborhood of the parameters defining the true conditional density g^0. Moreover all the entries of the transition matrix Q are supposed to be strictly positives. Since Ψ is compact this means that all the entries of the transition matrix Q are bounded by below.

T-3 There exists $\eta > 0$ such that for all $g \in \mathcal{G}$ with $\|g - g^0\|_{L^2(\mu)} \leq \eta$, $\left\|\frac{g}{g^0} - 1\right\|_{L^2(\mu)} < \infty$.

T-4 Assumptions **(H-5)** and **(H-6)** hold true. Note that **(H-4)** does not generally hold if the number of regimes is not known.

T-5 With the following notations :

$$l'_k := \frac{\partial l_{\psi_k}}{\partial \psi_k}\left(\psi_k^0\right),\ l''_k := \frac{\partial^2 l_{\psi_k}}{\partial {\psi_k}^2}\left(\psi_k^0\right)$$

we assume that for distinct $(\psi_k)_{1\leq k\leq K}$

$$\left\{\left(l_{\psi_k}\right)_{1\leq k\leq K^0}, \left(l'_k\right)_{1\leq k\leq K^0}, \left(l''_k\right)_{1\leq k\leq K^0}\right\}$$

are linearly independent in the Hilbert space $L^2(\mu)$.

Asymptotic behavior

Theorem

Consider the MS-INARCH(K, L) model and assume that **(T-1)–(T-5)** hold true. Then, the MLRT is represented by

$$\lambda_n = \sup_{d \in \mathcal{D}} \left(\max \left\{ \frac{1}{\sqrt{n}} \sum_{i=2}^{n} d(Y_1, \cdots, Y_L); 0 \right\} \right)^2 + o_P(1),$$

recalling that $\mathcal{D}$ denotes the limit-set of score functions. Hence, there exists a centered Gaussian process $\{W_S, S \in \mathbb{F}\}$ with continuous sample path and covariance kernel $P\left(W_{S_1} W_{S_2}\right) = P\left(S_1 S_2\right)$ such that

$$\lim_{n \to \infty} \lambda_n = \sup_{S \in \mathbb{F}} \left(\max(W_S, 0)\right)^2,$$

where the index set $\mathbb{F}$ of limit derivatives depends on the compact set $\boldsymbol{\Psi}$ of possible parameters.

- We can verify of assumptions **(T-1)–(T-5)** for the special case of linear MS-INARCH(K, 1) model.
- This theorem implies that asymptotic distribution is bounded in probability but it depends on the true model parameters. Therefore, the theorem cannot be used to applications for obtaining critical values.

estimating the true number of regimes (1)

- For $l \in \mathbb{N}_+$, denote

$$\mathcal{G}_l = \{\text{MS-INARCH}(l, L) \text{ with } l \text{ regimes and parameters set } \boldsymbol{\Psi}_l\}.$$

- For some fixed $M \in \mathbb{N}_+$ sufficiently large, we shall consider the following class of functions

$$\mathcal{G}_M = \bigcup_{l=1}^{M} \mathcal{G}_l.$$

- For every $g \in \mathcal{G}_M$ define the number of regimes as

$$l(g) = \min\{l \in \{1, \ldots, M\}, \ g \in \mathcal{G}_l\}.$$

Then, $l_0 = l\left(g^0\right)$ denotes the number of regimes of the true model. An estimate of the number of regimes, say $\hat{l}$ is defined as the integer $l \in \{1, \ldots, M\}$ which maximizes the following penalized criterion,

$$T_n(l) = \sup_{\boldsymbol{\psi} \in \boldsymbol{\Psi}_l} l_n(\psi) - \alpha_n(l),$$

where $\alpha_n(\cdot)$ is a suitably chosen penalization sequence.

estimating the true number of regimes (2)

Proposition

Consider the MS-INARCH(K, L) model and assume that **(T-1)–(T-5)** hold true. Let $(a_n(\cdot))$ be an increasing function of l such that $a_n(l_1) - a_n(l_2) \xrightarrow[n\to\infty]{} \infty$ for every $l_1 > l_2$ and $a_n(l)/n \xrightarrow[n\to\infty]{} 0$ for every l. Then, as $n \to \infty$, the estimator $\hat{l}$ converges in probability to the true number of regimes, i.e.

$$\hat{l} \xrightarrow[n\to\infty]{\mathbb{P}} K^0.$$

The BIC ($a_n(l) = \ln(n)l$) fulfills the assumptions of Proposition.

Simulation Study

- We simulate data from a MINARCH(2,1) model where the hidden process Z_t is distributed according to $\mathrm{P}[Z_t = 1] = 1 - \mathrm{P}[Z_t = 0] = 2/3$ and for each regime the Poisson intensities are given by $\lambda_{1,t} = 1.5 + 0.9Y_{t-1}$ and $\lambda_{2,t} = 2 + 0.5Y_{t-1}$. In this case, the marginal LRT coincides with the LRT.
- The goal of this exercise is to compare the AIC and BIC in terms of determining the true number of regimes by counting the percentage of times that the aforementioned model selection criteria choose correct that $K = 2$ after fitting models with $K = 1, 2, 3, 4$ regimes. Results are based on various sample sizes and 1000 runs.

TABLE – Percentage of times that AIC and BIC attain their minimum value among all models fitted to a MINACH(2,1) model generated using various sample sizes. Results are based on 1000 simulations.

Sample Size	$K=1$		$K=2$		$K=3$		$K=4$	
	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
250	61.7	35.5	0	**64.1**	22.6	0	15.7	0.4
500	63.2	19.7	0	**80.3**	11.3	0	25.5	0
1000	56.2	3.1	0	**96.9**	2.5	0	41.3	0
1500	52.4	0.5	0	**99.5**	0.5	0	47.1	0

A Data Example : disease cases caused by E.coli

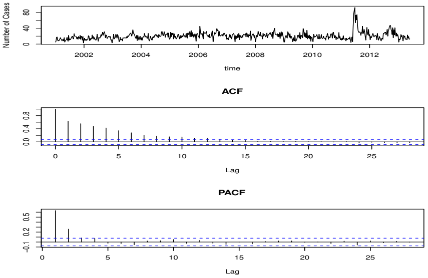

FIGURE – Weekly number of reported disease cases caused by E.coli in the state of North Rhine-Westphalia (Germany) from January 2001 to May 2013 and their sample autocorrelation and partial autocorrelation functions.

The data

- The time series of E.coli cases consist of 646 observations.
- We fit several mixture models by assuming that the hidden process (Z_t) is :
  1. An iid sequence.
  2. A Markov chain
  3. Its transition probabilities depends on past values of Y_t's.
- we employ the transformation $p_k = \exp(\theta_k)/(1 + \sum_{j=1}^{K} \exp(\theta_j))$, $k = 1, \ldots, K$ and $\theta_K = 0$ to avoid numerical instability.

Lag	Simple Poisson	Model 1	Model 2	Model 3
L	$K = 1$	$K = 2$		
1	4636.327	4364.537	4361.451	4363.568
2	4540.943	**4319.091**	4380.950	4351.380
3	4522.201	4328.674	4357.856	4351.560

TABLE – BIC values after fitting models to E.coli count time series.

Estimated models

- The BIC favors a mixture of a two-component simple mixture model where in each region Y_t is regressed on Y_{t-1}, Y_{t-2}.
- The parameter estimates and their standard errors for $\boldsymbol{\psi} = (\theta_1, \psi_{1,0}, \psi_{1,1}, \psi_{1,2}, \psi_{2,0}, \psi_{2,1}, \psi_{2,2})^T$ are $(-0.597, 9.475, 0.573, 0.262, 5.431, 0.344, 0.226)^T$ and $(0.068, 1.123, 0.003, 0.003, 0.641, 0.001, 0.002)^T$.
- These results are interpreted as follows. The probability of one regime is $p_1 = 0.36$ and the probability of the other regime is $p_2 = 1 - p_1 = 0.64$.
- In the region which has the smaller probability, the mean number of weakly cases is larger than the mean number of weakly cases in the region with higher probability.

Conclusion

- We introduced Markov switching non linear autoregressive Poisson models and study their properties.
- When the number of regimes is unknown, we showed that a marginal likelihood ratio test for testing the number of hidden regimes converges to a Gaussian process. Note that the computation of such function is much more easier when the observations are discrete. The generalization to real observations is not straightforward.
- The MS-INARCH(K, L) model can be extended by introducing a feedback process in the right hand side : MS-INGARCH(K, $L1$, $L2$) model.
- In this case, the log-likelihood function requires evaluation of all the possible states for $Z_1, \cdots, Z_n$ and therefore the complexity of such computation is of the order K^n.
- An alternative way to deal with this issue is to employ the infinite representation of the MS-INGARCH model in terms of an MS-INARCH model and truncate it at some finite order.
- However, We can provide a counterexample which shows that estimation of the true number of regimes, under a misspecified GARCH type count time series model, is not possible.

Bibliography

- *Inference in Hidden Markov Models.*,
  O. Cappé, E. Moulines, T. Rydén, Springer, 2005.
- *Asymptotic properties of the maximum likelihood estimator in autoregressive models with Markov regime.*,
  R. Douc, E. Moulines, T. Rydén, Annals of statistics, 32, 2254-2304, 2004.
- *nference and testing for structural change in general Poisson autoregressive models.*,
  P. Doukhan, W. Kengne, Electronic Journal of Statistics, 9, 1267–1314, 2015.
- *Mixtures of Nonlinear Poisson Autoregressions*,
  P. Doukhan, K. Fokianos, J. Rynkiewicz, Journal of time series analysis, (to appear).
- *Likelihood ratio inequalities with applications to various mixture*,
  E. Gassiat, Annals of Institute Henri Poincaré 38, 897-906, 2002.
- *Maximum–likelihood estimation for hidden Markov models.*,
  B.G. Leroux, Stochastic Processes and its Applications, 40, 127-143.
- *Asymptotic properties of autoregressive regime-switching models.*,
  M. Olteanu, J. Rynkiewicz, ESAIM P.S., 16, 25-47, 2012.
- *mixture integer-valued ARCH model.*,
  F.Q., Zhu, D. Wang, Journal of Statistical Planning and Inference, 140, 2025-2036, 2010.