Count and duration time series with conditional stochastic order equal to the conditional mean order

Abdelhakim Aknouche Christian Francq

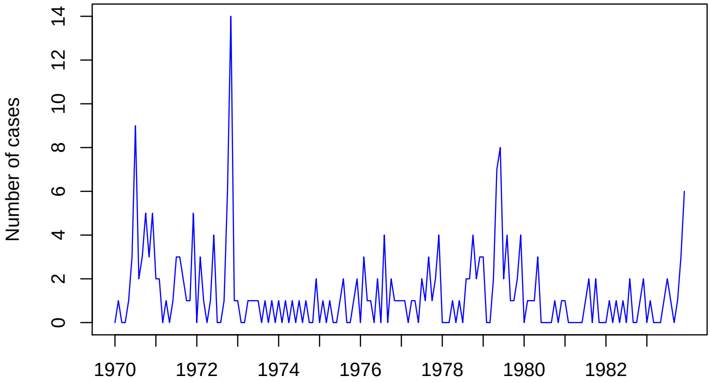
USTHB, CREST and University of Lille

Econometrics, Finance and Statistics Workshop, January 3 2019, Las Majadas de Pirque

Supported by the ANR via the Project MultiRisk (ANR-16-CE26-0015-02)

Count time serie example

Monthly number of poliomyelitis cases in the United States from 1970 to 1983



Standard models of count times series

Standard Poisson INGARCH models assume $Y_t \mid \mathscr{F}_{t-1} \sim \mathscr{P}(\lambda_t)$ with

$$\lambda_t = \omega_0 + \sum_{i=1}^{q} \alpha_{0i} Y_{t-i} + \sum_{j=1}^{p} \beta_{0j} \lambda_{t-j}$$

and $\mathscr{F}_{t-1} = \sigma(Y_u, u < t)$. One can also consider other conditional distributions, in particular the Negative Binomial INGARCH model.

The (first order) INAR model assumes

$$Y_t = \mathscr{B}(Y_{t-1}, \alpha) + \text{integer-valued distribution.}$$

Models for positive times series

When (Y_t) is valued in $[0,\infty)$, an ARMA-type model

$$Y_t = \lambda_t + \epsilon_t,$$

where $\lambda_t = E(Y_t \mid Y_u, u < t)$ and (ϵ_t) is a white noise, is not convenient (it is difficult to impose $Y_t \geq 0$).
Engle and Russell (1998) proposed the Autoregressive Conditional Duration (ACD) model

$$Y_t = \lambda_t z_t,$$

where (z_t) iid positive with $Ez_t = 1$.

Limitations of the multiplicative ACD form

The Multiplicative Error Model (MEM) form

$$Y_t = \lambda_t z_t,$$

where $\lambda_t \in \mathscr{F}_{t-1} = \sigma(Y_u, u < t)$, with

$$z_t \text{ and } \lambda_t \text{ independent,}$$

is generally impossible when Y_t is valued in $\mathbb{N}$. Even for durations (or volumes or any positive time series), the MEM structure is restritive. For instance, it implies that

$$\text{Var}(Y_t \mid \mathscr{F}_{t-1}) \propto \lambda_t^2.$$

Framework of the present paper

Let positive exogenous variables $\boldsymbol{X}_t = (x_{1,t},\dots,x_{r,t})$, the information set $\mathscr{F}_{t-1} = \sigma(Y_u, \boldsymbol{X}_u, u < t)$.
We relax the multiplicative structure (necessary for count time series), and assume that the condition distribution depends on a parametric time-varying conditional mean

$$\lambda_t(\theta_0) := E(Y_t \mid \mathscr{F}_{t-1}) = \lambda(Y_u, \boldsymbol{X}_u, u < t; \theta_0), \qquad t \in \mathbb{Z}.$$

For instance

$$\lambda_t(\theta) = \omega + \sum_{i=1}^{q} \alpha_i Y_{t-i} + \sum_{j=1}^{p} \beta_j \lambda_{t-j} + \sum_{i=1}^{r} \pi_i x_{i,t-1},$$

with

$$\theta = (\omega, \alpha_1, \dots, \beta_q, \pi_1, \dots, \pi_r) \in [0,\infty)^m, \quad m = p + q + r + 1.$$

First objective

We want conditions for stationarity and ergodicity.

The main difficulty is that, contrary to standard time series models,

- there exists no explicit solution $Y_t = f(\theta_0, z_t, z_{t-1}, \dots)$;
- the theory of the Markov chains with continuous state space does not apply.

Few references

- Ferland, Latour and Oraichi (2006) for Poisson-INGARCH;
- Neumann (2011) for absolute regularity of nonlinear Poisson autoregressions, and Doukhan and Neumann (2017) for a broader class;
- Franke (2010) and Doukhan, Fokianos and Tjostheim (2012, 2013) for weak dependence of nonlinear Poisson;
- Douc, Doukhan and Moulines (2013), Douc, Roueff and Sim (2015, 2016) and Sim, Douc and Roueff (2016) for observation-driven Markov processes;
- Gonçalves, Mendes-Lopes and Silva (2015) for stationarity and ergodicity of compound Poissont INGARCH;
- Davis and Liu (2016) for stationarity and mixing when the conditional distribution belongs to the one-parameter exponential family.

Methodology

- Davis and Liu (2016) builds explicit solutions as limits of functions of quantiles of an iid sequence;
- We adopt the same strategy, but
  - the conditional distribution is not restricted to the one-parameter exponential family;
  - the dynamics of the conditional mean is more general;
  - exogenous variables are allowed.

Central assumption
Stochastic-equal-mean order property

Let F_λ be a family of cdf indexed by the mean $\lambda = \int y dF_\lambda(y) \in \mathbb{R}$. Assume that the stochastic order is equal to the mean order:

$$\lambda \leq \lambda^* \quad \Rightarrow \quad F_\lambda(y) \geq F_{\lambda^*}(y), \quad \forall y \in \mathbb{R}.$$

Equivalently,

$$\lambda \leq \lambda^* \quad \Rightarrow \quad F_\lambda^-(u) \leq F_{\lambda^*}^-(u), \quad \forall u \in (0, 1).$$

where F_λ^- is the quantile function of cdf F_λ.

Examples of cdf with stochastic-equal-mean order

- any distribution belonging to the one-parameter linear exponential family

$$g_\lambda(y) = h(y)\exp\{\eta y - A(\eta)\} 1_{\{y\geq 0\}},$$

for some scalar natural parameter $\eta = \eta(\lambda)$;

- Negative Binomial NB(r,p)

$$P(Y=k) = \frac{\Gamma(k+r)}{k!\Gamma(r)} p^r (1-p)^k, \quad k \in \mathbb{N},$$

when $r = p\lambda/(1-p)$) and p is fixed;

Examples of stochastic-equal-mean order (continued)

- Gamma distributions: for fixed a
  - $\Gamma(a, a/\lambda)$ belongs to the exponential family;
  - $\Gamma(a\lambda, a)$ also satisfies the property, but does not belong to the exponential family.

  Remark: an ACD cannot have the distribution $Y_t \mid \mathscr{F}_{t-1} \sim \Gamma(a\lambda_t, a)$, because $\mathrm{Var}(Y_t \mid \mathscr{F}_{t-1}) = \lambda_t/a$.
- any zero-inflated version of a cdf satisfying the stochastic-equal-mean order property:

$$P(Y \leq y) = \tau + (1-\tau)F_\lambda(y), \qquad y \geq 0.$$

Stationarity and ergodicity in the INGARCH-X case

Stationarity condition

There exists a stationary (and ergodic) sequence (Y_t) such that $EY_t < \infty$ and the conditional cdf satisfies stochastic-equal-mean order property with mean

$$\lambda_t = \omega + \sum_{i=1}^{q} \alpha_i Y_{t-i} + \sum_{j=1}^{p} \beta_j \lambda_{t-j} + \sum_{i=1}^{r} \pi_i x_{i,t-1},$$

where $(\boldsymbol{X}_t)$ stationary and ergodic with $E\|\boldsymbol{X}_t\| < \infty$, if and only if

$$\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1.$$

idea of the proof

Note that the condition does not depend on $\boldsymbol{\pi}$.

Moments in the INGARCH(1,1) case

Moment conditions

Let the previous assumptions with $p = q = 1$ and $r = 0$. Assume that, for $Y \sim F_\lambda(y)$ and some integer $\ell \geq 2$, there exist nonnegative coefficients $a_j(0), a_j(1), \ldots, a_j(j)$ for all $j \leq \ell$ such that

$$EY^j = \sum_{i=0}^{j} a_j(i)\lambda^i \text{ for } j = 1, \ldots, \ell.$$

We have $EY_t^\ell < \infty$ if and only if

$$\sum_{j=0}^{\ell} a(j) \binom{\ell}{j} \alpha^j \beta^{\ell-j} < 1,$$

where $a(0) = a(1) = 1$ and $a(j) = a_j(j)$ for $j \geq 2$.

Examples

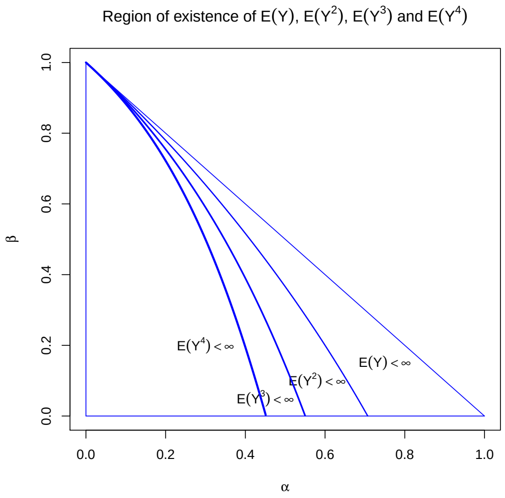
- If $Y_t \mid \mathscr{F}_{t-1} \sim NB(p\lambda_t/(1-p), p)$ then Y_t admits moments of any order iff $\alpha + \beta < 1$.
- If $Y_t \mid \mathscr{F}_{t-1} \sim NB(r, r/(\lambda_t + r))$ then Y_t admits a moment of

$$\text{order } 2 \quad \text{iff} \quad (\alpha+\beta)^2 + \frac{\alpha^2}{r} < 1,$$

$$\text{order } 3 \quad \text{iff} \quad (\alpha+\beta)^3 + \frac{3\alpha^2(\alpha+\beta)}{r} + \frac{2\alpha^3}{r^2} < 1,$$

$$\text{order } 4 \quad \text{iff} \quad (\alpha+\beta)^4 + \frac{6\alpha^2(\alpha+\beta)^2}{r} + \frac{\alpha^3(11\alpha+8\beta)}{r^2} + \frac{6\alpha^4}{r^3} < 1.$$

Moment conditions for the INGARCH(1,1) process
with NB$(1,p_t)$ conditional distribution



Extension to nonlinear conditional means

The stationarity conditions are the same if

$$\lambda_t = g(Y_{t-1},\dots,Y_{t-q},\lambda_{t-1},\dots,\lambda_{t-p}) + \pi(\boldsymbol{X}_{t-1}),$$

with

$$\left|g(y_1,\dots,y_q,\lambda_1,\dots,\lambda_p) - g(y_1',\dots,y_q',\lambda_1',\dots,\lambda_p')\right|$$
$$\leq \sum_{i=1}^{q}\alpha_i|y_i - y_i'| + \sum_{j=1}^{p}\beta_j|\lambda_j - \lambda_j'|.$$

Absolute regularity coefficients

Let $\mathscr{B}$ be the Borel sigma-algebra of $\mathbb{R}^\infty$, and let the β-mixing coefficient

$$\beta(h) = E \sup_{A \in \mathscr{B}} |P\{(Y_h, Y_{h+1}, \dots) \in A \mid Y_0, Y_{-1}, \dots\} - P\{(Y_h, Y_{h+1}, \dots) \in A\}|.$$

Mixing

Under the previous assuptions (stochastic-equal-mean order $+\sum \alpha_i + \beta_i < 1$), and if $Y_t(\Omega) = \mathbb{N}$,

$$\beta(h) \leq K\rho^h, \qquad h \geq 0.$$

for some $K > 0$ and $\rho \in (0,1)$.

Motivation for testing the MEM specification

Assuming

$$Y_t = \lambda_t z_t,$$

with $\lambda_t = E(Y_t \mid \mathscr{F}_{t-1})$ independent of z_t is

- impossible when $Y_t(\Omega) = \mathbb{N}$ (the support of z_t depends on λ_t);
- restrictice when $Y_t(\Omega) = [0, \infty)$ (the shape of the conditional distribution is time-constant);
  - $z_t = Y_t / \lambda_t$ and λ_t are always uncorrelated (when 2nd order moments exist);
  - z_t and λ_t may be dependent (when the conditional density of Y_t given $\mathscr{F}_{t-1}$ is not of the form $f(\cdot/\lambda_t)/\lambda_t$).

$\Rightarrow$ a test for nonlinear dependence

Distance covariance

Székely et al. (2007), Rizzo and Székely (2016), Davis et al. (2018)

Based on observations $Y_1, \dots, Y_n$, the null

$$H_0: \quad z_t \text{ and } \lambda_t \text{ are independent,}$$

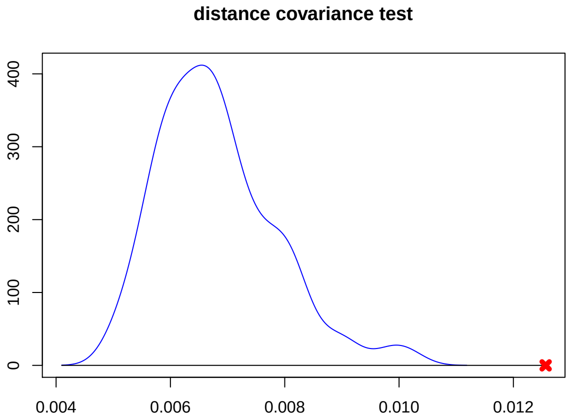
is rejected for large values of

$$\text{dCov}_n^2 = \int \left|\hat{\varphi}_{z,\lambda}(t,s) - \hat{\varphi}_z(t)\hat{\varphi}_\lambda(s)\right|^2 w(t,s)dtds,$$

where $\hat{\varphi}_{z,\lambda}$, $\hat{\varphi}_z$ and $\hat{\varphi}_\lambda$ are respectively empirical estimators of the characteristic functions of (z_t, λ_t), z_t and λ_t, and the weighting function $w(t,s)$ is, for instance, proportional to $t^{-2}s^{-2}$. The distribution under the null is approximated by a bootstrap procedure.

S&P 500 transaction volume (3/10/2013 to 3/10/2018)

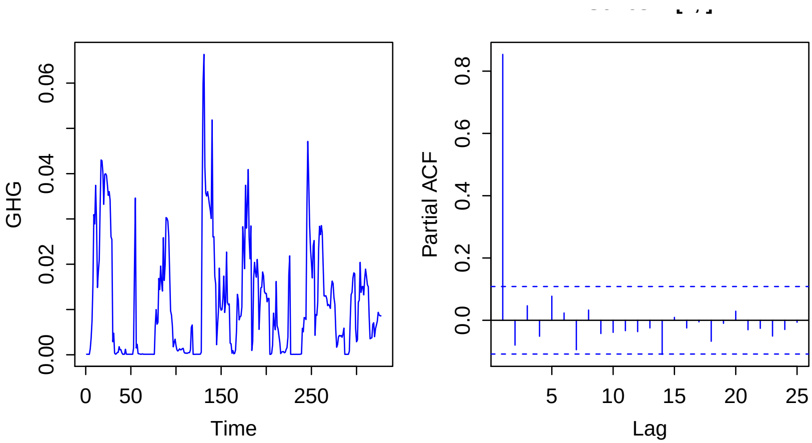
Testing the MEM structure of an ACD(2,2)



Bootstrap distribution and observed dCov (red cross)

Greenhouse gas concentrations

GHG every 6 hours from May 10 to July 31, 2010, and empirical PACF



Extended ACD model for the GHG series

The empirical PACF suggests an (extended) ACD(1,0) model. Several zero-inflated conditional Gamma distributions have been tried, leading to

$$\lambda_t = \omega + \alpha Y_{t-1}, \quad Y_t \mid \mathscr{F}_{t-1} \sim \tau\delta_0 + (1-\tau)\Gamma(\lambda_t b, b),$$

with maximum-likelihood estimates (MLE) $\hat{\omega} = 0.0024$, $\hat{\alpha} = 0.834$, $\hat{\tau} = 0.186$ and $\hat{b} = 245.2$.

Remark: The main interest is often on the conditional mean, but the MLE of the mean parameter may be sensitive to a misspecification of the conditional distribution.

Second objective

We want to estimate the mean parameter θ_0, but we want to be totally agnostic about the conditional distribution of the observations.
Indeed, there is no obvious choice for the conditional variance

$$v_t(\xi_0) := \text{Var}(Y_t \mid \mathscr{F}_{t-1}) = v(Y_u, \boldsymbol{X}_u, u < t; \xi_0).$$

In particular, for the Poisson conditional distribution we have $v_t = \lambda_t$ but count time series often exhibit (conditional) overdispersion.

We are thus interested in estimators that could be consistent even if the conditional variance is misspecified.

Existing misspecification-consistent estimators

Let $\widetilde{\lambda}_t(\theta) = \lambda(Y_{t-1}, \boldsymbol{X}_{t-1}, \dots, Y_1, \boldsymbol{X}_1, \widetilde{Y}_0, \widetilde{\boldsymbol{X}}_0, \dots; \theta)$ for given initial values $\widetilde{Y}_0, \widetilde{\boldsymbol{X}}_0, \dots$

Estimators based on the exponential family are generally consistent, in particular the Poisson Quasi-MLE (PQMLE)

$$\widehat{\theta}_{PQML} = \arg\max_{\theta\in\Theta} \sum_{t=1}^{n} \left\{ Y_t \log\left(\widetilde{\lambda}_t(\theta)\right) - \widetilde{\lambda}_t(\theta) \right\},$$

or the Negative Binomial QMLE (NBQMLE)

$$\widehat{\theta}_{NBQML} = \arg\max_{\theta\in\Theta} \sum_{t=1}^{n} Y_t \log\left(\frac{\widetilde{\lambda}_t(\theta)}{r_0 + \widetilde{\lambda}_t(\theta)} \right) - r_0 \log\left\{ r_0 + \widetilde{\lambda}_t(\theta) \right\},$$

studied by Ahmed and Francq (2016) and Aknouche, Bendjeddou and Touche (2018) (without exogenous variables).

More general QMLE and Estimating Functions QLE

A exponential family based QMLE satisfies

$$s_n(\widehat{\theta}) = 0, \qquad s_n(\theta) = \sum_{t=1}^{n} \frac{Y_t - \widetilde{\lambda}_t(\theta)}{\widetilde{\upsilon}_t(\theta)} \frac{\partial \widetilde{\lambda}_t(\theta)}{\partial \theta},$$

where $\upsilon_t(\theta)$ is the conditional variance of a given member of the exponential family (Wedderburn (1974) and Gouriéroux, Monfort and Trognon (1984)).
With the more general concept of optimal estimating functions of Godambe (1960, 1985), $\upsilon_t(\theta)$ may be a general conditional variance.

Motivations

PQMLE and NBQMLE are consistent for estimating θ_0 under very mild regularity conditions, but they may be inefficient when the conditional distribution is misspecified. Moreover, due to positivity constraints, their asymptotic distributions are not easily tractable when some coefficients are equal to zero.

The aim of this paper is to propose and study alternative estimators which enjoy the same consistency property as the QMLE's when the conditional distribution is misspecified, but have simpler asymptotic distributions when one or several coefficients are null and gain in efficiency when v_t is well specified.

Weighted LSE

Given a theoretical weight function $w_t = w(Y_{t-1}, \boldsymbol{X}_{t-1}, \dots) > 0$ and its observation-proxy

$$\widetilde{w}_t = w(Y_{t-1}, \boldsymbol{X}_{t-1} \dots, Y_1, \boldsymbol{X}_1, \widetilde{Y}_0, \widetilde{\boldsymbol{X}}_0 \dots) \geq \underline{w} > 0,$$

let the weighted least square estimator (WLSE)

$$\widehat{\theta}_{1WLS} = \arg\min_{\theta \in \Theta} \widetilde{L}_n(\theta, \widetilde{w}),$$

where

$$\widetilde{L}_n(\theta, \widetilde{w}) \quad = \quad \frac{1}{n} \sum_{t=1}^{n} \widetilde{l}_t(\theta, \widetilde{w}_t) \qquad \text{with} \qquad \widetilde{l}_t(\theta, w_t) = \frac{\left(Y_t - \widetilde{\lambda}_t(\theta)\right)^2}{w_t}.$$

The weighting sequence $\widetilde{w} = \{\widetilde{w}_t\}_{t \geq 1}$ allows the WLSE to be CAN without too strong moment conditions, and may reduce the asymptotic variance of the estimator.

Two-stage WLSE

It can be seen that the optimal choice of $\tilde{w}$ is (proportional to) υ. Assuming an hypothetical conditional variance of the form

$$\upsilon^*\left(Y_{t-1}, \boldsymbol{X}_{t-1}, \ldots; \xi_0^*\right) = \upsilon_t^*\left(\xi_0^*\right),$$

the optimal sequence of weights may be estimated by

$$\left\{\widehat{w}_{t,n}\right\}_t, \quad \widehat{w}_{t,n} = \upsilon^*\left(Y_{t-1}, \boldsymbol{X}_{t-1}, \ldots, Y_1, \boldsymbol{X}_1, \widetilde{Y}_0, \widetilde{\boldsymbol{X}}_0, \ldots; \widehat{\xi}_n\right),$$

where $\widehat{\xi}_n$ is a first-step estimator of ξ_0^* (which is often function of the estimator $\widehat{\theta}_{1WLS}$ of θ_0, and eventually of estimates of some extra parameter ς_0). This leads to a two-stage WLSE, defined by

$$\widehat{\theta}_{2WLS} = \arg\min_{\theta \in \Theta} \widetilde{L}_n\left(\theta, \left\{\widehat{w}_{t,n}\right\}_t\right).$$

Poisson-type 2WLSE

If a conditional variance approximately proportional to the conditional mean is expected, one can employ the two-stage estimator $\widehat{\theta}_{2WLS} = \widehat{\theta}^{(P)}_{2WLS}$ where

$$\widehat{\theta}^{(P)}_{2WLS} = \arg\min_{\theta \in \Theta} \sum_{t=1}^{n} \frac{\left(Y_t - \widetilde{\lambda}_t(\theta)\right)^2}{\widehat{w}_{t,n}}, \quad \widehat{w}_{t,n} = \widetilde{\lambda}_t\left(\widehat{\theta}_{1WLS}\right).$$

NB-type 2WLSE

If the conditional variance is expected to be approximately proportional to that of the NB$(r, r/(r+\lambda_t))$, one can consider the two-stage estimator $\widehat{\theta}_{2WLS} = \widehat{\theta}^{(NB)}_{2WLS}$ where

$$\widehat{\theta}^{(NB)}_{2WLS} = \arg\min_{\theta\in\Theta} \sum_{t=1}^{n} \frac{\left(Y_t - \widetilde{\lambda}_t(\theta)\right)^2}{\widehat{w}_{t,n}}, \quad \widehat{w}_{t,n} = \widehat{\lambda}_t \left(1 + \frac{\widehat{\lambda}_t}{\widehat{r}}\right),$$

with

$$\widehat{r} = \left(\frac{1}{n}\sum_{t=1}^{n} \frac{(Y_t - \widehat{\lambda}_t)^2 - \widehat{\lambda}_t}{\widehat{\lambda}_t^2}\right)^{-1}, \quad \widehat{\lambda}_t = \widetilde{\lambda}_t(\widehat{\theta}_{1WLS}).$$

Double-Poisson-type 2WLSE

If the conditional variance is expected to be inversely proportional the conditional mean, as for the Double-Poisson, one can also consider

$$\widehat{\theta}_{2WLS}^{(Inv)} = \arg\min_{\theta\in\Theta} \sum_{t=1}^{n} \frac{\left(Y_t - \widetilde{\lambda}_t(\theta)\right)^2}{\widehat{w}_{t,n}}, \quad \widehat{w}_{t,n} = 1/\widetilde{\lambda}_t\left(\widehat{\theta}_{1WLS}\right).$$

INARCH-type conditional mean

Assume the AR/INARCH-type conditional mean

$$\lambda_t(\theta) = \theta' \chi_t, \qquad \chi_t = \left(1, Y_{t-1}, ..., Y_{t-q}\right)'.$$

Example: The INAR model

$$Y_t = \alpha_{01} \circ Y_{t-1} + ... + \alpha_{0p} \circ Y_{t-p} + \varepsilon_t, \quad t \in \mathbb{Z},$$

where $\{\varepsilon_t, t \in \mathbb{Z}\}$ is an iid sequence of non-negative integer-valued random variables with mean $E(\varepsilon_t) = \omega_0 > 0$ and the symbol $\circ$ denotes the binomial thinning operator.

Explicit WLSEs

The WLSEs have explicit forms for estimating INARCH:

$$\widehat{\theta}_{1WLS} = \left(\sum_{t=1}^{n} \frac{\chi_t \chi_t'}{w_t}\right)^{-1} \sum_{t=1}^{n} \frac{Y_t \chi_t}{w_t}.$$

Similarly, we have the following explicit 2WLSE

$$\widehat{\theta}_{2WLS}^{(P)} = \left(\sum_{t=1}^{n} \frac{\chi_t \chi_t'}{\chi_t' \widehat{\theta}_{1WLS}}\right)^{-1} \sum_{t=1}^{n} \frac{Y_t \chi_t}{\chi_t' \widehat{\theta}_{1WLS}}$$

$$\widehat{\theta}_{2WLS}^{(NB)} = \left(\sum_{t=1}^{n} \frac{\chi_t \chi_t'}{\chi_t' \widehat{\theta}_{1WLS}\left(1 + \frac{\chi_t' \widehat{\theta}_{1WLS}}{\widehat{r}}\right)}\right)^{-1} \sum_{t=1}^{n} \frac{Y_t \chi_t}{\chi_t' \widehat{\theta}_{1WLS}\left(1 + \frac{\chi_t' \widehat{\theta}_{1WLS}}{\widehat{r}}\right)}$$

$$\widehat{\theta}_{2WLS}^{(Inv)} = \left(\sum_{t=1}^{n} \chi_t' \widehat{\theta}_{1WLS} \chi_t \chi_t'\right)^{-1} \sum_{t=1}^{n} \chi_t' \widehat{\theta}_{1WLS} Y_t \chi_t.$$

Assumptions for CAN of the WLS

Stationarity and ergodicity:

A1 Strict stationarity and ergodicity of $\{(Y_t, X_t), t \in \mathbb{N}\}$.

Regularity conditions on $\lambda_t(\cdot)$ and $w_t(\cdot)$, moments conditions:

A2–A8 technical assumptions linear INGARCH case

Boundary conditions:

A9 The true parameter θ_0 belongs to the interior of Θ.

Asymptotic distribution of the WLSE

CAN of the WLSE

Under the assumptions **A1**-**A5**,

$$\widehat{\theta}_{1WLS} \to \theta_0 \qquad \text{a.s.} \qquad \text{as} \qquad n \to \infty.$$

Under **A1**-**A9**, as $n \to \infty$

$$\sqrt{n}\left(\widehat{\theta}_{1WLS} - \theta_0\right) \xrightarrow{d} \mathcal{N}(0, \Sigma) \qquad \Sigma = J^{-1}(\theta_0, w)\, I(\theta_0, w)\, J^{-1}(\theta_0, w)\,.$$

$$I(\theta_0, w) = E\left(\frac{\upsilon_t}{w_t^2}\frac{\partial\lambda_t(\theta_0)\partial\lambda_t(\theta_0)}{\partial\theta\partial\theta'}\right), \quad J(\theta_0, w) = E\left(\frac{1}{w_t}\frac{\partial\lambda_t(\theta_0)\partial\lambda_t(\theta_0)}{\partial\theta\partial\theta'}\right)$$

Asymptotic distribution of the two-stage WLSE

Additional assumptions are needed because, contrary to w_t, $\widehat{w}_{t,n}$ is not $\mathscr{F}_{t-1}$-measurable.
Let $\widetilde{\upsilon}_t^*(\xi) = \upsilon^*\left(Y_{t-1}, \boldsymbol{X}_{t-1}, ..., \widetilde{Y}_0, \widetilde{\boldsymbol{X}}_{-1}, \ldots; \xi\right)$, so that $\widehat{w}_{t,n} = \widetilde{\upsilon}_t^*(\widehat{\xi}_n)$.
When $\widehat{\xi}_n \to \xi_0^*$ and some additions assumptions hold (technical assumptions), the 2WLSE has the asymptotic distribution of the WLSE with $w_t = \upsilon_t^*(\xi_0^*)$.

Asymptotic distribution of the 2WLSE

CAN of the 2WLSE

Under **A1**-**A3**, **A4*** and **A5** $\widehat{\theta}_{2WLS} \to \theta_0$ a.s. as $n \to \infty$.
If in addition **A6**, **A7***, **A8*** and **A9** hold,

$$\sqrt{n}\left(\widehat{\theta}_{2WLS} - \theta_0\right) \xrightarrow{d} \mathcal{N}(0, \Sigma) \qquad \Sigma = J^{-1}(\theta_0, w) I(\theta_0, w) J^{-1}(\theta_0, w).$$

Consistent estimators of J and I are

$$\widehat{J} = \frac{1}{n}\sum_{t=1}^{n} \frac{1}{\widehat{w}_{t,n}} \frac{\partial \widetilde{\lambda}_t(\widehat{\theta}_{2WLS})}{\partial \theta} \frac{\partial \widetilde{\lambda}_t(\widehat{\theta}_{2WLS})}{\partial \theta'},$$

$$\widehat{I} = \frac{1}{n}\sum_{t=1}^{n} \frac{\left\{X_t - \widetilde{\lambda}_t(\widehat{\theta}_{2WLS})\right\}^2}{\widehat{w}_{t,n}^2} \frac{\partial \widetilde{\lambda}_t(\widehat{\theta}_{2WLS})}{\partial \theta} \frac{\partial \widetilde{\lambda}_t(\widehat{\theta}_{2WLS})}{\partial \theta'}.$$

Two-step WLSE with minimum variance

Optimal 2WLSE

If in addition the conditional variance is well specified up to a positive constant, that is $\xi_0^* = \xi_0$ and $\upsilon^*(\cdot) = k\upsilon(\cdot)$ for some $k > 0$, then **A6** can be replaced by **A6*** and

$$\sqrt{n}\left(\widehat{\theta}_{2WLS} - \theta_0\right) \xrightarrow{d} \mathcal{N}\left(0, I^{-1}\right) \quad \text{as} \quad n \to \infty.$$

Moreover the matrix $\Sigma - I^{-1}$ is positive semi-definite.

$$I = I(\theta_0, \upsilon) = E\left(\frac{1}{\upsilon_t}\frac{\partial\lambda_t(\theta_0)\partial\lambda_t(\theta_0)}{\partial\theta\partial\theta'}\right)$$

Comparison with the PQMLE

Under **A1**-**A3**, assumptions similar to **A6**-**A8**, and **A9** with positivity constraints on $\widetilde{\lambda}(\cdot)$, Ahmad and Francq (2016) established CAN of the PQMLE when there is no exogenous variables, and obtained

$$\sqrt{n}\left(\widehat{\theta}_{PQML}-\theta_0\right) \overset{\mathcal{L}}{\underset{n\to\infty}{\to}} N\left(0,\Sigma_P\right), \quad \Sigma_P = J_P^{-1} I_P J_P^{-1}$$

with

$$I_P = E\left(\frac{\upsilon_t(\theta_0)}{\lambda_t^2(\theta_0)} \frac{\partial\lambda_t(\theta_0)\partial\lambda_t(\theta_0)}{\partial\theta\partial\theta'}\right) \text{ and } J_P = E\left(\frac{1}{\lambda_t(\theta_0)} \frac{\partial\lambda_t(\theta_0)\partial\lambda_t(\theta_0)}{\partial\theta\partial\theta'}\right).$$

Comparison with the PQMLE (continued)

Since $I_P = I(\theta_0, \omega)$ and $J_P = J(\theta_0, \omega)$ with $\omega = \{\lambda_t\}$, we deduce that

The optimal WLSE is never asymptotically less efficient than the PQMLE

If the conditional variance is well specified, the two-stage WLSE is asymptotically more efficient than the PQMLE, in the sense that the matrix $\Sigma_P - I^{-1}$ is positive semi-definite.

Optimality for linear exponential distributions

Recall that the set $\{F_\lambda,\ \lambda \in \Lambda\}$ constitutes a one-parameter linear exponential family if for all $\lambda \in \Lambda$

$$P(X = k) = h(k)e^{\eta(\lambda)k - a(\lambda)}, \quad k \in \mathbb{N},$$

Examples: $F_\lambda \sim \mathscr{P}(\lambda)$ (then $\lambda = e^\eta$), or $F_\lambda \sim \mathrm{NB}(r,p)$ with $p = r/(\lambda + r)$) and r is fixed.

Efficiency of the 2WLSE for the exponential family

Assume the MLE is CAN, the distribution of $Y_t \mid \{\lambda_t = \lambda\}$ has the previous linear exponential form, and $\lambda_t(\theta_0)$ belongs almost surely to the interior of Λ. The optimal two-stage WLSE is then asymptotically as efficient as the MLE of θ_0.

The WLS estimators avoid boundary problems

PQMLE and NBQMLE are CAN under similar assumptions. However, because of the presence of $\log\left(\widetilde{\lambda}_t(\theta)\right)$, the condition

$$\lambda : \mathbb{N}^{\infty} \times \Theta \to [\underline{\lambda}, \infty) \quad \text{for some } \underline{\lambda} > 0$$

is imposed for the QMLE. In the INGARCH(1,1) case

$$\lambda_t(\theta) = \omega + \alpha Y_{t-1} + \beta \lambda_{t-1}(\theta),$$

one has to impose $\omega \geq \underline{\lambda}$, $\alpha \geq 0$ and $\beta \geq 0$. When $\beta_0 = 0$ (INARCH case), **A9** is not satisfied. The PQMLE then has a nonstandard asymptotic distribution (see Ahmad and Francq, 2016).
For the WSLE, it is possible to have $\widetilde{\lambda}_t(\theta) < 0$ for some values of θ, and thus **A9** is not really restrictive.

MSE-like loss

Selecting the weighting sequence $\widehat{w}_{t,n}$ by minimizing in $(\widehat{w}_{t,n})$ the MSE-like loss

$$\min_{c} \sum_{t=1}^{n} \left\{ \left(Y_t - \widehat{\lambda}_t \right)^2 - c \widehat{w}_{t,n} \right\}^2 = \sum_{t=1}^{n} \left\{ \left(Y_t - \widehat{\lambda}_t \right)^2 - \hat{c}_n \widehat{w}_{t,n} \right\}^2 ,$$

with

$$\hat{c}_n = \frac{\sum_{t=1}^{n} \left\{ \left(Y_t - \widehat{\lambda}_t \right)^2 \widehat{w}_{t,n} \right\}^2}{\sum_{t=1}^{n} \widehat{w}_{t,n}^2},$$

does not work very well in practice, certainly because the existence of high-order moments is required.

The presence of $\hat{c}_n$ comes from the fact that the optimal weights are of the form $w_t = c\text{Var}(Y_t \mid \mathscr{F}_{t-1})$ with $c > 0$.

QLIKE loss

Inspired by Patton (2011), we thus selected the two-stage WLSE $\widehat{\theta}^*_{2WLS}$ of weighting sequence $\widehat{w}_{t,n}$ which minimizes the QLIKE loss

$$\sum_{t=1}^{n} \frac{\left(Y_t - \widehat{\lambda}_t\right)^2}{\hat{c}_n \widehat{w}_{t,n}} + \log\left(\hat{c}_n \widehat{w}_{t,n}\right), \qquad \hat{c}_n = \frac{1}{n} \sum_{t=1}^{n} \frac{\left(Y_t - \widehat{\lambda}_t\right)^2}{\widehat{w}_{t,n}}.$$

In agreement with Patton, we found that the method based on the QLIKE loss works better than that based on the MSE.

Double-Poisson INARCH(3)

$N = 1000$ replications of length $n = 500$

Table: Bias and RMSE of estimators of the mean parameters

	ω		α_2		α_3	
	Bias	RMSE	Bias	RMSE	Bias	RMSE
$\widehat{\theta}_{PQML}$	0.836	0.960	-0.011	0.053	-0.039	0.065
$\widehat{\theta}_{NBQML}$	1.130	1.294	-0.016	0.069	-0.053	0.089
$\widehat{\theta}_{1WLS}$	0.462	0.596	-0.006	0.043	-0.022	0.047
$\widehat{\theta}^{(P)}_{2WLS}$	0.851	0.984	-0.013	0.055	-0.039	0.066
$\widehat{\theta}^{(NB)}_{2WLS}$	1.006	1.200	-0.019	0.070	-0.044	0.080
$\widehat{\theta}^{(Inv)}_{2WLS}$	0.248	0.479	-0.004	0.041	-0.012	0.041
$\widehat{\theta}^{*}_{2WLS}$	0.248	0.479	-0.004	0.041	-0.012	0.041

Computation time for estimating an INARCH(q)

Sample size $n = 500$

Table: CPU time in seconds

	$q = 3$	$q = 6$	$q = 12$	$q = 24$
$\widehat{\theta}_{PQML}$	0.0242	0.0452	0.1044	0.3968
$\widehat{\theta}_{NBQML}$	0.0444	0.0992	0.2398	0.8440
$\widehat{\theta}_{1WLS}$	0.0052	0.0064	0.0052	0.0066
$\widehat{\theta}^{(P)}_{2WLS}$	0.0098	0.0094	0.0134	0.0202
$\widehat{\theta}^{(NB)}_{2WLS}$	0.0092	0.0106	0.0150	0.0198
$\widehat{\theta}^{(Inv)}_{2WLS}$	0.0092	0.0134	0.0146	0.0194
$\widehat{\theta}^{*}_{2WLS}$	0.0330	0.0384	0.0532	0.0740

Reliability of the asymptotic theory in finite samples

Sample size $n = 500$

Table: Comparing RMSE to Mean Asymptotic Standard Error (MASE), and PCT to the nominal level $\alpha = 5\%$ for $\alpha_{02} = 0$ and to 100% for $\beta_0 > 0$

	α_{02}			β_0		
	RMSE	MASE	PCT	RMSE	MASE	PCT
$\widehat{\theta}_{PQML}$	0.058	0.077	6.4	0.093	0.103	99.8
$\widehat{\theta}_{NBQML}$	0.059	0.078	6.4	0.094	0.105	99.5
$\widehat{\theta}_{1WLS}$	0.078	0.078	6.7	0.102	0.102	99.6
$\widehat{\theta}^{(P)}_{2WLS}$	0.076	0.076	6.4	0.099	0.100	99.8
$\widehat{\theta}^{(NB)}_{2WLS}$	0.078	0.077	6.9	0.103	0.101	99.3
$\widehat{\theta}^{(Inv)}_{2WLS}$	0.082	0.082	7.0	0.108	0.108	99.2
$\widehat{\theta}^{*}_{2WLS}$	0.082	0.082	7.0	0.108	0.108	99.2

Reliability of the asymptotic theory in finite samples

Sample size $n = 2000$

Table: Comparing RMSE to MASE, and (PouCenTage of Rejection) PCT to the nominal level $\alpha = 5\%$ for $\alpha_{02} = 0$ and to 100% for $\beta_0 > 0$

	α_{02}			β_0		
	RMSE	MASE	PCT	RMSE	MASE	PCT
$\widehat{\theta}_{PQML}$	0.028	0.038	5.6	0.042	0.048	100
$\widehat{\theta}_{NBQML}$	0.029	0.038	5.8	0.043	0.049	100
$\widehat{\theta}_{1WLS}$	0.038	0.038	5.9	0.047	0.048	100
$\widehat{\theta}^{(P)}_{2WLS}$	0.037	0.038	5.7	0.046	0.047	100
$\widehat{\theta}^{(NB)}_{2WLS}$	0.038	0.038	5.7	0.047	0.047	100
$\widehat{\theta}^{(Inv)}_{2WLS}$	0.041	0.041	5.9	0.050	0.051	100
$\widehat{\theta}^{*}_{2WLS}$	0.041	0.041	5.9	0.050	0.051	100

Conclusion

For positive time series with time-varying conditional mean, simple stationarity and ergodicity conditions exist under

- the stochactic-equal-mean order condition;
- moment and mixing conditions are also available;
- the multiplicative form of standard ACD-type models is questionable.

Conclusion

The WLS estimators do not require the whole knowledge of the conditional distribution. Compared to the Poisson QMLE or NB QMLE, the WLSE presents the advantages of

- being of higher efficiency in some situations;
- being asymptotically efficient when the conditional distribution belongs to the linear exponential family;
- having a standard asymptotic normal distribution even when one or several coefficients of the conditional mean are equal to zero;
- being explicit and requiring no optimisation routine for INARCH models.

Thanks for your attention ☺ !

Conclusion

The WLS estimators do not require the whole knowledge of the conditional distribution. Compared to the Poisson QMLE or NB QMLE, the WLSE presents the advantages of

- being of higher efficiency in some situations;
- being asymptotically efficient when the conditional distribution belongs to the linear exponential family;
- having a standard asymptotic normal distribution even when one or several coefficients of the conditional mean are equal to zero;
- being explicit and requiring no optimisation routine for INARCH models.

Thanks for your attention ☺ !

Assumptions for consistency of the WLS

This assumption is discussed in the last section:

A1 Strict stationarity and ergodicity of $\{(Y_t, X_t), t \in \mathbb{N}\}$.

Asymptotic irrelevance of the initial values:

A2 Letting $a_t = \sup_{\theta \in \Theta} \left| \widetilde{\lambda}_t(\theta) - \lambda_t(\theta) \right|$, a.s. $\lim_{t\to\infty} \left\{ \sup_{\theta\in\Theta} \lambda_t(\theta) + Y_t + 1 \right\} a_t = 0$.

Identifiability condition:

A3 $\lambda_t(\theta) = \lambda_t(\theta_0)$ a.s. if and only if $\theta = \theta_0$.

Asymptotic irrelevance of the initial values:

A4 Almost surely, as $t \to \infty$

$$|w_t - \widetilde{w}_t| \left\{ 1 + Y_t^2 + \sup_{\theta\in\Theta} \lambda_t^2(\theta) \right\} \to 0.$$

Choice of the weight function:

A5 $E\left(\frac{v_1}{w_1}\right) < \infty$ (with $v_t = \mathrm{Var}(Y_t \mid \mathscr{F}_{t-1})$).

Extra assumptions for AN of the WLS

Information matrices:

A6 The matrices $I(\theta_0, w) = E\left(\frac{\upsilon_t}{w_t^2} \frac{\partial\lambda_t(\theta_0)\partial\lambda_t(\theta_0)}{\partial\theta\partial\theta'}\right)$ and $J(\theta_0, w) = E\left(\frac{1}{w_t} \frac{\partial\lambda_t(\theta_0)\partial\lambda_t(\theta_0)}{\partial\theta\partial\theta'}\right)$ exist and $J(\theta_0, w)$ is invertible.

Smoothness of the condition mean and moments:

A7 Almost surely, the function $\lambda_t(\cdot)$ admits continuous second-order derivatives in a neighbourhood $V(\theta_0)$ of θ_0, and we have $Ew_t^{-1} \sup_{\theta \in V(\theta_0)} \{Y_t - \lambda_t(\theta)\}^2 < \infty$,

$$Ew_t^{-1} \sup_{\theta \in V(\theta_0)} \left\| \frac{\partial^2 \lambda_t(\theta)}{\partial\theta\partial\theta'} \right\|^2 < \infty, \quad Ew_t^{-1} \sup_{\theta \in V(\theta_0)} \left\| \frac{\partial\lambda_t(\theta)}{\partial\theta} \frac{\partial\lambda_t(\theta)}{\partial\theta'} \right\| < \infty.$$

Extra assumptions for AN of the WLS (continued)

Asymptotic irrelevance of the initial values:

A8 Letting $b_t = \sup_{\theta\in\Theta} \left\| \partial\widetilde{\lambda}_t(\theta)/\partial\theta - \partial\lambda_t(\theta)/\partial\theta \right\|$, the sequences

$$b_t \left\{ Y_t + \sup_{\theta\in\Theta} |\lambda_t(\theta)| \right\}, \quad a_t \sup_{\theta\in\Theta} \left\| \frac{\partial\lambda_t(\theta)}{\partial\theta} \right\|$$

and

$$|w_t - \widetilde{w}_t| \sup_{\theta\in\Theta} \left\| \frac{\partial\lambda_t(\theta)}{\partial\theta} \right\| \left\{ Y_t + \sup_{\theta\in\Theta} |\lambda_t(\theta)| \right\}$$

are a.s. of order $O(t^{-\kappa})$ for some $\kappa > 1/2$.

Boundary conditions:

A9 The true parameter θ_0 belongs to the interior of Θ. return

Extra assumptions for two-stage WLSE

Additional assumptions are needed because, contrary to w_t, $\widehat{w}_{t,n}$ is not $\mathscr{F}_t$-measurable.
Let $\widetilde{v}_t^*(\xi) = v^*\left(Y_{t-1}, X_{t-1}, ..., \widetilde{Y}_0, \widetilde{X}_{-1}, \ldots; \xi\right)$, so that $\widehat{w}_{t,n} = \widetilde{v}_t^*(\widehat{\xi}_n)$, and let $w_t = v_t^*(\xi_0^*)$.

A4* There exists $\underline{\sigma} > 0$ such that, almost surely, $w_t > \underline{\sigma}$ and $\widehat{w}_{t,n} > \underline{\sigma}$ for n large enough. Assume $\widehat{\xi}_n$ is a strongly consistent estimator of ξ_0^*, the function $v_t^*(\cdot)$ is almost surely continuously differentiable,

$$\sup_{\xi \in V(\xi_0^*)} \left|\widetilde{v}_t^*(\xi) - v_t^*(\xi)\right| \le K\rho^t \quad E\frac{1}{w_t} \sup_{\xi \in V(\xi_0^*)} \left\| \frac{\partial v_t^*(\xi)}{\partial \xi} \right\| \sup_{\theta \in \Theta} \{Y_t - \lambda_t(\theta)\}^2 < \infty,$$

where K is a positive random variable $\mathscr{F}_0$-measurable, $\rho \in [0,1)$, and $V(\xi_0^*)$ is a neighborhood of ξ_0^*. Moreover, assume

$$E \sup_{\theta \in \Theta} |Y_t - \lambda_t(\theta)|^s < \infty \quad \text{for some } s > 0.$$

Extra assumptions for two-stage WLSE (continued)

A6* The matrix $I = E\left(\frac{1}{\upsilon_t}\frac{\partial\lambda_t(\theta_0)\partial\lambda_t(\theta_0)}{\partial\theta\partial\theta'}\right)$ exists and is invertible.

A7*= **A7** + $\sqrt{n}\left(\widehat{\xi}_n - \xi_0^*\right) = O_P(1)$ and

$$E\frac{1}{w_t}\sup_{\xi\in V(\xi_0^*)}\left\|\frac{\partial\upsilon_t^*(\xi)}{\partial\xi}\right\|^2\left[1+\sup_{\theta\in V(\theta_0)}\{Y_t-\lambda_t(\theta)\}^2\right]<\infty.$$

A8*= **A8** by replacing $|\widetilde{w}_t - w_t|$ by $\sup_{\xi\in V(\xi_0^*)}|\widetilde{\upsilon}_t(\xi)-\upsilon_t(\xi)|$, for some neighborhood $V(\xi_0^*)$ of ξ_0^*. return

Invertibility and identifiability assumptions

$$\lambda_t = \omega + \sum_{i=1}^{q} \alpha_i Y_{t-i} + \sum_{j=1}^{p} \beta_j \lambda_{t-j} + \boldsymbol{\pi}' \boldsymbol{X}_{t-1},$$

In addition to the previous conditions, assume

- $\sum_{j=1}^{p} \beta_j < 1$ for all $\theta \in \Theta$,
- $q > 0$,

$$\mathcal{A}_{\theta_0}(z) := \sum_{i=1}^{q} \alpha_{0i} z^i \quad \text{and} \quad \mathcal{B}_{\theta_0}(z) := 1 - \sum_{i=1}^{p} \beta_{0i} z^i$$

  have no common root, at least one $\alpha_{0i} \neq 0$ for $i = 1, \ldots, q$, and $\beta_{0p} \neq 0$ if $\alpha_{0q} = 0$.
- for all h, $Y_t \mid \{\boldsymbol{X}_u, u < t+h; Y_u, u < t\}$ is not degenerated,
- $\boldsymbol{\pi}' X_t$ is not degenerated when $\pi \neq 0$.

Other assumptions

For instance, one can take the weighting sequence

$$\widetilde{w}_t = c + aY_{t-1} + b\widetilde{w}_{t-1}$$

with $c > 0$, $a > 0$ and $b \in (0,1)$.

All the assumptions required for consistency of the WLSE are satisfied.
If $EY_t^4 < \infty$, the assumptions for the CAN are satisfied.

The condition $EY_t^4 < \infty$ is implied by the stationarity condition $\sum_{i=1}^q \alpha_{0i} + \sum_{j=1}^p \beta_{0j} < 1$ and $E\|X_t\|^4 < \infty$

- in the Poisson case $Y_t \mid \mathscr{F}_{t-1} \sim \mathscr{P}(\lambda_t)$
- when $Y_t \mid \mathscr{F}_{t-1} \sim NB(r_t, p)$ with $r_t = \lambda_t p/(1-p)$

Moment condition for the NB distribution with fixed r

Consider the INGARCH(1,1) case

$$\lambda_t(\theta_0) = \omega_0 + \alpha_0 Y_{t-1} + \beta_0 \lambda_{t-1}(\theta_0).$$

When $Y_t \mid \mathscr{F}_{t-1} \sim NB(r, p_t)$ with $p_t = r/(r+\lambda_t)$ we have $EY_t^2 < \infty$ if and only if

$$(\alpha_0 + \beta_0)^2 + \frac{\alpha_0^2}{\varsigma_0} < 1,$$

and $EY_t^4 < \infty$ if and only if

$$(\alpha_0 + \beta_0)^4 + \frac{6\alpha_0^2(\alpha_0+\beta_0)^2}{r} + \frac{\alpha_0^3(11\alpha_0 + 8\beta_0)}{r^2} + \frac{6\alpha_0^4}{r^3} < 1.$$

return

Construction of the stationarity solution

in the case $p = q = 1$

Let (U_t) iid $\mathcal{U}_{[0,1]}$ and the quantile function F_λ^-.
For $t \in \mathbb{Z}$, let $Y_t^{(k)} = \lambda_t^{(k)} = 0$ when $k \leq 0$ and, for $k > 0$, let

$$Y_t^{(k)} = F_{\lambda_t^{(k)}}^-(U_t), \qquad \lambda_t^{(k)} = \omega + \alpha Y_{t-1}^{(k-1)} + \beta \lambda_{t-1}^{(k-1)} + \boldsymbol{\pi}^\top \boldsymbol{X}_{t-1}.$$

Under the stochastic-equal-mean order condition,

$$E\left|\lambda_t^{(k)} - \lambda_t^{(k-1)}\right| = (\alpha + \beta) E\left(\lambda_{t-1}^{(k-1)} - \lambda_{t-1}^{(k-2)}\right) = (\alpha + \beta)^{k-1} \omega.$$

It follows that the sequence $\left\{\lambda_t^{(k)}\right\}_k$ converges in L^1 and a.s to the stationary solution λ_t.

return