Soft-clipping INGARCH Models for Time Series of Bounded Counts

HELMUT SCHMIDT UNIVERSITÄT
Universität der Bundeswehr Hamburg

MATH STAT

Christian H. Weiß, Malte Jahn

Department of Mathematics & Statistics,

Helmut Schmidt University, Hamburg

HELMUT SCHMIDT UNIVERSITÄT
Universität der Bundeswehr Hamburg

MATH STAT

The Soft-clipping Approach for Bounded Counts

Motivation

Count processes (X_t): range contained in $\mathbb{N}_0$,

i. e., set of *non-negative* integers (see Weiß, 2018).

$\Rightarrow$ *exactly* linear count time series model possible

if restricting to non-negative autocorrelation function (ACF).

For practice, models also allowing for negative ACF required.

Possible to find *approximately* linear models?

Distinguish **two cases:**

- **unbounded** counts (X_t), i. e., full $\mathbb{N}_0$ as range;
- **bounded** counts (X_t), i. e., upper bound $n \in \mathbb{N}$,

  so range $\{0, \ldots, n\}$ also bounded from above!

INGARCH-type models for **unbounded** counts (X_t):

Let $M_t = E[X_t | \mathcal{F}_{t-1}]$, then recursive scheme

$$M_t \;=\; f\Big(a_0 + \textstyle\sum_{i=1}^{p} a_i\, X_{t-i} + \sum_{j=1}^{q} b_j\, M_{t-j}\Big).$$

Count X_t emitted by distribution with mean M_t,

for example, $X_t | \mathcal{F}_{t-1} \sim \text{Poi}(M_t)$.

Ordinary (linear) INGARCH model (Ferland et al., 2006):

response function $f = \text{id}$, but then constraints

$a_0 > 0$ and $a_1, \ldots, a_p, b_1, \ldots, b_q \geq 0$ to ensure $M_t > 0$

(plus constraint $a_1 + \ldots + b_q < 1$ for stationary solution).

Soft-clipping Approach for Bounded Counts

INGARCH-type models for **unbounded** counts (X_t):

Let $M_t = E[X_t|\mathcal{F}_{t-1}]$, then recursive scheme

$$M_t \;=\; f\Big(a_0 + \textstyle\sum_{i=1}^{p} a_i\, X_{t-i} + \sum_{j=1}^{q} b_j\, M_{t-j}\Big).$$

Maybe **ReLU INGARCH model**?

If choosing rectified linear unit function,

$$\mathrm{ReLU}(x) = \max\{0, x\},$$

then $M_t = 0$ might happen, so degenerate count distribution.

Furthermore, $\mathrm{ReLU}(x)$ not differentiable in 0.

Solution by Weiß et al. (2022):

softplus INGARCH model with response function $f = \mathrm{sp}_c$.

Softplus function

$\mathrm{sp}_c(x) = c \ln\big(1 + \exp(x/c)\big)$

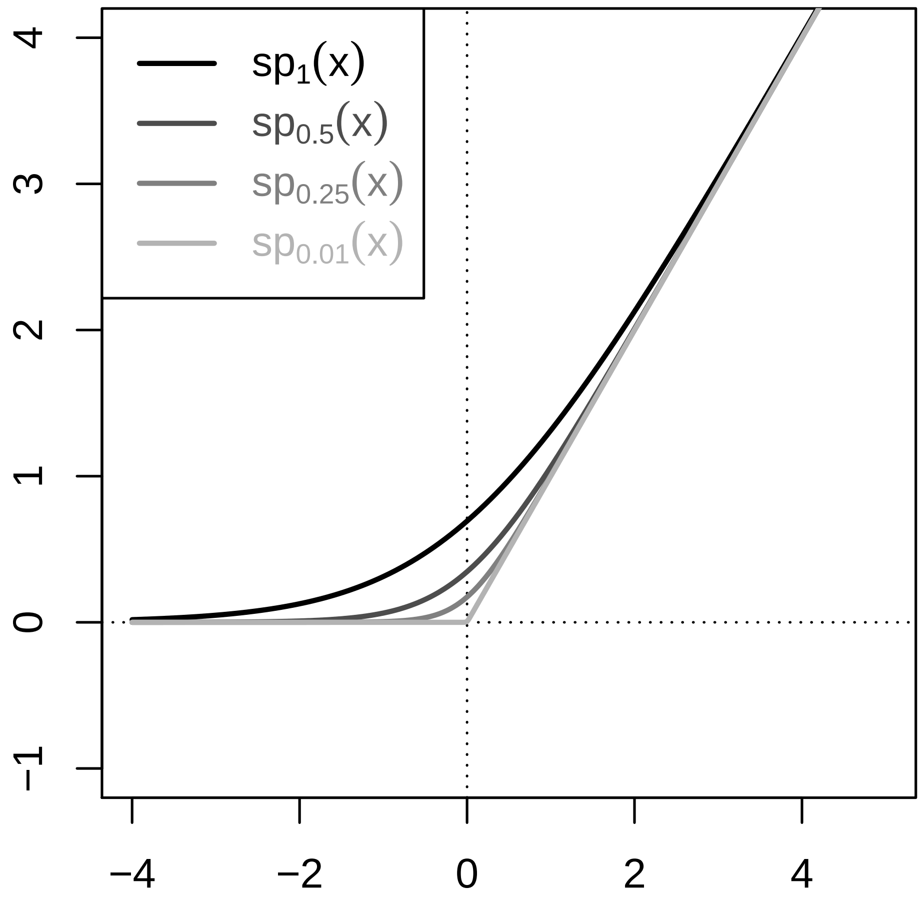
with $c > 0$;

(Mei & Eisner, 2017)

truly positive

and differentiable,

$\mathrm{sp}_c(x) \to \mathrm{ReLU}(x)$
for $c \to 0$.



INGARCH-type models for **bounded** counts (X_t), $\leq n \in \mathbb{N}$:

$$M_t \;=\; f\Big(n\,\alpha_0 + \textstyle\sum_{i=1}^{p} \alpha_i\, X_{t-i} + \sum_{j=1}^{q} \beta_j\, M_{t-j}\Big) \quad \text{with } f : \mathbb{R} \to (0;n),$$

or equivalently, using **normalized conditional mean**

$P_t = \frac{1}{n}\, M_t$ together with $f^*(x) = \frac{1}{n}\, f(n \cdot x)$:

$$P_t \;=\; f^*\Big(\alpha_0 + \textstyle\sum_{i=1}^{p} \alpha_i\, X_{t-i}/n + \sum_{j=1}^{q} \beta_j\, P_{t-j}\Big), \qquad f^* : \mathbb{R} \to (0;1).$$

Counts $X_t|\mathcal{F}_{t-1}$ emitted by, for example, $\text{Bin}(n, P_t)$.

Ristić et al. (2016): $f^* = \text{id} \Rightarrow$ strict parameter constraints;

Chen et al. (2020): $f^* = \text{logit}^{-1} \Rightarrow$ distinctly non-linear;

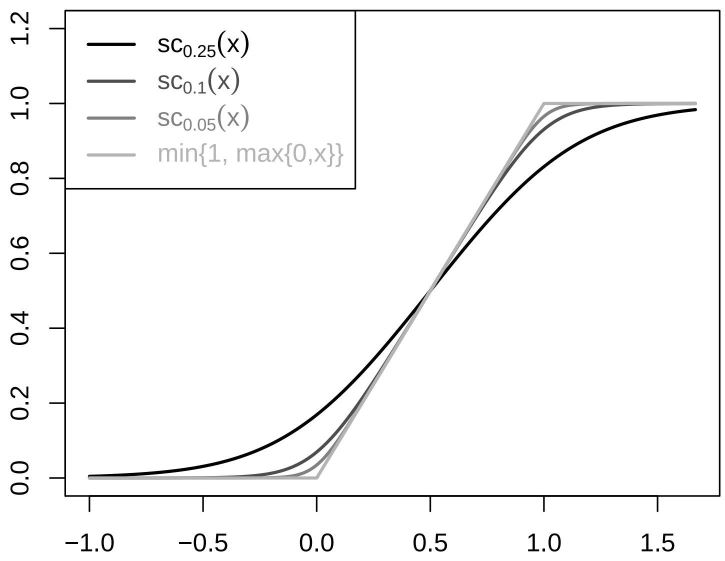
whereas softplus function not bounded from above.

Idea: use **soft clipping** (Klimek & Perelstein, 2020),

$$\mathrm{sc}_c : \mathbb{R} \to (0;1), \qquad \mathrm{sc}_c(x) \;=\; c\,\ln\left(\frac{1+\exp(\frac{x}{c})}{1+\exp(\frac{x-1}{c})}\right),$$

where $\mathrm{sc}_c(x) \;\to\; \mathrm{cReLU}(x) := \min\{1, \max\{0, x\}\}$

for $c \to 0$ (clipped ReLU, see Cai et al., 2017).



Soft-clipping INGARCH model
uses response function

$f^*(x) = \mathrm{sc}_c(x)$

for $P_t =$
$f^*\Big(\alpha_0 + \sum_{i=1}^{p} \alpha_i\, X_{t-i}/n + \sum_{j=1}^{q} \beta_j\, P_{t-j}\Big)$.

HELMUT SCHMIDT UNIVERSITÄT
Universität der Bundeswehr Hamburg

MATH STAT

Soft-clipping Binomial INGARCH Model

Properties

Let (X_t) have range $\{0,\ldots,n\}$.

The **soft-clipping BINGARCH** (scBINGARCH) model

assumes counts $X_t|P_t$ emitted by $\text{Bin}(n,P_t)$, where

$$\begin{aligned} X_t|\mathcal{F}_{t-1} &\sim \ \text{Bin}(n,P_t), \\ P_t &= \ \text{sc}_c\Big(\alpha_0+\textstyle\sum_{i=1}^{p}\alpha_i\,X_{t-i}/n+\sum_{j=1}^{q}\beta_j\,P_{t-j}\Big). \end{aligned}$$

Here, $|\alpha_i|,|\beta_j|<1$ for $i=1,\ldots,p$ and $j=1,\ldots,q$,

and we assume $\alpha_0\in(0;1+p+q)$.

Lemma 1 With aforementioned parameter constraints, and with $c > 0$, there exist $\epsilon_1, \epsilon_2 > 0$ such that

$$P_t \in [\epsilon_1; 1 - \epsilon_2] \qquad \text{for all } t.$$

Lemma 2 Let $n \in \mathbb{N}$ and $p_1 \geq p_2 \geq n/\delta > 0$ for some $\delta > 0$. Then, total variation (TV) distance satisfies

$$\begin{aligned} \mathsf{TV}\Big(\mathsf{Bin}(n, p_1), \mathsf{Bin}(n, p_2)\Big) &\leq 1 - \Big(1 - |p_1 - p_2|\Big)^n \\ &\leq 1 - \Big(\frac{p_2}{p_1}\Big)^n \\ &\leq 1 - \exp\Big(-\delta\,|p_1 - p_2|\Big). \end{aligned}$$

(Recall: $\mathsf{TV}(P_1, P_2) = \frac{1}{2} \sum_{x=0}^{\infty} \Big|P_1(x) - P_2(x)\Big|$.)

Lemmata 1 and 2 used to show that conditions (A1)–(A3) in Doukhan & Neumann (2019) hold.

Theorem 1 If scBINGARCH(p,q) satisfies $\sum_{j=1}^{q} |\beta_j| < 1$ and $\sum_{i=1}^{p} \max\{0,\alpha_i\} + \sum_{j=1}^{q} \max\{0,\beta_j\} < 1$, then:

(i) Markov process $(\boldsymbol{Z}_t)$, $\boldsymbol{Z}_t = (X_t,\ldots,X_{t-p+1},P_t,\ldots,P_{t-q+1})$, possesses unique stationary distribution;

(ii) stationary version of (X_t) absolutely regular with β-mixing coefficients bounded by $C\,\rho^{\sqrt{n}}$;

(iii) stationary version of (X_t, P_t) is ergodic.

HELMUT SCHMIDT UNIVERSITÄT
Universität der Bundeswehr Hamburg

MATH STAT

Soft-clipping Binomial INARCH Model

Properties

Soft-clipping Binomial INARCH(1) Model

Except for the i. i. d.-case $\alpha_1 = 0$,

$$X_t|\mathcal{F}_{t-1} \;\sim\; \text{Bin}(n, P_t) \quad \text{with } P_t \;=\; \text{sc}_c\big(\alpha_0 + \alpha_1\, X_{t-1}/n\big)$$

defines primitive and thus ergodic Markov chain,

so unique stationary solution and

ϕ-mixing with geometrically decreasing weights.

Exactly linear competitors: (equivalent for $n = 1$)

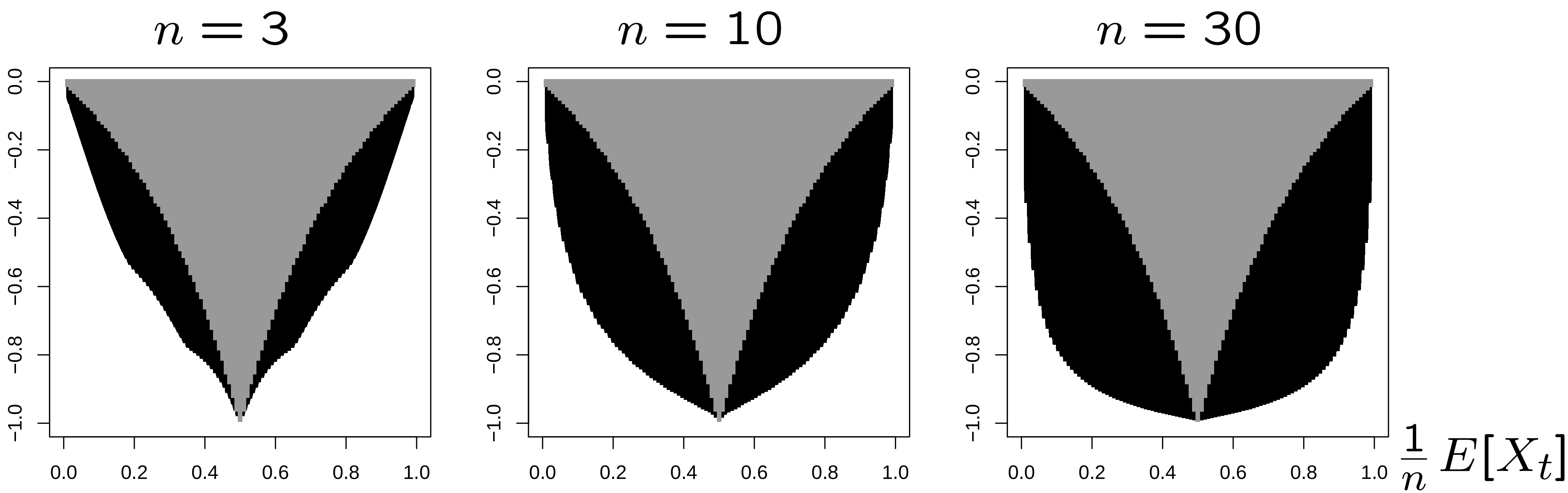
- binomial AR(1) (BinAR(1)) model of McKenzie (1985),
- ordinary BINARCH(1) model (Ristić et al., 2016),

both ACF$(k) = \rho^k$ with $\rho \in \Big(\max\{\frac{-\pi}{1-\pi}, \frac{1-\pi}{-\pi}\}\ ;\ 1\Big)$,

where $\pi = E[X_t/n]$.

Soft-clipping Binomial INARCH(1) Model

Plots of attainable pairs of ACF(1) against $\frac{1}{n}E[X_t]$ for different upper bounds n, where black region corresponds to scBINARCH(1) model with $c = 0.01$, and grey region to BinAR(1) or BINARCH(1) model, respectively:



Soft-clipping Binomial INARCH(1) Model

Which "extent of linearity" achieved by scBINARCH(1) model?

Numerical computation of normalized mean, BID, PACF 1–3.

"Linear values" by plugging-in into BINARCH(1)'s formulae (irrespective of any parameter constraints).

Values for $c = 0$ computed by response function cReLU(x).

Example results:

ρ	n	c	Mean $/n$ sc	Mean $/n$ lin	BID sc	BID lin	PACF(1) sc	PACF(1) lin	PACF(2) sc	PACF(2) lin	PACF(3) sc	PACF(3) lin
$\pi = 0.3$												
-0.5	3	0.1	0.308	0.300	1.150	1.200	-0.440	-0.500	0.004	0.000	0.000	0.000
		0.05	0.303	0.300	1.176	1.200	-0.472	-0.500	0.004	0.000	0.000	0.000
		0.025	0.302	0.300	1.183	1.200	-0.480	-0.500	0.004	0.000	0.000	0.000
		0.01	0.301	0.300	1.184	1.200	-0.482	-0.500	0.004	0.000	0.000	0.000
		0	0.301	0.300	1.184	1.200	-0.482	-0.500	0.004	0.000	0.000	0.000
-0.5	10	0.1	0.304	0.300	1.242	1.290	-0.465	-0.500	0.001	0.000	0.000	0.000
		0.05	0.300	0.300	1.282	1.290	-0.494	-0.500	0.001	0.000	0.000	0.000
		0.025	0.300	0.300	1.289	1.290	-0.499	-0.500	0.000	0.000	0.000	0.000
		0.01	0.300	0.300	1.290	1.290	-0.500	-0.500	0.000	0.000	0.000	0.000
		0	0.300	0.300	1.290	1.290	-0.500	-0.500	0.000	0.000	0.000	0.000
-0.5	30	0.1	0.304	0.300	1.274	1.319	-0.472	-0.500	0.000	0.000	0.000	0.000
		0.05	0.300	0.300	1.315	1.319	-0.498	-0.500	0.000	0.000	0.000	0.000
		0.025	0.300	0.300	1.319	1.319	-0.500	-0.500	0.000	0.000	0.000	0.000
		0.01	0.300	0.300	1.319	1.319	-0.500	-0.500	0.000	0.000	0.000	0.000
		0	0.300	0.300	1.319	1.319	-0.500	-0.500	0.000	0.000	0.000	0.000

Soft-clipping Binomial INARCH(1) Model

scBINARCH(1)'s conditional log-likelihood function:

$$\ell(\boldsymbol{\theta}) = \ln L(\boldsymbol{\theta}; x_1) \; = \; \textstyle\sum_{t=2}^{T} \ln p_{x_t|x_{t-1}}, \quad \text{where}$$

$$p_{j|i} \; = \; \binom{n}{j}\, \mathrm{sc}_c\big(\alpha_0 + \alpha_1 \tfrac{i}{n}\big)^j \Big(1 - \mathrm{sc}_c\big(\alpha_0 + \alpha_1 \tfrac{i}{n}\big)\Big)^{n-j}.$$

Theorem 2 If $c > 0$, existence of consistent CML estimator that is asymptotically normally distributed,

$$\sqrt{T}\,\big(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\big) \;\xrightarrow{d}\; \mathrm{N}\big(0,\; \mathbf{I}^{-1}(\boldsymbol{\theta})\big),$$

where $\mathbf{I}^{-1}(\cdot)$ denotes expected Fisher information, follows from Theorems 2.1 and 2.2 of Billingsley (1961).

Example simulations results (10,000 replications, $c = 0.01$):

α_0	α_1	n	T	Mean of $\hat{\alpha}_0$	Mean of $\hat{\alpha}_1$	Simulated s. e. of $\hat{\alpha}_0$	Simulated s. e. of $\hat{\alpha}_1$	Mean appr. s. e. of $\hat{\alpha}_0$	Mean appr. s. e. of $\hat{\alpha}_1$
0.45	-0.5	10	50	0.450	-0.498	0.043	0.116	0.044	0.115
			100	0.450	-0.499	0.030	0.080	0.030	0.079
			250	0.449	-0.498	0.019	0.049	0.019	0.049
			500	0.450	-0.500	0.013	0.035	0.013	0.034
0.15	0.5	10	50	0.164	0.450	0.042	0.127	0.041	0.128
			100	0.157	0.475	0.028	0.089	0.028	0.088
			250	0.152	0.491	0.017	0.054	0.017	0.054
			500	0.151	0.495	0.012	0.038	0.012	0.038

Extension to AR(p)-type scBINARCH model:

Existence of scBINARCH(p) model,

$$X_t|\mathcal{F}_{t-1} \ \sim\ \mathrm{Bin}(n, P_t) \quad \text{with } P_t \ =\ \mathrm{sc}_c\big(\alpha_0 + \textstyle\sum_{i=1}^{p} \alpha_i\, X_{t-i}/n\big),$$

derived from p-dimensional representation

$\boldsymbol{X}_t = (X_t, \ldots, X_{t-p+1})^\top$,

which again constitutes finite Markov chain.

Including a Feedback Term: scBINGARCH$(1, 1)$ Model

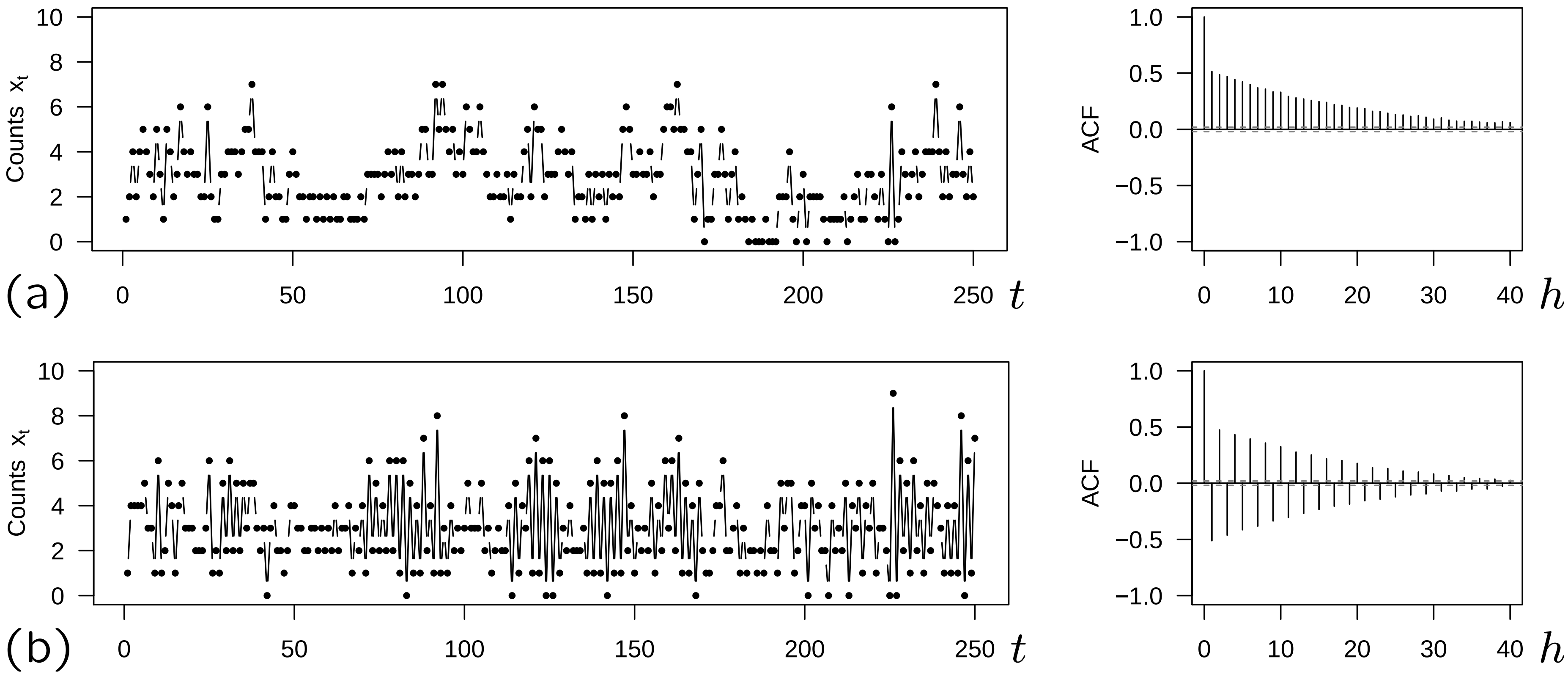
Properties

Feedback Term: scBINGARCH$(1,1)$ Model

Some kind of "long memory" via feedback term:

$$X_t|\mathcal{F}_{t-1} \sim \text{Bin}(n, P_t) \quad \text{with } P_t = \text{sc}_c\big(\alpha_0+\alpha_1\, X_{t-1}/n+\beta_1\, P_{t-1}\big).$$

(a) $\alpha_1 = 0.25,\ \beta_1 = 0.70$; (b) $\alpha_1 = -0.25,\ \beta_1 = -0.70$.



CML estimation using Davis & Liu (2016),

as conditional binomial distribution

belongs to one-parameter exponential family:

Theorem 3 If conditions of Theorem 1, then

scBINGARCH(1, 1)'s CML estimator $\hat{\boldsymbol{\theta}} \to \boldsymbol{\theta}$ almost surely,

and it is asymptotically normally distributed,

$$\sqrt{T}\,\big(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\big) \ \stackrel{d}{\to}\ \mathrm{N}\big(0,\ \mathbf{I}^{-1}(\boldsymbol{\theta})\big).$$

Simulations:

$T \geq 250$ required for good estimation performance,

especially if $|\alpha_1 + \beta_1|$ close to 1.

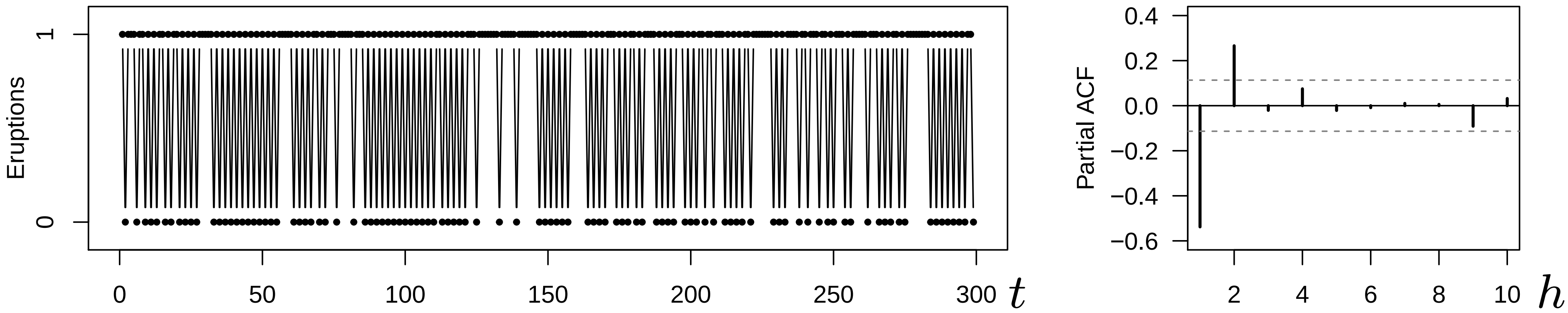
Application: Geyser Eruption Data

Data Example

Application: Geyser Eruption Data

$T = 299$ eruptions of Old Faithful Geyser, "1" if ≥ 3 min.

Time series plot and sample PACF of geyser eruption data:



Jentsch & Reichmann (2019) developed generalized *binary* ARMA (gbARMA) family to account for negative ACFs.

Jentsch & Reichmann (2019) select gbAR(2) model.

scBINARCH(2) nearly identical properties, but

scBINGARCH$(1, 1)$ better ability to slowly decaying ACF.

HELMUT SCHMIDT UNIVERSITÄT
Universität der Bundeswehr Hamburg

MATH STAT

Application: Air Quality Data

Data Example

Application: Air Quality Data

Liu et al. (2021): Daily air quality levels
for thirty Chinese cities (Dec. 2013–July 2019, $T = 2068$),
ordinal scale from $s_0 =$ "excellent" to $s_5 =$ "severely polluted".

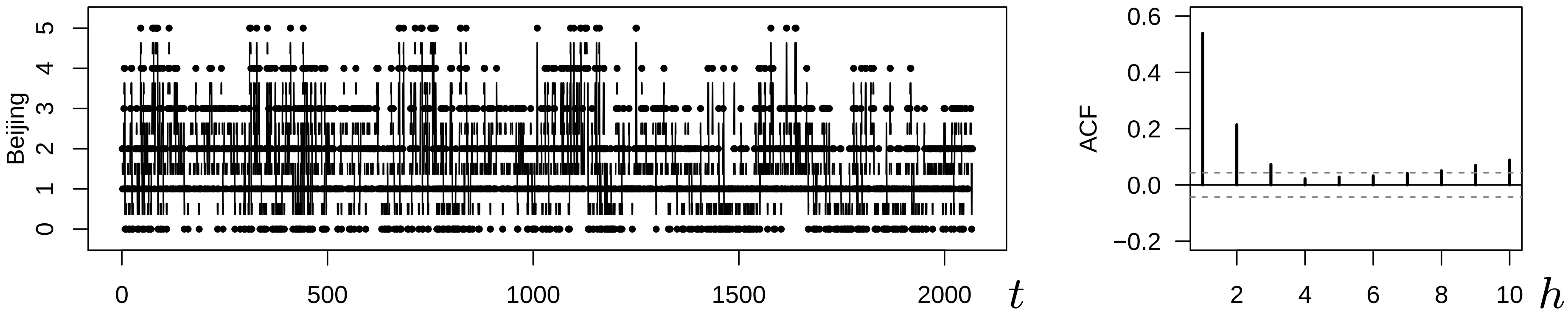
"Rank-count approach": $Y_{t,i} = s_{X_{t,i}}$ at time t in city i,
where $X_{t,i}$ bounded-counts r. v. with $n = 5$.

Liu et al. (2020): linear "ZOB Poisson" INGARCH$(1,1)$,
i. e., truncated Poisson with additional zero and one inflation;
does not allow for negative parameter values!

Thus, scBINGARCH$(1,1)$ sometimes better able
to mimic sample ACF: (. . .)

Application: Air Quality Data

Example: Air quality data for Beijing:



ACF for Beijing series decays rather quickly,

captured by negative β_1 (impossible for ZOB-INGARCH$(1,1)$):

$$(\widehat{\alpha}_0, \widehat{\alpha}_1, \widehat{\beta}_1) \approx (0.180, 0.595, -0.161).$$

scBINGARCH$(1,1)$ not fully able to mimic marginal.

Future extension: allow for zero and one inflation.

Conclusions & Future Research

- scBINGARCH family extends linear BINGARCH model, but gets by with less severe parameter restrictions.
- Behaves like linear BINGARCH for positive parameters, but allows for negative parameter and ACF values.
- Existence and mixing properties for scBINGARCH, consistency and asymptotic normality of CML estimator.

- **Future research:** spatio-temporal modeling of counts, beta-binomial scBINGARCH (extra-binomial variation), zero-(one-)inflated scBINGARCH.

Thank You

for Your Interest!

Christian H. Weiß

Department of Mathematics & Statistics

Helmut Schmidt University, Hamburg

`weissc@hsu-hh.de`

Literature

MATH
STAT

Billingsley (1961) *Statistical Inference for Markov Pr*. Univ. Chicago Press.

Cai et al. (2017) Deep learning with . . . *2017 IEEE CVPR*, 5406–5414.

Chen et al. (2020) Two . . . ARCH models. *Braz. J. Pr. St.* **34**, 685–711.

Davis & Liu (2016) Theory and inference . . . *Stat. Sinica* **26**, 1673–1707.

Doukhan & Neumann (2019) Absolute regularity of . . . *JAP* **56**, 91–115.

Ferland et al. (2006) Integer-valued GARCH processes. *JTSA* **27**, 923–942.

Jentsch & Reichmann (2019) Generalized binary . . . *Econometrics* **7**, 47.

Klimek & Perelstein (2020) Neural network-. . . *SciPost Physics* **9**.

Liu et al. (2021) Modeling normalcy-dominant ordinal . . . *JTSA*, forthc.

McKenzie (1985) Some simple models . . . *Water Res. Bull.* **21**, 645–650.

Mei & Eisner (2017) The neural Hawkes process . . . *NIPS'17*, 6757–6767.

Ristić et al. (2016) A binomial int. ARCH model. *Int. J. Biostatistics* **12**.

Weiß (2018) *An Introduction to Discrete-Valued Time Series*. Wiley.

Weiß et al. (2022) Softplus INGARCH models. *Statistica Sinica* **32**.