Semi-continuous time series for sparse data with volatility clustering

Šárka Hudecová and Michal Pešta

Charles University, Prague

Research Seminar – WU Vienna, Institute for Statistics and Mathematics

June 7, 2024

Work with Šárka Hudecová (CU) based on our paper in J. Time Series Analysis

(Hehuan Shan Dongfeng, 3245 m, Taiwan)

Illustration – series of claim amounts

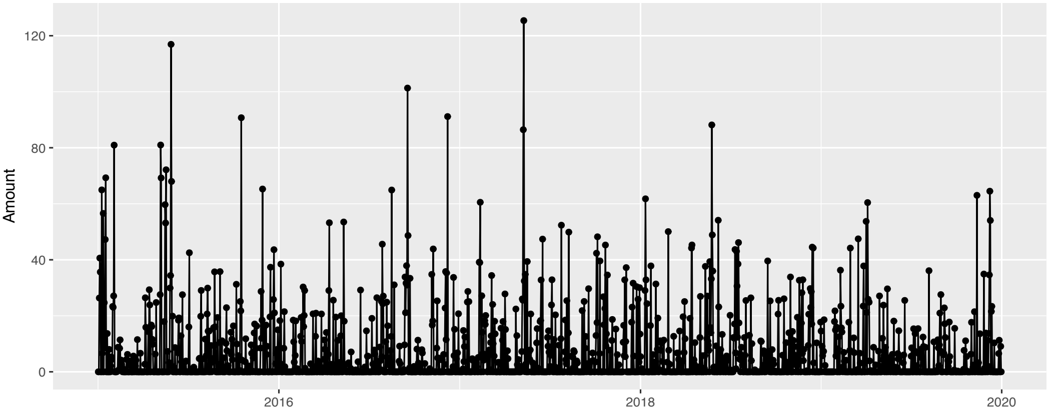


Illustration – series of claim amounts with zeros (detail)

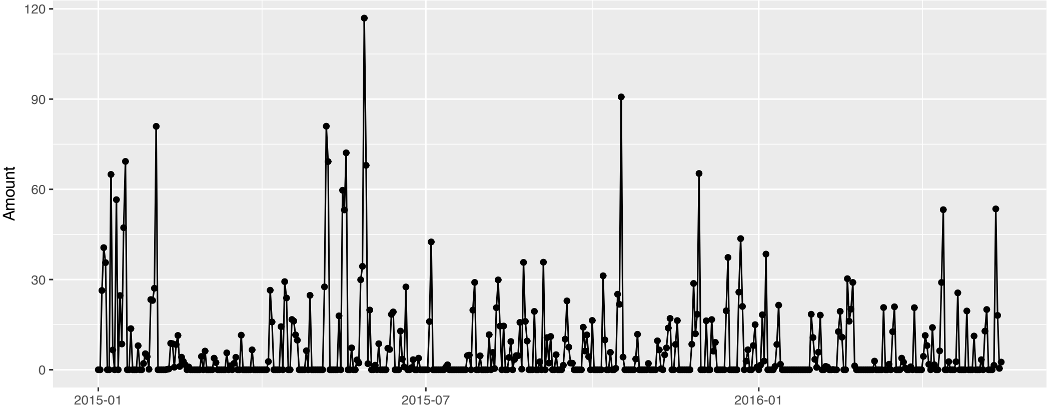
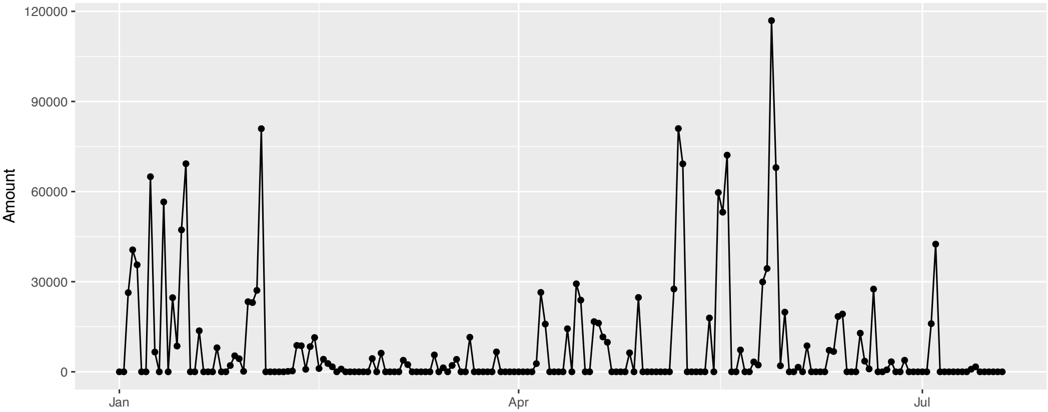


Illustration – volatility clustering (deeper detail)



Motivations

- **Financial markets:** not only returns, but also other market variables, e.g., *financial volumes, number of trades, or financial durations* (Gao et al., 2015) – conditionally heteroscedastic and cannot reach negative values (many zeros)
- **Health services:** an abundance of zeros in *medical expenditures* (Neelon et al., 2016) or (Hudecová et al., 2017)
- **Hydrology:** daily flows on *intermittent and ephemeral streams* Hutton (1990) – the occurrences of zero flows and the time-varying variability of flow
- **Meteorology:** *precipitation amounts* – a significant portion of zero amounts (Cuello et al., 2019)
- **Intermittent demands:** *consumption and production data* – the intermittent demands are erratic and lumpy & many non-zero demand sizes in retail enterprises (Petropoulos et al., 2016)

Introduction and pitfalls

- A *semi-continuous time series* contain a portion of observations equal to a single value (typically zero) and the remaining outcomes are positive.
- ↯ *Naive* approaches usually disregard the zeros, replace the zeros by surrogate small positive values, or aggregate the data, which all lead to loss of information. The logarithmic transformation is not feasible due to zeros.
- ↯ Common analytic procedures of the traditional time series analysis for low-base-rate outputs are often inappropriate for such data and may result in *biased results*.
- → That is why *non-negative* time series have long been a *challenging modeling problem*.

Hurdle GARCH

It is said that a random variable ε has a *hurdle distribution* if

- ε is non-negative having a distribution with a positive mass at point 0;
- the distribution of ε conditional on $\varepsilon > 0$ is continuous on $\mathbb{R}^+$.

Let $\{Y_t\}_{t\in\mathbb{Z}}$ be a time series following a hurdle GARCH(P, Q) model

$$Y_t = \sigma_t \varepsilon_t, \tag{1}$$

$$\sigma_t^2 = \omega + \sum_{i=1}^{P} \alpha_i Y_{t-i}^2 + \sum_{j=1}^{Q} \beta_j \sigma_{t-j}^2, \tag{2}$$

where

- $\omega, \alpha_i, \beta_j, 1 \leqslant i \leqslant P, 1 \leqslant j \leqslant Q$, are unknown parameters;
- $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a sequence of random innovations (not necessarily independent) with a hurdle distribution.

Hurdle distribution

The innovations $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ satisfy

$$\varepsilon_t = e_t E_t,$$

where

- $\{e_t\}_{t\in\mathbb{Z}}$ and $\{E_t\}_{t\in\mathbb{Z}}$ are independent;
- $\{e_t\}_{t\in\mathbb{Z}}$ are independent and identically distributed with a cumulative distribution function (cdf) $F_{\varepsilon|\varepsilon>0}$ and a density $f_{\varepsilon|\varepsilon>0}$ on the support $\mathbb{R}^+$ with respect to the Lebesgue measure;
- $\{E_t\}_{t\in\mathbb{Z}}$ is a strictly stationary irreducible Markov chain with state space $\{0, 1\}$.

Properties

- $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is *strictly stationary and ergodic*
- Every ε_t has a hurdle distribution such that

$$\mathsf{P}[\varepsilon_t \geqslant 0] = 1 \quad \text{and} \quad p := \mathsf{P}[\varepsilon_t = 0] = \mathsf{P}[E_1 = 0] \in (0,1)$$

- The transition probabilities $p(j|i) = \mathsf{P}[E_t = j|E_{t-1} = i], \quad 0 \leqslant i,j \leqslant 1$
- The irreducibility condition $\Rightarrow p(0|1) > 0, p(1|0) > 0, p(0|0) < 1, p(1|1) < 1$

$$p = \frac{p(0|1)}{p(0|1) + p(1|0)} = \frac{1 - p(1|1)}{2 - p(0|0) - p(1|1)}$$

- $\mathsf{E}\,\varepsilon_t = (1-p)\,\mathsf{E}\,e_1$, $\mathsf{E}\,\varepsilon_t^2 = (1-p)\,\mathsf{E}\,e_1^2$, $\mathsf{Var}\,\varepsilon_t = (1-p)\,\mathsf{Var}\,e_1 + (\mathsf{E}\,e_1)^2 p(1-p)$
- The parameters $p(0|0)$ and $p(1|1)$ models the *sparsity* of the non-zero values
- The density of ε_t with respect to a measure $\delta_0 + \lambda_+$ is

$$f_\varepsilon(x) = p\,\mathbb{1}\{x = 0\} + (1-p) f_{\varepsilon|\varepsilon>0}(x)\,\mathbb{1}\{x > 0\}$$

Conditional moments and conditional likelihood

- Let $\mathcal{F}_t = \sigma\{Y_t, \dots, Y_{t-p}, \sigma_t, \dots, \sigma_{t-q}\}$ and recall that $\mathbb{1}\{Y_{t-j} > 0\} = E_{t-j}, j \geqslant 0$

$$\mathsf{E}[Y_t|\mathcal{F}_{t-1}] = \mathsf{E}[\sigma_t e_t E_t|\mathcal{F}_{t-1}] = \sigma_t p(1|E_{t-1})\,\mathsf{E}\,e_1$$
$$\mathsf{E}[Y_t^2|\mathcal{F}_{t-1}] = \mathsf{E}[\sigma_t^2 e_t^2 E_t|\mathcal{F}_{t-1}] = \sigma_t^2 p(1|E_{t-1})\,\mathsf{E}\,e_1^2$$
$$\mathsf{Var}[Y_t^2|\mathcal{F}_{t-1}] = \sigma_t^2 p(1|E_{t-1})\left[\mathsf{Var}(e_1) + (\mathsf{E}\,e_1)^2 p(0|E_{t-1})\right]$$

- Even if $\mathsf{Var}\,\varepsilon_t = 1$, then σ_t^2 is generally not the conditional variance of Y_t, unless the variables $\{E_t\}$ are independent, i.e., $p(1|0) = p(1|1) = 1 - p$
- The conditional distribution of Y_t given $\mathcal{F}_{t-1}$ is hurdle with the cdf

$$F_{Y_t|\mathcal{F}_{t-1}}(y) = \mathsf{P}[Y_t \leqslant y|\mathcal{F}_{t-1}] = p(0|E_{t-1})\mathbb{1}\{y \geqslant 0\} + p(1|E_{t-1})F_{\varepsilon|\varepsilon>0}\left(\frac{y}{\sigma_t}\right)$$

- An *identification* condition on ε_t (or directly e_t) is needed
- Usually $\mathsf{E}\,\varepsilon_t^2 = 1$ (Francq and Zakoïan, 2004)

Conditional density

- $(P+Q+1)$-dimensional vector of the unknown GARCH parameters $\boldsymbol{\vartheta} = [\omega, \boldsymbol{\alpha}^\top, \boldsymbol{\beta}^\top]^\top$ with $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_P]^\top$, $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_Q]^\top$ and let $\boldsymbol{\eta} = [\boldsymbol{\vartheta}^\top, p(0|0), p(1|1)]^\top$
- Assume that the true (unknown) value of $\boldsymbol{\eta}$ is $\boldsymbol{\eta}_0$ satisfying

$$p_0(0|0) \in (0,1), p_0(1|0) \in (0,1), \omega_0 > 0, \alpha_{i,0} \geqslant 0, \beta_{j,0} \geqslant 0$$

- Let us bear in mind that $\sigma_t \equiv \sigma_t(\boldsymbol{\eta})$
- The conditional density of Y_t given $\mathcal{F}_{t-1}$ with respect to $\delta_0 + \lambda_+$ is

$$f_{Y_t|\mathcal{F}_{t-1}}(y_t;\boldsymbol{\eta}) = \{p(0|E_{t-1})\}^{\mathbb{1}\{y_t=0\}} \left\{ \frac{p(1|E_{t-1})}{\sigma_t} f_{\varepsilon|\varepsilon>0}\left(\frac{y_t}{\sigma_t}\right) \right\}^{\mathbb{1}\{y_t>0\}},$$

where the dependence of σ_t on $\boldsymbol{\eta}$ and $\mathcal{F}_{t-1}$ is given in (2)

Conditional log-likelihood

- Assume that $Y_1, \ldots, Y_T$ are the observed data
- Conditionally on the initial (unobservable) values $\boldsymbol{Y}_0, \ldots, \boldsymbol{Y}_{1-P}$ and $\boldsymbol{\sigma}_0, \ldots, \boldsymbol{\sigma}_{1-Q}$ $(\equiv \mathcal{F}_1)$, the *conditional log-likelihood* of $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_T$ becomes

$$\ell_{Y_1,\ldots,Y_T|\mathcal{F}_0}(\boldsymbol{\eta}) = \sum_{t=1}^{T} \Big[\log\{p(0|0)\}\,\mathbb{1}\{Y_t = Y_{t-1} = 0\} + \log\{1 - p(1|1)\}\,\mathbb{1}\{Y_t = 0 \wedge Y_{t-1} > 0\} + \log\{p(1|1)\}\,\mathbb{1}\{Y_t > 0 \wedge Y_{t-1} > 0\} + \log\{1 - p(0|0)\}\,\mathbb{1}\{Y_t > 0 \wedge Y_{t-1} = 0\} + \left\{ \log f_{\varepsilon|\varepsilon>0}\left(\frac{Y_t}{\sigma_t}\right) - \log \sigma_t \right\} \mathbb{1}\{Y_t > 0\} \Big].$$

Half normal distribution

- Distribution of $[\varepsilon|\varepsilon > 0] \sim |Z| \ldots\ Z \sim \mathcal{N}(0, \nu^2)$ for $\nu > 0$

$$f_{\varepsilon|\varepsilon>0}(x) = \sqrt{\frac{2}{\pi\nu^2}} \exp\left\{-\frac{x^2}{2\nu^2}\right\}, \quad x > 0$$

Partial quasi-maximum likelihood estimator

- Gaussian hurdle quasi-maximum likelihood estimator (*Gaussian HQMLE*) $\widehat{\boldsymbol{\eta}}_H = [\widehat{\omega}_H, \widehat{\boldsymbol{\alpha}}_H^\intercal, \widehat{\boldsymbol{\beta}}_H^\intercal, \widehat{p}_H(0|0), \widehat{p}_H(1|1)]^\intercal$ as

$$\arg\min_{\omega,\boldsymbol{\alpha},\boldsymbol{\beta},p(0|0),p(1|1)} \sum_{t=1}^{T} \Bigg[\left\{ (1-p)\frac{Y_t^2}{\widetilde{\sigma}_t^2} + \log\widetilde{\sigma}_t^2 - \log(1-p) \right\} \mathbb{1}\{Y_t > 0\}$$
$$-2\Big[\log\{p(0|0)\}\,\mathbb{1}\{Y_t = Y_{t-1} = 0\} + \log\{1-p(1|1)\}\,\mathbb{1}\{Y_t = 0 \wedge Y_{t-1} > 0\}$$
$$+ \log\{p(1|1)\}\,\mathbb{1}\{Y_t > 0 \wedge Y_{t-1} > 0\} + \log\{1-p(0|0)\}\,\mathbb{1}\{Y_t > 0 \wedge Y_{t-1} = 0\}\Big]\Bigg]$$

- $\widetilde{\sigma}_t^2 \equiv \big(\widetilde{\sigma}_t(\boldsymbol{\eta})\big)^2 := \omega + \sum_{i=1}^{P} \alpha_i Y_{t-i}^2 + \sum_{j=1}^{Q} \beta_j \widetilde{\sigma}_{t-j}^2, \quad t \geqslant 1$
- The initial values $\widetilde{\sigma}_{1-j} = Y_1, 1 \leqslant j \leqslant Q$ and $Y_{1-i} = Y_1, 1 \leqslant i \leqslant P$
- Thus, $\widehat{\sigma}_{H,t} = \widetilde{\sigma}_t(\widehat{\boldsymbol{\eta}}_H)$

Theorem – Strong consistency & AN (Hudecová and P., 2024)

Under regularity assumptions, $\widehat{\boldsymbol{\eta}}_H \xrightarrow[T\to\infty]{a.s.} \boldsymbol{\eta}_0$. *Under additional conditions,*

$$\sqrt{T}\,(\widehat{\boldsymbol{\eta}}_H - \boldsymbol{\eta}_0) \xrightarrow[T\to\infty]{D} \mathsf{N}_{p+q+3}(\mathbf{0}, \boldsymbol{J}_H^{-1}\boldsymbol{I}_H\boldsymbol{J}_H^{-1}),$$

where $\boldsymbol{J}_H = \boldsymbol{J}_0 + 2\boldsymbol{V}$ *and* $\boldsymbol{I}_H = \{(1-p_0)\kappa - 1\}\boldsymbol{J}_0 + 4\boldsymbol{V}$ *such that* $\boldsymbol{J}_0 = \mathsf{E}_{\boldsymbol{\eta}_0}\,\boldsymbol{X}_t\boldsymbol{X}_t^\top$,

$$\boldsymbol{X}_t = \begin{pmatrix} \dfrac{E_t}{\sigma_t^2(\boldsymbol{\eta}_0)}\dfrac{\partial\sigma_t^2}{\partial\boldsymbol{\vartheta}}(\boldsymbol{\eta}_0) \\ \dfrac{p_0}{1-p_0(0|0)}E_t \\ \dfrac{-p_0}{1-p_0(1|1)}E_t \end{pmatrix},$$

$$\boldsymbol{V} = \mathrm{diag}\left\{\mathbf{0}_{(p+q+1)}, \tfrac{p_0}{p_0(0|0)\{1-p_0(0|0)\}}, \tfrac{1-p_0}{p_0(1|1)\{1-p_0(1|1)\}}\right\}, \textit{ and } \kappa = \mathsf{E}\,\varepsilon_t^4.$$

Exponential distribution

- Distribution of $[\varepsilon | \varepsilon > 0] \sim \mathsf{Exp}(1 - p)$

! New identification condition $\mathsf{E}\,\varepsilon_t = 1$

Partial quasi-maximum likelihood estimator (revisited)

- Exponential hurdle quasi-maximum likelihood estimator (*Exponential HQMLE*) $\widehat{\boldsymbol{\eta}}_E = [\widehat{\omega}_E, \widehat{\boldsymbol{\alpha}}_E^\top, \widehat{\boldsymbol{\beta}}_E^\top, \widehat{p}_E(0|0), \widehat{p}_E(1|1)]^\top$ as

$$\arg\min_{\omega,\boldsymbol{\alpha},\boldsymbol{\beta},p(0|0),p(1|1)} \sum_{t=1}^{T} \left[\left\{ (1-p)\frac{Y_t}{\widetilde{\sigma}_t} + \log \widetilde{\sigma}_t - \log(1-p) \right\} \mathbb{1}\{Y_t > 0\} \right.$$
$$- \Big[\log\{p(0|0)\} \mathbb{1}\{Y_t = Y_{t-1} = 0\} + \log\{1-p(1|1)\} \mathbb{1}\{Y_t = 0 \wedge Y_{t-1} > 0\}$$
$$\left. + \log\{p(1|1)\} \mathbb{1}\{Y_t > 0 \wedge Y_{t-1} > 0\} + \log\{1-p(0|0)\} \mathbb{1}\{Y_t > 0 \wedge Y_{t-1} = 0\} \Big] \right]$$

- $\widetilde{\sigma}_t^2 \equiv \left(\widetilde{\sigma}_t(\boldsymbol{\eta})\right)^2 := \omega + \sum_{i=1}^{P} \alpha_i Y_{t-i}^2 + \sum_{j=1}^{Q} \beta_j \widetilde{\sigma}_{t-j}^2, \quad t \geqslant 1$
- The initial values $\widetilde{\sigma}_{1-j} = Y_1, 1 \leqslant j \leqslant Q$ and $Y_{1-i} = Y_1, 1 \leqslant i \leqslant P$
- Thus, $\widehat{\sigma}_{E,t} = \widetilde{\sigma}_t(\widehat{\boldsymbol{\eta}}_E)$

Exponential vs Gaussian quasi-likelihood – Reparametrization

! New identification condition $\mathsf{E}\,\varepsilon_t = 1$

↯ Reparametrize

$$\omega_0^E = \phi^2 \omega_0^G, \quad \boldsymbol{\alpha}_0^E = \phi^2 \boldsymbol{\alpha}_0^G, \quad \boldsymbol{\beta}_0^E = \boldsymbol{\beta}_0^G$$

under $\phi = \mathsf{E}\,\varepsilon_t \in (0, 1)$ for $\mathsf{E}\,\varepsilon_t^2 = 1$

Theorem – Strong consistency & AN – II (Hudecová and P., 2024)

Under regularity assumptions, $\widehat{\boldsymbol{\eta}}_{\mathsf{E}} \xrightarrow[\mathsf{T}\to\infty]{a.s.} \boldsymbol{\eta}_0$. *Under additional conditions,*

$$\sqrt{\mathsf{T}}\,(\widehat{\boldsymbol{\eta}}_{\mathsf{E}} - \boldsymbol{\eta}_0) \xrightarrow[\mathsf{T}\to\infty]{\mathsf{D}} \mathsf{N}_{p+q+3}(\mathbf{0}, \boldsymbol{J}_{\mathsf{E}}^{-1}\boldsymbol{I}_{\mathsf{E}}\boldsymbol{J}_{\mathsf{E}}^{-1}),$$

where $\boldsymbol{J}_{\mathsf{E}} = \boldsymbol{J}_{\mathsf{E},0} + \boldsymbol{V}$ *and* $\boldsymbol{I}_{\mathsf{E}} = \{(1-p_0)\varkappa - 1\}\boldsymbol{J}_{\mathsf{E},0} + \boldsymbol{V}$ *such that*
$\boldsymbol{J}_{\mathsf{E},0} = \mathsf{E}_{\boldsymbol{\eta}_0^{\mathsf{E}}}\, \boldsymbol{X}_{\mathsf{E},t}\boldsymbol{X}_{\mathsf{E},t}^{\mathsf{T}}$,

$$\boldsymbol{X}_{\mathsf{E},t} = \mathsf{E}_t\left(\frac{1}{\sigma_t(\boldsymbol{\eta}_0^{\mathsf{E}})}\frac{\partial\sigma_t}{\partial\boldsymbol{\vartheta}}(\boldsymbol{\eta}_0^{\mathsf{E}}),\ \frac{p_0}{1-p_0(0|0)},\ \frac{-p_0}{1-p_0(1|1)}\right)^{\mathsf{T}},$$

$\boldsymbol{V} = \mathrm{diag}\left\{\mathbf{0}_{(p+q+1)}, \frac{p_0}{p_0(0|0)\{1-p_0(0|0)\}}, \frac{1-p_0}{p_0(1|1)\{1-p_0(1|1)\}}\right\}$, *and* $\varkappa = \mathsf{E}\,\varepsilon_t^2$.

Bootstrap predictions

- To predict the future values of Y_{T+h} and the corresponding σ_{T+h}, $h = 1, \dots, \mathcal{H}$, from the available data $Y_1, \dots, Y_T$
- Besides the point prediction of Y_{T+h} and σ_{T+h}, an *interval prediction* for Y_{T+h} might be of interest
- The nature of the problem implies that to consider only the *upper predictive intervals* for Y_{T+h}

→ *Semi-parametric bootstrap* where the *zero-occurrence* $\{E_t\}$ is *bootstrapped parametrically* using the estimates of the transition probabilities $\widehat{p}(0|0)$ and $\widehat{p}(1|1)$, while the size of the *non-zero innovations* $\{e_t\}$ is *bootstrapped non-parametrically*

Semiparametric bootstrap – Algorithm

- Time series $\{Y_1, \ldots, Y_T\}$ and number of the bootstrap resamples B.

$\Rightarrow$ Compute $\widehat{\boldsymbol{\eta}}$ and $\{\widehat{\sigma}_t\}_{t=1}^{T+1}$. Define $\widehat{\varepsilon}_t = Y_t/\widehat{\sigma}_t$, $t = 1, \ldots, T$, and consider only the positive $\widehat{\varepsilon}_t$, denoted as $\widehat{e}_1, \ldots, \widehat{e}_{T^*}$, where $T^* = \sum_{t=1}^{T} \mathbb{1}\{Y_t > 0\}$.

For $b = 1$ to B

1: Simulate $\{\widehat{E}_{T+h}^{(b)}\}_{h=1}^{\mathcal{H}}$ as a realization of a Markov chain with the initial value $E_T = \mathbb{1}\{Y_T > 0\}$ and the transition probabilities $\widehat{p}(0|0)$ and $\widehat{p}(1|1)$.

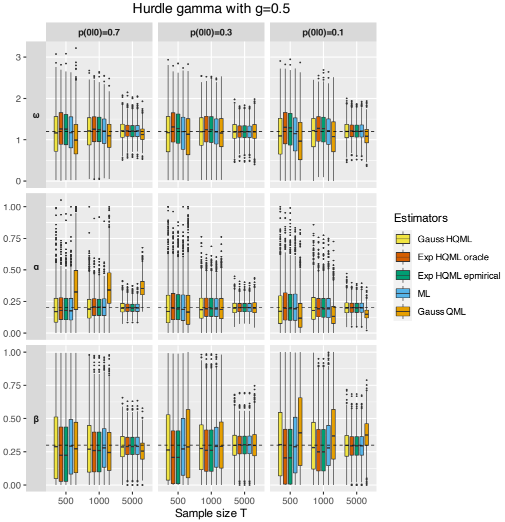
2: Generate $\{\widehat{e}_{T+h}^{(b)}\}_{h=1}^{\mathcal{H}}$ from a distribution with the cdf equal to the empirical distribution function of $\widehat{e}_1, \ldots, \widehat{e}_{T^*}$.

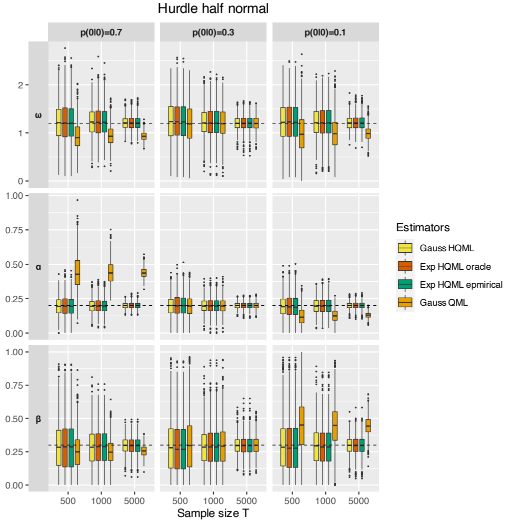
3: Calculate $\widehat{\varepsilon}_{T+h}^{(b)} = \widehat{E}_{T+h}^{(b)} \widehat{e}_{T+h}^{(b)}$.

4: $\widehat{Y}_{T+h}^{(b)} := \widehat{\sigma}_{T+h}^{(b)} \widehat{\varepsilon}_{T+h}^{(b)}$, $h = 1, \ldots, \mathcal{H}$, and $\left(\widehat{\sigma}_{T+h}^{(b)}\right)^2 := \widehat{\omega} + \sum_{i=1}^{P} \widehat{\alpha}_i \left(\widehat{Y}_{T+h-i}^{(b)}\right)^2 + \sum_{j=1}^{Q} \widehat{\beta}_j \left(\widehat{\sigma}_{T+h-j}^{(b)}\right)^2$, $h = 2, \ldots, \mathcal{H}$, where $\{\widehat{Y}_t^{(b)}\}_{t=1}^{T} \equiv \{Y_t\}_{t=1}^{T}$ and $\{\widehat{\sigma}_t^{(b)}\}_{t=1}^{T+1} \equiv \{\widehat{\sigma}_t\}_{t=1}^{T+1}$.

- Empirical (bootstrap) distribution of $\widehat{Y}_{T+h}^{(b)}$, $b = 1, \ldots, B$.

Simulation study – estimation





Synthetic data generating mechanism

- Hurdle GARCH(P, Q) model belongs to a general class of the hurdle models

$$Y_t = \sigma_t \varepsilon_t, \quad \sigma_t = f(Y_{t-1}, \ldots, Y_{t-P}, \sigma_{t-1}, \ldots, \sigma_{t-Q}),$$

which *involves also the hurdle MEM* type of models with various specifications for the conditional mean
- These differ from the hurdle GARCH models by the normalization condition, which, however, does not play a role in predictions
- Data are generated from a *misspecified model, a linear hurdle MEM model*

$$\sigma_t = \omega + \alpha Y_{t-1} + \beta \sigma_{t-1}, \quad \varepsilon_t = E_t e_t, \quad \mathsf{E}\, \varepsilon_t = 1,$$

where $\{E_t\}$ is a Markov chain, and e_t has a generalized $\Gamma(\alpha, \beta, \delta)$ distribution, proportional to $x^{\delta\alpha-1}\exp\{-(x/\beta)^{\delta}\}$, such that $\mathsf{E}\, \varepsilon_t = 1$
- $\Gamma(\alpha, \beta, 1)$ is the ordinary gamma with shape α and scale β, while $\Gamma(1/2, \beta, 2)$ is a half normal distribution

Bootstrap predictions under misspecification

- Comparison with models:

→ *linear hurdle MEM* model

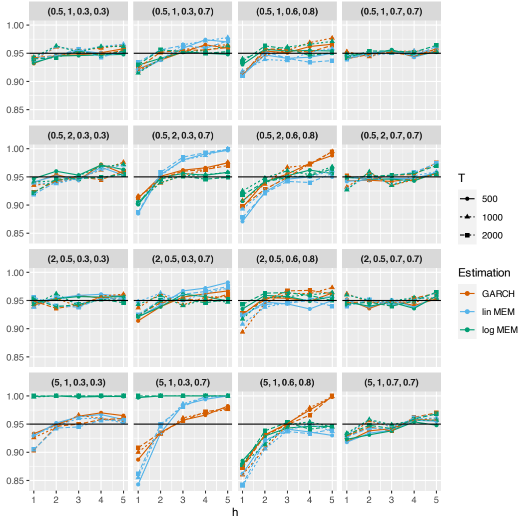
$$\sigma_t = \omega + \alpha Y_{t-1} + \beta \sigma_{t-1}, \quad \varepsilon_t = E_t e_t, \quad \mathsf{E}\,\varepsilon_t = 1,$$

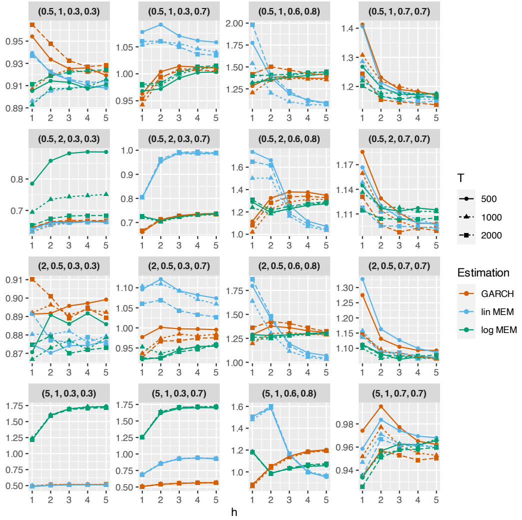
→ fully parametric *logarithmic hurdle MEM* approach from Hautsch et al. (2014)

$$\log \sigma_t = \omega + \alpha_1 \log \varepsilon_{t-1} \mathbb{1}\{Y_{t-1} > 0\} + \alpha_0 \mathbb{1}\{Y_{t-1} = 0\} + \beta \log \sigma_{t-1}$$

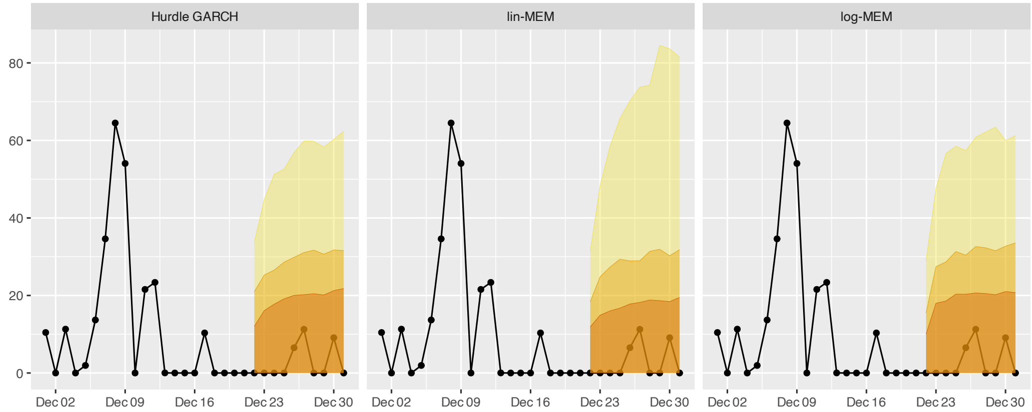
such that ε_t's are supposed to follow a hurdle generalized F distribution

Simulation study – prediction empirical coverage & interval length





Prediction – daily claim amounts – 10 days forecast



Conclusions

- Time series that contain *non-negligible portions of zeros* whereas the remaining observations are positive.
- *No parametric assumptions* on the distribution of the *innovations* are made, whereas the *temporal dependencies* of the series are *parametrized*.
- Our *main contributions* are:
  - proposition of a semi-parametric model for non-negative time series that exhibit *time-varying variability*;
  - proving estimators' *strong consistency* and *asymptotic normality*;
  - utilization of the practical *model selection criteria*;
  - providing *bootstrap predictions*.

Thank you for your attention !

Questions ?

References I

Cuello, W. S., J. R. Gremer, P. C. Trimmer, A. Sih, and S. J. Schreiber (2019). Predicting evolutionarily stable strategies from functional responses of Sonoran Desert annuals to precipitation. *Proceedings of the Royal Society B: Biological Sciences 286*(1894), 20182613.

Francq, C. and J.-M. Zakoïan (2004). Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli 10*(4), 605–637.

Gao, J., N. H. Kim, and P. W. Saart (2015). A misspecification test for multiplicative error models of non-negative time series processes. *J. Econometrics 189*(2), 346–359.

Hautsch, N., P. Malec, and M. Schienle (2014). Capturing the zero: A new class of zero-augmented distributions and multiplicative error processes. *J. Financ. Economet. 12*(1), 89–121.

Hudecová, Š. and M. P. (2024). Semi-continuous time series for sparse data with volatility clustering. *J. Time Ser. Anal. 10.1111/jtsa.12741*.

Hudecová, Š., M. P., and D. Hlubinka (2017). Modelling prescription behaviour of general practitioners. *Math. Slovaca 67*(1), 1–17.

Hutton, J. L. (1990). Non-negative time series models for dry river flow. *J. Appl. Probab. 27*(1), 171–182.

Neelon, B., A. J. O'Malley, and V. A. Smith (2016). Modeling zero-modified count and semicontinuous data in health services research Part 1: background and overview. *Stat. Med. 35*(27), 5070–5093.

Petropoulos, F., N. Kourentzes, and K. Nikolopoulos (2016). Another look at estimators for intermittent demand. *Int. J. Prod. Econ. 181*, 154–161.