LONG-TERM PREDICTION INTERVALS WITH MANY COVARIATES

{ SAYAR KARMAKAR, MAREK CHUDY AND WEI BIAO WU }

UNIVERSITY OF FLORIDA, UNIVERSITY OF VIENNA, UNIVERSITY OF CHICAGO

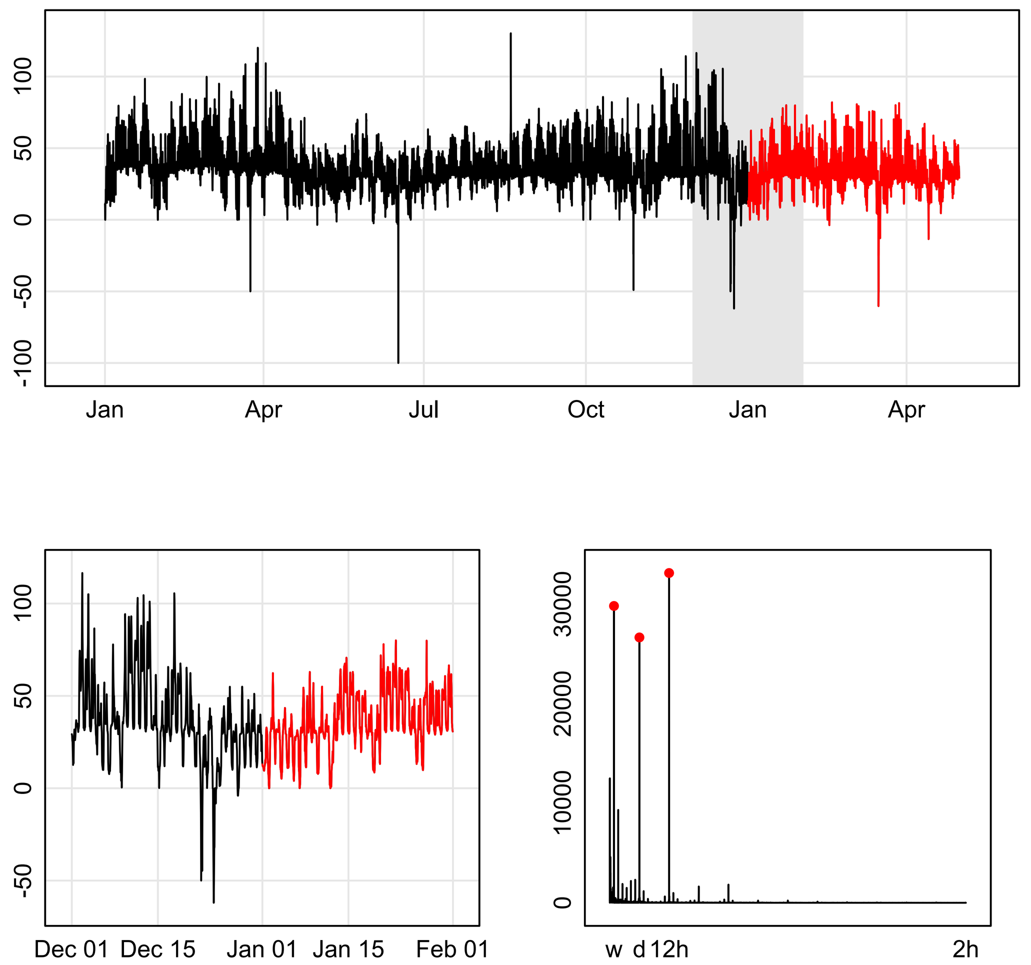
PROBLEM DESCRIPTION

$$y_t = \boldsymbol{x}_t^\top \boldsymbol{\beta} + e_t, \quad \boldsymbol{\beta} \in \mathbb{R}^p,$$

where $p \gg n$, e_i stationary, we want to build prediction intervals (p.i.) for $y_{n+1} + \ldots + y_{n+m}$.

Potential audience

- Policy-makers: US CBO, 75 year GDP
- Insurance: Long-term portfolio
- Investors: Derivatives (market strategy)
- Industry: Electricity pricing, Product plans



A. Full sample, B. Price drop, C. Periodogram

REVIEW

- For $\boldsymbol{\beta} = 0$ Zhou, Wu (2010) constructed p.i. for $e_{n+1} + \ldots + e_{n+m}$ where e_i is dependent linear process.
- When m is large, Chudy, Karmakar, Wu(2019) provided a bootstrap adjustment
- For non-zero β with small p, Zhou, Wu also used LAD residuals.

GOAL

- Goal 1: Does that extend for non-linear e_i?
- Goal 2: Extend for LASSO? i.e. $p \gg n$
- Goal 3: Also allow stochastic design?
- Goal 4: Empirical- Does the bootstrap for long horizon small sample still works?

ERROR SPECIFICATION

e_i is mean-0, stationary process:

- Tail: e_i could be light-tailed ($\mathbb{E}(|e_i|^2) < \infty$))/heavy-tailed ($\mathbb{E}(|e_i|^2) = \infty$)
- Dependence: $e_i = G(\mathcal{F}_i) = G(\eta_i, \eta_{i-1}, \ldots)$ where η are i.i.d.

$$\psi_{k,q} = \sup_{u \in \mathbb{R}} \|f_1(u|\mathcal{F}_k) - f_1(u|\mathcal{F}_k')\|_q.$$

- Short-range/long-range dependent
- e_i could possibly be non-linear.

METHODS

For the easy cases $\beta = 0$, short-range dependence, light tails, the CLT works.

$$[L, U] = \pm \hat{\sigma} Q^t_{\kappa-1}(\alpha/2)\sqrt{m},$$

can be used to predict $e_{n+1} + \ldots + e_{n+m}$. For more general case

- Estimate β by

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta} \in ^p} \frac{1}{n}\sum_{i=1}^{n}(y_i -^\top \boldsymbol{x}_i \boldsymbol{\beta})^2 + \lambda \sum_{j=1}^{p}|\beta_j|.$$

- Predict $e_{n+1} + \ldots + e_{n+m}$ by quantiles of $\hat{e}_i + \ldots + \hat{e}_{i+m-1}; i = 1, \ldots, n-m+1$.
- Then the forecast for $y_{n+1} + \ldots + y_{n+m}$ reads

$$\sum_{i=n+1}^{n+m} \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}} + \text{ Quantiles for } \sum_{i=n+1}^{n+m} \hat{e}_i.$$

LONG HORIZON PREDICTION ADJUSTMENT: BOOTSTRAP STEP

- Stationary bootstrap, replicate residuals $\hat{e}_t = y_t - \hat{y}_t$. Obtain $\hat{e}_t^b, t = 1, \ldots, n, b = 1, \ldots, B$.
- Compute $(\bar{e}^b_{t(m)}) = m^{-1}\sum_{i=1}^{m} e^b_{t-i+1} \forall \quad t$
- Estimate quantiles by Gaussian kernel density estimator from $\bar{e}^b_{n(m)}, b = 1, \ldots, B$.
- This keeps the quantile estimator consistent

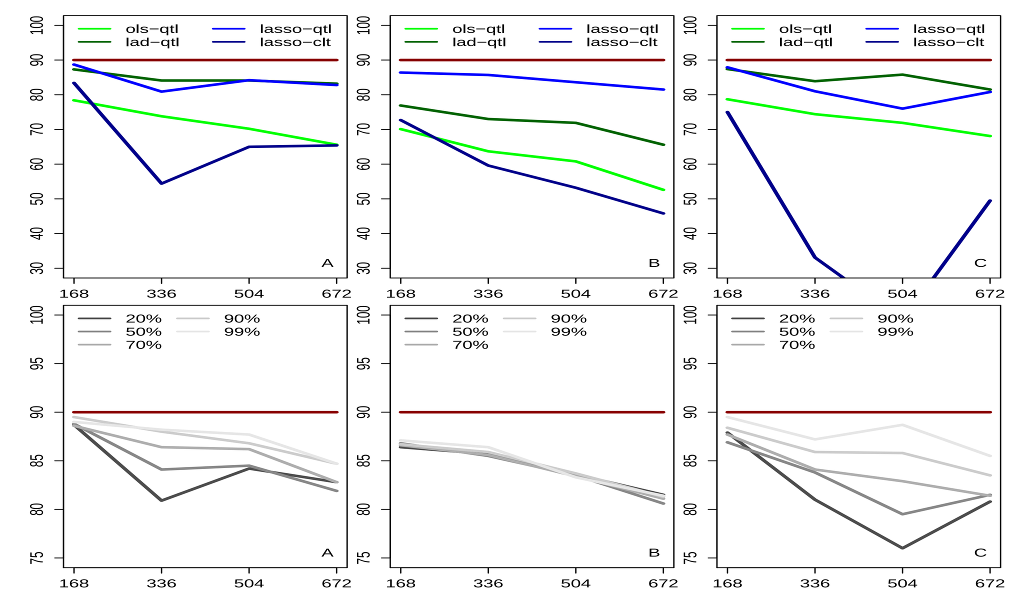
SIMULATION

A,D: Short-range $e_t = \phi_1 e_{t-1} + \sigma \epsilon_t$,

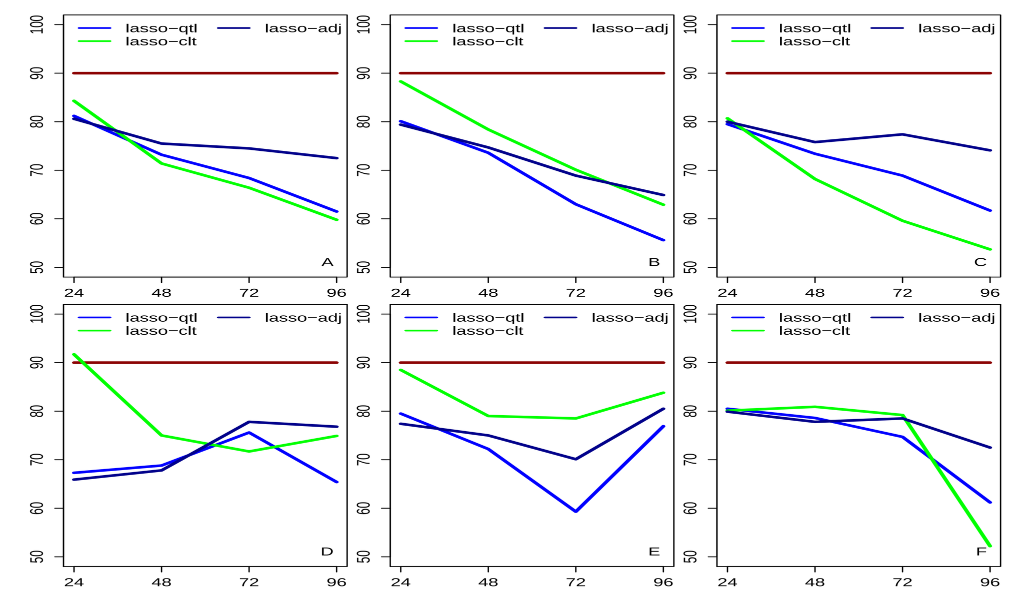
B,E: Long-range $e_t = \sigma \sum_{j=0}^{\infty}(j+1)^\gamma \epsilon_{t-j}$,

C,F: Non-linear $e_t = \phi_1 e_{t-1} + \sigma \epsilon_t + LSTAR$ where $LSTAR$ term is $G(e_{t-1}; \delta, \text{threshold})(\phi_2 e_{t-1})$

Low dimension

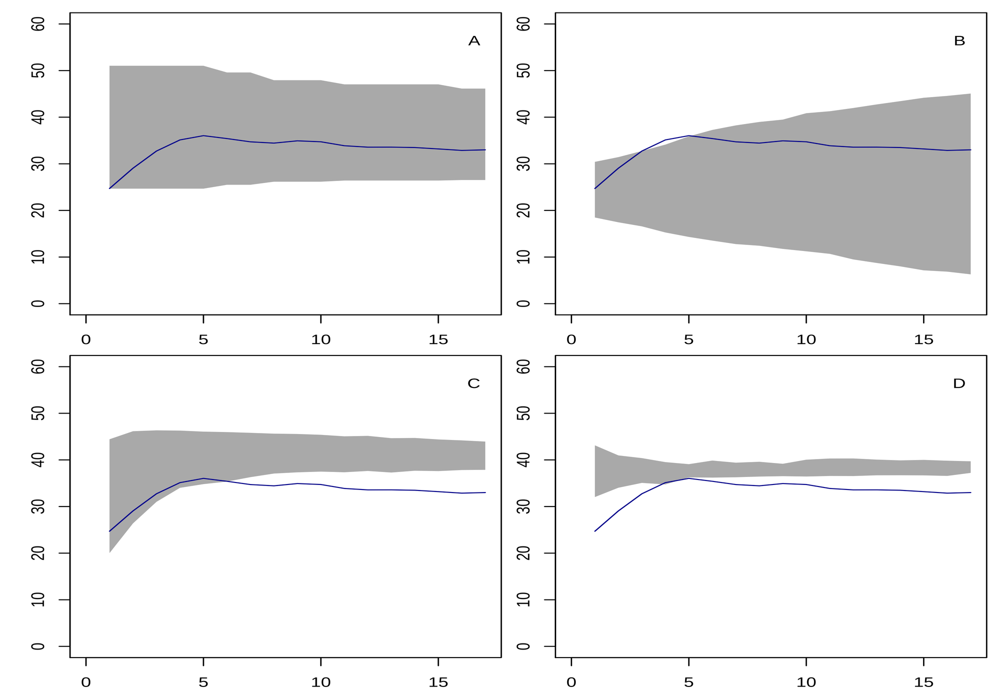


High dimension



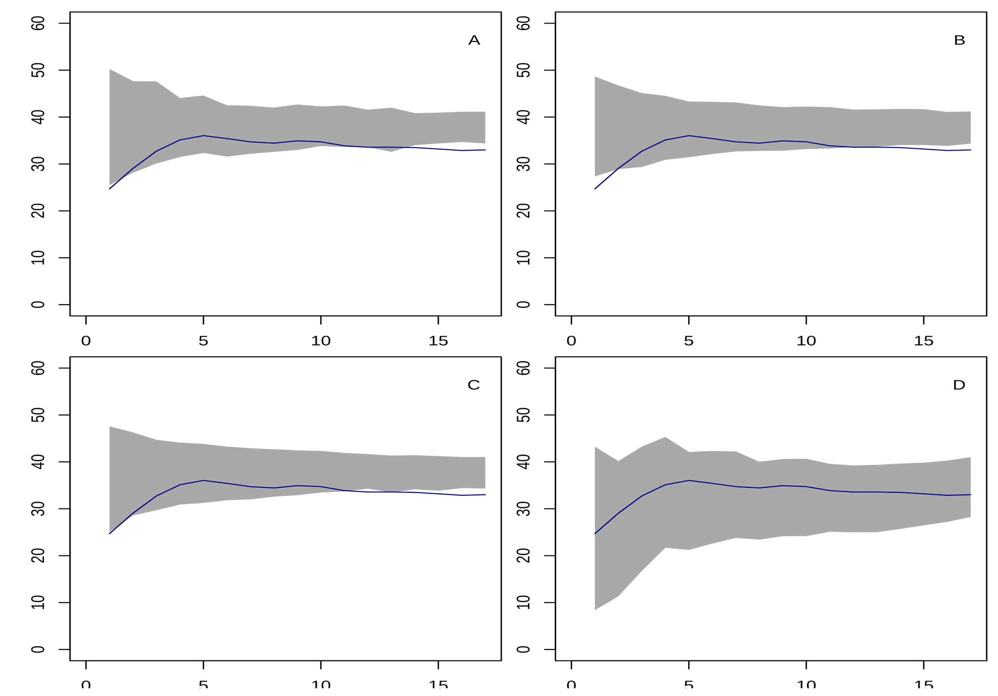
EMPIRICAL POOS

Alternative methods:



A. R. Bayes, B. Exp-smooth, C. NNAR, D. ARX

Our method:



A. 0 covar., B. 168 trigo., C. 4 more, D. all in.

EMPIRICAL QUANTILE CONSISTENCY FOR LASSO

Let $\hat{Q}_n(u)$ be the u-th empirical quantile of $(\hat{e}_i + \ldots + \hat{e}_{i+m-1})/H_m; i : m, .., n$ with $H_m = \sqrt{m}$ for light-tail. Assume (SRD) holds, $|X|_2 = (np)^{1/2}$, RE(s,κ) in Bickel et al. (2009) holds with constant $\kappa = \kappa(s, 3)$, where $s = \|\beta\|_0$ and $r = \lambda/2 = \max\{A\sqrt{n^{-1}\log p}\|e.\|_{2,\alpha}, B\|e.\|_{q,\alpha}|X|_q n^{-1+\min\{0,1/2-1/q-\alpha\}}\}$. Also

$$(\text{for light tails, i.e. } q \geq 2) s = o\left(\frac{m}{r^2 n}\right) \qquad (\text{for heavy tails, i.e. } 1 < q \leq 2), s = o\left(\frac{H_m^2 |l(n)|^2}{r^2 n^{2\gamma-1}}\right),$$

then $\hat{Q}_n(u)$ is a consistent estimator of $\bar{Q}_n(u)$, the true quantile of $(e_i + \ldots + e_{i+m-1})/H_m; i = m, \ldots, n$

Stochastic X: Results hold with $\mathbf{x}_i = G_x(\epsilon_i^x, \epsilon_{i-1}^x, \ldots), \delta^{\mathbf{x}}_{k,q} = \|\mathbf{x}_i - \mathbf{x}^*_{k,i}\|_q, \delta^{xe}_{k,q} = \max_{j \leq p}\|x_{lj}e_l - x^*_{k,lj}e^*_{k,l}\|_q$

REFERENCES/CONTACT

- Chudy, Karmakar, Wu (2020): Long-term prediction intervals of economic time series. *Empirical Economics*, 58 (1), 191-222
- Karmakar, Chudy,Wu (2021): Prediction intervals for high-dimensional regression. *To appear at Journal of Time-series Analysis*

Contact/Find me at

- sayarkarmakar@ufl.edu
- sayarkarmakar@github.edu