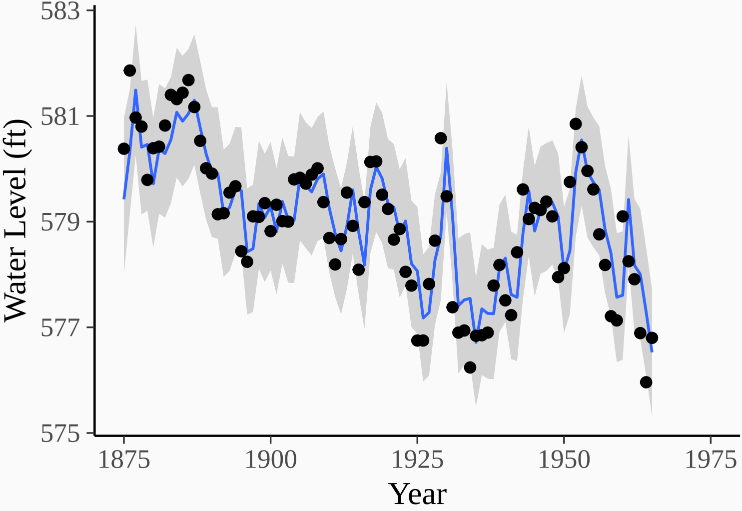
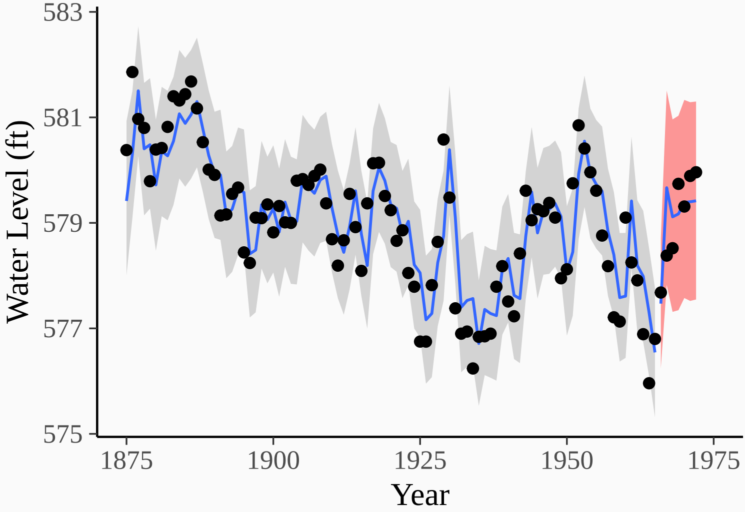
Approximate leave-future-out cross-validation for Bayesian time series models

Paul Bürkner, Jonah Gabry, Aki Vehtari

Water Level of Lake Huron



Water Level of Lake Huron: Predictions



Leave-Future-Out Cross-Validation (LFO-CV)

Perform M-step-ahead predictions (M-SAP) at observation i

$$p(y_{i+1}, \ldots, y_{i+M} \,|\, y_1, ..., y_i) =: p(y_{i+1:M} \,|\, y_{1:i})$$

Estimate expected M-SAP performance via LFO-CV

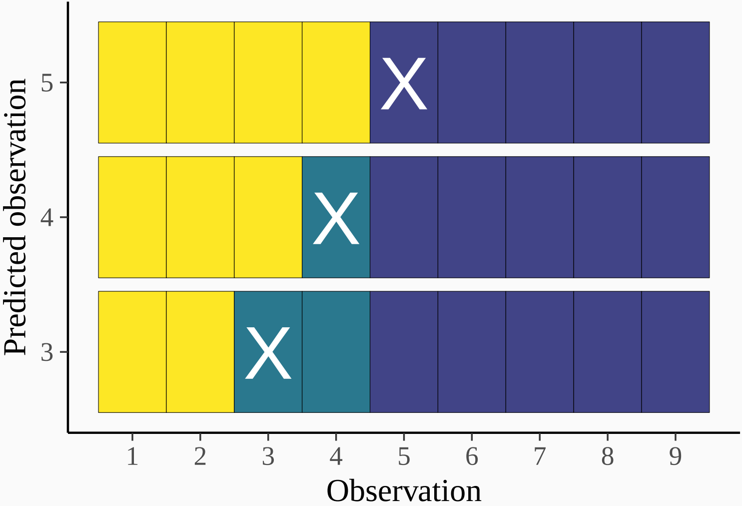
$$\mathrm{ELPD}_{\mathrm{LFO}} = \sum_{i=L}^{N-M} \log p(y_{i+1:M} \,|\, y_{1:i})$$

This requires fitting a separate model for each i

$$p(y_{i+1:M} \,|\, y_{1:i}) = \int p(y_{i+1:M} \,|\, y_{1:i}, \theta) \, p(\theta \,|\, y_{1:i}) \, \mathrm{d}\theta$$

Approximate M-Step-Ahead Predictions

We are moving **backwards** in time!



Pareto Smoothg Importance Sampling (PSIS) for LFO-CV

PSIS approximation of M-SAP:

$$p(y_{i+1:M} \mid y_{1:i}) \approx \frac{\sum_{s=1}^{S} w_i^{(s)} \, p(y_{i+1:M} \mid y_{1:i}, \theta^{(s)})}{\sum_{s=1}^{S} w_i^{(s)}}$$

Let's call J_i the index set of observations included in the approximating model but **not** in the target model

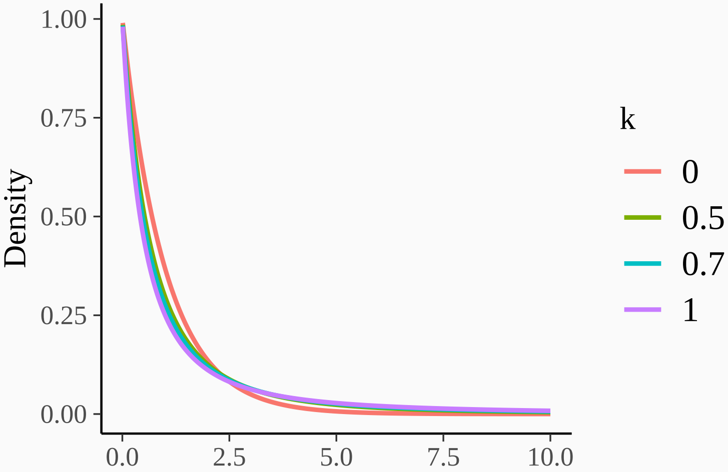
For observation i and posterior sample s we compute the importance ratio as

$$r_i^{(s)} = \frac{1}{\prod_{j \in J_i} p(y_j \mid \theta^{(s)})}$$

Stabilize $r_i^{(s)}$ via Pareto smoothing to obtain weights $w_i^{(s)}$

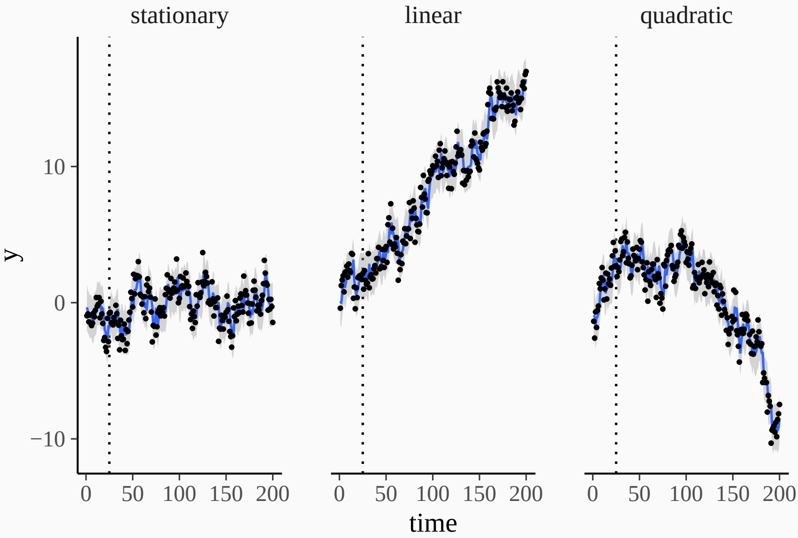
At what point do we have to refit the model?

The Generalized Pareto Distribution

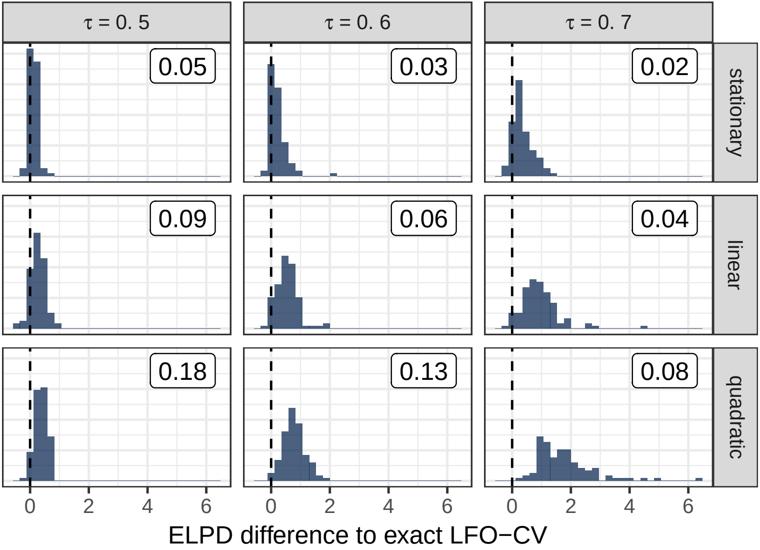


Refit the model if k exceeds a given threshold τ

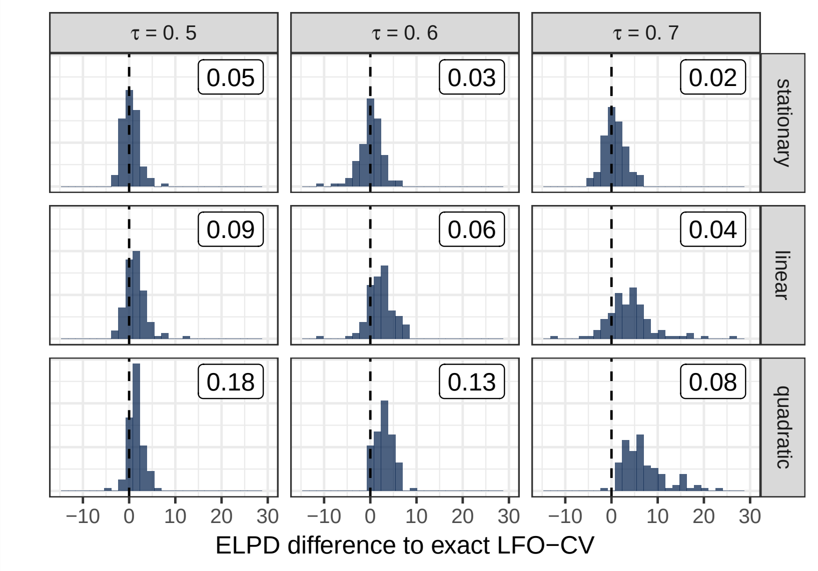
Simulation Conditions



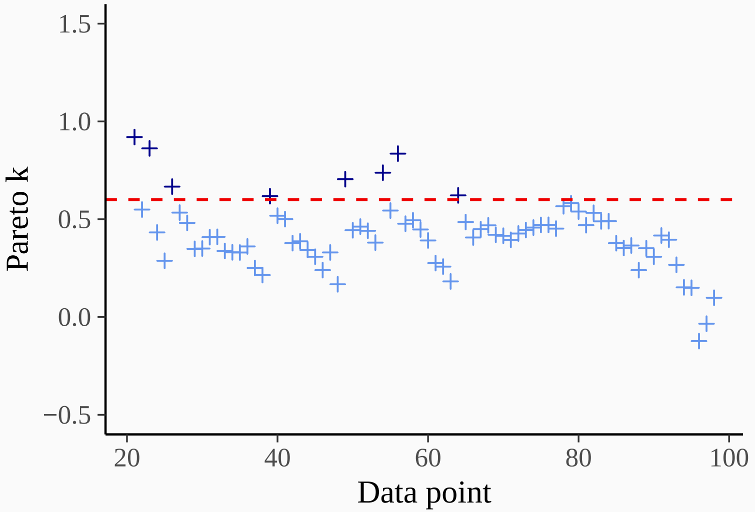
Simulation Results: ELPD of 1-SAP



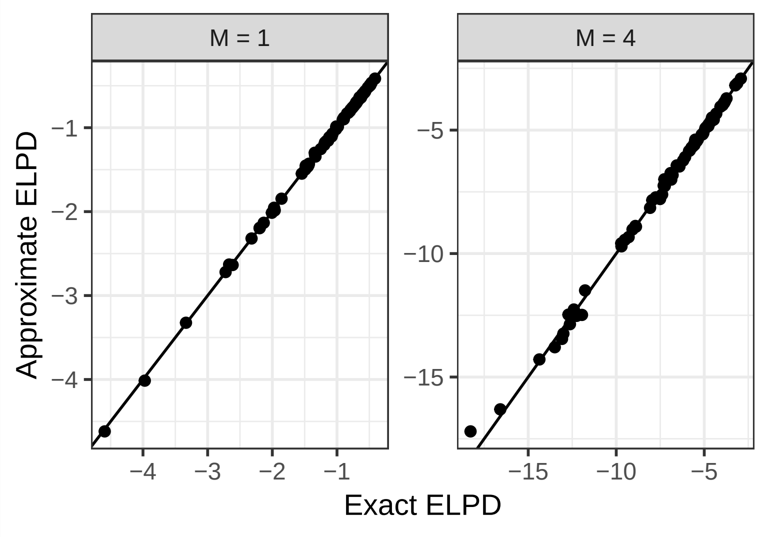
Simulation Results: ELPD of 4-SAP



Lake Huron Model: Pareto k Estimates



Lake Huron Model: ELPD Estimates



Conclusion

- CV has to respect the model's prediction task
- LFO-CV seems reasonable for time series models
- We can approximate LFO-CV via PSIS
- PSIS-LFO-CV provides a close approximation to exact LFO-CV
- PSIS-LFO-CV improves speed by an order of magnitude

Resources:

- Preprint: https://arxiv.org/abs/1902.06281
- GitHub: https://github.com/paul-buerkner/LFO-CV-paper
- Email: paul.buerkner@gmail.com

Thank You!

Importance Sampling

All we care about are expectations (over f):

$$\mathbb{E}_f[h(\theta)] = \int h(\theta) f(\theta)\, d\,\theta$$

Switch the distribution (from f to g) over which to integrate:

$$\mathbb{E}_f[h(\theta)] = \frac{\int h(\theta) r(\theta) g(\theta)\, d\,\theta}{\int r(\theta) g(\theta)\, d\,\theta}$$

with importance ratios

$$r(\theta) = \frac{f(\theta)}{g(\theta)}$$

Pareto Smoothed Importance Sampling (PSIS)

Suppose we can obtain samples $\theta^{(s)}$ from g and compute importance ratios $r(\theta^{(s)}) =: r^{(s)}$. Then we can approximate

$$\mathbb{E}_f[h(\theta)] \approx \frac{\sum_{s=1}^{S} r^{(s)} \, h(\theta^{(s)})}{\sum_{s=1}^{S} r^{(s)}}$$

Problem: The importance ratios $r^{(s)}$ tend to be highly unstable

Solution: Stabilize $r^{(s)}$ by applying Pareto Smoothing

- PSIS weights $w^{(s)}$ that replace $r^{(s)}$
- Diagnose accuracy via the Pareto shape parameter k

Approximate Block M-Step-Ahead Predictions