Efficient Accounting for Estimation Uncertainty in Coherent Forecasting of Count Processes

HELMUT SCHMIDT UNIVERSITÄT
Universität der Bundeswehr Hamburg

MATH STAT

C.H. Weiß, A. Homburg, G. Frahm

Department of Mathematics & Statistics,
Helmut Schmidt University, Hamburg

L.C. Alwan

Sheldon B. Lubar School of Business,
University of Wisconsin-Milwaukee

R. Göb

Department of Statistics,
University of Würzburg

HELMUT SCHMIDT UNIVERSITÄT
Universität der Bundeswehr Hamburg

MATH STAT

Coherent Forecasting of Count Processes

Introduction

Let $x_1, \ldots, x_T$ be **count time series**,

i. e., x_t from $\mathbb{N}_0 = \{0, 1, \ldots\}$, and $T \in \mathbb{N} = \{1, 2, \ldots\}$.

Underlying **count process** $(X_t)_{t\in\mathbb{Z}=\{\ldots,-1,0,1,\ldots\}}$. (Weiß, 2018)

Coherent forecasting: computed forecasts of count process should be count values themselves. (Freeland & McCabe, 2004)

Achieved for forecast horizon $h \in \mathbb{N}$ by first deriving full h-step-ahead conditional distribution of X_{T+h} (given the past $x_T, \ldots, x_1$), and by then computing $(\ldots)$

Coherent forecasting: (...)

- median (mode) of $X_{T+h}\,\big|\,x_T,\ldots,x_1$ as **central point forecast** (PF);
- extreme (upper) quantile of $X_{T+h}\,\big|\,x_T,\ldots,x_1$ as **non-central PF**; or
- finite subset of $\mathbb{N}_0$ satisfying coverage requirement as discrete **prediction interval** (PI) for $X_{T+h}\,\big|\,x_T,\ldots,x_1$;

see Homburg et al. (2019, 2020) for details.

Some authors (e. g., Kolassa, 2016) use full predictive probability mass function (PMF) as forecast value: **PMF forecasts**.

In practice, true predictive distribution of $X_{T+h}\,\big|\,x_T,\ldots,x_1$ unknown, so forecasts relying on fitted model.

But computed forecasts (using estimated model) might deviate from true forecasts (from true model in same situation).

Comprehensive study by Homburg et al. (2019, 2020): especially non-central PFs and PIs suffer a lot from estimation.

Since impossible to suppress estimation uncertainty in practice:
Is it possible to account for it in appropriate way?

HELMUT SCHMIDT UNIVERSITÄT
Universität der Bundeswehr Hamburg

MATH STAT

Resampling of Coherent Forecasts

Proposed Solution

Inspired by proposals of Freeland & McCabe (2004), Jung & Tremayne (2006), we suggest **resampling approach** to approximate distribution of coherent forecast value (random due to randomness of sample $X_1, \ldots, X_T$).

Instead of single forecast value or forecast interval, **ensemble of forecast values** (resulting from resampling) presented to practitioner in feasible way, to express effect of estimation uncertainty.

Proposed solution:

1. Fit model to data $x_1, \ldots, x_T$, compute parameter estimate $\widehat{\boldsymbol{\theta}}$.
2. Resampling approach to assess variability of $\widehat{\boldsymbol{\theta}}$, by
   (a) parametric bootstrap scheme, or
   (b) asymptotic resampling scheme; see details below.
3. Use resampled fitted models for coherent forecasts (PFs, PIs, PMFs) $\Rightarrow$ discrete frequency distribution of forecast values.
4. Use whole forecast's distribution to incorporate parameter uncertainty ("**ensemble forecasting**", see Palmer (2002)).

Step 2, resampling approach for $\widehat{\boldsymbol{\theta}}$:

Option (a): **parametric bootstrap** scheme.

Generate B times count time series from fitted model,

estimate model parameters $\boldsymbol{\theta}$ again for each replicate:

B bootstrap replicates $\widehat{\boldsymbol{\theta}}_1^*, \ldots, \widehat{\boldsymbol{\theta}}_B^*$ of $\widehat{\boldsymbol{\theta}}$.

Example: INAR bootstrap of Jentsch & Weiß (2019).

Disadvantages:

- individual bootstrap code for each type of DGP;
- computationally demanding, because estimation procedure for each bootstrap replication anew.

Step 2, resampling approach for $\widehat{\boldsymbol{\theta}}$:

Option (b): **asymptotic resampling** scheme.

For common count DGPs, ML estimator $\widehat{\boldsymbol{\theta}}_{\mathsf{ML}}$ behaves uniquely,

namely $\sqrt{T}\,(\widehat{\boldsymbol{\theta}}_{\mathsf{ML}} - \boldsymbol{\theta}) \sim \mathsf{N}\big(\mathbf{0}, \mathbf{I}^{-1}(\boldsymbol{\theta})\big)$.

Mean observed Fisher information $\frac{1}{T}\,\mathbf{J}(\boldsymbol{\theta})$ approximates $\mathbf{I}(\boldsymbol{\theta})$,

where $\mathbf{J}(\boldsymbol{\theta})$ Hessian of log-likelihood function.

We sample replicates $\widehat{\boldsymbol{\theta}}^{\star}_{\mathsf{ML},1}, \ldots, \widehat{\boldsymbol{\theta}}^{\star}_{\mathsf{ML},B}$ from $\mathsf{N}\big(\widehat{\boldsymbol{\theta}}_{\mathsf{ML}},\ \mathbf{J}^{-1}(\widehat{\boldsymbol{\theta}}_{\mathsf{ML}})\big)$,

used to compute corresponding B forecast replicates.

Note: neither required to generate B count time series,

nor to apply estimation to them. Saves lot of computing time.

HELMUT SCHMIDT UNIVERSITÄT
Universität der Bundeswehr Hamburg

MATH STAT

Performance of Resampling Approach

Simulation Study

Performance of Resampling Approach

For different types of DGP and several scenarios per DGP, $M = 1\,000$ count t. s. of various lengths T were simulated.

For each of these Monte–Carlo replicates, $\widehat{\boldsymbol{\theta}}_{\mathrm{ML}}$ computed.
Then, different types of coherent forecasts either using true $\boldsymbol{\theta}$, or estimate $\widehat{\boldsymbol{\theta}}_{\mathrm{ML}}$ $\quad\Rightarrow$ evaluate estimation uncertainty.

Next, for each Monte–Carlo t. s. and corresponding $\widehat{\boldsymbol{\theta}}_{\mathrm{ML}}$, resampling with $B = 500$ replicates, leading to either estimates $\widehat{\boldsymbol{\theta}}^{\star}_{\mathrm{ML},1}, \ldots, \widehat{\boldsymbol{\theta}}^{\star}_{\mathrm{ML},B}$ (if **asymptotic resampling**) or $\widehat{\boldsymbol{\theta}}^{*}_{\mathrm{ML},1}, \ldots, \widehat{\boldsymbol{\theta}}^{*}_{\mathrm{ML},B}$ (if **parametric bootstrap**). Then, **ensemble** of replicated forecast values for Monte–Carlo t. s.

Performance of Resampling Approach

Detailed simulation results in Weiß et al. (2021, open access).

Summary of main findings:

Both resampling approaches nearly same performance in coherent forecasting, but big difference in computing time. While simulations for one scenario of asymptotic resampling only few minutes, between 7–72 hours for parametric bootstrap. Therefore, asymptotic resampling clearly preferable for practice.

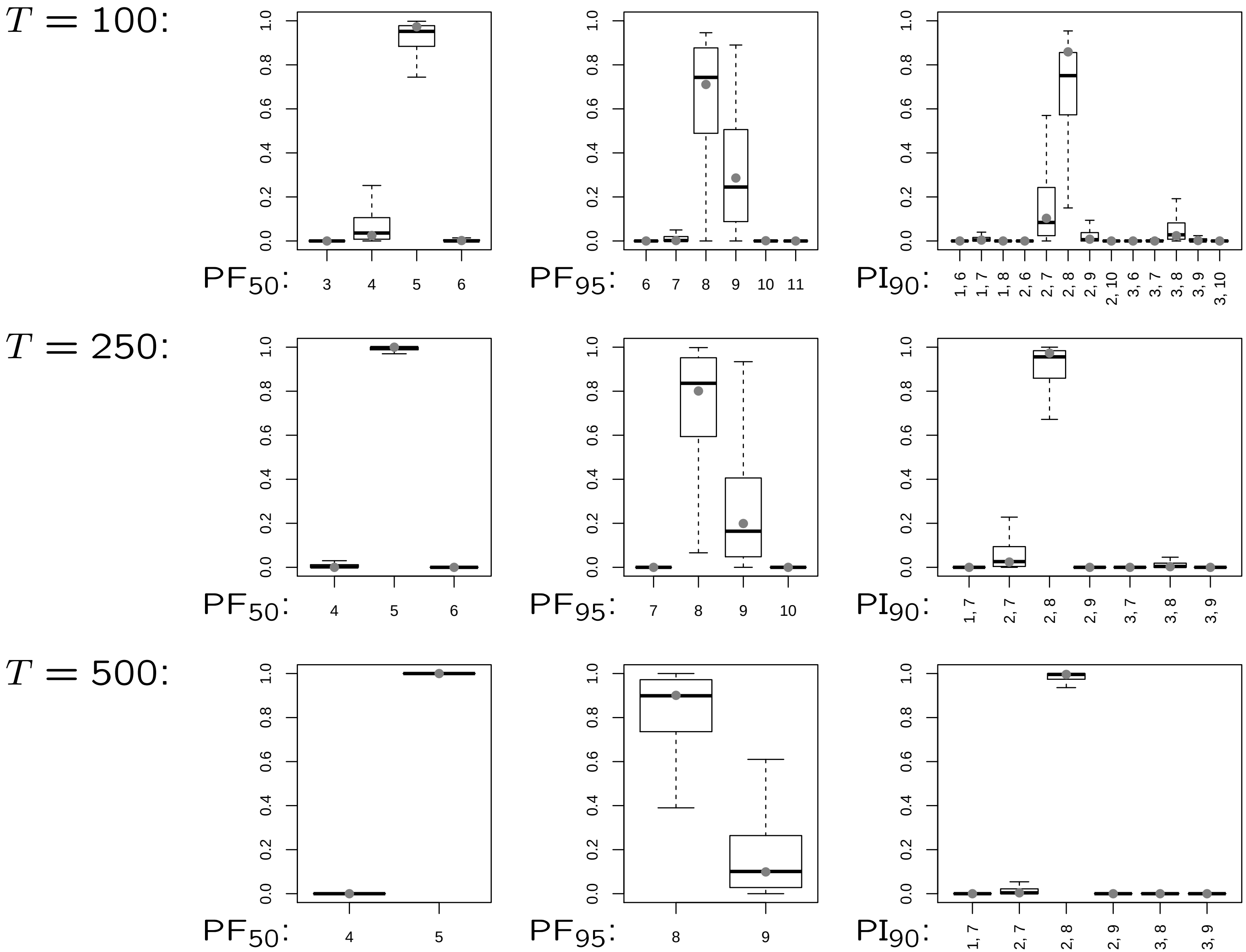
Ensemble forecasting leads to "blurred" forecasts.
But resampled forecasts' distribution ("blurredness") matches true forecast's distribution (estimation uncertainty) quite well.

Example:

Poi-INAR(1) with $\mu = 5$ and $\alpha = 0.50$.

True distrib. as gray dots,

boxplots of resampled forecasts.



HELMUT SCHMIDT UNIVERSITÄT
Universität der Bundeswehr Hamburg

MATH STAT

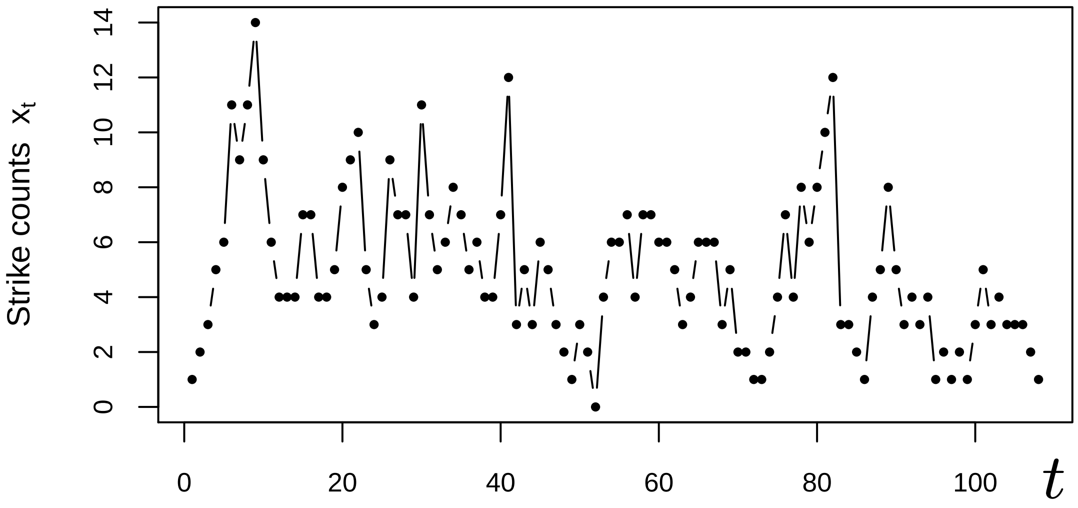
Application of Resampling Approach

Real-Data Example

Application of Resampling Approach

Real-world example: monthly numbers of "work stoppages" (strikes and lock-outs) of $\geq 1\,000$ workers in the U.S.

$T = 108$ counts from period 1994–2002 used for model fitting:



Weiß (2018) uses Poi-INARCH(1) model for learning sample.

ML estimates (s. e.)
$\hat{\mu}_{\text{ML}} \approx 4.981$ (0.593),
$\hat{\alpha}_{\text{ML}} \approx 0.636$ (0.081).

Out-of-sample forecasting for counts $t = 109, \ldots, 120$ in 2003.

Application of Resampling Approach

Numerical optimization routine `optim` in R leads to approximate bivariate normal distribution for $(\widehat{\mu}_{\mathsf{ML}}, \widehat{\alpha}_{\mathsf{ML}})$, namely

$$\mathsf{N}\left(\begin{pmatrix} 4.981 \\ 0.636 \end{pmatrix}, \begin{pmatrix} 0.352 & 0.016 \\ 0.016 & 0.007 \end{pmatrix}\right).$$

Using `mvrnorm` in R's `MASS` package, we simulate $B = 500$ replicates $(\widehat{\mu}^{\star}_{\mathsf{ML},b}, \widehat{\alpha}^{\star}_{\mathsf{ML},b})$ with $b = 1, \ldots, B$.

Then, forecasts for x_t with $t = 109, \ldots, 120$
once based on sample fit $(\widehat{\mu}_{\mathsf{ML}}, \widehat{\alpha}_{\mathsf{ML}})$,
and $B = 500$ times based on replicates $(\widehat{\mu}^{\star}_{\mathsf{ML},b}, \widehat{\alpha}^{\star}_{\mathsf{ML},b})$.

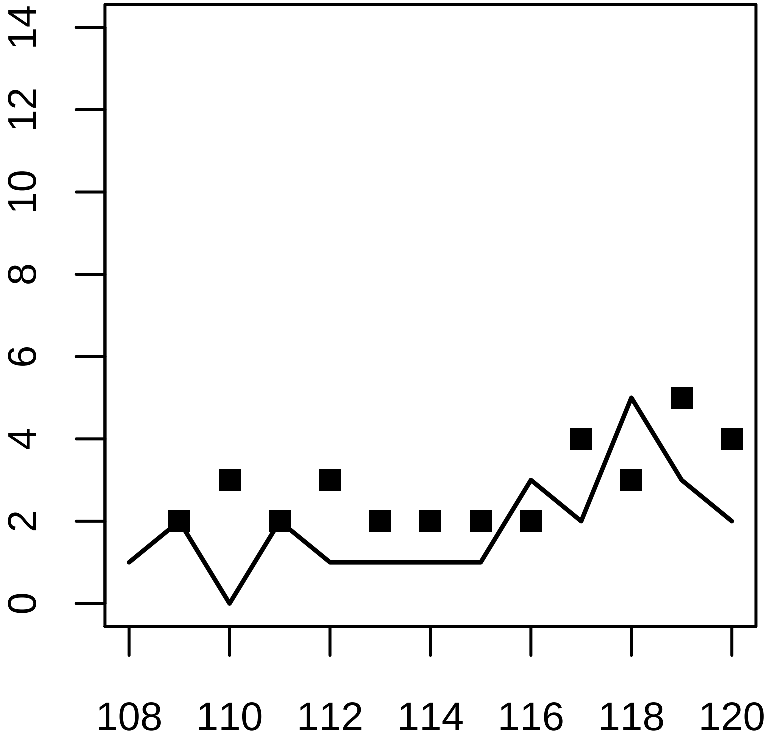
Example: forecasts for $x_{109} = 2$, given $x_{108} = 1$, are (absolute frequencies in parentheses)

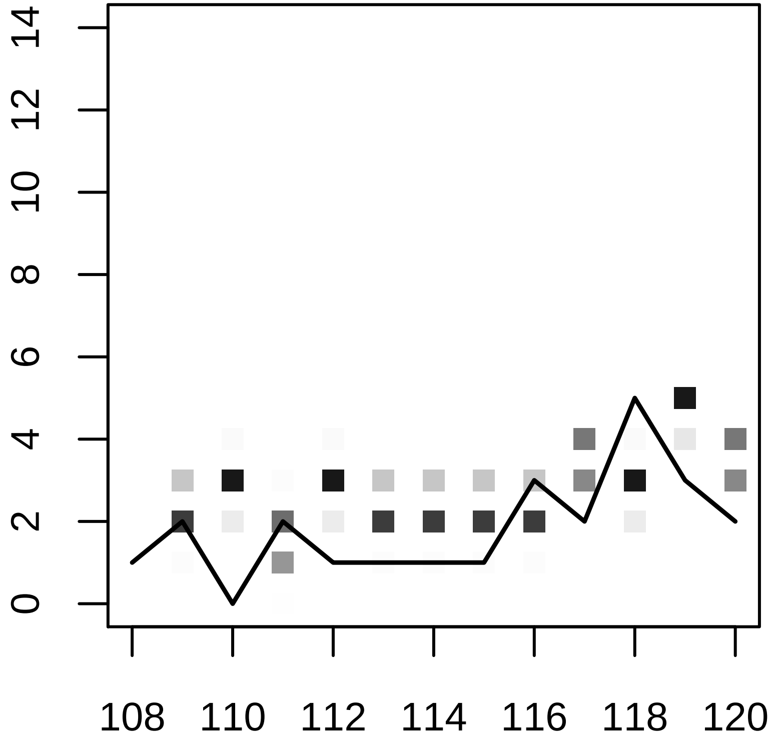
- PF_{50}: 2 if using fitted model, but 1 (6), 2 (382), 3 (112) under resampling;
- PF_{95}: 5 if using fitted model, but 4 (44), 5 (311), 6 (141), 7 (4) under resampling;
- PI_{90}: $\{0,\ldots,5\}$ if using fitted model, but $\{0,\ldots,3\}$ (12), $\{0,\ldots,4\}$ (254), $\{0,\ldots,5\}$ (207), $\{1,\ldots,6\}$ (27) under resampling.

So instead of unique forecast value, we get 3–4 values together with frequencies, expressing effect of estimation uncertainty.

Proposed **visual solution** for central point forecasting:

PF_{50} using fitted model resampling



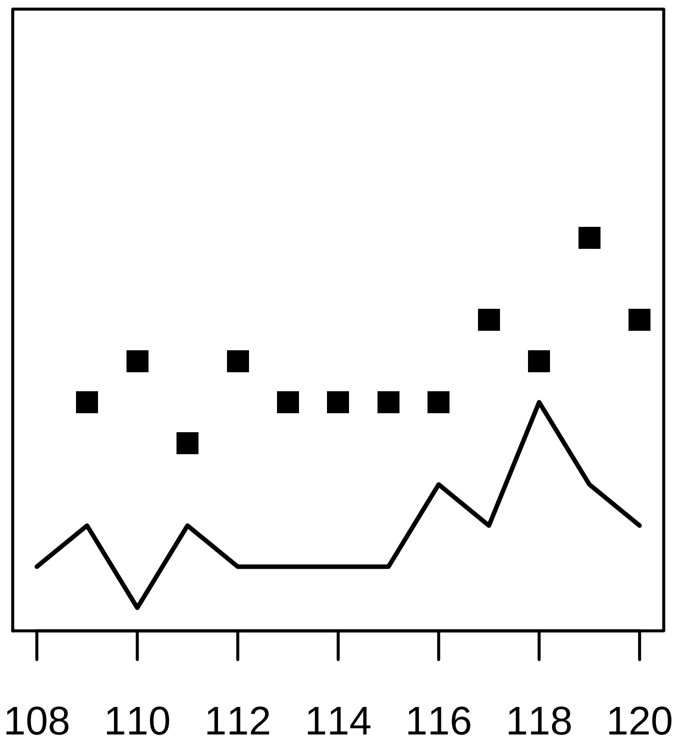


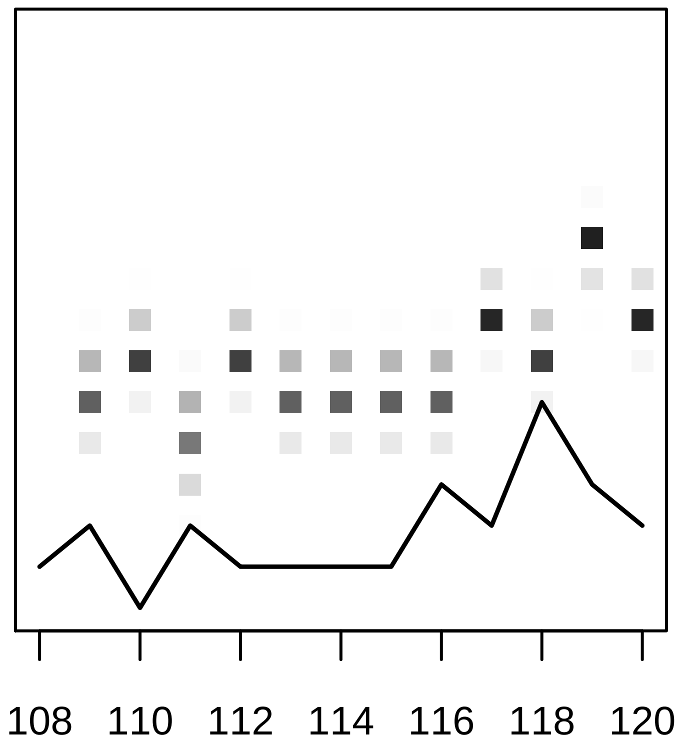
viewer believes
in exact inference

blurred structure reveals
estimation uncertainty

Proposed **visual solution** for non-central point forecasting:

PF_{95} using fitted model resampling



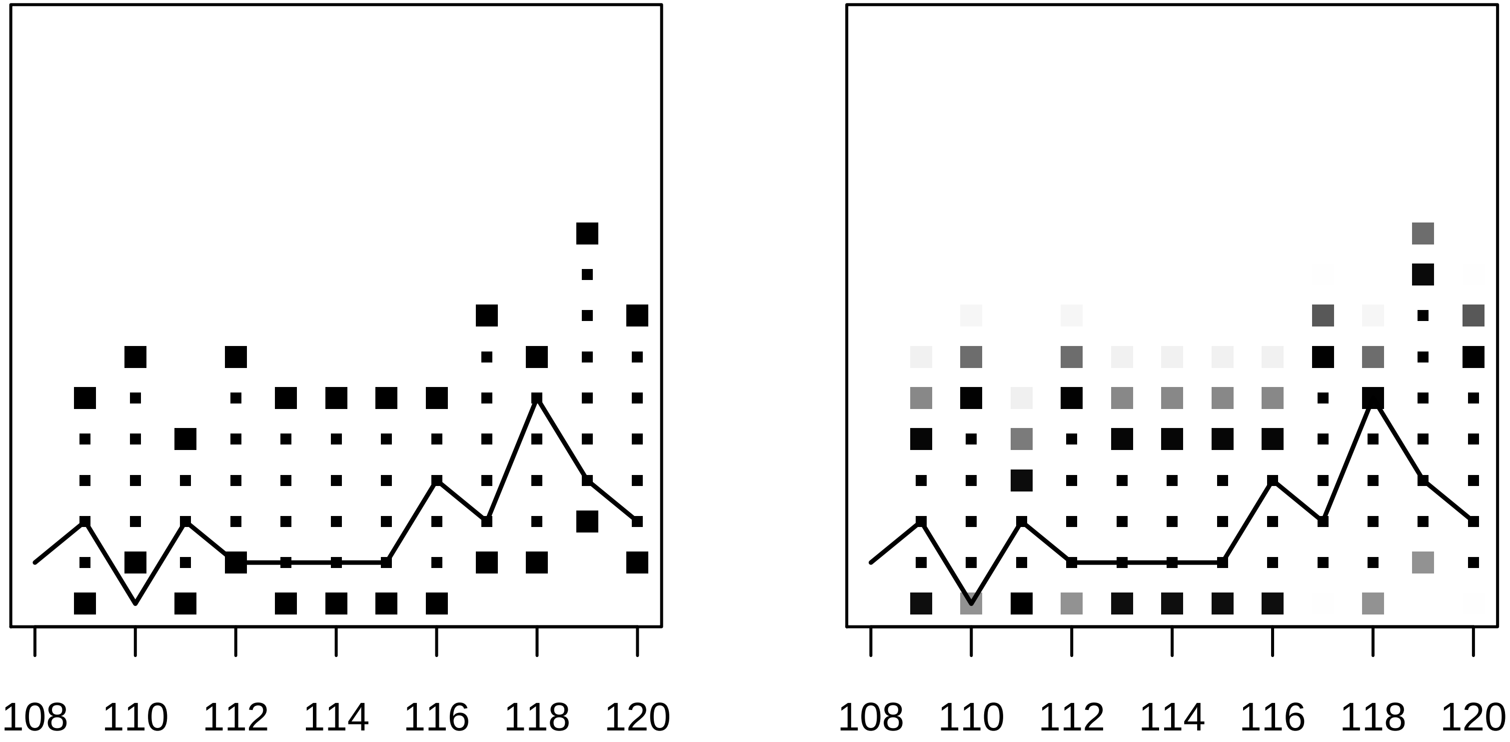


viewer believes
in exact inference

blurred structure reveals
estimation uncertainty

Proposed **visual solution** for interval forecasting:

PI_{90} using fitted model resampling



viewer believes in exact inference

blurred structure reveals estimation uncertainty

- Account for estimation uncertainty in coherent forecasting by resampling approach.
- Asymptotic resampling performs equally well as parametric bootstrap, but computationally much more efficient.
- Visual representation of resampled ensemble of forecasts using gray levels to simplify interpretation by practitioner.
- **Future research:**
  Types of coherent forecasting for ordinal processes, resampling approaches to account for estimation uncertainty.

Thank You

for Your Interest!

Christian H. Weiß

Department of Mathematics & Statistics

Helmut Schmidt University, Hamburg

weissc@hsu-hh.de

Literature

Weiß et al. (2021) Efficient accounting for estimation uncertainty in coherent forecasting of count processes.
***Journal of Applied Statistics*, in press.** (open access)

Freeland & McCabe (2004) Forec. discrete valued ... *IJF* **20**, 427–434.

Jentsch & Leucht (2016) Bootstrapping sample ... *AISM* **68**, 491–539.

Jentsch & Weiß (2019) Bootstrapping INAR ... *Bernoulli* **25**, 2359–2408.

Homburg et al. (2019) Evaluating approx. point ... *Econometrics* **7**, 30.

Homburg et al. (2020) A performance analysis of prediction ... *JF*, in press.

Jung & Tremayne (2006) Coherent forec. in integer ... *IJF* **22**, 223–238.

Kolassa (2016) Evaluating predictive count data ... *IJF* **32**, 788–803.

Palmer (2002) The economic value of ensemble ... *QJRMS* **128**, 747–774.

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

This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Projektnummer 394832307.