Conformal Prediction with Temporal Quantile Adjustments

Zhen Lin[1] Shubhendu Trivedi Jimeng Sun[1]

[1] University of Illinois at Urbana-Champaign

ILLINOIS

SunLab

Task: Construct prediction intervals for time series data with a cross-section

Applications: Healthcare; econometrics; science.

Notation:

- $\mathbf{S}_i = [Z_{i,1}\dots, Z_{i,t}, \dots, Z_{i,T}]$: The i-th time series
- $Z_{i,t} = (X_{i,t}, Y_{i,t})$: random variable (input and response) for the i-th time series at time t
- $\hat{C}_{i,t}$: Prediction interval for $Y_{i,t}$.

Cross-sectional and Longitudinal Validity

Definition

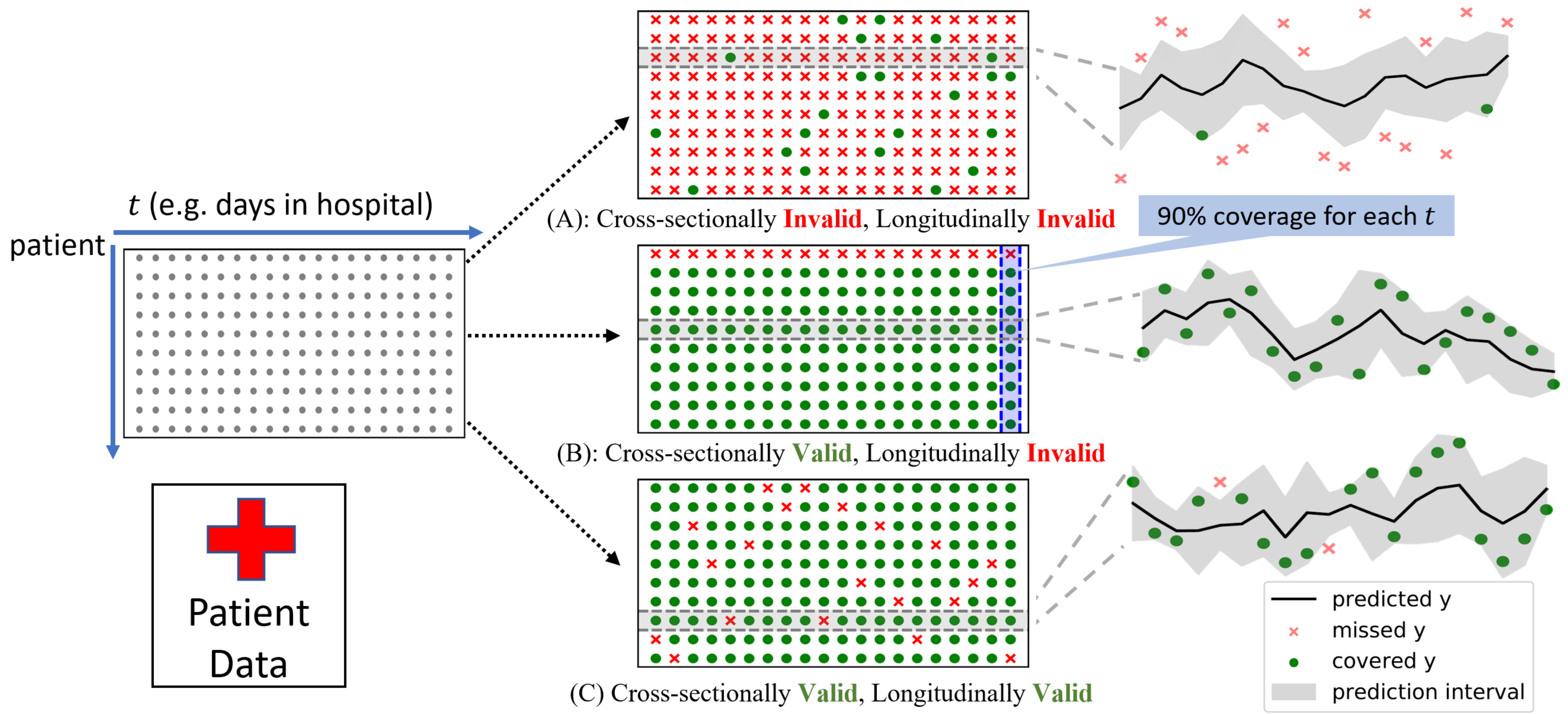
Prediction interval $\hat{C}_{\cdot,\cdot}$ is $1-\alpha$ cross-sectionally valid if, for any t,

$$\mathbb{P}_{\mathbf{S}_{N+1}}\{Y_{N+1,t} \in \hat{C}_{N+1,t}\} \geq 1-\alpha. \tag{1}$$

Definition

Prediction interval $\hat{C}_{\cdot,\cdot}$ is $1-\alpha$ longitudinally valid if for almost every time-series $\mathbf{S}_{N+1} \sim \mathcal{P}_S$ there exists a T_0 such that:

$$t > T_0 \implies \mathbb{P}_{Y_{N+1,t}|\mathbf{S}_{N+1,1:t-1}}\{Y_{N+1,t} \in \hat{C}_{N+1,t}\} \geq 1-\alpha. \tag{2}$$



Temporal Quantile Adjustments (TQA)

Preliminary: Split Conformal

Treading $\{\mathbf{S}_i\}_{i=1}^N$ as the calibration set,

$$\hat{C}^{split}_{N+1,t+1} := [\hat{y}-\hat{v}, \hat{y}+\hat{v}] \text{ where } \hat{v} := Q\Big(1-\alpha; \{|y_{i,t+1}-\hat{y}_{i,t+1}|\}_{i=1}^N \cup \{\infty\}\Big) \tag{3}$$

Here, $Q(\beta; A)$ means the β-quantile for the set A.

Validity: Assuming exchangeability, split conformal is cross-sectionally valid.

Limitation: If we already have evidence that $\mathbf{S}_{N+1}$ is "abnormal", we could adapt to this observation/belief.

Solution: In TQA, we replace α with a dynamic $a_{i,t} = \alpha - \hat{\delta}_{i,t}$. This could improve longitudinal coverage while maintaining cross-sectional validity.
Please find all theorems in our paper.

TQA-B: Quantile Budgeting

(i) Quantile Prediction:

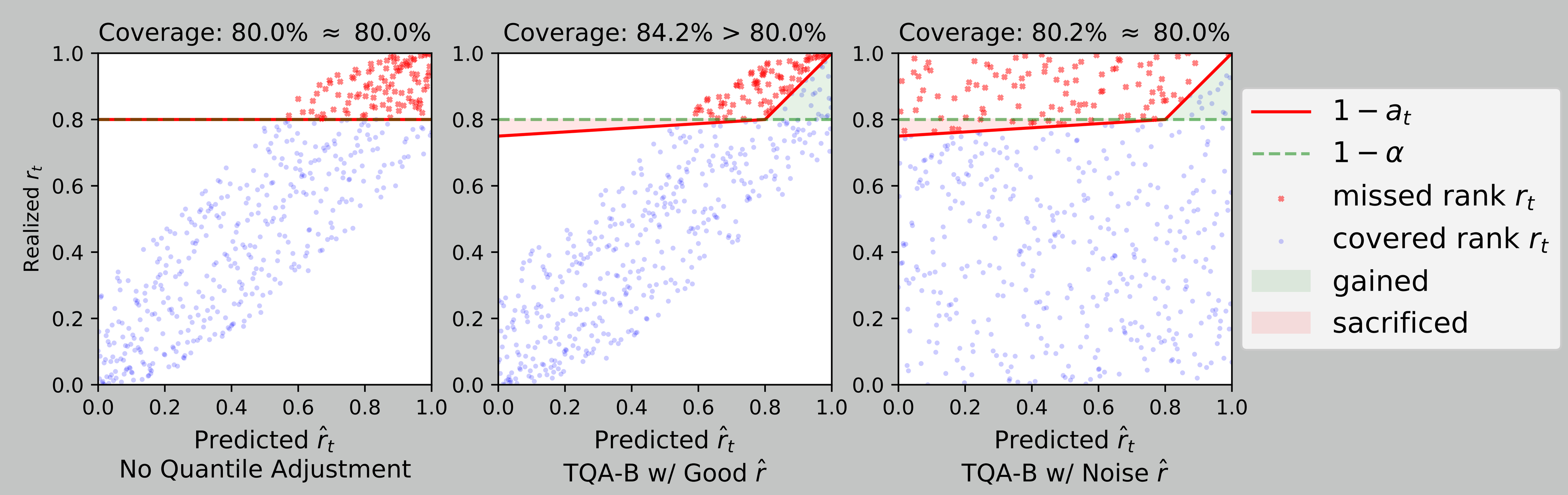
$$\hat{r}^{ms}_{i,t+1} := Q^{-1}(\bar{\epsilon}_{i,t}; \{\bar{\epsilon}_{j,t}\}_{j=1}^{N+1}) \text{ where } \bar{\epsilon}_{i,t} := \sum_{t'=1}^{t} \frac{|y_{i,t'}-\hat{y}_{i,t'}|}{t}\beta^{(t-t')}. \tag{4}$$

(ii) Budgeting:

$$\hat{\delta}^B_{i,t}(r;\alpha) := \begin{cases} C(r-(1-\alpha)) & (r < 1-\alpha) \\ (r-(1-\alpha)) & (r \geq 1-\alpha) \end{cases} \text{ where } C = \frac{(2\alpha N - \lfloor \alpha N \rfloor)(\lfloor \alpha N \rfloor + 1)}{\lceil (1-\alpha)N \rceil ((1-2\alpha)N + 1 + \lfloor \alpha N \rfloor)}. \tag{5}$$

TQA-E: Error-Based Adjustment

$$\hat{\delta}_{t+1} \leftarrow \begin{cases} \hat{\delta}_t + \gamma(err_t - \alpha) & (\hat{\delta}_t \geq \alpha - 1) \\ (1-\gamma)\hat{\delta}_t & (otherwise) \end{cases}. \tag{6}$$



Coverage profiles with hypothetical realized rank r condition on prediction $\hat{r}$, with $\alpha = 0.2$ for readability. ($Y_{i,t} \in \hat{C}_{i,t} \Leftrightarrow r_{i,t} \leq 1 - a_{i,t}$.) As $\hat{r}$ follows a uniform distribution, the proportion of dots below the red line represents the cross-sectional coverage probability. TQA-B generally improves coverage if $\hat{r}$ is correlated with the realized r (middle), and does not lose coverage otherwise (right). "Budgeting" refers to the constraint that sacrificed and gained have equal areas.

Experiments

Average Coverage: Frequency of $Y_{i,t}$ being in $\hat{C}_{i,t}$.
Tail Coverage: Average coverage of the least-covered 10% time series.
Inverse Efficiency: Average PI width divided by the average coverage.

Coverage	TQA-B	TQA-E	CFRNN (Split)	CQRNN	LASplit	QRNN	DPRNN
MIMIC	**91.31±1.32**	**91.19±0.48**	**90.06±1.73**	**90.15±1.24**	**90.33±1.54**	86.90±1.22	46.30±3.84
CLAIM	**91.19±0.49**	**91.56±0.35**	**90.21±0.56**	**90.15±0.68**	**90.20±0.64**	85.90±0.78	24.79±0.85
COVID	**90.79±1.45**	**91.73±0.85**	**90.25±1.69**	**90.08±1.62**	**90.18±1.46**	89.19±1.54	67.51±3.76
EEG	**90.73±1.21**	**90.63±0.75**	**89.92±1.44**	**89.99±1.76**	**89.80±1.15**	87.96±0.82	39.24±1.30
GEFCom	89.58±0.25	**90.94±0.14**	88.61±0.16	89.16±0.17	88.96±0.18	80.40±1.36	89.50±0.73
GEFCom-R	**90.56±0.64**	**90.72±0.45**	**89.92±0.78**	**90.07±0.63**	**89.95±0.72**	85.49±1.08	**91.03±0.76**

Tail Coverage Rate ↑	TQA-B	TQA-E	CFRNN (Split)	CQRNN	LASplit	QRNN	DPRNN
MIMIC	71.59±4.03	**80.68±1.74**	62.22±7.09	68.60±3.84	65.05±6.12	61.80±3.91	17.24±5.38
CLAIM	74.16±1.22	**81.53±0.77**	65.95±1.88	66.45±3.19	68.08±2.44	53.89±3.59	1.65±0.54
COVID	70.01±4.45	**82.39±1.28**	64.41±6.11	66.41±5.99	67.38±4.63	65.16±6.15	36.65±5.63
EEG	70.99±2.18	**79.03±1.22**	64.14±3.42	61.95±4.71	67.13±2.32	57.82±2.78	12.99±1.32
GEFCom	68.96±1.70	**81.77±0.36**	58.49±1.38	61.63±1.56	60.46±1.66	47.56±2.27	67.45±1.69
GEFCom-R	75.28±1.28	**81.80±0.69**	68.76±2.18	71.95±1.66	70.79±2.12	64.99±1.92	71.86±1.75

Inverse Efficiency ↓	TQA-B	TQA-E	CFRNN (Split)	CQRNN	LASplit	QRNN	DPRNN
MIMIC	1.990±0.165	2.382±0.265	1.964±0.170	**1.738±0.145**	2.072±0.223	1.623±0.146	1.258±0.132
CLAIM	3.020±0.045	3.279±0.074	3.003±0.052	**2.902±0.044**	3.009±0.064	2.691±0.035	2.401±0.205
COVID	**0.831±0.032**	1.167±0.337	**0.826±0.034**	0.908±0.091	**0.826±0.037**	0.888±0.096	0.744±0.050
EEG	**1.449±0.025**	1.749±0.125	**1.445±0.031**	1.586±0.052	**1.448±0.025**	1.497±0.042	1.061±0.027
GEFCom	0.238±0.005	0.280±0.013	**0.235±0.005**	0.242±0.005	0.238±0.005	0.211±0.005	0.636±0.009
GEFCom-R	**0.200±0.004**	0.222±0.010	**0.198±0.004**	0.207±0.004	0.201±0.004	0.193±0.004	0.590±0.009

This work was supported by NSF award SCH-2205289, DMS-1439786, SCH-2014438, IIS-1838042, NIH award R01 1R01NS107291-01.

zhenlin4@illinois.edu