预测科学研究中心
The Centre for Forecasting Science

Online Conformal Inference for Multi-step Time Series Forecasting

Xiaoqian Wang & Rob J Hyndman

11 December 2025

Outline

Uncertainty Estimation in Forecasting

Model-based Approaches

- Examples: ARIMA/state-space models
- Assumption: Explicit parametric error distribution
- Limitations: Sensitive to misspecification, coverage invalid under non-stationarity

Uncertainty Estimation in Forecasting

Model-based Approaches

- Examples: ARIMA/state-space models
- Assumption: Explicit parametric error distribution
- Limitations: Sensitive to misspecification, coverage invalid under non-stationarity

Resampling & Bayesian Approaches

- Bootstrap: Residual bootstrap, block bootstrap (autocorrelation)
- Bayesian: Posterior predictive distribution
- Limitations: Computationally heavy, depend on prior / resampling scheme, no finite-sample guarantees

Model-dependent / Heuristic Approaches

- Quantile methods: Quantile regression, ML-based quantile models
- Heuristic ML: Ensembles, MC dropout
- Limitations: Model misspecification impacts interval accuracy, no finite-sample guarantees, calibration challenging for time series

Model-dependent / Heuristic Approaches

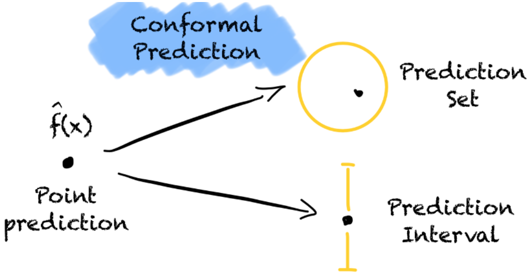
- Quantile methods: Quantile regression, ML-based quantile models
- Heuristic ML: Ensembles, MC dropout
- Limitations: Model misspecification impacts interval accuracy, no finite-sample guarantees, calibration challenging for time series

---

Conformal Prediction

- Distribution-free
- Model-agnostic
- Finite-sample coverage

Conformal Prediction (a.k.a. Conformal Inference)



Conformal prediction (Vovk et al., 2005) is an algorithm for uncertainty quantification that produces statistically valid prediction regions for any underlying point predictor only assuming exchangeability of the data.

(from Wikipedia)

Classical Conformal Prediction Methods

Split / Inductive Conformal Prediction

1. training data set: pre-trained model $\widehat{\mu} : \mathcal{X} \to \mathbb{R}$.
2. calibration / holdout set: nonconformity scores $R_i = |Y_i - \widehat{\mu}(X_i)|$, $i = 1, \ldots, n$.
3. prediction set for a given level α:

$$\widehat{C}_n(X_{n+1}) = \widehat{\mu}(X_{n+1}) \pm \mathrm{Q}_{1-\alpha}\left(\sum_{i=1}^{n} \frac{1}{n+1} \cdot \delta_{R_i} + \frac{1}{n+1} \cdot \delta_{+\infty}\right).$$

Drawback: The loss of accuracy due to sample splitting, sensitive to calibration set, the length of intervals is fixed.

Conformal Prediction Beyond Exchangeability

- Covariate shift (Lei & Candès, 2021; Tibshirani et al., 2019; Yang et al., 2024)
- Distribution drift (Gibbs & Candès, 2021; Zaffran et al., 2022)
- Spatial dependence (Mao et al., 2024)

Conformal Prediction Beyond Exchangeability

- Covariate shift (Lei & Candès, 2021; Tibshirani et al., 2019; Yang et al., 2024)
- Distribution drift (Gibbs & Candès, 2021; Zaffran et al., 2022)
- Spatial dependence (Mao et al., 2024)
- **Temporal dependence**
  - Ensemble batch prediction intervals (EnbPI, Xu & Xie, 2021)
  - Weighted / Locally exchangeable conformal prediction (Barber et al., 2023)
  - Adaptive conformal prediction and its extensions (Bastani et al., 2022; Gibbs & Candès, 2021; Gibbs & Candès, 2024; Zaffran et al., 2022)
  - Quantile tracking (Angelopoulos et al., 2023)

Limitations of Conformal Prediction for Time Series

- Assumption of (local) exchangeability
  - Choice of weights or window size affects reliability
- High-dimensional models
  - Nonconformity score definition becomes nontrivial
- **Multi-step forecasting challenges**
  - Recursive multi-step predictions accumulate errors
  - Temporal dependencies inherent in multi-step forecast errors

Outline

Problem Setup

- A time series $\{y_t\}_{t\geq 1}$ generated by an unknown DGP
- Exogenous predictors $\boldsymbol{x}_t = (x_{1,t}, \dots, x_{p,t})'$
- Data point $\{z_t = (\boldsymbol{x}_t, y_t)\}_{t\geq 1} \subseteq \mathbb{R}^p \times \mathbb{R}$
- Forecasting model $\hat{f}_t$, generating forecasts $\{\hat{y}_{t+h|t}\}_{h\in[H]}$
- Sequential split:
  - a *proper training set*: $\mathcal{D}_{\text{tr}} \subset \{1, \dots, t_r\}$
  - a *calibration set* $\mathcal{D}_{\text{cal}} \subset \{t_r + 1, \dots, t_r + t_c\}$, where $t_c \gg H$
- Nonconformity score:

$$s_{t+h|t} = \mathcal{S}(z_{1:t}, y_{t+h}) := y_{t+h} - \hat{f}_t(z_{1:t}, \boldsymbol{x}_{t+1:h}) = y_{t+h} - \hat{y}_{t+h|t}.$$

Framework: Online Learning with Sequential Splits

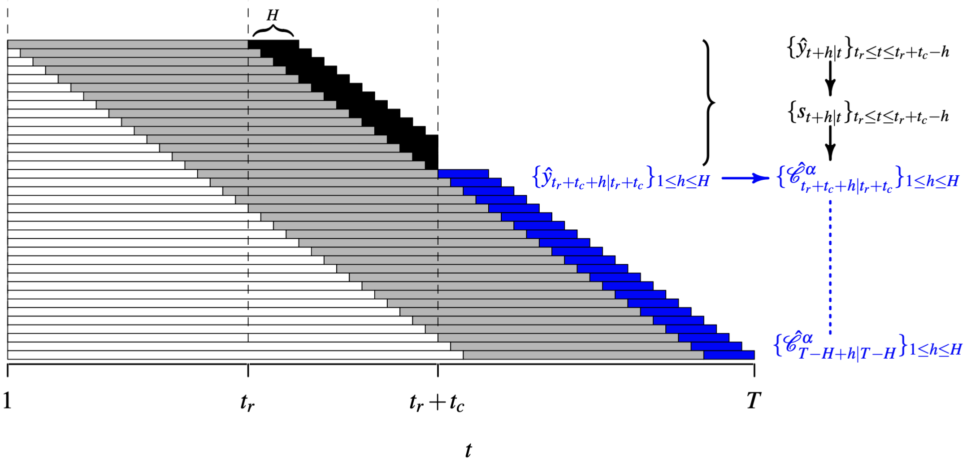


Figure 1: *Diagram of the online learning framework with sequential splits. White: unused data; Gray: training data; Black: forecasts in calibration set; Blue: forecasts in test set.*

Properties of Multi-step Forecast Errors

Assume a time series $\{y_t\}_{t\geq 1}$ generated by a non-stationary AR process:

$$y_t = f_t(y_{(t-d):(t-1)}, \boldsymbol{x}_{(t-k):t}) + \varepsilon_t, \tag{1}$$

where f_t is a nonlinear function, and ε_t is white noise.

- The sequence of model coefficients that parameterizes the function f is restricted to ensure that the stochastic process is locally stationary.

Based on Wold's representation theorem, for a zero-mean covariance-stationary time series, the optimal linear least-squares forecasts have h-step-ahead errors that are at most MA$(h-1)$ process (Diebold, 2024; Harvey et al., 1997).

Proposition 1 (MA($h-1$) process for h-step-ahead optimal forecast errors)

Let $\{y_t\}_{t\geq 1}$ be a time series generated by a general non-stationary autoregressive process as given in Equation (1), and assume that any exogenous predictors are known into the future. Then the forecast errors of optimal h-step-ahead forecasts follow an approximate MA($h-1$) process

$$e_{t+h|t} = \omega_{t+h} + \theta_1\omega_{t+h-1} + \cdots + \theta_{h-1}\omega_{t+1}.$$

where ω_t is white noise.

Proposition 2 (Autocorrelations of multi-step optimal forecast errors)

Let $\{y_t\}_{t\geq 1}$ be a time series generated by a general non-stationary autoregressive process as given in Equation (1), and assume that any exogenous predictors are known into the future. The forecast errors for optimal h-step-ahead forecasts can be approximately expressed as

$$e_{t+h|t} = \omega_{t+h} + \phi_1 e_{t+h-1|t} + \cdots + \phi_p e_{t+h-p|t},$$

where $p = \min\{d, h-1\}$, and ω_t is white noise. Therefore, the optimal h-step-ahead forecast errors are at most serially correlated to lag $(h-1)$.

Autocorrelated Multi-step Conformal Prediction (AcMCP)

For each $h \in [H]$, the iteration of the h-step-ahead quantile estimate is

$$q_{t+h|t} = \underbrace{q_{t+h-1|t-1} + \eta(\mathrm{err}_{t|t-h} - \alpha)}_{\text{quantile tracking}} + \underbrace{r_t\left(\sum_{i=h+1}^{t}(\mathrm{err}_{i|i-h} - \alpha)\right)}_{\text{error integration}} + \underbrace{\tilde{e}_{t+h|t}}_{\text{scorecasting}} .$$

where $\eta > 0$ is a constant learning rate, and r_t is a saturation function that

$$x \geq c \cdot g(t-h) \Rightarrow r_t(x) \geq b, \quad \text{and} \quad x \leq -c \cdot g(t-h) \Rightarrow r_t(x) \leq -b, \qquad (2)$$

for constant $b, c > 0$, and an admissible function g that is sublinear, nonnegative, and nondecreasing.

- $\tilde{e}_{t+h|t}$ is a forecast combination of an MA($h-1$) model and a linear regression model.

Outline

Coverage Guarantees

Proposition 3

Let $\{s_{t+h|t}\}_{t\in\mathbb{N}}$ be any sequence of numbers in $[-b, b]$ for any $h \in [H]$, where $b > 0$, and may be infinite. Assume that r_t is a saturation function obeying Equation (2), for an admissible function g. Then the iteration $q_{t+h|t} = r_t\left(\sum_{i=h+1}^{t}(\mathrm{err}_{i|i-h} - \alpha)\right)$ satisfies

$$\left|\frac{1}{T-h}\sum_{t=h+1}^{T}(\mathrm{err}_{t|t-h} - \alpha)\right| \leq \frac{c \cdot g(T-h) + h}{T-h}, \text{ for any } T \geq h+1.$$

Therefore the prediction intervals obtained by the iteration yield the correct long-run coverage; i.e., $\lim_{T\to\infty} \frac{1}{T-h}\sum_{t=h+1}^{T} \mathrm{err}_{t|t-h} = \alpha$.

Proposition 4

Let $\{s_{t+h|t}\}_{t\in\mathbb{N}}$ be any sequence of numbers in $[-b, b]$ for any $h \in [H]$, where $b > 0$, and may be infinite. Then the iteration $q_{t+h|t} = q_{t+h-1|t-1} + \eta(\mathrm{err}_{t|t-h} - \alpha)$ satisfies

$$\left| \frac{1}{T-h} \sum_{t=h+1}^{T} (\mathrm{err}_{t|t-h} - \alpha) \right| \leq \frac{b + \eta h}{\eta(T-h)}, \quad \text{for any } \eta > 0 \text{ and } T \geq h + 1.$$

Therefore the prediction intervals obtained by the iteration yield the correct long-run coverage; i.e., $\lim_{T\to\infty} \frac{1}{T-h} \sum_{t=h+1}^{T} \mathrm{err}_{t|t-h} = \alpha$.

Proposition 5

Let $\{\hat{q}_{t+h|t}\}_{t\in\mathbb{N}}$ be any sequence of numbers in $[-\frac{b}{2}, \frac{b}{2}]$, and $\{s_{t+h|t}\}_{t\in\mathbb{N}}$ be any sequence of numbers in $[-\frac{b}{2}, \frac{b}{2}]$, for any $h \in [H]$, $b > 0$ and may be infinite. Assume that r_t is a saturation function obeying Equation (2), for an admissible function g.

Then the prediction intervals obtained by the AcMCP iteration yield the correct long-run coverage; i.e., $\lim_{T\to\infty} \frac{1}{T-h} \sum_{t=h+1}^{T} \mathrm{err}_{t|t-h} = \alpha$.

Outline

Simulated Linear Autoregressive Process

Consider a simulated stationary ts generated from an AR(2) process:

$$y_t = 0.8y_{t-1} - 0.5y_{t-2} + \varepsilon_t,$$

where ε_t is white noise with error variance $\sigma^2 = 1$.

- $N = 5000$ data points
- $\mathcal{D}_{\text{tr}}$ and $\mathcal{D}_{\text{cal}}$, each with a length of 500
- $H = 3$
- Fit AR(2) models

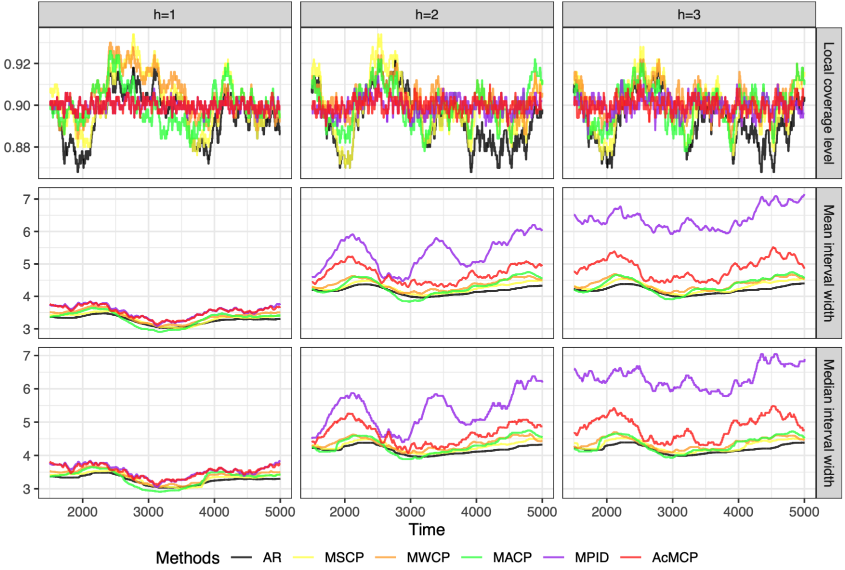


Figure 2: *AR(2) simulation results showing rolling coverage, mean and median interval width for each forecast horizon. The displayed curves are smoothed over a rolling window of size 500. The black dashed line indicates the target level of* $1-\alpha = 0.9$.

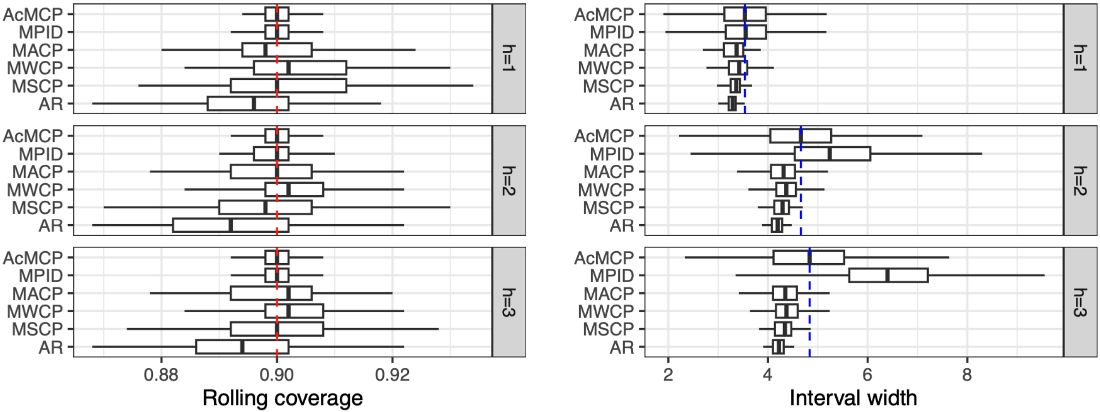


Figure 3: *AR(2) simulation results showing boxplots of the rolling coverage and interval width for each method across different forecast horizons. The red dashed lines show the target coverage level, while the blue dashed lines indicate the median interval width of the AcMCP method.*

Simulated Nonlinear Autoregressive Process

Consider a nonlinear data generation process:

$$y_t = \sin(y_{t-1}) + 0.5\log(y_{t-2} + 1) + 0.1y_{t-1}x_{1,t} + 0.3x_{2,t} + \varepsilon_t,$$

where $x_{1,t}$ and $x_{2,t}$ are uniformly distributed on $[0, 1]$, and ε_t is white noise with error variance $\sigma^2 = 0.1$.

- $N = 2000$ data points
- $\mathcal{D}_{\text{tr}}$ and $\mathcal{D}_{\text{cal}}$, each with a length of 500
- $H = 3$
- Fit feed-forward neural networks with a single hidden layer and lagged inputs

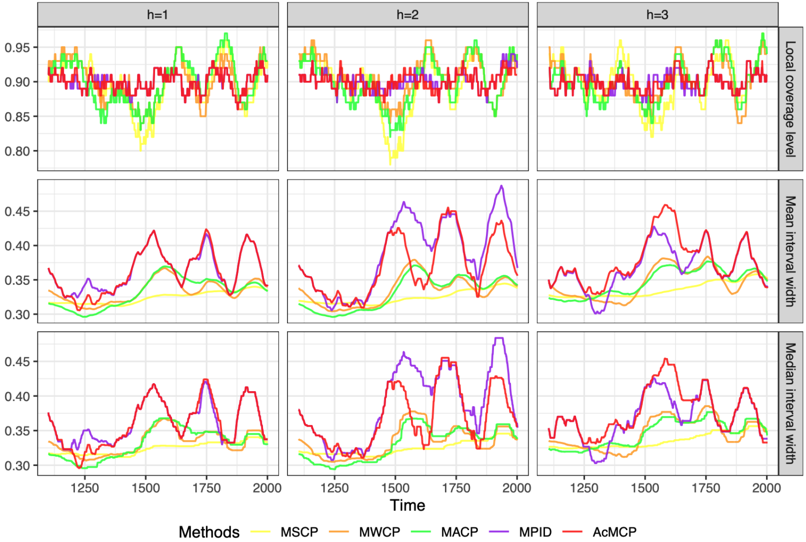


Figure 5: *Nonlinear simulation results showing rolling coverage, mean and median interval width for each forecast horizon. The displayed curves are smoothed over a rolling window of size* 100*. The black dashed line indicates the target level of* $1-\alpha = 0.9$.

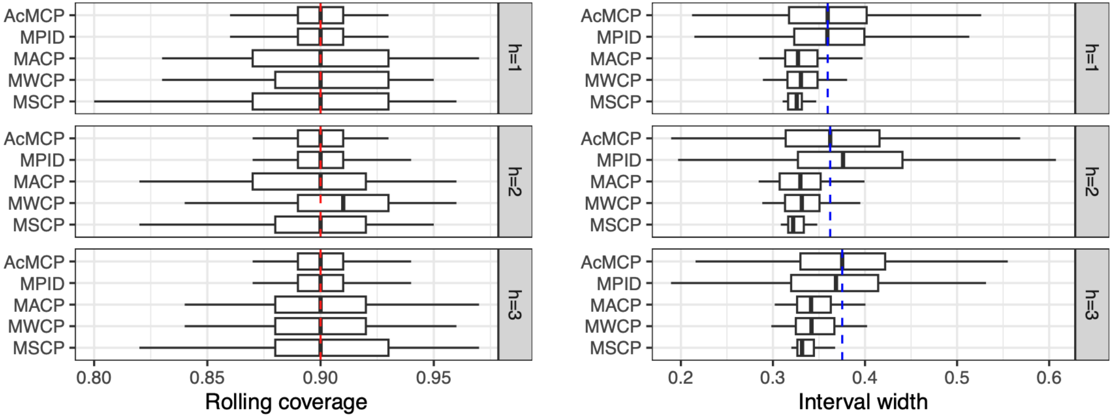


Figure 6: *Nonlinear simulation results showing boxplots of the rolling coverage and interval width for each method across different forecast horizons. The red dashed lines show the target coverage level, while the blue dashed lines indicate the median interval width of the AcMCP method.*

Eating Out Expenditure Data

The data involves **monthly** expenditure on cafes, restaurants and takeaway food services in Victoria from April 1982 up to December 2019.

Forecasting:

- $\mathcal{D}_{tr}$ = 20 years
- $\mathcal{D}_{cal}$ = 5 years
- $\mathcal{D}_{test}$ = 152 months
- H = 12
- Fit ARIMA with logarithmic transformation, ETS, and STL-ETS, and then output their simple average as final point forecasts

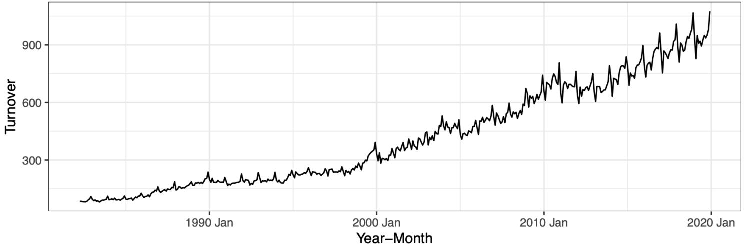


Figure 11: *Monthly expenditure on cafes, restaurants and takeaway food services in Victoria, Australia, from April 1982 to December 2019.*

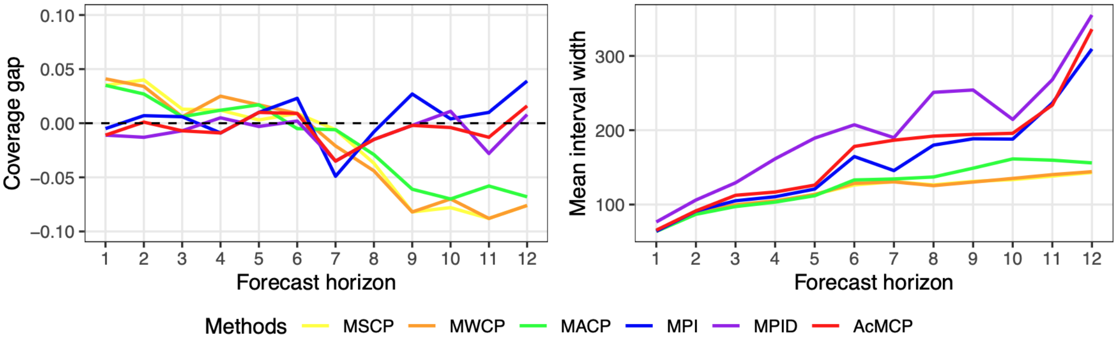


Figure 12: *Eating out expenditure data results showing coverage gap and interval width averaged over the entire test set for each forecast horizon. The black dashed line in the top panel indicates no difference from the 90% target level.*

Outline

Conclusion & Discussion

- A unified notation to formalize the mathematical representation of conformal prediction for time series
- Multi-step forecasting scenarios:
  - Extend simple conformal prediction methods
  - Autocorrelations of multi-step optimal forecast errors
  - Introduce AcMCP accounting for the autocorrelations
  - AcMCP can achieve long-run coverage guarantees without imposing assumptions regarding data distribution shifts

Discussion:

- Limited to ex-post forecasting
- Trade-off between coverage and interval width

More Information

- Wang, X., & Hyndman, R. J. (2025). Online conformal inference for multi-step time series forecasting. *arXiv preprint arXiv:2410.13115*.
- Wang, X., & Hyndman, R. J. (2025). `conformalForecast`: An R package for forecasting time series with conformal prediction. https://CRAN.R-project.org/package=conformalForecast.

Find me at ...

xqnwang.rbind.io

@xqnwang

xiaoqian.wang@amss.ac.cn

References i

Angelopoulos, A. N., Candès, E. J., & Tibshirani, R. J. (2023). Conformal PID control for time series prediction. *Advances in Neural Information Processing Systems*, *36*, 23047–23074.

Barber, R. F., Candès, E. J., Ramdas, A., & Tibshirani, R. J. (2023). Conformal prediction beyond exchangeability. *The Annals of Statistics*, *51*(2), 816–845. https://doi.org/10.1214/23-aos2276

Bastani, O., Gupta, V., Jung, C., Noarov, G., Ramalingam, R., & Roth, A. (2022). Practical adversarial multivalid conformal prediction. *Advances in Neural Information Processing Systems*, *35*, 29362–29373.

Diebold, F. X. (2024). *Forecasting: In economics, business, finance and beyond*. Department of Economics, University of Pennsylvania.

Gibbs, I., & Candès, E. (2021). Adaptive conformal inference under distribution shift. *Advances in Neural Information Processing Systems*, *34*, 1660–1672.

Gibbs, I., & Candès, E. J. (2024). Conformal inference for online prediction with arbitrary distribution shifts. *Journal of Machine Learning Research*, *25*(162), 1–36.

Harvey, D., Leybourne, S., & Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, *13*(2), 281–291.

References iii

Lei, L., & Candès, E. J. (2021). Conformal inference of counterfactuals and individual treatment effects. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *83*(5), 911–938.

Mao, H., Martin, R., & Reich, B. J. (2024). Valid model-free spatial prediction. *Journal of the American Statistical Association*, *119*(546), 904–914.

Tibshirani, R. J., Foygel Barber, R., Candes, E., & Ramdas, A. (2019). Conformal prediction under covariate shift. *Advances in Neural Information Processing Systems*, *32*.

Vovk, V., Gammerman, A., & Shafer, G. (2005). *Algorithmic learning in a random world*. Springer-Verlag. https://doi.org/10.1007/b106715

References iv

Xu, C., & Xie, Y. (2021). Conformal prediction interval for dynamic time-series. *Proceedings of the 38th International Conference on Machine Learning*, *139*, 11559–11569.

Yang, Y., Kuchibhotla, A. K., & Tchetgen Tchetgen, E. (2024). Doubly robust calibration of prediction sets under covariate shift. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *86*(4), 943–965.

Zaffran, M., Féron, O., Goude, Y., Josse, J., & Dieuleveut, A. (2022). Adaptive conformal predictions for time series. *Proceedings of the 39th International Conference on Machine Learning*, *162*, 25834–25866.