Robust Nonparametric Distribution Forecast with Backtest-based Bootstrap and Adaptive Residual Selection

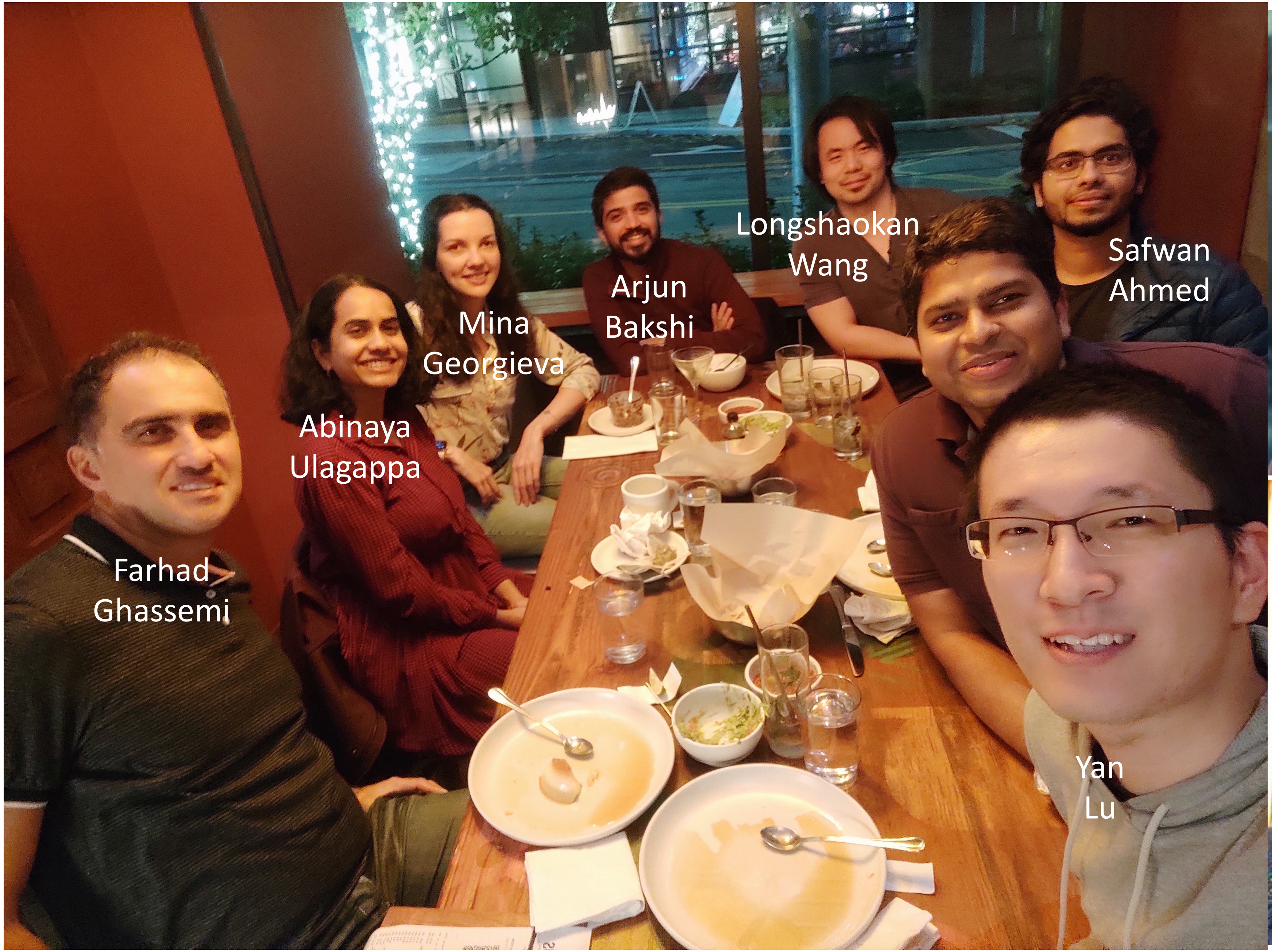
Presenter: Longshaokan (Marshall) Wang, longsha@amazon.com

Authors: Longshaokan Wang, Lingda Wang, Mina Georgieva, Paulo Machado, Abinaya Ulagappa, Safwan Ahmed, Yan Lu, Arjun Bakshi, Farhad Ghassemi

04-19-22

Amazon

IEEE ICASSP 2022


Longshaokan
Wang
Safwan
Ahmed
Arjun
Bakshi
Mina
Georgieva
Abinaya
Ulagappa
Farhad
Ghassemi
Yan
Lu


Lingda
Wang


Paulo
Machado

Outline

- **Introduction**
- Method
- Experiments
- Summary

Motivation

- Planners and optimization systems often require distribution forecast
  - Product manufacturing
  - Inventory Allocation
- Quantifying uncertainty associated with point forecast
- **Goal**: Develop accurate and efficient method for generating distribution forecast at scale

Summary

- Proposed a flexible plug-and-play framework that can extend an arbitrary Point Forecast model to produce Distribution Forecast
- Extended bootstrapping predictive residuals with backtest and covariate sampling
- Proposed an adaptive residual selector
- Proposed a new formula for applying bootstrapped residuals
- Empirical evaluation on real-world data

Summary

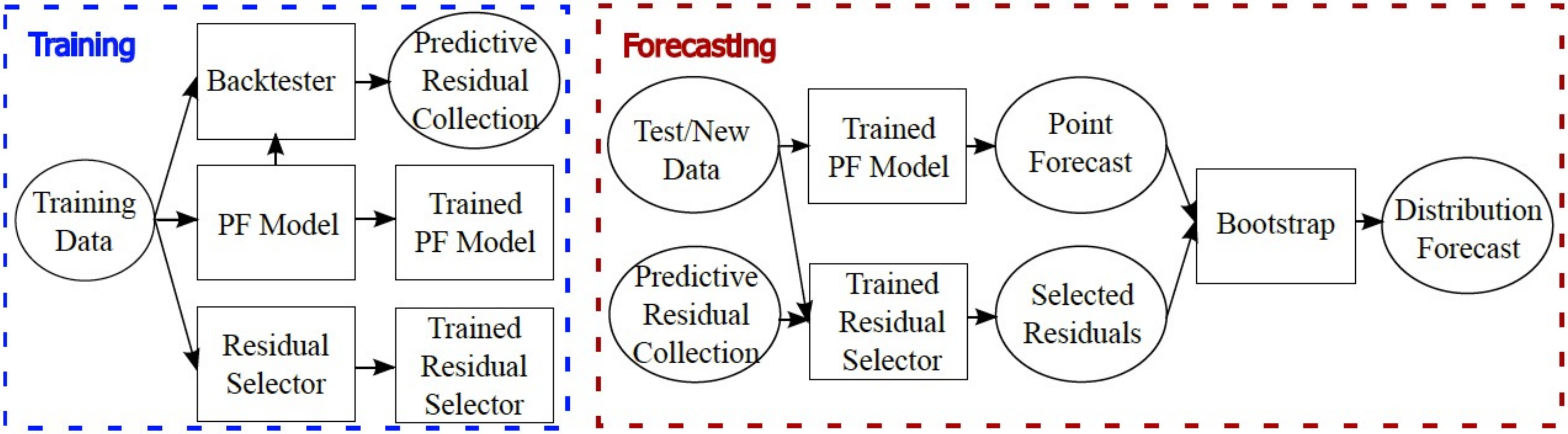
- The proposed Distribution Forecast framework has the following advantages:
  - Incorporates different sources of forecast uncertainty by design
  - Integrates well with an arbitrary PF model to produce DF
  - Is robust to model misspecification
  - Has negligible inference time latency
  - Retains interpretability for model diagnostics
  - State-of-the-art (SOTA) performance on internal and public datasets
  - Can provide more accurate point forecast through Bagging

amazon

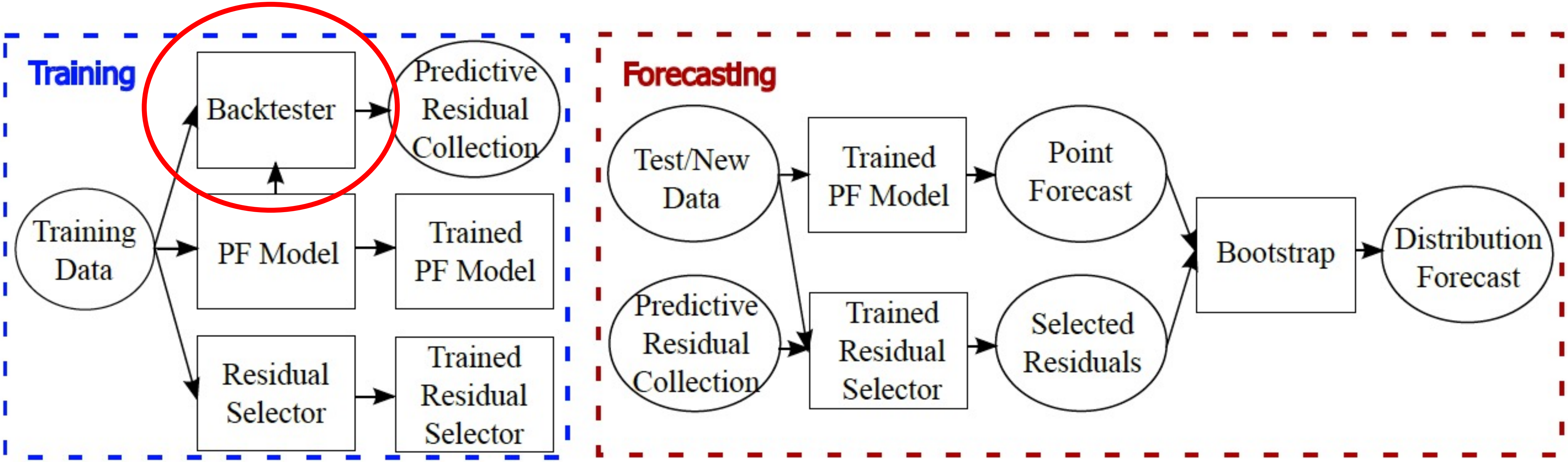
Outline

- Introduction
- **Method**
- Experiments
- Summary

Overview


Training
Training Data
Backtester
Predictive Residual Collection
PF Model
Trained PF Model
Residual Selector
Trained Residual Selector
Forecasting
Test/New Data
Predictive Residual Collection
Trained PF Model
Trained Residual Selector
Point Forecast
Selected Residuals
Bootstrap
Distribution Forecast

Backtesting



Backtesting (cont.)

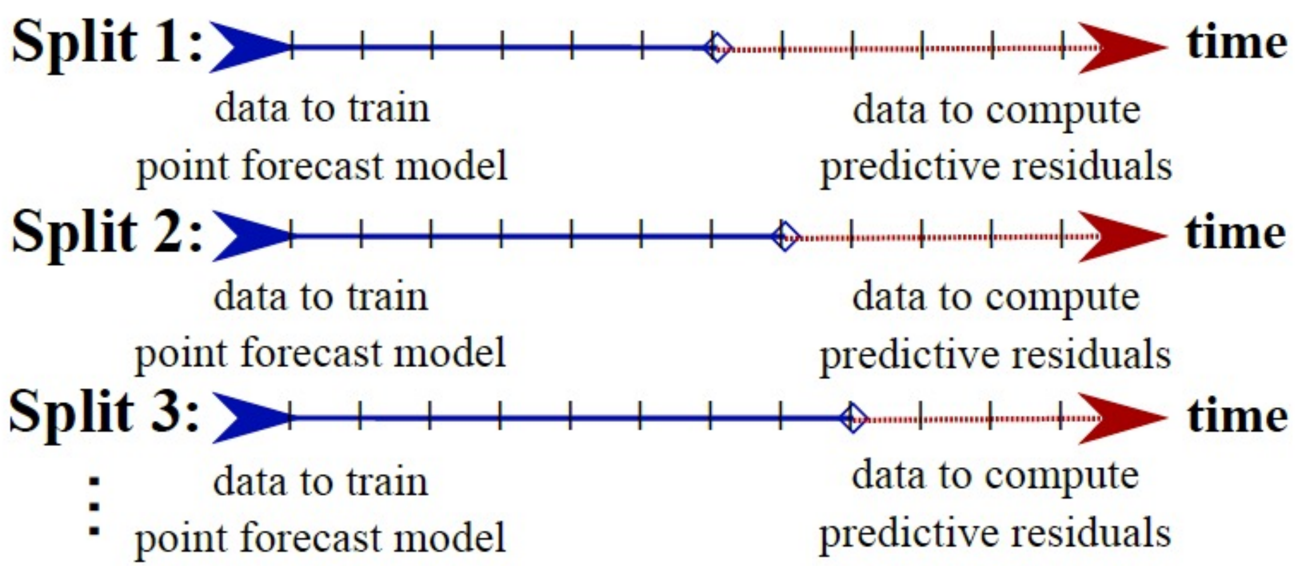
Training data: $\mathcal{D} = \{(\mathbf{X}_i^t, Y_i^t)\}_{i=1,2,\ldots,n}^{t=s_i,s_i+1,\ldots,d_i}$

Split points: $j = a, a+l, a+2l, \ldots, \max_i(d_i) - 1$

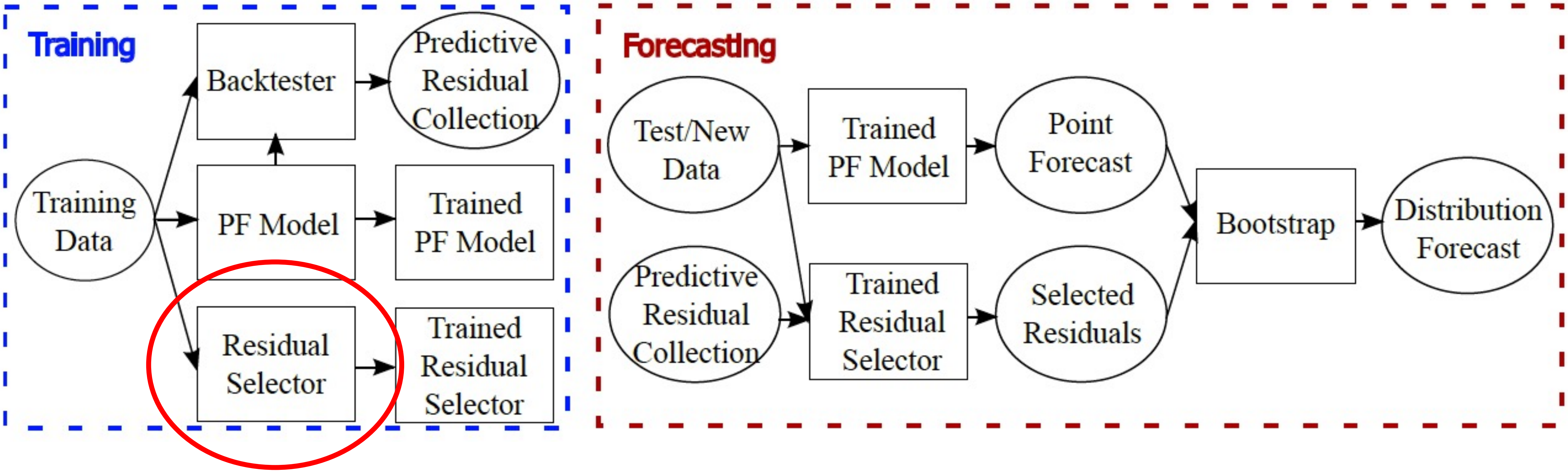
Training split: $\mathcal{A}_j = \{(\mathbf{X}_i^t, Y_i^t) \in \mathcal{D} \,|\, t \leq j\}$

Test split: $\mathcal{B}_j = \{(\mathbf{X}_i^t, Y_i^t) \in \mathcal{D} \,|\, t > j\}$

Predictive residuals from one split: $\{Y_i^t - \widehat{f}_j(Y_i^{s_i:j}, \mathbf{X}_i^{s_i:j}, \mathbf{X}_i^{(j+1):t}) | (\mathbf{X}_i^t, Y_i^t) \in \mathcal{B}_j\}$



Residual Selection

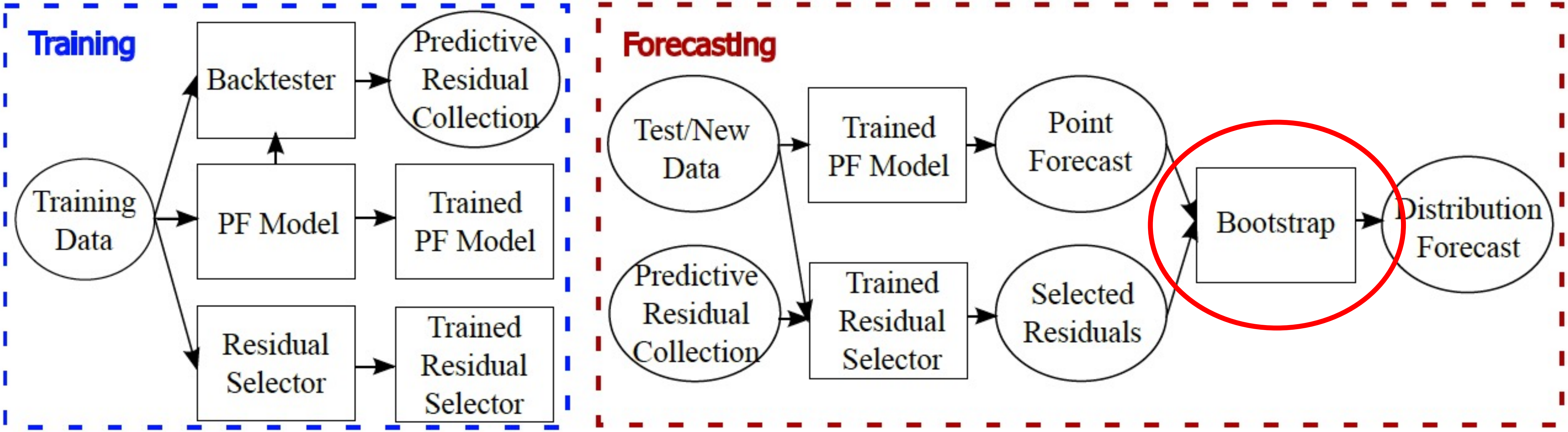


Residual Selection (cont.)

- Heuristics-based residual selection:
  - time series ID: $g(\mathcal{E}, \mathcal{M}, \mathcal{M}_{\text{future}}) = \{\varepsilon_{l,j}^t \in \mathcal{E} \,|\, l = i\}$ for time series *i*
  - time gap: $g(\mathcal{E}, \mathcal{M}, \mathcal{M}_{\text{future}}) = \{\varepsilon_{l,j}^t \in \mathcal{E} \,|\, t - j = k_i\}$
  - PF magnitude: $g(\mathcal{E}, \mathcal{M}, \mathcal{M}_{\text{future}}) = \{\varepsilon_{l,j}^t \in \mathcal{E} \,|\, \widehat{Y}_{l,j}^t \in (\widehat{Y}_i^{d_i+k_i} \cdot \frac{1}{\lambda}, \widehat{Y}_i^{d_i+k_i} \cdot \lambda)\}$
  - discount ratio, price…
- Algorithm-based residual selection:
  - dCor + threshold search + Kolmogorov-Smirnov test
  - Fit a model to predict residuals from meta information

amazon

Bootstrapping



amazon

Bootstrapping (cont.)

- First obtain point forecast $\widehat{Y}_i^{d_i+1} = \widehat{f}(Y_i^{s_i:d_i}, \mathbf{X}_i^{s_i:d_i}, \mathbf{X}_i^{d_i+1})$ and selected residuals $\mathcal{G} = \widehat{g}(\mathcal{E}, \mathcal{M}, \mathcal{M}_i^{d_i+1})$
- For $b = 1, 2, \ldots, B$, draw $\varepsilon_b \in \mathcal{G}$
- Generate 1-step bootstrap forecast:
  - Backtest-Additive:

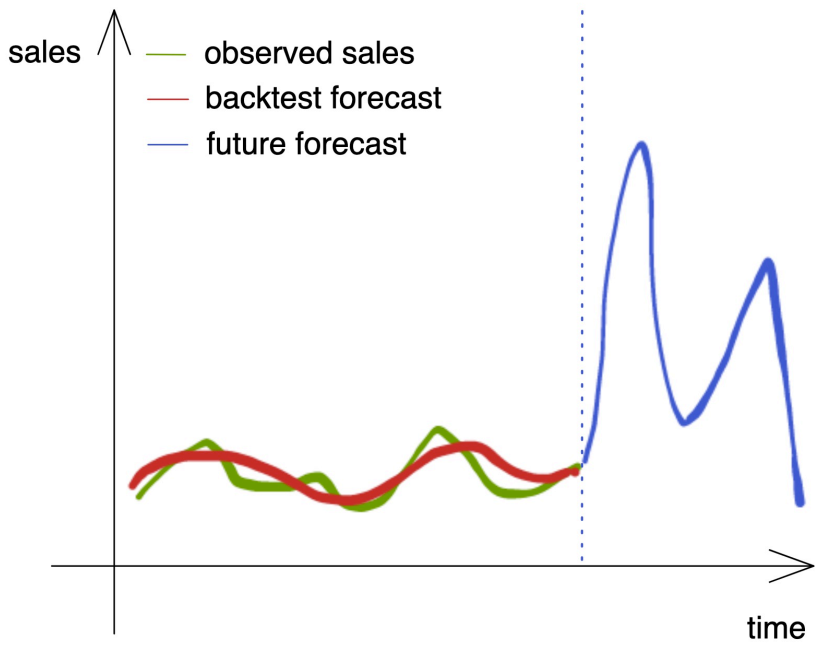
$$\widehat{Y}_{i,b,\mathrm{Add.}}^{d_i+1} = \widehat{Y}_i^{d_i+1} + \varepsilon_b$$

  - Backtest-Multiplicative:

$$r_b = \varepsilon_b / \widehat{Y}_b$$

$$\widehat{Y}_{i,b,\mathrm{Multi.}}^{d_i+1} = \widehat{Y}_i^{d_i+1} \cdot (1 + r_b) = \widehat{Y}_i^{d_i+1} + \widehat{Y}_i^{d_i+1} / \widehat{Y}_b \cdot \varepsilon_b$$

Motivation behind Backtest-Multi.



Practical Considerations

- Backtest and residual selection steps can be efficiently parallelized
- Negligible inference latency to obtain distribution forecast given point forecast
- Can generate quantile forecast for arbitrary quantiles w/o retraining
- Retains interpretability for model diagnostics

amazon

Outline

- Introduction
- Method
- **Experiments**
- Summary

Setup

- Data:
  - Sales data from Amazon.com
    - Between 01/01/2017 and 01/10/2021
    - 76 products
    - 147 covariates capturing information on pricing, supply constraints, trend, seasonality, special events, and product attributes
  - M4-hourly competition data (Makridakis 2018)
- 100-fold backtest for evaluation, separate from backtest for computing residuals
- Evaluation metric: Absolute Coverage Error (ACE):

$$\mathrm{CO}(\mathcal{D}_{\text{test}};\tau) = \frac{1}{|\mathcal{D}_{\text{test}}|}\sum\nolimits_{\mathcal{D}_{\text{test}}} I\{Y_i^t \leq \widehat{Y}_{i(\tau)}^t\}$$

$$\mathrm{ACE}(\mathcal{D}_{\text{test}};\tau) = |\mathrm{CO}(\mathcal{D}_{\text{test}};\tau) - \tau|$$

- Results averaged across backtest folds, 24-week/48-hour horizon for Sales/M4 data, 10 seeds for deep learning models, and target quantiles 0.1, 0.2, …, 0.9.

amazon

Comparison Against Classic Bootstrap Approaches

- Compare the proposed Backtest-Additive (BA) and Backtest-Multiplicative (BM) with bootstrap with fitted residuals (FR) (Hyndman 2018) and boostrap with fitted models (FM) (Pan 2016).

Table 1: ACE comparison of different bootstrap DF approaches integrated with different PF models.

Bootstrap\PF	Ridge	SVR	RF	NN
FR	$0.102(-0\%)$	$0.195(-0\%)$	$0.207(-0\%)$	$0.176(-0\%)$
FM	$0.095(-7\%)$	$0.218(+12\%)$	$0.171(-17\%)$	$0.125(-29\%)$
BA	$0.069(-32\%)$	$0.065(-67\%)$	$0.055(-73\%)$	$0.077(-56\%)$
BM	$\mathbf{0.038}(\mathbf{-63}\%)$	$\mathbf{0.061}(\mathbf{-69}\%)$	$\mathbf{0.027}(\mathbf{-87}\%)$	$\mathbf{0.048}(\mathbf{-73}\%)$

amazon

Comparison Against SOTA Approaches

- Compare the proposed bootstrap methods with SOTA approaches including Quantile Lasso, Quantile Gradient Boosting, DeepAR (Salinas 2020), Deep Factors (Wang 2019), MQ-CNN (Wen 2017), DSSM (Rangapuram 2018), and TFT (Lim 2021).

Table 2: ACE comparison of backtest-based bootstrap integrated with the median forecast vs the default DF.

DF\Model	QLasso	QGB	DeepAR	DFact	MQCNN	DSSM	TFT
Default	0.188	0.119	0.102	0.098	0.092	0.136	0.067
Median + BA	0.114	0.078	**0.100**	**0.067**	0.078	0.124	**0.058**
Median + BM	**0.039**	**0.036**	0.104	0.070	**0.071**	**0.112**	0.060

Robustness Against Model Assumptions

Table 3: ACE comparison of backtest-based bootstrap integrated with the median forecast vs the default DF from DeepAR under different pre-specified output distributions.

DF\Output Dist.	Neg. Bin.	Student's t	Normal	Gamma	Laplace	Poisson
Default	0.102	0.192	0.162	**0.138**	0.114	0.134
Median + BA	**0.100**	0.169	0.116	0.157	0.094	0.128
Median + BM	0.104	**0.165**	**0.111**	0.156	**0.088**	**0.125**

Improving Accuracy of Point Forecast via Bagging

Table 4: Relative change in MAPE for Bagging PF compared to the original PF.

Bootstrap\PF Model	Ridge	SVR	RF	NN
FR	+0.8%	+6.5%	+0.2%	+0.7%
FM	+0.4%	+6.6%	−3.8%	+2.6%
BA	−12.3%	−21.0%	**−10.0%**	+1.5%
BM	**−22.1%**	**−31.8%**	−5.3%	**−13.4%**

Outline

- Introduction
- Method
- Experiments
- **Summary**

Summary

- Proposed a Distribution Forecast framework with the following advantages:
  - Incorporates different sources of forecast uncertainty by design
  - Integrates well with an arbitrary PF model to produce DF
  - Is robust to model misspecification
  - Has negligible inference time latency
  - Retains interpretability for model diagnostics
  - State-of-the-art (SOTA) performance on internal and public datasets
  - Can provide more accurate point forecast through Bagging

Thank you!

longsha@amazon.com