Dynamic distributional forecasting using machine learning

Jozef Baruník[1] **Luboš Hanus**[1]

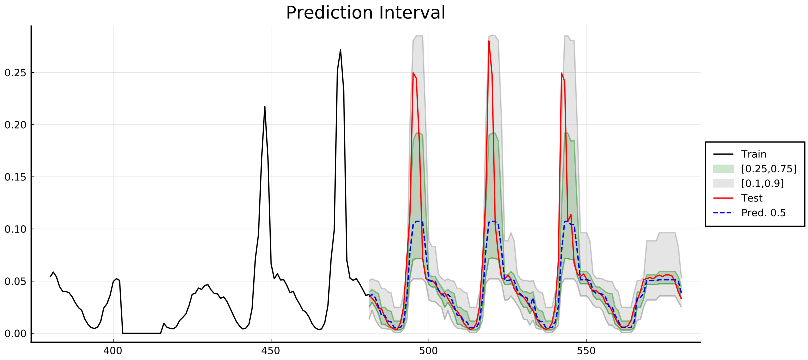
[1]The Czech Academy of Sciences
Institute of Information Theory and Automation

AFE
Novemeber 18, 2020

In **Twenty years of time series econometrics in ten pictures**:

"A fundamental problem of economic forecasting is that many economic variables are inherently very difficult to forecast, and despite advances in data availability, theory, and computational power, we have not seen dramatic improvements in forecast accuracy over the past decades."

- Stock and Watson (2017)


Prediction Interval
0.00
0.05
0.10
0.15
0.20
0.25
400
450
500
550
Train
[0.25,0.75]
[0.1,0.9]
Test
Pred. 0.5

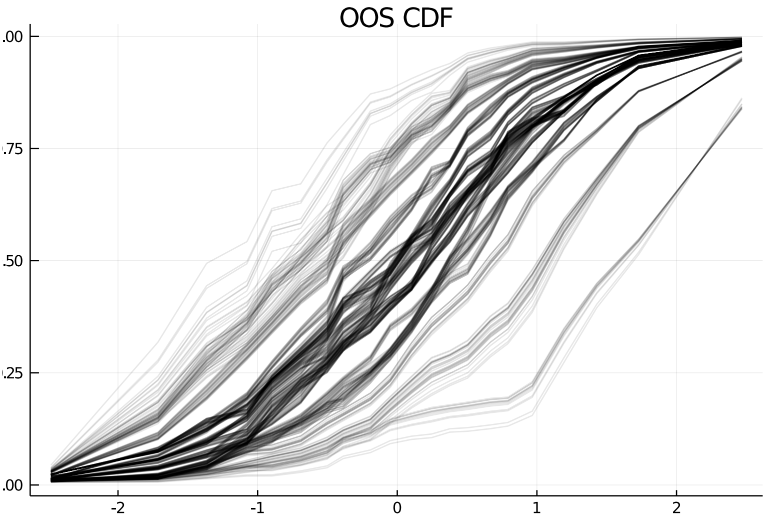


Figure: Example distribution functions.

Outline

Motivation and current state

- Large literature and interest in the first two moments of distributions.
  - Predictability of expected value of stock returns.
  - Simple volatility can predict the conditional second moment.
- **Quantiles** contrast to the limiting mean-variance analysis.
- Research focusing on forecasting **distributions** – limited.
- Modelling the **uncertainty**.
-

Probabilistic forecasting

- Point forecast, and using residuals to estimate a distribution.
  - Distribution of data is not considered while modeling
- Parametric modeling
  - Example of Gaussian: we estimate μ and σ
  - Small number of parameters
  - Assumption on data distribution $\rightarrow$ lack of complexity
- Non/semi-parametric modeling
  - Not that strong in assumptions $\rightarrow$ complex distributions
  - To many parameters $\rightarrow$ extensive modeling

Forecasting in quantiles

- Large economic literature looking at quantiles - started by Koenker and Bassett Jr (1978).
- Understandable problem and move our thoughts from average man in utility, tail events in finance, etc.
- Description of distributions using **conditional quantile** of returns, e.g. Engle and Manganelli (2004), Žikeš and Baruník (2016),...
- Quantile vs. Distribution regressions (Leorato & Peracchi, 2015)

Ways to look at a distribution

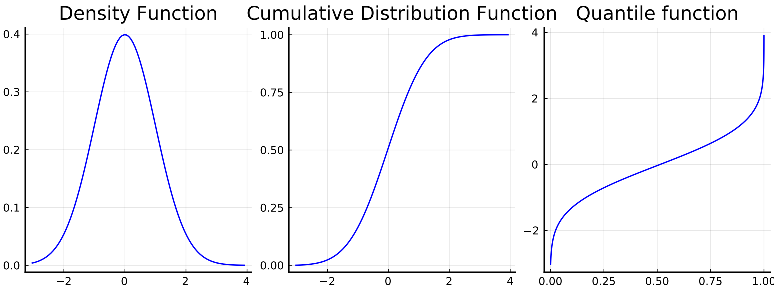


Figure: Functions/representations of a distribution

Choice of CDF over Quantiles because of Anatolyev and Baruník (2019) :) [later]

Distribution forecasting

A collection of conditional probabilities to describe cumulative distribution of excess returns, Foresi and Peracchi (1995)

- $F(y|x) = \Pr(Y \leq y|X = x)$
- Extension to multiple cutoffs J to estimate distinct functions
  $F_1(x), \ldots, F_J(x)$ for $-\infty < y_1 < \cdots < y_J < \infty$
  - Set of individual logits. Unordered model. Monotonicity problem with logits.

Ordered binary choice model for conditional return distribution - Anatolyev and Baruník (2019)

- $F_{t+1}(r) = \Pr\{r_{t+1} \leq c_j|\mathcal{I}_t\}$
- MLE model for all $j = 1, ..., J$ - fewer parameters $\rightarrow$ dimensionality reduction
- Monotonicity fulfilled - BUT imposed (measurement error).

Distribution illustration of approximation points

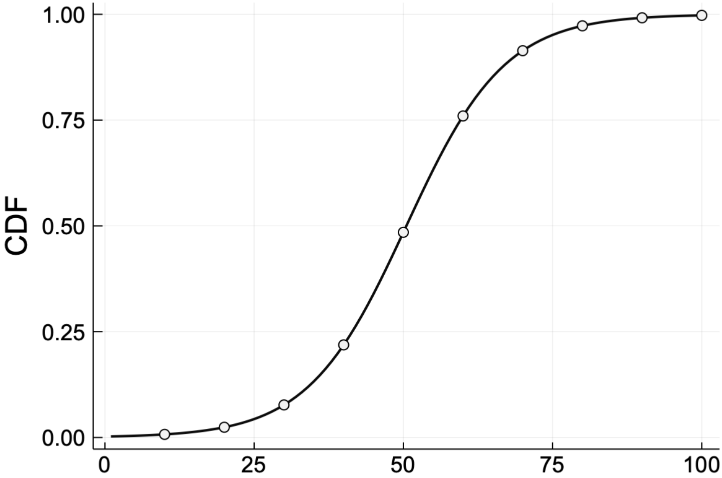


Figure: Cumulative distribution function

Does (not only) AB(2019) survive in Big Data?

- However, many parameters and large uncertainty.
- Which of explanatory variables are informative?
- Can we use two to three variables to explain the distribution?
- Tons of data to choose from.
- Can we use all of the data?

ML(E?)

Examples of ML in economics

- Mullainathan and Spiess (2017) - Machine learning: an applied econometric approach
- Sirignano et al., 2016 - Deep learning for mortgage risk
- Heaton et al., 2017 - Deep learing for finance: deep portfolios
- Gu et al. (2020) - Empirical asset pricing via machine learning
- Iworiso and Vrontos (2019) - Directional Predictability of Equity Premium
- Hronec (2020)
- ...

Literature related to probabilistic forecasting

- Salinas et al., 2020 - DeepAR: Probabilistic forecasting with autoregressive recurrent networks - Amazon Research Germany
  - assumptions on distributions - estimation of parameters: $N(\mu, \sigma^2)$
- Duan et al., 2019 NGBoost: Natural Gradient Boosting for Probabilistic Prediction
- Wen et al., 2017 - A Multi-Horizon Quantile Recurrent Forecaster
- Hu et al., 2019 - Distribution-Free Probability Density Forecast Through Deep Neural Networks
- Lim and Gorse, 2020 Deep Probabilistic Modelling of Price Movements for High-Frequency Trading
- ...

Distributional forecasting - ML

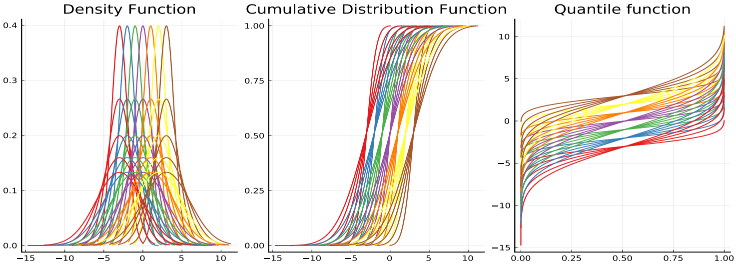


Figure: Functions/representations of a distribution

Requirements/restrictions/convenience to model

- CDF must be non-decreasing and on [0,1]
- Quantile function - only non-decreasing, but in the literature, authors model quantiles separetely mostly.
- Choice: Non-decreasing BOTH, CDF on [0,1] - output layer sigmoid $\rightarrow$ no difference.

Distributional forecasting - ML

- Building on the previous knowledge we use ML to estimate the distribution:

$$F_{t+1|t}(y) = P(y_{t+1} \leq y^{\tau} | \mathcal{I}_t)) = \mathfrak{g}(\cdot), \tag{1}$$

  where y^{τ} is an empirical quantile corresponding to a particular τ level.
- Quantile forecast for prediction intervals is inversely related to the CDF.

$$q_{y_{t+1}|\mathcal{I}_t}(\tau) = F^{-1}_{t+1|t}(y) \tag{2}$$

Model - network

- Predicting a set of conditional probabilities as output of the network:

$$\left(F^{(c_1)}_{y_{t+1}}, \ldots, F^{(c_J)}_{y_{t+1}}\right) = \mathfrak{g}_{W,b}(x_t) = g^{(L)}_{W^{(L)},b^{(L)}} \circ \ldots \circ g^{(1)}_{W^{(1)},b^{(1)}}(x_t)\,,$$

  where, in our case, $g^{(L)}(\cdot) = \sigma(\cdot)$.
- Loss function with monotonicity penalty:

$$Loss = \underbrace{\frac{1}{T}\sum_{t}^{T}(y_t - \hat{y}_t)^2}_{\text{MSE}} + \underbrace{\beta_m \sum_{t}^{T}\left(\hat{y}_{t,j} - \hat{y}_{t,j+1}\right)_+}_{\text{monotonicity check}} \tag{3}$$

  where $(y)_+$ is a rectified linear unit function, $relu(y) = max(y, 0)$.
- We use both types of networks MLP and RNN.

Distributional forecasting - ML

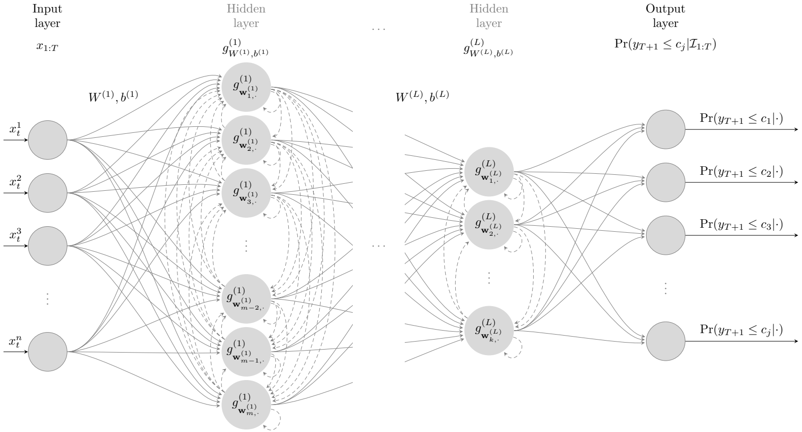


Figure: Recurrent Neural Network - probabilistic output

Procedure to obtain CDFs and prediction intervals

1. Parameters to choose: J, split size train/test
2. Create indicator based on **train** data

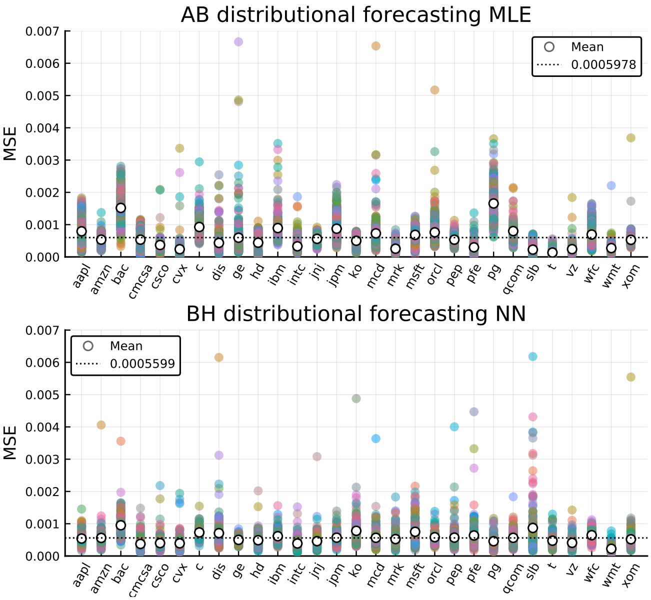
$$\mathbb{I}\{r_{1:T} \leq c_j\} = \begin{bmatrix} 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & 1.0 \\ 1.0 & 1.0 & 1.0 & 1.0 & 1.0 & 1.0 \\ \dots & \dots & \dots & \dots & \dots & \dots \end{bmatrix}$$

3. Estimation: Hyperoptimization with Cross-validation to find η, λ, p on train.
4. Bag of ensembles on with best model -> average prediction on test
5. Interpolation of estimated conditional distributions Fritsch and Carlson, 1980
   - $J = \{11, 21, 41, ...\}$ points approximated by 600 points, for instance
6. Find quantiles defining prediction intervals for desired probabilities

Finacial time-series study

- Data (2005:2018): 29 most liquid US stocks
- Predictors of each asset's equation (lag=1):
  - Log volatility, and Indicator: $j = 1, \ldots J$
- Train/Test ratio is 0.9 $\implies$ 1:3016 / 3017:3352
- Rolling estimates - forecast horizon $\{1, 2\}$
- Cross-validation: 5-folds
- Hyperparameter optimization: number of searches, RandomSearch
  - Dropout(p), $p = [0.0, 0.01, 0.05, 0.1, 0.2, 0.3]$
  - ADAMW(learning rate, λ), $\lambda = [0.005, 0.001, 0.00075, 0.0005, 0.00025, 0.0001]$
- Network size: [128, 64, 32] and [256, 128, 64, 32]
- Activation functions: ReLU, Sigmoid (output)
- No early stopping, instead taking $\hat{y}$ at minimum of cost functions.

Out-of-sample comparison: 29 assets, j=19



Forecast of prediction interval

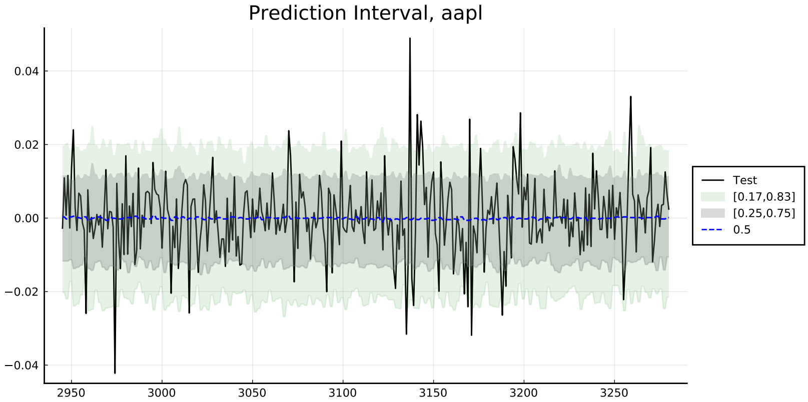


Figure: Rolling OOS prediction of aapl, H=2, js=41, MLP

Rolling forecast by NN, j=19

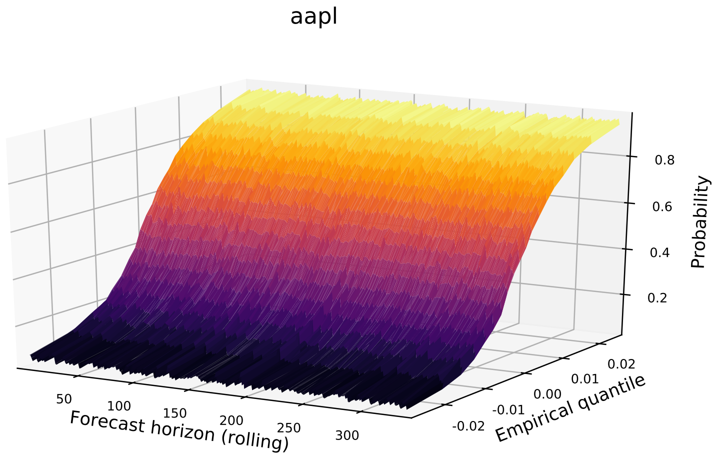


Figure: One asset [appl] - rolling forecasts, not interpolated

Macroeconomic data predictions: Δy_t, π_t, u_t, i_t

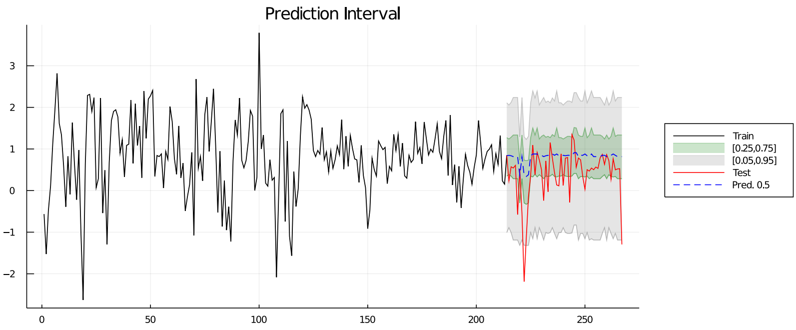


Figure: GDP growth: LSTM(64,32), J=19, batchsize=2

Macroeconomic data predictions: Δy_t, π_t, u_t, i_t

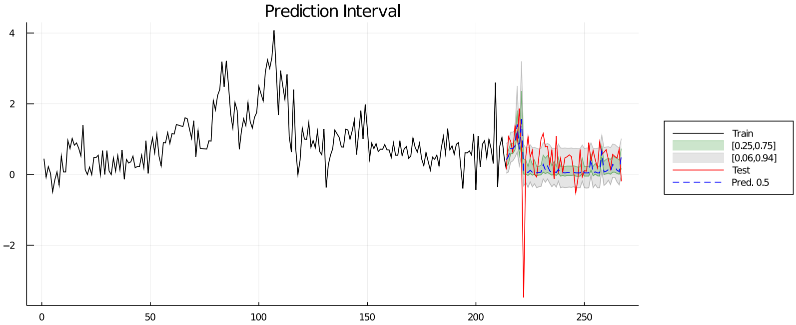


Figure: Inflation: LSTM(64,32), J=25, batchsize=2

Macroeconomic data predictions: Δy_t, π_t, u_t, i_t

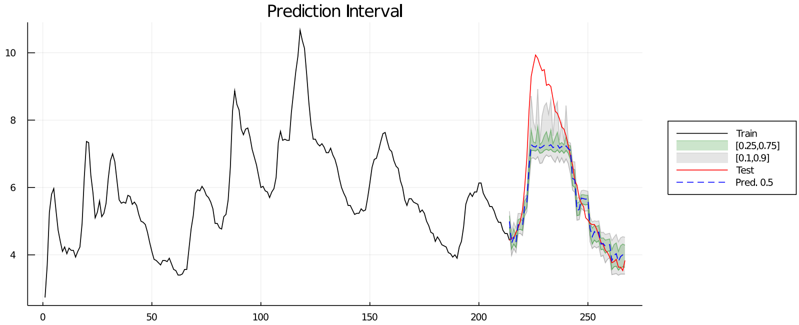


Figure: Unemployment: LSTM(64,32), J=19, batchsize=2

Macroeconomic data predictions: Δy_t, π_t, u_t, i_t

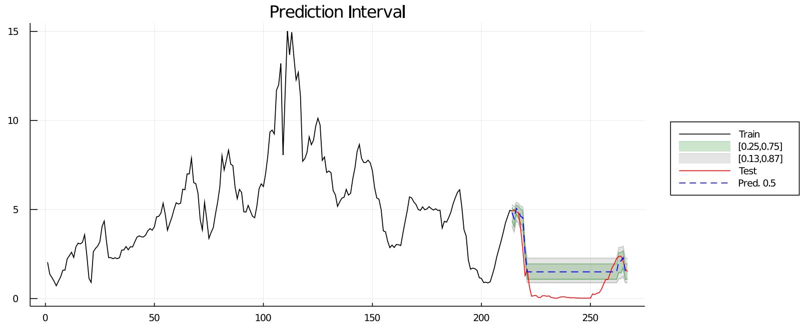


Figure: FFRate: LSTM(64,32), J=19, batchsize=2

- Observation: empirical quantiles were set prior to prediction.

Prediction intervals

Data set: Literature

- Ex. web traffic flow

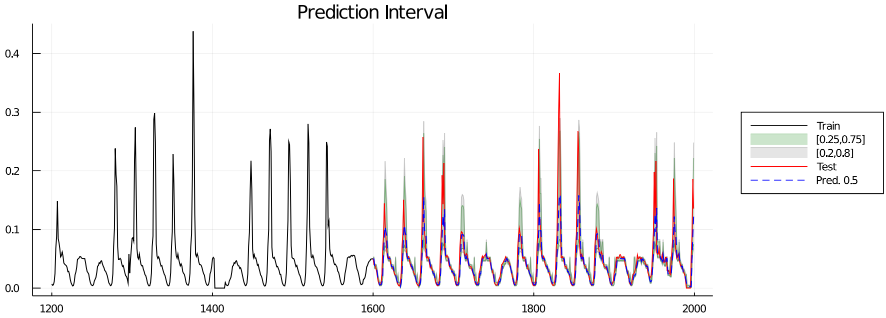


Figure: 1-steps ahead forecast, batchsize=2, LSTM(64,32)

Prediction intervals

Data set: Literature

- Ex. Exchange rate

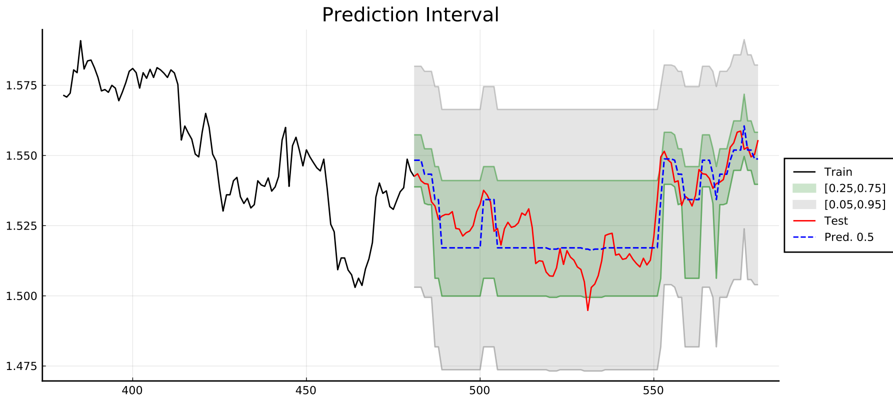


Figure: batchsize=64, LSTM(64,32)

Prediction intervals

Data set: Literature

- Ex. Exchange rate, DIFFERENT BATCHSIZE

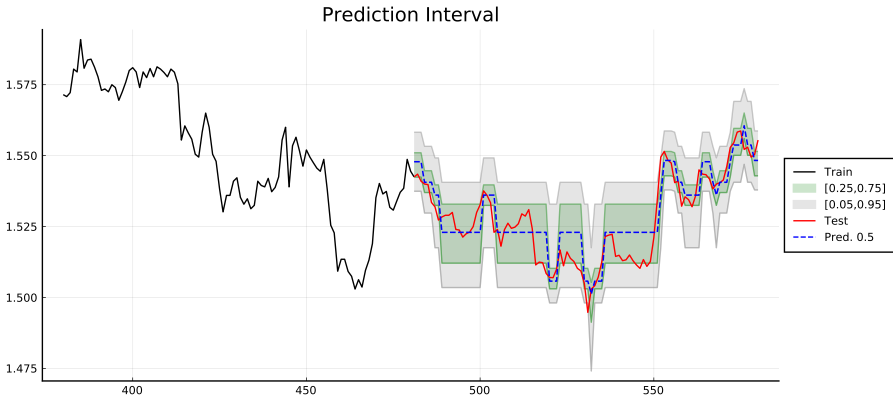


Figure: batchsize=2, LSTM(64,32)

Conclusion

- Crucial benefit: complete distribution
- Q?: Sometimes upper or lower quantiles are not at limits.
- Q?: Still questions about data preparation and parameters to set.
- Q?: Seasonality and stationarity?
- Modeling approached by learning models
  - Going from data to distribution
  - Data-driven functional form that predicts
- Flexibility of ML needs carefulness.

Bibliography I

- Anatolyev, S., & Baruník, J. (2019). Forecasting dynamic return distributions based on ordered binary choice. *International Journal of Forecasting*, *35*(3), 823–835.
- Duan, T., Avati, A., Ding, D. Y., Basu, S., Ng, A. Y., & Schuler, A. (2019). Ngboost: Natural gradient boosting for probabilistic prediction. *arXiv preprint arXiv:1910.03225*.
- Engle, R. F., & Manganelli, S. (2004). Caviar: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics*, *22*(4), 367–381.
- Foresi, S., & Peracchi, F. (1995). The conditional distribution of excess returns: An empirical analysis. *Journal of the American Statistical Association*, *90*(430), 451–466.

Bibliography II

- Fritsch, F. N., & Carlson, R. E. (1980). Monotone piecewise cubic interpolation. *SIAM Journal on Numerical Analysis*, *17*(2), 238–246.
- Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, *33*(5), 2223–2273.
- Heaton, J. B., Polson, N. G., & Witte, J. H. (2017). Deep learning for finance: Deep portfolios. *Applied Stochastic Models in Business and Industry*, *33*(1), 3–12.
- Hu, T., Guo, Q., Li, Z., Shen, X., & Sun, H. (2019). Distribution-free probability density forecast through deep neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, *31*(2), 612–625.

Bibliography III

- Iworiso, J., & Vrontos, S. (2019). On the directional predictability of equity premium using machine learning techniques. *Journal of Forecasting*, *n/a*(n/a). https://doi.org/10.1002/for.2632
- Koenker, R., & Bassett Jr, G. (1978). Regression quantiles. *Econometrica: journal of the Econometric Society*, 33–50.
- Leorato, S., & Peracchi, F. (2015). Comparing distribution and quantile regression.
- Lim, Y.-S., & Gorse, D. (2020). Deep probabilistic modelling of price movements for high-frequency trading. *arXiv preprint arXiv:2004.01498*.
- Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, *31*(2), 87–106.

Bibliography IV

- Salinas, D., Flunkert, V., Gasthaus, J., & Januschowski, T. (2020). Deepar: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting*, *36*(3), 1181–1191.
- Sirignano, J., Sadhwani, A., & Giesecke, K. (2016). Deep learning for mortgage risk. *arXiv preprint arXiv:1607.02470*.
- Stock, J. H., & Watson, M. W. (2017). Twenty years of time series econometrics in ten pictures. *Journal of Economic Perspectives*, *31*(2), 59–86.
- Wen, R., Torkkola, K., Narayanaswamy, B., & Madeka, D. (2017). A multi-horizon quantile recurrent forecaster. *arXiv preprint arXiv:1711.11053*.

Bibliography V

Žikeš, F., & Baruník, J. (2016). Semi-parametric conditional quantile models for financial returns and realized volatility. *Journal of Financial Econometrics*, *14*(1), 185–226.