ECB Workshop on Forecasting Techniques

“The Anatomy of Out-of-Sample Forecasting Accuracy" by Borup, Coulombe, Rapach, Schütte and Schwenk-Nebbe

Discussion by Michel van der Wel

Erasmus U Rotterdam

June 13, 2023

Paper Summary

- A lot of interest in machine learning

Paper Summary

- A lot of interest in machine learning
- Subject to “black-box” criticism $\rightarrow$ Limits use for policymaking

Paper Summary

- A lot of interest in machine learning
- Subject to “black-box” criticism $\rightarrow$ Limits use for policymaking
- Paper develops Shapley-based metrics for interpreting models
  - Two metrics for importance of individual predictors for predicted target values
  - New metric (PBSV) for contribution of individual predictors for loss in sequence of fitted models

Paper Summary

- A lot of interest in machine learning
- Subject to “black-box” criticism $\rightarrow$ Limits use for policymaking
- Paper develops Shapley-based metrics for interpreting models
  - Two metrics for importance of individual predictors for predicted target values
  - New metric (PBSV) for contribution of individual predictors for loss in sequence of fitted models
- Empirical study of forecasting US inflation provides sensible leading predictors (oil, components of CPI) and discrepancies between in-sample and out-of-sample importance

Shapley values

- Model: $y_{t+1} = f(x_t) + \varepsilon_{t+1}$, with x_t dimension P; Collect all predictor indices in set $S = \{1, 2, \ldots, P\}$

Shapley values

- Model: $y_{t+1} = f(x_t) + \varepsilon_{t+1}$, with x_t dimension P; Collect all predictor indices in set $S = \{1, 2, \ldots, P\}$

- Shapley value of p-th predictor is comparison of all possible models including p to all without p; in my sloppy notation:

$$\phi_p = \sum_{Q \subseteq S \setminus \{p\}} \frac{|Q|!(P-|Q|-1)!}{P!} \left(E[f|Q \cup \{p\}] - E[f|Q] \right)$$

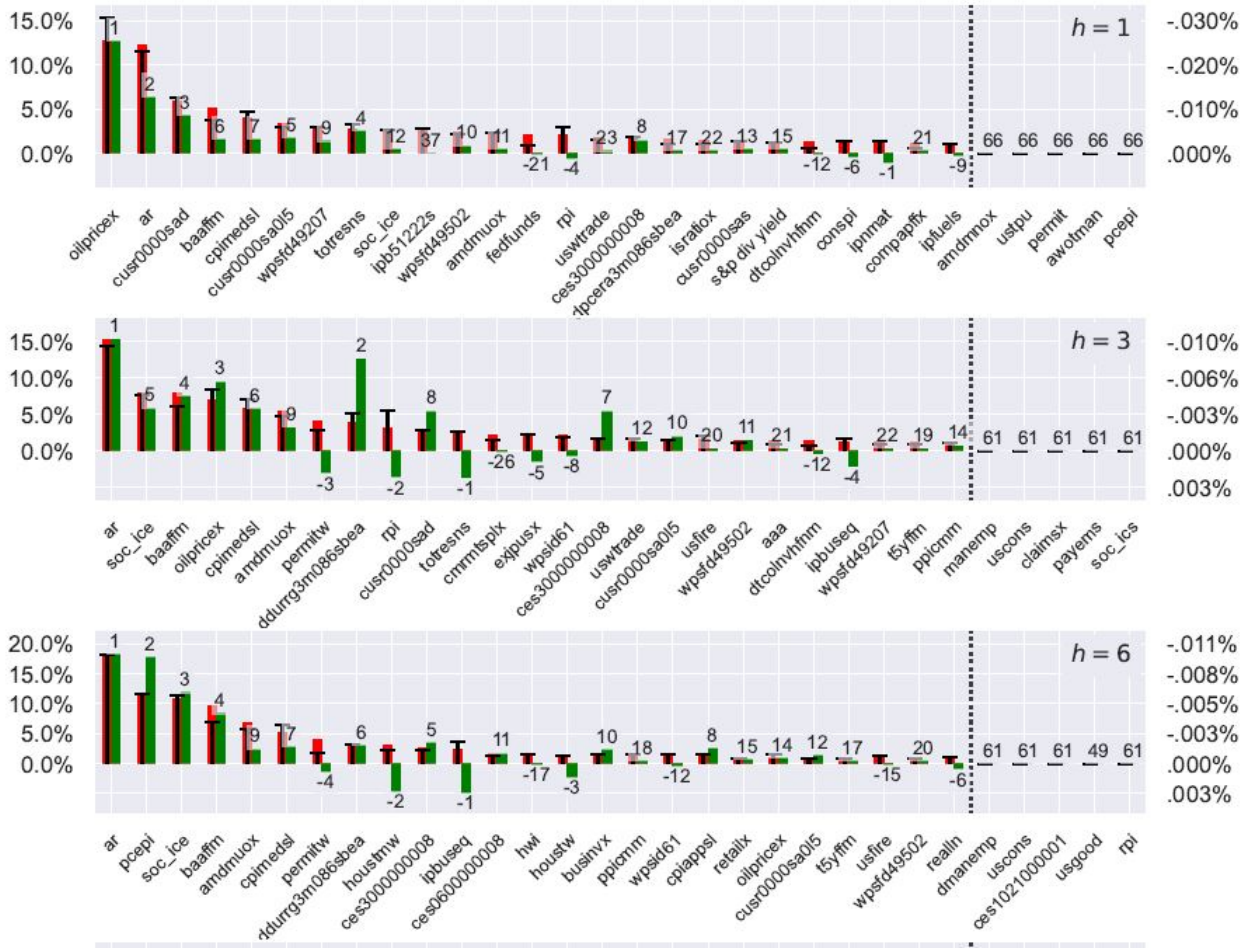
Shapley values

- Model: $y_{t+1} = f(x_t) + \varepsilon_{t+1}$, with x_t dimension P; Collect all predictor indices in set $S = \{1, 2, \ldots, P\}$
- Shapley value of p-th predictor is comparison of all possible models including p to all without p; in my sloppy notation:

$$\phi_p = \sum_{Q \subseteq S \setminus \{p\}} \frac{|Q|!(P-|Q|-1)!}{P!} \left(E[f|Q \cup \{p\}] - E[f|Q]\right)$$

- Paper cleverly adjusts setting for
  - Dealing with large number of predictors (use sampling)
  - Expanding samples (not one value; take average)
  - Retraining of the samples $\rightarrow$ (i/o)Shapley-VI$_p$
  - Loss-function effects rather than predicted values $\rightarrow$ PBSV$_p$

1. Empirical findings (Figure 1)



- Stability of findings (pcepi least to 2nd most important $h = 1$ to $h = 6$)?
- Why not more correlated results (similar series in FRED-MD)?

2. Empirical application

- Curious to robustness regarding
  - Forecasting y_{t+h} rather than $\frac{1}{h}\sum_{k=1}^{h} y_{t+k}$
  - Including predictors and moving average of predictors
  - Selection of L (AR-lag) and q (MA-order)

2. Empirical application

- Curious to robustness regarding
  - Forecasting y_{t+h} rather than $\frac{1}{h}\sum_{k=1}^{h} y_{t+k}$
  - Including predictors and moving average of predictors
  - Selection of L (AR-lag) and q (MA-order)

- Main benchmark is AR(k) model. Consider smaller/more targeted model with explanatory variables, survey data, etc? Particularly given somewhat modest (short-horizon) improvement of ML approaches (7% for $h = 1$; 19% for $h = 12$)

2. Empirical application

- Curious to robustness regarding
  - Forecasting y_{t+h} rather than $\frac{1}{h}\sum_{k=1}^{h} y_{t+k}$
  - Including predictors and moving average of predictors
  - Selection of L (AR-lag) and q (MA-order)

- Main benchmark is AR(k) model. Consider smaller/more targeted model with explanatory variables, survey data, etc? Particularly given somewhat modest (short-horizon) improvement of ML approaches (7% for $h = 1$; 19% for $h = 12$)

- For PCA, possible to compare with significance?

3. Lazy/Free-riding discussant

Overes and Van der Wel (Computational Economics; 2023) also use Shapley values (for driving factors of sovereign credit ratings). From referee process:

- Is iteratively omitting variables possible/useful?

3. Lazy/Free-riding discussant

Overes and Van der Wel (Computational Economics; 2023) also use Shapley values (for driving factors of sovereign credit ratings). From referee process:

- Is iteratively omitting variables possible/useful?
- Can you extend the results with more machine learning techniques? [Ensembles mentioned; Included in your work!]

3. Lazy/Free-riding discussant

Overes and Van der Wel (Computational Economics; 2023) also use Shapley values (for driving factors of sovereign credit ratings). From referee process:

- Is iteratively omitting variables possible/useful?
- Can you extend the results with more machine learning techniques? [Ensembles mentioned; Included in your work!]
- Can you take the panel nature into account? [You do!]

3. Lazy/Free-riding discussant

Overes and Van der Wel (Computational Economics; 2023) also use Shapley values (for driving factors of sovereign credit ratings). From referee process:

- Is iteratively omitting variables possible/useful?
- Can you extend the results with more machine learning techniques? [Ensembles mentioned; Included in your work!]
- Can you take the panel nature into account? [You do!]
- Compare to `scikit-learn` package (which also provides feature importance estimates)

3. Lazy/Free-riding discussant

Overes and Van der Wel (Computational Economics; 2023) also use Shapley values (for driving factors of sovereign credit ratings). From referee process:

- Is iteratively omitting variables possible/useful?
- Can you extend the results with more machine learning techniques? [Ensembles mentioned; Included in your work!]
- Can you take the panel nature into account? [You do!]
- Compare to `scikit-learn` package (which also provides feature importance estimates)
- Closer comparison of findings with existing literature and evaluation also of signs

Final point and conclusion

Final point: What audience do you see for this paper?

- In-between ML and econometrics, currently

Final point and conclusion

Final point: What audience do you see for this paper?

- In-between ML and econometrics, currently
- If more econometric, perhaps some (small-scale!) simulation study is possible? And further breakdown/analysis/comparison with existing models/metrics for linear model?

Final point and conclusion

Final point: What audience do you see for this paper?

- In-between ML and econometrics, currently
- If more econometric, perhaps some (small-scale!) simulation study is possible? And further breakdown/analysis/comparison with existing models/metrics for linear model?

In conclusion:

- Opens the black box with clever adaptations to time series setting
- Great work!

ECB Workshop on Forecasting Techniques

“The Anatomy of Out-of-Sample Forecasting Accuracy" by Borup, Coulombe, Rapach, Schütte and Schwenk-Nebbe

Discussion by Michel van der Wel

Erasmus U Rotterdam

June 13, 2023