Session 1B, Machine Learning for Central Banks

Discussant: Nathan Palmer[1]

November 11, 2020

[1] Federal Reserve Board. These remarks represent my own views and not necessarily those of the Federal Reserve.

Preliminaries: Benefits & Cautions of Machine Learning

Benefits:

1. Discover non-linearities and interactions
2. Handles "big data" or "big parameters" efficiently

Cautions:

3. Overfitting: model selection/validation?
4. Interpretation:
   a. Black box interpretation?
   b. Model inference?

Forecasting UK Inflation Bottom Up by Joseph, Kalamara, Potjagailo, Kapetanios

Discussed by Nathan Palmer

November 11, 2020

Motivation: Improve Central Bank Inflation Forecasting

- Use item-level CPI component series to forecast headline CPI
  - plus macro series
- Horse race a set of inflation forecasting models:
  - traditional dimensionality reduction (PCA and others)
  - shrinkage (Ridge, LASSO, elastic net)
  - non-linear ML (SVM, NN, RF)
- Interpret results with model-agnostic Shapley regressions

Forecasting Problem

LHS:

- UK headline CPI, yoy
- UK core CPI, ex food and energy, yoy
- UK core CPI, services, yoy

RHS regressors:

- 581 (491) item-level CPI components series, levels
- 46 macro series (robustness check)
- lags of LHS

Results

- Shrinkage models and ML models improved most over AR(p)
- Many statistically significant improvements at 6, 9, 12 months
- Adding macro data: models largely improved
  - ML models and ridge regression improved less (already good)

Comments

- Usage of component series: improves forecasting and narrative
- Appreciate the interpretation and inference via Shapley regressions
- A question about the micro data

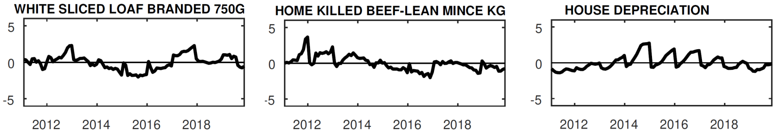
A Question About Data

The stdev portion of this plot leapt out:

Figure 1: Temporal dynamics of item-level index statistics of year-on-year changes for selected item sample. The shown changes are limited to ±25% for clearer presentation with a small number of changes beyond this range. Source: ONS and authors' calculation.

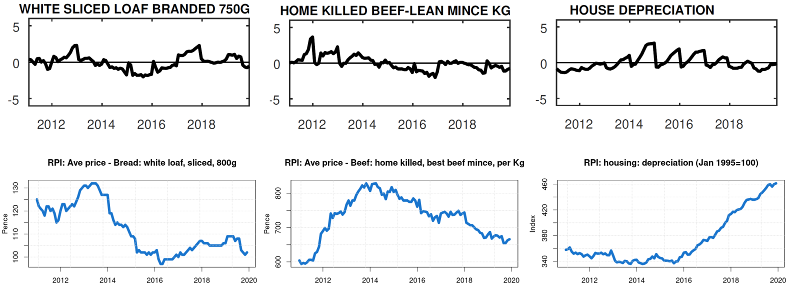
A Question About Data

Digging into the micro series



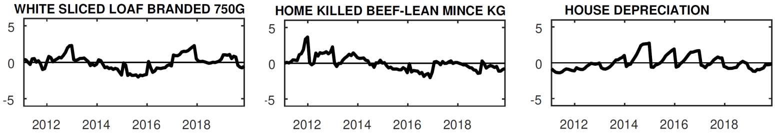
A Question About Data

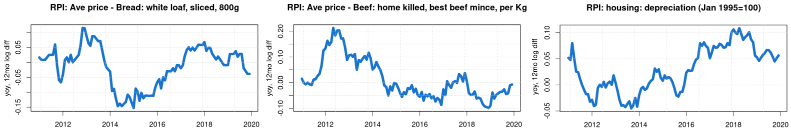
Digging into the micro series



A Question About Data

Digging into the micro series





A Question About Data

1. The stationarity of item-level series, in levels, comes from chaining and rebasing
2. Item-level series internally correlated, but perhaps 'breaks' structure with other macro series
   - Improve effect of additional macro series
3. Is it possible to use y-o-y change in nominal micro series?

The Macroeconomy as a Random Forest by Philippe Goulet Coulombe

Discussed by Nathan Palmer

November 11, 2020

Motivation: Generalize Non-linear Time Series Models

There are many frameworks for capturing time-varying β_t

- threshold/switching regressions
- smooth transitions
- structural breaks
- random walk time-varying parameters

Motivation: Generalize Non-linear Time Series Models

There are many frameworks for capturing time-varying β_t

- threshold/switching regressions
- smooth transitions
- structural breaks
- random walk time-varying parameters

Contribution:

- *by construction* random forests can generalize the above models
  - capture both latent & observable β_t time-variation
- Demonstrate inference, interpretation of these models
- Horse races: simulated data, empirical forecasting

Basic Autoregressive Random Forest (ARRF)

Consider:

$$y_t = X_t \beta_t + \epsilon_t$$
$$\beta_t = F_\beta(S_t)$$

F_β is generated by a random forest with modified objective:

Basic Autoregressive Random Forest (ARRF)

Consider:

$$y_t = X_t\beta_t + \epsilon_t$$
$$\beta_t = F_\beta(S_t)$$

F_β is generated by a random forest with modified objective:

$$\min_{j\in\mathcal{J}^-,\ c\in\mathbb{R}} \left[\min_{\beta_1} \sum_{t\in\S_1^{RW}(j,c)} w(t;\zeta)\left(y_t - X_t\beta_1\right)^2 + \lambda\|\beta_1\|_2 \right.$$
$$\left. + \min_{\beta_2} \sum_{t\in\S_2^{RW}(j,c)} w(t;\zeta)\left(y_t - X_t\beta_2\right)^2 + \lambda\|\beta_2\|_2 \right]$$

1. Replace $\bar{y}$ with $X\beta$
2. λ: local linear forest regularization
3. ζ: smooth β_t over time (shrink β_t to β_{t-1} vs 0)
4. $X_t \subset S_t$ implies additional regularization

Basic Autoregressive Random Forest (ARRF)

Consider:

$$y_t = X_t \beta_t + \epsilon_t$$
$$\beta_t = F_\beta(S_t)$$

F_β is generated by a random forest with modified objective:

$$\min_{j \in \mathcal{J}^-,\ c \in \mathbb{R}} \left[\min_{\beta_1} \sum_{t \in \S_1^{RW}(j,c)} w(t;\zeta)\left(y_t - X_t\beta_1\right)^2 + \lambda\|\beta_1\|_2 \right.$$
$$\left. + \min_{\beta_2} \sum_{t \in \S_2^{RW}(j,c)} \boxed{w(t;\zeta)}\left(y_t - \boxed{X_t\beta_2}\right)^2 + \boxed{\lambda\|\beta_2\|_2} \right]$$

1. Replace $\bar{y}$ with $X\beta$
2. λ: local linear forest regularization
3. ζ: smooth β_t over time (shrink β_t to β_{t-1} vs 0)
4. $X_t \subset S_t$ implies additional regularization

Feature Engineering in S_t

In general $X_t \subset S_t$ – for the forecasting exercise:

1. 8 lags of y_t
2. t for structural breaks/exogenous time-variation
3. 2 lags of all variables in FRED[1]
4. 8 lags of 5 factors extracted from FRED by PCA
5. for each $Z \in$ FRED, two moving-average factors via PCA

[1] 248 quarterly series, or 134 monthly series

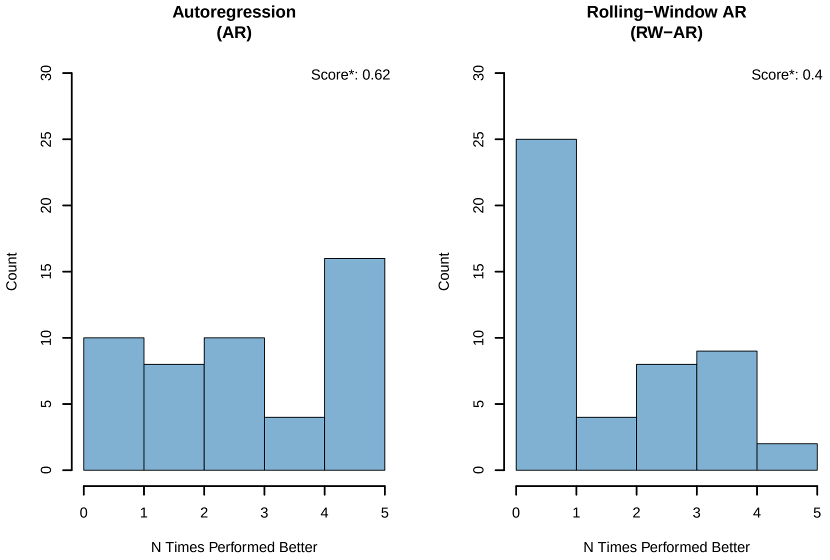
Proposed Inference

- Forests often have 100s of trees; thus 100s of $\hat{\beta}_t$ vectors
- Each tree is a posterior draw from an approximate Bayesian bootstrap on the tree functional
- Thus can construct (block) Bayesian bootstraps on the time series of $\hat{\beta}_t$

Simulated Data Horse Race

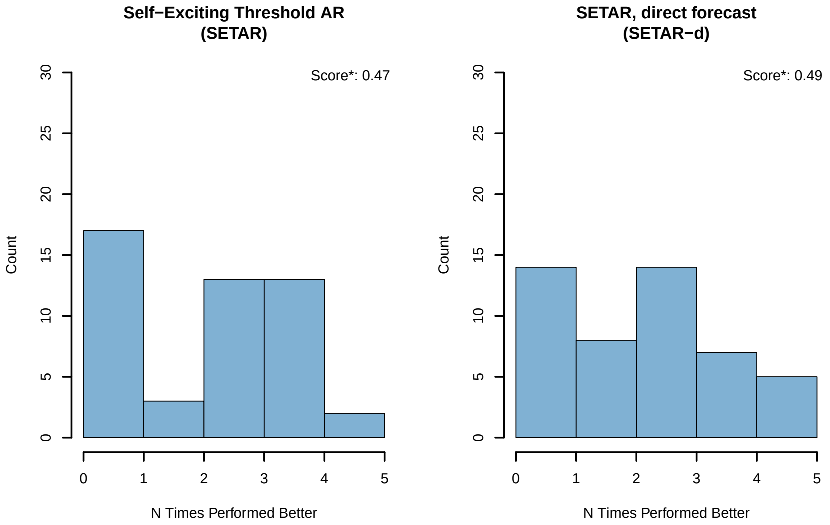
- Six increasingly "hard" DGPs:
  - AR, ..., SETAR that collapses to AR via structural break
- 2 data quantity regimes (150, 300 obs)
- 4 forecasting horizons $h \in \{1, 2, 3, 4\}$
- Six "racing" models: AR, RF, ARRF, 2 SETARs, rolling-window AR
- Simple question: how often does ARRF, RF beat AR, others?
  - max attainable: "beat 5 models, 48 times"

Simulated Data Horse Race



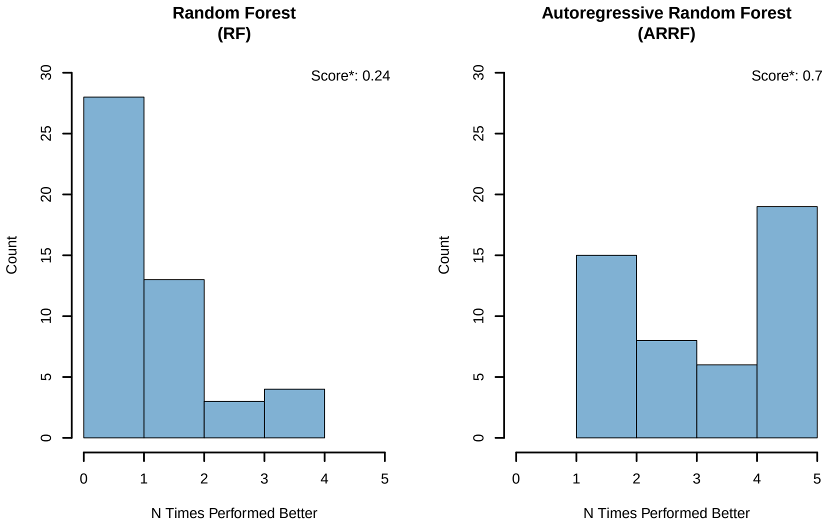
*Fraction of maximum possible score: 240 = 4*5*12

Simulated Data Horse Race



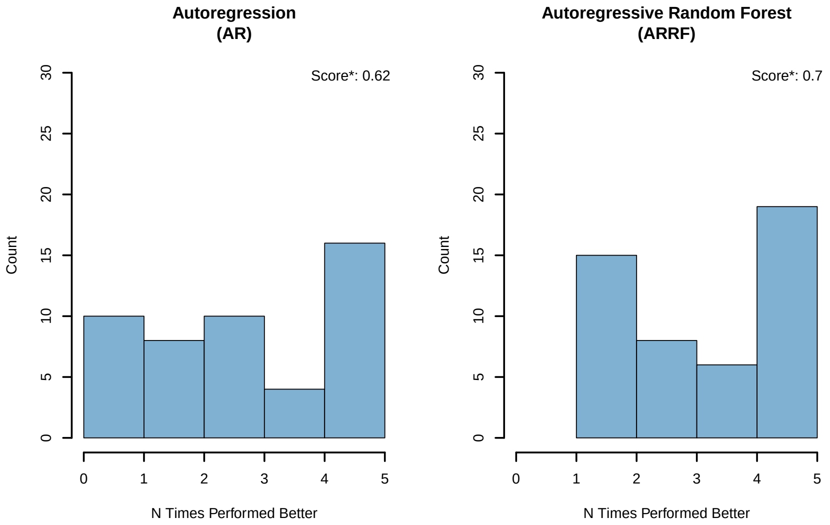
*Fraction of maximum possible score: 240 = 4*5*12

Simulated Data Horse Race



*Fraction of maximum possible score: 240 = 4*5*12

Histograms: Simulated Data Horse Race



*Fraction of maximum possible score: 240 = 4*5*12

Comments

Straightforward simulation story:

- AR does very well (reassuring)
- ARRF does better than AR
- RF does worst of all

Comments

Straightforward simulation story:

- AR does very well (reassuring)
- ARRF does better than AR
- RF does worst of all

Comments:

- Great econometric extension; great use of inference, interpretation, forecast improvement
- Would love discussion, visualization, on what gets selected from S_t
  - For example, what elements of S_t selected for FAARRF in Figure 5?
- Q: what mistakes will make my ARRF look like RF?
  - Promising use case: "first tool to grab"

The Consequences of the Covid-19 Job Losses by Gulyas and Pytka

Discussed by Nathan Palmer

November 11, 2020

Motivation: Examine Job Loss under a Pandemic

- Use a random forest + the universe of Austrian UI records, predict:
  - cumulative earnings loss
  - number of years lost employment
- Propose a wage-replacement policy via CART

Why a Random Forest?

- Treatment effect of job loss may be heterogeneous; RF can capture this heterogeneity

Why a Random Forest?

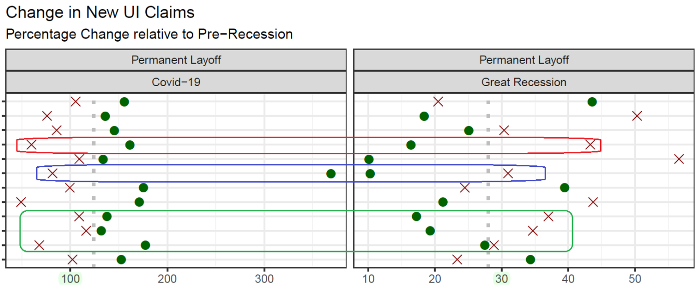
- Treatment effect of job loss may be heterogeneous; RF can capture this heterogeneity
- Typical RF chooses splits along variable z to minimize a loss

Why a Random Forest?

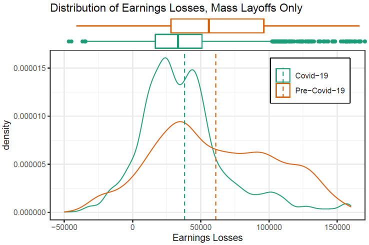
- Treatment effect of job loss may be heterogeneous; RF can capture this heterogeneity
- Typical RF chooses splits along variable z to minimize a loss
- Modified RF chooses splits to *maximize* difference between estimated income loss effects τ for individuals in splits $\underline{z}$, $\overline{z}$:

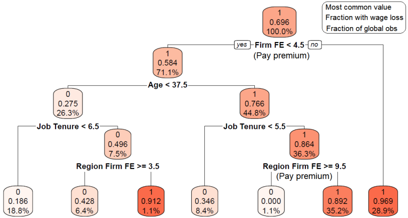
$$(\tau_{\underline{z}} - \tau_{\overline{z}})^2 \frac{n_{\underline{z}} n_{\overline{z}}}{N}$$

Why Heterogeneity in Effects Might Matter

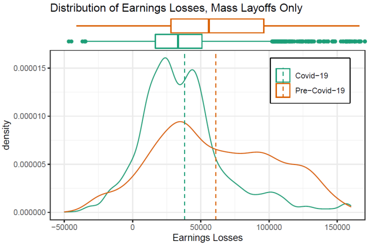


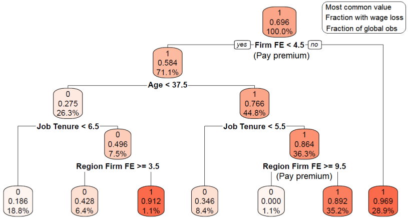
A Reason for Targeting Policy





A Reason for Targeting Policy





Policy in words:

1. Workers displaced from employers paying above the median
2. Workers with a relatively long job tenure, displaced from low paying firms, in regions with fewer good jobs on the market

Comments

1. RF is attractive, given potential heterogeneities in effects. Figure 1 is striking
   - Would like to see RF compared to a more traditional model (or more detailed discussion of why this is a bad idea)
2. The policy tree is very interesting – would like to see more regarding its specific construction
3. Would be interested in discussion of variance on estimates of τ
4. Do you have a measure of HH wealth?
   - May be useful to include – high income people may also be high-wealth