Discussion of:

Using The Yield Curve To Forecast Economic Growth

Parley Ruogu Yang

Michael Puglia

Federal Reserve Board of Governors

Modelling with Big Data and Machine Learning: Interpretability and Model Uncertainty

November 5, 2019

Why Are We Here?

- Specifically I mean, at this conference…
- From the call for papers:

> *Our events aim both to discuss recent developments and, crucially, focus on particular aspects of Big Data and ML approaches which are of increased interest to applied researchers. Two such aspects form the focus of this two-day conference. The first relates to a commonly cited weakness of ML methods when applied to economic problems and data, which is lack of* ***interpretability of ML model outputs****. This makes the adoption of such models difficult for economists who wish to have a more structural understanding of the underlying economic issues. The second, and related, focus is on the* ***estimation and/or calibration of the uncertainty associated with model outputs****. Both these matters have not received as much attention in the mainstream ML literature as economists would like to.*

- In short, we need new standards of comparison to discuss our ML results

Summary of Contributions

- To measure forecast performance in a macro time-series setting, use moving regressions, not regressions
  - In-sample R^2 is not a proxy for out-of-sample forecast performance
- Proposes the "p-statistic" as a measure of estimation uncertainty
  - Results appeal to intuition
- Novel methods for measuring time-series forecast performance
  - Optimizes performance over a lookback window *à la* local projections
  - Constructs voting ensembles for time series forecasters with memory

Related Literature

- Variance Adjudication vs. Forecast Performance
  - In-sample R^2 and out-of-sample MSE [Lopez de Prado, 2019]

$$y_t = \alpha + \sum_{i=1}^{I} \beta_i X_{t,i} + \sum_{j=1}^{J} \gamma_j Z_{t,j} + \varepsilon_t$$

- Cross-sectional vs. Time-series Data
  - i.i.d. vs serially correlated data [Varian, 2014]

- Modeling systems vs. targeting objects
  - VARs vs. Local Projections [Jorda, 2005]

Hypothesis Testing I

- Three tests of significance:

Test	H_0	H_1
1	$\beta = 0$	$\beta \neq 0$
2	$\beta_1 = -\beta_2$	$\beta_1 \neq -\beta_2$
3	$\varphi = 0$	$\varphi \neq 0$

- Define the "p-statistic" as:

$$p = \Phi\left(\frac{\widehat{\beta_1}}{sd\left(\widehat{\beta_1}\right)}\right)$$

- CDF-scaled version of a t-statistic

Hypothesis Testing II

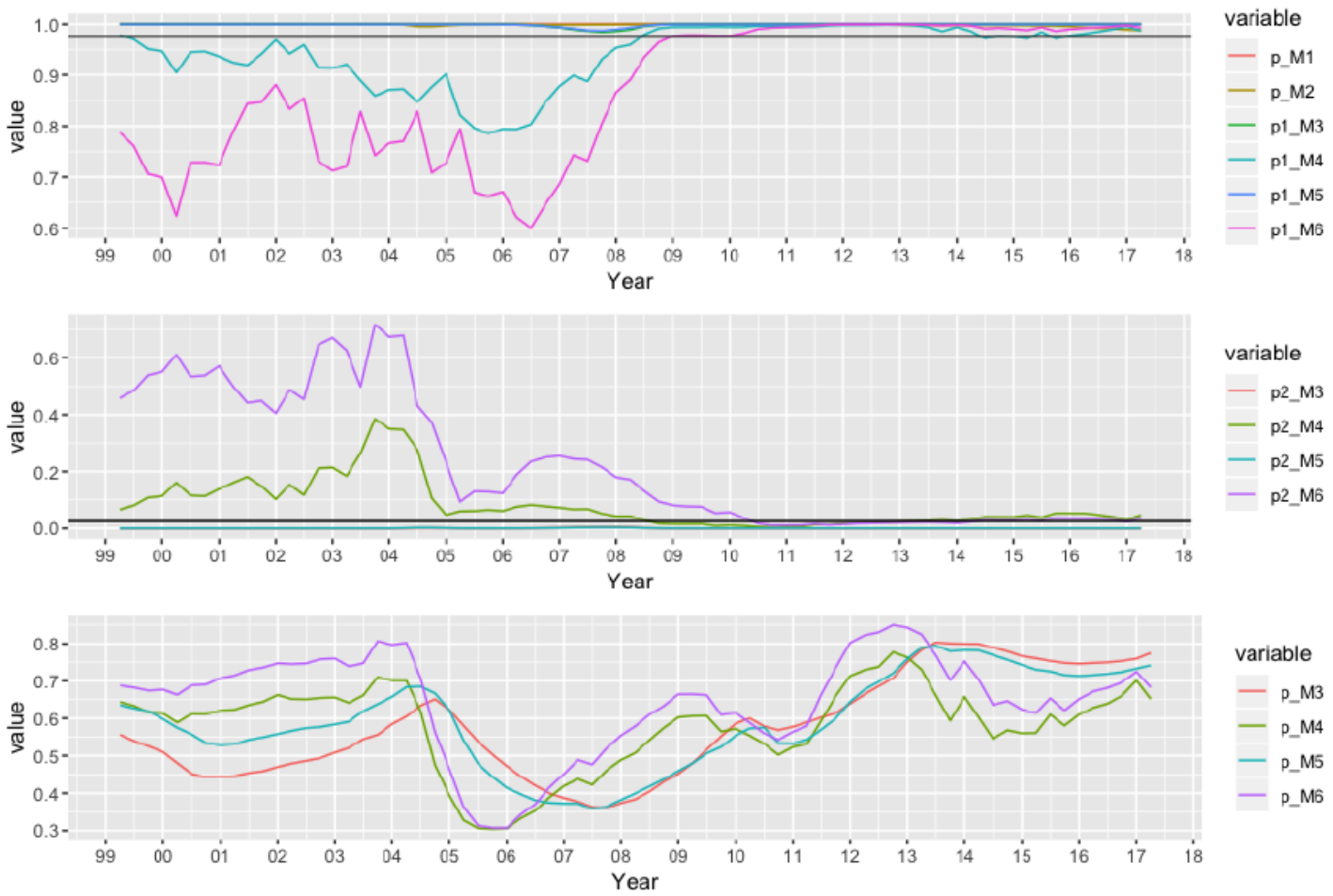


Figure 7: Top to bottom: p-statistic for the test on the first hypothesis for β in models 1 and 2 and β_1 in models 3 to 6, with the black line indicating 0.975; p-statistic for the test on the first hypothesis for β_2 in models 3 to 6, with the black line indicating 0.025; p-statistic for the test on the second hypothesis.

Hypothesis Testing III

- p-statistic passes sanity checks
  - The power of terms spreads to forecast GDP growth is time-varying (De Pace 2013)
  - The significance of term spreads at long horizons is absorbed by lagged GDP growth
  - GDP growth at long horizons is persistent

- It's a novel approach to unpacking the "black box"
  - If the black box contains model "semi-parameters," can we mine them for something?

- Interesting questions arise:
  - Asymptotics?
  - Bootstraps from the full sample of coefficients?
  - Quantifying joint uncertainty over all moving regressions → Moving regression table

- Other Thoughts:
  - What about Diebold-Mariano tests?
  - Do we use this for model selection, algorithm selection? To interpret tell a story?
  - How might extend this technique to other ML methods containing "semi-parameters" inside the algorithm?

Ensembles I

- The performances of the models vary over time

- Ensemble techniques can be used to aggregate outputs (i.e. stacking)
  - ML Classification → voting
  - ML Regression → averaging

- Paper proposes an ensemble rule amenable to the time-series setting
  - Demonstrates that forecast performance can be improved

Ensembles II

- For each horizon *k*:
  - For each time *t*:
    - Estimate and forecast over all models, windows and short/long rate combinations
    - Aggregate forecasts via a voting function min(*L*())
    - Call the selected output the *k*-horizon forecast at time *t*

$$L\left(\left\{g_{\tau,\tau+k}-\hat{g}_{\tau,\tau+k}\right\}_{w+k+1\leq\tau\leq t-k}\right)=\sum_{w+k+1\leq\tau\leq t-k}\lambda^{\,t-k-\tau}l\left(\left|g_{\tau,\tau+k}-\hat{g}_{\tau,\tau+k}\right|\right)$$

ι	**Ensemble**	**Voting Function**	λ
7	No lookback	min(Single period loss)	$\lambda \to 0^{+}$
8	Full-history lookback	min(Cumulative loss)	1
9	Filtered lookback	min(Cumulative loss w/ decay)	(0,1]

Ensembles III

- Work highlights that techniques developed for work on cross-sectional data are useful, but need to be amended for application on time-series data sets
  - Serial correlation
  - Time ordering
  - Structural breaks

- Alternative specifications for voting loss function *l*() explored
  - Existing loss functions (L1, L2, etc.) generally work pretty well
  - Replacing the curse of dimensionality with the curse of hyperparameters?

Future Directions

- Work in the line of Local Projections might be fruitful
  - Do LPs nest these methods?
  - Can something be mined from or contributed to the LP literature?
  - Helps make the methods more accessible to economists
- Work in the line of ML cross-validation might be fruitful
  - Using the language in addition to the methods of ML would make paper more accessible as well
- Other algorithms?
  - Lasso/Ridge regression (so as to include more features)
  - Neural networks (to capture non-linearities)
- Other features?
  - Credit and financial conditions are likely strong predictors
  - Leading Economics Indicators and macro indexes

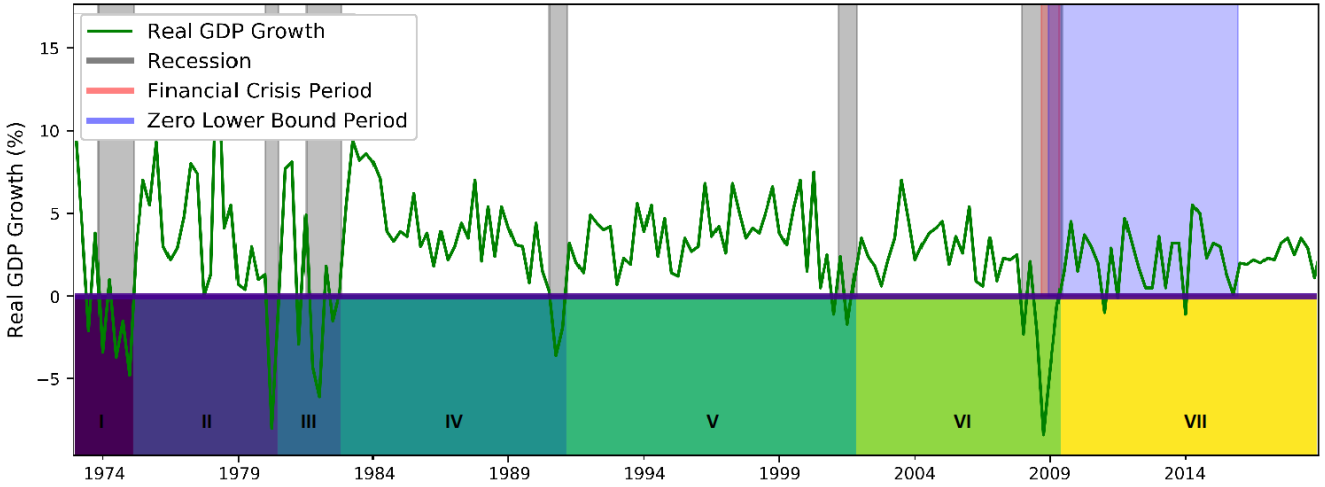
Recap

- A lot of promising new ideas in this paper to be explored
  - Sits at the intersection of a few strands of research
- Points on how to properly measure forecast performance well taken
  - We should all take heed
- Creative use of model “semi-parameters” to measure uncertainty
- Novel method for ensembling time-series forecasters
- I enjoyed the time I spent with the paper and thank you!

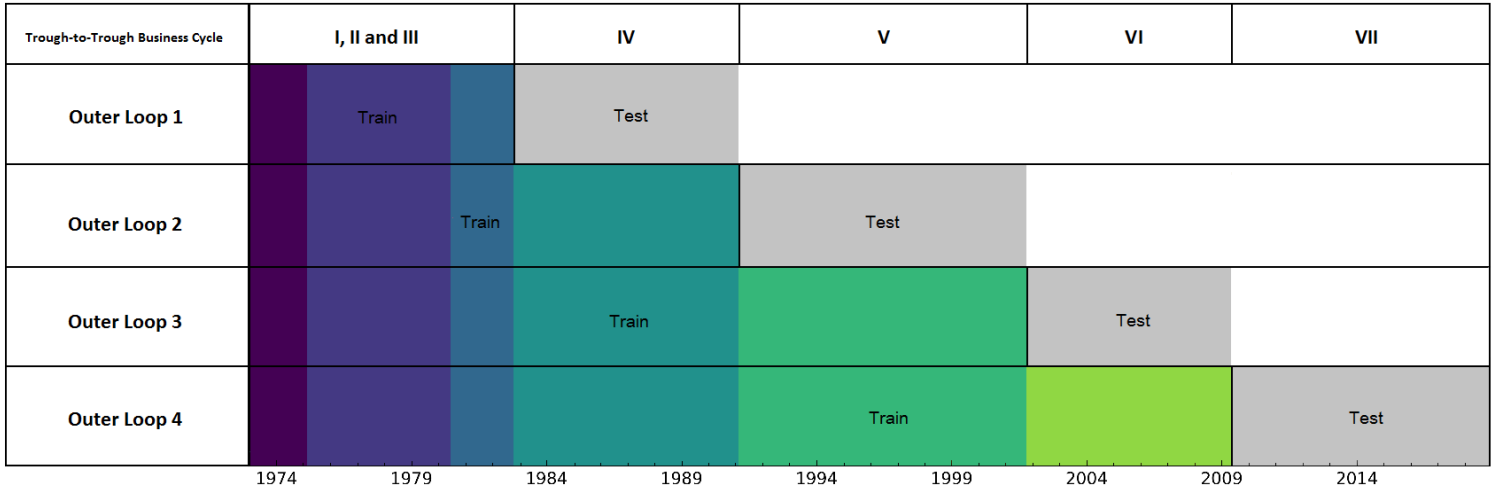
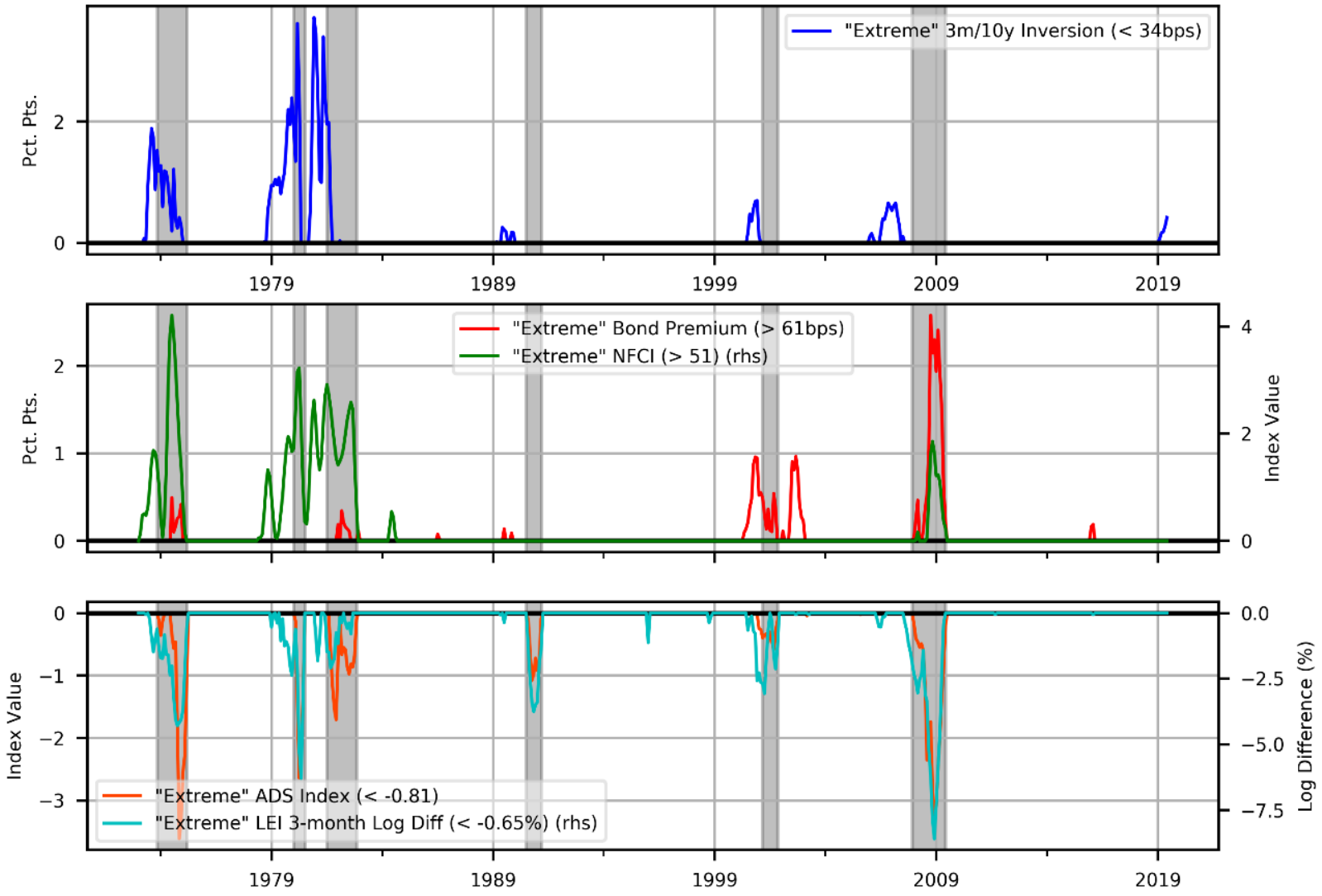
Appendix

Data and Forecasts

Real Output Growth and Recession



Nested Time-Series Cross-Validation



Term Spreads and Credit Conditions