A Comprehensive Look at the Empirical Performance of Forecast Error Prediction

Kenneth Eva and Fabian Winkler

ASSA 2023 Annual Meeting
January 6, 2023

The views expressed in this presentation are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or any other person associated with the Federal Reserve System.

Motivation

- Unresolved debate in finance: Are aggregate stock returns predictable?

Motivation

- Unresolved debate in finance: Are aggregate stock returns predictable?
- Simple method:
  - Regress future returns on a predictor (e.g. P/D)
  - Frequently find strong statistical significance
  - But an in-sample predictive regression can see the future

Motivation

- Unresolved debate in finance: Are aggregate stock returns predictable?
- Simple method:
  - Regress future returns on a predictor (e.g. P/D)
  - Frequently find strong statistical significance
  - But an in-sample predictive regression can see the future
- Goyal and Welch (2007): Can we also forecast returns out of sample?
  - *No.* The out-of-sample performance is dismal.
  - The models would not have helped an investor to make a profit.

Motivation

- Unresolved debate in finance: Are aggregate stock returns predictable?
- Simple method:
  - Regress future returns on a predictor (e.g. P/D)
  - Frequently find strong statistical significance
  - But an in-sample predictive regression can see the future
- Goyal and Welch (2007): Can we also forecast returns out of sample?
  - *No.* The out-of-sample performance is dismal.
  - The models would not have helped an investor to make a profit.

- Return predictability is weak at best.

This paper

- Unresolved debate in macroeconomics: Are expectations rational?

This paper

- Unresolved debate in macroeconomics: Are expectations rational?
- Simple method:
  - Regress forecast errors on a predictor (e.g. forecast revisions)
  - Frequently find strong statistical significance
  - But an in-sample predictive regression can see the future

This paper

- Unresolved debate in macroeconomics: Are expectations rational?
- Simple method:
  - Regress forecast errors on a predictor (e.g. forecast revisions)
  - Frequently find strong statistical significance
  - But an in-sample predictive regression can see the future
- This paper: Can we also predict forecast errors out of sample?
  - *Mostly no.* The out-of-sample performance is mostly dismal.
  - Most models would not have helped a forecaster to improve their forecasts.
  - But there are important exceptions.

This paper

- Unresolved debate in macroeconomics: Are expectations rational?
- Simple method:
  - Regress forecast errors on a predictor (e.g. forecast revisions)
  - Frequently find strong statistical significance
  - But an in-sample predictive regression can see the future
- This paper: Can we also predict forecast errors out of sample?
  - *Mostly no.* The out-of-sample performance is mostly dismal.
  - Most models would not have helped a forecaster to improve their forecasts.
  - But there are important exceptions.

- Most biases are weak at best. Research should focus on those that are strong.

Models of expectations

A person predicts y_{t+h} as $\hat{y}_{t+h|t}$ at time t. Their information set is $\mathcal{F}_t$. Consider:

$$y_{t+h} - \hat{y}_{t+h|t} = \beta x_t + e_{t+h}, \mathbb{E}\left[e_{t+h} \mid \mathcal{F}_t\right] = 0$$

- Rational expectations model: $\beta = 0$ for any x_t in $\mathcal{F}_t$.
- Behavioral models: $\beta \neq 0$ for some x_t in $\mathcal{F}_t$.

Measure of predictive performance

- In-sample (IS) behavioral forecast:

$$\hat{y}^*_{t+h|t,IS} = \hat{y}_{t+h|t} + \hat{\beta}_T x_t$$

- Out-of-sample (OOS) behavioral forecast:

$$\hat{y}^*_{t+h|t,OOS} = \hat{y}_{t+h|t} + \hat{\beta}_t x_t$$

where $\hat{\beta}_t$ is the OLS coefficient estimated with data up to time t.

Measure of predictive performance

- In-sample (IS) behavioral forecast:

$$\hat{y}^*_{t+h|t,IS} = \hat{y}_{t+h|t} + \hat{\beta}_T x_t$$

- Out-of-sample (OOS) behavioral forecast:

$$\hat{y}^*_{t+h|t,OOS} = \hat{y}_{t+h|t} + \hat{\beta}_t x_t$$

where $\hat{\beta}_t$ is the OLS coefficient estimated with data up to time t.

Difference in cumulative sum of squared errors (SSE)

$$\Delta_{tj} = \sum_{s=T_0}^{t} \left[\left(y_{t+h} - \hat{y}_{t+h|t,j} \right)^2 - \left(y_{t+h} - \hat{y}^*_{t+h|t,j} \right)^2 \right], \; j = IS, OOS$$

An increase in Δ_{tj} means the behavioral model performs better.

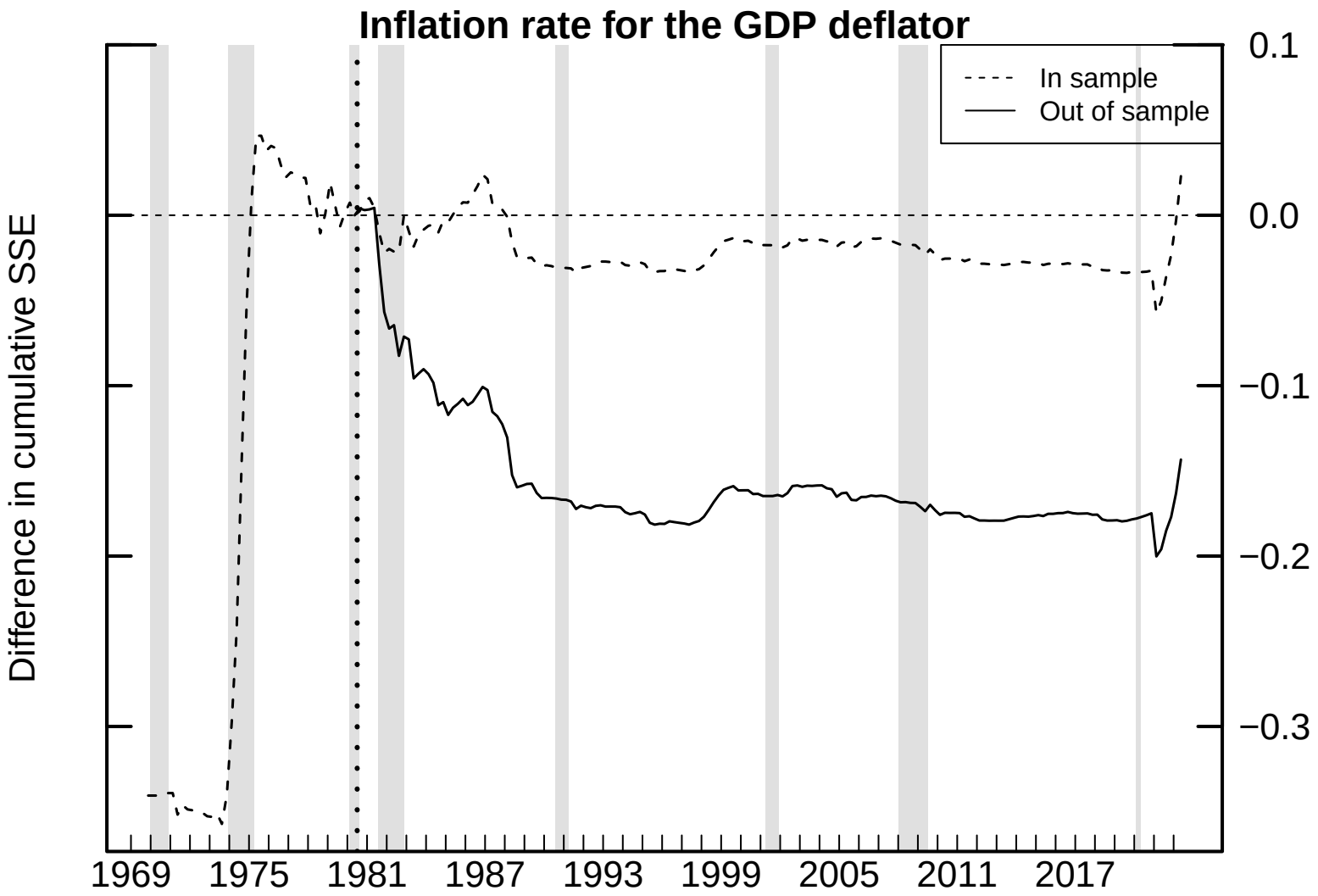
Underreaction of aggregate forecasts

- Coibion and Gorodnichenko (2015, "CG") propose:

$$y_{t+h} - \hat{y}_{t+h|t} = \alpha + \beta \left(\hat{y}_{t+h|t} - \hat{y}_{t+h|t-1} \right) + e_{t+h}$$

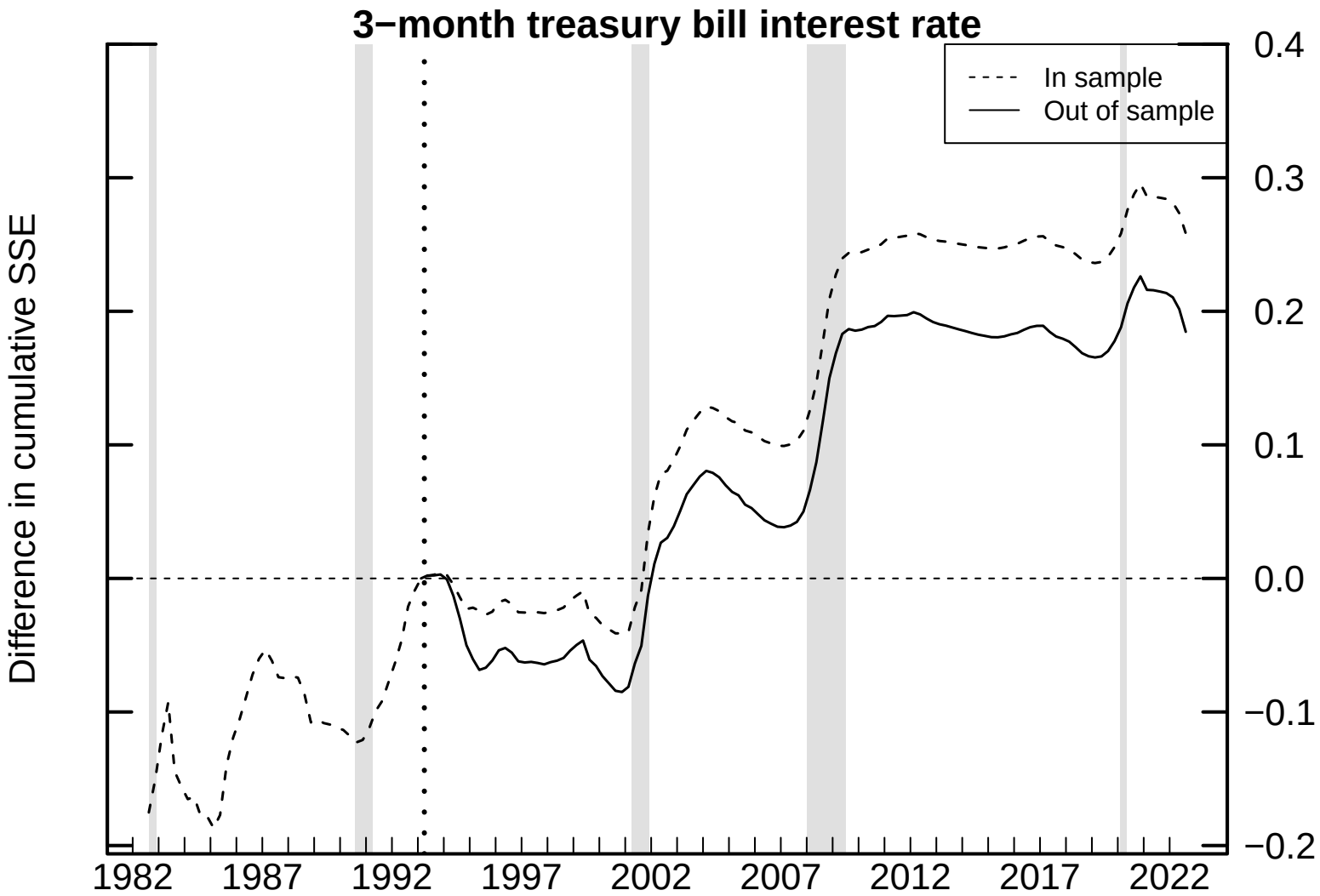
- Evaluate this for all variables in the Survey of Professional Forecasters
- 3-quarter ahead predictions
- Extend their data to 2021Q4

CG: Inflation (GDP deflator)



CG: GDP growth

CG: Short-term interest rate



CG: Across SPF variables, weak predictability on average

	$\Delta_{T,IS}$	$\Delta_{T,OOS}$
Inflation (deflator)	0.023*	-0.143
Inflation (CPI)	-0.008	-0.083
Real GDP	0.003*	-0.145
Industrial Production	0.119**	0.032**
Nominal GDP	0.037*	-0.157
Unemployment rate	0.003	-0.274
Consumption	0.020	-0.049
Non-residential inv.	0.014	-0.125
Residential inv.	0.027	-0.088
Federal govt.	-0.001	-0.075
Non-federal govt.	0.026	-0.134
Housing starts	0.132***	0.021**
3-month yield	0.258***	0.185***
10-year yield	0.450***	0.115**
AAA yield	0.279***	0.220***
Term spread	0.102	0.014
Bond spread	0.022	-0.126

- We bootstrap the distribution of $\Delta_{T,IS}$, $\Delta_{T,OOS}$ under the null of RE for critical values (***, **, *: 1, 5, 10 percent)
- Predictability is weak even in sample
- However, RE is solidly rejected for interest rates
- The results are similar if we regress only on a constant

Overreaction of individual forecasts

- Bordalo et al. (2020, BGMS) apply the CG model to individual forecasts:

$$y_{t+h} - \hat{y}_{t+h|it} = \alpha + \beta \left(\hat{y}_{t+h|it} - \hat{y}_{t+h|it-1} \right) + e_{it+h}$$

- Repeat out-of-sample tests, summing squared errors across individuals
- Results:
  - Still no predictability despite cross-sectional data dimension
  - Interest rates are still the only variables that are OOS-predictable

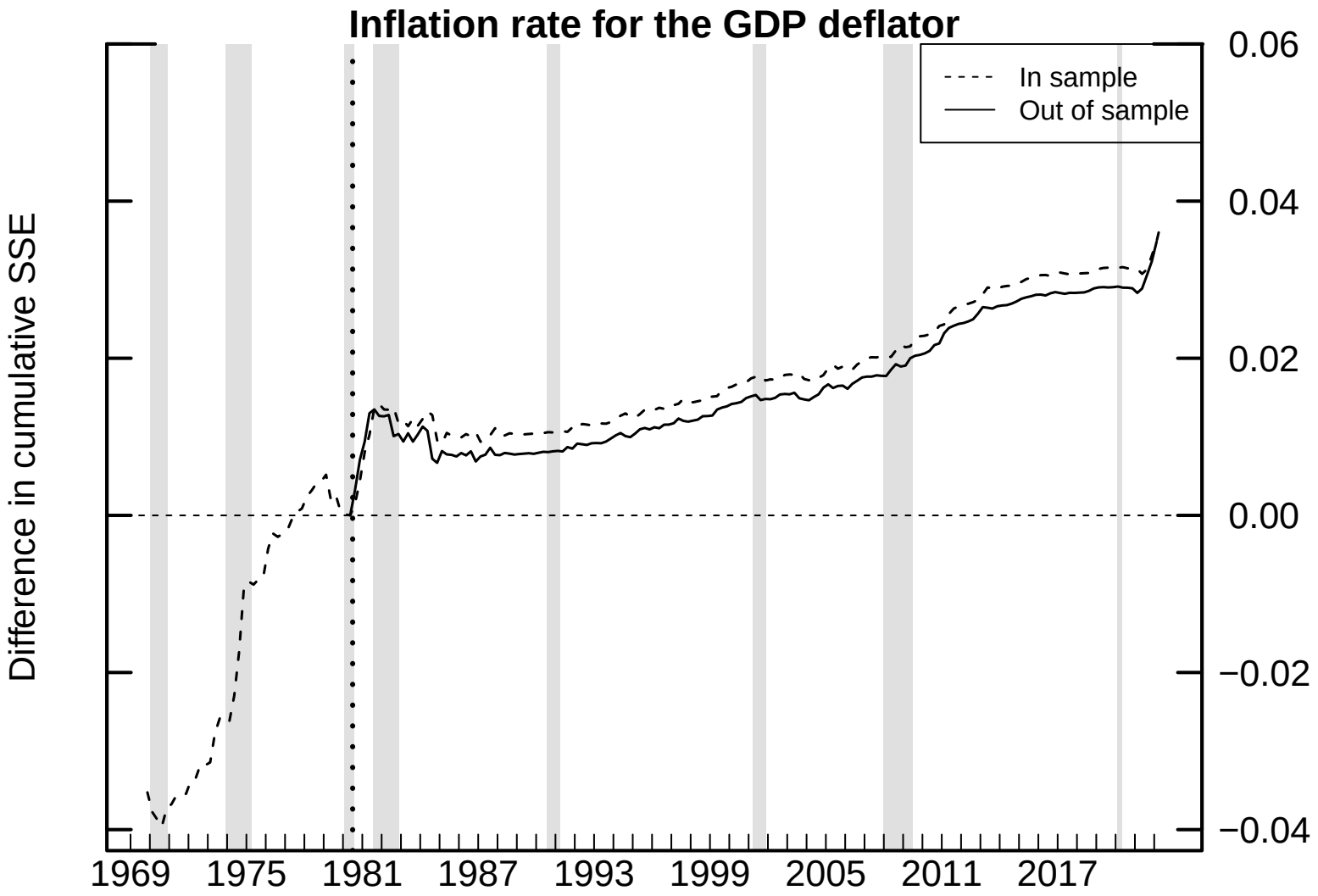
Forecast combination bias

- The advantages of forecast combination are well documented in the forecasting literature (e.g. Timmermann, 2006)
- This implies that forecasters could do better by averaging their forecasts
- Here we propose:

$$y_{t+h} - \hat{y}_{t+h|it} = \beta \left(\bar{y}_{t+h-1|t-1} - \hat{y}_{t+h-1|it-1} \right) + e_{it+h}$$

  - Combination is based only on lagged values, which are known to forecasters in real time

Combination bias: Inflation (GDP deflator)



Forecast combination

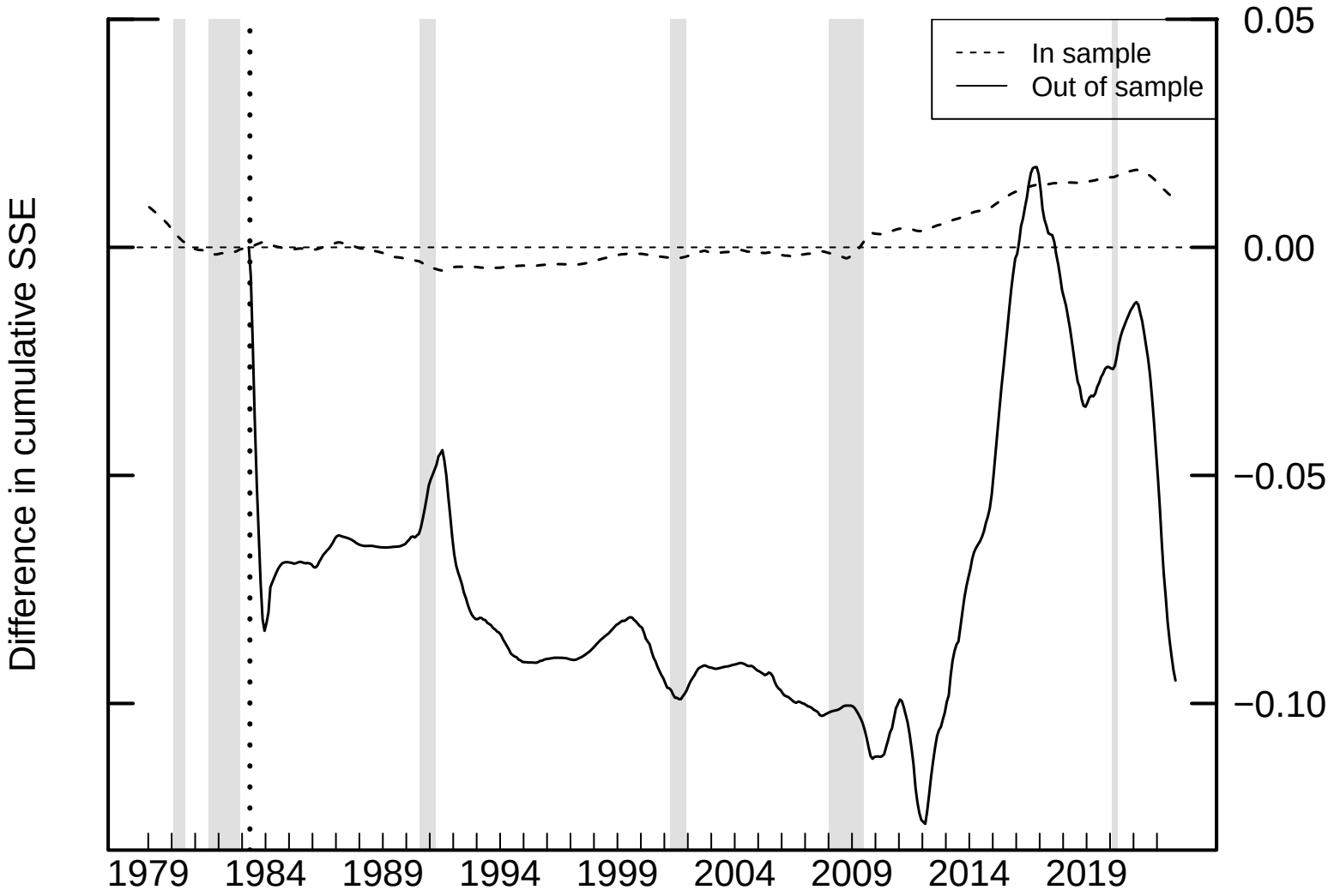
	$\Delta_{T,IS}$	$\Delta_{T,OOS}$
Inflation (deflator)	0.037^{***}	0.035^{***}
Inflation (CPI)	0.014^{***}	0.013^{***}
Real GDP	0.005^{**}	0.005^{***}
Industrial Production	0.011^{***}	0.010^{***}
Nominal GDP	0.006^{*}	0.006^{***}
Unemployment rate	0.018^{***}	0.012^{***}
Consumption	0.002	0.001^{**}
Non-residential inv.	0.023^{***}	0.021^{***}
Residential inv.	0.049^{***}	0.043^{***}
Federal govt.	0.018^{***}	0.011^{***}
Non-federal govt.	0.038^{***}	0.029^{***}
Housing starts	0.105^{***}	0.103^{***}
3-month yield	0.057^{***}	0.056^{***}
10-year yield	0.055^{***}	0.052^{***}
AAA yield	0.102^{***}	0.100^{***}
Term spread	0.058^{***}	0.057^{***}
Bond spread	0.099^{***}	0.091^{***}

- Forecast combination bias is highly significant in and out of sample
- Strategic interactions between forecasters (e.g. Gemmi and Valchev, 2021) could be an explanation

Households

- Households are less sophisticated forecasters, so our OOS tests may reject the null more easily
- Use median inflation expectations in the Michigan survey
- Focus on mean bias (e.g. d'Acunto et al., 2022), rolling window regressions

Mean bias in Michigan inflation expectations



Additional tests

- We examine other tests:
  - Simple mean bias
  - Autocorrelation of forecast errrors
  - Mincer-Zarnovitz
  - Nordhaus
- Our results are robust to different specifications:d
  - Different forecast horizons
  - Rolling-window instead of recursive regressions
  - Data transformations (levels, logs, cumulative or per-period growth rates)
- Using BlueChip data produces similar results.
  - Interest rates across the yield curve show solid mean bias.

Conclusion

- Most documented biases in macroeconomic expectations do not survive OOS tests.
- The models could not have been used to make better predictions.
- But some biases hold up out of sample, in particular:
  - Mean bias in interest rate expectations
  - Forecast combination bias
- Efforts to model expectation formation should focus on the latter biases.