A Comprehensive Empirical Evaluation of Biases in Expectation Formation

Kenneth Eva and Fabian Winkler

NBER Summer Institute
July 14, 2023

Disclaimer

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

- Debate in finance: Are aggregate stock returns predictable?

Motivation

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

Motivation

- Debate in finance: Are aggregate stock returns predictable?
- Simple method:
  - Regress future aggregate returns on a predictor (e.g. P/D)
  - Find statistical significance
  - But an in-sample predictive regression can see the future
- Welch and Goyal (2008): Can we also forecast returns out of sample?
  - *No.* The out-of-sample performance of the models in the literature is typically weak.
  - The models would not have helped an investor make a profit.

Motivation

- Debate in finance: Are aggregate stock returns predictable?
- Simple method:
  - Regress future aggregate returns on a predictor (e.g. P/D)
  - Find statistical significance
  - But an in-sample predictive regression can see the future
- Welch and Goyal (2008): Can we also forecast returns out of sample?
  - *No.* The out-of-sample performance of the models in the literature is typically weak.
  - The models would not have helped an investor make a profit.

- Return predictability is weak at best.

This paper

- Debate in macroeconomics: Are expectations rational?

This paper

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

This paper

- Debate in macroeconomics: Are expectations rational?
- Simple method:
  - Regress forecast errors on a predictor (e.g. forecast revisions)
  - Find 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 typically weak.
  - The models would not have helped a forecaster improve their forecasts.
  - *But* there are a number of important exceptions.

This paper

- Debate in macroeconomics: Are expectations rational?
- Simple method:
  - Regress forecast errors on a predictor (e.g. forecast revisions)
  - Find 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 typically weak.
  - The models would not have helped a forecaster improve their forecasts.
  - *But* there are a number of important exceptions.

- Research should focus on models of biases that are useful to improve prediction.

Related literature

- **Defenses of rationality:** Andolfatto, Hendry and Moran (2008); Elliott, Komunjer and Timmermann (2008); Farmer, Nakamura and Steinsson (2021); Hajdini and Kurmann (2022), ...
- **IS vs OOS:** Pearce (1987), Bonham and Dacy (1991), Shmueli (2010), Bianchi, Ludvigsson and Ma (2022), ...
- **Predictability in finance:** Meese and Rogoff (1983), Lettau and van Nieuwerburgh (2007), Welch and Goyal (2008), Campbell and Thompson (2008), Farmer and Timmermann (2023) ...

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:

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

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 $\mathcal{F}_t$.
  - Forecast errors are *unpredictable*.
  - Forecasts are *optimal*: $\mathbb{E}\left(y_{t+h} - \hat{y}_{t+h|t}\right)^2$ is minimal among all forecasts in $\mathcal{F}_t$.
- Behavioral models: $\beta \neq 0$ for some x_t in $\mathcal{F}_t$.
  - Forecast errors are predictable.
  - Forecasts are suboptimal: $\mathbb{E}\left(y_{t+h} - \left[\hat{y}_{t+h|t} + \beta x_t\right]\right)^2 < \mathbb{E}\left(y_{t+h} - \hat{y}_{t+h|t}\right)^2$.

Measure of predictive performance

- In-sample (IS) bias-adjusted forecast:

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

- Out-of-sample (OOS) bias-adjusted forecast:

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

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

Measure of predictive performance

- In-sample (IS) bias-adjusted forecast:

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

- Out-of-sample (OOS) bias-adjusted forecast:

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

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

Difference in cumulative sum of squared errors (SSE)

$$\Delta SSE_t = \frac{\sum_{s=T_0}^{t} \left[\left(y_{t+h} - \hat{y}_{t+h|t} \right)^2 - \left(y_{t+h} - \hat{y}^{*}_{t+h|t} \right)^2 \right]}{\sum_{s=T_0}^{T} \left(y_{t+h} - \hat{y}_{t+h|t} \right)^2}$$

An increase in ΔSSE_t means the behavioral model performs better.

- We use a bootstrap for critical values.

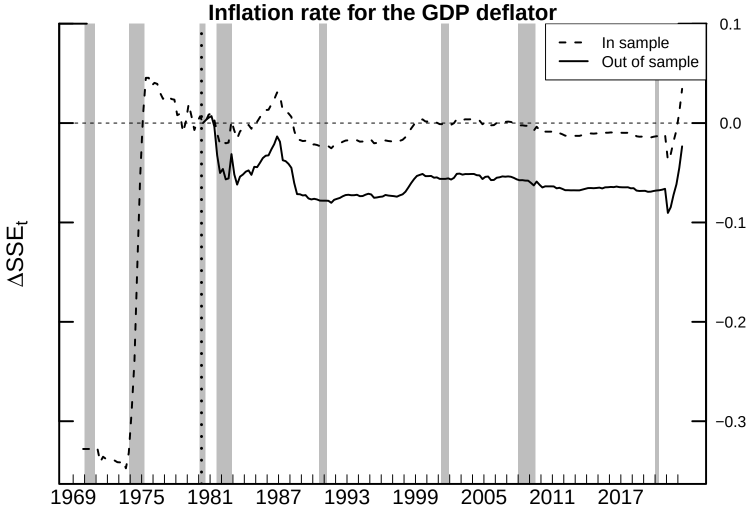
Consensus professional forecasts

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

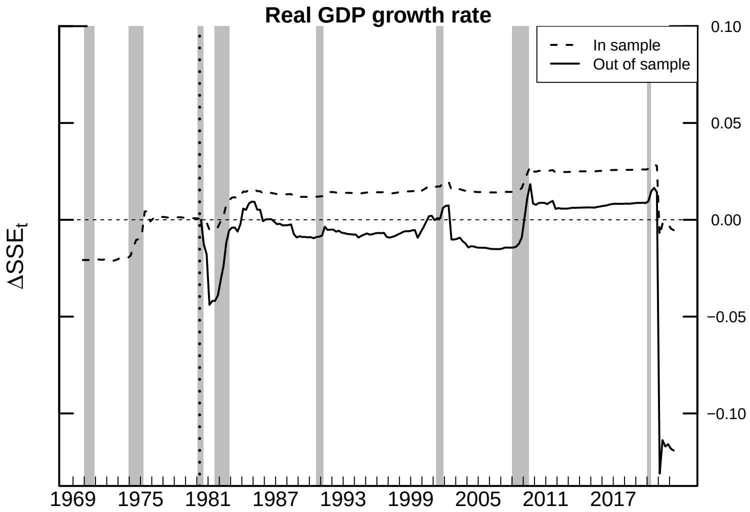
$$y_{t+h} - \hat{y}_{t+h|t} = \beta \left(\hat{y}_{t+h|t} - \hat{y}_{t+h|t-1} \right) + e_{t+h}$$

- Evaluate this for macro variables in the Survey of Professional Forecasters
  and interest rates in BlueChip
  - 3-quarter ahead predictions
  - Extend CG data to 2022Q2
  - Omit the constant to improve power

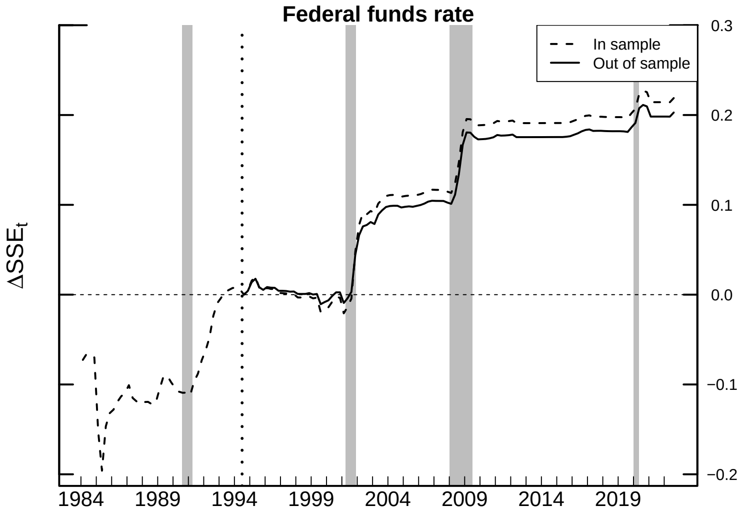
Coibion and Gorodnichenko (2015) model



Coibion and Gorodnichenko (2015) model



Coibion and Gorodnichenko (2015) model



Macro variables: little OOS predictability

ΔSSE_T OOS	(1) CG	(2) Mean bias	(3) Autocorrelation	(4) Mincer-Zarnowitz
Inflation (deflator)	-0.023	-0.416	0.122***	-0.476
Inflation (CPI)	0.015**	-0.073	-0.040	-0.043
Real GDP	-0.119	-0.040	0.005	-0.153
Industrial Production	0.035***	0.007	-0.008	0.008
Nominal GDP	-0.101	-0.029	-0.028	-0.052
Unemployment rate	-0.251	-0.029	-0.032	-0.041
Consumption	-0.031	-0.011	-0.172	0.010
Non-residential inv.	-0.061	-0.065	-0.035	-0.112
Residential inv.	-0.026	-0.063	-0.007	-0.306
Federal govt.	-0.012	-0.062	-0.035	0.027
Non-federal govt.	-0.027	-0.105	0.026	-0.103
Housing starts	0.047***	-0.043	0.117***	-0.100

Interest rates: stable mean bias

	(1)	(2)	(3)	(4)
ΔSSE_T OOS	CG	Mean bias	Autocorrelation	Mincer-Zarnowitz
Federal funds rate	0.203***	0.061**	0.047**	-0.007
3-month yield	0.181***	0.129***	0.089***	0.089**
6-month yield	0.211***	0.157***	0.129***	0.115**
1-year yield	0.196***	0.135**	0.037	0.043
2-year yield	0.112***	0.154***	-0.022	0.046**
10-year yield	-0.001	0.295***	-0.046	0.054**
Aaa yield	0.067***	0.225***	0.016	-0.125
Baa yield	0.052**	0.402***	0.108**	0.238***
1y-3m spread	-0.009	-0.118	0.011	-0.148
10y-2y spread	0.056***	-0.061	-0.022	-0.057
Aaa-Baa spread	-0.011	-0.139	0.071	0.036

Models for individual forecasts

- Forecast errors at the individual level may be more easily predictable
- Bordalo et al. (2020, BGMS) apply the CG model to individual forecasts:

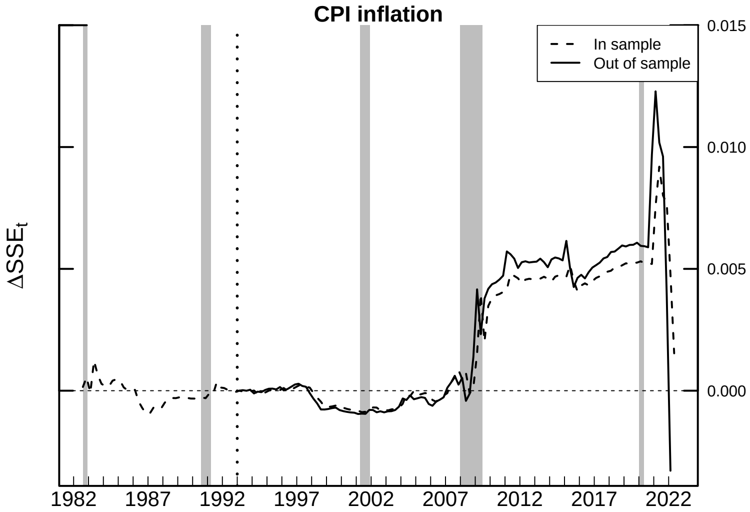
$$y_{t+h} - \hat{y}_{t+h|it} = \beta \left(\hat{y}_{t+h|it} - \hat{y}_{t+h|it-1} \right) + e_{it+h}$$

- We also test the Kohlhas-Walther (2021, KW), autocorrelation, Mincer-Zarnovitz (MC), Nordhaus models
- We also propose a test based on forecast combination (e.g. Timmermann, 2006):

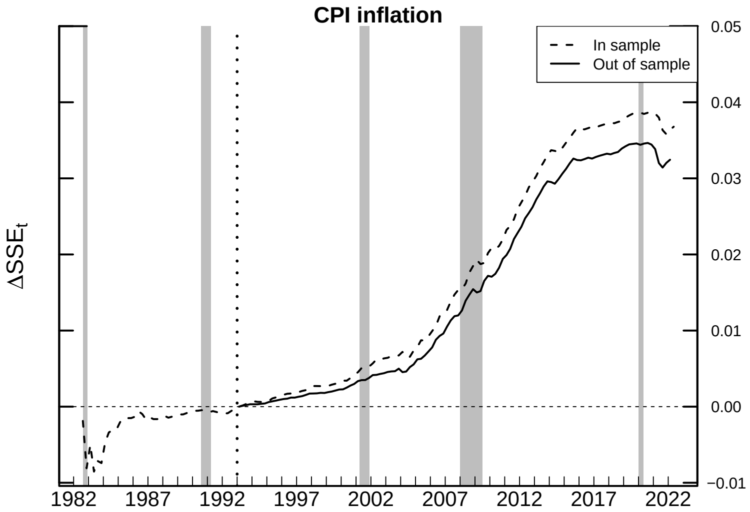
$$y_{t+h} - \hat{y}_{t+h|it} = \beta \left(\bar{y}_{t+h-1|t-1} - \hat{y}_{t+h-1|it-1} \right) + e_{it+h}$$

  - Combination is based on lagged values known to forecasters.

Bordalo et al. (2020) model



Forecast combination model



Macro variables: Forecast combination works best

	(1)	(2)	(3)	(4)	(5)
ΔSSE_T OOS	BGMS	Autocorr	MZ	KW	Forecast combination
Inflation (deflator)	0.007***	0.114***	-0.513	-0.323	0.141***
Inflation (CPI)	-0.003	-0.036	0.043**	-0.036	0.034***
Real GDP	0.023***	-0.029	0.042**	0.013	0.060***
Industrial Production	-0.010	-0.016	0.036**	0.002	0.049***
Nominal GDP	0.011***	-0.060	-0.071	-0.061	0.061***
Unemployment rate	-0.102	-0.048	0.010	0.024	0.013***
Consumption	0.062***	-0.166	-0.003	-0.090	0.036***
Non-residential inv.	-0.018	-0.035	-0.055	-0.102	0.064***
Residential inv.	-0.018	0.046***	-0.037	-0.092	0.108***
Federal govt.	0.090***	0.063***	-0.009	-0.027	0.148***
Non-federal govt.	0.140***	0.039***	0.126***	-0.048	0.206***
Housing starts	0.004	0.208***	-0.099	-0.054	0.108***

Interest rates: Less predictability at individual level

ΔSSE_T OOS	(1) BGMS	(2) Autocorr	(3) MZ	(4) KW	(5) Forecast combination
Federal funds rate	0.080***	0.075***	-0.070	0.035**	0.035***
3-month yield	0.071***	0.132***	0.025**	0.123***	0.043***
6-month yield	0.108***	0.145***	0.091***	0.132***	0.031***
1-year yield	0.085***	0.070***	-0.007	0.039**	0.041***
2-year yield	0.040***	0.032***	-0.015	0.04**	0.044***
10-year yield	-0.003	-0.016	0.008	0.099***	0.069***
Aaa yield	0.000	0.005	-0.181	-0.160	0.068***
Baa yield	0.001	0.197***	0.012	0.248***	0.130***
1y-3m spread	0.084***	-0.025	0.232***	-0.233	0.087***
10y-2y spread	-0.002	-0.004	0.029**	0.002	0.053***
Aaa-Baa spread	0.008	-0.084	0.513***	0.066	0.067***

Households

- Households are less sophisticated forecasters, so our OOS tests may reject the null more easily
- We use the Michigan survey (1978M1–2022M12) and the SCE from FRBNY (2013M6-2022M12)
- We test average and median expectations as well as individual expectations of (CPI) inflation.

Household tests

	(1) Mean bias	(2) Revisions	(3) Autocorrelation	(4) Mincer-Zarnovitz	(5) Forecast combination
Michigan avg.	0.028**	0.002	0.063***	0.008**	–
Michigan med.	-0.226	-0.005	-0.043	-0.252	–
Michigan ind.	-0.014	–	–	0.617***	0.889***
SCE avg.	-0.104	-0.002	0.032	-1.491	–
SCE med.	-0.562	0.050**	-0.061	-2.139	–
SCE ind.	-0.113	–	–	0.632***	0.799***

Interpretation

Why do so many models not survive our OOS tests?

1. Low power: The null is harder to reject using OOS tests because of estimation error
   - Preliminary results indicate that the power of our tests is quite good.
   - Raising the bar is useful to distinguish stronger and weaker biases.
2. Time-varying parameters: The predictive relationship exists but is unstable.
   - Indeed, the real-time estimated parameters vary considerably over time where our tests fail.

Conclusion

- Many documented biases in macroeconomic expectations do not pass OOS tests.
- The models could not have been used to make better predictions in real time.
- But some biases hold up *very well* out of sample:
  - Mean bias in interest rate forecasts
  - Forecast combination bias
- These biases should receive larger attention in future research.