Belief Distortions and Macroeconomic Fluctuations

By FRANCESCO BIANCHI, SYDNEY C. LUDVIGSON, AND SAI MA*

This document contains supplementary material for the paper "Belief Distortions and Macroeconomic Fluctuations."
JEL: E7, E27, E32, E17, G4.

Keywords: beliefs, biases, machine learning, expectations

ONLINE APPENDIX

Additional Results

This section contains additional results not reported in the main text. Table A1 reports results for the mean of each survey. Table A2 shows results for a machine specification that uses k-fold cross validation. Table A3 shows results when we use the historical data based on the end-of-sample vintage, rather than using real-time data. Table A4 shows MSE ratios for the median GDP growth forecast from each survey, when we shut off the switching component of the algorithm. The SPF median MSE ratio for GDP growth forecasts over the period 1995:Q1-2018:Q2 rises to 0.96 from the baseline 0.85, mostly because the machine would have had no ability to see the coming downturn in the 2001 recession. The switching option is completely irrelevant for the forecasting performance over more recent subsamples, since the only triggered switch in the forecast sample occurs immediately prior to the 2001 recession. For the full sample results, the MSE ratio of 0.96, obtained by shutting down the machine's switching option, fails to accurately reflect all the information that was available in real time to survey respondents and pertinent to the accuracy of the foreast. This machine model is thus not a good benchmark against which to measure any forecaster bias and in our view it does not give an accurate account of it.

* Bianchi: Department of Economics, Johns Hopkins University, Wyman Park Building 613, 3100 Wyman Park, Baltimore, MD 21211, Department of Economics, Duke University, Centre for Economic and Policy Research (CEPR), and National Bureau of Economic Research (NBER) Ludvigson: Department of Economics, New York University, 19 W.4th St, 6th Floor, New York, NY 10012, Center for Economic and Policy Research (CEPR), and National Bureau of Economic Research (NBER) (sydney.ludvigson@nyu.edu); Ma: Federal Reserve Board of Governors, C Ave & 20th Street NW, Washington, DC 20551. (sai.ma@frb.gov). The views expressed are those of the authors and do not necessarily reflect those of the Federal Reserve Board or the Federal Reserve System.

Table A1— Machine Learning versus Survey Forecasts: Mean Forecasts

Machine Learning versus Survey Forecasts

ML: $y_{j,t+h} = \alpha_{jh}^{(\mu)} + \beta_{jh\mathbb{F}}^{(\mu)} \mathbb{F}_t^{(\mu)} \left[y_{j,t+h} \right] + \mathbf{B}_{jh\mathcal{Z}}^{(\mu)} \mathcal{Z}_{jt} + \epsilon_{jt+h}$

Mean Forecast

	Inflation			**GDP**	
	SPF	SOC	BC	SPF	BC
$MSE_{\mathbb{E}}/MSE_{\mu}$	0.95	0.42	0.84	0.89	0.76
OOS R^2	0.05	0.58	0.16	0.11	0.24

Machine v.s. survey mean-square-forecast errors. $MSE_{\mathbb{E}}$ and $MSE_{\mathbb{F}}$ denote the machine and survey mean-squared-forecast-errors, respectively, for 4-quarter-ahead forecasts, averaged over the evaluation sample. The out-of-sample Rsquared, OOS R^2, is defined as 1-$MSE_{\mathbb{E}}/MSE_{\mathbb{F}}$. The vintage of observations on the variable being forecast is the one available four quarters after the period being forecast. The evaluation period for the Survey of Professional Forecasters (SPF) is 1995:Q1 to 2018:Q2; for the Michigan Survey of Consumers (SOC) is 1996:Q4 to 2018:Q2; and for the Bluechip (BC) survey is 1997:Q3 to 2018:Q2.

Table A2— Machine Learning: K-Fold Estimation

Median SPF Inflation

Center	Baseline	10-Fold , $\hat{T}_E$	10 Fold, Expanding Window
$MSE_{\mathbb{E}}/MSE_{\mathbb{F}}$	0.85	1.04	1.10

This table reports the MSE ratios between the machine forecast and survey median inflation forecast. The column "baseline" reports the ratio when machine forecasts are trained using optimal in-sample subsample of size T_E and training and validation subsample of size T_V. The second column reports the ratio when the machine forecasts are trained using 10-Fold cross validations over a rolling window of size T_E. The third column reports the ratio when the machine forecasts are trained using 10-Fold cross validations over an expanding window.

Data

This appendix describes our data.

VAR DATA

REAL GDP:

The real Gross Domestic Product is obtained from the US Bureau of Economic Analysis. It is in billions of chained 2012 dollars, quarterly frequency, seasonally adjusted, and at annual rate. We take the log of this variable. The source is from Bureau of Economic Analysis (BEA code: A191RX). The sample spans 1960:Q1 to 2019:Q3.

Table A3— Machine forecasts: Real-time vs. 2018:Q2 vintage

Machine vs Survey: Real-time vs 2018:Q2 Vintage

ML: $y_{j,t+h} = \alpha_{jh}^{(\mu)} + \beta_{jhF}^{(\mu)} F_t^{(\mu)} \left[y_{j,t+h} \right] + B_{jhZ}^{(\mu)} Z_{jt} + \epsilon_{jt+h}$

SPF GDP Forecasts							
Percentile	Median	Mean	5th	10th	20th	25th	30th
Real-time	0.88	0.91	0.70	0.81	0.80	0.84	0.87
Vintage 2018:Q2	0.79	0.83	0.68	0.76	0.75	0.77	0.80
Percentile	40th	60th	70th	75th	80th	90th	95th
Real-time	0.88	0.85	0.81	0.79	0.80	0.69	0.64
Vintage 2018:Q2	0.81	0.76	0.71	0.70	0.71	0.60	0.57
SPF Inflation Forecasts							
Percentile	Median	Mean	5th	10th	20th	25th	30th
Real-time	0.85	0.95	0.56	0.74	0.83	0.90	0.88
Vintage 2018:Q2	0.79	0.89	0.56	0.69	0.75	0.82	0.79
Percentile	40th	60th	70th	75th	80th	90th	95th
Real-time	0.89	0.74	0.70	0.67	0.59	0.55	0.47
Vintage 2018:Q2	0.81	0.71	0.66	0.64	0.52	0.56	0.39

Machine v.s. survey mean-square-forecast errors. $\mathrm{MSE}_{\mathbb{E}}$ and $\mathrm{MSE}_{\mathbb{F}}$ denote the machine and survey mean-squared-forecast-errors, respectively, for 4-quarter-ahead forecasts, averaged over the evaluation sample. The vintage of observations on the variable being forecast is specified in the first column, either real-time as in the main text or using 2018Q2 vintage data. The evaluation period is 1995:Q1 to 2018:Q2.

Table A4— No-switching Specification

Survey of Professional Forecasters (SPF)		
	1995:Q1-2018:Q2	2013:Q2-2018:Q2
$MSE_{\mathbb{E}}/MSE_{\mathbb{F}}$	0.96	0.67
Michigan Survey of Consumers (SOC)		
	1995:Q1-2018:Q2	2013:Q2-2018:Q2
$MSE_{\mathbb{E}}/MSE_{\mathbb{F}}$	0.98	0.71
Blue Chip Financial Forecasts (BC)		
	1995:Q1-2018:Q2	2013:Q2-2018:Q2
$MSE_{\mathbb{E}}/MSE_{\mathbb{F}}$	0.97	0.69

Machine v.s. survey mean-square-forecast errors. $\mathrm{MSE}_{\mathbb{E}}$ and $\mathrm{MSE}_{\mathbb{F}}$ denote the machine and survey mean-squared-forecast-errors, respectively, for 4-quarter-ahead GDP growth forecasts, averaged over the evaluation sample. The vintage of observations on the variable being forecast is the one available four quarters after the period being forecast.

REAL PERSONAL CONSUMPTION EXPENDITURES:

The real Personal Consumption Expenditures is obtained from the US Bureau of Economic Analysis. It is in billions of chained 2012 dollars, quarterly frequency, seasonally adjusted, and at annual rate. We take the log of this variable. The source is from Bureau of Economic Analysis (BEA code: DPCERX). The sample spans 1960:Q1 to 2019:Q3.

GDP price deflator:

The Gross Domestic Product: implicit price deflator is obtained from the US Bureau of Economic Analysis. Index base is 2012=100, quarterly frequency, and seasonally adjusted. We take the log of this variable. The source is from Bureau of Economic Analysis (BEA code: A191RD). The sample spans 1960:Q1 to 2019:Q3.

Real investment:

The real Gross Private Domestic Investment is obtained from the US Bureau of Economic Analysis. It is in billions of chained 2012 dollars, quarterly frequency, seasonally adjusted, and at annual rate. We take the log of this variable. The source is from Bureau of Economic Analysis (BEA code: A006RX). The sample spans 1960:Q1 to 2019:Q3.

Real wage:

We obtain real wages by dividing the Average Hourly Earnings of Production and Nonsupervisory Employees: Manufacturing over the Personal Consumption Expenditures (implicit price deflator). Average Hourly Earnings of Production and Nonsupervisory Employees: Manufacturing is obtained from the US Bureau of Labor Statistics; it is in dollars per hour, quarterly frequency (average), and seasonally adjusted. BLS Account Code: CES3000000008. Personal Consumption Expenditures (implicit price deflator) is obtained from the US Bureau of Economic Analysis. Index base is 2012=100, quarterly frequency, and seasonally adjusted. We take the log of the ratio of these variables. The source is from Bureau of Economic Analysis (BEA code: DPCERD). The sample spans 1960:Q1 to 2019:Q3.

S&P 500 stock market index:

The S&P 500 is obtained from the S&P Dow Jones Indices LLC. It is the quarterly average of the daily index value at market close. We take the log of this variable. The sample spans 1960:Q1 to 2019:Q3.

Federal funds rate (FFR):

The Effective Federal Funds Rate is obtained from the Board of Governors of the Federal Reserve System. It is in percentage points, quarterly frequency (average), and not seasonally adjusted. The sample spans 1960:Q1 to 2019:Q3.

Survey Data

All details on survey data and survey forecast construction here, with links to data sources.

SURVEY OF PROFESSIONAL FORECASTERS

The SPF is conducted each quarter by sending out surveys to professional forecasters, defined as forecasters. The number of surveys sent varies over time, but recent waves sent around 50 surveys each quarter according to officials at the Federal Reserve Bank of Philadelphia. Only forecasters with sufficient academic training and experience as macroeconomic forecasters are eligible to participate. Over the course of our sample, the number of respondents ranges from a minimum of 9, to a maximum of 83, and the mean number of respondents is 37. The surveys are sent out at the end of the first month of each quarter, and they are collected in the second or third week of the middle month of each quarter. Each survey asks respondents to provide nowcasts and quarterly forecasts from one to four quarters ahead for a variety of variables. Specifically, we use the SPF micro data on individual forecasts of the price level, long-run inflation, and real GDP.[1] Below we provide the exact definitions of these variables as well as our method for constructing nowcasts and forecasts of quarterly and annual inflation and GDP growth for each respondent.[2]

The following variables are used on either the right- or left-hand-sides of forecasting models:

1) Quarterly and annual inflation (1968:Q4 - present): We use survey responses for the level of the GDP price index (PGDP), defined as

"Forecasts for the quarterly and annual level of the chain-weighted GDP price index. Seasonally adjusted, index, base year varies. 1992-1995, GDP implicit deflator. Prior to 1992, GNP implicit deflator. Annual forecasts are for the annual average of the quarterly levels."

Since advance BEA estimates of these variables for the current quarter are unavailable at the time SPF respondents turn in their forecasts, four quarter-ahead inflation and GDP growth forecasts are constructed by dividing the forecasted level by the survey respondent-type's nowcast. Let $\mathbb{F}_t^{(i)}\left[P_{t+h}\right]$ be forecaster i's prediction of PGDP h quarters ahead and $\mathbb{N}_t^{(i)}\left[P_t\right]$ be forecaster i's nowcast of PGDP for the current quarter. Annualized inflation forecasts for forecaster i are

$$\text{(A1)} \qquad \mathbb{F}_t^{(i)}\left[\pi_{t+h,t}\right] = (400/h) \times \ln\left(\frac{\mathbb{F}_t^{(i)}\left[P_{t+h}\right]}{\mathbb{N}_t^{(i)}\left[P_t\right]}\right),$$

where $h = 1$ for quarterly inflation and $h = 4$ for annual inflation. Similarly, we

[1] Individual forecasts for all variables can be downloaded at https://www.philadelphiafed.org/research-and-data/real-time-center/survey-of-professional-forecasters/historical-data/individual-forecasts.

[2] The SPF documentation file can be found at https://www.philadelphiafed.org/-/media/research-and-data/real-time-center/survey-of-professional-forecasters/spf-documentation.pdf?la=en.

construct quarterly and annual nowcasts of inflation as

$$\mathbb{N}_t^{(i)}\left[\pi_{\,t,t-h}\right] = (400/h) \times \ln\left(\frac{\mathbb{N}_t^{(i)}\left[P_t\right]}{P_{t-h}}\right),$$

where $h = 1$ for quarterly inflation and $h = 4$ for annual inflation, and where P_{t-1} is the BEA's advance estimate of PGDP in the previous quarter observed by the respondent in time t, and P_{t-4} is the BEA's most accurate estimate of PGDP four quarters back. After computing inflation for each survey respondent, we calculate the 5th through the 95th percentiles as well as the average, variance, and skewness of inflation forecasts across respondents.

2) Long-run inflation (1991:Q4 - present): We use survey responses for 10-year-ahead CPI inflation (CPI10), which is defined as

   "Forecasts for the annual average rate of headline CPI inflation over the next 10 years. Seasonally adjusted, annualized percentage points. The "next 10 years" includes the year in which we conducted the survey and the following nine years. Conceptually, the calculation of inflation is one that runs from the fourth quarter of the year before the survey to the fourth quarter of the year that is ten years beyond the survey year, representing a total of 40 quarters or 10 years. The fourth-quarter level is the quarterly average of the underlying monthly levels."

   Only the median response is provided for CPI10, and it is already reported as an inflation rate, so we do not make any adjustments and cannot compute other moments or percentiles.

3) Real GDP growth (1968:Q4 - present): We use the level of real GDP (RGDP), which is defined as

   *"Forecasts for the quarterly and annual level of chain-weighted real GDP. Seasonally adjusted, annual rate, base year varies. 1992-1995, fixed-weighted real GDP. Prior to 1992, fixed-weighted real GNP. Annual forecasts are for the annual average of the quarterly levels. Prior to 1981:Q3, RGDP is computed by using the formula NGDP / PGDP * 100."*

   Quarterly and annual growth rates are constructed the same way as for inflation, except RGDP replaces PGDP.

In order to generate out-of-sample forecasts that could have been made in real time, it is necessary to take a stand on the information set of the forecasters when each forecast was made. We assume that forecasters could have used all data released before the survey deadlines. Table A5 lists the survey deadlines that are available, beginning with the 1990:Q3 survey. Before 1990:Q3, we make the conservative assumption that respondents only had data released by the first day of the second month of each quarter.

Table A5— SPF Survey Deadlines[3]

Survey	Deadline Date	Survey	Deadline Date	Survey	Deadline Date
1990:Q1	Unknown	1991:Q1	2/16/91	1992:Q1	2/22/92
Q2	Unknown	Q2	5/18/91	Q2	5/15/92
Q3	8/23/90	Q3	8/18/91	Q3	8/21/92
Q4	11/22/90	Q4	11/16/91	Q4	11/20/92
1993:Q1	2/19/93	1994:Q1	2/21/94	1995:Q1	2/21/95
Q2	5/20/93	Q2	5/18/94	Q2	5/22/95
Q3	8/19/93	Q3	8/18/94	Q3	8/22/95
Q4	11/23/93	Q4	11/18/94	Q4	11/20/95
1996:Q1	3/2/96	1997:Q1	2/19/97	1998:Q1	2/18/98
Q2	5/18/96	Q2	5/17/97	Q2	5/16/98
Q3	8/21/96	Q3	8/16/97	Q3	8/15/98
Q4	11/18/96	Q4	11/19/97	Q4	11/14/98
1999:Q1	2/16/99	2000:Q1	2/12/00	2001:Q1	2/14/01
Q2	5/15/99	Q2	5/13/00	Q2	5/12/01
Q3	8/14/99	Q3	8/12/00	Q3	8/15/01
Q4	11/13/99	Q4	11/11/00	Q4	11/14/01
2002:Q1	2/12/02	2003:Q1	2/14/03	2004:Q1	2/14/04
Q2	5/13/02	Q2	5/12/03	Q2	5/14/04
Q3	8/14/02	Q3	8/16/03	Q3	8/13/04
Q4	11/13/02	Q4	11/14/03	Q4	11/13/04
2005:Q1	2/9/05	2006:Q1	2/8/06	2007:Q1	2/8/07
Q2	5/12/05	Q2	5/10/06	Q2	5/9/07
Q3	8/11/05	Q3	8/9/06	Q3	8/8/07
Q4	11/8/05	Q4	11/8/06	Q4	11/7/07
2008:Q1	2/7/08	2009:Q1	2/10/09	2010:Q1	2/9/10
Q2	5/8/08	Q2	5/12/09	Q2	5/11/10
Q3	8/7/08	Q3	8/11/09	Q3	8/10/10
Q4	11/10/08	Q4	11/10/09	Q4	11/9/10
2011:Q1	2/8/11	2012:Q1	2/7/12	2013:Q1	2/11/13
Q2	5/10/11	Q2	5/8/12	Q2	5/7/13
Q3	8/8/11	Q3	8/7/12	Q3	8/12/13
Q4	11/8/11	Q4	11/6/12	Q4	11/18/13
2014:Q1	2/10/14	2015:Q1	2/10/15	2016:Q1	2/9/16
Q2	5/11/14	Q2	5/12/15	Q2	5/10/16
Q3	8/11/14	Q3	8/11/15	Q3	8/9/16
Q4	11/10/14	Q4	11/10/15	Q4	11/8/16

[3]SPF survey deadlines are posted online at https://www.philadelphiafed.org/-/media/research-and-data/real-time-center/survey-of-professional-forecasters/spf-release-dates.txt?la=en.

Table A5 (Cont'd)

Survey	Deadline Date	Survey	Deadline Date	Survey	Deadline Date
2017:Q1	2/7/17	2018:Q1	2/6/18		
Q2	5/9/17	Q2	5/8/18		
Q3	8/8/17	Q3	8/7/18		
Q4	11/7/17	Q4	11/6/18		

Michigan Survey of Consumers (SOC)

We construct MS forecasts of annual inflation and GDP growth of respondents answering at time t. Each month, the SOC contains approximately 50 core questions, and a minimum of 500 interviews are conducted by telephone over the course of the entire month, each month. We use two questions from the monthly survey for which the time series begins in January 1978, and convert to quarterly observations as explained below.

1) Annual CPI inflation: To get a point forecast, we combine the information in the survey responses to questions A12 and A12b.

   - Question A12 asks (emphasis in original): *During the next 12 months, do you think that prices in general will go up, or go down, or stay where they are now?*
   - A12b asks (emphasis in original): *By about what percent do you expect prices to go (up/down) on the average, during the next 12 months?*

2) Annual real GDP growth: We use survey responses to question A7, which asks (emphasis in original):

   And how about a year from now, do you expect that in the country as a whole business conditions will be better, or worse than they are at present, or just about the same?

   Respondents select one of three options: "better a year from now," "about the same," or "worse a year from now." There is a long history of using survey data as a proxy for spending and output (see, for example, Ludvigson - "Consumer Confidence and Consumer Spending" - Journal of Economic Perspectives - 2004). Using a companion question in the SOC that asks about contemporaneous business conditions, Curtin (2019) and the SOC survey documentation suggest constructing a "balance score" to generate a contemporaneous measure of real GDP growth. The *balance score* equals the percentage of respondents who expected that the economy to improve minus the percentage that expected it to worsen + 100. Applying this methodology to question A7.

   The balance score is obtained monthly and we use the observation for the middle month of each quarter as our quarterly observation. We convert the score to a quantitative survey-based measure of real GDP growth using a simple linear regression. Specifically, at time s, we assume that GDP growth, $y_{j,s+4}$, is related to the contemporaneous Michigan Survey balance sore, M_s, by:

$$y_{j,s+4} = \beta_0 + \beta_1 M_s + \epsilon_s.$$

This equation is estimated using OLS and the real-time vintage data, and then the forecast is constructed as $\mathbb{F}_{j,t}[y_{j,t+4}] = \hat{\beta}_0 + \hat{\beta}_1 M_t$

Specifically, we first estimate the coefficients of this regression over the sample 1978:Q1-1994:Q1. Using the estimated coefficients and the balance score from 1995:Q1 gives us the point forecast of inflation for 1995:Q1-1996:Q1. We then re-estimate this equation, recursively, adding one observation to the end of the sample at a time, and storing the fitted values. This results in a time series of forecasts $\mathbb{F}_{j,t}[y_{j,t+4}]$.

As with the SPF, we take a stand on the information set of consumers when each forecast was made, and we assume that consumers could have used all data released before they completed the survey. For the SOC interviews are conducted monthly over the course of an entire month. We set the interview response deadline for each survey as the first day of the survey month. For example, we set the deadline to February 1st, 2019, for the February 2019 Survey of Consumers, while in reality, the interview period was from February 2 to February 29, 2019. In other months, the true interview start period may be near the end of the previous month, such as in February 2019, when it was January 31st, 2019. To align the SOC more closely with the SPF deadline for survey completion (end of the second or third week of the middle month of the quarter), we use the middle month of each quarter as our quarterly observation for the SOC.

BLUECHIP DATA

We obtain Blue Chip expectation data from Blue Chip Financial Forecasts. The surveys are conducted each month by sending out surveys to forecasters in around 50 financial firms such as Bank of America, Goldman Sachs & Co., Swiss Re, Loomis, Sayles & Company, and J.P. Morgan Chase. The participants are surveyed around the 25th of each month and the results published a few days later on the 1st of the following month. The forecasters are asked to forecast the average of the level of U.S. interest rates over a particular calendar quarter, e.g. the federal funds rate and the set of H.15 Constant Maturity Treasuries (CMT) of the following maturities: 3-month, 6-month, 1-year, 2-year, 5-year and 10-year, and the quarter over quarter percentage changes in Real GDP, the GDP Price Index and the Consumer Price Index, beginning with the current quarter and extending 4 to 5 quarters into the future.

In this study, we look at a subset of the forecasted variables. Specifically, we use the Blue Chip micro data on individual forecasts of the quarter-over-quarter (Q/Q) percentage change in the Real GDP, the GDP Price Index and the CPI, and convert to quarterly observations as explained below.

1) Quarterly and annual PGDP inflation (1986:Q1 - 2018:Q3): We use survey responses for the quarter-over-quarter percentage change in the GDP price index, defined as:

"Forecasts for the quarter-over-quarter percentage change in the GDP Chained Price Index. Seasonally adjusted annual rate (SAAR). 1992 Jan. to 1996 June, Q/Q % change (SAAR) in GDP implicit deflator. 1986 Jan. to 1991 Dec., Q/Q % change (SAAR) in GNP implicit deflator."

Quarterly and annual inflation forecasts are constructed as follows. Let $\mathbb{F}_t^{(i)}\left[gP_{t+h}^{(Q/Q)}\right]$ be forecaster i's prediction of Q/Q % change in PGDP h quarters ahead. $\mathbb{F}_t^{(i)}\left[gP_{t+h}^{(Q/Q)}\right]$ are reported at annual rates in percentage points, so we convert to quarterly raw units before compounding. Annualized inflation forecasts for forecaster i in the next quarter are:

$$\mathbb{F}_t^{(i)}\left[\pi_{t+1,t}\right] = 400 \times \ln\left(1 + \frac{\mathbb{F}_t^{(i)}\left[gP_{t+1}^{(Q/Q)}\right]}{100}\right)^{\frac{1}{4}}$$

Annual Inflation forecasts are:

$$\mathbb{F}_t^{(i)}\left[\pi_{t+4,t}\right] = 100 \times \ln\left(\prod_{h=1}^{4}\left(1 + \frac{\mathbb{F}_t^{(i)}\left[gP_{t+h}^{(Q/Q)}\right]}{100}\right)^{\frac{1}{4}}\right)$$

Quarterly nowcasts of inflation are constructed as:

$$\mathbb{N}_t^{(i)}\left[\pi_{t,t-1}\right] = 400 \times \ln\left(1 + \frac{\mathbb{N}_t^{(i)}\left[gP_t^{(Q/Q)}\right]}{100}\right)^{\frac{1}{4}}$$

where $\mathbb{N}_t^{(i)}\left[gP_t^{(Q/Q)}\right]$ is forecaster i's nowcast of Q/Q % change in PGDP for the current quarter. Annual nowcasts of inflation for forecaster i are:

$$\mathbb{N}_t^{(i)}\left[\pi_{t,t-4}\right] = 100 \times \ln\left(\frac{\mathbb{N}_t^{(i)}\left[P_t\right]}{P_{t-4}}\right),$$

where P_{t-4} is the BEA's most accurate estimate of PGDP four quarters back and $\mathbb{N}_t^{(i)}[P_t]$ is forecaster i's nowcast of PGDP for the current quarter which is constructed as: $\mathbb{N}_t^{(i)}[P_t] = \exp\left(\mathbb{N}_t^{(i)}\left[\pi_{t,t-1}\right]/400 + \ln P_{t-1}\right)$. Similarly, we also calculate the 5th through the 95th percentiles as well as the average, variance, and skewness of inflation forecasts across respondents.

2) Real GDP growth (1984:Q3 - 2018:Q3): We use quarter-over-quarter percentage change in the Real GDP, which is defined as

   "Forecasts for the quarter-over-quarter percentage change in the level of chain-weighted real GDP. Seasonally adjusted, annual rate. Prior to 1992, Q/Q % change (SAAR) in real GNP."

   Quarterly and annual growth rates are constructed the same way as for inflation, except RGDP replaces PGDP.

3) CPI inflation (1984:Q3 - 2018:Q3): We use quarter-over-quarter percentage change in the consumer price index, which is defined as

   "Forecasts for the quarter-over-quarter percentage change in the CPI (consumer prices for all urban consumers). Seasonally adjusted, annual rate."

   Quarterly and annual CPI inflation are constructed the same way as for PGDP inflation, except CPI replaces PGDP.

The surveys are conducted right before the publication of the newsletter. Each issue is always dated the 1st of the month and the actual survey conducted over a two-day period almost always between 24th and 28th of the month. The major exception is the January issue when the survey is conducted a few days earlier to avoid conflict with the Christmas holiday. Therefore, we assume that the end of the last month (equivalently beginning of current month) is when the forecast is made. For example, for the report in 2008 Feb, we assume that the forecast is made on Feb 1, 2008. To convert monthly forecasts to quarterly forecasts, we use the forecasts in the middle month of each quarter as the quarterly forecasts. This is to align the Blue Chip more closely with the SPF deadline for survey completion, similar to what we do for the SOC.

REAL-TIME MACRO DATA

At each forecast date in the sample, we construct a dataset of macro variables that could have been observed on or before the day of the survey deadline. We use the Philadelphia Fed's Real-Time Data Set to obtain vintages of macro variables.[4] These vintages capture changes to historical data due to periodic revisions made by government statistical agencies. The vintages for a particular series can be available at the monthly and/or quarterly frequencies, and the series have monthly and/or quarterly observations. In cases where a variable has both frequencies available for its vintages and/or its observations, we choose one format of the variable. For instance, nominal personal consumption expenditures on goods is quarterly data with both monthly and quarterly vintages available; in this case, we use the version with monthly vintages.

Following Coibion and Gorodnichenko (2015), to construct forecasts and forecast errors, we use the vintage of inflation and GDP growth data that is available four quarters after the period being forecast. For example, the forecast error for a survey forecast of

[4] The real-time data sets are available at https://www.philadelphiafed.org/research-and-data/real-time-center/real-time-data/data-files.

P in 2017:Q2 that is made based on data as $t = 2016\text{:Q2}$ is computed by comparing the survey forecast $\mathbb{F}^{(i)}_{2016:\mathrm{Q2}}\left[P_{2017:\mathrm{Q2}}\right]$ with the actual value of $P_{2017:\mathrm{Q2}}$ given in the 2018:Q2 vintage of the real-time dataset.

REAL-TIME REGRESSANDS

Following CG, all regressions are run and forecast errors computed using forecasts of real-time inflation and GDP data available four quarters after the period being forecast. Following Faust and Wright (2013), we use continuous time compounding of inflation and GDP growth. For example, four quarter inflation is computed as

$$\pi_{t+4,t} = (100) \times \ln\left(\frac{P_{t+h}}{P_t}\right),$$

where P_t is the time t price level.

REAL-TIME REGRESSORS

For the regressors we need to combine all of the data observed at the time of a forecast date, and know the specific day that the data in each vintage are released. It is not sufficient to align vintage dates with forecast dates because the time t vintage might include data released after the time t forecast was made. The series-specific documentation on the Philadelphia Fed's website provides details on the timing of the vintages for each series. For some series, exact release dates are known, and thus the vintages reflect the data available at the time of the data release. When this is the case, we download the release dates from the relevant statistical agency and compare each vintage release date to the corresponding survey deadline to determine whether a particular vintage can be included in a survey respondent's information set.

For other variables, we only know that vintages contain data available in the middle of a month or quarter, but not the exact day. A subset of these variables come from the BEA National Income and Product Accounts, which are released at the end of each month. Since NIPA series are released at the end of each month, and vintages reflect data available in the middle of each month, a survey respondent making a forecast in the middle of a month includes the current month's vintage of NIPA data in her information set. However, there is another subset of variables with unknown release dates, for which we must make the conservative assumption that a forecaster at time t observes at most the time $t-1$ vintage of data. An Excel Workbook containing the known release dates and timing assumptions is available on the authors' websites.

In addition to the macro variables with different vintages that we obtain from the Philadelphia Fed, we include a measure of residential real estate prices from the Case-Shiller/S&P index deflated by the Consumer Price Index, and energy prices from the U.S. Bureau of Labor Statistics (BLS). Energy prices do not get revised, so they do not have multiple vintages. Instead there is just one historical version of the data.

After combining all of the series that are known by the forecasters at each date, we convert monthly data to quarterly by using either the beginning-of-quarter or end-of-quarter values. The decision to use beginning-of-quarter or end-of-quarter depends on the survey deadline of a particular forecast date. If the survey deadline is known to be in the middle of the second month of quarter t, then it is conceivable that the forecasters would have information about the first month of quarter t. Therefore, we use the first month of that quarter's values. A few anomalous observations have unknown survey deadlines (e.g., the SPF deadlines for 1990:Q1). In such cases, we allow only information up to quarter $t-1$ to enter the model. Thus, we use the last month of the previous quarter's values in these cases.

Table A6 gives the complete list of real-time macro variables. Included in the table is the first available vintages for each variable that has multiple vintages. We do not include the last vintage because most variables have vintages through the present.[5] Table A6 also lists the transformation applied to each variable to make them stationary before generating factors. Let X_{it} denote variable i at time t after the transformation, and let X_{it}^A be the untransformed series. Let $\Delta = (1-L)$ with $LX_{it} = X_{it-1}$. There are seven possible transformations with the following codes:

1 Code lv: $X_{it} = X_{it}^A$
2 Code Δlv: $X_{it} = X_{it}^A - X_{it-1}^A$
3 Code $\Delta^2 lv$: $X_{it} = \Delta^2 X_{it}^A$
4 Code ln: $X_{it} = \ln(X_{it}^A)$
5 Code Δln: $X_{it} = \ln(X_{it}^A) - \ln(X_{it-1}^A)$
6 Code $\Delta^2 ln$: $X_{it} = \Delta^2 \ln(X_{it}^A)$
7 Code $\Delta lv/lv$: $X_{it} = (X_{it}^A - X_{it-1}^A)/X_{it-1}^A$

Table A6— List of Macro Dataset Variables

No.	Short Name	Source	Tran	Description	First Vintage
			Group 1: Output and Income		
1	IPMMVMD	Philly Fed	Δln	Ind. production index - Manufacturing	1962:M11
2	IPTMVMD	Philly Fed	Δln	Ind. production index - Total	1962:M11
3	CUMMVMD	Philly Fed	lv	Capacity utilization - Manufacturing	1979:M8
4	CUTMVMD	Philly Fed	lv	Capacity utilization - Total	1983:M7
5	NCPROFATMVQD	Philly Fed	Δln	Nom. corp. profits after tax without IVA/CCAdj	1965:Q4
6	NCPROFATWMVQD	Philly Fed	Δln	Nom. corp. profits after tax with IVA/CCAdj	1981:Q1
7	OPHMVQD	Philly Fed	Δln	Output per hour - Business sector	1998:Q4
8	NDPIQVQD	Philly Fed	Δln	Nom. disposable personal income	1965:Q4
9	NOUTPUTQVQD	Philly Fed	Δln	Nom. GNP/GDP	1965:Q4
10	NPIQVQD	Philly Fed	Δln	Nom. personal income	1965:Q4
11	NPSAVQVQD	Philly Fed	Δlv	Nom. personal saving	1965:Q4
12	OLIQVQD	Philly Fed	Δln	Other labor income	1965:Q4
13	PINTIQVQD	Philly Fed	Δln	Personal interest income	1965:Q4

[5]For variables BASEBASAQVMD, NBRBASAQVMD, NBRECBASAQVMD, and TRBASAQVMD, the last available vintage is 2013:Q2.

Table A6 (Cont'd)

No.	Short Name	Source	Tran	Description	First Vintage
14	PINTPAIDQVQD	Philly Fed	Δln	Interest paid by consumers	1965:Q4
15	PROPIQVQD	Philly Fed	Δln	Proprietors' income	1965:Q4
16	PTAXQVQD	Philly Fed	Δln	Personal tax and nontax payments	1965:Q4
17	RATESAVQVQD	Philly Fed	Δlv	Personal saving rate	1965:Q4
18	RENTIQVQD	Philly Fed	Δlv	Rental income of persons	1965:Q4
19	ROUTPUTQVQD	Philly Fed	Δln	Real GNP/GDP	1965:Q4
20	SSCONTRIBQVQD	Philly Fed	Δln	Personal contributions for social insurance	1965:Q4
21	TRANPFQVQD	Philly Fed	Δln	Personal transfer payments to foreigners	1965:Q4
22	TRANRQVQD	Philly Fed	Δln	Transfer payments	1965:Q4
23	CUUR0000SA0E	BLS	$\Delta^2 ln$	Energy in U.S. city avg., all urban consumers, not seasonally adj	
				Group 2: Employment	
24	EMPLOYMVMD	Philly Fed	Δln	Nonfarm payroll	1946:M12
25	HMVMD	Philly Fed	lv	Aggregate weekly hours - Total	1971:M9
26	HGMVMD	Philly Fed	lv	Agg. weekly hours - Goods-producing	1971:M9
27	HSMVMD	Philly Fed	lv	Agg. weekly hours - Service-producing	1971:M9
28	LFCMVMD	Philly Fed	Δln	Civilian labor force	1998:M11
29	LFPARTMVMD	Philly Fed	lv	Civilian participation rate	1998:M11
30	POPMVMD	Philly Fed	Δln	Civilian noninstitutional population	1998:M11
31	ULCMVQD	Philly Fed	Δln	Unit labor costs - Business sector	1998:Q4
32	RUCQVMD	Philly Fed	Δlv	Unemployment rate	1965:Q4
33	WSDQVQD	Philly Fed	Δln	Wage and salary disbursements	1965:Q4
				Group 3: Orders, Investment, Housing	
34	HSTARTSMVMD	Philly Fed	Δln	Housing starts	1968:M2
35	RINVBFMVQD	Philly Fed	Δln	Real gross private domestic inv. - Nonresidential	1965:Q4
36	RINVCHIMVQD	Philly Fed	Δlv	Real gross private domestic inv. - Change in private inventories	1965:Q4
37	RINVRESIDMVQD	Philly Fed	Δln	Real gross private domestic inv. - Residential	1965:Q4
38	CASESHILLER	S&P	Δln	Case-Shiller US National Home Price index/CPI	1987:M1
				Group 4: Consumption	
39	NCONGMMVMD	Philly Fed	Δln	Nom. personal cons. exp. - Goods	2009:M8
40	NCONHHMMVMD	Philly Fed	Δln	Nom. hh. cons. exp.	2009:M8
41	NCONSHHMMVMD	Philly Fed	Δln	Nom. hh. cons. exp. - Services	2009:M8
42	NCONSNPMMVMD	Philly Fed	Δln	Nom. final cons. exp. of NPISH	2009:M8
43	RCONDMMVMD	Philly Fed	Δln	Real personal cons. exp. - Durables	1998:M11
44	RCONGMMVMD	Philly Fed	Δln	Real personal cons. exp. - Goods	2009:M8
45	RCONHHMMVMD	Philly Fed	Δln	Real hh. cons. exp.	2009:M8
46	RCONMMVMD	Philly Fed	Δln	Real personal cons. exp. - Total	1998:M11
47	RCONNDMVMD	Philly Fed	Δln	Real personal cons. exp. - Nondurables	1998:M11
48	RCONSHHMMVMD	Philly Fed	Δln	Real hh. cons. exp. - Services	2009:M8
49	RCONSMMVMD	Philly Fed	Δln	Real personal cons. exp. - Services	1998:M11
50	RCONSNPMMVMD	Philly Fed	Δln	Real final cons. exp. of NPISH	2009:M8
51	NCONGMVQD	Philly Fed	Δln	Nom. personal cons. exp. - Goods	2009:Q3
52	NCONHHMVQD	Philly Fed	Δln	Nom. hh. cons. exp.	0209:Q3
53	NCONSHHMVQD	Philly Fed	Δln	Nom. hh. cons. exp. - Services	2009:Q3
54	NCONSNPMVQD	Philly Fed	Δln	Nom. final cons. exp. of NPISH	2009:Q3
55	RCONDMVQD	Philly Fed	Δln	Real personal cons. exp. - Durable goods	1965:Q4

Table A6 (Cont'd)

No.	Short Name	Source	Tran	Description	First Vintage
56	RCONGMVQD	Philly Fed	Δln	Real personal cons. exp. - Goods	2009:Q3
57	RCONHHMVQD	Philly Fed	Δln	Real hh. cons. exp.	2009:Q3
58	RCONMVQD	Philly Fed	Δln	Real personal cons. exp. - Total	1965:Q4
59	RCONNDMVQD	Philly Fed	Δln	Real pesonal cons. exp. - Nondurable goods	1965:Q4
60	RCONSHHMVQD	Philly Fed	Δln	Real hh. cons. exp. - Services	2009:Q3
61	RCONSMVQD	Philly Fed	Δln	Real personal cons. exp. - Services	1965:Q4
62	RCONSNPMVQD	Philly Fed	Δln	Real final cons. exp. of NPISH	2009:Q3
63	NCONQVQD	Philly Fed	Δln	Nom. personal cons. exp.	1965:Q4
			Group 5: Prices		
64	PCONGMMVMD	Philly Fed	$\Delta^2 ln$	Price index for personal cons. exp. - Goods	2009:M8
65	PCONHHMMVMD	Philly Fed	$\Delta^2 ln$	Price index for hh. cons. exp.	2009:M8
66	PCONSHHMMVMD	Philly Fed	$\Delta^2 ln$	Price index for hh. cons. exp. - Services	2009:M8
67	PCONSNPMMVMD	Philly Fed	$\Delta^2 ln$	Price index for final cons. exp. of NPISH	2009:M8
68	PCPIMVMD	Philly Fed	$\Delta^2 ln$	Consumer price index	1998:M11
69	PCPIXMVMD	Philly Fed	$\Delta^2 ln$	Core consumer price index	1998:M11
70	PPPIMVMD	Philly Fed	$\Delta^2 ln$	Producer price index	1998:M11
71	PPPIXMVMD	Philly Fed	$\Delta^2 ln$	Core producer price index	1998:M11
72	PCONGMVQD	Philly Fed	$\Delta^2 ln$	Price index for personal. cons. exp. - Goods	2009:Q3
73	PCONHHMVQD	Philly Fed	$\Delta^2 ln$	Price index for hh. cons. exp.	2009:Q3
74	PCONSHHMVQD	Philly Fed	$\Delta^2 ln$	Price index for hh. cons. exp. - Services	2009:Q3
75	PCONSNPMVQD	Philly Fed	$\Delta^2 ln$	Price index for final cons. exp. of NPISH	2009:Q3
76	PCONXMVQD	Philly Fed	$\Delta^2 ln$	Core price index for personal cons. exp.	1996:Q1
77	CPIQVMD	Philly Fed	$\Delta^2 ln$	Consumer price index	1994:Q3
78	PQVQD	Philly Fed	$\Delta^2 ln$	Price index for GNP/GDP	1965:Q4
79	PCONQVQD	Philly Fed	$\Delta^2 ln$	Price index for personal cons. exp.	1965:Q4
80	PIMPQVQD	Philly Fed	$\Delta^2 ln$	Price index for imports of goods and services	1965:Q4
			Group 6: Trade and Government		
81	REXMVQD	Philly Fed	Δln	Real exports of goods and services	1965:Q4
82	RGMVQD	Philly Fed	Δln	Real government cons. and gross inv. - Total	1965:Q4
83	RGFMVQD	Philly Fed	Δln	Real government cons. and gross inv. - Federal	1965:Q4
84	RGSLMVQD	Philly Fed	Δln	Real government cons. and gross. inv. - State and local	1965:Q4
85	RIMPMVQD	Philly Fed	Δln	Real imports of goods and services	1965:Q4
86	RNXMVQD	Philly Fed	Δlv	Real net exports of goods and services	1965:Q4
			Group 7: Money and Credit		
87	BASEBASAQVMD	Philly Fed	$\Delta^2 ln$	Monetary base	1980:Q2
88	M1QVMD	Philly Fed	$\Delta^2 ln$	M1 money stock	1965:Q4
89	M2QVMD	Philly Fed	$\Delta^2 ln$	M2 money stock	1971:Q2
90	NBRBASAQVMD	Philly Fed	$\Delta lv/lv$	Nonborrowed reserves	1967:Q3
91	NBRECBASAQVMD	Philly Fed	$\Delta lv/lv$	Nonborrowed reserves plus extended credit	1984:Q2
92	TRBASAQVMD	Philly Fed	$\Delta^2 ln$	Total reserves	1967:Q3
93	DIVQVQD	Philly Fed	Δln	Dividends	1965:Q4

MONTHLY FINANCIAL FACTOR DATA

The 147 financial series in this data set are versions of the financial dataset used in Jurado et al. (2015) and Ludvigson et al. (2019). It consists of a number of indicators measuring the behavior of a broad cross-section of asset returns, as well as some aggregate financial indicators not included in the macro dataset. These data include valuation ratios such as the dividend-price ratio and earnings-price ratio, growth rates of aggregate dividends and prices, default and term spreads, yields on corporate bonds of different ratings grades, yields on Treasuries and yield spreads, and a broad cross-section of industry equity returns. Following Fama and French (1992), returns on 100 portfolios of equities sorted into 10 size and 10 book-market categories. The dataset X^f also includes a group of variables we call "risk-factors," since they have been used in cross-sectional or time-series studies to uncover variation in the market risk-premium. These risk-factors include the three Fama and French (1993) risk factors, namely the excess return on the market MKT_t, the "small-minus-big" (SMB_t) and "high-minus-low" (HML_t) portfolio returns, the momentum factor UMD_t, and the small stock value spread $R15 - R11$.

The raw data used to form factors are always transformed to achieve stationarity. In addition, when forming forecasting factors from the large macro and financial datasets, the raw data (which are in different units) are standardized before performing PCA. When forming common uncertainty from estimates of individual uncertainty, the raw data (which are in this case in the same units) are demeaned, but we do not divide by the observation's standard deviation before performing PCA.

Throughout, the factors are estimated by the method of static principal components (PCA). Specifically, the $T \times r_F$ matrix $\hat{F}_t$ is $\sqrt{T}$ times the r_F eigenvectors corresponding to the r_F largest eigenvalues of the $T \times T$ matrix $xx'/(TN)$ in decreasing order. In large samples (when $\sqrt{T}/N \rightarrow \infty$), Bai and Ng (2006) show that the estimates $\hat{F}_t$ can be treated as though they were observed in the subsequent forecasting regression.

All returns and spreads are expressed in logs (i.e. the log of the gross return or spread), are displayed in percent (i.e. multiplied by 100), and are annualized by multiplying by 12, i.e., if x is the original return or spread, we transform to $1200\ln(1 + x/100)$. Federal Reserve data are annualized by default and are therefore not "re-annualized." Note: this annualization means that the annualized standard deviation (volatility) is equal to the data standard deviation divided by $\sqrt{12}$. The data series used in this dataset are listed below by data source. Additional details on data transformations are given below the table.

We convert monthly data to quarterly by using either the beginning-of-quarter or end-of-quarter values. The decision to use beginning-of-quarter or end-of-quarter depends on the survey deadline of a particular forecast date. If the survey deadline is known to be in the middle of the second month of quarter t, then it is conceivable that the forecasters would have information about the first month of quarter t. Therefore, we use the first month of that quarter's values. Alternatively, a few anomalous observations have unknown survey deadlines (e.g., the SPF deadlines for 1990:Q1). In such cases, we allow only information up to quarter $t - 1$ to enter the model. Thus, we use the last month of the previous quarter's values in these cases.

Let X_{it} denote variable i observed at time t after e.g., logarithm and differencing transformation, and let X_{it}^A be the actual (untransformed) series. Let $\Delta = (1 - L)$ with $LX_{it} = X_{it-1}$. There are six possible transformations with the following codes:

1 Code lv: $X_{it} = X_{it}^A$.

2 Code Δlv: $X_{it} = X_{it}^A - X_{it-1}^A$.

3 Code $\Delta^2 lv$: $X_{it} = \Delta^2 X_{it}^A$.

4 Code ln: $X_{it} = ln(X_{it}^A)$.

5 Code Δln: $X_{it} = ln(X_{it}^A) - ln(X_{it-1}^A)$.

6 Code $\Delta^2 ln$: $X_{it} = \Delta^2 ln X_{it}^A$.

7 Code $\Delta lv/lv$: $\left(X_{it}^A - X_{it-1}^A\right)/X_{it-1}^A$

Table A7— List of Financial Dataset Variables

No.	Short Name	Source	Tran	Description
		Group 1: Prices, Yield, Dividends		
1	D_log(DIV)	CRSP	Δln	Δ log D_t^* see additional details below
2	D_log(P)	CRSP	Δln	Δ log P_t see additional details below
3	D_DIVreinvest	CRSP	Δln	Δ log $D_t^{re,*}$ see additional details below
4	D_Preinvest	CRSP	Δln	Δ log $P_t^{re,*}$ see additional details below
5	d-p	CRSP	ln	$\log(D_t^*) - \log P_t$ see additional details below
		Group 2: Equity Risk Factors		
6	R15-R11	Kenneth French	lv	(Small, High) minus (Small, Low) sorted on (size, book-to-market)
7	Mkt-RF	Kenneth French	lv	Market excess return
8	SMB	Kenneth French	lv	Small Minus Big, sorted on size
9	HML	Kenneth French	lv	High Minus Low, sorted on book-to-market
10	UMD	Kenneth French	lv	Up Minus Down, sorted on momentum
		Group 3: Industries		
11	Agric	Kenneth French	lv	Agric industry portfolio
12	Food	Kenneth French	lv	Food industry portfolio
13	Beer	Kenneth French	lv	Beer industry portfolio
14	Smoke	Kenneth French	lv	Smoke industry portfolio
15	Toys	Kenneth French	lv	Toys industry portfolio
16	Fun	Kenneth French	lv	Fun industry portfolio
17	Books	Kenneth French	lv	Books industry portfolio
18	Hshld	Kenneth French	lv	Hshld industry portfolio
19	Clths	Kenneth French	lv	Clths industry portfolio
20	MedEq	Kenneth French	lv	MedEq industry portfolio
21	Drugs	Kenneth French	lv	Drugs industry portfolio
22	Chems	Kenneth French	lv	Chems industry portfolio
23	Rubbr	Kenneth French	lv	Rubbr industry portfolio
24	Txtls	Kenneth French	lv	Txtls industry portfolio
25	BldMt	Kenneth French	lv	BldMt industry portfolio

Table A7 (Cont'd)

No.	Short Name	Source	Tran	Description
26	Cnstr	Kenneth French	lv	Cnstr industry portfolio
27	Steel	Kenneth French	lv	Steel industry portfolio
28	Mach	Kenneth French	lv	Mach industry portfolio
29	ElcEq	Kenneth French	lv	ElcEq industry portfolio
30	Autos	Kenneth French	lv	Autos industry portfolio
31	Aero	Kenneth French	lv	Aero industry portfolio
32	Ships	Kenneth French	lv	Ships industry portfolio
33	Mines	Kenneth French	lv	Mines industry portfolio
34	Coal	Kenneth French	lv	Coal industry portfolio
35	Oil	Kenneth French	lv	Oil industry portfolio
36	Util	Kenneth French	lv	Util industry portfolio
37	Telcm	Kenneth French	lv	Telcm industry portfolio
38	PerSv	Kenneth French	lv	PerSv industry portfolio
39	BusSv	Kenneth French	lv	BusSv industry portfolio
40	Hardw	Kenneth French	lv	Hardw industry portfolio
41	Chips	Kenneth French	lv	Chips industry portfolio
42	LabEq	Kenneth French	lv	LabEq industry portfolio
43	Paper	Kenneth French	lv	Paper industry portfolio
44	Boxes	Kenneth French	lv	Boxes industry portfolio
45	Trans	Kenneth French	lv	Trans industry portfolio
46	Whlsl	Kenneth French	lv	Whlsl industry portfolio
47	Rtail	Kenneth French	lv	Rtail industry portfolio
48	Meals	Kenneth French	lv	Meals industry portfolio
49	Banks	Kenneth French	lv	Banks industry portfolio
50	Insur	Kenneth French	lv	Insur industry portfolio
51	RlEst	Kenneth French	lv	RlEst industry portfolio
52	Fin	Kenneth French	lv	Fin industry portfolio
53	Other	Kenneth French	lv	Other industry portfolio
		Group 4: Size/BM		
54	1_2	Kenneth French	lv	(1, 2) portfolio sorted on (size, book-to-market)
55	1_4	Kenneth French	lv	(1, 4) portfolio sorted on (size, book-to-market)
56	1_5	Kenneth French	lv	(1, 5) portfolio sorted on (size, book-to-market)
57	1_6	Kenneth French	lv	(1, 6) portfolio sorted on (size, book-to-market)
58	1_7	Kenneth French	lv	(1, 7) portfolio sorted on (size, book-to-market)
59	1_8	Kenneth French	lv	(1, 8) portfolio sorted on (size, book-to-market)
60	1_9	Kenneth French	lv	(1, 9) portfolio sorted on (size, book-to-market)
61	1_high	Kenneth French	lv	(1, high) portfolio sorted on (size, book-to-market)
62	2_low	Kenneth French	lv	(2, low) portfolio sorted on (size, book-to-market)
63	2_2	Kenneth French	lv	(2, 2) portfolio sorted on (size, book-to-market)
64	2_3	Kenneth French	lv	(2, 3) portfolio sorted on (size, book-to-market)
65	2_4	Kenneth French	lv	(2, 4) portfolio sorted on (size, book-to-market)
66	2_5	Kenneth French	lv	(2, 5) portfolio sorted on (size, book-to-market)
67	2_6	Kenneth French	lv	(2, 6) portfolio sorted on (size, book-to-market)
68	2_7	Kenneth French	lv	(2, 7) portfolio sorted on (size, book-to-market)
69	2_8	Kenneth French	lv	(2, 8) portfolio sorted on (size, book-to-market)
70	2_9	Kenneth French	lv	(2, 9) portfolio sorted on (size, book-to-market)
71	2_high	Kenneth French	lv	(2, high) portfolio sorted on (size, book-to-market)

Table A7 (Cont'd)

No.	Short Name	Source	Tran	Description
72	3_low	Kenneth French	lv	(3, low) portfolio sorted on (size, book-to-market)
73	3_2	Kenneth French	lv	(3, 2) portfolio sorted on (size, book-to-market)
74	3_3	Kenneth French	lv	(3, 3) portfolio sorted on (size, book-to-market)
75	3_4	Kenneth French	lv	(3, 4) portfolio sorted on (size, book-to-market)
76	3_5	Kenneth French	lv	(3, 5) portfolio sorted on (size, book-to-market)
77	3_6	Kenneth French	lv	(3, 6) portfolio sorted on (size, book-to-market)
78	3_7	Kenneth French	lv	(3, 7) portfolio sorted on (size, book-to-market)
79	3_8	Kenneth French	lv	(3, 8) portfolio sorted on (size, book-to-market)
80	3_9	Kenneth French	lv	(3, 9) portfolio sorted on (size, book-to-market)
81	3_high	Kenneth French	lv	(3, high) portfolio sorted on (size, book-to-market)
82	4_low	Kenneth French	lv	(4, low) portfolio sorted on (size, book-to-market)
83	4_2	Kenneth French	lv	(4, 2) portfolio sorted on (size, book-to-market)
84	4_3	Kenneth French	lv	(4, 3) portfolio sorted on (size, book-to-market)
85	4_4	Kenneth French	lv	(4, 4) portfolio sorted on (size, book-to-market)
86	4_5	Kenneth French	lv	(4, 5) portfolio sorted on (size, book-to-market)
87	4_6	Kenneth French	lv	(4, 6) portfolio sorted on (size, book-to-market)
88	4_7	Kenneth French	lv	(4, 7) portfolio sorted on (size, book-to-market)
89	4_8	Kenneth French	lv	(4, 8) portfolio sorted on (size, book-to-market)
90	4_9	Kenneth French	lv	(4, 9) portfolio sorted on (size, book-to-market)
91	4_high	Kenneth French	lv	(4, high) portfolio sorted on (size, book-to-market)
92	5_low	Kenneth French	lv	(5, low) portfolio sorted on (size, book-to-market)
93	5_2	Kenneth French	lv	(5, 2) portfolio sorted on (size, book-to-market)
94	5_3	Kenneth French	lv	(5, 3) portfolio sorted on (size, book-to-market)
95	5_4	Kenneth French	lv	(5, 4) portfolio sorted on (size, book-to-market)
96	5_5	Kenneth French	lv	(5, 5) portfolio sorted on (size, book-to-market)
97	5_6	Kenneth French	lv	(5, 6) portfolio sorted on (size, book-to-market)
98	5_7	Kenneth French	lv	(5, 7) portfolio sorted on (size, book-to-market)
99	5_8	Kenneth French	lv	(5, 8) portfolio sorted on (size, book-to-market)
100	5_9	Kenneth French	lv	(5, 9) portfolio sorted on (size, book-to-market)
101	5_high	Kenneth French	lv	(5, high) portfolio sorted on (size, book-to-market)
102	6_low	Kenneth French	lv	(6, low) portfolio sorted on (size, book-to-market)
103	6_2	Kenneth French	lv	(6, 2) portfolio sorted on (size, book-to-market)
104	6_3	Kenneth French	lv	(6, 3) portfolio sorted on (size, book-to-market)
105	6_4	Kenneth French	lv	(6, 4) portfolio sorted on (size, book-to-market)
106	6_5	Kenneth French	lv	(6, 5) portfolio sorted on (size, book-to-market)
107	6_6	Kenneth French	lv	(6, 6) portfolio sorted on (size, book-to-market)
108	6_7	Kenneth French	lv	(6, 7) portfolio sorted on (size, book-to-market)
109	6_8	Kenneth French	lv	(6, 8) portfolio sorted on (size, book-to-market)
110	6_9	Kenneth French	lv	(6, 9) portfolio sorted on (size, book-to-market)
111	6_high	Kenneth French	lv	(6, high) portfolio sorted on (size, book-to-market)
112	7_low	Kenneth French	lv	(7, low) portfolio sorted on (size, book-to-market)
113	7_2	Kenneth French	lv	(7, 2) portfolio sorted on (size, book-to-market)
114	7_3	Kenneth French	lv	(7, 3) portfolio sorted on (size, book-to-market)
115	7_4	Kenneth French	lv	(7, 4) portfolio sorted on (size, book-to-market)
116	7_5	Kenneth French	lv	(7, 5) portfolio sorted on (size, book-to-market)
117	7_6	Kenneth French	lv	(7, 6) portfolio sorted on (size, book-to-market)
118	7_7	Kenneth French	lv	(7, 7) portfolio sorted on (size, book-to-market)

Table A7 (Cont'd)

No.	Short Name	Source	Tran	Description
119	7_8	Kenneth French	lv	(7, 8) portfolio sorted on (size, book-to-market)
120	7_9	Kenneth French	lv	(7, 9) portfolio sorted on (size, book-to-market)
121	8_low	Kenneth French	lv	(8, low) portfolio sorted on (size, book-to-market)
122	8_2	Kenneth French	lv	(8, 2) portfolio sorted on (size, book-to-market)
123	8_3	Kenneth French	lv	(8, 3) portfolio sorted on (size, book-to-market)
124	8_4	Kenneth French	lv	(8, 4) portfolio sorted on (size, book-to-market)
125	8_5	Kenneth French	lv	(8, 5) portfolio sorted on (size, book-to-market)
126	8_6	Kenneth French	lv	(8, 6) portfolio sorted on (size, book-to-market)
127	8_7	Kenneth French	lv	(8, 7) portfolio sorted on (size, book-to-market)
128	8_8	Kenneth French	lv	(8, 8) portfolio sorted on (size, book-to-market)
129	8_9	Kenneth French	lv	(8, 9) portfolio sorted on (size, book-to-market)
130	8_high	Kenneth French	lv	(8, high) portfolio sorted on (size, book-to-market)
131	9_low	Kenneth French	lv	(9, low) portfolio sorted on (size, book-to-market)
132	9_2	Kenneth French	lv	(9, 2) portfolio sorted on (size, book-to-market)
133	9_3	Kenneth French	lv	(9, 3) portfolio sorted on (size, book-to-market)
134	9_4	Kenneth French	lv	(9, 4) portfolio sorted on (size, book-to-market)
135	9_5	Kenneth French	lv	(9, 5) portfolio sorted on (size, book-to-market)
136	9_6	Kenneth French	lv	(9, 6) portfolio sorted on (size, book-to-market)
137	9_7	Kenneth French	lv	(9, 7) portfolio sorted on (size, book-to-market)
138	9_8	Kenneth French	lv	(9, 8) portfolio sorted on (size, book-to-market)
139	9_high	Kenneth French	lv	(9, high) portfolio sorted on (size, book-to-market)
140	10_low	Kenneth French	lv	(10, low) portfolio sorted on (size, book-to-market)
141	10_2	Kenneth French	lv	(10, 2) portfolio sorted on (size, book-to-market)
142	10_3	Kenneth French	lv	(10, 3) portfolio sorted on (size, book-to-market)
143	10_4	Kenneth French	lv	(10, 4) portfolio sorted on (size, book-to-market)
144	10_5	Kenneth French	lv	(10, 5) portfolio sorted on (size, book-to-market)
145	10_6	Kenneth French	lv	(10, 6) portfolio sorted on (size, book-to-market)
146	10_7	Kenneth French	lv	(10, 7) portfolio sorted on (size, book-to-market)
147	VXO	Fred MD	lv	VXOCLSx

CRSP Data Details

Value-weighted price and dividend data were obtained from the Center for Research in Security Prices (CRSP). From the Annual Update data, we obtain monthly value-weighted returns series vwretd (with dividends) and vwretx (excluding dividends). These series have the interpretation

$$VWRETD_t = \frac{P_{t+1} + D_{t+1}}{P_t}$$
$$VWRETX_t = \frac{P_{t+1}}{P_t}$$

From these series, a normalized price series P, can be constructed using the recursion

$$P_0 = 1$$
$$P_t = P_{t-1} \cdot VWRETX_t.$$

A dividend series can then be constructed using

$$D_t = P_{t-1}(VWRETD_t - VWRETX_t).$$

In order to remove seasonality of dividend payments from the data, instead of D_t we use the series

$$D_t^* = \frac{1}{12}\sum_{j=0}^{11} D_{t-j}$$

i.e., the moving average over the entire year. For the price and dividend series under "reinvestment," we calculate the price under reinvestment, P_t^{re}, as the normalized value of the market portfolio under reinvestment of dividends, using the recursion

$$P_0^{re} = 1$$
$$P_t^{re} = P_{t-1} \cdot VWRETD_t$$

Similarly, we can define dividends under reinvestment, D_t^{re}, as the total dividend payments on this portfolio (the number of "shares" of which have increased over time) using

$$D_t^{re} = P_{t-1}^{re}(VWRETD_t - VWRETX_t).$$

As before, we can remove seasonality by using

$$D_t^{re,*} = \frac{1}{2}\sum_{j=0}^{11} D_{t-j}^{re}.$$

Five data series are constructed from the CRSP data as follows:

- D_log(DIV): $\Delta \log D_t^*$.
- D_log(P): $\Delta \log P_t$.
- D_DIVreinvest: $\Delta \log D_t^{re,*}$
- D_Preinvest: $\Delta \log P_t^{re,*}$
- d-p: $\log(D_t^*) - \log(P_t)$

Kenneth French Data Details

The following data are obtained from the data library of Kenneth French's Dartmouth website (http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html):

- Fama/French Factors: From this dataset we obtain the data series RF, Mkt-RF, SMB, HML.
- 25 Portfolios formed on Size and Book-to-Market (5 x 5): From this dataset we obtain the series R15-R11, which is the spread between the (small, high book-to-market) and (small, low book-to-market) portfolios.
- Momentum Factor (Mom): From this dataset we obtain the series UMD, which is equal to the momentum factor.
- 49 Industry Portfolios: From this dataset we use all value-weighted series, excluding any series that have missing observations from Jan. 1960 on, from which we obtain the series Agric through Other. The omitted series are: Soda, Hlth, FabPr, Guns, Gold, Softw.
- 100 Portfolios formed in Size and Book-to-Market: From this dataset we use all value-weighted series, excluding any series that have missing observations from Jan. 1960 on. This yields variables with the name X_Y where X stands for the index of the size variable (1, 2, ..., 10) and Y stands for the index of the book-to-market variable (Low, 2, 3, ..., 8, 9, High). The omitted series are 1_low, 1_3, 7_high, 9_9, 10_8, 10_9, 10_high.

Daily Financial Data

Daily Data and construction of daily factors

The daily financial series in this data set are from the daily financial dataset used in Andreou et al. (2013). We create a smaller daily database which is a subset of the large cross-section of 991 daily series in their dataset. Our dataset covers five classes of financial assets: (i) the Commodities class; (ii) the Corporate Risk category; (iii) the Equities class; (iv) the Foreign Exchange Rates class and (v) the Government Securities.

The dataset includes up to 87 daily predictors in a daily frequency from 23-Oct-1959 to 24-Oct-2018 (14852 trading days) from the above five categories of financial assets. We remove series with fewer than ten years of data and time periods with no variables observed, which occurs for some series in the early part of the sample. For those years, we have less than 87 series. There are 39 commodity variables which include commodity indices, prices and futures, 16 corporate risk series, 9 equity series which include major US stock market indices and the 500 Implied Volatility, 16 government securities which include the federal funds rate, government treasury bills of securities from three months to ten years, and 7 foreign exchange variables which include the individual foreign exchange rates of major five US trading partners and two effective exchange rate. We choose these daily predictors because they are proposed in the literature as good predictors of economic growth.

We construct daily financial factors in a quarterly frequency in two steps. First, we use these daily financial time series to form factors at a daily frequency. The raw data used to form factors are always transformed to achieve stationarity and standardized before performing factor estimation (see generic description below). We re-estimate factors at each date in the sample recursively over time using the entire history of data available in real time prior to each out-of-sample forecast.

In the second step, we convert these daily financial indicators to quarterly weighted variables to form quarterly factors by selecting an optimal weighting scheme according to the method described below (see the weighting scheme section).

The data series used in this dataset are listed below in Table A8 by data source. The tables also list the transformation applied to each variable to make them stationary before generating factors. The transformations used to stationarize a time series are the same as those explained in the section "Monthly financial factor data".

Table A8— List of Daily Financial Dataset Variables

No.	Short Name	Source	Tran	Description
				Group 1: Commodities
1	GSIZSPT	Data Stream	Δln	S&P GSCI Zinc Spot - PRICE INDEX
2	GSSBSPT	Data Stream	Δln	S&P GSCI Sugar Spot - PRICE INDEX
3	GSSOSPT	Data Stream	Δln	S&P GSCI Soybeans Spot - PRICE INDEX
4	GSSISPT	Data Stream	Δln	S&P GSCI Silver Spot - PRICE INDEX
5	GSIKSPT	Data Stream	Δln	S&P GSCI Nickel Spot - PRICE INDEX
6	GSLCSPT	Data Stream	Δln	S&P GSCI Live Cattle Spot - PRICE INDEX
7	GSLHSPT	Data Stream	Δln	S&P GSCI Lean Hogs Index Spot - PRICE INDEX
8	GSILSPT	Data Stream	Δln	S&P GSCI Lead Spot - PRICE INDEX
9	GSGCSPT	Data Stream	Δln	S&P GSCI Gold Spot - PRICE INDEX
10	GSCTSPT	Data Stream	Δln	S&P GSCI Cotton Spot - PRICE INDEX
11	GSKCSPT	Data Stream	Δln	S&P GSCI Coffee Spot - PRICE INDEX
12	GSCCSPT	Data Stream	Δln	S&P GSCI Cocoa Index Spot - PRICE INDEX
13	GSIASPT	Data Stream	Δln	S&P GSCI Aluminum Spot - PRICE INDEX
14	SGWTSPT	Data Stream	Δln	S&P GSCI All Wheat Spot - PRICE INDEX
15	EIAEBRT	Data Stream	Δln	Europe Brent Spot FOB U$/BBL Daily
16	CRUDOIL	Data Stream	Δln	Crude Oil-WTI Spot Cushing U$/BBL - MID PRICE
17	LTICASH	Data Stream	Δln	LME-Tin 99.85% Cash U$/MT
18	CWFCS00	Data Stream	Δln	CBT-WHEAT COMPOSITE FUTURES CONT. - SETT. PRICE
19	CCFCS00	Data Stream	Δln	CBT-CORN COMP. CONTINUOUS - SETT. PRICE
20	CSYCS00	Data Stream	Δln	CBT-SOYBEANS COMP. CONT. - SETT. PRICE
21	NCTCS20	Data Stream	Δln	CSCE-COTTON #2 CONT.2ND FUT - SETT. PRICE
22	NSBCS00	Data Stream	Δln	CSCE-SUGAR #11 CONTINUOUS - SETT. PRICE
23	NKCCS00	Data Stream	Δln	CSCE-COFFEE C CONTINUOUS - SETT. PRICE
24	NCCCS00	Data Stream	Δln	CSCE-COCOA CONTINUOUS - SETT. PRICE
25	CZLCS00	Data Stream	Δln	ECBOT-SOYBEAN OIL CONTINUOUS - SETT. PRICE
26	COFC01	Data Stream	Δln	CBT-OATS COMP. TRc1 - SETT. PRICE
27	CLDCS00	Data Stream	Δln	CME-LIVE CATTLE COMP. CONTINUOUS - SETT. PRICE
28	CLGC01	Data Stream	Δln	CME-LEAN HOGS COMP. TRc1 - SETT. PRICE
29	NGCCS00	Data Stream	Δln	CMX-GOLD 100 OZ CONTINUOUS - SETT. PRICE
30	LAH3MTH	Data Stream	Δln	LME-Aluminium 99.7% 3 Months U$/MT
31	LED3MTH	Data Stream	Δln	LME-Lead 3 Months U$/MT
32	LNI3MTH	Data Stream	Δln	LME-Nickel 3 Months U$/MT

Table A8 (Cont'd)

No.	Short Name	Source	Tran	Description
33	LTI3MTH	Data Stream	Δln	LME-Tin 99.85% 3 Months U$/MT
34	PLNYD	www.macrotrends.net	Δln	Platinum Cash Price (U$ per troy ounce)
35	XPDD	www.macrotrends.net	Δln	Palladium (U$ per troy ounce)
36	CUS2D	www.macrotrends.net	Δln	Corn Spot Price (U$/Bushel)
37	SoybOil	www.macrotrends.net	Δln	Soybean Oil Price (U$/Pound)
38	OATSD	www.macrotrends.net	Δln	Oat Spot Price (US$/Bushel)
39	WTIOilFut	US EIA	Δln	Light Sweet Crude Oil Futures Price: 1St Expiring Contract Settlement ($/Bbl)
		Group 2: Equities		
40	S&PCOMP	Data Stream	Δln	S&P 500 COMPOSITE - PRICE INDEX
41	ISPCS00	Data Stream	Δln	CME-S&P 500 INDEX CONTINUOUS - SETT. PRICE
42	SP5EIND	Data Stream	Δln	S&P500 ES INDUSTRIALS - PRICE INDEX
43	DJINDUS	Data Stream	Δln	DOW JONES INDUSTRIALS - PRICE INDEX
44	CYMCS00	Data Stream	Δln	CBT-MINI DOW JONES CONTINUOUS - SETT. PRICE
45	NASCOMP	Data Stream	Δln	NASDAQ COMPOSITE - PRICE INDEX
46	NASA100	Data Stream	Δln	NASDAQ 100 - PRICE INDEX
47	CBOEVIX	Data Stream	lv	CBOE SPX VOLATILITY VIX (NEW) - PRICE INDEX
48	S&P500toVIX	Data Stream	Δln	S&P500/VIX
		Group 3: Corporate Risk		
49	LIBOR	FRED	Δlv	Overnight London Interbank Offered Rate (%)
50	1MLIBOR	FRED	Δlv	1-Month London Interbank Offered Rate (%)
51	3MLIBOR	FRED	Δlv	3-Month London Interbank Offered Rate (%)
52	6MLIBOR	FRED	Δlv	6-Month London Interbank Offered Rate (%)
53	1YLIBOR	FRED	Δlv	One-Year London Interbank Offered Rate (%)
54	1MEuro-FF	FRED	lv	1-Month Eurodollar Deposits (London Bid) (% P.A.) minus Fed Funds
55	3MEuro-FF	FRED	lv	3-Month Eurodollar Deposits (London Bid) (% P.A.) minus Fed Funds
56	6MEuro-FF	FRED	lv	6-Month Eurodollar Deposits (London Bid) (% P.A.) minus Fed Funds
57	APFNF-AANF	Data Stream	lv	1-Month A2/P2/F2 Nonfinancial Commercial Paper (NCP) (% P. A.) minus 1-Month Aa NCP (% P.A.)
58	APFNF-AAF	Data Stream	lv	1-Month A2/P2/F2 NCP (% P.A.) minus 1-Month Aa Financial Commercial Paper (% P.A.)
59	TED	Data Stream, FRED	lv	3Month Tbill minus 3-Month London Interbank Offered Rate (%)
60	MAaa-10YTB	Data Stream	lv	Moody Seasoned Aaa Corporate Bond Yield (% P.A.) minus Y10-Tbond
61	MBaa-10YTB	Data Stream	lv	Moody Seasoned Baa Corporate Bond Yield (% P.A.) minus Y10-Tbond
62	MLA-10YTB	Data Stream, FRED	lv	Merrill Lynch Corporate Bonds: A Rated: Effective Yield (%) minus Y10-Tbond
63	MLAA-10YTB	Data Stream, FRED	lv	Merrill Lynch Corporate Bonds: Aa Rated: Effective Yield (%) minus Y10-Tbond
64	MLAAA-10YTB	Data Stream, FRED	lv	Merrill Lynch Corporate Bonds: Aaa Rated: Effective Yield (%) minus Y10-Tbond
		Group 4: Treasuries		

Table A8 (Cont'd)

No.	Short Name	Source	Tran	Description
65	FRFEDFD	Data Stream	Δlv	US FED FUNDS EFF RATE (D) - MIDDLE RATE
66	FRTBS3M	Data Stream	Δlv	US T-BILL SEC MARKET 3 MONTH (D) - MIDDLE RATE
67	FRTBS6M	Data Stream	Δlv	US T-BILL SEC MARKET 6 MONTH (D) - MIDDLE RATE
68	FRTCM1Y	Data Stream	Δlv	US TREASURY CONST MAT 1 YEAR (D) - MIDDLE RATE
69	FRTCM10	Data Stream	Δlv	US TREASURY CONST MAT 10 YEAR (D) - MIDDLE RATE
70	6MTB-FF	Data Stream	lv	6-month treasury bill market bid yield at constant maturity (%) minus Fed Funds
71	1YTB-FF	Data Stream	lv	1-year treasury bill yield at constant maturity (% P.A.) minus Fed Funds
72	10YTB-FF	Data Stream	lv	10-year treasury bond yield at constant maturity (% P.A.) minus Fed Funds
73	6MTB-3MTB	Data Stream	lv	6-month treasury bill yield at constant maturity (% P.A.) minus 3M-Tbills
74	1YTB-3MTB	Data Stream	lv	1-year treasury bill yield at constant maturity (% P.A.) minus 3M-Tbills
75	10YTB-3MTB	Data Stream	lv	10-year treasury bond yield at constant maturity (% P.A.) minus 3M-Tbills
76	BKEVEN05	FRB	lv	US Inflation compensation: continuously compounded zero-coupon yield: 5-year (%)
77	BKEVEN10	FRB	lv	US Inflation compensation: continuously compounded zero-coupon yield: 10-year (%)
78	BKEVEN1F4	FRB	lv	BKEVEN1F4
79	BKEVEN1F9	FRB	lv	BKEVEN1F9
80	BKEVEN5F5	FRB	lv	US Inflation compensation: coupon equivalent forward rate: 5-10 years (%)
		Group 5: Foreign Exchange (FX)		
81	US_CWBN	Data Stream	Δln	US NOMINAL DOLLAR BROAD INDEX - EXCHANGE INDEX
82	US_CWMN	Data Stream	Δln	US NOMINAL DOLLAR MAJOR CURR INDEX - EXCHANGE INDEX
83	US_CSFR2	Data Stream	Δln	CANADIAN $ TO US $ NOON NY - EXCHANGE RATE
84	EU_USFR2	Data Stream	Δln	EURO TO US$ NOON NY - EXCHANGE RATE
85	US_YFR2	Data Stream	Δln	JAPANESE YEN TO US $ NOON NY - EXCHANGE RATE
86	US_SFFR2	Data Stream	Δln	SWISS FRANC TO US $ NOON NY - EXCHANGE RATE
87	US_UKFR2	Data Stream	Δln	UK POUND TO US $ NOON NY - EXCHANGE RATE

From Daily to Quarterly Factors: Weighting Schemes

After we obtain daily financial factors $\mathbf{G}_{D,t}$, we use some weighting schemes proposed in the literature about Mixed Data Sampling (MIDAS) regressions to form quarterly factors, $\mathbf{G}^{Q}_{D,t}$. Denote by G^D_t a factor in a daily frequency formed from the daily financial dataset and denote by G^Q_t a quarterly aggregate of the corresponding daily factor time series. Let $G^D_{N_D-j,d_t,t}$ denote the value of a daily factor in the j^{th} day counting backwards from the survey deadline d_t in quarter t. Hence, the day d_t of quarter t corresponds with $j = 0$ and is therefore $G^D_{N_D,d_t,t}$. For simplicity, we suppress the subscript d_t thus

$G^D_{N_D-j,d_t,t} \equiv G^D_{N_D-j,t}$.

We compute the quarterly aggregate of a daily financial factor as a weighted average of observations over the N_D business days before the survey deadline. This means that the forecasters's information set includes daily financial data up to the previous N_D business days before the survey deadline. G_t^Q is defined as:

$$G_t^Q(\mathbf{w}) \equiv \sum_{j=1}^{N_D} w_i G^D_{N_D-j,t}$$

where w_i is a weight. We consider the following three types of weighting schemes to convert daily factor observations to quarterly. Each weighting scheme weights information by some function of the number of days prior to the survey deadline.

1. $w_i = 1$ for $i = 1$ and $w_i = 0$ otherwise. This weighting scheme places all weight on data in the last business day before the survey deadline for that quarter and zero weight on any data prior to that day.

2. $w_i = \frac{\theta^j}{\sum_{j=1}^{N_D}\theta^j}$ where we consider a range of θ^j for $\theta^j = (0.1, 0.2, 0.3, 0.7, 0.8, 0.9, 1)'$. The smaller is θ^j, the more rapidly information prior to the survey deadline day is down-weighted. This down-weighting is progressive but not nonmonotone. $\theta^j = 1$ is a simple average of the observations across all days in the quarter.

3. The third parameterization has two parameters, or $\theta^D = (\theta_1, \theta_2)'$ and allows for nonmonotone weighting of past information:

$$w(i; \theta_1, \theta_2) = \frac{f\left(\frac{i}{N_D}, \theta_1; \theta_2\right)}{\sum_{j=1}^{N_D} f\left(\frac{j}{N_D}, \theta_1; \theta_2\right)}$$

where:

$$f(x, a, b) = \frac{x^{a-1}(1-x)^{b-1}\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$$

$$\Gamma(a) = \int_0^\infty e^{-x}x^{a-1}dx$$

The weights $w(i; \theta_1, \theta_2)$ are the Beta polynomial MIDAS weights of Ghysels et al. (2007), which are based on the Beta function. This weighting scheme is flexible enough to generate a range of possible shapes with only two parameters.

We consider these possible weighting schemes and choose the optimal weighting scheme $\mathbf{w}^*$ from 24 weighting schemes for a daily financial factor G_t^D by minimizing the sum of square residuals in a regression of $y_{j,t+h}$ on $G_t^Q(\mathbf{w})$:

$$y_{j,t+h} = a + b \cdot \underbrace{\sum_{i=1}^{N_D} w_i G^D_{N_D-i,t}}_{G_t^Q(\mathbf{w})} + u_{t+h}.$$

This is done in real time using recursive regressions. We re-estimate the weights at each date in the sample recursively over time using the entire history of data available in real time prior to each out-out-sample forecast.

We assume that $N_D = 14$ which implies that forecasters use daily information in at most the past two weeks before the survey deadline. The process is repeated for each daily financial factor in $\mathbf{G}_{D,t}$ to form quarterly factors $\mathbf{G}^Q_{D,t}$.

Estimation and Machine Learning

MACHINE ALGORITHM DETAILS

The model to be estimated is

$$y_{j,t+h} = \mathcal{X}_t' \boldsymbol{\beta}^{(i)}_{jh} + \epsilon_{jt+h}.$$

It should be noted that the most recent observation on the left-hand-side is generally available in real time only with a one-period lag, thus the forecasting estimations can only be run with data over a sample that stops one period later than today in real time. $\mathcal{X}_t$ always denotes the most recent data that would have been in real time prior to the date on which the forecast was submitted. The coefficients $\boldsymbol{\beta}^{(i)}_{jh,t}$ are estimated using the Elastic Net (EN) estimator, which depend on regularization parameter parameters $\boldsymbol{\lambda}^{(i)}_t = \left(\lambda^{(i)}_{1t}, \lambda^{(i)}_{2t}\right)'$ (See the next section for a description of EN). The procedure involves iterating on the steps given in the main text.

We allow the machine to additionally learn about whether the coefficient on the survey forecast should be shrunk toward zero or toward unity. Recall that the machine forecast for the ith percentile is

$$\mathbb{E}^{(i)}_t \left(y_{j,t+h}\right) \equiv \hat{\alpha}^{(i)}_{jh} + \hat{\beta}^{(i)}_{j\mathbb{F}} \mathbb{F}^{(i)}_t \left[y_{j,t+h}\right] + \hat{\mathbf{B}}^{(i)\prime}_{j\mathcal{Z}} \mathcal{Z}_{jt}.$$

If the machine model is implemented as an estimation with using forecast errors as the dependent variable, i.e.,

$$y_{j,t+h} - \mathbb{F}^{(i)}_t \left[y_{j,t+h}\right] = \alpha^{(i)}_{jh} + \beta^{(i)}_{jh\mathbb{F}} \mathbb{F}^{(i)}_t \left[y_{j,t+h}\right] + \mathbf{B}^{(i)\prime}_{jh\mathcal{Z}} \mathcal{Z}_t + \epsilon_{jt+h}, \tag{A2}$$

the machine efficient benchmark is characterized by $\beta^{(i)}_{j\mathbb{F}} = 0$; $\mathbf{B}^{(i)}_{jh\mathcal{Z}} = \mathbf{0}$;$\alpha^{(i)}_{jh} = 0$. Because EN shrinks estimated coefficients toward zero, this results in shrinkage of $\beta^{(i)}_{jh\mathbb{F}}$ toward unity. In this case the machine forecast is given by

$$\mathbb{E}^{(i)}_t \left(y_{j,t+h}\right) \equiv \hat{\alpha}^{(i)}_{jh} + \left(\hat{\beta}^{(i)}_{j\mathbb{F}} + 1\right) \mathbb{F}^{(i)}_t \left[y_{j,t+h}\right] + \hat{\mathbf{B}}^{(i)\prime}_{j\mathcal{Z}} \mathcal{Z}_{jt}.$$

By contrast, if the machine forecast is implemented by running the specification

$$y_{j,t+h} = \alpha^{(i)}_{jh} + \beta^{(i)}_{jh\mathbb{F}} \mathbb{F}^{(i)}_t \left[y_{j,t+h}\right] + \mathbf{B}^{(i)\prime}_{j\mathcal{Z}} \mathcal{Z}_t + \epsilon_{jt+h},$$

then $\beta^{(i)}_{jh\mathbb{F}}$ is shrunk toward zero and the algorithm will typically place less weight on the survey forecast than the specification (A2). In the implementation, we allow the machine to choose which specification to run over time by having it pick the one that that minimizes the mean-square loss function $\mathcal{L}\left(\boldsymbol{\lambda}^{(i)}_t, T_E, T_V\right)$ over psuedo out-of-sample forecast errors in every training sample.

On rare occasions, one or more of the explanatory variables used in the machine forecast specification assumes a value that is order of magnitudes different from its historical value. This is usually indicative of a measurement problem in the raw data. We therefore program the machine to detect in real-time whether its forecast is an extreme outlier, and in that case to discard the forecast replacing it with the survey forecast. Specifically, at each t, the machine forecast $\mathbb{E}^{(i)}_t\left(y_{j,t+h}\right)$ is set to be the survey forecast $\mathbb{F}^{(i)}_t\left[y_{j,t+h}\right]$ whenever the former is five or more standard deviations above its own rolling mean over the most recent 20 quarters. This outlier option is rarely used in the estimation.

Elastic Net Estimator

We use the Elastic Net (EN) estimator, which combines Least Absolute Shrinkage and Selection Operator (LASSO) and ridge type penalties. LASSO. Suppose our goal is to estimate the coefficients in the linear model:

$$y_{j,t+h} = \alpha_{jh} + \beta_{jh\mathbb{F}}\mathbb{F}^{(i)}_t\left[y_{j,t+h}\right] + \underbrace{\mathbf{B}_{j\mathcal{Z}}}_{qr\times qr}\mathcal{Z}_{jt} + \epsilon_{jt+h}$$

Collecting all the independent variables and coefficients into a single matrix and vector, the model can be written as:

$$y_{j,t+h} = \mathcal{X}'_{tj}\boldsymbol{\beta}_{jh} + \epsilon_{jt+h}$$

where $\mathcal{X}_t = \left(1, \mathcal{X}_{1t,...,}\mathcal{X}_{Kt}\right)'$ collects all the independent variable observations $\left(\mathbb{F}^{(i)}_t\left[y_{j,t+h}\right], \mathcal{Z}_{jt}\right)$ into a vector with "1" and $\boldsymbol{\beta}_{jh}\ \boldsymbol{\mathcal{D}}\ \left(\alpha_{jh}, \beta_{jh\mathbb{F}}, \text{vec}\left(\mathbf{B}_{jh\mathcal{Z}}\right)\right)' \equiv \left(\beta_0, \beta_1, ...\beta_K\right)'$ collects all the coefficient. It is customary to standardize the elements of $\mathcal{X}_t$ such that sample means are zero and sample standard deviations are unity. The coefficient estimates are then put back in their original scale by multiplying the slope coefficients by their respective standard deviations, and adding back the mean (scaled by slope coefficient over standard deviation.)

The EN estimator incorporates both an L_1 and L_2 penalty:

$$\hat{\boldsymbol{\beta}}^{\text{EN}} = \underset{\beta_0,\beta_1,...,\beta_k}{\text{argmin}}\left\{\sum_{\tau=1}^{T}\left(y_{j,\tau+h} - \mathcal{X}'_\tau\boldsymbol{\beta}^{(i)}_{jh}\right)^2 + \underbrace{\lambda^{(i)}_1\sum_{j=1}^{k}\left|\boldsymbol{\beta}_{jh}\right|}_{\text{LASSO}} + \underbrace{\lambda^{(i)}_2\sum_{j=1}^{k}\boldsymbol{\beta}^2_{jh}}_{\text{ridge}}\right\}$$

By minimizing the MSE over the training samples, we choose the optimal $\lambda_1^{(i)}$ and $\lambda_2^{(i)}$ values simultaneously.

Dynamic Factor Estimation

Let $x_t^C = (x_{1t}^C, \ldots, x_{Nt}^C)'$ generically denote a dataset of economic information in some category C that is available for real-time analysis. It is assumed that x_t^C has been suitably transformed (such as by taking logs and differencing) so as to render the series stationary. We assume that x_{it}^C has an approximate factor structure taking the form

$$x_{it}^C = \Lambda_i^{C\prime}\mathbf{G}_t^C + e_{it}^X,$$

where $\mathbf{G}_t^C$ is an $r_G \times 1$ vector of latent common factors ("diffusion indexes"), Λ_i^C is a corresponding $r_C \times 1$ vector of latent factor loadings, and e_{it}^X is a vector of idiosyncratic errors.[6] The number of factors r_G is typically significantly smaller than the number of series, N, which facilitates the use of very large datasets. Additional factors to account for nonlinearities are formed by including polynomial functions of $\mathbf{G}_t^C$, and by including factors formed from polynomials of the raw data.

We re-estimate factors at each date in the sample recursively over time using the entire history of data available in real time prior to each out-out-sample forecast. Let x_{it} denote the ith variable in a large dataset. The following steps are taken in forming the macro, financial, and daily factors:

1) Remove outlier values from a series, defined as values whose distance from the median is greater than ten times the interquartile range.

2) Scale each series according to the procedure proposed by Huang et al. (2017). We run the following regression for each variable x_{it}:

$$y_{jt+h} = \beta_{jh,i,0} + \beta_{jh,i,x} x_{it} + \nu_{j,i,t+h}.$$

Then, we form a new dataset of variables $\hat{\beta}_{jh,i,x} x_{it}$ where $\hat{\beta}_{jh,i,x}$ denotes the OLS estimate of $\beta_{jh,i,x}$. These "scaled" variables are standardized and denoted $\widetilde{x}_{it}$.

3) Throughout, the factors are estimated over $\widetilde{x}_{it}$ by the method of static principal components (PCA). The approach we consider is to posit that $\widetilde{x}_{it}$ has a factor structure taking the form

$$\widetilde{x}_{it} = \lambda_i' \mathbf{G}_t + e_{it}, \tag{A3}$$

where $\mathbf{G}_t$ is a $r \times 1$ vector of latent common factors, λ_i is a corresponding $r \times 1$ vector of latent factor loadings, and e_{it} is a vector of idiosyncratic errors.[7] Specifically, the $T \times r$ matrix $\hat{g}_t$ is $\sqrt{T}$ times the r eigenvectors corresponding to the r

[6] In an approximate dynamic factor structure, the idiosyncratic errors e_{it}^X are permitted to have a limited amount of cross-sectional correlation.

[7] We consider an *approximate* dynamic factor structure, in which the idiosyncratic errors e_{it} are permitted to have a

largest eigenvalues of the $T \times T$ matrix $\widetilde{x}\widetilde{x}'/(TN_{\widetilde{x}})$ in decreasing order, where T is the number of time periods and $N_{\widetilde{x}}$ is the number of variables in the large dataset. The optimal number of common factors, r is determined by the panel information criteria developed in Bai and Ng (2002). To handle missing values in any series, we use an expectation-maximization (EM) algorithm by filling with an initial guess and forming factors, using (A3) to update the guess with $\mathbb{E}(\widetilde{x}_{it}) = \mathbb{E}\left(\lambda_i' \hat{g}_t\right)$, and iterating until the successive values for $\mathbb{E}(\widetilde{x}_{it})$ are arbitrarily close.

4) Collect the common factors into the matrix $\mathbf{G}_{raw}$, where each principle component is a column.

5) Square the raw variables and repeat steps 2 through 5. Collect the common factors from squared data into a matrix $\mathbf{G}_{sqr}$, where component is a column.

6) Square the first factor in $\mathbf{G}_{raw}$, and call this $\mathbf{G}^2_{raw1}$.

7) Our matrix of factors is $[\mathbf{G}_{raw}, \mathbf{G}_{sqr1}, \mathbf{G}^2_{raw1}]$, where $\mathbf{G}_{sqr1}$ is the first column of $\mathbf{G}_{sqr}$.

For macro factors, we use all of the variables listed in Table A6. After step 1 above, an additional step of removing missing variables and observations is needed for the macro variables. We remove series with fewer than seven years of data and time periods with less than fifty-percent of variables observed, which occur in the early part of the sample. Furthermore, we lag variables with missing data in the final observation whenever more than twenty-percent of variables are missing data in the last observation.[8]

For the financial factors, we use all of the variables listed in Table A7, and no additional steps are performed beyond those described above.

ECONOMIC NAMES OF FACTORS

Any labeling of the factors is imperfect because each is influenced to some degree by all the variables in the large dataset and the orthogonalization means that no one of them will correspond exactly to a precise economic concept like output or unemployment. Following Ludvigson and Ng (2009), we relate the factors to the underlying variables in the large dataset. For each time period in our evaluation sample, we compute the marginal R^2 from regressions of each of the individual series in the panel dataset onto each factor, one at a time. Each series $\widetilde{x}_{it}$ is assigned the group name in the data appendix tables naming all series, e.g., non-farm payrolls are part of the Employment group (EMP). If

limited amount of cross-sectional correlation. The approximate factor specification limits the contribution of the idiosyncratic covariances to the total variance of x as N gets large:

$$N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|E\left(e_{it}e_{jt}\right)\right| \leq M,$$

where M is a constant.

[8]Even though the EM algorithm is designed to estimate missing observations, it does not perform well when there are too many missing observations at a single point in time.

series $\widetilde{x}_{it}$ has the highest average marginal R^2 over all evaluation periods for factor G_{kt}, we label G_{kt} according to the group to which $\widetilde{x}_{it}$ belongs, e.g., G_{kt} is an Employment factor. We further normalize the sign of each factor so that an increase in the factor indicates an increase in $\widetilde{x}_{it}$. Thus, in the example above, an increase in G_{kt} would indicate a rise in non-farm payrolls. Table A9 reports the series with largest average marginal R^2 for each factor of each large dataset.

Table A9— Economic Interpretation of the Factors

	Series with Largest R^2	
	Macro Factors	Label
$G_{1,M,t}$	Nonfarm Payrolls	Macro Factor: Employment
$G_{2,M,t}$	Interest paid by consumers	Macro Factor: Money and Credit
$G_{3,M,t}$	Agg. Weekly hours - Service-producing	Macro Factor: Employment.
$G_{4,M,t}$	Agg. Weekly hours - Good-producing	Macro Factor: Employment
$G_{5,M,t}$	Nonborrowed Reserves	Macro Factor: Money and Credit
$G_{6,M,t}$	Housing Starts	Macro Factor: Housing
$G_{7,M,t}$	Change in private inventories	Macro Factor: Orders and Investment
$G_{8,M,t}$	PCE: Service	Macro Factor: Consumption
	Financial Factors	
$G_{1,F,t}$	D_log(P)	Financial Factor: Prices, Yield, Dividends
$G_{2,F,t}$	SMB	Financial Factor: Equity Risk Factors
$G_{3,F,t}$	HML	Financial Factor: Equity Risk Factors
$G_{4,F,t}$	R15_R11	Financial Factor: Equity Risk Factors
$G_{5,F,t}$	D_DIVreinvest	Financial Factor: Prices, Yield, Dividends
$G_{6,F,t}$	Smoke	Financial Factor: Industries
$G_{7,F,t}$	UMD	Financial Factor: Equity Risk Factors
$G_{8,F,t}$	Telcm	Financial Factor: Industries
	Daily Factors	
$G_{1,D,t}$	ECBOT-SOYBEAN OIL	Daily Factor: Commodities
$G_{2,D,t}$	A Rated minus Y10 Tbond	Daily Factor: Corporate Risk
$G_{3,D,t}$	6-month US T-bill	Daily Factor: Treasuries
$G_{4,D,t}$	6-month treasury bill minus 3M-Tbills	Daily Factor: Treasuries
$G_{5,D,t}$	CBT-MINI DOW JONES	Daily Factor: Equities
$G_{6,D,t}$	Corn	Daily Factor: Commodities
$G_{7,Dt}$	APFNF-AAF	Daily Factor: Corporate Risk
$G_{8,D,t}$	US nominal dollar broad index	Daily Factor: FX

Note: This table reports the series with largest marginal R^2 for the factor specified in the first column. The marginal R^2 is computed from regressions of each of the individual series onto the factor, one at a time, for the time period that the factor shows up as relevant for the median bias.

PREDICTOR VARIABLES

The vector $\mathbf{Z}_{jt} \equiv \left(y_{j,t}, \mathbf{G}'_t, \mathbf{W}'_{jt}\right)'$ is an $r = 1 + r_G + r_W$ vector which collects the data at time t with $\mathcal{Z}_{jt} \equiv \left(y_{j,t}, ..., y_{j,t-p_y}, \mathbf{G}'_t, ..., \mathbf{G}'_{t-p_G}, \mathbf{W}'_{jt}, ..., \mathbf{W}'_{jt-p_W}\right)'$ a vector of contemporaneous and lagged values of $\mathbf{Z}_{jt}$, where p_y, p_G, p_W denote the total number of lags of $y_{j,t}$, $\boldsymbol{\Theta}'_t$, $\mathbf{W}'_{jt}$, respectively. Superscript (i) refers to the ith forecaster, where i denotes either the mean "*mean*" or an ith percentile value of the forecast distribution, i.e., "65" is the 65th percentile. The predictors below are listed as elements of $y_{j,t}$, $\boldsymbol{\Theta}'_{jt}$, or $\mathbf{W}'_{jt}$ for different surveys and variables.

SPF INFLATION

For y_j equal to inflation the forecasting model considers the following variables.
In $\mathbf{W}'_{jt}$:

1) $\mathbb{F}^{(i)}_{jt-k}\left[y_{jt+h-k}\right]$, lagged values of the ith type's forecast, where $k = 1, 2, ...$
2) $\mathbb{F}^{(s\neq i)}_{jt-1}\left[y_{jt+h-1}\right]$, lagged values other type's forecasts, $s \neq i$
3) $\text{var}_N\left(\mathbb{F}^{(\cdot)}_{t-1}\left[y_{jt+h-1}\right]\right)$, where $\text{var}_N(\cdot)$ denotes the cross-sectional variance of lagged survey forecasts
4) $\text{skew}_N\left(\mathbb{F}^{(\cdot)}_{t-1}\left[y_{jt+h-1}\right]\right)$, where $\text{skew}_N(\cdot)$ denotes the cross-sectional skewness of lagged survey forecasts
5) Trend inflation measured as $\overline{\pi}_{t-1} = \begin{cases} \rho\overline{\pi}_{t-2} + (1-\rho)\pi_{t-1}, \ \rho = 0.95 & \text{if } t < \text{1991:Q4} \\ \text{CPI10}_{t-1} & \text{if } t \geq \text{1991:Q4}, \end{cases}$
where CPI10 is the median SPF forecast of annualized average inflation over the current and next nine years. Trend inflation is intended to capture long-run trends. When long-run forecasts of inflation are not available, as is the case pre-1991:Q4, we us a moving average of past inflation.
6) $\widetilde{GDP}_{t-1}$ = detrended gross domestic product, defined as the residual from a regression of GDP_{t-1} on a constant and the four most recent values of GDP as of date $t-8$. See Hamilton (2018).
7) $\widetilde{EMP}_{t-1}$ = detrended employment, defined as the residual from a regression of EMP_{t-1} on a constant and the four most recent values of EMP as of date $t-8$. See Hamilton (2018).
8) $\mathbb{N}^{(i)}_t\left[\pi_{t,t-h}\right]$ = Nowcast as of time t of the ith percentile of inflation over the period $t-h$ to t.

Lags of the dependent variable:

1) $y_{t-1,t-h-1}$ one quarter lagged annual inflation.

The factors in $\boldsymbol{\Theta}'_{jt}$ include factors formed from three large datasets separately:

1) $\mathbf{G}_{M,t-k}$, for $k = 0, 1$ are factors formed from a real-time macro dataset $\mathcal{D}^M$ with 92 real-time macro series; includes both monthly and quarterly series, with monthly series converted to quarterly according to the method described in the data appendix.

2) $\mathbf{G}_{F,t-k}$, for $k = 0, 1$ are factors formed from a financial data set $\mathcal{D}^F$ with 147 monthly financial series.

3) $\mathbf{G}^Q_{D,t}$, are quarterly factors formed from a daily financial dataset $\mathcal{D}^D$ of 87 daily financial indicators. The raw daily series are first converted to daily factors $\mathbf{G}_{D,t}(\mathbf{w})$ and the daily factors are aggregated up to quarterly observations $\mathbf{G}^Q_{D,t}(\mathbf{w})$ using a weighted average of daily factors, with the weights $\mathbf{w}$ dependent on two free parameters that are chosen to minimize the sum of squared residuals in a regression of $y_{j,t+h}$ on $\mathbf{G}_{D,t}(\mathbf{w})$.

The 92 macro series in $\mathcal{D}^M$ are selected to represent broad categories of macroeconomic time series. The majority of these are real activity measures: real output and income, employment and hours, consumer spending, housing starts, orders and unfilled orders, compensation and labor costs, and capacity utilization measures. The dataset also includes commodity and price indexes and a handful of bond and stock market indexes, and foreign exchange measures. The financial dataset $\mathcal{D}^f$ is an updated monthly version of the of 147 variables comprised solely of financial market time series used in Ludvigson and Ng (2007). These data include valuation ratios such as the dividend-price ratio and earnings-price ratio, growth rates of aggregate dividends and prices, default and term spreads, yields on corporate bonds of different ratings grades, yields on Treasuries and yield spreads, and a broad cross-section of industry, size, book-market, and momentum portfolio equity returns.[9] The 87 daily financial indicators in $\mathcal{D}^D$ include daily time series on commodities spot prices and futures prices, aggregate stock market indexes, volatility indexes, credit spreads and yield spreads, and exchange rates.

SPF GDP Growth

For y_j equal to GDP growth the forecasting model considers the following variables. In $\mathbf{W}'_{jt}$

1) $\mathbb{F}^{(i)}_{jt-k}\left[y_{jt+h-k}\right]$, lagged values of the ith type's forecast, where $k = 1, 2, ...$

2) $\mathbb{F}^{(s\neq i)}_{jt-1}\left[y_{jt+h-1}\right]$, lagged values other type's forecasts, $s \neq i$

3) $\mathrm{var}_N\left(\mathbb{F}^{(\cdot)}_{t-1}\left[y_{jt+h-1}\right]\right)$, where $\mathrm{var}_N(\cdot)$ denotes the cross-sectional variance of forecasts

[9]A detailed description of the series is given in the Data Appendix of the online supplementary file at www.sydneyludvigson.com/s/ucc_data_appendix.pdf

4) $\text{skew}_N\left(\mathbb{F}_{t-1}^{(\cdot)}\left[y_{jt+h-1}\right]\right)$, where $\text{skew}_N(\cdot)$ denotes the cross-sectional skewness of forecasts

5) $\widetilde{GDP}_{t-1}$ = detrended gross domestic product, defined as the residual from a regression of GDP_{t-1} on a constant and the four most recent values of GDP as of date $t-8$. See Hamilton (2018).

6) $\widetilde{EMP}_{t-1}$ = detrended employment, defined as the residual from a regression of EMP_{t-1} on a constant and the four most recent values of EMP as of date $t-8$. See Hamilton (2018).

7) $\mathbb{N}_t^{(i)}\left[y_{t,t-h}\right]$ = Nowcast as of time t of the ith percentile of GDP growth over the period $t-h$ to t.

8) VXO_t, defined as CBOE S&P 100 volatility index. We also include its squared and cubic terms, VXO_t^2, and VXO_t^3.

Lags of the dependent variable:

1) $y_{j,t-1,t-h-1}, y_{j,t-2,t-h-2}$ one and two quarter lagged annual GDP growth.

The factors in $\boldsymbol{\Theta}'_{jt}$ include factors formed from three large datasets separately:

1) $\mathbf{G}_{M,t-k}$, for $k=0,1$ are factors formed from a real-time macro dataset $\mathcal{D}^M$ with 92 real-time macro series; includes both monthly and quarterly series, with monthly series converted to quarterly according to the method described in the data appendix.

2) $\mathbf{G}_{F,t-k}$, for $k=0,1$ are factors formed from a financial data set $\mathcal{D}^F$ with 147 monthly financial series.

3) $\mathbf{G}_{D,t}^Q$, are quarterly factors formed from a daily financial dataset $\mathcal{D}^D$ of 87 daily financial indicators. The raw daily series are first converted to daily factors $\mathbf{G}_{D,t}(\mathbf{w})$ and the daily factors are aggregated up to quarterly observations $\mathbf{G}_{D,t}^Q(\mathbf{w})$ using a weighted average of daily factors, with the weights $\mathbf{w}$ dependent on two free parameters that are chosen to minimize the sum of squared residuals in a regression of $y_{j,t+h}$ on $\mathbf{G}_{D,t}(\mathbf{w})$.

The 92 macro series in $\mathcal{D}^M$ are selected to represent broad categories of macroeconomic time series. The majority of these are real activity measures: real output and income, employment and hours, consumer spending, housing starts, orders and unfilled orders, compensation and labor costs, and capacity utilization measures. The dataset also includes commodity and price indexes and a handful of bond and stock market indexes, and foreign exchange measures. The financial dataset $\mathcal{D}^f$ is an updated monthly version of the of 147 variables comprised solely of financial market time series used in Ludvigson and Ng (2007). These data include valuation ratios such as the dividend-price ratio and earnings-price ratio, growth rates of aggregate dividends and prices, default and term spreads, yields on corporate bonds of different ratings grades, yields on Treasuries and yield spreads, and a broad cross-section of industry, size, book-market, and momentum portfolio equity returns.[10] The 87 daily financial indicators in $\mathcal{D}^D$ include daily time series on commodities spot prices and futures prices, aggregate stock market indexes, volatility indexes, credit spreads and yield spreads, and exchange rates.

SOC INFLATION

For consistency, the predictors for the SOC inflation forecasts are constructed similarly to those of the SPF inflation forecasts. Again, consider the following forecast regression,

$$y_{j,t+h} = \alpha_{jh} + \beta_{jh\mathbb{F}}\mathbb{F}_{j,t}^{MS,(i)}\left[y_{j,t+h}\right] + \underbrace{\mathbf{B}_{jh\mathcal{Z}}}_{1xq}\mathcal{Z}_{jt} + \epsilon_{jt+h},$$

where the variables are defined as above, and i is either the mean "*mean*" or an ith percentile value of the forecast distribution. We denote forecasts from the SPF using $\mathbb{F}_{js}^{SPF,(i)}[\cdot]$ and from the Michigan Survey using $\mathbb{F}_{js}^{MS,(i)}[\cdot]$.

In $\mathbf{W}'_{jt}$:

1) $\mathbb{F}_{jt-1}^{SPF,(\mu)}\left[y_{jt+h-1}\right]$, the mean SPF forecast for CPI.
2) $\mathbb{F}_{jt-1}^{SPF,(50)}\left[y_{jt+h-1}\right]$, the $50th$ percentile SPF forecast for CPI.
3) $\mathbb{F}_{jt-1}^{SPF,(25)}\left[y_{jt+h-1}\right]$, the $25th$ percentile SPF forecast for CPI.
4) $\mathbb{F}_{jt-1}^{SPF,(75)}\left[y_{jt+h-1}\right]$, the $75th$ percentile SPF forecast for CPI.
5) $\mathrm{var}_N\left(\mathbb{F}_{t-1}^{SPF,(\cdot)}\left[y_{jt+h-1}\right]\right)$, the cross-sectional variance of SPF forecasts of CPI.
6) $\mathrm{skew}_N\left(\mathbb{F}_{t-1}^{SPF,(\cdot)}\left[y_{jt+h-1}\right]\right)$, the cross-sectional skewness of SPF forecasts of CPI.
7) Trend inflation measured as $\overline{\pi}_{t-1} = \begin{cases} \rho\overline{\pi}_{t-2} + (1-\rho)\pi_{t-1}, \ \rho = 0.95 & \text{if } t < \text{1991:Q4} \\ \text{CPI10}_{t-1} & \text{if } t \geq \text{1991:Q4}, \end{cases}$ where CPI10 is the median SPF forecast of annualized average inflation over the current and next nine years. Trend inflation is intended to capture long-run trends. When long-run forecasts of inflation are not available, as is the case pre-1991:Q4, we us a moving average of past inflation.
8) $\widetilde{GDP}_{t-1}$ = detrended gross domestic product, defined as the residual from a regression of GDP_{t-1} on a constant and the four most recent values of GDP as of date $t-8$. See Hamilton (2018).

[10] A detailed description of the series is given in the Data Appendix of the online supplementary file at www.sydneyludvigson.com/s/ucc_data_appendix.pdf

9) $\widetilde{EMP}_{t-1}$ = detrended employment, defined as the residual from a regression of EMP_{t-1} on a constant and the four most recent values of EMP as of date $t-8$. See Hamilton (2018).

Lags of dependent variables:

1) $y_{t-1,t-h-1}$ one quarter lagged annual CPI inflation.

The factors in $\boldsymbol{\Theta}'_{jt}$ include factors formed from three large datasets separately:

1) $\mathbf{G}_{M,t-k}$, for $k = 0, 1$ are factors formed from a real-time macro dataset $\mathcal{D}^M$ with 92 real-time macro series; includes both monthly and quarterly series, with monthly series converted to quarterly according to the method described in the data appendix.

2) $\mathbf{G}_{F,t-k}$, for $k = 0, 1$ are factors formed from a financial data set $\mathcal{D}^F$ with 147 monthly financial series.

3) $\mathbf{G}^Q_{D,t}$, are quarterly factors formed from a daily financial dataset $\mathcal{D}^D$ of 87 daily financial indicators. The raw daily series are first converted to daily factors $\mathbf{G}_{D,t}(\mathbf{w})$ and the daily factors are aggregated up to quarterly observations $\mathbf{G}^Q_{D,t}(\mathbf{w})$ using a weighted average of daily factors, with the weights $\mathbf{w}$ dependent on two free parameters that are chosen to minimize the sum of squared residuals in a regression of $y_{j,t+h}$ on $\mathbf{G}_{D,t}(\mathbf{w})$.

The 92 macro series in $\mathcal{D}^M$ are selected to represent broad categories of macroeconomic time series. The majority of these are real activity measures: real output and income, employment and hours, consumer spending, housing starts, orders and unfilled orders, compensation and labor costs, and capacity utilization measures. The dataset also includes commodity and price indexes and a handful of bond and stock market indexes, and foreign exchange measures. The financial dataset $\mathcal{D}^f$ is an updated monthly version of the of 147 variables comprised solely of financial market time series used in Ludvigson and Ng (2007). These data include valuation ratios such as the dividend-price ratio and earnings-price ratio, growth rates of aggregate dividends and prices, default and term spreads, yields on corporate bonds of different ratings grades, yields on Treasuries and yield spreads, and a broad cross-section of industry, size, book-market, and momentum portfolio equity returns.[11] The 87 daily financial indicators in $\mathcal{D}^D$ include daily time series on commodities spot prices and futures prices, aggregate stock market indexes, volatility indexes, credit spreads and yield spreads, and exchange rates.

SOC GDP Growth

For y_j equal to GDP growth the forecasting model considers the following variables In $\mathbf{W}'_{jt}$:

[11] A detailed description of the series is given in the Data Appendix of the online supplementary file at www.sydneyludvigson.com/s/ucc_data_appendix.pdf

1) $\mathbb{F}^{SPF,(\mu)}_{jt-1}\left[y_{jt+h-1}\right]$, the mean SPF forecast for GDP growth.

2) $\mathbb{F}^{SPF,(50)}_{jt-1}\left[y_{jt+h-1}\right]$, the 50$th$ percentile SPF forecast for GDP growth.

3) $\mathbb{F}^{SPF,(25)}_{jt-1}\left[y_{jt+h-1}\right]$, the 25$th$ percentile SPF forecast for GDP growth.

4) $\mathbb{F}^{SPF,(75)}_{jt-1}\left[y_{jt+h-1}\right]$, the 75$th$ percentile SPF forecast for GDP growth.

5) $\text{var}_N\left(\mathbb{F}^{SPF,(\cdot)}_{t-1}\left[y_{jt+h-1}\right]\right)$, the cross-sectional variance of SPF forecasts for GDP growth.

6) $\text{skew}_N\left(\mathbb{F}^{SPF,(\cdot)}_{t-1}\left[y_{jt+h-1}\right]\right)$, the cross-sectional skewness of SPF forecasts for GDP growth.

7) $\widetilde{GDP}_{t-1}$ = detrended gross domestic product, defined as the residual from a regression of GDP_{t-1} on a constant and the four most recent values of GDP as of date $t-8$. See Hamilton (2018).

8) $\widetilde{EMP}_{t-1}$ = detrended employment, defined as the residual from a regression of EMP_{t-1} on a constant and the four most recent values of EMP as of date $t-8$. See Hamilton (2018).

9) VXO_t, defined as CBOE S&P 100 volatility index. We also include its squared and cubic terms, VXO_t^2, and VXO_t^3.

Lags of dependent variables:

1) $y_{j,t-1,t-h-1}$, $y_{j,t-2,t-h-2}$ one and two quarter lagged annual GDP growth.

The factors in $\boldsymbol{\Theta}'_{jt}$ include factors formed from three large datasets separately:

1) $\mathbf{G}_{M,t-k}$, for $k=0,1$ are factors formed from a real-time macro dataset $\mathcal{D}^M$ with 92 real-time macro series; includes both monthly and quarterly series, with monthly series converted to quarterly according to the method described in the data appendix.

2) $\mathbf{G}_{F,t-k}$, for $k=0,1$ are factors formed from a financial data set $\mathcal{D}^F$ with 147 monthly financial series.

3) $\mathbf{G}^Q_{D,t}$, are quarterly factors formed from a daily financial dataset $\mathcal{D}^D$ of 87 daily financial indicators. The raw daily series are first converted to daily factors $\mathbf{G}_{D,t}(\mathbf{w})$ and the daily factors are aggregated up to quarterly observations $\mathbf{G}^Q_{D,t}(\mathbf{w})$ using a weighted average of daily factors, with the weights $\mathbf{w}$ dependent on two free parameters that are chosen to minimize the sum of squared residuals in a regression of $y_{j,t+h}$ on $\mathbf{G}_{D,t}(\mathbf{w})$.

The 92 macro series in $\mathcal{D}^M$ are selected to represent broad categories of macroeconomic time series. The majority of these are real activity measures: real output and income, employment and hours, consumer spending, housing starts, orders and unfilled orders, compensation and labor costs, and capacity utilization measures. The dataset also includes commodity and price indexes and a handful of bond and stock market indexes, and foreign exchange measures. The financial dataset $\mathcal{D}^f$ is an updated monthly version of the of 147 variables comprised solely of financial market time series used in Ludvigson and Ng (2007). These data include valuation ratios such as the dividend-price ratio and earnings-price ratio, growth rates of aggregate dividends and prices, default and term spreads, yields on corporate bonds of different ratings grades, yields on Treasuries and yield spreads, and a broad cross-section of industry, size, book-market, and momentum portfolio equity returns.[12] The 87 daily financial indicators in $\mathcal{D}^D$ include daily time series on commodities spot prices and futures prices, aggregate stock market indexes, volatility indexes, credit spreads and yield spreads, and exchange rates.

Blue Chip Inflation

For consistency, the predictors for the BC inflation (PGDP inflation and CPI inflation) forecasts are constructed analogously to those of the SPF inflation forecasts. The only differences are that for own-survey forecasting variables (including nowcasts), e.g. $\mathbb{F}_t^{(i)}\left[y_{jt+h}\right]$ in $\mathbf{W}'_{jt}$, we now use survey forecasts from Blue Chip, instead of SPF.

Blue Chip GDP Growth

For y_j equal to GDP growth the forecasting model considers the same variables as in the SPF GDP growth forecasts with SPF forecasts replaced with Blue Chip Forecasts.

Private and Public Signals

This section derives the coefficient estimates given in the text for a model in which forecasters are presumed to combine statistical predictive models using public information (i.e., public signals) with their own judgemental forecast (i.e., a private signal) to form an overall prediction.

For reference, the machine learning model is

$$y_{j,t+h} = \alpha_{jh}^{(i)} + \beta_{jh\mathbb{F}}^{(i)}\mathbb{F}_t^{(i)}\left[y_{j,t+h}\right] + \mathbf{B}_{j\mathcal{Z}}^{(i)\prime}\mathcal{Z}_{jt} + \epsilon_{jt+h}, \qquad h \geq 1.$$

To set the stage, suppose that we have a jointly normal random vector split into two pieces, X_1 and X_2:

$$\left[\begin{array}{c} X_1 \\ X_2 \end{array}\right] \sim N\left(\left[\begin{array}{c} \mu_1 \\ \mu_2 \end{array}\right], \left[\begin{array}{cc} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{array}\right]\right)$$

[12] A detailed description of the series is given in the Data Appendix of the online supplementary file at www.sydneyludvigson.com/s/ucc_data_appendix.pdf

Denote the conditional distribution of X_1 given X_2 as $X_1|X_2$. With joint normality the optimal updating rule is

$$X_1|X_2 \sim N\left(\mu_1 + \beta\left(X_2 - \mu_2\right), \Omega\right)$$

$$\beta = \Sigma_{12}\Sigma_{22}^{-1}, \qquad \Omega = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}.$$

Below we use $\mathbb{E}_o[\cdot]$ to denote the conditional expectation implied by the optimal updating rule, i.e., $\mathbb{E}_o[X_1|X_2] = \mu_1 + \beta(X_2 - \mu_2)$ in the above. These results are used below.

Let x be publicly available information and let z be a private signal about an unknown variable y. These variables are related to one another according to the system

$$x \sim iid\left(0, \sigma_x^2\right)$$

(A4) $$y = \alpha x + u_2, \ u_2 \sim N\left(0, \sigma_2^2\right)$$

(A5) $$z = y + u_1, \ u_1 \sim N\left(0, \sigma_1^2\right),$$

where α is a known parameter describing the mapping from x to y, and x, u_1, and u_2 are i.i.d. and mutually uncorrelated with one another.

Consider the optimal forecast of y in this setting, when one combines a statistical model using x with a private signal z. From (A4) and (A5), conditional on observing αx we have

$$\begin{aligned} y|x &= \alpha x + u_2, \ u_2 \sim N\left(0, \sigma_2^2\right) \\ z &= (y|x) + u_1 = \alpha x + u_2 + u_1, \ u_1 \sim N\left(0, \sigma_2^2\right) \end{aligned}$$

$$\begin{bmatrix} y|x \\ z|x \end{bmatrix} \sim N\left(\begin{bmatrix} \alpha x \\ \alpha x \end{bmatrix}, \begin{bmatrix} \sigma_2^2 & \sigma_2^2 \\ \sigma_2^2 & \sigma_1^2 + \sigma_2^2 \end{bmatrix}\right)$$

Now suppose that the agent combines the information in αx with a private signal z.[13] Conditional on both private and public signals, the optimal forecast is

$$\begin{aligned} \mathbb{E}_o\left[y|x, z\right] &= \alpha x + \sigma_2^2\left(\sigma_1^2 + \sigma_2^2\right)^{-1}(z - \alpha x) \\ &= \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} z + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}\alpha x \\ &= \gamma z + (1 - \gamma)\alpha x \end{aligned}$$

where

$$\gamma \equiv \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}.$$

The optimal weight on the private signal z versus the public signal αx depends on the

[13]The result below is unchanged if one assumes that the agent first recieves z and then combines it with αx.

true precision σ_1^{-2} of the private signal relative the precision σ_2^{-2} of the public signal.

We can also compute:

$$\begin{aligned}\mathbb{V}_o\left[y|x,z\right] &= \sigma_2^2-\sigma_2^2\sigma_2^2\left(\sigma_1^2+\sigma_2^2\right)^{-1}\\ &= \sigma_2^2\left[1-\sigma_2^2\left(\sigma_1^2+\sigma_2^2\right)^{-1}\right]\\ &= \sigma_2^2\sigma_1^2\left(\sigma_1^2+\sigma_2^2\right)^{-1}.\end{aligned}$$

Forecaster. The forecaster assigns weights to the private and public signal as follows:

$$\mathbb{F}=\gamma^F z+\left(1-\gamma^F\right)\alpha^F x$$

The survey response is here interpreted as a forecast based partly on a respondent's statistical model using public information ($\alpha^F z_2$) in combination with a private signal z.

Machine. The machine forecast of y given by

$$\mathbb{E}=\widehat{\beta}\mathbb{F}+\widehat{B}x=\widehat{\beta}\left[\gamma^F z+\left(1-\gamma^F\right)\alpha^F x\right]+\widehat{B}x,$$

where $\widehat{\beta}$ and $\widehat{B}$ are estimated coefficients. The machine uses maximum likelihood to estimate the coefficients $b\equiv(\beta,B)'$ from a regression of y on $f\equiv(\mathbb{F},x)'$, with $b=\text{cov}\left(f,\,f'\right)^{-1}\text{cov}(f,y)$. This estimator results in the values:

$$\begin{aligned}\widehat{\beta} &= \frac{\gamma}{\gamma^F}\\ \widehat{B} &= (1-\gamma)\,\alpha-\left(\frac{\gamma}{\gamma^F}-\gamma\right)\alpha^F.\end{aligned}$$

Proof:

Writing the covariances in matrix form, we have that

$$b=\begin{bmatrix}\beta\\ B\end{bmatrix}=\begin{bmatrix}\text{var}(\mathbb{F}) & \text{cov}(\mathbb{F},x)\\ \text{cov}(\mathbb{F},x) & \text{var}(x)\end{bmatrix}^{-1}\begin{bmatrix}\text{cov}(\mathbb{F},y)\\ \text{cov}(x,y)\end{bmatrix}. \tag{A6}$$

First,

$$\begin{aligned}\text{var}(\mathbb{F}) &= \text{var}\left(\gamma^F z+\left(1-\gamma^F\right)\alpha^F x\right)\\ &= \text{var}\left(\gamma^F(\alpha x+u_2+u_1)+\left(1-\gamma^F\right)\alpha^F x\right)\\ &= \left(\gamma^F\right)^2\left(\alpha^2\sigma_x^2+\sigma_2^2+\sigma_1^2\right)+\left[\left(1-\gamma^F\right)\alpha^F\right]^2\sigma_x^2\\ &= \left(\gamma^F\right)^2\left(\sigma_2^2+\sigma_1^2\right)+\left(\gamma^F\alpha+\left(1-\gamma^F\right)\alpha^F\right)^2\sigma_x^2 \qquad \text{(A7)}\end{aligned}$$

Second,

$$\begin{aligned} \operatorname{cov}(\mathbb{F}, x) &= \operatorname{cov}\left(\left(\gamma^F \alpha + \left(1-\gamma^F\right)\alpha^F\right)x + \gamma^F\left(u_2+u_1\right), x\right) \\ &= \left(\gamma^F \alpha + \left(1-\gamma^F\right)\alpha^F\right)\sigma_x^2 \end{aligned} \tag{A8}$$

Third,

$$\operatorname{var}(x) = \sigma_x^2 \tag{A9}$$

Fourth,

$$\begin{aligned} \operatorname{cov}(\mathbb{F},y) &= \operatorname{cov}\left(\left(\gamma^F \alpha + \left(1-\gamma^F\right)\alpha^F\right)x + \gamma^F\left(u_2+u_1\right), \alpha x + u_2\right) \\ &= \left(\gamma^F \alpha + \left(1-\gamma^F\right)\alpha^F\right)\alpha\sigma_x^2 + \gamma^F\sigma_2^2 \end{aligned} \tag{A10}$$

Last,

$$\begin{aligned} cov(x,y) &= cov(x, \alpha x + u_2) \\ &= \alpha\sigma_x^2. \end{aligned} \tag{A11}$$

Plugging equation (A7) to (A11) to equation (A6), we have

$$\begin{aligned} b &= \underbrace{\begin{bmatrix} \left(\gamma^F\right)^2\left(\sigma_2^2+\sigma_1^2\right) + \left(\gamma^F\alpha + \left(1-\gamma^F\right)\alpha^F\right)^2\sigma_x^2 & \left(\gamma^F\alpha + \left(1-\gamma^F\right)\alpha^F\right)\sigma_x^2 \\ \left(\gamma^F\alpha + \left(1-\gamma^F\right)\alpha^F\right)\sigma_x^2 & \sigma_x^2 \end{bmatrix}^{-1}}_{\equiv B_1} \\ &\quad \times \begin{bmatrix} \left(\gamma^F\alpha + \left(1-\gamma^F\right)\alpha^F\right)\alpha\sigma_x^2 + \gamma^F\sigma_2^2 \\ \alpha\sigma_x^2. \end{bmatrix} \\ &= \frac{1}{\Delta}\begin{bmatrix} \sigma_x^2 & -\left(\gamma^F\alpha + \left(1-\gamma^F\right)\alpha^F\right)\sigma_x^2 \\ -\left(\gamma^F\alpha + \left(1-\gamma^F\right)\alpha^F\right)\sigma_x^2 & \left(\gamma^F\right)^2\left(\sigma_2^2+\sigma_1^2\right) + \left(\gamma^F\alpha + \left(1-\gamma^F\right)\alpha^F\right)^2\sigma_x^2 \end{bmatrix} \\ &\quad \times \begin{bmatrix} \left(\gamma^F\alpha + \left(1-\gamma^F\right)\alpha^F\right)\alpha\sigma_x^2 + \gamma^F\sigma_2^2 \\ \alpha\sigma_x^2. \end{bmatrix} \end{aligned}$$

where Δ is the determinant of the matrix B_1.

We then have,

$$\begin{aligned} \Delta\beta &= \left[\left(\gamma^F\alpha + \left(1-\gamma^F\right)\alpha^F\right)\alpha\sigma_x^4 + \gamma^F\sigma_2^2\sigma_x^2 - \alpha\left(\gamma^F\alpha + \left(1-\gamma^F\right)\alpha^F\right)\sigma_x^2\sigma_x^2\right] \\ &= \gamma^F\sigma_2^2\sigma_x^2 \end{aligned} \tag{A12}$$

Solving for Δ,

$$\begin{aligned} \Delta &= \sigma_x^2\left(\gamma^F\right)^2\left(\sigma_2^2+\sigma_1^2\right) + \left(\gamma^F\alpha + \left(1-\gamma^F\right)\alpha^F\right)^2\sigma_x^4 - \left(\gamma^F\alpha + \left(1-\gamma^F\right)\alpha^F\right)^2\sigma_x^4 \\ &= \sigma_x^2\left(\gamma^F\right)^2\left(\sigma_2^2+\sigma_1^2\right). \end{aligned} \tag{A13}$$

Plugging Δ back to equation (A12), we have

$$\beta = \frac{\gamma^F \sigma_2^2 \sigma_x^2}{\sigma_x^2 \left(\gamma^F\right)^2 \left(\sigma_2^2 + \sigma_1^2\right)} = \frac{\sigma_2^2}{\gamma^F \left(\sigma_2^2 + \sigma_1^2\right)} = \frac{\gamma}{\gamma^F}$$

Now solving for B,

$$\begin{aligned}
\Delta B &= -\left(\gamma^F \alpha + \left(1-\gamma^F\right)\alpha^F\right)\sigma_x^2\left[\left(\gamma^F \alpha + \left(1-\gamma^F\right)\alpha^F\right)\alpha\sigma_x^2 + \gamma^F\sigma_2^2\right] \\
&\quad + \alpha\sigma_x^2\left(\left(\gamma^F\right)^2\left(\sigma_2^2+\sigma_1^2\right) + \left(\gamma^F\alpha + \left(1-\gamma^F\right)\alpha^F\right)^2\sigma_x^2\right) \\
&= -\left(\gamma^F \alpha + \left(1-\gamma^F\right)\alpha^F\right)\sigma_x^2\gamma^F\sigma_2^2 + \alpha\sigma_x^2\left(\gamma^F\right)^2\left(\sigma_2^2+\sigma_1^2\right) \qquad \text{(A14)}
\end{aligned}$$

Plugging Δ back to equation (A14), we have

$$\begin{aligned}
B &= \frac{-\left(\gamma^F \alpha + \left(1-\gamma^F\right)\alpha^F\right)\sigma_x^2\gamma^F\sigma_2^2 + \alpha\sigma_x^2\left(\gamma^F\right)^2\left(\sigma_2^2+\sigma_1^2\right)}{\sigma_x^2\left(\gamma^F\right)^2\left(\sigma_2^2+\sigma_1^2\right)} = \alpha - \left(\alpha - \alpha^F + \frac{\alpha^F}{\gamma^F}\right)\gamma \\
&= (1-\gamma)\,\alpha - \left(\frac{\gamma}{\gamma^F} - \gamma\right)\alpha^F
\end{aligned}$$

as stated.

Coibion-Gorodnichenko Regressions

To construct SPF forecasts of annual inflation, forecasters at time t are presumed to use an advance estimate of $t-1$ price level combined with their survey respondent forecast of that price level at $t+3$ to form a forecast of π_{t+3}.

$$\underbrace{\pi_{t+3} - \mathbb{F}_t^{(\mu)}\left[\pi_{t+3}\right]}_{\text{Forecast Error}} = \alpha + \beta\left(\underbrace{\mathbb{F}_t^{(\mu)}\left[\pi_{t+3}\right] - \mathbb{F}_{t-1}^{(\mu)}\left[\pi_{t+3}\right]}_{\text{Forecast Revision}}\right) + \epsilon_{t+3} \qquad \text{(A15)}$$

where the annual inflation at time $t+3$ is defined as,

$$\pi_{t+3} = 100 \times \left(\frac{P_t}{P_{t-1}} \times \frac{P_{t+1}}{P_t} \times \frac{P_{t+2}}{P_{t+1}} \times \frac{P_{t+3}}{P_{t+2}} - 1\right). \qquad \text{(A16)}$$

Following CG, regressions are run and forecast errors computed using forecasts of real-time inflation data available four quarters after the period being forecast.

The survey forecast is constructed as follows

$$\mathbb{F}_t\left[\pi_{t+3}\right] = 100 \times \left(\frac{P_t^{avg}}{P_{t-1}} \times \frac{P_{t+1}^{avg}}{P_t^{avg}} \times \frac{P_{t+2}^{avg}}{P_{t+1}^{avg}} \times \frac{P_{t+3}^{avg}}{P_{t+2}^{avg}} - 1\right),$$

where $P_{t+h}^{avg} = \frac{1}{N_{t+h}}\sum_{i=1}^{N_{t+h}} P_{t+h}^i$, for $h = 0, \ldots, 3$, i represents an individual forecaster,

N_{t+h} is the number of forecasters at time time $t+h$, and P_{t-1} is the BEA's advance estimate at t for prices in $t-1$.

FORECAST ERRORS

The forecast error on the LHS of the regressions (A15) is constructed in the following way:

$$\begin{aligned}\pi_{t+3,t} - \mathbb{F}_t^{(\mu)}\left[\pi_{t+3,t}\right] \equiv 100 \times \Bigg[&\left(\frac{\pi_{t,t-1} - \mathbb{F}_t^{(\mu)}\left[\pi_{t,t-1}\right]}{400} + 1\right) \\ &\times\left(\frac{\pi_{t+1,t} - \mathbb{F}_t^{(\mu)}\left[\pi_{t+1,t}\right]}{400} + 1\right) \\ &\times\left(\frac{\pi_{t+2,t+1} - \mathbb{F}_t^{(\mu)}\left[\pi_{t+2,t+1}\right]}{400} + 1\right) \\ &\times\left(\frac{\pi_{t+3,t+2} - \mathbb{F}_t^{(\mu)}\left[\pi_{t+3,t+2}\right]}{400} + 1\right) - 1\Bigg]\end{aligned} \tag{A17}$$

In brackets is the product of quarterly forecast errors from the nowcast to $h=3$ quarters ahead.

IN-SAMPLE ANALYSIS

Table A10 presents the replication for CG, as well as results from extending the sample size to 2018:Q2. Panel A replicates the numbers from columns (1) and (2) of Table 1 Panel B of CG. Panel B presents the results for the extended sample.

Table A12 presents the results from CG regressions when we replace the survey forecast with our machine forecast for SPF mean inflation. More specifically, we estimate is the following regression:

$$\underbrace{\pi_{t+3,t} - \mathbb{E}_{t+3|t}^{(\mu)}}_{\text{Machine Forecast Errors}} = \alpha + \beta\left(\underbrace{\mathbb{E}_t^{(\mu)}\left[\pi_{t+3,t}\right] - \mathbb{E}_{t-1}^{(\mu)}\left[\pi_{t+3,t}\right]}_{\text{Machine Forecast Revision}}\right) + \delta\pi_{t-1} + \epsilon_{jt+3}$$

where $\mathbb{E}_t^{(\mu)}\left[\pi_{t+3,t}\right]$ is the machine mean forecast made at time t and $\mathbb{E}_{t-1}^{(\mu)}\left[\pi_{t+3,t}\right]$ is the machine forecast made at time $t-1$.

OUT-OF-SAMPLE ANALYSIS

We construct a series of real-time out-of-sample forecasts using the CG model:

$$\pi_{t+3} - \mathbb{F}_t^{(\mu)}\left[\pi_{t+3}\right] = \alpha^{(\mu)} + \beta^{(\mu)}\left(\mathbb{F}_t^{(\mu)}\left[\pi_{t+3}\right] - \mathbb{F}_{t-1}^{(\mu)}\left[\pi_{t+3}\right]\right) + \epsilon_{t+3}$$

Table A10— CG In-Sample Regressions of Forecast Errors on Forecast Revisions (Survey)

Regression: $\pi_{t+3,t} - \mathbb{F}_t\left[\pi_{t+3,t}\right] = \alpha + \beta\left(\mathbb{F}_t\left[\pi_{t+3,t}\right] - \mathbb{F}_{t-1}\left[\pi_{t+3,t}\right]\right) + \delta\pi_{t-1,t-2} + \epsilon_t$				
	(1)	(2)	(3)	(4)
	Panel A: Sample: 1969:Q1 - 2014:Q4		**Panel B**: Sample: 1969:Q1 - 2018:Q2	
Constant	0.001	-0.077	-0.022	-0.116
t-stat	(0.005)	(-0.442)	(-0.167)	(-0.758)
$\mathbb{F}_t\left[\pi_{t+3,t}\right] - \mathbb{F}_{t-1}\left[\pi_{t+3,t}\right]$	1.194**	1.141**	1.186**	1.116**
t-stat	(2.496)	(2.560)	(2.478)	(2.532)
$\pi_{t-1,t-2}$		0.021		0.027
t-stat		(0.435)		(0.574)
$\bar{R}^2$	0.195	0.197	0.193	0.195

Table A11— Notes: The annual inflation is defined as $\pi_{t+3,t} = \frac{P_t}{P_{t-1}} \times \frac{P_{t+1}}{P_t} \times \frac{P_{t+2}}{P_{t+1}} \times \frac{P_{t+3}}{P_{t+2}}$, the covariate $\mathbb{F}_t\left[\pi_{t+3,t}\right]$ is the SPF of annual inflation with information in period t and $\mathbb{F}_{t-1}\left[\pi_{t+3,t}\right]$ is the SPF mean forecast of the same annual inflation but with information in $t-1$. Panel A presents the sample in Coibion and Gorodnichenko (2015) and Panel B updates the sample to 2018:Q2. Regressions are run and model evaluated using real-time data with observation on $\pi_{t+3,t}$ available 4 quarters after the advance estimate of it. Newey-West corrected (t-statistics) with lags = 4. Newey-West HAC: *sig. at 10%. **sig. at 5%. ***sig. at 1%.

To do so, we estimate the model over an initial sample, forecast out one period, roll (or recurse) forward, and repeat estimation and forecast. The regression estimation uses the latest vintage of inflation in real time and, following CG, computes forecast errors using real-time data available four quarters after the period being forecast. The CG model forecast for π_{t+3} is thus

$$\widehat{\pi}^{(\mu)}_{t+3} = \widehat{\alpha}^{(\mu)}_t + \left(1 + \widehat{\beta}^{(\mu)}_t\right)\mathbb{F}^{(\mu)}_t\left[\pi_{t+3}\right] - \widehat{\beta}^{(\mu)}_t\mathbb{F}^{(\mu)}_{t-1}\left[\pi_{t+3}\right]$$

For the rolling procedure, we try windows of sizes $w = 5$, 10, and 20 years. For the recursive procedure that uses expanding windows over time, we try initial window sizes of 5, 10, and 20 years.

The survey and model errors are recorded as:

$$\begin{aligned}\text{survey error}_t &= \mathbb{F}^{(\mu)}_t\left[\pi_{t+3}\right] - \pi_{t+3}\\ \text{CG model error}_t &= \widehat{\pi}^{(\mu)}_{t+3} - \pi_{t+3}\end{aligned}$$

We then compute MSEs for the survey respondent and those for the CG over the different forecast samples of size P as

$$\begin{aligned}\text{MSE}_{\mathbb{F}} &= \frac{1}{P}\sum_{s=1}^{P}\left(\text{survey error}_{t+s}\right)^2\\ \text{MSE}_{\text{CG}} &= \frac{1}{P}\sum_{s=1}^{P}\left(\text{CG model error}_{t+s}\right)^2.\end{aligned}$$

These are reported in Table A14.

Table A12— CG Regressions of Forecast Errors on Forecast Revisions (Machine)

Regression: $\pi_{t+3,t} - \mathbb{E}_t\left[\pi_{t+3,t}\right] = \alpha + \beta\left(\mathbb{E}_t\left[\pi_{t+3,t}\right] - \mathbb{E}_{t-1}\left[\pi_{t+3,t}\right]\right) + \delta\pi_{t-1,t-2} + \epsilon_t$		
	(1)	(2)
Constant	-0.12	-0.13
t-stat	(-1.21)	(-0.94)
$\mathbb{E}_t\left[\pi_{t+3,t}\right] - \mathbb{E}_{t-1}\left[\pi_{t+3,t}\right]$	-0.04	-0.04
t-stat	(-0.22)	(-0.24)
$\pi_{t-1,t-2}$		0.00
t-stat		(0.08)
$\bar{R}^2$	0.0008	0.0008

Table A13— Notes: The annual inflation is defined as $\pi_{t+3,t} = \frac{P_t}{P_{t-1}} \times \frac{P_{t+1}}{P_t} \times \frac{P_{t+2}}{P_{t+1}} \times \frac{P_{t+3}}{P_{t+2}}$, the covariate $\mathbb{E}_t\left[\pi_{t+3,t}\right]$ is the machine mean forecast of annual inflation with information in period t and $\mathbb{E}_{t-1}\left[\pi_{t+3,t}\right]$ is the machine mean forecast of the same annual inflation but with information in $t-1$. Regressions are run and model evaluated using real-time data with observation on $\pi_{t+3,t}$ available 4 quarters after the advance estimate of it. Newey-West corrected (t-statistics) with lags = 4. Newey-West HAC: *sig. at 10%. **sig. at 5%. ***sig. at 1%. The sample is 1995:Q1 to 2018:Q2.

Dynamic Responses to Cyclical Shocks–Local Projection

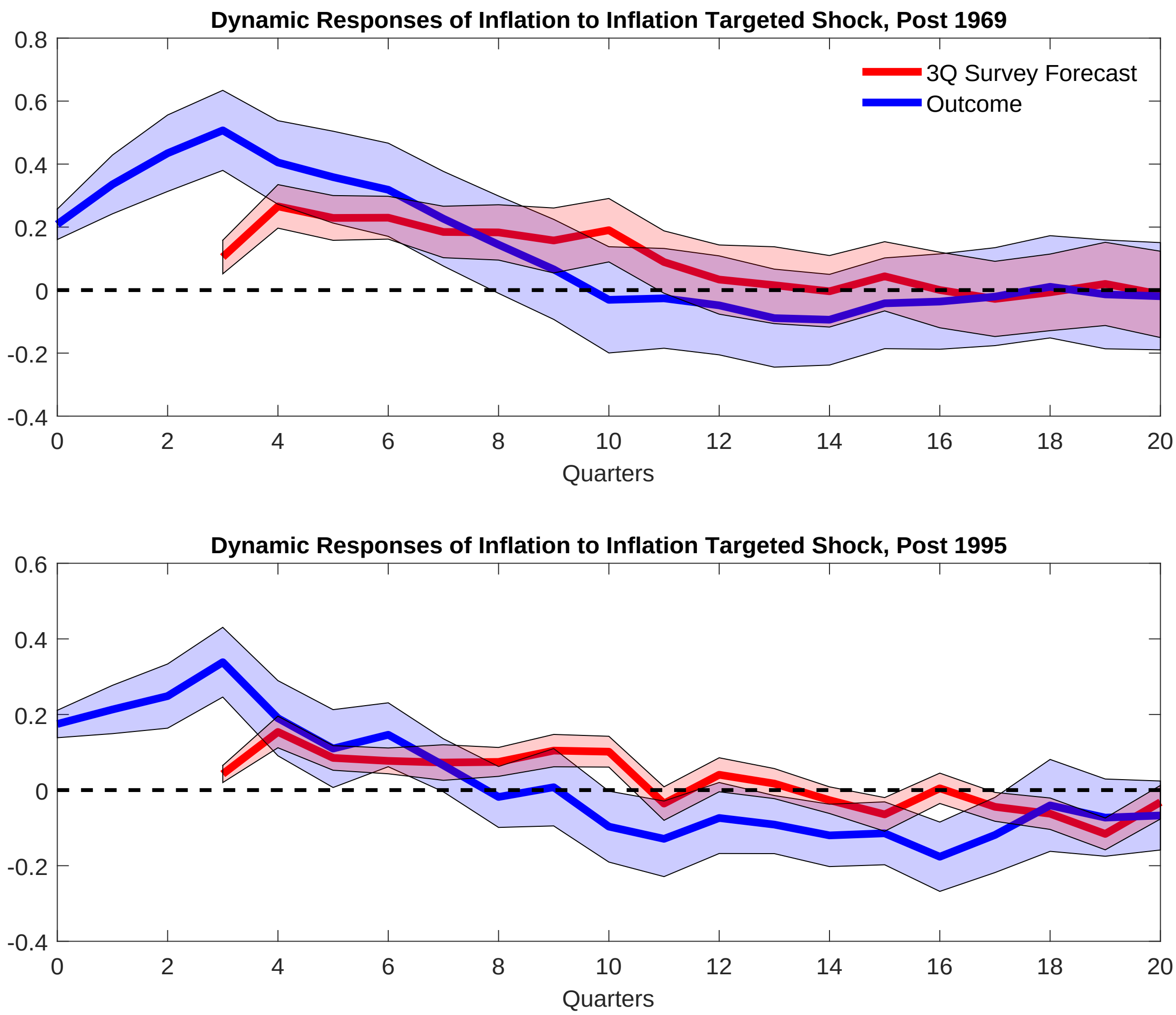
We follow Angeletos et al. (2020) (AHS) and estimate the dynamic responses to inflation or GDP growth shocks from Angeletos et al. (2018) via local projection using a series of single equation regressions, one for each horizon $0 \leq h \leq H$ taking the form

$$z_{t+h} = \alpha_h + \beta_h \varepsilon_t + \gamma' W_t + u_{t+h} \tag{A18}$$

where z_t is either the outcome variable at t, the survey forecast made at t, $\mathbb{F}_t^{(i)}\left[y_{j,t+h}\right]$, or the machine forecast made at time t, $\mathbb{E}_t^{(i)}\left[y_{j,t+h}\right]$. The dynamic responses plotted in the figures of the main text and below are given by the sequence of coefficients $\{\beta\}_{h=0}^{H}$, where W_t is a vector of control variables that are the same as those used in AHS and include one lag each of the outcome and survey forecast. We consider two outcome variables: inflation and real GDP growth. Following Angeletos et al. (2020), we plot forecasts and outcome variables so that $\mathbb{F}_t^{(50)}\left[y_{j,t+h}\right]$ is lined up with $y_{j,t+h}$ along a vertical slice and the difference between the two is the forecast error. On the left-hand-side the forecasts are made at time t for period $t+h$, while the shock occurs at t. We compute the heteroskedasticity and autocorrelation robust (HAC) standard errors with a 4-quarter Bartlett kernel to calculate standard errors for the impulse responses. The ± 1 standard error bands are reported.

Top panel of Figure A1 shows that we replicate the dynamic responses of inflation to an inflation targeted shock over the same sample used in Angeletos et al. (2020). The bottom panel of Figure A1 shows the dynamic responses are similar using the local projection estimation over our evaluation sample 1995:1-2018:Q2.

Figure A1. Dynamic Responses: Forecast and Outcome



Dynamic responses of GDP and inflations. The shaded areas are 68% confidence intervals based on HAC standard errors with a Bartlett kernel and 4 lags. The x-axis denotes quarters from the shock. The outcome variable is inflation π_t and the shock is the inflation-targeted shock. The survey forecast is $\mathbb{F}_t^{(50)}\left[y_{t+3}\right]$. The shock time series are from Angeletos et al. (2018). In the first row, the impulse responses are estimated over sample 1969:Q1 to 2018:Q2. In the second row, the impulse responses are estimated over sample 1995:Q1 to 2018:Q2. In both rows, we "align" the forecast responses such that, at a given vertical slice of the plot, the outcome and forecast responses are measured over the same horizon, and the difference between the two is the forecast error. The vintage of observations on the outcome variable is final-release data.

Table A14— Mean Square Errors for the CG Model and SPF

Forecast model: $\widehat{\pi}^{(\mu)}_{t+3} = \widehat{\alpha}^{(\mu)}_t + \left(1+\widehat{\beta}^{(\mu)}_t\right) \mathbb{F}^{(\mu)}_t \left[\pi_{t+3}\right] - \widehat{\beta}^{(\mu)}_t \mathbb{F}^{(\mu)}_{t-1} \left[\pi_{t+3}\right]$			
	$\text{MSE}_{CG}/\text{MSE}_{\mathbb{F}}$		
Method	Quarterly Compound	Continuous Compound	CG Sample
Rolling 5 years	1.38	1.38	1.39
Rolling 10 years	1.29	1.29	1.29
Rolling 20 years	1.31	1.30	1.34
Recursive 5 years	1.69	1.68	1.71
Recursive 10 years	1.60	1.59	1.59
Recursive 20 years	1.33	1.30	1.34

Notes: The table reports the ratio of MSEs of the CG model forecast over the survey forecast. The regression estimation uses the latest vintage of inflation in real time and, following CG, computes forecast errors real-time data available four quarters after the period being forecast. The sample spans the period 1969:Q1 - 2018:Q2. The CG sample refers to the sample in Coibion and Gorodnichenko (2015) that ends in 2014:Q4.

Machine Forecasts with Bloomberg Survey Data

To form an estimate of the median SPF machine forecast $\mathbb{E}^{(50)}_t \left[y_{j,t+h}\right]$ for four-quarter ahead GDP growth that does not use the median type's time t observation $\mathbb{F}^{(i)}_t \left[y_{j,t+h}\right]$, we can use similar professional forecasts from more timely survey data available prior to the time t SPF survey deadline. The Bloomberg (BBG) US consensus forecasts are updated daily (except for weekends and holidays). We use the median forecast from the Bloomberg Terminal for GDP on the closest day before the SPF survey deadline. The terminal reports daily quarter-over-quarter real GDP growth forecasts starting in 2003:Q1. To be consistent with the SPF forecasts, we construct the annual GDP growth forecast as follows.

Let $gY^{(Q/Q)}_{t+h}$ denote annualized quarter-over-quarter GDP growth in percent, h quarters ahead, and let $\mathbb{B}^{(50)}_t \left[gY^{(Q/Q)}_{t+h}\right]$ be the median BBG forecaster's prediction of this variable made at time t, where time t is the closest day before the SPF survey deadline listed in Table A.1. Bloomberg $\mathbb{B}^{(50)}_t \left[gY^{(Q/Q)}_{t+h}\right]$ are reported at annual rates in percentage points, so we convert to quarterly raw units before compounding. Let $y_{t+4,t}$ denote four-quarter real GDP growth. We construct the four-quarter real GDP growth BBG forecast from $gY^{(Q/Q)}_{t+h}$ as:

$$\mathbb{B}^{(50)}_t \left[y_{t+4,t}\right] = 100 \times ln\left(\prod_{h=1}^{4}\left(1+\frac{\mathbb{B}^{(50)}_t \left[gY^{(Q/Q)}_{t+h}\right]}{100}\right)^{\frac{1}{4}}\right).$$

$\mathbb{B}^{(50)}_t \left[y_{t+4,t}\right]$ exhibits a correlation with the SPF median annual GDP forecast $\mathbb{F}^{(50)}_t \left[y_{t+4,t}\right]$

of 96.4% over the common sample from 2003:Q1 to 2018:Q2.

To form an estimate of the median SPF machine forecast $\mathbb{E}_t^{(50)}\left[y_{j,t+h}\right]$ for four-quarter ahead GDP growth that does not use the median type's time t survey forecast $\mathbb{F}_t^{(i)}\left[y_{j,t+h}\right]$, we instead use the BBG professional consensus survey forecast which is publicly available on the closest day before time t SPF survey deadline. The estimation is the same as in the baseline estimation with two exceptions. First, we replace the survey deadline observation of the time t SPF median forecast series $\mathbb{F}_t^{(50)}\left[y_{j,t+h}\right]$ with the time t observation on the median forecast from the BBG survey, $\mathbb{B}_t^{(50)}\left[y_{t+4,t}\right]$. Second, the machine forecast is estimated over a shorter sample starting from 2003:Q1, when BBG data are available. The evaluation sample is for this estimation spans 2010:Q1-2018:Q2.

Table A15— Results with Bloomberg Forecasts

ML: $y_{j,t+h}=\alpha_{jh}^{(i)}+\beta_{jFt}^{(i)}\mathbb{F}_t^{(i)}\left[y_{j,t+h}\right]+B_{jZ}^{(i)}Z_{jt}\epsilon_{jt+h}$		
SPF GDP Median Forecast		
Replace survey deadline observation $\mathbb{F}_t^{(50)}$ with $\mathbb{B}_t^{(50)}$		
	Use $\mathbb{F}_t^{(50)}$ for all t	Replace survey deadline $\mathbb{F}_t^{(50)}$ with $\mathbb{B}_t^{(50)}$
MSE_E/MSE_F	0.81	0.85
OOS R^2	0.19	0.15

Notes: This table reports the MSE ratios with and without using Bloomberg consensus forecasts. The second column reports the results when SPF median forecasts are used for all quarters. The second column reports the results when the current-quarter SPF median forecast of GDP growth is replaced by the Bloomberg median forecast and include one lag of SPF forecasts of all types. $\text{MSE}_{\mathbb{E}}$ and $\text{MSE}_{\mathbb{F}}$ denote the machine and SPF survey mean-squared-forecast-errors, respectively, for 4-quarter-ahead forecasts, averaged over the evaluation sample. The out-of-sample Rsquared, OOS R^2, is defined as 1-$\text{MSE}_{\mathbb{E}}/\text{MSE}_{\mathbb{F}}$. The evaluation period is 2010:Q1 to 2018:Q2

*

REFERENCES

Andreou, Elena, Eric Ghysels, and Andros Kourtellos, "Should macroeconomic forecasters use daily financial data and how?," *Journal of Business & Economic Statistics*, 2013, *31* (2), 240–251.

Angeletos, George-Marios, Fabrice Collard, and Harris Dellas, "Business cycle anatomy," Technical Report, National Bureau of Economic Research 2018.

_ **, Zhen Huo, and Karthik A Sastry**, "Imperfect macroeconomic expectations: evidence and theory," 2020. NBER Working paper No. 27308.

Bai, Jushan and Serena Ng, "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, January 2002, *70* (1), 191–221.

_ **and** _ , "Confidence Intervals for Diffusion Index Forecasts and Inference for Factor-Augmented Regressions," *Econometrica*, 2006, *74* (4), 1133–50.

Coibion, Olivier and Yuriy Gorodnichenko, "Information rigidity and the expectations formation process: A simple framework and new facts," *American Economic Review*, 2015, *105* (8), 2644–78.

Curtin, Richard Thomas, *Consumer Expectations: Micro Foundations and Macro Impact*, Cambridge University Press, 2019.

Fama, Eugene F. and Kenneth R. French, "The Cross-Section of Expected Returns," *Journal of Finance*, 1992, *47* (2), 427–65.

_ **and** _ , "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics*, 1993, *33*, 3–56.

Faust, Jon and Jonathan H Wright, "Forecasting inflation," in "Handbook of economic forecasting," Vol. 2, Elsevier, 2013, pp. 2–56.

Ghysels, Eric, Arthur Sinko, and Rossen Valkanov, "MIDAS regressions: Further results and new directions," *Econometric Reviews*, 2007, *26* (1), 53–90.

Hamilton, James D, "Why you should never use the Hodrick-Prescott filter," *Review of Economics and Statistics*, 2018, *100* (5), 831–843.

Huang, Dashan, Fuwei Jiang, and Guoshi Tong, "Real time macro factors in bond risk premium," 2017. Unpublished paper, Lee kong Chian School of Business.

Jurado, Kyle, Sydney C. Ludvigson, and Serena Ng, "Measuring Uncertainty," *The American Economic Review*, 2015, *105* (3), 117–1216.

Ludvigson, Sydney C. and Serena Ng, "The Empirical Risk-Return Relation: A Factor Analysis Approach," *Journal of Financial Economics*, 2007, *83* (1), 171–222.

_ **and** _ , "Macro Factors in Bond Risk Premia," *The Review of Financial Studies*, 2009, *22* (12), 5027–67.

_ **, Sai Ma, and Serena Ng**, "Uncertainty and Business Cycles: Exogenous Impulse or Endogenous Response?," *American Economic Journal: Macroeconomics*, forthcoming, 2019.