Online Appendix

Data

This appendix describes our data.

5.0.1 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.

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.

5.0.2 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.[7] 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.[8]

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."

   Quarterly and annual inflation forecasts are constructed as follows. Let $\mathbb{F}_t^{(i)}[P_{t+h}]$ be forecaster i's prediction of PGDP h quarters ahead and $\mathbb{N}_t^{(i)}[P_t]$ be forecaster i's nowcast of PGDP for the current quarter. Annualized inflation forecasts for forecaster i are

   $$\mathbb{F}_t^{(i)}[\pi_{t+h,t}] = (400/h) \times \ln\left(\frac{\mathbb{F}_t^{(i)}[P_{t+h}]}{\mathbb{N}_t^{(i)}[P_t]}\right), \tag{A.11}$$

   where $h = 1$ for quarterly inflation and $h = 4$ for annual inflation. Similarly, we construct quarterly and annual nowcasts of inflation as

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

   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

[7] 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.

[8] 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.

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 OOS 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 A.1 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 A.1: SPF Survey Deadlines[9]

Survey	Deadline Date	Survey	Deadline Date	Survey	Deadline Date
1990:Q1	NA	1991:Q1	2/16/91	1992:Q1	2/22/92
Q2	NA	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

[9] 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 A.1 (Cont'd)

Survey	Deadline Date	Survey	Deadline Date	Survey	Deadline Date
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
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: We use survey responses to question A12b, which asks (emphasis in original):

   By about what percent do you expect prices to go (up/down) on the average, during the next 12 months?

   Respondents provide a numerical value to the interviewer and the SOC provides the mean, median, and 25th and 75th percentiles. Since this is already reported as an inflation rate, we do not make any adjustments.

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. 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:

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

   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 $\mathbb{N}_t^{(i)}\left[P_t\right]$ is forecaster i's nowcast of PGDP for the current quarter which is constructed as: $\mathbb{N}_t^{(i)}\left[P_t\right] = \exp\left(\mathbb{N}_t^{(i)}\left[\pi_{t,t-1}\right]/400 + \ln P_{t-1}\right)$. 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.[10] 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.

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

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

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 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 beginning-of-quarter values. Alternatively, if the survey deadline is unknown we allow only information up to quarter $t-1$ to enter the model. Thus, we use end-of-quarter values in these cases.

Table A.2 lists the Philadelphia Fed variables as well as the energy prices data from the BLS. 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.[11] Table A.2 also lists the transformation applied to each variable to make them stationary before

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

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 A.2: 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
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 and Investment		
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
			Group 4: Consumption		
38	NCONGMMVMD	Philly Fed	Δln	Nom. personal cons. exp. - Goods	2009:M8
39	NCONHHMMVMD	Philly Fed	Δln	Nom. hh. cons. exp.	2009:M8
40	NCONSHHMMVMD	Philly Fed	Δln	Nom. hh. cons. exp. - Services	2009:M8

Table A.2 (Cont'd)

No.	Short Name	Source	Tran	Description	First Vintage
41	NCONSNPMMVMD	Philly Fed	Δln	Nom. final cons. exp. of NPISH	2009:M8
42	RCONDMMVMD	Philly Fed	Δln	Real personal cons. exp. - Durables	1998:M11
43	RCONGMMVMD	Philly Fed	Δln	Real personal cons. exp. - Goods	2009:M8
44	RCONHHMMVMD	Philly Fed	Δln	Real hh. cons. exp.	2009:M8
45	RCONMMVMD	Philly Fed	Δln	Real personal cons. exp. - Total	1998:M11
46	RCONNDMVMD	Philly Fed	Δln	Real personal cons. exp. - Nondurables	1998:M11
47	RCONSHHMMVMD	Philly Fed	Δln	Real hh. cons. exp. - Services	2009:M8
48	RCONSMMVMD	Philly Fed	Δln	Real personal cons. exp. - Services	1998:M11
49	RCONSNPMMVMD	Philly Fed	Δln	Real final cons. exp. of NPISH	2009:M8
50	NCONGMVQD	Philly Fed	Δln	Nom. personal cons. exp. - Goods	2009:Q3
51	NCONHHMVQD	Philly Fed	Δln	Nom. hh. cons. exp.	0209:Q3
52	NCONSHHMVQD	Philly Fed	Δln	Nom. hh. cons. exp. - Services	2009:Q3
53	NCONSNPMVQD	Philly Fed	Δln	Nom. final cons. exp. of NPISH	2009:Q3
54	RCONDMVQD	Philly Fed	Δln	Real personal cons. exp. - Durable goods	1965:Q4
55	RCONGMVQD	Philly Fed	Δln	Real personal cons. exp. - Goods	2009:Q3
56	RCONHHMVQD	Philly Fed	Δln	Real hh. cons. exp.	2009:Q3
57	RCONMVQD	Philly Fed	Δln	Real personal cons. exp. - Total	1965:Q4
58	RCONNDMVQD	Philly Fed	Δln	Real pesonal cons. exp. - Nondurable goods	1965:Q4
59	RCONSHHMVQD	Philly Fed	Δln	Real hh. cons. exp. - Services	2009:Q3
60	RCONSMVQD	Philly Fed	Δln	Real personal cons. exp. - Services	1965:Q4
61	RCONSNPMVQD	Philly Fed	Δln	Real final cons. exp. of NPISH	2009:Q3
62	NCONQVQD	Philly Fed	Δln	Nom. personal cons. exp.	1965:Q4
			Group 5: Prices		
63	PCONGMMVMD	Philly Fed	$\Delta^2 ln$	Price index for personal cons. exp. - Goods	2009:M8
64	PCONHHMMVMD	Philly Fed	$\Delta^2 ln$	Price index for hh. cons. exp.	2009:M8
65	PCONSHHMMVMD	Philly Fed	$\Delta^2 ln$	Price index for hh. cons. exp. - Services	2009:M8
66	PCONSNPMMVMD	Philly Fed	$\Delta^2 ln$	Price index for final cons. exp. of NPISH	2009:M8
67	PCPIMVMD	Philly Fed	$\Delta^2 ln$	Consumer price index	1998:M11
68	PCPIXMVMD	Philly Fed	$\Delta^2 ln$	Core consumer price index	1998:M11
69	PPPIMVMD	Philly Fed	$\Delta^2 ln$	Producer price index	1998:M11
70	PPPIXMVMD	Philly Fed	$\Delta^2 ln$	Core producer price index	1998:M11
71	PCONGMVQD	Philly Fed	$\Delta^2 ln$	Price index for personal. cons. exp. - Goods	2009:Q3
72	PCONHHMVQD	Philly Fed	$\Delta^2 ln$	Price index for hh. cons. exp.	2009:Q3
73	PCONSHHMVQD	Philly Fed	$\Delta^2 ln$	Price index for hh. cons. exp. - Services	2009:Q3
74	PCONSNPMVQD	Philly Fed	$\Delta^2 ln$	Price index for final cons. exp. of NPISH	2009:Q3
75	PCONXMVQD	Philly Fed	$\Delta^2 ln$	Core price index for personal cons. exp.	1996:Q1
76	CPIQVMD	Philly Fed	$\Delta^2 ln$	Consumer price index	1994:Q3
77	PQVQD	Philly Fed	$\Delta^2 ln$	Price index for GNP/GDP	1965:Q4
78	PCONQVQD	Philly Fed	$\Delta^2 ln$	Price index for personal cons. exp.	1965:Q4
79	PIMPQVQD	Philly Fed	$\Delta^2 ln$	Price index for imports of goods and services	1965:Q4
			Group 6: Trade and Government		
80	REXMVQD	Philly Fed	Δln	Real exports of goods and services	1965:Q4
81	RGMVQD	Philly Fed	Δln	Real government cons. and gross inv. - Total	1965:Q4
82	RGFMVQD	Philly Fed	Δln	Real government cons. and gross inv. - Federal	1965:Q4
83	RGSLMVQD	Philly Fed	Δln	Real government cons. and gross. inv. - State and local	1965:Q4
84	RIMPMVQD	Philly Fed	Δln	Real imports of goods and services	1965:Q4
85	RNXMVQD	Philly Fed	Δlv	Real net exports of goods and services	1965:Q4
			Group 7: Money and Credit		
86	BASEBASAQVMD	Philly Fed	$\Delta^2 ln$	Monetary base	1980:Q2
87	M1QVMD	Philly Fed	$\Delta^2 ln$	M1 money stock	1965:Q4
88	M2QVMD	Philly Fed	$\Delta^2 ln$	M2 money stock	1971:Q2
89	NBRBASAQVMD	Philly Fed	$\Delta lv/lv$	Nonborrowed reserves	1967:Q3
80	NBRECBASAQVMD	Philly Fed	$\Delta lv/lv$	Nonborrowed reserves plus extended credit	1984:Q2
91	TRBASAQVMD	Philly Fed	$\Delta^2 ln$	Total reserves	1967:Q3
92	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, Ludvigson, and Ng (2015) and Ludvigson, Ma, and Ng (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 $\sqrt{T}/N \to \infty$), 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.

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 A.3: List of Financial Dataset Variables

No.	Short Name	Source	Tran	Description
		Group 1: Prices, Yield, Dividends		
1	D_log(DIV)	CRSP	Δln	$\Delta \log D_t^*$ see additional details below
2	D_log(P)	CRSP	Δln	$\Delta \log P_t$ see additional details below
3	D_DIVreinvest	CRSP	Δln	$\Delta \log D_t^{re,*}$ see additional details below
4	D_Preinvest	CRSP	Δln	$\Delta \log P_t^{re,*}$ 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
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)

Table A.3 (Cont'd)

No.	Short Name	Source	Tran	Description
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)
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)
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)

Table A.3 (Cont'd)

No.	Short Name	Source	Tran	Description
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_libra

- 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.

5.0.3 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, Ghysels, and Kourtellos (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. The raw daily data are also standardized before performing factor estimation (see generic description below). We estimate factors at each daily date in the sample using the entire history (from 23-Oct-1959) of variables observed in real time.

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

The data series used in this dataset are listed below in Table A.4 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 A.4: 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

Table A.4 (Cont'd)

No.	Short Name	Source	Tran	Description
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
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

Table A.4 (Cont'd)

No.	Short Name	Source	Tran	Description
64	MLAAA-10YTB	Data Stream, FRED	lv	Merrill Lynch Corporate Bonds: Aaa Rated: Effective Yield (%) minus Y10-Tbond
		Group 4: Treasuries		
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, $\mathbf{G}_{D,t}^{Q}$. Denote by G_t^D a factor in a daily frequency formed from the daily financial dataset and denote by G_t^Q a quarterly aggregate of the corresponding daily factor time series. Let $G_{N_D-j,d_t,t}^D$ 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 $G_{N_D,d_t,t}^D$. For simplicity, we suppress the subscript d_t thus $G_{N_D-j,d_t,t}^D \equiv G_{N_D-j,t}^D$.

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. G_t^Q is defined as:

$$G_t^Q(\mathbf{w}) \equiv \sum_{i=1}^{N_D} w_i G_{N_D-i,t}^D$$

where $\mathbf{w}$ is a vector of weights. 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 downweighted. 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 non-monotone 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, Sinko, and Valkanov (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 these 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 $G_t^Q(\mathbf{w})$:

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

This is done in real time using recursive regressions and an initial in-sample estimation window that matches the timing described below for the data-dependent choice of tuning parameter in the machine learning estimation (see the section on Estimation and Machine Learning).

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 $\mathbf{G}_{D,t}^Q$.

Estimation and Machine Learning

The model to be estimated is

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

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 $\boldsymbol{\beta}_{j,t}^{(i)}$ are estimated using the Elastic Net (EN) estimator, which depend on regularization parameter parameters $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)'$ (See the next section for a description of EN).

The procedure involves iterating on the following steps.

1. **Sample partitioning**: At time t, a prior sample of size $\widetilde{T}$ is partitioned into two subsample windows: an "in-sample" estimation subsample consisting of the first T_{IS} observations, and a hold-out "training" subsample of T_{TS} subsequent observations, i.e., $\widetilde{T} = T_{IS} + T_{TS}$.

2. **In-sample estimation**: Initial estimates of $\boldsymbol{\beta}^{(i)}$ are obtained using the EN estimator using observations $1, ..., T_{IS}$, given an arbitrary fixed (non-random) starting value for $\boldsymbol{\lambda}$. Denote this initial estimate $\boldsymbol{\beta}_{T_{IS}}^{*(i)}(\mathbf{X}_{T_{IS}}, \boldsymbol{\lambda})$, where "*" denotes the value of the estimator given an arbitrary $\boldsymbol{\lambda}$.

3. **Training and cross-validation**: The regularization parameter $\boldsymbol{\lambda}$ is estimated by minimizing mean-square loss $\mathcal{L}(\boldsymbol{\lambda}, T_{IS}, T_{TS})$ over pseudo-out-of-sample forecast errors generated from rolling regressions using only the most recent T_{IS} observations. That is, the first rolling prediction uses data from 1 to T_{IS}, the second rolling prediction uses data from 2 to $T_{IS} + 1$, etc., where

$$\mathcal{L}(\boldsymbol{\lambda}, T_{IS}, T_{TS}) \equiv \frac{1}{T_{TS} - h} \sum_{\tau = T_{IS}}^{T_{IS}+T_{TS}-h} \left(\mathcal{X}_\tau' \boldsymbol{\beta}_{j,\tau}^{*(i)}(\mathbf{X}_{T_{IS}}, \boldsymbol{\lambda}) - y_{j,\tau+h} \right)^2,$$

and where $\boldsymbol{\beta}_{j,\tau}^{*(i)}(\mathbf{X}_{T_{IS}}, \boldsymbol{\lambda})$ is the time τ EN estimate of $\boldsymbol{\beta}_j^{(i)}$ given $\boldsymbol{\lambda}$ and data through time τ in a sample of size T_{IS}.

4. Steps 1-3 are repeated over a grid of estimation and training sample window lengths T_{IS}^* and T_{TS}^* such that alternative partitions satisfy $T_{IS}^* + T_{TS}^* \leq \widetilde{T}$, where shorter window lengths remove consecutive observations at the start of the prior sample. The final machine estimate of $\boldsymbol{\beta}_{j,t}^{(i)}(\mathbf{X}_{\widetilde{T}}, \boldsymbol{\lambda})$ uses $\left\{ \widehat{\boldsymbol{\lambda}}, \widehat{T}_{IS}, \widehat{T}_{IS} \right\} = \underset{\boldsymbol{\lambda}, T_{IS}^*, T_{TS}^*}{\operatorname{argmin}} \mathcal{L}(\boldsymbol{\lambda}, T_{IS}^*, T_{TS}^*)$ and is denoted $\widehat{\boldsymbol{\beta}}_{j,t}^{(i)}\left(\mathbf{X}_{\widetilde{T}}, \widehat{\boldsymbol{\lambda}}\right)$.

5. **Out-of-sample prediction**: The values of the regressors at time t are used to make a true out-of-sample prediction of y_{t+h}, using $\widehat{\boldsymbol{\beta}}_{j,t}^{(i)}\left(\mathbf{X}_{\widetilde{T}}, \widehat{\boldsymbol{\lambda}}\right)$, and the machine forecast error $y_{t+h} - \mathcal{X}_t' \widehat{\boldsymbol{\beta}}_{j,t}^{(i)}\left(\mathbf{X}_{\widetilde{T}}, \widehat{\boldsymbol{\lambda}}\right)$ stored.

6. **Roll forward and repeat**: The initial in-sample subperiod is rolled forward one period and uses data from $2, ..., T_{IS} + 1$ and steps 2-5 are repeated until the final out-of-sample forecast is made for $y_{j,T}$, where T is the last period of our sample.

Averaging across the forecast errors from step 5 above, gives $\text{MSE}_{\mathbb{E}}$ over the *evaluation sample* $t = (T_{IS} + T_{TS} + h)$,...,$T$ that we use to assess whether belief distortions are present.

Any such distortions are quantified by looking at the ratio $\text{MSE}_{\mathbb{E}}/\text{MSE}_{\mathbb{F}}$ over the evaluation sample.

The procedure can be understood for a specific example. Let $h = 4$ quarters. The initial in-sample period is 1969:Q1-1973:Q4, corresponding to $T_{IS} = 20$. Suppose $T_{TS} = 16$. The initial in-sample regression is run for t =1969:Q1-1973:Q4 (dependent variable from 1970:Q1-1973:Q4, independent variable from 1969:Q1-1972:Q4), and the values of the regressors at t = 1973:Q4 are used to forecast y_j for 1973:Q4-1974:Q4, which serves as the first training forecast. All parameters are then reestimated from 1969:Q2-1974:Q1 and forecasts are recomputed for y_j for 1974:Q1-1975:Q1 and so on, until the final training forecast is made for y_j for 1977:Q3 to 1978:Q3. The squared forecast errors from the 16 training forecasts are stored and the regularization parameters $\boldsymbol{\lambda}$ are chosen to minimize the mean-square-forecast error over the training sample, with this chosen value denoted $\boldsymbol{\lambda}^*$. Next, the regressors at t =1978:Q3 are multiplied by $\hat{\boldsymbol{\beta}}_j^{(\tau)}(\boldsymbol{\lambda}^*)$ to produce an out-of-sample forecast of y_j for 1978:Q3-1979:Q3, which is the first observation in our "evaluation" sample. We repeat this on a grid of (T_{IS},T_{TS}), and the optimal combination of T_{IS} and T_{TS} minimizes $MSE_{j\mathbb{E}}(\boldsymbol{\lambda}^*, T_{IS}, T_{TS})$. The entire procedure is repeated by rolling forward to the next in-sample period 1969:Q1-1974:Q1.

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

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

One possibility is to run the machine model as a regression of forecast errors on time t information:

$$y_{j,t+h} - \mathbb{F}_t^{(i)}[y_{j,t+h}] = \alpha_j^{(i)} + \beta_{j\mathbb{F}}^{(i)}\mathbb{F}_t^{(i)}[y_{j,t+h}] + \mathbf{B}_{j\mathcal{Z}}^{(i)\prime}\mathcal{Z}_t + \epsilon_{jt+h}, \tag{A.12}$$

where for this case the machine efficient benchmark is characterized by $\beta_{j\mathbb{F}}^{(i)} = 0$; $\mathbf{B}_{j\mathcal{Z}}^{(i)} = \mathbf{0}$;$\alpha_j^{(i)} = 0$. Because elastic net shrinks estimated coefficients toward zero, this centering of the left-hand-side variable allows the machine to shrink $\beta_{j\mathbb{F}}^{(i)}$ toward one. In this case the machine forecast is given by

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

By contrast, if the machine forecast is implemented by running

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

then $\beta_{j\mathbb{F}}^{(i)}$ is shrunk toward zero and the algorithm will typically place less weight on the survey forecast than if the estimation (A.12) is run. 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 $\mathcal{L}(\boldsymbol{\lambda}, T_{IS}, T_{TS})$ in every training sample.

To capture non-linearities, the machine forecasts follow a simple switching model. In most periods, the forecast is based on the "normal-times" statistical model just described. To cope with rapid economic change, as in a recession, the machine forecast is permitted to switch to a simpler model based on a recession indicator. We use as the recession indicator the term spread, defined as the difference between the 10-year Treasury bond rate and the 3-month Treasury bill rate. When the term spread at time t is at or below the real time sample 10th percentile value, the machine forecast of $t+4$ is switched to a recession-model forecast which is based solely on a dummy indicator I_{t-4}, which takes the value 1 when the term spread at $t-4$

is below a threshold. The precise threshold used is the one that minimizes the mean-square loss function in the relevant training sample prior to the actual forecast. The machine chooses among thresholds that represent the real time sample 10th, 5th, or 1st percentile values for the term spread. The recession model forecast is the fitted value from a regression of real time real GDP growth at time t on the 4-quarter lagged value of I_{t-4}.

Elastic Net

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:

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

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

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

where $\mathcal{X}_t = (1, \mathcal{X}_{1t,\ldots,}\mathcal{X}_{Kt})'$ collects all the independent variable observations $\left(\mathbb{F}_t^{(i)} \left[y_{j,t+h} \right], \mathcal{Z}_{jt}\right)$ into a vector with "1" and $\boldsymbol{\beta}_j = \left(\alpha_j, \beta_{j\mathbb{F}}, \text{vec}\left(\mathbf{B}_{j\mathcal{Z}}\right)\right)' \equiv (\beta_0, \beta_1, \ldots\beta_K)'$ 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:

$$\hat{\boldsymbol{\beta}}^{\text{EN}} = \underset{\beta_0, \beta_1, \ldots, \beta_k}{\text{argmin}} \left\{ \sum_{t=1}^{T} \left(y_{j,t+h} - \mathcal{X}_t' \boldsymbol{\beta}_j \right)^2 + \underbrace{\lambda_1 \sum_{j=1}^{k} \left| \boldsymbol{\beta}_j \right|}_{\text{LASSO}} + \underbrace{\lambda_2 \sum_{j=1}^{k} \boldsymbol{\beta}_j^2}_{\text{ridge}} \right\}$$

By minimizing the MSE over the training samples, we choose the optimal λ_1 and λ_2 values simultaneously.

Dynamic Factor Estimation

We re-estimate factors at each date in the sample using the entire history of variables observed in real time. 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, Jiang, and Tong (2017). We run the following regression for each variable x_{it}:

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

Then, we form a new dataset of variables $\hat{\beta}_{j,i,x} x_{it}$ where $\hat{\beta}_{j,i,x}$ denotes the OLS estimate of $\beta_{j,i,x}$. 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

$$\widetilde{x}_{it} = \lambda_i' \mathbf{G}_t + e_{it}, \tag{A.13}$$

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.[12] Specifically, the $T \times r$ matrix $\hat{g}_t$ is $\sqrt{T}$ times the r eigenvectors corresponding to the r 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 (A.13) 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 A.2. 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.[13]

For the financial factors, we use all of the variables listed in Table A.3, and no additional steps are performed beyond those described above.

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

[13] 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.

Economic Interpretation of the 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 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 A.5 reports the series with largest average marginal R^2 for each factor of each large dataset.

Table A.5: 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: Orders and Investment
$G_{7,M,t}$	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

Let $\mathbf{Z}_{jt} \equiv \left(y_{j,t}, \hat{\mathbf{G}}'_{jt}, \mathbf{W}'_{jt}\right)'$ be a $r = 1 + r_G + r_W$ vector which collects $y_{j,t}$, the r_G estimated factors, and the r_W additional predictors, and define $\mathcal{Z}_{jt} \equiv \left(y_{j,t}, ..., y_{j,t-p_y}, \hat{\mathbf{G}}'_{jt}, ..., \hat{\mathbf{G}}'_{jt-p_G}, \mathbf{W}'_{jt}, ..., \mathbf{W}'_{jt-p_W}\right)'$, where p_y, p_G, p_W are lags of $y_{j,t}$, $\hat{\mathbf{G}}'_{jt}$, $\mathbf{W}'_{jt}$, respectively. We consider the following machine forecasting regression

$$y_{j,t+h} = \alpha_j + \beta_{j\mathbb{F}} \mathbb{F}_t^{(i)} [y_{j,t+h}] + \underbrace{\mathbf{B}_{j\mathcal{Z}}}_{1\times q} \mathcal{Z}_{jt} + \epsilon_{jt+h}$$

where $q = r + p_y + p_G + p_W$. Let superscript (i) refer 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 $y_{j,t}$, $\hat{\mathbf{G}}'_{jt}$, 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. $\mathbb{F}_{jt-k}^{(i)} [y_{jt+h-k}]$, where $k = 1, \ldots, 2$
2. $\mathbb{F}_{jt-1}^{(s \neq i)} [y_{jt+h-1}]$, where $s = mean$, 50, 25, 75 for all $s \neq i$
3. $\text{var}_N \left(\mathbb{F}_{t-1}^{(\cdot)} [y_{jt+h-1}]\right)$, where $\text{var}_N (\cdot)$ denotes the cross-sectional variance of lagged survey forecasts
4. $\text{skew}_N \left(\mathbb{F}_{t-1}^{(\cdot)} [y_{jt+h-1}]\right)$, where $\text{skew}_N (\cdot)$ denotes the cross-sectional skewness of lagged survey forecasts
5. Trend inflation measured as $\overline{\pi}_{t-1} = \begin{cases} \rho \overline{\pi}_{t-2} + (1-\rho)\pi_{t-1}, \ \rho = 0.95 & \text{if } t < \text{1991:Q4} \\ \text{CPI10}_{t-1} & \text{if } t \geq \text{1991:Q4} \end{cases}$ 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. $\mathbb{N}_t^{(i)} [\pi_{t,t-h}]$ = 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 $\hat{\mathbf{G}}'_{jt}$ 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.[14] 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. $\mathbb{F}^{(i)}_{jt-k}[y_{jt+h-k}]$, where $k = 1, 2$

2. $\mathbb{F}^{(s\neq i)}_{jt-1}[y_{jt+h-1}]$, where $s = mean$, 50, 25, 75 for all $s \neq i$

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

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

[14] 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

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. $\mathbb{N}_t^{(i)}[y_{t,t-h}]$ = 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 $\hat{\mathbf{G}}'_{jt}$ 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.[15] 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.

[15] 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

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,

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

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

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

7. Trend inflation measured as $\overline{\pi}_{t-1} = \begin{cases} \rho\overline{\pi}_{t-2} + (1-\rho)\pi_{t-1},\ \rho = 0.95 & \text{if } t < \text{1991:Q4} \\ \text{CPI10}_{t-1} & \text{if } t \geq \text{1991:Q4} \end{cases}$ 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).

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 $\hat{\mathbf{G}}'_{jt}$ 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.[16] 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 $\mathbf{W}'_{jt}$:

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

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

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

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

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

6. $\text{skew}_N\left(\mathbb{F}^{SPF,(\cdot)}_{t-1}[y_{jt+h-1}]\right)$, 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).

[16] 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

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 $\hat{\mathbf{G}}'_{jt}$ 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.[17] 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. $\mathbb{F}_t^{(i)}[y_{jt+h}]$ in $\mathbf{W}'_{jt}$, we now use survey forecasts from Blue Chip, instead of SPF.

[17]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

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.

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}.

$$\underbrace{\pi_{t+3} - \mathbb{F}_t^{(\mu)}\left[\pi_{t+3}\right]}_{\text{Forecast Error}} = \alpha + \beta\left(\underbrace{\mathbb{F}_t^{(\mu)}\left[\pi_{t+3}\right] - \mathbb{F}_{t-1}^{(\mu)}\left[\pi_{t+3}\right]}_{\text{Forecast Revision}}\right) + \epsilon_{t+3} \tag{A.14}$$

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

$$\pi_{t+3} = 100 \times \left(\frac{P_t}{P_{t-1}} \times \frac{P_{t+1}}{P_t} \times \frac{P_{t+2}}{P_{t+1}} \times \frac{P_{t+3}}{P_{t+2}} - 1\right). \tag{A.15}$$

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 Error

The forecast error on the LHS of the regressions (A.14) is constructed in the following way:

$$\begin{aligned}\pi_{t+3,t} - \mathbb{F}_t^{(\mu)}\left[\pi_{t+3,t}\right] \equiv 100 \times \Bigg[& \left(\frac{\pi_{t,t-1} - \mathbb{F}_t^{(\mu)}\left[\pi_{t,t-1}\right]}{400} + 1\right) \\ & \times \left(\frac{\pi_{t+1,t} - \mathbb{F}_t^{(\mu)}\left[\pi_{t+1,t}\right]}{400} + 1\right) \\ & \times \left(\frac{\pi_{t+2,t+1} - \mathbb{F}_t^{(\mu)}\left[\pi_{t+2,t+1}\right]}{400} + 1\right) \\ & \times \left(\frac{\pi_{t+3,t+2} - \mathbb{F}_t^{(\mu)}\left[\pi_{t+3,t+2}\right]}{400} + 1\right) - 1\Bigg]\end{aligned} \tag{A.16}$$

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

In-sample analysis

Table A.6 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 A.6: CG In-Sample Regressions of Forecast Errors on Forecast Revisions

Regression: $\pi_{t+3,t} - \mathbb{F}_t[\pi_{t+3,t}] = \alpha + \beta(\mathbb{F}_t[\pi_{t+3,t}] - \mathbb{F}_{t-1}[\pi_{t+3,t}]) + \delta\pi_{t+2,t-1} + \epsilon_t$				
	(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)
$\mathbb{F}_t[\pi_{t+3,t}] - \mathbb{F}_{t-1}[\pi_{t+3,t}]$	1.194**	1.141**	1.186**	1.116**
t-stat	(2.496)	(2.560)	(2.478)	(2.532)
$\pi_{t+2,t-1}$		0.021		0.027
t-stat		(0.435)		(0.574)
$\bar{R}^2$	0.195	0.197	0.193	0.195

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 $\mathbb{F}_t[\pi_{t+3,t}]$ is the SPF of annual inflation with information in period t and $\mathbb{F}_{t-1}[\pi_{t+3,t}]$ 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%.

Out-of-Sample Analysis

We seek to construct a series of real-time OOS forecasts using the model:

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

We estimate 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 real-time data available four quarters after the period being forecast. The CG model forecast for π_{t+3}

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

For the rolling procedure, we try windows of sizes $w = 5$, 10, and 20 years. For the recursive procedure, we try initial window sizes of 5, 10, and 20 years as well.

The survey and model errors are

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

We also compute rolling MSEs over different forecast samples of size P as

$$\text{MSE}_{\mathbb{F}} = \frac{1}{P}\sum_{s=1}^{P}\left(\text{survey error}_{t+s}\right)^2$$

$$\text{MSE}_{\text{CG}} = \frac{1}{P}\sum_{s=1}^{P}\left(\text{CG model error}_{t+s}\right)^2$$

Table A.7: Mean Square Errors for the CG Model and SPF

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

Notes: The table reports the ratio of MSEs of the CG model forecast over the survey forecast. The regression estimation uses the lastest 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.