Working Papers

FEDERAL RESERVE BANK OF BOSTON

No. 25-4

Forecasting U.S. Economic Activity with a Small Information Set

Daniel H. Cooper, Giovanni P. Olivei, and Hannah Rhodenhiser

Abstract:
We provide a parsimonious setup for forecasting U.S. GDP growth and the unemployment rate based on a few fundamental drivers. This setup yields forecasts that are reasonably accurate compared with private-sector and Federal Reserve forecasts over the 1984–2019 and post–COVID-19 pandemic periods. This result is achieved by jointly estimating the processes for GDP growth and the unemployment rate, with the constraint that GDP and unemployment follow Okun's law in first differences. This setup can be easily extended to replace the variables in the information set with factors that might better capture the underlying fundamentals.

JEL Classifications: E27, E37
Keywords: Macroeconomic forecasting, small information set, forecast accuracy

---

Danile H. Cooper (Daniel.Cooper@bos.frb.org) is a vice president and economist in the Federal Reserve Bank of Boston Research Department. Giovanni P. Olivei (Giovanni.Olivei@bos.frb.org) is a senior vice president and the deputy director of research at the Federal Reserve Bank of Boston. Hannah Rhodenhiser (hmrh@bu.edu) is an economics PhD candidate at Boston University.

The authors thank Ella Steciuk for excellent research assistance with this project.

This version: June 2025 **https://doi.org/10.29412/res.wp.2025.04**

1. Introduction

In the large literature assessing the properties and accuracy of private sector and Federal Reserve Board forecasts of U.S. macroeconomic conditions, the results vary depending on the chosen benchmark forecast and the sample period considered. Nevertheless, it is not uncommon to find evidence that univariate autoregressive models and/or atheoretical methods involving a large amount of data can perform as well as (and sometimes better than) private sector and Federal Reserve Board (hereafter Federal Reserve) forecasts – at least for certain variables and at forecast horizons beyond the very near term.[1] Univariate and large-data forecasting approaches are important because they generate projections that are statistically well grounded and also replicable, unlike the judgmental forecasts from the private sector and Federal Reserve. However, these non-judgmental approaches, especially those based on univariate autoregressive methods, may not be well suited to providing forecast narratives based on macroeconomic fundamentals – a desirable property, especially when producing forecasts to inform monetary policymakers.

In this paper, we show how a simple framework based on a very small set of macroeconomic variables can produce forecasts for the unemployment rate and GDP growth that are comparable to forecasts from the private sector and the Federal Reserve. This parsimonious approach to forecasting performs well for the 1984–2019 period and for the post–COVID-19 pandemic period of 2021 through 2024. The information set we use includes the unemployment rate, a measure of consumer sentiment, the federal funds rate, and a corporate credit risk spread, with the relationship between the unemployment rate and GDP constrained to follow Okun's law in first differences.

[1] See, for example, Faust and Wright (2009), Chang and Hanson (2016), Crump et al. (2025), and Bernanke and Boivin (2003).

Not surprisingly, some previous studies evaluate forecasts against benchmarks that, even in reduced form, are based on a simple set of variables that capture important macroeconomic dynamics. Sims (2002), for example, compares Federal Reserve forecasts with forecasts from a five-variable, reduced-form vector autoregression (VAR) that includes real GDP, the GDP deflator, an index of commodity prices, the federal funds rate, and the three-year Treasury yield. This VAR setup generally underperformed Federal Reserve forecasts. Clark and McCracken (2008) illustrate how the forecasting accuracy of a small-scale VAR depends on a host of choices – such as lag selection, detrending methods, the estimation window, and accounting for breaks – which can differ depending on the variable being forecast. Another large literature focuses on forecasts based on medium-scale dynamic stochastic general equilibrium (DSGE) models and "semi-structural" approaches using theoretical restrictions. With these models, too, the forecasting accuracy relative to the private sector and/or the Federal Reserve is mixed.[2]

Despite the many studies comparing the performance of different forecasting methods, there is limited consensus on a small set of macroeconomic variables that would generate forecasts that, across horizons and over time, are generally as accurate as those from private-sector forecasters and the Federal Reserve. That said, much of the variance of many macroeconomic variables can be captured by a small set of factors (Stock and Watson 2002, 2017), the dynamics of which can be used to back out reasonably accurate forecasts of key variables. Yet, the economic interpretation of these factors is not always straightforward, limiting the usefulness of dynamic factor models in policy settings where a forecast narrative is important.

This paper makes three main contributions to the literature. First, it presents a parsimonious benchmark for forecasting U.S. economic activity based on important fundamental drivers that can be used to construct a broad narrative while yielding forecasts that are reasonably accurate compared with

[2] For a review of the literature, see Giacomini (2015).

private sector and Federal Reserve forecasts. Our approach also embeds long-run values for the federal funds rate, the unemployment rate, and GDP growth, all of which are crucial components of the forecasting process from the perspective of a policymaker. Second, with respect to pure forecasting method performance, the variables in our information set provide guidance for extensions that could replace some (or all) of those variables with factors. In this regard, one takeaway from our exercise is that information about the labor market, household balance sheets, the interest rate environment, and credit market sentiment is important in forecasting GDP growth and the unemployment rate. It is possible that using factors that better summarize relevant information along these four dimensions yields even more accurate forecasts. Third, our results demonstrate that exploiting some well-established empirical relationships, such as the one described by Okun's law, can lead to more accurate forecasts. Indeed, we show that forecasting GDP growth with a parsimonious set of variables is challenging unless constraints are imposed on the joint behavior of GDP growth and the unemployment rate, a result that likely also applies to other forecasting methods.

These contributions are relevant – to varying degrees – to different strands of the forecasting literature, which is too vast to fully summarize. Nevertheless, the analysis relates most closely to studies that assess forecast accuracy relative to a simple benchmark.[3] Research by Faust and Wright (2009) is a prominent example of comparing Federal Reserve forecasts of inflation and GDP growth with the forecasts from several atheoretical methods ranging from simple univariate benchmarks (such as random walk or autoregressive forecasting) to approaches that use many predictors. Faust and Wright find that simple univariate autoregressive forecasts can be more accurate than Federal Reserve forecasts for GDP growth after the Federal Reserve's "nowcasting" advantage is factored in. In a follow-up paper, Chang

[3] This literature is closely tied to the influential work of Meese and Rogoff (1983), which examines the forecasting performance of several exchange rate models vis-à-vis a simple random walk.

and Hanson (2015) consider a broader set of real variables forecast by the Federal Reserve. Their results yield a similar conclusion, that at least one of their forecasting benchmarks performed as well as the Federal Reserve forecasts at the one-year horizon.[4]

Despite some evidence of forecasting success, simple univariate benchmarks do not seem to have found their way into policy discussions. There may be many reasons for this, although one cannot rule out that information from simple forecast benchmarks is factored into the judgmental outlook. Nevertheless, the lack of references to simple univariate benchmarks likely also reflects the aforementioned limited usefulness of these benchmarks for building a forecast narrative and the possibility that the relevant benchmark changes over time. Indeed, there are often questions regarding forecast stability with reduced-form specifications – concerns that also apply to our forecasting approach. However, the fact that our framework is based on a parsimonious set of fundamental drivers for overall macroeconomic conditions could make it more stable and reliable over time.

The rest of the paper is structured as follows. Section 2 describes our setup, some criteria for forecast comparison, and the data used in the analysis. Section 3 discusses our estimation methods, and Section 4 illustrates the main results of our analysis, together with some robustness checks. We provide concluding comments in Section 5.

2. A Simple Reference Framework for Evaluating Forecasts

We evaluate Federal Reserve and private sector forecasts relative to the results from a simple benchmark that generates direct multistep projections for the change in the unemployment rate and the

[4] The Federal Reserve's forecasts of the unemployment rate, however, were more accurate than those of univariate benchmarks.

change in GDP from two equations that are estimated jointly and take the following general form at horizons $i \geq 1$:

$$\begin{aligned} \Delta u_{t+i,t-1} &= \boldsymbol{\beta}_i \mathbf{X}_{t-1} + \delta_i + \varepsilon_{u,t+i} \\ dy_{t+i,t-1} &= dy^{POT}_{t+i,t-1} + \gamma_i \Delta u_{t+i,t-1} + \varepsilon_{y,t+i}. \end{aligned} \tag{1}$$

The variable $\Delta u_{t+i,t-1}$ is the difference between the unemployment rate at time $t+i$ and $t-1$, while $dy_{t+i,t-1}$ is the annualized percentage growth rate of real GDP from $t-1$ to $t+i$, and $dy^{POT}_{t+i,t-1}$ is an estimate of potential GDP growth over the same horizon. The coefficients δ_i, γ_i, and in the vector $\boldsymbol{\beta}_i$ are indexed by i, as their values will vary with the horizon of the multistep projection. The setup links future unemployment rate changes to a vector of variables $\mathbf{X}_{t-1}$ known at time t when the forecast is made. Future GDP growth relative to potential growth is linked to the projected change in the unemployment rate by means of a simple Okun's law relationship in first differences. According to this relationship, whenever the unemployment rate is projected to decline (increase), GDP is expected to grow above (below) potential.

We consider time t forecasts of the change in the unemployment rate and the change in GDP over the next $t+i$ quarters, where i ranges from 1 to 7. Since $t-1$ is the reference quarter for computing changes, this generates GDP and unemployment rate projections two to eight quarters out. We do not evaluate current-quarter (time t) forecasts or "nowcasts," as this setup is not well suited to such a purpose.[5]

The intercept δ_i in the unemployment rate equation is constant in our baseline specification but is time-varying in an alternative specification in which we assume the intercept follows a random walk. The motivation for considering a time-varying intercept is that $\mathbf{X}_{t-1}$ will include variables for which a sample

[5] For a review of nowcasting methods, see Cascaldi-Garcia, Luciani, and Modugno (2023).

average may not be a good approximation of the variable's longer-run (equilibrium) value at the time of the forecast.

We rely on a small information set for $\mathbf{X}_{t-1}$. Specifically, we include the unemployment rate, consumer sentiment, a measure of the Federal Reserve's monetary policy stance, and a credit spread. The current state of the labor market is informative about future changes in the unemployment rate due to the cyclical nature of the economy. If the unemployment rate is sufficiently high (low), it can be expected to decline (increase) over the forecast horizon.[6] In addition, a long-standing literature documents the value of consumer sentiment as an indicator of a broader set of economic fundamentals (see Fuhrer 1993; Ludvigson 2004), which the other (limited) variables in $\mathbf{X}_{t-1}$ may not capture entirely. Consumer sentiment has the advantage of being a timely summary statistic that does not revise and has some forecasting power. Recently, the ability of some consumer sentiment measures to provide an adequate summary of household fundamentals has come into question – an issue we return to later. Nevertheless, at least over the pre–COVID-19 period, which is the focus of most of our analysis, sentiment had predictive power for future economic activity.

Even if consumer sentiment correlates more with future household spending than with business spending, it is still a highly relevant forecast variable for an economy such as the United States, in which consumer outlays represent a predominant share of GDP. Moreover, our information set includes the deviation of the federal funds rate from its estimated "neutral" rate, a gauge of the monetary policy stance that is important for predicting the evolution of the interest-sensitive components of demand, including business capital formation. Finally, the information set also includes a measure of the corporate bond spread, which is important because, as the literature documents, credit spreads can be predictive of

[6] Wages are one element of this equilibrating mechanism, as compensation rises when the unemployment rate is low, thus reducing firms' incentive to hire, and vice versa when the unemployment rate is high.

economic activity.[7] The equilibrium values for some of the variables in $\mathbf{X}_{t-1}$ – for example, the federal funds rate and the unemployment rate – can change over time (Laubach and Williams 2003; Shimer 1999), which explains why we allow δ_i to be time-varying in an alternative specification.

2.1 Forecasts and Their Actual Values

We compare the accuracy of quarterly forecasts generated by the simple two-equation setup in (1) with the accuracy of forecasts from the Federal Reserve, the Survey of Professional Forecasters, and the Blue Chip Economic Indicators. The Federal Reserve forecasts are featured in the Greenbook or, more recently, the Tealbook (GB/TB hereafter) and are prepared by the Federal Reserve Board staff before each Federal Open Market Committee meeting. There are eight GB/TB forecasts per year over our sample period, and for comparison purposes, we use the final forecast made in each quarter, which is typically from the third month of the quarter.

The Survey of Professional Forecasters (SPF), currently conducted by the Federal Reserve Bank of Philadelphia, solicits professional private-sector forecasters' outlook for GDP, the unemployment rate, and other macroeconomic variables in the second month of each quarter. We compare forecasts from (1) with the median SPF forecasts. The Blue Chip Economic Indicators (BC) is also a survey of private-sector forecasters, some of whom are also included in the SPF sample. The BC survey is conducted at the beginning of each month, and we compare our results with the median BC "consensus" forecast published in the third month of each quarter.

Figure 1 shows the relative timing of the GB/TB, SPF, and BC forecasts that we use in a given quarter *t*, based on the months in each quarter. Typically, the SPF forecast is available first, followed by

[7] See, among others, Philippon (2009) and Gilchrist and Zakrajšek (2012).

BC and then GB/TB. In some instances, however, the GB/TB forecast is released in the second month of the quarter, so the timing advantage of the GB/TB forecast in terms of seeing additional data over the course of the quarter is not always present.

Forecasts obtained from our simple framework in (1) use $t-1$ information that is available at time t, when the forecast is made. Given this timing, we use the first (or "advance") estimate of real GDP in quarter $t-1$, which is released near the end of the first month of quarter t, together with $t-1$ values for the unemployment rate, consumer sentiment, the federal funds rate, and the credit spread. Information about these variables is released in a timely fashion, which means that their $t-1$ values are all available before the end of the first month of quarter t.

Due to this timing, our benchmark forecasts do not benefit from more information that becomes available over the course of quarter t and is used, to different degrees, by Federal Reserve Board staff and professional forecasters. Including more timely information would lead to only minimal improvement in the accuracy of our benchmark forecasts; nevertheless, to close some of the information gap, we include the value of the unemployment rate in the first month of quarter t in addition to its $t-1$ average in our information set.[8] Given that information about the unemployment rate prevailing in the first month of a given quarter becomes available very early in the second month of the quarter, the timing of available information for our forecast can be thought of as coinciding roughly with the timing of the SPF forecast. Indeed, including time t information about the unemployment rate suggests that our forecast is not overly penalized relative to the others.

In addition, we use the most recent vintages of real GDP and the unemployment rate available as of 2025:Q1 to calculate forecast errors. Any differences in the unemployment rate between its most recent

[8] For the other variables – consumer sentiment, the federal funds rate, and the credit spread – we use only time $t-1$ information.

vintage and earlier ones are due to small revisions to monthly seasonal factors. As a result, discrepancies between unemployment rate vintages tend to be negligible, especially at quarterly frequency. Revisions to GDP over time, however, can be substantial. Moreover, the Bureau of Economic Analysis (BEA) shifted from fixed to chained weights in the mid-1990s, and it expanded coverage of expenditures, such as including intellectual property investment late in our sample period. This creates some tradeoffs in determining the "actual" value of GDP to use in accessing forecast accuracy. Many forecast evaluation studies use a value that was reported not long after the forecast was made. For example, Romer and Romer (2000) use the BEA's second GDP release, which is published two months after the end of the quarter for which the GDP is being measured. Their rationale is that interest in a forecast is typically higher around the time the forecast is made. Moreover, such "real-time" forecast errors do not penalize forecasters for a later change in GDP methodologies and accounting. At the same time, as Sims (2002) notes, the most recent data vintage should provide the most reliable assessment of the evolution of GDP over time, and it is not clear why a forecast should be penalized for a miss in real time that is later overturned by updated data.

While we are aware that choosing a particular data vintage for assessing forecast accuracy has advantages and drawbacks, we follow Sims (2002) and use the most recent GDP vintage available at the time of writing.[9] This approach increases the forecast accuracy of our simple framework to some extent, since Okun's law errors in real time have some predictive power for future GDP revisions.[10] That is, if real-time GDP growth falls short of what Okun's law would have predicted, later vintages will tend to show stronger growth, and vice versa. Given this property of GDP revisions, our framework will better predict GDP growth when actual values are taken from later vintages.

[9] Importantly, this choice pertains to the value we use for *realized* GDP. The GDP data we use to generate forecasts based on (1) are the most recent vintage available at the time of the forecast.
[10] See, for example, Barnes et al. (2012) and Jordà et al. (2020).

2.2 Forecasting Horizons and Data

Our forecasting exercise mimics, as best as possible, a forecast made in real time, meaning that we use only information that was available before the quarter in which the forecast was made. We try to adhere to this pseudo out-of-sample forecasting method by using real-time data whenever possible.

We begin our forecast comparisons with 1984:Q1, and based on our real-time-data approach, this means that in constructing forecasts as of 1984:Q1, we consider only data available through 1983:Q4.[11] Using this information, we construct our first two-quarter benchmark forecast covering the period 1983:Q4 to 1984:Q2, our first three-quarter forecast covering the period 1983:Q4 to 1984:Q3, and so on up to the first eight-quarter forecast, which covers 1983:Q4 to 1985:Q4. The last quarter included in our benchmark forecasts is 2019:Q4 – right before the start of the COVID-19 pandemic. As a result, the last two-quarter forecast is made as of 2019:Q3 and covers 2019:Q2 to 2019:Q4, while the last eight-quarter forecast covers 2017:Q4 to 2019:Q4.

Our main forecast comparison exercise stops before the onset of the COVID-19 pandemic because the unique nature of the pandemic made our setup – a simple reduced-form forecasting framework based on historical regularities with a limited information set – inadequate for predicting macroeconomic outcomes. Nevertheless, we briefly consider how our framework would have performed during the pandemic recovery, as this also provides an opportunity to discuss issues related to the recent behavior of consumer sentiment.

[11] Our estimation period starts in 1966:Q1, as discussed in Section 3.

While our framework can always generate forecasts over the full two- to eight-quarter-ahead horizon, private-sector and Federal Reserve forecasts can be less frequent.[12] In all cases, we limit the comparison of the GB/TB, SPF, and BC forecasts with the forecasts from our framework to dates when both forecasts were available.

Regarding data sources, real-time GDP vintages come from the Real-Time Research Data Center at the Federal Reserve Bank of Philadelphia. The series includes real fixed-weight GNP for all vintages before 1992; real fixed-weight GDP from 1992 through 1995; and real chain-weight GDP from 1996 onward. The final quarterly observation for each of these vintages is the BEA's first or "advance" release for the quarter. For example, our forecasts as of 1984:Q1 use the vintage GDP data that have the 1983:Q4 advance GDP release as their final value. We also use the most recent GDP vintage at the time of the forecast to estimate potential GDP growth, $d\hat{y}^{POT}_{t-1,t-i-1}$, in (1), with a sample starting date of 1954:Q1 whenever possible. This estimate is computed based on the Hodrick-Prescott (HP) trend for the log level of real GDP, with a smoothing parameter set at 10^4. This choice produces a "rigid" estimated trend, which reduces the influence of data endpoints at the cost of potentially delaying the detection of persistent trend changes. We take the endpoint estimate for $d\hat{y}^{POT}_{t-1,t-i-1}$ as the growth of potential GDP that is expected to prevail over the forecast horizon.

We measure the unemployment rate as the quarterly average of the reported monthly rates for civilians aged 16 years and older. As discussed earlier, the information set, $\mathbf{X}_{t-1}$, includes both the quarterly unemployment rate as of $t-1$ and the rate prevailing in the first month of quarter t. The other variables in $\mathbf{X}_{t-1}$ include the index of consumer sentiment from the University of Michigan Survey of

[12] SPF quarterly forecast horizons are no longer than five quarters. BC forecasts decrease in number as the forecast horizon expands. Specifically, forecasts for six, seven, and eight quarters out are available three times, two times, and one time per calendar year, respectively. In earlier periods, the longer horizon GB/TB forecasts also are slightly less frequent.

Consumers, which is a real-time measure of consumer attitudes based on five different questions. Here, too, we take quarterly averages of the monthly data because even though measured consumer sentiment is persistent, the monthly data are noisy.[13] $\mathbf{X}_{t-1}$ also includes a measure of interest rates, defined as the difference between the federal funds rate and the so-called neutral rate, which we estimate as the sum of an equilibrium real federal funds rate and long-run inflation expectations. These expectations proxy for the evolution of the Federal Reserve's inflation target over time.[14] Long-run inflation expectations are taken from a series in the Federal Reserve Board's FRB/US model that corresponds to the SPF's real-time long-term inflation forecasts. These expectations are available starting in 1979. Before that date, the series is back-cast using a learning model that mimics the behavior of SPF inflation expectations (see Brayton and Tinsley 1996).[15] In addition, we do not directly include the equilibrium real federal funds rate in $\mathbf{X}_{t-1}$, which is absorbed by the intercept δ_i in (1). When δ_i is constant, our framework effectively assumes that the equilibrium real federal funds rate is unchanged over the sample period for which (1) is being estimated. With the real portion of the neutral federal funds rate absorbed by the intercept δ_i, the interest rate in $\mathbf{X}_{t-1}$ becomes the nominal federal funds rate less long-run inflation expectations. These expectations are available at a quarterly frequency only, and we take the nominal federal funds in the third month as the quarterly value.[16]

[13] Indeed, using more up-to-date monthly sentiment data, such as the index value in the third month of quarter t - 1 or the first month of quarter t, does not improve forecast accuracy.
[14] The FOMC did not make its 2 percent inflation target explicit until 2012, and the target was likely higher in earlier parts of our sample.
[15] The series, with mnemonics "PTR," is retrievable at https://www.federalreserve.gov/econresdata/frbus/us-models-package.htm. It is measured on a PCE deflator basis and therefore features a small downward adjustment relative to the SPF measure, which is CPI-based.
[16] The reason for using the last month of the quarter value is that the federal funds rate is highly persistent, and there are forecasting accuracy gains from using more up-to-date values. Given the timing of the forecasts, one could also use the value of the nominal federal funds rate in the first month of quarter t, as we do with the unemployment rate. While such a choice offers a small forecast improvement, it does not materially alter our findings.

The last variable in $\mathbf{X}_{t-1}$ is the Gilchrist and Zakrajšek (2012) credit spread, or GZ spread. This measure, which is available at a monthly frequency, is constructed from secondary market prices of nonfinancial firms' debt instruments. Unlike a standard credit-spread measure, such as the BAA corporate bond index relative to the 10-year Treasury yield, the GZ spread carefully matches each individual corporate debt instrument to a synthetic risk-free security that mimics exactly its fixed coupon schedule. This approach avoids the duration mismatch present in standard credit-spread measures.[17]

The GZ spread is produced by the Federal Reserve and is updated monthly, with each monthly update extending the series through the preceding calendar month.[18] The data are timely, with updates typically posted by the fourth business day of each month. We include the GZ spread from the third month of quarter $t-1$ in $\mathbf{X}_{t-1}$. Importantly, while the GZ spread is a good example of data available in real time for forecasting, the series was not compiled and posted until late in our sample period. Therefore, the series is "new" from a forecasting perspective, which yields the potential for so-called look-ahead bias – a caveat that should be kept in mind when interpreting our results.[19]

[17] The synthetic risk-free securities are priced off the Treasury yield curve as estimated by Gürkaynak, Sack, and Wright (2007). As a result, the GZ spread is subject to revisions over time, as there may be changes in the panel of bonds being used as reference points and in the estimates of the Treasury yield curve. These revisions, however, appear to be minor. For example, the correlation between the GZ spread series as reported in Gilchrist and Zakrajšek's (2012) original article and the most recent vintage at the time of writing is around 0.99. Lacking real-time vintages for this series, we use the most recently available vintage.

[18] A data file with the GZ spread series can be downloaded from https://www.federalreserve.gov/econresdata/notes/feds-notes/2016/files/ebp_csv.csv. The series is a research product from the Federal Reserve Board.

[19] Since values for the GZ spread series start in 1973, and for the estimation of (1) we use data going back to 1966, we back-cast the series using monthly values of the standard BAA–10-year Treasury yield spread.

3. Estimation Approach

We estimate (1) via full information maximum likelihood.[20] The sample period is an expanding window with the starting date fixed at 1966:Q1. As mentioned, the first forecasts we consider are as of 1984:Q1, which means that the first (and shortest) estimation window is 1966:Q1 to 1983:Q4. Considering shorter estimation windows to extend our forecast evaluation exercise further back in time would generate unstable coefficient estimates, possibly reflecting small-sample bias. However, even with our expanding estimation horizon, the estimated effect of the credit spread on economic activity can be unstable; in a few instances, its coefficient is the opposite sign of what one would expect. Whenever this is the case, we estimate (1) without the credit spread in $\mathbf{X}_{t-1}$ because it is unlikely that a forecaster would use such a variable at a given point in time if its impact on activity is counterintuitive. This restriction, however, has little impact on our findings, as the credit-spread effect's coefficient consistently has the expected sign once the estimation window includes observations from the 1990s.

In addition, when we allow for a time-varying intercept in the unemployment rate equation in (1), we assume it evolves following a random walk:

$$\delta_{i,t} = \delta_{i,t-1} + \eta_{i,t}\,, \tag{2}$$

where $\eta_{i,t}$ is uncorrelated with the error terms $\varepsilon_{u,i}$ and $\varepsilon_{y,i}$ in (1). Maximum likelihood estimates of (1) – (2) implemented by the Kalman filter will result in estimates for the standard deviation ($\sigma_{\eta,i}$) of the innovation in (2) that are biased toward zero. We use the Stock and Watson (1998) median unbiased estimator to address this so-called pile-up problem (see the appendix for further details).

[20] Maximum likelihood estimation will produce consistent parameter estimates despite the serial correlation that will be present in the error terms $\varepsilon_{u,i}$ and $\varepsilon_{y,i}$ in (1). Standard errors for the estimated parameters, however, will not be consistent (Levine 1983), but hypothesis testing is not the focus of this exercise.

4. Empirical Results

Our baseline results rely on estimating (1) with an intercept δ_i that is *not* time varying. Therefore, variation in δ_i, as with the other parameters in (1), comes only from the expanding window over which (1) is estimated. Figure 2 shows point estimates over time and at different forecast horizons for the most relevant coefficients: the unemployment rate, consumer sentiment, the interest rate, the credit spread, and the Okun's law coefficient.[21] There is some instability at the beginning of the sample – likely due to the short estimation window – and during the Great Recession for all reported parameters. However, with the important exception of the credit spread, the parameter estimates are relatively stable, especially for such a simple estimated reduced-form relationship. That said, the estimated parameter for Okun's law exhibits some low-frequency variation, which may be due to the very simple first-difference specification that we use. In particular, we do not explicitly model lagged dynamics, including the possibility of "error-correction" of GDP to the unemployment rate in levels.

By contrast, the credit-spread effect varies significantly over time. It is smaller in earlier periods[22] but becomes larger when the 2001 recession and, especially, the Great Recession are included in the estimation window. This result speaks to the difficulty, documented in the literature, of finding financial variables that have a stable effect on activity over time.[23] The result also points to the possibility of nonlinear effects, the modeling of which goes beyond the scope of this analysis. Still, consistent with Gilchrist and Zakrajšek (2012), inclusion of the GZ spread is important for reducing forecasting errors, especially during the Great Recession.

[21] The figure omits estimates of the constant value of the unemployment rate and the value at the first month of the quarter (expressed as a difference relative to the average value prevailing in the previous quarter).
[22] As mentioned earlier, we drop the credit spread from estimates of (1) when its coefficient has a counterintuitive sign.
[23] See, for example, Stock and Watson (2003).

Table 1 reports our baseline findings regarding forecast accuracy, showing for each forecast horizon the root mean squared error (RMSE) from the SPF, BC, and GB/TB forecasts relative to the RMSE from our simple forecasting framework. Results are reported for different sample periods to ensure that our results are not driven by specific episodes. A value greater than 1 implies less forecast accuracy relative to our benchmark, and vice versa when the value is less than 1. The table also includes asterisks denoting whether we can reject the null that the forecasts being compared are equally accurate based on the Diebold-Mariano (1995) test.[24] The top panel in the table reports results from forecasts of the unemployment rate, while the bottom panel shows results from GDP growth forecasts.

Most of the reported ratios in the table are greater than 1, especially for the unemployment rate forecasts. This implies that over the 1984–2019 period, forecasts of the unemployment rate and GDP growth generated using our simple framework tend to have a lower RMSE than the equivalent SPF, BC, and GB/TB forecasts. The results also show that among the comparison forecasts, the GB/TB forecast almost uniformly features an RMSE that is lower than that of the SPF or BC forecast, though the difference becomes smaller in more recent samples (especially compared with BC), as previously documented by Reifschneider and Tulip (2019).

In addition, the GB/TB forecasts tend to be more accurate than ours at shorter horizons, especially for GDP growth. This finding is broadly consistent with previous work (such as Faust and Wright 2009), documenting the GB/TB advantage in forecasting near-term economic activity. As Faust and Wright (2009) highlight, the Federal Reserve's "nowcasting" of GDP relies on replicating "key elements of the data construction machinery of the Bureau of Economic Analysis" along with important judgment concerning "large transitory events … such as dock strikes or hurricanes." However, this information

[24] For the test, we use a quadratic loss function. The Diebold-Mariano statistic is calculated by regressing the loss differential on a constant, using heteroskedasticity and autocorrelation consistent (HAC) standard errors. See Diebold (2012).

advantage largely disappears beyond the three-quarter horizon. Indeed, at policy-relevant horizons of four to six quarters, our simple forecasting framework fares no worse than the GB/TB forecasts.

The table also illustrates how the in-sample RMSE differences do not translate into clear-cut differences in a statistical sense. Indeed, one can rarely reject (at standard confidence levels) the null that the accuracy of the forecasts is the same based on a Diebold-Mariano test. One likely reason is that the number of independent forecast observations being considered, while covering approximately 25 years, is still small.

Panel A of Figure 3 compares four-quarter-ahead forecasts from our framework with the GB/TB four-quarter-ahead forecasts – the two types of forecasts in the sample that, overall, appear to be relatively more accurate. It is evident, from both the unemployment rate and GDP charts, that during the 1990–1991 and 2001 recessions, the GB/TB was more accurate. However, during the Great Recession, the forecasting performances were similar, as the GZ spread helped our forecast accuracy considerably. Still, while the Federal Reserve staff, like other forecasters, may have a difficult time predicting recessions, their forecasts often outperform our simple benchmark (and those of private forecasters) once they realize that the economy is in a recession and apply relevant judgment. Panel B of the figure repeats the exercise in Panel A but compares our four-quarter-ahead forecasts with those of private forecasters. For the unemployment rate, the SPF and BC forecasts, like the GB/TB, are more accurate than ours early in the sample period, when our benchmark is estimated over a relatively short window and exhibits some of its aforementioned instability. However, after the mid-1980s, there are no long stretches in which private unemployment rate forecasts are clearly more accurate than ours. And while private forecasts for GDP growth were more accurate during the 1990 recession, it is again difficult to point to an extended period outside that window in which they represent an obvious improvement.

4.1 Results from Single-Equation Estimation

Our main results are based on estimating the two equations in (1) jointly using maximum likelihood. However, it is instructive to see how the results differ when we estimate the two equations separately using ordinary least squares (OLS). This approach does not impose cross-equation restrictions and is less efficient than our baseline estimates because it ignores potential correlation between the error terms ε_u and ε_y. Without these restrictions, it matters in which order the two equations are estimated – that is, whether we predict the unemployment rate first with $\mathbf{X}_{t-1}$ and GDP second given the unemployment rate forecast, or vice versa. We show that with a limited information set, predicting the unemployment rate first yields more accurate forecasts.

Table 2 reports the results of this exercise. It shows the RMSE of the forecast from our simple framework estimated via OLS relative to the RMSE of our forecast estimated via maximum likelihood.[25] The table shows results for 1984 through 2019, as our findings do not differ materially when we consider different subsamples. As panel A illustrates, forecasting the change in the unemployment rate first with the information set $\mathbf{X}_{t-1}$ and then using the result to project GDP growth via Okun's law in differences closely approximates estimating (1) jointly via maximum likelihood.

Suppose instead we estimate the following two equations separately using OLS for each of the $t+i$ horizons:

$$\begin{aligned} dy_{t+i,t-1} &= dy^{POT}_{t+i,t-1} + \boldsymbol{\psi}_i \mathbf{X}_{t-1} + \theta_i + \varepsilon_{y,t+i} \\ \Delta u_{t+i,t-1} &= \zeta_i (dy_{t+i,t-1} - dy^{POT}_{t+i,t-1}) + \varepsilon_{u,t+i}, \end{aligned}$$

Here, we use the information set $\mathbf{X}_{t-1}$ to predict GDP growth relative to potential GDP growth and then

---

[25] We are aware of the debate in the literature on the appropriateness of the Diebold-Mariano test in comparing the forecasting performance of alternative models (see, for example, West 2006; Clark and McCracken 2013). We use the test in Table 2 and in later tables as a rough gauge of relative forecasting performance.

use the result to forecast the unemployment rate via Okun's law, after having inverted the dependent and independent variables. This results in forecasts for both GDP growth and the unemployment rate that are significantly less accurate than our baseline forecasts (see Table 2, Panel B). Given the results in Panel A, this also means that these forecasts are much less accurate than projections using available information to first forecast the unemployment rate and then forecast GDP growth via Okun's law.

These results illustrate why the simple forecasting approach used in this paper is more successful than previous attempts at forecasting GDP growth with a small information set over an extended sample period. A small information set like ours can be used to forecast the unemployment rate relatively well, which in turn helps to forecast GDP growth. In other words, a more accurate forecast of GDP growth requires an intermediate step of first forecasting the unemployment rate using available information and then translating that result into a GDP forecast via Okun's law. This is essentially the opposite of the process underlying GB/TB forecasts, where the focus is on forecasting the components of GDP and then using the implied GDP forecast to inform unemployment rate forecast. The same is true even for model-driven forecasts such as those from the Federal Reserve's FRB/US. Forecasting GDP growth first works well for the GB/TB because the Federal Reserve staff employs significant resources to obtain an accurate assessment of near- and medium-term demand, which in turn helps inform their unemployment rate forecasts. However, with a limited information set, forecasting the unemployment rate first then backing out a GDP forecast using that prediction yields more accurate forecasts.

4.2 Results from Estimating a Time-Varying Intercept

Our baseline results are based on estimating a constant intercept δ_i in (1), which depends on the average of $\mathbf{X}_{t-1}$ over the forecast estimation window. Changes in the equilibrium value for some variables in $\mathbf{X}_{t-1}$, such as the unemployment rate and the federal funds rate, are a concern in real time and would invalidate the notion of a constant intercept. Therefore, we consider if and how the accuracy of forecasts generated from (1) changes when we allow for a time-varying intercept $\delta_{i,t}$. (We discuss the criterion for estimating the variance of the innovations in $\delta_{i,t}$ from one quarter to the next in the appendix.)

Table 3 reports the results of this exercise, in which we compare the RMSE from forecasts generated by our framework estimated with a time-varying intercept with the RMSE from our baseline estimates with a constant intercept. The results show that allowing for a time-varying intercept that follows the law of motion in (2) does not improve the performance of our simple forecasting framework. Indeed, the accuracy of the forecasts with a time-varying intercept tends to deteriorate as the forecast horizon expands. One possible reason for this result, which we discuss more in the appendix, is small-sample bias. The bias works in the direction of identifying a time-varying intercept when the estimation window is relatively short. However, as the size of the estimation window increases, evidence of time variation diminishes, reversing the results based on shorter estimation windows. This issue is more severe at longer forecast horizons due to fewer independent observations in the estimation equation, which increases the potential for small-sample bias.

Still, these findings do not necessarily imply that changes in equilibrium values for the unemployment rate or the federal funds rate are immaterial for forecasting purposes. Indeed, our findings are specific to the way we introduce time variation, and they also point to well-known difficulties in assessing potential changes in equilibrium values of macroeconomic variables in real time. The results further serve as a reminder that while our simple forecasting framework compares favorably to the set of forecasts we consider, it still benefits from look-ahead bias. That is, it would be natural to consider a time-varying intercept when forecasting with our framework in real time even if, ex post, our simple approach to addressing the potential for changing equilibrium values for some of the variables in $\mathbf{X}_{t-1}$ is not particularly successful.

Indeed, time variation in the equilibrium values of the unemployment rate and the federal funds rate that is not adequately captured by the simple random walk in (2) could help improve our forecasts. In particular, Table 4 reports the RMSE of forecasts in which the unemployment rate and the federal funds rate in $\mathbf{X}_{t-1}$ are expressed as deviations from their respective ex post estimated time-varying equilibrium values relative to the RMSE of our baseline forecasts.[26] The results in the table highlight that having this ex post information on equilibrium values in real time would have led to more accurate forecasts, though the difference between the two outcomes is not statistically significant in most instances. Nevertheless, the results tell a very different story from the ones in Table 3 and highlight how the accuracy of forecasts generated by our simple framework using a time-varying intercept depends on how the time variation is modeled.

4.3 Robustness

We consider the robustness of our findings along two dimensions. The first concerns the choice of the smoothing parameter to estimate trend GDP in real time using a HP filter. Our baseline estimates rely on a rigid trend due to the filter's well-known issues with endpoints, which are especially relevant in a forecasting setting. Table 5 reports results when we employ less rigid trends for the filter, that is, a smoothing parameter (λ) equal to 1,600 or 6,400, or an even more rigid one (16,000) relative to the baseline (10,000). The table shows that a smoothing parameter of 1,600, which other studies in the

[26] We use the Congressional Budget Office (CBO) estimate of the natural rate of unemployment for the equilibrium value of the unemployment rate and the Laubach and Williams (2003) natural rate of interest for the equilibrium interest rate. (For both series, we use the most recent vintage available at the time of writing.) More specifically, the nominal federal funds rate is expressed as a deviation from its equilibrium value given by the sum of Laubach and Williams' estimate of the real rate and long-run inflation expectations. The unemployment rate is expressed as a deviation from the CBO estimate of the natural rate. In addition, consumer sentiment and the GZ spread are expressed as deviations from a fixed estimated equilibrium value obtained from regressing each variable on a constant, the unemployment rate gap, and the federal funds rate gap. The estimated equilibrium value for each series is the constant from the respective regression. The sample period for the estimation is 1984:Q1 through 2019:Q4 but excludes 2008 through 2013 to avoid biases from influential observations during the Great Recession.

literature commonly use to isolate business-cycle frequencies in quarterly data, results in a deterioration of the GDP forecasts.[27] However, as the value for the smoothing parameter increases, differences across forecasts relative to our baseline results become negligible.

The deterioration in forecasting performance when trend GDP is computed with the conventional smoothing parameter of 1,600 appears related to the aforementioned endpoint issues, as this more flexible trend tends to bias estimates toward actual realizations of the data at the beginning and the end of the sample horizon.[28] Indeed, when we estimate trend GDP using our full sample, the difference between the forecasts using a smoothing parameter of 1,600 and those using 10,000 is negligible at all horizons and, if anything, slightly more precise with the less rigid smoothing parameter.[29] Overall, while a larger smoothing parameter may slow the detection of a change in trend GDP (and thus trend GDP growth) in real time, such a cost – at least over the sample horizon we consider – appears to be outweighed by the benefit of having a trend estimate that is less sensitive to endpoint bias.

We also check the robustness of our results to our measure of interest rates, which in our baseline estimates is given by the deviation of the nominal federal funds rate from an estimate of its equilibrium value. Alternatively, one could use a measure of the real federal funds rate based on the difference between the nominal rate and four-quarter realized inflation, where realized inflation proxies for short-term inflation expectations. Such an approach yields forecasts that are very close to our baseline results, as shown in panel A of Table 6. [30, 31] Indeed, the results suggest that using realized four-quarter core CPI

[27] We report results only over our full forecasting sample because the findings do not vary substantially across subsamples.
[28] See, for example, Mise et al. (2005) for a discussion on the suboptimality of the HP filter at time-series endpoints.
[29] Results are available upon request.
[30] To capture realized inflation, we use the four-quarter change in core CPI inflation. Note that CPI inflation is not revised over time beyond a seasonal-factors update. For this reason, we use the most recent data vintage in this forecasting exercise, as this series is a close approximation of the data available in real time. In addition, the four-quarter inflation measure, by construction, should not be affected by revisions to the seasonal factors.
[31] For simplicity, the table shows only results at the forecast-relevant four-quarter-ahead horizon.

inflation rather than long-term inflation expectations to determine the real federal funds rate could yield slightly more accurate forecasts.

Despite this finding, we use long-term inflation expectations in our baseline specification because if we consider a weighted average of the two inflation expectations measures (long-term and realized inflation), the estimates place a noticeably larger weight on long-term expectations. For example, at the four-quarter-ahead forecast horizon, the estimated weight on long-term inflation expectations is always greater than 0.7 across different estimation windows. We therefore believe it is more realistic to mimic a real-time forecasting exercise using long-term expectations. It is also worth noting that the forecast improvement from using short-run realized inflation as a proxy for inflation expectations should be viewed with caution, as it is concentrated during the Great Recession. Outside that period, forecasts using our baseline specification tend to be more accurate.[32]

In addition, our results do not depend on whether we include a short-term versus long-term interest rate in our information set. As an example, panel B in Table 6 shows results comparing forecasts using the real 10-year Treasury yield.[33] The RMSE at different horizons using the 10-year Treasury yield is similar to those from our baseline forecasts, with a slightly more accurate unemployment rate forecast – though the improvement is not statistically significant.

In sum, we believe that our framework's use of parsimonious and well-defined fundamentals for forecasting economic activity is one of its strengths. Our goal is to demonstrate that our choice of variables provides a useful starting point for forecasting with a small information set, though one could certainly consider alternative variables that capture the same fundamentals and yield potentially more accurate forecasts.

[32] Results are available upon request.
[33] In both specifications, we subtract long-run inflation expectations from the nominal interest rate variable.

4.4 A Post-Pandemic Exercise

As a final exercise, we briefly illustrate how our simple forecasting framework fares for the period following the height of the COVID-19 pandemic, and we highlight some of the challenges related to its usefulness going forward.[34] One such challenge is how to best capture consumer sentiment following the pandemic. Indeed, recently the University of Michigan Index of Consumer Sentiment has decoupled noticeably from macroeconomic fundamentals. The reasons for this shift are not entirely obvious and to some extent appear to predate the onset of the pandemic. The signal from the University of Michigan's sentiment index also seems to have become less informative about the macroeconomy than the signal from the Conference Board's Consumer Confidence Index. Figure 4 plots the average quarterly value for both indices from 1984 through 2024. While the two measures tracked each other relatively well over most of the sample, a large gap opened following the pandemic. Indeed, according to the Michigan survey, consumer sentiment at the end of our sample period (2024:Q4) was at about the same level as it was in late 2009 and early 2010, when the unemployment rate was close to 10 percent (compared with 4.1 percent in 2024:Q4). In comparison, consumer attitudes as measured by the Conference Board index were notably higher than their 2009 and 2010 levels.

To account for the fact that the Michigan consumer sentiment measure appears to have recently lost its information content about macroeconomic developments, we use a predicted measure of consumer sentiment in our forecasting framework for the post-pandemic period. This measure is built off the Conference Board index, four-quarter total CPI inflation, the four-quarter change in the Employment Cost Index (ECI), and the two-quarter change in the Wilshire 5000 (a broad stock market index) and in housing prices (as measured by the Federal Housing Finance Agency All-Transactions House Price Index). These

[34] We do not consider forecasts for the earlier stages of the pandemic because a simple framework based on historical relationships cannot be expected to perform well during such a unique economic environment.

variables explain roughly 90 percent of the variation in the Michigan consumer sentiment index from 1984:Q1 through 2015:Q4. Importantly, each variable's contribution is statistically significant at standard confidence levels, implying that the Michigan measure responds to fundamentals somewhat differently compared with the way the Conference Board measure responds – even though the two indices are highly correlated. The dashed line in Figure 5 plots the in-sample and out-of-sample (post-2015) fitted values from this prediction exercise, and the sold line plots the actual value of the Michigan sentiment index. The figure shows a persistent divergence between the actual and predicted series that began shortly before the onset of the pandemic and became more pronounced after 2020.

Figure 6 shows actual and predicted values for the four-quarter change in the unemployment rate and four-quarter GDP growth from 2021:Q2 through 2024:Q4 when we use this proxy for sentiment in our forecasting framework.[35] The first forecast uses information up to 2020:Q3 – one quarter after the start of the pandemic. The figure shows that, despite our framework's reliance on a limited information set, its predictions are not consistently less accurate than the SPF and BC forecasts over this period.[36] Indeed, the RMSE of our benchmark forecast is below the RMSE of the SPF and BC forecasts over this sample period, even with the omission of the 2021:Q2–2021:Q4 period, when the quadratic loss from the forecast miss tends to be especially large and potentially biases the RMSE computation.

Overall, the results in this section highlight that the performance of our simple framework remains comparable with other forecasts following the pandemic, despite its reliance on a limited information set. This, however, requires proxying for consumer sentiment using some additional variables. While we rely

[35] Using this proxy measure for sentiment (from 1984 onward and actual values of the Michigan index before 1984) would lead to forecasts that are very similar to the ones obtained with actual sentiment in the information set and no loss of accuracy over the pre-pandemic forecast horizons we already considered.
[36] Since GB/TB forecasts are made public with a five-year lag, this comparison includes only SPF and BC forecasts. For this forecasting exercise, our estimation sample runs from 1966:Q1 through 2019:Q4.

on a small set of household fundamentals to predict sentiment, we nevertheless need a wider information set for our model overall – something that our forecasting approach was intentionally designed to avoid.

5. Conclusions

Our analysis shows how a parsimonious set of variables that captures important macroeconomic dynamics can generate forecasts for the unemployment rate and GDP growth that, over relatively long periods, are as accurate as private sector and Federal Reserve forecasts. However, such an outcome requires imposing an Okun's law relationship on the joint behavior of the unemployment rate and GDP – a restriction that could also prove relevant for improving forecasting performance in other settings. Future research could explore the extent to which alternative measures for each of the four fundamentals we consider in our small information set – the labor market, the sentiment/attitude of consumers, the interest rate environment, and financial-risk appetite – yield more accurate forecasts. Factor-based measures that summarize information along these four dimensions could be a particularly promising avenue to consider, especially recently, given the decoupling of movements in the University of Michigan's Index of Consumer Sentiment from those fundamentals.

The simplicity of our setup notwithstanding, several ex post choices are consequential for forecast accuracy. Using the GZ spread as a measure of financial risk appetite yields more accurate forecasts than do other measures, such as a more traditional corporate bond spread, as highlighted in Gilchrist and Zakrajšek (2012). However, as we note earlier, the GZ spread could be viewed as a "new" series, since it is not available in real time for most of the sample period we consider. The choice of a rigid versus more flexible trend for determining potential GDP also matters, and it is always difficult in real time to assess the extent to which a recent change in GDP reflects a change in potential output that might affect the evolution of GDP over the forecast horizon. In our pseudo-out-of-sample forecasting exercise, we resolve ex post some of the tradeoffs a forecaster faces in real time by taking a conservative view of changes in trend GDP. However, this may not be the best approach in all circumstances.

Similarly, equilibrium values for other variables, such as the federal funds rate and the unemployment rate, can change and are difficult to capture in real time. The fact that using sample averages to proxy for equilibrium values of the unemployment rate and the equilibrium federal funds rate produces more accurate forecasts than our simple approach to modeling time variation in these equilibrium values is an ex post result that hardly could have been known ex ante and may not necessarily hold going forward. Considering better ways to accommodate potential time variation in these equilibrium values could be a worthwhile extension of our approach. Still, while there are relevant areas to explore further, our forecasting exercise using a simple framework with a limited information set provides a parsimonious approach to generating forecasts for key real activity variables that are (at least ex post) reasonably accurate when compared with other forecasts.

Bibliography

Andrews, D. W. K., and Ploberger, W. 1994. "Optimal Tests When a Nuisance Parameter Is Present Only under the Alternative." *Econometrica* 62 (6): 1383–414.

Barnes, M. L., Gumbau-Brisa, F., and Olivei, G.P. 2013. "Do Real-time Okun's Law Errors Predict GDP Data Revisions?" Federal Reserve Bank of Boston Research Department Working Papers No. 13-3.

Bernanke, B. S., and Boivin, J. 2003 "Monetary Policy in a Data-rich Environment." *Journal of Monetary Economics* 50 (3): 525–46.

Brayton, F., and Tinsley, P. A. 1996. "A Guide to FRB/US: A Macroeconomic Model of the United States." *Finance and Economics Discussion Series* 96-42, Board of Governors of the Federal Reserve System.

Carroll, C., Fuhrer, J. C., and Wilcox, D. W. 1994. "Does Consumer Sentiment Forecast Household Spending? If So, Why?" *American Economic Review* 84 (5): 1397–408.

Cascaldi-Garcia, D., Luciani, M., and Modugno, M. 2023. "Lessons from Nowcasting GDP across the World." *International Finance Discussion Papers* 1385, Board of Governors of the Federal Reserve System.

Chang, A. C., and Hanson, T. J. 2016. "The Accuracy of Forecasts Prepared for the Federal Open Market Committee." *Journal of Economics and Business* 83: 23–43.

Clark, T. E., and McCracken, M. W. 2008. "Forecasting with Small Macroeconomic VARs in the Presence of Instabilities." *Finance and Economics Discussion Series*, Board of Governors of the Federal Reserve System.

Clark, T. E., and McCracken, M. W. 2013. "Advances in Forecast Evaluation." In *Handbook of Economic Forecasting*, edited by G. Elliott, C. Granger, and A. Timmermann (eds.). 1(2):1107--1201.

Crump R. K., Eusepi, S., Giannone, D., Qian, E., and Sbordone, A. M. 2025. "A Large Bayesian VAR of the United States Economy." *International Journal of Central Banking* 21 (2): 351–409.

Diebold, F. X. 2015. "Comparing Predictive Accuracy, Twenty Years Later: A Personal Perspective on the Use and Abuse of Diebold-Mariano Tests." *Journal of Business & Economic Statistics* 33 (1): 1–9.

Diebold, F. X., and Mariano, R. S. 1995. "Comparing Predictive Accuracy." *Journal of Business & Economic Statistics* 13 (3): 253–63.

Faust, J., and Wright, J. H. 2009. "Comparing Greenbook and Reduced Form Forecasts Using a Large Realtime Dataset." *Journal of Business & Economic Statistics* 27 (4): 468–79.

Fuhrer, J. C. 1993. "What Role Does Consumer Sentiment Play in the U.S. Macroeconomy?" *New England Economic Review*, Federal Reserve Bank of Boston: 32–44.

Giacomini, R. 2015. “Economic Theory and Forecasting: Lessons from the Literature.” *The Econometrics Journal* 18 (2): C22–C41.

Gilchrist, S., and Zakrajšek, E. 2012. “Credit Spreads and Business Cycle Fluctuations.” *American Economic Review* 102 (4): 1692–720.

Gürkaynak, R. S., Sack, B., and Wright, J. H. 2007. “The U.S. Treasury Yield Curve: 1961 to the Present.” *Journal of Monetary Economics* 54 (8): 2291–304.

Jordà, O., Kouchekinia, N., Merrill, C., and Sekhposyan, T. 2020. “The Fog of Numbers.” *FRBSF Economic Letter*, Federal Reserve Bank of San Francisco 2020 (20): 1–5.

Laubach, T., and Williams, J. C. 2003. “Measuring the Natural Rate of Interest.” *The Review of Economics and Statistics* 85 (4): 1063–70.

Levine, D. 1983. “A Remark on Serial Correlation in Maximum Likelihood.” *Journal of Econometrics* 23 (3): 337–42.

Ludvigson, S. C. 2004. “Consumer Confidence and Consumer Spending.” *Journal of Economic Perspectives* 18 (2): 29–50.

Meese, R. A. and Rogoff, K. 1983. “Empirical Exchange Rate Models of the Seventies : Do They Fit out of Sample?” *Journal of International Economics* 14 (1–2): 3–24.

Mise, E., Kim, T-H., and Newbold, P. 2005. “On Suboptimality of the Hodrick-Prescott Filter at time series endpoints.” *Journal of Macroeconomics* 27 (1): 53–67.

Philippon, T. 2009. “The Bond Market's q.” *Quarterly Journal of Economics* 124 (3): 1011–56.

Quandt, R. E. 1960. “Tests of the Hypothesis That a Linear Regression System Obeys Two Separate Regimes.” *Journal of the American Statistical Association* 55 (290): 324–30.

Reifschneider, D., and Tulip, P. 2019. “Gauging the Uncertainty of the Economic Outlook Using Historical Forecasting Errors: The Federal Reserve’s Approach.” *International Journal of Forecasting* 35 (4): 1564–582.

Shimer, R. 1998. “Why Is the U.S. Unemployment Rate So Much Lower?” *NBER Macroeconomics Annual* 13: 11–74.

Romer, C. D., and Romer, D. H. 2000. “Federal Reserve Information and the Behavior of Interest Rates.” *American Economic Review* 90 (3): 429–57.

Sims, C. A. 2002. “The Role of Models and Probabilities in the Monetary Policy Process.” *Brookings Papers on Economic Activity* 33 (2): 1–62.

Stock, J. H., and Watson, M. W. 1998. “Median Unbiased Estimation of Coefficient Variance in a Time-Varying Parameter Model.” *Journal of the American Statistical Association* 93 (441): 349–58.

Stock, J. H., and Watson, M. W. 2002. “Macroeconomic Forecasting Using Diffusion Indexes.” *Journal of Business & Economic Statistics* 20 (2): 147–162.

Stock, J. H., and Watson, M. W. 2003. “Forecasting Output and Inflation: The Role of Asset Prices.” *Journal of Economic Literature* 41 (3): 788–829.

Stock, J. H., and Watson, M. W. 2017. “Twenty Years of Time Series Econometrics in Ten Pictures.” *Journal of Economic Perspectives* 31 (2): 59–86.

West, K. D. 2006. “Forecast Evaluation.” In *Handbook of Economic Forecasting*, edited by G. Elliott, C. Granger, and A. Timmermann. 1(1): 99–134.

Figure 1

Monthly Timing of Selected GB/TB, SPF, and BC Forecasts within Quarter *t*

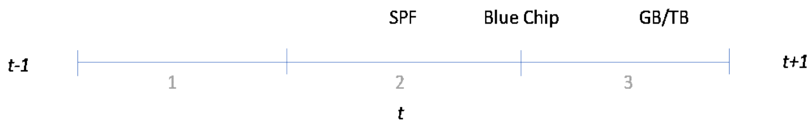
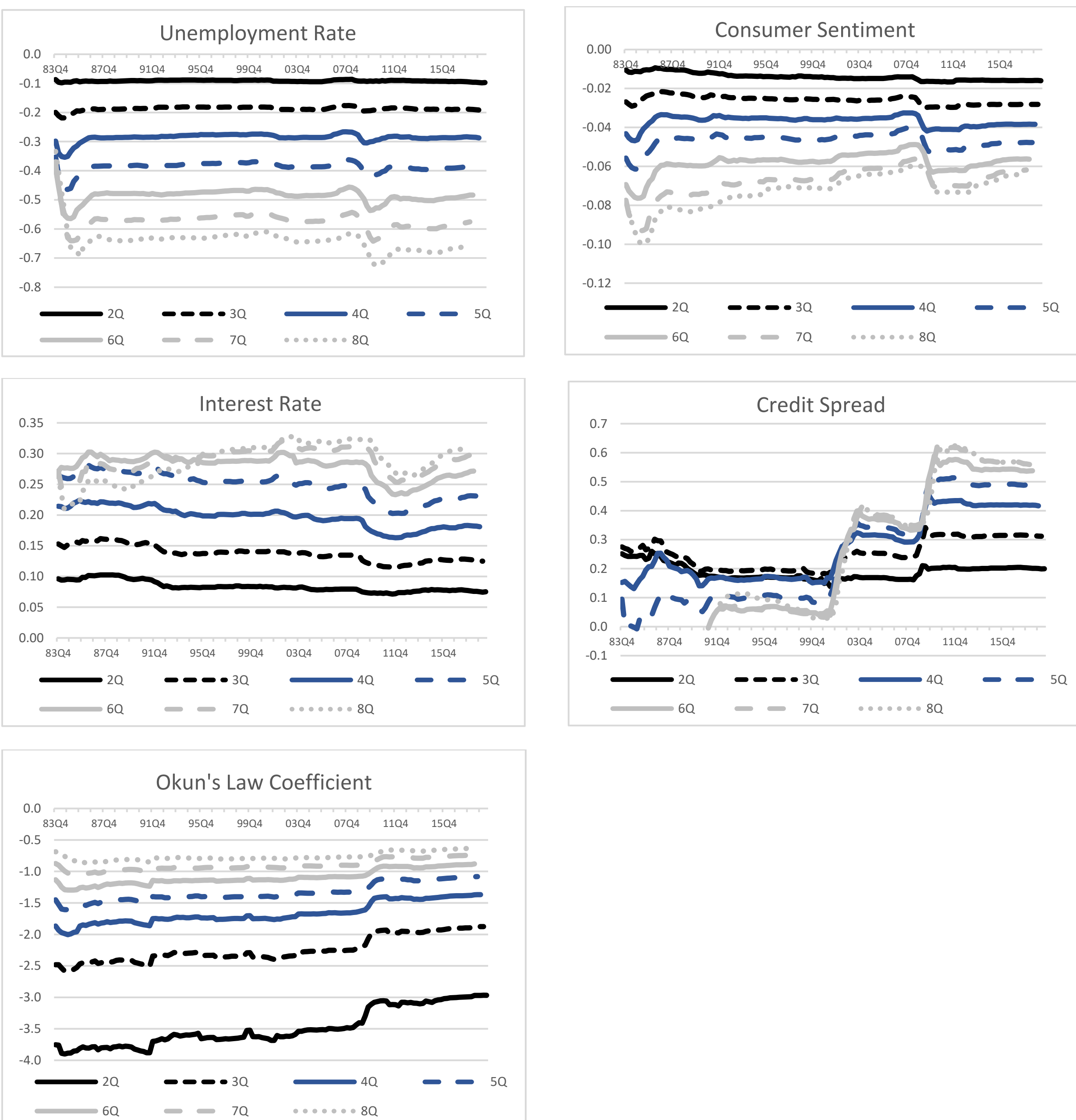


Figure 2

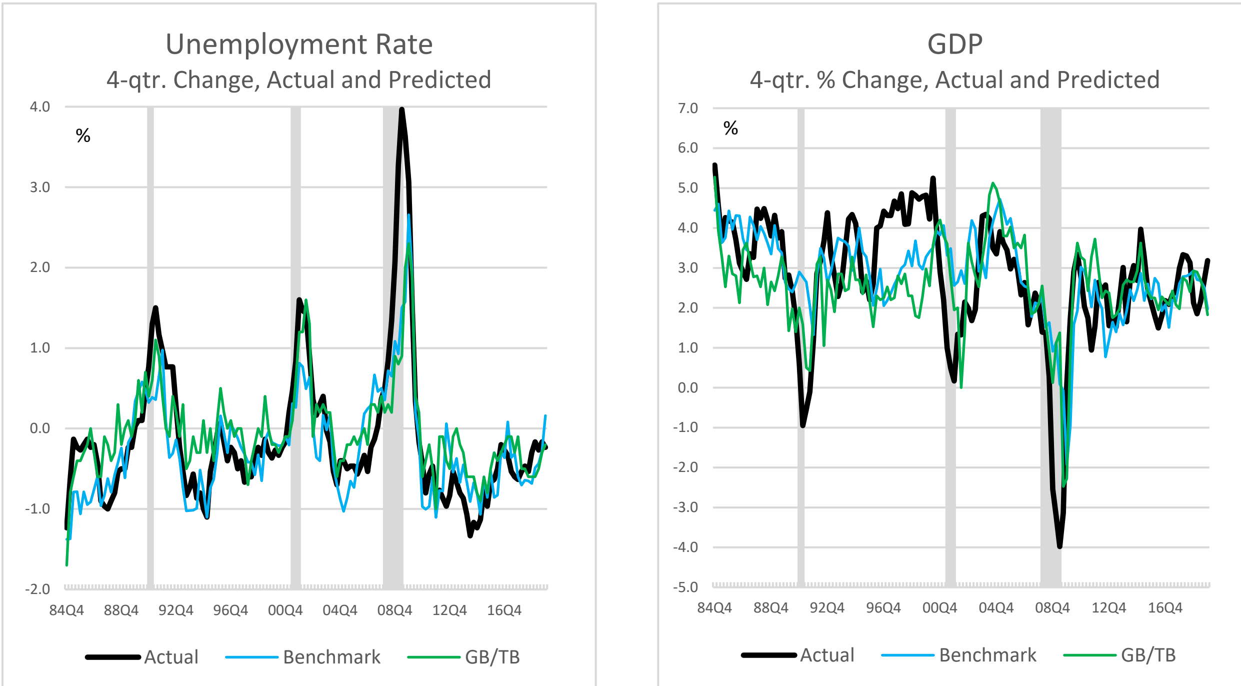
Selected Coefficient Estimates in Equation (1)



Source: Authors' calculations.

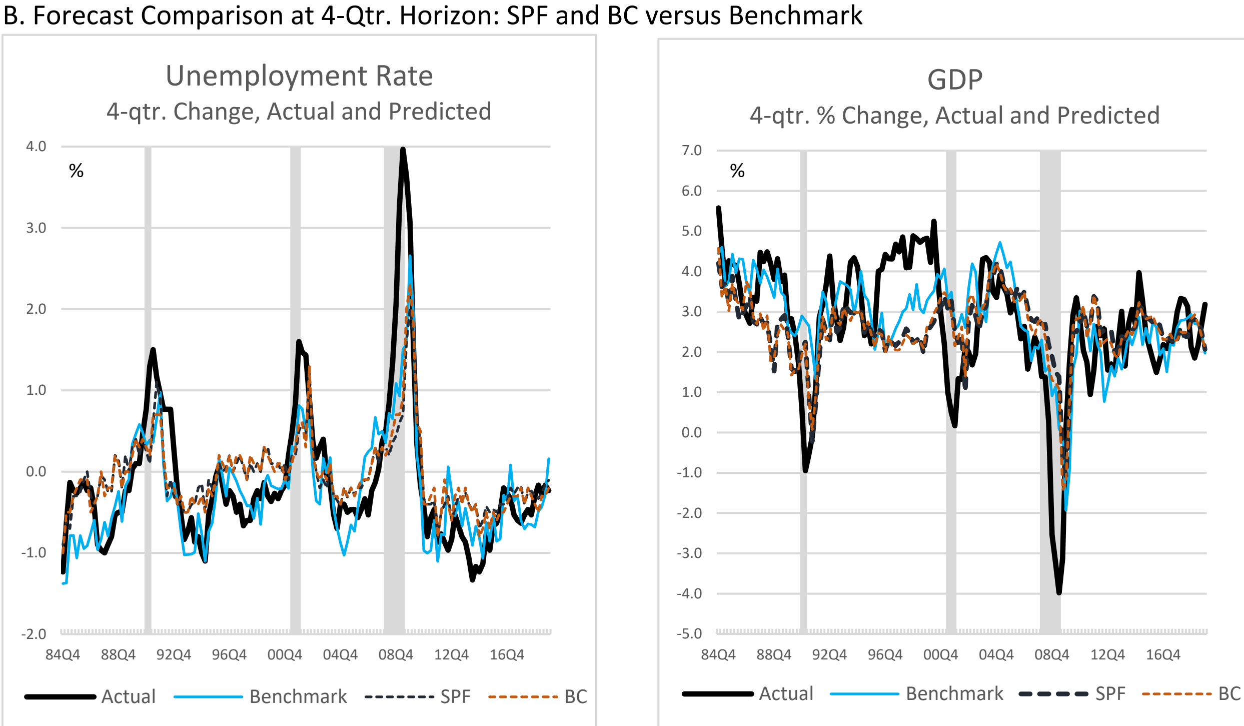
Figure 3

A. Forecast Comparison at 4-qtr. Horizon: GB/TB versus Benchmark



Sources: Bureau of Labor Statistics/Bureau of Economic Analysis/Haver Analytics, authors' calculations.

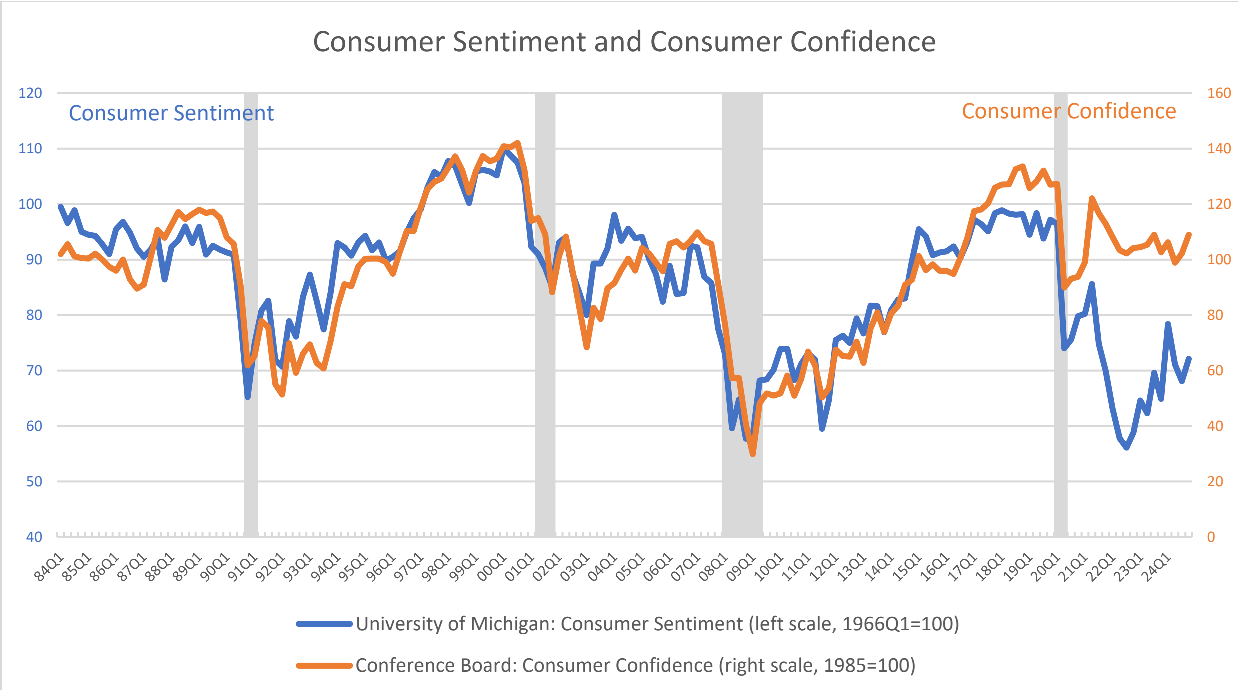
B. Forecast Comparison at 4-Qtr. Horizon: SPF and BC versus Benchmark



Sources: Bureau of Labor Statistics/Bureau of Economic Analysis/Haver Analytics, authors' calculations.

Figure 4

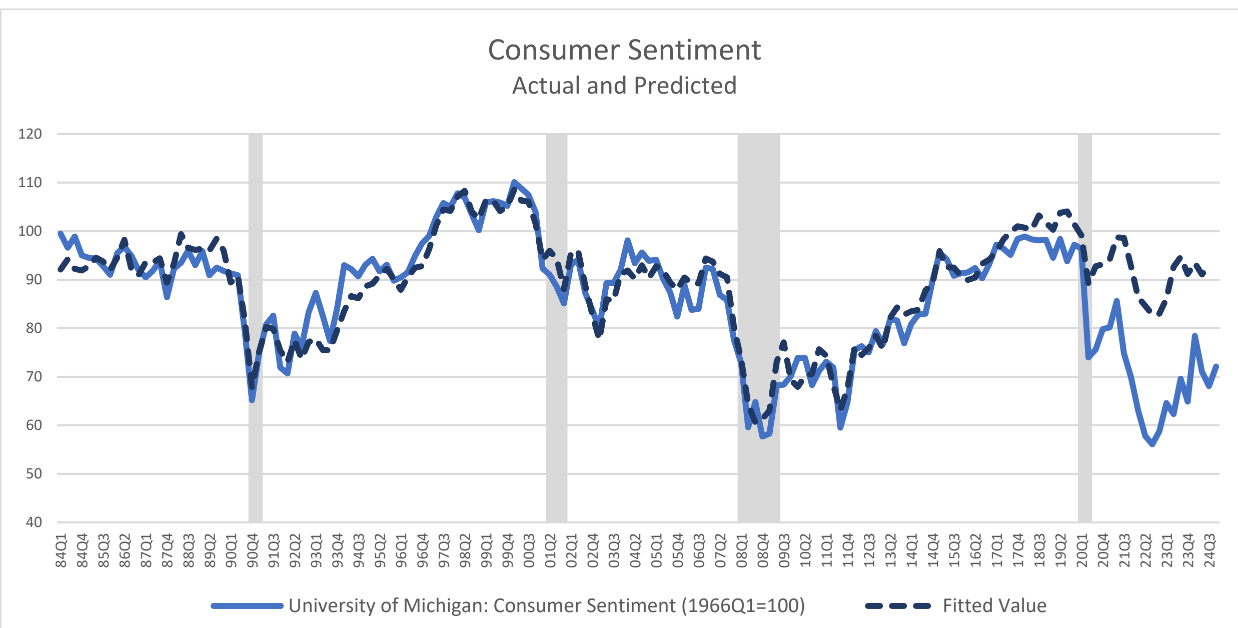
Indices of Consumer Attitudes



Source: University of Michigan/ Conference Board/Haver Analytics, authors' calculations.

Figure 5

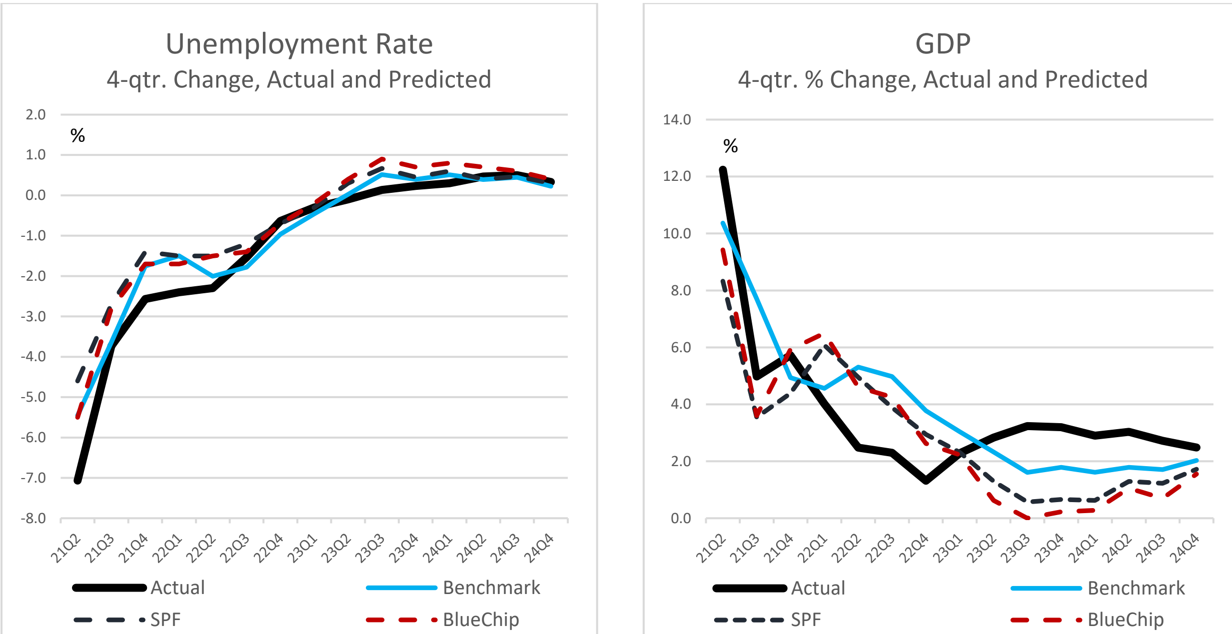
A Proxy Measure for Consumer Sentiment



Sources: University of Michigan/Haver Analytics, authors' calculations.

Figure 6

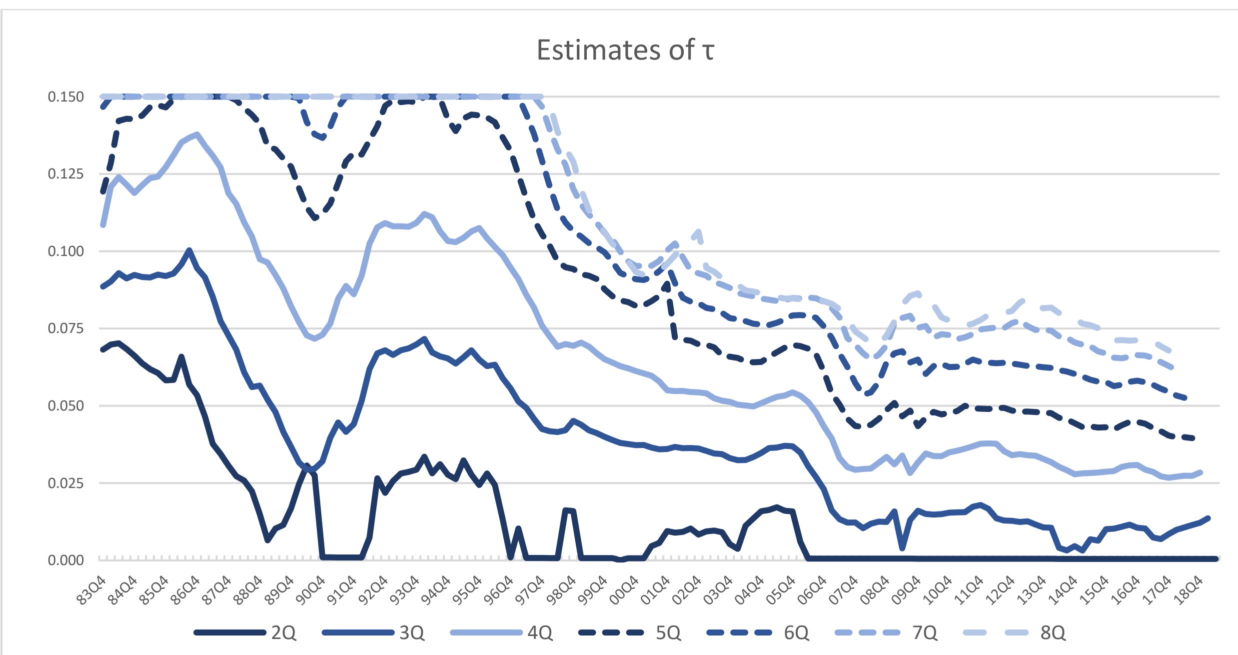
Post-pandemic Forecast Comparison at 4-qtr. Horizon: SPF and Blue Chip versus Benchmark



Sources: Bureau of Labor Statistics/Bureau of Economic Analysis/Haver Analytics, authors' calculations.

Figure A.1

Estimates of $\hat{\tau} = \hat{\sigma}_\eta / \hat{\sigma}_u$ in (1) When Intercept δ Is Time-varying



Source: Authors' calculations.

Table 1

Forecasts' RMSE as a Ratio of Benchmark's RMSE

		A. Unemployment Rate Change						
		Horizon (Quarters)						
		2	3	4	5	6	7	8
1984–2019								
	SPF	1.21*	1.18*	1.16	1.10			
	BC	1.11	1.12	1.10	1.06	1.08	1.08	1.31*
	GB/TB	0.94	1.05	1.03	1.00	1.01	1.02	1.03
1994–2019								
	SPF	1.24	1.24*	1.22*	1.13			
	BC	1.14	1.18**	1.15*	1.07	1.10	1.08	1.32*
	GB/TB	0.99	1.11	1.08	1.02	1.02	1.04	1.05
2004–2019								
	SPF	1.22	1.18	1.17	1.13			
	BC	1.09	1.10	1.09	1.06	1.12	1.10	1.38*
	GB/TB	0.97	1.08	1.08	1.06	1.07	1.08	1.08

		B. GDP Growth						
		Horizon (Quarters)						
		2	3	4	5	6	7	8
1984–2019								
	SPF	1.01	1.06	1.07	1.06			
	BC	0.97	1.03	1.05	1.04	1.04	1.01	1.03
	GB/TB	0.90	0.99	1.03	1.03	1.01	1.01	1.01
1994–2019								
	SPF	1.02	1.07	1.08	1.08			
	BC	0.96	1.02	1.04	1.04	1.06	1.05	1.13
	GB/TB	0.90	0.99	1.03	1.04	1.05	1.04	1.09
2004–2019								
	SPF	0.94	0.99	1.07	1.14			
	BC	0.86	0.91	0.98	1.03	1.10	1.08	1.28
	GB/TB	0.84	0.91	0.97	1.00	1.03	0.93	1.07

Note: * and ** indicate that a Diebold Mariano test of the hypothesis that the two forecasts being compared are equally accurate is rejected at the 90 and 95 percent confidence level, respectively.

Table 2

OLS-based versus FIML-based Benchmark Forecasts

	A. X Forecasts the Change in the Unemployment Rate						
	Horizon (Quarters)						
1984–2019	2	3	4	5	6	7	8
Unemployment rate change	0.99	1.00	1.01	1.02	1.04*	1.04	1.04
GDP growth	1.00	1.00	1.00	1.00	1.01	1.01	1.00

	B. X Forecasts GDP Growth						
	Horizon (Quarters)						
1984–2019	2	3	4	5	6	7	8
Unemployment rate change	1.45	1.42	1.36*	1.30*	1.26*	1.20	1.19
GDP growth	1.11**	1.14**	1.18**	1.21**	1.19**	1.15**	1.12*

Note: * and ** indicate that a Diebold Mariano test of the hypothesis that the two forecasts being compared are equally accurate is rejected at the 90 and 95 percent confidence level, respectively.

Table 3

Time-varying Intercept Forecasts versus Constant Intercept Benchmark Forecasts

	A. Unemployment Rate Change						
	Horizon (Quarters)						
	2	3	4	5	6	7	8
1984–2019	1.01**	1.03*	1.06**	1.09**	1.11**	1.13**	1.16**
1994–2019	1.01*	1.03**	1.08**	1.12**	1.14**	1.16**	1.17**
2004–2019	1.00	1.01	1.04*	1.08**	1.12**	1.15**	1.18**

	B. GDP Growth						
	Horizon (Quarters)						
	2	3	4	5	6	7	8
1984–2019	1.00	1.02*	1.04*	1.06	1.05	1.06*	1.05*
1994–2019	1.01*	1.02*	1.05*	1.08*	1.08*	1.08**	1.07**
2004–2019	1.00	1.00	0.99	0.99	1.00	1.03	1.06

Note: * and ** indicate that a Diebold Mariano test of the hypothesis that the two forecasts being compared are equally accurate is rejected at the 90 and 95 percent confidence level, respectively.

Table 4

RMSE of Forecasts Using the CBO Estimate of the Equilibrium Unemployment Rate and Laubach and Williams (2003) Estimate of the Equilibrium Real Rate of Interest Relative to RMSE of Benchmark Forecasts with Constant Intercept

	A. Unemployment Rate Change						
	Horizon (Quarters)						
	2	3	4	5	6	7	8
1984–2019	0.99	0.98	0.95	0.94	0.91	0.90*	0.90*
1994–2019	1.00	0.98	0.96	0.95	0.94	0.93	0.94
2004–2019	1.00	0.99	0.98	0.99	0.99	0.99	0.99

	B. GDP Growth						
	Horizon (Quarters)						
	2	3	4	5	6	7	8
1984–2019	0.98*	0.97	0.96	0.94	0.93	0.93	0.93
1994–2019	0.99*	0.97	0.96	0.95	0.95	0.95	0.96
2004–2019	0.98*	0.97*	0.95*	0.95	0.96	0.97	0.98

Note: * indicates that a Diebold Mariano test of the hypothesis that the two forecasts being compared are equally accurate is rejected at the 90 percent confidence level.

Table 5

Forecast Comparison Using Alternative Values of the Smoothing Parameter λ in the Computation of HP-filter for GDP (λ = 10,000 in Benchmark)

	A. Unemployment Rate Change						
	Horizon (Quarters)						
Sample: 1984–2019	2	3	4	5	6	7	8
λ = 1,600	0.99	0.98	0.97	0.97*	0.98	0.99	1.00
λ = 6,400	1.00	1.00	0.99*	1.00	0.99	1.00	1.00
λ = 16,000	1.00	1.00	1.00	1.01	1.00	1.00	1.00

	B. GDP Growth						
	Horizon (Quarters)						
Sample: 1984–2019	2	3	4	5	6	7	8
λ = 1,600	1.05	1.09**	1.10**	1.10**	1.11**	1.11**	1.12**
λ = 6,400	1.01	1.01	1.01	1.01	1.01	1.02	1.02
λ = 16,000	1.00	0.99	0.99	0.99	0.99	0.99	0.99

Note: Forecasts are expressed in terms of their RMSE relative to the RMSE of the benchmark specification. * and ** indicate that a Diebold Mariano test of the hypothesis that the two forecasts being compared are equally accurate is rejected at the 90 and 95 percent confidence level, respectively.

Table 6

A. RMSE of 4-quarter Forecasts with Alternative Measure of Monetary Policy Stance Relative to RMSE of 4-quarter Benchmark Forecasts: Short-term Inflation Expectation Proxy

	Unemployment Rate Change	GDP Growth
1984–2019	0.99	0.99
1994–2019	0.98	0.98
2004–2019	0.92	0.92

B. RMSE of 4-quarter Forecasts with Alternative Measure of Monetary Policy Stance Relative to RMSE of 4-quarter Benchmark Forecasts: 10-year Treasury Yield

	Unemployment Rate Change	GDP Growth
1984–2019	0.97	1.03
1994–2019	0.98	0.96
2004–2019	0.94	0.95

Appendix

Estimation with a Time-Varying Intercept

We outline our estimation procedure when we assume the intercept is time-varying and follows a random walk. The intercept evolves following equation (2) in the main text, and we estimate the resulting system of (1) and (2) using the maximum likelihood estimation method implemented by a Kalman filter. However, this approach yields estimates for the standard deviation (σ_η) of the innovation $\eta_{i,t}$ in (2) that are biased toward zero.

Stock and Watson (1998) address this issue by exploiting the fact that the realization of a structural break test can be used to infer the variance of a time-varying parameter such as the intercept δ_t in (2). In particular, they provide a look-up table for the construction of the median unbiased estimator of $\lambda \equiv T\dfrac{\sigma_\eta}{\sigma_u}$, where T is the number of observations in the estimation window, from the realization of a structural break test (Table 3 on page 354). As that table shows, different structural break tests give rise to different median unbiased estimates of λ. We consider three such tests based on the Quandt (1960) likelihood ratio statistic and the Andrews and Ploberger (1994) mean and exponential Wald statistics. The structural break tests are performed on the unemployment rate equation in (1). We cap the implied maximum value for the ratio of the estimated standard deviations $\dfrac{\hat{\sigma}_\eta}{\hat{\sigma}_u}$ at 0.15. Provided this constraint is not binding, we choose the minimum estimated value of λ generated by these three statistics, meaning that we consider the lowest estimate of $\dfrac{\hat{\sigma}_\eta}{\hat{\sigma}_u}$. Simply put, we take a conservative assessment of the amount of time variation in the intercept δ_t. With a given $\hat{\lambda}$, all the parameters in (1) can then be estimated using the maximum likelihood method.

As discussed in the text in Section 4.2, a time-varying intercept does not improve on our baseline forecasts that assume a constant intercept. One explanation for a lack of improved accuracy with a time-varying intercept can be gleaned from the estimates of $\hat{\tau} \equiv \hat{\lambda} / T = \hat{\sigma}_{\eta} / \hat{\sigma}_{u}$, which determine the standard deviation of the innovation of the time-varying intercept relative to the standard deviation of the error in the unemployment rate equation. Figure A.1 shows these estimates for the different sample periods and forecasting horizons over which we estimate the system of equations determined by (1) and (2). In the figure, $\hat{\tau}$ declines across all estimated forecast horizons because more data are included in the estimation window. It is also apparent that the role for a time-varying intercept is smaller at shorter forecast horizons, where the number of independent observations over which the system (1) – (2) is estimated is larger. One possible interpretation of these findings is that evidence for structural breaks over shorter estimation periods is impacted by small sample bias. As more data become available, there is less evidence of breaks and, as a result, less of a role for a time-varying intercept.