The Reliability of the Nominal GDP Expectations Gap*

Andrew B. Martinez,[a,b] Alexander D. Schibuola,[b] and David Beckworth[c]
[a]Department of Economics, American University
[b]Office of Macroeconomic Analysis, U.S. Department of the Treasury
[c]Mercatus Center, George Mason University

Arguments for nominal income targeting are often dismissed because it is an unreliable measure. To assess these concerns, we compare the real-time performance of several nominal and real measures of economic slack. We find that the nominal GDP expectations gap—the difference between nominal GDP and average projections thereof from surveys of professional forecasters—performs well as a measure of economic slack: its historical revisions are two to three times smaller than other measures, it significantly improves real-time forecasts of inflation since the pandemic, and it makes monetary policy rules up to 40 percent less volatile. Overall, concerns about nominal income targets are misplaced.

JEL Codes: C53, E32, E37, E47.

1. Introduction

Many economists have argued that a nominal income target can produce better economic outcomes than other monetary policy frameworks; see, for example, Hall and Mankiw (1994), Frankel and Chinn (1995), Orphanides (2003), Sheedy (2014), Bullard and Singh (2020), and Beckworth and Hendrickson (2020). However, there are also

*The authors have no relevant financial or nonfinancial interests to disclose. The views expressed here are those of the authors and do not necessarily represent those of the U.S. Department of the Treasury or of the U.S. government. Thanks to IJCB managing editor Chris Waller and also to Tara Sinclair, Mike Kiley, Yunjong Eo, Simon van Norden, and two anonymous referees for helpful comments and discussions. Author emails: amartinez@american.edu (corresponding author); Alexander.Schibuola@treasury.gov; Dbeckworth@mercatus.gmu.edu.

critiques of a nominal GDP (NGDP) targeting framework; see, for example, Ball (1999) and Svensson (2020a). One concern is whether NGDP estimates are reliable enough to be targeted. For example, when Federal Reserve officials discussed alternative monetary frameworks during the Federal Open Market Committee (FOMC) meeting in November 2011, then-president of the Federal Reserve Bank of Philadelphia Charles Plosser, voicing the concerns of several FOMC members, argued that under an NGDP targeting framework, if monetary policymakers "misestimate the so-called gap that needs to be closed, we could make serious policy errors, and those would accumulate over time."[1] Thus, these policy errors could have lasting impacts.

Concern about the real-time measurement of policy-relevant gaps is not unique to NGDP. Potential GDP is unobservable and difficult to measure in real time. Estimates of the output gap vary greatly at any given time and are subject to substantial revisions. This has had important implications for predicting inflation and the implementation of monetary policy; see, for example, Orphanides (2001), Orphanides and van Norden (2002), and Orphanides and van Norden (2005). Recent evaluations of the performance of policy-relevant output gaps indicate that while these concerns have diminished, they continue to persist; see, for example, Edge and Rudd (2016), Champagne, Poulin-Bellisle, and Sekkel (2018), Berge (2023), and Furlanetto et al. (2023). This has spurred the search for reliable real-time estimates; see, for example, Kamber, Morley, and Wong (2018), Quast and Wolters (2022), and Barbarino, Berge, and Stella (2024).

Despite the extensive literature on the reliability of output gaps and how to improve them, real-time performance has not been evaluated in nominal terms. This paper assesses the reliability of NGDP gaps and compares their performance with real gaps. We focus on the NGDP expectations gap proposed by Beckworth (2020) and extended by Schibuola and Martinez (2021), which uses historical surveys of professional forecasters' NGDP projections to construct a gap that measures the difference between what NGDP was and what

[1]Excerpted from the transcript of the Meeting of the FOMC on November 1–2, 2011. Available online from the Federal Reserve Board at https://www.federalreserve.gov/monetarypolicy/files/FOMC20111102meeting.pdf (last accessed October 4, 2022).

it was expected to be. This measure differs from typical NGDP rules and is similar to forecast targeting in that it allows the target to grow at a nonconstant rate; see Svensson (2020b). We compare the NGDP expectations gap against the real-time performance of nominal and real measures of the output gap from the Federal Reserve Board (FRB), the Congressional Budget Office (CBO), and forecast-based measures.

The NGDP expectations gap performs well. First, its historical revisions are two to three times smaller than measures produced by the FRB staff or the CBO and are comparable with or slightly better than statistical output gap measures. Second, it significantly improves forecasts of core personal consumption expenditures (PCE) inflation following the COVID-19 pandemic. Finally, the implied interest rate targets from standard Taylor rules are up to 40 percent less volatile in real time when using the NGDP expectations gap than when using output gaps. Overall, this indicates that the concerns associated with the real-time implementation of NGDP targets and the implications for policymaking are misplaced.

The rest of the paper proceeds as follows: Section 2 describes existing forecast-based measures of economic slack. Section 3 presents the NGDP expectations gap. Section 4 compares the real-time properties of the NGDP expectations gap with other nominal and real gap measures. Subsections 4.2 and 4.3 evaluate how well the various gap measures predict future inflation, and interest rate targets. Section 5 concludes.

2. Forecast-Based Measures of Economic Slack

There are many ways to measure economic slack, with differing implications for economic policy. As Kiley (2013) argues, the various approaches can be broadly classified into three definitions of what output has deviated from: (i) an underlying trend, (ii) the implied production-consistent level of output, and (iii) the "natural" level of output. While all three definitions have important considerations, this paper focuses on the first definition.[2] This section sets the stage for describing our methods by briefly reviewing two existing

[2]For example, see Blanchard and Quah (1989), Justiniano and Pimiceri (2008), and Shackleton (2018) for discussions on the implications of the other definitions.

forecast-based output gap measures and discussing the conceptual similarities and differences between them.

2.1 Beveridge-Nelson Decomposition

A forecast-based measure of the output gap can be obtained from the Beveridge and Nelson (1981) decomposition. According to Morley (2002), for some integrated series, y_t, the Beveridge-Nelson (BN) trend is the minimum mean square error forecast of the long-run level of the series excluding the deterministic drift. This can be represented as the contemporaneous level of the series plus the sum of the H-period-ahead forecasts of the first differences of y_t:

$$\mathrm{BN}_t \equiv \lim_{H\to\infty} \mathbb{E}_t\left[y_{t+H} - H\mu\right] = y_t + \lim_{H\to\infty} \sum_{h=1}^{H} \mathbb{E}_t\left[\Delta y_{t+h} - \mu\right], \quad (1)$$

where μ represents the deterministic drift, $\Delta y_t = y_t - y_{t-1}$, and $\mathbb{E}_t[\cdot]$ captures the expected value based on information at time period t. The difference between the BN trend and the data is calculated as the implied forecast error by subtracting y_t from Equation (1) to obtain the BN cyclical component. This implied BN output gap appears in numerous studies; see, for example, Morley (2011) and Berger, Morley, and Wong (2020).

The BN cyclical component, i.e., the second term on the right-hand side of Equation (1), is obtained after assuming which model generates the most accurate long-run forecast. For example, if an autoregressive model with p lags, i.e. an AR(p), produces the most accurate long-run forecasts of y_t, then the implied output gap from the BN decomposition becomes

$$y_t - \mathrm{BN}_t = -\left(\frac{\alpha(L)}{1-\alpha(L)}\right)(\Delta y_t - \mu), \quad (2)$$

where $\alpha(L) = \sum_{j=1}^{p} a_j L^{j-1}$ and where L is the lag operator such that $L^p y_t = y_{t-p}$ and a_j are the autoregressive parameters. Parameters are estimated using Δy_t, even though it may not produce the best forecast for y_t at longer horizons; see, for example, Clements and Hendry (1993, 2001a, 2001b).

BN estimates often indicate cycles with smaller amplitudes than other measures. Kamber, Morley, and Wong (2018) impose restrictions on the parameter estimates to increase the amplitude of BN cycles. They show that increases in the persistence of the estimated autoregressive parameters align BN estimates more closely with policy-relevant measures of the output gap while retaining the reliability of the estimation procedure in real time.

2.2 Hamilton Filter

The Hamilton (2018) filter is a forecast-based measure developed due to concerns about the Hodrick and Prescott (1997) filter.[3] Like the BN decomposition, the Hamilton filter is based on long-run forecasts. Specifically, Hamilton (2018) captures the underlying trend using eight-period-ahead forecasts from an AR(4) model of the level of y_t. Quast and Wolters (2022) modify this approach by averaging over forecasts from multiple horizons to better capture a wider range of business cycle frequencies:

$$MHF_t \equiv y_t - \frac{1}{9}\sum_{h=4}^{12} \hat{\mathbb{E}}_{t-h}\left[y_t\right], \tag{3}$$

where the forecast generated for time period t using an AR(4) model is represented by $\hat{\mathbb{E}}_{t-h}\left[y_t\right]$.

This modified Hamilton filter (MHF) averages over past forecasts ranging between 4 and 12 quarters prior to the period of interest. However, Quast and Wolters (2022) also show that estimates of the trend centered around the 10-quarter-ahead forecast horizon would allow for the inclusion of longer financial/credit cycles in the standard business cycle; see Beaudry, Galizia, and Portier (2020). They also show that wider windows are more closely correlated with National Bureau of Economic Research (NBER) recession definitions and policy-relevant output gaps such as those from the FRB and the CBO.

[3]Phillips and Shi (2021) and others show how to improve the performance of the HP filter.

3. The NGDP Expectations Gap

As noted above, this paper focuses on the deviation from trends approach to measuring economic slack as outlined by Kiley (2013). The BN decomposition and the Hamilton filter employ this approach, but assume parameter constancy and that the data generation process of y_t is well represented by the assumed model. However, the appropriate model at any given time is generally unknown and the large fluctuations in output over the past several decades indicate clear evidence of structural breaks. This section presents the expectations gap as an alternative approach that addresses these challenges.

3.1 Expectations Gap

One way to circumvent the concerns surrounding parameter constancy and model reliability while still capturing the expected trend is to use estimates from a set of models rather than any single procedure.[4] Following Beckworth (2020) and Schibuola and Martinez (2021), we use surveys of professional forecasters to measure deviations from the expected trend.[5] The idea is that forecasters generate projections using the best available methods at the time to represent their expectations about future trends. Moreover, aggregating forecasts reduces the influence of each model and allows for the possibility that the best model for predicting y_t may evolve over time.

We combine professional forecasts of the growth rate at each horizon with the latest observed log-level of y_t to obtain the expected level:[6]

$$\tilde{\mathbb{E}}_{t-h}\left[y_{i,t}\right] = y_{t-h-1} + \sum_{j=0}^{h} ln\left(1 + \tilde{g}_{i,t-h+j\|t-h}\right), \tag{4}$$

[4]Other studies have considered model-based combinations; see, for example, Morley and Piger (2012), Guérin, Maurin, and Mohr (2015), Pichette et al. (2019).

[5]This is related to an approach in Coibion, Gorodnichenko, and Ulate (2018) where the output gap is based on survey forecasts of long-run growth.

[6]In practice we use the level of output. The log-level is used here to facilitate comparisons with existing approaches.

where $\tilde{\mathbb{E}}_t[\cdot]$ is an estimate of the expected value given the information available at time t, $\tilde{g}_{i,t\|t-h}$ is the expected quarterly output growth rate from a given forecaster or group of forecasters, i, generated from the survey at period $t-h$ for all quarters up through period t, and y_{t-h-1} is the last available log-level of output that would have been observable when the forecast was generated.

While we allow y_{t-h-1} to vary with new data vintages, forecasts of the growth rates are fixed across vintages since these are only available in real time. This means that the expectations gap is inherently quasi-real-time. It also circumvents one of the largest sensitivities of output gap measures by limiting how much forecasts change across data vintages; see Orphanides and van Norden (2002).[7]

The measure of expectations in Equation (4) is similar to those measured by the BN trend in Equation (1) or the Hamilton filter in Equation (3).[8] While the BN trend is based on forecasts using information up through time period t, the main difference is the choice of models. For example, if professional forecasters all relied on the same AR(4) model to predict future the expected value would be essentially identical to the trend derived using a modified Hamilton filter.

We construct an average expected trend by averaging estimates of expected output across a range of forecast horizons. Beckworth (2020) refers to this estimate as the "neutral level" of output. The gap between observed output y_t and its average expected value at time period t is the expectations gap

$$EGAP_t \equiv y_t - \frac{1}{20}\sum_{h=0}^{19} \tilde{\mathbb{E}}_{t-h}\left[y_{i,t}\right], \tag{5}$$

[7]Using fixed growth rate forecasts can be interpreted as a restricted version of Quast and Wolters (2022) where the expected growth rates are estimated over shorter samples and older vintages of data. We can analyze how estimates of expected output are revised in real time with changes to the initial level of output by adding additional notation such that the initial level of output for a given vintage of data would be denoted as $y_{t-h-1,v}$ where v specifies the vintage.

[8]Robust prediction methods suggest links across measures; see Martinez, Castle, and Hendry (2022). Showing this is left to future research.

where the 20-quarter range implies that the estimate is centered around 10 quarters ahead, which is longer than the modified Hamilton filter. While the extended range means that some cyclical aspects of the forecasts could enter estimates of the trend at shorter horizons (see Isiklar and Lahiri 2007), an extended range centered around 10 quarters ahead is able to filter out longer financial cycles, consistent with the results in Quast and Wolters (2022).

The choice of forecaster or group of forecasters is important for the expectations gap. While it is possible to construct estimates of the expected trend, as in Equation (5), for individual forecasters, constructing a consensus estimate based on the mean or median forecaster helps robustify against model choice.[9] As the average of all individual forecasters represents the central tendency, it excludes and can be distorted by information that is contained in the full range of forecasters. For example, forecaster disagreement has been linked to the business cycle; see, for example, Patton and Timmermann (2010) and Bürgi and Sinclair (2021). A large dispersion around the central tendency also indicates greater uncertainty about the expected trend; see Lahiri and Sheng (2010). This type of uncertainty can be quantified by constructing Equations (4) and (5) using forecasters that represent a specific range, e.g. by generating a measure for all i where $i \in \{\text{10th percentile}, \text{ median}, \text{ 90th percentile}\}$.[10]

3.2 The Expectations Gap Using Survey Forecasts

The expectations gap uses surveys of professional forecasters in its construction. The two surveys we focus on are the Survey of Professional Forecasters (SPF) and the Blue Chip Economic Indicators (BCEI). The SPF is a quarterly survey of about 40 academic and business forecasters first published in 1968 as the American Statistical Association/National Bureau of Economic Research Outlook Survey. Management of the survey was transferred to the Federal

[9]A long literature shows that pooling forecasts can be more accurate in the context of structural change and model uncertainty; see, for example, Bates and Granger (1969), Hendry and Clements (2004), Timmermann (2006), Genre et al. (2013).

[10]Another way to measure uncertainty is to use a factor stochastic volatility model of real-time estimates; see Berge (2023).

Reserve Bank of Philadelphia in 1990. The BCEI is a monthly survey of more than 50 business economists that began in 1976 and is maintained by the firm Wolters Kluwer.

The SPF is published in the second month of every quarter with quarterly forecasts that extend up to four quarters ahead, annual forecasts that extend out two years, and forecasts of annual average growth rates that extend out 10 years. The BCEI is published around the 10th day of each month with quarterly forecasts that extend up to eight quarters ahead, annual forecasts that extend out two years, and forecasts of annual averages that go to 10 years.

The quarterly forecasts from both surveys directly provide the near-term growth rate forecasts necessary to construct Equation (4). However, since we extend the level forecasts 20 quarters ahead, it is necessary to impose an assumption about forecasters' expectations about longer-term growth rates.

The simplest assumption is that the last available quarterly growth rate forecast is the expected long-run equilibrium growth rate. For the SPF, this implies that the expected growth rates in Equation (4) would be $\tilde{g}_{i,t-h+j\|t-h} = \tilde{g}_{i,t-h+4\|t-h}$ for the $j \geq 4$ forecast horizons. While this assumption is generally innocuous, since forecasters typically do not have much information beyond the one- to two-year forecast horizon, it may be problematic around turning points if forecasters are expecting a recession at the end of the forecast horizon or a slow recovery, which would distort the implied long-run expected growth rate. Thus, we choose this option only when there are no longer-term forecasts available.

Our preferred assumption is that the annual average growth rate forecasts can serve as a proxy for quarterly growth rates at horizons beyond one year ahead and that these expected long-run growth rates prevail immediately after the last available quarterly growth forecast. This allows for the quarterly growth forecasts to have cyclical information but longer-run growth forecasts capture the long-run equilibrium growth rates.

In practice, we use a mix of these assumptions across both surveys. The BCEI updates its long-run GDP growth rate forecasts on a semiannual basis. Thus, it meets the conditions for our preferred assumption and we extend the quarterly growth rate forecasts beyond eight quarters ahead with the last available annual growth rate forecasts. However, the SPF only produces long-run forecasts of

GDP growth and CPI inflation, which it updates at an annual and quarterly frequency respectively. Furthermore, the SPF's long-run forecasts are only available from 1990 onward.

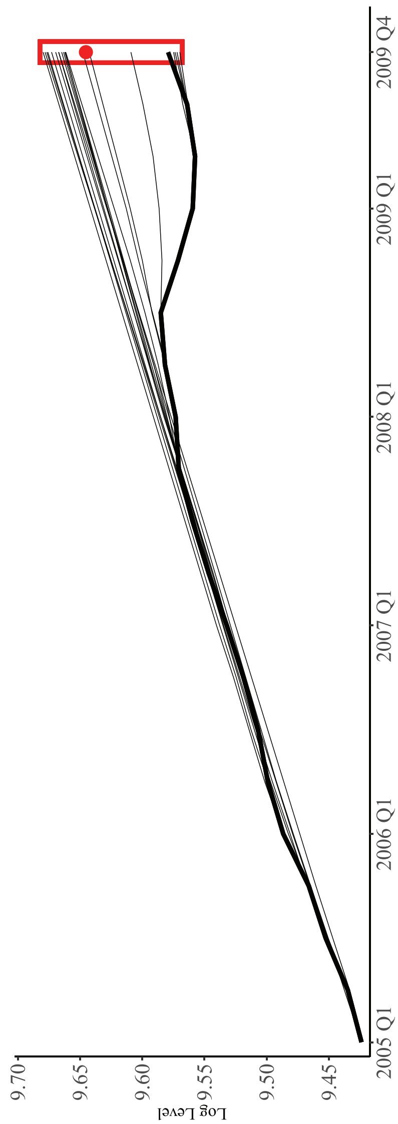
To circumvent these limitations for the SPF, we construct a proxy long-run NGDP growth rate forecast by combining the expected long-run real GDP growth rate forecasts with the long-run CPI inflation forecasts adjusted for the expected gap between CPI and GDP inflation that prevails at shorter forecast horizons and then use that to extend out the quarterly growth rate forecasts beyond four quarters ahead. We also extend the SPF back prior to 1990 using the assumption that the last available quarterly forecast growth is the expected long-run growth rate. These choices ultimately have very little impact from an empirical or practical standpoint since we average across many different forecasts and forecast horizons. This makes estimates of the expectations gap less sensitive to how any given forecast is constructed.

Figure 1 illustrates our approach for generating the SPF's expected level of NGDP for 2009:Q4. We start by constructing a series of log-level NGDP forecasts for 2009:Q4 using quarterly SPF NGDP growth rate forecasts for each quarter since 2005:Q1 (i.e., 19 quarters prior to 2009:Q4). Each thin black line represents an NGDP forecast path constructed according to Equation (4). In total there are 20 thin lines which terminate in 2009:Q4. The rectangular box denotes the range of these forecasts. Next, we take the average of these forecasts for 2009:Q4, as represented by the large red dot. The expectations gap in Equation (5) is then constructed by taking the difference between the dot and the actual value as represented by the thick black line. We repeat this process every quarter and generate a series of expectations gaps.

We can also construct a similar series using the BCEI. However, since the BCEI is published monthly, there is more than one forecast available for each quarterly horizon. To account for this, we average across all available monthly forecasts for the horizons of interest. Thus, since the forecast horizon of interest in Equation (5) extends up to 20 quarters, we average over 60 monthly forecasts rather than 20 quarterly forecasts.

To capture the uncertainty around the expectations gap, we use information on the dispersion across individual forecasts. For the SPF, which publishes individual forecasts, it is possible to calculate

Figure 1. Estimating the SPF's Expected NGDP Level Forecast for 2009:Q4



forecaster percentiles and a range of forecaster disagreement. A cruder proxy of forecaster disagreement is available using the BCEI averages of the top and bottom 10 forecasters, which are available beginning in 1990.

3.3 Estimates

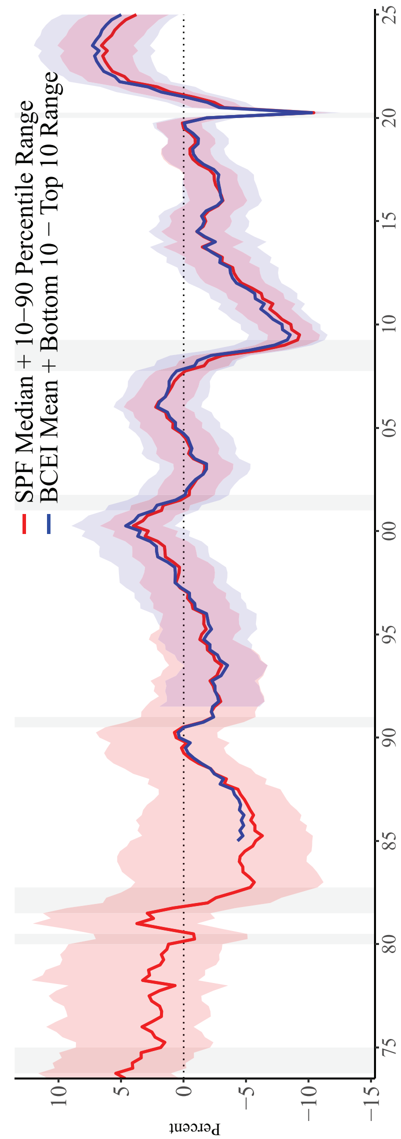
Figure 2 shows the NGDP expectations gaps and their ranges from the SPF and the BCEI. To extend the SPF estimates as far back as possible, we fill in missing long-run growth forecasts with the longest available quarterly growth rate forecast at the time as a proxy for the long-run growth rate up to five years ahead as described above. Doing so allows us to extend estimates for the SPF back to 1974.

Figure 2 shows that, except for the mid-to-late 1980s, the BCEI and SPF measures are very similar and appear to capture important features associated with economic cycles. Except for the 1973–75 recession, the gaps always turn negative during a recession. The gaps also capture the sharp economic downturns surrounding the 1981–82 recession, the 2008–09 Great Recession, and the COVID-19 pandemic-induced recession in 2020. The gaps also indicate episodes of a boom economy around periods associated with the late 1990s dot-com bubble, the housing bubble of the mid-2000s, and the inflationary period of the pandemic recovery in 2021–24.

The measures of forecaster disagreement surrounding the estimates of the expectations gap indicate five episodes coinciding with important economic periods during which the economy clearly performed substantially above or below most individual forecasters' expectations. The first episode occurred during the dot-com era boom; the second during and after the 2008–09 recession, when the economy grew much slower than expected; the third during the so-called invisible recession of 2015–16;[11] and the fourth episode during the recession in 2020 induced by the COVID-19 pandemic. Finally, during the inflationary period of the pandemic recovery in 2021–24 the entire range moved above zero. Each of these episodes corresponds to important economic events and supports

[11]See Irwin (2018). Alternatively, Baumeister, Leiva-Léon, and Sims (2024) refer to this episode as the 2014–16 oil price slump.

Figure 2. NGDP Expectations Gap



Note: Shaded bars denote NBER recession dates.

the interpretation of the NGDP expectations gap as an indicator of changes in the business cycle.

The range of disagreement also conveys how the surveys' designs and how forecasters views of the economy have evolved. In the 1970s and 1980s, there was large forecaster disagreement about the state of the economy, with the range only ever fully going above zero in 1974 but otherwise showing a large spread for most of the 1970s. The widespread disagreement among forecasters about the underlying economic trends persisted through the 1980s. This result is robust even if the post-1990 measures are constructed using the furthest available short-run growth rate forecast as a proxy measure for the long-run growth rates. However, it may be sensitive to changes in how the survey has been conducted since 1990 or the Federal Reserve's adoption of an implicit and then explicit inflation target in the 2000s. Notably, while in 2021–24 economic growth and inflation returned to a period reminiscent of the early 1980s, forecasts do not signal a return to the pre-1990s level of disagreement.

4. Nominal and Real Gaps in Real Time

This section compares and contrasts various real-time estimates of nominal and real gaps, starting with a visual comparison of alternative measures and then a formal examination of the revision properties. Next, we analyze the performance of the various measures in real time by comparing their ability to predict future inflation. Finally, we examine the implied interest rate target from Taylor rules using the various real-time gap estimates. While there are similarities across the measures, the NGDP expectations gap performs well. Its historical revisions are two to three times smaller than standard output gap measures and it significantly improves near-term forecasts of core PCE inflation following the COVID-19 pandemic. Finally, when used inside of a Taylor rule, the NGDP expectations gap generates implied interest rate targets that are up to 40 percent less volatile than those implied from standard output gap measures.

This analysis compares the real-time estimates of the nominal and real expectations gaps with four alternative real-time measures: (i) the output gaps and the implied nominal and real expectations gaps constructed from the FRB staff's forecasts, (ii) the CBO's nominal and real output gaps, (iii) the gaps generated using the modified

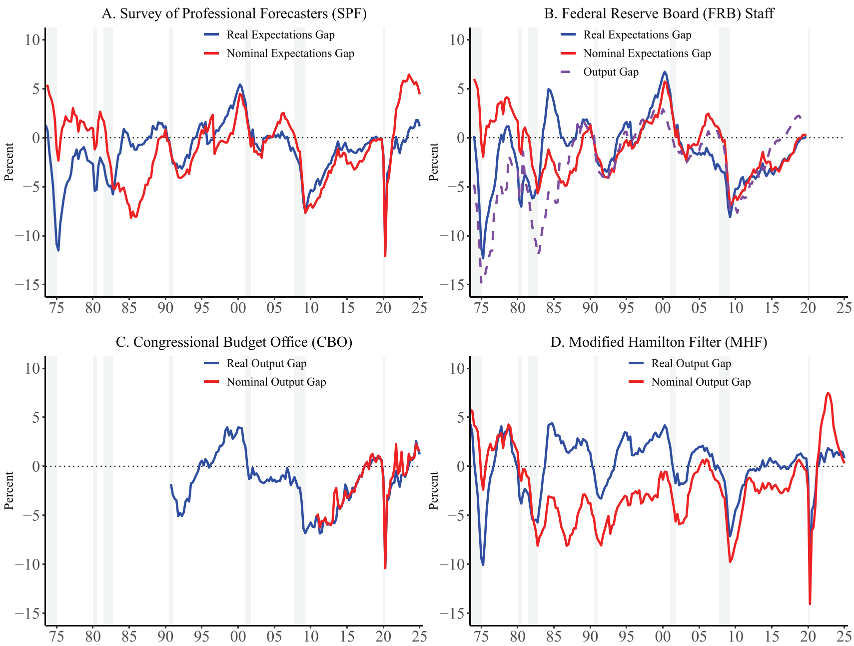
Hamilton filter of Quast and Wolters (2022), and (iv) the gaps generated using the modified BN decomposition of Kamber, Morley, and Wong (2018). All estimates are constructed based on real-time monthly vintages of nominal and real GDP from the Federal Reserve Bank of Philadelphia's Real-Time Dataset for Macroeconomists; see Croushore and Stark (2001). We treat the first month in which a particular quarter's advance estimate of GDP was available as the real-time estimate. This is motivated by the special emphasis given to the advance estimate in economic policy settings. However, the results are robust to using the second or third release of GDP.

Panel A of Figure 3 plots the real-time nominal and real expectations gaps derived from the SPF. Note that the real gap is constructed using the same procedure as the nominal gap above, relying solely on the real GDP forecasts. The real-time nominal gap is directly comparable to the latest estimate in Figure 2 and suggests that the real-time estimates are not substantially different from the latest vintage, as will be shown quantitatively below. For much of the sample, the nominal and real gaps follow each other fairly closely. However, there are several important deviations. First, in the late 1990s the real gap signals overheating more quickly than the nominal gap does. Second, in the mid-2000s, when home prices and construction costs were rising, the nominal gap shows stronger signs of overheating. Finally, the nominal gap indicates more slack in the mid-2010s, during the invisible recession.

Prior to the 1990s there were much larger and prolonged differences between the nominal and real gaps. For much of the stagflationary period of the 1970s, the nominal gap generally remains positive (except around recessions), whereas the real gap is persistently negative until 1985, after which the measures gradually converge. These differences illustrate that the nominal and real measures can convey very different signals about slack during high inflationary periods. This is also the case in 2022, when the real measure turned negative while the nominal measure neared record highs.

Panel B of Figure 3 plots the implied real-time real and nominal expectations gaps constructed from the FRB staff GDP growth forecasts (as computed using our methodology) and the real-time output gap through 2019 (the last publicly available vintage of the staff's

Figure 3. Real-Time Comparison of Nominal and Real Gaps



Note: Shaded bars denote NBER recession dates. Real-time vintages use advance estimates of GDP.

forecasts).[12] The FRB staff's expectations gaps are similar to those derived from the SPF for most of the period, except for the mid-1980s, when there is a larger spike in the FRB real gap, and the late 1990s, when both the FRB's nominal and real gaps indicate overheating. The real-time output gap starts out more closely aligned with the real expectations gap in the 1970s but does not exhibit the same spike as the real expectations gap does in the mid-1980s. By 1990, all three measures move together closely. Interestingly, the output gap indicates less overheating than both expectations gaps in the late 1990s but more overheating than the real expectations

[12]Federal Reserve staff forecasts and output gaps come from the Federal Reserve Bank of Philadelphia's real-time data center. Real-time output gaps are extended using the historical data set from Edge and Rudd (2016).

gap, which indicates additional slack in the mid-2000s. By 2019, the real-time output gap indicates overheating, whereas both expectations gaps are neutral. While there are many periods in which the real-time expectations gaps and output gap provided similar signals, the expectations gaps, especially the nominal variant, often provide evidence of overheating (or the lack thereof) in real time just prior to economic downturns.

Panel C of Figure 3 plots the real-time estimate of the CBO's nominal and real output gaps, which the CBO constructs to be identical on purpose. Differences between the gaps arise in real time since the CBO only updates its estimates of potential on a 6- to 12-month basis.[13] While the CBO's real-time output gap detected overheating during the dot-com boom, it failed to detect any overheating during the mid-2000s, which are commonly associated with a housing bubble. Since 2021 the CBO's nominal gap has become positive, while the real gap turned positive in 2023.

Panel D of Figure 3 plots real-time measures of the real and nominal output gaps using the modified Hamilton filter (MHF) of Quast and Wolters (2022). While the MHF appears to capture the business cycle using real GDP, the NGDP measure is less convincing. The implied real-time real output gap turns negative for all recessions and indicates overheating in the late 1970s, the late 1990s, and the mid-2000s. It also indicates overheating since 2021. On the other hand, the implied nominal output gap never signals slack in the 1970s and then spends most of the 1990s and 2000s in negative territory before surging following the pandemic recession. This suggests that the AR(4) model underlying the MHF is misspecified. Consequently, it does not correctly capture the trend or the dynamics in NGDP in the 1970s and 1980s, which illustrates the risk of using a single misspecified model to estimate the gaps.

Overall, the real-time nominal expectations gap captures the most important features of the business cycle. Our analysis also suggests that there are important differences between the nominal and real expectations gaps. The stagflationary episode in the 1970s, the housing bubble in the mid-2000s, the invisible recession

[13] Historical vintages of the CBO's estimates of real and nominal potential GDP are from the Federal Reserve Bank of St. Louis' FRED and ALFRED databases.

in 2015–16, and more recently the pandemic recovery period suggest that including nominal fluctuations provides a clearer picture of overheating and underperformance associated with episodes of broader price changes which might otherwise be muddled by slow real growth. Conversely, both the real expectations gap and conventional output gaps have missed or underestimated these economic developments.

4.1 Revision Properties

This subsection quantifies the reliability of the various real-time gap estimates.[14] Reliability is measured based on the size of revisions by: the standard deviation (SD) and the root mean square (RMS). These measures are normalized using the standard deviation of the ex post estimate of the gap for each method to facilitate comparability. A "noise-to-signal ratio" (NSR) captures how large the variation of the revision is relative to the true variation of the data. Methods with lower NSRs indicate smaller revisions. We also calculate the frequency with which the real-time estimates have the same sign as the ex post estimate. Methods that have a large share of same-sign estimates are considered to be more reliable.

Two sample periods are considered in order to capture the properties of as many measures as possible. The first sample spans 1975–2018 and uses the November 2019 vintage as the ex post estimates. This allows comparison against the FRB staff estimates, which were available through 2019 at the time of this analysis. The second sample spans 1991–2023 and uses the December 2024 vintages as the ex post estimates for that sample. This enables a direct comparison against the CBO estimates, which are available from 1991.

Table 1 presents the results for the real gaps. The 1975–2018 sample indicates the following: First, the FRB staff's output gap (O-Gap) has the highest NSRs. Edge and Rudd (2016) and Barbarino, Berge, and Stella (2024) show that this is driven by the volatility in the earlier part of the sample. Interestingly, the NSR for the FRB staff's implied real expectations gap (E-Gap), which is based on the same information the FRB staff had when they

[14] Following the approaches in Orphanides and van Norden (2002), Edge and Rudd (2016), and Barbarino, Berge, and Stella (2024).

Table 1. Revision Properties for Real Gaps

	1975–2018 (Obs: 176)			**1991–2023 (Obs: 132)**		
	NSR		**% sign agree**	**NSR**		**% sign agree**
	SD	**RMS**		**SD**	**RMS**	
CBO O-Gap	·	·	·	0.71	0.71	73
FRB O-Gap	1.13	1.20	82	·	·	·
FRB E-Gap	0.33	0.41	86	·	·	·
SPF E-Gap	0.39	0.47	78	0.36	0.42	73
MHF O-Gap	**0.28**	**0.29**	**94**	**0.32**	**0.32**	**94**
MBN O-Gap	0.42	0.47	86	0.34	0.37	92

Note: Bolded values denote best performance by metric and sample. MHF is the modified Hamilton filter from Quast and Wolters (2022). MBN is the modified Beveridge-Nelson decomposition from Kamber, Morley, and Wong (2018). NSR: Noise-to-signal ratio. RMS: Root mean square deviation. SD: Standard deviation.

generated their O-Gap, is just over one-third of their O-Gap. The SPF's real E-Gap performs roughly in line with the FRB's E-Gap. However, the measure with the lowest NSR this period is the MHF, which is one-fourth the FRB's O-Gap, and also has the highest share of agreeing signs at 94 percent.

The results for the 1991–2023 sample are similar. The CBO's estimates have the largest NSR and smallest share of sign agreements. The SPF's real E-Gap NSRs are roughly half the size of the CBO's. The SPF's real E-Gap also performs better than in the first sample. The MHF has the lowest NSR and the highest share of estimates with the same sign, followed by the modified BN decomposition (MBN) from Kamber, Morley, and Wong (2018). These results are consistent with Barbarino, Berge, and Stella (2024)'s findings for a broader set of methods and across different subsamples and are also consistent with Kamber, Morley, and Wong (2018)'s findings that simpler models are subject to fewer revisions because they have fewer parameters to estimate and are less likely to overfit the data.

Table 2 presents the performance of the nominal gaps. In contrast to real gaps, the MHF is the worst performer across both subsamples. In the first subsample, the NSRs for the FRB staff's nominal E-Gap are essentially unchanged from their real E-Gap performance and

Table 2. Revision Properties for Nominal Gaps

	1975–2018 (Obs: 176)			1991–2023 (Obs: 132)		
	NSR		% sign agree	NSR		% sign agree
	SD	RMS		SD	RMS	
BCEI E-Gap	·	·	·	0.28	0.31	93
FRB E-Gap	0.34	0.40	**93**	·	·	·
SPF E-Gap	**0.31**	**0.35**	91	**0.26**	**0.29**	92
MHF O-Gap	0.75	1.23	60	0.48	0.61	78
MBN O-Gap	**0.31**	0.50	84	0.29	0.35	**94**

Note: Bolded values denote best performance by metric and sample. MHF is the modified Hamilton filter from Quast and Wolters (2022). MBN is the modified Beveridge-Nelson decomposition from Kamber, Morley, and Wong (2018). NSR: Noise-to-signal ratio. RMS: Root mean square deviation. SD: Standard deviation.

the SPF exhibits lower NSRs than for the real E-Gap. The SPF's nominal E-Gap and the MBN have the lowest NSRs for the first sample period, while the FRB staff's E-Gap has the largest share of sign agreement. In the second sample period, the SPF has the lowest NSRs while the MBN has the highest share of sign agreement. The SPF was also on par with the best performance of the real gaps in Table 1. The BCEI NGDP E-Gap also performs as well as these measures. These results, and those for the real gaps, are robust to the choice of data vintage; see the appendix.

Overall, the nominal gaps perform as well as or better than real gap measures. All of the NGDP E-Gaps have NSRs that are two to three times smaller than the FRB or CBO real O-Gaps. Furthermore, both the nominal and real measures of the FRB E-Gap dramatically outperform the FRB O-Gap despite being constructed from the same information set. While the nominal E-Gaps do not consistently outperform simple statistical procedures, they are more robust to those methods in cases where the underlying model may be misspecified.

4.2 Forecasting Core PCE Inflation

Another way to evaluate competing gap measures is to assess their effectiveness in forecasts of inflation using the Phillips curve. This exercise utilizes Yellen (2015, 2017)'s model of core PCE inflation, which extends the typical autoregressive distributed lag model used in previous evaluation exercises—see Orphanides and van Norden (2005)—that has been widely used in macroeconomic policy settings. The unrestricted model is specified as

$$\pi_t^c = \beta_1 \pi_t^e + \beta_2 \pi_{t-1}^c + \beta_3 \pi_{t-2}^c + \beta_4 \mathrm{GAP}_t + \beta_5 \mathrm{RPIM}_t + \epsilon_t, \qquad (6)$$

where π_t^c denotes the real-time vintage of the growth rate (expressed in annualized log differences) of core PCE prices;[15] π_t^e is expected long-run inflation proxied by the latest available real-time forecast of long-run PCE inflation reported in the SPF, RPIM_t is a control for the effect of changes in the relative price of core imported goods and is defined as the annualized growth rate of the price index for core imported goods (excluding petroleum, natural gas, computers, and semiconductors), less the lagged four-quarter change in core PCE inflation, all multiplied by the share of nominal core imported goods in NGDP, and GAP_t captures the level of measured resource utilization.

In the model's original formulation, GAP_t is approximated by the unemployment rate gap (U-Gap) calculated as the unemployment rate less the CBO's measure of the long-run natural rate of unemployment. This exercise compares three alternative measures. The first of these is the FRB staff's real-time measure of the O-Gap and forecasts thereof for all vintages from 1996 until 2019. The sample is extended with real-time CBO measures of potential output and the BCEI's real-time monthly forecasts of real GDP. Second, real-time measures of the U-Gap are derived from the FRB staff's estimates of the natural rate and real-time forecasts of the unemployment rate. The sample is extended using real-time measures of the CBO's natural rate of unemployment and the BCEI's real-time monthly forecasts of the unemployment rate. Third, the real-time monthly vintages of the BCEI's NGDP E-Gap are considered along with forward-looking estimates of this gap. The

[15] Real-time vintages of core PCE are from the Federal Reserve Bank of Philadelphia's real-time database; see Croushore and Stark (2001).

forward-looking estimates are constructed by reformulating Equation (5) as a j-quarter-ahead forecast of the expected gap,

$$\mathbb{E}_t\left[EGAP_{t+j}\right] = \tilde{\mathbb{E}}_t\left[y_{i,t+j}\right] - \frac{1}{20}\sum_{h=0}^{19-j}\left(\tilde{\mathbb{E}}_{t-h}\left[y_{i,t+j}\right] + j\tilde{\mathbb{E}}_t\left[y_{i,t+j}\right]\right), \tag{7}$$

where at each quarter ahead additional weight is given to the latest estimate of expected output. This is consistent with the notion that estimates of the underlying trend follow a random walk such that the most recent forecast should receive the most weight; see Morley (2011). This formulation also implies that when evaluated against forecasts of NGDP, any notion of slack or overheating is expected to dissipate within 20 quarters or five years by construction.

The model in Equation (6) is reestimated in real time for each alternative GAP measure every month starting in February 1996 using an expanding estimation window starting in 1990:Q1. For each monthly vintage, a conditional forecast of quarterly core PCE inflation is generated for the current quarter up through four quarters ahead based on the estimated model. It is assumed that π_t^e and RPIM_t follow a random walk such that forecasts are simply extrapolations of the last available real-time observation.[16] Finally, the forecasts are evaluated against the third estimate of core PCE inflation, as is common in the literature; see, for example, Sinclair and Stekler (2013).

Table 3 presents the out-of-sample forecast accuracy results for the alternative GAP measures. This analysis treats the forecasts using the real-time FRB/CBO O-Gap as a relative baseline and uses it to compare the forecasts derived from the FRB/CBO U-Gap and the BCEI NGDP E-Gap respectively. The results are presented in terms of the relative root mean square forecast errors (RMSEs) across three different samples: the Great Moderation sample from 1996:Q1 through 2007:Q4, the Great Recession sample from 2008:Q1 through 2019:Q4, and the pandemic sample of 2020:Q1 through 2024:Q4. This allows for evaluation of forecast performance across several different economic cycles. A value less than one implies that

[16]Note that our measure of RPIM_t is only quasi-real-time in that we use the latest vintage without accounting for data revisions.

Table 3. Core PCE Inflation Forecasts: RMSEs Relative to the O-Gap

	Great Moderation 1996:Q1–2007:Q4		Great Recession 2008:Q1–2019:Q4		Pandemic 2020:Q1–2024:Q4	
h	**U-Gap**	**E-Gap**	**U-Gap**	**E-Gap**	**U-Gap**	**E-Gap**
0	0.99	**0.96**	**0.98**	1.00	0.99	**0.96***
1	0.98	**0.97**	**0.98**	1.03	1.00	**0.94***
2	0.98	**0.96**	1.01	1.03	1.01	**0.95**
3	0.96	**0.95**	**0.98**	1.00	1.01	**0.96**
4	**0.94***	0.97	**0.98**	0.98	1.01	**0.97**

Note: Bold terms denote best performance by sample/horizon. Stars show statistically significant differences: ** has a p-value less than 1 percent and * has a p-value less than 5 percent. The Great Moderation sample has 131 observations, the Great Recession sample has 141, and the pandemic sample has 62.

a measure performs better than the FRB/CBO O-Gap measure, while a value greater than one indicates that it performs worse. Asterisks denote whether the differences are statistically significant using standard Diebold and Mariano (1995) tests.

The results for the Great Moderation sample in Table 3 indicate that all three measures of slack perform similarly. Forecasts using the O-Gap have the largest RMSEs across all horizons, while the E-Gap generally has the smallest errors until the longest horizon. Only the U-Gap produces significantly better forecasts at the longest horizon. The results flip during the Great Recession sample, when the E-Gap generally has the largest RMSEs, although none of the differences are statistically significant.

The pandemic era results indicate that forecasts based on the E-Gap perform best. The E-Gap has a smaller RMSE than both the O-Gap and the U-Gap at every forecast horizon. Furthermore, the differences are significant at the 2 percent level for the nowcast and the one-quarter-ahead horizon and at the 9 percent level (not shown) for the two- to four-quarter-ahead horizons. This illustrates that real-time estimates of the forward-looking NGDP expectations gap perform as well as the FRB/CBO's measures of the O-Gap and the U-Gap in normal times and that the NGDP expectations gap is able to outperform other measures during periods of high inflation.

4.3 Real-Time Taylor Rule Reliability

This last exercise analyzes the sensitivity of Taylor rules for setting interest rates using various real-time measures of slack. To do so, we extend the analysis done in Orphanides (2001). We use a simple Taylor rule following the original specification in Taylor (1993):

$$i_t = r_t^* + \pi_t + 0.5\left(\pi_t - \pi_t^*\right) + 0.5\mathrm{GAP}_t, \tag{8}$$

where i_t is the predicted nominal federal funds rate, r_t^* is the natural real rate of interest, π_t is a measure of 12-month inflation, π_t^* is the inflation target, and GAP_t represents a measure of economic slack. When considering the O-Gap, then $\mathrm{GAP}_t = y_t - y_t^*$ and $\pi_t^* = 2$. For the NGDP E-Gap, if we note that $\pi_t = p_t - p_{t-1}$ and $n_t = y_t + p_t$ is nominal income, then for a general inflation target $\pi_t^* = p_t^* - p_{t-1}$, (8) simplifies to

$$i_t = r_t^* + \pi_t + 0.5\left(n_t - n_t^*\right), \tag{9}$$

which is the standard NGDP targeting rule. The change in n_t^* is generally presumed to be the sum of the inflation target and the growth rate of potential real GDP. In the NGDP E-Gap framework, n_t^* is more flexible and develops according to forecasters' expectations as they evolve over time.

Following Knotek et al. (2016), real-time estimates of the natural real rate of interest are obtained from the FOMC's quarterly summary of economic projections (SEP) median projections of the longer-run federal funds rate minus long-run median expectations of core PCE inflation. For periods when these were not available, we use a fixed 2 percent real interest rate until 2009 as a real-time proxy and then use the long-run Blue Chip Financial Forecasters surveys' estimates of the real federal funds rate thereafter.

Table 4 presents the revision properties for the FRB/CBO O-Gap and the BCEI NGDP E-Gap. The NGDP expectations gap has the smallest real-time revisions and volatility across all metrics and all measures of inflation. This is consistent with the results in subsection 4.1 and reconfirms earlier findings by Orphanides (2001) that the volatility in monetary policy rules is driven in large part by instability in estimates of slack. Our results indicate that despite improvements in output gap measurement over time, the NGDP expectations gap continues to outperform these estimates.

Table 4. Revision Properties of Taylor Rules by Measure of Inflation

	GDP Deflator			PCE			Core PCE		
		NSR			NSR			NSR	
Slack	**s.d.**	**SD**	**RMS**	**s.d.**	**SD**	**RMS**	**s.d.**	**SD**	**RMS**
O-Gap	0.94	0.39	0.25	0.90	0.36	0.24	1.09	0.53	0.32
E-Gap	**0.59**	**0.22**	**0.15**	**0.57**	**0.21**	**0.14**	**0.67**	**0.27**	**0.18**

Note: Bold terms denote best performance by sample/horizon. The sample is 1996:Q1–2024:Q4.

Table 5. Revision Properties of Forward-Looking Taylor Rules by Horizon

	h=0			h=2			h=4		
		NSR			NSR			NSR	
Slack	**s.d.**	**SD**	**RMS**	**s.d.**	**SD**	**RMS**	**s.d.**	**SD**	**RMS**
O-Gap	1.10	0.46	0.27	1.91	0.80	0.47	2.52	1.05	0.61
E-Gap	**0.88**	**0.32**	**0.22**	**1.73**	**0.63**	**0.40**	**2.39**	**0.88**	**0.54**

Note: Bold terms denote best performance by sample/horizon. The sample is 1996:Q1–2024:Q4.

Next, the real-time performance of a forward-looking Taylor rule is considered. The analysis uses forecasts of the GDP deflator inflation measure, since these are available from both the FRB staff and BCEI forecasters. Table 5 presents the results for forecast horizons from the nowcast up through four quarters ahead. The results again show that the Taylor rule using the E-Gap has smaller real-time revisions than when using the O-Gap, even at longer horizons.

Overall, the NGDP E-Gap generates interest rate targets that are up to 40 percent less volatile than those suggested by standard O-Gap measures. There are two obvious reasons for these gains. First, the results in subsection 4.1 clearly show that the NGDP E-Gap is less sensitive to revisions in real time, in part because the forecasts are not revised for a given survey publication date. Second, the NGDP E-Gap combines the inflation and real O-Gaps into a single NGDP gap measure that further reduces the volatility of estimates from a Taylor rule. These results are also consistent with the results for other studies using alternative monetary policy rules; see, for example, Beckworth and Hendrickson (2020) and Beckworth and Horan (2022).

5. Conclusion

This paper compares the real-time performance of NGDP gaps against other measures of economic slack. To do so, it uses the NGDP expectations gap, which is constructed as the difference between observed NGDP and what surveys of professional forecasters expected it to be, to compare the real-time performance of nominal and real gap measures from the FRB staff, the CBO, and several forecast-based measures. The paper evaluates the measures in the context of several policy-relevant applications, including real-time forecasts of core PCE inflation and interest rate targets.

The analysis shows that the NGDP expectations gap performs as well as or better than other measures. First, its historical revisions are two to three times smaller than the output gap measures produced by the FRB staff and the CBO. Moreover, they are comparable with or slightly better than other statistical output gap measures such as Kamber, Morley, and Wong (2018) and Quast and Wolters (2022). Second, forecasts of core PCE inflation following the COVID-19 pandemic that are based on the NGDP expectations gap are significantly more accurate than those based on the output gap or unemployment rate gap measures. Finally, Taylor rules for setting interest rates are up to 40 percent less volatile in real time when using the NGDP expectations gap.

The results of this paper indicate that the fears associated with real-time implementation of nominal targets are overblown. While all measures of economic slack are subject to mismeasurement and revisions, the NGDP expectations gap performs among the best across the available measures and provides a useful measure of economic slack in real-time settings. This is especially true in the high-inflation environment following the COVID-19 pandemic, when economic slack was of particular concern for policymakers. Thus, while the most appropriate target for monetary policy is still open for debate, the concern associated with the misestimation of NGDP targets is misplaced.

Appendix

Table A.1. Revision Properties for Real Gaps across Multiple Data Vintages: Advance, Second, and Third

	1975–2018 (Obs: 528)			**1991–2023 (Obs: 396)**		
	NSR		**% sign agree**	**NSR**		**% sign agree**
	SD	**RMS**		**SD**	**RMS**	
CBO O-Gap	·	·	·	0.68	0.68	75
FRB O-Gap	1.13	1.20	84	·	·	·
FRB E-Gap	0.33	0.40	87	·	·	·
SPF E-Gap	0.39	0.47	78	0.36	0.41	72
MHF O-Gap	**0.28**	**0.29**	**94**	**0.32**	**0.32**	**94**
MBN O-Gap	0.41	0.46	87	0.36	0.33	92

Note: Bolded values denote best performance by metric and sample. MHF is the modified Hamilton filter from Quast and Wolters (2022). MBN is the modified Beveridge-Nelson decomposition from Kamber, Morley, and Wong (2018). NSR: Noise-to-signal ratio. RMS: Root mean square deviation. SD: Standard deviation.

Table A.2. Revision Properties for Nominal Gaps across Multiple Data Vintages: Advance, Second, and Third

	1975–2018 (Obs: 528)			1991–2023 (Obs: 396)		
	NSR		% sign agree	NSR		% sign agree
	SD	RMS		SD	RMS	
BCEI E-Gap	·	·	·	0.27	0.29	93
FRB E-Gap	0.33	0.39	**93**	·	·	·
SPF E-Gap	**0.30**	**0.34**	91	**0.25**	**0.28**	92
MHF O-Gap	0.75	1.22	61	0.48	0.60	78
MBN O-Gap	0.31	0.49	83	0.28	0.34	**94**

Note: Bolded values denote best performance by metric and sample. MHF is the modified Hamilton filter from Quast and Wolters (2022). MBN is the modified Beveridge-Nelson decomposition from Kamber, Morley, and Wong (2018). NSR: Noise-to-signal ratio. RMS: Root mean square deviation. SD: Standard deviation.

References

Ball, L. 1999. "Efficient Rules for Monetary Policy." *International Finance* 2 (1): 63–83.

Barbarino, A., T. J. Berge, and A. Stella. 2024. "The Stability and Economic Relevance of Output Gap Estimates." *Journal of Applied Econometrics* 39 (6): 1065–81.

Bates, J. M., and C. W. Granger. 1969. "The Combination of Forecasts." *Journal of the Operational Research Society* 20 (4): 451–68.

Baumeister, C., D. Leiva-León, and E. Sims. 2024. "Tracking Weekly State-Level Economic Conditions." *Review of Economics and Statistics* 106 (2): 483–504.

Beaudry, P., D. Galizia, and F. Portier. 2020. "Putting the Cycle Back into Business Cycle Analysis." *American Economic Review* 110 (1): 1–47.

Beckworth, D. 2020. "The Stance of Monetary Policy: The NGDP Gap." Mercatus Policy Brief (April). Arlington, VA: Mercatus Center at George Mason University.

Beckworth, D., and J. R. Hendrickson. 2020. "Nominal GDP Targeting and the Taylor Rule on an Even Playing Field." *Journal of Money, Credit and Banking* 52 (1): 269–86.

Beckworth, D., and P. Horan. 2022. "The Fate of FAIT: Salvaging the FED's Framework." Mercatus Center Working Paper (October). Arlington, VA: Mercatus Center at George Mason University.

Berge, T. J. 2023. "Time-Varying Uncertainty of the Federal Reserve's Output Gap Estimate." *Review of Economics and Statistics* 105 (5): 1191–206.

Berger, T., J. Morley, and B. Wong. 2020. "Nowcasting the Output Gap." *Journal of Econometrics* 232 (1): 18–34.

Beveridge, S., and C. R. Nelson. 1981. "A New Approach to Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to Measurement of the 'Business Cycle'." *Journal of Monetary Economics* 7 (2): 151–74.

Blanchard, O., and D. Quah. 1989. "The Dynamic Effects of Aggregate Demand and Supply Disturbances." *American Economic Review* 79 (4): 655–73.

Bullard, J., and A. Singh. 2020. "Nominal GDP Targeting with Heterogeneous Labor Supply." *Journal of Money, Credit and Banking* 52 (1): 37–77.

Bürgi, C., and T. M. Sinclair. 2021. "What Does Forecaster Disagreement Tell Us about the State of the Economy?" *Applied Economics Letters* 28 (1): 49–53.

Champagne, J., G. Poulin-Bellisle, and R. Sekkel. 2018. "The Real-Time Properties of the Bank of Canada's Staff Output Gap Estimates." *Journal of Money, Credit and Banking* 50 (6): 1167–88.

Clements, M. P., and D. F. Hendry. 1993. "On the Limitations of Comparing Mean Square Forecast Errors." *Journal of Forecasting* 12 (8): 617–37.

———. 2001a. *Forecasting Non-Stationary Economic Time Series.* Cambridge, MA: MIT Press.

———. 2001b. "Forecasting with Difference-Stationary and Trend-Stationary Models." *Econometrics Journal* 4 (1): 1–19.

Coibion, O., Y. Gorodnichenko, and M. Ulate. 2018. "The Cyclical Sensitivity in Estimates of Potential Output." *Brookings Papers on Economic Activity* (Fall): 343–441.

Croushore, D., and T. Stark. 2001. "A Real-Time Data Set for Macroeconomists." *Journal of Econometrics* 105 (1): 111–30.

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

Edge, R. M., and J. B. Rudd. 2016. "Real-Time Properties of the Federal Reserve's Output Gap." *Review of Economics and Statistics* 98 (4): 785–91.

Frankel, J., and M. Chinn. 1995. "The Stabilizing Properties of a Nominal GNP Rule." *Journal of Money, Credit and Banking* 27 (2): 318–34.

Furlanetto, F., K. Hagelund, F. Hansen, and Ø. Robstad. 2023. "Norges Bank Output Gap Estimates: Forecasting Properties, Reliability, Cyclical Sensitivity and Hysteresis." *Oxford Bulletin of Economics and Statistics* 85 (1): 238–67.

Genre, V., G. Kenny, A. Meyler, and A. Timmermann. 2013. "Combining Expert Forecasts: Can Anything Beat the Simple Average?" *International Journal of Forecasting* 29 (1): 108–21.

Guérin, P., L. Maurin, and M. Mohr. 2015. "Trend-Cycle Decomposition of Output and Euro Area Inflation Forecasts: A Real-Time Approach Based on Model Combination." *Macroeconomic Dynamics* 19 (2): 363–93.

Hall, R. E., and N. G. Mankiw. 1994. "Nominal Income Targeting." In *Monetary Policy*, ed. N. G. Mankiw, 71–94. University of Chicago Press.

Hamilton, J. D. 2018. "Why You Should Never Use the Hodrick-Prescott Filter." *Review of Economics and Statistics* 100 (5): 831–43.

Hendry, D. F., and M. P. Clements. 2004. "Pooling of Forecasts." *Econometrics Journal* 7 (1): 1–31.

Hodrick, R. J., and E. C. Prescott. 1997. "Postwar US Business Cycles: An Empirical Investigation." *Journal of Money, Credit and Banking* 29 (1): 1–16.

Irwin, N. 2018. "The Most Important Least-Noticed Economic Event of the Decade." *New York Times*, September 29. https://www.nytimes.com/2018/09/29/upshot/mini-recession-2016-little-known-big-impact.html.

Isiklar, G., and K. Lahiri. 2007. "How Far Ahead Can We Forecast? Evidence from Cross-Country Surveys." *International Journal of Forecasting* 23 (2): 167–87.

Justiniano, A., and G. E. Pimiceri. 2008. "Potential and Natural Output." Mimeo, Northwestern University.

Kamber, G., J. Morley, and B. Wong. 2018. "Intuitive and Reliable Estimates of the Output Gap from a Beveridge-Nelson Filter." *Review of Economics and Statistics* 100 (3): 550–66.

Kiley, M. T. 2013. "Output Gaps." *Journal of Macroeconomics* 37: 1–18.

Knotek, E. S., R. J. Verbrugge, C. Garciga, C. Treanor, and S. Zaman. 2016. "Federal Funds Rates Based on Seven Simple Monetary Policy Rules." Economic Commentary No. 2016-07 (July), Federal Reserve Bank of Cleveland.

Lahiri, K., and X. Sheng. 2010. "Measuring Forecast Uncertainty by Disagreement: The Missing Link." *Journal of Applied Econometrics* 25 (4): 514–38.

Martinez, A. B., J. L. Castle, and D. F. Hendry. 2022. "Smooth Robust Multi-Horizon Forecasts." In *Essays in Honor of M. Hashem Pesaran: Prediction and Macro Modeling*, Vol. 43A, ed. A. Chudik, C. Hsiao, and A. Timmermann, 143–65. Emerald Publishing Limited.

Morley, J. C. 2002. "A State–Space Approach to Calculating the Beveridge-Nelson Decomposition." *Economics Letters* 75 (1): 123–27.

———. 2011. "The Two Interpretations of the Beveridge–Nelson Decomposition." *Macroeconomic Dynamics* 15 (3): 419–39.

Morley, J. C., and J. Piger. 2012. "The Asymmetric Business Cycle." *Review of Economics and Statistics* 94 (1): 208–21.

Orphanides, A. 2001. "Monetary Policy Rules Based on Real-Time Data." *American Economic Review* 91 (4): 964–85.

———. 2003. "The Quest for Prosperity without Inflation." *Journal of Monetary Economics* 50 (3): 633–63.

Orphanides, A., and S. van Norden. 2002. "The Unreliability of Output-Gap Estimates in Real Time." *Review of Economics and Statistics* 84 (4): 569– 83.

———. 2005. "The Reliability of Inflation Forecasts Based on Output Gap Estimates in Real Time." *Journal of Money, Credit and Banking* 37 (3): 583–601.

Patton, A. J., and A. Timmermann. 2010. "Why Do Forecasters Disagree? Lessons from the Term Structure of Cross-Sectional Dispersion." *Journal of Monetary Economics* 57 (7): 803–20.

Phillips, P. C., and Z. Shi. 2021. “Boosting: Why You Can Use the HP Filter.” *International Economic Review* 62 (2): 521–70.

Pichette, L., M.-N. Robitaille, M. Salameh, and P. St-Amant. 2019. “Dismiss the Output Gaps? To Use with Caution Given Their Limitations.” *Economic Modelling* 76: 199–215.

Quast, J., and M. H. Wolters. 2022. “Reliable Real-Time Output Gap Estimates Based on a Modified Hamilton Filter.” *Journal of Business and Economic Statistics* 40 (1): 152–68.

Schibuola, A., and A. B. Martinez. 2021. “The Expectations Gap: An Alternative Measure of Economic Slack.” Mercatus Center Working Paper (September), Arlington, VA: Mercatus Center at George Mason University.

Shackleton, R. 2018. “Estimating and Projecting Potential Output Using CBO’s Forecasting Growth Model.” CBO Working Paper No. 2018-03, Washington, D.C.: Congressional Budget Office.

Sheedy, K. D. 2014. “Debt and Incomplete Financial Markets: A Case for Nominal GDP Targeting.” *Brookings Papers on Economic Activity* 2014 (1): 301–73.

Sinclair, T. M., and H. O. Stekler. 2013. “Examining the Quality of Early GDP Component Estimates.” *International Journal of Forecasting* 29 (4): 736– 50.

Svensson, L. E. O. 2020a. “Monetary Policy Strategies for the Federal Reserve.” *International Journal of Central Banking* 16 (1): 133–93.

———. 2020b. “What Rule for the Federal Reserve? Forecast Targeting.” *International Journal of Central Banking* 16 (6): 39–95.

Taylor, J. B. 1993. “Discretion versus Policy Rules in Practice.” *Carnegie Rochester Conference Series on Public Policy* 39 (December): 195–214.

Timmermann, A. 2006. “Forecast Combinations.” In *Handbook of Economic Forecasting*, Vol. 1, ed. A. Timmermann, C. Granger, and E. Graham, 135–96. Elsevier.

Yellen, J. L. 2015. “Inflation Dynamics and Monetary Policy.” Speech delivered at the Philip Gamble Memorial Lecture, University of Massachusetts Amherst, Amherst, Mass. Available online at: https://www.federalreserve.gov/newsevents/speech/files/yellen20150924a.pdf. Last accessed: August 26, 2022. Washington, D.C.: Board of Governors of the Federal Reserve System.

———. 2017. "Inflation, Uncertainty, and Monetary Policy." *Business Economics* 52 (4): 194–207.