**Finance and Economics Discussion Series**
**Divisions of Research & Statistics and Monetary Affairs**
**Federal Reserve Board, Washington, D.C.**

# Which Output Gap Estimates Are Stable in Real Time and Why?

**Alessandro Barbarino, Travis J. Berge, Han Chen, Andrea Stella**

**2020-102**

Please cite this paper as:
Barbarino, Alessandro, Travis J. Berge, Han Chen, and Andrea Stella (2020). "Which Output Gap Estimates Are Stable in Real Time and Why?," Finance and Economics Discussion Series 2020-102. Washington: Board of Governors of the Federal Reserve System, https://doi.org/10.17016/FEDS.2020.102.

NOTE: Staff working papers in the Finance and Economics Discussion Series (FEDS) are preliminary materials circulated to stimulate discussion and critical comment. The analysis and conclusions set forth are those of the authors and do not indicate concurrence by other members of the research staff or the Board of Governors. References in publications to the Finance and Economics Discussion Series (other than acknowledgement) should be cleared with the author(s) to protect the tentative character of these papers.

# Which Output Gap Estimates Are Stable in Real Time and Why?

Alessandro Barbarino*
Federal Reserve Board

Travis J. Berge
Federal Reserve Board

Han Chen
Federal Reserve Board

Andrea Stella
Federal Reserve Board

December 11, 2020

**Abstract**

Output gaps that are estimated in real time can differ substantially from those estimated after the fact. We aim to understand the real-time instability of output gap estimates by comparing a suite of reduced-form models. We propose a new statistical decomposition and find that including a Okun's law relationship improves real-time stability by alleviating the end-point problem. Models that include the unemployment rate also produce output gaps with relevant economic content. However, we find that no model of the output gap is clearly superior to the others along each metric we consider.

*JEL classifications*: E24, E32, E52.
*Keywords*: Output gap; real-time data; trend-cycle decomposition; unobserved components model.

*We thank without implication Gianni Amisano for helpful comments and suggestions. Michael Kister and Niko Paulson provided excellent research assistance. We also thank seminar participants at CBF 2018, SETA 2019, and IAAE 2019 for their comments. The views expressed here are our own and do not necessarily reflect the views of the Board of Governors of the Federal Reserve or the Federal Reserve System.

# 1 Introduction

The output gap—the deviation of observed output from its potential level—is a concept fundamental to the conduct of economic policy. Indeed, the cyclical state of the economy is often at the forefront of policymakers' thinking, and they often refer to the output gap when discussing policy decisions.[1] Yet the output gap is never observed and must be estimated. And in an influential paper, Orphanides and van Norden (2002) document that output gap estimates are unreliable in real time, throwing into question their adequacy for policymakers.

In light of these findings, a large literature has taken on the challenge of improving the stability of real-time output gap estimates. Garratt, Lee, Mise and Shields (2008) ameliorate the end-point problem by detrending output data that has had forecasts appended to it, while Clements and Galvão (2012) explicitly model data revisions alongside the output gap. Another approach is to include highly cyclical variables in multivariate models. Trimbur (2009) finds manufacturing capacity utilization improves the real-time properties of output gap estimates. Fleischman and Roberts (2011) extract an output gap estimate from a large number of macroeconomic variables; their model also improves real-time stability. Finally, Edge and Rudd (2016) consider the real-time properties of the Federal Reserve staff's judgmental output gap estimates. They show that, in contrast to model-based approaches, the Board staff produced stable output gap estimates in real time between the mid-1990s and the mid-2000s (when their sample ends).

In this paper, we aim to understand what contributes to the real-time stability of an output gap estimate. To do so, we produce output gap estimates from a suite of models, including the models considered in Orphanides and van Norden (2002), the relatively new de-trending methods from Mueller and Watson (2017) and Hamilton (2018), the judgmental output gap from the Federal Reserve's staff, as well as a number of univariate and multi-variate unobserved components models. We next propose a decomposition of the estimates from models that can be written as a linear Gaussian state-space system. The decomposition estimates the contribution of each observable series to the real-time instability of the output gap from three distinct sources: revisions to the underlying data, parameter instability, and the end-point problem. Lastly, we consider the performance of the various output gaps in terms of their economic value by comparing them to a judgmental estimate produced by the staff of the Federal Reserve, their ability to forecast inflation, and whether they are consistent with ex-post monetary policy decisions.

[1] See, for example, the January 19, 2017 speech of Janet Yellen "The Economic Outlook and the Conduct of Monetary Policy," made when she was Chair of the Federal Open Market Committee of the Federal Reserve.

The results confirm the difficulty of estimating the output gap in real-time. First, we find that the Federal Reserve staff's output gap is as stable as any of the model-based estimates we consider, reconfirming Edge and Rudd (2016) with a dataset that now includes the Great Recession. We also find that the Tealbook output gap has economic content, in the sense that it forecasts inflation no worse than any of the other models we consider and implies an interest rate policy that is relatively close to the actual outcome.[2] Among the models we consider, those that include an Okun's law relationship, relating the unemployment rate's deviation from its trend to the output gap, are overall quite similar to the Tealbook output gap estimate. They provide a stable and reliable estimate of the cyclical position of the economy, and are consistently among the best performers along a number of economic considerations, suggesting that they have the ability to inform policy in a meaningful way.

A recent related literature has focused on estimating the output gap using large macroeconomic panels. Aastveit and Trovik (2014) and Barigozzi and Luciani (2018) use large datasets to estimate the output gap via a dynamic factor model. Morley and Wong (2019) estimate the output gap with a Bayesian VAR comprising a large macroeconomic dataset. Importantly, their methodology allows them to inspect the contribution of VAR forecast errors to the cycle. They find that the unemployment rate is the single most important variable that contributes to the cycle. However, these papers do not focus on the real-time stability of the output gap estimate.

The next section introduces the real-time data and various models used to estimate the cyclical state of the economy. Section 3 documents the real-time stability of the suite of models we consider, while in Section 4 we decompose the model's relative instability. In Section 5 we evaluate the economic relevance of the output gap instabilities we document, and the final section provides a concluding discussion.

# 2 Empirical design

## 2.1 Real-time data

We use real-time vintage data throughout the paper. Our primary data source is the Real-time Dataset for Macroeconomic of the Federal Reserve Bank of Philadelphia, where we obtain real-time data vintages of real

[2]The Federal Reserve Board staff estimate is contained in a document known as the Tealbook, prepared prior to each Federal Open Market Committee (FOMC) meeting. The Tealbook was previously known as the Greenbook. Throughout, however, we will refer to the Federal Reserve staff forecast as the Tealbook. Tealbook forecasts can be found at: `https://www.philadelphiafed.org/research-and-data/real-time-center/greenbook-data/`.

GDP, the unemployment rate, the GDP price deflator, and manufacturing capacity utilization.[3] The database contains quarterly real-time vintages for each series. That is, the real-time vintage gives the data series as it would have been seen by a practitioner as of the middle of each quarter. Vintage $T+1$ contains data that typically begins in 1947:Q1 and continues through quarter $T$. For real GDP, inflation, and the unemployment rate, our first vintage is 1965:Q4, while the first available vintage for manufacturing capacity utilization is 1985:Q3. The final vintage in our dataset is 2019:Q1, for 214 vintages in total. The 1996:Q1 real-time GDP vintage is missing because of a federal government shutdown. [4]

We also consider the real-time stability of the output gap estimate produced by the staff of the Federal Reserve Board. The staff estimate is contained in a document known as the Tealbook, prepared prior to each Federal Open Market Committee (FOMC) meeting. Tealbook output gap estimates since March 1996 are available from the Federal Reserve Bank of Philadelphia. Following Orphanides and van Norden (2002) we use the first Tealbook estimate available each quarter. However, Edge and Rudd (2016) provide real-time output gap nowcasts—that is, estimates of the output gap in the previous quarter—since 1950:Q4, which we use to extend our sample earlier into history. Because Tealbook forecasts are made public with a five year lag, the last Tealbook output gap estimate in our dataset is from the fourth quarter of 2013.[5]

## 2.2 Models of the output gap

We first consider several univariate trend-cycle decompositions explored by Orphanides and van Norden (2002) and Edge and Rudd (2016). Let $y$ denote 100 times the log of real GDP. We first decompose log real GDP into a trend ($y^*$) and a cycle ($C$) using a linear time trend, a quadratic trend, and the HP filter. For completeness, we add other more recently developed models. The first is the trend-cycle model from Garratt et al. (2008) (GLMS). The GLMS methodology aims to reduce the end-point problem by applying an HP filter to a data series that has been augmented with a forecast from an AR model. The Mueller and Watson (2017) (MW) extracts a trend from a univariate series by projecting the series onto a small number of trigonometric functions intended to isolate the low-frequency information in the series. Finally, Hamilton (2018) provides a new regression-based approach to trend extraction. Each of these consider only real GDP

[3] See table A1. Croushore and Stark (2001) and the Federal Reserve Bank of Philadelphia's website (`https://www.philadelphiafed.org/research-and-data/real-time-center/real-time-data`) provide additional details.

[4] We set output gaps to missing in 1995:Q4 since many of the models cannot produce a gap estimate without additional assumptions. There are then 213 observations when we work with real-time data.

[5] See Cleve, Luke Van and Laforte, Jean-Philippe, and Stella, Andrea (2019) for additional details about vintages of the Federal Reserve staff's output gap estimate.

and are straightforward to implement.

Our primary focus, however, is on output gap estimates produced by unobserved components (UC) models, because we can cast them in state space form which offers a way to formally assess properties of the output gap estimates. The UC models decompose the log of real GDP into a trend, cycle, and idiosyncratic error ($e_y$): $y_t = y_t^* + C_t + e_{yt}$. Trend output follows random walk with drift, $\tau$, which itself follows a random walk:

$$y_t^* = y_{t-1}^* + \tau_{t-1} + \varepsilon_{y^*t}$$
$$\tau_t = \tau_{t-1} + \varepsilon_{\tau t}.$$

The drift term captures slow-moving changes to the growth rate of trend output due to, for example, changes to productivity growth or demographics. The error terms, $\varepsilon_{y^*t}$ and $\varepsilon_{\tau t}$, are distributed as standard normals. The output gap, $C$, follows a stationary AR(2) process

$$C_t = \phi_1 C_{t-1} + \phi_2 C_{t-2} + \varepsilon_{Ct}.$$

where the error term $\varepsilon_{Ct}$ is distributed as a standard normal. We consider several nested versions of this framework. The first is the univariate model above, first studied in Clark (1987). We then gradually enlarge the information set of the model to include other important macroeconomic indicators. We add to the framework above the unemployment rate, capacity utilization, and inflation one at a time. Considering results from bivariate UC models allows us to single out the individual contributions of the various macroeconomic indicators. In each case, the only change to the model is the inclusion of an additional observation equation and state-transition equation to the model. Consider the model that adds variable $x$. This variable relates to the output gap:

$$x_t = x_t^* + \alpha_0 C_t + \alpha_1 C_{t-1} + e_{xt},$$

where the parameters $\alpha_0$ and $\alpha_1$ measure the observable's response to changes in the output gap $C$. In the bivariate model with the unemployment rate, these would be Okun's law coefficients, whereas in the model with inflation, these would be Phillips curve parameters instead (Kuttner 1994). The other trend component, $x_t^*$, again follows a random walk.

The largest model includes real GDP, unemployment and inflation. While we will not explore the various

trends that result from our models, in principle, they are of interest. For example, the unemployment rate trend, $u^*$, could proxy the NAIRU (Staiger, Stock and Watson 1997), and the trend of inflation, $\pi^*$, could be interpreted as a proxy for inflation expectations (see Stock and Watson (2007), Kim, Manopimoke and Nelson (2014), and Clark and Doh (2014)). The fundamental aspect of our small multivariate models, whether they include unemployment rate or inflation, is the presence of a common cycle that affects the deviation of each macroeconomic variable from its trend.

For each data vintage, all models are estimated on an estimation sample that begins in 1960:Q1 and which expands as time goes by. The UC models are cast as linear Gaussian state-space models. We estimate these models in a Bayesian way using a Gibbs sampler. Since the estimation is standard, we leave the details to Appendix A2.1. However, in our experience, Bayesian unobserved components models can be sensitive to the priors placed on the variance of shocks to model trends.[6] Because we produce real-time estimates, we require a method for setting model priors in a systematic, data-driven manner. We take the following approach. For a given observable and data vintage, we produce an estimate of the trend using an HP filter on a pre-sample period, 1947:Q1–1959:Q4. We then center our Inverse Gamma prior about the variance implied by this pre-sample trend. We set the degrees of freedom parameter in the IG distribution to $T_v/20$, where $T_v$ denotes the number of observations in data vintage $v$. Priors for other model parameters are conjugate and diffuse. More details can be found in Appendix A2.2.

# 3 Estimating the output gap in real time

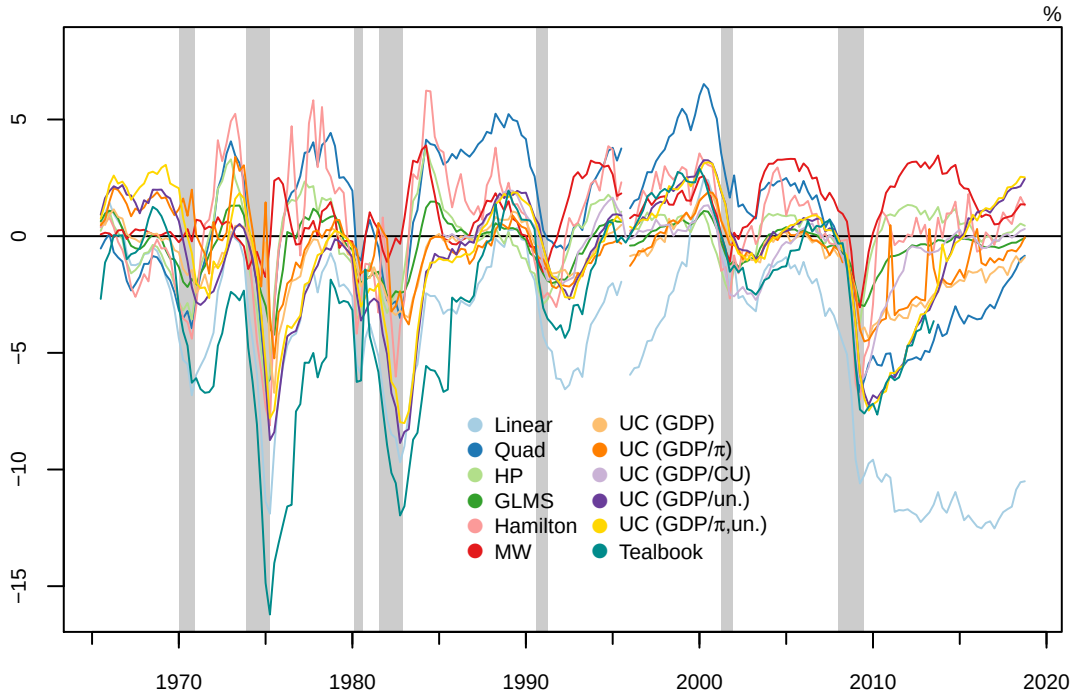
Following Orphanides and van Norden (2002) and Edge and Rudd (2016), for each data vintage, define the real-time output gap, $C_t^R$, as the estimate of the output gap in the previous quarter. So, for example, the real-time output gap associated with the 2019:Q1 vintage is the estimate of the output gap in 2018:Q4. Because the first estimate of quarterly NIPA data is released with roughly a one-month lag, this assumption ensures that a model estimated at vintage $v$ has the first NIPA estimates for output and inflation in quarter $v-1$, as well as the monthly estimates of the unemployment rate and capacity utilization.

Figure 1 shows the sequence of real-time output gap estimates. While each output gap estimate is broadly pro-cyclical, at any point in time the models deliver a wide range of estimates. Table 1 provides summary statistics, which reinforce the observation that output gap estimates differ dramatically across

[6] When the models are estimated via maximum likelihood, in our experience this sensitivity manifests itself as sensitivity to starting values.

models: the real-time estimates have notably different means and volatilities, and often have different first derivatives and signs.

Figure 1: Sequence of real-time output gap estimates, 1965:Q3–2018Q4.




Notes: Figure shows the sequence of real-time output gap estimates for the suite of models we consider. GLMS refers to Garratt et al. (2008), and MW to Mueller and Watson (2017). Grey shaded areas denote NBER-defined recession. Final Tealbook estimates for 2013:Q3. See text for details.

Table 1: Summary statistics of real-time output gap estimates, 1965:Q3–2018Q4.

| | N.obs | Mean | Std dev | Min | Max | $\% > 0$ |
|---|---|---|---|---|---|---|
| Linear | 213 | -4.55 | 4.06 | -12.53 | 2.47 | 8 |
| Quadratic | 213 | 0.41 | 3.24 | -6.71 | 6.52 | 57 |
| HP | 213 | -0.11 | 1.62 | -6.63 | 3.84 | 56 |
| GLMS | 213 | -0.40 | 1.06 | -4.22 | 1.48 | 39 |
| Hamilton | 213 | 0.42 | 2.49 | -8.11 | 6.24 | 62 |
| MW | 213 | 1.01 | 1.36 | -3.05 | 3.89 | 77 |
| UC (GDP) | 213 | -0.93 | 1.40 | -6.72 | 1.91 | 25 |
| UC (GDP/$\pi$) | 213 | -0.55 | 1.50 | -5.24 | 3.40 | 36 |
| UC (GDP/CU) | 134 | -0.49 | 1.54 | -6.31 | 1.97 | 43 |
| UC (GDP/u.) | 213 | -0.89 | 2.73 | -8.86 | 3.26 | 46 |
| UC (GDP/u./$\pi$) | 213 | -0.73 | 2.66 | -8.01 | 3.17 | 46 |
| Tealbook | 192 | -2.99 | 3.84 | -16.21 | 2.90 | 27 |

Notes: Sample period is 1965:Q3–2018:Q4 except for UC(GDP/CU) (begins in 1985:Q2) and Tealbook (ends in 2013:Q3). 1995:Q4 set to missing for all models. $\% > 0$ denotes the percent of quarters when the output gap is positive. See text for details.

Table 2 describes the revisions of each output gap model we consider. We define the output gap revision in quarter $t$ to be the difference between the in-sample output gap estimate from our last vintage, $V$=2019:Q1, and the real-time estimate. The revision proxies for uncertainty that surrounds any given real-time output gap estimate. The columns labeled "NSR" report noise-to-signal ratios: The first is defined as the ratio of the standard deviation of the model revision to the standard deviation of the final gap estimate, and the second is the ratio of the root mean square of the revisions to the standard deviation of the final gap estimate. The column "% sign agree" measures the fraction of observations where the real-time and final estimate have the same sign, i.e., whether the economy is above or below potential.

Table 2: Summary statistics of output gap revisions, 1965:Q3–2018:Q4.

| | N | Mean | Std dev | RMSE | NSR (SD) | NSR (RMSE) | % sign agree |
|---|---|---|---|---|---|---|---|
| Linear | 213 | 5.48 | 2.04 | 5.85 | **0.43** | 1.22 | 34 |
| Quadratic | 213 | -0.21 | 3.73 | 3.73 | 1.11 | 1.11 | 62 |
| HP | 213 | **0.16** | 1.53 | 1.54 | 1.05 | 1.05 | 59 |
| GLMS | 213 | 0.44 | **0.99** | **1.08** | 0.67 | 0.74 | 73 |
| Hamilton | 213 | -0.48 | 1.25 | 1.34 | 0.48 | 0.51 | **84** |
| MW | 213 | -1.01 | 1.72 | 1.99 | 1.11 | 1.28 | 57 |
| UC (GDP) | 213 | 1.91 | 2.08 | 2.82 | 0.73 | 1.00 | 65 |
| UC (GDP/$\pi$) | 213 | 0.80 | 1.31 | 1.53 | 0.70 | 0.81 | 77 |
| UC (GDP/CU) | 134 | 0.54 | 1.23 | 1.34 | 0.72 | 0.78 | 76 |
| UC (GDP/u.) | 213 | 0.82 | 1.38 | 1.60 | 0.49 | 0.56 | 81 |
| UC (GDP/u./$\pi$) | 213 | 0.71 | 1.22 | 1.41 | **0.43** | **0.50** | 83 |
| Tealbook | – | – | – | – | – | – | – |

Notes: Revision defined as final output gap estimate less real-time estimate, $C_t^V - C_t^R$ and $C_t^V$ denotes the output gap estimate using the $V$=2019:Q1 data vintage. Bold face entries indicate best performing model according to that measure. All statistics are for 1965:Q3–2018:Q4 except for UC (GDP/CU), which begins in 1985:Q2. Tealbook revision statistics not available because Tealbook 2019:Q1 output gap estimate is not yet available. See text for details.

Most univariate models fare quite poorly in terms of their real-time stability. With the exception of the GLMS and Hamilton models, the univariate models have noise-signal ratios that almost always exceed one. The procedure of Garratt et al. (2008) does improve real-time stability. We are among the first to evaluate the real-time properties of the newer detrending methods—see also Quast and Wolters (2019) and Berge (2020)—and find that while Hamilton's (2018) method is relatively stable, the MW decomposition is less stable than many other univariate models in real time. Among the UC models, UC (GDP) and UC (GDP/$\pi$) are less stable than the other UC models, but are about as stable as the univariate models. Comparing the UC (GDP/$\pi$) model to the UC (GDP), and comparing UC (GDP/u./$\pi$) to UC (GDP/u.), adding inflation

does not improve real-time stability, in contrast to Planas and Rossi (2004). Finally, we note that output gap estimates of models that include the unemployment rate have smaller revisions, less variable revisions, smaller noise-signal ratios, and tend not to change signs. Gonzalez-Astudillo and Roberts (2016) show that the unemployment rate helps identify the cyclical component of GDP; these results indicate it can also improve the real-time stability of that estimated cycle, despite the additional parameters that must be estimated.

Table 3: Summary statistics of output gap revisions, selected subsamples.

| | 1975:Q1–1997:Q4 | | | | 1998:Q1–2013:Q3 | | | |
|---|---|---|---|---|---|---|---|---|
| | N | NSR (SD) | NSR (RMSE) | % sign agree | N | NSR (SD) | NSR (RMSE) | % sign agree |
| Linear | 91 | 0.57 | 2.31 | 24 | 63 | **0.27** | 0.67 | 48 |
| Quadratic | 91 | 0.75 | 1.68 | 45 | 63 | 0.42 | 0.53 | 95 |
| HP | 91 | 1.11 | 1.10 | 55 | 63 | 0.88 | 0.88 | 65 |
| GLMS | 91 | 0.72 | 0.74 | 75 | 63 | 0.67 | 0.76 | 73 |
| Hamilton | 91 | **0.40** | 0.50 | **93** | 63 | 0.28 | **0.28** | 90 |
| MW | 91 | 1.17 | 1.33 | 51 | 63 | 0.71 | 1.13 | 68 |
| UC (GDP) | 91 | 0.75 | 0.88 | 58 | 63 | 0.53 | 1.07 | 62 |
| UC (GDP/$\pi$) | 91 | 0.72 | 0.74 | 76 | 63 | 0.55 | 1.12 | 63 |
| UC (GDP/CU) | 50 | *0.63* | *0.64* | *72* | 63 | 0.44 | 0.77 | 78 |
| UC (GDP/u.) | 91 | 0.53 | 0.53 | 79 | 63 | 0.29 | 0.64 | 81 |
| UC (GDP/u./$\pi$) | 91 | 0.45 | **0.45** | 85 | 63 | 0.29 | 0.61 | 78 |
| Tealbook | 91 | 1.51 | 1.85 | 77 | 63 | 0.43 | 0.49 | **92** |

Notes: Revision defined as 2013:Q4 output gap estimate less real-time estimate, $C_t^{2013:Q4} - C_t^R$. Bolded entries indicate best performing model according to that metric. Real-time vintages for UC(GDP/CU) begin in 1985:Q2; statistics for UC(GDP/CU) are in italics for the first subperiod because of the shorter period over which they are calculated. First subsample begins in 1975:Q1 because the 2013:Q4 Tealbook output gap estimate begins in that quarter. See text for details.

In Tables 3 and 4 we compute these summary statistics for selected subperiods. Table 3 splits the sample into two roughly equal subperiods, 1975:Q3–1997:Q4 and 1998:Q1–2013:Q3, following Edge and Rudd (2016).[7] We end the sample in 2013 so that we can include the Tealbook output gap estimates. Edge and Rudd (2016) document that the Federal Reserve's output gap estimates become much more stable beginning in the 1990s, and we find that remains true even after including the Great Recession period. The model-based estimates are also more stable in the second half of the sample, but the improvement in stability is not as pronounced as in the case of Tealbook estimates. Among the UC models, the models using unemployment

[7] Edge and Rudd (2016) consider two subsamples, 1980Q1:1992Q4 vs 1994Q1:2006Q4 and 1966Q1:1997Q4 vs 1998Q1:2006Q4. Our main takeaways are robust to splitting our sample in 1993 instead of 1997 and to starting the sample in 1980 instead of 1975.

are again the most stable.

Table 4 measures the stability of the output gap estimates at turning points, presumably the most important periods of time for policymakers. We compute output gap revisions for the 12 quarters prior to NBER peaks and troughs. Table 4 is broadly similar to the previous ones in that the Hamilton-trend implied output gap and the output gap estimated with the UC models using unemployment are very stable at turning points. Interestingly, the Tealbook has produced relatively less stable output gap estimates prior to turning points, especially business cycle peaks.

Table 4: Real-time output gap revisions at turning points.

| | NBER peaks | | | | NBER troughs | | | |
|---|---|---|---|---|---|---|---|---|
| | N | NSR (SD) | NSR (RMSE) | % sign agree | N | NSR (SD) | NSR (RMSE) | % sign agree |
| Linear | 91 | 0.65 | 3.23 | 15 | 91 | 0.41 | 1.75 | 30 |
| Quadratic | 91 | 1.48 | 1.47 | 69 | 91 | 1.00 | 1.00 | 73 |
| HP | 91 | 1.13 | 1.39 | 55 | 91 | 0.75 | 1.08 | 57 |
| GLMS | 91 | 0.86 | 1.18 | 66 | 91 | 0.64 | 0.86 | 67 |
| Hamilton | 91 | 0.61 | **0.64** | **87** | 91 | 0.43 | **0.49** | **90** |
| MW | 91 | 1.16 | 1.15 | 64 | 91 | 0.89 | 0.90 | 74 |
| UC (GDP) | 91 | 0.86 | 2.24 | 49 | 91 | 0.73 | 1.31 | 54 |
| UC (GDP/$\pi$) | 91 | 0.82 | 1.20 | 73 | 91 | 0.76 | 0.84 | 73 |
| UC (GDP/CU) | 39 | *1.34* | *2.65* | *85* | 39 | *0.44* | *0.85* | *85* |
| UC (GDP/u.) | 91 | 0.59 | 1.08 | 78 | 91 | 0.41 | 0.66 | 75 |
| UC (GDP/u./$\pi$) | 91 | **0.54** | 0.99 | 85 | 91 | **0.38** | 0.58 | 79 |
| Tealbook | 65 | *1.99* | *2.47* | *63* | 66 | *0.97* | *1.19* | *80* |

Notes: Revision defined as 2013:Q4 output gap estimate less real-time estimate, $C_t^{2013:Q4} - C_t^R$. Bolded entries indicate best performing model according to that metric. Italicized entries indicate the model has fewer observations. Real-time vintages for UC(GDP/CU) begin in 1985:Q2. There are fewer Tealbook observations because the 2013:Q4 Tealbook output gap estimate begins only in 1975:Q1. See text for details.

# 4 Decomposing output gap revisions

We now turn to understanding why some real-time estimates of the output gap are more stable than others. We propose a decomposition that measures the contribution of each observable series to the instability of the output gap from three distinct components: revisions to the underlying data, parameter instability, and the end-point problem.

### 4.1 Sources of instability

The model parameters from a model estimated on data vintage $v$ and periods $t = 1, ..., \tau$ are denoted $\Theta^{v}_{1:\tau}$. Also assume that $T+1$ is the last available vintage. We can then define the following objects.

- The **final** output gap estimate in period $t$, $C^{F}_{t|T}(\Theta^{T+1}_{1:T})$, is the smoothed estimate of the output gap in period $t$ when the model is estimated using the complete final data vintage, $T+1$.
- The **quasi-final** output gap estimate, $C^{QF}_{t|t}(\Theta^{T+1}_{1:T})$, is the filtered estimate of the output gap, again using parameters estimated from the full sample of the final data vintage.
- The **quasi-real time** estimate, $C^{QR}_{t|t}(\Theta^{T+1}_{1:t})$, is the filtered estimate of the output gap produced using parameter estimates from the final data vintage, but estimated using data only through period $t$.
- The **real-time** output gap estimate, $C^{R}_{t|t}(\Theta^{t+1}_{1:t})$, is the filtered estimate of the gap using all available data from vintage $t+1$.

The revision $C^{F}_{t|T} - C^{R}_{t|t}$ that was used in Section 3 to compute real-time stability statistics can be decomposed:

$$\begin{aligned} C^{F}_{t|T} - C^{R}_{t|t} &= C^{F}_{t|T}(\Theta^{T+1}_{1:T}) - C^{QF}_{t|t}(\Theta^{T+1}_{1:T}) && \text{Effect of } \textbf{\textit{end-point}} \textbf{ problem} \\ &+ C^{QF}_{t|t}(\Theta^{T+1}_{1:T}) - C^{QR}_{t|t}(\Theta^{T+1}_{1:t}) && \text{Effect of } \textbf{parameter instability} \\ &+ C^{QR}_{t|t}(\Theta^{T+1}_{1:t}) - C^{R}_{t|t}(\Theta^{t+1}_{1:t}) && \text{Effect of } \textbf{data revisions} \end{aligned} \tag{1}$$

The difference $C^{F}_{t|T} - C^{QF}_{t|t}$ is due to the *end-point* problem, since we are comparing two sets of estimates obtained with exactly the same parameter estimates and the same data vintages, but data samples of different lengths. The difference $C^{QF}_{t|t} - C^{QR}_{t|t}$ is due to parameter instability since the only difference between the two estimates is whether we use data through $T$ or $t$ to estimate the parameters. Finally, the difference $C^{QR}_{t|t} - C^{R}_{t|t}$ is due to data revisions, since the quasi-real estimate is obtained with the final vintage and the real-time estimate with the real-time vintage, but both estimates are obtained filtering the data through period $t$ and using parameters estimated with data through period $t$.

We can further track the source of each element of the decomposition in (1) into the effects from each observable included in the model. The unobserved states from a model estimated in state-space form are weighted averages of the observables, where the weights are functions of the parameters (Koopman and Harvey 2003). For the purposes of illustration, consider the decomposition of the UC (GDP/u.) model (of

course, the algebra is generalizable). The various estimates of the output gap in period $t$ are:

$$
\begin{aligned}
C_t^F &= \sum_{\tau=1}^{T} [w_\tau^u(\Theta_{1:T}^{T+1})u_\tau + w_\tau^y(\Theta_{1:T}^{T+1})y_\tau] \\
C_t^{QF} &= \sum_{\tau=1}^{t} [w_\tau^u(\Theta_{1:t}^{T+1})u_\tau + w_\tau^y(\Theta_{1:t}^{T+1})y_\tau] \\
C_t^{QR} &= \sum_{\tau=1}^{t} [w_\tau^u(\Theta_{1:t}^{T+1})u_\tau + w_\tau^y(\Theta_{1:t}^{T-1})y_\tau] \\
C_t^{R} &= \sum_{\tau=1}^{t} [w_\tau^u(\Theta_{1:t}^{t+1})u_\tau + w_\tau^y(\Theta_{1:t}^{t+1})y_\tau].
\end{aligned}
\tag{2}
$$

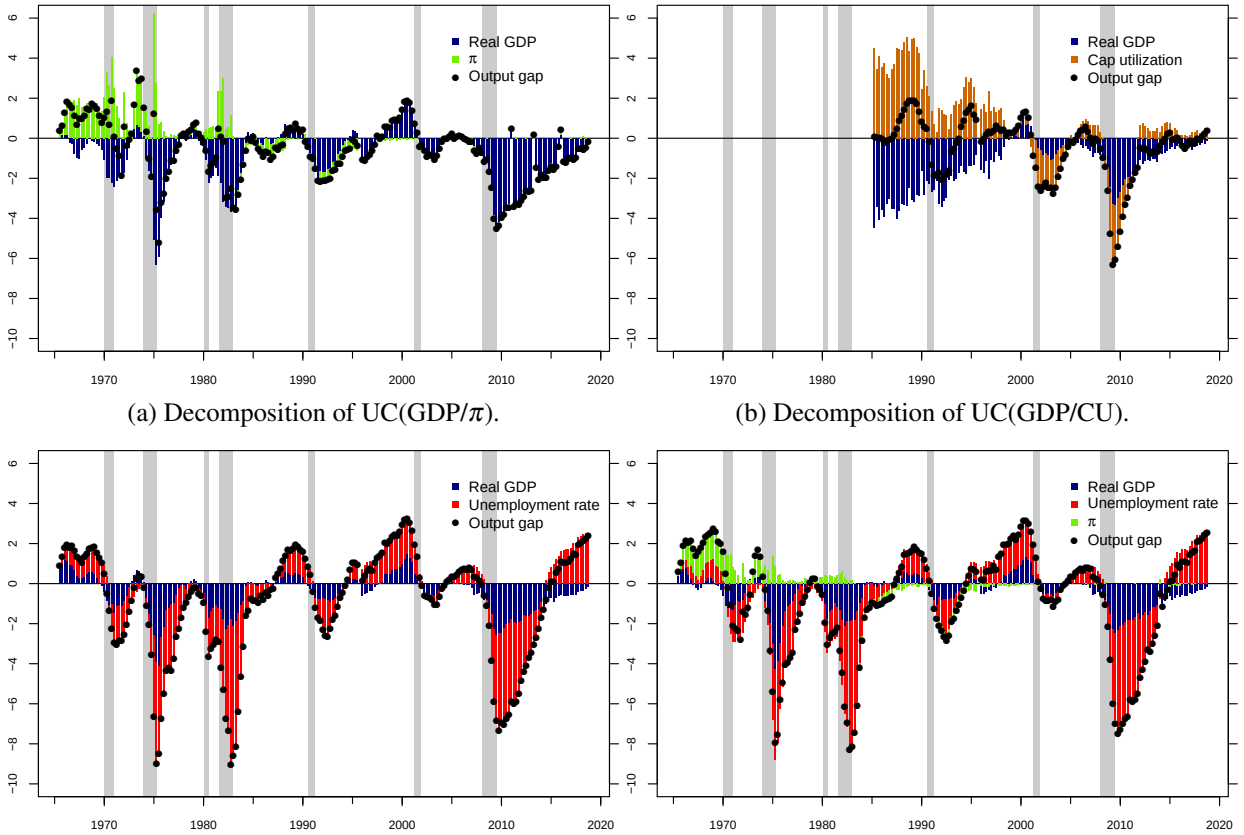
The equations in (2) decompose the level of any given estimate of the output gap. We can now write output gap estimate instability in terms of the observables by implementing a variance decomposition of the output gap revisions. By plugging (2) into (1) and taking the standard deviation, we can decompose the standard deviation of $C_{t|T}^F - C_{t|t}^R$ into the contributions of the *end-point* problem, parameter instability, and data revisions, and decompose each component into the contributions of the observables series.

### 4.2 Decomposition results

To motivate the decomposition of the revisions, Figure 2 plots the sequence of real-time output gap estimates from the multivariate unobserved component models, and decomposes the level of those estimates into the contribution of the observables following equation (2). As shown in the upper-left panel, the UC (GDP/$\pi$) model relies heavily on inflation to make inference about the output gap in the 1960s and early 1970s, but almost exclusively relies on the information coming from real GDP to estimate the output gap since the mid-1990s. After 1980, inflation carries very little information about the output gap estimate in either model that includes it. The UC models that condition on the unemployment rate or capacity utilization heavily rely on them to estimate the cyclical state of the economy. Real GDP also influences the output gap estimate, but its influence is relatively small, especially when compared to the cyclical information in the unemployment rate.

Table 5 decomposes the change in the output gap estimate from the real-time estimate to the final. We decompose the noise-to-signal ratio that uses standard deviations. To do so, we take the standard deviation of both sides of equation (1), and then divide by the standard deviation of the final estimate. The *end-point* problem explains the bulk of the revisions to the output gap for all models. Indeed, it often explains all or

Figure 2: Decomposition of real-time output gap estimates, selected UC models.




(a) Decomposition of UC(GDP/$\pi$).

(b) Decomposition of UC(GDP/CU).

(c) Decomposition of UC(GDP/u.).

(d) Decomposition of UC(GDP/u./$\pi$).

Notes: Figures show decomposition of real-time output estimate from selected UC models. Contribution of observable (in p.p.) shown as shaded bar, which sum to output gap estimate. See text for details.

more than all of the noise-to-signal ratio. Decomposing the variance of equation (1) involves computing the covariances among the three components, which we summarize in the residual at the bottom of Table 5; these covariances turn out to be mostly negative and explain for a few models large portions of the noise-to-signal ratio. For instance, when we compare the GLMS to the HP filter, the improvement in noise-to-signal ratio does not seem to come from a reduction in the *end-point* problem, bur rather a larger negative residual, which is driven by a negative correlation between the *end-point* problem and the data revision components. The contribution of parameter instability is largest for the UC (GDP/$\pi$) model, and especially from inflation.

In previous sections we showed that models that include labor market data produce a more stable output gap in real time. Table 5 shows that the improvement in stability is obtained by reducing the influence of the *end-point* problem. The contribution of data revisions does not become smaller when other observables are added to the univariate model, and the contribution of parameter instability may become larger. For

example, comparing the variability of the UC(GDP) output gap to that from UC(GDP/$\pi$/u.), we see that the variability as measured by the noise-to-signal ratio is more than a third lower. Most of that improvement is due to a reduction in the end-point problem.

Table 5: Decomposing the noise-to-signal ratio into observables.

| | UC (GDP) | HP (GDP) | GLMS (GDP) | UC (GDP/$\pi$) | UC (GDP/CU) | UC (GDP/u.) | UC (GDP/$\pi$/u.) |
|---|---|---|---|---|---|---|---|
| NSR (SD) | 0.7 | 1.0 | 0.7 | 0.7 | 0.7 | 0.5 | 0.4 |
| *Of which is due to:* | | | | | | | |
| Data Revisions | 0.2 | 0.4 | 0.7 | 0.4 | 0.2 | 0.1 | 0.2 |
| GDP | *0.2* | *0.4* | *0.1* | *0.3* | *0.3* | *0.1* | *0.1* |
| $\pi$ | | | | *0.4* | | | *0.1* |
| CU | | | | | *0.3* | | |
| u. | | | | | | *0.1* | *0.2* |
| Parameter instability | 0.1 | 0.0 | 0.0 | 0.4 | 0.1 | 0.1 | 0.3 |
| GDP | *0.1* | *0.0* | *0.0* | *0.3* | *0.5* | *0.1* | *0.1* |
| $\pi$ | | | | *0.5* | | | *0.2* |
| CU | | | | | *0.4* | | |
| u. | | | | | | *0.1* | *0.2* |
| End-point problem | 0.7 | 1.0 | 1.0 | 0.7 | 0.7 | 0.5 | 0.5 |
| GDP | *0.7* | *1.0* | *1.0* | *0.7* | *0.6* | *0.5* | *0.4* |
| $\pi$ | | | | *0.1* | | | *0.0* |
| CU | | | | | *0.3* | | |
| u. | | | | | | *0.3* | *0.2* |
| Residual | -0.3 | -0.3 | -1.0 | -0.8 | -0.3 | -0.3 | -0.5 |

Notes: Table decomposes the standard deviation of the revisions of the output gap estimates and divides it by the standard deviation of the final estimate. Residual is due to covariances among components. All statistics are for 1965:Q3–2018:Q4 except for UC (GDP/CU), which begins in 1985:Q2. See text for details.

# 5 Economic relevance

In addition to being stable in real-time, output gap estimates should also be meaningful: a model that estimates the output gap to be a constant is stable but entirely useless. We examine three practical uses of output gap estimates. First, we compare model output gap estimates to the Federal Reserve Board's judgmental one. We also evaluate whether output gaps are useful in the context of Phillips curve forecasts of inflation, and if they can describe a benchmark interest rate for policymakers.

## 5.1 How do the model estimates compare to Tealbook estimates?

Since Edge and Rudd (2016) show that the Fed's output gap measure is a reliable input to the FOMC's decision-making process, in Table 6, we compare the real-time output gap estimates to the real-time output gap estimate from the Tealbook. Table 7 compares them to the October 2013 Tealbook estimate instead.

With the exception of the linear trend in the more recent sub-sample, model-based output gaps are, on average, lower than the Tealbook estimate, although the standard deviation of that difference is usually large enough that one would not reject a null hypothesis that it is zero. In Section 3, we found that some univariate models produced stable real-time output gap estimates, for example, the output gap from Hamilton was very stable in real-time across different subsamples. However, the Hamilton filter's output gap is very different from the Tealbook's estimate. Indeed, when the final Tealbook estimate is used as "truth," Table 7, the noise-to-signal ratios are higher than one in the first half of the sample and more than twice the magnitude of many of the other models in the second half.

In terms of noise-to-signal ratios, the UC models that include the unemployment rate tend to perform best. Close inspection of Figure 1 shows that UC (GDP/u.) and UC (GDP/u./$\pi$) are both very close to the Tealbook output gap estimate starting from the late 1980s. Prior to that, the two model estimates are quite similar, but the model that includes inflation has a slightly higher output gap estimate throughout the 1970s.

Table 6: Evaluating model-based estimates relative to real-time Tealbook estimates.

| | 1975:Q1–1997:Q4 | | | | | 1998:Q1–2013:Q3 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | N | Mean | SD | NSR (SD) | NSR (RMSE) | N | Mean | SD | NSR (SD) | NSR (RMSE) |
| Linear | 91 | **-0.03** | 3.15 | 0.72 | **0.72** | 63 | 2.75 | 2.57 | 0.83 | 1.22 |
| Quadratic | 91 | -5.78 | 3.15 | 0.72 | 1.51 | 63 | -1.87 | 1.66 | 0.54 | 0.81 |
| HP | 91 | -3.84 | 4.20 | 0.96 | 1.30 | 63 | -1.73 | 3.07 | 1.00 | 1.14 |
| GLMS | 91 | -3.49 | 3.82 | 0.88 | 1.18 | 63 | -1.31 | 2.39 | 0.77 | 0.88 |
| Hamilton | 91 | -4.76 | 4.11 | 0.94 | 1.44 | 63 | -2.04 | 2.33 | 0.76 | 1.00 |
| MW | 91 | -4.82 | 4.36 | 1.00 | 1.49 | 63 | -3.33 | 3.05 | 0.99 | 1.46 |
| UC (GDP) | 91 | -3.00 | 3.38 | 0.78 | 1.03 | 63 | -0.73 | 1.60 | 0.52 | 0.57 |
| UC (GDP/$\pi$) | 91 | -3.07 | 3.78 | 0.87 | 1.11 | 63 | -0.90 | 1.75 | 0.57 | 0.64 |
| UC (GDP/CU) | 50 | *-1.31* | *1.62* | *0.72* | *0.93* | 63 | **-0.56** | 1.93 | 0.63 | 0.65 |
| UC (GDP/u.) | 91 | -2.38 | **2.08** | **0.48** | **0.72** | 63 | -0.69 | 0.73 | 0.24 | 0.33 |
| UC (GDP/u./$\pi$) | 91 | -2.47 | 2.31 | 0.53 | 0.77 | 63 | -0.64 | **0.72** | **0.23** | **0.31** |

Notes: Table shows summary statistics for difference between real-time model-based output gap estimate relative to real-time Tealbook estimate, $C_t^{R,TB} - C_t^{R,i}$. Bolded entries denote best performing model according to each metric. Real-time estimates from UC (GDP/CU) begin in 1985:Q2; shortened sample indicated with italics. See text for details.

Table 7: Evaluating model-based estimates relative to October 2013 Tealbook.

| | 1975:Q1–1997:Q4 | | | | | 1998:Q1–2013:Q3 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | N | Mean | SD | NSR (SD) | NSR (RMSE) | N | Mean | SD | NSR (SD) | NSR (RMSE) |
| Linear | 91 | 2.41 | **1.33** | **0.59** | 1.22 | 63 | 3.34 | 2.65 | 1.09 | 1.76 |
| Quadratic | 91 | -3.33 | 1.49 | 0.66 | 1.61 | 63 | -1.28 | 1.95 | 0.81 | 0.96 |
| HP | 91 | -1.39 | 2.33 | 1.03 | 1.20 | 63 | -1.14 | 2.38 | 0.98 | 1.08 |
| GLMS | 91 | -1.04 | 1.69 | 0.75 | 0.88 | 63 | -0.72 | 1.72 | 0.71 | 0.77 |
| Hamilton | 91 | -2.32 | 2.46 | 1.09 | 1.49 | 63 | -1.45 | 1.77 | 0.73 | 0.94 |
| MW | 91 | -2.38 | 2.64 | 1.17 | 1.57 | 63 | -2.74 | 2.34 | 0.97 | 1.48 |
| UC (GDP) | 91 | -0.56 | 1.57 | 0.70 | 0.73 | 63 | -0.14 | **0.97** | **0.40** | **0.40** |
| UC (GDP/$\pi$) | 91 | -0.62 | 1.66 | 0.73 | 0.78 | 63 | -0.31 | 1.21 | 0.50 | 0.51 |
| UC (GDP/CU) | 50 | *-1.24* | *1.06* | *0.73* | *1.12* | 63 | **0.03** | 1.41 | 0.58 | 0.58 |
| UC (GDP/u.) | 91 | 0.07 | 1.74 | 0.77 | 0.77 | 63 | -0.10 | 1.12 | 0.46 | 0.46 |
| UC (GDP/u./$\pi$) | 91 | **-0.03** | 1.50 | 0.66 | **0.66** | 63 | -0.05 | 1.13 | 0.47 | 0.46 |
| Tealbook | 91 | 2.44 | 3.41 | 1.51 | 1.85 | 63 | 0.59 | 1.05 | 0.43 | 0.49 |

Notes: Table shows summary statistics for difference between real-time model-based output gap estimate relative to real-time Tealbook estimate, $C_t^{F,TB} - C_t^{R,i}$. Bolded entries denote best performing model according to each metric. Real-time estimates from UC (GDP/CU) begin in 1985:Q2; shortened sample indicated with italics. See text for details.

## 5.2 Forecasting inflation

We now consider the importance of real-time instability to predictions of inflation. We fit Phillips curve models to each output gap estimate that we consider:

$$\pi_t^{t+h} = \alpha + \sum_{i=1}^{6} \beta_i \pi_{t-i} + \gamma C_{t-1}^i + e_t \tag{3}$$

where $\pi_t = 400 \times \log(P_t / P_{t-1})$ is the annualized growth rate of the core PCE price index and $\pi_t^{t+h}$ is the average inflation rate from quarter $t$ to $t+h$; $C_{t-1}^i$ is the real-time output gap estimate from model $i$ that is available in quarter $t$. We constrain the coefficients on lagged inflation to sum to one and estimate equation (3) via ordinary least squares. We forecast three horizons: two, four, and eight quarters ahead. For each horizon-model pair, we consider the sequence of real-time, quasi-real time, and the 2013:Q4 output gap estimate. Because we are interested in isolating the impact of the output gap when forecasting inflation, we use the current vintage of inflation data. To keep the estimation sample the same for all models, our first out-of-sample forecast is made using models estimated from 1985:Q2, the first available real-time estimate from the UC (GDP/CU) model, and the first out-of-sample forecast is produced for the first quarter of 1990. After that, we expand the window used to estimate (3) until our final forecast quarter, 2013:Q3.

Table 8 shows the ratio of the RMSE for each model relative to the RMSE from the model that uses the October 2013 Tealbook output gap. The raw RMSE for the October 2013 Tealbook is shown in bolded text. For other models, values less than one indicate superior inflation forecasts relative to the null model. We test for superior forecast ability using a Diebold and Mariano (1995) test and using Newey-West standard errors.

Table 8: Phillips curve based forecasts of core PCE inflation, 1990:Q1–2013:Q3.

| | Two quarters ahead | | | Four quarters ahead | | | Eight quarters ahead | | |
|---|---|---|---|---|---|---|---|---|---|
| | RT | QRT | 2013:Q4 | RT | QRT | 2013:Q4 | RT | QRT | 2013:Q4 |
| Linear | 1.02 | 1.03 | 1.02 | 1.02 | 1.02 | 1.03 | 1.01 | 1.02 | 1.02 |
| Quadratic | 1.05 | 1.04 | 1.01 | 1.04 | 0.97 | 1.00 | 1.03 | 0.99 | 1.00 |
| HP | 0.99 | 0.96 | 0.98 | 0.88 | 0.89 | 0.91 | 0.86 | 0.93 | 0.93 |
| GLMS | 1.05 | 1.04 | 0.98 | 0.96 | 0.95 | 0.91 | 0.92 | 0.92 | 0.93 |
| Hamilton | 1.01 | 0.97 | 0.95 | 0.91 | 0.92 | 0.87 | 0.89 | 0.97 | 0.88 |
| MW | 1.07 | 1.05 | 1.04 | 1.03 | 1.01 | 0.98 | 0.99 | 0.98 | 1.02 |
| UC (GDP) | 1.08 | 1.07 | 1.01 | 1.05 | 1.01 | 1.00 | 1.01 | 0.99 | 1.00 |
| UC (GDP/$\pi$) | 1.02 | 1.07 | 1.03 | 1.00 | 1.03 | 1.00 | 0.98 | 0.99 | 0.98 |
| UC (GDP/CU) | 1.05 | 1.07 | 1.00 | 1.05 | 1.06 | 1.01 | 1.06 | 1.06 | 1.02 |
| UC (GDP/u.) | 1.07 | 1.07 | 1.02 | 1.06 | 1.06 | 1.02 | 1.05 | 1.05 | 1.03 |
| UC (GDP/u./$\pi$) | 1.06 | 1.07 | 1.03 | 1.05 | 1.07 | 1.03 | 1.05 | 1.05 | 1.03 |
| Tealbook | 1.04 | – | **0.46** | 1.02 | – | **0.44** | 1.04 | – | **0.46** |

Notes: Absolute RMSE of inflation forecast from equation 3 using October 2013 Tealbook output gap estimate shown as bolded values. All other entires are the ratio of that RMSE to the bolded value. Out-of-sample period is 1990:Q1–2013:Q3 and models are estimated beginning in 1985:Q2 using an expanding window. Asterisks indicate dominance of indicated model over the null Tealbook estimate is statistically significant at 10 ($^{*}$) or 5 ($^{**}$) percent level and with Newey-West standard errors. See text for details.

As has been shown in other empirical applications—e.g., Stock and Watson (2009), Berge (2018), and Dotsey, Fujita and Stark (2018), among many others—no single output gap estimate produces clearly superior Phillips curve based inflation forecasts. Table 8 shows a period wherein inflation has been quiescent (Coibion and Gorodnichenko 2015). In this period, although some output gap estimates may slightly outperform the null Tealbook output gap-based forecast, the improvement is never statistically meaningful. However, when we evaluate other out-of-sample periods, and if we substitute total PCE price inflation for core, we find that no output gap estimate clearly dominates the others. (See Appendix A3.)

Unsurprisingly, focusing on each model's individual results, the best inflation forecast is produced when using the 2013:Q4 output gap estimate. Forecast performance tends to deteriorate when real-time or quasi-real-time estimates are used to forecast inflation instead. The reduction in forecast ability, however, is rather small. For example, when we focus on the four-quarter ahead forecasts, the deterioration in forecast ability from using real-time output gap estimates to the 2013:Q4 output gap estimate is around 2-3 percentage

points relative to the Tealbook estimate, and in fact, for some of the univariate filters, the real-time estimate actually produces superior inflation forecasts to the 2013:Q4 estimate.

### 5.3 Output gaps as input to policy

Lastly, we evaluate the implications of the different output gap estimates for monetary policy. Because policy is not set in a rules-based manner, and because it is likely that policymakers implicitly consider an output gap that may be described as some combination of the output gaps presented above, comparing the Taylor-rule implied interest rates to the actual federal funds rate is a transparent and easy-to-understand way to understand the relative performance of the output gaps. For each output gap, we compute the short-term interest rate prescribed by the Taylor (1993) rule:

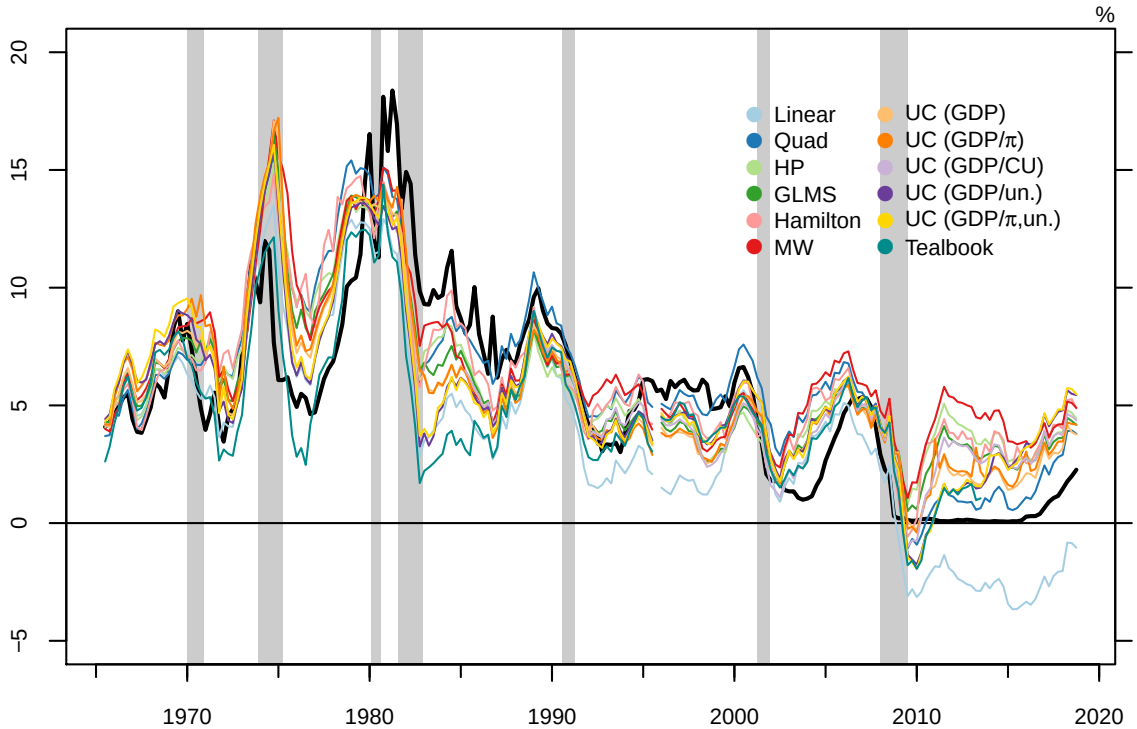
$$r_t = r_t^* + \pi_t + w_C C_t + w_\pi(\pi - \pi_t^*), \tag{4}$$

where $r_t^*$ is the neutral interest rate, $\pi_t$ is percent change in inflation from one year ago, $y_t$ is the output gap, and $\pi_t^*$ is the trend rate of inflation. In the 1993 paper, Taylor sets $w_C = w_\pi = 1/2$, and assumes that $r_t^*$ and $\pi_t^*$ are time-invariant and equal to two percent. Results using a "balanced" Taylor rule with $w_C = 1$; $w_\pi = 1/2$ are presented in Section A4 of the Appendix.

Figure 3 plots the Taylor-rule implied policy rates using our various real-time output gap estimates. The Taylor rule uses time-invariant estimates of $\pi^*$ and $r^*$, so that the prescribed interest rates tend to be below realized federal funds rate in the first half of the sample and above it in the second half. However, the pattern of Taylor-rule-implied interest rates broadly follow that of the realized effective funds rate—given the pattern of inflation, the rules prescribe high interest rates in the 1970s followed by a steady move lower. The prescribed interest rates also tend to move lower during or immediately following NBER-defined recessions. Each of the output gaps imply notable downward movements to the Federal funds rate in the 1990, 2001 and 2007-2008 recessions. Notably, and in contrast to the actual policy rate, some output gaps imply an increase in the Federal funds rate at the end of the Great Recession. For example, the Mueller-Watson output gap implies a minimum value of the federal funds rate of roughly 1 percent in 2009:Q3 before increasing to 5¾ percent in 2011:Q3. The interest rate implied by the Tealbook's output gap estimate increases much more gradually after the Great Recession, and is just 1.5 percent in the final available period, 2013:Q3.

Table 9 calculates the root-mean-square error for each Taylor (1993) rule implied interest rate relative to

Figure 3: Taylor (1993) rule implied federal funds rate, 1965:Q3–2018:Q4.




Notes: Figure shows prescribed interest rate from Taylor 1993 rule, calculated as in equation 4 with $w_\pi = w_C = 1/2$, $\pi$ equal to the four quarter change in the real-time GDP price deflator, $y_t$ as the real-time output gap estimate, and $\pi^* = r^* = 2$ percent. Effective federal funds rate shown as thick black line. See text for details.

the effective federal funds rate, again for the full sample and a roughly equal split sample. As before, we test for statistically meaningful differences relative to the interest rate implied by the October 2013 Tealbook output gap estimate. The real-time Tealbook output gap produces an implied interest rate that is closest to policy in the most recent period; in fact, the real-time output gaps imply an interest rate path that is closer to actual policy than the output gap estimate from 2013:Q4. However, in the first half of the sample, the interest rate implied by Tealbook is among the least similar to actual policy.

No model clearly outperforms the others in terms of delivering an interest rate path that matches actual policy. Many of the models that were relatively stable in real time would have implied interest rate policies that were ultimately ignored by policymakers. Again consider the Hamilton model as an example, since it delivers stable output gaps in real time. Yet in late 2010, this model would have implied an interest rate policy more than 300 basis points higher than the actual federal funds rate, and nearly 350 basis points higher than the policy implied by the Tealbook's output gap. Similarly, the linear detrending of real GDP

resulted in relatively stable real-time output gap estimates, but would have prescribed sizable and persistent negative policy rates through at least 2018.

Table 9: RMSE of prescribed interest rate from Taylor (1993).

| | 1965:Q3–2013:Q3 | | | 1975:Q1–1997:Q4 | | | 1998:Q1–2013:Q3 | | |
|---|---|---|---|---|---|---|---|---|---|
| | RT | QRT | 2013:Q4 | RT | QRT | 2013:Q4 | RT | QRT | 2013:Q4 |
| Linear | 2.84 | 2.66 | 2.86 | 3.56 | 3.34 | 3.11 | 2.41 | 2.17 | 2.22 |
| Quadratic | 2.28 | 2.32 | 3.14 | 2.79 | 2.88 | 3.44 | 1.68 | 1.59 | 2.51 |
| HP | 2.58 | 2.61 | 2.59 | 2.83 | 2.89 | 2.88 | 2.55 | 2.49 | 2.27 |
| GLMS | 2.49 | 2.47 | 2.57 | 2.83 | 2.83 | 2.88 | 2.23 | 2.22 | 2.23 |
| Hamilton | 2.56 | 2.59 | 2.53 | 2.86 | 2.94 | 2.92 | 2.44 | 2.36 | 2.24 |
| MW | 2.90 | 2.88 | 2.56 | 2.88 | 2.88 | 2.77 | 3.11 | 3.04 | 2.33 |
| UC (GDP) | 2.34 | 2.38 | 3.17 | 2.83* | 2.87 | 3.20 | 1.76 | 1.74 | 2.66 |
| UC (GDP/$\pi$) | 2.51 | 2.48 | 2.75 | 2.94* | 2.84* | 3.00 | 1.90 | 1.89 | 2.35 |
| UC (GDP/CU) | – | 2.46 | 2.74 | – | 2.90** | 3.08 | 2.08 | 1.98 | 2.08 |
| UC (GDP/u.) | 2.47 | 2.51 | 2.90 | 2.92 | 2.98 | 3.24 | 2.02 | 2.00 | 2.24 |
| UC (GDP/u./$\pi$) | 2.51 | 2.55 | 2.89 | 2.92* | 2.99 | 3.23 | 2.04 | 2.00 | 2.18 |
| Tealbook | 2.59 | – | 2.67 | 3.49 | – | 3.23 | 1.27 | – | 1.54 |

Notes: Table shows RMSE of implied interest rate from Taylor (1993) rule calculated using each output gap estimate relative to actual effective federal funds rate. UC (GDP/CU) is excluded from the real-time evaluation in the first two sub-samples due to shortened sample availability. Asterisks denote that DMW test statistic for dominance of indicated model over the null October 2013 Tealbook output gap is statistically significant at 10 (*) or 5 (**) percent level. See text for details.

# 6 Conclusions

We consider several statistical models of the U.S. economy and evaluate their estimates of the output gap along several dimensions. First, we document that output gap estimates can be wide ranging and that real-time estimates from these models can be heavily revised after the fact. We propose a statistical decomposition of output gap revisions, and find that the *end-point* problem is the primary reason that univariate models fail to achieve real-time stability. The finding suggests that methods aimed at reducing the impact of the *end-point* problem may be particularly useful to policymakers.

While researchers such as Garratt et al. (2008) or Clements and Galvão (2012) have tackled this problem explicitly, we find another possible avenue: the use of labor market information when producing output gap estimates. The models we consider that contain an Okun's law relationship are able to provide an estimate of the output gap that is relatively stable in real time, in spite of the fact that they contain more parameters to be estimated. These output gap estimates also have useful economic properties. Models that include

the unemployment rate are consistently among the best-in-class of the models we consider in terms of their usefulness as a harbinger of future inflation and as a gauge for the stance of monetary policy. And since their output gap estimates are relatively stable, this performance does not substantially deteriorate when used in real time relative to when used *ex-post*.

Finally, the real-time stability of the output gap—as was emphasized by Orphanides and van Norden (2002)—is not the only metric by which to judge output gap estimates. No output gap estimate we consider is clearly superior to the others along each metric. While the Mueller-Watson gap produced relatively accurate forecasts between 1984 and 2013, when used in a Taylor rule, it also prescribed a Federal Funds rate of more than 5 percent only a few years after the end of the Great Recession. Similarly, while the Garratt et al. (2008) procedure does reduce the *end-point* problem inherent when filtering in real time, the output gap produced by that procedure is often outperformed in terms of economic significance. Of course, at the time of writing the U.S. economy has been subject to an unprecedented economic and social trauma: the outbreak of the COVID-19 in late 2019 and 2020. It will be interesting to continue to monitor the real-time behavior of these various output gap estimates and their usefulness to policymakers.

## References

**Aastveit, Knut Are and Tørres Trovik**, "Estimating the output gap in real time: A factor model approach," *The Quarterly Review of Economics and Finance*, 2014, *54* (2), 180 – 193.

**Barigozzi, Matteo and Matteo Luciani**, "Measuring US Aggregate Output and Output Gap Using Large Datasets," *SSRN*, 2018.

**Berge, Travis J.**, "Understanding survey-based inflation expectations," *International Journal of Forecasting*, 2018, *34* (4), 788–801.

____ , "Time-varying uncertainty of the Federal Reserve's output gap estimate," Finance and Economics Discussion Series 2020-012, Board of Governors of the Federal Reserve System, Washington, D.C. February 2020.

**Clark, Peter K.**, "The Cyclical Component of U.S. Economic Activity," *Quarterly Journal of Economics*, 1987, *102* (4), 797–814.

**Clark, Todd E. and Taeyoung Doh**, "Evaluating Alternative Models of Trend Inflation," *International Journal of Forecasting*, 2014, *30* (3), 426–48.

**Clements, Michael P. and Ana Beatriz Galvão**, "Improving Real-Time Estimates of Output and Inflation Gaps With Multiple-Vintage Models," *Journal of Business & Economic Statistics*, 2012, *30* (4), 554–562.

**Cleve, Luke Van and Laforte, Jean-Philippe, and Stella, Andrea**, "Real-time Historical Estimates of the Output Gap," FEDS Notes, Board of Governors of the Federal Reserve System, Washington, D.C. 2019.

**Coibion, Olivier and Yuriy Gorodnichenko**, "Is the Phillips Curve Alive and Well after All? Inflation Expectations and the Missing Disinflation," *American Economic Journal: Macroeconomics*, January 2015, *7* (1), 197–232.

**Croushore, Dean and Tom Stark**, "A Real-Time Data Set for Macroeconomists," *Journal of Econometrics*, 2001, *105* (1), 111–130.

**Diebold, Francis X and Roberto S Mariano**, "Comparing Predictive Accuracy," *Journal of Business & Economic Statistics*, July 1995, *13* (3), 253–263.

**Dotsey, Michael, Shigeru Fujita, and Tom Stark**, "Do Phillips Curves Conditionally Help to Forecast Inflation?," *International Journal of Central Banking*, September 2018, *14* (4), 43–92.

**Durbin, J. and S.J. Koopman**, "A simple and efficient simulation smoother for state space time series analysis," *Biometrika*, August 2002, *89* (3), 603–616.

**Edge, Rochelle M. and Jeremy B. Rudd**, "Real-time properties of the Federal Reserve's output gap," *The Review of Economics and Statistics*, 2016, *98* (4), 785–91.

**Fleischman, Charles A. and John M. Roberts**, "From Many Series, One Cycle: Improved Estimates of the Business Cycle from A Multivariate Unobserved Components Model," Finance and Economics

Discussion Series 2011-46, Board of Governors of the Federal Reserve System, Washington, D.C. 2011.

**Garratt, Anthony, Kevin Lee, Emi Mise, and Kalvinder Shields**, "Real-time Representations of the output gap," *The Review of Economics and Statistics*, 2008, *90* (4), 792–804.

**Gonzalez-Astudillo, Manuel and John M. Roberts**, "When Can Trend-Cycle Decompositions Be Trusted?," Finance and Economics Discussion Series 2016-99, Board of Governors of the Federal Reserve System, Washington, D.C. 2016.

**Hamilton, James D.**, "Why You Should Never Use the Hodrick-Prescott Filter," *The Review of Economics and Statistics*, 2018.

**Jarociński, Marek**, "A note on implementing the Durbin and Koopman simulation smoother," *Computational Statistics & Data Analysis*, 2015, *91* (C), 1–3.

**Kim, Chang-Jin and Charles R. Nelson**, *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications*, MIT Press, 1999.

**——, Pym Manopimoke, and Charles R. Nelson**, "Trend inflation and the Nature of Structural Breaks in the New Keynesian Phillips Curve," *Journal of Money, Credit and Banking*, 2014, *46* (2-3), 253–66.

**Koopman, Siem Jan and Andrew Harvey**, "Computing observation weights for signal extraction and filtering," *Journal of Economic Dynamics & Control*, 2003, *27*, 1317–1333.

**Kuttner, Kenneth**, "Estimating Potential Output as a Latent Variable," *Journal of Business and Economic Statistics*, 1994, *12* (3), 361–36.

**Morley, James and Benjamin Wong**, "Estimating and Accounting for the Output Gap with Large Bayesian Vector Autoregressions," Working Papers 2018-04, University of Sydney, School of Economics September 2019.

**Mueller, U. and M. Watson**, "Low-Frequency Econometrics," *Advances in Economics and Econometrics: Eleventh World Congress of the Econometric Society*, 2017, *2*, 53–94.

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

**Planas, Christophe and Alessandro Rossi**, "Can inflation data improve the real-time reliability of output gap estimates?," *Journal of Applied Econometrics*, 2004, *19*, 121–33.

**Quast, Josefine and Maik H. Wolters**, "Reliable real-time output gap estimates based on a modified Hamilton filter," IMFS Working Paper Series 133, Goethe University Frankfurt, Institute for Monetary and Financial Stability (IMFS) 2019.

**Staiger, Douglas, James H. Stock, and Mark W. Watson**, "The NAIRU, Unemployment and Monetary Policy," *The Journal of Economic Perspectives*, 1997, *11* (1), 33–49.

**Stock, James and Mark W. Watson**, "Phillips Curve Inflation Forecasts," Technical Report 2009.

**Stock, James H. and Mark W. Watson**, "Why Has U.S. Inflation Become Harder to Forecast?," *Journal of Money, Credit and Banking*, 2007, *Supplement to Vol. 39* (1), 3–33.

**Taylor, John B.**, "Discretion versus policy rules in practice," *Carnegie-Rochester Conference Series on Public Policy*, 1993, *39*, 195–214.

**Trimbur, Thomas M.**, "Improving Real-Time Estimates of the Output Gap," Finance and Economics Discussion Series 2009-32, Board of Governors of the Federal Reserve System, Washington, D.C. 2009.

# Appendix

## A1 Data details

Table A1: Description of real-time data vintages

| Variable | First real-time vintage | Final real-time vintage | Source |
|---|---|---|---|
| Real GDP | 1965:Q4 | 2019:Q1 | FRB-Ph RT |
| GDP deflator | 1965:Q4 | 2019:Q1 | FRB-Ph RT |
| Unemployment rate | 1965:Q4 | 2019:Q1 | FRB-Ph RT |
| Man. capacity utilization | 1985:Q3 | 2019:Q1 | FRB-Ph RT |
| Tealbook gap | 1975:Q1 | 2013:Q4 | FRB-Ph GB |

Notes: FRB-Ph RT denotes real-time dataset from the Federal Reserve Bank of Philadelphia, available at `https://www.philadelphiafed.org/research-and-data/real-time-center/real-time-data`. FRB-Ph GB denotes Greenbook/Tealbook dataset from Federal Reserve Bank of Philadelphia, available at `https://www.philadelphiafed.org/research-and-data/real-time-center/real-time-data/greenbook`. The Philadelphia Fed Greenbook data has real-time output gap estimates since 1975:Q1. To this data we append earlier Federal Reserve staff nowcasts beginning in 1950:Q4, obtained from Edge and Rudd (2016).

## A2 Estimation details

### A2.1 Overview of Estimation

The UC models can be cast as linear Gaussian state-space models, the estimation of which can be achieved using standard methods. Here we outline the estimation procedure.

Group the observables into the vector $Y_t$ and the unobserved states into the vector $s_t$. For example, the UC(3) model would have $Y_t = (y_t, u_t, \pi_t)'$ and $s_t = (y_t^*, \tau_t^*, u_t^*, \pi_t^*, C_t, C_{t-1})'$. The model is then compactly written in state-space form:

$$Y_t = F s_t + e_t; \quad e_t \sim \mathrm{N}(0, \Sigma) \tag{1}$$

$$s_t = G s_{t-1} + \varepsilon_t; \quad \varepsilon_t \sim \mathrm{N}(0, \Omega) \tag{2}$$

Measurement errors are uncorrelated with each other as well as with the shocks to the states.

We estimate the model using Bayesian methods and using a Gibbs sampler. The intuition of the sampler is straightforward: conditional on the data and states, equations 1 and 2 are a set of independent linear, Gaus-

sian regressions for which parameters can be sampled from their known posterior distribution. Conditional on the model's parameters, the states can be drawn using the Kalman filter. Thus, we estimate the model using a Gibbs sampler alternates between draws of model parameters that are conditioned on the states and draws of the states that are conditioned on model parameters.

Each step of the sampler is standard; see, e.g., Kim and Nelson (1999). Our priors are conjugate so that all the posteriors are known in closed form. The algorithm consists the following five steps:

1. Sample $F$. Conditional on the data, states and $\Sigma$, draw the slope terms in equation 1. Because $\Sigma$ is diagonal, we do this equation-by-equation.

2. Sample $\Sigma$. Conditional on the data, states, and $F$, draw the variances of the measurement errors.

3. Sample $G$. Conditional on states and $\Omega$, draw the AR parameters of the cycle. We impose stationarity of the autoregressive process via rejection sampling.

4. Sample $\Omega$. Conditional on the data ($y^T$), states ($s^T$) and $G$, draw the variances of the shocks to the states.

5. Sample the states. Conditional on the data and parameters, we draw the states using the Durbin and Koopman (2002) and as implemented in Jarociński (2015).

We produce 12,000 draws from this sampler, ignoring the first 2,000 and using a thinning factor of 10 for a total of 1,000 draws each model's posterior distribution.

### A2.2 Priors

We set our priors in the following data-driven manner. As described in the text, in general our priors follow the typical Normal-Inverse Gaussian scheme. Our priors are set for each vintage as follows.

- For the priors of the variance for innovations to trends (e.g., $\varepsilon_{y^*}$, $\varepsilon_{u^*}$, $\varepsilon_{\pi^*}$), we set the scale parameter of the distribution, which controls the tightness of the prior, to be $T^{vintage}/20$, a somewhat informative distribution. We then HP filter a pre-sample (1947–1959) of the vintage data. The shape parameter is then set so that the prior distribution is centered about the HP-filter-implied trend variance.

- We follow an analogous procedure for setting the prior for the variance of the innovations to the drift in trend output, $\varepsilon_\tau$.

- The prior for the variance of the innovations to the cycle is centered about the variance of the pre-sample's HP-filter implied cycle, and with scale parameter equal to one, a much less informative prior.

- Measurement error innovations are very loosely specified, assumed to be IG(1,1).

- Our priors for "slope coefficients" are Gaussian and not tightly specified.

  - Reflecting our belief that the cycle is quite persistent, our priors for the AR coefficients of the cycle, $\phi_1$ $\phi_2$ are centered at (3/2, -2/3)' with variance-covariance matrix equal to the identity matrix.

  - The $\alpha$ parameters are given economically sensible priors with variance-covariance matrix equal to the identity. For inflation, the mean of $\alpha_0$ and $\alpha_1$ are set to 1/10 and 1/20. When the observable is unemployment, they are centered about -½ and -¼, and for capacity utilization they are centered at 0 and much more diffuse, with variance-covariance matrix equal to ten times the identity.

Table A2 shows the priors and posteriors for the 2019:Q1 vintage of the UC (GDP/u./$\pi$) model.

Table A2: Priors and posterior distribution of UC (GDP/u./$\pi$) model.

| | Priors | | | Posteriors | | |
|---|---|---|---|---|---|---|
| Parameter | 5% | Mean | 95% | 5% | Mean | 95% |
| $\alpha_0$ | -2.14 | -.5 | 1.14 | -.65 | -.41 | -.23 |
| $\alpha_1$ | -1.89 | -.25 | 1.39 | -.14 | .01 | .15 |
| $\beta_0$ | -1.54 | .10 | 1.74 | -.32 | -.03 | .27 |
| $\beta_1$ | -1.59 | .05 | 1.69 | -.07 | .22 | .49 |
| $\phi_1$ | -.14 | 1.50 | 3.14 | 1.54 | 1.68 | 1.81 |
| $\phi_2$ | -2.31 | -.67 | .98 | -.84 | -.72 | -.58 |
| $\sigma_{y^*}$ | .03 | .24 | .44 | .20 | .34 | .57 |
| $\sigma_\tau$ | .00 | .02 | .05 | .02 | .02 | .03 |
| $\sigma_{u^*}$ | .00 | .07 | .16 | .13 | .21 | .30 |
| $\sigma_{\pi^*}$ | .00 | .08 | .19 | .43 | .52 | .65 |
| $\sigma_C$ | .29 | 1.08 | 1.82 | .35 | .45 | .56 |
| $\sigma_{e,y}$ | .23 | 1.00 | 1.73 | .23 | .29 | .36 |
| $\sigma_{e,u}$ | .23 | 1.00 | 1.73 | .13 | .15 | .17 |
| $\sigma_{\pi,u}$ | .23 | 1.00 | 1.73 | .56 | .65 | .73 |

Notes: This table shows the priors and posteriors of UC(GDP/u./$\pi$) model estimated on the for 2019:Q1 vintage. $\sigma$ denotes standard deviation of shock process.

## A3 Alternative inflation forecast experiments

Table A3: Phillips curve based forecasts of total PCE inflation, 1990:Q1–2013:Q3.

| | Two quarters ahead | | | Four quarters ahead | | | Eight quarters ahead | | |
|---|---|---|---|---|---|---|---|---|---|
| | RT | QRT | 2013:Q4 | RT | QRT | 2013:Q4 | RT | QRT | 2013:Q4 |
| Linear | 1.01 | 0.99 | 0.98 | 1.00 | 0.98 | 0.98 | 1.00 | 0.96 | 0.98 |
| Quadratic | 1.01 | 0.96 | 1.04 | 1.00 | 0.94 | 1.04 | 0.98 | 0.95 | 1.04 |
| HP | 0.97 | 0.95 | 0.99 | 0.92 | 0.92 | 0.97 | 0.89 | 0.90 | 0.98 |
| GLMS | 0.97 | 0.97 | 0.99 | 0.93 | 0.93 | 0.97 | 0.92 | 0.93 | 0.99 |
| Hamilton | 0.98 | 0.94 | 0.94 | 0.94 | 0.94 | 0.92 | 0.90 | 0.93 | 0.89 |
| MW | 1.03 | 1.00 | 1.02 | 1.02 | 0.98 | 1.00 | 0.99 | 0.95 | 1.01 |
| UC (GDP) | 1.02 | 1.00 | 1.03 | 1.01 | 0.99 | 1.02 | 0.98 | 0.97 | 1.02 |
| UC (GDP/$\pi$) | 1.01 | 0.99 | 1.02 | 1.01 | 0.97 | 1.01 | 1.00 | 0.97 | 1.00 |
| UC (GDP/CU) | 1.03 | 1.01 | 1.01 | 1.04 | 1.02 | 1.02 | 1.04 | 1.02 | 1.03 |
| UC (GDP/u.) | 1.05 | 1.04 | 1.04 | 1.05 | 1.05 | 1.05 | 1.04 | 1.04 | 1.05 |
| UC (GDP/u./$\pi$) | 1.05 | 1.04 | 1.04 | 1.05 | 1.04 | 1.05 | 1.04 | 1.03 | 1.05 |
| Tealbook | 1.04 | – | **1.20** | 1.03 | – | **1.00** | 1.04 | – | **0.88** |

Notes: Absolute RMSE of inflation forecast from equation 3 using October 2013 Tealbook output gap estimate shown as bolded values. All other entires are the ratio of that RMSE to the bolded value. Out-of-sample period is 1990:Q1–2013:Q3 and models are estimated beginning in 1985:Q2 using an expanding scheme. Asterisks denote that DM test statistic for dominance of indicated model over the null Tealbook estimate is statistically significant at 10 (*) or 5 (**) percent level. See text for details.

Table A4: Phillips curve based forecasts of core PCE inflation, 1984:Q1–2013:Q3.

| | Two quarters ahead | | | Four quarters ahead | | | Eight quarters ahead | | |
|---|---|---|---|---|---|---|---|---|---|
| | RT | QRT | 2013:Q4 | RT | QRT | 2013:Q4 | RT | QRT | 2013:Q4 |
| Linear | 1.07 | 1.06 | 1.09 | 1.05 | 1.08 | 1.11 | 1.02 | 1.08 | 1.12 |
| Quadratic | 1.06 | 1.05 | 1.03 | 1.05 | 1.06 | 1.05 | 1.03 | 1.06 | 1.10 |
| HP | 0.98 | 0.96 | 1.01 | 0.96 | 0.96 | 1.00 | 0.93 | 0.95 | 0.95 |
| GLMS | 1.00 | 1.01 | 1.01 | 0.98 | 1.00 | 1.00 | 0.99 | 1.00 | 0.95 |
| Hamilton | 0.98 | 0.96 | 0.96 | 0.95 | 0.94 | 0.93 | 0.92 | 0.95 | 0.94 |
| MW | 0.96 | 0.93 | 1.01 | 0.99 | 0.94 | 0.99 | 1.03 | 0.97 | 0.92 |
| UC (GDP) | 0.98 | 1.05 | 1.03 | 0.98 | 1.07 | 1.04 | 1.01 | 1.08 | 1.08 |
| UC (GDP/$\pi$) | 0.94 | 1.05 | 1.04 | 0.92 | 1.05 | 1.06 | 0.97 | 1.04 | 1.08 |
| UC (GDP/CU) | – | 1.05 | 1.01 | – | 1.05 | 1.02 | – | 1.06 | 1.03 |
| UC (GDP/u.) | 1.07 | 1.09 | 1.06 | 1.08 | 1.11 | 1.09 | 1.05 | 1.10 | 1.12 |
| UC (GDP/u./$\pi$) | 1.08 | 1.10 | 1.06 | 1.08 | 1.12 | 1.09 | 1.05 | 1.10 | 1.11 |
| Tealbook | 1.03 | – | **0.63** | – 0.99 | – | **0.66** | – 0.92 | – | **0.75** |

Notes: Absolute RMSE of inflation forecast from equation 3 using October 2013 Tealbook output gap estimate shown as bolded values. All other entires are the ratio of that RMSE to the bolded value. Out-of-sample period is 1984:Q1–2013:Q3 and models are estimated beginning in 1975:Q1 using an expanding scheme. Real-time UC (GDP/CU) model omitted due to short sample. Asterisks denote that DM test statistic for dominance of indicated model over the null Tealbook estimate is statistically significant at 10 (*) or 5 (**) percent level. See text for details.

Table A5: Phillips curve based forecasts of total PCE inflation, 1984:Q1–2013:Q3.

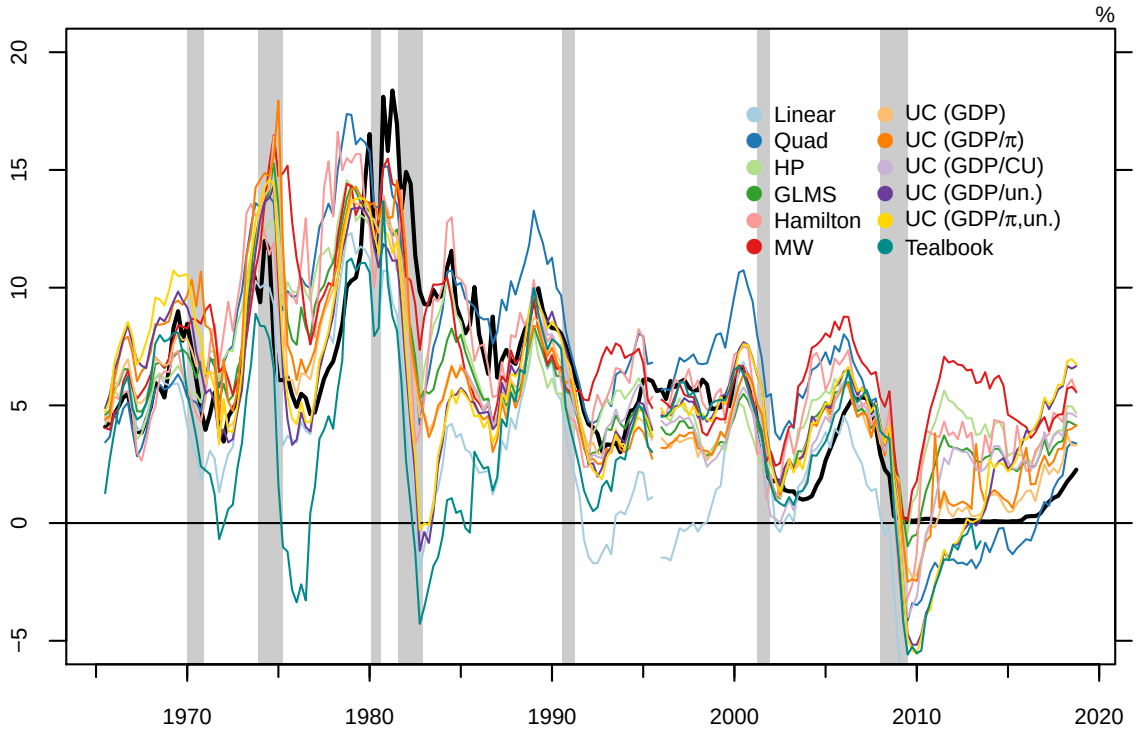
| | Two quarters ahead | | | Four quarters ahead | | | Eight quarters ahead | | |
|---|---|---|---|---|---|---|---|---|---|
| | RT | QRT | 2013:Q4 | RT | QRT | 2013:Q4 | RT | QRT | 2013:Q4 |
| Linear | 1.04 | 1.04 | 1.04 | 1.02 | 1.05 | 1.06 | 1.00 | 1.06 | 1.07 |
| Quadratic | 1.04 | 1.03 | 1.03 | 1.02 | 1.03 | 1.05 | 1.01 | 1.03 | 1.10 |
| HP | 1.02 | 1.00 | 1.02 | 0.99 | 0.98 | 1.01 | 0.96 | 0.97 | 0.97 |
| GLMS | 1.02 | 1.03 | 1.01 | 1.00 | 1.01 | 1.01 | 0.98 | 0.99 | 0.97 |
| Hamilton | 1.01 | 0.99 | 1.01 | 0.98 | 0.98 | 0.99 | 0.94 | 0.96 | 0.97 |
| MW | 1.03 | 1.00 | 1.02 | 1.04 | 1.01 | 1.00 | 1.06 | 1.03 | 0.93 |
| UC (GDP) | 1.02 | 1.05 | 1.03 | 0.99 | 1.05 | 1.04 | 0.99 | 1.06 | 1.07 |
| UC (GDP/$\pi$) | 1.01 | 1.04 | 1.04 | 0.98 | 1.04 | 1.05 | 0.97 | 1.03 | 1.07 |
| UC (GDP/CU) | – | 1.04 | 1.01 | 0.78 | 1.04 | 1.02 | – | 1.04 | 1.03 |
| UC (GDP/u.) | 1.06 | 1.07 | 1.05 | 1.06 | 1.08 | 1.07 | 1.04 | 1.08 | 1.11 |
| UC (GDP/u./$\pi$) | 1.06 | 1.07 | 1.05 | 1.07 | 1.08 | 1.08 | 1.04 | 1.08 | 1.10 |
| Tealbook | 1.02 | – | **1.36** | 0.98 | – | **1.26** | 0.89 | – | **1.27** |

Notes: Absolute RMSE of inflation forecast from equation 3 using October 2013 Tealbook output gap estimate shown as bolded values. All other entires are the ratio of that RMSE to the bolded value. Out-of-sample period is 1984:Q1–2013:Q3 and models are estimated beginning in 1975:Q1 using an expanding scheme. Asterisks denote that DM test statistic for dominance of indicated model over the null Tealbook estimate is statistically significant at 10 ($^{*}$) or 5 ($^{**}$) percent level. See text for details.

## A4 Alternative Taylor rule implied interest rates

Here, we repeat the analysis of section 5.3 but use instead a rule with $w_C = 1$ and $w_\pi = 1/2$.

Figure A1: "Balanced" Taylor rule implied federal funds rate, 1965:Q3–2018:Q4.




Notes: Figure shows prescribed interest rate from Taylor 1993 rule, calculated as in equation 4 with $w_\pi = 1/2$, $w_C = 1$, $\pi$ equal to the four quarter change in the real-time GDP price deflator, $y_t$ as the real-time output gap estimate, and $\pi^* = r^* = 2$ percent. Thick black line denotes actual effective federal funds rate. See text for details.

Table A6: Performance of prescribed interest rate from ' 'balanced" Taylor rule.

| | 1965:Q3–2013:Q3 | | | 1975:Q1–1997:Q4 | | | 1998:Q1–2013:Q3 | | |
|---|---|---|---|---|---|---|---|---|---|
| | RT | QRT | 2013:Q4 | RT | QRT | 2013:Q4 | RT | QRT | 2013:Q4 |
| Linear | 5.24 | 4.80 | 4.21 | 5.27 | 4.62 | 3.77 | 6.18 | 5.80 | 4.72 |
| Quadratic | 2.92 | 3.02 | 4.29 | 3.43 | 3.68 | 4.39 | 2.61 | 2.50 | 3.82 |
| HP | 2.74 | 2.77 | 2.73 | 3.04 | 3.11 | 3.08 | 2.73 | 2.66 | 2.32 |
| GLMS | 2.42 | 2.39 | 2.70 | 2.90 | 2.90 | 3.08 | 2.04 | 2.03 | 2.23 |
| Hamilton | 3.03 | 3.06 | 3.08 | 3.46 | 3.60 | 3.56 | 2.79 | 2.70 | 2.69 |
| MW | 3.33 | 3.31 | 2.74 | 3.05 | 3.09 | 2.87 | 3.93 | 3.82 | 2.57 |
| UC (GDP) | 2.28** | 2.30 | 4.27 | 2.98** | 2.99 | 3.78 | 1.37 | 1.29 | 3.87 |
| UC (GDP/$\pi$) | 2.60 | 2.54 | 3.13 | 3.14 | 2.93** | 3.30 | 1.62 | 1.60 | 2.74 |
| UC (GDP/CU) | – | 2.67 | 3.25 | – | 3.27* | 3.63 | 2.06 | 1.94 | 2.25 |
| UC (GDP/u.) | 3.02 | 3.04 | 3.81 | 3.69 | 3.69 | 4.03 | 2.50 | 2.50 | 3.23 |
| UC (GDP/u./$\pi$) | 3.05 | 3.07 | 3.78 | 3.57 | 3.66 | 4.00 | 2.55 | 2.40 | 3.13 |
| Tealbook | 4.25 | – | 3.23 | 5.75 | – | 4.01 | 1.90 | – | 1.58 |

Notes: Table shows RMSE of implied interest rate from "balanced" Taylor rule ($w_C = 1$, $w_\pi = 1/2$) calculated using each output gap estimate relative to actual effective federal funds rate. UC (GDP/CU) is excluded from the real-time evaluation in the first two sub-samples because of limited sample. Asterisks denote that DMW test statistic for dominance of indicated model over the null October 2013 Tealbook output gap is statistically significant at 10 (*) or 5 (**) percent level. See text for details.