ISSN 2279-9362

Research
Education
Outreach
CCA

Real-time Reliable Output Gap Estimates based on a Forecast Augmented Hodrick-Prescott Filter

Giacomo Porcellotti

No. 718
May 2024

Carlo Alberto Notebooks

www.carloalberto.org/research/working-papers

© 2024 by Giacomo Porcellotti . Any opinions expressed here are those of the authors and not those of the Collegio Carlo Alberto.

Real-time Reliable Output Gap Estimates based on a Forecast Augmented Hodrick-Prescott Filter

Giacomo Porcellotti *

May 6, 2024

Abstract

Although its poor end-of-sample performance and the resulting real-time unreliability, Hodrick and Prescott (1997) filter is the most popular detrending method. Here I exploit forecasts by Survey of Professional Forecasters (SPF) to construct a real-time reliable output gap estimate based on a forecast augmented HP filter. I show that the new filter has a sharp better forecasting properties for output growth, unemployment rate and inflation by a standard Phillips curve relation. It is much more correlated with Policy Institutions output gaps and also slightly outperforms one-sided Hamilton (2018) filtered output gaps in these comparisons. An analytical formulation of the forecast augmented HP filter is provided. The good real-time properties of forecast augmentation are extended also to Butterworth et al. (1930), Christiano and Fitzgerald (2003) band-pass filters and Canada's country case.

Keywords: Business cycle measurement, trend-cycle decomposition, real-time reliability, output growth forecasting, inflation forecasting

JEL Codes: C15, C22, E32, E37

*University of Turin-Collegio Carlo Alberto, giacomo.porcellotti@carloalberto.org.
I am greatful to Fabio Bagliano and Roberto Golinelli for crucial comments. I also thank Josefine Quast, Luca Gambetti, Federico Ravenna, Giovanni Pellegrino and all the other participants to the 11^{th} SIdE WEEE and also participants to 8^{th} International PhD Meeting in Salonicco, 64^{th} RSA of SIE in L'Aquila.

Introduction

The seminal work by Orphanides and Norden (2002) showed that widely used trend-cycle decomposition methods lead to real-time unreliable output gap estimates, meaning that estimates are highly revised when new incoming data are available. Since policy decisions require reliable estimates of output gap in real-time, following literature has primarily focused on the real-time performance of competing filtering techniques. In this debate, the filter that was most criticized for its poor end-of-sample performance was undoubtedly the Hodrick and Prescott (1997) filter (Orphanides and Norden (2002), Edge and Rudd (2016), Hamilton (2018), Jönsson (2020), Quast and Wolters (2022) and Kamber et al. (2018, 2024)). Other than US, its real-time *un*reliability has been documented in Japan (Kamada, 2005), Norway (Bernhardsen et al., 2005), Euro Area (Marcellino and Musso, 2011), Brazil (Cusinato et al., 2013), Canada (Cayen and Van Norden, 2005; Champagne et al., 2018b), Germany and UK (Quast and Wolters, 2022). In particular, Hamilton (2018) observed that because it's two side nature, filtered values at the end of the sample significantly differ from those in the middle and proposed a one side approach with better real-time properties (Jönsson (2020), Quast and Wolters (2022)). Although its poor real-time performance, the filter remains very popular either in academia either in policy institutions as IMF (De Resende (2014)), Fed (Edge and Rudd (2016)) and BIS. In a theoretical perspective, Canova (2023) proposed a horse race of competing output gap estimates to study their properties in leading DSGE New Keynesian models and found that HP filter ranks well in different DGP (including the Smets and Wouters (2007) model). With a simulated data approach, Hodrick (2020) shows that Hodrick and Prescott (1997) and Baxter and King (1999) filters perform better than Hamilton (2018) one when considering more complex time series, with distinct growth and cyclical components, (as real GDP) while the reverse is true for simple time series, such as a random walk.

In consideration of this longlasting popularity, I hence propose a modification of "Orphanides and Norden (2002) procedure" to construct real-time output gap estimates which also exploit GDP forecasts by Survey of Professional Forecasters (SPF).

I found that this combination improves the real-time performance of HP filter and leads to more economic meaningful output gap estimates. In particular, to asses real-time reliability I re-compute traditional Orphanides and Norden (2002) revision statistics and show that when augmenting each GDP data vintage with median SPF forecasts, conclusions on real-time (un)reliability of HP filter are totally different. Then, by forecasting exercises and correlations with other cyclical measures, I show that the improved real-time performance of the filter is effective in delivering more economic meaningful output gap estimates.

Finally, I extend the validity of my results to others filtering techniques and I show that also older band-pass approaches (Baxter and King (1999) and Butterworth et al. (1930)) benefit from my proposed forecast augmentation. Conversely, the more recent Hamilton (2018) filter, since its one side nature, doesn't improve its real-time performance by forecast augmentation, but notwithstanding still remains the leader filter for real-time reliability. However, when comparing the economic meaningfulness of the resulting output gap estimates I find slight improvements for forecasting accuracy and correlations when using the SPF augmented HP filter than Hamilton (2018) filter and its Quast and Wolters (2022) modified version. Hence, marginal contribution of this paper is to get HP filter back on track for the horse race on output gap estimation. Actually, real-time reliability should be only a necessary prerequisite for policymaker purposes, but it is not informative on the quality of the resulting business cycle estimate. I conclude that when improving the end-of-sample performance of HP filter with SPF forecasts, the resulting output gap has correlations and forecasting properties very similar to Hamilton (2018) approach and so conclusions of recent contributions (Hamilton (2018), Champagne et al. (2018b), Jönsson (2020) and Quast and Wolters (2022)) on HP real-time unreliability are highly revised.

Forecast augmentation of GDP vintages is not a totally new idea. Kaiser and Maravall (1999) first made use of ARIMA forecasts to improve the performance of HP filter. Garratt et al. (2008) made use of forecasts of current and future post-revision output levels to obtain more precisely estimated measures of the gap. Quast and Wolters (2022) proposed a AR(4) process to forecast augment GDP vintages, but the resulting HP output gap still yields a worse performance in terms of real-time reliability, forecasting and correlations than Hamilton (2018) filtered real-time output gaps. Eo and Morley (2023) used just SPF forecasts, but to improve real-time detection of business cycle turning points. In this papers I explore these different alternatives of forecast augmentations, including Greenbook forecast augmentation, showing that SPF is the best solution to maximize the filter's real-time performance. To clear the ground from misunderstandings, it is important to stress what the paper does not do. My work is not concerned with developing a limit theory for HP filter λ parameter as the sample size tends to infinity as in Phillips and Jin (2021); nor about adjusting the λ parameter when changing the frequency of observations as in Cooley and Ohanian (1991), Correia et al. (1992), Baxter and King (1999) and Ravn and Uhlig (2002). Finally, the paper is also not concerned with discussing potential adjustments to the one-sided HP filter to better align its cyclical properties to those of the two-sided HP filter as in Wolf et al. (forthcoming).

The paper is organized as follows. Section 1 describes the original one-sided HP filter and the source of its end-of-sample bias. Section 2 presents the new forecast augmented version. Section 3 provides data sources for historical data and explores Survey Professional Forecasters (3.1), Greenbook (3.2) and model forecast augmentations (3.3). Section 4 reports its revision statistics on real-time reliability with respect to the original HP filter. Section 5 evaluates the performance in terms of forecasting (5.1) and correlations with final revised institutional output gaps (5.2). Section 6 reports comparison with Hamilton (2018) filtered output gaps (6.1), with Greenbook's staff estimates (6.2) and extends forecast augmentation also to band-pass filters (6.3). Section 7 extends the validity of my results to Canada's country case. Finally, Section 8 concludes.

1 The Hodrick-Prescott filter

The HP filter decomposes a time series $y = (y_1, ..., y_T)'$ into an additive cyclical component, $c = (c_1, ..., c_T)'$, and a trend component $\tau = (\tau_1, ..., \tau_T)'$:

$$y_t = \tau_t + c_t \tag{1}$$

with T denoting the full sample size. In its original two-sided formulation the HP filter takes the form:

$$\{\tau_{1|T,\lambda}, ..., \tau_{T|T,\lambda}\} = argmin_{\tau_1,..,\tau_t} \left(\sum_{t=1}^{T} (y_s - \tau_s)^2 + \lambda \sum_{s=2}^{T-1} (\tau_{s+1} - 2\tau_s + \tau_{s-1})^2 \right) \tag{2}$$

where the first summation penalizes a poor fitting, while the second one penalizes trend variability. The prior assumption behind the minimization problem of equation (2) is that $\tau = (\tau_1, ..., \tau_T)'$ varies "smoothly" over time and the measure of the smoothness is the sum of squares of its second difference. λ is the *smoothing parameter* penalizing the variability in the trend component. The larger the value of λ, the smoother the trend component and the greater the variability of the cycle. As λ approaches infinity, the trend component corresponds to a linear time trend. For quarterly data, Hodrick and Prescott (1997) propose setting λ equal to 1600 and it is also nowadays a common choice.

Notice that the two-sided filter exploits observation on *both sides*, i.e. a generic trend observation τ_s depends on the λ parameter and on the full sample size (1,..,T).

However, a serious shortcoming with this formulation is that for policymaking purposes we need trend (cycle) estimates in real-time, i.e. at the end of the entire sample. Building on this idea, the relevant case for real-time estimation is the so called *one-sided* HP filter, which decomposes a time series y_t into trend τ_t and cycle c_t exploiting only observations up to t:

$$\hat{\tau}_{t|t,\lambda} = argmin_{\tau_t} \left(\min_{\tau_1,...,\tau_{t-1}} \left(\sum_{s=1}^{t} (y_s - \tau_s)^2 + \lambda \sum_{s=2}^{t-1} (\tau_{s+1} - 2\tau_s + \tau_{s-1})^2 \right) \right) \tag{3}$$

Notice that this procedure is equivalent to applying the two-sided filter recursively on an expanding sample and keeping, from each recursion step, only the trend estimate for the latest period (see for instance Stock and Watson (1999), Mehra (2004), Phillips and Jin (2021) and Wolf et al. (forthcoming)). However, the big drawback is that previously estimated trend and cycle components, for a specific point in time, change to a large extent as more data become available and the decomposition is re-computed based on a longer time series. This drawback has been documented by Orphanides and Norden (2002), Mise et al. (2005) and Jönsson (2020) among others. In order to make clear this recursive instability, equation (3) can be conveniently rewritten in matrix form as:

$$\tau_t = (I_t + \lambda S'S)^{-1} y \tag{4}$$

$$\underset{txt}{S} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & ... & 0 \\ 1 & -2 & 1 & 0 & 0 & ... & 0 \\ 0 & 1 & -2 & 1 & 0 & ... & 0 \\ \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & ... & 0 & 1 & -2 & 1 & 0 \\ 0 & ... & 0 & 0 & 1 & -2 & 1 \\ 0 & ... & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{5}$$

where $\tau_t = (\tau_1, ..., \tau_T)$ in equation (4) is the estimated trend, y_t is the time series and S is a second-difference matrix. The cyclical component $c_t = (c_1, ..., c_T)$ is the difference between y_t and τ_t[1]. The solution for (4) implies:

$$\hat{\tau}_{t|t,\lambda} = \sum_{s=1}^{t} \omega_{t|t,s,\lambda} \cdot y_s \tag{6}$$

and

$$\hat{c}_{t|t,\lambda} = y_t - \sum_{s=1}^{t} \omega_{t|t,s,\lambda} \cdot y_s \tag{7}$$

[1]Notice that such matrix formulation can be extended also to the original two-sided case just changing notation t with T in (4) and (5)

where $\omega_{t|t,s,\lambda}$ are the elements of the t-th row of the matrix $(I_t + \lambda S'S)^{-1}$. Cornea-Madeira (2017) and Hamilton (2018) derive explicit formulas for such weights in finite samples.

The reason for the recursive instability of the HP filter is that weights $\omega_{t|t,s,\lambda}$ that are used to extract the trend and cycle components change as observations are added to the sample and the components $\hat{\tau}_{t|t,\lambda}$ and $\hat{c}_{t|t,\lambda}$ are re-estimated. How the weights change depend on how the position s of the time series observation under consideration relates to the full sample t. The weights used to calculate the trend component for the last observation t change to a large extent as more data become available and the time series observation under consideration moves further away from the sample end. Using these insights, in next section I derive an analytical expression for the forecast augmented HP filter mitigating the end-of-sample bias affecting the widespread one-sided method.

2 A Forecast Augmented Hodrick-Prescott Filter

In order to mitigate the end-of-sample bias of the one-sided HP filter, I feed in equation (3) all information available at time t, including the real-time forecast observations available at that time:

$$\hat{\tau}_{t|t,f_1,\ldots,f_h,\lambda} = argmin_{\tau_t} \left(\min_{\tau_1,\ldots,\tau_{t-1}} \left(\sum_{s=1}^{f_h} (y_s - \tau_s)^2 + \lambda \sum_{s=2}^{f_h-1} (\tau_{s+1} - 2\tau_s + \tau_{s-1})^2 \right) \right) \tag{8}$$

with $f_1, \ldots, f_h$ being all the forecasts available after the last observation t up to forecast horizon h. (i.e. if h=5 as for SPF forecasts, f_5 is the last forecast observation for horizon h=5). The polynomial form of equations (6) and (7) hence becomes:

$$\hat{\tau}_{t|t,f_1,\ldots,f_h,\lambda} = \sum_{s=1}^{t} \omega_{t|t,f_1,\ldots,f_h,s,\lambda} \cdot y_s \tag{9}$$

and

$$\hat{c}_{t|t,f_1,\ldots,f_h,\lambda} = y_t - \sum_{s=1}^{t} \omega_{t|t,f_1,\ldots,f_h,s,\lambda} \cdot y_s \tag{10}$$

where weights $\omega_{t|t,f_1,\ldots,f_h,s,\lambda}$ now also depends on the number of forecasts available at time t and s is the position of the time series observation under consideration in the full sample *including forecasts*.

In one-sided HP filter weights used to calculate the trend component for the last observation t change to a large extent as more data become available. In the forecast augmented HP filter adding forecast observations to the sample makes changes in the weights of equations (6) and (7) the smaller (a) the more observations are added to the sample (b) the lower the forecast errors of appended forecasts. Both (a) and (b), in turn, implies that revisions of the trend (cycle) estimate are smaller as well. Intuitively, the lower forecast errors, the less weights of equations (6) and (7) used to calculate the trend (cycle) component for the last observation will change as more data become available. Similarly, as weights of the the last observation change to a large extent, the more forecast observations are added the less its weights will change. Such relation between number of appended forecast observations and weights' change (cycle revision) will be pretty clear in Figure 1, where I show the marginal effect of adding an extra forecast observation on cycle revision for US output gap. However and in a general perspective, as changes in weights of equations (9) and (10) are lower than those of (6) and (7), the resulting effect will be that the cyclical (trend) component of forecast augmented HP filter better matches the one of two-sided HP filter than what does the component(s) estimated with the original one-sided method (3).

3 Data and Forecasts

I obtain data on quarterly US real GDP from the Federal Reserve Bank of Philadelphia's real-time dataset- hereafter FRB. The real-time dataset provides vintages of real GDP at given dates in the past, as they would have been available at that time. The first data vintage is 1965Q4 and the vintage series predominantly starts in 1947Q1 and ends one quarter before the publication date of the data vintage, i.e.

in 1965Q3 for the first [2]. Following literature starting by Orphanides and Norden (2002), real-time estimates of one-sided HP filter are constructed in the following two-step procedure. First, I detrend each and every vintage of log real GDP available to construct an ensemble of output gap series. Of course, earlier vintage output gap series are shorter than later vintages since the output series on which they are based end earlier. Next, I use these different vintages to construct a new series which consists entirely of the output gap estimate for the latest period of every vintages. This new series is the real-time estimate of the output gap and represents the most timely estimate of the output gap which policy makers could have constructed at any point in time.

The forecast augmented HP filter is constructed following the same two-step procedure, but using real GDP vintages in conjunction with real-time forecasts for each vintage series. There are several possibilities to forecast augment historical vintages and consists on professional, institutional or model forecasts. The following paragraphs review such options of forecast augmentation and discuss some pros and cons.

3.1 Survey of Professional Forecasters

The Survey of Professional Forecasters - hereafter SPF - is the oldest quarterly survey of macroeconomic forecasts in the United States. It asks professional economists to provide forecasts for several macroeconomic variables, as well as over different forecasting horizons. The forecast augmented HP filter is constructed using real GDP vintages in conjunction with the corresponding median predictions. The main advantage of SPF predictions is that they perfectly fit the filter, as the survey's timing is aligned to the Bureau of Economic Analysis's (BEA) advance release of

[2]Since the FRBP dataset contain a few missing observations, some imputations have to be made in order to investigate real-time properties throughout the entire sample period. For the vintages 1992Q1–1992Q4 data up to 1958Q4 are missing. Data from the 1993Q1 vintage are used to impute these missing data. Similarly, data up to 1959Q2 from the 1997Q2 vintage are used for imputations for the 1996Q1–1997Q1 vintages, and data up to 1958Q4 from the 2000Q2 vintage are used for the 1999Q4–2000Q1 vintages. Finally, the 1995Q4 observation for the same quarter's vintage is imputed by using the value from the 1996Q1 vintage.

the data and hence with the FRB real-time dataset (see Stark et al., 2010). In particular, I augment each vintage with real-time nowcast [3] of the current quarter and forecasts [4] for the following 4 quarters. For example, for vintage 1968Q4 I augment the series with nowcast for 1968Q4 and forecasts up to 1969Q4.

SPF doesn't provide quarter-to-quarter forecasts for longer time horizon. Hence the main limitation of SPF augmentation is that we can rely only on 5 additional observations. A second shortcoming may be due to information rigidities of SPF predictions documented by Coibion and Gorodnichenko (2012, 2015). As forecasting accuracy plays a role in the forecast augmented HP filter, the fact that predictions about the future level of output are not always particularly accurate or even unbiased may reflect on filter's real-time reliability. However, Eo and Morley (2023) show that SPF nowcasts and forecasts are effective in detecting the troughs of business cycles, as the degree of information rigidity significantly decreases one year after the start of a recession (Coibion and Gorodnichenko (2015)).

Hereafter, I will refer to this filter as *SPF augmented HP* filter. Because SPF forecasts are available from 1968Q4, the new real-time SPF augmented output gap is available from 1968Q3 onwards.

3.2 Greenbook Forecasts

Greenbook - hereafter GB - projections are produced before each meeting of the Federal Open Market Committee (FOMC) by research staff at the Board of Governors. There are a couple of shortcomings with using GB real GDP growth forecasts in conjunction with FRB historical values. First, projections are released to the public with a five years-lag, and hence available in real-time only to members of the FOMC committee and its research staff. Because this five-year ban I adopt as

[3]SPF also provides *backcast* for the quarter prior to the quarter in which the survey is conducted, which corresponds to the last observation of each vintages available in Federal Reserve Bank of Philadelphia's real-time dataset. I chose to change such observation with SPF corresponding backcast. However, results are insensitive to this choice since most of the professional forecasters don't revise it. See Survey of Professional Forecasters pag. 19 for further details.

[4]For vintages 1969Q1 1969Q2 1969Q3 1970Q1 1974Q3 the one-year-ahead forecast of GDP is missing, so I adopt imputation and replace the four quarter ahead forecast with the three quarters ahead forecast of the following vintage.

final revised estimate the most recent available estimate that the staff of the Federal Reserve Board constructed for the meeting of the the last 07/12/2018.

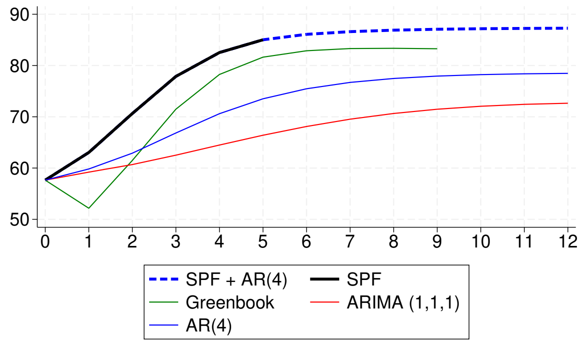
Notwithstanding the real-time unavailability, I found augmenting historical vintages with GB forecasts can be an useful exercise to construct a *GB augmented HP filter*, as additional instrument in FED's toolbox and studying how it performs in comparison to Greenbook staff output gap estimate (see Section 6.2). Second, as FOMC's meetings are 8-12 yearly depending to decades and historical periods, some arbitrary choices are needed to reasonably match GB projections with historical vintages. In particular, when matching the two datasets is important that matched information were actually available at the same time. This issue has been already addressed by Orphanides (2004) and Edge and Rudd (2016) when switching Greenbook output gap from FOMC-meeting to quarterly frequency. In order to make reasonable my GB augmentation I follow its approach and collect observations corresponding to the Greenbook prepared during (or when not available by) the middle of the quarter (February, May, August, November). This choice makes the matching feasible, because observations of historical vintages represent data exactly as they were known in the middle of each quarter.

With regard to the number of forecasts observations, it ranges between a minimum of four projections for each observation from 1973q3 to a maximum of even 9 predictions for some observations from 1989q4 onwards. In order to make possible a comparison with SPF augmentation I take the 1973q3-2018q4 series as benchmark and explore the longer time series in robustness checks.

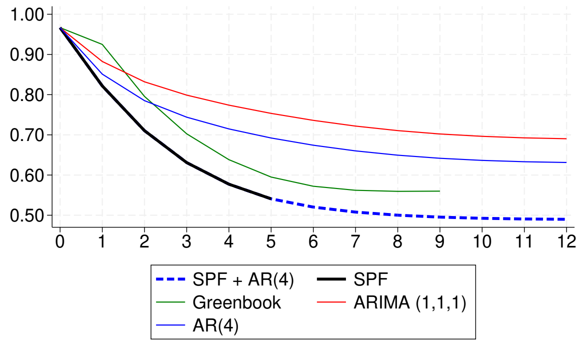
3.3 Model Forecasts

Someone may argue why using just SPF or GB forecasts, which has the serious drawback of being available only for a limited number of forecast horizons. Unlike, model forecasts are still available in real-time and can also be extended for longer horizons as needed. Actually, vintages forecast augmentation is not totally a new idea. Some previous contributions by Kaiser and Maravall (1999), Mise et al. (2005) and Quast and Wolters (2022) (which buils on Stock and Watson (2007)) also proposed to forecast augment data vintages but using ARIMA or AR forecasts. Figure (1) plots HP filter real-time reliability according to h ahead vintages forecast

Figure (1) Model Forecasts vs SPF Forecasts



(a) Correlations



(b) Noise-to-signal ratio

augmentation. In particular, Panel (1a) reports how correlations between real-time and final revised estimates changes if vintages are augmented with an additional forecast observation. Panel (1b) shows the marginal gain for the Noise-to-signal ratio (NSR), which is the ratio between the standard deviations of Total Difference (i.e. the difference between final revised and real-time estimates) and final revised estimate.

Although all forecast augmentations share the favorable property to increase real-time reliability, the magnitude of improvement differ across methods and the marginal gains are decreasing the longer the forecast horizon. Model forecast augmentations typically display a lower performance than professional/institutional augmentations by SPF and GB. In particular, AR(4) augmentation suggested by Stock and Watson (2007), Quast and Wolters (2022) (and similarly Mise et al. (2005)) outperform ARIMA augmentation à là Kaiser and Maravall (1999). The latter approach meets some computational challenges, as maximum likelihood doesn't find convergence for all vintages, forcing to keep a parsimonious model. For both model forecast augmentations I follow Marcellino et al. (2006) and adopt iterated forecasts instead of direct forecasts, as they found the former being more accurate than the latter. Consistently with this finding, AR iterated forecasts augmentation further boosts the real-time performance than direct one[5].

Overall, SPF augmentation presents the highest real-time reliability and wins the horse race among alternative end-of-sample bias mitigations. I also try to further boost its performance by adding additional AR forecast observations, but the effect on real-time reliability is almost negligible, with correlation further increasing from 85 % to 87 % and NSR further decreasing from 0.54 to 0.49. Similarly to SPF, also GB augmentation display a nice real-time performance, but its gains are also decreasing with 8^{th} and 9^{th} forecast observations being even irrelevant. As a result, the superior performance of professional/institutional augmentations vs model augmentations is due to the higher forecasting accuracy of the former's initial forecast observations, although the latter offer more predictions. In particular, SPF is the most successful augmentation despite the lowest number of forecast appended to the sample, meaning that its performance relies on the accuracy of the five predictions appended.

Because SPF augmentation displays the best performance, hereafter I will use SPF predictions to construct the (*forecast*) *SPF augmented HP filter*, therefore results of Sections 4, 5 and 6 relies on the specification of the filter obtained with SPF forecasts.

4 Real-Time Reliability

Despite real-time reliability doesn't provide any evidence on the economic meaningfulness of competing output gap estimates, it is a crucial feature for policy making purposes. Because policy decisions require reliable estimates of output gap in real-time, real-time reliability is a desirable property for a good trend-cycle decomposition method.

In this section I investigate the real-time reliability of the SPF augmented HP filter in comparison with its original one-sided version. In particular, what is commonly meant for real-time reliability is that output gap computed with real-time data vin-

[5]In particular, I found correlation between real-time and final revised estimate increasing from 71 % with direct forecasts to 75 % with iterated ones

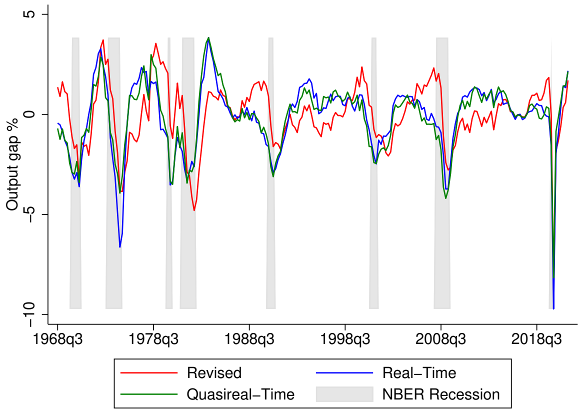
tages is as close as possible to that computed with final revised data. However, a potential source of differences between the real-time and the final estimates is the ongoing revision of published data. Therefore, following Orphanides and Norden (2002) I isolate the importance of this factor by computing the *quasireal-time* estimates. These are constructed in the following way. First, I construct an ensemble of "rolling" estimates of the output gap, that is I take the final vintage of the GDP series and I use only the observations up to and including 1968Q3 in order to compute the quasireal-time estimate for 1968Q3. Next, I extend the sample period by one observation and I repeat the detrending, continuing in this way until I have used the full sample period of the final GDP series. In this way I obtain a full set of quasireal-time output gap vintages.

Next, as when I constructed the real-time series, I collect the first available estimate of the output gap at each point in time from the various series I constructed in step one. This sequence of output gaps is the *quasireal-time* series. The difference between the real-time and the quasireal-time series is entirely due to the effects of data revision, since estimates in the two series at any particular point in time are based on data samples covering exactly the same time period. Because SPF forecasts are available only in real-time, for the forecast augmented HP filter I augment quasireal-time vintages using forecasts of median GDP growth rate.

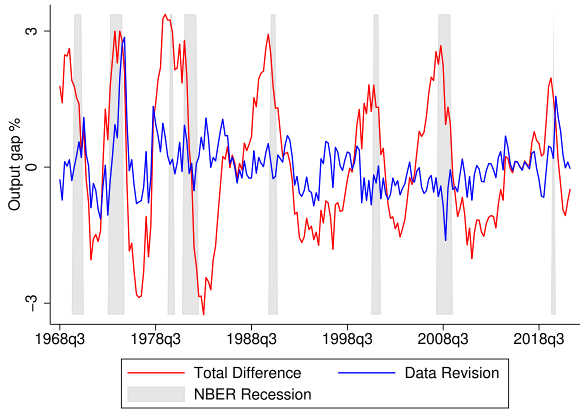
Finally, to make sure that a comparison of real-time and revised data is not biased by the last data vintages in which real-time and revised data converge, I discard the last quarters of observations. Hence, my results for the revised gaps are based on the 2024Q1 vintage with gaps being estimated from 1968q3 until 2021Q4. Figure (2) shows HP final Revised, real-time and quasireal-time output gap estimates with and without SPF forecast augmentation. To account for the role of data revisions on the total difference, the right panel of each figure reports Total Difference and Data Revisions, respectively measured as the difference between revised and real-time estimates and quasireal-time and real-time estimates.

Consistently with Figure (1), both Table (1) and Figure (2) show that the improvement by forecast augmentation is sharp. Correlation between Real-time and Final

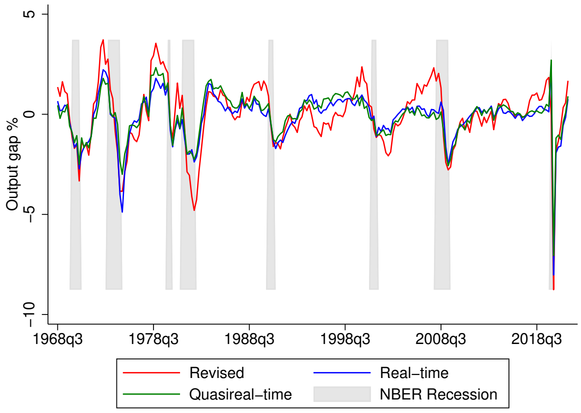
Figure (2) Output Gaps and Revisions



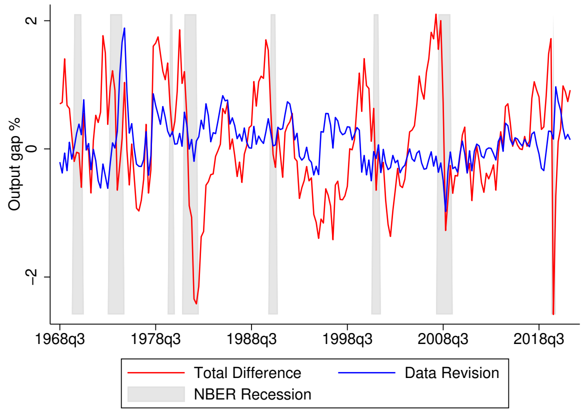
(a) Hodrick-Prescott-Output Gaps



(b) Hodrick-Prescott-Revision



(c) SPF Augmented HP-Output Gaps



(d) SPF Augmented HP-Revision

Revised estimates increases by 26 % without contribution of Data Revision. Indeed, correlation between Quasireal-time and Final Revised estimates (i.e. correlation between Real-time and Final Revised estimates *in absence* of Data Revision) remains approximately constant across augmentation or not. The bottom part of Table (1) also report traditional revision statistics on Opposite Sign Frequency - the ratio of observation for which final revised and real-time gap have the opposite sign - and Noise-to-signal ratio, already discussed in Section 3. Also looking at such statistics I found an improvement in the real-time properties of HP filter by forecast augmentation. The Opposite Sign Frequency reduces by 17 % and NSR for Total Difference almost halved. For robustness check, I also test if results hold when the trend variability is penalized by an higher λ parameter of 16000 and actually I found improvements similar to those of Table (1).

Table (1) Output Gap Estimates Revision Statistics

	HP_{SPF}	HP	*Difference*
Final/Real-time	**0.85**	0.58	0.26***
Quasi-Real/Real-time	**0.95**	0.94	0.01
Quasi-Real/Final	**0.84**	0.58	0.26***
Total Difference/Data Revision	0.14	**0.31**	-0.17***
Opposite Sign Frequency	22 %	**40 %**	-17 %
NSR-Total Difference	0.54	**0.97**	-0.43***
NSR-Data Revision	**0.23**	0.36	-0.13***

Notes: Bold denotes the most positive value. *Difference* reports the difference with respect to the SPF augmented filtered statistical real-time gap. *, **, and *** on correlation coefficients denote significant differences on the 10, 5, and 1 % level based on Fisher et al. (1921)'s z transformation. Inference on NSRs is made by the RMSE-based noise-to-signal ratio as in Edge and Rudd (2016).

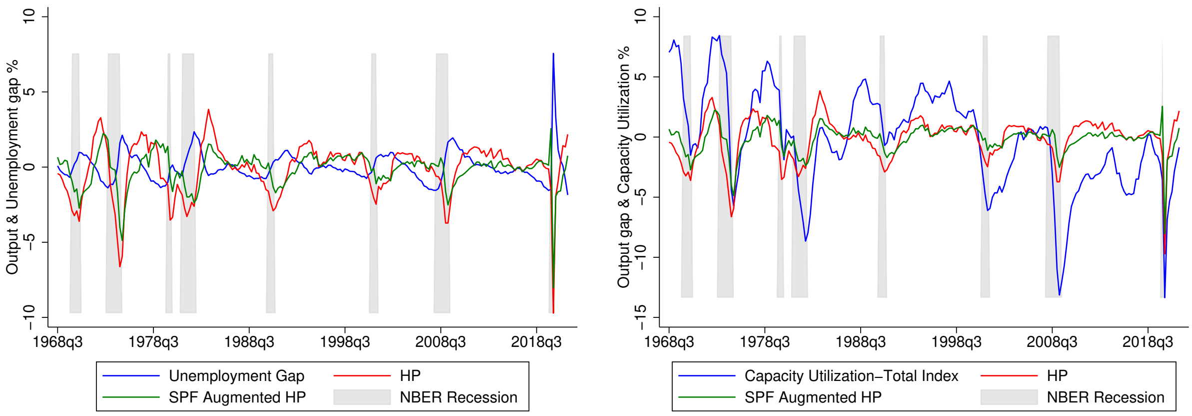
To sum up, I found solid evidence that the forecast augmented HP filter has better real-time properties than its original version. However, despite real-time reliability is a crucial feature for policy making purposes, it doesn't say anything on economic meaningfulness of the output gap estimate. Therefore, in Section 5 I show that SPF forecast augmentation improves the performance of HP filter for many purposes.

5 Economic Meaningfulness of Output Gap Estimates

Since the SPF augmented HP filter is more real-time reliable, here I present the resulting effects in terms of quality of output gap estimate. However, evaluating the meaningfulness of competing output gap estimates in this regard is difficult, because there is no a professional consensus on what business cycle is (Comin and Gertler (2006), Beaudry et al. (2020)) and the use of detrending methods is also source of debate (Hamilton (2018), Kamber et al. (2018), Morley and Panovska (2020) and Quast and Wolters (2022) just recently and just on univariate methods). Hence I adopt a plurality of assessment criteria to test the better performance of the forecast augmented HP filter in terms of economic meaningfulness.

Stationarity is a first basic requirement. SPF augmented HP filtered output gap is stationary, as follows by a Dickey and Fuller (1979) stationarity test. As further preliminary analysis I test how the two output gap estimates fit data into a standard Okun (1963) law and if they are positively correlated with measures of capacity utilization. Figure (3) plots the two gap estimates with (final revised and HP detrend) unemployment gap and a demeaned index of capacity utilization for total economy. As can be seen from Figure (3) the forecast augmented HP filter fits data much

Figure (3) Unemployment Gap and Capacity Utilization



more better than the original HP filter. The Relative RMSE for the Okun's law is 0.67, implying an improvement of 33 % of data fitting [6]. Moreover, I found a β coefficient of -0.68, much more consistent with recent literature on Okun's law (Ball et al. (2017)) than -0.25 of the original HP filter.

Although the correlation of output gap measures with capacity utilization should not be overemphasized, such measures aimed capturing the concept of sustainable maximum output, i.e. the greatest level of output a plant can maintain within the framework of a realistic work schedule, after factoring in normal downtime and assuming sufficient availability of inputs to operate the capital in place. A similar comparison was proposed by Camba-Mendez and Rodriguez-Palenzuela (2003) and I also retain being a useful indication on quality of output gap estimates. Correla-

[6] If I include a couple of lags for the output gap term as suggested by Ball et al. (2017) I found a similar sharp improvement.

tion with the demeaned Total Index of capacity utilization is 0.61 for the forecast augmented HP filter and 0.36 for the original HP. When I restrict to the index for Manufacturing sector only I found similar values.

5.1 Forecasting Performance

Nelson (2008) first proposed the evaluation of competing output gap measures via their output growth forecasting performance. In principle, a high-quality measure of output gap can be a useful predictor of GDP growth because capturing trend reversion of GDP. In particular, if an output gap is negative, we would expect above average output growth rates in the future so that output reverts back to trend. Conversely, if the output gap is positive, output growth should be below average some time in the future. In order to model this basic intuition I use a standard forecasting equation in which output growth h periods ahead is predicted using the real-time output gap vintage:

$$y_{t+h} - y_t = \alpha + \beta \tilde{y}_t + \epsilon_{t+h|t} \tag{11}$$

where y denotes log real GDP, $\tilde{y}_t$ the estimated real-time output gap vintage and $\epsilon_{t+h|t}$ the forecast error. I expect β to be negative, meaning the predictive power of output gap estimate to forecast the sign of the output growth.

Second, because output gap has a predictive content for output growth, I propose a similar forecasting equation also for unemployment rate. In principle, changes in unemployment rate and GDP are mediated by the famous Okun (1963) law, so that if GDP will increase some time in the future unemployment rate should decreases with few lags of time [7]. Therefore, similarly to equation (11), unemployment rate h periods ahead is predicted using the real-time output gap vintage:

$$u_{t+h} - u_t = \alpha + \beta \tilde{y}_t + \epsilon_{t+h|t} \tag{12}$$

[7] Notice that equation (12) is *not* the Okun (1963) law. Rather, it captures the forecasting power of output gap on unemployment rate mediated by the Okun (1963) law.

where u denotes unemployment rate, $\tilde{y}_t$ the estimated real-time output gap vintage and $\epsilon_{t+h|t}$ the forecast error.

Finally, since macroeconomics theory postulates a causal relationship between output gap and inflation by the Phillips curve relation, here I consider a Phillips curve forecasting model to evaluate alternative output gap estimates. In principle, an high quality measure of potential GDP should be an inflationary barrier and so provide useful information for inflation forecasting. Orphanides and Van Norden (2005) showed the effect of filters revision on inflation forecastability and argued that Phillips Curve models with real-time econometric estimates of the output gap perform worse than Phillips Curve models based on final estimates of the gap as well as univariate models of inflation. They conclude that real-time estimates of the output gap are not reliable for forecasting inflation in the United States.

Following Stock and Watson (1999), Orphanides and Van Norden (2005), Clark and McCracken (2006), Stock and Watson (2007), Kamber et al. (2018) and Quast and Wolters (2022) I use an Autoregressive Distributed Lag (ADL) Phillips Curve forecasting equation:

$$\pi_{t+h} - \pi_t = \alpha + \sum_{i=0}^{p} \beta_i \Delta\pi_{t-i} + \sum_{i=0}^{q} \gamma_i \tilde{y}_t + \epsilon_{t+h|t} \tag{13}$$

where π_t is the annualized core CPI inflation rate, $\tilde{y}$ the real-time output gap estimate and $\epsilon_{t+h|t}$ the forecast error. I use inflation vintages for the right side of the equation and I set lag lenght p=8 and q=4 [8].

Equation (11) (12) and (13) are all estimated by OLS and have final revised data on the left side of the equation. The initial sample runs from 1947Q1 to 1968Q3 and it's recursively expanded quarter by quarter plugging recursively the estimated real-time output gap vintages, following the *pseudo-out-of-sample* forecasting methodology *à*

[8]For simplicity, I use the same lag lengths on inflation and output gap across forecast horizons and all time periods. Particularly, I conduct Breuch-Godfrey and Durbin-Watson tests to verify absence of serial correlation with p=8 and q=4. Tests are conducted by setting h=1 for the entire sample period. By adding more lags for inflation and output gaps results are qualitatively the same. Moreover, Clark and McCracken (2006) also allowed the lag lengths to be chosen at each point in time as forecasting proceeds, and obtained qualitatively similar results.

là Stock and Watson (2007). In order to rely only on final revised data for the left hand side of equation (9), I discard the last two years of observations, hence the time sample is recursively expanded from 1968q3 to 2021q4. Table (2) reports the

Table (2) Relative RMSE for forecasting

Horizon	GDP growth	Unemployment	Core Inflation
1	0.92***	1.01*	1.01
2	0.86***	0.97***	0.98
3	0.80***	0.92***	0.94**
4	0.78***	0.87***	0.93***
5	0.78***	0.85***	0.91***
6	0.79***	0.83***	0.91**
7	0.80***	0.82***	0.94
8	0.82***	0.82***	0.97
9	0.83***	0.83***	1.00
10	0.84***	0.83***	1.02
11	0.86***	0.83***	1.03
12	0.87***	0.84***	1.03

Notes: *, **, and *** denote significance on the 10, 5, and 1% level based on a two-sided Diebold and Mariano (1995) test.

relative root mean squared error (RRMSE) for all three forecasting equation, by plugging competing output gap estimates into the model. In particular RRMSE is the ratio between RMSE of the model with SPF augmented output gap and the model with the one-sided one. Therefore, when RRMSE is below one means that forecast augmented HP filter has a better forecasting performance than original HP. When RRMSE is above one the opposite is true. Forecasts are performed for 1 to 12 periods ahead and their significance respect to the competing one is evaluated by the two sided Diebold and Mariano (1995) test.

Overall, I found evidence for a sharp better forecasting performance of the SPF augmented HP filter. For GDP growth, improvements of forecasting accuracy are particularly sharp, also in the order of 20 % and 22 %. The β coefficient is negative as expected. Notably, forecasting accuracy increases also for horizons higher than one year, despite I augmented vintages with SPF forecasts only until four quarters

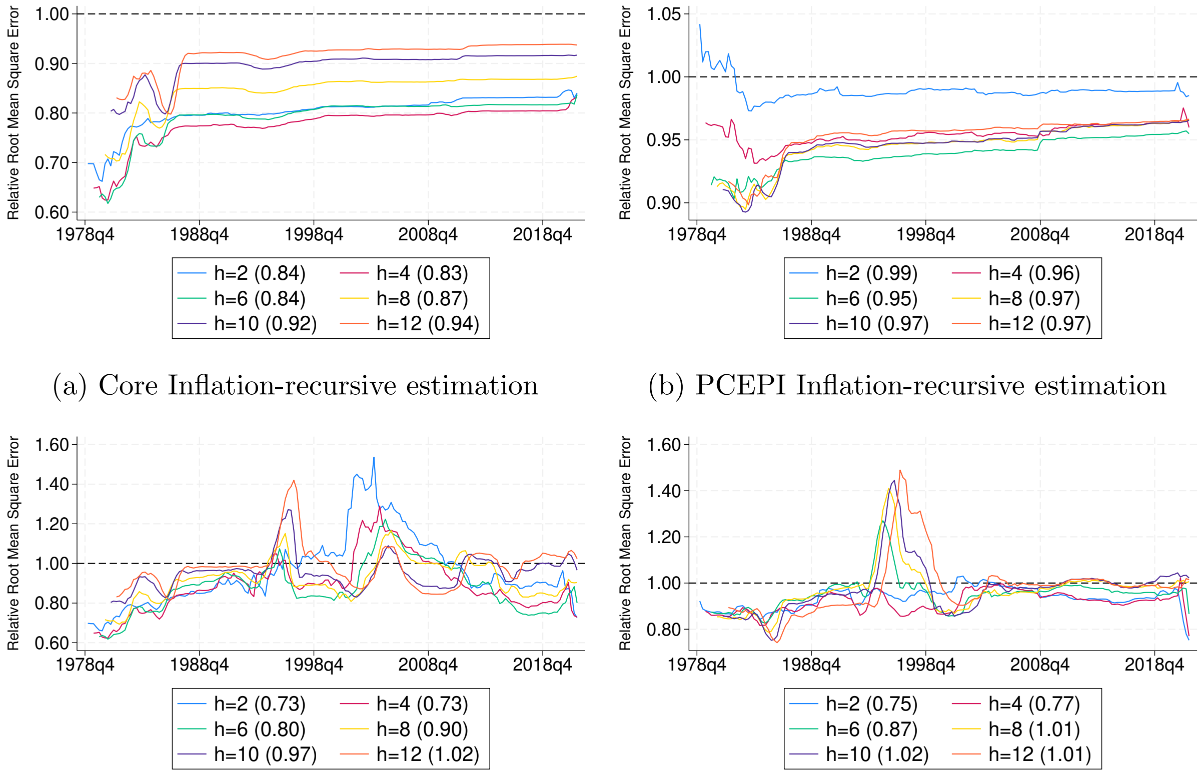
ahead. When I extend equation (11) and additionally control for the first difference of the output gap to account for changes in the level and the dynamics of the gap separately, as in Nelson (2008), I found very similar results. Whit regard to unemployment rate, I found a better performance of SPF augmented HP filter for horizons higher than h=1. It makes sense, since output gap predicts future output growth and the resulting effects on the labor market take place few quarters ahead. Both equations (11) and (12) contain output gap vintages just for historical values, however when instead plugging forecast augmented output gap vintages for the SPF augmented HP output gap I don't find any modifications to results of Table (2).
Finally, for inflation forecasting improvements are quite low and limited to first eight quarters-forecast horizons. Such marginal improvements are not robust when considering headline measures of inflation, as PCE or GDPDEF inflation, or when plugging SPF augmented output gap vintages in equation (13). Both modifications lead to RRMSE ranging unity. In particular, headline measures of inflation also absorb supply factors as oil ones and hence weakening the Phillips curve relation.
This difficulty for inflation forecasting is in line with previous literature documenting instability and flattening of Phillips Curve relationship. Among others, Stock and Watson (2007), Fuhrer et al. (2009), Faust and Wright (2013), Edge and Rudd (2016), Kamber et al. (2018) and Quast and Wolters (2022) showed that it is generally difficult to beat univariate inflation forecast models through conditioning on output gaps. In line with this literature, I test if conditioning equation (13) to the SPF augmented HP output gap improves inflation forecasts than the same model omitting the gap term. Figure (4) plots how the relative RMSE of the two models evolved over time both in a recursive and rolling estimation [9]. Because literature documents instability of the Phillips Curve and its weakening in the last decades, reporting how the RRMSE evolved over time makes the idea of the marginal gains for inflation forecasting when conditioning to the output gap term rather than using an AR process. Looking at recursive estimation it's clear how the Phillips Curve has weakened over decades, with RRMSE steadly trending towards unity, even if remain-

[9] For the rolling (recursive) estimation, parameters are estimated in a 21,5 years-window starting from 1947q1-1968q3 and then rolling (recursively) expanded, whereas RMSEs are measured in a constant 10 year-window (recursively expanded window).

ing significantly below one for core inflation. However rolling estimation is probably more informative. Across forecast horizons, it shows how inflation forecastability by Phillips Curve is decreased up to Great Recession - with a negative record in decade 1995-2005 when parameters were entirely estimated during the Great Moderation - and then improved in the last 15 years. In square brackets I report RRMSEs computed in the last decade 2011q4-2021q4, hence with parameters estimated in 1990q2-2011q3. For horizon up to h=8 they are significantly below unity, for both core and PCEPI inflation and the figure shows a further strengthening of the curve for the post-COVID period. These results are somehow in contrast with Quast and Wolters (2022), who found that Hamilton (2018) filtered output gaps don't beat the univariate model when computing the RRMSE in the full sample. Similarly, Kamber et al. (2018) showed that gains in changing the output gap estimates for forecasting inflation can be marginal at best and are generally not significant. More consistently with my results, Edge and Rudd (2016) in a more recent sample found that omitting the Greenbook real-time output gap causes a noticeable deterioration in forecasting performance.

To sum up, SPF augmentation improves inflation forecasts than original HP filter and also not trivially beats AR forecasts when the Phillips Curve relation holds. Therefore, although instability of the curve, such results show that SPF augmentation provides a better real-time measure of economic slack also as pertains inflation.

Figure (4) Stability of Phillips Curve



(a) Core Inflation-recursive estimation

(b) PCEPI Inflation-recursive estimation

(c) Core Inflation-rolling estimation

(d) PCEPI Inflation-rolling estimation

5.2 Correlation with Policy Institutions

There are four policy institutions estimating US output gaps: Congressional Budget Office (CBO), Federal Reserve (Fed), International Monetary Found (IMF) and the Organization for Economic Cooperation and Development (OED). Despite recent contributions highlight that real-time versions of these estimates are subject to heavy revisions (Coibion et al. (2018) and Gonzalez-Astudillo (2017)), there are still several good reasons for comparing statistical output gap estimates with final revised ones of policy institutions. First, non-statistical estimates are widely used by policy institutions to make countries' recommendations, particularly for budget and fiscal policies, and to formulate long-term forecast on their future output growth. Second, since the true cycle is unknown and the spirit of the paper is agnostic on that, a variety of benchmarks is a good news. Indeed, although the methodologies of all policy institutions are based on a mix of models and statistical approaches, they all combine these approaches in a different way. Third, comparison with CBO and OECD output gap is particularly interesting because their methodology relies on a production-function approach and hence typically impose more structure in the estimation than statistical methods. In particular, CBO estimates potential output with different methods for five sectors in the economy [10]. The main one is the nonfarm business (NFB) sector, which represents approximately 75 percent of the U.S. economy. The remaining four smaller sectors are agriculture and forestry, households, nonprofit organizations serving households, and government. For the bulk of the economy - the nonfarm business sector - CBO uses the Solow growth model taking the form of a Cobb-Douglass production function. Also OECD uses a production-function approach, but assumptions on inputs projections are very judgemental and TFP is assumed to converge to a certain degree among different countries in the medium-run (see Beffy et al. (2006)). Fed measure of output gap is judgmental and is the results from a number of estimation techniques that varied over time, including statistical filters and more structural model-based procedures. Additionally, Edge and Rudd (2016) showed that Fed's output gap has been more reliable than those based on statistical methods over the last 20 years. Finally, for IMF there is a considerable methodological variation across countries in how estimates of potential output are generated. In this sense the approach is methodologically similar to Fed's one, since they use a combination of statistical methods (including the HP filter), multivariate models (as Phillips Curve) and structural methods (although the production-function approach is not widely used by IMF economists [11]).

Unfortunately, there are some limitations regarding available samples and frequencies of the estimates. While for Fed and CBO quarterly data are available, samples for IMF and OECD are shorter and only rely on annual data. Data for Fed are available from 1975Q1 and ends in 2018Q3, since are based on the Greenbook[12] (see Section 3.2). For CBO the widest time sample is available from 1947q1 to 2023q4,

[10] Shackleton (2018) is a nice guide on CBO methodology for potential output methodology.

[11] See De Resende (2014) for a complete picture on the bunch of methodologies employed by IMF

[12] I adopt the most recent available estimate of potential GDP that the staff of the Federal Reserve Board constructed for the meeting of the Federal Open Market Committee (FOMC) the last 07/12/2018.

however to avoid biases from data revision I discard the last two years of observations as in Section 4. Finally, time samples for IMF and OECD respectively starts in 1980 and 1985 and finishes in 2022.

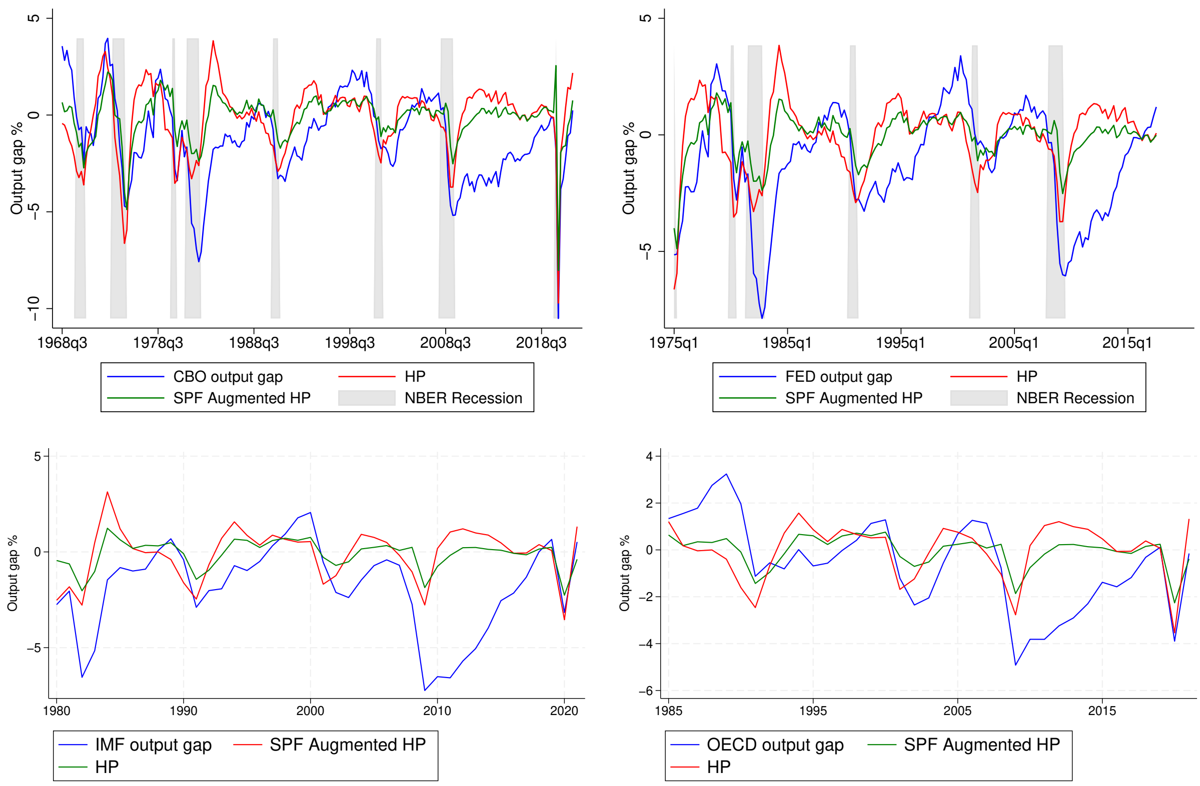
As policy institutions gaps differ for data frequency, I adopt alternative λ parameters to estimate real-time HP filtered output gap estimates. In particular, Hodrick and Prescott (1997) originally propose λ=100 for annual data, as ratio between cycle and trend variability, whereas in a frequency domain perspective Ravn and Uhlig (2002) suggest λ=6.25 for the same frequency of observations. Table (7) in Appendix A reports correlations of real-time statistical output gap estimates, obtained by detrending (the log of) US GDP vintages, with final revised policy institution gap measures. For annual frequency estimates I use only vintages for the first quarter of each year and I aggregate them by average. For forecast augmentation I use SPF annual-average forecasts for the current year (the year in which the survey is conducted) and the following two years[13].

Across λ values and observations frequency, when augmenting vintages with SPF forecasts, the increasing in correlations is sharp for all policy institutions, with even more than double correlations for FED gap. Interestingly, when I measure these correlations using final revised data I found correlation indexes of 0.75, 0.82, 0.92 and 0.90 respectively for CBO, FED, IMF and OECD. It means that in absence of the end-of-sample bias the HP filter successfully reflect the ex post expert evaluation of US business cycle. With this regard, correlation with CBO is particularly informative, as it only adopts a production function approach to estimate potential. Hence the comparison with the latter cannot be biased as for FED and IMF which used also HP filter itself to estimate their output gap measures.

Notwithstanding the choice of λ value and time observations frequency, we found a further confirmation that the strong mitigation of the end-of-sample bias induced by SPF forecast augmentation is effective in delivering more economic meaningful output gap estimates, as more correlated with final revised policy institution output gaps estimates.

[13] Actually the third forecast observation is available just 2010 onwards. See Survey of Professional Forecasters pag. 19 for further details.

Figure (5) Statistical Real-time and Policy Institutions Final Revised Output Gap



6 Comparison to Others Methods

6.1 Comparison with Hamilton filtered output gaps

In addition to the literature on the real-time reliability of output gaps, the article by Quast and Wolters (2022) is closely related to my work. They show that Hamilton filter output gap has better real-time and forecasting properties than the original HP filter and propose a modified Hamilton filter leading to a much better coverage of typical business cycle frequencies and a smooth estimated trend.

The original Hamilton (2018) cyclical component is the 8 quarter forecast error of a local projection based on an AR(4) model, where the 8 quarter horizon is chosen because cyclical factors, such as a recession, typically occurs over the next 2 years rather than large trend changes and are the primary reason for forecast errors over such a horizon. The version modified by Quast and Wolters (2022) consists in a an equally weighted average of forecast errors based on 4 to 12 quarter ahead projections.

Because their one side nature, when I apply the SPF forecast augmentation to both versions of the filter, I don't find any improvements in real-time reliability and hence in the economic meaningfulness of the resulting output gap estimates. However, when looking at Orphanides and Norden (2002) revision statistics, Hamilton (2018) filter still yields the cyclical component less revised with new incoming data and confirms being the leader filter with real-time data. Comparing Table (4) with results by Jönsson (2020) and Quast and Wolters (2022) shows that despite SPF augmentation largely contributes to reduce the gap of real-time performance, real-time Hamilton (2018) filtered output gaps has an overlap with their final revised measure of 0.96 against 0.85 for the SPF augmented HP filter.

Notwithstanding, real-time reliability itself doesn't say anything on the economic meaningfulness of output gap estimate. I hence repeat forecasting exercises and correlations with Policy Institutions output gaps to evaluate if Hamilton (2018) filtered output gaps still outperform than HP filter, despite the SPF augmentation, as argued in Quast and Wolters (2022). Table (8) in Appendix B reports RRMSE in the spirit of Section 5.1, but comparing the SPF augmented HP filter with original Hamilton (2018) (Panel A) and with the version modified by Quast and Wolters (2022) (Panel B)[14]. Although always below the unity, results are almost never statistically significant, suggesting that the forecasting performance of the two filters is approximately the same. When I extend equation (11) and additionally control for the first difference of the output gap to account for changes in the level and the dynamics of the gap separately, as in Nelson (2008), I found very similar results. As in Section 5.1, when instead plugging forecast augmented vintages lead to the same results for GDP growth and unemployment and RRMSE very close to one for core CPI.

With regard to correlations with policy institutions estimates, results are quite mixed. Table 7 in Appendix A shows that Quast and Wolters (2022) output gap

[14]For sake of consistency, results of Table (8) and Table (7) are based on output gap vintages computed with forecast augmentation for both filters. However, when I don't apply forecast augmentation to Hamilton (2018) filter the implications for results are negligible, since its one side approach.

present higher correlation indexes with CBO and Fed output gaps, while SPF augmented HP better approximates yearly output gaps IMF and OECD. However, except for the Fed, both differences are not statistically significant.

To sum up, extending comparisons of Section 5 to Hamilton (2018) filtered output gaps shows that SPF augmented HP filter has a slightly better forecasting performance and the same correlation properties with policy institutions output gaps of Hamilton (2018) type filters. Such results closely follow Quast and Wolters (2022) findings on Hodrick and Prescott (1997) and Hamilton (2018) forecasting and correlation properties and give a very different picture of the economic meaningfulness of these two estimates in a real-time setting.

6.2 Comparison with Greenbook Output Gap

Among others, another recent paper telling the HP filter real-time unreliability is the one by Edge and Rudd (2016). They consider revision properties of Greenbook output gap estimates after the mid-1990s and found a marked improvement of such properties in the last twenty years. Conversely, statistical univariate approaches[15] didn't display the same improvements (with Hodrick and Prescott (1997) and Christiano and Fitzgerald (2003) reductions of noise to signal ratios being quite small). They complement the paper with a dynamic pseudo-out-of sample inflation forecasting exercise similar to that of Section 5.1 and found that Phillips Curve models conditioning to GB real-time output gap display a forecasting performance similar to those conditioning to the revised version of the gap and only slightly increases forecasting accuracy when compared with AR inflation forecasting models.

However, as a cross-methods comparison is missing in Edge and Rudd (2016) paper, a natural question arising in light of the forecast augmented HP filter is how GB output gap predict inflation in real-time respect to the new forecast augmented HP filter. This information is particularly valuable for two reasons. First, since GB

[15] Edge and Rudd (2016) consider deterministic filters as linear, broken-linear, quadratic trend; unobserved components approaches including Hodrick and Prescott (1997), Harvey (1985), Watson (1986), Clark (1987); Christiano and Fitzgerald (2003) band-pass filter and Beveridge and Nelson (1981) decomposition.

output gap estimates are judgemental and includes also HP filter among others in their estimation, is interesting to know if such judgemental[16] estimate outperform or not a pure bias-corrected HP filter approach. Second, while describing the evolution of measuring potential output by the Fed, Orphanides (2004) observes that such different estimation techniques 'were meant to correspond to a concept of *non-inflationary* full employment'. In this regard, repeating the pseudo-out-of-sample forecasting exercise performed in Section 5.1 and 6.1 allows to test if SPF augmented HP filter represents an improvement in GB output gap as benchmark of inflationary barrier.

Similarly to GB GDP projections, GB gap estimates are produced before each meeting of the Federal Open Market Committee (FOMC), hence meeting observations are aggregated at quarterly frequency according to principles described in Section 3.2, in order to make feasible a comparison with the quarterly SPF augmented HP filter. The first GB vintage is referred to August 1996, with the vintage series starting in 1975q1. As gap estimates are available with a five-year lag, the most recent gap estimate is from December 2018. However, I use final revised version of outcome variables as in previous sections, hence I can extend the forecast evaluation also to predictions formulated in correspondence of the last vintage 2018q4[17]. For the rest, dynamic out-of-sample simulations work exactly as in sections 5.1 and 6.1. Panel C of Table 8 in Appendix B reports results in terms of relative RMSE.

The superior performance of SPF augmented HP filter than GB output gap for inflation forecasting is sharp across forecast horizons, with improvements in forecasting accuracy also in the order of 30 %. Such marked improvements in forecasting accuracy suggest that SPF augmented HP filter represents a better tool than GB gap estimate as inflationary barrier. Notice that these results cannot be guided by the time varying GB output gap real-time reliability documented by Edge and Rudd (2016), as this forecasting exercise just relies on forecasts formulated from 1996q3

[16] Fed output gap estimate is not explicitly derived from a single model of the economy, but is rather a judgmental weight of a number of estimation techniques. See among others Orphanides (2004), Fleischman and Roberts (2011) and Coibion et al. (2018) for a brief review of time-varying popularity of methodologies employed by GB staff to estimate potential.

[17] I found it being reasonable, as for this forecasting exercise I can only conduct it using newer data set of real-time GB gaps, which runs from 1996:Q2 onwards.

onwards, period during which GB output gap performs well in a real-time setting. Hence, the higher forecasting accuracy of SPF augmented HP output gap is rather due to the fact that it really represents a better proxy of inflationary barrier than GB output gap. With regard to GDP growth and unemployment rate the picture is pretty different. For the former, SPF augmented HP output gap still outperforms than GB estimates, but improvements are very low and never statistically significant. Conversely, for unemployment rate GB is a better alternative, with statistically significant improvements in the order of 20 % for the last four forecast horizons. Usual robustness checks on specifications of equations (11) and (12) to additionally control for the first difference of the output gap lead to the same results. A potential explanation of this superior performance of GB gap vs SPF augmented HP, at odds with the other outcomes of forecasting exercises, is the intimate connection of GB output gap with Okun (1963) law. While describing the evolution of measuring potential output by the Fed, Orphanides (2004) explains that 'the specific construction methods and assumptions varied over time. During the 1960s and until 1976, the starting point was Okun's (1962) analysis'. Furthermore, more recently Fleischman and Roberts (2011) describe a methodology to compute potential output using a multivariate unobserved components model that is taken into account by the Federal Reserve Board when producing its judgmental estimates of potential output. Its procedure embeds some parts of different methodologies, including the relationship between cyclical fluctuations in output and unemployment. It is hence reasonable that GB output gap performs very well for unemployment rate forecasting, as Okun (1963) law has been a crucial input for its estimation.

Interestingly, when I repeat the same forecasting exercises performing the GB augmentation described in Section 3.2 for HP filter, I found very similar results. It means that FOMC's research staff can conveniently use its own GDP growth projections to adjust HP filter end-of-sample bias and use the resulting output gap as proxy of inflationary barrier instead its own GB output gap estimates. Alternatively, it can also use SPF predictions to adjust the filter, as are available to the public in real-time. To sum up, SPF (GB) augmented HP filter outperforms FED gap estimate to predict inflation and also displays some low improvements in GDP growth forecasting. Conversely, for unemployment rate, GB output gap is a better predictor. These results, closely follow Edge and Rudd (2016) paper on GB output gap real-time reliability, but give a very different take for what pertains its forecasting power on inflation and suggest FOMC's research staff to add a new tool to proxy inflationary barrier.

6.3 Band-Pass Filters

Unlike Hamilton (2018) filter, band-pass methods as Butterworth et al. (1930) (BW) and Baxter and King (1999) (BP) obviously also benefit of SPF augmentation to mitigate its end-of-sample bias. Table (3) and Figures (6), (7) show the effect of SPF augmentation respectively on output gap estimates and revision statistics in the spirit of Section 4.

Table (3) Output Gap Estimates Revision Statistics

	CF_{SPF}	CF	*Difference*
Final/Real-time	**0.85**	0.79	0.06*
Quasi-Real/Real-time	**0.96**	0.94	0.02
Quasi-Real/Final	**0.84**	0.77	0.07*
Total Difference/Data Revision	**0.10**	0.06	0.04
Opposite Sign Frequency	24 %	**33 %**	-9 %
NSR-Total Difference	0.53	**0.62**	-0.09***
NSR-Data Revision	**0.23**	0.23	0.00
	BW_{SPF}	BW	*Difference*
Final/Real-time	**0.85**	0.53	0.32***
Quasi-Real/Real-time	0.94	**0.94**	0.00
Quasi-Real/Final	**0.85**	0.54	0.31***
Total Difference/Data Revision	0.19	**0.36**	0.17*
Opposite Sign Frequency	22 %	**42 %**	-20 %
NSR-Total Difference	0.53	**1.03**	-0.50***
NSR-Data Revision	0.25	**0.37**	0.12***

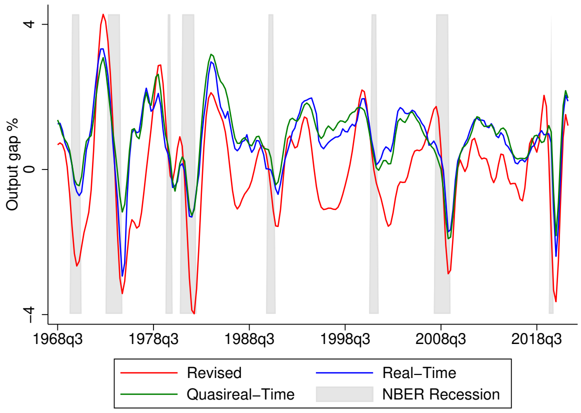
Notes: Bold denotes the most positive value. *Difference* reports the difference with respect to the SPF augmented filtered statistical real-time gap. *, **, and *** on correlation coefficients denote significant differences on the 10, 5, and 1 % level based on Fisher et al. (1921)'s z transformation. Inference on NSRs is made by the RMSE-based noise-to-signal ratio as in Edge and Rudd (2016).

BW filter is a low pass filters widely used in electrical engineering in its one-sided form. Gomez (2001) shows that HP filter may be regarded as a member of the Butterworth family of filters. Specifically, under certain assumptions, the gain function of the two-sided BW filter is the same of that of HP filter[18]. As a consequence, it is not surprising that its revision statistics are very similar to those of HP filter.

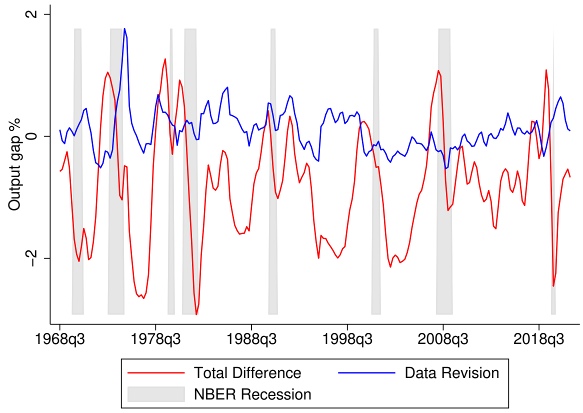
Baxter and King (1999) is a band-pass filter applying a two-sided weighted moving average to the data, where the moving average is computed using symmetric weights with a fixed length. As I deal with real-time data, for BP filter I consider

[18]See Gomez (2001) and Álvarez and Gómez-Loscos (2018) for a complete explanation.

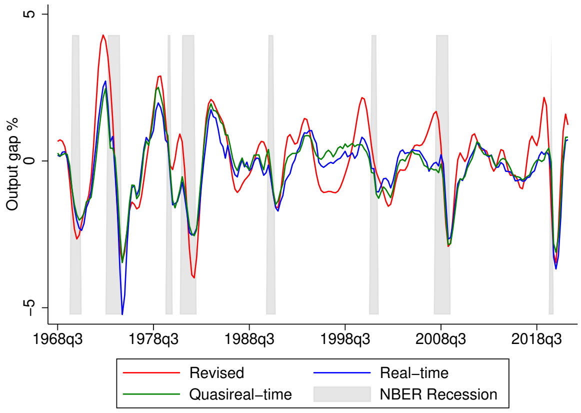
Figure (6) Christiano-Fitzgerald



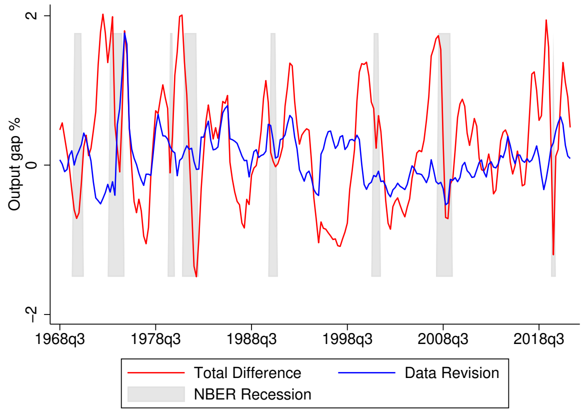
Christiano-Fitzgerald-Output Gaps



Christiano-Fitzgerald-Revision



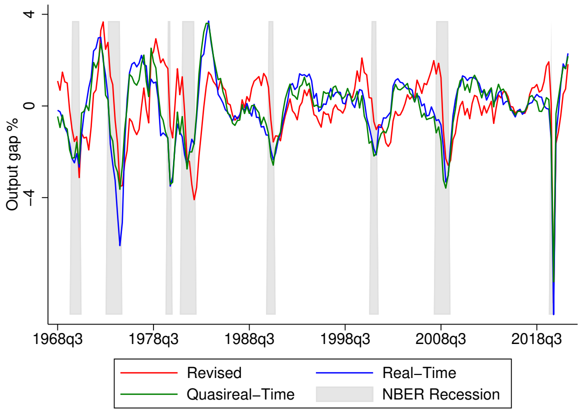
SPF Augmented CF-Output Gaps



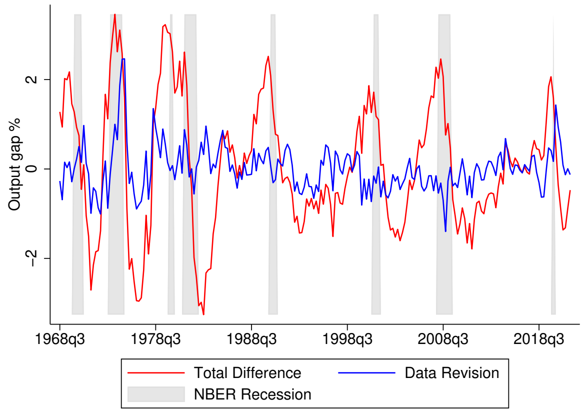
SPF Augmented CF-Revision

the implementation by Christiano and Fitzgerald (2003) (CF), who develop one-sided approximations in order to avoid losing observations at the sample end-points. Christiano and Fitzgerald (2003) showed that the Baxter and King (1999) filter end-of-sample bias vanishes only slowly as more data becomes available. Unlike of HP filter, weights of the CF version of the BP filter are based on the available sample, with end-of-sample weights being chosen optimally, hence appending forecasts to the sample more hardly reduces measurement error than in HP filter. Consistently, when Quast and Wolters (2022) augment vintages with AR(4) forecasts they don't find significant improvements. Looking at Figure (6), the real-time estimate of the gap lies generally below the final estimate, implying a positive bias for CF filter. Another way to see this is that potential output is systematically underestimated in real time. When augmenting the filter with SPF projections the positive bias

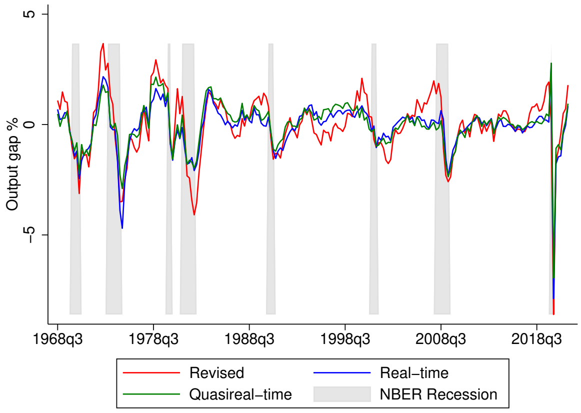
Figure (7) Butterworth



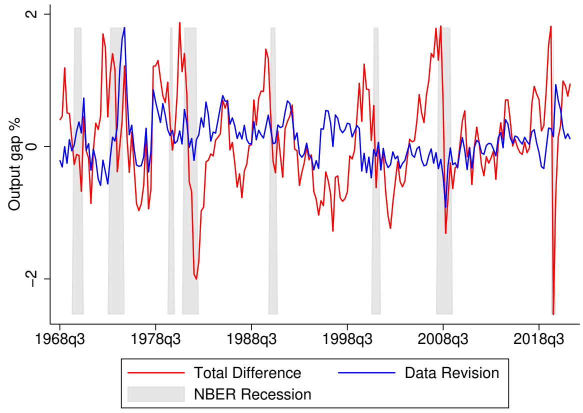
Butterworth-Output Gaps



Butterworth-Revisions



SPF Augmented BW-Output Gaps



SPF Augmented BW-Revisions

is partially absorbed, with correlation between real-time and final revised estimate increasing from 79 % to 85 % and noise to signal ratio reducing from 0.62 to 0.53.

7 The case of Canada

To check whether my results are specific to United States or have rather an external validity, I estimate a forecast augmented HP output gap also for Canada and test its real-time, correlation and forecasting properties with respect to competing estimates. The choice of Canada is to some extent forced, as to the best of my knowledge no other such extensive real-time dataset exists, apart Philadelphia's FRB, for such a long time period and inclusive of Bank of Canada (hereafter BoC) staff forecasts[19]. Although the choice of country is in some way driven by data availability, the real-time dataset provided by BoC is a great tool to extend the validity of my results, as it includes accurate GDP predictions for long forecast horizons, sometimes up to 11 years. Therefore, Champagne et al. (2018b) show that such projections are more accurate and informative than the average forecasts from Consensus Economics and those from standard econometric models. Hence, building on BoC real-time dataset inclusive of projections and historical values, I estimate a *BoC forecast augmented* HP filter. However, the main limitation with BoC forecast augmentation is due to real-time availability of projections and historical values. Indeed, similarly to Greenbook dataset there is a five year ban on access to staff data, preventing the public to construct a BoC forecast augmented HP filter in real-time. Notwithstanding such limitation, Bank of Canada' staff and Governing Council have these projections in real-time and can hence conveniently adjust the HP filter to mitigate its end-of-sample bias.

Similarly to US, BoC's staff also prepares official output gap estimates for Governing Council; its estimation is judgmental, in the sense that output gap estimates are based on different sources of information and economic models, including HP filter [20]. Hence, the BoC augmented HP filter I propose in this chapter can be a reliable tool also for boosting BoC's staff official gap measure, as well as improving the real-time performance of HP filter itself.

With regard to Canadian output gap estimation, Cayen and Van Norden (2005) first

[19] Champagne et al. (2018a) is an excellent guide on the potential of Bank of Canada real-time dataset.

[20] See Pichette et al. (2015) for a complete picture of recent methods

investigated the reliability of statistical detrending methods and showed that real-time output gap estimates are subject to large revisions as in U.S. More recently, Champagne et al. (2018b) has exploited the BoC real-time dataset and showed that the real-time performance of the BoC's staff official output gap has increased since the early 2000s.

Table (4) reports revision statistics for HP filter, the *BoC augmented* version and the official output gap estimate prepared by Bank of Canada's staff. Real-time estimates are constructed as usual with vintages available from 1986q4 onward in a time sample starting from 1967q1[21]. Because Champagne et al. (2018b) show that

Table (4) Output Gap Estimates Revision Statistics

	HP_{BoC}	HP	BoC Staff
1986q4–2000q1			
Final/Real-time	0.84	0.28***	0.88
Opposite Sign Frequency	15 %	**44%**	9%
NSR-Total Difference	0.69	**1.20*****	1.13***
2000q2–2016q4			
Final/Real-time	0.90	0.61***	0.94
Opposite Sign Frequency	27%	**36%**	15 %
NSR-Total Difference	0.55	**0.92*****	0.39***
1986q4–2016q4			
Final/Real-time	0.84	0.42***	0.65***
Opposite Sign Frequency	22 %	**40%**	17 %
NSR-Total Difference	0.63	1.09***	**1.10*****

Notes: Bold denotes the most positive value. *, **, and *** denote significant differences respect to the augmented HP filter on the 10, 5, and 1 % level based on Fisher et al. (1921)'s z transformation. Inference on NSRs is made by the RMSE-based noise-to-signal ratio as in Edge and Rudd (2016).

[21] Because of data missing, some imputations have to be made as for FRB dataset. For vintages 2001q2-2002q2 data up to 1980q4 are missing, hence I use observations of 2002q3 vintage to impute these missing data.

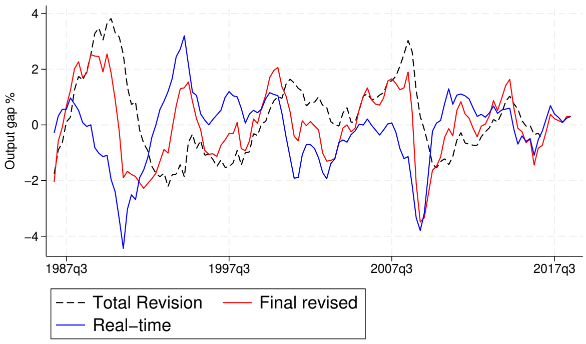
real-time reliability of BoC's staff output gap has increased over time, I report revision statistics for two subsamples 1986q4–2000q1 and 2000q2–2016q4, as well as for the full sample. In order to align as much as possible my results to theirs, in Table (4) I report Noise-to signal-ratio defined as the ratio between root mean square of total revision and standard deviation of final revised estimate. Notice that, for 1986q4–2000q1 subsample the final revised estimate is 2002q2 output gap vintage, whereas for 2000q2-2016q4 and for the full sample is the last available data vintage, which is 2018q4. As for US, I discard the last two years of observations in order to rely only on final revised data , hence my full sample in 1986q4-2016q4. Since BoC projections are available in real-time and in level, for Canada I cannot simulate quasireal-time estimates and infer the role of data revision for BoC's staff output gap and BoC augmented HP filter.

In line with US, HP filter display the worst real-time performance across subsamples, but BoC augmentation successfully mitigates its end-of-sample bias and halved the NSR. The BoC's staff real-time output gap, as shown in Figure (8), has decreased over time its downward bias and improved its real-time performance.

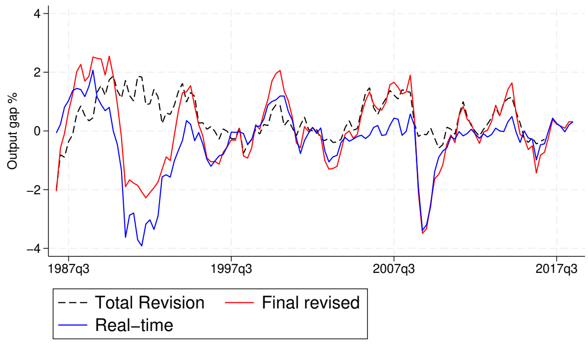
Table (5) reports relative RMSEs as in previous Tables (2), comparing the BoC augmented HP filter with its original version. I adopt the same forecasting models of equations (11), (12) and (13) respectively for GDP growth, unemployment and inflation. As for US, on the left hand side of each equation I use final revised data. This choice allows me to extend the forecasts evaluation also to predictions formulated in the last vintage 2018q4. The initial sample runs from 1967Q1 to 1986Q3 and it's recursively expanded quarter by quarter[22] plugging recursively the estimated real-time output gap vintages, following the pseudo-out-of-sample forecasting methodology à là Stock and Watson (2007). As for US, I found forecast augmentation improving the forecasting accuracy for all outcomes, particularly for GDP growth and unemployment rate. For GDP growth specification, controlling

[22] Although imputations described in previous footnote there are some missing data for first observations in later GDP vintages. The effect on the forecasting exercise is somehow to estimate parameters in a rolling widow. For example, for vintages 1987q1-1997q3 the first observation is 1967q1, for vintages 1997q4-2002q3 is 1968q4, for 2002q4-2010q4 is 1971q1, for 2011q1-2015q4 is 1973q1 and last vintages up to 2018q4 start in 1982q2. See Staff economic projections real-time dataset for further details.

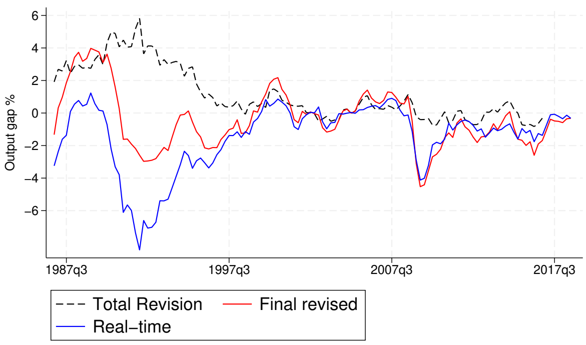
Figure (8) Output Gaps and Total Revision



(a) Hodrick-Prescott



(b) BoC Augmented HP



(c) BoC's staff

for the first difference of the gap delivers very similar results. When I compare the performance of BoC augmented HP filter with BoC official gap and with Hamilton (2018) filtered output gaps I found all relative RMSEs ranging very close to unity, meaning that the three gap estimates display a very similar forecasting performance. Last, in the spirit of Section 5.2 I compute correlations between real-time statistical

Table (5) Relative RMSE for forecasting

Horizon	GDP growth	Unemployment	Core Inflation
1	0.93***	1.06***	0.97
2	0.89***	0.99	0.96**
3	0.86***	0.92***	0.98**
4	0.85***	0.87***	0.95**
5	0.83***	0.83***	0.97
6	0.83**	0.81***	0.98
7	0.85**	0.83***	0.98
8	0.84**	0.82***	0.98

Notes: *, **, and *** denote significance on the 10, 5, and 1% level based on a two-sided Diebold and Mariano (1995) test.

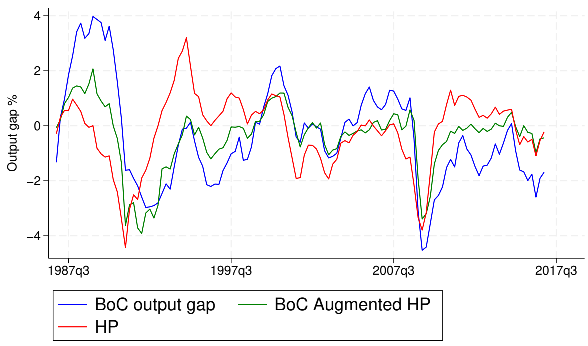
output gap estimates and final revised policy institutions output gaps. As for US, IMF (OECD) estimates output gap at annual frequency from 1980 (1985) onwards and hence also large difference in correlation indexes are not statistically significant because the small sample. As explained above, BoC's staff output gap estimate is judgemental, similarly to Greenbook's one and estimated at quarterly frequency for 1982q2-2018q3 time sample. BoC Augmented HP filter has the highest correlation indexes with all policy institutions output gaps, also respect to Hamilton (2018) filtered output gaps. Such results are somehow not totally surprising, as widespread popularity of HP filter among policy institutions (see Section 5.2). Indeed, as already seen for US in absence of end-of-sample distortion HP filter successfully reflect the ex post expert evaluation of business cycle.

Table (6) Correlations with Policy Institutions Final Revised Output Gaps

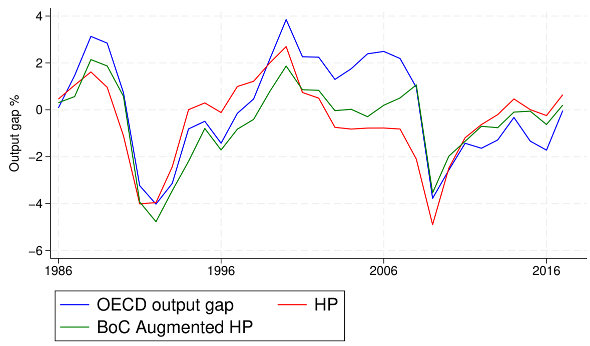
Output Gap	BoC Staff	IMF	OECD
BoC-$HP_{\lambda=1600/100}$	**0.79**	**0.86**	**0.89**
Quast et al. (2022)	0.69*	0.74	0.78
Hamilton (2018)	0.65	0.72	0.74
$HP_{\lambda=1600/100}$	0.18	0.67	0.71
BoC-$HP_{\lambda=6.25}$	-	0.85	0.78
$HP_{\lambda=6.25}$	-	0.30	0.21
Diff. BoC-$HP_{\lambda=1600/100}$ - $HP_{\lambda=1600/100}$	0.61***	0.19*	0.18**
Diff. BoC-$HP_{\lambda=6.25}$ - $HP_{\lambda=6.25}$	-	0.63***	0.57***
Diff. BoC-$HP_{\lambda=1600/100}$ - Quast et al.(2022)	0.10*	0.12	0.12

Notes: Bold denotes the most positive correlation. $_{\lambda=1600/100}$ means Hodrick and Prescott (1997) λ parameters equals to 1600 for quarterly data and 100 for annual frequency. *, **, and *** denote significant differences respect to the augmented HP filter on the 10, 5, and 1 % level based on Fisher et al. (1921)'s z transformation.

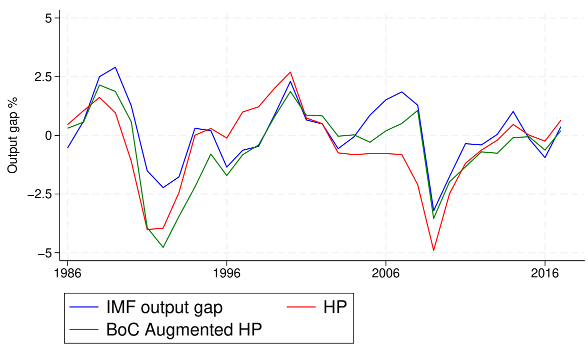
Figure (9) Statistical Real-time and Policy Institutions Final Revised Output Gap



(a) Bank of Canada Staff



(b) OECD



(c) IMF

8 Conclusion

I construct a forecast augmented HP filter to mitigate the end-of-sample bias of the popular Hodrick and Prescott (1997) detrending method. In particular, this paper proposes an application of the new forecast augmented HP filter for US output gap estimation. After exploring different alternatives of forecast augmentations, I found using GDP vintages in conjunction with median SPF forecasts is effective not only to improve the real-time reliability of output gap estimate, but also to improve its forecasting performance for output growth, unemployment rate and inflation by Phillips curve. The performance of the SPF augmented HP filter is also compared to Hamilton (2018) one-sided filter and its modified version by Quast and Wolters (2022). Both forecasting exercises and correlation with policy institutions final revised output gaps show that the two methods display similar performances, with SPF augmented HP output gap slightly outperforms Hamilton (2018) filtered gaps for output growth forecasting. Above all, results of recent contributions (Hamilton (2018), Champagne et al. (2018b), Jönsson (2020), Quast and Wolters (2022)) arguing a better performance of Hamilton (2018) filtered output gaps are highly revised. Therefore, marginal contribution of this paper is to get HP filter (and others two-side methods) back on track for the horse race on output gap estimation. It is particularly important for two reasons. First, the filter is very popular both in academia and in policy institutions estimating potential GDP, so proposing a new approach which preserves its real-time performance is useful for policy purposes. Second, since the issue of real-time reliability is now highly mitigated, future research has now a more sizeable extensive margin to focus on providing *economic meaningful* business cycle estimates instead of *real-time reliable* ones. With this regard, in this paper I evaluate the performance of competing output gap estimates considering their predictive content for inflation and GDP growth or measuring correlations with final revised institutional output gaps. However, alternative assessment criteria on the meaningful of output gap has been proposed by Coibion et al. (2018), who study how potential output estimates respond to economic shocks or Canova (2023), who investigates the gap properties in theoretical models.

References

L. J. Álvarez and A. Gómez-Loscos. A menu on output gap estimation methods. *Journal of Policy Modeling*, 40(4):827–850, 2018.

L. Ball, D. Leigh, and P. Loungani. Okun's law: Fit at 50? *Journal of Money, Credit and Banking*, 49(7):1413–1441, 2017.

M. Baxter and R. G. King. Measuring business cycles: approximate band-pass filters for economic time series. *Review of economics and statistics*, 81(4):575–593, 1999.

P. Beaudry, D. Galizia, and F. Portier. Putting the cycle back into business cycle analysis. *American Economic Review*, 110(1):1–47, 2020.

P.-O. Beffy, P. Ollivaud, P. Richardson, and F. Sédillot. New oecd methods for supply-side and medium-term assessments: a capital services approach. 2006.

T. Bernhardsen, Ø. Eitrheim, A. S. Jore, and Ø. Røisland. Real-time data for norway: challenges for monetary policy. *The North American Journal of Economics and Finance*, 16(3):333–349, 2005.

S. Beveridge and C. R. Nelson. 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): 151–174, 1981.

S. Butterworth et al. On the theory of filter amplifiers. *Wireless Engineer*, 7(6): 536–541, 1930.

G. Camba-Mendez and D. Rodriguez-Palenzuela. Assessment criteria for output gap estimates. *Economic modelling*, 20(3):529–562, 2003.

F. Canova. Faq: How do i estimate the output gap? 2023.

J.-P. Cayen and S. Van Norden. The reliability of canadian output-gap estimates. *The North American Journal of Economics and Finance*, 16(3):373–393, 2005.

J. Champagne, G.-P. Bellisle, and R. Sekkel. Evaluating the bank of canada staff economic projections using a new database of real-time data and forecasts. Technical report, Bank of Canada Staff Working Paper, 2018a.

J. Champagne, G. Poulin-Bellisle, and R. Sekkel. The real-time properties of the bank of canada's staff output gap estimates. *Journal of Money, Credit and Banking*, 50(6):1167–1188, 2018b.

L. J. Christiano and T. J. Fitzgerald. The band pass filter. *international economic review*, 44(2):435–465, 2003.

P. K. Clark. The cyclical component of us economic activity. *The Quarterly Journal of Economics*, 102(4):797–814, 1987.

T. E. Clark and M. W. McCracken. The predictive content of the output gap for inflation: Resolving in-sample and out-of-sample evidence. *Journal of Money, Credit and Banking*, pages 1127–1148, 2006.

O. Coibion and Y. Gorodnichenko. What can survey forecasts tell us about information rigidities? *Journal of Political Economy*, 120(1):116–159, 2012.

O. Coibion and Y. Gorodnichenko. Information rigidity and the expectations formation process: A simple framework and new facts. *American Economic Review*, 105(8):2644–2678, 2015.

O. Coibion, Y. Gorodnichenko, and M. Ulate. The cyclical sensitivity in estimates of potential output. *Brookings Papers on Economic Activity*, 2018(2):343–434, 2018.

D. Comin and M. Gertler. Medium-term business cycles. *American Economic Review*, 96(3):523–551, 2006.

T. F. Cooley and L. E. Ohanian. The cyclical behavior of prices. *Journal of Monetary Economics*, 28(1):25–60, 1991.

A. Cornea-Madeira. The explicit formula for the hodrick-prescott filter in a finite sample. *Review of economics and statistics*, 99(2):314–318, 2017.

I. H. Correia, J. C. d. Neves, and S. Rebelo. Business cycles in portugal: Theory and evidence. In *The Portuguese economy towards 1992*, pages 1–64. Springer, 1992.

R. T. Cusinato, A. Minella, and S. da Silva Pôrto Júnior. Output gap in brazil: a real-time data analysis. *Empirical Economics*, 44:1113–1127, 2013.

C. De Resende. An assessment of imf medium-term forecasts of gdp growth. *IEO Background Paper No. BP/14/01 (Washington: Independent Evaluation Office of the IMF)*, 2014.

D. A. Dickey and W. A. Fuller. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, 74 (366a):427–431, 1979.

F. X. Diebold and R. S. Mariano. Com paring predictive accu racy. *Journal of Business and Economic Statistics*, 13(3):253–263, 1995.

R. M. Edge and J. B. Rudd. Real-time properties of the federal reserve's output gap. *Review of Economics and Statistics*, 98(4):785–791, 2016.

Y. Eo and J. Morley. Does the survey of professional forecasters help predict the shape of recessions in real time? *Economics Letters*, 233:111419, 2023.

J. Faust and J. H. Wright. Forecasting inflation. In *Handbook of economic forecasting*, volume 2, pages 2–56. Elsevier, 2013.

R. A. Fisher et al. 014: On the" probable error" of a coefficient of correlation deduced from a small sample. 1921.

C. A. Fleischman and J. M. Roberts. From many series, one cycle: improved estimates of the business cycle from a multivariate unobserved components model. 2011.

J. C. Fuhrer, Y. K. Kodrzycki, J. S. Little, G. P. Olivei, and P. A. Samuelson. *Understanding inflation and the implications for monetary policy: A Phillips curve retrospective.* MIT Press Cambridge, MA, 2009.

A. Garratt, K. Lee, E. Mise, and K. Shields. Real-time representations of the output gap. *The Review of Economics and Statistics*, 90(4):792–804, 2008.

V. Gomez. The use of butterworth filters for trend and cycle estimation in economic time series. *Journal of Business & Economic Statistics*, 19(3):365–373, 2001.

M. Gonzalez-Astudillo. Gdp trend-cycle decompositions using state-level data. 2017.

J. D. Hamilton. Why you should never use the hodrick-prescott filter. *Review of Economics and Statistics*, 100(5):831–843, 2018.

A. C. Harvey. Trends and cycles in macroeconomic time series. *Journal of Business & Economic Statistics*, 3(3):216–227, 1985.

R. J. Hodrick. An exploration of trend-cycle decomposition methodologies in simulated data. Technical report, National Bureau of Economic Research, 2020.

R. J. Hodrick and E. C. Prescott. Postwar us business cycles: an empirical investigation. *Journal of Money, credit, and Banking*, pages 1–16, 1997.

K. Jönsson. Real-time us gdp gap properties using hamilton's regression-based filter. *Empirical Economics*, 59(1):307–314, 2020.

R. Kaiser and A. Maravall. Estimation of the business cycle: A modified hodrick-prescott filter. *Spanish Economic Review*, 1:175–206, 1999.

K. Kamada. Real-time estimation of the output gap in japan and its usefulness for inflation forecasting and policymaking. *The North American Journal of Economics and Finance*, 16(3):309–332, 2005.

G. Kamber, J. Morley, and B. Wong. Intuitive and reliable estimates of the output gap from a beveridge-nelson filter. *Review of Economics and Statistics*, 100(3): 550–566, 2018.

G. Kamber, J. Morley, and B. Wong. Trend-cycle decomposition after covid cama working paper 24/2024 march 2024. 2024.

M. Marcellino and A. Musso. The reliability of real-time estimates of the euro area output gap. *Economic Modelling*, 28(4):1842–1856, 2011.

M. Marcellino, J. H. Stock, and M. W. Watson. A comparison of direct and iterated multistep ar methods for forecasting macroeconomic time series. *Journal of econometrics*, 135(1-2):499–526, 2006.

Y. P. Mehra. The output gap, expected future inflation and inflation dynamics: another look. *Contributions in Macroeconomics*, 4(1):20121030, 2004.

E. Mise, T.-H. Kim, and P. Newbold. On suboptimality of the hodrick–prescott filter at time series endpoints. *Journal of Macroeconomics*, 27(1):53–67, 2005.

J. Morley and I. B. Panovska. Is business cycle asymmetry intrinsic in industrialized economies? *Macroeconomic Dynamics*, 24(6):1403–1436, 2020.

C. R. Nelson. The beveridge–nelson decomposition in retrospect and prospect. *Journal of Econometrics*, 146(2):202–206, 2008.

A. M. Okun. *Potential GNP: its measurement and significance.* Cowles Foundation for Research in Economics at Yale University, 1963.

A. Orphanides. Monetary policy rules, macroeconomic stability, and inflation: A view from the trenches. *Journal of Money, Credit and Banking*, pages 151–175, 2004.

A. Orphanides and S. v. Norden. The unreliability of output-gap estimates in real time. *Review of economics and statistics*, 84(4):569–583, 2002.

A. Orphanides and S. Van Norden. The reliability of inflation forecasts based on output gap estimates in real time. *Journal of Money, Credit and Banking*, pages 583–601, 2005.

P. C. Phillips and S. Jin. Business cycles, trend elimination, and the hp filter. *International Economic Review*, 62(2):469–520, 2021.

L. Pichette, P. St-Amant, B. Tomlin, and K. Anoma. Measuring potential output at the bank of canada: The extended multivariate filter and the integrated framework. Technical report, Bank of Canada Discussion Paper, 2015.

J. Quast and M. H. Wolters. Reliable real-time output gap estimates based on a modified hamilton filter. *Journal of Business & Economic Statistics*, 40(1): 152–168, 2022.

M. O. Ravn and H. Uhlig. On adjusting the hodrick-prescott filter for the frequency of observations. *Review of economics and statistics*, 84(2):371–376, 2002.

R. Shackleton. *Estimating and projecting potential output using CBO's forecasting growth model*. Congressional Budget Office, 2018.

F. Smets and R. Wouters. Shocks and frictions in us business cycles: A bayesian dsge approach. *American economic review*, 97(3):586–606, 2007.

T. Stark et al. Realistic evaluation of real-time forecasts in the survey of professional forecasters. *Federal Reserve Bank of Philadelphia Research Rap, Special Report*, 1, 2010.

J. H. Stock and M. W. Watson. Business cycle fluctuations in us macroeconomic time series. *Handbook of macroeconomics*, 1:3–64, 1999.

J. H. Stock and M. W. Watson. Why has us inflation become harder to forecast? *Journal of Money, Credit and banking*, 39:3–33, 2007.

M. W. Watson. Univariate detrending methods with stochastic trends. *Journal of monetary economics*, 18(1):49–75, 1986.

E. Wolf, F. Mokinski, and Y. S. Schüler. On adjusting the one-sided hodrick-prescott filter. *Journal of Money, Credit and Banking*, forthcoming.

A Appendix

Table (7) Correlations between statistical real-time and Policy Institutions final revised output gap estimates.

Output Gap	CBO	FED	IMF	OECD
SPF-$HP_{\lambda=1600/100}$	0.70	0.68	**0.82**	**0.87**
$HP_{\lambda=1600/100}$	0.43	0.32	0.73	0.67
Diff.	0.26***	0.36***	0.09	0.20**
SPF-$HP_{\lambda=6.25}$	-	-	0.64	0.75
$HP_{\lambda=6.25}$	-	-	0.30	0.28
Diff.	-	-	0.34**	0.47**
Quast et al. (2022)	**0.75**	**0.78**	0.75	0.78
Hamilton (2018)	0.70	0.72	0.65	0.68
Diff. with Quast et al.(2022)	-0.05	-0.10**	0.07	0.09
Diff. with Hamilton (2018)	0.00	-0.04	0.17	0.19**

Notes: Bold denotes the most positive correlation. SPF-$HP_{\lambda=1600/100}$ denotes output gap estimated with original Hodrick and Prescott (1997) λ parameters equals to 1600 for quarterly data (correlations with CBO and Fed) and 100 for annual frequency (correlations with IMF and OECD).*, **, and *** denote significant differences on the 10, 5, and 1 % level, based on Fisher et al. (1921)'s z transformation, respect to SPF-HP.

B Appendix

Table (8) Relative RMSE

Panel A: Hamilton (2018)			
horizon	GDP growth	Unemployment rate	Core CPI inflation
1	0.94	0.98	0.96
2	0.96	0.97	0.94*
3	0.96	0.96	0.93
4	0.97	0.96	0.94
5	0.97	0.95	0.94
6	0.97	0.94	0.95
7	0.98	0.98*	0.96
8	0.98	0.97*	0.99
9	0.99	0.98	1.00
10	1.00	0.98	1.01
11	1.00	0.97	1.01
12	1.01	0.97	1.00
Panel B: Quast and Wolters (2022)			
1	0.95	0.98	0.99
2	0.96	0.97	0.97
3	0.97	0.96	0.96
4	0.97	0.96	0.97
5	0.96**	0.95	0.97
6	0.96**	0.95	0.98
7	0.96	0.98	0.98
8	0.96	0.98	1.00
9	0.97	0.98	1.02
10	0.98	0.98	1.03
11	0.99	0.98	1.03
12	1.00	0.98	1.03

Panel C: Greenbook Output Gap			
horizon	GDP growth	Unemployment rate	Core CPI inflation
1	1.01	1.00	0.85*
2	0.99	1.02*	0.78*
3	0.97	1.03	0.73*
4	0.96	1.05	0.68*
5	0.94	1.06	0.65*
6	0.94	1.08	0.66*
7	0.95	1.10**	0.65*
8	0.95	1.13**	0.65*
9	0.95	1.16**	0.64*
10	0.94	1.20***	0.62*
11	0.93	1.22***	0.63*
12	0.92	1.25***	0.66*

Notes: The three panels reports relative RMSE of the SPF augmented HP filter respect to Hamilton (2018), Quast and Wolters (2022) and greenbook output gaps. *, **, and *** denote significance on the 10, 5, and 1% level based on a two-sided Diebold and Mariano (1995) test.