

KIEL WORKING PAPER

**Reliable Real-time
Output Gap Estimates
Based on a Modified
Hamilton Filter**



No. 2158 June 2020

Josefine Quast and Maik H. Wolters

ABSTRACT

RELIABLE REAL-TIME OUTPUT GAP ESTIMATES BASED ON A MODIFIED HAMILTON FILTER

Josefine Quast and Maik H. Wolters

We propose a simple modification of Hamilton's (2018) time series filter that yields reliable and economically meaningful real-time output gap estimates. The original filter relies on 8 quarter ahead forecast errors of a simple autoregression of real GDP. While this approach yields a cyclical component that is hardly revised with new incoming data due to the one-sided filtering approach, it does not cover typical business cycle frequencies evenly, but mutes short and amplifies medium length cycles. Further, as the estimated trend contains high frequency noise, it can hardly be interpreted as potential GDP. A simple modification based on the mean of 4 to 12 quarter ahead forecast errors shares the favorable real-time properties of the Hamilton filter, but leads to a much better coverage of typical business cycle frequencies and a smooth estimated trend. Based on output growth and inflation forecasts and a comparison to revised output gap estimates from policy institutions, we find that real-time output gaps based on the modified and the original Hamilton filter are economically much more meaningful measures of the business cycle than those based on other simple statistical trend-cycle decomposition techniques, such as the HP or bandpass filter, and should thus be used preferably.

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

JEL classification: C18, E32, E37

Josefine Quast

Julius-Maximilians-Universität Würzburg
Sanderring 2
D-97070 Würzburg
Email: josefine.quast@uni-wuerzburg.de
www.wiwi.uni-wuerzburg.de

Maik Wolters

Julius-Maximilians-Universität Würzburg
Sanderring 2
D-97070 Würzburg
Email: maik.wolters@uni-wuerzburg.de
www.wiwi.uni-wuerzburg.de

The paper in its present form is accepted for publication in the Journal of Business & Economic Statistics. The authors thank the Associate Editor, two anonymous referees, Tino Berger, Marco Del Negro, Daniel Fehrlé, James D. Hamilton, Katja Heinisch, Gary Koop, Helmut Lütkepohl, Dieter Nautz, Christian Pigorsch, Pierre-Alain Pionnier, Franck Portier, Tara M. Sinclair, Benjamin Wong, Christopher Zuber, and participants at several workshops and conferences for helpful comments and discussions. Further, we thank Christoph Schleicher and Mark Watson for sharing their codes.

The responsibility for the contents of this publication rests with the author, not the Institute. Since working papers are of a preliminary nature, it may be useful to contact the author of a particular issue about results or caveats before referring to, or quoting, a paper. Any comments should be sent directly to the author.

1 Introduction

Ever since Orphanides and van Norden (2002) have provided evidence for the poor real-time performance of commonly used output gap estimation methods, the debate regarding the reliability of output gap estimates has been vivid. Recently, Edge and Rudd (2016) and Champagne et al. (2018) have shown that the reliability of Federal Reserve and Bank of Canada staff output gap estimates has increased since the mid-1990s, while the reliability of purely statistical detrending procedures, like, for example, the Hodrick-Prescott (HP) filter, continues to be poor.

In a recent article, Hamilton (2018) has proposed a new regression based filter for detrending time series as an alternative to the HP filter. Hamilton proposes to use the 8 quarter forecast error of a projection based on an AR(4) model as the cyclical component of a macroeconomic time series. The filter produces a stationary cycle for a wide range of time series, but suffers much less from end-of-sample bias than the HP filter and the other trend-cycle decomposition methods considered by Orphanides and van Norden (2002). Further, the creation of spurious cycles and an ad hoc choice of the smoothing parameter are avoided (see, e.g., King and Rebelo, 1993; Harvey and Jaeger, 1993; Cogley and Nason, 1995; Canova, 1998, for problems with the HP filter).

Hamilton applies the filter to US GDP and shows that its cyclical component turns negative during NBER defined recessions and positive during expansions. Hence, it is potentially not only a useful detrending method, but also applicable for output gap estimation. This would solve the long-standing problem of the unreliability of real-time output gap estimates, as the Hamilton filter is a one-sided filter. Based on spectral density analysis we show, however, that the filter does not cover typical business cycle frequencies from 6 to 32 quarters evenly. Cycles of lengths between 10 and 20 quarters are substantially amplified relative to longer and shorter cycles. The latter are muted almost completely. Further, the extracted GDP trend is not smooth. We show that this is caused by the forecast-based filter that maps high frequency noise, generally not associated with economically meaningful fluctuations, in GDP 8 quarters ago into current trend estimates. Hence, the Hamilton filter yields a very noisy measure of potential GDP.

We propose a simple modification that shares the favorable real-time properties of the Hamilton filter, but leads to a more even coverage of typical business cycle frequencies and a smooth trend. Rather than using a fixed 8 quarter forecast horizon, we take a simple average of forecast errors based on forecast horizons ranging from 4 to 12 quarters. Through this, short, medium, and long

business cycles are covered more evenly. The modified filter further avoids spikes in the cyclical component of GDP and yields a smooth trend of GDP. Hence, a clear interpretation of the trend as potential GDP and of the cyclical component as the output gap is possible. The modified filter is still centered around the 8 quarter horizon proposed by Hamilton and it is similarly easy to compute as Hamilton's original approach.

We analyze the reliability of US output gaps computed with the modified Hamilton filter and compare it to output gaps computed with the original one, the HP, and the bandpass (BP) filter as examples of commonly used simple statistical trend-cycle decomposition techniques. In particular, we answer two research questions. First, are the different real-time output gap estimates reliable, i.e. are subsequent revisions small? Second, are they economically meaningful measures of the business cycle?

To evaluate the revision properties, we compute output gaps using real-time data vintages and compare them to those based on revised data. We find that the revisions of output gaps based on the modified and the original Hamilton filter are small and mainly due to revisions in the underlying data. There are two reasons for the small revisions. First, the Hamilton filter is primarily a one-sided filter, though Hamilton recommends to use the whole available sample to estimate the AR(4) parameters. Second, parameters of univariate AR models for log real GDP are particularly stable in comparison to parameters of more complicated multivariate models. Contrary to that, HP and BP filter estimates suffer from large end-point problems due to their two-sided nature and exhibit revisions as large as the gaps themselves.

Evaluating the meaningfulness of output gap estimates and comparing competing output gap estimates in this regard is difficult because there is no clear benchmark and the true cycle is unknown. We analyze the meaningfulness of output gaps from different angles to achieve nevertheless convincing results.

First, we compare real-time output gap estimates to revised output gap estimates of the Federal Reserve, the Congressional Budget Office, the IMF, and the OECD. While there is no true cycle that can be used as a benchmark, important policy institutions should have, at least in retrospect, expert knowledge on the size and length of past business cycle phases. At the very least such expert benchmarks can be helpful in measuring business cycle characteristics that matter from a practitioner's perspective. We find that the correlations of the ex post institutional output gaps with the two Hamilton based real-time output gaps are significantly stronger compared to those

based on the HP and BP filter.

Second, we test whether an output gap estimate has predictive content for output growth and inflation. For example, if output was below potential, this implies that output growth should increase in the future so that output reverts back to potential. We find significant improvements in output growth forecasting accuracy when Hamilton-type real-time output gaps are used instead of HP or BP filtered ones. Further, according to Phillips curve models, output gaps should have implications for future inflation. Evaluating out-of-sample inflation forecasts by means of a standard Phillips curve forecasting equation, manifests findings in the literature that no statistically meaningful distinction can be made between differently filtered gap measures (see, e.g., Edge and Rudd, 2016; Champagne et al., 2018; Kamber et al., 2018). This reflects the general difficulty of beating univariate inflation forecast models with output gap based models (Atkeson and Ohanian, 2001; Fisher et al., 2002; Orphanides and van Norden, 2005; Stock and Watson, 2007, 2008; Faust and Wright, 2013) rather than output gap measurement problems.

In order to check whether our results are specific to US data, we repeat all evaluation exercises based on data for the UK and Germany. Overall the results are similar to the US case, though the BP filter does not perform significantly worse than the modified Hamilton filter when considering their correlation with ex post output gap estimates by policy institutions.

Overall, we find that output gap estimates based on the modified and the original Hamilton filter have favorable real-time properties and are meaningful measures of the business cycles with much fewer drawbacks than those based on other simple statistical trend-cycle decomposition methods. The original Hamilton filter is in general useful for detrending and in particular for consistent comparisons of theoretical stationary models and nonstationary observed data. The proposed modification makes the Hamilton filter applicable to output gap measurement, as it leads to a more even coverage of business cycle frequencies and a smooth trend that can be interpreted as measuring potential output. Our analysis of the applicability of the Hamilton filter to business cycle measurement and the proposed modification are of high relevance since many authors do not use the Hamilton filter merely for producing a stationary time series, but for interpreting trend and cycle measures economically (see, among others, López-Salido et al., 2017; Van Zandweghe, 2017; Bordo and Siklos, 2018; Danielsson et al., 2018; Richter et al., 2019; Hamilton, 2019; Ahn and Hamilton, 2020; Richter et al., 2020). According to our example of output gap estimation, it is advisable in such cases to check whether the filter produces trend and cycle that are in line with

theoretical concepts and adjust the filter if needed.

2 A Simple Modification of the Hamilton Filter

Hamilton (2018) proposes to use the 8 quarter forecast error of a projection based on an AR(4) model as the cyclical component of a macroeconomic time series. The 8 quarter horizon is chosen because cyclical factors, such as whether a recession occurs over the next 2 years rather than large trend changes, are the primary reason for forecast errors over such a horizon. Hence, to detrend a macroeconomic time series, y_t , the following simple autoregression can be estimated by OLS:

$$y_t = \beta_0 + \beta_1 y_{t-8} + \beta_2 y_{t-9} + \beta_3 y_{t-10} + \beta_4 y_{t-11} + v_t. \quad (1)$$

The cyclical component is given by the residual \hat{v}_t :

$$\hat{v}_t = y_t - \hat{\beta}_0 - \hat{\beta}_1 y_{t-8} - \hat{\beta}_2 y_{t-9} - \hat{\beta}_3 y_{t-10} - \hat{\beta}_4 y_{t-11}. \quad (2)$$

Applied to log quarterly real GDP, it is tempting to interpret \hat{v}_t as an output gap. However, Hamilton (2018) and Schüller (2018) remark that cycles of 8, 4, 8/3 and 2 quarters are muted and even completely eliminated for the special case of the difference filter to which the filter reduces when being applied to a random walk. Hence, typical business cycle frequencies between 6 and 32 quarters (Burns and Mitchell, 1946; Stock and Watson, 1999a) are not covered evenly and especially short business cycles of around 2 years lengths are eliminated or considerably dampened.

To get an output gap that covers business cycle frequencies from 6 to 32 quarters more evenly, we propose a simple modification of the Hamilton filter. Rather than using a fixed 8 quarter horizon, we propose using an equally weighted average of forecast errors based on 4 to 12 quarter ahead projections to estimate the output gap, \tilde{y}_t :

$$\tilde{y}_t = 1/9 \sum_{i=4}^{12} \hat{v}_{t,i}, \quad \text{with} \quad (3)$$

$$\hat{v}_{t,i} = y_t - \hat{\beta}_{0,i} - \hat{\beta}_{1,i} y_{t-i} - \hat{\beta}_{2,i} y_{t-i-1} - \hat{\beta}_{3,i} y_{t-i-2} - \hat{\beta}_{4,i} y_{t-i-3}. \quad (4)$$

In the following, we show based on analyses in the frequency domain that this modification changes the cyclical properties of the Hamilton filter in a way that makes it applicable for output gap estimation.

The spectral density of a covariance stationary process y_t at frequency $\omega \in [0, \pi]$ is given by the Fourier transform of the autocovariance function (see, e.g., Canova, 2011, p. 19):

$$SD_y(\omega) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} ACF_y(\tau) e^{-i\omega\tau}, \quad (5)$$

where $i = \sqrt{-1}$ and ω is measured in radians. Further, a time-invariant filter with absolutely summable weights can be written as a two-sided moving average:

$$x_t = \sum_{j=-\infty}^{\infty} b_j y_{t-j}. \quad (6)$$

The Power Transfer Function (PTF) $|B(\omega)|^2 = |b(e^{-i\omega})|^2$ measures the squared gain of such a linear filter, i.e. the variance each frequency ω contributes to the filtered series x_t compared to its variance contribution in the original series y_t :

$$SD_x(\omega) = |B(\omega)|^2 SD_y(\omega). \quad (7)$$

To filter business cycles, the PTF should take a value of 1 for the business cycle frequencies, i.e. cycles of 6 to 32 quarters length corresponding to radians of $\omega = 2\pi/6$ to $\omega = 2\pi/32$, and a value of zero for all other frequencies.

Figure 1 shows the PTFs for the original and the modified Hamilton filter. The regression coefficients of the AR(4) process have been computed based on quarterly US log real GDP data from 1947Q1 to 2019Q4. The gray shaded areas indicate cycle lengths of 6 to 32 quarters that are typically associated with business cycle fluctuations. We plot the PTFs in the standard frequency representation as well as in a version that shows the cycle length in quarters on the horizontal axis (panel B) to facilitate the reading of the figure. The standard frequency representation might give the false impression that high frequency fluctuations are particularly important as these take up more than half of the graph (white area to the right of the gray shaded area), while it gets clear in the representation showing the cycle length that these are irregular fluctuations, i.e. noise, with little practical relevance.

It becomes apparent that the original Hamilton filter eliminates business cycles between 6 and 10 quarters almost completely. Hence, short business cycle frequencies are not present in a Hamilton filtered output gap. On the other hand, medium and long term cycles with a duration longer than 10 and up to 32 quarters are substantially amplified compared to shorter business cycles. Further,

the Hamilton filter does not only pass through high frequency noise, but even strongly amplifies many of these high frequency cycles. Panel C shows PTFs of the Hamilton filter for different forecast horizons. Short horizons emphasize short business cycles and mute long business cycles in the output gap, while it is the other way around for long horizons. Hence, taking an average over different horizons leads to a more even coverage of business cycle frequencies as shown by our proposed modified filter in panels A and B. Further, the proposed modified filter avoids the amplification of high frequency noise as best seen in panel A.

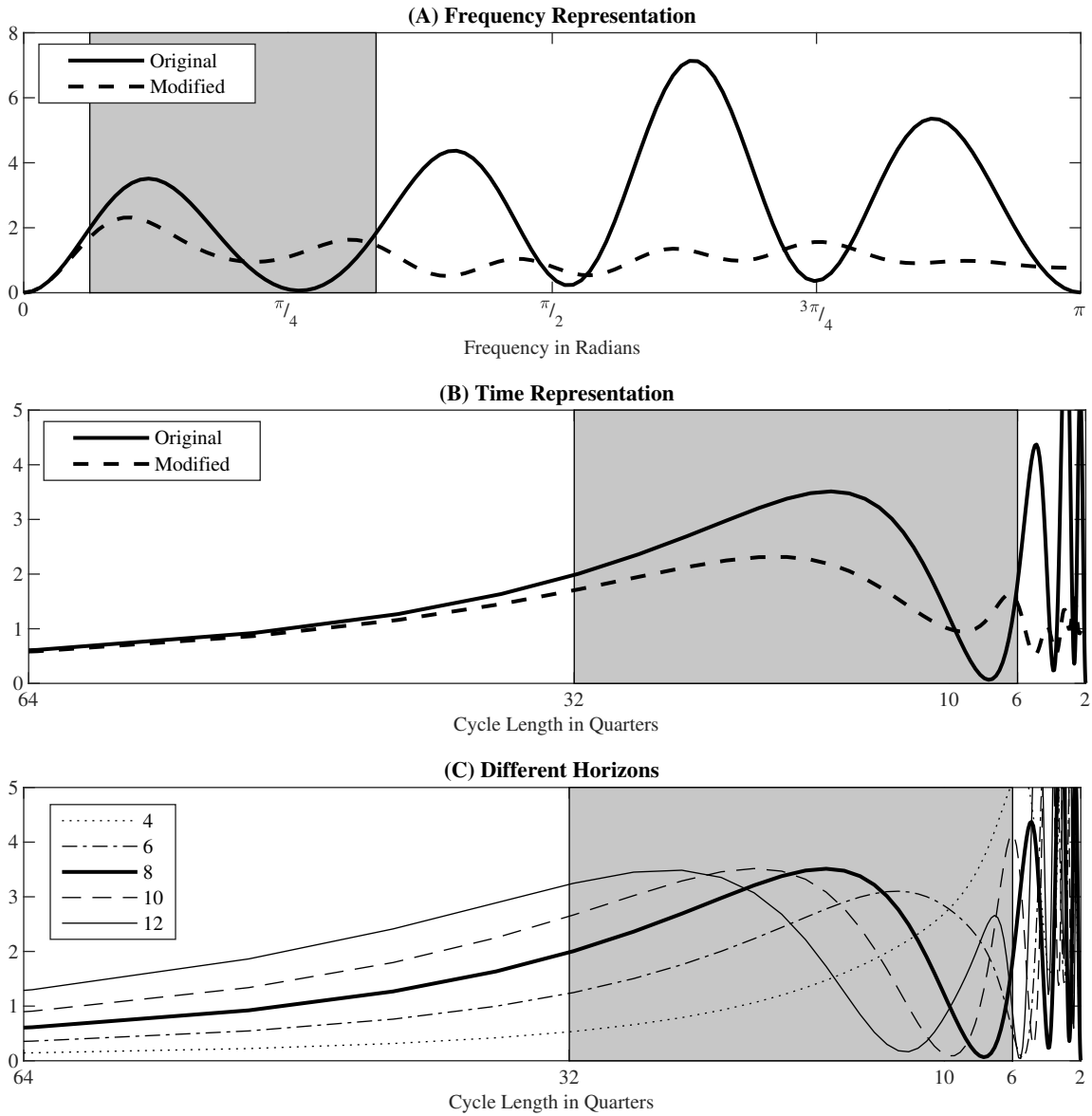


Figure 1: PTFs for the Gap Components in Frequency and Time Representation

Based on a standard definition of business cycles, only frequencies between 6 and 32 quarters

should be present in the output gap. While the modified Hamilton filter is successful in achieving a more even coverage of these frequencies compared to the original one, both versions pass through some longer cycles. In light of recent evidence emphasizing the importance of business cycles longer than 32 quarters (Comin and Gertler, 2006; Beaudry et al., 2020), the decaying PTFs, where longer lasting cycles also pass through to the output gap with increasing suppression, might even be considered as advantageous.

Regarding, the muting of short cycles based on the original Hamilton filter, one might argue based on NBER business cycle dates that this is not of high relevance. There is only one cycle shorter than 10 quarters in the early 1980s in post WWII data. However, cycles extracted by the filtering techniques considered here, do not necessarily consist of an expansion and a recession phase, but could alternatively capture periods of above or below average growth, so that a full NBER cycle could include several cycles of higher frequency. In the online appendix we show cycles ranging from 6 to 10 quarters extracted with the BP filter. Based on this measure, one can observe that the importance of short cycles has decreased over time, but that they are still important during periods of high GDP volatility like, for example, in the early 1980s or around the Global Financial Crisis. Hence, accounting for short cycles via the proposed modification of the Hamilton filter is particularly advantageous for a precise output gap measurement during periods of high volatility.

Figure 2 compares the output gap and trend estimates for the original and the modified Hamilton filter and reveals that the modified Hamilton filter is more successful in filtering out high frequency noise. This can be best seen for the estimated trend growth rate (Panel C). For the original Hamilton filter, the variance of the trend growth rate is similar to that of the growth rate of GDP itself. Further, the period t growth rate of trend GDP and the period $t - 8$ GDP growth rate are very similar (correlation coefficient: 0.94). This is not surprising given that the trend growth rate is computed as $\Delta\hat{y}_t = \hat{\beta}_1\Delta y_{t-8} + \hat{\beta}_2\Delta y_{t-9} + \hat{\beta}_3\Delta y_{t-10} + \hat{\beta}_4\Delta y_{t-11}$, and the fact that the estimate of $\hat{\beta}_1$ is close to 1, while the other AR-coefficients are closer to zero. This implies that movements in GDP 8 quarters ago have a substantial impact on the current trend estimate of the original Hamilton filter. The PTFs of the trend component in Figure 3 illustrate why the trend implied by the modified Hamilton filter is much smoother. While the original Hamilton passes cycles from 6 to 32 quarters as well as amplified high frequency cycles through to the trend, the modified filter is merely characterized by long cycles.

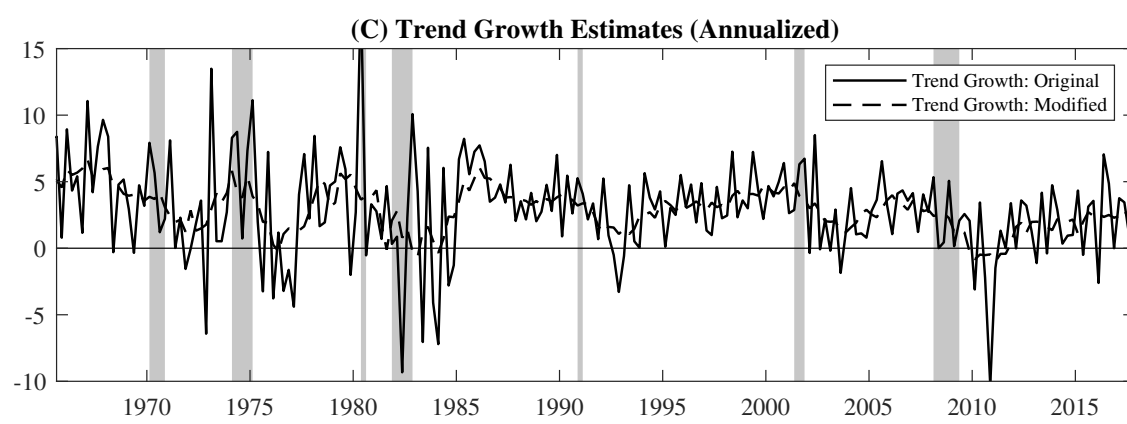
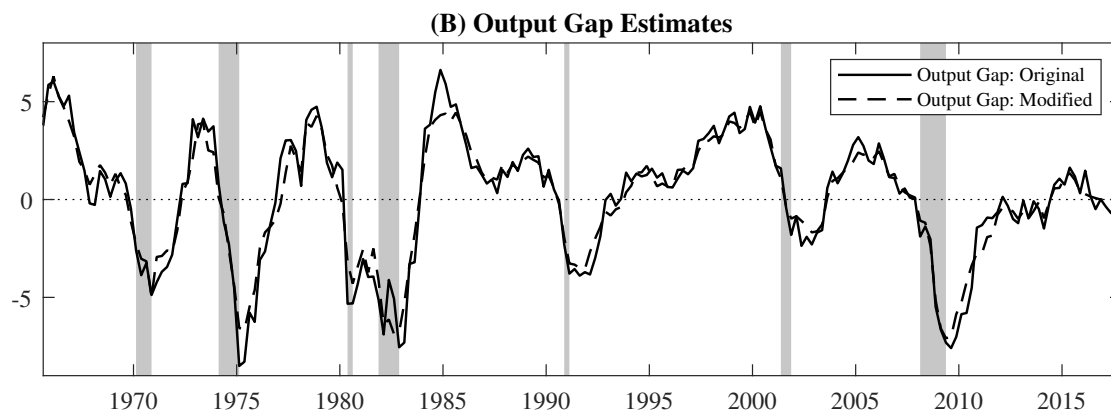
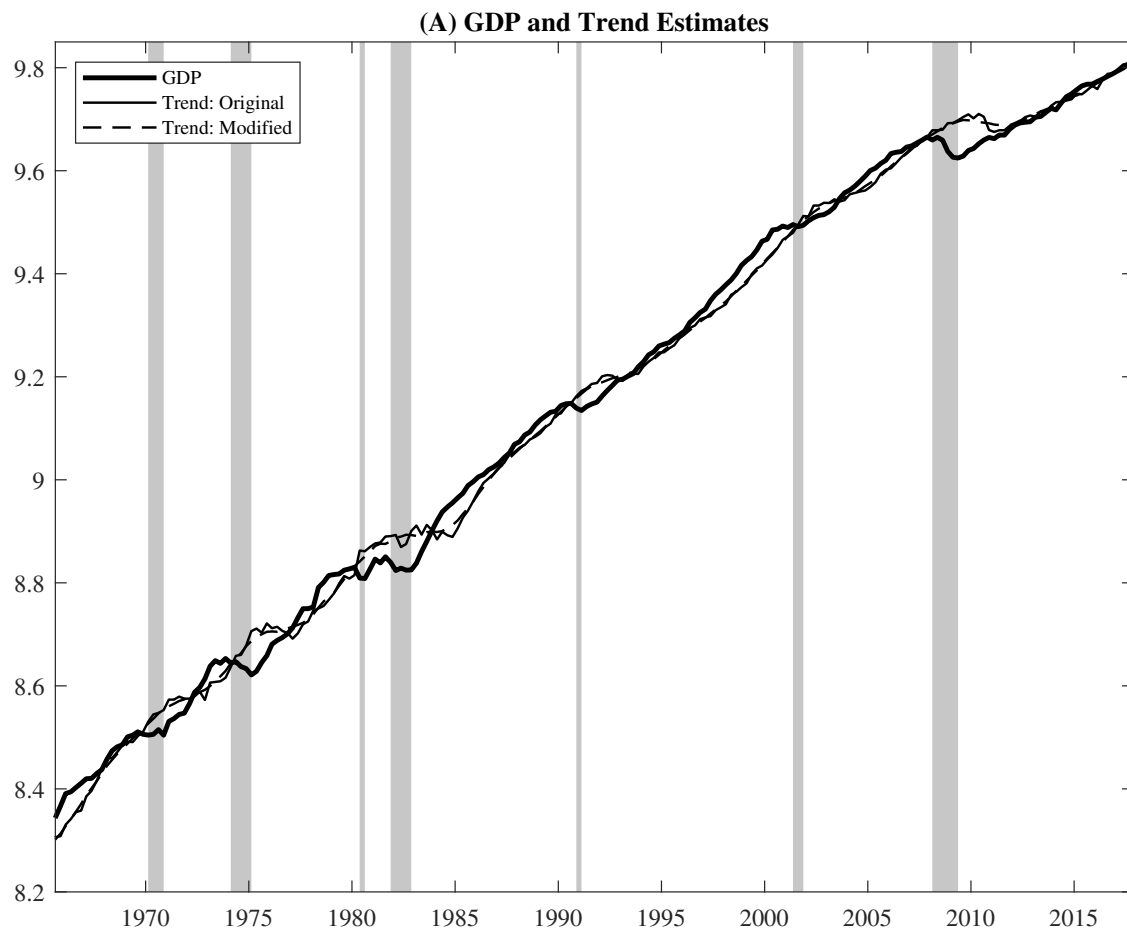


Figure 2: Trend, Output Gap, and Trend Growth: Original and Modified Hamilton Filter

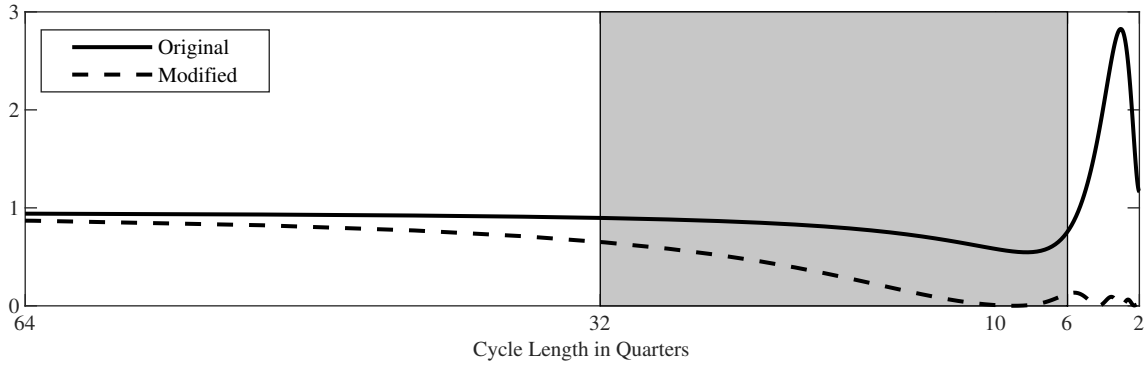


Figure 3: PTFs for the Trend Components

The high frequency fluctuations of the trend of the original Hamilton filter make its economic interpretation difficult as the high frequency movements do not reflect fast changes in the productive capacity in the economy, but are an artifact of the filter being based on a single forecast horizon. By contrast, the extracted trend based on the modified Hamilton filter is very smooth and its growth rate varies slowly over time like those of other often used trend measures based, for example, on the HP or BP filter. Further, the smooth trend of the modified Hamilton filter leads to a smoother output gap compared to the original Hamilton filter.

Such smoother trend and gap estimates may be desirable for three reasons. First, policy makers usually associate potential GDP with the medium- to long-term level of sustainable real output that is driven by low-frequency movements in population growth, labor force participation, and the capital stock (see, e.g., ECB, 2011; Wolters, 2018; Hodrick, 2020). Consequently, potential output growth variability is expected to be rather low, too (ECB, 2000). Second, since both, potential output and the output gap, are used for monetary and fiscal policy making, there is a “pragmatic desire for ‘smooth’ estimates” (St-Amant and van Norden, 1997) to avoid erratic policy changes. For example, potential output estimates are essential in the European Union’s fiscal surveillance framework (Langedijk and Larch, 2011; ECB, 2011). They are deployed to compute the cyclically-adjusted budget balance (CAB) that is used to assess the Member States’ fiscal policy stances. Erratic changes in the CAB would make it difficult to come up with appropriate fiscal policies, in particular given implementation and transmission lags. Third, DSGE model-based estimates of potential GDP are often highly volatile (Vetlov et al., 2011), but recent research shows that it is questionable whether such flexible-price based estimates serve as good benchmarks for desirable characteristics of potential GDP measures. Justiniano and Primiceri (2008) and Coenen et al.

(2009) show that the high volatility can predominantly be attributed to wage and price mark-up shocks. Justiniano and Primiceri (2008) show that the mark-up shocks not only lack structural interpretation (see also Chari et al., 2009), but are also tightly linked to measurement error. Recent firm-level suggests that price mark-ups move rather slowly (Korinek and Ng, 2018; De Loecker and Eeckhout, 2020), so that models with more precisely measured mark-up shocks could yield smoother flex-price potential output estimates than in standard DSGE models.

While the clean economic narrative of Hamilton’s original filter might seem to be obscured by taking an average of 4 to 12 quarter ahead forecast errors rather than focusing on a single forecast horizon, this is not the case. As taking an average of forecast errors makes the interpretation more fuzzy, it has the advantage of substantially improving the treatment of high frequency noise in the obtained output gap and trend measures. The trend estimate implied by the usage of a single forecast horizon depends very much on the starting point, i.e. the trend estimates vary from quarter to quarter due to high frequency noise in GDP at the forecast starting point, whereas the modified filter excludes this noise by combining different forecast starting points, yielding a more precise trend and cycle measurement. In this sense, when applied to business cycle measurement, the economic narrative of the modified filter becomes cleaner as a smoother trend estimate can be more easily interpreted as potential output and the cycle as an output gap.

Overall, the proposed modified filter achieves a very good real-time reliability due to its one-sided nature, but covers business cycle frequencies more evenly than the original approach by Hamilton (2018) and yields a smoother trend estimate. Yet, it is still centered around the 8 quarter horizon proposed by Hamilton and is very simple to compute.

3 Real-time Reliability

We analyze the real-time reliability of different output gaps in the spirit of Orphanides and van Norden (2002) by computing output gaps based on real-time data vintages and comparing them to those based on revised data. We obtain data on quarterly real GDP from the Federal Reserve Bank of Philadelphia’s real-time data set. The first data vintage is from 1965Q4 and the last data vintage is from 2020Q1. Data predominantly starts in 1947Q1 and ends one quarter before the publication date of the data vintage, i.e. in 1965Q3 for the first and in 2019Q4 for the last data vintage. We take logs of real GDP and apply the various filtering techniques to the real GDP

vintages to get real-time output gap vintages.

We use the last data vintage as our measure of final revised data and define the output gap revision as the difference between the final revised and the real-time estimate. 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, we follow Orphanides and van Norden (2002) and discard the last 2 years of observations. Hence, our results for the revised gaps are based on the 2020Q1 vintage with gaps being estimated until 2017Q4. This proceeding is reasonable since most revisions take place in the first 2 years after the initial publication of GDP (Orphanides and van Norden, 2002; Edge and Rudd, 2016). Hence, the sample of output gaps that we study starts in 1965Q3 and ends in 2017Q4.

The Hamilton filter is one-sided by definition. For the revised output gap estimates we estimate the AR-parameters based on the full sample, while for the real-time estimates the AR-parameters are based on the respective real-time subsamples. For the HP and BP filter we use two-sided versions for the revised output gap estimates, while we use one-sided versions for the real-time estimates. Hence, we follow exactly the same approach as in Orphanides and van Norden (2002) that is also used in other papers on this topic (see, e.g., Cayen and van Norden, 2005; Marcellino and Musso, 2011; Edge and Rudd, 2016; Kamber et al., 2018). For the one-sided HP filter, we use the Kalman filter implementation described in, e.g., Stock and Watson (1999b) or Hamilton (2018). While the literature predominantly follows this approach, Wolf et al. (2020) discuss potential adjustments to the one-sided HP filter to better align its cyclical properties to those of the two-sided HP filter. Regarding the BP filter, we use the implementation of Christiano and Fitzgerald (2003), in which the weights are adjusted according to the sample length as discussed in, e.g., van Norden (2002), Christiano and Fitzgerald (2003) or Watson (2007).

To further distinguish between data revisions and revisions due to the filter-induced structure, we compute quasi real-time output gap estimates. That is we estimate one-sided output gap measures as in the real-time estimation, but based on the final revised data. By that we are able to isolate the impact of pure data revisions—defined as the quasi real-time output gap minus the real-time output gap—as the estimates of the real-time and the quasi real-time series cover the exact same time periods.

Figure 4 shows that revisions to output gaps computed with the original and the modified Hamilton filter are small relative to the amplitude of the output gap and are mainly caused by

data revisions as also shown by Jönsson (2019) for the original Hamilton filter. By contrast, final and real-time output gap estimates of the HP and BP filter markedly differ from each other. Revisions for these are of the same sizes as the output gaps themselves confirming the results by Orphanides and van Norden (2002) for an updated sample. The revisions for the HP and BP filter are to a large extent filter-induced revisions, while data revisions only play a minor role.

Table 1 presents summary statistics for the output gap revisions. The upper part of the table shows statistics on total revisions. The Hamilton filtered output gap has the smallest mean error with a value of 0.03 and the BP filter the largest one with a value of 0.29, which is still not very large given that the BP filtered output gap fluctuates between ± 4 . The standard deviation is smallest for the modified Hamilton filter taking a value of 0.84 and almost twice as large for the HP filter with the original Hamilton and the BP filter being in between. However, this value is not very informative without comparing it to the standard deviation of the output gap because the amplitudes of the two versions of the Hamilton filtered output gap are quite a bit larger than the amplitudes of the other two output gaps. Such a comparison is given by the noise-to-signal ratios. The first measure of the noise-to-signal ratio compares the standard deviation of the revision to the standard deviation of the final revised output gap estimate. This value is only 0.29 for the two versions of the Hamilton filtered output gap, while it is much larger for the BP and HP filtered output gaps with values of 0.59 and almost 0.97. In principle, a noise-to-signal ratio based on dividing the RMSE of the output gap revision by the standard deviation of the final revised output gap is more informative because it reflects biases in the real-time output gaps. However, the noise-to-signal ratios based on both measures are almost the same as the biases of all four real-time output gap estimates are small. Further, we report the differences between the noise-to-signal ratios relative to the results obtained for the modified Hamilton filter in columns 6 and 8. To analyze whether they are statistically significant, we follow Edge and Rudd (2016) and compute empirical distributions of the noise-to-signal ratios based on a naive block bootstrapping procedure with replacement. The HP and BP filter based noise-to-signal ratios are significantly larger than the one of the modified Hamilton filter which is almost identical to the one of the original Hamilton filter. Finally, we also report the fraction of observations in which the final revised and the real-time output gap estimates have opposite signs. This is the case for only 5% and 6% of observations for the modified and the original Hamilton filter, respectively, but for 20% and 31% of observations for the BP and HP filtered output gaps, respectively.

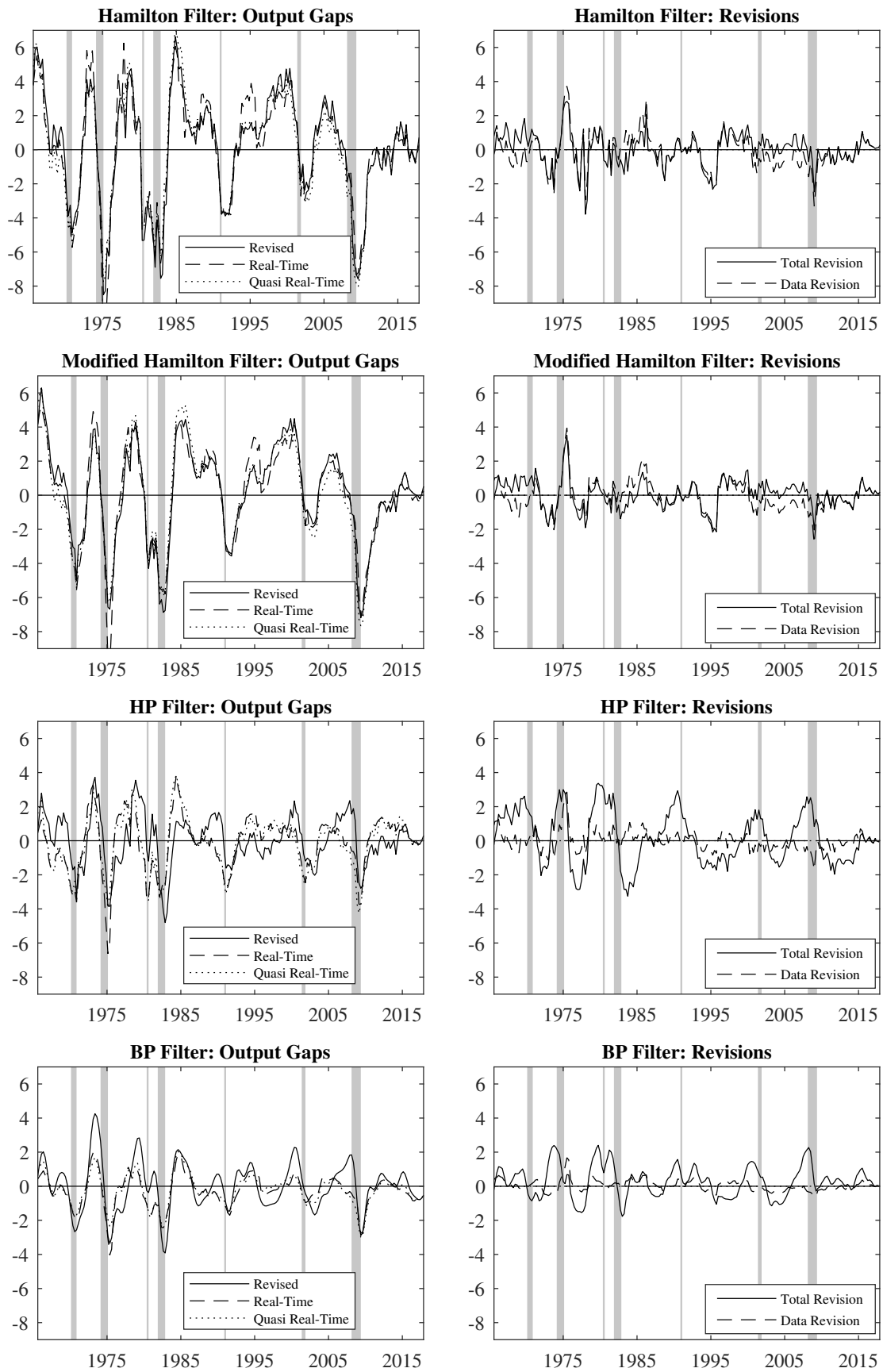


Figure 4: Output Gaps and Revisions

Table 1: Output Gap Revision Statistics

	Mean	SD	RMSE	Noise-Signal Ratios				Opposite Sign
				SD	<i>Diff</i>	RMSE	<i>Diff</i>	
Total Revisions								
Hamilton	0.03	0.98	0.97	0.29	0.00	0.29	0.00	0.06
Modified	0.05	0.84	0.84	0.29	-	0.29	-	0.05
HP	0.17	1.55	1.56	0.97	0.68***	0.97	0.68***	0.31
BP	0.29	0.91	0.96	0.59	0.30***	0.62	0.33***	0.20
Data Revisions								
Hamilton	-0.08	1.04	1.05	0.31	0.00	0.31	0.00	0.06
Modified	-0.07	0.90	0.90	0.31	-	0.31	-	0.07
HP	0.01	0.58	0.58	0.36	0.05	0.36	0.05	0.04
BP	0.05	0.33	0.33	0.21	-0.10**	0.21	-0.10**	0.08

Notes: *, **, and *** denote significance on the 10, 5, and 1% significance level based on a naive block bootstrapping procedure with 5000 replications and a block size of 4. *Diff* refers to differences in the noise-to-signal ratios in the previous column relative to the one of the modified Hamilton filter.

The lower part of the table shows statistics on the share of the revisions that is caused by data revisions. This share is computed by subtracting the real-time output gap estimates from the quasi real-time output gap estimates. For the two versions of the Hamilton filter, the standard deviations and the noise-to-signal ratios based on data revisions and on total revisions are very similar. This reflects that almost all revisions of these output gap measures are due to data revisions. For the HP and BP filter, the standard deviations, RMSEs, and the two noise-to-signal ratio measures are much smaller than those for total revisions, reflecting that data revisions are relatively unimportant. The differences in the noise-to-signal ratios associated with the data revisions are insignificant or even in favor of the BP based measure. Thus, the significantly higher presence of noise in the total revision series of the HP and BP filter compared to the modified Hamilton filter in the upper part of the table is merely due to the different filtering procedures rather than due to data revisions. Further, there are relatively few sign switches between the real-time and the quasi real-time output gap estimates for all four methods.

Overall, we find that the output gaps based on the modified and the original Hamilton filter are

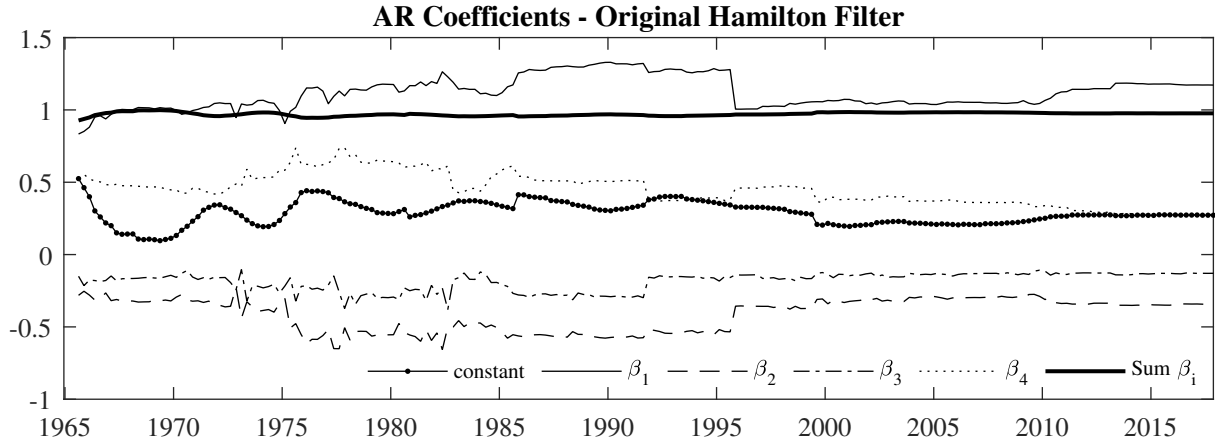


Figure 5: Recursive Estimates of the AR(4) Parameters of the Original Hamilton Filter

by far the most reliable real-time output gap measures. It is also worth noting that the BP filter performs better than the popular HP filter that shows the least reliable performance among the four output gap measures. A likely cause for the large difference between the BP and HP filter is that the Christiano-Fitzgerald (CF) version of the BP filter adjusts weights based on the number of available observations, while those of the HP filter are independent of the sample length. As noted in van Norden (2002), the optimal end-of-sample one-sided BP filter provides lower bounds on measurement error of current trends and cycles estimated with univariate filters. This is in line with our findings that the CF version of the BP filter outperforms the HP filter in terms of being less prone to data revisions.

The reason for the high reliability of the two Hamilton filtered output gaps with respect to revisions is that they rely on one-sided filters. Revisions can occur only for two reasons: Data revisions and changes in the estimated parameters of the AR(4) processes. As discussed above, data revisions only lead to small revisions of the output gap estimates. Regarding the estimated AR(4) parameters, Figure 5 shows how they change over time via the recursive estimation procedure and how they converge to the full sample estimates for the original Hamilton filter. While there are some changes, overall the parameters are rather stable. When looking at the sum of the AR-coefficients, there are almost no changes at all. GDP follows a unit-root or near unit-root process for all subsamples. Regarding the estimated constant, there are some changes until the subsample ending in 1985. Afterwards, the estimates decrease somewhat reflecting the growth slowdown in the 1980s.

For the remainder of the analysis, we solely use the one-sided real-time output gap estimates

based on the different filtering techniques as they reflect the available and relevant information set. Revised and quasi real-time gap measures were only estimated for the purpose of assessing the real-time reliability.

4 Economic Meaningfulness of Output Gap Estimates

While there is evidence for a superior real-time performance of the two versions of the Hamilton filtered output gap compared to the HP and BP filtered ones, it is yet unclear whether the Hamilton filtered gaps are economically meaningful. Stationarity is a first basic requirement. Output gap estimates based on the Hamilton filter are stationary as long as the original time series is integrated of order 4 or less (Hamilton, 2018). This is clearly the case for real GDP. Further, at least the modified Hamilton filter yields a smooth trend estimate and an output gap that covers the relevant business cycle frequencies relatively evenly.

A second requirement for a meaningful output gap is that the method used is able to successfully disentangle trend and cycle. Hamilton's method is able to do so to the extent that no large trend changes occur between periods $t - 8$ and t . Hamilton (2018) justifies the exclusion of trend changes over a 2 year horizon arguing that the primary reason for forecast errors at this horizon are transitory factors such as whether a recession occurs and the timing of recoveries. Other output gap estimation methods are based on a similar a priori belief that the cyclical is much larger than the volatility in the second difference of the trend component. For example, the standard smoothing parameter of 1600 of the HP filter is based on the assumption that the variance in the cyclical component's innovations is $\sigma_c^2 = 5$, while the one in the second difference of the trend component is $\sigma_v^2 = 0.125$ (see, e.g., Canova, 2011, pp. 83-84; Hamilton, 2018). A statistical formalization of the choice of λ leads to a value of 0.245 for US quarterly GDP (Hamilton, 2018). Similarly, when using the Beveridge-Nelson (BN) decomposition, only a version with a strong prior belief that the volatility of the cyclical component is much higher than the one of the trend component leads to sensible output gap estimates, while changes in the trend component dominate without imposing such a prior belief (Kamber et al., 2018). Hence, the a priori choice that over a 2 year horizon trend changes are much less important than cyclical fluctuations is in line with other output gap estimation methods and the modified version of the Hamilton filter accounts for trend changes to a certain extent already after 1 rather than 2 years.

After all, the appropriateness of a filtering technique depends on the researcher’s objective. For output gap estimations, the objective is often to match important historical business cycle episodes. Computing correlations of the quarter-on-quarter change of the different real-time output gap measures with a dummy variable that takes the value of 1 during NBER defined expansions and 0 during NBER defined recessions, yields a correlation of 0.45 for the original Hamilton, 0.55 for the modified Hamilton, 0.41 for the HP and 0.47 for the BP filtered output gaps. Overall, the analysis yields first indications that the modified Hamilton filtered output gap is economically meaningful. More systematic evaluations are provided below.

4.1 Correlation with Output Gaps from Policy Institutions

First, we compare real-time output gap estimates to revised US output gap estimates of the Federal Reserve (Fed), the Congressional Budget Office (CBO), the IMF, and the OECD. Revised output gap estimates of policy institutions should be useful benchmarks for three reasons. First, they entail economic considerations regarding past courses of the US business cycle that include a substantial amount of economic expertise rather than being based only on statistical models. Second, the assessments of policy institutions should capture output gap dynamics that are deemed important from a practitioner’s perspective. Finally, recent papers show that output gaps from policy institutions have been more reliable than those based on statistical methods over the last 20 years (see Edge and Rudd, 2016, for the Fed’s output gap and Champagne et al., 2018, for the Bank of Canada’s output gap).

For all four policy institutions, models or statistical approaches build the foundation for the potential output estimates (see, e.g., Coibion et al., 2018, for a detailed overview on how the institutions estimate potential output). Yet, all institutions combine these with a large amount of judgment. The Fed’s estimates rely on a judgmental pooling of results from different statistical and structural methods and models (Mishkin, 2007; Edge and Rudd, 2016). The CBO focuses on a sectoral production function approach where “a substantial degree of judgment” is applied, for example, to the projections of potential TFP, potential output of the household sector, or federal employment (Shackleton, 2018). The OECD also uses a production function approach, but assumptions regarding the future NAIRU, working age population, rates of participation, and productivity or capital and wage shares are implemented on a judgmental basis (Beffy et al., 2006). Lastly, the IMF’s production function framework (De Masi, 1997) is also augmented. Here

the judgment of desk economists and mission chiefs plays a key role in evaluating a country’s potential output (De Resende, 2014; Rosnick, 2016).

There are some limitations regarding available samples and frequencies of output gap estimates from policy institutions. While for the Fed and the CBO quarterly data is available, samples for the IMF and the OECD rely on annual data. Further, the Fed data ends in 2013Q4 as it is based on the Greenbook which is made available to the public with a lag of about 5 years. Data for the Fed and the CBO is available from the start of our real-time output gap sample in 1965Q3, while data for the IMF (OECD) starts in 1980 (1985).

Table 2 shows correlations for the quarterly output gap series covering the period 1965Q3 until 2013Q4 on the left and for the annual series for data from 1985 to 2013 on the right. The results show that the two Hamilton filtered real-time output gaps are highly correlated with all revised institutional output gaps with correlation coefficients ranging from 0.57 to 0.83. The differences in the correlation coefficients between the original and the modified Hamilton filter are not statistically significant so that both reflect ex post expert evaluations similarly well. The HP and BP filtered real-time output gap correlations with the policy institutions’ gaps are significantly lower ranging from 0.11 to 0.55, with those of the HP filter being particularly low.

Due to the similarity of Hamilton-type real-time and full sample output gap estimates, we unsurprisingly find a similar correlation coefficient between the original Hamilton filtered output gap and the CBO gap in our real-time setting as Hodrick (2020) does in his full-sample setting. Adopting the latter, we also find a similarly high correlation coefficient between the two-sided HP filtered output gap and the one of the CBO as reported in Hodrick (2020). However, this perfectly showcases the unreliability issues associated with the HP filter. Our results indicate that these correlations deteriorate as soon as one refrains from using full sample information and applies the one-sided HP filter in a real-time setting.

Interestingly, for all four statistical output gap measures the correlations are higher with the output gaps of the Fed and the CBO—the two US institutions—than with the IMF and OECD—the two international institutions. Overall, these results imply that the real-time Hamilton filtered output gaps are able to reflect the ex post expert evaluation of the US business cycle to a considerable degree in real time, while the BP and in particular the HP filter do this to a much smaller extent.

Table 2: Correlations between Statistical Real-Time and Expert Ex-Post Output Gap Estimates

	Quarterly Data (65Q3-13Q4)				Annual Data (1985-2013)							
	Fed	<i>Diff</i>	CBO	<i>Diff</i>	Fed	<i>Diff</i>	CBO	<i>Diff</i>	IMF	<i>Diff</i>	OECD	<i>Diff</i>
Hamilton	0.57	-0.05	0.66	-0.05	0.78	-0.05	0.78	0.04	0.64	-0.04	0.68	-0.05
Modified	0.62	—	0.71	—	0.83	—	0.82	—	0.68	—	0.73	—
HP	0.20	-0.42***	0.33	-0.38***	0.31	-0.52***	0.36	-0.46***	0.28	-0.40**	0.11	-0.62***
BP	0.46	-0.16**	0.55	-0.16***	0.50	-0.33**	0.51	-0.31**	0.46	-0.22	0.36	-0.37**

Notes: *Diff* reports the respective difference in the correlation coefficients with respect to the modified Hamilton filtered statistical real-time gap. **, and *** denote significant differences on the 5, and 1% significance level based on Fisher’s z transformation (Fisher, 1921).

4.2 Forecasting Performance: Output Growth

The previous analysis shows that the assessment of the meaningfulness of output gap estimates is to some extent subjective depending on the researcher’s objective regarding what the output gap should measure. However, with the assessment of the predictive content of output gap estimates also a more objective criterion is available.

Nelson (2008) proposed the evaluation of competing output gap measures via their output growth forecasting performance. If an output gap was negative, one would expect above average output growth rates in the future so that output reverts back to trend. Conversely, if the output gap was positive, output growth should be below average some time in the future. We use a standard forecast equation in which output growth h periods ahead is predicted using the real-time output gap vintage:

$$y_{t+h} - y_t = \alpha + \beta \hat{c}_t + \varepsilon_{t+h|t}, \quad (8)$$

where y denotes log real GDP, \hat{c} the estimated real-time output gap vintage, and $\varepsilon_{t+h|t}$ the forecast error. The equation is estimated with OLS. The initial sample runs from 1965Q3 to 1975Q2, the sample is recursively expanded quarter-by-quarter and forecasts are computed for horizons 1 to 12. Based on the intuition developed above, we expect $\beta < 0$ at some horizon, essentially indicating the ability of the output gap to predict trend-reverting tendencies of the output growth series.

Table 3 reports RMSEs for the original Hamilton and the modified Hamilton filtered output gaps relative to the HP and BP filtered ones as well as to one another. In the last column we

Table 3: Output Growth Forecast Evaluation Based on Statistical Real-Time Output Gaps

Horizon	Relative RMSE					
	Ham/HP	Mod/HP	Ham/BP	Mod/BP	Mod/Ham	BP/HP
1	0.91***	0.91***	0.98*	0.97*	0.99	0.94***
2	0.85***	0.83***	0.94**	0.93**	0.99	0.90***
3	0.79***	0.78***	0.91***	0.89***	0.98*	0.87***
4	0.76***	0.75***	0.89***	0.87***	0.98*	0.86***
5	0.75***	0.74***	0.88***	0.87***	0.98*	0.85***
6	0.76***	0.74***	0.89***	0.87***	0.98*	0.85***
7	0.77***	0.76***	0.90**	0.88**	0.98**	0.86***
8	0.79***	0.77***	0.90**	0.89**	0.98**	0.87***
9	0.80***	0.79***	0.90**	0.89**	0.98**	0.88***
10	0.82***	0.80***	0.91**	0.89**	0.98**	0.90***
11	0.84***	0.82***	0.91**	0.89**	0.98**	0.92**
12	0.85***	0.84***	0.91**	0.90**	0.98**	0.93**

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

additionally report the RMSEs for the BP relative to the HP filtered one. The relative RMSEs reveal that output growth forecasts based on the two Hamilton-type filtered real-time output gap measures are significantly more accurate compared to using an HP or BP filtered gap measure. These accuracy gains amount up to 26% (13%) for the modified Hamilton filter relative to the HP (BP) filtered gap. We also find slightly more accurate forecasts based on the modified compared to the original Hamilton filter. However, compared to the gains obtained with respect to the HP or BP filter, they remain extremely small.

Except for the first horizon of the Hamilton filtered gaps, where the slope coefficient is zero, all slope coefficients have negative signs. This indicates that all four output gap measures predict trend-reverting output growth rates. They are significantly different from zero from horizon 4 onward for the two Hamilton and from horizon 1 onward for the HP and BP filtered output gaps. When we extend equation (8) 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), we find very similar results.

Overall, the results from this exercise show that output gaps estimated with the original or

the modified Hamilton filter increase output growth forecasting accuracy relative to using output gaps estimated with the HP or BP filter. This indicates that they might be comparatively more informative about the stance of the business cycle.

4.3 Forecasting Performance: Inflation

Since theory predicts that output gap estimates should be useful predictors for forecasting inflation, we also consider a Phillips curve type forecasting model to evaluate the competing output gap measures. We follow Stock and Watson (1999b), Clark and McCracken (2006), Stock and Watson (2008), and Kamber et al. (2018) in specifying an autoregressive distributed lag (ADL) Phillips Curve forecasting equation:

$$\pi_{t+h} - \pi_t = \alpha + \sum_{i=0}^p \beta_i \Delta \pi_{t-i} + \sum_{i=0}^q \gamma_i \hat{c}_{t-i} + \varepsilon_{t+h|t}, \quad (9)$$

where π_t denotes US PCE inflation, \hat{c}_t the recursively estimated real-time output gap vintage, and $\varepsilon_{t+h|t}$ the forecast error. While our baseline specification is based on changes in inflation, i.e. imposes a unit root in inflation, forecasting results are very similar when we estimate models in terms of inflation levels or specifications that include a relative import price inflation term. We use final revised data for inflation. The lag lengths p for inflation and q for the output gap are determined based on the entire sample using the SIC. We consider $p \in [0, 12]$ and $q \in [0, 12]$. In the recursive forecast evaluation, we assume that the optimal lag order is known a priori. As above, the initial sample runs from 1965Q3 to 1975Q2, the sample is recursively expanded quarter-by-quarter, and forecasts are computed for horizons 1 to 12. For comparison, we also compute results for a model that omits the output gap, but is otherwise identical to equation (9).

Table 4 shows root mean squared inflation forecast errors based on the two Hamilton filtered output gaps relative to those based on the HP, the BP filter, one another, and those based on the specification without output gap. The differences between the forecasting models are marginal and mostly insignificant. Further, models that condition on an output gap measure do not significantly improve upon a univariate inflation forecast. While these results unfortunately do not help in evaluating competing output gap measures, they are fully in line with the literature. Among others, Stock and Watson (2007, 2008), Faust and Wright (2013), Edge and Rudd (2016), and Kamber et al. (2018) find that it is generally difficult to beat univariate inflation forecast models through conditioning on output gaps. We find very similar results when conditioning on final

Table 4: Inflation Forecast Evaluation Based on Statistical Real-Time Output Gaps

Horizon	Relative RMSE						
	Ham/HP	Mod/HP	Ham/BP	Mod/BP	Mod/Ham	Ham/No Gap	Mod/No Gap
1	0.92***	0.99***	0.87***	0.93*	1.07***	1.02*	1.10***
2	1.02*	1.01*	0.88***	0.87***	0.98*	0.91**	0.89**
3	0.98	0.98*	0.98	0.98	1.00	1.01	1.01
4	1.02	1.02	1.02	1.02	1.00	1.00	1.00
5	1.03***	1.02**	1.03	1.01	0.98	1.01	1.00
6	1.01	1.01	1.04	1.04	1.00	0.96	0.97
7	1.01	1.01	1.04**	1.04**	1.00	0.96	0.97
8	1.02	1.01	1.05**	1.05**	1.00	0.96	0.96
9	1.01	1.01	1.03	1.03	1.00	0.96	0.96
10	0.99*	1.01	1.04**	1.06***	1.02*	0.95	0.97
11	1.00**	1.01	1.05**	1.06***	1.02**	0.96	0.98
12	1.00	1.00	1.01	1.01	1.00	0.96	0.96

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

revised output gaps. Hence, these results are due to the general decline in the forecastability of inflation in recent decades (Stock and Watson, 2007, 2008) rather than to real-time output gap measurement problems.

5 Alternative Specifications and Robustness

Alternative Forecast Horizon Ranges The modified Hamilton filter is based on a simple mean of forecast errors of horizons from 4 to 12 quarters ahead. This specification is close to Hamilton’s original proposal because it is centered around the 8 quarter horizon proposed by Hamilton (2018). To test the sensitivity of our results, we also compute results for the different output gap evaluations for other forecast horizon ranges. Specifically, we consider tighter (6 – 10 quarters) and wider (2 – 14) bands of forecast horizons. While these two specifications are still centered around the 8 quarter horizon, we also compute results for tight and wide bands centered around $h = 6$ (4 – 8; 2 – 10) and $h = 10$ (8 – 12; 6 – 14).

We find that real-time output gaps based on all these different specifications have a much

higher correlation with NBER dated recessions and revised output gaps from policy institutions than real-time output gaps based on the HP or the BP filter. Among the different forecast horizon specifications, it turns out that including longer forecast horizons compared to the baseline—either via increasing the range of considered forecast horizons or by centering around a horizon of 10 instead of 8 quarters ahead—increases correlations, while tighter bands or centering around a 6 quarter ahead forecast horizons decreases correlations. However, when looking at the results of the spectral density analysis, it becomes clear that including longer forecast horizons leads to a less even coverage of typical business cycle frequencies. Short cycles are muted and cycles that are longer than 8 years, which is typically considered as the upper bound for business cycles, are amplified. Hence, if one thinks that policy institutions have the best assessment of which dynamics should be taken into account, then including higher horizons might be preferable. This is in line with the observation of Beaudry et al. (2020) that the classical range of business cycles up to a length of 8 years might not be sufficient and that longer cycles should be included in the definition of business cycles. If one wants to stick to a standard definition of business cycle frequencies instead, then our baseline specification is preferable. With respect to the output growth forecasting exercise, differences across specifications are very small. Finally, for inflation forecasts, the 4-8 band, centered around a forecast horizon of 6 quarters, yields small accuracy gains compared to the baseline range, though differences are only significant for 1 quarter ahead inflation forecasts. Based on these exercises, a trade-off gets apparent: Including longer horizons helps the real-time output gap measure to match NBER recessions and the revised expert output gaps better, while including shorter horizons increases inflation forecasting accuracy. Overall, these results show that a range centered around a forecast horizon of 8 quarters seems to be a good choice when considering the results across all the different evaluation exercises.

Optimized Weights of the Modified Hamilton Filter The modified Hamilton filter is based on an unweighted average of forecast errors of different horizons. While this makes the computation particularly simple and transparent, we also analyze the extent to which the output gap measurement can be improved by choosing an optimized weighting scheme instead. First, we choose weights in order to achieve a PTF that is as close as possible to the ideal PTF that takes a value of 1 for cycle lengths from 6 to 32 quarters and a value of zero otherwise. Second, we choose weights in order to get as close as possible to a PTF that takes the average value of the original Hamilton

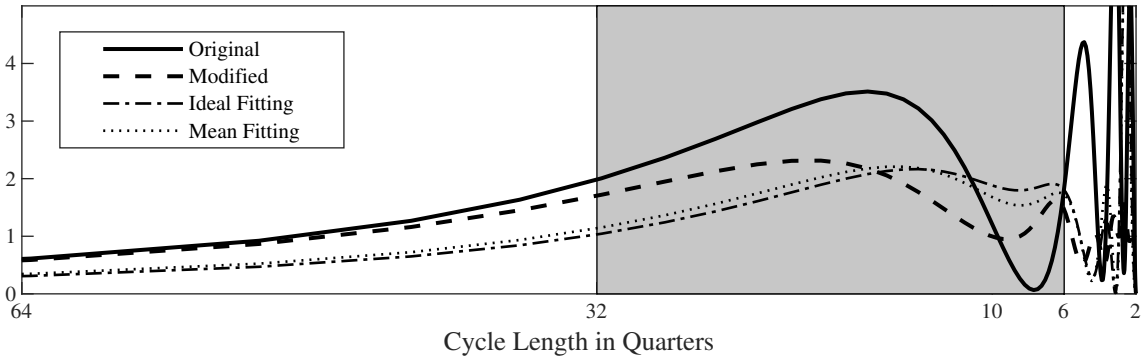


Figure 6: Original, Modified and Fitted Power Transfer Functions

filter’s PTF over the cycles from 6 to 32 quarters length and zero otherwise. This average is close to 1.7, i.e. almost double the size of an ideal PTF reflecting the higher amplitude of output gaps based on Hamilton-type filters. For both exercises we use the full sample for the optimization of weights, consider forecasting horizons between 4 and 12 quarters, and a quadratic loss function. We restrict the weights across the 9 considered horizons to be positive, to sum to 1, and cut off the optimization after a maximum cycle length of 128 quarters. If a shorter (64 quarters) or longer (256 quarters) cutoff point is chosen, we obtain the same results up to rounding precision.

In both optimized versions, only 3 of the 9 horizons have weights that are larger than zero. These are horizons 4 (weight: 0.39), 6 (0.43), and 9 (0.18) for fitting the ideal PTF and 4 (0.33), 6 (0.42), and 9 (0.25) for fitting the average PTF of the business cycle frequencies based on the original Hamilton filter. Hence, the results indicate that it is sensible to include shorter as well as longer horizons. Above we found that including higher horizons increases correlations with revised expert output gaps. The higher weights on short compared to long horizons in the optimization exercise confirm, however, that this comes at the cost of deviating from standard definitions of business cycle frequencies by giving too much weight to medium- to long-run cycles.

Figure 6 shows the PTFs for Hamilton’s original filter, the modified one, and the filters with optimized weights. It gets apparent that the cyclical properties of our simple rule-of-thumb modification are most similar to the one where we optimize weights to fit the average PTF of the original Hamilton filter for cycles between 6 and 32 quarters.

In all evaluation exercises, the results are very similar for the versions with optimized weights (for simplicity determined based on the full sample) and the one based on a simple average of forecast errors of different horizons. The correlations of our baseline modified Hamilton filter with

the different revised expert gaps are higher than those of the versions with optimized weights, though the differences are significant only in some cases. Regarding output growth forecasting accuracy, differences between the original, the baseline modified, and optimized modified Hamilton filters are marginal and insignificant. Lastly, differences in inflation forecasts remain small and inconclusive, even if optimized weights are considered.

Forecast-Augmented Output Gap Estimates One possibility to mitigate the end-of-sample distortions of two-sided filters is to use forecast-augmented data series (see, e.g., Kaiser and Maravall, 1999; Mise et al., 2005; Garratt et al., 2008; Kaiser and Maravall, 2012). To evaluate the implications towards our results, we augment each log real GDP vintage with 4 quarters of forecasts based on a univariate AR(4) process as in Stock and Watson (2007, 2008). The forecast-augmentation has only minor implications for the Hamilton-type filters as the forecasts have only a small effect on the estimated AR-parameters. Apart from that, they do not affect the output gap estimates due to the one-sided filtering approach. Regarding the HP filter, we find in line with Garratt et al. (2008) that the forecast augmentation reduces uncertainty around HP filtered end-point estimates distinctively. The noise-to-signal ratio decreases by 27%. Notably, for the BP filter we do not find such a large effect as the results based on the forecast augmented output gap are very similar to the baseline results. Christiano and Fitzgerald (2003) show that the BP filter’s end-of-sample bias vanishes only slowly as more data becomes available. Weights of the CF version of the BP filter are based on the available sample, with end-of-sample weights being chosen optimally, so that appending forecasts to the sample hardly reduces measurement error. Weights of the HP filter are chosen independent of the available sample, so that end-of-sample measurement error is higher. The HP filter’s end-of-sample bias declines considerably within the first year. Hence when forecast-augmented series are used, the HP filter results align with those obtained from using the BP filter. Yet, there is still significantly more noise present in the real-time HP and BP filter based output gaps compared to the Hamilton-type ones. The expert output gap based evaluation shows a significantly better performance of the Hamilton filtered real-time gaps compared to the HP and BP filtered ones. In the output growth forecast exercise, the two Hamilton filtered real-time output gaps still show distinctively greater forecast accuracy compared to the HP filtered gap, though forecasts based on the HP filter have improved via the forecast augmentation and are of similar accuracy as those of the BP filter. The results of the inflation forecast exercise

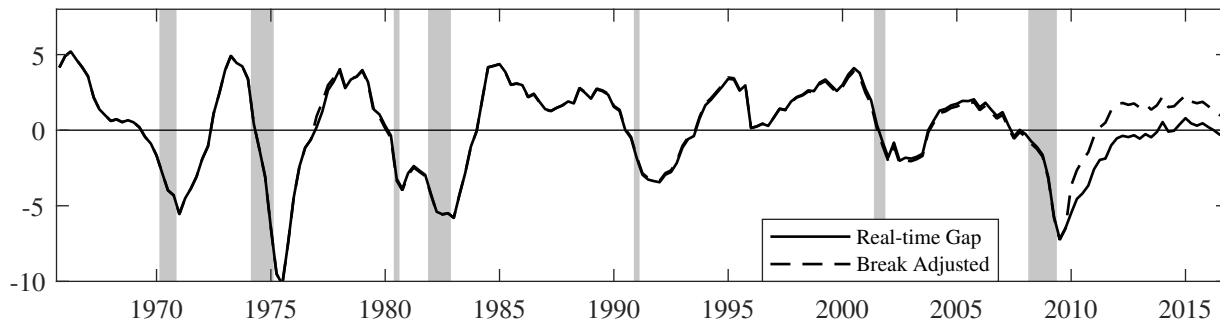


Figure 7: Real-time and Break Adjusted Real-time Modified Hamilton Filtered Output Gap

are as inconclusive as before.

Breaks in Trend Growth Both, Hamilton’s original as well as our proposed modified procedure, rely on constant coefficients. To study whether accounting for structural breaks can improve the real-time output gap measures, we follow the literature and account for two possible structural changes. While Orphanides and van Norden (2002) and Perron and Wada (2009, 2016) provide evidence for a structural change in 1973Q1 associated with the productivity slowdown during the 1970s, Kamber et al. (2018) find evidence for a break in the long-run growth rate in 2006Q1. Since structural changes are especially hard to detect at the end of the sample, we follow Orphanides and van Norden (2002) and assume that the respective potential change is reflected in vintages starting 4 years after occurrence.

We then re-estimate the Hamilton-filtered output gaps for each data vintage accounting for the detected structural breaks in the constant. Figure 7 shows the unadjusted and adjusted real-time output gap estimates based on the modified Hamilton filter. Overall, the break-point adjusted output gap is higher than the baseline output gap in the aftermath of the Great Recession. Apart from that, the dynamics are the same. The reason is that the magnitude of the estimated breaks—and in particular the one in the 1970s—is relatively small compared to the unconditional variance of GDP growth, so that the estimated autoregressive coefficients are little impacted by allowing for breaks. Results are very similar for the original Hamilton filter. Therefore, the original and modified Hamilton filtered output gaps are rather robust to accounting for structural change in trend growth.

Comparison to the Beveridge-Nelson decomposition In addition to the literature on the real-time reliability of output gaps, the paper by Kamber et al. (2018) is closely related to our

work. They use a modified BN decomposition to estimate output gaps. The method is related to the Hamilton filter as it also uses an autoregression to estimate the trend. The approach of Kamber et al. (2018) is on the one hand less ad hoc because they estimate the trend based on long-horizon conditional expectations rather than imposing a fixed horizon of 8 quarters. Further, the long-horizon conditional expectations are measured using data until period t rather than $t - 8$, so that trend changes between $t - 8$ and t are accounted for. By contrast, when using the Hamilton filter one implicitly assumes that there are no relevant trend changes during this period. On the other hand, the computation of the Hamilton filter and also the modified version proposed in this paper are much simpler and more intuitive. The standard BN decomposition yields an output gap that is completely at odds with standard business cycle facts and therefore an algorithm with several steps has to be run to get the modified BN decomposition that makes sure that most GDP dynamics are attributed to the cycle rather than trend changes. For the Hamilton filter the computation of forecast errors based on simple autoregressions is instead sufficient. See Hodrick (2020) for further discussions of similarities and differences between Hamilton (2018) and the BN decomposition (Beveridge and Nelson, 1981).

The BN output gap proposed by Kamber et al. (2018) and Hamilton-type output gaps are highly correlated ex post and in real time (with correlation coefficients of around 0.85). Hence, it is not surprising that the different output gap evaluation exercises yield overall similar results for the real-time BN and Hamilton filtered output gaps. In the expert output gap evaluation exercise, correlations are predominantly higher when the modified Hamilton filtered gap is used. Yet, this is only significant at the 10% level for the quarterly correlations with the CBO gap. In the output growth forecasting exercise, the results are very similar without significant differences over the first four forecasting horizons. From horizon 5 onward, both Hamilton-type gaps provide significantly more accurate forecasts compared to the BN filtered gap. In the inflation forecasting exercise, no differences between the gap measures can be detected as we obtain almost the same results for the Hamilton and BN filtered gaps.

The Cases of Germany and the United Kingdom In order to check whether the results are specific to the US, we compile similar quarterly vintage data sets on real GDP for Germany and the UK and repeat all output gap evaluation exercises. We provide a short description of the results in the following and refer to the online appendix for details. Real-time data vintages

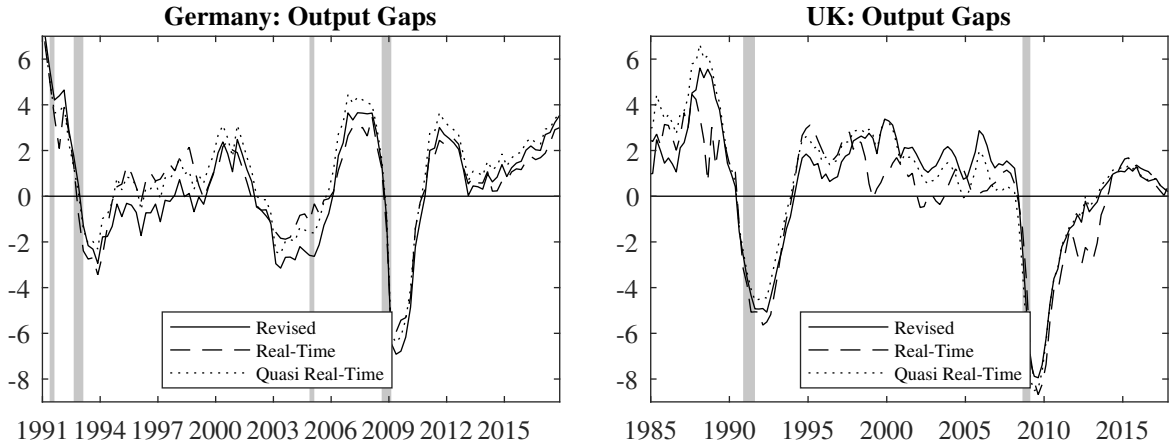


Figure 8: Estimated Output Gaps for Germany and the UK

for Germany are obtained from the Deutsche Bundesbank’s real-time dataset, while they come from the Bank of England for the UK. Vintages for Germany (the UK) are available from 1971Q4 (1976Q1) onward and contain quarterly values starting from 1962Q1 (1955Q1). For Germany, we splice level adjusted data for West Germany before the German reunification with data for whole Germany afterwards. We use UK data from 1985Q1 onward for all evaluation exercises as all ex post output gap estimates from policy institutions are available since then. For Germany we use data from 1991Q1 onward, i.e. post-reunification data. As before, we choose the 2020Q1 vintage as our final series and disregard the last 2 years. Thus our sample ends in 2017Q4. Figure 8 shows the real-time, quasi real-time, and revised output gap estimates as obtained from the modified Hamilton filter. Gray shades show recessions based on the Bry-Boschan business cycle dating algorithm (Bry and Boschan, 1971; Harding and Pagan, 2002).

All three estimated output gaps are very similar. Since the quasi real-time and revised series are almost identical, it gets apparent that the remaining revisions are driven by data revisions. First, we find significantly less noise left in the Hamilton-type real-time gap estimates compared to those obtained from the HP or BP filter. Also, similarly to the US case, we find that there are relatively fewer sign switches among the real-time and the revised measures. Hence, Hamilton-type output gap measures for Germany and the UK are reliable. Second, we find that the modified Hamilton filter covers standard business cycle frequencies more evenly than the original Hamilton filter. Indeed, the results are very similar to those in Figure 1 for the US. Third, we again find that our modification yields much smoother trend growth estimates compared to Hamilton’s originally proposed approach, facilitating the interpretation of the obtained trend as potential output and

Table 5: Correlations between Statistical Real-Time and Expert Ex Post Output Gap Estimates

	Germany (1991-2019)						UK (1985-2019)					
	EC	<i>Diff</i>	IMF	<i>Diff</i>	OECD	<i>Diff</i>	EC	<i>Diff</i>	IMF	<i>Diff</i>	OECD	<i>Diff</i>
Hamilton	0.74	-0.04	0.77	-0.04	0.72	-0.05	0.51	-0.02	0.56	-0.03	0.45	-0.03
Modified	0.79	—	0.81	—	0.76	—	0.53	—	0.59	—	0.48	—
HP	0.45	-0.34**	0.57	-0.24*	0.34	-0.43**	0.12	-0.42*	0.10	-0.49**	0.00	-0.48**
BP	0.79	-0.10	0.72	-0.09	0.63	-0.13	0.25	-0.29	0.22	-0.37*	0.13	-0.35

Notes: *Diff* reports the difference in the correlation coefficients with respect to the modified Hamilton filtered statistical real-time gap. *, **, and *** denote significant differences on the 10, 5, and 1% level based on Fisher’s z transformation (Fisher, 1921).

the cycle as output gap.

To evaluate the economic content of the real-time output gaps for Germany and the UK, we first compute their correlations with revised output gaps of the European Commission (EC), the IMF, and the OECD published as of autumn 2019. Since the expert gaps are only available at the annual frequency, we compute annual averages of the quarterly real-time output gaps. For Germany (UK), all three expert gaps are jointly available from 1991 (1985) onward. Table 5 shows the results.

Similar to the US case, Hamilton based output gaps are highly correlated with the German and UK ex post expert output gaps. Again, differences between the original and the modified Hamilton filter are not statistically significant. While only one correlation (UK, IMF) is significantly smaller when using the BP filtered real-time gap compared to the modified Hamilton filter, all correlations between the HP filtered real-time and the ex post expert output gaps are significantly smaller for both countries.

As in the analysis of revision statistics, we start evaluating forecasts later than in the US case. For both exercises, evaluation starts in 1991Q1 (1985Q1) for Germany (the UK). We again find that using Hamilton-type real-time gaps yields significantly greater output growth forecast accuracy compared to using the HP filter. These accuracy gains range between 16% – 25% for Germany and 21% – 25% for the UK. All gains are significant either at the 1% or 5% level. Compared to the originally proposed filter, our modification shows small advantages for both countries over some horizons. While, the forecasts are statistically significantly more accurate for Germany at shorter

and longer horizons, for the UK this holds for $h = 5$. We also find accuracy gains with respect to the BP filter. They amount up to 22% and are significant at the 1% level over all horizons for Germany. For the UK, they amount up to 14%. Here, gains are statistically significant at the 10% level over most horizons from horizon 4 onward. Similar to the US case, the results in the inflation forecasting exercise are inconclusive in terms of choosing a particular real-time output gap measure.

Overall, conducting the analysis for Germany and the UK underlines the benefits of using Hamilton-type real-time output gap measures compared to alternative simple time series filters as they are both, robust in real-time and economically more meaningful. As for the US, the modified Hamilton filter shows some advantages in the various evaluation exercises compared to the original one on top of the conceptual advantages demonstrated based on the spectral density analysis.

6 Conclusion

We have proposed a modified version of the Hamilton filter for the estimation of reliable and economically meaningful real-time output gaps. It shares the favorable real-time properties of the original Hamilton filter and is similarly easy to compute. However, it has a much better coverage of typical business cycle frequencies and yields a smooth estimated trend, while the original approach does not. The original approach is very useful for detrending, whereas the modified version allows for a meaningful economic interpretation of the cyclical component as an output gap and of the trend as potential GDP. This is particularly important as existing papers applying the Hamilton filter use it not merely for detrending, but attach an economic interpretation to the filtered time series. Compared to other simple statistical trend-cycle decomposition techniques, such as the HP or the BP filter, the real-time output gaps based on the modified and the original Hamilton filter show a much higher correlation with ex post assessments of output gaps from important policy institutions. Hence, the methods yield real-time output gaps that capture business cycle dynamics that are deemed important from a practitioner's perspective. They also yield greater output growth forecasting accuracy compared to forecasts that are based on output gaps computed with other simple statistical trend-cycle decomposition techniques. Hence, our results suggest that the modified and the original Hamilton filter should be preferred over the HP or BP filter as simple decomposition techniques in context of output gap estimation. Our results do not only hold for

US real-time output gap estimation but to a large extent also for estimations based on data for the UK and Germany.

References

- Ahn, H. J. and J. D. Hamilton (2020). Heterogeneity and unemployment dynamics. *Journal of Business and Economic Statistics*. forthcoming.
- Atkeson, A. and L. E. Ohanian (2001). Are Phillips curves useful for forecasting inflation? *Federal Reserve Bank of Minneapolis Quarterly Review* 25(1), 2–11.
- Beaudry, P., D. Galizia, and F. Portier (2020). Putting the cycle back into business cycle analysis. *American Economic Review* 110(1), 1–47.
- Beffy, P. O., P. Ollivaud, P. Richardson, and F. Sédillot (2006). New OECD methods for supply-side and medium-term assessments. *OECD Economics Department Working Papers* (482).
- Beveridge, S. and C. R. Nelson (1981). A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the 'business cycle'. *Journal of Monetary Economics* 7(2), 151–174.
- Bordo, M. D. and P. L. Siklos (2018). Central banks: Evolution and innovation in historical perspective. In T. J. Rodney Edvinsson and D. Waldenström (Eds.), *Sveriges Riksbank and the History of Central Banking*, Studies in Macroeconomic History, pp. 26–89. Cambridge, MA: Cambridge University Press.
- Bry, G. and C. Boschan (1971). Programmed selection of cyclical turning points. In *Cyclical Analysis of Time Series: Selected Procedures and Computer Programs*, pp. 7–63. NBER.
- Burns, A. F. and W. C. Mitchell (1946). *Measuring business cycles*. NBER Books.
- Canova, F. (1998). Detrending and business cycle facts. *Journal of Monetary Economics* 41(3), 475–512.
- Canova, F. (2011). *Methods for Applied Macroeconomic Research*. Princeton University Press.

- Cayen, J.-P. and S. van Norden (2005). The reliability of Canadian output-gap estimates. *North American Journal of Economics and Finance* 16(3), 373–393.
- Champagne, J., G. Poulin-Bellisle, and R. Sekkel (2018). The real-time properties of the Bank of Canada’s staff output gap estimates. *Journal of Money, Credit and Banking* 50(6), 1167–1188.
- Chari, V. V., P. J. Kehoe, and E. R. McGrattan (2009). New Keynesian models: Not yet useful for policy analysis. *American Economic Journal: Macroeconomics* 1(1), 242–266.
- Christiano, L. J. and T. J. Fitzgerald (2003). The band pass filter. *International Economic Review* 44(2), 435–465.
- Clark, T. E. and M. W. McCracken (2006). The predictive content of the output gap for inflation: Resolving in-sample and out-of-sample evidence. *Journal of Money, Credit and Banking* 38(5), 1127–1148.
- Coenen, G., F. Smets, and I. Vetlov (2009). Estimation of the Euro Area output gap using the NAWM. *Working Paper* (No. 5). Lietuvos Bankas.
- Cogley, T. and J. M. Nason (1995). Effects of the Hodrick-Prescott filter on trend and difference stationary time series implications for business cycle research. *Journal of Economic Dynamics and Control* 19(1-2), 253–278.
- Coibion, O., Y. Gorodnichenko, and M. Ulate (2018). The cyclical sensitivity in estimates of potential output. *BPEA Conference Draft, Fall*, 343–441.
- Comin, D. and M. Gertler (2006). Medium-term business cycles. *American Economic Review* 96(3), 523–551.
- Danielsson, J., M. Valenzuela, and I. Zer (2018). Learning from history: Volatility and financial crises. *The Review of Financial Studies* 31(7), 2774–2805.
- De Loecker, J. and J. Eeckhout (2020). The rise of market power and the macroeconomic implications. *The Quarterly Journal of Economics*. forthcoming.
- De Masi, P. (1997). IMF estimates of potential output: Theory and practice. IMF Working Paper 97/177.

- De Resende, C. (2014). An assessment of IMF medium-term forecasts of GDP growth. IEO Background Paper 14/01.
- ECB (2000). Potential output growth and output gaps: Concept, uses and estimates. *Monthly Bulletin*, European Central Bank. October 2000, 37-47.
- ECB (2011). Trends in potential output. *Monthly Bulletin*, European Central Bank. January 2011, 73-85.
- Edge, R. M. and J. B. Rudd (2016). Real-time properties of the Federal Reserve's output gap. *Review of Economics and Statistics* 98(4), 785–791.
- Faust, J. and J. H. Wright (2013). Forecasting inflation. In *Handbook of economic forecasting*, Volume 2A, pp. 2–56. Elsevier.
- Fisher, J. D., C. T. Liu, and R. Zhou (2002). When can we forecast inflation? *Federal Reserve Bank of Chicago Economic Perspectives* 1, 30–42.
- Fisher, R. A. (1921). On the probable error of a coefficient of correlation deduced from a small sample. *Metron* 1, 3–32.
- Garratt, A., K. Lee, E. Mise, and K. Shields (2008). Real-time representations of the output gap. *The Review of Economics and Statistics* 90(4), 792–804.
- Hamilton, J. D. (2018). Why you should never use the Hodrick-Prescott filter. *Review of Economics and Statistics* 100(5), 831–843.
- Hamilton, J. D. (2019). Measuring global economic activity. *Journal of Applied Econometrics*. forthcoming.
- Harding, D. and A. Pagan (2002). Dissecting the cycle: A methodological investigation. *Journal of Monetary Economics* 49(2), 365–381.
- Harvey, A. C. and A. Jaeger (1993). Detrending, stylized facts and the business cycle. *Journal of Applied Econometrics* 8(3), 231–247.
- Hodrick, R. J. (2020). An exploration of trend-cycle decomposition methodologies in simulated data. *NBER Working Paper Series* (No. 26750).

- Jönsson, K. (2019). Real-time US GDP gap properties using Hamilton’s regression-based filter. *Empirical Economics*, forthcoming.
- Justiniano, A. and G. E. Primiceri (2008). The time-varying volatility of macroeconomic fluctuations. *American Economic Review* 93(3), 604–641.
- Kaiser, R. and A. Maravall (1999). Estimation of the business cycle: A modified Hodrick-Prescott filter. *Spanish Economic Review* 1(2), 175–206.
- Kaiser, R. and A. Maravall (2012). *Measuring business cycles in economic time series*. Springer Science & Business Media.
- Kamber, G., J. Morley, and B. Wong (2018). Intuitive and reliable estimates of the output gap from a Beveridge-Nelson filter. *Review of Economics and Statistics* 100(3), 550–566.
- King, R. G. and S. T. Rebelo (1993). Low frequency filtering and real business cycles. *Journal of Economic Dynamics and Control* 17(1-2), 207–231.
- Korinek, A. and D. X. Ng (2018). Digitization and the macroeconomics of superstars. *Working paper*. Johns Hopkins University and University of Virginia.
- Langedijk, S. and M. Larch (2011). Testing EU fiscal surveillance: How sensitive is it to variations in output gap estimates? *International Review of Applied Economics* 25(1), 39–60.
- López-Salido, D., J. C. Stein, and E. Zakrajšek. (2017). Credit-market sentiment and the business cycle. *The Quarterly Journal of Economics* 132(3), 1373–1426.
- Marcellino, M. and A. Musso (2011). The reliability of real-time estimates of the Euro Area output gap. *Economic Modelling* 28(4), 1842–1856.
- Mise, E., T.-H. Kim, and P. Newbold (2005). On suboptimality of the Hodrick-Prescott filter at time series endpoints. *Journal of Macroeconomics* 27(1), 53–67.
- Mishkin, F. S. (2007). Estimating potential output. Speech at the Conference on Price Measurement for Monetary Policy, May 24, Federal Reserve Bank of Dallas.
- Nelson, C. R. (2008). The Beveridge-Nelson decomposition in retrospect and prospect. *Journal of Econometrics* 146(2), 202–206.

- Orphanides, A. and S. van Norden (2002). The unreliability of output-gap estimates in real time. *Review of Economics and Statistics* 84(4), 569–583.
- Orphanides, A. and S. van Norden (2005). The reliability of inflation forecast based on output gap estimates in real time. *Journal of Money, Credit and Banking* 37, 583–600.
- Perron, P. and T. Wada (2009). Let’s take a break: Trends and cycles in US real GDP. *Journal of Monetary Economics* 56(6), 749–765.
- Perron, P. and T. Wada (2016). Measuring business cycles with structural breaks and outliers: Applications to international data. *Research in Economics* 70(2), 281–303.
- Richter, B., M. Schularick, and I. Shim (2019). The costs of macroprudential policy. *Journal of International Economics* 118, 263–282.
- Richter, B., M. Schularick, and P. Wachtel (2020). When to lean against the wind. *Journal of Money, Credit, and Banking*. forthcoming.
- Rosnick, D. (2016). Potential for trouble: The IMF’s estimates of potential GDP. Washington, DC: Center for Economic and Policy Research.
- Schüler, Y. S. (2018). On the cyclical properties of Hamilton’s regression filter. Deutsche Bundesbank Discussion Paper 03/2018.
- Shackleton, R. (2018). Estimating and projecting potential output using CBO’s forecasting growth model. Congressional Budget Office, Working Paper 3.
- St-Amant, P. and S. van Norden (1997). Measurement of the output gap: A discussion of recent research at the Bank of Canada. *Technical Report* (No. 79), 1–62.
- Stock, J. H. and M. W. Watson (1999a). Business cycle fluctuations in US macroeconomic time series. *Handbook of Macroeconomics* 1, 3–64.
- Stock, J. H. and M. W. Watson (1999b). Forecasting inflation. *Journal of Monetary Economics* 44(2), 293–335.
- Stock, J. H. and M. W. Watson (2007). Why has U.S. inflation become harder to forecast? *Journal of Money, Credit and Banking* 39(s1), 3–33.

- Stock, J. H. and M. W. Watson (2008). Phillips curve inflation forecasts. In J. Fuhrer, Y. K. Kodrzycki, J. S. Little, and G. P. Olivei (Eds.), *Understanding Inflation and the Implications for Monetary Policy: A Phillips Curve Retrospective*, pp. 101–186. Cambridge, MA: MIT Press.
- van Norden, S. (2002). Filtering for current analysis. *Working Paper 2002-28*. Bank of Canada.
- Van Zandweghe, W. (2017). The changing cyclicity of labor force participation. *Federal Reserve Bank of Kansas City Economic Review Q III*, 5–34.
- Vetlov, I., T. Hlédik, M. Jonsson, H. Kucsera, and M. Pisani (2011). Potential output in DSGE models. *ECB Working Paper* (No. 1351).
- Watson, M. W. (2007). How accurate are real-time estimates of output trends and gaps? *FRB Richmond Economic Quarterly* 93(2), 143–161.
- Wolf, E., F. Mokinski, and Y. Schüler (2020). On adjusting the one-sided Hodrick-Prescott filter. *Deutsche Bundesbank Discussion Paper 11/2020*.
- Wolters, M. H. (2018). How the baby boomers’ retirement wave distorts model-based output gap estimates. *Journal of Applied Econometrics* 33(5), 680–689.