JOSEFINE QUAST
MAIK H. WOLTERS

Reliable Real-time Output Gap Estimates

Based on a Modified Hamilton Filter

Institute for Monetary and Financial Stability

GOETHE UNIVERSITY FRANKFURT

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

Josefine Quast[*]

University of Jena

Maik H. Wolters[†]

University of Jena, Kiel Institute for the World Economy, IMFS at Goethe University Frankfurt

March 26, 2019

Abstract

We propose a simple modification of the time series filter by Hamilton (2018) 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 log real GDP. While this approach yields a cyclical component of GDP 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 short business cycles are muted and medium length business cycles are amplified. Further, the estimated trend is as volatile as GDP itself and can thus hardly be interpreted as potential GDP. A simple modification that is 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 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 the Bandpass filter.

---

[*]josefine.quast@uni-jena.de

[†]maik.wolters@uni-jena.de

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.[1] 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 (2018b) has proposed a new regression based filter for detrending time series as an alternative to the HP filter. Hamilton proposes using 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 the primary reason for forecast errors over such a horizon is cyclical factors such as whether a recession occurs over the next two years rather than large trend changes. 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), because it is a one-sided filter.[2]

Hamilton applies the filter to GDP and shows that its cyclical component turns negative during NBER defined recessions and positive during expansions. Hence, it is a potentially useful measure of the business cycle and the approach can possibly solve the long-standing problem of the unreliability of real-time output gap estimates. 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 amplified substantially relative to longer and shorter cycles. The latter are muted almost completely. Further, the extracted GDP trend is not smooth, but is as volatile as GDP itself, so that it cannot be interpreted as measuring potential GDP.

We propose a simple modification of the Hamilton filter that shares its favorable real-time properties, but leads to a more even coverage of typical business cycle frequencies. 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. In this way short, medium and long business cycles are covered more evenly, so that the original detrending method by Hamilton is turned into one that is useful for both detrending and business cycle filtering.[3] The modified filter further avoids unreasonable spikes and an unreasonably large amplitude of the cyclical component of GDP and yields a smooth trend of GDP. Hence, an interpretation of the trend as potential GDP and of the cyclical component as an 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 U.S. output gaps computed with the modified Hamilton filter and compare it to output gaps computed with the original one and the HP and the Bandpass (BP) filter as examples of commonly used simple statistical trend-cycle decomposition techniques. In particular, we answer two

---

[1]Other papers have confirmed the unreliability of real-time output gap estimates produced with common statistical methods for other economies (see, e.g., Bernhardsen et al., 2005; Cayen and van Norden, 2005; Marcellino and Musso, 2011).

[2]Hamilton (2018b) argues that that his approach does not only solve the end-of-sample bias of the HP filter, but also two additional drawbacks: the creation of spurious cycles and an ad hoc smoothing parameter that is at odds with the one chosen by a formal statistical procedure (see also Baxter and Stockman, 1989; Baxter, 1991; King and Rebelo, 1993; Harvey and Jaeger, 1993; Cogley and M.Nason, 1995; Canova, 1998, for problems with the HP filter).

[3]The Hamilton filter has already been used in the literature for detrending various macroeconomic variables (see, e.g., Ahn and Hamilton, 2016; Bordo and Siklos, 2017; López-Salido et al., 2017; Richter et al., 2017; Van Zandweghe, 2017; Danielsson et al., 2018; Richter et al., 2018). In all these studies the obtained trends and gaps are interpreted economically. For this, an even coverage of the relevant frequencies seems to be important. Hamilton himself uses the filter to construct a cyclical measure of global economic activity (Hamilton, 2018a).

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? It goes without saying that a positive answer to the first question needs not necessarily entail a positive answer to the second one. Fortunately, we find that the answer to both questions is yes for the modified Hamilton filter.

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 multiple ex post output gaps from policy institutions with the real-time output gap based on the modified Hamilton filter are significantly stronger than with those based on the HP or the BP filter.

Second, we test whether an output gap estimate has predictive content for output growth and inflation. For example, if output is below potential, this implies that output growth should increase in the future so that output reverts back to potential. Further, output gaps should have according to Phillips curve models implications for future inflation.

While we do not find significant differences in output growth forecasting accuracy for the full sample, the output gap based on the modified Hamilton filter leads to significantly better forecasts compared to the HP and BP filter when excluding the Great Recession. Evaluating out-of-sample inflation forecasts by means of a standard Phillips curve forecasting equation manifests the finding 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), rather than output gap measurement problems. It is, however, noteworthy that the modified Hamilton filtered real-time output gap shows small though insignifcant gains in inflation forecasting accuracy compared to the HP and BP filtered output gaps.

Overall, we find that output gap estimates based on the modified 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.

2 A Simple Modification of the Hamilton Filter

Hamilton (2018b) proposed using the 8-quarter forecast error of a projection based on an AR(4) model as the cyclical component of a macroeconomic time series. Hence, to detrend a macroeconomic time series, y_t, the following simple autoregression can be estimated by OLS:

$$y_t = \beta_0 + \beta_1 y_{t-8} + \beta_2 y_{t-9} + \beta_3 y_{t-10} + \beta_4 y_{t-11} + \upsilon_t. \tag{1}$$

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

$$\hat{\upsilon}_t = y_t - \hat{\beta}_0 - \hat{\beta}_1 y_{t-8} - \hat{\beta}_2 y_{t-9} - \hat{\beta}_3 y_{t-10} - \hat{\beta}_4 y_{t-11}. \tag{2}$$

Applied to log quarterly real GDP it is tempting to interpret $\hat{\upsilon}_t$ as an output gap. However, Hamilton (2018b) and Schüler (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 two 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, $\tilde{y}_t$:

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

$$\hat{\upsilon}_{t,i} = y_t - \hat{\beta}_{0,i} - \hat{\beta}_{1,i} y_{t-i} - \hat{\beta}_{2,i} y_{t-i-1} - \hat{\beta}_{3,i} y_{t-i-2} - \hat{\beta}_{4,i} y_{t-i-3}. \tag{4}$$

In the following, we show based on analysis in the frequency domain that this indeed mitigates the problematic cyclical properties of the original filter.

The spectral density of a covariance stationary process y_t at frequency $\omega \in [0, 2\pi]$ is given by the Fourier transform of the autocovariance function (see, e.g., Canova, 2011):

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

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:

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

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:

$$SD_x(\omega) = |B(\omega)|^2 SD_y(\omega). \tag{7}$$

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 $\omega = 2\pi/6$ to $\omega = 2\pi/32$, and a value of zero for all other frequencies.

The frequency response for linear filters as in equation (6) is given by:

$$B(\omega) = b_0 + \sum_{j=-\infty, j\neq 0}^{\infty} b_j cos(\omega j). \qquad (8)$$

For the Hamilton filter the weights $b_{-1}, \ldots, b_{-4}$ are given by the estimated coefficients $\hat{\beta}_1, \ldots, \hat{\beta}_4$ from equation (2), $b_0 = 1$ and all other weights equal zero. From this the Power Transfer Function can directly be computed as the square modulus of $B(\omega)$.

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 U.S. log real GDP data from 1947Q1 to 2017Q4. The grey 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 (panel A) as well as a version that shows the cycle length in quarters on the horizontal axis (panel B) to facilitate the reading of the figure.[4] 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 in the representation showing the cycle length it is clear that these are irregular fluctuations, i.e. noise, with little practical relevance.

It is apparent that the original Hamilton filter eliminates business cycles ranging from 6 to 10 quarters almost completely. Hence, short business cycle frequencies are not present in a Hamilton filtered output gap. On the other hand, cycles with a duration of 10 to 20 quarters are amplified substantially compared to short cycles and long cycles of 20 to 32 quarters. 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. Ideally, only frequencies between 6 and 32 quarters would be present in the output gap. Yet, only the two-sided Bandpass filter is able to get close to such an ideal filter (see Christiano and Fitzgerald, 2003, for PTFs of the approximate Bandpass filter). However, such is only obtainable in retrospect using two-sided filtering, but not in real-time applications.

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 (2018b). Yet, it is very simple to compute. Further, the smoothing of different PTF peaks and troughs improves substantially the treatment of high frequency noise in the obtained output gap and trend measures. While the gap and the trend include erratic movements each quarter for the original Hamilton filter, those from the modified Hamilton filter are very smooth as shown in Figure 2 (Panel A and B). This difference is most clearly visible for the estimated trend growth rate (Panel C) that is as volatile as the growth rate of GDP itself for the original Hamilton filter. Due to the structure of equation (1) the period t growth rate of trend GDP and the period $t-8$ GDP growth rate are very similar (correlation coefficient: 0.93).[5] Hence, the trend measure extracted via the original Hamilton filter can hardly be interpreted as potential GDP, which in turn makes the economic interpretation of the resulting output gap difficult. By contrast,

---

[4] A Fourier frequency ω is associated with cycle length $p = 2\pi/\omega$.

[5] For the original Hamilton filter trend growth is given by: $\Delta\hat{y}_t = \hat{\beta}_1\Delta y_{t-8} + \hat{\beta}_2\Delta y_{t-9} + \hat{\beta}_3\Delta y_{t-10} + \hat{\beta}_4\Delta y_{t-11}$.
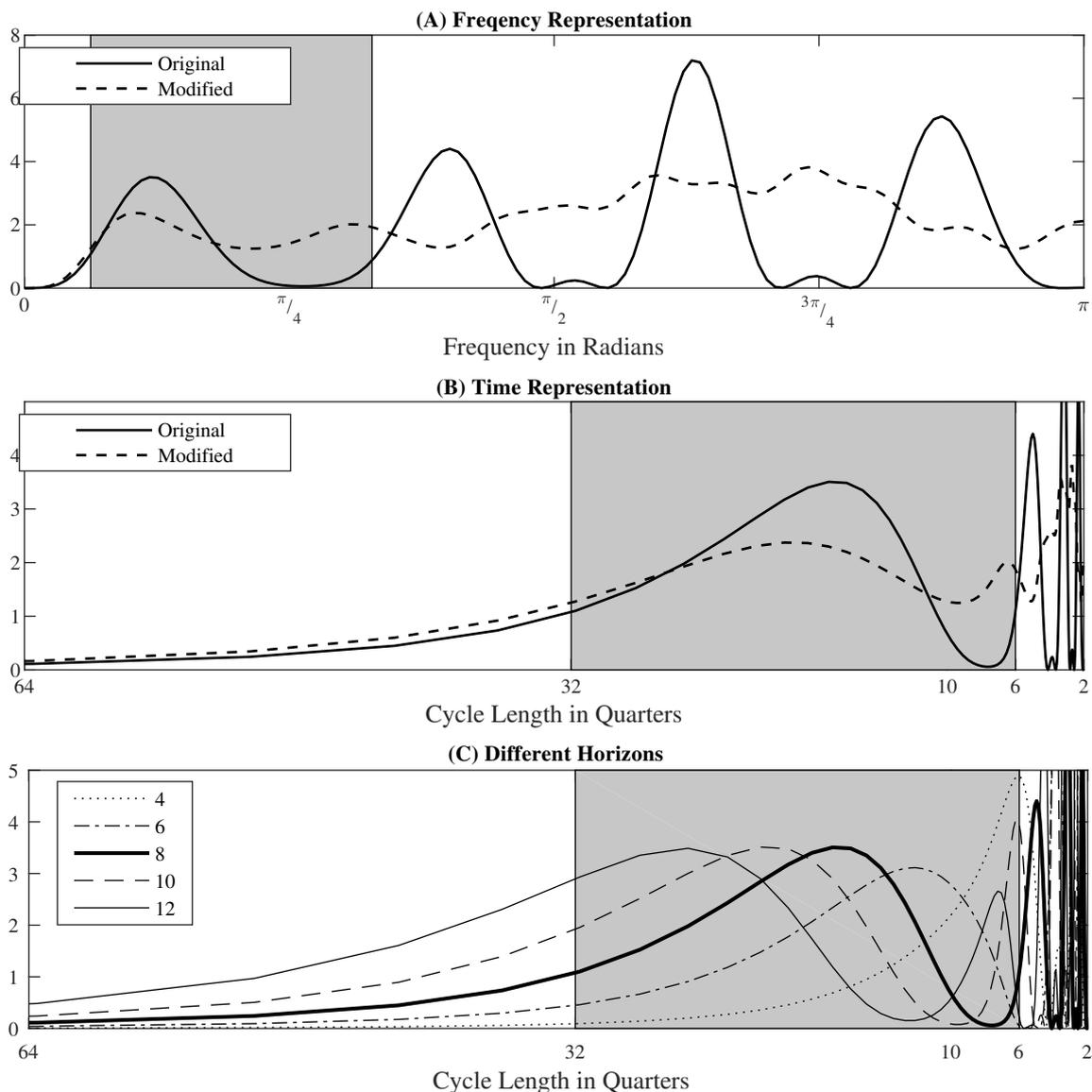
Figure 1: Power Transfer Functions in Frequency and Time Representation

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 as, for example, based on the HP or the BP filter. Further, the smooth trend of the modified Hamilton filter leads to a smoother output gap compared to the original Hamilton filter.

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 real-time data on quarterly real GDP from the Federal Reserve Bank of Philadelphias's real-time data set. The first data vintage is from 1965Q4 and the last data vintage is from 2018Q1. Data in all vintages starts in 1947Q1 and ends one quarter before the publication date of the data vintage, i.e. in 1965Q3 for the first and in 2017Q4 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 then collect the end-point estimates of these vintages to get a real-time output gap series
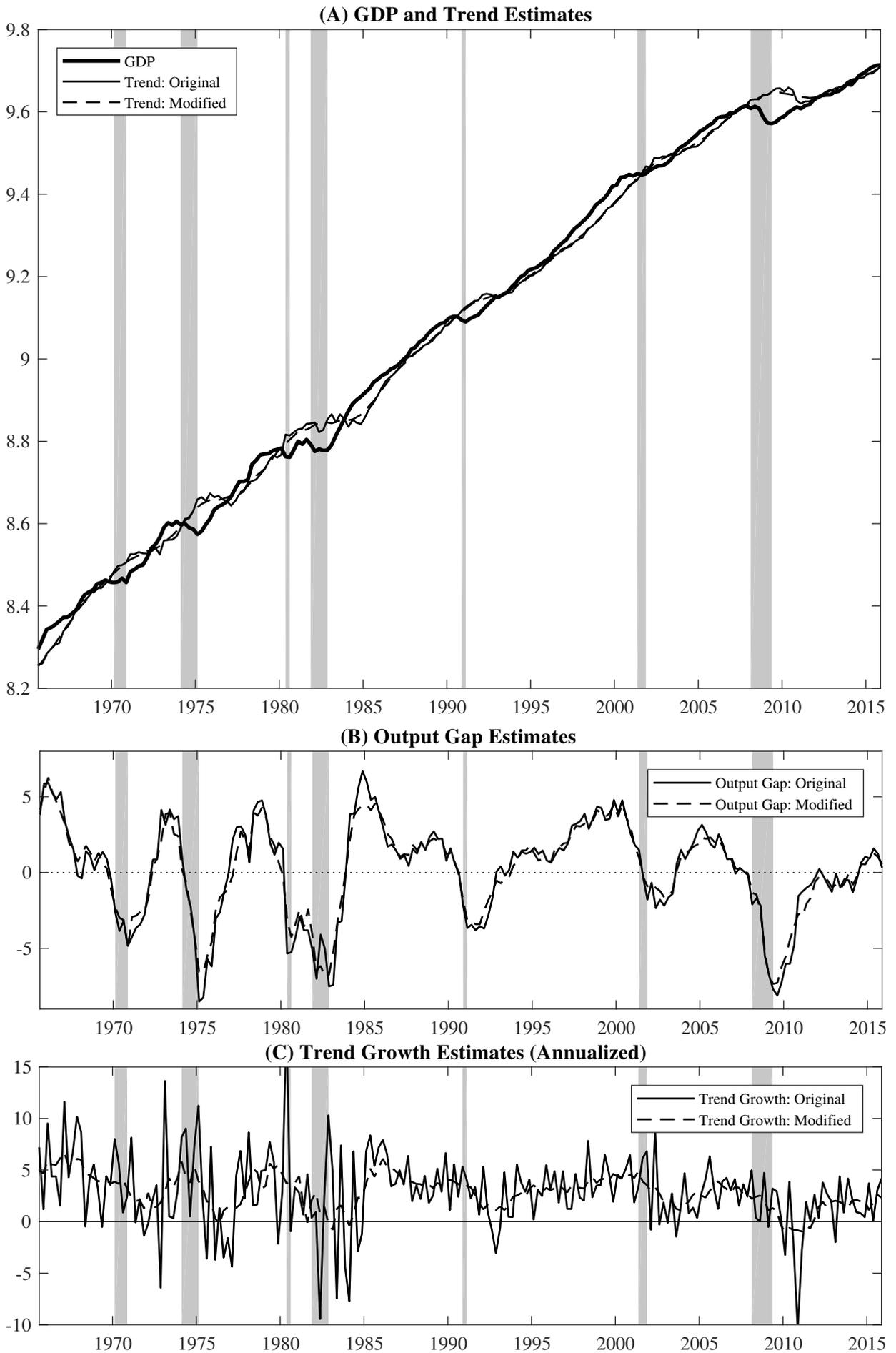
Figure 2: Trend, Output Gap and Trend Growth: Original and Modified Hamilton Filter

for each filtering method.

We use the last data vintage from 2018Q1 as our measure of final revised data and define the output gap revision as the difference between the final 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 are very close to each other, we follow Orphanides and van Norden (2002) and discard the last two years of observations. Hence, our results for the revised gaps are based on the 2018Q1 vintage with gaps being estimated until 2015Q4. This proceeding is reasonable since most revisions take place in the first two 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 2015Q4.

To distinguish between data revisions and revisions due to the filter-induced structure, we compute a quasi real-time series. That is, we estimate output gaps recursively 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 are based on data covering the exact same time periods.

Figure 3 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 the BP filter differ markedly 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 extend filter-induced revisions, while data revisions play only 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 0.04 and the BP filter the largest 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 0.86 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 revised estimate 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 0.3 for the two versions of the Hamilton filtered output gap, while it is with values of 0.6 and almost 1.0 much larger for the BP and HP filtered output gaps. In principle, a noise-to-signal ratio based on dividing the RMSE of the 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, because 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 (Kilian and Lütkepohl, 2017).[6] The HP and BP filter based noise-to-signal ratios are significantly larger than the one of the modified Hamilton filter. Finally, we also report the fraction of observations in which the final

---

[6]We use 5000 bootstrap replications and a block size of 4 to compute the empirical distributions of the noise-to-signal ratios.
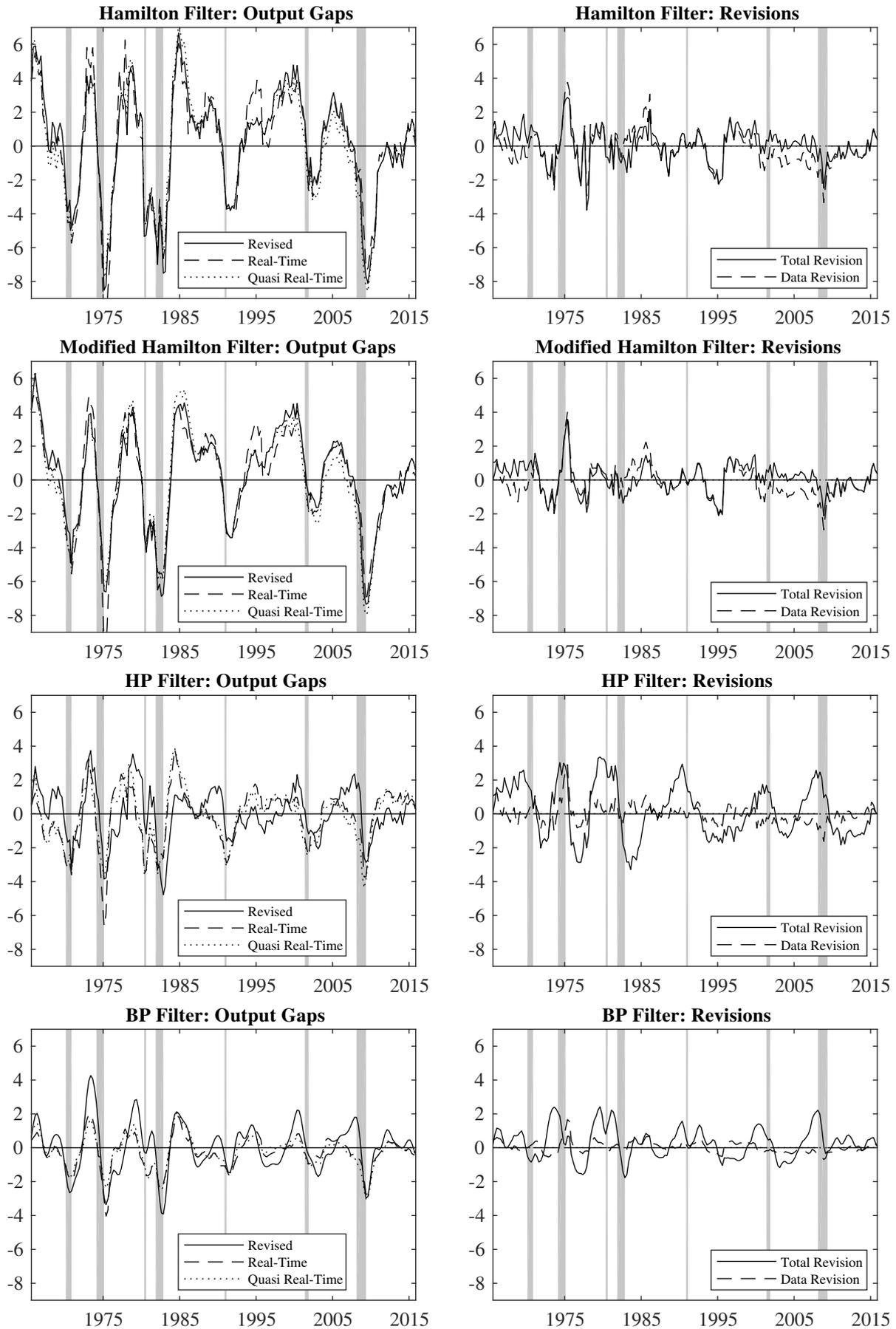
7

Figure 3: Output Gaps and Revisions

revised and the real-time output gap estimates have opposite signs. This is the case for only 3% and 6% of observations for the modified and the original Hamilton filter, respectively, but for 19% and 31% of observations for the BP and HP filtered output gaps, respectively.

Table 1: Statistics of Total Revisions and Data Revisions of Output Gaps

| | Mean | SD | RMSE | Noise-Signal Ratios | | | | Opposite Sign |
				SD	Diff	RMSE	Diff	
				Total Revisions				
Hamilton	0.04	1.00	1.00	0.30	0.01	0.29	0.01	0.06
Modified	0.06	0.86	0.86	0.29	-	0.29	-	0.03
HP	0.17	1.58	1.59	0.97	0.68***	0.97	0.68***	0.31
BP	0.29	0.92	0.96	0.59	0.30***	0.62	0.33***	0.19
				Data Revisions				
Hamilton	-0.12	1.10	1.10	0.32	0.00	0.33	0.00	0.07
Modified	-0.11	0.96	0.96	0.32	-	0.32	-	0.06
HP	0.01	0.60	0.60	0.37	0.05	0.37	0.05	0.04
BP	0.04	0.34	0.34	0.22	-0.10 **	0.22	-0.10 **	0.08

Notes: *, **, and *** denote significance on the 10, 5, and 1% significance level. 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 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 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. Hence, 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 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 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.

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. Data revisions lead only to small revisions of the output gap estimates as discussed above. Regarding the estimated AR(4) parameters, Figure 4 shows for the original Hamilton filter how they change over time via the recursive estimation procedure and how they converge to the full sample estimates. 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.

Determining the current output gap is an important element in the policy process at policy institutions like, for example, central banks. To understand to which extent the revisions of output gaps matter in the policy process of a central bank one can think about the implied interest rate distortions. For example,
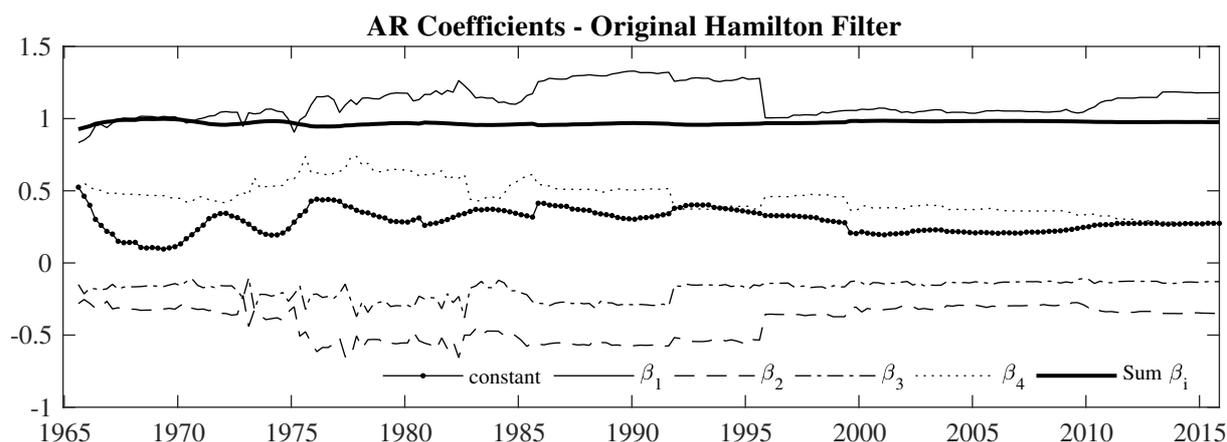
Figure 4: Recursive estimates of the AR(4) parameters of the original Hamilton filter

when using the standard Taylor rule, the interest rate reaction to the output gap is 0.5. Hence, the interest rate distortions through output gap revisions are given by multiplying the output gap revisions with 0.5. The usage of real-time compared to revised output gap estimates would imply interest rates that are on average 0.08 percentage points too low for the HP filter, 0.14 percentage points too low for the BP filter, but only 0.02-0.03 percentage points too low for the two Hamilton filter based output gaps. More interesting might be the maximal mistake when using real-time compared to revised output gap estimates. The numbers (1.90, 1.79, 1.68, and 1.20 percentage points for the original Hamilton, the modified Hamilton, the HP and the BP filter, respectively) are however not very meaningful. Due to the higher amplitude of the two versions of the Hamilton filter, the output gap coefficients in a policy rule would be chosen smaller. The standard deviations of the revised Hamilton filtered output gaps are approximately twice as large (3.26 and 2.92 for the original and the modified version) compared to the HP filtered (1.52) and the BP filtered (1.45) output gaps. Hence, we adjust the numbers for the BP and the two Hamilton filter to make them comparable to the HP filter by dividing by the standard deviation of the respective output gap and multiplying the result with the standard deviation of the HP filtered output gap. This results in maximal interest rate setting mistakes of 1.68 percentage points for the HP filtered (no adjustment), 1.26 for the BP filtered output gap (1.20/1.45*1.52) and only 0.89 (1.90/3.26*1.52) for the original and 0.93 (1.79/2.92*1.52) for the modified Hamilton filtered output gap. Thus, the maximal interest rate mistake implied by output gap revisions would be almost twice as large for the HP filtered compared to using one of the two Hamilton filter based output gaps.

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. A first basic requirement of the output gap being stationary is fulfilled as long as the original time series is integrated of order 4 or less (Hamilton, 2018b).[7] 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

---

[7]Some well-known output gap estimation methods fail to do so. Besides deterministic detrending procedures also DSGE model-based output gaps can be nonstationary (Wolters, 2018).

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 (2018b) justifies the exclusion of trend changes over a two-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 volatility is much larger than the volatility of the trend. An example is the standard smoothing parameter of 1600 of the HP filter. A statistical formalization of the choice of the HP smoothing parameter would result in a much lower cyclical relative to trend volatility.[8] Similarly, when using the Beveridge-Nelson 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 without imposing such a prior belief, changes in the trend component dominate (Kamber et al., 2018). Hence, the a-priori choice that over a two-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 one rather than two 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 0.46 for the original Hamilton, 0.54 for the modified Hamilton, 0.40 for the HP and 0.41 for the BP filtered output gaps.[9] 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 U.S. output gap estimates of the Federal Reserve, 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 U.S. business cycle that include a substantial amount of economic expertise rather than being based only on statistical models. Second, the assessment 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. Yet, all institutions combine this 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,

---

[8]The standard HP smoothing parameter of 1600 is based on the assumption that the volatility of the cyclical component is $\sigma_c = 5$ and the volatility of the trend component is $\sigma_t = 0.125$ resulting in $\lambda = \sigma_c^2/\sigma_t^2 = 1600$. A statistical formalization of the choice of λ leads to a value of 0.245 for U.S. quarterly GDP (Hamilton, 2018b).

[9]One would expect that output falls below potential during a recession implying a decreasing output gap, while one would expect it to rise above potential during a boom implying an increasing output gap. Based on this, we compute correlations with NBER recession dates based on the changes in the output gap rather than the output gap itself.

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 of 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. For the Fed and the CBO quarterly data is available, while for the IMF and the OECD only annual data is available. Further, for the Fed data is only available until 2012Q3. The data is based on the Greenbook which is made available to the public with a 5 year lag. 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 and the OECD starts in 1985.

Table 2 shows correlations for the quarterly output gap series covering the period 1965Q3 until 2012Q3 on the left and for the annual series for data from 1985 to 2012 on the right.[10] It gets apparent that the two Hamilton filtered real-time output gaps are highly correlated with all policy institutions' output gaps with correlation coefficients ranging from 0.55 to 0.86. The output gap based on the modified Hamilton filter has an even higher correlation with all four expert output gap estimates than the one based on the original filter. The HP and BP filtered real-time output gap correlations with the policy institutions' gaps are much lower ranging from 0.16 to 0.59, with those of the HP filter being particularly low. Both Hamilton filtered output gaps have a significantly higher correlation with the expert output gaps compared to the output gaps based on the HP or the BP filter. Interestingly, for all four statistical output gap measures the correlations are higher with the output gaps of the FED and the CBO—the two U.S. institutions—than with the IMF and OECD—the two international institutions. Overall, these results imply that the real-time Hamilton filtered output gaps—and in particular the version based on the modified Hamilton filter—are able to reflect the ex post expert evaluation of the U.S. 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: Evaluation Based on Expert Output Gap Estimates: Correlations

	Quarterly Data (1965Q3-2012Q3)		Annual Data (1985-2012)			
	FED	CBO	FED	CBO	IMF	OECD
Hamilton	0.59	0.68	0.78	0.82	0.69	0.55
Modified	0.64	0.73	0.83	0.86	0.73	0.60
HP	0.25***	0.36***	0.33***	0.43***	0.29**	0.16**
BP	0.49**	0.59**	0.51**	0.57**	0.49	0.37

Notes: *, **, and *** denote significant differences relative to the respective correlation with the modified Hamilton filter on the 10, 5, and 1% significance level.

---

[10]The sample end is chosen to make the results for the FED's output gap comparable to the other ones. Extending the sample for the other output gap series yields very similar results. For simplicity, we base the annualized series for the real-time gap measures on yearly averages of the real-time series. When we account for within year revisions for quarters one to three of a given year to mimic the information set of a policy maker we obtain very similar results, both regarding correlations as well as p-values of differences between correlations.

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 is negative, one 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. We use a standard forecast equation in which output growth h periods ahead is predicted using the real-time output gap:

$$y_{t+h} - y_t = \alpha + \beta \hat{c}_t + \varepsilon_{t+h|t}, \tag{9}$$

where y denotes log real GDP, $\hat{c}$ the estimated real-time output gap and $\varepsilon_{t+h|t}$ the forecast error. The equation is estimated with OLS. The first subsample goes from 1965Q3 to 1975Q2 covering 10 years of data. The sample is then recursively extended quarter-by-quarter. Forecasts are computed for horizons of 1 to 12 quarters ahead. 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 shows RMSEs and the estimated slope coefficient β averaged over all subsamples. The RMSEs are reported for the Hamilton, modified Hamilton and BP filtered output gaps relative to the HP filtered one. A number smaller than 1 indicates a superior forecasting performance of the Hamilton, modified Hamilton or BP filtered output gap, while a number larger than 1 indicates that the HP output gap dominates. The RMSEs of output growth forecasts based on the four different real-time output gap measures are almost identical. The small differences are insignificant based on the Dieboldt-Mariano test. The signs of the slope coefficients show that all four output gap measures predict trend-reverting output growth rates. The slope coefficients are negative and significant from horizon 6 onwards for the two Hamilton and the BP filtered output gap and from horizon 11 onwards for the HP filtered output gap. The size of the slope coefficients cannot be directly compared, because of the different amplitudes of the three output gap measures.

The results change when leaving out the period of the Great Recession and its aftermath. When ending the forecast evaluation in 2007Q4 the RMSEs of the two Hamilton filtered output gaps relative to the HP filtered one drop to values around 0.9 for horizons 8 to 12 and the difference is significant on the 10% level. The RMSEs relative to the BP filter drop to values around 0.97 for horizons 8 to 12, but differences remain insignificant. One reason why the Hamilton filtered output gap's relative performance improves when leaving out the post 2007 period is that it adjusts to structural breaks in the level of GDP late. The Hamilton filter needs about three years to adjust to the level shift of GDP after the Great Recession. This is why the output gap of the Hamilton filter is highly negative until the end of 2010. The HP and BP filter based output gap measures are more flexible with respect to trend adjustments if there are very large business cycle fluctuations, so that the output gap goes back to zero more quickly after the crisis. The recovery after the crisis was slow, so that there are no above average output growth rates after 2010 as one would expect from the prolonged highly negative output gap of the Hamilton filter until the end of 2010.

Table 3: Output Growth Forecast Evaluation

Horizon	Relative RMSE			Slope Coefficient (p-value)			
	Hamilton/HP	Modified/HP	BP/HP	Hamilton	Modified	HP	BP
1	1.02	1.02	1.02	0.03 (0.24)	0.03 (0.22)	0.13 (0.02)	0.08 (0.34)
2	1.01	1.01	1.01	0.02 (0.53)	0.03 (0.54)	0.18 (0.05)	0.04 (0.60)
3	1.01	1.01	1.01	-0.02 (0.67)	-0.02 (0.69)	0.17 (0.14)	-0.08 (0.70)
4	1.02	1.01	1.02	-0.06 (0.37)	-0.07 (0.34)	0.13 (0.27)	-0.26 (0.26)
5	1.01	1.00	1.02	-0.12 (0.13)	-0.15 (0.10)	0.05 (0.45)	-0.47 (0.07)
6	1.01	0.99	1.01	-0.18 (0.04)	-0.22 (0.03)	-0.05 (0.63)	-0.69 (0.02)
7	0.99	0.98	1.00	-0.24 (0.01)	-0.29 (0.01)	-0.15 (0.57)	-0.89 (0.00)
8	0.98	0.97	0.99	-0.30 (0.00)	-0.36 (0.00)	-0.27 (0.29)	-1.07 (0.00)
9	0.97	0.97	0.98	-0.35 (0.00)	-0.41 (0.00)	-0.36 (0.15)	-1.21 (0.00)
10	0.97	0.97	0.97	-0.39 (0.00)	-0.45 (0.00)	-0.44 (0.07)	-1.33 (0.00)
11	0.97	0.97	0.97	-0.42 (0.00)	-0.49 (0.00)	-0.51 (0.03)	-1.45 (0.00)
12	0.97	0.97	0.97	-0.44 (0.00)	-0.53 (0.00)	-0.60 (0.01)	-1.54 (0.00)

Notes: *, **, and *** denote significance on the 10, 5, and 1% significance level based on a two-sided Dieboldt-Mariano (1995) test (refers to the relative RMSEs in columns 2-4).

Overall, the results from this exercise are inconclusive in judging which output gap measure has favorable properties. On the one hand the two Hamilton filtered output gaps perform somewhat better than the HP filtered one, but on the other hand a drawback of the Hamilton filter is its inflexibility to adjust to structural breaks early on. Differences to the BP filter with respect to the prediction of trend-reverting output growth rates are small.[11]

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:

$$\pi_{t+h} - \pi_t = \alpha + \sum_{i=0}^{p} \beta_i \Delta\pi_{t-i} + \sum_{i=0}^{q} \gamma_i \hat{c}_{t-i} + \varepsilon_{t+h|t}, \qquad (10)$$

where π_t denotes U.S. PCE inflation, $\hat{c}_t$ the estimated real-time output gap and $\varepsilon_{t+h|t}$ the forecast errors.[12] 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 $q \in [0, 12]$. When doing 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 extended 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 (10).

Table 4 shows root mean squared inflation forecast errors based on the two Hamilton filtered output gaps relative to those based on the HP and the BP filter and based on the specification without output gap.

---

[11] When conducting the same exercise with final revised output gaps, the RMSEs of the Hamilton filtered output gaps are significantly larger than the ones of the HP and BP filter. However, this is not a suitable criterion to conclude that for historical analyses revised HP and BP filtered output gaps are preferable. The low RMSEs merely reflect the two-sided nature of the HP and BP filter. When using future information to measure the output gap, then it is not surprising that the predictive content regarding these future realizations is higher compared to using a one-sided filter like the Hamilton filter.

[12] This specification is based on changes in inflation, i.e. it imposes a unit root in inflation. We also estimated models in terms of inflation levels and specifications that include a relative import price inflation term. The forecasting results are very similar.

Table 4: Inflation Forecast Evaluation

| Horizon | Relative RMSE | | | | | |
	Hamilton/HP	Modified/HP	Hamilton/BP	Modified/BP	Hamilton/No Gap	Modified/No Gap
1	1.02	1.01	1.02	1.02	1.10***	1.10***
2	1.01	0.98	0.99	0.97	1.01	0.99
3	1.00	0.99	1.00	0.99	1.04	1.03
4	1.00	1.01	1.00	1.01	1.01	1.02
5	1.01	1.00	0.99	0.98	0.99	0.97
6	1.01	1.00	0.97	0.96	0.96	0.95
7	0.97	0.97	0.97	0.97	0.93	0.92
8	1.01	1.01	1.01	1.01	0.96	0.95
9	1.01	1.01	1.00	1.00	0.97	0.96
10	1.01	1.01	1.00	1.00	0.97	0.96
11	1.01	1.01	1.02**	1.02**	0.96	0.95
12	1.01	1.01	1.01	1.00	0.96	0.95

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

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), 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 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

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. 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 $(4 - 8$ quarters) and wider $(2 - 10)$ 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 higher 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 higher 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. If one wants to stick to a standard definition of business cycle frequencies, then our baseline specification is preferable. Regarding output growth forecasts differences across specifications are very small. Finally, for inflation forecasts the bands centered around a forecast horizon of 6 quarters show small gains in accuracy compared to the baseline range, though differences are only significant for 1- and 2-quarter-ahead inflation forecasts. Based on these exercises, a trade-off gets 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.

Optimizing 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 to which extent the output gap measurement can be improved by choosing an optimized weighting scheme instead. First, we choose weights in order achieve a PTF that is as close as possible to the ideal PTF that takes a value of one 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 filter's PTF over the cycles from 6 to 32 quarters length and zero otherwise. This average is close to 2, 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 and to sum to one.

In both optimized versions, only 4 of the 9 horizons have weights that are larger than zero. These are horizons 4 (weight 0.45), 6 (0.24), 9 (0.17) and 12 (0.14) for fitting the ideal PTF and 4 (0.27), 6 (0.26), 9 (0.32) and 12 (0.15) for fitting the average PTF of the business cycle frequencies based on the original Hamilton filter.[13] Hence, the results indicate that it is sensible to include shorter as well as longer horizons. They also indicate that $h = 8$ serves as a good center for the band of forecasting horizons considered in our modification. 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 5 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 very 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.

We also repeated all other analyses with the output gaps based on optimized weights.[14] In all evaluation exercises, the results are very similar for the versions with optimized weights and the one based on a simple average of forecast errors of different horizons. The correlations of the simple modified Hamilton filter with revised expert gaps are slightly higher than those of the modified Hamilton filter

---

[13]We cut off the optimization after a maximum cycle length of 128 quarters. The results are the same up to rounding precision if we chose a shorter (64 quarters) or longer (256) cutoff point.

[14]To keep the exercise tractable, we determine the optimal weights based on the full sample and assume that these are known a priori.
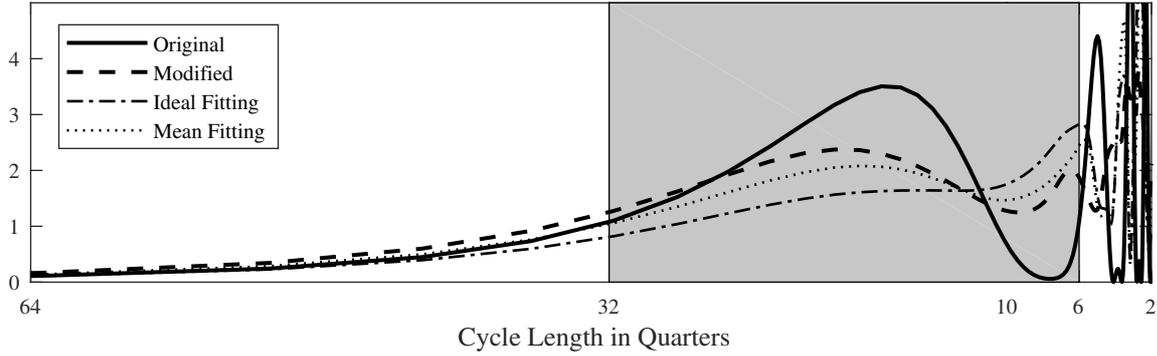
Figure 5: Original, Modified and Fitted Power Transfer Functions in Time Representation

with optimized weights, there are no differences with respect to the output growth forecasts and only small differences for inflation forecasts. For the latter the output gaps based on optimized weights have a slightly higher forecast accuracy for short horizons of 1 and 2 quarters ahead.

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 implement Bai-Perron procedures (Bai and Perron, 1998, 2003) to account for multiple structural changes in U.S. real GDP.

We mimic the information set available in real time and use Bai-Perron procedures recursively to detect structural changes in the estimated trend growth rate represented by $\hat{\beta}_0$ for each data vintage.[15] We find evidence for a first structural change in trend growth in the beginning of the 1960s. Subsequent changes are then associated with the productivity slowdown during the 1970s, the start of the Great Moderation during the mid-1980s and the run-up to the dot-com bubble starting in the mid-1990s. While breaks are detected reliably about 5 to 6 years after their occurrence, exact dates of the determined break points are sensitive to the specific data vintage used. Afterwards, we re-estimate the Hamilton-filtered output gaps for each data vintage accounting for the detected structural breaks in the constant.

Figure 6 shows the unadjusted and the Bai-Perron adjusted real-time output gap estimates. Overall, the break-point adjusted output gap is higher than the baseline output gap from the mid-1980s onwards. Apart from that, the dynamics are very similar. The reason is that the magnitude of the estimated breaks is relatively small compared to the unconditional variance of GDP growth, so that the estimated autoregressive coefficients are little impacted by allowing for breaks. Therefore, the original and modified Hamilton filtered output gaps are robust to accounting for structural change in trend growth.[16]

Comparison to 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 Beveridge-Nelson decomposition to estimate output gaps. The method is related to the Hamilton filter, because 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

---

[15]We compute the break dates based on Hamilton's original procedure, test for changes in the constant and use a 15% trimming as minimum distance between two potential breaks in the respective sample.

[16]When accounting for structural breaks, the size of revisions increases somewhat, because a breakpoint is first detected about five years after it occured. The correlations of real-time and revised Hamilton filtered output gaps drop from about 0.95 to about 0.81.
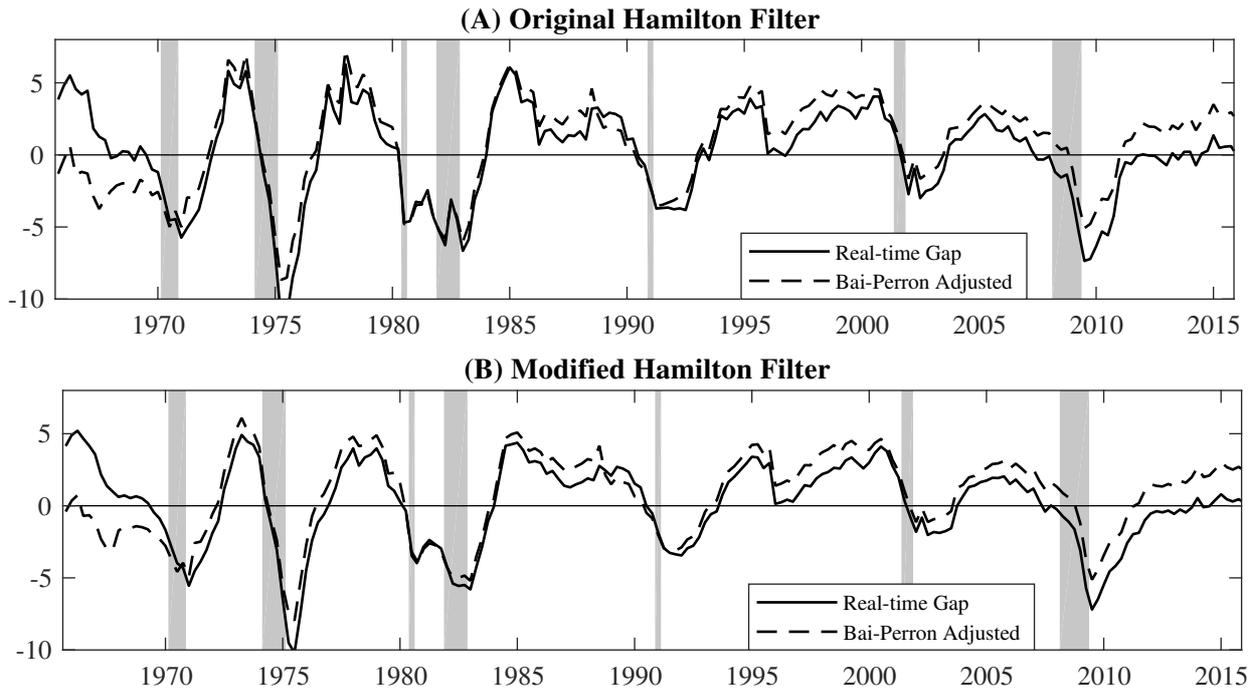
(A) Original Hamilton Filter

(B) Modified Hamilton Filter

Figure 6: Real-time and Bai-Perron Break-Point Adjusted Gaps

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 is much simpler and more intuitive. The standard Beveridge-Nelson 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 Beveridge-Nelson 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.

The Beveridge-Nelson (BN) output gap proposed by Kamber et al. (2018) and Hamilton type output gaps are highly correlated (correlation coefficients around 0.9), both ex post and in real time. Hence, it is not surprising that the different output gap evaluation exercises yield overall similar results for the real-time BN output gap as for the real-time Hamilton filtered output gaps.[17] The only difference is in the output growth forecasting exercise. While the modified Hamilton filter showed significant gains over the HP filtered output gap in forecasting accuracy for the longer horizons when excluding the Great Recession from the sample, the difference between the BN and HP filtered output gaps is insignificant.

6 Conclusions

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 Hamilton filter and is similarly easy to compute, but has a much better coverage of typical business cycle frequencies and yields a smooth estimated trend, while the original approach does not. While the original approach is very useful for detrending, the modified version allows a meaningful economic interpretation of the

---

[17]We follow the specification in Kamber et al. (2018) and choose AR(12) and $\delta = 0.24$ for the noise-to-signal ratio.

cyclical component as an output gap and of the trend as potential GDP. This is particular important, because the 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 gap based on the modified Hamilton filter shows a much higher correlation with ex post assessments of output gaps from important policy institutions. Hence, the method yields a real-time output gap that captures business cycle dynamics that are deemed important from a practitioner's perspective.

References

Ahn, H. J. and J. D. Hamilton (2016). Heterogeneity and unemployment dynamics. NBER Working Paper 22451.

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.

Bai, J. and P. Perron (1998). Estimating and testing linear models with multiple structural changes. *Econometrica 66*, 47–78.

Bai, J. and P. Perron (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics 18*(1), 1–22.

Baxter, M. (1991). Business cycles, stylized facts, and the exchange rate regime: Evidence from the United States. *Journal of International Money and Finance 10*(1), 71–88.

Baxter, M. and A. C. Stockman (1989). Business cycles and exchange rate regime: Some international evidence. *Journal of Monetary Economics 23*(3), 377–400.

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).

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

Bordo, M. D. and P. L. Siklos (2017). Central banks: Evolution and innovation in historical perspective. NBER Working Paper 23847.

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.

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.

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.

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 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.

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.

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.

Hamilton, J. D. (2018a). Measuring global economic activity. Working Paper, University of California at San Diego.

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

Harvey, A. C. and A. Jaeger (1993). Detrending, stylized facts and the business cycle. *Journal of Applied Econometrics 8*(3), 231–247.

Jönsson, K. (2019). Real-time US GDP gap properties using Hamiltons regression-based filter. *Empirical Economics*, forthcoming.

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.

Kilian, L. and H. Lütkepohl (2017). *Structural vector autoregressive analysis*. Cambridge University Press.

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.

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.

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.

Richter, B., M. Schularick, and I. Shim (2018). The costs of macroprudential policy. NBER Working Paper 24989.

Richter, B., M. Schularick, and P. Wachtel (2017). When to lean against the wind. CEPR Discussion Paper 12188.

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 Hamiltons 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.

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 Zandweghe, W. (2017). The changing cyclicality of labor force participation. *Economic Review Q III*, 5–34.

Wolters, M. H. (2018). How the baby boomers' retirement wave distorts model-based output gap estimates. *Journal of Applied Econometrics 33*(5), 680–689.

IMFS WORKING PAPER SERIES

Recent Issues

132 / 2019	Josefine Quast Maik H. Wolters	Reliable Real-time Output Gap Estimates Based on a Modified Hamilton Filter
131 / 2019	Philipp Lieberknecht Volker Wieland	On the Macroeconomic and Fiscal Effects of the Tax Cuts and Jobs Act
130 / 2019	Eduard Hofert	Regulating Virtual Currencies
129 / 2018	Olga Goldfayn-Frank Johannes Wohlfart	How Do Consumers Adapt to a New Environment in their Economic Forecasting? Evidence from the German Reunification
128 / 2018	Christopher Roth Johannes Wohlfart	How Do Expectations About the Macroeconomy Affect Personal Expectations and Behavior?
127 / 2018	Michael Haliassos Thomas Jansson Yigitcan Karabulut	Financial Literacy Externalities
126 / 2018	Felix Strobel	The Government Spending Multiplier, Fiscal Stress and the Zero Lower Bound
125 / 2018	Alexander Meyer-Gohde Daniel Neuhoff	Generalized Exogenous Processes in DSGE: A Bayesian Approach
124 / 2018	Athanasios Orphanides	The Boundaries of Central Bank Independence: Lessons from Unconventional Times
123 / 2018	Karl-Heinz Tödter Gerhard Ziebarth	Zinsen, Effektivpreise und Lebenskosten – Ein Beitrag zur Konstruktion eines intertemporalen Preisindex
122 / 2018	Helmut Siekmann	Legal Tender in the Euro Area
121 / 2018	Maik H. Wolters	How the Baby Boomers' Retirement Wave Distorts Model-Based Output Gap Estimates
120 / 2017	Helmut Siekmann	Die Einstandspflicht der Bundesrepublik Deutschland für die Deutsche Bundesbank und die Europäische Zentralbank
119 / 2017	Gregor Boehl	Monetary Policy and Speculative Stock Markets
118 / 2017	Gregor Boehl Thomas Fischer	Can Taxation Predict US Top-Wealth Share Dynamics?
117 / 2017	Tobias H. Tröger	Why MREL Won't Help Much

116 / 2017	Tobias H. Tröger	Too Complex to Work – A Critical Assessment of the Bail-in Tool under the European Bank Recovery and Resolution Regime
115 / 2017	Guenter W. Beck Volker Wieland	How to Normalize Monetary Policy in the Euro Area
114 / 2017	Michael Binder Jorge Quintana Philipp Lieberknecht Volker Wieland	Model Uncertainty in Macroeconomics: On the Implications of Financial Frictions
113 / 2017	Mewael F. Tesfaselassie Maik Wolters	The Impact of Growth on Unemployment in a Low vs. a High Inflation Environment
112 / 2017	Gerhard Rösl Franz Seitz Karl-Heinz Tödter	Doing away with cash? The welfare costs of abolishing cash
111 / 2017	Jinhyuk Yoo	Capital Injection to Banks versus Debt Relief to Households
110 / 2017	Robert C. M. Beyer Volker Wieland	Instability, imprecision and inconsistent use of equilibrium real interest rate estimates
109 / 2016	Helmut Siekmann	Replacing or Supplementing the Euro in Member States whose Currency is the Euro
108 / 2016	Helmut Siekmann	Restricting the Use of Cash in the European Monetary Union
107 / 2016	Volker Wieland Elena Afanasyeva Meguy Kuete Jinhyuk Yoo	New Methods for Macro-Financial Model Comparison and Policy Analysis
106 / 2016	Helmut Siekmann	Konstruktionsfehler bei der Einlagensicherung auf EU-Ebene
105 / 2016	Athanasios Orphanides	Fiscal Implications of Central Bank Balance Sheet Policies
104 / 2016	Helmut Siekmann	Preis- und Finanzstabilität: der Primat der Politik, der rechtliche Rahmen und das „ökonomische Gesetz"
103 / 2016	John B. Taylor Volker Wieland	Finding the Equilibrium Real Interest Rate in a Fog of Policy Deviations
102 / 2016	Tilman Bletzinger Volker Wieland	Forward guidance and "lower for longer": The case of the ECB