MAGKS

Joint Discussion Paper
Series in Economics

by the Universities of

Aachen · Gießen · Göttingen
Kassel · Marburg · Siegen

ISSN 1867-3678

No. 21-2024

Lars-H. R. Siemers

On the Hamilton-HP Filter Controversy: Evidence from German Business Cycles

Coordination: Bernd Hayo • Philipps-University Marburg
School of Business and Economics • Universitätsstraße 24, D-35032 Marburg
Tel: +49-6421-2823091, Fax: +49-6421-2823088, e-mail: hayo@wiwi.uni-marburg.de

On the Hamilton-HP Filter Controversy: Evidence from German Business Cycles

Lars-H. R. Siemers[†]

This version: 12 December, 2024

Abstract

James Hamilton put doubt on the quality of the HP filter estimates, and proposed an alternative regression approach to decompose trend and cycle of time series (H filter). We investigate the new H filter in detail and compare it to the HP filter. We apply both to German GDP time series. We find that in times of huge shocks the regular H filter produces unreliable trends. Its Quast-Wolters modification (QWH filter), in contrast, does not suffer from this issue. Checking expert benchmark congruency, we find that this modification outperforms all parameter constellations of the standard H filter. With a benchmark-specific adequate choice of the smoothing factor, in turn, the HP filter outperforms the H filter. The H filter, however, outperforms the HP filter with regard to correlation with expert benchmark recession dating. And the H filter uniquely outperforms the HP filter in capturing the gap-inflation link, an issue especially important for central banks. Overall, our results suggest using the QWH filter among the H filter options, and a smoothing factor of 38 for the HP filter.

Keywords: Hamilton filter • HP filter • expert-benchmark congruency • gap-inflation link • spectral analysis

JEL codes: C18; C22; E31; E32; H60

[†] University of Siegen, Department of Economics, Unteres Schloß 3, 57072 Siegen, Germany.
Email: lars.siemers@uni-siegen.de

1 Introduction

The valuation of the position within a business cycle is crucial in many economic contexts. The Keynesian idea of stabilising the economy via counter-cyclical economic policy, for instance, requires knowledge of whether the economy produces above or below its potential, typically captured by positive or negative output gaps. Misjudgement regularly causes harmful cyclical interventions.[1] The challenge is that potential output is unobservable and has to be estimated. There are many heterogeneous methods developed from different perspectives on the problem, consequently estimating different levels of trend or potential output by nature. If a researcher simply wants to de-trend a variable for econometric reasons, for instance to prevent spurious regression, this may not be of too much concern. But if the estimates are used to estimate gaps which are applied for evaluating and planning the public budget, a meaningful annual estimate is important, for it can have crucial consequences for economic policy decisions. Based on Keynes' idea of deficit spending, debt brakes and regulations of public debt have been established in public finance, for, in practice, politics regularly increased expenditures in downturns without decreasing them in upswings – with the consequence of accumulating public debt from business cycle to business cycle (for Germany, see e.g. GCEE 2007). In this context, the output gap of a particular year determines the structural financial position of a budget, mainly measured as cyclically-adjusted budget balance (CAB), for instance, in the context of the medium-term budget objectives (MTOs) in the European Union (EU), or the German federal (as well as many states') debt regulation.

Given tense cyclically-adjusted budget positions require tough consolidations, in democracy the process of how such an evaluation is determined must be transparent, traceable, and, at least for economists, reproducible in a relatively easy way. Therefore, there is always a trade-off between precision (requiring complicated methods and high input of resources) and sufficient plainness. The Swiss government, for instance, therefore decided to use a simple, modified Hodrick-Prescott (HP) filter (Hodrick and Prescott 1981, 1997) when they introduced the Swiss debt brake (Bruchez 2003). In the EU debt regulation, in contrast, the more complex EU production function approach (EUPF) is chosen. In the context of the European debt crisis in the aftermath of the Great Financial Crisis, EU member states (with the exception of UK and Czech Republic) decided that each member has to implement a mandatory balanced-budget rule aiming at annual structural balance, so a cyclically-adjusted version of the budget (ECB 2012, Treaty on Stability, Coordination and

[1] Galí and Perotti (2003) and Lane (2003) discuss cyclical properties of fiscal policy in OECD countries. Mohr (2001), for instance, cites pro-cyclical fiscal policy in Germany in the 1990s. In developing economies, pro-cyclical fiscal policy is widespread (Alesina et al. 2008; Gavin and Perotti 1997; Kaminski et al. 2004; Talvi and Vegh 2005).

Governance in the Economic and Monetary Union) was necessary. The federal German government chose the EUPF especially for reasons of consistency with EU regulation; but many German states, in contrast, introduced simpler regulations. Following the idea of the plainness requirement, many EU members may also follow the strategy of using simpler methods such as the HP filter, as the Swiss government did. While drawbacks of the HP filter are well-known, at least some of these have been healed by some modifications, which means it is still used intensively. But Hamilton (2018) again put conceivable doubt on the reliability of the HP filter, and started a Hamilton-HP filter controversy. We scrutinise Hamilton's suggestion not to use the HP filter to estimate GDP trend and output gap, but instead to use a simple autoregression to estimate the cyclical and trend component of GDP development (H filter). We compare several estimates of trend and cyclical component via H and HP filters, rooted in different filter parameters used in practice or determined by different perspectives from theory. Since potential output is unobservable even ex post, we cannot compare the estimates with true ex post values as is possible in forecast issues. We try to overcome this problem by comparing the results of these simple filters with those of more complex methods, which do not only rely on univariate GDP development but on a variety of relevant macro variables and theory – and are provided by institutions with a lot of reputation and expertise in this task. This allows the deduction of the reliability of the simple methods vis-à-vis those more complex benchmarks, but especially the comparison of the competing H and HP filters. We contribute several new insights to the Hamilton-HP controversy.

2 Relation to the Literature

Our research is especially related to two strands of literature: the business-cycle and filter literature, as well as the general literature on how to define "natural," or potential output.

2.1 Business-Cycle Literature

In the tradition of Burns and Mitchell (Mitchell 1913; Burns and Mitchell 1936, 1946) and, for instance, Harding and Pagan (2002), we estimate potential output, cyclical component as well as output gaps to identify cycles based on patterns of aggregate economic activity. Burns and Mitchell (1946) defined oscillations in business data with recurring periods of between 1.5 and 8 years as business-cycle fluctuations – this definition became the commonly applied one. That is, business cycles are only those that occur at this range of duration. Each method separating the trend (or structural) and cyclical components of an economic time-series ought to capture all fluctuations outside this range by its trend component, and all others within the range by its cyclical component. Focusing on the H versus HP filter controversy, we restrict ourselves to those univariate approaches.

2.1.1 The HP Filter Literature

Regarding the HP filter, it is well known that it suffers from a (start- as well as) end-point bias.[2] Kaiser and Maravall (1999) hence suggest including proper forecasts and backcasts, and Bruchez (2003) modifies the weights for the final three years to reduce the end-point bias. There is particular controversy about the correct value of the smoothing parameter, labelled λ. Hodrick and Prescott (1997) state the assumption that the development of cyclical component and economic growth can be described as two independent random processes that have a certain relationship in size; based on these premises, they show that their HP filter is an optimal filter if $\sqrt{\lambda}$ equals the assumed relation in sizes. They assume that a cyclical component of 5% is "moderately large", in the order of ⅛ of a 1% rate of economic growth in a quarter. Consequently, with quarterly data, they suggest $\lambda = \left[{}^{5}/_{(1/8)}\right]^2 = 1600$. For annual data, in turn, they suggest a value of 100. However, the premises the inventors stated did not reach a consensus at all, and the factor for annual data, though used widespread in practice (e.g. Backus and Kehoe 1992),[3] turned out to be inconsistent with the premises of Hodrick and Precott themselves. Based on Hodrick and Prescott's (1997) value for quarterly data (1600), Ravn and Uhlig (2002) analytically deduce that the optimal filter requires a value of only ${}^{1600}/_{256}$ = 6.25 for annual data. However, Reeves et al. (2000) also cast doubt on the plausibility of Hodrick and Prescott's premises and the resulting quarterly value of 1600 in practice, for it is only shown to be applicable for white-noise cyclical components, and the empirical sample estimate of the parameter deviates from their prior belief. Alternatively, if applying spectral analysis to macroeconomic time series (cf. e.g. Granger 1966; Iacobucci 2005), the researcher has to choose a length for the reference cycle. In line with Mitchell (1913) as well as Burns and Mitchell (1936, 1946) a typical business-cycle length of eight years is assumed. Kaiser and Maravall (1999) establish the theoretical relationship of reference cycle length and smoothing parameter λ. They confirm the value of 1600 for quarterly data and a reference length of eight years, but deduce, in line with Ravn and Uhlig (2002), a value of $\lambda = 7$ for annual data, instead of 100, which is found to be consistent with a longer reference cycle of 16 years. Baxter and King (1999) show, in contrast, that for a reference cycle length of eight years an HP smoothing value of $\lambda = 10$ is consistent with their high-pass filter. In spectral analysis (e.g., Polock 2000), it is stated that filter parameters ought

[2] In early years, the HP filter was criticised for generating artificial cycles (Cogley and Nason 1995; Harvey and Jaeger 1993; King and Rebelo 1993). These distortions, however, are unavoidable and a feature of even any optimal filter (Ehlgen 1998).

[3] Cooley and Ohanian (1991) as well as Correia et al. (1992) suggest a higher value of 400 (Ravn and Uhlig 2002).

to be set as functions of the cut-off frequency, and that at the cut-off frequency the parameter(s) should produce a gain function value of ½. When this rule is applied, the required smoothing parameter λ is 677.13 for quarterly data when we, as usual, assume a cut-off frequency rooted in cycles of a length of eight years; this also challenges the quarterly data parameter of 1600. Yet another spectral theory study of Pedersen (2001) deduces optimal values for quarterly data in a range of 1007 to 1269, depending on assumptions, but argues that 1600 is close to the optimal value for cycles shorter than nine years; the annual counterpart range is 3.73 to 5.03. Given the conclusion regarding 1600, we conclude that the Ravn-Uhlig counterpart of 6.25 is at least close to optimal.

Mohr (2001) emphasises, however, that the correct reference cycle length (directly related to the cut-off frequency[4]) depends on the specific issue analysed. While eight years might be a proper choice in the context of pure business cycle analyses, for other issues shorter or longer cycle length can be more appropriate. Aiming at structural budget positions, the goal is for there to be a balance of positive and negative business cycle effects. In the EU treaties, it is stated that these are measured via output gaps, that potential special year effects ought to be accounted for, and that the balance ought to be established over an unspecific "medium term." A structural budget position is directly linked to the structural or trend component of (price-adjusted) GDP. Cycle-driven deficits ought to be balanced by cycle-based surpluses within a manageable period of time; this would ask for a comparatively short reference cycle length, and thus a relatively low HP smoothing factor (Mohr 2001: 25). This, however, involves the risk of a pro-cyclically fiscal rule, because automatic stabilisers may not function effectively in such a setting. Hence, there is a trade-off between effective automatic stabilisers and structural balanced budget in a reasonable period of time. Kaiser and Maravall (1999) determine the value of λ such that for the predetermined frequency of the chosen reference cycle length it maximises the estimated sample spectral. As a result, the variance of the cyclical component is especially determined by frequencies in the range predetermined by the reference length. Mohr (2001) argues that this method implicitly assumes an equality of theoretical and actually estimated spectral – which is known to be only true for the fictional case of infinitely long time series. Therefore, Mohr (2001) chooses the smoothing factor by maximising the actually estimated spectral (instead of the theoretical) via λ. For the German nominal GDP he derives $\lambda = 20$: the maximum is reached for a frequency of 0.76, roughly in line with the targeted eight years

[4] Spectral analysis is rooted in periodic movements and wave theory, where an object or wave performs a periodic movement and reaches its original state again after T periods. The periodic ups and downs of waves are used as a model for business cycles. The frequency of waves is in an inverted relationship to duration T and can be described with trigonometric functions, which are determined by frequency $2 \cdot \pi / T$. The duration, in turn, is stated in the unit of time measurement. Assuming a reference cycle length of eight years, T is 8 (annual data) or 32 (quarterly data) (cf., e.g., Granger 1966; Iacobucci 2005).

reference length.[5] Within the harmonised method of the European System of Central Banks (ESCB), which considers the aggregate of the EU member states as a whole, the chosen λ for price-adjusted time series is 30 (Bouthevillain et al. 2001). In contrast, Swiss law determined the modified HP filter with $\lambda = 100$ for the Swiss debt brake.

An often ignored issue is the time horizon the estimates lean on. In most cases authors simply use the time span available, though this choice may also affect the result. An exception to this rule is the Swiss modified HP filter, where Swiss law determined a time horizon of 30 years, i.e. a moving window. Overall, therefore, the HP filter approach involves a high degree of haziness: there is a high degree of freedom in choosing the smoothing factor, the time frame included, and whether and how many future periods to include as forecasted values. We hence focus on a comparison of common practice as well as theoretical requirements from the different perspectives of statistics and spectral theory, and analyse the HP filter for many different smoothing parameters in order to contribute to this literature. Additionally, we add forecasts as well as backcasts to address the start- and end-point bias issue. Those HP filter results are then used in a comparison to the results of the H filter.

2.1.2 The H Filter Literature

Hamilton (2018) puts considerable doubt on the reliability of the HP filter. He argues that the common practice of setting the smoothing parameter for the HP filter cannot obtain optimal decompositions into trend and cycle – even worse, the HP filter is generally not based on the respective true data-generating processes and hence generates spurious results for the statistical trend. He demonstrates that the HP filter results for the cycle can be shown to be identical to assuming white noise cyclical components in a state-space model – that is, that they would be random and exhibit no discernible pattern. His conclusion is that the estimated cyclical pattern does not reflect any true dynamics at all but is an entirely spurious feature of the applied HP filter.

Hamilton offers an alternative approach to deducing reliable estimates of the cyclical component. He demonstrates that a comparably simple regression of a variable's value vector in a period $t+h$, labelled y_{t+h}, on a constant and the p most recent realization vectors of that variable as of date t (so lags y_t, y_{t-1}, …, up to y_{t-p}) allows the consistent estimation of the cyclical component reasonably well for a wide range of non-stationary data-generating processes. The choice of h and p depends on the kind of data used (monthly, quarterly, or annual observations) and the kind of cycles interested in

[5] For $\lambda = 100$, reference cycle lasts about 13 years (frequency 0.48), and with $\lambda = 8$, e.g., this cycle would last only about six years.

(business cycles or longer cycles such as financial crises cycles in the context of early warning indicators), that is, again on the length of the reference cycle.

However, the Hamilton approach may also suffer from weaknesses. Following Quast and Wolters (2022) as well as Schüler (2018) it amplifies long, medium-term cycles, while muting short cycles of two years or below in length in the context of business-cycle analysis. Schüler (2018) finds that it also generates artificial cycles, suffers from small-sample bias, smooths structural breaks – as the HP filter does – and the choice of the lag structure via parameters h and p is similarly ad hoc as the choice of λ. Based on Monte-Carlo simulations, Jönsson (2020a) demonstrates that even if the data are generated without any cyclical dynamics, the Hamilton regression produces cyclical component dynamics. That is, the Hamilton filter modifies the original cyclical structure of the data as well (as the HP filter). In contrast to Schüler (2018), Quast and Wolters (2022) find that the trend estimates are not smooth, and thus hardly can be interpreted as a measure of potential output, which is found to be rooted in remaining high-frequency noise, which is not associated with business fluctuations. We find that this is an issue in the German context in times of great crisis. Overall, the Hamilton regression approach is found to be weaker for business cycle analysis due to overemphasising cycles longer than typical business cycles.[6] In line, Hall and Thomson (2021) did not find any advantage in using the Hamilton regression in comparison to HP or Baxter-King filters for New Zealand macro data. Schüler as well as Quast and Wolters, however, reveal that the H filter suffers much less from bias at the tails of the time series (cf. also Jönsson 2020b), which is a highly important advantage for economic policy in practice, where data revision often changes recent gap estimates significantly – a major challenge for budget planning, especially when being forced to behave in line with a debt brake, as in Germany. To benefit from this advantage without suffering from the above mentioned problem of still unreliable cycle-trend decomposition, Quast and Wolters (2022) suggest a modification of the H filter. They recommend checking the properties of the H filter and modifying it if needed. We provide evidence that their modification is very helpful for trend-cycle decomposition of German GDP.

In our analysis, we consider huge GDP shocks that provide a very similar challenge for the filters as structural breaks. In such situations we find, in contrast to Schüler (2018), evidence that the Hamilton regression may not smooth structural breaks, as the HP filters do; this is in line with Quast and Wolters (2022), who find that Hamilton gaps are rather robust in checking for structural breaks.

[6] Schüler (2018) states that the h=5 Hamilton regression (annual data) outperforms the Basel-III credit-to-GDP gap (based on HP filtering) in identifying imbalances in the financial markets, where longer cycles are to be considered. In contrast, Drehmann and Yetman (2018) find that the Hamilton filter can be outperformed by the HP filter referring the credit-to-GDP gap in predicting crises in financial markets.

We find, however, that the H filter turns out to suffer from unreliable properties in times of huge shock for the German GDP – and thus probably also in the context of structural breaks. We show that this problem can be healed by the modification suggested by Quast and Wolters (2022). In line with Hall and Thomson (2021), we also find that the Hamilton regression involves, in comparison, very volatile output gaps.

In the H versus HP filter controversy it is also interesting to note that in the COVID-19 pandemic the Swiss mHP filter produced instable trend values that shrank relatively sharply with a time lag, with problematic pro-cyclical estimates of the business-cycle factor of the Swiss debt brake in 2022 (Schmassmann and Wiedmer 2022). The conclusion of Swiss authorities was that smoothing methods like the (m)HP filter fail in times of huge crisis. We add that the H filter also fails in such times. The Swiss system now substituted the mHP filter with the EU production function approach (EUPF). Therefore, we also use the official EU approach (see below) as a benchmark method in evaluating both the H and HP filter. Moreover, we compare several different results of the methods on the basis of some additional benchmark indicators, which, to the best of our knowledge, has not been done so far. While Quast and Wolters (2022), Schüler (2018), and Hall and Thomson (2021) alike compare Hamilton's regression only with the HP filter with the standard smoothing parameter of 100, we include further values suggested in the literature, and we not only investigate the regular H filter but additional versions of it. Using two different expert benchmarks (like the EU method) and an inflation indicator, we find that the HP filter with a parameter of 100 may not be the best choice for the smoothing factor. Simulations show that, depending on the targeted benchmark, lower or higher values are optimal. We show that the conclusion on whether the H or HP filter is preferred is highly conditional on the smoothing factor.

2.1.3 The Benchmark Methods: Production Function Approach

Within the European Union (EU) system the trend-cycle decomposition is especially important in evaluating the general economic situation and the public finances of the member states. GDP potential and output gaps are estimated within the EU by a production function method (EUPF, Havik et al. 2014). This method is applied by the European Commission for all member states. The EU method is also exactly required for all national stability reports of the member states. The German federal as well as many regional state governments (Bundeslaender) apply it in the context of the constitutional German debt brake. Within the context of the Swiss debt brake, the government used, until very recently, the Swiss modified HP filter (mHP, Bruchez 2003), but it has now been substituted with a national production function approach exactly in line with this EU method. Other versions of a production function approach are internationally used by the OECD and IMF, by the

Congressional Budget Office (CBO 1997) or other US organisations (Stiroh 1998), or by other OECD countries (Gern et al. 2008). Especially for Germany, it is applied by the German Central Bank (Bundesbank 1973, 1981, 1995), the German research institutes in the context of the biannual joint report for the German federal government (JEF 2024: 60–65), or the German Council of Economic Experts (Breuer et al. 2022). While the HP and H filters represent purely data-driven statistical methods, the production-function approach is more theory-driven and is rooted in a supply-side perspective. Based on macroeconomic theory, it estimates the unobservable theoretical variable of potential or natural output (see also below) by assuming total output can be caught by the theoretical concept of a macroeconomic production function, with the inputs of capital, labour, and technical progress (cf., e.g., Denis et al. 2002; Denis et al. 2006; D'Auria et al. 2010; Havik et al. 2014). In doing so, the "normal" level of capital, labour, and total factor productivity is again estimated with statistical filter methods.[7] The EU approach explicitly assumes that the output gap is zero in the final period forecast, which is the year five years in advance. Mohr (2001) argues that an advantage of the HP filter vis-à-vis production function approaches is that HP filters, in contrast to the production function method, have the property that positive and negative gaps are balanced in a relatively short time segment. He considers this feature to be especially important for using gaps to cyclically adjust the fiscal budget position.

2.1.4 Alternative Decomposition Methods

Generally, potential output can be estimated using (i) smoothing and de-trending methods such as the HP filter, the split-time trend method, the log-linear time trend method, or the moving-average method, (ii) econometric estimates such as the Okun-law regression (Okun 1962; CBO 2004), the production function method or Hamilton's regression (H filter), or (iii) survey data on capacity utilisation (see, e.g., Fisher et al. 1997). Although we focus on univariate filter methods and restrict ourselves to the H-HP controversy, we briefly want to embed our study in the broader literature and discuss alternative methods. In spectral analysis, ideal filters involve weighting frequencies depending on cut-off frequencies. There is a range of frequencies that is considered to cover the cycles looked for. These frequencies ought to be weighted by unity while other frequencies should be weighted by zero. Within the range, the cycle frequencies are assigned to the cyclical component, the others to the trend component. Such ideal characteristics are, however, only possible in infinitely long time-series. In practice, with a finite number of observations over time, filters try to

[7] The used potential of capital input is the stock of capital, measured by the perpetual inventory method, the potential of labour input involves application of an HP filter and estimating NAWRU, and the trend of the efficiency of the factor inputs (total factor productivity) involves a bivariate Kalman filter model (since 2010, before also the HP filter was applied).

fulfil the ideal filter properties as precisely as possible. There is a low and high cut-off frequency, identifying the minimum and maximum length of typical business cycles; for business cycles the standard range is 1.5 to 8 years. Low-pass filters only account for the low cut-off frequency, trying to weight only below it by zero; band-pass filters, in contrast, try to account for both thresholds. There are many methods to estimate potential output / trend and output gaps, e.g., the BK filter (Baxter and King 1999), the Beveridge and Nelson (1981) decomposition, the CF filter (Christiano and Fitzgerald 2003), or univariate time-series models of unobserved components (Harvey 1989). While the Beveridge-Nelson decomposition is criticised for producing implausible implications (Mc Morrow and Roeger 2001), the BK filter and the models of unobserved components, being more complex and easier to manipulate, generate quite similar results to the HP filter (Mc Morrow and Roeger 2001; Mohr 2001). The BK and CF filters are both band-pass filters, but its properties are very much the same (Pedersen 2001). In fact, the HP filter is equivalent to a high-pass filter and comparable with such band-pass filters (King and Rebelo 1993). We refrain from adding the more complicated multivariate filter methods, such as the Kalman filter (e.g., Kálmán 1960; Kálmán and Bucy 1961; Stratonovich 1959), which takes a broader economic context into account. Similarly to the production function method these perform potentially better than the simple univariate methods investigated, and hence also qualify to function as a benchmark. In contrast to the production function methods, however, there are no official estimates published by reputable expert institutions. Therefore, our own estimations applying these techniques would be less convincing benchmarks.

2.2 Literature on Natural Output and Other Related Issues

Theoretically the decomposition of trend and cycle is a decomposition of economic growth and the business cycle. In a static macro model the structural component is the unobservable potential output, i.e. the situation when there is – given all market frictions and constraints – "full" employment of all production factors. It represents the long-term equilibrium, labelled "natural" output, to which economic activity (GDP) returns to after fluctuations (business cycle).[8] That is, the stated qualification of existing frictions, constraints, and asymmetric information – causing involuntary and search unemployment etc. – causes an ordinary economic-capacity use of the economy that is deployed. This level of output is supply-determined and often modelled, for instance in the (new) neoclassical synthesis, as if it is independent of the price level or inflation.

[8] Other long-term equilibrium levels, such as unemployment or interest rate, are also labelled as being at its natural rate. This idea of a natural long-term equilibrium goes at least back to Wicksell (1898, 1907) who discussed economic fluctuations rooting in a temporary deviation of money interest rate and the ordinary natural rate of interest (marginal productivity of capital).

Hence, it is related to a situation where there is no inflation pressure. This aspect is captured within the EUPF by estimating the non-accelerating wage rate of unemployment (NAWRU), that is, the level of wages that does not generate any effect on inflation (wage-inflation spiral), when estimating the "natural" level of labour input. Hence, the ordinary level of resource use and inflation stability are the two core indicators for potential output. In this perspective, the dynamics of potential output is rooted in growth which is driven by investment, including innovations, as well as effective labour resources – and considered to develop smoothly. Woodford (2009) challenges this perspective, and emphasises that modern empirical DSGE models, such as Smets and Wouters (2007), involve short-term deviations from the "natural" rate of GDP where these business-cycle movements typically also shift the natural rate of economic activity, with the consequence that the potential output dynamics does not at all necessarily involve a smooth trend. As Quast and Wolters (2022) argue, these volatile natural output estimates may in fact be rooted in mark-up shocks tightly linked to measurement error (Chari et al. 2009; Justiniano and Primiceri 2008). Recent results addressing this issue indeed find much more smoothing development of potential output (de Loecker et al. 2020). Overall, this literature suggests that potential output is related to an ordinary, normal, and therefore natural level of capacity use and stable inflation. The ideas of a macroeconomic production function and NAWRU, followed by the EUPF method, are exactly in line with this insight. Therefore, we use institutional production function estimates as benchmarks to evaluate the quality of the analysed filter methods. Moreover, we investigate the link between estimated output gaps and the change in inflation as a further criterion to evaluate the quality of analysed H and HP filter methods.

Finally, we also contribute to the literature of business-cycle dating for Germany. Schirwitz (2009) presents a comprehensive German classical business cycle dating from 1970 to 2006, and Breuer et al. (2022) recently provided a business cycle dating back to 1950. In contrast to Breuer et al. (2022), we focus on the much simpler univariate GDP-based approaches of H and HP filter methods used in practice on the basis of annual data – and apply a new five-requirement dating rule. Comparing our results with those of Breuer et al. (2022), based on a much more sophisticated multidimensional set of indicators, allows us to learn about the loss of quality when applying those simpler techniques. This is an important task, given the trade-off of precision and simplicity required for traceability for the public in a democracy – an issue that led Swiss authorities to prefer the simple HP filter, in modified form. In line with Schirwitz (2009: 290), we believe that the advantages of methods like the H and HP filters are clearly simplicity, higher transparency, long-term uniformity, and (international) comparability vis-á-vis expert committee or complicated institutional approaches.

3 Modus Operandi and Data

3.1 Approach

We provide a comprehensive evaluation of several kinds of H and HP filters over the years 1980-2023, based on several statistical indicators that are benchmark anchored. Along the lines of Quast and Wolters (2022), we use output gap estimations of institutions that invest a lot of manpower, expert knowledge and other resources as benchmarks. These institutions apply a much more sophisticated method than univariate filters do.[9] In Germany, two benchmarks in particular are available:[10] (i) the EUPF estimates by the experts of the Federal Ministries of Finance and Economics; (ii) the expert estimates of the German Council of Economic Experts (GCEE), which is a comparatively independent council of five economics professors, equipped with a staff of 18 further economists, and institutionalised support from the Federal Statistical Office, the Central Bank (Bundesbank), and further institutions. The GCEE also applies a production function method (Breuer and Elstner 2020; Ochsner et al. 2024). All benchmark indicators are rooted in the same general idea: potential output reflects a situation in which an economy deploys its capacity at its "normal" rate. As a widely used theory-driven approach to directly estimating potential output in a structural model, the EUPF and GCEE production function results are used as an important benchmark for comparison with the H and HP filters in practice. The benchmark comparison allows the evaluation of the extent to which these benchmark movements of potential output are captured by the H and HP filters. Our major focus in doing so will be the estimated output gaps.

At first we compare the all-time standard deviation and the all-time mean of absolute gaps. Both are indicators for the dispersion of the gaps of the respective filter. The absolute value mean provides additional information about the mean size of positive or negative gaps of the respective filter. Both statistics are also deduced for the benchmarks for comparison. We then apply a benchmark congruency analysis where we check to what extent the filter gap estimates are in line with those of the benchmarks. The term "in line with" is operationalised by comparing (i) the sign of the filter and benchmark gaps, (ii) Pearson's correlation coefficient of the filter and benchmark gaps, and (iii) the root mean squared error (RMSE), given by

[9] We exclude international institutions, because those suffer from distance and less national expert knowledge. This premise is confirmed by the results of Quast and Wolters (2022) for the US.

[10] We also checked the latest estimates of the experts of the German research institutes (including some Austrian partners) in the context of their Joint Economic Forecast (project group Gemeinschaftsdiagnose, spring forecast JEF 2024). These data are only internally available. They also cover EUPF estimates, but additionally the modified version MODEM. It turns out that the MODEM and official EUPF estimates are so highly correlated that including it was not adding value.

$$RMSE^{filter}(benchmark) = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(gap_t^{benchmark} - gap_t^{filter}\right)^2}, \tag{1}$$

where t is the respective observed year and T the number of observed years. Additionally, we check the adequacy of the filter gaps to predict the benchmark gap by the following regression (unbiasedness test):

$$gap_t^{benchmark} = \alpha + \beta \cdot gap_t^{filter} + \varepsilon \tag{2}$$

If gap_t^{filter} is a good predictor for the $gap_t^{benchmark}$ we will find $\alpha = 0$ and $\beta = 1$. This is checked via an F-test.

All these quality indicators are standard techniques used to evaluate the quality of forecasts (cf. Fildes and Steckler 2002). But in our context we have to substitute the true realized value, which was estimated in advance, with the envisaged $gap_t^{benchmark}$. The term "error" in the context of RMSE has to be interpreted in this sense. Additionally, theory and the definition of natural output requires that there is no change in inflation when the economy produces at normal capacity (gap of zero). In times of positive gaps, inflation is expected to rise, and vice versa. We check this property for the competing H and HP filters again via comparing sign congruency and Pearson's correlation coefficient of the competing filters with the change in inflation of consumer prices, measured by the deflator of private consumption (first difference of deflator). We also estimate regression model (2) for this indicator. However, in contrast to the test with respect to benchmarks EUPF and GCEE, left-hand and right-hand side variables of equation (2) do not measure the same indicator "output gap" – we regress change in inflation on the respective output gap. Therefore, the unbiasedness F-test is inadequate and we instead focus on R^2, root mean squared error (MSE), and the slope coefficient β. The first two indicators inform about how much of the variation of the change in inflation is explained by the respective estimated gap (R^2) and the modified standard deviation of the residuals (root MSE), that is, how close the estimated relationship between gap and change in inflation is. If the expected link between gap and inflation is covered coefficient β got to be positive and statistically significant.

Following the idea of Quast and Wolters (2022), we also investigate the correlation of the gap estimates with the recession dating of a benchmark institution that invests a substantial amount of expert knowledge as well as substantial resources – and is thus not solely based on statistical models such as the investigated filters. We choose the GCEE dating published in 2022 and generate expansion dummies that equal unity in all dated expansion periods and zero in dated recessions. The premise here is that the external expert dating provides the most convincing dating. To evaluate the consistency of the gap estimates we compute Pearson's correlation coefficient of the change of the respective filter gaps with our expansion period dummy. It is expected that in times of expansion the gap rises, while it decreases in recession periods. We then also determine our own dating rule, based on observed output gaps. Following this rule, we deduce the respective turning points (peaks and troughs), and the resulting business-cycle dating for the second benchmark EUPF. Given this EUPF dating, we again generate an expansion dummy and calculate the respective correlation coefficients for this second benchmark. Given we focus on univariate GDP time-series methods, our goal is explicitly not to identify the turning points in the best possible way, which would require the inclusion of additional cycle indicators such as investment, unemployment etc. Hence, we follow a simple but not simplistic rule-based identification for comparison, based on the estimated gaps. We deploy a five-requirement approach: we identify a business-cycle peak if (1) the gap is positive, (2) higher than 1%, (3) the highest since the preceding peak, (4) since the preceding peak there has been a negative gap in between, that is, in absolute terms, at least higher than 1%, too, and (5) it is followed by another trough, that is, a negative gap below –1%. Correspondingly, a trough requires the following: (1) the gap is negative, (2) lower than –1%, (3) is the lowest since the preceding trough, (4) since the preceding trough there has been a positive gap in between that is at least higher than 1%, too, and (5) it is followed by another peak, that is, a positive gap bigger than 1%. Applying this dating rule to the official EUPF gaps, we date recession periods as years running from peak to trough turning points, and expansion periods in the residual years.

Applying our dating rule also to the investigated H and HP filter gaps, we also determine the resulting business cycles in order to compare average cycle lengths. Given average cycle length is not identical to what is typically meant by the length of a filter's reference cycle, we deduce it, as usual, via spectral analysis. We determine each filter's spectral density and cumulative spectral distribution, and investigate to what degree natural frequencies that are in line with typical business cycles are covered. This allows the comparison of the specific filters' range of sensibility for business-cycle frequencies. Finally, we also put a special focus on the filters' crisis resilience, that is, we evaluate how convincingly the estimated output gaps and trends behave in times of huge shocks.

In detail, we investigate four different kinds of H filters in comparison to four different HP filters. The general H filter idea, developed in Hamilton (2018), is the following autoregression model:

$$y_{t+h} = \beta_0 + \beta_1 \cdot y_t + \beta_2 \cdot y_{t-1} + \cdots + \beta_{p+1} \cdot y_{t-p} + \varepsilon_{t+h} \tag{3}$$

Variable of interest y in a period h periods ahead from t ought to be explained by an autoregression on its value in t and, from t on, p additional lags. He states the general rule that h and p has to be integer multiples of the number of observations per year, and argues that within a period of two years macro data such as GDP are still affected by cyclical factors and the timing of recovery processes. For business-cycle analysis, Hamilton (2018: 838) thus recommends, applied to annual data, combining h=2 (distance to lags) with p=1 (number of lags). The regular annual-data H filter regression model is therefore as follows:

$$y_{t+2} = \beta_0 + \beta_1 \cdot y_t + \beta_2 \cdot y_{t-1} + \varepsilon_{t+2}, \quad (4)$$

with y being price-adjusted GDP. Quast and Wolters (2022) criticize that this H filter does not cover typical business-cycle frequencies evenly and that the trend still transports high-frequency noise, so that it cannot be interpreted as potential output. Hence, we also estimate their modified H filter (QWH), and estimate the three general H(h, p) filters H(1,1), H(2,1), and H(3,1). The QWH trend and QWH cycle component is then determined by the average of these three trend and cycle estimates; Quast and Wolters (2022: 157) state that their modification (i) produces a more even coverage of the typical cycle frequencies and (ii) less dependence on the starting point.[11] Hamilton (2018) states that a higher value of h means aiming for longer cycle length. While the H filter implicitly excludes trend changes within the chosen two-year window (h), the QWH modification allows for trend changes within one year, weighted by a relatively high fraction of ⅓.

For the HP filter we use, based on our literature review, four different smoothing factors: (a) 2.914235, applying the rule of Pollock (2000) that filter parameters ought to be chosen such that the value of the gain function of the filter is 0.5 at the cut-off frequency (here, as usual for business cycles, 8 years); (b) the Ravn and Uhlig (2002) value of 6.25 (which implicitly assumes that 1600 is the correct value for quarterly data); (c) the value of 20 deduced especially for Germany by the Bundesbank (Mohr 2001);[12] and (d) the widespread standard annual-data value of 100 (Hodrick and Prescott 1981, 1997). Finally, we determine four optimal smoothing factors λ in simulations. The objective is: (i) maximising Pearson's correlation coefficient (PCC) of the HP gap with the two

[11] In Quast and Wolter (2022) it seems that they suggest to (1) do the H-filter regression, to estimate the, in our application, three forecast errors, (2) calculate the average of the forecast errors, and (3) then to determine the H-trend by the difference of observed GDP and this average error as estimated cycle component. We deviate slightly from this by directly averaging estimated cycle components (error) and the respective trend terms. This may generate a possibly even better smoothing of the average H-filter.

[12] The harmonized value of 30, agreed on within the European System of Central Banks, is very close to 20 and thus producing very similar results, which would not add much insight in comparison.

different benchmark gaps; (ii) minimising the respective RMSE of the HP gap. As stated, the benchmarks are the EUPF and GCEE gaps.

3.2 Data

3.2.1 Data Base of the Study

For our core evaluation analysis, we use the spring 2024 economic forecast of the German federal government (FedGov 2024). The government data include the official estimates of potential output, cycle component, and output gap, required in the context of the fiscal surveillance procedure of the EU. Those values determine the officially estimated cyclically-adjusted budget position within the EU system as well as for the German constitutional debt brake. In both procedures the EUPF method must be applied. The EUPF estimates for German potential GDP, cycle component, and output gap are provided from 1980 until 2023 and prolonged by their forecasted values up to 2028. This is our EUPF benchmark data, and we directly use them without any adjustment in order to compare the results with their H and HP filter counterparts. The data also cover price-adjusted GDP (2015=100) for the years 1980-2023, based on the Federal Statistical Office, and the government's forecast GDP values for the current year and the future years up to 2028, which were used for the government's EUPF estimates. The forecasted numbers are important, for they are required to be included to overcome the end-point problem of the HP filter. Given this problem, in fact, also involves a starting-point problem, we additionally imputed GDP data for the three years before 1980, which, to the best of our knowledge, has not been done so far in the literature. For reasons of autoregression within the H filter, this method also loses the first three years at the beginning, i.e. trend and cycle for 1980 to 1982. To overcome this issue, it is required to impute three additional former years as well (1977-1979). As we also use the Quast-Wolters modification of the H filter, which involves alternative variations of parameters h and p, it is necessary to add GDP data up to 1976. We take these data from exactly the same National Account data that the government's data come from (NA 2024) in order to prevent a lack of comparability. For this reason, we also only include those observations that are really required to estimate trend and cycle starting from 1980, but not additional pre-years. Therefore, our comparisons are not biased by comparing different time windows or by the start-end-point bias of the HP filter, which might have biased the comparisons in former studies. Using exactly these GDP data also for our H and HP filter estimates guarantees the comparability of our results with the EUPF.

The GCEE benchmark data on potential output and output gap are publicly available on the institution's homepage.[13] Both variables are available for the complete years investigated. The GCEE turning point and recession benchmark dating is rooted in Breuer et al. (2022: table 1). For the GCEE data it is important to recognise that the recession dating in Breuer et al. (2022) and the used up-to-date gap estimates are not fully comparable, since the GCEE production function method was adjusted recently.[14] The years investigated, 1980–2023, allow the evaluation of how the filters perform in times of huge shocks. It covers the German reunification on 3 October 1990,[15] which provides a significant positive economic shock, because East Germans were, for the first time, able to consume the complete range of Western goods, and their wealth in East German marks was exchanged for West German DM at a quite beneficial exchange rate. This caused an extraordinary boom, especially for the West German economy. Additionally, there were the Great Financial Crisis (in Germany especially in 2009) and the COVID-19 pandemic with massive lockdowns.

3.2.2 Annual versus Quarterly Data

Business-cycle researchers, rather interested in evaluating the position in the business cycle in progress and recognizing turning points as soon as possible, for good reasons typically identify business cycles on the basis of multiple monthly or quarterly business-cycle indicators, i.e. a system of indicators more sophisticated than even the production function approach (e.g. Breuer et al. 2022; GCEE 2017, Chapt. 3: 133-34; Gehringer and Mayer 2021; RWI 2017).[16] Without dispute, using quarterly or even monthly data allows the identification of peaks and troughs more precisely within a year. GDP values of two consecutive quarters are linked by significant autocorrelation, which qualifies quarterly GDP data to forecast future values via its own former development. This autocorrelation link mostly disappears for two consecutive annual observations (Schröder et al. 2008: 223).[17] In our ex post analysis without forecasting issues, this aspect poses no problems. On the contrary, this is even an advantage since, when using annual data, distorting issues like seasonal and calendar effects as well as autocorrelation are of much less relevance. The choice of data frequency thus depends on the aim of an investigation. As stated by Schirwitz (2009), for ex post

[13] Data homepage: https://www.sachverstaendigenrat-wirtschaft.de/en/publications/data.html (Economic Growth data)
[14] While the former method is published in Breuer and Elstner (2020), the new is still not published but stated as forthcoming working paper on the homepage; a probably fitting description is given in Ochsner et al. (2024).
[15] Statistically it was implemented for the year 1991. The two time-series of West-Germany and reunified Germany were chained via the 1991 annual average. Hence there ought to be no statistical structural break.
[16] Pioneering works in the field of cycle indicators are, e.g., Hamilton (1989) as well as Stock and Watson (1989,1991). A further recent contribution is, e.g., Kyo et al. (2022). Krüger (2021) investigates the performance of leading indicators of the German business cycle, Pažický (2021) the yield curve in Germany and the US.
[17] Consequently, Hamilton (2018) suggests $h = 2$ for annual data.

business-cycle analysis as well as comparison, which is our focus, aggregated GDP data is an adequate candidate, while the evaluation of the current cycle and forecasting requires multivariable approaches and cycle indicators, published monthly or at least quarterly and before GDP.

The focus of our study is not on forecasting issues, but rather on issues based on ex-post data, which allows the historical determination of the cycle position of specific years as a whole and, for instance, the calculation of cyclically adjusted values of macroeconomic variables such as public budget balances. For the EU MTO and German debt brake the annual budget numbers are combined with annual output gaps estimated via EUPF. That is, in public finance as well as other contexts, annual business-cycle data are required. Another application in the context of the EU and German debt regulation is a suggestion made by Gebhard and Siemers (2020). They criticise the official EU approach for using real-time interest rates. When there are abnormal levels of interest rates, e.g. in low-interest-rate periods, EUPF-based estimates of the structural budget position are, as a consequence, biased by unusual interest payments, in addition to business-cycle distortions. Hence, they propose an extension of the rules by substituting abnormal interest rates with non-distorted cyclically-adjusted annual interest rates. This requires the use of average annual interest rates over former business cycles – a task only possible to accomplish knowing the former completed business cycles on an annual basis. Furthermore, our evaluation approach involves a comparison with the benchmarks EUPF and GCEE. Those are available only on an annual basis – research institutions, governments, and central banks typically publish potential GDP estimates on an annual basis.[18,19,20]

4 Results

Figure 1 shows that our two benchmarks follow a similar cycle pattern. The gap estimates are very close to each other in the 1990s, in the second half of the 2000s, and since 2020. The GCEE gaps, however, produce considerably lower gaps in the 1980s, and also in the 2010s, though to a lower extent; in 1999 and the first half of the 2000s, in contrast, the GCEE gaps are estimated to be higher than those of its EUPF counterparts. Figure 2 shows GDP and estimated trend development for the

[18] For the US investigation in Quast and Wolters (2022) there were only quarterly data from the central bank (Fed) and the Congressional Budget Office (CBO) available. The other used international benchmarks in their study were all published at annual basis.

[19] Gebhardt and Siemers (2020) find that using ex post annual gap data and identifying peaks and troughs (and thus business cycles) simply via gap peaks and troughs – so a less sophisticated dating rule than our five-requirement approach – produces similar results referring the years of turning points and length of business cycles, compared to more sophisticated approaches based on monthly or quarterly data.

[20] Being historically interested on annual basis whether a year suffered from economic problems or not, monthly or quarterly downturns that last only for a short period of time do not necessarily signal a year of economic, fiscal, or social pressure if it is smoothed by much better business activity in the rest of the year. Thus troughs and peaks based on quarterly data can be misleading in such an investigation.

three different H filters H(1,1), H(2,1), and H(3,1). It shows that the H filter, given it is based on autoregression, is producing phase-delayed trends of GDP development by construction, and parameter *h* is determining the degree of delay. The H(1,1) trend follows GDP relatively closely and delays GDP ups and downs by one year, the regular H(2,1) trend by two years, and the H(3,1) trend by three years. The delayed ups and downs, in turn, become smaller with increasing parameter *h*, that is, the trend is smoothed to a higher degree. Therefore, the smoothing property is rooted in phase delay. This suggests that the *h* parameter is qualitatively related to the HP smoothing factor λ. The QWH filter modifies the regular H filter H(2,1) by using the average across those three H filters – so the QWH trend (see Figure 3) is simply a horizontal equally weighted linear combination of the three trend values in each year, similarly to expected utility being a weighted average of the single potential utility outcomes.[21]

4.1 Statistical Comparison and Evaluation

Table 1 reports all numeric results of our investigation for the H filters, and Table 2 for the HP filters, including the benchmark results, respectively. As already seen in Figure 1, the GCEE gaps involve the higher volatility among the two benchmarks: the GCEE standard deviation (SD) is 2.3 versus 1.7 for the EUPF gaps. The latter is not only the lower SD among the two benchmarks but is also the lowest compared to all investigated H filters. Accordingly, the absolute average gap size of EUPF (1.3) is the lowest in Table 1. Both the regular H(2,1) and H(3,1) display higher dispersion than the GCEE benchmark, the QWH modification about the same dispersion as GCEE. It turns out that both statistical indicators, SD as well as gap size, rise in parameter *h* for the H filters. While the H(1,1) indicators are closest to the EUPF values, the QWH indicator results are very close to the GCEE values.

Comparing these H filter results with those of the HP filters in Table 2, we find that the H filters generate higher gap SDs as well as average gap sizes than the HP filters, on average. While the H filter SD values range from 2.0 (H(1,1) filter) to 3.2 (H(3,1) filter), the HP values are located in bracket 1.2 to 1.8 only. The H filters' average absolute gap sizes, in turn, move within a range of 1.4 to 2.5, whereas the HP values' range is 0.9 to 1.4. As with parameter *h* in the H filter context, the HP filters involve a rising SD as well as average gap size for higher λ, as expected: 1.2 for λ=2.91 and 1.8 for λ=100 (SD), and 0.9 for λ=2.91 and 1.4 for λ=100. That is, for both species of filter alike we find that both indicators rise in the respective smoothing factor, i.e. in *h* for the H filter and in λ for the HP filter.

[21] Table A.1, in the appendix, reports all detailed gap estimates by year (H-filters and HP filters), plus the turning points based on our dating rule above (marked by bold numbers), including those of the benchmark indicators.

On analysing congruency of the H filter gaps with the benchmarks EUPF gap and GCEE gap (second section of Table 1) it turns out that sign congruency is best among the two benchmarks themselves (86%); it is lowest for the H(1,1) (70% for both benchmarks alike), followed by H(3,3) with 75% for the EUPF and 77% for GCEE. The regular H(2,1) and the QWH filters both have 80% sign congruency, with both benchmarks. There is no systematic linear relationship between parameter *h* and the sign congruency with the benchmarks. Comparing these results to the HP filter results in Table 2, we find that the regular and the QWH filters can only outperform the HP2.91 and HP6.25; the latter is still equally good for the EUPF benchmark. The HP2.91 is similarly congruent to H(3,1), the H(1,1) sign congruency is lower than even the worst HP filter's. In line with our finding with regard to parameter *h*, we do not find a systematic linear relationship between the HP filter's sign-congruency and its smoothing factor λ. We find the highest sign congruency for the HP20, and the second-best HP100 outperforms the best-performing H filters (the regular and the QWH), too.
Turning to RMSE, we find that the benchmarks reveal a value of 1.61 to each other. With regard to EUPF, all but H(3,1) can outperform the GCEE benchmark: we find the lowest RMSE(EUPF) among the H filters for QWH, at only 1.12; the regular H(2,1) displays 1.55, which is similar to H(1,1) (1.54). With regard to RMSE(GCEE), in contrast, none of the H filters can outperform EUPF (1.61). The best value is again QWH (1.71), followed by the regular H filter (2.03). Turning to the HP filter results (Table 2) for benchmark EUPF, we see that all HP filters involve an RMSE(EUPF) that is below unity, i.e. better than the GCEE's RMSE(EUPF). HP2.91 is clearly the worst among the HP filters, but with an RMSE(EUPF) of 0.91 it is still better than any H filter. We find the lowest RMSE(EUPF) for HP20, at only 0.61, followed by HP100 with 0.66. Turning to RMSE(GCEE), the highest error is again found for HP2.91 with 1.98. This high value, however, is again still lower than any of the RMSE(GCEE) of the three standard H filters, i.e. the HP filters clearly outperform the regular H filters for this indicator. But the QWH (1.71) is able to outperform all HP filters but HP100 (1.47) with regard to RMSE(GCEE). It turns out that RMSE(GCEE) is decreasing in smoothing factor λ; we do not find this pattern among the H filters for parameter *h*.
Evaluating the next part of the summary tables (correlation indicator), we find that all analysed filters are statistically significantly correlated with each other, albeit to different degrees. With benchmark EUPF all H filters but H(1,1) are more highly correlated than the alternative benchmark GCEE (0.80), though both benchmarks are rooted in the production function approach. We find the highest correlation for QWH (0.90) and the regular H filter (0.88). For the GCEE benchmark, the correlations of the H(1,1) as well as of the regular H filter (0.76) are lower than that of EUPF (0.8), but the other two H filters, H(3,1) and QWH, also reveal a correlation of 0.8. The correlation with

GCEE increases with parameter h. Checking these H filters' EUPF correlations against those of the HP filters (Table 2), we see that only the regular H (0.88) and the QWH filter (0.9) can outperform the worst HP filter, so HP2.91 (0.86). HP20 (0.94) and HP100 (0.93) outperform the best H filter QWH (0.90); the latter is equally highly correlated as the HP6.25. Turning to the GCEE correlations, we find a similar pattern with respect to λ to that we found for h: the correlation with benchmark GCEE rises with growing smoothing parameter. The regular H filter is equally highly correlated with GCEE as HP20, but higher than the HP filter with lower λ. QWH and H(3,1) are correlated even more strongly with GCEE, but less than HP100, which has the highest correlation with GCEE of all (0.85).

Evaluating the correlations of the H and HP filters among each other (Table 3), we find, as expected, that the H(1,1) is closest to HP2.91, the regular H(2,1) closest to HP20 (but there are no noteworthy differences to the other HP filters), and the H(3,1) with the standard HP100. Interestingly, the QWH filter has highest correlation with the standard HP100, but the differences, especially to HP20, are not big. The correlation between regular and modified QWH filter is very high (0.97).

4.2 Evaluation of Gap-Inflation Link

Turning to the important link of output gap and inflation change (fourth part of Table 1), we find that H(1,1) and QWH display the highest sign congruency among the H filters: 66% and 64%; the regular H filter is third with 61%. Compared to the HP filters (fourth part of Table 2), the regular H filter is as good as the best HP filter (HP20) with regard to sign congruency to inflation change (0.61); H(1,1) and QWH are better. The standard HP100 performs worse (59%); the worst congruency among the filters we find for HP2.91 and HP6.25 with 0.5 and below. Comparing the filter results with that of the EUPF benchmark, we find that the best H filter H(1,1) displays equally high sign congruency as EUPF (both 0.66), while all HP filters perform worse. The GCEE benchmark (0.59), in contrast, is outperformed by all H filters but H(3,1); among the HP filters, HP100 is equally good and HP20 better (0.61). Turning to Pearson's correlation coefficient of output gap and the change in inflation (next line in tables), we find that it is statistically significant for all H filters but H(3,1), while only HP100 displays a significant correlation coefficient among the HP filters (0.26). We find the by far best indicator value for H(1,1) with 0.42, followed by QWH (0.33), and regular H(2,1) with 0.30. For comparison, the PCCs of both production-function benchmarks are statistically significant; but the EUPF displays only 0.29, and GCEE 0.33. Thus the QWH filter performs equally good as the GCEE benchmark (both 0.33) but better than the EUPF (only 0.29), which is still better than any of the HP filters. That is, while all H filters but H(3,1) and both benchmarks are able to capture the theoretically predicted link of output gap and change in inflation, only the standard HP100 filter is able to do so. With only 26% of perfect positive correlation, however, it performs worse than all H filters but H(3,1). Therefore, all HP filters alike are worse in capturing the gap-inflation link than the two benchmarks, or than H(1,1), H(2,1), and QWH. All H filters but H(3,1) are better at capturing the gap-inflation link than the EUPF method. The GCEE (0.33), however, is better at doing so than the regular H filter (0.30). Interestingly, both sign congruency and PCC decrease in parameter h, and we find that the H(1,1) is the best estimator for capturing the gap-inflation link. This finding is especially important for central banks. Within HP filters, the best choice is smoothing factor 20 (sign congruency) and 100 (correlation). Overall, we find that the gap-inflation link is clearly better captured by the H filters.

4.3 Unbiased Benchmark Predictor Test and Inflation Link Regression

Our tests for unbiased prediction of the benchmark indicators suggest that the H filters are no unbiased predictors of the two benchmarks (Table 1, last but two section): the p values of the F test are for both benchmarks all below any standard significance level, so that we reject the null of unbiased predictor. Turning to the HP filters (Table 2), we find that this also holds for the GCEE benchmark. Looking at the EUPF results of the HP filters, in contrast, we obtain a mixed result. Demanding a confidence level of only 90%, we reject the null of unbiased predictor for the HP20 and HP100 filters, but we cannot reject the null that HP2.91 and HP6.25 are unbiased predictors of the EUPF benchmark. Demanding the standard confidence level of 95%, however, we have to reject this null for HP6.25, too; and demanding even high confidence of 99%, we even have to reject the null for HP2.91, too – that is, at significance level 1% we reject each null that states that the considered HP filters are unbiased predictors. Based on the marginal significance level (p value) ranking, the best predictor is HP20: with a p value of 0.48 it is relatively clear that the null can be accepted. All other p values are below 12%, in contrast. That is, none of the considered H and HP filters are unbiased predictors of the GCEE benchmark, but HP20 in particular is an unbiased predictor of the EUPF benchmark, while none of the H filters are. Evaluating the p values of the two benchmarks EUPF and GCEE themselves, we find that both p values are more or less zero: we can reject the null of unbiased predictor of EUPF for GCEE, and vice versa. In the remaining block on the regression of the change in inflation on the output gap (Diff(Inflation)) we report R^2, root MSE, the slope coefficient β and the p-value of the t-test for statistical significance of this coefficient, respectively (Tables 1 and 2). We find that benchmark GCEE explains more of the inflation variation than EUPF (R^2 of 0.11 versus 0.085) and that the root MSE of GCEE is smaller, too (1.139 vs. 1.154). Both slope terms are positive and statistically different from zero, as required. An increase of the EUPF gap has a stronger effect on inflation. Comparing those benchmark results with the H filter values (Table 1), we find that all H filter gaps but H(3,1) explain more inflation variation than EUPF, and that H(1,1) as well as QWH also explain more inflation variation than GCEE. The same holds pertaining to the root MSE results. All H filter coefficients β are positive and significant at the 7.8% significance level. The highest coefficient is found for H(1,1), followed by EUPF and QWH. Moreover, while EUPF and GCEE are significant at levels of 1.5% and 1.2% QWH and H(1,1) are significant at levels of 0.9% and even 0.0%. Thus QWH and especially H(1,1) outperform the benchmarks with respect to the gap-inflation link; the regular H filter has slightly higher R^2 and slightly lower root MSE than EUPF, but β is significant at a higher significance level. We also find that R^2 as well as β decrease and root MSE as well as the t-test p-values increase in parameter h. Comparing these findings with the HP filter results in Table 2 it turns out that the HP filters can capture the gap-inflation link worse than the benchmarks and all H filters but H(3,1). All R^2 are lower, all root MSE are higher than those of the benchmarks or these H filters. Nonetheless, all β are positive and weakly significant; none of the β is significant at the 5% level, however. The best performance we find for the conventional HP100 filter. Overall, the gap-inflation link is best captured by the H(1,1) and QWH, closely followed by GCEE. Consequently, the H filter outperforms the HP filter in capturing the gap-inflation link.

4.4 Business Cycles Adequacy

Turning to the correlations between the change of the gap and the series of benchmark expansion dummies, these are positive and significant for all methods, as required. The correlation with the official recession dating of the GCEE increases with parameter h for the H filters, ranging from 0.36 for H(1,1) and 0.52 for H(3,1); interestingly, averaging those three filters, the QWH filter performs even slightly better (0.53). Again, the HP filter reveals the same pattern with λ: the PCC increases from 0.357 (HP2.91) to 0.429 (HP100). But the coefficients are lower, and even HP100 can only outperform H(1,1), which means the H filters perform better. Turning to the expansion-phase dummies based on our own dating rule for benchmark EUPF, [22] we find exactly the same pattern:

[22] Besides the benchmark dating of the GCEE (Breuer et al. 2022), there are, of course, other dating results available, though less up-to-date, partly also discussed by Breuer et al. (2022). Gebhardt and Siemers (2020: Appendix B) evaluate several business-cycle dating results, respectively based on monthly and quarterly multidimensional business data, in a kind of meta-analysis. They conclude that much simpler univariate filter methods applied to more aggregated annual GDP data, may produce more or less similar ex post business cycle dating. A crucial point is the treatment of intermediate cycles in between main cycles. An instance is that the GCEE (Breuer et al. 2022) does not consider 2011/12 as an end of the business cycle starting in 2008; they also reject identifying a recession in 2012 to 2013; as others, they also find a cycle in this period, but, based on other indicators, they draw the conclusion that this period represents only an interruption of the business cycle which continues afterwards. Applying our dating rule, requiring five conditions to hold, include the precondition that turning points are only sufficient to be identified as main turning point if there are in advance and afterwards gaps of at least 1% in absolute terms. With our rule, only H(1,1), the regular H(2,1), and HP2.91 do identify in 2011 a main peak that ends a short cycle after the Great Financial Crisis; all other dating results are in accord with the GCEE opinion. That is, our five-requirement dating rule may represent a simple substitute for more complicated rules for researchers that do not have resources and men-power comparable to GCEE.

the correlation increases in h and λ, and the H filters outperform HP. The regular H filter (0.58) performs better than all HP filters but HP100 (0.60), which is only slightly better; but QWH (0.64) and H(3,1), at 0.71, display the highest correlation and are more highly correlated with the EUPF expansion-phase dummies. This suggests that recession-dating adequacy increases with higher sensibility for longer cycles. Eventually analysing the recession-dating adequacy among the two benchmarks, we find that the correlation between the official dating of GCEE and the change in gap of the EUPF is 0.484, and that that of the GCEE gaps used in our analysis is 0.485. This is a bit less than that of the regular H filter H(2,1), and higher than any of the HP filter's correlations. Turning to the adequacy of dating with the EUPF benchmark dating with its own gaps changes, we find a PCC of 0.621, and the GCEE gap changes reveal a PCC of 0.642, which is the highest PCC of all methods analysed. Both benchmarks are more highly correlated than any HP filter. The highest PCC with the EUPF dating is found for H(3,1), followed by GCEE, QWH, and EUPF itself.

4.5 Spectral Analysis

Our results so far suggest that both a rising parameter h and a rising smoothing factor λ do increase the sensibility for longer cycles. While the latter is well known, we use spectral analysis to prove this for the H filters as well. Spectral analysis allows the deduction of the "typical" cycle length and cycle-length sensibility of our H and HP filters. In a first step, as a reference point to our spectral analysis, we computed for each considered method all corresponding average cycle lengths, based on our dating rule (final row of Table 1 and 2). We find similar average cycle lengths among the H filters (Table 1), ranging from 7 (regular H filter) to 9 (H(3,1) and QWH) – and in line with our finding that QWH may be more sensitive to longer cycles than the regular H filter. This is also in accord with the benchmark results of 9 (EUPF) and 8 (GCEE). While the comparison of H(1,1) and H(3,1) fits to the hypothesis that the reference cycles rise in h, the shortest length of the regular H filter with h=2 falls a bit out of this pattern. Turning to the HP filter results (Table 2), however, we find respective lengths in accord of the hypothesis with regard to the smoothing factor: the length weakly increases in λ, from 7 (HP2.91) to 9 (HP20), and to an extraordinary 14 years of HP100.[23] The latter confirms the results of Mohr (2001), who argues that smoothing factor 100 involves unusually short frequencies, or conversely an unusually long reference cycle of about 13 years. Turning to our spectral analysis, we calculated the respective cumulative spectral distribution (CSD) and the corresponding spectral density functions (SDF) for all analysed filters (cf. Figures 4 to 7). In all figures we mark the lower cut-off frequency for the typical business-cycle length of 8 years –

[23] Kufenko and Geiger (2017) found for the OECD a range of 4.2 to 7.4 years. This would be more in line with the regular H- and the HP2.91 filters (7 years).

i.e. a natural frequency of 1/8 – by a red vertical line.[24] Turning to Figure 4 (H filters), we find that the CSD starts steeper the higher parameter h is; consequently, the H(1,1) CSD accumulates 15.9%, regular H(2,1) 37%, and H(3,1) 57% of the distribution already at the lower cut-off frequency 1/8. That is, the higher we choose h the more sensitive the H filter is for longer cycle durations. This is also confirmed by the SDF in Figure 5: the H(1,1) and H(2,1) peaks are both at $0.\overline{18}$ (5½ years), but for lower frequencies the densities of the regular H(2,1) are higher, and for higher frequencies the H(1,1) densities are relatively evenly distributed, whereas the H(2,1) densities decrease step by step. That is, the sensibility of the regular H(2,1) for longer cycle durations is higher. The H(3,1), in contrast, has its peak more or less at $0.0\overline{45}$ (22 years) and $0.\overline{18}$ (5½ years) alike, and for frequencies above 0.2 (or below 5 years) the densities decrease markedly. Therefore, the reference cycle length increases with parameter h, as is already known for the HP smoothing factor λ. The latter known result is reproduced for our German sample in Figures 6 and 7. While the CSD for HP2.91 and HP6.25 remains more or less zero up to frequency 0.09, it already rises at frequency 0.068 for HP20, and for HP100 already at frequency 0.045 (Figure 6). At the 8-years threshold, HP2.91 accumulates only about 10%, the HP6.25 about 13%, rising to 25% for HP20 and as high as about 40% of the distribution for HP100. All HP filters have their respective density peak at $0.\overline{18}$ (5½ years), but obviously the densities decrease in λ (Figure 7). For lower frequencies, densities rise in λ, and vice versa.

Finally, the CSD of the modified H filter (QWH in Figure 4) is similar to the regular H filter's, but accumulates a higher part of the distribution at the 8-years threshold, that is, QWH is more sensitive for longer cycles (Figure 4). Accordingly, in Figure 5, the peak at $0.0\overline{45}$ (22 years) is more pronounced in comparison, and that at $0.\overline{18}$ (5½ years) is lower. We can thus confirm the critique of Quast and Wolters (2022) that the regular H filter distributes the typical cycle frequencies above 1/8 less evenly than the QWH for our German sample, but, as demonstrated, it also covers more cycles with durations above 8 years. So the finding in Quast and Wolters (2022) as well as Schüler (2018), that the regular H filter amplifies longer and mutes short cycles, is to be qualified: the QWH indeed has higher densities for shorter cycles, but, as we show, also for longer cycles; the cumulative density distribution reveals that the mass of distribution on cycles longer than 8 years is higher for the QWH in comparison to the regular H filter. This is rooted in the less even distribution – the regular H filter is more centred around medium-long cycles.

[24] Based on Burns and Mitchell (1946) or Stock and Watson (1999), typical business-cycle frequencies are located in the range of 1.5 to 8 years.

4.6 Crisis Resilience

The sample time-frame includes three significant shocks for the German economy: the German reunification (1989 to 1991) was a positive consumption demand shock, while the Great Financial Crisis (in Germany in 2009) and the recent COVID-19 pandemic were negative shocks. Figure 3 shows GDP as well as H and QWH trends in these crises in comparison. In combination with Figure 2, we see that the H filter produces non-reliable trend patterns in times of crisis. In the reunification years from 1989 to 1991, German economic growth rates were extraordinarily high and increased from an already high 3.9% in 1989 to 5.3 (1990) and 5.1 (1991), because the East German consumers were able to consume all the West German goods for the first time in their lives, and enjoyed an income and wealth effect due to a very beneficial fixed exchange rate between the East German Mark and the West German Deutschmark (DM).[25] In 1992 and especially 1993, this overheating resulted in a drop of economic growth, down to –1.0% in 1993. This GDP pattern is, delayed by two years, followed by the regular H filter, that is, there is a rise in trend up to 1994, followed by a drop in 1995, which is not a convincing development of potential output (Figure 3). Turning to the big financial crisis, the regular H filter, as a lagged-dependent-variable regression, also produces an instable up and down of potential output in the years 2003 to 2012, and suggests that the slope of the potential-output curve even increases during the financial crisis, which is odd – so (*i*) a small boom around the millennium causes an implausible rise of trend in the years 2003 to 2004, though the economy experiences a weak sideward movement of GDP; and (*ii*) potential output increases markedly just in the global crisis years 2007 to 2010. The QWH trend, in contrast, more smoothly follows GDP, with a much less curious rise and drop. The Great Financial Crisis and COVID-19 pandemic both hit the German economy in times of an expansion in 2008 and a weakened expansion in 2019 respectively, so that, due to two years' delay via autoregression, the H trend continues to grow until 2010 and 2021 respectively, despite severe crisis; and as a delayed echo of the crises the trend drops severely after a duration of two years in 2011 and 2022. That is, by construction of the autoregression, estimated potential output levels increase in significant crisis and decrease in the aftermath in times of recovery; the opposite happens in times of significant booms followed by significant consolidation. Therefore, the H filter causes odd results when an economy is hit by a positive or negative shock, that is, experiencing a significant ad-hoc boom with

[25] The DM became de facto the unofficial medium of exchange directly after the Fall of the Berlin Wall and the unofficial end of the inner border in 1989. These days the exchange rate was about 1:20. Beginning at January 1990, East-Germans were able to open DM accounts; the official fixed exchange rate was 1:5. In May 1990, an economic and currency union was determined. Wages, pensions, rent payments etc. are now even exchanged 1:1. Accounts and cash is exchanged at rates between 1:1 and 1:2. Obviously, these exchange rates were determined politically, and not by the market.

a consolidation to normal including a period of weak and negative growth, and vice versa. Again the QWH trend suffers much less from this odd characteristic in both the Great Financial Crisis and pandemic period, but smooths the development of the potential output proxy much more realistically, as intended by Quast and Wolters (2022) with their modification. That is, the QWH filter demonstrates its advantage in full strength especially in years of crisis and huge turning points of GDP. The QWH trend is a much more reasonable description of the true development.

Figure 8 shows the results of the two most extreme HP filters analysed for the same crises; HP2.91 has the smallest smoothing factor and HP100 the greatest. Given HP0 would be the GDP development itself and HP∞ the linear trend, HP2.91 follows GDP relatively closely and HP100 a linear trend relatively closely. Consequently, the other two HP filters run in between the two illustrated trends, HP20 closer to HP100, HP6.25 closer to HP2.91. That is, the HP trend, as an estimate for potential output, is quite elastic to current GDP movements for low smoothing factors, medium elastic for factors like 20, and least elastic for 100 and higher. The consequence is that in situations of massive shocks potential output is assumed to be quite stable in the short and medium run for high λ. With a factor of only 2.91 or 6.25, accordingly, this involves the odd assumption that potential output increases quite elastically upwards in times of massive booms and downwards in huge contractions. Beyond this property, Figure 8 reveals that the criticised smoothing of statistical structural breaks also causes additional trouble in times of strong ups and downs: the higher λ is, the more this up and down (reunification) or down and up (Great Financial Crisis and COVID-19) causes smoothing of the trend before and after the shock. In the upper part of Figure 8 we observe that the HP100 already responds to the boom in 1989-1991 in the pre-years by an increase of the trend: as far back as 1982 the trend starts to increase and runs above the observed GDP development: there are continuously negative gaps until 1989, which is odd, most of all for 1989; in these pre-boom years the trend growth is increasing continuously, which does not make sense. The counter effect of this “pre-smoothing” is a “post-smoothing” in the following years, where the trend growth, despite massive boom, now already begins to decrease in 1990, before the downturn in 1993; the gap remains, as a consequence, positive even in the contraction. We observe something similar in the two negative shocks with inverted sign: despite boom time before the Great Financial Crisis and, at a weaker degree, also before the recent pandemic, the trend is not rising. This is caused by the later recovery afterward, where the GDP returns to normal level: the recovery is also smoothed. So it is like laying a thread over a hill or over a hole and then stretching it from both sides: the result is a more and more straight line, depending on how tightly you stretch. In the HP analogy, the strength of stretching is determined by λ. So this characteristic is rooted in the HP filter, which is a two-sided filter, and the effect is stronger the higher smoothing factor λ is. The user, therefore, has to evaluate the trade-off between smoothing and odd behaviour in times of huge shocks: the higher λ the more convincing the annual trend development in times of crisis but the less convincing the trend development in pre- and post-years due to pre- and post-smoothing. In this challenge, a smoothing factor of about 20 to 40 seems to represent an adequate compromise between the two extremes illustrated. In comparison, the H filter benefits from its one-sided filter property: there is no pre- and post-smoothing; but the regular H filters, in contrast to HP, suffers from the autoregression-immanent construction of delayed developments. This drawback, however, is overcome by the modified version of QWH, which means that this is the preferred filter in times of huge shocks.

4.7 Simulations: Optimising the HP filter via Smoothing Factor

Our results suggest a preference for QWH among the single Hamilton filter variants. Among the investigated HP filters, however, the conclusion depends on the preference regarding the benchmark: for EUPF the preferred one is HP20, while it is HP100 for GCEE. The HP filters allow more flexibility due to smoothing factor $\lambda \in (0, \infty)$. We therefore also simulated Pearson's correlation coefficient and RMSE for the two benchmarks (the results are illustrated in Figures A.1 to A.4 in the appendix) and deduce the optimal $\lambda \in [1{,}2000]$. Regarding the EUPF benchmark, we find that the optimal λ is 38 for both maximizing PCC(EUPF) and minimizing RMSE(EUPF). That is, neither the German-specific value of 20 (Mohr 2001) nor that chosen for the harmonized value within ESCB (30) is the optimal choice if the EUPF is the desired benchmark and correlation or RMSE with it the objective – but both values are close to it. Comparing the overall gain in correlation with EUPF we find that it is economically insignificant: PCC(HP20)=0.935, PCC(HP30)=0.940, and PCC(HP38)=0.941. Regarding RMSE(EUPF), it is 0.610 (HP20) versus 0.584 (HP30) and 0.579 (HP38). For the GCEE benchmark, in contrast, we find the upper limit of our analysed bracket, i.e. the highest simulated value 2000, for both indicators PCC(GCEE) and RMSE(GCEE) to perform best. Believing the GCEE estimates are the best benchmark, researchers ought to use very high smoothing factors far beyond 100, even for annual data. Analysing Figures A.3 and A.4 we find that the marginal gain of increasing the smoothing factor diminishes significantly in a range of about 400 to 500. PCC(GCEE) is 0.765 for HP20, 0.788 (HP30), and 0.852 (HP100) versus 0.890 (HP400), 0.893 (HP500), and 0.909 (HP2000). The RMSE(GCEE) diminishes from 1.739 (HP20) to 1.677 (HP30) to 1.468 (HP100), versus 1.301 (HP400) to 1.290 (HP500) to 1.248 (HP2000). Nonetheless, high values such as 2000 produce trends that are already close to linear trend estimations. Moreover, Mohr (2001) notes that there is a trade-off between including too low frequencies (passband problem) and having too low spectral densities for relevant business-cycle frequencies (compression problem). Accounting for this insight, spectral analysis suggests that the GCEE result of extremely high optimal smoothing factor for the HP filter is problematic. Higher λ moves the spectral transfer function to lower frequencies (longer cycles) and hence increases the passband problem (while diminishing the compression problem). Accordingly, we found the reference cycle length of HP100 to be already far beyond the conventional upper limit of eight years. In the context of restricting public debt, for instance, deficits and surpluses ought to balance over the cycles. Being strict would require setting a relatively short reference cycle, i.e. low values of λ. On the other hand, too low reference cycle length involves the risk of not allowing automatic stabilisers to function effectively in practice. This discussion causes Mohr (2002) to suggest $\lambda = 20$ for Germany. In combination with our findings we suggest using higher values, even a little bit higher than the harmonised European value of 30 for Germany.

5 Conclusion

The Hamilton autoregression (H filter) generates smoothing by delaying the development of the considered dependent variable. Hence, the smoothing factor of the H filter is the h parameter, i.e. the amount of forward periods used in autoregression. The H filter's degree of smoothing turns out to be insufficient in times of huge shocks: the Quast-Wolters modification (QWH filter) solves this H filter problem and produces a convincing estimate of potential output development – in contrast to the Hodrick-Prescott method (HP filter). The latter is found to pre- and post-smooth huge shocks. This generates unreliable trend developments in times of big boom and contraction periods; this is directly related to the known HP problem of smoothing structural breaks, and there is no available modification for HP filters to address the issue. An adequate second-best compromise could be a choice of the smoothing factor at around 20 or 40. In times of huge GDP shocks, the QWH filter is the preferred choice.

Compared to expert institutions' benchmarks, it turns out that the QWH is also the preferred H filter: it outperforms the regular H filter, or other versions of the H filter, at least in the context of estimating output gaps and potential output. Among the HP filters we find that HP38, i.e. the use of smoothing factor 38, is optimal for accomplishing a maximum congruency with the EU production function approach (EUPF). However, aiming at being as close as possible to the estimates of the German Council of Economic Experts (GCEE), we find that it is optimal to choose very high values, such as 2000. It turns out that, in comparison, the HP filters outperform even the QWH filter for both benchmarks with respect to sign congruency, statistical correlation, and root mean squared error – but with different smoothing factors. The H filters can only outperform HP filters with very low factors (e.g. 3) for the EUPF benchmark; with typically used factors like 6.25, 20, or 100 the benchmark congruency of the HP filter is better. Regarding similarity with the second benchmark, QWH can slightly outperform HP20, but is also beaten by higher smoothing factors of the HP filter. However, we find that the H filter gaps are more highly correlated with official recession dating results from expert institutions than the HP filters. Our spectral analysis proves a significant degree of similarity between H filter's parameter h and HP's smoothing factor λ. Their sensibilities increase similarly for longer cycles. We confirm the finding that the QWH filter distributes typical business-cycle lengths more evenly than the regular H filter. The latter indeed mutes shorter cycles in comparison. But we also find that the QWH is more sensitive to longer cycles in comparison to the regular H filter. The regular H filter is more centred and hence more sensitive to medium cycle lengths. We add another important finding: the H filter method is the better alternative vis-à-vis the HP filter in terms of capturing the link between output gap and change in inflation, an aspect especially relevant for central banks in controlling inflation. While the benchmark gap estimates of EUPF as well as GCEE perform better than all analysed HP filter versions, the regular H filter performs slightly better than EUPF and only slightly worse than GCEE. The QWH filter performs better than the EUPF benchmark and the regular H filter, and just as well as the GCEE benchmark. The best performance of all is found for H(1,1), that is, the H filter with only one year forward instead of two.

The overall conclusion is that the answer to the Hamilton-HP filter controversy is complex. The better method depends greatly on what exactly the researcher's focus is. In the context of estimating potential GDP and output gaps, researchers ought to choose the Quast-Wolters modification of the Hamilton filter among the H filter options, or HP38 among the HP filters. While the QWH filter will produce gaps that are more highly correlated with official recession dating, the HP38 filter will produce higher gap congruency with expert institutions' estimates based on EUPF. With a focus on the link between output gaps and inflation, the preferred choice is definitely the H(1,1)-filter, followed by QWH. In times of huge GDP shocks, again the QWH filter is best. Regarding external validity, our findings are primarily conditional on Germany. Future research has to investigate whether our findings are valid for other regions as well.

References

Alesina, A., G. Tabellini and F. R. Campante (2008), "Why Is Fiscal Policy Often Procyclical?" *Journal of the European Economic Association* 6(5): 1006–1036.

Backus, D. K. and P. J. Kahoe (1992), "International Evidence on the Historical Properties of Business Cycles," *American Economic Review* 82(4): 864–888.

Baxter, M. and R. G. King (1999), "Measuring business cycles: Approximate band-pass filters for economic time series," *Review of Economics and Statistics* 81: 575–593.

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.

Bouthevillain, C., P. Cour-Thimann, G. van den Dool, P. Hernández de Cos, G. Langenus, M. Mohr, S. Momigliano, and M. Tujula (2001), "Cyclically Adjusted Budget Balances: an Alternative Approach," Working Paper No. 77, European Central Bank.

Breuer, S. and S. Elstner (2020), "Germany's Growth Prospects Against the Backdrop of Demographic Change," *Journal of Economics and Statistics* 240(5): 565–605.

Breuer, S., S. Elstner, F. Kirsch, and V. Wieland (2022), "Konjunkturzyklen in Deutschland – die Datierung durch den Sachverständigenrat", *Perspektiven der Wirtschaftspolitik* 23(3): 200–240.

Bruchez, P.-A. (2003), "A Modification of the HP Filter Aiming at Reducing the End-Point Bias," Working Paper – 18 August 2003, Swiss Federal Finance Administration.

Bundesbank (1995), "Production potential in Germany and its determinants," *Monthly Report* August: 39–53. Frankfurt/Main: Deutsche Bundesbank.

Bundesbank (1981), "Recalculation of the production potential of the Federal Republic of Germany," *Monthly Report* October: 30–36. Frankfurt/Main: Deutsche Bundesbank.

Bundesbank (1973), "Production potential in the Federal Republic of Germany," *Monthly Report* October: 27–32. Frankfurt/Main: Deutsche Bundesbank.

Burns, A.F. and W. C. Mitchell (1936), *Production during the American business cycle of 1927-1933*. NBER Bulletin 61(9), http://www.nber.org/books/mitc36-1.

Burns, A. F. and W. C. Mitchell (1946), *Measuring Business Cycles*. NBER Books, National Bureau of Economic Research.

CBO – Congressional Budget Office (1997), An Economic Model for Long-run Budget Simulations, Washington D.C.

CBO – Congressional Budget Office (2004), A Summary of Alternative Methods for Estimating Potential GDP, Background Paper, Washington D.C.

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: 435–465.

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.

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

Correia, I. H., J. L. Neves, and S. T. Rebelo (1992), "Business Cycles from 1850–1950 – New Facts about Old Data," *European Economic Review* 36(2/3): 459–467.

D'Auria, F., Denis, C., Havik, K., McMorrow, K., Planas, C., Raciborski, R., Röger, W. and A. Rossi (2010), "The Production Function Methodology for Calculating Potential Growth Rates and Output Gaps," ECFIN Economic Paper No. 420, European Commission.

De Loecker, J., J. Eeckhout, and G. Unger (2020), "The Rise of Market Power and the Macroeconomic Implications," *Quarterly Journal of Economics* 135(2): 561–644.

Denis,C., McMorrow, K. and W. Röger (2002), "Production Function Approach to Calculating Potential Growth and Output Gaps – Estimates for the EU Member States and the US," Economic Papers 176, European Commission.

Denis,C., Grenouilleau, D., McMorrow, K. and W. Röger (2006), "Calculating Potential Growth Rates and Output Gaps – a Revised Production Function Approach," Economic Papers 247, European Commission.

Drehmann, M. and J. Yetman (2018), "Why You Should Use the Hodrick-Prescott Filter – At Least to Generate Credit Gaps," *BIS Working Paper* No. 744. Bank of International Settlements.

ECB – European Central Bank (2012), Main Elements of the Fiscal Compact (Box 12), *ECB Monthly Bulletin* March 2012: 101–102.

Ehlgen, J. (1998), "Distortionary effects of the optimal Hodrick–Prescott filter," *Economics Letters* 61(3): 345–349.

FedGov (2024), *Gesamtwirtschaftliches Produktionspotenzial und Konjunkturkomponenten*, Datengrundlagen und Ergebnisse der Schätzungen der Bundesregierung, Spring Forecast of the Federal Government (24 April, 2024), Federal Ministry for Economic Affairs and Energy joint with Federal Ministry of Finance, Germany.

Fildes, R. and H. Steckler (2001), "The State of Macroeconomic Forecasting," *Journal of Macroeconomics* 24(4): 435–468.

Fisher, P. G., L. Mahadeva and J. D. Whitley (1997), "The output gap and inflation – experience at the Bank of England," in: Bank of International Settlement, *BIS Conference Papers* 4: 68–90.

Galí, J. and R. Perotti (2003), "Fiscal Policy and Monetary Policy Integration in Europe," *Economic Policy* 18(37): 533–572.

Gavin, M. and R. Perotti (1997), "Fiscal Policy in Latin America," in: B. Bernanke and J. Rotemberg (eds.), *NBER Macroeconomics Annual* 1997, MIT Press.

GCEE – German Council of Economic Experts (2007), *Staatsverschuldung wirksam begrenzen*, Occasional Report, 12 March 2007, Wiesbaden.

GCEE – German Council of Economic Experts (2017), *Towards a Forward-Looking Economic Policy*, Annual Report 2017/18, Wiesbaden.

Gebhardt, H. and L.-H. Siemers (2020), „Die strukturelle Besserung der Länderfinanzen in der Niedrigzinsphase: die trügerische Leichtigkeit des Scheins“, *Perspektiven der Wirtschaftspolitik* 21(1): 54–78.

Gehringer, A. and T. Mayer (2021), “Measuring the Business Cycle Chronology with a Novel Business Cycle Indicator for Germany.” *Journal of Business Cycle Research* 17(1): 71–89.

Gern, K.-J., T. Schmidt and M. Schröder (2008), *Mittelfristige gesamtwirtschaftliche Prognosen. Ein internationaler Vergleich der Modelle und Prognosegüte*. Wiley-VCH Press.

Granger, C. W. J. (1966), “The Typical Spectral Shape of an Economic Variable,” *Econometrica* 34(1): 150–161.

Hall, V. B. and P. Thomson (2021), “Does Hamilton’s OLS regression provide a “better alternative” to the Hodrick-Prescott Filter? A New Zealand Business Cycle Perspective,” *Journal of Business Cycle Research* 17(2): 151–183.

Hamilton, J. D. (1989), “A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle,” *Econometrica* 57(2): 357–384.

Hamilton, J. D. (2018), “Why You Should Never Use the Hodrick-Prescott Filter.” *Review of Economics and Statistics* 100(5): 831–843.

Harding, D. and A. Pagan (2002), “Dissecting the cycle: a methodological investigation,” *Journal of Monetary Economics* 49(2): 365–381.

Harvey, A. C. (1989), *Forecasting, structural time series models and the Kalman Filter*. Cambridge University Press.

Harvey, A. C. and A. Jaeger (1993), “Detrending, Stylized Facts and the Business Cycle,” *Journal of Applied Econometrics* 8(3): 231–247.

Havik, K., K. McMorrow, F. Orlandi, C. Planas, R. Raciborski, W. Röger, A. Rossi, A. Thum-Thysen, V. Vandermeulen (2014), “The Production Function Methodology for Calculating Potential Growth Rates & Output Gaps,” Economic Papers 535, European Commission.

Hodrick, R. J. and E. C. Prescott (1981), “Post-War U.S. Business Cycles: An Empirical Investigation,” Discussion Papers 451, Northwestern University, Center for Mathematical Studies in Economics and Management Science.

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

Iacobucci, A. (2005), “Spectral Analysis for Economic Time Series,” in: Leskow, J., M. Puchet, and L. F. Punzo (eds.), *New Tools of Economic Dynamics*, Springer Press: 203–219 (Chapt. 12).

JEF – Joint Economic Forecast Project Group (Projektgruppe Gemeinschaftsdiganose) (2024), Deutsche Wirtschaft kränkelt – Reform der Schuldenbremse kein Allheilmittel, *Gemeinschaftsdiagnose* #1-2024. www.gemeinschaftsdiagnose.de

Jönsson, K. (2020a), “Cyclical Dynamics and Trend/Cycle Definitions: Comparing the HP and Hamilton Filters,” *Journal of Business Cycle Research* 16(2): 151–162.

Jönsson, K. (2020b), “Real-time US GDP gap properties using Hamilton’s regression-based filter,” *Empirical Economics* 59(1): 307–314.

Justiniano, A. and G. E. Primiceri (2008), “The Time-Varying Volatility of Macroeconomic Fluctuations,” *American Economic Review* 98(3): 604–641.

Kaiser, R. and A. Maravall (1999), "Estimation of the business cycle: A modified Hodrick-Prescott filter," *Spanish Economic Review* 1: 175–206.

Kálmán, R. E. (1960), "A new approach to linear filtering and prediction problems," *Journal of Basic Engineering* 82 (1): 35–45.

Kálmán, R. E. and R. S. Bucy (1961), "New Results in Linear Filtering and Prediction Theory," *Journal of Basic Engineering* 83: 95–108.

Kaminski, G., C. Reinhart, and C. Vegh (2004), "When It Rains It Pours: Procyclical Capital Flows and Macroeconomic Policies," in: Gertler, M. and K. Rogoff (eds.), *NBER Macroeconomic Annual* 2004, MIT 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.

Krüger, J. J. (2021), "A Wavelet Evaluation of Some Leading Business Cycle Indicators for the German Economy," *Journal of Business Cycle Research* 17(3): 293–319.

Kufenko, V. and N. Geiger (2017), "Stylized Facts of the Business Cycle: Universal Phenomenon, or Institutionally Determined?" *Journal of Business Cycle Research* 13(2): 165–187.

Kyo, K., H. Noda, and G. Kitagawa (2022), "Co-movement of Cyclical Components Approach to Construct a Coincident Index of Business Cycles," *Journal of Business Cycle Research* 18(1): 101–127.

Lane, P. (2003), "The Cyclical Behavior of Fiscal Policy: Evidence from the OECD," *Journal of Public Economics* 87(12): 1661–1675.

Mc Morrow, K. and W. Roeger (2001), "Potential Output: Measurement Methods, "New" Economy Influences and Scenarios for 2001-2010. A Comparison of the EU15 and the US," Economic Working Papers No. 150, European Commission, Directorate-General for Economic and Financial Affairs.

Mitchell, W. C. (1913), *Business Cycles*. University of California Press, Berkeley.

Mohr, M. (2001), "Ein disaggregierter Ansatz zur Berechnung konjunkturbereinigter Budgetsalden für Deutschland: Methoden und Ergebnisse," Discussion Paper No. 13/01, Deutsche Bundesbank, Frankfurt/Main.

NA (2024), Volkswirtschaftliche Gesamtrechnungen [National Account Data], *Fachserie* 18, Reihe 1.5, Wiesbaden: Federal Statistical Office Germany.

Ochsner, C., L. Other, E. Thiel and C. Zuber (2024), "Demographic Aging and Long-Run Economic Growth in Germany," Working Paper 02/2024, German Council of Economic Experts.

Okun, A. M. (1962), "Potential GNP: Its measurement and significance," in: Proceedings of the Business and Economics Statistics Session, American Statistical Association, 98-103. Washington: American Statistical Association.

Pažický, M. (2021), "Predicting Recessions in Germany Using the German and the US Yield Curve," *Journal of Business Cycle Research* 17(3): 263–291.

Pedersen, T. M. (2001), "The Hodrick-Prescott Filter, the Slutzky Effect, and the Distortionary Effect of Filters," *Journal of Economic Dynamics and Control* 25(8): 1081–1101.

Pollock, D. S. G. (2000), "Trend estimation and de-trending via rational square-wave filters," *Journal of Econometrics* 99(2): 317–334.

Quast, J. and M. H. Wolters (2022), “Reliable Real-time Output Gap Estimates Based on a Modified Hamilton Filter,” *Journal of Business & Economic Statistics* 40(1): 152–168.

Ravn, M. O. and H. Uhlig (2002), “On Adjusting the Hodrick-Prescott Filter for the Frequency of Observations,” *Review of Economics and Statistics* 84(2): 371–376.

Reeves, J. J., C. A. Blyth, C. M. Triggs and J. P. Small (2000), “The Hodrick-Prescott Filter, a Generalization, and a New Procedure for Extracting an Empirical Cycle from a Series,” *Studies in Nonlinear Dynamics & Econometrics* 4(1): article 1.

RWI – Leibniz-Institut für Wirtschaftsforschung (2017), Die wirtschaftliche Entwicklung im Ausland und Inland zur Jahreswende 2016/17, *RWI Konjunkturberichte* 68(1), Essen.

Schmassmann, S. and S. Wiedmer (2022), *Analyse alternativer Verfahren zur Berechnung des Konjunkturfaktors*, Final Report, Swiss Ministry of Finance, Swiss Conferderation.

Schirwitz, B. (2009), “A comprehensive German business cycle chronology,” *Empirical Economics* 37(2): 287–301.

Schröder, M., M. Kappler and A. Schrimpf (2008), „Dritter Teil: Studie des ZEW Mannheim,“ in: Gern. K.-J., T. Schmidt and M. Schröder (eds.), *Mittelfristige gesamtwirtschaftliche Prognosen*, Research Report of ifW Kiel, RWI Essen, and ZEW Mannheim. Wiley Press.

Schüler, Y. S. (2018), “On the cyclical properties of Hamilton‘s regression filter,” Discussion Paper No. 03/2018, Deutsche Bundesbank, Frankfurt/Main.

Smets, F. R. and R. Wouters (2007), “Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach,” *American Economic Review* 97(3): 586–606.

Stratonovich, R. L. (1959), “Optimum nonlinear systems which bring about a separation of a signal with constant parameters from noise,” *Radiofizika* 2(6): 892–901.

Stiroh, K. J. (1998), “Long-Run Growth Projections and the Aggregate Production Function: a Survey of Models Used by the U.S. Government,” *Contemporary Economic Policy* 16(4): 467–479.

Stock, J. H. and M. W. Watson (1989), “New indexes of coincident and leading macroeconomic indicators,” in: Blanchard, O. and S. Fischer (eds.), *NBER macroeconomics annual.* MIT Press: 351–394.

Stock, J. H. and M. W. Watson (1991), “A probability model of the coincident economic indicators,” in: Lahiri, K. and G. Moore (eds.), *Leading economic indicators: New approaches and forecasting records.* Cambridge University Press: 63–89.

Stock, J. H. and M. W. Watson (1999), “Business cycle fluctuations in US macroeconomic time series,” *Handbook of Macroeconomics* 1: 3–64.

Talvi, E. and C. Vegh (2005), “Tax Base Variability and Procyclicality of Fiscal Policy,” *Journal of Development Economics* 78(1): 156–190.

Woodford, M. (2009), “Convergence in Macroeconomics: Elements of the New Synthesis,” *American Economic Journal: Macroeconomics* 1(1): 267–79.

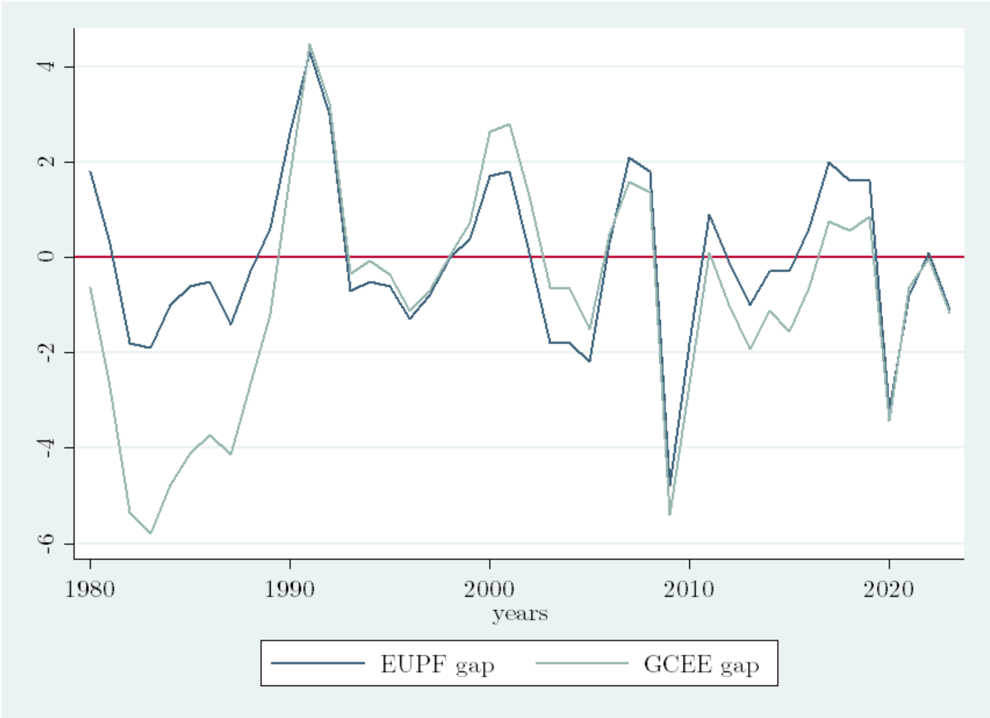

Figure 1: Gaps of the two benchmark indicators (EUPF and GCEE)

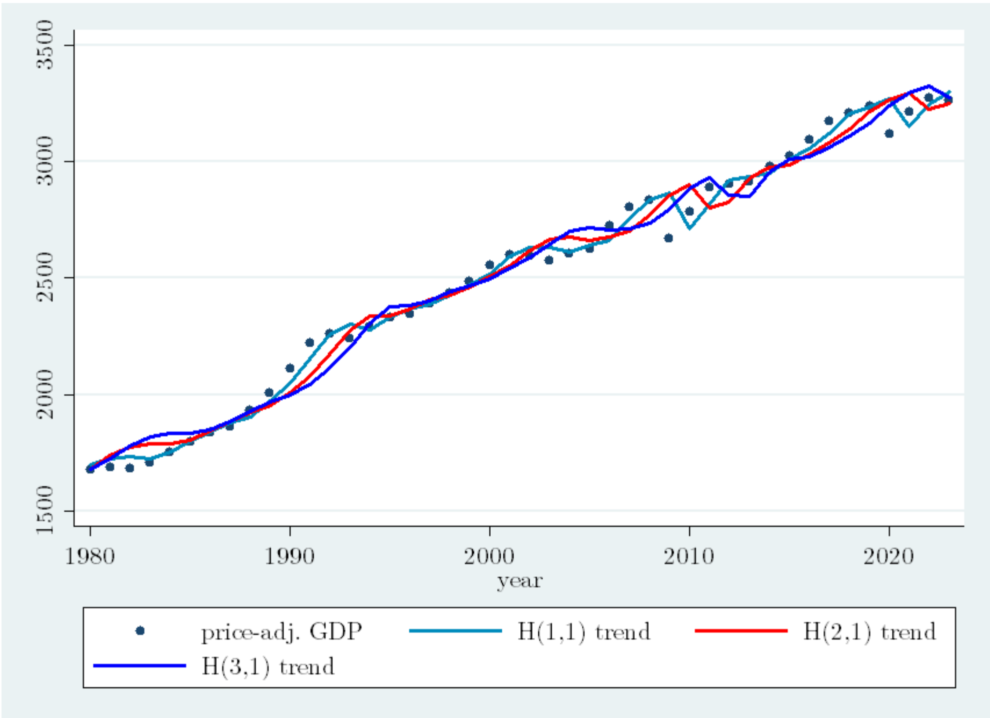

Figure 2: Price-adjusted GDP in comparison to the three estimated Hamilton-filter trends

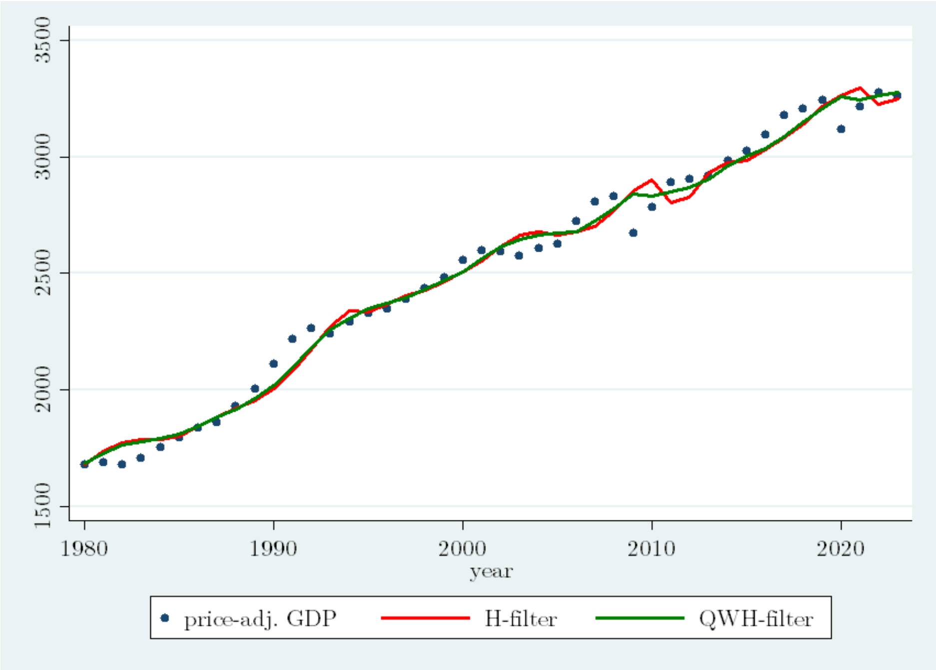

Figure 3: Comparison of regular H filter and modified QWH filter

Table 1: Results table for the Hamilton filters in comparison

indicator	EUPF	GCEE	H(1,1)	H	H(3,1)	QWH
SD	1.7	2.3	2.0	2.8	3.2	2.4
av. gap size	1.3	1.9	1.4	2.2	2.5	1.8
			congruancy with benchmark indicators			
sign congr.(EUPF)	1	0.86	0.70	0.80	0.75	0.80
sign congr.(GCEE)	0.86	1	0.70	0.80	0.77	0.80
RMSE(EUPF)	0	1.61	1.54	1.55	1.91	1.12
RMSE(GCEE)	1.61	0	2.24	2.03	2.08	1.71
			correlation matrix of gaps (PCC)			
EUPF	1					
GCEE	0.80*	1				
H(1,1)	0.65*	0.53*	1			
H	0.88*	0.76*	0.70*	1		
H(3,1)	0.85*	0.80*	0.50*	0.88*	1	
QWH	0.90*	0.80*	0.77*	0.97*	0.92*	1
			gap-inflation link			
sign[Diff(inflation)]	0.66	0.59	0.66	0.61	0.57	0.64
Diff(inflation)	0.29*	0.33*	0.42*	0.30*	0.23	0.33*
			Unbiasedness Test Regressions (*p*-value of *F*-test H_0: α=0 and β=1)			
EUPF	1	0.000	0.000	0.000	0.000	0.000
GCEE	0.003	1	0.002	0.000	0.000	0.000
Diff(inflation)						
R^2	0.085	0.110	0.175	0.091	0.054	0.112
root MSE	1.154	1.139	1.096	1.151	1.173	1.137
β	0.202	0.064	0.255	0.127	0.087	0.166
p-value test β=0	0.015	0.012	0.000	0.038	0.078	0.009
			PCC of 1st difference of gap with expansion-phase dummy			
GCEE (Breuer et al. 2022)	0.484*	0.485*	0.360*	0.494*	0.521*	0.529*
EUPF (own dating)	0.621*	0.642*	0.361*	0.582*	0.713*	0.635*
Average cycle length[a]	9	8	8	7	9	9

Source: Our own calculations of H filter based on price-adjusted GDP (FedGov 2024); EU approach numbers directly taken from FedGov (2024); GCEE (German Council of Economic Experts) gaps directly taken from GCEE homepage. – [a] average cycle length based on our dating rule and starting in 1991.

Table 2: Results table for the Hodrick-Prescott filters in comparison

indicator	EUPF	GCEE	HP2.91	HP6.25	HP20	HP100
SD	1.7	2.3	1.2	1.3	1.5	1.8
av. gap size	1.3	1.9	0.9	1.0	1.2	1.4
		congruancy with benchmark indicators				
sign congr.(EUPF)	1	0.86	0.77	0.80	0.91	0.89
sign congr.(GCEE)	0.86	1	0.75	0.77	0.84	0.82
RMSE(EUPF)	0	1.61	0.91	0.77	0.61	0.66
RMSE(GCEE)	1.61	0	1.98	1.89	1.74	1.47
		correlation matrix of gaps (PCC)				
EUPF	1					
GCEE	0.80*	1				
HP2.91	0.86*	0.66*	1			
HP6.25	0.90*	0.70*	0.99*	1		
HP20	0.94*	0.76*	0.95*	0.98*	1	
HP100	0.93*	0.85*	0.86*	0.91*	0.97*	1
		gap-inflation link				
sign[Diff(inflation)]	0.66	0.59	0.48	0.50	0.61	0.59
PCC with Diff(inflation)	0.29*	0.33*	0.24	0.23	0.24	0.26*
		Unbiasedness Test Regressions (*p*-value of *F*-test H_0: α=0 and β=1)				
EUPF	1	0.000	0.016	0.053	0.483	0.116
GCEE	0.003	1	0.002	0.002	0.002	0.000
Diff(inflation)						
R^2	0.085	0.110	0.056	0.055	0.057	0.066
root MSE	1.154	1.139	1.172	1.173	1.172	1.166
β	0.202	0.064	0.241	0.213	0.189	0.170
p-value test β=0	0.015	0.012	0.052	0.055	0.057	0.067
		PCC of 1[st] difference of gap with expansion-phase dummy				
GCEE (Breuer et al. 2022)	0.484*	0.485*	0.357*	0.380*	0.403*	0.429*
EUPF (own dating)	0.621*	0.642*	0.472*	0.511*	0.558*	0.600*
Average cycle length[a]	9	8	7	9	9	14

Source: Our own calculations of H filter based on price-adjusted GDP (FedGov 2024); EU approach numbers directly taken from FedGov (2024); GCEE (German Council of Economic Experts) gaps directly taken from GCEE homepage. – [a] average cycle length based on our dating rule and starting in 1990, for comparability.

Table 3: Correlations between Hamilton and Hodrick-Prescott filters

	H(1,1) gap	H gap	H(3,1) gap	QWH gap	HP2.91 gap	HP6.25 gap	HP20 gap	HP100 gap
H(1,1) gap	1							
H gap	0.70*	1						
H(3,1) gap	0.50*	0.88*	1					
QWH gap	0.77*	0.97*	0.92*	1				
HP2.91 gap	0.63*	0.74*	0.61*	0.73*	1			
HP6.25 gap	0.60*	0.76*	0.66*	0.75*	0.99*	1		
HP20 gap	0.56*	0.77*	0.72*	0.77*	0.95*	0.98*	1	
HP100 gap	0.51*	0.76*	0.78*	0.78*	0.86*	0.91*	0.97*	1

Source: Our own calculations of H and HP filters based on price-adjusted GDP (FedGov 2024).

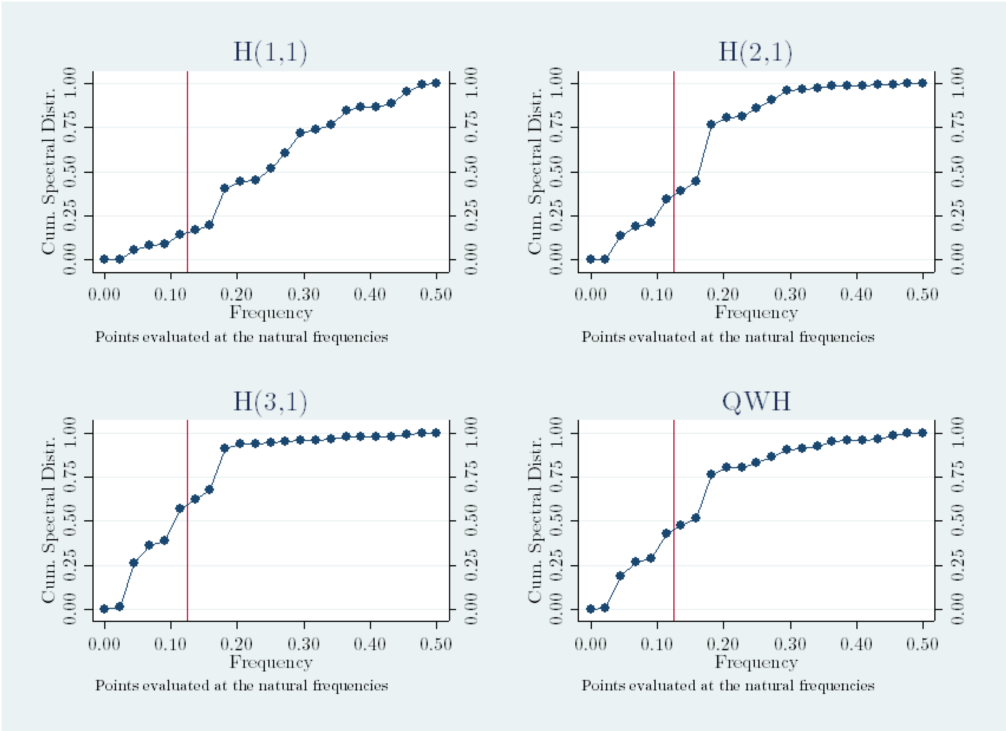

Figure 4: Cumulative spectral distribution of the four Hamilton filters. Red vertical lines at 8-years natural frequency $\frac{1}{8}$=0.125.

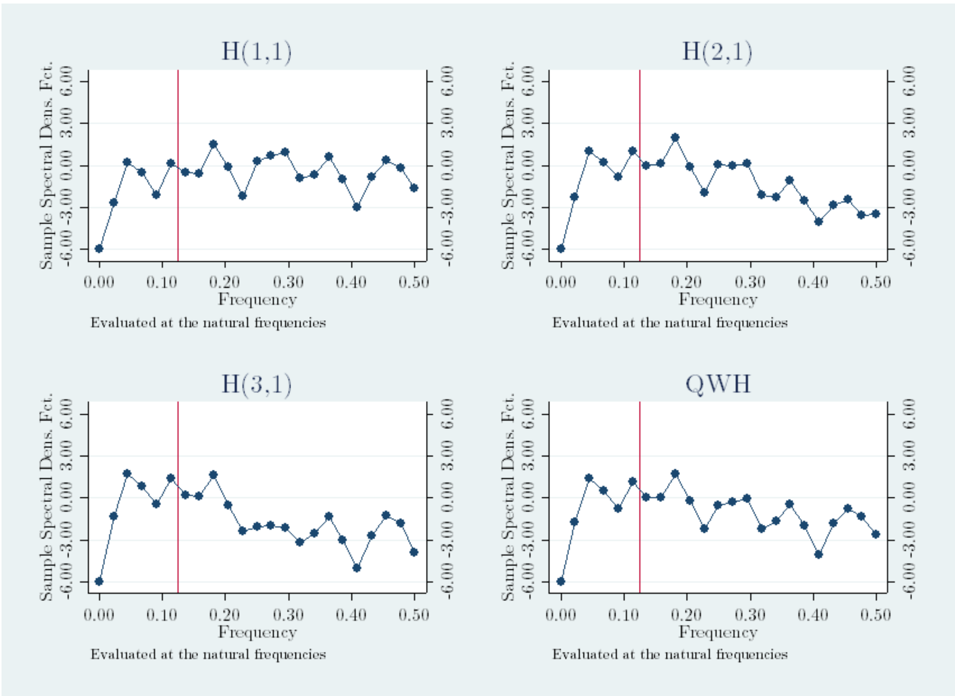

Figure 5: Spectral density function of the four Hamilton filters. Red vertical lines at 8-years natural frequency $\frac{1}{8}$=0.125.

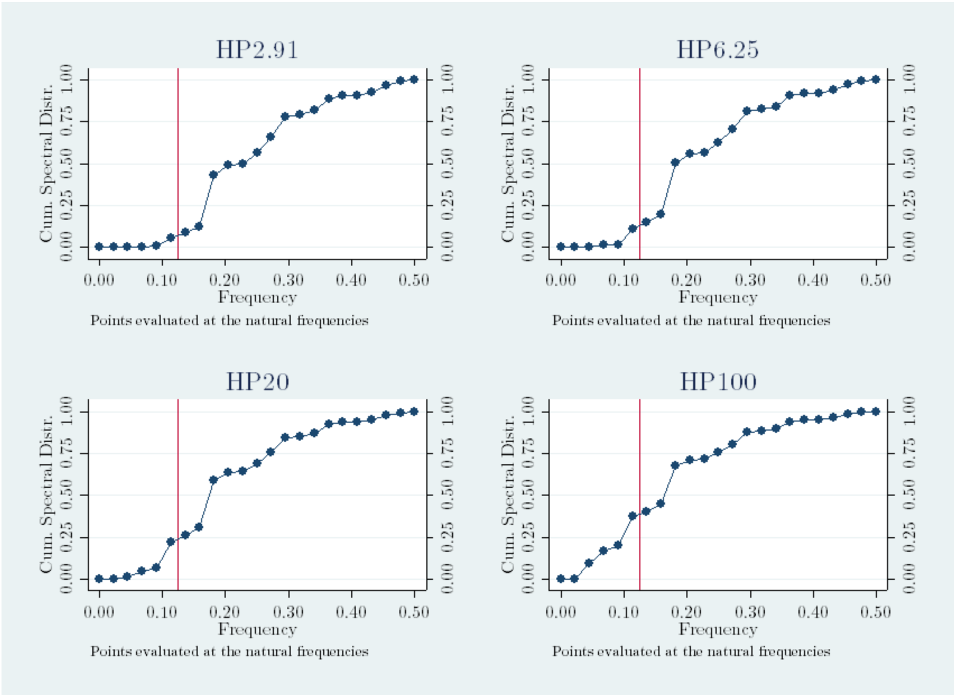

Figure 6: Cumulative spectral distribution of the four Hodrick-Prescott filters. Red vertical lines at 8-years natural frequency $\frac{1}{8}$=0.125.

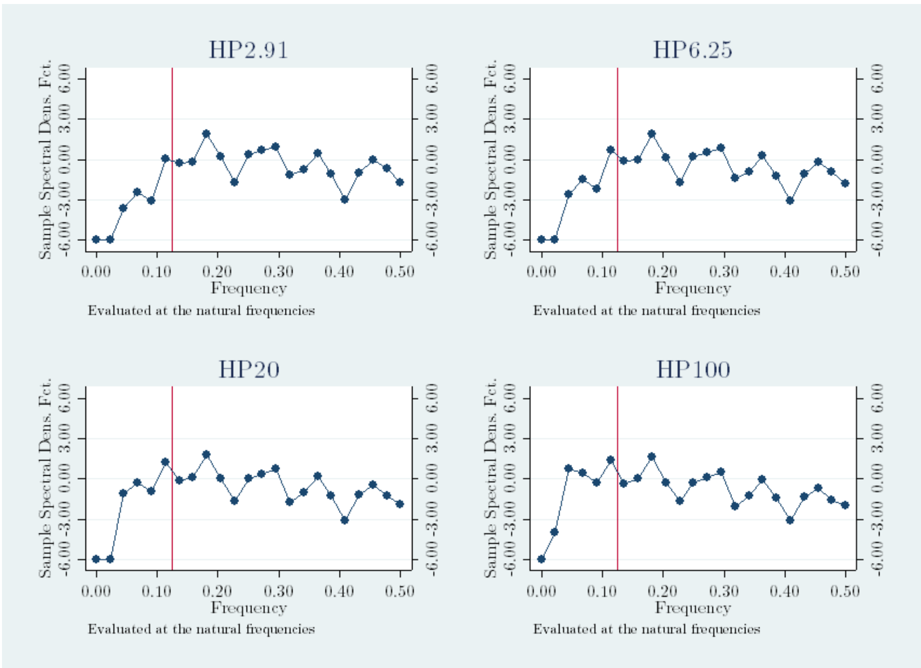

Figure 7: Spectral density function of the four Hodrick-Prescott filters. Red vertical lines at 8-years natural frequency $\frac{1}{8}$=0.125.

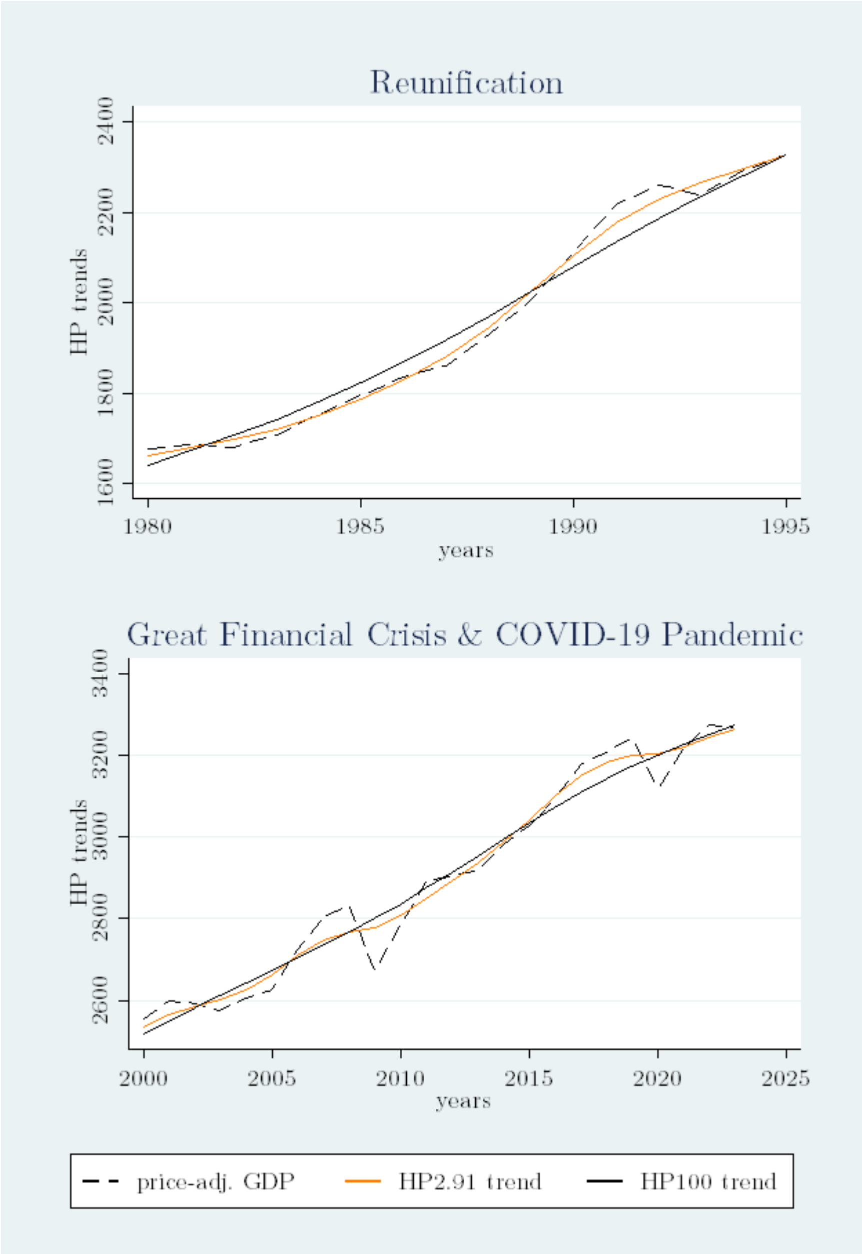

Figure 8: Comparison of HP2.91 and HP100 in times of shocks

Appendix

Table A.1: Estimated output gaps of benchmarks, H filters, and HP filters in %, 1980-2023

year	EUPF	GCEE	H(1,1)	H	H(3,1)	QWH	HP2.91	HP6.25	HP20	HP100
1980	**1.8**	-0.6	-1.3	-0.2	-0.1	-0.5	1.0	**1.3**	**1.9**	**2.2**
1981	0.3	-2.7	-2.1	-3.0	-2.2	-2.5	0.3	0.5	0.9	0.8
1982	-1.8	-5.4	**-3.0**	**-5.4**	-5.6	**-4.7**	-1.0	-1.0	-1.1	-1.5
1983	**-1.9**	**-5.8**	-1.2	-4.6	**-6.3**	-4.1	-0.8	-0.9	-1.2	-2.0
1984	-1.0	-4.8	0.2	-1.8	-4.4	-2.1	0.3	0.1	-0.3	-1.5
1985	-0.6	-4.1	-0.2	-0.4	-2.2	-0.9	0.5	0.3	-0.3	-1.5
1986	-0.5	-3.7	-0.2	-0.3	-0.6	-0.4	0.4	0.1	-0.6	-1.8
1987	-1.4	-4.1	-1.0	-1.2	-1.2	-1.1	**-1.0**	**-1.4**	**-2.0**	**-3.0**
1988	-0.3	-2.7	1.3	0.4	0.2	0.6	-0.8	-1.1	-1.5	-2.1
1989	0.6	-1.2	1.7	2.7	2.0	2.1	-0.9	-0.9	-0.9	-1.0
1990	2.6	1.7	3.1	5.2	5.7	4.7	0.3	0.5	1.0	1.4
1991	**4.3**	**4.5**	**3.2**	**6.8**	**8.5**	**6.1**	**1.9**	**2.4**	**3.1**	**3.9**
1992	3.0	3.2	0.2	4.1	6.9	3.7	1.4	1.8	2.5	3.4
1993	-0.7	-0.4	**-2.7**	-1.6	1.6	-0.9	**-1.2**	**-1.0**	-0.6	0.1
1994	-0.5	-0.1	0.6	**-1.9**	-0.6	-0.7	-0.1	-0.2	-0.1	0.4
1995	-0.6	-0.4	-0.1	-0.4	**-2.0**	-0.8	0.1	-0.1	-0.3	0.1
1996	**-1.3**	**-1.1**	-0.8	-0.8	-1.5	**-1.0**	-0.5	-0.8	**-1.1**	-0.9
1997	-0.8	-0.7	0.2	-0.7	-0.5	-0.4	-0.4	-0.6	-0.9	-0.8
1998	0.0	0.1	0.4	0.3	-0.3	0.2	-0.3	-0.3	-0.5	-0.4
1999	0.4	0.7	0.4	0.8	0.6	0.6	-0.3	-0.2	-0.1	0.0
2000	1.7	2.6	**1.4**	**1.9**	**2.2**	**1.9**	0.8	1.1	1.3	1.5
2001	**1.8**	**2.8**	0.3	1.8	2.2	1.4	**1.2**	**1.5**	**1.7**	1.9
2002	0.1	1.3	-1.5	-0.9	0.2	-0.7	0.3	0.3	0.4	0.5
2003	-1.8	-0.6	**-2.1**	**-3.4**	-2.7	**-2.7**	-1.0	-1.2	-1.3	-1.4
2004	-1.8	-0.7	-0.3	-2.7	**-3.5**	-2.2	-0.8	-1.0	-1.2	-1.4
2005	**-2.2**	**-1.5**	-0.6	-1.5	-3.4	-1.8	**-1.4**	**-1.6**	**-1.6**	**-1.8**
2006	0.3	0.5	**2.4**	1.8	0.7	1.6	0.6	0.7	0.9	0.8
2007	**2.1**	**1.6**	1.8	**3.9**	3.4	**3.0**	2.1	2.4	**2.7**	**2.6**
2008	1.8	1.4	-0.2	2.2	**3.5**	1.8	**2.3**	**2.5**	2.6	2.3
2009	**-4.8**	**-5.4**	**-6.8**	**-6.5**	**-4.5**	**-5.9**	**-3.8**	**-4.0**	**-4.2**	**-4.6**
2010	-1.8	-2.6	2.7	-4.0	-3.5	-1.7	-0.8	-1.0	-1.4	-1.9
2011	0.9	0.1	**2.8**	**3.2**	-1.4	1.5	**1.5**	1.4	1.1	0.6
2012	-0.1	-1.0	-0.6	2.8	1.6	1.2	0.4	0.3	0.1	-0.3
2013	-1.0	-1.9	-0.7	-0.5	2.4	0.4	-0.7	-0.8	-1.0	-1.2
2014	-0.3	-1.1	1.1	0.2	0.9	0.7	-0.2	-0.3	-0.4	-0.4
2015	-0.3	-1.6	0.5	1.3	0.6	0.8	-0.5	-0.5	-0.5	-0.2
2016	0.6	-0.6	1.2	2.0	2.4	1.9	-0.2	0.0	0.2	0.7
2017	**2.0**	0.8	1.7	3.0	**3.8**	**2.8**	0.9	1.1	1.6	2.2
2018	1.6	0.6	0.1	2.1	3.0	1.7	0.8	1.1	1.5	2.1
2019	1.6	0.8	0.2	0.7	2.3	1.1	1.4	**1.5**	**1.7**	**2.2**
2020	**-3.2**	-3.4	**-4.7**	**-4.5**	**-3.8**	**-4.3**	**-2.7**	**-2.8**	**-2.8**	**-2.6**
2021	-0.8	-0.6	**2.1**	-2.4	-2.3	-0.9	-0.1	-0.2	-0.4	-0.3
2022	0.1	0.0	1.0	1.5	-1.5	0.3	1.0	0.9	0.8	0.7
2023	-1.1	-1.2	-1.1	0.4	-0.3	-0.4	0.0	0.0	-0.2	-0.3

Source: Our own calculations of filters based on price-adjusted GDP (FedGov 2024); EU approach numbers directly taken from FedGov (2024); GCEE (German Council of Economic Experts) gaps directly taken from GCEE homepage. Numbers in bold mark turning points of the business cycle, based on our five-requirement dating rule.

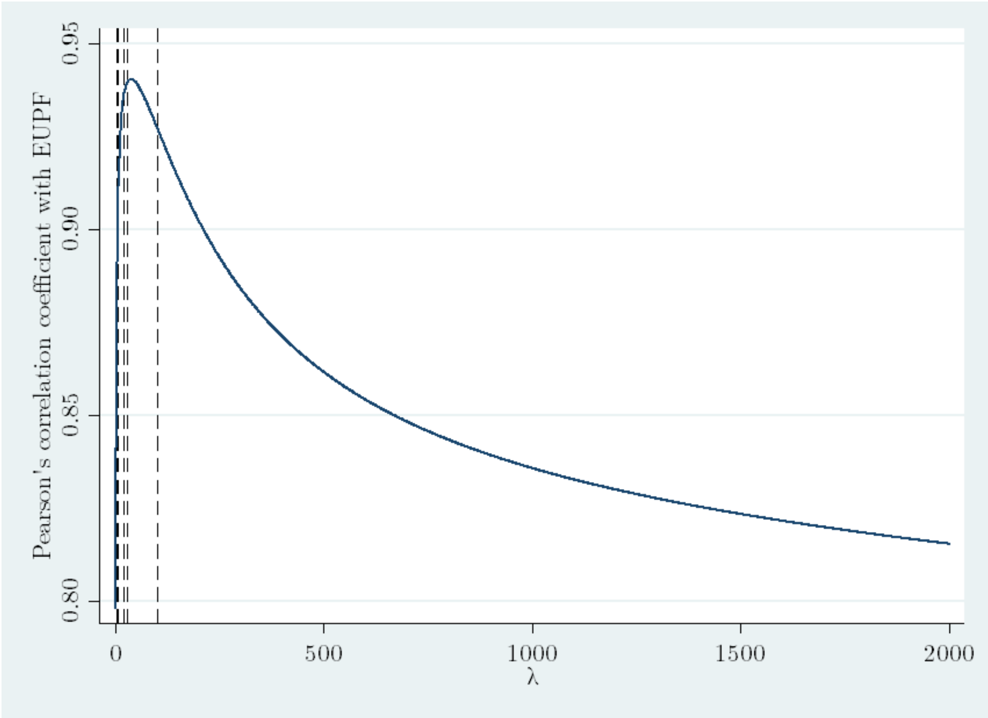

Figure A.1: Correlation coefficient of HP filter with the EUPF gap for varying smoothing factor λ. The vertical dashed lines mark, from the left: λ=2.91, λ=6.25, λ=20, λ=30, and λ=100.

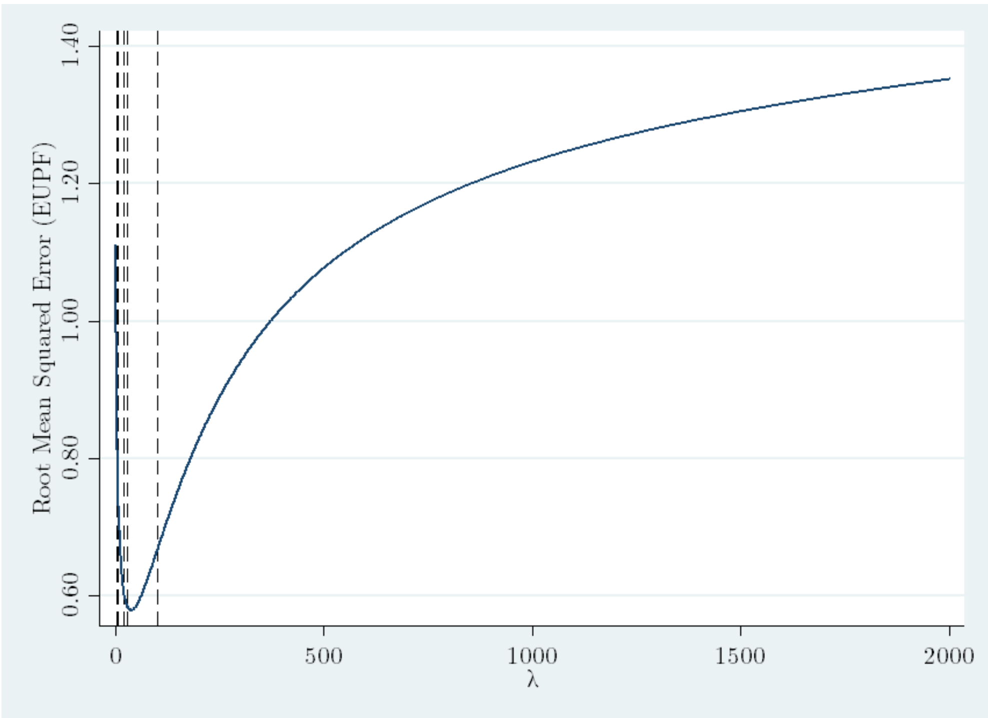

Figure A.2: Root Mean Squared Error of HP filter with EUPF gap for varying smoothing factor λ. The vertical dashed lines mark, from the left: λ=2.91, λ=6.25, λ=20, λ=30, and λ=100.

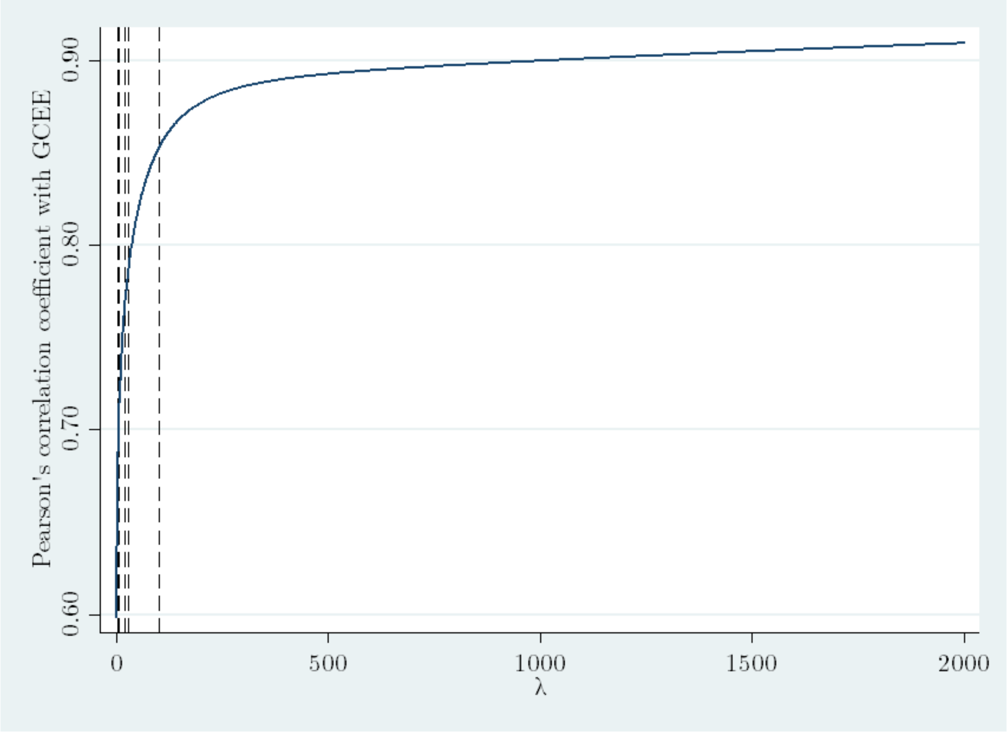

Figure A.3: Correlation coefficient of HP filter with the GCEE gap for varying smoothing factor λ. The vertical dashed lines mark, from the left: λ=2.91, λ=6.25, λ=20, λ=30, and λ=100.

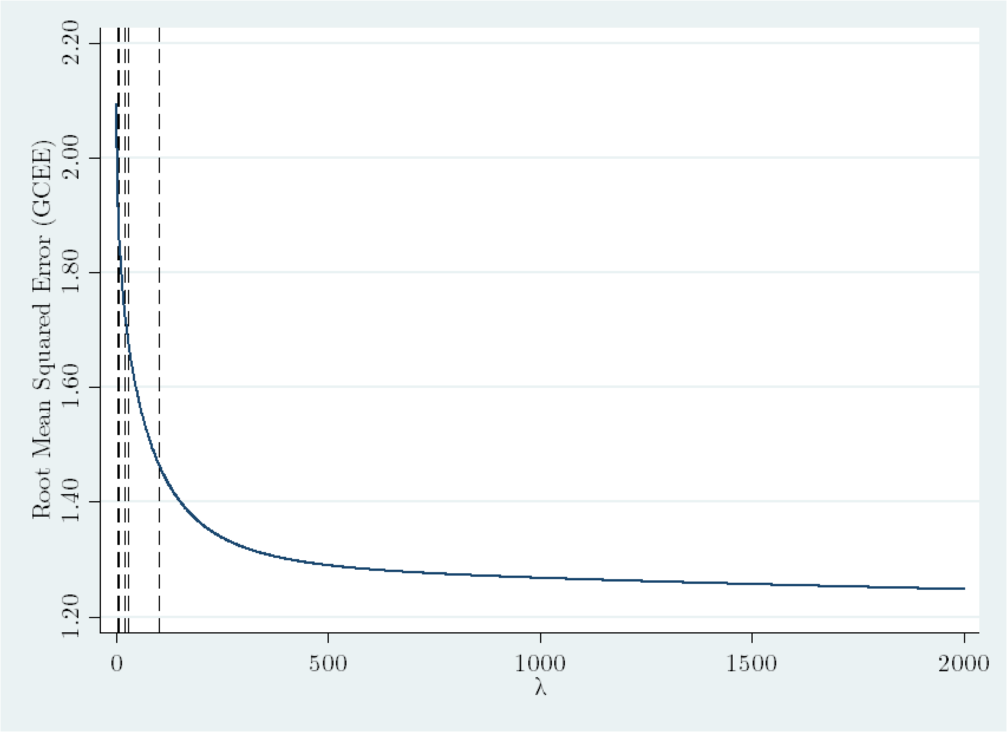

Figure A.4: Root Mean Squared Error of HP filter with GCEE gap for varying smoothing factor λ. The vertical dashed lines mark, from the left: λ=2.91, λ=6.25, λ=20, λ=30, and λ=100.