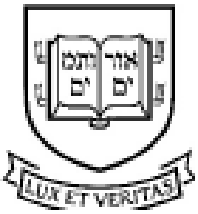
THE BOOSTED HP FILTER IS MORE GENERAL THAN YOU MIGHT THINK

By

Ziwei Mei, Peter C. B. Phillips and Zhentao Shi

September 2022

COWLES FOUNDATION DISCUSSION PAPER NO. 2348



COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.yale.edu/

The boosted HP filter is more general than you might think*

Ziwei Mei[a], Peter C. B. Phillips[b,c,d,e] and Zhentao Shi[a,f]

[a]The Chinese University of Hong Kong

[b]University of Auckland, [c]Yale University,

[d]Singapore Management University, [e]University of Southampton

[f]Georgia Institute of Technology

Abstract

The global financial crisis and Covid recession have renewed discussion concerning trend-cycle discovery in macroeconomic data, and boosting has recently upgraded the popular HP filter to a modern machine learning device suited to data-rich and rapid computational environments. This paper sheds light on its versatility in trend-cycle determination, explaining in a simple manner both HP filter smoothing and the consistency delivered by boosting for general trend detection. Applied to a universe of time series in FRED databases, boosting outperforms other methods in timely capturing downturns at crises and recoveries that follow. With its wide applicability the boosted HP filter is a useful automated machine learning addition to the macroeconometric toolkit.

Key words: Boosting, Business cycle, Machine learning, Macroeconomics, Recession

JEL codes: C22 Time Series Models, C55 Large Data Sets, C43 Index Numbers and Aggregation

*Phillips acknowledges research support from the NSF under Grant No. SES 18-50860 at Yale University and a Kelly Fellowship at the University of Auckland. Ziwei Mei: zwmei@link.cuhk.edu.hk; Peter C. B. Phillips: peter.phillips@yale.edu; Zhentao Shi: zhentao.shi@gatech.edu.

1 Introduction

Understanding long-term trends and short-term business cycles in economic activity was a primary pursuit of the foundational researchers of the econometrics profession especially during the years of the Great Depression where these matters figured prominently and prompted the development of new econometric approaches (Frisch, 1933; Tinbergen, 1939). At that time with very few exceptions data were scarce. By contrast vast datasets are now available to researchers covering multiple decades of quarterly and monthly time series observations across a wide range of economic variables. Long trajectories of data provide rich information about many aspects of economic activity and wellbeing, including the impact of technical progress on growth and the course and consequences of intermittent slowdowns and recessions. Analysis of the information carried in such time series provides useful indicators of the passage and present state of economic activity, which in turn helps to shape assessments of policymakers, regulators, corporate executives, and consumers in guiding decision making. The 2008 global financial crisis (GFC) and its aftermath and most recently the gobal economic impact of the Covid-19 pandemic are timely reminders that the work launched by Frisch (1933) and Tinbergen (1939) remains an ongoing mission for the econometrics community.

Modern econometric approaches often focus on decomposing time series observations of a variable y_t into additive components that represent long run trending behavior, f_t, and cyclical activity, c_t, as

$$y_t = f_t + c_t. \tag{1}$$

The trend embodies the long run general course or tendency in the data and the cycle reflects periodic fluctuations in economic activity in which businesses, labor markets and consumer behavior alternately expand and contract. Trends in many economic aggregates like a nation's real GDP are primarily determined by the impact of production technologies, the size and quality of labor forces, the accumulation of physical and human capital, and entrepreneurship, whereas business cycles are affected by shorter term internal and external influences including contractionary or expansionary fiscal, monetary, and political policies, combined with the dynamic propagation of these forces within an economy. As such, trend and cycle may be considered latent elements in the data which are to be identified and estimated by econometric methods through decomposition or direct modeling.

Since its introduction the Hodrick-Prescott (HP) filter (Hodrick and Prescott, 1997) has become a convenient and highly popular off-the-shelf choice for trend-cycle decomposition. Given an observed time series $y = (y_1, y_2, \cdots, y_n)'$, the HP filter finds $\widehat{f}^{\text{HP}} = (\widehat{f}_1^{\text{HP}}, \widehat{f}_2^{\text{HP}}, \ldots, \widehat{f}_n^{\text{HP}})'$ via a penalized least squares criterion

$$\widehat{f}^{\text{HP}} = \arg\min_{(f_t)} \left\{ \sum_{t=1}^{n} (y_t - f_t)^2 + \lambda \sum_{t=3}^{n} (\Delta^2 f_t)^2 \right\}, \tag{2}$$

where the second difference $\Delta^2 f_t = \Delta f_t - \Delta f_{t-1} = f_t - 2f_{t-1} + f_{t-2}$ of the trend component provides a measure of fluctuations, whose degree is controlled through the tuning parameter λ governing the extent of the penalty in the second component of the extremum criterion (2). Penalization plays

a key role in determining the outcome of the filter, with larger λ imposing a greater penalty on roughness thereby favoring smoothness in the outcome $\widehat{f}^{\mathrm{HP}}$.

Formally developed a century ago by Whittaker (1923) in pathbreaking work on penalized estimation, ideas for 'graduating' data have a long history in actuarial science and subsequent work in statistics and engineering, as reviewed in Phillips and Jin (2021, hereafter, PJ). When $\lambda \to \infty$ the solution for (2) satisfies $\Delta^2 f_t = 0$, giving a linear trend function $f_t = a + bt$ for some constant coefficients a and b. The conventional choice for quarterly data is $\lambda = 1600$, suggested by Hodrick and Prescott (1997) based on empirical experimentation with macroeconomic time series. Corresponding settings of λ for empirical work with monthly and annual data were given in Ravn and Uhlig (2002).

A poignant newspaper column posted by Krugman (2012) concerning the importance of long run trend identification during the GFC raised substantial interest on trend determination and the HP filter, spurring an influx of opinions, theory investigations, novel proposals, and empirical evidence. De Jong and Sakarya (2016), Cornea-Madeira (2017), and Sakarya and de Jong (2020) explored the HP filter's finite-sample algebraic properties. Arguing that the HP filter is a nonparametric procedure in which tuning parameters are typically sample size dependendent to achieve consistent estimation, PJ analyzed its asymptotic features via operator calculus, showing that the common setting $\lambda = 1600$ is too large to completely remove stochastic trends in time series of the length usually encountered in empirical work. Instead, the HP filter smooths those trends into paths that are asymptotically differentiable forms of Brownian motion. Several revised or modified versions based on the HP filter have recently emerged. For example, Yamada (2020) and Yamada (2022) generalize the HP filter to overcome data imperfections, and with the advent of Covid-19 Lee et al. (2021) use an L_1-type penalized HP filter to identify the turning points in the inflection rates of the virus.

Retaining the squared penalization scheme of the original HP filter, Phillips and Shi (2021, hereafter, PS) proposed a *boosted* HP filter (bHP hereafter) designed to upgrade the procedure to a machine learning device that uses the data more intensively to improve its properties and performance. The bHP filter is a repeated application of the HP filter to the residual extracted in the last iteration (see Algorithm 1 below) in which the number of iterations, m, controls the intensity of reusage. In practice, PS suggested monitoring a stopping criteria to terminate the iteration in a data-driven manner (see Algorithm 2), which makes bHP automated in application, as envisaged in Phillips (2005). In three empirical examples, PS fed 127 individual time series of various lengths and trending patterns through the bHP machine. The trend and cycle estimates returned after a few iterations in this algorithm largely confirmed and refined the existing findings in the literature for these series. Further simulation experiments and empirical applications (e.g., Hall and Thomson (2021, 2022)) have been conducted and the robustness continues to hold. These outcomes are indicative that, as a machine learning device which is agnostic about the data generating mechanism, the bHP filter satisfactorily accommodates trend processes that are much more general than the unit root processes studied in the asymptotic theory of PS. The main contribution of the present

paper is to provide analytic, simulation, and empirical support for this important extension.

In doing so we provide a preparatory result (Lemma 1 in Section 3.1) that characterizes the shrinkage effect of the operational form of the HP filter. We keep using the tuning parameter formula $\lambda = \mu n^4$ for some constant $\mu > 0$, which is an expansion rate extensively studied in PJ. For sample sizes that are typical in quarterly economic data, this expansion rate approximates well the actual form of the filter in practical work with the common choice $\lambda = 1600$ for the HP filter. In view of the two-sided nature of the filter, the HP operator is a function of lead and lag operators. For a class of complex numbers $a \in \mathbb{C}$ such that a^4 is a non-negative real number, Lemma 1 shows that the HP residual operator shrinks the complex exponential e^{ax} towards zero by the factor $\mu a^4/(\mu a^4+1) \in [0,1)$, which is a pseudo-differential operator extension of the elementary property $D_x^m \mathrm{e}^{ax} = a^m \mathrm{e}^{ax}$ of the usual differential operator $D_x = \mathrm{d}/\mathrm{d}x$. Repeating the operation m times gives the power factor $(\mu a^4/(\mu a^4+1))^m \mathrm{e}^{ax}$, which tends to zero as $m \to \infty$.

This lemma enables a unified development of asymptotics of the HP and bHP filters for a variety of nonstationary trend processes that include unit root $I(1)$ time series, higher order integrated $I(q)$ processes with integer $q \geq 2$, and local-to-unity (LUR) processes, which are among the most widely used models for nonstationary data. In each case, boosting enables consistent estimation of the trend whereas single implementation of the HP filter is inconsistent, producing a smoothed version of the original trend process. Upon standardization these trends all have asymptotic stochastic process representations in terms of convergent series of trigonometric functions with complex exponential forms that are amenable to analysis by operator methods and thereby deliver the respective asymptotic forms of the filter operation. Similar methods apply to general deterministic trend functions with convergent Fourier series representations. Taken together, the asymptotic results span a wide range of trend models commonly used in econometric practice.

In machine learning terminology the HP and bHP filters are *unsupervised learning* methods which seek to extract key features of the data but do not use regressors to fit the dependent variables or 'labels'. In contrast to this methodology, Hamilton (2018) firmly advocated that the HP filter should be replaced in empirical work by an autoregression (AR(p)), with specific order $p = 4$ recommended for quarterly data. Regressions of this type fall into the category of simple *supervised learning* methods in which a few lagged observations are trained to predict a future target. Unlike nonparametric approaches such as HP and bHP where tuning parameters are unavoidable and play a central role in consistent estimation, parametric autoregressions typically bypass tuning parameters, although users still need to decide on the number of lags (Quast and Wolters, 2022). Hamilton (2018)'s paper has stirred considerable discussion and debate. The issues raised bear directly on empirical econometric practice and they affect economic policy analysis in fundamental ways concerning the manner in which observed economic indicators can be intepreted as indicative of long term trend behavior as distinct from cyclical fluctuations, the very considerations that motivated Krugman (2012)'s public post.

Amongst this commentary, we draw attention to Schüler (2021) who pointed out that Hamilton's AR(4) regression filter fails to reproduce the standard chronology of US business cycles and

'emphasizes cycles that exceed the duration of regular business cycles (i.e., longer than 8 years), and completely mutes certain short-term fluctuations.' Cogley and Nason (1995) earlier pointed to some similar distortionary effects of HP filter estimates of cycles using frequency domain methods to assess the filter's transfer function and gain at various frequencies. Knight (2021) provided a new frequency domain analysis of both the HP and bHP filters, showing that the latter has some additional 'free pass' effects at low frequencies over that of the HP filter, giving it improved recovery properties for trends with frequencies in an interval around the origin. Recent empirical works by Drehmann and Yetman (2018), Hall and Thomson (2021) and Jönsson (2020) compared the HP filter and Hamilton's regression filter in real data examples and the accumulated empirical evidence from these studies favors the former. In subsequent work Hall and Thomson (2022) provided transfer function analysis and studied the empirical performance of the HP and bHP filters with various stopping rules, recommending a 'twicing' version of bHP with a second iteration (2HP) for trend and growth cycle analysis with New Zealand quarterly macroeconomic data.

PS provided a detailed response to Hamilton (2018)'s critiques of the HP filter and advocated the use of an automated bHP filter. The present paper adds further simulation and real data testimony concerning these two different approaches. We apply the HP-based methods and AR(p) to the open-source FRED-QD database (McCracken and Ng, 2020) and FRED-MD database (McCracken and Ng, 2016). Our empirical findings provide a fairly consistent message about the performance of the HP and bHP filters in relation to the AR filter. Both a small-scale test drive of the procedures on US real GDP and industrial production and a large-scale deployment to the entire databases are conducted. In brief, the HP filter is able to capture most historical business cycles although the fitted trends tend to be overly flattened or smoothed. The bHP filter is more adaptive to various generating mechanisms and patterns in the trend. On the other hand, AR regression filters typically seek to reduce observed series to martingale difference residuals and to offer useful mechanisms for prediction and impulse response analysis, but do not provide a method for identifying and estimating general trend and cycle processes, particularly those for which irregularity is a prominent feature, thereby failing to capture key trend and cycle elements of a macroeconomy.

bHP is an L_2-boosting method applied to time series, drawing on ideas of boosting by repetitive use in the computer science literature (Freund and Schapire, 1995) with statistical roots that go back to Tukey (1977)'s introduction of the *twicing* technique mentioned above. The key notion is to gradually 'boost' an ensemble of many weak learners into a more powerful fitting machine. Boosting has evolved into a very successful machine learning method, with many variants proposed along the way, for instance *adaboost* (Friedman et al., 2000), componentwise boosting (Bühlmann, 2006), and L_2-boosting (Bühlmann and Yu, 2003).

In addition to the above, some useful theoretical developments and applications of boosting have occurred in the econometric literature (Bai and Ng, 2009; Shi, 2016; Yousuf and Ng, 2021; Kueck et al., 2022). PS was the first paper to employ boosting methods in nonstationary time series, much of its asymptotic analysis being built on the foundation of functional limit theory (Phillips, 1986, 1987a,b). Careful interpretation of regression findings with nonstationary time series and panels

has relied on orthonormal series representations of limiting stochastic processes (Phillips, 1998). These methods offer an understanding of nonstationarity in terms of coordinate basis functions and have, in turn, proved useful in analyzing the asymptotic properties of both the HP and bHP filters. Recent years have also witnessed the proliferation in econometric work of machine learning methods, including cross-sectional studies (Belloni et al., 2014; Caner and Kock, 2018; Farrell et al., 2021; Athey et al., 2021), time series (Shi et al., 2023; Masini and Medeiros, 2021; Babii et al., 2021), and panel modeling (Su et al., 2016; Moon and Weidner, 2018), to name a few.

The present paper contributes to the literature in several ways by broadening our understanding of the capabilities of the boosted HP filter. It considerably expands the types of stochastic and deterministic trend mechanisms that bHP is able to consistently estimate. In doing so it complements ongoing work that considers the use of boosting methods for time series with long range dependence (Biswas et al., 2022). In addition, the present work provides further numerical evidence of the robustness and versatility of bHP in simulations and various real data applications. Overall, the findings reveal new analytic properties and empirical performance that support the boosted filter as a useful machine learning method for extracting low frequency components of macroeconomic time series, thereby contributing to the comprehension of trend phenomena.

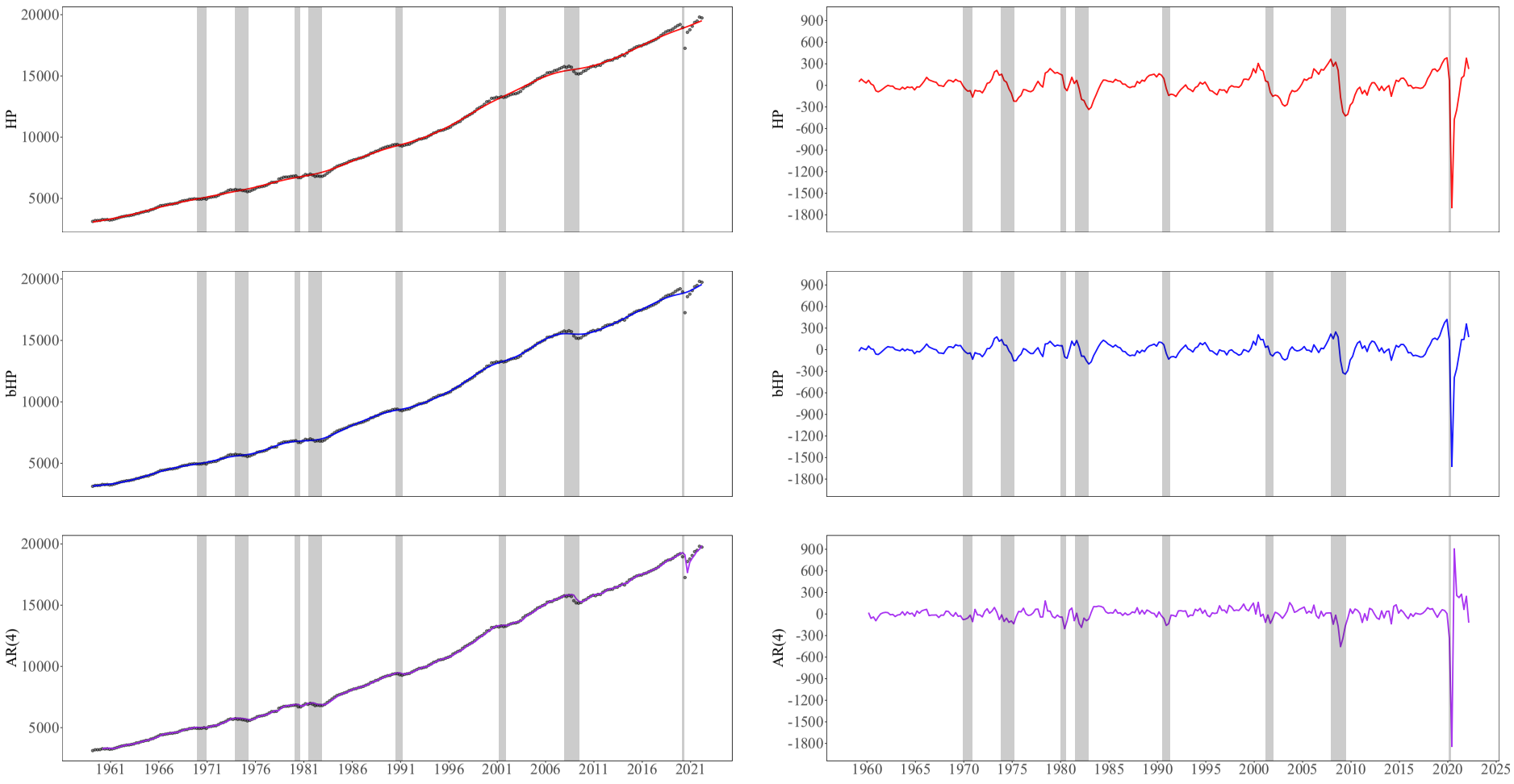
The rest of the paper is organized as follows. Section 2 gives an initial demonstration with HP, bHP and AR methods each applied to quarterly and monthly time series of two US aggregates. After formally introducing the HP and bHP filters and their operational forms, Section 3 presents the basic asymptotic approximation. It is followed in Section 4 by three applications of the theory to unit root, higher order integrated, and LUR time series on the smoothing properties of the usual HP filter and the consistency of bHP. The theory is supported by simulation exercises in Section 5. Section 6 provides a 'big data' empirical implementation of the methods that explores FRED-QD or -MD database to study trends and cycles in the US economy. Proofs and additional simulations are given in the Appendix.

2 An Empirical Appetizer

Prior to a formal analysis two small-scale applications are conducted to compare the methods discussed in the Introduction. At the time of writing, US real quarterly GDP in the FRED-QD database consists of 253 quarterly observations from 1959:Q1 to 2022:Q1. In Figure 1 the shaded time periods mark recessions identified by the National Bureau of Economic Research (NBER). The 2001 Internet bubble, the 2008 GFC, and the 2020 Covid-19 pandemic triggered the three recessions in the 21th century, and this subsample is shown in Figure 2 for clearer visualization.

In the left panels,[1] the black dots in the graphics display the raw data, and the colored solid lines are the trends as estimated by the HP filter (red line, top), bHP (blue line, middle), and AR(4)

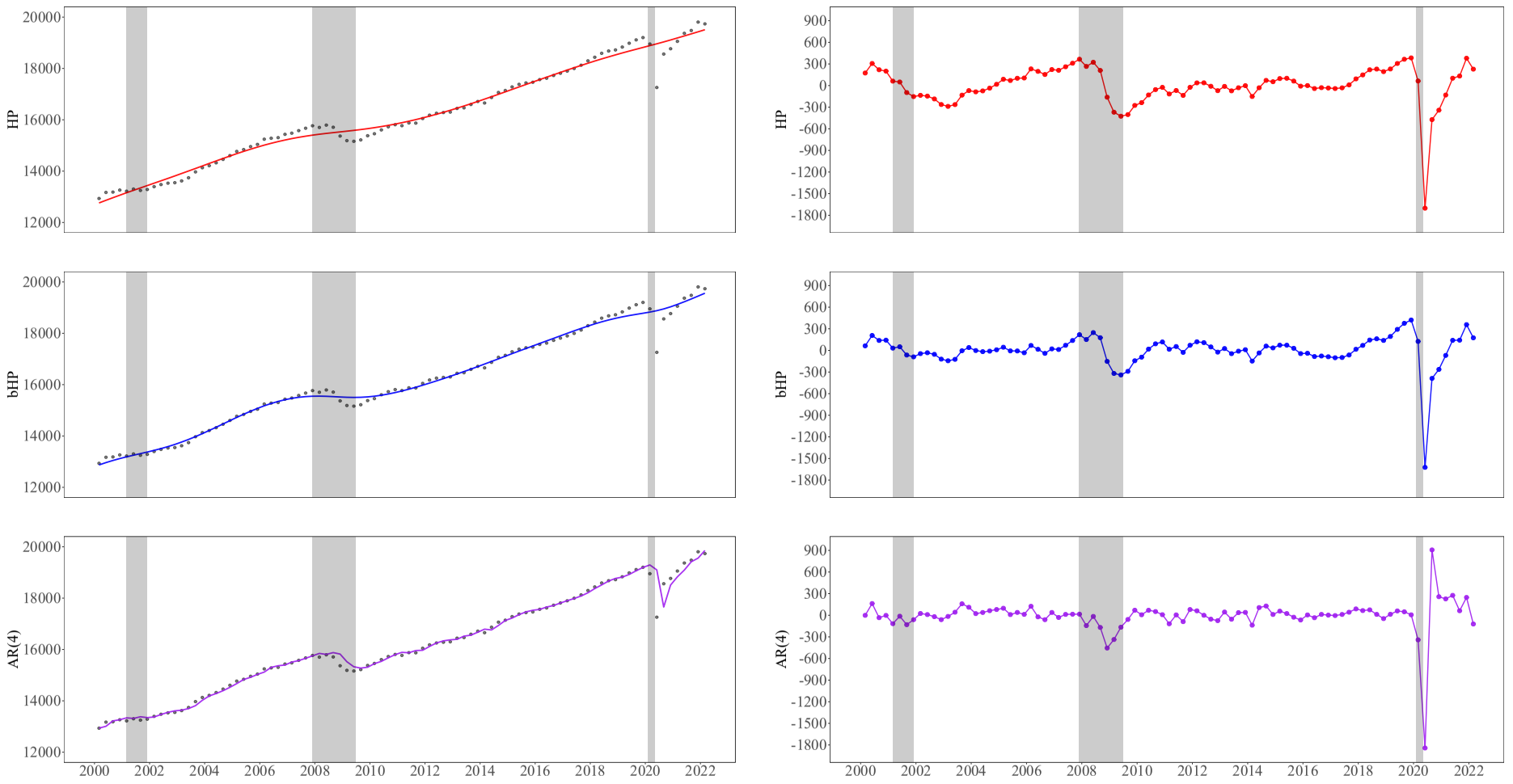
[1]Figures 1 and 2 are left-rotated 90° to landscape mode in the paper. So the left (right) panel of the paper appears in the lower (upper) position in the rotated display. Similarly, the top, center and bottom panels refer to the unrotated graphics, whereas they are shown in 90° rotated view in the paper. The descriptions in the text refer to the original unrotated orientation of the graphics.



(a) Estimated Trends

(b) Estimated Cycles

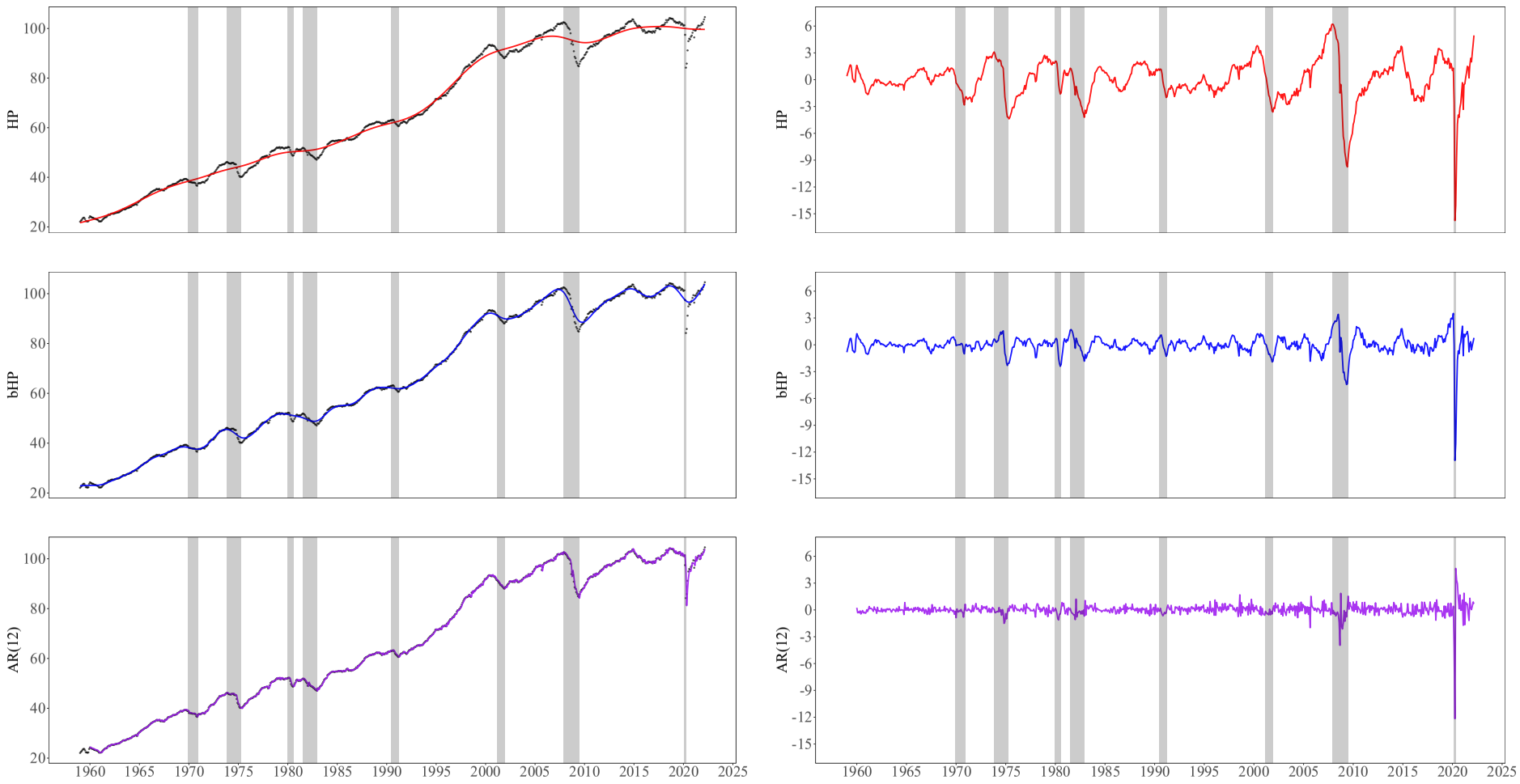
Figure 1: Trends, Cycles, and Raw Data of US Real GDP



(a) Estimated Trends

(b) Estimated Cycles

Figure 2: Zoomed-in version of Figure 1 after 2000



(a) Estimated Trends

(b) Estimated Cycles

Figure 3: Trends, Cycles, and Raw Data of US Monthly Industrial Production Index

(violet line, lower).[2] The HP filter produces a smooth, monotonically increasing trend. bHP is more responsive with a slight dip during the long recession in 2008–09, leaving the residual to capture that long cycle; and the trends from HP do not decline in the short Covid-19 recession despite the enormous single period drop in the raw data in 2020:Q2, again leaving this dynamic to the cycle. By contrast, the AR(4) yields a trend that traces the observed data closely, with a major fall in 2020:Q3 (a one period delay following the actual drop in the raw data), and then proceeds with an immediate recovery that again follows the data closely.

More revealing are the estimated cycles shown in the right panels, which measure the gap between the raw data and the estimated trend. Residuals from HP and bHP exhibit clear business cycles that capture the trough in all recessions. The GFC was the most severe prior to 2020, but it was dwarfed in magnitude by the Covid-19 recession, with its two-year recovery period. The cyclical component produced by AR(4) is in many cases counterintuitive. It shows that GDP usually recovered in the middle of many recessions, and after the colossal slump in 2020:Q2 it returned to an unprecedented historical peak in 2020:Q3. This reversion is due to the preceding major downturn which, with a typical autoregressive one-period delay in the AR(4) fit, is translated into a major upturn in the next quarter. Wild swings of adjustment in two consecutive quarters are symptomatic of compensating under-prediction and over-prediction that can occur with autoregressive data fits, which in turn accentuate the swings in actual economic activity.

As no monthly GDP is reported in FRED-MD, to cross-check the above empirical findings we performed a parallel analysis with the US monthly industrial production index from FRED-MD, consisting of 759 monthly observations from 1959:1 to 2022:3. Figure 3 delivers a similar picture. In the left-lower panel, estimated cycles obtained from an AR(12) — one year's past information as equivalent to AR(4) for quarterly data — barely differ between recessions and normal periods, and the dramatic U-turn in 2020 takes only two months. By contrast both HP and bHP characterize the state of the US economy in a manner that is more in line with the NBER recession assessments and broad public, official and media perceptions during those periods.

In these two empirical examples, the HP and bHP filtered GDP data appear relatively smooth whereas the the industrial production index shows more evidence of stochastic trend behavior and this is correspondingly more pronounced in the filtered series. Our later empirical analysis employs 246 quarterly series in the FRED-QD database and 127 series in the FREQ-MD database. This broad range of time series includes a mixture of trend behaviors. The next section develops filter asymptotic theory to assist in understanding their potential for capturing such phenomena and to shed light on possible finite sample properties.

[2]The 2HP 'twicing' filter was also calculated but the results are not displayed in the graphs as they are very close to those of HP.

3 Smoothing effects of the HP filter

The squared penalty in the HP filter (2) leads to the explicit solution

$$\hat{f}^{\text{HP}} = Sy,$$

where $S = (I_n + \lambda D_n D_n')^{-1}$, D_n' is the rectangular $(n-2) \times n$ matrix with the second differencing vector $d = (1, -2, 1)'$ along the leading tri-diagonals and I_n is the $n \times n$ identity matrix. The fitted cyclical component (residual) is $\hat{c}^{\text{HP}} = (\hat{c}_1^{\text{HP}}, \hat{c}_2^{\text{HP}}, \ldots, \hat{c}_n^{\text{HP}})' = y - \hat{f}^{\text{HP}}$. These compact expressions are useful in studying the effects of boosting.

Algorithm 1. *The boosted HP filter (in m iterations)*

Step1 *Specify the smoothing parameter $\lambda > 0$ and the number of iterations $m \geq 2$. Set $\hat{c}^{(1)} = \hat{c}^{\text{HP}}$ and $j = 1$.*

Step2 *If $j < m$, set $\widehat{c}^{(j+1)} = (I_n - S)\widehat{c}^{(j)}$ and then update $j = j + 1$.*

Step3 *Repeat Step2 until $j = m$. Save the fitted cyclical component $\widehat{c}^{(m)}$ and the estimated trend $\widehat{f}^{(m)} = y - \widehat{c}^{(m)}$.*

The recursive form of the above algorithm is simply $\widehat{c}^{(m)} = (I_n - S)^m y$ and $\widehat{f}^{(m)} = B_m y$, where $B_m := I_n - (I_n - S)^m$. The bHP filter is easy to implement. For instance, if the original HP filter is called by the `hpfilter` function in the `R` package `mFilter`, then bHP with m iterations can be carried out with a single line of code in the following `tidyverse` style

```
purrr::rerun(m, y %<>% mFilter::hpfilter(lambda, "lambda") %>% .$cycle)
```

where `y` is the observed time series, `m` controls the number of iterations, and `lambda` sets the tuning parameter.

3.1 Shrinkage Effects

With tuning parameter $\lambda = \mu n^4$ for some constant $\mu > 0$ and lag operator $\mathbb{L}$, the asymptotic approximation (as $n \to \infty$) of the HP filtered trend has the operator form

$$G_\lambda = \frac{1}{\lambda \mathbb{L}^{-2}(1-\mathbb{L})^4 + 1} = \frac{1}{\mu \mathbb{L}^{-2}\left[n(1-\mathbb{L})\right]^4 + 1},$$

and the corresponding residual operator is $1 - G_\lambda = \dfrac{\mu \mathbb{L}^{-2}\left[n(1-\mathbb{L})\right]^4}{\mu \mathbb{L}^{-2}\left[n(1-\mathbb{L})\right]^4 + 1}$. The estimated trends of the HP and bHP (m iterations) filters then have the asymptotic forms $\widehat{f}_t^{\text{HP}} = G_\lambda y_t$ and $\widehat{f}_t^{(m)} = \left[1 - (1 - G_\lambda)^m\right] y_t$, respectively.

The following lemma reveals the effect of the operator when it is applied to a class of complex exponential functions of the form $\exp(at/n)$. We introduce two notations. For a real number

$x > 1$, define a set $\mathcal{A}(x) := \left\{a \in \mathbb{C} : a^4 \in [0, (\log x)^2]\right\}$, and for two real numbers x and y, let $x \wedge y := \min\{x, y\}$.

Lemma 1. *Let $\lambda = \mu n^4$ for some real constant $\mu > 0$ and $t \leq n$ be a positive integer.*

(a) For any fixed $m \in \mathbb{N} := \{1, 2, \ldots\}$, as $n \to \infty$ we have

$$\sup_{1 \leq t \leq n,\, a \in \mathcal{A}(n)} \left| \left[(1 - G_\lambda)^m - \left(\frac{\mu a^4}{\mu a^4 + 1} \right)^m \right] \mathrm{e}^{at/n} \right| \to 0. \tag{3}$$

(b) As $m, n \to \infty$ we have

$$\sup_{1 \leq t \leq n,\, a \in \mathcal{A}(n \wedge m)} \left| (1 - G_\lambda)^m \, \mathrm{e}^{at/n} \right| \to 0.$$

Remark 1. The set $\mathcal{A}(\cdot)$ is constructed to ensure that a is either a real number θ or a purely imaginary number $\theta\mathbf{i}$ for some $\theta \in \mathbb{R}$, where $\mathbf{i} := \sqrt{-1}$ is the imaginary unit. Lemma 1 (a) shows that the operator $(1 - G_\lambda)$ works as if $\mathrm{e}^{at/n}$ is multiplied by a real factor $\frac{\mu a^4}{\mu a^4 + 1} \in [0, 1)$ at each operation. For a finite m, the set $\mathcal{A}(n)$ bounds $|a^2| = \theta^2 \leq \log n$. This upper bound, diverging sufficiently slowly, guarantees the asymptotic validity of (3) as $n \to \infty$. Part (b) extends this approximation to include large $m \to \infty$, so that the factor $\left(\frac{\mu a^4}{\mu a^4 + 1}\right)^m$ vanishes asymptotically. When m passes to infinity, the effect of the iterations of the operator is controlled by $\mathcal{A}(n \wedge m)$ in which $\theta^2 \in [0, \log(n \wedge m)]$. The expansion rates of $\mathcal{A}(n)$ and $\mathcal{A}(n \wedge m)$ are devised (for finite m and large m, respectively) to regularize the domains over a where uniform convergence is achievable.

Lemma 1 shows the asymptotic effect of the operator $(1 - G_\lambda)^m$ on a simple exponential function. PJ use the exponential function as an intermediate step in studying the effect of the HP operator involving $(1 - G_\lambda)$. In the numerator of the residual operator $(1 - G_\lambda)$ the scaled differencing operator $[n(1 - \mathbb{L})]$ acts like a differential operator and the lag operator $\mathbb{L}^{-1}$ acts like the identity, asymptotically when $n \to \infty$, so the effect of these operations applied to $\mathrm{e}^{at/n}$ is straightforward. The HP operator involves these elementary operators in the nonlinear operator $[\mu \mathbb{L}^{-2} [n(1 - \mathbb{L})]^4 + 1]^{-1}$ and PJ showed how this complex operator may be analyzed as a pseudo-differential operator. Using this machinery PS studied repeated applications of the operator to the trigonometric functions present in the Karhunen-Loève (KL) representation of Brownian motion — see (5) below. This approach is used in developing the asymptotic analysis in the next section.

4 Applications

This section illustrates the use of Lemma 1 for trends that involve unit roots, higher order integrated time series and local unit root time series.

4.1 Unit Root Processes

To demonstrate how the bHP filter moderates the residual component in the trend fitting process, we begin with simple unit root time series that was fully analyzed in PS. The following discussion is

heuristic to reveal the manner in which the moderation operates. Consider a time series x_t generated as an $I(1)$ stochastic trend from a unit root model $x_t = x_{t-1} + u_t$ from initialization $x_0 = o_p(\sqrt{n})$, for which the normalized series satisfies the functional law (Phillips and Solo, 1992)

$$X_n(\cdot) := n^{-1/2} x_{t=\lfloor n\cdot \rfloor} \rightsquigarrow B(\cdot), \tag{4}$$

where $B(\cdot) = BM(\omega^2)$ is a Brownian motion with (long run) variance ω^2, and $\lfloor \cdot \rfloor$ is the integer floor function. The KL representation of this Brownian motion over the interval [0,1] is

$$B(r) = \sqrt{2} \sum_{k=1}^{\infty} \frac{\sin\left[(k-\frac{1}{2})\pi r\right]}{(k-\frac{1}{2})\pi} \xi_k = \sum_{k=1}^{\infty} \sqrt{\lambda_k} \varphi_k(r) \xi_k, \tag{5}$$

where $\xi_k \sim i.i.d.\ N(0, \omega^2)$ are the random coefficients, $\lambda_k = 1/[(k-\frac{1}{2})\pi]^2$ are the eigenvalues, and $\{\varphi_k(r) = \sqrt{2}\sin[(k-\frac{1}{2})\pi r] = \sqrt{2}\sin(r/\sqrt{\lambda_k})\}_{k=1}^{\infty}$ is an orthonormal system of corresponding eigenfunctions in $L_2[0,1]$. The series (5) converges almost surely and uniformly over $r \in [0,1]$.

When the innovation u_t follows a linear process as in

$$u_t = C(L)\varepsilon_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j}, \quad \sum_{j=0}^{\infty} j\,|c_j| < \infty, \quad C(1) \neq 0 \tag{6}$$

with $\varepsilon_t = i.i.d.\left(0, \sigma_\varepsilon^2\right)$ and $E\left(|\varepsilon_t|^p\right) < \infty$ for some $p > 4$, we construct an expanded probability space with a Brownian motion $B(\cdot)$ for which uniform convergence holds almost surely (Phillips, 2007, Lemma 3.1), viz.,

$$\sup_{0 \le t \le n} \left| \frac{x_t}{\sqrt{n}} - B\left(\frac{t}{n}\right) \right| = o_{a.s.}\left(\frac{1}{n^{1/2-1/p}}\right). \tag{7}$$

In this space the convergence (4) takes the strong form

$$n^{-1/2} x_{\lfloor nr \rfloor} - B(r) = o_{a.s.}(1).$$

In what follows and unless otherwise stated, we assume that we are working in this expanded probability space. In the original space the results translate, as usual, into weak convergence mirroring (4).

Write $\varphi_k(t/n) = \sqrt{2}\sin(\frac{t/n}{\sqrt{\lambda_k}}) = \sqrt{2}\,\mathrm{Im}\left[\exp(\frac{\mathbf{i}t/n}{\sqrt{\lambda_k}})\right]$, where $\mathrm{Im}[\cdot]$ gives the imaginary part of the argument. When the operator $(1-G_\lambda)^m$ is applied to the kth term of (5) for any fixed k, Lemma 1 gives

$$(1-G_\lambda)^m \,\mathrm{Im}\left[\mathrm{e}^{\frac{\mathbf{i}t/n}{\sqrt{\lambda_k}}}\right] \approx \left[\frac{\mu}{\mu+\lambda_k^2}\right]^m \mathrm{Im}\left[\mathrm{e}^{\frac{\mathbf{i}t/n}{\sqrt{\lambda_k}}}\right] = \left[\frac{\mu}{\mu+\lambda_k^2}\right]^m \sin\left(\frac{t/n}{\sqrt{\lambda_k}}\right), \tag{8}$$

so that when m is large we have

$$(1-G_\lambda)^m \sin\left(\frac{t/n}{\sqrt{\lambda_k}}\right) \approx \exp\left(-\frac{m\lambda_k^2}{\mu+\lambda_k^2}\right) \sin\left(\frac{t/n}{\sqrt{\lambda_k}}\right) \to 0, \tag{9}$$

as m and n pass to infinity. With careful handling of a finite-term approximation to the infinite series in (5), PS showed that when $\lambda = \mu n^4$ the residual of the bHP filter becomes

$$n^{-1/2}\widehat{c}^{(m)}_{\lfloor nr \rfloor} \approx (1 - G_\lambda)^m \, n^{-1/2} x_{t=\lfloor nr \rfloor} \rightsquigarrow 0,$$

thereby recovering the consistency of the trend component $n^{-1/2}\widehat{f}^{(m)}_{\lfloor nr \rfloor} \rightsquigarrow B(r)$, as there is no cycle component in the unit root model.

4.2 Higher Order Integrated Processes

Many other types of nonstationary time series besides $I(1)$ processes occur in macroeconomic data. For instance, aggregate measures of the money supply and nominal price series are often well modeled by higher order integrated time series, particularly by $I(2)$ processes (Johansen, 1995; Haldrup, 1998). The KL series of the limiting Brownian motion process involves orthonormal series of sine functions but it is equally clear from (8) and (9) that similar shrinking factors apply to cosine series and more general trigonometric polynomial functions. Such functions figure in series representations of higher order integrated processes.

Suppose the observed time series y_t follows a higher order integrated process $I(q)$, for some integer $q \in \{2, 3, \ldots\}$, of the form

$$(1 - \mathbb{L})^q \, y_t = u_t,$$

where u_t is a linear process satisfying (6). The repeated summation form of y_t from initialization at $t = 0$ gives $y_t = \sum_{j_q=1}^{t} \sum_{j_{q-1}=1}^{j_q} \cdots \sum_{j_1=1}^{j_2} u_{j_1} + p_{q-1}(t)$ where $p_{q-1}(t)$ is a polynomial in t of degree $q-1$ with constant coefficients. Standard weak convergence methods lead to the following limit process after rescaling

$$\frac{y_{t=\lfloor n \cdot \rfloor}}{n^{q-0.5}} \rightsquigarrow B_q(\cdot) := \int_0^{\cdot} \int_0^{s_{q-1}} \cdots \int_0^{s_3} \int_0^{s_2} B(s_1) ds_1 ds_2 \cdots ds_{q-2} ds_{q-1}, \tag{10}$$

as $n \to \infty$. Uniform convergence in (7) ensures that in the expanded probability space we have the corresponding result for $y_t / n^{q-0.5}$, viz.,

$$\sup_{0 \le t \le n} \left| \frac{y_t}{n^{q-0.5}} - B_q\left(\frac{t}{n}\right) \right| = o_{a.s.}(1). \tag{11}$$

The orthonormal series representation of $B_q(r)$ is obtained by termwise integration in view of the uniform and almost sure convergence of the KL series for Brownian motion in (5), giving

$$\begin{aligned} B_q(r) &= \sum_{k=1}^{\infty} \xi_k \sqrt{\lambda_k} \int_0^r \int_0^{s_{q-1}} \cdots \int_0^{s_3} \int_0^{s_2} \varphi_k(s_1) ds_1 ds_2 \cdots ds_{q-2} ds_{q-1} \\ &= \sqrt{2} \sum_{k=1}^{\infty} \left\{ \sum_{\ell=1}^{\lfloor q/2 \rfloor} (-1)^{\ell-1} \lambda_k^{\ell} \frac{r^{q-2\ell}}{(q-2\ell)!} + \lambda_k^{q/2} \mathrm{Im}\left[(-\mathbf{i})^{q-1} \, \mathrm{e}^{(\mathbf{i}/\sqrt{\lambda_k})r} \right] \right\} \xi_k. \end{aligned} \tag{12}$$

The braces in (12) includes two terms: $\sum_{\ell=1}^{\lfloor q/2\rfloor}(-1)^{\ell-1}\lambda_k^{\ell}r^{q-2\ell}/(q-2\ell)!$ is a $(q-2)$th order polynomial, and $\lambda_k^{q/2}\mathrm{Im}\left[(-\mathbf{i})^{q-1}\,\mathrm{e}^{(\mathbf{i}/\sqrt{\lambda_k})r}\right]$ alternates between a sine (for odd q) and a cosine (for even q).

Example 1. Setting $q=2$ makes y_t an $I(2)$ process. Rescaling by $n^{3/2}$ and letting $n\to\infty$ we have

$$Y_n(\cdot)=\frac{y_{t=\lfloor n\cdot\rfloor}}{n^{3/2}}\rightsquigarrow B_2(\cdot)\equiv\int_0^{\cdot}B(s)ds,\quad\text{with} \tag{13}$$

$$B_2(r)=\sqrt{2}\sum_{k=1}^{\infty}\frac{1-\cos\left[(k-\frac{1}{2})\pi r\right]}{[(k-\frac{1}{2})\pi]^2}\xi_k=\sum_{k=1}^{\infty}\lambda_k[\sqrt{2}-\psi_k(r)]\xi_k, \tag{14}$$

where $\{\psi_k(r)=\sqrt{2}\cos\left[(k-\frac{1}{2})\pi r\right]=\sqrt{2}\cos\left[r/\sqrt{\lambda_k}\right]\}_{k=1}^{\infty}$ is an orthonormal system of cosine series on $L_2[0,1]$. The series (14) of $B_2(r)$ converges faster than that of $B(r)$ since the decay rate of the coefficients λ_k exceeds that of $\sqrt{\lambda_k}$ as $k\to\infty$. Correspondingly, $B_2(r)$ is a smooth (once differentiable) Gaussian stochastic process in contrast to the Brownian motion $B(r)$. The first derivative of $B_2(r)$ is $B(r)$ and, like B, B_2 has a zero initial value at the origin.

It was long held as conventional wisdom that the HP filter removed up to four unit roots, thereby detrending integrated processes up to the fourth order (King and Rebelo, 1993). This claim was disproved by PJ for $I(1)$ processes under the expansion rate $\lambda=\mu n^4$, which was shown to match common quarterly time series applications in macroeconomics. The next result does the same for $I(2)$ processes, showing that the smoothing property of the HP filter continues to apply in this case as $n\to\infty$, giving a limit representation that is a smoothed version of $B_2(r)$ rather than $B_2\,(r)$ itself.

Proposition 1. *If y_t satisfies the functional law (13) and $\lambda=\mu n^4$, then the HP filtered series has the following limit form in the extended probability space as $n\to\infty$:*

$$\frac{\hat{f}^{\mathrm{HP}}_{\lfloor nr\rfloor}}{n^{3/2}}\to_{a.s.}\sum_{k=1}^{\infty}\lambda_k\left[\sqrt{2}-\frac{\lambda_k^2}{\mu+\lambda_k^2}\psi_k\,(r)\right]\xi_k=:f^{\mathrm{HP}}(r). \tag{15}$$

Under $\lambda=\mu n^4$, Proposition 1 shows that the HP filtered $I(2)$ trend approaches a limiting stochastic process $f^{\mathrm{HP}}(r)$ which deviates from the limiting trend process $B_2(r)$ and is therefore inconsistent for this expansion rate of λ. Correspondingly, the estimated $\widehat{c}_t^{\mathrm{HP}}$ of the cycle component c_t has the following limiting functional form

$$\frac{\hat{c}^{\mathrm{HP}}_{\lfloor nr\rfloor}}{n^{3/2}}=n^{-3/2}\left(y_{\lfloor nr\rfloor}-\hat{f}^{\mathrm{HP}}_{\lfloor nr\rfloor}\right)\to_{a.s.}\sum_{k=1}^{\infty}\frac{\mu\lambda_k}{\mu+\lambda_k^2}\psi_k\,(r)\,\xi_k=:c^{\mathrm{HP}}(r) \tag{16}$$

upon standardization in the expanded space. This limit function is a stochastic process that inherits some of the stochastic trend properties of the limiting process $B_2(r)$. It is therefore to be expected that with a smoothing parameter that approximates $\lambda=\mu n^4$, the HP filter will fail to remove all the trend properties of the $I(2)$ process and the imputed business cycle estimate $\hat{c}_t^{\mathrm{HP}}$ will inevitably

carry some of these 'spurious' characteristics. Note also that at the origin the filtered series limit function is $f^{\text{HP}}(0) = \sum_{k=1}^{\infty} \frac{\mu\lambda_k}{\mu+\lambda_k^2}\psi_k(0)\xi_k = \sqrt{2}\mu\sum_{k=1}^{\infty}\frac{\lambda_k}{\mu+\lambda_k^2}\xi_k = c^{\text{HP}}(0) \neq 0$, a random, mean zero, initialization different from $B_2(0) = 0$.

Remark 2. Given the fact $\lambda_k \asymp 1/k^2$,[3] the coefficients in the series representation (15) satisfy $\lambda_k^3/(\mu+\lambda_k^2) \asymp k^{-6}$, from which we deduce that the limit process in (15) is a Gaussian stochastic process that is continuously differentiable to the fifth order, with fifth derivative

$$[f^{\text{HP}}(r)]^{(5)} = \sum_{k=1}^{\infty}\left[\frac{\sqrt{\lambda_k}}{\mu+\lambda_k^2}\varphi_k(r)\right]\xi_k,$$

which is a non-differentiable Gaussian process for all $\mu > 0$ similar to the Brownian motion $B(r) = \sum_{k=1}^{\infty}\sqrt{\lambda_k}\varphi_k(r)\xi_k$. Thus, the trend extracted by the HP filter when $\lambda = \mu n^4$ is a very smooth function.

The inconsistency of the HP filter estimate of $B_2(r)$ in (15) is anticipated in view of PJ's earlier findings for HP filtering of an $I(1)$ stochastic trend. The next result shows that boosting restores consistency to the HP filter for $I(2)$ and higher order integrated time series. For practical applications, the results for $I(2)$ time series are clearly the most relevant.

Theorem 1. *Suppose that y_t satisfies the functional law (10). Given $\lambda = \mu n^4$, the bHP filter has the following standardized limit theory*

$$n^{0.5-q}\hat{f}^{(m)}_{\lfloor nr\rfloor} \rightsquigarrow B_q(r)$$

for all positive integers $q \in \mathbb{N}$ as $m, n \to \infty$.

When $q = 1$ this result includes the unit root $I(1)$ case with $B_1 = B$. The generalization for $q \geq 2$ follows by use of Lemma 1 and the asymptotic representation of repeated applications of the HP operator on the exponential function.

4.3 Local Unit Root Processes

While models with unit roots provide a prototypical framework for capturing persistence in time series data, these models have modifications designed to capture a wider class of time series behavior in which the autoregressive roots are not restricted to unity as they are with integrated processes. An important subclass of more general models with near unit roots (Phillips, 2023) is the class of LUR models

$$(1 - \mathrm{e}^{c/n}\mathbb{L})\, y_t = u_t, \;\; t = 1, 2, \cdots, n, \text{ with } y_0 = o_p(\sqrt{n}), \tag{17}$$

in which the autoregressive root $\mathrm{e}^{c/n} \approx 1 + c/n$ is local to unity for some constant $c \in \mathbb{R}$ and large n. These models were developed in (Phillips, 1987b; Chan and Wei, 1987) and have been used for

[3] For any two positive sequences a_n and b_n, we use $a_n \asymp b_n$ to signify that $C^{-1}b_n \leq a_n \leq Cb_n$ for some positive constant $C \in (1, \infty)$ as n is sufficiently large.

power analyses and in empirical research to provide robustness against pure unit root specifications. Time series generated by (17) are nonstationary and, after suitable standardization, y_t converges to a linear diffusion, or Ornstein-Uhlenbeck (OU), process

$$Y_n(r) := \frac{y_{\lfloor nr \rfloor}}{\sqrt{n}} = \int_0^r \mathrm{e}^{(r-s)c} dX_n(s) + O(n^{-1/2}) \rightsquigarrow J_c(r) := \int_0^r \mathrm{e}^{(r-s)c} dB(s), \tag{18}$$

as $n \to \infty$, where $B(\cdot)$ is Brownian motion with variance ω^2 as in (4). When $c < 0$ the limiting OU process J_c is stationary and mean-reverting; when $c > 0$ the process is explosive (Phillips, 1987b; Phillips and Magdalinos, 2007).

Using Lemma 3.1 of Phillips (2007), the uniform convergence law (7) holds, ensuring that

$$\sup_{0 \le t \le n} |Y_n(t/n) - J_c(t/n)| = o_{a.s.}(1) \tag{19}$$

in the expanded probability space. A convenient series representation of $J_c(r)$ is

$$J_c(r) = \sum_{k=1}^{\infty} \frac{1}{\lambda_k c^2 + 1} \left[\sqrt{\lambda_k} \varphi_k(r) + \sqrt{2} c \lambda_k \mathrm{e}^{cr} - c \lambda_k \psi_k(r) \right] \xi_k, \tag{20}$$

as derived by (A.8) in the Appendix. Note that the first component of (20) has the form

$$\sum_{k=1}^{\infty} \frac{\sqrt{\lambda_k}}{\lambda_k c^2 + 1} \varphi_k(r) \xi_k = \sum_{k=1}^{\infty} \left(\sqrt{\lambda_k} - \frac{\lambda_k^{3/2} c^2}{\lambda_k c^2 + 1} \right) \varphi_k(r) \xi_k$$

in which $\sum_{k=1}^{\infty} \sqrt{\lambda_k} \varphi_k(r) \xi_k = B(r)$, corresponding to the leading (non-differentiable) Brownian motion component in the decomposition $J_c(r) = B(r) + c \int_0^r \mathrm{e}^{(r-s)c} B(s) ds$. The remaining terms of (20) provide a series representation of the smooth component $c \int_0^r \mathrm{e}^{(r-s)c} B(s) ds$ of $J_c(r)$, one of which is the exponential term $\sqrt{2} c \nu \mathrm{e}^{cr}$ with a random Gaussian coefficient $\nu := \sum_{k=1}^{\infty} \frac{\lambda_k}{1 + c^2 \lambda_k} \xi_k \sim N\left(0, \sigma_\nu^2\right)$, where $\sigma_\nu^2 = \omega^2 \sum_{k=1}^{\infty} \left(\frac{\lambda_k}{1 + c^2 \lambda_k} \right)^2$.

The following proposition shows that the limit representation of the HP filtered LUR time series is inconsistent, yielding a smoothed version of the diffusion $J_c(r)$.

Proposition 2. *If y_t satisfies the functional law (18) and $\lambda = \mu n^4$, then the HP filtered series has the following limiting form as $n \to \infty$:*

$$\frac{\hat{f}^{HP}_{\lfloor nr \rfloor}}{n^{1/2}} \to_{a.s.} \sum_{k=1}^{\infty} \frac{1}{\lambda_k c^2 + 1} \left[\frac{\sqrt{2} c \lambda_k \mathrm{e}^{cr}}{\mu c^4 + 1} + \frac{\lambda_k^2}{\mu + \lambda_k^2} \left(\sqrt{\lambda_k} \varphi_k(r) - c \lambda_k \psi_k(r) \right) \right] \xi_k =: f^{\mathrm{HP}}_{\mathrm{LUR}}(r). \tag{21}$$

Remark 3. Since $\lambda_k \asymp 1/k^2$, the coefficients associated with the sine and cosine waves in (21) satisfy

$$\frac{\lambda_k^{5/2}}{(\lambda_k c^2 + 1)(\mu + \lambda_k^2)} \asymp \frac{1}{k^5}, \quad \frac{c \lambda_k^3}{(\lambda_k c^2 + 1)(\mu + \lambda_k^2)} \asymp \frac{1}{k^6},$$

respectively. The real exponential function component $\sqrt{2}c\mathrm{e}^{cr}\sum_{k=1}^{\infty}\frac{\lambda_k\xi_k}{(\lambda_k c^2+1)(\mu c^4+1)}$ has a random coefficient and is infinitely differentiable. The limit process is therefore a Gaussian stochastic process continuously differentiable to the fourth order with the fourth derivative given by

$$[f_{\mathrm{LUR}}^{HP}(r)]^{(4)}=\sum_{k=1}^{\infty}\frac{1}{\lambda_k c^2+1}\left[\frac{\sqrt{2}c^5\lambda_k\mathrm{e}^{cr}}{\mu c^4+1}+\frac{1}{\mu+\lambda_k^2}\left(\sqrt{\lambda_k}\varphi_k(r)-c\lambda_k\psi_k(r)\right)\right]\xi_k,$$

which is a convergent series.

Remark 4. The limits of the HP estimated trend (21) and the HP estimated cycle

$$\frac{\hat{c}_{\lfloor nr\rfloor}^{\mathrm{HP}}}{n^{1/2}}\to_{a.s.}\sum_{k=1}^{\infty}\frac{1}{\lambda_k c^2+1}\left[\frac{\sqrt{2}\mu c^5\lambda_k\mathrm{e}^{cr}}{\mu c^4+1}+\frac{\mu}{\mu+\lambda_k^2}\left(\sqrt{\lambda_k}\varphi_k(r)-c\lambda_k\psi_k(r)\right)\right]\xi_k \tag{22}$$

have an exponential trend component e^{cr}, scaled respectively by the positively correlated zero mean Gaussian coefficients $\frac{\sqrt{2}c^5}{1+\mu c^4}\nu$ and $\frac{\sqrt{2}\mu c^5}{1+\mu c^4}\nu$ whose covariance is $2\mu\omega^2\left(\frac{c^5}{1+\mu c^4}\right)^2\sigma_\nu^2>0$ for $\mu>0$. For $c<0$ the deterministic factor e^{cr} induces exponential decay in both the limiting HP fitted trend and fitted cycle. When $c>0$ the factor e^{cr} induces exponential growth in these components. When $c=0$ the limit in (21) corresponds to

$$\frac{\hat{f}_{\lfloor nr\rfloor}^{\mathrm{HP}}}{n^{1/2}}\to_{a.s.}\sum_{k=1}^{\infty}\frac{\lambda_k^2}{\mu+\lambda_k^2}\sqrt{\lambda_k}\varphi_k(r)\xi_k,\quad\frac{\hat{c}_{\lfloor nr\rfloor}^{\mathrm{HP}}}{n^{1/2}}\to_{a.s.}\sum_{k=1}^{\infty}\frac{\mu}{\mu+\lambda_k^2}\sqrt{\lambda_k}\varphi_k(r)\xi_k,$$

giving the same findings as in PJ.

Remark 5. These latter two properties of the limiting residual process in (22) contrast with those in (16) for the HP fitted residual of $B_2(r)$, where the HP filter removes the polynomial (in this case, intercept) component in the residual, leaving only the trigonometric functions. The HP filter's elimination of the intercept of $B_2(r)$ in the fitted residual is explained by the fact that the HP filter removes time polynomial functions up to the third degree, thereby including the case of the intercept in the representation of B_2. This facility does not include the exponential function $\mathrm{e}^{cr}=\sum_{j=0}^{\infty}(cr)^j/j!$, which exceeds the capacity of the HP filter. Nonetheless, when $c<0$ the exponential decay factor e^{cr} diminishes the magnitude of this component in the residual. Simulation evidence given in Appendix B corroborates the sign effects of c on the estimation error of the HP filter.

Whereas the HP filter fails to fully capture an exponential trend function, this objective is fulfilled by the bHP filter, as confirmed in the next result.

Theorem 2. *If y_t satisfies the functional law (18) and $\lambda=\mu n^4$, then the bHP filter is consistent with $n^{-1/2}\widehat{f}_{\lfloor nr\rfloor}^{(m)}\rightsquigarrow J_c(r)$ as $m,n\to\infty$.*

For any finite $c \in \mathbb{R}$, Theorem 2 shows that boosting the HP filter removes stochastic trend components in the residual and provides consistent recovery of a local to unity trend. With respect to the arguments given above in Remark 5 regarding the effects of filtering on a finite degree time polynomial trend, the limit theory of the bHP filter in PS (Theorem 2, p. 555) shows that the bHP filter in m iterations removes a polynomial trend of degree $(4m-1)$ from the residual cycle. Passing m to infinity ensures that boosting captures all the terms in a power series expansion of an exponential trend, corroborating the consistency of the bHP filter given in Theorem 2 here.

4.4 Deterministic Trends and Structural Breaks

Time series models sometimes explicitly include deterministic trends and structural breaks as constituent trend components. Using similar arguments to those in PJ and PS, such components may be analyzed for the wider class of stochastic trend functions considered in this paper. In particular, consider a deterministic trend given by the time polynomial

$$d_n(t) = \alpha_n + \beta_{n,1} t + \cdots + \beta_{n,J} t^J, \tag{23}$$

and a polynomial trend with break given by

$$g_n(t) = \begin{cases} \alpha_n^0 + \beta_{n,1}^0 t + \cdots + \beta_{n,J}^0 t^J, & t < \tau_0 = \lfloor n r_0 \rfloor, \\ \alpha_n^1 + \beta_{n,1}^1 t + \cdots + \beta_{n,J}^1 t^J, & t \geq \tau_0 = \lfloor n r_0 \rfloor. \end{cases} \tag{24}$$

The following is a direct corollary that bHP consistently estimates the trend if the underlying stochastic trend is accompanied by an additive polynomial deterministic trend with possible structural breaks.

Corollary 1. *Suppose that y_t^0 satisfies the functional law $\frac{y^0_{\lfloor nr \rfloor}}{n^{q-0.5}} \rightsquigarrow L_q(r)$.*

(a) Let $y_t = y_t^0 + d_n(t)$ where $d_n(t)$ is given by (23) and the coefficients satisfy $\alpha_n / n^{q-0.5} \to \alpha$ and $n^{j-q+0.5}\beta_{n,j} \to \beta_j$ for $j = 1, 2, \cdots, J$. Given $\lambda = \mu n^4$, the asymptotic form of the bHP filtered trend is

$$\frac{\hat{f}^{(m)}_{\lfloor nr \rfloor}}{n^{q-0.5}} \rightsquigarrow B_q(r) + d(r),$$

as $m, n \to \infty$, where $d(r) = \alpha + \beta_1 r + \cdots + \beta_J r^J$.

(b) Let $y_t = y_t^0 + g_n(t)$ where the coefficients in (24) satisfy $\alpha_n^s / n^{q-0.5} \to \alpha$ and $n^{j-q+0.5}\beta^s_{n,j} \to \beta_j^s$ for $j = 1, 2, \cdots, J$ and $s = 1, 2$. Given $\lambda = \mu n^4$, the asymptotic form of the bHP filtered trend is

$$\frac{\hat{f}^{(m)}_{\lfloor nr \rfloor}}{n^{q-0.5}} \rightsquigarrow L_q(r) + \begin{cases} g(r), & r \neq r_0, \\ 0.5[g(r_0^-) + g(r_0^+)], & r = r_0, \end{cases}$$

as $m, n \to \infty$ *with* $m/n \to 0$, *where*

$$g(r) = \begin{cases} \alpha^0 + \beta_1^0 r + \cdots + \beta_J^0 r^J, & r < r_0, \\ \alpha^1 + \beta_1^1 r + \cdots + \beta_J^1 r^J, & r \geq r_0. \end{cases}$$

Similar to PS, Part (b) of Corollary 1 requires an additional condition that $m/n \to \infty$, which is useful in establishing the limit behavior of the filtered series around the break point $r = r_0$. A parallel result holds giving consistency of the bHP filter for an LUR time series with deterministic time polynomial drifts and structural breaks. Simply set $q = 1$ and replace $B_q(r)$ by $J_c(r)$ in Corollary 1 and the results follow. A full statements is omitted.

We have shown consistency of the boosted HP filter estimates of $I(q)$ and LUR trends, coupled with deterministic trends and breaks. It is easy to see that finitely additive normalized combinations of these trends are correspondingly included. For example, suppose

$$y_{t,n} = \frac{y_t^{(1)}}{n} + y_t^{(2)} + \sqrt{n}\alpha + \frac{\beta}{\sqrt{n}} t$$

where $y_t^{(1)}$ is an $I(2)$ trend as in (13), $y_t^{(2)}$ is an LUR trend as in (18), and the respective innovations of these time series $(u_t^{(1)}, u_t^{(2)})'$ are potentially correlated random variables satisfying a bivariate version of (6). Then the boosted HP filter reproduces the asymptotic form of this combined trend process so that

$$n^{-1/2} \widehat{f}_{\lfloor nr \rfloor}^{(m)} \rightsquigarrow B_2(r) + J_c(r) + (\alpha + \beta r)$$

given $\lambda = \mu n^4$ as $m, n \to \infty$. Such results demonstrate the versatility of the boosted filter in dealing with complex forms of nonstationary time series.

5 Simulations

Practical implementation of boosting requires two tuning parameters λ and m to be selected. For low frequency macroeconomic data, PS suggested that the conventional choice $\lambda = 1600$ be used for quarterly data, so that the first iteration gives the HP filter. This choice provides a benchmark that can be adjusted in the case of annual or monthly data (Ravn and Uhlig, 2002). To determine the boosting number m PS proposed a data-driven stopping rule based on a Bayesian-type information criterion (BIC)

$$IC(m) = \frac{\widehat{c}^{(m)\prime} \widehat{c}^{(m)}}{\widehat{c}^{\mathrm{HP}\prime} \widehat{c}^{\mathrm{HP}}} + \log n \frac{\mathrm{tr}(B_m)}{\mathrm{tr}(I_n - S)}, \tag{25}$$

which is motivated by a bias-variance trade-off.[4] Unless explicitly stated otherwise, the numerical work of bHP in this paper employs the following algorithm.

[4] In addition to the BIC (25), PS also suggested a stopping rule based on an augmented Dickey-Fuller unit root test, according to the notion that cyclical behavior should not exhibit unit root behavior. Simulations in PS showed that (25) generally provided better finite sample performance and has therefore been used here.

Algorithm 2. *The boosted HP filter (with the BIC stopping rule)*

Step1 Set a maximum number of iterations, say, $m_{\max} = 200$. Specify the smoothing parameter $\lambda = 1600$ for quarterly data, $\lambda = 129600$ for monthly data, or $\lambda = 6.25$ for annual data.

Step2 Run Algorithm 1 with $m = m_{\max}$, and compute $IC(m)$ in (25) at each iteration.

Step3 Given the minimizer $\widehat{m} := \min_{m \leq m_{\max}} IC(m)$, compute $\widehat{f}^{(\widehat{m})}$ and $\widehat{c}^{(\widehat{m})}$.

Code to implement the above algorithm is provided. In `R` software simply call

```
BoostedHP(y, lambda = 1600, stopping = "BIC", Max_Iter = 200)
```[5]

where `y` is the observed time series, and this function will return the estimated trend and cyclical components, along with the sequence of values of $IC(m)$ and the number of iterations $\widehat{m}$.

5.1 Data Generating Processes

In the following numerical exercises, we generate time series in the general form of (1). Along with various stochastic and deterministic trend processes, we specify a cyclical component c_t satisfying a stationary AR(2)

$$(1 - \cos(\phi)\mathbb{L} + 0.25\mathbb{L}^2)\, c_t = e_t, \quad \text{where } e_t \sim i.i.d.\ N(0, \sigma_e^2).$$

with two complex conjugate roots $2\left[\cos(\phi) \pm \mathbf{i}\sin(\phi)\right]$. To mimic five-year business cycles, we set $\phi = \pi/10$ for quarterly data to produce a periodicity of $2\pi/\phi = 20$ quarters, and $\phi = \pi/30$ for monthly data correspondingly.

We first design five data generating processes (DGPs) based on an $I(2)$ trend. Let $d_n^{\text{lin}}(t) := 100t/n - 50$ be a linear trend and $d_n^{\text{snd}}(t) := 5\,(100t/n)^{1/5}\cos[0.05\pi\,(100t/n)^{0.9}]$ be a sinusoidal trend. In the following list DGP1 is a simple $I(2)$ process and DGPs 2–5 involve an $I(2)$ model combined with additional trend features.

DGP1. $f_t^{(1)} = f_{t-1}^{(1)} + X_t$, where $X_t = X_{t-1} + v_t$, and $v_t \sim i.i.d.\ N(0,1)$.

DGP2. $f_t^{(2)} = f_t^{(1)} + d_n^{\text{snd}}(t)$. A sinusoidal trend is added to DGP1.

DGP3. $f_t^{(3)} = f_t^{(1)} + 500\,(t/n)^3$. A third degree polynomial trend is added to DGP1.

DGP4. $f_t^{(4)} = w_t \cdot 1\{t \leq 0.5n\} + \left((d_n^{\text{lin}}(t))^2 + \sum_{s=0.5n+1}^{t}\sum_{r=0.5n+1}^{s} v_r\right) \cdot 1\{t > 0.5n\}$, where $w_t \sim i.i.d.\ N(0, \sigma_e^2)$. The first half of the sample is white noise; a structural break occurs in the middle of the sample; and the second half of the sample is an $I(2)$ stochastic trend combined with a deterministic quadratic drift.

[5] An `R` package can be accessed at https://github.com/zhentaoshi/bHP_R_pkg (Chen and Shi, 2021), and this command line is the default setting of the arguments in the function. Parallel open-source functions are also available in `Matlab` and `Python` at https://github.com/zhentaoshi/Boosted_HP_filter.

Table 1: MSE of the Estimated Trend: $I(2)$

Quarterly data						Monthly data					
DGP	n	HP	2HP	bHP	AR	DGP	n	HP	2HP	bHP	AR
	100	33.46	25.63	25.68	77.11		300	567.76	279.64	58.95	90.93
1	200	33.70	25.60	24.72	85.37	1	600	597.37	288.88	64.64	92.56
	300	33.76	25.51	24.36	90.02		900	595.67	286.44	67.11	93.29
	100	34.41	25.78	25.66	77.71		300	568.72	279.85	59.01	90.94
2	200	33.72	25.60	24.73	85.46	2	600	597.40	288.89	64.63	92.56
	300	33.76	25.51	24.36	90.03		900	595.69	286.45	67.12	93.29
	100	34.30	25.95	25.63	79.90		300	568.84	280.01	59.08	90.96
3	200	33.73	25.61	24.72	85.70	3	600	597.39	288.91	64.64	92.56
	300	33.76	25.51	24.36	90.09		900	595.69	286.44	67.11	93.29
	100	109.24	55.62	41.27	139.09		300	377.14	179.65	54.14	107.48
4	200	36.88	32.51	32.80	117.24	4	600	318.44	161.43	50.97	106.57
	300	34.76	31.81	32.14	108.94		900	315.60	159.84	52.14	107.59
	100	108.85	57.53	41.61	137.84		300	376.69	181.61	54.29	107.14
5	200	37.02	32.56	32.82	115.63	5	600	318.58	161.49	50.99	106.53
	300	34.78	31.81	32.15	107.95		900	315.63	159.85	52.14	107.58

DGP5. $f^{(5)}(t) = f_t^{(4)} + d_n^{\text{snd}}(t)$. The sinusoidal trend from DGP3 is added to DGP4.

While we set $\sigma_e = 5$ for DGPs 1–5 so that the stationary component c_t is non-negligible in finite samples in relation to the $I(2)$ trend, we reduce σ_e to 1 to match the settings in PS where unit root trends were employed, for LUR processes have less prominent trend behavior than $I(2)$ processes. The following designs in DGPs 6–10 are used with a base LUR model parallel to DGPs 1–5. LURs from three groups (nearly explosive, unit root, near stationary) are used with $c \in \{3, 0, -3\}$.

DGP6. $f_t^{(6)} = \mathrm{e}^{c/n} f_{t-1}^{(6)} + v_t$.

DGP7. $f^{(7)} = f^{(6)}(t) + d_n^{\text{snd}}(t)$.

DGP8. $f_t^{(8)} = f^{(6)}(t) + 500\,(t/n)^3$.

DGP9. $f_t^{(9)} = w_t \cdot 1\{t \leq 0.5n\} + (d_n^{\text{lin}}(t) + f_{t-0.5n}^{(6)}) \cdot 1\{t > 0.5n\}$, where the second half after the structural break features a linear upward drift plus the LUR.

DGP10. $f^{(10)}(t) = f_t^{(9)} + d_n^{\text{snd}}(t)$.

Following Algorithm 2, $\lambda = 1600$ is used for quarterly data. Sample sizes are $n = 100$ (25 years) as a baseline, while $n = 200$ (50 years) and 300 (75 years) are comparable in size to the empirical application, where most time series in the FRED-QD database have 253 quarters. For monthly data $\lambda = 129600$ is used and sample sizes are scaled up by a factor of three, giving $n \in \{300, 600, 900\}$ which is in line with the FREQ-MD database of 759 months.

5.2 Simulation Results

The original HP filter, 'twicing' $m = 2$ iterated HP filter (2HP) (Hall and Thomson, 2022), bHP and the autoregressive method are applied to trended time series based on $I(2)$ and LUR. For each

replication, the mean squared error (MSE) of the estimated trend is calculated as $(n-8)^{-1}\sum_{t=5}^{n-4}(\widehat{f}_t - f_t)^2$.

We report the empirical MSE averaged over 5000 replications. Table 1 displays the MSEs for DGPs 1–5 where the stochastic trends are based on $I(2)$ time series with additional trend, sinusoid, and break components. It is evident that the iterated methods (2HP and bHP) can both improve the original HP filter. The lowest MSEs are obtained uniformly by either of the iterated filters. The MSEs for bHP for each sample size are all close across DGPs 1–3, suggesting that the additional deterministic trend components are estimated accurately and the estimation errors stem primarily from the $I(2)$ stochastic trend. The same is true for DGPs 4 and 5 in which the mid-sample structural break provides a greater challenge in trend detection. bHP produces the lowest MSEs in most cases in quarterly data, and evidently outperforms 2HP in monthly data, where the 'golden rule' $\lambda = 129600$ is inadequate for 2HP to remove the trend. For quarterly data, the MSEs of bHP for DGPs 4 and 5 decrease considerably as n increases from 100 to 300, indicating that the filter narrows down the point of the structural change with greater information, corroborating asymptotic theory of consistent date estimation of the break. When the tuning parameter λ rises from 1600 to 129600 for monthly data, the MSEs for HP and 2HP are inflated in many cases tenfold or more. Evidently, the conventional choice $\lambda = 129600$ for monthly data seems much too large for HP trend detection in monthly $I(2)$ time series. By contrast the same large tuning parameter has a considerably muted effect on the trend detection performance of bHP.

For the autoregressive approach using an AR(p) model, $p = 4$ is set for quarterly data and $p = 12$ for monthly data. The MSEs from these AR filters appear stable across sample sizes with the lag order $p = 12$. Larger lag order parameters p tend to produce smaller MSEs, as might be expected in some cases like the $I(2)$ model with a structural break (and third degree polynomial trend) because these are fitted more easily with a long autoregression. The MSEs from the AR models are far larger than those from the bHP filter, revealing some of the limitations of AR specifications as a flexible trend detection device.

Similar phenomena and rankings of MSEs are observed in Table 2 for DGPs 6–10 when the underlying stochastic trend is LUR. When $n = 100$ for quarterly data and $n = 300$ for monthly data, the near-explosive case with $c = 3$ yields larger MSEs for HP. Under other sample sizes the results are stable across the values $c \in \{3, 0, -3\}$. The MSEs from bHP are only slightly affected when λ is raised from 1600 to 129600. Again, the bHP filter is uniformly superior to the competitors.

In summary, the simulation results show that, at least in terms of MSEs, the conventional choices of λ for quarterly and monthly data appear too large for good trend detection by the HP filter. In only a few LUR cases does the autoregressive approach deliver smaller MSEs in trend detection than HP. Both these methods have poor performance compared with bHP, in some cases with MSEs that are several times higher. The iterative mechanism and its data-determined m clearly make the initial choice of the penalty parameter λ less critical to performance, compensating for its shortcomings particularly with more complex trend processes, thereby providing a more robust approach to general trend estimation than the other methods.

Table 2: MSE of the Estimated Trends: LUR

		$c=3$				$c=0$				$c=-3$			
Quarterly data													
DGP	n	HP	2HP	bHP	AR	HP	2HP	bHP	AR	HP	2HP	bHP	AR
	100	2.77	2.16	1.85	3.46	2.11	1.93	1.81	3.12	2.12	1.94	1.83	3.04
6	200	2.10	1.91	1.80	3.44	2.11	1.93	1.81	3.27	2.14	1.95	1.83	3.20
	300	2.10	1.91	1.81	3.43	2.12	1.93	1.82	3.31	2.13	1.94	1.83	3.27
	100	3.78	2.31	1.90	4.32	3.12	2.07	1.83	4.12	3.13	2.09	1.85	4.14
7	200	2.12	1.91	1.80	3.74	2.13	1.93	1.81	3.65	2.16	1.95	1.83	3.65
	300	2.11	1.91	1.81	3.58	2.12	1.93	1.82	3.52	2.13	1.94	1.83	3.52
	100	3.61	2.45	1.97	4.31	2.93	2.21	1.89	4.37	2.95	2.23	1.91	4.40
8	200	2.12	1.92	1.80	3.78	2.14	1.94	1.81	3.81	2.16	1.96	1.83	3.81
	300	2.11	1.91	1.81	3.61	2.12	1.93	1.82	3.63	2.13	1.94	1.83	3.63
	100	7.87	6.48	4.97	7.44	2.94	2.58	2.35	4.50	2.32	2.10	2.01	4.12
9	200	2.79	2.54	2.37	4.27	2.23	2.09	2.03	3.95	2.03	1.93	1.91	3.82
	300	2.27	2.12	2.07	3.90	2.06	1.96	1.93	3.79	1.96	1.88	1.86	3.71
	100	8.66	6.62	5.06	8.09	3.73	2.72	2.38	5.16	3.11	2.23	2.04	4.84
10	200	2.81	2.54	2.37	4.55	2.25	2.10	2.03	4.20	2.05	1.94	1.91	4.08
	300	2.27	2.12	2.07	4.05	2.07	1.96	1.93	3.91	1.96	1.88	1.86	3.84
Monthly data													
DGP	n	HP	2HP	bHP	AR	HP	2HP	bHP	AR	HP	2HP	bHP	AR
	300	5.25	4.27	2.40	3.67	5.02	4.20	2.36	3.58	5.02	4.21	2.36	3.56
6	600	5.12	4.24	2.42	3.69	5.13	4.25	2.43	3.64	5.17	4.28	2.43	3.62
	900	5.15	4.26	2.48	3.69	5.17	4.28	2.48	3.66	5.19	4.29	2.49	3.65
	300	6.29	4.42	2.50	3.92	6.07	4.36	2.46	3.85	6.06	4.36	2.46	3.85
7	600	5.14	4.25	2.42	3.76	5.15	4.26	2.43	3.73	5.19	4.29	2.44	3.72
	900	5.16	4.26	2.48	3.72	5.17	4.28	2.48	3.70	5.19	4.29	2.49	3.70
	300	6.05	4.54	2.59	3.94	5.85	4.48	2.58	3.96	5.83	4.48	2.59	3.96
8	600	5.15	4.25	2.43	3.78	5.16	4.26	2.44	3.79	5.20	4.29	2.44	3.79
	900	5.16	4.26	2.48	3.73	5.17	4.28	2.48	3.74	5.19	4.29	2.49	3.74
	300	4.95	4.10	2.57	4.12	4.31	3.60	2.36	3.99	3.91	3.30	2.24	3.90
9	600	3.77	3.23	2.25	3.82	3.63	3.12	2.20	3.79	3.55	3.06	2.16	3.76
	900	3.51	3.03	2.19	3.74	3.48	3.00	2.17	3.73	3.45	2.98	2.16	3.72
	300	5.78	4.26	2.63	4.38	5.13	3.75	2.41	4.23	4.74	3.45	2.28	4.16
10	600	3.79	3.24	2.25	3.92	3.65	3.12	2.20	3.87	3.57	3.06	2.16	3.85
	900	3.52	3.03	2.19	3.79	3.48	3.00	2.17	3.77	3.45	2.98	2.16	3.76

6 Further Empirics

Section 2 demonstrated preliminary findings for trend and cycle detection with real GDP series from the FRED-QD database and monthly industrial production series from the FRED-MD database. This section gives a 'big data' application, making full use of the time series in these two databases to draw an overall picture of trend and growth cycles in the US economy.

The quarterly data findings are presented first. FRED-QD is a comprehensive collection of macroeconomic time series, covering 14 categories of economic activity concerning, for example, the national income and product accounts, the labor market covering unemployment, earnings and productivity, and financial markets covering interest rates, money and credit. Most of the 246 individual time series are complete, with 253 quarters spanning 1959:Q1 to 2022:Q1. A few sequences have missing values and the shortest one has 121 quarters. Many of the time series show strong trend behavior, such as the real GDP, the consumer price index, and stock market indices such as the S&P 500. Other series have random wandering behavior with a more limited range, such as the unemployment rate and interest rates, although the historical patterns vary considerably over the sample period. Our approach to the analysis of the time series in these databases is agnostic, using the HP, bHP and AR methodologies to filter each individual time series without any pre-processing.[6] Regardless of the sample size, we maintain the use of the standard setting $\lambda = 1600$, just as in the simulations.

These time series have their own distinctive scales of measurement. GDP, for instance, is in trillions of US dollars whereas interest rates are reported in small percentage terms. To make the time series comparable for group assessment we take the filtered residuals and standardize each estimated residual sequence to have unit sample variance. These standardized residuals provide rough individual measures of the nature and form of business cycles over the historical period.

Of particular interest are the shapes of the estimated cycles around periods of recession. Some of the time series require reorientation for this purpose as their natural movements reverse normal directions during recessions, e.g., by uptrending in recession periods. These are manually identified and their signs changed. Thus, while most variables decline as economic activity diminishes, unemployment rates rise and so flipping the sign of the unemployment rate to negative brings the direction of movement in line with the majority of the other variables. Accordingly, upon inspection minus signs are assigned to ID 58-72 and 197 (unemployment) and ID 158-162 (money stocks) in the database.

The mass of thin green lines in Figure 4 show the cyclical components of the time series as estimated by the HP filter (upper panel), the bHP (middle panel), and an AR(4) regression (lower panel). These individual residual series are evidently noisy, and the collection of 246 lines in a single graph merge together into a dense green shading where individual movements are hardly visible. We add solid lines (red for HP, blue for bHP, violet for AR(4)) to aggregate the individual sequences by simple cross section averaging into ensemble indices. These indices provide summary indicator

[6] 2HP calculations were recorded in all cases and found to produce outcomes close to those of HP and are therefore not included in the reported results.

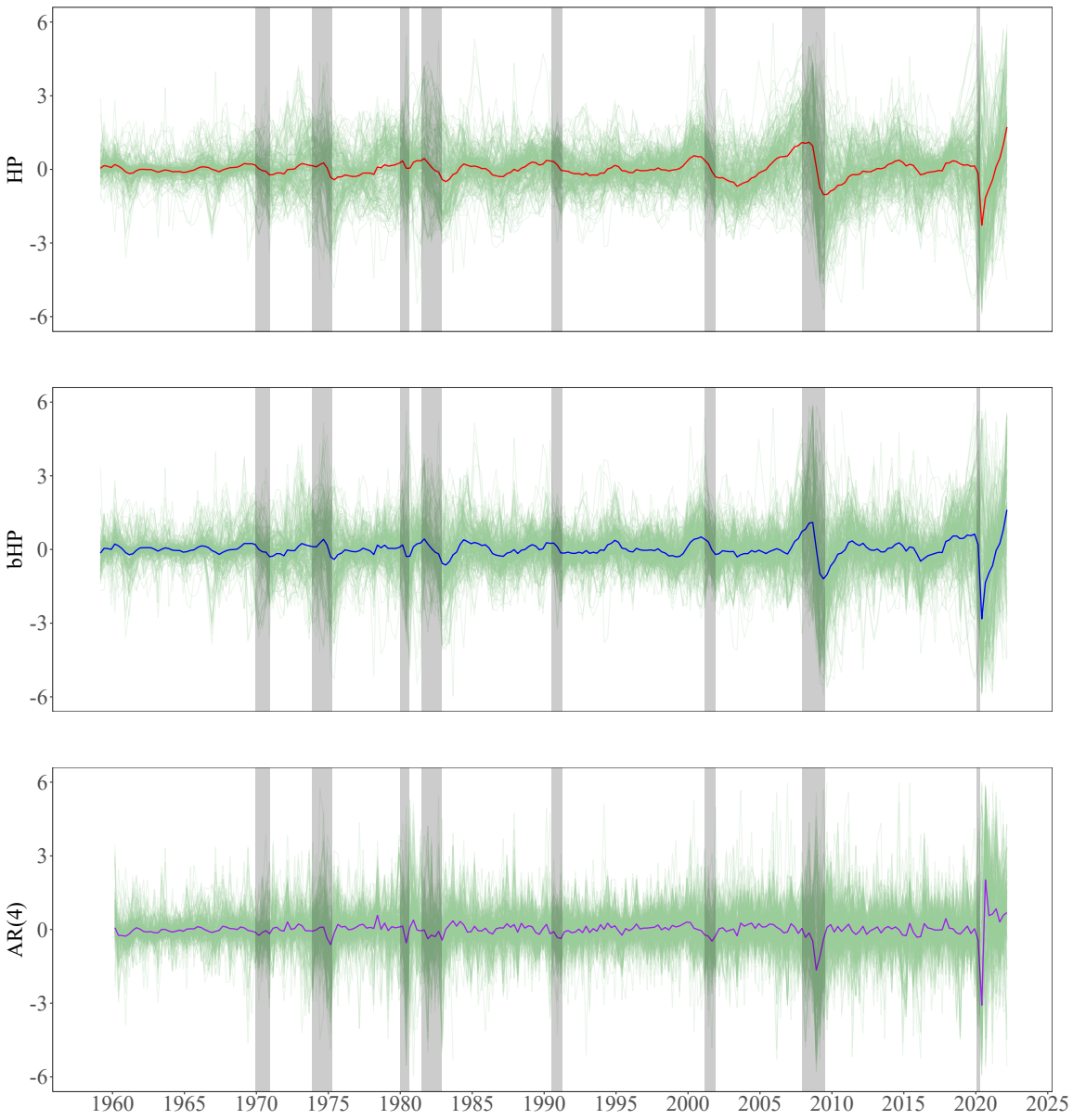


Figure 4: Aggregate Cyclical Indices from FRED-QD

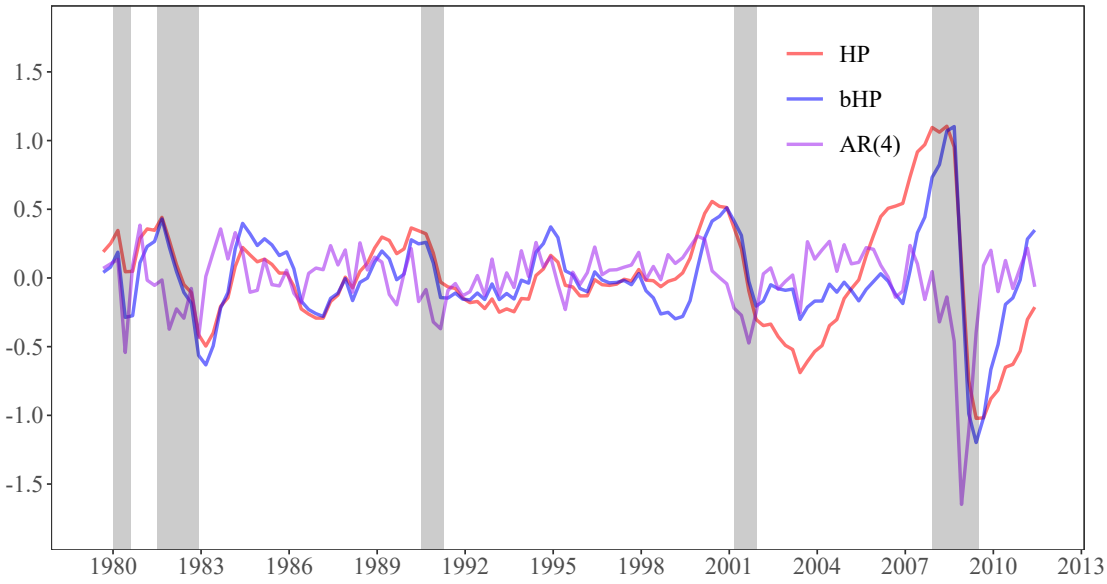


Figure 5: Aggregate Cyclical Indices from FRED-QD during 1979–2011

measures of the business cycle status of the economy that broadly reflect the movements of the individual time series estimates.

The average indices generated by the HP and bHP filters move in a fairly synchronized way, particularly around the recession periods identified by the NBER which are marked by the grey shaded areas. For instance, after the dot-com bubble burst in March 2000 these indices each decline to the trough during the 2001 recession of the early 2000s, matching the contraction phase of the business cycle. The bHP index then begins to rise again with the recovery, whereas the HP index is slower and takes much longer to recover. For the great recession of 2008–2009, these two indices again decline to a trough as the financial sector turmoil spreads through the real economy during the economic contraction and then subsequently go through a slow return to normalcy matching the slow recovery with the HP index again taking longer to recover. The recent Covid-19 recession was of much greater scale and the recovery followed swiftly. Outside of the recession periods, business and growth cycles are still visible in the movement of the indices, such as the long expansion in 1990s and the strong economic activity prior to the Covid pandemic. Overall, from the historical data these two indices appear in line with consensus perceptions of US economic activity. Although the differences between these two indices are subtle, the bHP filtered index seems more responsive to historical movements in the data and more in line with the chronology of the recession phases than the HP filter.

The results for the AR(4) regression filter are similar to those in Figure 1, bottom-right panel, for real GDP and again somewhat counterintuitive. First, little systematic cyclical behavior is observed over the entire time span. Second, the index recovered in the middle of the great recession and outperformed the pre-crisis peak before NBER dating indicated termination of the recession in 2009. Third, large oscillating shocks are recorded by the AR index over the two quarters 2020:Q2

and Q3 in which the index first slumps and then immediately recovers and bounces back to an unprecedented high level. Overall the AR index appears either lacking in sensitivity for much of the period or over-sensitive and mistimed in others.

Figure 5 aligns the four aggregate indices for comparison within a single graph, zooming in on the time period 1979–2011, which excludes the immense Covid-19 downturn for better visualization. The HP and bHP indices appear close during recession periods and in line with NBER dating, but show some differences during normal and expansionary periods. For example, only the bHP filter shows a clear downturn during the Russian debt default in August 1998 and bailout of the hedge fund Long-Term Capital Management, even though this period was not classified as a recession. Also, the HP filter produces strong uptrends over 2003–2008 whereas this upswing is not so evident in bHP until 2007.

To check the robustness of these findings from quarterly data, parallel exercises are conducted for the 127 individual time series in the monthly FRED-MD database. The tuning parameter $\lambda = 129600$ is used for HP and bHP, and an AR(12) regression filter is employed. Negative signs are given to series 25–31 (unemployment) and series 70–73 (money stock). The scale-standardized residuals are plotted in Figure 6 again in thin green lines and the colored solid lines are simple averages of these. The residuals of a few of the time series are so large at the inception of the Covid-19 pandemic that they reach magnitudes of ± 20 standard errors and are not visible in the displayed figure. Instead, to more clearly visualize the fluctuations of the aggregate indices over the whole period, the y-axis of each subfigure is restricted to ± 6 standard errors. Similar patterns are observed from these monthly data and the filtered indices are now evident with fine grain monthly movements. For example, in the recovery episode after Covid-19, the aggregate HP filtered index monotonically increases whereas the bHP index shows small fluctuations that accord with waves of new Covid variants and the effects of lockdowns on economic activity. The AR(12) index appears, with exceptions around the GFC and Covid-19 recessions, to be closer to a martingale difference sequence than a growth or business cycle index.

To further illustrate this finding, we conduct autocorrelation tests for the indices generated by these filters. We implement the robust testing method by Dalla et al. (2020) against the joint null hypothesis $\rho_1 = \rho_2 = \cdots = \rho_K = 0$, where $(\rho_k)_{k=1}^K$ is the autocorrelation function of a time series under consideration. We examine the autocorrelation functions up to one and a half years by setting $K = 6$ for the quarterly aggregates and $K = 18$ for the monthly data, and the results are plotted in Figure 7a and 7b, respectively. A test statistic above the critical value (5% test level, in black crosses) rejects the null. In all cases, only the AR index fails to reject the null, suggesting little time persistence. In comparison, the aggregate cyclical indices from HP and bHP exhibit significant autocorrelations and dynamics over time.

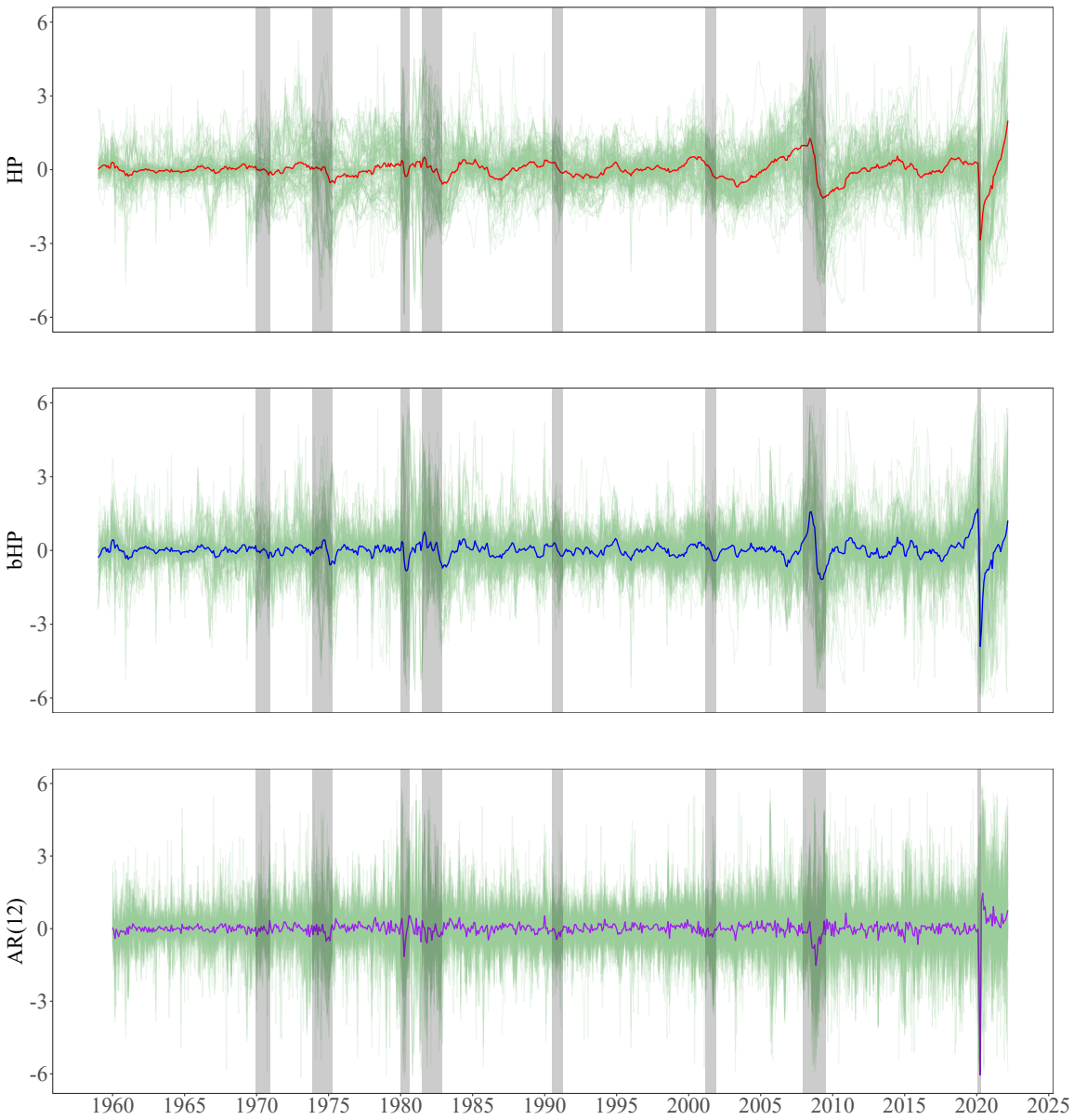
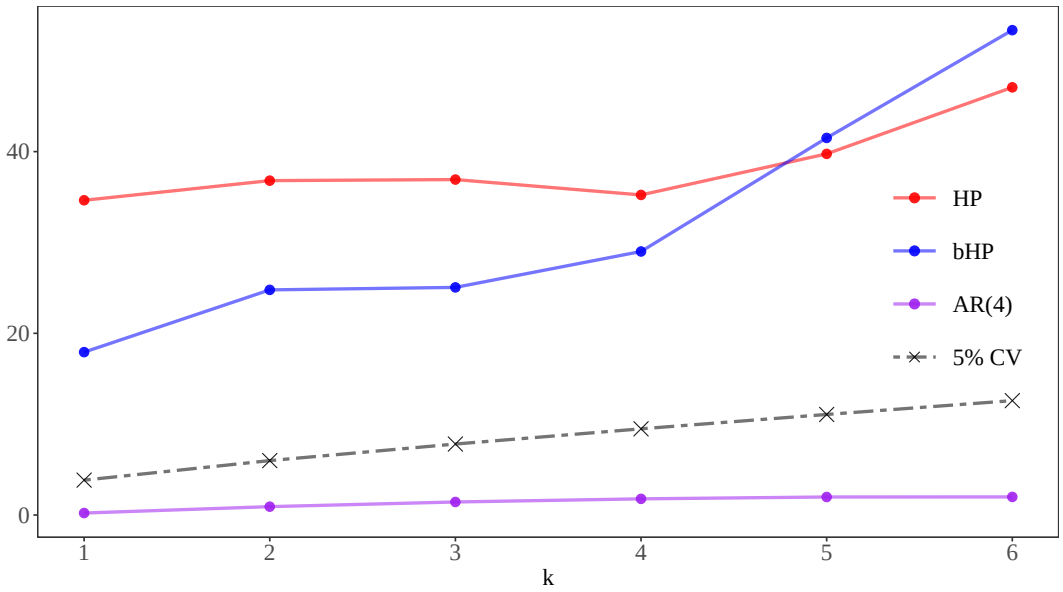
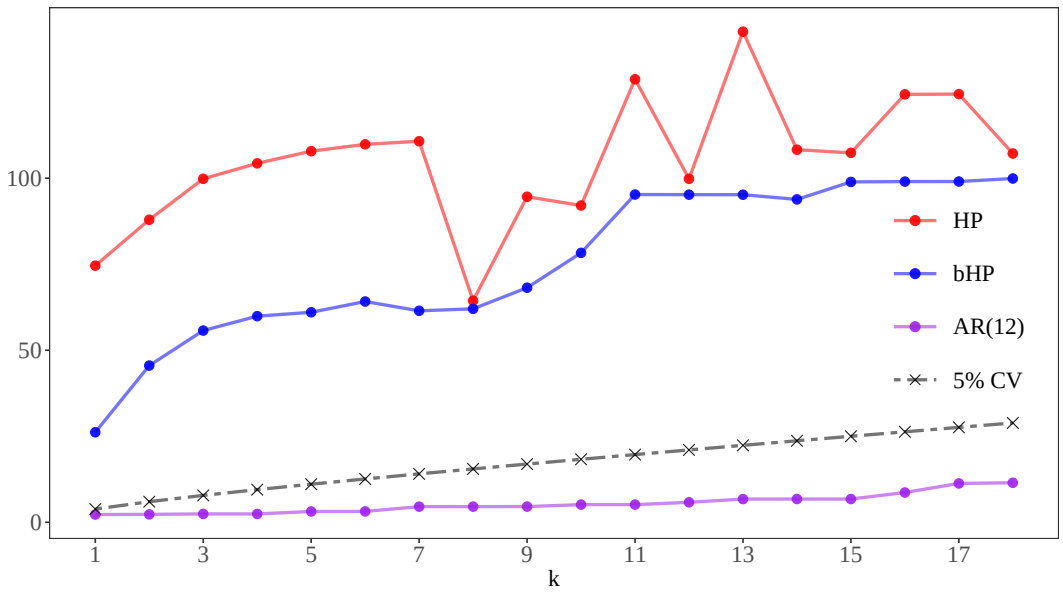


Figure 6: Aggregate Cyclical Indices from FRED-MD



(a) Quarterly Indices



(b) Monthly Indices

Figure 7: Test Statistics Against the Null $\rho_1 = \cdots = \rho_K = 0$

7 Conclusion

This paper extends the analysis of the boosted HP filter to higher order integrated processes and time series with roots that are local to unity, possibly coupled with polynomial time trends and structural breaks. The primary asymptotic effect of the HP filter is to smooth the underlying stochastic trend, a property that typically leads to inconsistent estimation but still provides a general picture of how the trend has evolved historically. Boosting the HP filter by repeated application ensures consistent estimation of the full limiting trend process and allows for a rich combination of possible trend behavior including breaks. Practical implementation is facilitated by a data-driven procedure that gives researchers an automated machine learning tool of empirical analysis.

The HP and bHP methods of trend estimation are motivated by the penalized likelihood approach initiated by Whittaker (1923) which seeks in the context of a model such as (1) to find the 'most probable' trend function f_t over a historical period. Trends, as is now well understood, encompass a vast range of possible behavior in which future behavior is not always predictable from the past. It is for this reason that economists are so often wrong in assessing macroeconomic and financial market tendencies, to wit: whether inflation will be transient or sustained, whether there is stock market exuberance and potentially serious consequences of collapse,[7] when a recession will occur, how extensive it will be, how long a recovery will take and whether an earlier trend path will be resumed. In a model of the form (1) representing behavior over an historical period, such characteristics are implicitly embodied, whereas in a purely predictive model such as an autoregression for which the innovations are assumed to be martingale differences, such departures inevitably amount to model misspecification.

The alternative paradigm of pure autoregressive modeling was strongly advocated by Hamilton (2018) and motivated by the following alternative characterization

> "Here I suggest an alternative concept of what we might mean by the cyclical component of a possibly nonstationary series: How different is the value at date $t+h$ from the value that we would have expected to see based on its behavior through date t?" (Hamilton, 2018, p. 836)

This conceptualization acknowledges the possibility of model misspecification in that future paths, including cycles but also trends, may differ from those that may be anticipated from past observations. In doing so, the view implicitly accepts that predictive models such as autoregressions may well be incompatible with historical models such as (1) that seek to graduate the data to understand the trends and cycles that have occurred in the past.

Trends and cycles in real world economies are complex and often evolve in unanticipated ways even though the mechanisms and behaviors that drive them may well have some common characteristics (Reinhart and Rogoff, 2009). As PS (p. 551) remarked:

[7] Recall Queen Elizabeth II's timely and famously insightful comment at the London School of Economics (5 November, 2008) on the 2008 GFC collapse that 'It's awful — why did nobody see it coming?'

"... the period and intensity of business cycles and recessions are so noted for their irregularity that these features are embodied in the many popular descriptive terminologies that are given to them, among which we may mention the terms great depression, great moderation, great recession, short sharp recession, and long recovery."

Such phenomena are difficult to faithfully capture by a supervised learning method like an autoregression that uses only lagged observations to predict future outcomes. By contrast, bHP is a nonparametric unsupervised machine learning approach that is capable of extracting a wide class of underlying trends and cycles from the data. In doing so, the method accommodates historical decomposition of the type (1) and recognizes the existence of underlying trend and cycle formations that may take irregular and unpredictable forms. The asymptotic theory, simulations and empirical applications reported here all corroborate these advantages.

References

Athey, S., M. Bayati, N. Doudchenko, G. Imbens, and K. Khosravi (2021). Matrix Completion Methods for Causal Panel Data Models. *Journal of the American Statistical Association 116*(536), 1716–1730.

Babii, A., E. Ghysels, and J. Striaukas (2021). Machine Learning Time Series Regressions with an Application to Nowcasting. *Journal of Business & Economic Statistics*, 1–23.

Bai, J. and S. Ng (2009). Boosting Diffusion Indices. *Journal of Applied Econometrics 24*(4), 607–629.

Belloni, A., V. Chernozhukov, and C. Hansen (2014). Inference on Treatment Effects After Selection Among High-Dimensional Controls. *The Review of Economic Studies 81*(2), 608–650.

Biswas, E., F. Sabzikar, and P. C. B. Phillips (2022). Boosting the hp filter for trending time series with long range dependence. *Working Paper, Yale University*.

Bühlmann, P. (2006). Boosting for High-Dimensional Linear Models. *The Annals of Statistics*, 559–583.

Bühlmann, P. and B. Yu (2003). Boosting with the L2 Loss: Regression and Classification. *Journal of the American Statistical Association 98*(462), 324–339.

Caner, M. and A. B. Kock (2018). Asymptotically Honest Confidence Regions for High Dimensional Parameters by the Desparsified Conservative Lasso. *Journal of Econometrics 203*(1), 143–168.

Chan, N. H. and C. Z. Wei (1987). Asymptotic inference for nearly nonstationary AR(1) processes. *The Annals of Statistics 15*(3), 1050 – 1063.

Chen, Y. and Z. Shi (2021, June). R Package Vignette for Boosted HP Filter. Technical report, DOI:10.13140/RG.2.2.33986.66249.

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.

Cornea-Madeira, A. (2017). The Explicit Formula for the Hodrick-Prescott Filter in a Finite Sample. *Review of Economics and Statistics 99*(2), 314–318.

Dalla, V., L. Giraitis, and P. C. B. Phillips (2020). Robust tests for white noise and cross-correlation. *Econometric Theory*, 1–29.

De Jong, R. M. and N. Sakarya (2016). The Econometrics of the Hodrick-Prescott Filter. *Review of Economics and Statistics 98*(2), 310–317.

Drehmann, M. and J. Yetman (2018). Why You Should Use the Hodrick-Prescott Filter — At Least to Generate Credit Gaps. *BIS Working Papers 744*.

Farrell, M. H., T. Liang, and S. Misra (2021). Deep Neural Networks for Estimation and Inference. *Econometrica 89*(1), 181–213.

Freund, Y. and R. E. Schapire (1995). A Desicion-Theoretic Generalization of On-line Learning and an Application to Boosting. In *European Conference on Computational Learning Theory*, pp. 23–37. Springer.

Friedman, J., T. Hastie, and R. Tibshirani (2000). Additive Logistic Regression: A Statistical View of Boosting. *The Annals of Statistics 28*(2), 337–407.

Frisch, R. (1933). Propagation Problems and Impulse Problems in Dynamic Economics. In *Economic Essays in Honour of Gustav Cassel*, pp. 171–205. George Allen & Unwin.

Haldrup, N. (1998). An Econometric Analysis of I(2) Variables. *Journal of Economic Surveys 12*(5), 595–650.

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.

Hall, V. B. and P. Thomson (2022). A boosted hp filter for New Zealand business cycle analysis. *Victoria University of Wellington, Working Paper*, 1–39.

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

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

Johansen, S. (1995). A Stastistical Analysis of Cointegration for I(2) Variables. *Econometric Theory 11*(1), 25–59.

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

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.

Knight, K. (2021). The Boosted Hodrick-Prescott Filter, Penalized Least Squares, and Bernstein Polynomials. *Working Paper, University of Toronto*.

Krugman, P. (2012). Filters and Full Employment (Not Wonkish, Really). http://krugman.blogs.nytimes.com/2012/07/11/filters-and-full-employment-not-wonkish-really.

Kueck, J., Y. Luo, M. Spindler, and Z. Wang (2022). Estimation and Inference of Treatment Effects with L2-Boosting in High-Dimensional Settings. *Journal of Econometrics*.

Lee, S., Y. Liao, M. H. Seo, and Y. Shin (2021). Sparse HP filter: Finding Kinks in the COVID-19 Contact Rate. *Journal of Econometrics 220*(1), 158–180.

Masini, R. and M. C. Medeiros (2021). Counterfactual Analysis With Artificial Controls: Inference, High Dimensions, and Nonstationarity. *Journal of the American Statistical Association 116*(536), 1773–1788.

McCracken, M. and S. Ng (2020). Fred-qd: A Quarterly Database for Macroeconomic Research. Technical report, National Bureau of Economic Research.

McCracken, M. W. and S. Ng (2016). FRED-MD: A Monthly Database for Macroeconomic Research. *Journal of Business & Economic Statistics 34*(4), 574–589.

Moon, H. R. and M. Weidner (2018). Nuclear Norm Regularized Estimation of Panel Regression Models. *arXiv preprint arXiv:1810.10987*.

Phillips, P. C. B. (1986). Understanding spurious regressions in econometrics. *Journal of Econometrics 33*(3), 311–340.

Phillips, P. C. B. (1987a). Time series regression with a unit root. *Econometrica*, 277–301.

Phillips, P. C. B. (1987b). Towards a Unified Asymptotic Theory for Autoregression. *Biometrika 74*(3), 535–547.

Phillips, P. C. B. (1998). New Tools for Understanding Spurious Regressions. *Econometrica*, 1299–1325.

Phillips, P. C. B. (2005). Automated Discovery in Econometrics. *Econometric Theory 21*(1), 3–20.

Phillips, P. C. B. (2007). Unit Root Log Periodogram Regression. *Journal of Econometrics 138*(1), 104–124.

Phillips, P. C. B. (2023). Estimation and inference with near unit roots. *Econometric Theory*.

Phillips, P. C. B. and S. Jin (2021). Business cycles, Trend Elimination, and the HP Filter. *International Economic Review 62*(2), 469–520.

Phillips, P. C. B. and T. Magdalinos (2007). Limit theory for moderate deviations from a unit root. *Journal of Econometrics 136*(1), 115–130.

Phillips, P. C. B. and Z. Shi (2021). Boosting: Why You Can Use the HP Filter. *International Economic Review 62*(2), 521–570.

Phillips, P. C. B. and V. Solo (1992). Asymptotics for Linear Processes. *The Annals of Statistics*, 971–1001.

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.

Reinhart, C. M. and K. S. Rogoff (2009). *This Time is Different*. Princeton University Press.

Sakarya, N. and R. M. de Jong (2020). A Property of the Hodrick-Prescott Filter and Its Application. *Econometric Theory 36*(5), 840–870.

Schüler, Y. S. (2021). On the Cyclical Properties of Hamilton's Regression Filter. *Available at SSRN 3559776*.

Shi, Z. (2016). Econometric Estimation with High-dimensional Moment Equalities. *Journal of Econometrics 195*(1), 104–119.

Shi, Z., L. Su, and T. Xie (2023). L2-relaxation: With applications to forecast combination and portfolio analysis. *Review of Economics and Statistics*.

Su, L., Z. Shi, and P. C. B. Phillips (2016). Identifying Latent Structures in Panel Data. *Econometrica 84*(6), 2215–2264.

Tinbergen, J. (1939). *Business Cycles in the United States of America: 1919-1932*. League of Nations.

Tukey, J. W. (1977). *Exploratory Data Analysis*. Reading, PA: Addison-Wesley.

Whittaker, E. T. (1923). On a New Method of Graduation. *Proceedings of the Edinburgh Mathematical Society 41*, 63–75.

Yamada, H. (2020). A Smoothing Method That Looks Like the Hodrick–Prescott Filter. *Econometric Theory 36*(5), 961–981.

Yamada, H. (2022). Trend Extraction From Economic Time Series with Missing Observations by Generalized Hodrick–Prescott Filters. *Econometric Theory 38*, 419–453.

Yousuf, K. and S. Ng (2021). Boosting High Dimensional Predictive Regressions with Time Varying Parameters. *Journal of Econometrics 224*(1), 60–87.

Online Appendix

A Proofs

A.1 Preparatory Results

To simplify the proof of Lemma 1, we start with the following preliminary result for the simple case where $a \neq 0$ and $m = 1$. For any $a \in \mathcal{A}(n) \setminus \{0\}$, define the two complex numbers

$$\delta = \mathrm{e}^{2a/n}\left[\frac{n}{a}\left(1-\mathrm{e}^{-a/n}\right)\right]^4 - 1,$$
$$\zeta = \frac{\mu a^4(1+\delta)}{\mu a^4(1+\delta)+1} - \frac{\mu a^4}{1+\mu a^4}.$$

Lemma A.1. *Let $\lambda = \mu n^4$ and $a \in \mathcal{A}(n) \setminus \{0\}$. Then*

$$(1 - G_\lambda)\,\mathrm{e}^{at/n} = \left(\frac{\mu a^4}{\mu a^4 + 1} + \zeta\right)\mathrm{e}^{at/n}, \tag{A.1}$$

and the real part and the imaginary part of ζ are both of order $O(|a|/n)$.

Proof of Lemma A.1. Using the operator calculus in PJ (p. 510) we have

$$\begin{aligned} G_\lambda \mathrm{e}^{at/n} &= \frac{1}{1+\mu\mathbb{L}^{-2}[n(1-\mathbb{L})]^4}\mathrm{e}^{at/n} = \int_0^\infty \mathrm{e}^{-\{\mu\mathbb{L}^{-2}[n(1-\mathbb{L})]^4+1\}s}\mathrm{e}^{at/n}\mathrm{d}s \\ &= \int_0^\infty \mathrm{e}^{-s}\sum_{j=0}^\infty \frac{(-s)^j}{j!}[\mu\mathbb{L}^{-2}[n(1-\mathbb{L})]^4]^j \mathrm{e}^{at/n}\mathrm{d}s. \end{aligned} \tag{A.2}$$

Notice that

$$\begin{aligned} & [\mu\mathbb{L}^{-2}[n(1-\mathbb{L})]^4]^j \mathrm{e}^{at/n} \\ = \;& \mu^j n^{4j}(1-\mathbb{L})^{4j}\mathrm{e}^{a(t+2j)/n} = \mu^j n^{4j}\sum_{k=0}^{4j}(-1)^k\binom{4j}{k}\mathrm{e}^{a(t+2j-k)/n} \\ = \;& \mu^j n^{4j}\mathrm{e}^{a(t+2j)/n}\sum_{k=0}^{4j}(-1)^k\binom{4j}{k}\mathrm{e}^{-ak/n} = \mu^j n^{4j}\mathrm{e}^{a(t+2j)/n}(1-\mathrm{e}^{-a/n})^{4j} \\ = \;& \left[\mu\mathrm{e}^{2a/n}n^4(1-\mathrm{e}^{-a/n})^4\right]^j \mathrm{e}^{at/n}, \end{aligned}$$

using the binomial expansion $(x+y)^{4j} = \sum_{k=0}^{4j}\binom{4j}{k}x^k y^{4j-k}$. Evaluation of (A.2) gives

$$\begin{aligned} G_\lambda \mathrm{e}^{at/n} &= \int_0^\infty \mathrm{e}^{-s}\sum_{j=0}^\infty \frac{(-s)^j}{j!}\left[\mu\mathrm{e}^{2a/n}n^4(1-\mathrm{e}^{-a/n})^4\right]^j \mathrm{e}^{at/n}\mathrm{d}s \\ &= \int_0^\infty \mathrm{e}^{-\{\mu\mathrm{e}^{2a/n}n^4(1-\mathrm{e}^{-a/n})^4+1\}s}\mathrm{e}^{at/n}\mathrm{d}s \end{aligned}$$

$$= \frac{\mathrm{e}^{at/n}}{\mu \mathrm{e}^{2a/n} n^4 (1-\mathrm{e}^{-a/n})^4 + 1} = \frac{\mathrm{e}^{at/n}}{\mu a^4 (1+\delta) + 1},$$

by the definition of δ. The residual operator yields

$$(1 - G_\lambda)\, \mathrm{e}^{at/n} = \left(\frac{\mu a^4 (1+\delta)}{\mu a^4 (1+\delta) + 1} \right) \mathrm{e}^{at/n} = \left(\frac{\mu a^4}{\mu a^4 + 1} + \zeta \right) \mathrm{e}^{at/n}$$

by the definition of ζ, verifying (A.1).

To establish the order of ζ, we first check the order of δ by working on its two factors $\mathrm{e}^{a/n}$ and $\frac{n}{a}\left(1 - \mathrm{e}^{-a/n}\right)$ one by one. Taylor expansion of $\mathrm{e}^{a/n}$ for $a \in \mathcal{A}(n)$ gives

$$\mathrm{e}^{a/n} = 1 + \sum_{j=1}^{\infty} \frac{1}{j!} \left(\frac{a}{n}\right)^j = 1 + O\left(\frac{|a|}{n}\right).$$

Similar expansion of $\mathrm{e}^{-a/n}$ gives

$$\begin{aligned} \frac{n}{a}\left(1 - \mathrm{e}^{-a/n}\right) &= \frac{n}{a}\left(1 - \sum_{j=0}^{\infty} \frac{(-1)^j}{j!} \left(\frac{a}{n}\right)^j\right) = -\sum_{j=1}^{\infty} \frac{(-1)^j}{j!} \left(\frac{a}{n}\right)^{j-1} \\ &= 1 - \sum_{j=2}^{\infty} \frac{(-1)^j}{j!} \left(\frac{a}{n}\right)^{j-1} = 1 + O\left(\frac{|a|}{n}\right). \end{aligned}$$

Substituting these terms into δ we have

$$\delta = \left[\mathrm{e}^{a/n}\right]^2 \left[\frac{n}{a}\left(1 - \mathrm{e}^{-a/n}\right)\right]^4 - 1 = \left(1 + O\left(\frac{|a|}{n}\right)\right)^6 - 1 = O\left(\frac{|a|}{n}\right). \tag{A.3}$$

Next, decomposing ζ we have

$$\begin{aligned} \zeta &= \frac{\mu a^4 \delta}{\mu a^4 (1+\delta) + 1} + \left(\frac{\mu a^4}{\mu a^4 (1+\delta) + 1} - \frac{\mu a^4}{1 + \mu a^4} \right) \\ &= \frac{\delta}{1 + \delta + 1/(\mu a^4)} + \left(\frac{1}{1 + \delta + 1/(\mu a^4)} - \frac{1}{1 + 1/(\mu a^4)} \right) =: s_1 + s_2. \end{aligned}$$

By the triangle inequality and (A.3), the modulus of the denominator of s_1 satisfies

$$|1 + \delta + 1/(\mu a^4)| \geq \left|1 + 1/(\mu a^4)\right| - |\delta| \geq 1 - |\delta| = 1 - O\left(|a|/n\right) \geq 1/2,$$

for n sufficiently large, bounded away from the origin. So the modulus of s_1 has order no larger than its numerator, which implies that $|s_1| = O(\delta) = O(|a|/n)$. Similarly, we have

$$|s_2| = \left| \frac{\delta}{(1 + \delta + 1/(\mu a^4))\,(1 + 1/(\mu a^4))} \right| \leq \left| \frac{\delta}{1 + \delta + 1/(\mu a^4)} \right| = |s_1| = O\left(|a|/n\right),$$

thereby confirming that $\zeta = O\left(|a|/n\right)$ since $|\zeta| \leq |s_1| + |s_2| = O\left(|a|/n\right)$. □

We extends the operator algebra in Lemma A.1 to general m to prove Lemma 1.

Proof of Lemma 1. When $a = 0$, we have for $m \geq 1$

$$(1 - G_\lambda)^m \, \mathrm{e}^{at/n} = (1 - G_\lambda)^m \, 1 = \left(\frac{\mu n^4 \mathbb{L}^{-2}}{1 + \mu \mathbb{L}^{-2} [n(1 - \mathbb{L})]^4} \right)^m (1 - \mathbb{L})^{4m} 1 = 0,$$

as well as

$$\left[(1 - G_\lambda)^m - \left(\frac{\mu a^4}{\mu a^4 + 1} \right)^m \right] \mathrm{e}^{at/n} = \left[(1 - G_\lambda)^m - 0 \right] 1 = 0.$$

We now focus on the case when $a \neq 0$.

Part (a). When the operator $(1 - G_\lambda)$ is repeatedly applied m times with m a fixed integer, (A.1) yields

$$\begin{aligned} (1 - G_\lambda)^m \mathrm{e}^{at/n} &= \left[\frac{\mu a^4}{1 + \mu a^4} + \zeta \right]^m \mathrm{e}^{at/n} \\ &= \left[\left(\frac{\mu a^4}{1 + \mu a^4} \right)^m + \sum_{j=1}^{m} \binom{m}{j} \left(\frac{\mu a^4}{1 + \mu a^4} \right)^{m-j} \zeta^j \right] \mathrm{e}^{at/n}, \end{aligned} \tag{A.4}$$

where the second equality holds by binomial expansion. Rearranging this equation and taking the modulus gives

$$\begin{aligned} & \left| \left[(1 - G_\lambda)^m - \left(\frac{\mu a^4}{1 + \mu a^4} \right)^m \right] \mathrm{e}^{at/n} \right| \\ = \; & \left| \sum_{j=1}^{m} \binom{m}{j} \left(\frac{\mu a^4}{1 + \mu a^4} \right)^{m-j} \zeta^j \mathrm{e}^{at/n} \right| \leq \left| \sum_{j=1}^{m} \binom{m}{j} \left(\frac{\mu a^4}{1 + \mu a^4} \right)^{m-j} \zeta^j \right| \left| \mathrm{e}^{at/n} \right| \\ \leq \; & \mathrm{e}^{|a|} \sum_{j=1}^{m} \binom{m}{j} \left(\frac{\mu a^4}{1 + \mu a^4} \right)^{m-j} |\zeta|^j \leq \mathrm{e}^{|a|} \sum_{j=1}^{m} \binom{m}{j} |\zeta|^j \\ \leq \; & m \, |\zeta| \, \mathrm{e}^{|a|} \, (1 + o(1)) = O\left(n^{-1} |a| \mathrm{e}^{|a|} \right), \end{aligned}$$

as $t \leq n$, $0 < \dfrac{\mu a^4}{1 + \mu a^4} \leq 1$ and $\zeta = O(|a|/n)$. Since $|a| \leq \sqrt{\log n}$ for any $a \in \mathcal{A}(n)$, it follows that

$$\begin{aligned} n^{-1} |a| \mathrm{e}^{|a|} &\leq n^{-1} \sqrt{\log n} \exp\left(\sqrt{\log n} \right) \\ &= \exp\left(\sqrt{\log n} + \log \sqrt{\log n} - \log n \right) \to 0 \end{aligned}$$

as $n \to \infty$, giving Part (a) for fixed m.

Part (b). Taking the modulus of (A.4), we have

$$\left| (1 - G_\lambda)^m \mathrm{e}^{at/n} \right| = \left| \left[\frac{\mu a^4}{1 + \mu a^4} + \zeta \right]^m \mathrm{e}^{at/n} \right| \leq \left| \frac{\mu a^4}{1 + \mu a^4} + \zeta \right|^m \left| \mathrm{e}^{at/n} \right|$$

$$\leq \left|\frac{\mu a^4}{1+\mu a^4} + |\zeta|\right|^m \mathrm{e}^{|a|} = \left|1 - \frac{1-|\zeta|(1+\mu a^4)}{1+\mu a^4}\right|^m \mathrm{e}^{|a|}.$$

For any $a \in \mathcal{A}(n \wedge m) \subseteq \mathcal{A}(n)$ the order of ζ in Lemma A.1 remains valid and thus

$$1 - |\zeta|(1+\mu a^4) = 1 - O(|a|/n + |a|^5/n) \geq 1/2$$

for m, n sufficiently large. The above inequality, together with $|a| \leq \log(m \wedge n) \leq \log m$, implies

$$\left|(1-G_\lambda)^m \mathrm{e}^{at/n}\right| \leq \left(1 - \frac{1/2}{1+\mu(\log m)^2}\right)^m \mathrm{e}^{\sqrt{\log m}}$$

uniformly for $a \in \mathcal{A}(n \wedge m)$ and $t \leq n$ with m, n sufficiently large, so that

$$\left(1 - \frac{1/2}{1+\mu(\log m)^2}\right)^m \mathrm{e}^{\sqrt{\log m}} \to \lim_{m\to\infty} \exp\left(-\frac{m/2}{1+\mu(\log m)^2} + \sqrt{\log m}\right) = 0,$$

giving the stated result for Part (b). □

The following Corollary A.1 is an immediate implication of Lemma 1 as the HP residual operator is repeatedly applied to trigonometric functions with increasingly higher frequencies.

Corollary A.1. *Suppose* $\lambda = \mu n^4$.

(a) For any fixed $m \in \mathbb{N}$, *if* $K_n = \lfloor \pi^{-1}\sqrt{\log n} \rfloor$, *then*

$$\begin{aligned}
\sup_{1\leq t\leq n,\, k\leq K_n} \left|\left[(1-G_\lambda)^m - \left(\frac{\mu}{\mu+\lambda_k^2}\right)^m\right] \varphi_k\left(\frac{t}{n}\right)\right| &\to 0 \\
\sup_{1\leq t\leq n,\, k\leq K_n} \left|\left[(1-G_\lambda)^m - \left(\frac{\mu}{\mu+\lambda_k^2}\right)^m\right] \psi_k\left(\frac{t}{n}\right)\right| &\to 0
\end{aligned}$$

as $n \to \infty$.

(b) If $K_{n,m} = \left\lfloor \pi^{-1}\sqrt{\log(n \wedge m)} \right\rfloor$, *then*

$$\begin{aligned}
\sup_{1\leq t\leq n,\, k\leq K_{n,m}} \left|(1-G_\lambda)^m \varphi_k(t/n)\right| &\to 0 \\
\sup_{1\leq t\leq n,\, k\leq K_{n,m}} \left|(1-G_\lambda)^m \psi_k(t/n)\right| &\to 0
\end{aligned}$$

as $n, m \to \infty$.

Proof of Corollary A.1. **Part (a).** The definitions of $\psi_k(\cdot)$ and $\varphi_k(\cdot)$ give

$$(1-G_\lambda)^m \left[\psi_k\left(\frac{t}{n}\right) + \mathbf{i}\varphi_k\left(\frac{t}{n}\right)\right] = \sqrt{2}\,(1-G_\lambda)^m \,\mathrm{e}^{\frac{\mathbf{i}(t/n)}{\sqrt{\lambda_k}}}.$$

Let $a = \mathbf{i}/\sqrt{\lambda_k}$. We verify

$$a^4 = \lambda_k^{-2} = [(k - 1/2)\,\pi]^4 \leq K_n^4 \pi^4 \leq (\log n)^2$$

satisfies the condition $a \in \mathcal{A}(n)$, and then Lemma 1 (a) ensures that for any fixed m

$$\sup_{1\leq t\leq n,\, k\leq K_n} \left| \left[(1 - G_\lambda)^m - \left(\frac{\mu}{\mu + \lambda_k^2} \right)^m \right] \mathrm{e}^{\frac{\mathbf{i}(t/n)}{\sqrt{\lambda_k}}} \right| \to 0$$

as $n \to \infty$. We complete the proof by separating the imaginary and the real parts of $\exp\left(\frac{\mathbf{i}(t/n)}{\sqrt{\lambda_k}}\right)$, respectively.

Part (b). Similarly, for $a = \mathbf{i}/\sqrt{\lambda_k}$ we verify that $a^4 \leq K_{n,m}^4 \pi^4 \leq (\log(n \wedge m))^2$. The fact $a \in \mathcal{A}(n \wedge m)$ allows us to invoke Lemma 1 (b):

$$\sup_{1\leq t\leq n\, k\leq K_{n,m}} \left| (1 - G_\lambda)^m \, \mathrm{e}^{\frac{\mathbf{i}(t/n)}{\sqrt{\lambda_k}}} \right| \to 0 \tag{A.5}$$

as $m, n \to \infty$, and then the results follow. □

Remark A.1. When setting $m = 1$, Lemma 1 and Corollary A.1 immediately imply

$$\left(G_\lambda - \frac{1}{\mu c^4 + 1} \right) \mathrm{e}^{ct/n} \to 0$$

for any $c \in \mathbb{R}$ uniformly over all $t \leq n$, and

$$\begin{aligned} \left(G_\lambda - \frac{\lambda_k^2}{\mu + \lambda_k^2} \right) \varphi_k\left(\frac{t}{n} \right) &\to 0, \\ \left(G_\lambda - \frac{\lambda_k^2}{\mu + \lambda_k^2} \right) \psi_k\left(\frac{t}{n} \right) &\to 0. \end{aligned} \tag{A.6}$$

uniformly for all $k \leq K_n$ under consideration. These real exponential functions, sine waves and cosine waves are the building blocks of the series representations of the higher order integrated processes and LUR processes.

A.2 Main Results

Proof of Proposition 1. The KL representation of $L_2(r)$ specified in (14) converges almost surely and uniformly in $[0, 1]$. Let the K_n-term finite KL representation be $B^{K_n}(r) := \sum_{k=1}^{K_n} \lambda_k(\sqrt{2} - \psi_k(r))\xi_k$. When $K_n \to \infty$ as $n \to \infty$, $\sup_{0\leq t\leq n} \left| B_2(t/n) - B_2^{K_n}(t/n) \right| = o_{a.s.}(1)$ and by uniform convergence

$$\sup_{0\leq t\leq n} \left| Y_n(t/n) - B_2^{K_n}(t/n) \right| = o_{a.s.}(1).$$

It follows that $Y_n(t/n)$ is almost surely uniformly well approximated by $B^{K_n}(t/n)$ for $t \leq n$ as $n \to \infty$. Hence the HP filtered trend has the following approximation

$$\begin{aligned}\frac{\widehat{f}_t^{\mathrm{HP}}}{n^{3/2}} &= G_\lambda \frac{y_t}{n^{3/2}} = G_\lambda \left[L^{K_n}\left(\frac{t}{n}\right) + o_{a.s.}(1) \right] \\ &= \sum_{k=1}^{K_n} \lambda_k \left[G_\lambda \left(\sqrt{2} - \psi_k\left(\frac{t}{n}\right)\right)\right] \xi_k + o_{a.s.}(1) \\ &= \sum_{k=1}^{K_n} \sqrt{2}\lambda_k \xi_k - \sum_{k=1}^{K_n} \lambda_k \xi_k G_\lambda \psi_k\Big(\frac{t}{n}\Big) + o_{a.s.}(1) \end{aligned} \tag{A.7}$$

as $n \to \infty$. The $o_{a.s.}(1)$ in (A.7) holds because the two sided moving average filter produced by the operator G_λ is an absolutely summable weighted moving average with stable geometric decay, which preserves the error order by majorization.

The asymptotic form of the HP filter can be written, according to (A.7), as

$$\begin{aligned}\frac{\hat{f}_{t,K_n}^{\mathrm{HP}}}{n^{3/2}} &= \sum_{k=1}^{K_n} \sqrt{2}\lambda_k \xi_k - \sum_{k=1}^{K_n} \lambda_k \xi_k \left[\frac{\lambda_k^2}{\mu + \lambda_k^2} \psi_k\left(\frac{t}{n}\right) + o(1)\right] + o_{a.s.}(1) \\ &= \sum_{k=1}^{K_n} \sqrt{2}\lambda_k \xi_k - \sum_{k=1}^{K_n} \frac{\lambda_k^3}{\mu + \lambda_k^2} \psi_k\left(\frac{t}{n}\right) \xi_k + o_{a.s.}(1).\end{aligned}$$

Note that $\lambda_k = 1/[(k-\frac{1}{2})\pi]^2$ and $\frac{\lambda_k^3}{\mu+\lambda_k^2} = O(k^{-6})$. Hence, the series $\sum_{k=1}^{\infty} \lambda_k \xi_k$ and $\sum_{k=1}^{\infty} \frac{\lambda_k^3}{\mu + \lambda_k^2} \psi_k\left(\frac{t}{n}\right) \xi_k$ converge almost surely and uniformly as $K_n \to \infty$. When $K_n = \lfloor \pi^{-1}\sqrt{\log n} \rfloor$ as $n \to \infty$, by Corollary A.1 (a) we obtain the following asymptotic form of the HP filter trend as

$$\frac{\hat{f}_t^{\mathrm{HP}}}{n^{3/2}} = \sum_{k=1}^{\infty} \left[\sqrt{2}\lambda_k - \frac{\lambda_k^3}{\mu + \lambda_k^2} \psi_k\left(\frac{t}{n}\right)\right] \xi_k + o_{a.s.}(1).$$

The proof is completed. □

Proof of Theorem 1. When $q = 1$, the convergence is already established in Theorem 1 of PS. In this proof, we focus on $q \geq 2$. By the uniform convergence law (11) and the KL representation of the $I(q)$ process in (12), the bHP estimated cycle has the following approximation

$$\begin{aligned}\frac{\widehat{c}_t^{(m)}}{n^{q-0.5}} &= (1 - G_\lambda)^m \frac{y_t}{n^{q-0.5}} = -(1 - G_\lambda)^m \left[L_q^{K_{n,m}}\left(\frac{t}{n}\right) + o_{a.s.}(1)\right] \\ &= -\sqrt{2} \sum_{k=1}^{K_{n,m}} \xi_k (1 - G_\lambda)^m \left[\sum_{\ell=1}^{\lfloor q/2 \rfloor} (-1)^{\ell-1} \lambda_k^j \frac{(t/n)^{q-2\ell}}{(q-2\ell)!} + \lambda_k^{q/2} \mathrm{Im}\left[(-\mathbf{i})^{q-1} \mathrm{e}^{\frac{\mathbf{i}(t/n)}{\sqrt{\lambda_k}}}\right]\right] \\ &\quad + o_{a.s.}(1)\end{aligned}$$

as $n \to \infty$ with $K_{n,m} = \left\lfloor \pi^{-1}\sqrt{\log(n \wedge m)} \right\rfloor$ by Corollary A.1 (b).

When $4m \geq q$ and $1 \leq \ell \leq \lfloor q/2 \rfloor$, the polynomial component is

$$\begin{aligned}
& (1-G_\lambda)^m (t/n)^{q-2\ell} \\
= & \frac{1}{\Gamma(m)} \int_0^\infty s^{m-1} \mathrm{e}^{-s\left(1+\lambda \mathbb{L}^{-2}(1-\mathbb{L})^4\right)} \left[\lambda \mathbb{L}(1-\mathbb{L})^4\right]^m \left(\frac{t}{n}\right)^{q-2\ell} ds \\
= & \frac{1}{\Gamma(m)} \int_0^\infty s^{m-1} \mathrm{e}^{-s} \sum_{j=0}^\infty \frac{(-1)^j s^j}{j!} \left[\lambda \mathbb{L}^{-2}(1-\mathbb{L})^4\right]^{m+j} \left(\frac{t}{n}\right)^{q-2\ell} ds \\
= & \frac{1}{\Gamma(m)} \int_0^\infty s^{m-1} \mathrm{e}^{-s} \sum_{j=0}^\infty \frac{(-1)^j s^j}{j!} \left[\lambda (1-\mathbb{L})^4\right]^{m+j} \left(\frac{t+2(m+j)}{n}\right)^{q-2\ell} ds \\
= & 0.
\end{aligned}$$

For the cyclical functions, since $\mathrm{Im}\left((-\mathbf{i})^{q-1} \mathrm{e}^{\frac{\mathbf{i}(t/n)}{\sqrt{\lambda_k}}}\right)$ is either $\pm\cos\left(t/\left(n\sqrt{\lambda_k}\right)\right)$ or $\pm\sin\left(t/\left(n\sqrt{\lambda_k}\right)\right)$ and $\lambda_k^{q/2} = O(k^{-2})$, the series $\sum_{k=1}^\infty \lambda_k^{q/2}\xi_k$ converges almost surely and uniformly for all $t \leq n$. Thus, according to (A.5) we have

$$\begin{aligned}
& \sum_{k=1}^{K_{n,m}} \xi_k (1-G_\lambda)^m \left[\sum_{\ell=1}^{\lfloor q/2 \rfloor} (-1)^{\ell-1} \lambda_k^j \frac{(t/n)^{q-2\ell}}{(q-2\ell)!} + \lambda_k^{q/2} \mathrm{Im}\left[(-\mathbf{i})^{q-1} \mathrm{e}^{\frac{\mathbf{i}(t/n)}{\sqrt{\lambda_k}}}\right]\right] \\
= & \sum_{k=1}^{K_{n,m}} \lambda_k^{q/2} \xi_k \mathrm{Im}\left((-\mathbf{i})^{q-1}\left[(1-G_\lambda)^m \mathrm{e}^{\frac{\mathbf{i}(t/n)}{\sqrt{\lambda_k}}}\right]\right) = \sum_{k=1}^{K_{n,m}} \lambda_k^{q/2} \xi_k \cdot o(1) = o_{a.s.}(1)
\end{aligned}$$

when $K_{n,m} = \left\lfloor \pi^{-1}\sqrt{\log(n \wedge m)} \right\rfloor$. This confirms that $\widehat{c}_t^{(m)}/n^{q-0.5} = o_{a.s.}(1)$ uniformly over $t \leq n$, and thus $n^{0.5-q} \cdot \hat{f}^{(m)}_{\lfloor nr \rfloor} \rightsquigarrow L_q(r)$ as stated. □

Before establishing the results for the LUR case it is convenient to derive the following series representation

$$\begin{aligned}
J_c(r) &= B(r) + c\int_0^r \mathrm{e}^{(r-s)c} B(s) ds \\
&= B(r) + \sqrt{2}c\mathrm{e}^{cr} \sum_{k=1}^\infty \xi_k \sqrt{\lambda_k} \int_0^r \mathrm{e}^{-sc} \sin\left(\frac{s}{\sqrt{\lambda_k}}\right) ds \\
&= B(r) + \sqrt{2}c\mathrm{e}^{cr} \sum_{k=1}^\infty \xi_k \frac{\lambda_k^{3/2}}{\lambda_k c^2 + 1} \left(\frac{1}{\sqrt{\lambda_k}} - \mathrm{e}^{-cr} c \sin\left(\frac{r}{\sqrt{\lambda_k}}\right) - \frac{\mathrm{e}^{-cr}}{\sqrt{\lambda_k}} \cos\left(\frac{r}{\sqrt{\lambda_k}}\right)\right) \\
&= \sqrt{2} \sum_{k=1}^\infty \xi_k \frac{\sqrt{\lambda_k}}{c^2\lambda_k + 1} \sin\left(\frac{r}{\sqrt{\lambda_k}}\right) + \sqrt{2}c \sum_{k=1}^\infty \xi_k \frac{\lambda_k}{\lambda_k c^2 + 1} \left(\mathrm{e}^{cr} - \cos\left(\frac{r}{\sqrt{\lambda_k}}\right)\right) \\
&= \sum_{k=1}^\infty \frac{\sqrt{2}c\lambda_k \mathrm{e}^{cr} + \sqrt{\lambda_k}\varphi_k(r) - c\lambda_k \psi_k(r)}{\lambda_k c^2 + 1} \xi_k, \qquad \text{(A.8)}
\end{aligned}$$

as in Phillips (1998). This representation is needed in the following proofs.

Proof of Proposition 2. The series presentation (20) converges almost surely and uniformly over r. It is approximated by the K_n-term representation

$$J_c^{K_n}(r) = \sum_{k=1}^{K_n} \frac{\sqrt{2}c\lambda_k \mathrm{e}^{cr} + \sqrt{\lambda_k}\varphi_k(r) - c\lambda_k\psi_k(r)}{\lambda_k c^2 + 1}\xi_k$$

in the sense of $\sup_{0\le t\le n}\left|J_c(r) - J_c^{K_n}(r)\right| = o_{a.s.}(1)$ when $K_n \to \infty$ as $n \to \infty$, so that in the expanded space $\sup_{0\le t\le n}\left|n^{-1/2}y_{\lfloor nr\rfloor} - J_c^{K_n}(r)\right| = o_{a.s.}(1)$ by the uniform convergence (19). The HP estimated trend is then approximated as

$$\begin{aligned}\frac{\hat{f}_t^{\mathrm{HP}}}{n^{1/2}} &= G_\lambda \frac{y_t}{n^{1/2}} = G_\lambda\left[J_c^{K_n}\left(\frac{t}{n}\right) + o_{a.s.}(1)\right] \\ &= \sum_{k=1}^{K_n}\left[G_\lambda \frac{\sqrt{2}c\lambda_k \mathrm{e}^{ct/n} + \sqrt{\lambda_k}\varphi_k(t/n) - c\lambda_k\psi_k(t/n)}{\lambda_k c^2 + 1}\right]\xi_k + o_{a.s.}(1).\end{aligned} \tag{A.9}$$

In view of Remark A.1, when $K_n = \left\lfloor \pi^{-1}\sqrt{\log n}\right\rfloor$ we have

$$\sum_{k=1}^{K_n} G_\lambda \frac{\sqrt{2}c\lambda_k\xi_k \mathrm{e}^{ct/n}}{\lambda_k c^2+1} = \sum_{k=1}^{K_n}\frac{\sqrt{2}c\lambda_k\xi_k}{\lambda_k c^2+1}\left(\frac{\mathrm{e}^{ct/n}}{\mu c^4+1} + o(1)\right) = \frac{\sqrt{2}c\mathrm{e}^{ct/n}}{\mu c^4+1}\sum_{k=1}^{K_n}\frac{\lambda_k}{\lambda_k c^2+1}\xi_k + o_{a.s.}(1), \tag{A.10}$$

since $\sum_{k=1}^{K_n}\frac{\lambda_k}{\lambda_k c^2+1}\xi_k \sim N\left(0, \omega^2\sum_{k=1}^{K_n}\frac{\lambda_k^2}{(\lambda_k c^2+1)^2}\right)$ with variance bounded by

$$\omega^2\sum_{k=1}^{K_n}\frac{\lambda_k^2}{(\lambda_k c^2+1)^2} \le \omega^2\sum_{k=1}^{\infty}\frac{\lambda_k^2}{(\lambda_k c^2+1)^2} \le \omega^2\sum_{k=1}^{\infty}\lambda_k^2 = \frac{\omega^2}{6}. \tag{A.11}$$

Similarly,

$$\begin{aligned}\sum_{k=1}^{K_n} G_\lambda\frac{\sqrt{\lambda_k}\varphi_k(t/n)}{\lambda_k c^2+1} &= \sum_{k=1}^{K_n}\frac{\sqrt{\lambda_k}\xi_k}{\lambda_k c^2+1}\left(\frac{\lambda_k^2}{\mu+\lambda_k^2}\varphi_k\left(\frac{t}{n}\right) + o(1)\right) \\ &= \sum_{k=1}^{K_n}\frac{\lambda_k^2}{\mu+\lambda_k^2}\cdot\frac{\sqrt{\lambda_k}}{\lambda_k c^2+1}\varphi_k\left(\frac{t}{n}\right)\xi_k + o_{a.s.}(1)\end{aligned} \tag{A.12}$$

and

$$\begin{aligned}\sum_{k=1}^{K_n} G_\lambda\frac{c\lambda_k\xi_k\psi_k(t/n)}{\lambda_k c^2+1} &= \sum_{k=1}^{K_n}\frac{c\lambda_k\xi_k}{\lambda_k c^2+1}\left(\frac{\lambda_k^2}{\mu+\lambda_k^2}\psi_k\left(\frac{t}{n}\right) + o(1)\right) \\ &= c\sum_{k=1}^{K_n}\frac{\lambda_k^2}{\mu+\lambda_k^2}\cdot\frac{\lambda_k}{\lambda_k c^2+1}\psi_k\left(\frac{t}{n}\right)\xi_k + o_{a.s.}(1)\end{aligned} \tag{A.13}$$

as $n \to \infty$ by virtue of uniform almost surely convergence. Substituting (A.10), (A.12), and (A.13)

into (A.9) yields

$$\frac{\hat{f}_t^{\mathrm{HP}}}{n^{1/2}} = \sum_{k=1}^{K_n} \frac{1}{\lambda_k c^2 + 1} \left[\frac{\sqrt{2} c \lambda_k \mathrm{e}^{ct/n}}{\mu c^4 + 1} + \frac{\lambda_k^2}{\mu + \lambda_k^2} \left(\sqrt{\lambda_k} \varphi_k(t/n) - c \lambda_k^2 \psi_k(t/n) \right) \right] \xi_k + o_{a.s.}(1)$$

uniformly for all $t \leq n$. The limiting expression (21) follows as K_n passes to infinity as $n \to \infty$. □

Proof of Theorem 2. By virtue of the uniform convergence (19) the estimated residual in the expanded probability space is

$$\begin{aligned} \frac{\widehat{c}_t^{(m)}}{n^{1/2}} &= (1 - G_\lambda)^m \left[J_c^{K_{n,m}} \left(\frac{t}{n} \right) + o_{a.s.}(1) \right] \\ &= \sum_{k=1}^{K_{n,m}} (1 - G_\lambda)^m \frac{\sqrt{2} c \lambda_k \mathrm{e}^{\frac{ct}{n}} + \sqrt{\lambda_k} \varphi_k(\frac{t}{n}) - c \lambda_k \psi_k(\frac{t}{n})}{\lambda_k c^2 + 1} \xi_k + o_{a.s.}(1). \end{aligned} \tag{A.14}$$

In view of Lemma 1 and Corollary A.1, when $K_{n,m} = \left\lfloor \pi^{-1} \sqrt{\log(n \wedge m)} \right\rfloor$ as $m, n \to \infty$, we have

$$\sum_{k=1}^{K_{n,m}} (1 - G_\lambda)^m \frac{\sqrt{2} c \lambda_k \xi_k \mathrm{e}^{ct/n}}{\lambda_k c^2 + 1} = \sqrt{2} c \sum_{k=1}^{K_{n,m}} \frac{\lambda_k}{\lambda_k c^2 + 1} \xi_k \cdot o(1) = o_{a.s.}(1), \tag{A.15}$$

$$\sum_{k=1}^{K_{n,m}} (1 - G_\lambda)^m \frac{\sqrt{\lambda_k} \varphi_k(t/n)}{\lambda_k c^2 + 1} = \sum_{k=1}^{K_{n,m}} \frac{\sqrt{\lambda_k}}{\lambda_k c^2 + 1} \xi_k \cdot o(1) = o_{a.s.}(1), \tag{A.16}$$

$$\sum_{k=1}^{K_{n,m}} (1 - G_\lambda)^m \frac{c \lambda_k \xi_k}{\lambda_k c^2 + 1} \psi_k(\frac{t}{n}) = c \sum_{k=1}^{K_{n,m}} \frac{\lambda_k}{\lambda_k c^2 + 1} \xi_k \cdot o(1) = o_{a.s.}(1), \tag{A.17}$$

uniformly over $t \leq n$, as in (A.16) the random component

$$\sum_{k=1}^{K_{n,m}} \frac{\sqrt{\lambda_k}}{\lambda_k c^2 + 1} \xi_k \sim N \left(0, \omega^2 \sum_{k=1}^{K_{n,m}} \frac{\lambda_k}{(\lambda_k c^2 + 1)^2} \right)$$

has a finite variance

$$\omega^2 \sum_{k=1}^{K_{n,m}} \frac{\lambda_k}{(\lambda_k c^2 + 1)^2} \leq \omega^2 \sum_{k=1}^{\infty} \frac{\lambda_k}{(\lambda_k c^2 + 1)^2} \leq \omega^2 \sum_{k=1}^{\infty} \lambda_k = \frac{\omega^2}{2},$$

and the orders in (A.15) and (A.17) are controlled by an argument as in (A.11). We thus conclude that the leading term in (A.14) is also $o_{a.s.}(1)$, that is, $n^{-1/2} \widehat{c}_t^{(m)} = o_{a.s.}(1)$ uniformly for all $t \leq n$. It follows that $\sup_{0 \leq t \leq n} \left| n^{-1/2} \hat{f}_t^{(m)} - J_c(t/n) \right| = o_{a.s.}(1)$ in the expanded probability space and $n^{-1/2} \hat{f}_{\lfloor nr \rfloor}^{(m)}$ weakly converges to $J_c(r)$ in the original space. □

Table B.1: MSE of the Estimated Trends with $\lambda = 1.6 \times 10^{-5} n^4$: LUR

DGP	n	$c=3$				$c=0$				$c=-3$			
		HP	2HP	bHP	AR	HP	2HP	bHP	AR	HP	2HP	bHP	AR
6	100	2.77	2.16	1.85	3.46	2.11	1.93	1.81	3.12	2.12	1.94	1.83	3.04
	200	3.81	2.99	2.01	3.44	3.42	2.89	1.98	3.27	3.44	2.92	2.00	3.20
	300	5.35	4.15	2.24	3.43	4.88	4.06	2.18	3.31	4.88	4.06	2.19	3.27
7	100	3.78	2.31	1.90	4.32	3.12	2.07	1.83	4.12	3.13	2.09	1.85	4.14
	200	4.81	3.15	2.08	3.74	4.46	3.06	2.06	3.65	4.48	3.09	2.08	3.65
	300	6.40	4.33	2.35	3.58	5.95	4.24	2.30	3.52	5.93	4.24	2.31	3.52
8	100	3.61	2.45	1.97	4.31	2.93	2.21	1.89	4.37	2.95	2.23	1.91	4.40
	200	5.00	3.29	2.14	3.78	4.67	3.21	2.15	3.81	4.70	3.24	2.18	3.81
	300	6.90	4.50	2.44	3.61	6.49	4.43	2.44	3.63	6.45	4.42	2.44	3.63
9	100	7.87	6.48	4.97	7.44	2.94	2.58	2.35	4.50	2.32	2.10	2.01	4.12
	200	4.57	3.80	2.63	4.27	3.44	2.91	2.17	3.95	3.02	2.58	2.01	3.82
	300	4.77	3.93	2.39	3.90	4.15	3.45	2.19	3.79	3.77	3.17	2.07	3.71
10	100	8.66	6.62	5.06	8.09	3.73	2.72	2.38	5.16	3.11	2.23	2.04	4.84
	200	5.37	3.95	2.68	4.55	4.26	3.07	2.22	4.20	3.85	2.74	2.05	4.08
	300	5.62	4.10	2.46	4.05	5.00	3.63	2.25	3.91	4.62	3.34	2.13	3.84

B Additional Numerical Results

Remarks 4 and 5 following Proposition 2 predict that when $\lambda = \mu n^4$ the residual from the HP filter will retain a near explosive component involving the factor e^{cr} when $c > 0$, which suggests that MSEs should be larger at the localizing coefficient $c = 3$ than at $c = -3$, ceteris paribus. This outcome is observed in Table 2 when $n = 100$ for quarterly data and when $n = 300$ for monthly data, but is unclear in the larger sample sizes because the tuning parameter λ is kept to $\lambda = 1600$ and $\lambda = 129600$ in Table 2. In further confirmation of Remarks 4 and 5, Table B.1 reports MSE results for the same DGPs and estimation methods as in Table 2 but using the tuning parameter $\lambda = 1.6 \times 10^{-5} n^4$. For $n = 100$ or 300, we have $\lambda = 1600$ and 129600 and the results in these cases in Table B.1 are the same as those in the corresponding cells of Table 2. The same holds for the AR filter because λ is irrelevant in the autoregression.

Consider the LUR case of DGP6. The MSE of the HP filter under $c = 3$ is the largest, followed by $c = -3$ which in turn exceed those of $c = 0$. These outcomes are fully consistent with theory as the exponential factor e^{cr} is present in both the near explosive ($c > 0$) and near stationary ($c < 0$) cases in (22). The HP filter fails to completely catch the exponential factor effects because it removes only polynomial trends up to the third order. In contrast, when $c = 0$ the exponential function factor is no longer present in (22), so that the HP filter MSE slightly improves when $c = 0$ relative to $c = -3$. The HP MSEs have similar rankings over c for DGPs 7–8, whereas the MSEs in DGPs 9–10 are primarily affected by the presence of a structural break. The MSEs of the bHP filter show much smaller differences between near explosive, unit root, and near stationary cases. These results further confirm the robustness of bHP's capabilities in trend-cycle determination in LUR models.