Yale University

EliScholar – A Digital Platform for Scholarly Publishing at Yale

Cowles Foundation Discussion Papers | Cowles Foundation

5-1-2019

Boosting the Hodrick-Prescott Filter

Peter C.B. Phillips

Zhentao Shi

Follow this and additional works at: https://elischolar.library.yale.edu/cowles-discussion-paper-series

 Part of the Economics Commons

Recommended Citation

Phillips, Peter C.B. and Shi, Zhentao, "Boosting the Hodrick-Prescott Filter" (2019). *Cowles Foundation Discussion Papers*. 60.
https://elischolar.library.yale.edu/cowles-discussion-paper-series/60

This Discussion Paper is brought to you for free and open access by the Cowles Foundation at EliScholar – A Digital Platform for Scholarly Publishing at Yale. It has been accepted for inclusion in Cowles Foundation Discussion Papers by an authorized administrator of EliScholar – A Digital Platform for Scholarly Publishing at Yale. For more information, please contact elischolar@yale.edu.

BOOSTING THE HODRICK-PRESCOTT FILTER

By

Peter C.B. Phillips and Zhentao Shi

May 2019

COWLES FOUNDATION DISCUSSION PAPER NO. 2192



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

http://cowles.yale.edu/

arXiv:1905.00175v1 [econ.EM] 1 May 2019

Boosting the Hodrick-Prescott Filter

Peter C.B. Phillips

Yale University, University of Auckland

University of Southampton and Singapore Management University

Zhentao Shi

The Chinese University of Hong Kong

Abstract

The Hodrick-Prescott (HP) filter is one of the most widely used econometric methods in applied macroeconomic research. The technique is nonparametric and seeks to decompose a time series into a trend and a cyclical component unaided by economic theory or prior trend specification. Like all nonparametric methods, the HP filter depends critically on a tuning parameter that controls the degree of smoothing. Yet in contrast to modern nonparametric methods and applied work with these procedures, empirical practice with the HP filter almost universally relies on standard settings for the tuning parameter that have been suggested largely by experimentation with macroeconomic data and heuristic reasoning about the form of economic cycles and trends. As recent research has shown, standard settings may not be adequate in removing trends, particularly stochastic trends, in economic data. This paper proposes an easy-to-implement practical procedure of iterating the HP smoother that is intended to make the filter a smarter smoothing device for trend estimation and trend elimination. We call this iterated HP technique the *boosted HP filter* in view of its connection to L_2-boosting in machine learning. The paper develops limit theory to show that the boosted HP filter asymptotically recovers trend mechanisms that involve unit root processes, deterministic polynomial drifts, and polynomial drifts with structural breaks – the most common trends that appear in macroeconomic data and current modeling methodology. In doing so, the boosted filter provides a new mechanism for consistently estimating multiple structural breaks. A stopping criterion is used to automate the iterative HP algorithm, making it a data-determined method that is ready for modern data-rich environments in economic research. The methodology is illustrated using three real data examples that highlight the differences between simple HP filtering, the data-determined boosted filter, and an alternative autoregressive approach. These examples show that the boosted HP filter is helpful in analyzing a large collection of heterogeneous macroeconomic time series that manifest various degrees of persistence, trend behavior, and volatility.

Key words: Boosting, Cycles, Empirical macroeconomics, Hodrick-Prescott filter, Machine learning, Nonstationary time series, Trends, Unit root processes

JEL codes: C22, C55, E20

Peter C. B. Phillips: peter.phillips@yale.edu. Zhentao Shi: zhentao.shi@cuhk.edu.hk. Phillips acknowledges research support from the Kelly Foundation at the University of Auckland and an LKC Fellowship at Singapore Management University. Shi acknowledges support from the Hong Kong Research Grants Council Early Career Scheme No. 24614817. We thank Yang Chen for excellent research assistance and Zheng Song for helpful comments.

The principle adopted here in the construction of a trend for a time series consists in minimizing a linear combination of two sums of squares, of which one refers to the second differences of the trend values, the other to the deviations of the observations from the trend values ... this procedure seems a particularly natural one when dealing with economic time series. The resulting family of trends may be described as quasi-linear trends. Leser (1961)

Our statistical approach does not utilize standard time series analysis. The maintained hypothesis based on growth theory considerations is that the growth component of aggregate economic times series varies smoothly over time. Hodrick and Prescott (1997)

1 Introduction

Two prominent features of macroeconomic data are trending long-run growth in aggregate economic activity and a cyclical component that represents fluctuations in this activity over shorter periods known as business cycles. Modern macroeconomic theory of the business cycle, evident in the vast literature on RBC and DSGE modeling, seeks to explain the cyclical movement and co-movement of macroeconomic variables about long-run trends. Both aspects of economic activity are important in economic analysis and policy making. Trends are an intrinsic element in determining long term economic prospects and the overall health of an economy. Cyclical behavior is especially important to policy makers, who are interested in understanding past contractions with a view to foreseeing the onset of future recessions and minimizing their impact on employment and business activity.

To analyze business cycles in observed data it is necessary to isolate the cyclical component from the trend. This decomposition is commonly achieved by regression or filtering. The latter is primarily motivated by the prior view that a trend is distinguished as a more smoothly varying component in relation to the observed data, a concept reflected in the header quotation from Hodrick and Prescott (1997) (hereafter HP), leading to the so-called HP filter. This filter belongs to the statistical approach that was introduced by Whittaker (1922) and Whittaker and Robinson (1924), who provided a probabilistic framework of penalized maximum likelihood estimation to deliver a quantitative measure of trend (or graduation in their terminology). The explicit form of the smoothness measure penalty involving squared second differences in HP (1997) was used in earlier work by Leser (1961), who emphasized its relevance to trend determination with economic data on the grounds of its 'quasi-linear' trend-producing properties, as indicated in the primary header quotation.

The HP method is now widely used in applied macroeconomic work by economists in central banks, international economic agencies, industry, and government. Its use in academic work is less extensive, partly because it has been subject over many years to considerable criticism and analyses that have revealed a myriad of its limitations for empirical studies in economics, an early example being Cogley and Nason (1995). For recent discussion of the merits and demerits of the filter, see Phillips and Jin (2015) (PJ, henceforth) and Hamilton (2018) and the many references cited therein.

Given time series data $(x_t : t = 1, \ldots, n)$, the HP method decomposes the series into two additive components — a trend component (f_t), and a residual or cyclical component (c_t), estimated as

$$\left(\widehat{f}_t^{\,\mathrm{HP}}\right) = \arg\min_{(f_t)} \left\{ \sum_{t=1}^{n} (x_t - f_t)^2 + \lambda \sum_{t=2}^{n} \left(\Delta^2 f_t\right)^2 \right\}, \text{ and } \quad \left(\widehat{c}_t^{\,\mathrm{HP}}\right) = \left(x_t - \widehat{f}_t^{\,\mathrm{HP}}\right) \tag{1}$$

where $\Delta f_t = f_t - f_{t-1}$, $\Delta^2 f_t = \Delta f_t - \Delta f_{t-1} = f_t - 2f_{t-1} + f_{t-2}$, and $\lambda \geq 0$ is a tuning parameter that controls the extent of the penalty. As is apparent from this criterion, the method is nonparametric and the choice of λ inevitably plays a major role in determining the shapes of the fitted trend and cycle functions. If λ is selected too large, the fitted trend becomes nearly linear, as implied by

Leser's (1961) characterization, and a linear trend is indeed the solution as $\lambda \to \infty$. In consequence, if the trend is nonlinear and λ is too large, the HP fitted trend produces a residual trend that contaminates the cyclical component. If λ is selected too small, the fitted trend becomes highly flexible, so that it closely tracks the data and thereby embodies elements of short-term fluctuations. In the absence of an underlying model that defines and enables direct quantification of trend and cycle, some degree of cycle distortion with the HP filter is inevitable. Of course, precisely the same criticism applies to other smoothing techniques, as well as regression methods of trend extraction when the trend model is misspecified, as is nearly always the case in practical work.

In the application of modern nonparametric methods, serious efforts are normally made to choose tuning parameters based on some well-defined optimization criterion with data-determined versions of these criteria that can be implemented in empirical work. This approach is greatly facilitated by the use of an underlying model representing the generating mechanism. By contrast, empirical practice with the HP filter almost universally relies on standard settings for the tuning parameter that have been suggested largely by experimentation with macroeconomic data and heuristic reasoning about the form of economic cycles and trends. For quarterly data the standard choice is $\lambda = 1600$, as recommended by HP (1997) based on their experimentation with US data. This value has served as a gold standard in converting the tuning parameter to other sampling frequencies such as annual or monthly data (Ravn and Uhlig, 2002). Importantly, these HP filter smoothing parameter settings are normally employed irrespective of the sample size of the data in contrast to standard nonparametric methods.

HP filtering retains an agnostic position with regard to the exact functional form of the trend and cycle. This approach has the advantage of generality, but the central weakness is that its properties with respect to consistent trend estimation are poorly understood, especially with regard to the relevant choice of the tuning parameter. Recent research by PJ (2015) has shown that the performance of the filter as a potentially consistent estimator of trends of certain types can be assessed directly in relation to the choice of the tuning parameter, just as in standard nonparametric analysis as the sample size $n \to \infty$. That work revealed the constraints that are needed on λ in order to achieve consistent trend estimation for polynomial trends and stochastic process trends. The paper also analyzed the effect of the filter on trends with structural breaks. In doing so, PJ showed how the standard HP filter settings may not be adequate in removing trends — particularly stochastic trends — in economic data, a conclusion that reverses earlier thinking (King and Rebelo, 1993) to the effect that the HP filter removes up to four unit roots in the original data. PJ further showed that the common choice of the tuning parameter $\lambda = 1600$ for quarterly data is typically too large to completely remove stochastic process trends given the length of the macroeconomic time series usually encountered in practice. It is likely therefore that remnants of stochastic trends linger in the fitted cycles of much applied research, a conclusion supported by the recent examples exhibited in Hamilton (2018) and the earlier work by Cogley and Nason (1995).

The present paper proposes an easy-to-implement modification that is designed to make the HP filter more effective in trend fitting and trend elimination, and is formulated with data-determined smoothing choices to aid practical application. The idea is simple. Since the cyclical component may often retain trend elements, we feed the data into the filter again to clean the leftover elements. The notion of refitting the residual in statistical applications goes back to Tukey (1977) under the name *twicing*, where procedures were employed twice to assist in data-cleaning exercises. The notion of twicing can be continued into "*n*–ing", as discussed in Buja, Hastie, and Tibshirani (1989). This approach motivates the present proposal of repeated application of the HP filter. Since the solution of the HP filter can be explicitly expressed as a linear operation, which will be made clear in Section 2, repeated HP fitting is closely related to the L_2-boosting (Bühlmann and Yu, 2003) procedure which is now commonly used in machine learning.

The existing statistical theory on boosting is developed mostly for environments with independent identically distributed (iid) data. We are unaware of any earlier work concerned with the use of

L_2-boosting on stochastically trending data or nonstationary time series. A primary contribution of the present paper is to apply and analyze the idea of repeated fitting algorithms in machine learning to trend detection and nonstationary time series environments. In particular, we establish asymptotic theory to justify the use of the boosted version of the HP filter, extending earlier work by PJ on the asymptotics of the HP filter. We find that if the number of iterations slowly diverges with the sample size, the boosted HP filter (BHP filter, hereafter) can recover the underlying trend irrespective of whether the time series contains a stochastic process trend, a deterministic polynomial drift, or a polynomial drift with a structural break. An interesting implication of the asymptotic theory in the latter case is that the BHP filter effectively delivers a consistent estimator of the break point. This result extends to the case of multiple structural breaks, so the boosted filter provides a new device for consistently estimating multiple break points, delivered automatically without additional methods of detection.

Use of machine learning methods in econometrics has grown quickly in recent years with a particular focus on applied microeconomics where large cross sectional and wide panel datasets are available. The low-frequency nature of most macroeconomic time series, on the other hand, means that the volume of today's macroeconomic databases is much smaller by comparison and is, of course, completely dwarfed by those generated from Internet communication.[1] Nonetheless, the phenomena that macroeconomists study are no less complex than those that confront scientists working with 'big data' and the scope for machine learning methods is large. For example, in empirical analysis using country-level macro indicators, macroeconomists handle data collected from a large number of heterogeneous economies in highly varying stages of economic development. In such cases, where there are many possible determining factors and diverse time series trajectories, automated econometric procedures (e.g., Phillips (2005)) can be of tremendous appeal in practical work, as has been argued recently by many authors since the development of high dimensional regression methods such as Lasso (Tibshirani, 1996).

This line of thinking partly motivates the present econometric implementation of machine learning methods in which we propose two data-driven stopping rules to terminate iterations of the BHP filter. One rule ceases the iteration according to the outcome of a unit root test, and is appropriate when stationary time series is considered a prerequisite for further investigation, such as business cycle analysis. The second rule relies on a new version of the Bayesian information criterion (BIC) that is developed for the present BHP filter framework to take into account sample fit and effective degrees of freedom after each iteration. The latter approach accords with the now common practice of using BIC-type information criteria as stopping rules in econometric work.

We conduct three real data applications of the HP and BHP filters to study the impact of our iterative fitting algorithm. The first application revisits Okun's law, which posits an empirical association of comovement between real GDP and unemployment, in an international cross country setting. Ball, Leigh, and Loungani (2017) extend the scope of Okun (1962)'s original focus on the United States to 20 OECD countries, applying the standard HP filter to each time series to obtain deviations from long run levels. Accordingly, when the standard filter fails to remove the time trend, the resulting Okun law regression can suffer spurious regression effects. The BHP filter helps to mitigate the effects of potential contamination from long run influences.

A second example conducts a cross country comparison of business cycles. In an extensive application of HP methods to remove trend, Aguiar and Gopinath (2007) suggest that cyclical components are more persistent and volatile in emerging economies than in developed countries. In revisiting this application using the BHP filter, we find that the time series collected from the emerging economies are much shorter than those from the developed ones, and so the use of the standard $\lambda = 1600$ tuning parameter uniformly across all countries tends to over-penalize the

[1]For instance, the size of the latest version of FRED-MD (as of April 2019) is smaller than 600kb. FRED-MD is the Monthly Databases for Macroeconomic Research (`https://research.stlouisfed.org/econ/mccracken/fred-databases/`). See McCracken and Ng (2016) for details.

shorter series. Repeated fitting helps to regularize the unbalanced panel and robustify the finding by Aguiar and Gopinath (2007) of the empirical distinction in cyclical behavior between emerging and developed economies.

The third application examines US industrial production over the last century from 1919 to 2018. Like many other macroeconomic time series this series displays strong trend characteristics with some major fluctuations over subperiods that include two world wars and the great depression during the early part of the period, and the financial crisis and great recession over the latter period. As such, the series presents challenges in trend determination that include the complex issue of whether such subperiods are better interpreted as part of the trend or part of the evolving cyclical processes of modern industrialized economies. In this application, we provide a detailed comparison of the HP and BHP filters with the alternative autoregressive modeling approach recently advocated by Hamilton (2018).

We close this introduction with a brief discussion of related literature on filtering and boosting. First, there are now many competing methods of data filtering to remove trend such as the band pass filter methods of Baxter and King (1999), Christiano and Fitzgerald (2003), and Corbae and Ouliaris (2006). Most of these share many common characteristics with the HP filter. Nonetheless, the HP filter remains the most popular[2] in practical work and serves as a benchmark for other agnostic methods of trend extraction. In addition, there has been renewed recent interest in the theoretical properties of the HP filter, useful algebraic representations, and computational algorithms. Phillips (2010) and PJ (2015) provide exact matrix and operator representations and new asymptotics. Cornea-Madeira (2017) gives an explicit algebraic formula for the HP filter in finite samples. De Jong and Sakarya (2016) and Sakarya and de Jong (2017) provide further finite sample results, including another representation of the HP filter as a symmetric weighted average plus some adjustments. Hamilton (2018) provides a cautionary note concerning the limitations of the HP filter approach, reinforcing earlier warnings and suggesting the alternative of scalar autoregression.

Second, machine learning methods have been employed to generate new statistical procedures specifically tailored for economic applications in recent work by Belloni, Chen, Chernozhukov, and Hansen (2012), Belloni, Chernozhukov, and Hansen (2014), Chernozhukov, Hansen, and Spindler (2015), Fan, Liao, and Yao (2015), Hirano and Wright (2017) and Caner and Kock (2018), to name a few. Boosting is one of the most successful machine learning methods. Originally proposed for classification problems (Freund and Schapire, 1995), boosting has given rise to many useful variants (e.g. Hastie, Tibshirani, and Friedman (2009)). Bühlmann and Yu (2003) extended the idea of refitting to linear regression with the L_2 norm for the residuals, opening up a wide range of potential applications. In high dimensional regression, component-wise boosting is also related to forward stage selection and the greedy algorithm (Bühlmann, 2006). The idea of refitting (specifically twicing) was introduced by Tukey (1977) and appeared in econometrics in Newey, Hsieh, and Robins (2004). In high-dimensional regression, Bai and Ng (2009) employed boosting in macroeconomic forecasting. Shi (2016) used boosting to select relevant moments in structural models defined by many moment conditions. Ng (2014) and Luo and Spindler (2017) applied boosting to recession forecasting and other economic examples.

The rest of the paper is organized as follows. Section 2 introduces the iterative algorithm for boosting the HP filter and develops asymptotic theory that characterizes the behavior of the boosted filter, giving conditions for consistent estimation of stochastic process and deterministic polynomial trends with possible structural breaks. Stopping rules are provided to automate the procedure for practical work. Simulations are conducted in Section 3 to reveal the effect of boosting along with the stopping rules. Section 4 reports three empirical applications of the BHP methodology. Section 5 concludes.

[2]As of April, 2019, the article by Hodrick and Prescott (1997) had 8,540 listed citations and the article by Baxter and King (1999) 3,559 citations in Google Scholar.

2 The Boosted HP Filter

2.1 The Boosting Algorithm

The optimization problem (1) leading to the HP filter and related criteria for general filters of this type have closed-form algebraic solutions in convenient matrix form.[3] In the HP case, if D' is the rectangular $(n-2) \times n$ matrix with second differencing vector $d = \begin{pmatrix} 1 & -2 & 1 \end{pmatrix}'$ along the leading tri-diagonals and I_n is the $n \times n$ identity matrix, the explicit form of the trend solution is

$$\widehat{f}^{\mathrm{HP}} = Sx, \tag{2}$$

where $S = (I_n + \lambda DD')^{-1}$ is a deterministic operator and $x = (x_1, ..., x_n)'$ is the sample data. The smoothed component $\widehat{f}^{\mathrm{HP}}$ is interpreted as the estimated trend and

$$\widehat{c}^{\mathrm{HP}} = x - \widehat{f}^{\mathrm{HP}} = (I_n - S)\, x \tag{3}$$

as the estimated cyclical or stationary component.

The behavior and asymptotic properties of the estimated trend $\widehat{f}^{\mathrm{HP}}$ crucially depend on the choice of the tuning parameter and the underlying generating mechanism of x_t. For macroeconomic data the mechanism may reasonably be expected to involve a stochastic trend, possibly accompanied by some deterministic drift component that may be well modeled by a low order polynomial or a similar deterministic function subject to breaks. In the prototypical case of a unit root process, x_t satisfies under quite general conditions the functional law (Phillips and Solo, 1992)

$$\frac{x_{\lfloor nr \rfloor}}{\sqrt{n}} \rightsquigarrow B(r) \tag{4}$$

where $\lfloor \cdot \rfloor$ is the floor function, B is Brownian motion with variance ω^2 given by the long run variance of Δx_t, and $\rightsquigarrow$ signifies weak convergence, here on the Skorohod space $D[0,1]$.

The problem of consistent HP filter estimation of the trend then amounts to whether as $n \to \infty$ we have $\frac{\widehat{f}^{\mathrm{HP}}_{\lfloor nr \rfloor}}{\sqrt{n}} \rightsquigarrow B(r)$, in which case the filter asymptotically captures the underlying stochastic process trend in x_t. PJ show that this reproduction of the asymptotic form of the trend holds only under special restrictions on the smoothing parameter λ that ensure it does not diverge too quickly. In particular, if $\lambda = O(n^4)$ or greater as $n \to \infty$, then the HP filter trend $\widehat{f}^{\mathrm{HP}}_{\lfloor nr \rfloor}$ is inconsistent. In such cases, $\frac{\widehat{f}^{\mathrm{HP}}_{\lfloor nr \rfloor}}{\sqrt{n}} \rightsquigarrow f^{\mathrm{HP}}(r)$ where $f^{\mathrm{HP}}(r)$ is a smooth stochastic process different from $B(r)$, which implies that the cyclical component $\widehat{c}^{\mathrm{HP}} = x - \widehat{f}^{\mathrm{HP}}$ inevitably inherits elements of the stochastic trend even in the limit. Similar issues arise in the case of time series with stochastic trends coupled with deterministic drift or deterministic drift with breaks (see PJ for details). In all these cases, the HP filter fails to recover the underlying trend in x_t asymptotically. The limit theory therefore confirms much informal commentary in the literature concerning the shortcomings of the HP filter as a suitable trend determination mechanism for economic data.

We propose an easy remedy to establish consistent estimation of stochastic process and deterministic trends in the data. If the cyclical component $\widehat{c}^{\mathrm{HP}}_t$ still exhibits trending behavior after HP filtering, we continue to apply the HP filter to $\widehat{c}^{\mathrm{HP}}$ to remove the leftover trend residual. After a second fitting, the cyclical component can be written as

$$\widehat{c}^{(2)} = (I_n - S)\, \widehat{c}^{\mathrm{HP}} = (I_n - S)^2\, x,$$

[3] See Phillips (2010), Phillips and Jin (2015), De Jong and Sakarya (2016), and Cornea-Madeira (2017) for recent work on exact matrix forms and other exact representations of the HP and related filters.

where the superscript "(2)" indicates that the HP filter is fitted twice. The corresponding trend component becomes

$$\widehat{f}^{(2)} = x - \widehat{c}^{(2)} = \left(I_n - (I_n - S)^2\right) x.$$

If $\widehat{c}^{(2)}$ continues to exhibit trend behavior, the filtering process may be continued for a third or further time. After m repeated applications of the filter, the cyclical and trend component are

$$\begin{aligned} \widehat{c}^{(m)} &= (I_n - S)\,\widehat{c}^{(m-1)} = (I_n - S)^m\, x \\ \widehat{f}^{(m)} &= x - \widehat{c}^{(m)} = B_m x, \end{aligned}$$

where $B_m = I_n - (I_n - S)^m$. We call this iterated process the *boosted HP filter*, in view of its similarity to L_2-boosting in terms of numerical implementation.

Boosting is so-called because it has the capacity to enhance the flexibility of what is called in the machine learning literature a *weak base learner* as the starting point. In machine learning language, this process is known as a mechanism for achieving the 'strength of weak learnability' (Schapire, 1990). The intuition behind the asymptotic validity of the BHP filter is the observation that the HP filter, with a conventional choice of the tuning parameter λ, serves as a weak base learner.[4] In consequence, the crude HP filter is too weak by itself to fully capture the underlying trend, particularly when the trend involves a stochastic process. Initiating from the conventional HP filter, we iterate the procedure to strengthen this filter as a weak base learner. As discussed in the following section, under certain conditions on the number of iterations m the boosted version is able to restore the trend as $n \to \infty$, even if the base learner itself is too weak for consistency.

2.2 Asymptotic Theory

The criterion underlying the HP filter is agnostic about the data generation process. The key element in controlling the capacity of the filter to capture underlying trend behavior of various forms lies in the choice of the smoothing parameter λ and controls that are implemented on its asymptotic behavior in relation to sample size. The latter is particularly important and is presently almost universally neglected in empirical work. As we now discuss, suitable controls may be implemented on the boosted HP filter to ensure that underlying trend behavior is captured consistently.

It will be convenient to start with the case where the time series x_t has a stochastic trend and satifies the functional law (4). In this case, Theorem 3 of PJ (2015) shows that if $\lambda = \mu n^4$ for some fixed constant $\mu > 0$ independent of the sample size n, then

$$\frac{\widehat{f}^{\mathrm{HP}}_{\lfloor nr \rfloor}}{\sqrt{n}} \rightsquigarrow f^{\mathrm{HP}}(r) = \sum_{k=1}^{\infty} \frac{\lambda_k^2}{\mu + \lambda_k^2} \sqrt{\lambda_k} \varphi_k(r)\, \xi_k, \tag{5}$$

where $\xi_k \sim$ iid $N\left(0, \omega^2\right)$, $\varphi_k(r) = \sqrt{2}\sin\left(r/\sqrt{\lambda_k}\right)$ and $\lambda_k = \left[\left(k - \frac{1}{2}\right)\pi\right]^{-2}$. This asymptotic form of the HP filtered data is deduced by analyzing the asymptotic impact of the HP operator on the Karhunen-Loève (KL) representation[5] of the Brownian motion limit function given in (4), viz,

$$B(r) = \sum_{k=1}^{\infty} \sqrt{\lambda_k} \varphi_k(r)\, \xi_k \tag{6}$$

[4] In quarterly economic time series for example, accumulating evidence has shown that the setting $\lambda = 1600$ is often too large given the length of the time series typically encountered in empirical macroeconomics (Schlicht, 2005; Phillips and Jin, 2015; Hamilton, 2018).

[5] Readers may refer to Phillips (1998) for further details of KL representations and their relevance in the asymptotic analysis of nonstationary time series.

The representations (5) and (6) are orthonormal series in the trigonometric polynomials $\varphi_k(r)$ as well as the random coefficients ξ_k and the series converge almost surely and uniformly for $r \in [0,1]$. But whereas Brownian motion is everywhere non-differentiable, the asymptotic form of the HP filter given in (5) is differentiable to the fourth order and converges almost surely and uniformly for $r \in [0,1]$. As discussed in PJ, for typical time series of quarterly macroeconomic data, the limit form in (5) produces a smoothed version of the time series that closely matches output from a HP filter with $\lambda = 1600$ when the constant μ is set so that $\mu = 1600/n^4$ to ensure comparability of the tuning parameter with the standard setting that is used in practical work with quarterly data. Thus, (5) may be regarded as an asymptotic approximation to the trend output from HP filtering typical quarterly macroeconomic time series. The upshot is that when the HP filter is conducted under standard settings for λ, it fails to deliver a consistent estimate of an underlying stochastic trend in the data.

The limit formula on the right-hand side of (5) is particularly convenient as a starting point in understanding the effects of repeated HP fitting. The following theorem confirms that, in contrast to (5), the BHP filter $\widehat{f}_t^{(m)}$ captures the stochastic trend in the data when the number of iterations m in the boosted filter is allowed to diverge and the primary tuning parameter setting is $\lambda = \mu n^4$.

Theorem 1. *Suppose that x_t satisfies the functional limit law (4) and the HP filter is iterated m times according to the boosted HP algorithm with $\lambda = \mu n^4$ and μ fixed. If $m \to \infty$ as $n \to \infty$ then*

$$\frac{\widehat{f}_{\lfloor nr \rfloor}^{(m)}}{\sqrt{n}} \rightsquigarrow B(r). \tag{7}$$

Remarks

(i) This result shows that repeated application of the HP filter to the cyclical component residual from each pass of the filter is successful in asymptotically eliminating remnants of the stochastic trend from the estimated cyclical component of the time series. The BHP filter algorithm thereby assures consistent estimation of a stochastic process trend like Brownian motion. Importantly and distinct from the asymptotic result in PJ, consistent estimation of the stochastic process trend applies for the boosted filter even with the primary filter setting retained as $\lambda = \mu n^4$.

(ii) The heuristic explanation of (7) is as follows. Both series representations (5) and (6) converge almost surely and uniformly in r and therefore admit further linear operations associated with the boosted filter. Successive operations of the filter then proceed to remove the remaining stochastic trend components from the cycle, leading to consistent estimation of the stochastic trend. The proof of the theorem makes use of the operator $G_\lambda = \frac{1}{\lambda L^{-2}(1-L)^4+1}$, which is the asymptotic form (apart from end corrections) of the HP operator on the time series. As shown in PJ, the operator G_λ may be interpreted as a pseudo-integral operator, which facilitates the analysis of its asymptotic properties. The corresponding operator that delivers the cyclical component is $1 - G_\lambda$ and m successive operations in the boosted HP filter then lead to the operator $(1 - G_\lambda)^m$. The asymptotic result is obtained by using the approximation $(1 - G_\lambda)\varphi_k\left(\frac{t}{n}\right) \approx \frac{\mu}{\mu+\lambda_k^2}\varphi_k\left(\frac{t}{n}\right)$, which holds with a well-controlled approximation error. Repeated fitting leads to

$$(1 - G_\lambda)^m \varphi_k\left(\frac{t}{n}\right) \approx \left(\frac{\mu}{\mu + \lambda_k^2}\right)^m \varphi_k\left(\frac{t}{n}\right) \to 0, \quad \text{as } m \to \infty.$$

Then, $[1 - (1 - G_\lambda)^m]\varphi_k\left(\frac{\lfloor nr \rfloor}{n}\right) \approx \varphi_k(r)$, as $m, n \to \infty$. Pursuing this line of argument,

the proof of Theorem 1 establishes (7) rigorously by verifying that the approximation errors accumulated throughout the series summation are asymptotically negligible.

(iii) When it is viewed as a special case of linear penalized spline smoothing, the HP filter places knots on the $n-2$ observed time points $x_t,\ t=2,\ldots,n-1$ omitting the first and last observations (Paige and Trindade, 2010, Eq.(2.2)). The nonstationary nature of the knots in trending time series cases requires new technical tools in analyzing the asymptotic behavior of the boosting procedure. In this respect, the present results go beyond the scope of existing work, such as Bühlmann and Yu (2003)'s Section 3.2 which is concerned with boosting nonparametric mean models based on penalized spline smoothing with fixed or iid knots.

We next proceed to consider the effect of boosting the HP filter when the time series involves a deterministic trend. PJ have shown that the HP filter itself asymptotically preserves a polynomial trend up to the 3rd order. In consequence, the boosted HP filter also asymptotically maintains the presence of a polynomial trend up to the 3rd order. The following result further shows that the boosted HP filter consistently estimates any higher order polynomial trends that may accompany a stochastic process trend in the data, as well as the limiting stochastic trend itself, thereby capturing the full limiting trend process.

Theorem 2. *Let $x_t = \alpha_n + \beta_{n,1}t + \cdots + \beta_{n,J}t^J + x_t^0$ where x_t^0 follows the functional limit law (4) and the coefficients in the polynomial $\alpha_n/\sqrt{n} \to \alpha$, $n^{j-1/2}\beta_{n,j} \to \beta_j$ for $j = 1,\ldots,J$. Suppose the HP filter is iterated m times with $\lambda = \mu n^4$ and μ fixed. If $m, n \to \infty$, then*

$$\frac{x_{\lfloor nr \rfloor}}{\sqrt{n}}, \frac{\widehat{f}_{\lfloor nr \rfloor}^{(m)}}{\sqrt{n}} \rightsquigarrow \alpha + \beta_1 r + \cdots + \beta_J r^J + B(r). \tag{8}$$

The polynomial component of x_t in Theorem 2 is specified with sample size dependent coefficients, assuring that the standardized time series $\frac{x_{\lfloor nr \rfloor}}{\sqrt{n}}$ satisfies the functional law (8), giving a limit stochastic process trend with polynomial drift of degree J. The result extends Theorem 4 of PJ by showing that boosting the HP filter with primary tuning parameter setting $\lambda = \mu n^4$ asymptotically preserves a polynomial trend of any finite order as $m \to \infty$. The implication is that when using the conventional $\lambda = 1600$ setting of the smoothing parameter for quarterly macroeconomic time series, the boosting algorithm ensures that the filter delivers a consistent estimator of the limiting form of the trend in a time series that has a stochastic process trend with a finite order polynomial drift.

A closely related result might be expected in the case of boosting the HP filter applied to a time series that has a stochastic process limiting trend function accompanied by a deterministic drift that is piecewise continuous with a finite number of break points. Theorem 5 of PJ showed by using the Fourier series representation of the drift function that the asymptotic effect of the HP filter is to smooth a piecewise continuous limit drift function into a smooth curve in which the breaks in the deterministic trend are represented by smooth transitions over adjacent neighborhoods. In view of Theorem 2, we might expect that boosting the HP filter would enable the filter under some conditions to capture the continuous parts of a polynomial trend as $m, n \to \infty$. The following result provides asymptotic theory and conditions for the case of a time series with a stochastic trend and time polynomial drift with a single break point.

To fix ideas, suppose $g_n(t)$ is a trend break polynomial with a single break point at $\tau_0 = \lfloor nr_0 \rfloor$ with $r_0 \in (0,1)$ that takes the form

$$g_n(t) = \begin{cases} \alpha_n^0 + \beta_{n,1}^0 t + ... + \beta_{n,J}^0 t^J & t < \tau_0 = \lfloor nr_0 \rfloor \\ \alpha_n^1 + \beta_{n,1}^1 t + ... + \beta_{n,J}^1 t^J & t \geq \tau_0 = \lfloor nr_0 \rfloor \end{cases},$$

with $\frac{\alpha_n^\delta}{\sqrt{n}} \to \alpha^\delta$ and $\left\{ n^{j-\frac{1}{2}} \beta_{n,j}^\delta \to \beta_j^\delta : j = 1, .., J \right\}$ for $\delta = 0, 1$. The limiting form of this polynomial break function is

$$n^{-1/2} g_n \left(\lfloor nr \rfloor \right) \to g\left(r \right) = \left\{ \begin{array}{ll} \alpha^0 + \beta_1^0 r + ... + \beta_J^0 r^J & r < r_0 \\ \alpha^1 + \beta_1^1 r + ... + \beta_J^1 r^J & r \geq r_0 \end{array} \right. ,$$

giving a piecewise continuous polynomial function with a single break at $r = r_0 \in (0, 1)$. If the generating mechanism of the observed data x_t is $x_t = g_n\left(t \right) + x_t^0$, where x_t^0 is a stochastic trend that satisfies the functional limit law (4), then the normalized process $n^{-1/2} x_{t=\lfloor nr \rfloor}$ has the following limit

$$n^{-1/2} x_{t=\lfloor nr \rfloor} \rightsquigarrow g\left(r \right) + B\left(r \right) =: B_g(r) \tag{9}$$

as $n \to \infty$. PJ (2015) explored the asymptotic form of the HP filter applied to such a time series x_t, showing that when $\lambda = \mu n^4$ the limiting form of the normalized HP filtered time series is a smooth function $B_g^{\mathrm{HP}}(r) := g^{\mathrm{HP}}\left(r \right) + B^{\mathrm{HP}}(r)$ where $g^{\mathrm{HP}}\left(r \right)$ is a continuous function approximation to $g\left(r \right)$ that smooths over the break point of $g\left(r \right)$ at r_0 and $B^{\mathrm{HP}}(r)$ is a smooth functional approximation to the Brownian motion $B\left(r \right)$ of the same form as (5).

The following result shows that the boosted HP filter can consistently estimate the limit function $B_g(r)$ for $r \neq r_{0.}$, thereby capturing the polynomial drift function at all points except the break point, as well as the stochastic trend process.

Theorem 3. *Let $x_t = g_n\left(t \right) + x_t^0$ where x_t satisfies $n^{-1/2} x_{t=\lfloor nr \rfloor} \rightsquigarrow g\left(r \right) + B\left(r \right)$. Suppose the HP filter is iterated m times with $\lambda = \mu n^4$ and μ fixed. If $\frac{1}{m} + \frac{m}{n} \to 0$ as $n \to \infty$, then*

$$\frac{\widehat{f}_{\lfloor nr \rfloor}^{(m)}}{\sqrt{n}} \rightsquigarrow g^{\mathrm{bHP}}\left(r \right) + B\left(r \right) := \left\{ \begin{array}{ll} g\left(r \right) + B\left(r \right), & r \neq r_0 \\ \frac{1}{2}\left\{ g\left(r_0^- \right) + g\left(r_0^+ \right) \right\} + B\left(r_0 \right), & r = r_0 \end{array} \right. . \tag{10}$$

for each $r \in [0, 1]$ and $r_0 \in (0, 1)$.

Compared with Theorem 2, this result imposes the additional condition that $m/n \to 0$ as $m, n \to \infty$. The extra condition is useful in the proof in deriving the limit behavior of the boosted filter around the break point $r = r_0$. When $r \approx r_0$, the HP filter and boosted filter both smooth the time series trajectory of x_t using observations on either side of the break point. Like the Fourier series approximation of a piecewise continuous function (and the Gibbs phenomenon), the HP filter and boosted filter do not converge to the true value, $B_g(r_0)$, of the limit function at the break point $r = r_0$. PJ show that when $\lambda = \mu n^4$ the HP filter converges to a smoothed version of the limit process $B_g(r)$ for all $r \in (0, 1)$. The above result shows that for the same tuning parameter setting of λ the boosted filter provides a substantial enhancement by consistently estimating $B_g(r)$ for all $r \neq r_0$ in the limit as $m, n \to \infty$ when $m/n \to 0$. An implication of this result is that the BHP filter provides a consistent estimate of the break point r_0 in an arbitrary polynomial trend. By consistently estimating $B_g(r) = g(r) + B(r)$ for all $r \neq r_0$ the boosted filter effectively reveals the break point r_0 by virtue of the fact that the deterministic limit function $g(r)$ has a finite right limit to the value $g(r_0^+) = g(r_0)$ on the right and has a finite left limit to a value $g(r_0^-) \neq g(r_0)$. The asymptotic form of $\widehat{f}_{\lfloor nr \rfloor}^{(m)} / \sqrt{n}$ captures this same behavior, showing that at the break point $r = r_0$ the limit of the BHP filter is the simple average of the left and right limits, viz., $\frac{1}{2}\left\{ g\left(r_0^- \right) + g\left(r_0^+ \right) \right\}$, thereby mimicking the behavior of the Fourier series representation of the function $g(r)$ at the break point r_0.

Theorem 3 is proved for a polynomial of arbitrary finite order with a single break point. The proof of this theorem reveals that under the same conditions the result may be extended to any piecewise continuous polynomial function with a finite number of break points. In effect, the extended result shows that the boosted HP filter can consistently estimate a stochastic trend together

with a breaking polynomial drift that has multiple structural breaks. Consistency applies for all points in the domain with the exception of the break points themselves. But in the same manner as Theorem 3 the fact that consistency holds almost everywhere with exceptions at the break points ensures that the boosted HP filter delivers consistent estimates of the break points themselves. These results hold in the presence of discrete break points. The present asymptotic development does not provide for local break point departures with breaks that decay with the sample size. The analysis of the asymptotic properties of the BHP filter in such cases is left as a topic of future research.

2.3 Numerical Illustrations with the Boosted HP Filter

It is not uncommon for modern machine learning methods — such as boosting, random forest and artificial neural network — to have multiple tuning parameters. The BHP filter has been formulated with two tuning parameters, one primary (λ) and one secondary (m). The near-universal choice for the primary smoothing parameter is $\lambda = 1600$ for quarterly data and Theorems 1 and 2 show that with this choice, assuming that $\lambda = \mu n^4 = 1600$, the boosted filter can successfully consistently estimate and extract both stochastic and deterministic trends. This setting therefore means that there is good reason to continue using the standard value $\lambda = 1600$ with quarterly data and to focus attention on a suitable choice for the secondary parameter m. This approach matches Bühlmann and Yu (2003)'s recommendation of using a relatively large primary parameter for smoothing and designating the boosting parameter as the sole tuning parameter. This regularization method of terminating the algorithm after several iterations is called *early stopping* in statistical learning theory.

As the results of the last section show, the asymptotic effect of increasing m is similar to reducing the value of λ in the simple HP filter, as both approaches can lead to consistent estimation of stochastic trends. In particular, PJ's results imply that consistent trend estimation is restored if smaller values of λ are used so that λ/n^4 shrinks to 0 fast enough as $n \to \infty$. But implementing such a scheme would require a grid system $\left(\lambda^{(1)}, \lambda^{(2)}, \ldots, \lambda^{(m)}\right)$ to be specified and the performance of the HP filter over these choices to be monitored and evaluated by some other criterion. Such a regularization scheme itself is an iterative procedure that is conceptually no more appealing than early stopping and is difficult to implement absent suitable criteria for the selection process and supporting asymptotic theory.

Macroeconomic time series are now available internationally in great abundance and computation is therefore a relevant consideration in all such big data applications when machine learning methods are employed (Aruoba, Diebold, Kose, and Terrones, 2010; Aruoba and Diebold, 2010). A key computational advantage in the use of a BHP filter and early stopping procedure is its lower computational complexity and higher numerical stability in comparison to choosing λ values on a grid system for the simple HP filter (Raskutti, Wainwright, and Yu, 2014; Blanchard, Hoffmann, and Reiß, 2017). Early stopping computes only once the inverse matrix operator $S = S(\lambda) = (I_n + \lambda DD')^{-1}$ for a given λ. Using the same matrix $S(\lambda)$ stored in computer memory, successive iterations in the boosted HP filter involve simple matrix-scalar multiplication $(I_n - S(\lambda))\widehat{c}^{(m-1)}$, which amounts to $2n^2 - n$ linear operations. In contrast, searching for a suitable λ on a grid system involves inverting an $n \times n$ matrix to obtain $S(\lambda)$ or carrying out QR decomposition[6] for every value of λ on the grid, compared to which the computational cost of matrix-scalar multiplication is negligible.

We discuss various stopping rules for determining the number m of boosting iterations in Section 2.4 below. Before doing so, we conduct a numerical exercise to observe the effects of repeated fitting

[6] Instead of directly computing the matrix inverse it is common to use a QR decomposition of the $(2n-2) \times n$ matrix $[I_n \mid \lambda D]'$ to reduce computation cost. However, even for such a QR decomposition the leading term of the number of linear operations is $\frac{10}{3}n^3$ by Householder transformation.

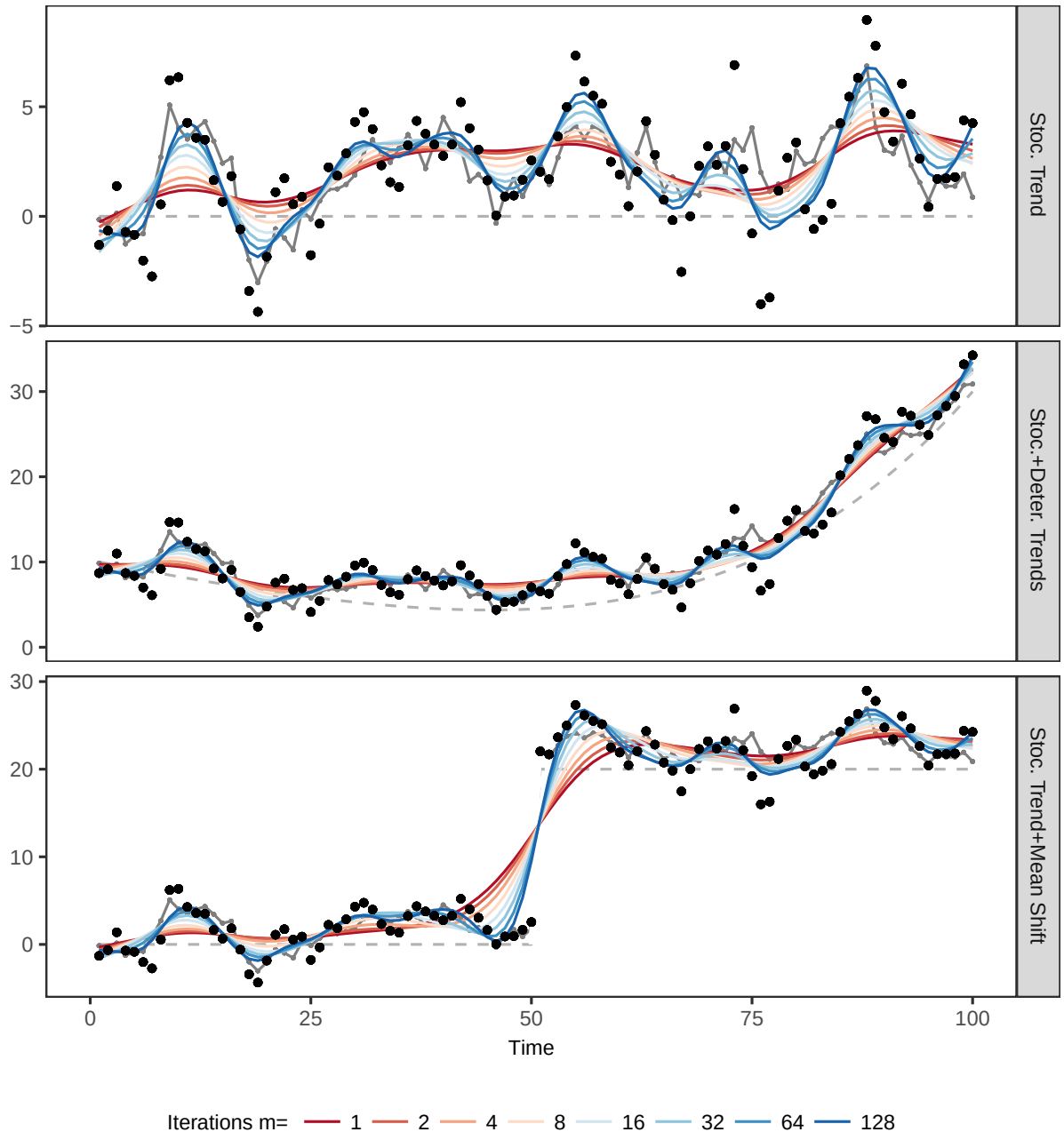


Figure 1: In each panel, observations x_t are shown in black dots, the underlying trend process $\tilde{x}_t$ by the solid gray lines, the deterministic trend $g_n(t)$ by the dashed gray lines, and the fitted trend lines obtained by the HP filter (red line) and BHP filters (colored shades of orange progressing to shades of blue).

in three prototypical cases involving a stochastic trend, and a stochastic trend with a drift and a mean break. Let $u_t^{(z)} \sim$ iid $N(0,1)$ and $u_t^{(e)} \sim$ iid $N(0,1)$ be independent innovation sequences, and $g_n(t)$ be a deterministic sequence. Define

$$\begin{aligned} z_t &= z_{t-1} + u_t^{(z)}, \\ e_t &= 0.5e_{t-1} + u_t^{(e)} + u_{t-1}^{(e)}, \\ \tilde{x}_t &= g_n(t) + z_t, \\ x_t &= \tilde{x}_t + e_t, \end{aligned} \tag{11}$$

where (z_t) is a random walk, (e_t) is an ARMA(1,1) stationary process, $(\tilde{x}_t)$ is a trend consisting of a non-random drift component $g_n(t)$ and the stochastic trend component z_t, and (x_t) is the observed time series, which allows for stationary deviations or measurement error in observations of $\tilde{x}_t$.

Given the same realized $x_t^0 = z_t + e_t$ of length $n = 100$, we generate the observations $x_t = g_n(t) + x_t^0$ shown in the three panels of Figure 1 by varying the deterministic component as follows: $g_n(t) = 0$ in the top panel; a 4th order polynomial trend $g_n(t) = 10^{-3} \cdot (n-t)^2 + 3 \cdot 10^{-7} \cdot t^4$ in the middle panel; and a mean shift $g_n(t) = 20 \cdot 1\{t \geq 0.5n + 1\}$ in the lower panel. The observations x_t are represented by the black scattered dots, the trend $\tilde{x}_t$ by the solid gray line, and the deterministic trend $g_n(t)$ by the dashed gray line. The HP filter with $\lambda = 1600$ is used to extract the trend and is shown as the red curve ($m = 1$) in the figure. The other curves are the fitted trends obtained by iterating the HP filter m times (the m values are given in the figure legend) according to the boosted filter.

Both the upper and the middle panels reveal that the repeated fitting tracks the random wandering behavior of the random walk uniformly better than the HP filter.[7] The curves are insensitive to the number of iterations once m becomes large. This phenomenon is known as boosting's 'resistance to overfitting'.[8] It is corroborated analytically in the proof of Theorem 1 where it is shown that as m becomes large for given n the boosted filter stabilizes and approximates a finite number of terms in the orthonormal series representation of the limiting trend process. This finite term orthonormal representation is a smooth approximation to the true limit process, explaining the smooth form of the boosted HP filter.

In the lower panel, it is evident from the plots that boosting the filter goes a long way towards enhancing performance in the region of the structural break by eliminating a substantial amount of the transition smoothing in the HP filter around the break point. For large $m \approx n$, the boosted filter trajectory is strongly suggestive of a structural break around observation $t = 50$ with a mid-point estimate of the value at the break point, corroborating the implications of Theorem 3.

2.4 Stopping Criterion

The residual component after trend extraction by smoothing methods such as the HP filter has long been a building block for applied macroeconomists in studying business cycles and the interactions between macroeconomic aggregates and indicators. By definition, the cyclical component is a time series that exhibits no long run trending behavior, so that its spectrum has no unit root or deterministic trend asymptote at the zero frequency. In practice this criterion can be implemented by the elimination of all low frequency elements, an approach that band-pass filter methods use directly in filtering the data (Baxter and King, 1999; Christiano and Fitzgerald, 2003; Corbae and Ouliaris, 2006).

A natural and somewhat analogous approach in the present context is to refilter the data until

[7] The BHP filter gives a superior fit to the true trend (represented by the solid gray line in Figure 1) with a smaller L_2 distance to the gray line than the HP filter ($m = 1$) for all values of $2 \leq m \leq 128$.

[8] Further evidence of resistance to overfitting is given later by Figure 3 in the simulation study.

there is no evidence of a non-stationary zero frequency asymptote. This can be conveniently achieved by monitoring the outcome of unit root tests on the residual series $\widehat{c}_t^{(m)}$. Standard procedures for unit root testing such as the augmented Dickey-Fuller (ADF) or Phillips-Perron (Phillips and Perron, 1988) tests can be used and the boosting iterations can be continued until the test statistic is smaller than a specified p-value, such as 0.05 or 0.01. Such test-based stopping criteria are easy to implement and are well-tailored to existing applied macroeconomic practice, echoing Kozbur (2017)'s test-based stopping criterion for forward selection, and Diebold and Kilian (2000)'s test-based forecasting approach. Relatedly, Hodrick and Prescott (1997) used unit root tests to assist in determining an appropriate setting for the primary smoothing parameter λ. In our simulations and empirical examples, we will use the ADF test conducted with significance level 0.05 to illustrate implementation of this approach. The boosted HP filter that results from this ADF test-based selection will be denoted BHP-ADF.

Information criteria offer an alternative approach to a stopping criterion. These criteria are routinely employed in statistics to achieve bias-variance trade-offs and to prevent overfitting in modeling and forecasting. We therefore consider the following Bayesian-type information criterion (BIC) for the selection of the stopping time for m

$$IC\left(m\right)=\frac{\widehat{c}^{(m)\prime}\widehat{c}^{(m)}}{\widehat{c}^{\mathrm{HP}\prime}\widehat{c}^{\mathrm{HP}}}+\log\left(n\right)\frac{\mathrm{tr}\left(B_m\right)}{\mathrm{tr}\left(I_n-S\right)}. \tag{12}$$

Similar to BIC, this criterion penalizes fit by adding $\log(n)$ times a term that quantifies the relative weight of the m additional iterations that are involved in the boosted filter. The first term of (12) measures the residual sum of squares fit of the boosted HP filter, $\widehat{c}^{(m)\prime}\widehat{c}^{(m)}$, relative to the HP filter itself, $\widehat{c}^{\mathrm{HP}\prime}\widehat{c}^{\mathrm{HP}}$. The penalty term involves the usual $\log(n)$ scale factor multiplied by a ratio that measures the effective degrees of freedom of the boosted filter after m iterations to the effective degrees of freedom of the HP filter. To interpret this ratio, it is useful to think of the linear operator (I_n-S) that produces the residual cyclical component $\widehat{c}^{\mathrm{HP}}=(I_n-S)\,x$ of the HP filter as analogous to a linear regression projector or hat matrix, so that $\mathrm{tr}\left(I_n-S\right)$ is analogous to the degrees of freedom in the sample after projection. The operator $B_m=I_n-(I-S(\lambda))^m$ is the m-fold operator corresponding to the boosted filter and the quantity $\mathrm{tr}(B_m)$ may therefore be interpreted in a similar way as the effective degrees of freedom after successive fitting by the boosted HP. This interpretation corresponds to usage in the machine learning literature (Tutz and Binder, 2006). It is convenient from now on to refer to the criterion $IC\left(m\right)$ simply as BIC and to the boosted HP filter that results from this selection rule as BHP-BIC. .

It is shown in the Appendix that $\mathrm{tr}(B_m)$ can be asymptotically approximated by the following simple analytic expression as $n\to\infty$

$$\mathrm{tr}(B_m)=\mathrm{tr}(I_n-(I_n-S(\lambda))^m)=n-\sum_{k=1}^{n-2}\frac{(\lambda\delta_k^2)^m}{(1+\lambda\delta_k^2)^m}\left\{1+o\left(1\right)\right\},\quad \delta_k^2=4\left(1-\cos\frac{k\pi}{n-1}\right)^2. \tag{13}$$

Figure 2 graphs $\mathrm{tr}(B_m)$ against this approximation as a function of m, showing how the penalty term coefficient $\mathrm{tr}(B_m)$ increases monotonically and nonlinearly with m for any given value of the sample size n. Differentiating (13) with respect to m gives

$$\frac{\partial\mathrm{tr}\left(B_m\right)}{\partial m}=\log\left(1+\frac{1}{\lambda\delta_k^2}\right)\sum_{k=1}^{n-2}\left(\frac{1}{1+1/\left(\lambda\delta_k^2\right)}\right)^m\left\{1+o\left(1\right)\right\}>0,$$

so that $\mathrm{tr}(B_m)$ is increasing in m with decreasing derivative as m increases, as is evident in Figure 2. Moreover, as is clear from formula (13) and the graph, the penalty coefficient $\mathrm{tr}(B_m)\to 2$ as $\lambda\delta_k^2\to\infty$, which happens for all $k=\lfloor nr\rfloor$ with $r\in(0,1]$ and m fixed when $\lambda=\mu n^4$. Thus, the

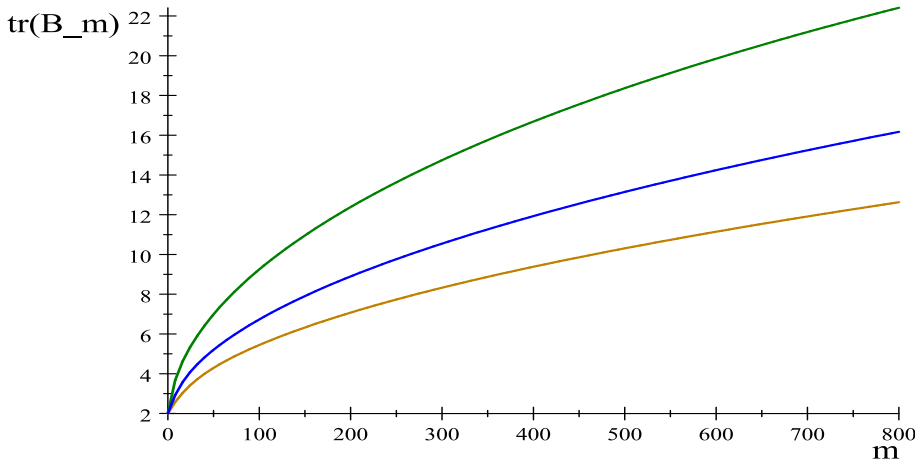


Figure 2: Plots of $\text{tr}(B_m) = n - \sum_{k=1}^{n-2} \frac{(\lambda\delta_k)^m}{(1+\lambda\delta_k)^m} \{1 + o(1)\}$ for $n = 50, 75, 100$ (green, blue, sienna).

impact of the penalty on the choice of m is attenuated as $n \to \infty$.

With the implementation of one of these stopping rules, the boosted HP fitting algorithm is automated and data-determined, making it ready for practical use like other non-parametric procedures with data-determined bandwidth selectors. The following sections assess the performance of these stopping rules in simulations and provide two real data applications.

3 Simulations

We conduct simulation exercises with six data generating processes to observe the finite sample performance of the boosted HP filter in practice when the trend process involves both stochastic and deterministic elements. Similar to the models used in Section 2.3, in DGPs 1 and 2 below we add a stationary component to the deterministic trends. With the addition of this component to the data iterating an excessive number of times in the boosting process in finite samples can potentially lead to overfitting. The experimental design therefore reveals the bias-variance tradeoff that occurs in such cases and the effectiveness of the two stopping criteria in preventing saturation fitting in practical applications of the boosted filter.

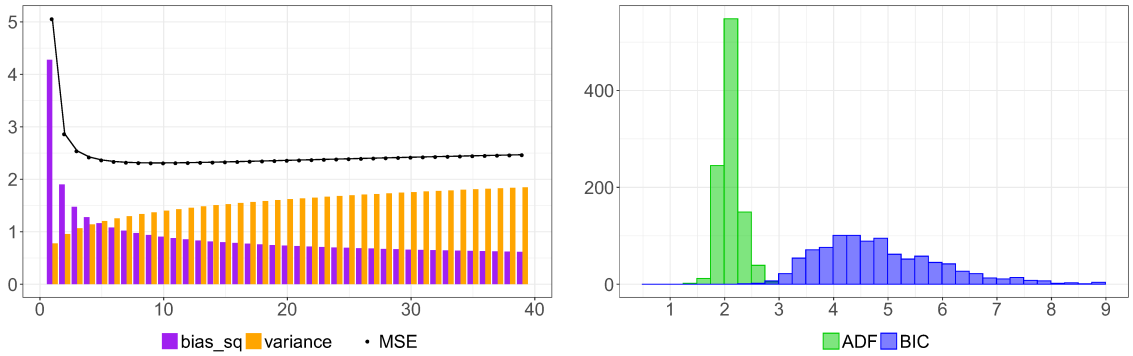
DGPs 3-6 focus on fitting a trend under alternative plausible generating mechanisms that include a pure random walk, a structural break, a sinusoidal trend, and various combinations of these trends. The experiments also provide performance comparisons of the boosted filter approach to trend extraction with the autoregressive model estimation approach advocated in Hamilton (2018).

3.1 The Bias-Variance Tradeoff in Boosting

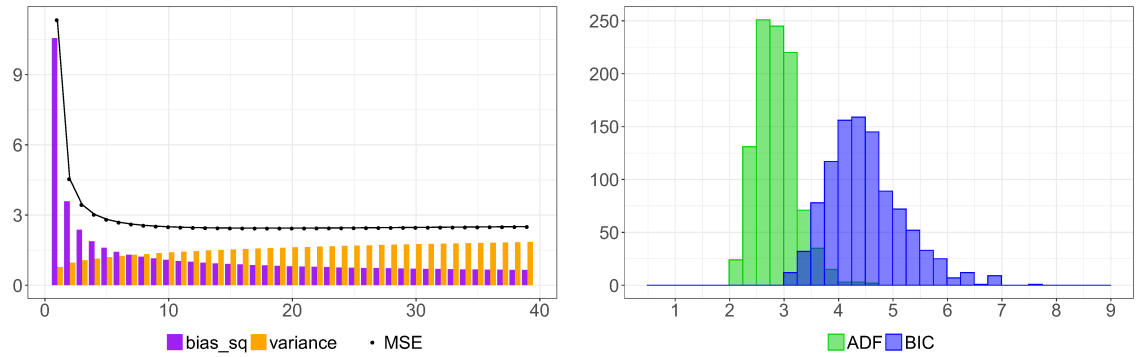
According to Theorem 2, the boosted HP filter can asymptotically remove any finite-order polynomial drift, whereas the HP filter can only handle a polynomial drift up to the 3rd order. Higher order time polynomials are known to be useful in modeling the nonlinear growth of both macroeconomic and microeconomic time series and as sieve approximations to more general nonlinear trend functions (Baek, Cho, and Phillips (2015); Cho and Phillips (2018)). We are therefore interested in the capability of the boosting mechanism to enable the HP filter to capture these general deterministic trend elements in addition to stochastic trends.

The following two experimental designs involve finite degree polynomial drift functions, $g_n(t)$, to accompany the stochastic trend generating mechanism as in (11). The specification illustrates

the potential gains that can be obtained in trend determination by boosting the HP filter even in the presence of simple deterministic drifts.



(a) DGP 1: stochastic trend with 3rd-order polynomial drift



(b) DGP 2: stochastic trend with 4th-order polynomial drift

Figure 3: Bias-variance tradeoffs in trend estimation (left panel) and distributions of stopping times calculated by the ADF (green) and BIC (blue) criteria (right panel) for the HP ($m = 1$) and boosted HP filters ($m > 1$). The number of iterations (m) is shown on the horizontal axis in each figure.

DGP 1 Set the sample size $n = 100$ (25 years in quarterly data), and the deterministic trend $g_n(t) = 0.0005t^3$. Step 1: Generate a stochastic trend plus drift process $\tilde{x}_t = g_n(t) + z_t$ as defined in (11). Step 2: Given the realized trend $\tilde{x}_t$ in Step 1, simulate the stationary random component e_t, also defined in (11), to produce the measured observation $x_t = \tilde{x}_t + e_t$. Step 3: Repeat Step 2 50 times (calling this the *inner loop*) in order to compute the bias and variance of the filters given the trend $\tilde{x}_t$. Step 4: Repeat Steps 1-3 for 1000 replications (we call this the *outer loop*) to average the bias and variance over the various realizations of the trend process.

DGP 2 This experimental design is identical to DGP 1 except for the fact that the deterministic trend component is generated by a 4th order polynomial $g_n(t) = 5 \cdot 10^{-6} \cdot t^4$ rather than a 3rd order polynomial.

Given a realized $\tilde{x}_t$, in the inner loop of 50 replications we compute for fixed t and m the empirical versions of the bias $\mathbb{B}^{(m)} = E[\widehat{f}_t^{(m)}] - \tilde{x}_t$ and the variance $\text{var}[\widehat{f}_t^{(m)}]$. Then over the realized trend trajectory $\tilde{x} = (\tilde{x}_t)_{t=1}^n$ we calculate the squared-bias $\mathbb{Q}^{(m)} = \frac{1}{n}\sum_{t=1}^n \left[\mathbb{B}^{(m)}\right]^2$ and the variance $\mathbb{V}^{(m)} = \frac{1}{n}\sum_{t=1}^n \text{var}[\widehat{f}_t^{(m)}]$. Finally, in the outer loop for each m we average over the 1000

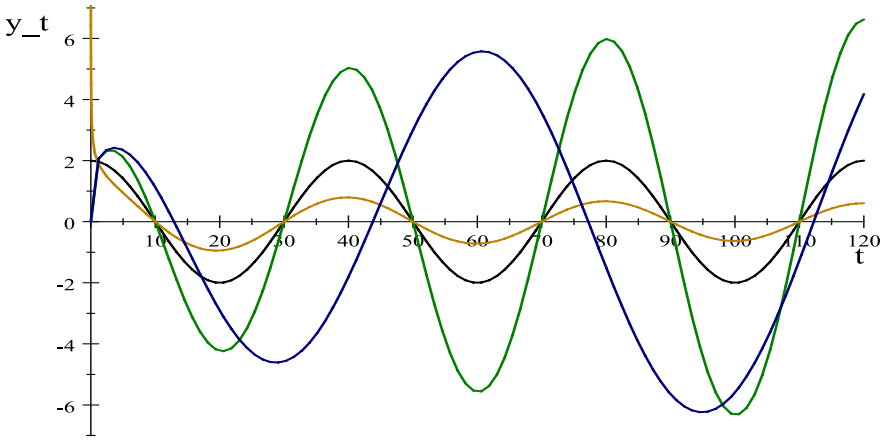


Figure 4: Plots of various sinusoidal trend functions y_t: trigonometric $2\cos(0.05\pi t)$ (black); trending trigonometric $2t^{0.25}\cos(0.05\pi t)$ (green); evaporating trigonometric $2t^{-0.25}\cos(0.05\pi t)$ (sienna); and evolving duration trigonometric $2t^{0.25}cos(0.05\pi t^{0.90})$ (blue).

replications of $\mathbb{Q}^{(m)}$ and $\mathbb{V}^{(m)}$. The squared bias and the variance are displayed in the left subgraph of Figure 3 for each $m = 1, \ldots, 40$. The black dotted line above the bars sums the underlying two bars and gives the mean squared error (MSE).

In both DGPs, similar patterns of bias-variance tradeoff are evident. Initiating the iteration process from the HP filter ($m = 1$), we observe a sizable drop in the squared bias and MSE in the first few iterations of boosting. The squared bias continues to decrease as the iterations proceed, whereas variance slowly increases. After it reaches a minimum, the MSE remains insensitive as a rather flat curve as m continues to grow, which reflects the boosting saturation that occurs in finite samples.

To evaluate the effect of the data-driven stopping criteria, we save the number of iterations in each instance and take the sample average in the inner loop. The outer loops provide 1000 such average stopping times and histograms of these average stopping times are shown in the right subgraph of Figure 3. In DGP 1 the 3rd-order polynomial trend can be asymptotically removed by fitting the HP filter only once. Setting the test size to be 0.05, we find that only 25.9% of the average ADF stopping times are smaller than two, indicating that some remnants of the stochastic trend appear in the residual cyclical component with nontrivial probability. The BIC criterion requires at least two iterations in all replications and often three or four fittings. The effect of these fittings is evident in the large reductions in the squared bias as observed in the left panel.

The stopping time data is more intriguing in DGP 2 where we replace DGP 1's cubic trend by a 4th-order time polynomial. According to the limit theory, without the use of boosting the HP filter cannot asymptotically remove such a higher order polynomial trend. This asymptotic theory is clearly supported in the finite sample computations. The bottom-right subgraph of Figure 3 shows that the average ADF stopping criterion is at least two and the BIC criterion requires at least three fittings and often as many as four or five.

3.2 Goodness of Trend Determination

In the previous subsection, DGPs 1 and 2 were designed as mechanisms to produce a polynomial trend plus a stochastic trend. Whatever the precise nature of the trend, conditional on its realized form computations of the bias and variance of the boosted HP filter reveal the tradeoff that occurs in these measures of fit as the number m of iterations in the boosted filter rises. As the results with DGPs 1 and 2 show, bias typically falls quickly as m begins to rise, demonstrating immediate gains

from boosting. But with increasing m bias reductions diminish and variance rises to a point where mean squared error stabilizes. Thus, in finite samples there are limits to what can be accomplished by boosting just as in any nonparametric procedure.

Many empirical studies model time series data in terms of integrated or near-integrated processes augmented with various complementary mechanisms such as polynomial drifts, similar drifts with breaks, sinusoidal trends, or trends induced by time varying coefficients, all of which are intended to improve harmony with the observed data but with no certainty concerning the true specification of its generating mechanism. This section considers the performance of the BHP filter in such cases and compares the performance of the BHP filter with Hamilton (2018)'s alternative recommendation of the use of autoregressive (AR) modeling with a small number of lags, typically an AR(4) which is expected to be well suited to quarterly data applications.

The following four models are used to illustrate the performance characteristics of these approaches. The pure random walk case is used as a baseline in DGP 3 and DGPs 4-6 couple this integrated process with various other complementary trend specifications that progressively enhance the complexity of the generating mechanism. The notation follows the framework of (11).

DGP 3 The observed time series is $x_t^{(3)} = z_t$, a random walk with independent Gaussian increments.

DGP 4 Real economic activity may involve long duration cycles that are time-dependent and evolve in a non-replicative manner, for example with varying magnitudes or cycle lengths. We use a deterministic sinusoidal trend of the form $g_n(t) = 5t^{1/5}\cos(0.05\pi t^{0.9})$ to embody this type of complexity. Figure 4 graphs the form of various expanding and decaying sinusoidal trends of this type. The observed time series is expressed in the form $x_t^{(4)} = g_n(t) + x_t^{(3)}$.

DGP 5 This model serves as a simple prototype of GDP takeoff that can be used to represent a successful emerging economy growth trajectory. The model has a structural break in the middle of the sample and takes the form

$$x_t^{(5)} = u_t^{(z)} \cdot \mathbf{1}\{t < 0.5n\} + \big(t - 0.5n + \sum_{s=0.5n}^{t} u_s^{(z)}\big) \cdot \mathbf{1}\{t \geq 0.5n\}.$$

The first half of the sample is a stationary sequence and the second half is an integrated process with a linear upward drift.

DGP 6 This model is formed from the composition of the deterministic sinusoidal trend $g_n(t) = 5t^{1/5}\cos(0.05\pi t^{0.9})$ of DGP 4 with the structural break model in DGP 5 leading to the time series $x_t^{(6)} = g_n(t) + x_t^{(5)}$.

The goal in the simulation exercise is to determine the trend from data generated by these different mechanisms using the HP filter, the BHP filter, and the AR(4) regression technique of Hamilton (2018). In each replication the observed time series is filtered or regressed to obtain the corresponding fitted trend estimate $\widehat{f}_t$. Deviation from the underlying trend $\tilde{x}_t$ is measured in terms of the MSE calculated as $\mathbb{M}_n = \frac{1}{n-8}\sum_{t=5}^{n-4}(\widehat{f}_t - \tilde{x}_t)^2$. The end points are trimmed in $\mathbb{M}_n$ to accommodate start-up in the AR(4) process and end points in the HP filter. Calculating the MSE using $\mathbb{M}_n$ without trimming did not materially affect the results reported below. The trend processes $\tilde{x}_t$ are produced from the generating processes prescribed above so that $\tilde{x}_t^{(3)} = x_t^{(3)}$, $\tilde{x}_t^{(4)} = g_n(t) + \tilde{x}^{(3)}$, $\tilde{x}_t^{(5)} = (t - 0.5n + \sum_{s=0.5n}^{t} u_s^{(z)}) \cdot \mathbf{1}\{t \geq 0.5n\}$, and $\tilde{x}_t^{(6)} = g_n(t) + \tilde{x}^{(5)}$ for each corresponding DGP.

Table 1: MSE and Number of Iterations

DGP	HP	ADF	BIC	AR(4)
		MSE		
3	1.5982	1.5033	0.8540	0.9295
4	2.6204	1.4697	0.9943	1.1536
5	1.0719	0.9001	0.5787	1.0091
6	1.8795	0.8913	0.6329	1.2881
	Average number of iterations			
3		1.2342	9.4858	
4		2.1012	5.7384	
5		1.5456	5.3348	
6		2.3284	4.9120	

Note: Use of the fitted AR(4) autoregression follows Hamilton (2018)'s recommended approach, namely $\widehat{f}_t = \widehat{\beta}_0 + \sum_{k=1}^{4} \widehat{\beta}_k x_{t-k}$, with coefficients $(\widehat{\beta}_k)_{k=0}^{4}$ obtained from an AR(4) regression with fitted intercept.

Table 1 reports the empirical average of $\mathbb{M}_n$ and the observed number of iterations in the BHP filter over 5000 replications. In DGP 3, where the time series is generated from a random walk, the fitted AR(4) is particularly well suited since the regression model includes the true generating mechanism. Unlike the AR(4) regression which is based only on past information in forecasting the trend, the two-sided nature of the HP filter uses all sample information, including future observations to determine the current period trend value. There are notable differences in the results between the ADF selected and BIC selected stopping times for the iteration. These differences reveal the importance of iterating the filter. The BIC selector leads to a substantially lower MSE in trend determination from the boosted filter. The ADF selector tends to stop the iteration too early to achieve optimal improvement with an average number of iterations of 1.23, which is close to the HP filter itself (with $m = 1$) and substantially lower than the average number of 9.49 iterations for the BIC selector. With the BIC selector the BHP filter provides a substantial reduction in MSE over the HP filter. The BHP-BIC filter also produces a smaller MSE to the underlying trend than the AR(4) regression, an interesting result given that the AR(4) regression model encompasses the simple random walk model DGP 3 and the BHP has none of these explicit features.

The HP filter methods are all nonparametric in nature and, as the asymptotic theory suggests, when the tuning parameters are chosen appropriately these methods can adapt to complex trend processes and generating mechanisms. The simulation evidence supports this theory. In particular, once a slowly moving smooth deterministic trend is added to the random walk in DGP 4, the differences in performance are magnified and the MSE of the AR(4) regression deteriorates more than the BHP-BIC filter. Interestingly, the presence of a deterministic trend triggers more iterations in the BHP-ADF filter and it reduces the MSE to 1.47 from the value 1.50 in DGP 3.

Since the first half of the DGP 5 sample is a white noise for which the constant level trend function is easy to predict in a nonparametric method, the filter methods each obtain a smaller MSE than their counterparts in DGP 3. However, as a global parametric method, the AR(4) regression is inevitably misspecified when this structural break from an I(0) to an I(1) process is present in the observed series. In this case, the MSEs of the BHP-ADF and BHP-BIC filters are both substantially smaller than that of the AR(4). DGP 6 raises the level of trend complexity further by including an evolving sinusoidal trend. For this DGP, the boosted filter again provides much better trend determination. In fact, BHP-BIC has MSE less than half that of the AR(4). Comparison of the results for DGP 4 and DGP 6 shows that the boosted HP filter provides a very effective tool that adapts well to increasing complexity in the underlying trend mechanism. The

HP filter, on the other hand, has MSE that is almost three times the size of that of the BHP-BIC filter.

Figure 5 displays typical time series paths for DGPs 3-6, the corresponding estimated trends and the residual cyclical/stationary components obtained by the HP filter, the BHP-BIC filter, and the fitted AR(4).[9] In each case, after applying the BIC criterion for stopping, the estimated cyclical component is in general significantly smaller than that of the HP filter. In particular, for DGP 5 and 6 the HP filter incurs large estimation errors in the middle of the series around the structural break point, whereas the impact of the structural shift is ameliorated by the BHP-BIC filter in terms both of magnitude and duration. On the other hand, the AR(4) does not fit well in the stationary episode of DGP 5 and 6, most likely due to its nonsmooth nature and tendency to track the observations rather than the trend. Moreover, when the time series has no stationary episode as in DGP 4, the AR(4) fits also encounter difficulty in adapting to turning points in the time series. The latter is a feature that will be echoed later in one of the empirical examples in Section 4.3 where the filter's behavior during economic crises is studied. Once the more complex deterministic mechanism is removed, the BHP filter and the AR(4) both fit well, as is evident in the results for DGP 3.

4 Empirical Examples

We illustrate the effects of using the boosted HP filter in three real data examples. The first example revisits empirical support for Okun's law across 20 OECD countries. The second explores business cycle behavior in a panel of 78 heterogeneous time series covering emerging and developed markets with various degree of persistence and volatility. The third studies the behavior of the filters in trend determination using US industrial production data over the past century. In this last application we use the HP and BHP filters as well as the AR(4) parametric approach.

4.1 Okun's Law

Okun's law (Okun, 1962) posits an empirical association between output and the unemployment rate that has received wide attention among practicing economists and policy makers as well as academic economists and authors of undergraduate economics texts. For the United States, Okun's law is stated as relating a 1% increase in GDP (relative to potential GDP) to a 0.5% reduction in the unemployment rate (relative to the natural rate of unemployment). Following the original formulation by Okun, Ball, Leigh, and Loungani (2017) specify the empirical model in terms of the following empirical regression equation

$$U_t - U_t^* = \beta \left(Y_t - Y_t^* \right) + \varepsilon_t, \tag{14}$$

where U_t is the unemployment rate, Y_t is the logarithm of GDP, and U_t^* and Y_t^* are the natural rate of unemployment and the potential GDP. The sign and the magnitude of β signify the direction and strength of the relationship. In view of its potential policy implications, Okun's law has been extensively tested over time and cross countries. Most recently, Ball, Leigh, and Loungani (2017) testify to its robustness in 20 advanced economies. These authors, as many others, estimate the long-run levels of U_t^* and Y_t^* by means of the HP filter under the primary parameter setting $\lambda = 100$ for annual data. Equation (14) is therefore a simple regression between two cyclical components produced by the HP filter.

The primary motivation of using trend extraction techniques prior to the regression (14) is to focus on cyclical variates. A secondary motivation is to eliminate the possibility of spurious

[9] The BHP-ADF filter is omitted to avoid overcrowding the graphics. The BHP-ADF curves generally lie between the conventional HP and BHP-BIC versions of the filter.

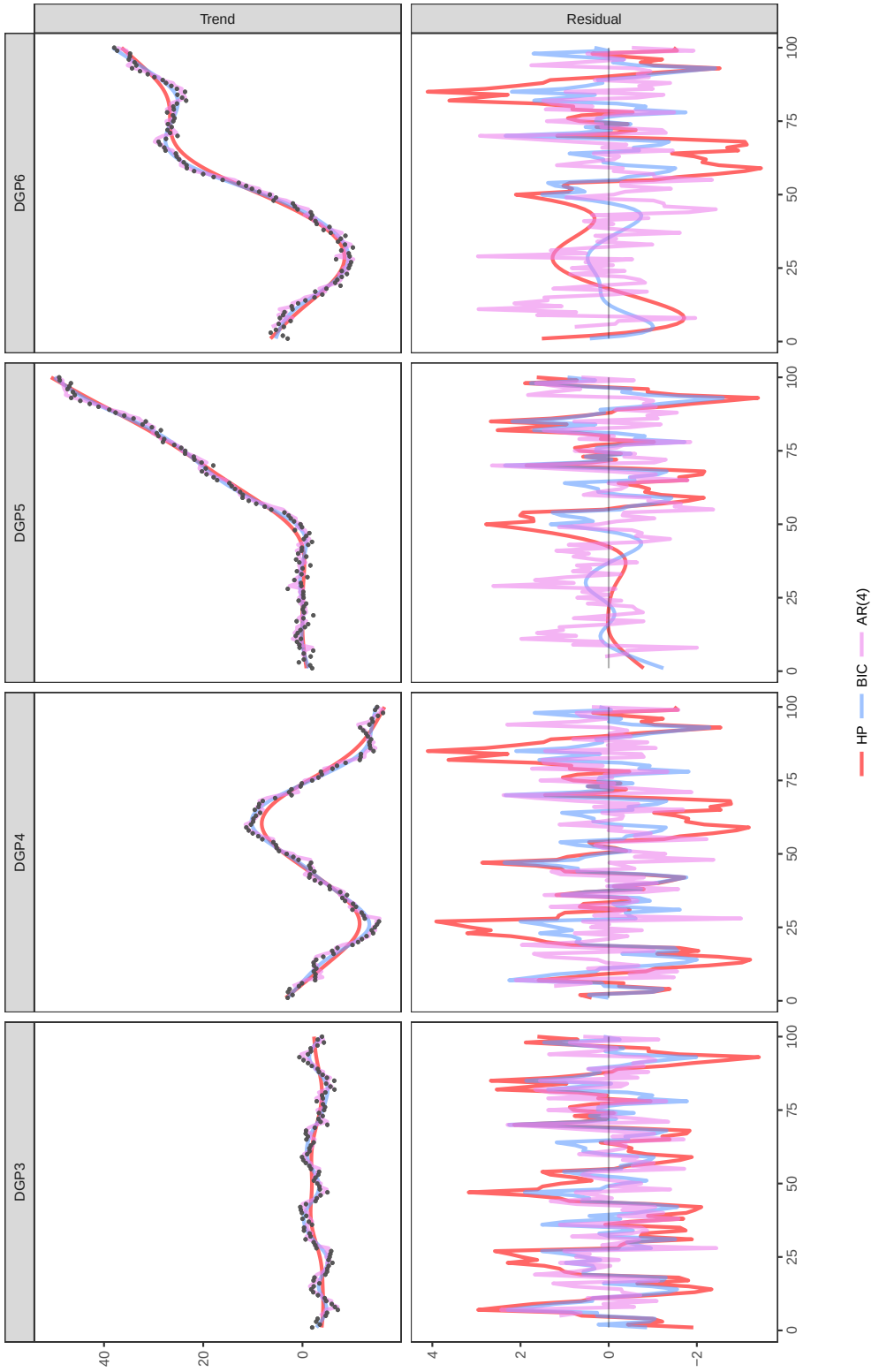


Figure 5: A typical path of DGP 3, 4, 5 or 6 (black dots), estimated trend, and cyclical component (HP: red, BIC: blue, AR(4): violet).



Figure 6: Fitted OLS coefficients and R^2 statistics for equation (14) with data obtained by simple HP and boosted HP filtering using ADF and BIC tuning parameter selection. Annual data over 1980 to 2016.

regression in the variables, which would distort inference (Granger and Newbold, 1974; Phillips, 1986) unless there is strong justification for residual stationarity and a cointegrating relationship between the variables. Use of the ADF test in the implementation of the boosted filter assists in addressing both these issues and rationalizing the regression.

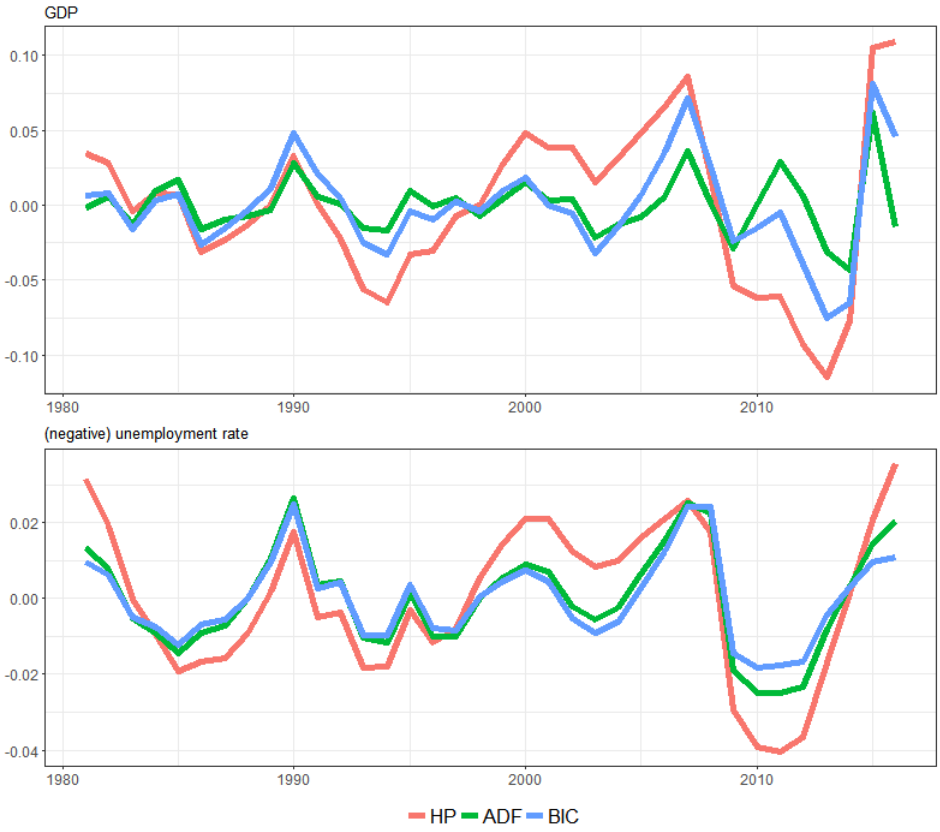


Figure 7: Cyclical components of GDP and the unemployment rate in Ireland. The negative unemployment rate is shown in the second panel. Annual data over 1980 to 2016.

We collect annual GDP data from the OECD (`OECD.stat`) and annual unemployment rates from the World Bank. We follow Ball, Leigh, and Loungani (2017) in studying the same 20 economies over the period 1980 to 2016. The dataset is mostly balanced, except for a few countries with 1 or 2 missing values at the beginning of the time period. We maintain the primary parameter setting of $\lambda = 100$ for the simple HP filter, and we apply the boosted HP filter based on the same tuning parameter λ. For each country, Figure 6 reports the OLS coefficient estimate of β in the upper panel, and the regression R^2 in the lower panel. The regressions are conducted with cyclical components extracted by the simple HP filter (shown by red bars), the boosted HP filter with iterations stopped by (i) ADF test outcomes at the 5% level (shown by green bars), and (ii) use of the information criterion (12) (shown by blue bars).

For most countries, the fitted coefficients and R^2 are similar across the filtering methods. For example, in United States the coefficient is approximately -0.5, and R^2 is around 0.8. These figures accord with established results for the USA and the recent findings of Ball, Leigh, and Loungani (2017). In particular, the results from using the boosted filter tend to confirm the conclusion of the latter authors that 'Okun's law is a strong relationship in most countries'.

One country where there is a surprisingly large contrast among the methods is Ireland, where the equation R^2 is 0.71 after simple HP filtering but only 0.10 after BHP-ADF filtering and 0.43

after BHP-BIC filtering. To explore these differences, we display the relevant data for Ireland in Figure 7. The upper panel graphs the estimated cyclical components of GDP obtained by HP, BHP-ADF and BHP-BIC. The red line produced by the HP filter shows a long upward trend from the mid 1990s to 2007, followed by a sustained slump until 2013. These trends are evident in the data from inspection and it is apparent that the HP filter fails to remove them in estimating the cycle. In fitting the boosted filter using the ADF procedure to select the boosting tuning parameter 19 iterations of the filter were needed, the largest number of iterations among all the 40 series in this experiment. The associated cyclical component is represented by the green line. This GDP cyclical component fluctuates around the mean in a smaller range, shows no evidence of a residual trend, and it appears much more stable than the cycle determined by the HP filter. The shape of the cycle obtained by using BIC selection is very similar after $m = 5$ iterations.

The lower panel displays the three fitted curves of the (negative) unemployment rate for Ireland. The negative rate is used in the figure to better visualize the association with the GDP fitted cycles shown in the upper panel. For the unemployment rate series, the boosted filter is stopped by ADF after 2 iterations and by BIC after 5 iterations. In both cases, the use of repeated filtering clearly mitigates residual trend behavior in the unemployment rate in comparison with the HP filter. The mitigation is more evident in the case of BHP-BIC filtering where the fitted cycle in the unemployment rate appears even more regular than after BHP-ADF filtering, especially during the aftermath of the 2007-2008 financial crisis. These adjustments in the fitted cycle from BHP filtering appear not to reduce the evident association with the GDP cycle and the fitted regression coefficients after BHP filtering indeed have similar values, as shown in the upper panel of Figure 6.

In sum, this application continues to support the robustness of Okun's law across developed countries, thereby reinforcing the conclusion of Ball, Leigh, and Loungani (2017). But the results also expose the insufficiency of the standard HP filter to remove stochastic trend components in the case of Ireland. Repeated fitting in this case helps to isolate the cyclical component in each time series.

4.2 International Business Cycles: Emergent and Developed Economies

The HP filter was originally motivated in Hodrick and Prescott (1997) through its usefulness in the empirical study of business cycles in the USA. In an influential paper with a similar thematic concerning international evidence of business cycles, Aguiar and Gopinath (2007) find that emerging markets (represented by 13 economies) in general are more persistent in the cyclical components of the three series they consider (GDP, consumption and investment) than those of the developed markets (represented by another group of 13 countries). In summarizing their study they declared that "[for emerging markets] the cycle is the trend."

We revisit this conclusion using the methods of the present paper to analyze the same data that the authors provide online.[10] Within each country, the three time series have the same length but across countries the length of the time series varies considerably. For example, the median length is 52 quarters for the emerging economies, with Argentina the shortest (1993Q1–2002Q4, 40 quarters), whereas the median is 94 quarters for the developed countries, with Australia, Finland, Netherlands and Norway the longest (1979Q3—2003Q2, 95 quarters). The authors established their empirical results after HP-filtering all 78 time series with the standard setting $\lambda = 1600$. As discussed earlier in the paper, the analysis in Phillips and Jin (2015) shows that the implied penalty from using this standard setting is heavier for shorter time series, making stochastic trend identification difficult in international comparisons with series of differing lengths. As our asymptotic theory shows, the boosted HP filter provides a mechanism for adapting the standard setting to account for shorter and longer sample sizes. We employ the iterated procedure to the logarithm of GDP, consumption

[10] Downloadable at `https://scholar.harvard.edu/gopinath/pages/data-and-codes`.

Table 2: Number of iterations and some moments (median in each group)

	HP		BHP-ADF		BHP-BIC	
	emerging	developed	emerging	developed	emerging	developed
			Number of iterations			
GDP (Y)	1	1	2	3	10	7
Consumption (C)	1	1	2	2	10	7
Investment (I)	1	1	4	2	12	6
			Variance and correlation coefficient			
$\sigma(Y_t)$	0.0251	0.0134	0.0228	0.0094	0.0173	0.0076
$\sigma(C_t)$	0.0339	0.0127	0.0325	0.0093	0.0233	0.0070
$\sigma(I_t)$	0.0960	0.0413	0.0786	0.0332	0.0607	0.0268
$\rho(C_t, Y_t)$	0.7592	0.7234	0.6284	0.4772	0.6610	0.4370
$\rho(I_t, Y_t)$	0.8327	0.7024	0.7527	0.5435	0.7177	0.5135
$\rho(Y_t, Y_{t-1})$	0.7608	0.7528	0.6201	0.5624	0.5444	0.4475

and investment to study whether the cyclical patterns noted by Aguiar and Gopinath (2007) in the two groups of countries remain distinguishable.[11]

For each of the 78 time series, we apply the HP filter and automated BHP filters, all with the same $\lambda = 1600$ setting, to extract trend and save the cyclical component. In general, the emerging economies need more iterations than the developed countries to isolate trend, manifesting the differences in persistence. Table 2 displays within each group of 13 countries the medians of the number of iterations, standard deviations, and correlation coefficients. The standard deviations typically become smaller as boosting progresses, while the relative magnitude between the emerging and developed markets remains stable. Similar relative sizes are observed in the correlation coefficients. The repeated filtering changes absolute values, but the relative magnitudes of the volatility and persistence are largely maintained in the two groups of countries.

Figure 8 shows the cyclical components of each time series, with the developed nations in green and the emerging nations in black. Despite the small number of iterations involved, the BHP-ADF filter provides noticeably greater smoothing of the time series. With a only few more iterations taken by the BHP-BIC filter, the cyclical components appear more stable around the mean. The contrast in the volatility of the two groups of countries is strongly manifest in the graph. Overall, this application of the boosted filter therefore confirms that Aguiar and Gopinath (2007)'s findings are robust when machine learning methods are used to assist in compensating for the differing lengths of the time series across countries.

4.3 US Industrial Production Index

In this final application of our methods, we analyze a single macroeconomic time series of industrial production that has visually evident trend and (somewhat irregular) cyclical components over a long historical period. The US *industrial production index* used here is an indicator of aggregate economic activity that measures real production output of manufacturing, mining, and utility industries based on hundreds of individual time series. The series is seasonally adjusted, covers the last century from 1919:Q1–2018:Q2, and comprises 398 observations. It is one of the longest US quarterly macroeconomic series available from the Federal Reserve data base.[12]

[11] Aguiar and Gopinath (2007) report the moments after processing the cyclical components in a macroeconomic structural model. We directly work with the raw data here to keep the comparison as simple and straightforward as possible.

[12] Downloadable at `https://fred.stlouisfed.org/series/IPB50001SQ`.

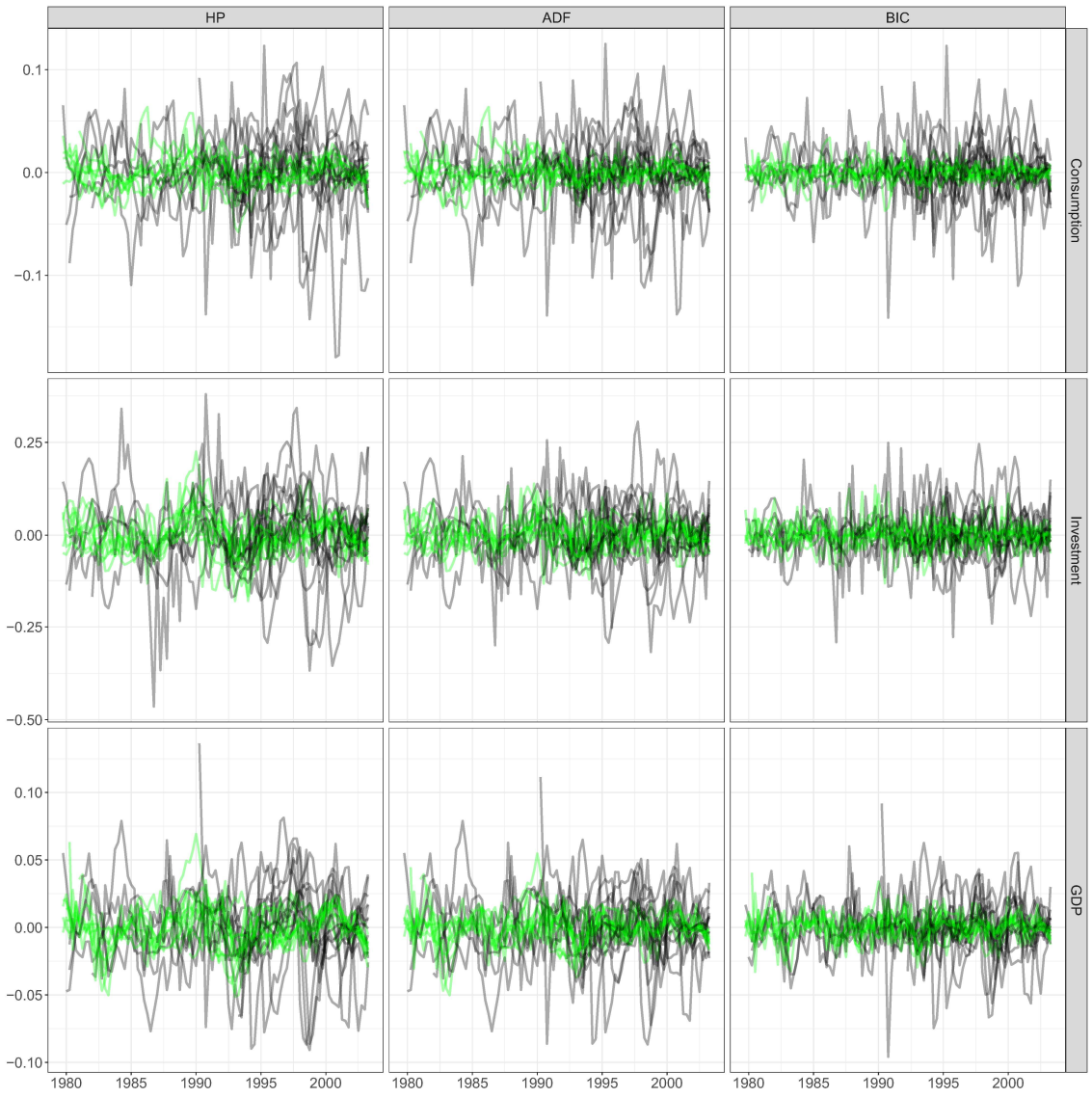


Figure 8: Cyclical components of GDP, Consumption, and Investment obtained by the HP, BHP-ADF and BHP-BIC filters, the latter with data-determined stopping. Developed nations are displayed in green and emerging nations are shown in black.

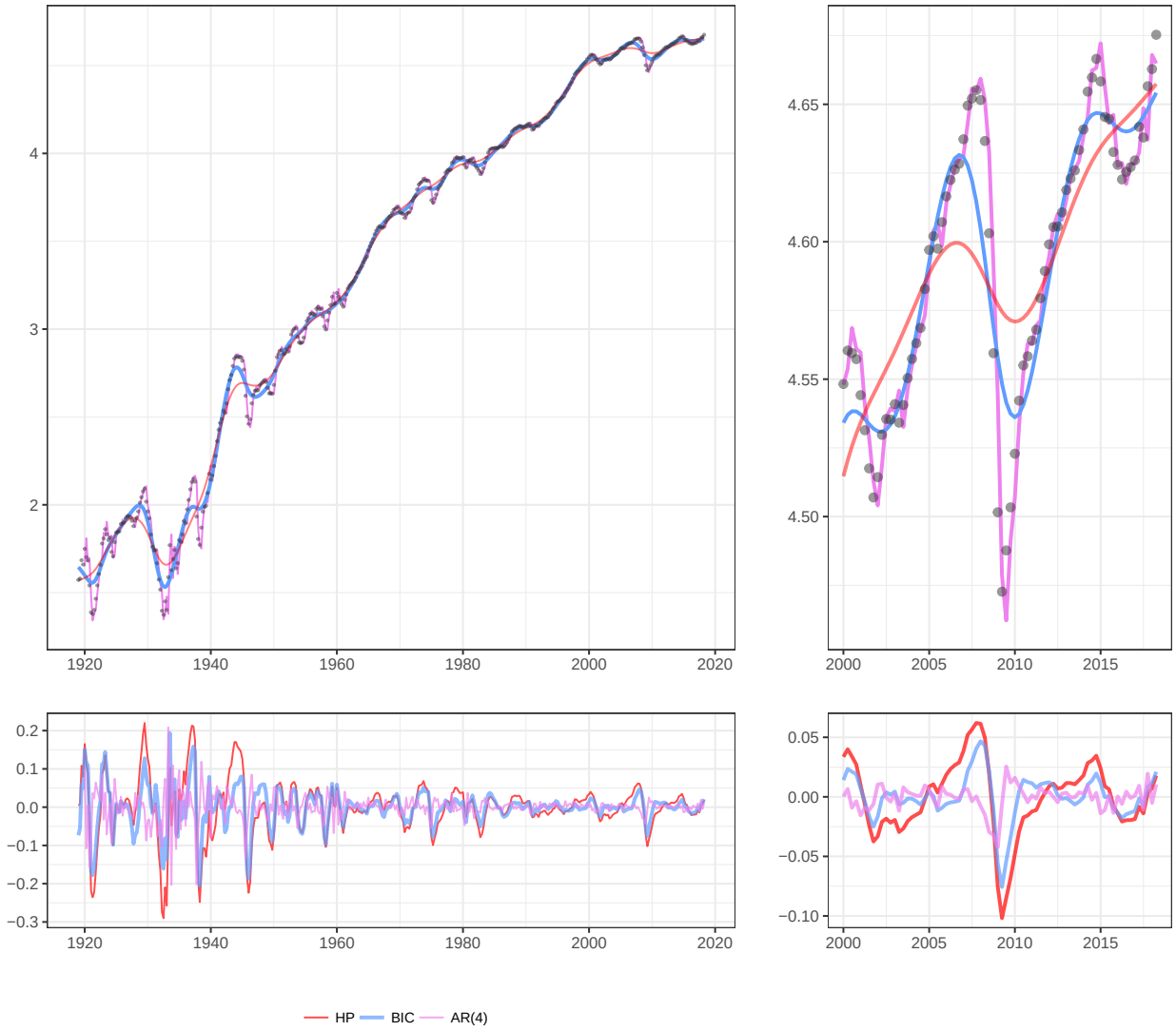


Figure 9: The upper two panels plot the time series and the fitted trend estimates. The lower two panels plot the fitted cyclical components. The left panel plots the time series (black dots), the estimated trend (upper panel) and cyclical (lower panel) components obtained by the HP (red) filter, the BHP-BIC (blue) filter, and AR(4) (violet) regression fitting. The right panel zooms in on the period 2000:Q1–2018:Q2 from the left panel to highlight the comparisons.

In Figure 9, the black dots in the top-left subgraph plots the logarithm of the raw time series, in which both the Great Depression and the recent Great Recession are clearly visible. The index is very volatile before the Second World War. Following the Second World War, fluctuations around the upward path of the index moderate but occur regularly until the end of the 20th century. The top-right subfigure zooms in on the more recent and more dramatic period of 21st century experience over 2000:Q1–2018:Q2.

It is common for macroeconomists, for example Romer (1999), to study the many changing features of long time series of this type by analyzing subperiods and comparing their defining characteristics across such periods. The HP filter approach, as well as other forms of trend extraction, seeks a unified econometric representation of the trend component of the entire series. The HP filter, shown by the red line in Figure 9, is created with smoothing parameter setting $\lambda = 1600$, and this filter accordingly smooths out the peaks and valleys of the index. When we fit the time series with the boosted filter, the BHP-ADF test rejects the null after a single HP-filter, thereby producing exactly the same results as the standard HP filter. Such early stopping based on a unit root test is unsurprising given the evidence of the HP filter's residual plot shown in the left-bottom subgraph. Despite the large error magnitude in the early years, the HP filter cyclical component oscillates around zero and little evidence is observed of a tendency to drift away from the mean. In contrast, the BHP-BIC filter involves 7 iterations and the resulting cyclical component (blue line) is much closer to zero with fewer large fluctuations. Moreover, compared to the HP filter, it is more responsive to the downturn of industrial production during the episode of the financial crisis. For example, in the top-right subgraph, the HP filter completely ignored the dot-com bubble collapse whereas the BHP-BIC filter declines in 2001 and subsequently reflects the serious impact of the Great Recession on the upward trend path of production. As a measure of potential industrial production, the estimated trend from the boosted HP filter is more consistent with the fundamental deterioration that many macroeconomists, such as Krugman (2012), perceived to have occurred in the aftermath of the financial crisis.

Figure 9 also reports results from the alternative AR(4) approach to filter this time series into trend and cycle. Noticeably, the fitted trend (shown by the violet line) closely tracks all observations in typical autoregressive fashion and there is little evidence of cycles in the residual. Moreover, during both crashes in the first decade of the 21st century shown in the right-top panel, the AR(4) fitted trend overshoots the extremes of the realized observations both before the burst of the financial bubble and at the end of the collapse that produced the downturn in the real economy that is reflected in the production index. This phenomenon is a typical feature of highly autoregressive (near unit root) fitting. In the present case the fitted AR(4) has long run autoregressive coefficient 0.9978 which is virtually unity.[13] Autoregressive fitting of time series tends to capture the fine grain as well as the global features of a time series trajectory, thereby removing most of the variation in the time series and reducing the residual to a series with martingale difference characteristics. This approach seems too rigid in its goal of removing variation to separate slow moving trend components from cyclical and stationary elements in time series. In particular, as this example demonstrates, the AR approach seems unable to effectively differentiate between trends and cycles in the economic environment, especially during episodes of extraordinary change that impact both trend and cyclical elements in economic activity. In the present case, the slow economic recovery following the crisis may be viewed as a distinguishing feature of the great recession cycle and the overall downturn that was induced may be considered as an inevitable impact on the trend (c.f., Krugman, 2012).

[13] The estimated coefficients of the AR(4) model intercept and the first to fourth lags are $0.011, 1.421, -0.514, 0.216$ and -0.1252, respectively. The long run autoregressive coefficient, based on the sum of the four autoregressive coefficients, is 0.9978. The characteristic polynomial has roots $[0.6123, 0.9959, -0.0936 \pm 0.4433i]$ and the complex roots correspond to a damped (amplitude 0.4531) cycle of 4.61 quarters or 1.15 years. The seasonally unadjusted industrial production series was also analyzed and produced very similar results to those given here, so they are not reported.

Table 3: Residual variance (×1000) from HP and BHP filtering and AR(4) autoregression

residual variance	HP	BIC	AR(4)
$x_t - \widehat{f}_t$	4.945	2.438	1.298
$x_{t+1} - \widehat{f}_t$	5.078	2.713	0.286

Our assessment of these findings of trend and cycle extraction with a century-long time series is that the BHP-BIC filter works well during periods of normal growth and mild cyclical activity and is also capable of capturing more complex heterogeneous features of the downturns in trend and slow variable recovery from recessions that arose in earlier and later years of this long historical period.

This application reveals some major differences between the filter and autoregressive approaches to cycle estimation. Table 3 shows the sample variance of the fitted cyclical components. In the first row of the table, it is clear that the autoregression has very small residual variance, amounting to approximately one half of that of the cyclical residual of the BHP-BIC filter and one quarter of that of the HP filter residual. This substantial gain in fitting the time series so well compared with the filtering methods arises from the capacity of the autoregression to capture a stochastic trend component in the time series parametrically (effectively by means of a near-unit root autoregression – here with a long run autoregressive coefficient 0.9978 that is so close to unity) – and to employ a parametric damped cycle of amplitude 0.4531 with period around 1.15 years that is induced by a pair of complex autoregressive roots in the characteristic equation of the autoregression. This fitted cycle has high frequency and is much closer to representing a damped seasonal oscillation in the production series than a business cycle, even though the time series that is used here is seasonally adjusted.

A more dramatic illustration of the forecast-oriented and fit-driven nature of the autoregression is apparent with a simple one-period phase shift. In particular, if we move the observed time series one period forward, then in the second row of the table the sample variance of $x_{t+1} - \widehat{f}_t$ is only about 20% that of the unshifted fit in the first row. Thus, the AR(4) is far better at predicting past values than future values of the trending trajectory, just as may be expected from a unit root autoregression fit of a far more complex time series trajectory. Such a one-period-ahead forecasting mechanism is highly effective in reducing residual variation. But this success in the AR mechanism can compromise its role in capturing more slowly varying (multi-period) trend effects in a time series that involve features over many time periods. Of course, by construction, the parametric AR(4) filter is fundamentally different from the HP filter and has different traditional modeling goals, such as forecasting using past information and studying the impact of past shocks on the system variable.[14]

By comparison, phase shifting in the case of the HP and BHP-BIC filters leads naturally to deterioration in residual variation. As seen in the table, the sample variance grows in the second row of the table for both HP filters. This is explained by the fact that these filters are designed as nonparametric penalized and smoothly varying best fits to each individual observation in the

[14]In his pioneering study of business cycles in the United States between 1919-1932, Tinbergen (1939, p.140) employed a fourth-order difference equation to describe the "systematic cyclical forces" in the US economy, analyzing the amplitude and period of the cycle, which turned out to be 4.8 years. Tinbergen claimed that the influence of further lags was "found to be small", a conclusion that matches the recommendation of Hamilton (2018) in using the AR(4) approach. In our present example, there is some evidence that extending the number of lags is beneficial in capturing in-sample fluctuation. Use of an AR(6), for example, reduces the residual variance of 1.298 (shown in the first row of Table 3) of the AR(4) to 1.087. Further, autoregressive lag determination by the standard BIC criterion clearly prefers AR(6) (BIC = -6.75) to AR(4) (BIC = -6.58) for the industrial production time series. The roots of the characteristic equation of the estimated AR(6) are $\{0.0256, 0.9971, -0.3934 \pm 0.6608i, 0.6360 \pm 0.4128i\}$, in which the two sets of complex roots indicate damped cycles of periods 1.51 years and 2.73 years.

time series trajectory, taking into account what is happening in neighboring observations. Such a mechanism seems more closely aligned with the general objective of historical trend determination when the concept of trend is based on the hypothesis articulated by Hodrick and Prescott (1997) in the header quotation that "the growth component of aggregate economic time series varies smoothly over time".

5 Conclusion

This paper explores the use of a machine learning modification of the HP filter that is designed for trend extraction in studying business cycles in macroeconomic data. The algorithm is based on the idea of repeated HP filtering and is linked to L_2-boosting methods that are now commonly used in machine learning approaches to linear regression. The boosted HP filter allows empirical investigators to continue to use a primary tuning parameter setting such as the standard $\lambda = 1600$ setting for quarterly data applications but alleviates the concern of using a single choice of this parameter in filtering time series of various lengths and persistence. To enhance the asymptotic performance of the HP filter, the boosted filter introduces a secondary tuning parameter that controls the degree of boosting while holding the primary parameter λ fixed. In practical applications, data-determined methods that rely on nonstationarity tests or information criteria may be used to select this secondary parameter in a convenient way. Asymptotic theory shows that the boosted HP filter has the capacity to consistently estimate, and thereby remove, a stochastic trend with time polynomial drift as well as a stochastic trend with deterministic polynomial drift and multiple structural breaks. These results seem relevant for many empirical applications in which standard HP tuning parameter settings are currently used.

The limit theory reveals some of the capabilities of HP filter methods as trend fitting and extraction processes for practical work that have heretofore been little understood. The methodology allows empirical researchers to rely on existing software for standard implementation of the HP filter in the boosting environment. Like ordinary least squares and vector autoregressions, empirically convenient methods such as the HP filter are unlikely ever to be put out of business, in spite of concerns that have been repeatedly raised over many years about their usefulness and their effects on subsequent analysis.

It is hoped that the present contribution will help empirical researchers to better understand the capabilities and limitations of the HP filter and to guide the implementation of a simple machine-learning vehicle for its improvement in applications, thereby mitigating some of the limitations of the HP filter itself. Our position is therefore more optimistic than that of Hamilton (2018). As we have shown, a key advantage of the boosted filter is that it provides a new device for consistently estimating in a nonparametric manner a wide class of trending mechanisms including processes that involve structural breaks, while remaining agnostic about the precise form of the trend nonstationarity in the data. These are central advantages of the boosted filter that justify some optimism, support continuing empirical use, and clearly distinguish the methods from competitor approaches that are typically reliant on correct model specification.

To implement the automated boosted HP filters in practical work, we have developed a documented R function `BoostedHP` along with a test example to assist empirical researchers. These programs may be downloaded from the following website
`https://github.com/zhentaoshi/Boosted_HP_filter`.

References

Aguiar, M., and G. Gopinath (2007): "Emerging market business cycles: The cycle is the trend," *Journal of Political Economy*, 115(1), 69–102.

Aruoba, S. B., and F. X. Diebold (2010): "Real-time macroeconomic monitoring: Real activity, inflation, and interactions," *American Economic Review*, 100(2), 20–24.

Aruoba, S. B., F. X. Diebold, M. A. Kose, and M. E. Terrones (2010): "Globalization, the Business Cycle, and Macroeconomic Monitoring," in *NBER International Seminar on Macroeconomics 2010*, pp. 245–286. University of Chicago Press.

Baek, Y. I., J. S. Cho, and P. C. Phillips (2015): "Testing linearity using power transforms of regressors," *Journal of Econometrics*, 187(1), 376–384.

Bai, J., and S. Ng (2009): "Boosting diffusion indices," *Journal of Applied Econometrics*, 24(4), 607–629.

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

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

Belloni, A., D. Chen, V. Chernozhukov, and C. Hansen (2012): "Sparse models and methods for optimal instruments with an application to eminent domain," *Econometrica*, 80(6), 2369–2429.

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.

Blanchard, G., M. Hoffmann, and M. Reiss (2017): "Optimal adaptation for early stopping in statistical inverse problems," *arXiv preprint arXiv:1606.07702*.

Bühlmann, P. (2006): "Boosting for high-dimensional linear models," *The Annals of Statistics*, pp. 559–583.

Bühlmann, P., and B. Yu (2003): "Boosting with the L 2 loss: regression and classification," *Journal of the American Statistical Association*, 98(462), 324–339.

Buja, A., T. Hastie, and R. Tibshirani (1989): "Linear smoothers and additive models," *The Annals of Statistics*, pp. 453–510.

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.

Chernozhukov, V., C. Hansen, and M. Spindler (2015): "Post-selection and post-regularization inference in linear models with many controls and instruments," *American Economic Review: Papers and Proceedings*, 105(5), 486–90.

Cho, J. S., and P. C. Phillips (2018): "Sequentially testing polynomial model hypotheses using power transforms of regressors," *Journal of Applied Econometrics*, 33(1), 141–159.

Christiano, L. J., and T. J. Fitzgerald (2003): "The band pass filter," *International Economic Review*, 44(2), 435–465.

Cogley, T., and J. M. Nason (1995): "Effects of the Hodrick-Prescott filter on trend and difference stationary time series Implications for business cycle research," *Journal of Economic Dynamics and control*, 19(1-2), 253–278.

Corbae, D., and S. Ouliaris (2006): "Extracting cycles from nonstationary data," *Econometric Theory and Practice: Frontiers of Analysis and Applied Research*, pp. 167–77.

Cornea-Madeira, A. (2017): "The Explicit Formula for the Hodrick-Prescott Filter in a Finite Sample," *Review of Economics and Statistics*.

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

Diebold, F. X., and L. Kilian (2000): "Unit-root tests are useful for selecting forecasting models," *Journal of Business & Economic Statistics*, 18(3), 265–273.

Fan, J., Y. Liao, and J. Yao (2015): "Power Enhancement in High-Dimensional Cross-Sectional Tests," *Econometrica*, 83(4), 1497–1541.

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.

Granger, C. W., and P. Newbold (1974): "Spurious regressions in econometrics," *Journal of Econometrics*, 2(2), 111–120.

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

Hastie, T., R. Tibshirani, and J. Friedman (2009): *The elements of statistical learning: Data mining, inference, and prediction*. Springer-Verlag.

Hirano, K., and J. H. Wright (2017): "Forecasting with model uncertainty: Representations and risk reduction," *Econometrica*, 85(2), 617–643.

Hodrick, R. J., and E. C. Prescott (1997): "Postwar US business cycles: an empirical investigation," *Journal of Money, Credit, and Banking*, pp. 1–16.

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.

Kozbur, D. (2017): "Testing-Based Forward Model Selection," *American Economic Review: Papers and Proceedings*, 107(5), 266–269.

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.

Leser, C. (1961): "A simple method of trend construction," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 91–107.

Luo, Y., and M. Spindler (2017): "L2-Boosting for Economic Applications," *arXiv:1702.03244*.

McCracken, M. W., and S. Ng (2016): "FRED-MD: A monthly database for macroeconomic research," *Journal of Business & Economic Statistics*, 34(4), 574–589.

Newey, W. K., F. Hsieh, and J. M. Robins (2004): "Twicing kernels and a small bias property of semiparametric estimators," *Econometrica*, 72(3), 947–962.

Ng, S. (2014): "Viewpoint: boosting recessions," *Canadian Journal of Economics/Revue canadienne d'économique*, 47(1), 1–34.

Noschese, S., L. Pasquini, and L. Reichel (2013): "Tridiagonal Toeplitz matrices: properties and novel applications," *Numerical linear algebra with applications*, 20(2), 302–326.

Okun, A. M. (1962): *Potential GNP: its measurement and significance*, no. 190. Yale University, Cowles Foundation for Research in Economics.

Paige, R. L., and A. A. Trindade (2010): "The Hodrick-Prescott Filter: A special case of penalized spline smoothing," *Electronic Journal of Statistics*, 4, 856–874.

Phillips, P. C., and V. Solo (1992): "Asymptotics for linear processes," *The Annals of Statistics*, pp. 971–1001.

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

——— (1998): "New tools for understanding spurious regressions," *Econometrica*, 66(6), 1299–1325.

——— (2005): "Automated discovery in econometrics," *Econometric Theory*, 21(1), 3–20.

——— (2007): "Unit root log periodogram regression," *Journal of Econometrics*, 138(1), 104–124.

——— (2010): "Two New Zealand pioneer econometricians," *New Zealand Economic Papers*, 44(1), 1–26.

Phillips, P. C. B., and S. Jin (2015): "Business Cycles, Trend Elimination, and the HP Filter," Discussion paper, Yale University, Cowles Foundation Discussion Paper 2005.

Phillips, P. C. B., and P. Perron (1988): "Testing for a unit root in time series regression," *Biometrika*, pp. 335–346.

Raskutti, G., M. J. Wainwright, and B. Yu (2014): "Early stopping and non-parametric regression: an optimal data-dependent stopping rule.," *Journal of Machine Learning Research*, 15(1), 335–366.

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.

Romer, C. D. (1999): "Changes in business cycles: evidence and explanations," *Journal of Economic Perspectives*, 13(2), 23–44.

Sakarya, N., and R. M. de Jong (2017): "A property of the Hodrick-Prescott filter and its application," in *Technical Report*. The Ohio State University.

Schapire, R. E. (1990): "The strength of weak learnability," *Machine learning*, 5(2), 197–227.

Schlicht, E. (2005): "Estimating the smoothing parameter in the so-called Hodrick-Prescott filter," *Journal of the Japan Statistical Society*, 35(1), 99–119.

Shi, Z. (2016): "Econometric estimation with high-dimensional moment equalities," *Journal of Econometrics*, 195(1), 104–119.

Tibshirani, R. (1996): "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267–288.

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.

Tutz, G., and H. Binder (2006): "Generalized Additive Modeling with Implicit Variable Selection by Likelihood-Based Boosting," *Biometrics*, 62(4), 961–971.

Whittaker, E. T. (1922): "On a new method of graduation," *Proceedings of the Edinburgh Mathematical Society*, 41, 63–75.

Whittaker, E. T., and G. Robinson (1924): *The Calculus of Observations: A Treatise on Numerical Mathematics*. Blackie & Son.

A Proofs

Proof of Theorem 1. We apply the approach used in the proof of Theorem 3 of PJ (2015) with some new modifications. The idea is to use the KL series representations (5) and (6) and the fact that these series converge almost surely and uniformly in r so that successive pseudo-integral operations may be applied to them to obtain the asymptotic form of the boosted filter. The derivations here primarily relate to the asymptotic impact of boosting.

Write $X_n(r) = n^{-1/2} x_{\lfloor nr \rfloor}$. By Lemma 3.1 of Phillips (2007) it is known that an expanded probability space can be constructed with a corresponding limiting Brownian motion for which the following uniform convergence holds almost surely

$$\sup_{0 \leq t \leq n} \left| X_n\left(\frac{t}{n}\right) - B\left(\frac{t}{n}\right) \right| = o_{a.s.}(1).$$

In what follows calculations are made in this expanded space where almost sure convergence applies and in the original space the results translate as usual into weak convergence.

Since the KL series representation (6) of $B(r)$ converges almost surely and uniformly in r we may use a finite series KL approximation $B^{K_n}(r) = \sum_{k=1}^{K_n} \sqrt{\lambda_k} \varphi_k\left(\frac{t}{n}\right) \xi_k$ with the property that for $K_n \to \infty$ we have $\sup_{0 \leq r \leq 1} \left| B^{K_n}(r) - B(r) \right| = o_{a.s.}(1)$. Then

$$\sup_{0 \leq t \leq n} \left| X_n\left(\frac{t}{n}\right) - B^{K_n}\left(\frac{t}{n}\right) \right| = o_{a.s.}(1), \tag{15}$$

if $K_n \to \infty$ as $n \to \infty$. It follows that $X_n\left(\frac{t}{n}\right)$ is almost surely uniformly well approximated by $B^{K_n}\left(\frac{t}{n}\right)$ for all $t \leq n$ as $n \to \infty$. We therefore consider the effect of the HP trend operator $G_\lambda = \frac{1}{\lambda L^{-2}(1-L)^4 + 1}$ directly upon $B^{K_n}\left(\frac{t}{n}\right) = \sum_{k=1}^{K_n} \sqrt{\lambda_k} \varphi_k\left(\frac{t}{n}\right) \xi_k$ as $n \to \infty$.

Noting that $\sqrt{\lambda_k} = \left[\left(k - \frac{1}{2}\right)\pi\right]^{-1}$ and $\varphi_k\left(\frac{t}{n}\right) = \sqrt{2} \sin\left(\frac{t/n}{\sqrt{\lambda_k}}\right)$, we have, analogous to PJ (2015),

$$\begin{aligned}
n(1-L)\varphi_k\left(\frac{t}{n}\right) &= \sqrt{2} n \left[\sin\left(\frac{\frac{t}{n}}{\sqrt{\lambda_k}}\right) - \sin\left(\frac{\frac{t-1}{n}}{\sqrt{\lambda_k}}\right) \right] \\
&= \sqrt{2} \cdot 2n \cdot \sin\left(\frac{1}{2n\sqrt{\lambda_k}}\right) \cos\left(\frac{\frac{2t-1}{n}}{2\sqrt{\lambda_k}}\right) \\
&= \frac{\sqrt{2}}{\sqrt{\lambda_k}} \cdot \frac{\sin\left(\frac{1}{2n\sqrt{\lambda_k}}\right)}{\frac{1}{2n\sqrt{\lambda_k}}} \cos\left(\frac{\frac{2t-1}{n}}{2\sqrt{\lambda_k}}\right) = O\left(\frac{1}{\sqrt{\lambda_k}}\right)
\end{aligned} \tag{16}$$

where the result holds uniformly in $t \leq n$ and uniformly for all $k \geq 1$ since $\frac{\sin x}{x}$ and $\cos(x)$ are both uniformly bounded. It follows from (16) that

$$n(1-L)\varphi_k\left(\frac{t}{n}\right) = \frac{\sqrt{2}}{\sqrt{\lambda_k}} \cos\left(\frac{\frac{t}{n}}{\sqrt{\lambda_k}}\right) \{1 + o(1)\}, \quad \text{as } n \to \infty, \tag{17}$$

uniformly in $t \leq n$ and uniformly in $k \leq K_n$ with $n\sqrt{\lambda_{K_n}} \to \infty$, which holds when $K_n/n \to 0$. By repeated argument as in (16) we find that

$$L^{-2}\left[n(1-L)\right]^4 \varphi_k\left(\frac{t}{n}\right) = \frac{\sqrt{2}}{\lambda_k^2} \sin\left(\frac{\frac{t}{n}}{\sqrt{\lambda_k}}\right) \{1 + o(1)\} = \frac{\varphi_k\left(\frac{t}{n}\right)}{\lambda_k^2} \{1 + o(1)\}, \tag{18}$$

again uniformly in $t \leq n$ and uniformly for all $k \leq K_n$ whenever $K_n/n \to 0$. Similarly, as in (16), we have $L^{-2}\left[n\left(1-L\right)\right]^4 \varphi_k\left(\frac{t}{n}\right) = O\left(1/\lambda_k^2\right)$ for all $k \geq 1$ and $t \leq n$.

Next observe that when $\lambda = \mu n^4$ for some $\mu > 0$ we have

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

Using (18) and the operational calculus in PJ (2015) we find that

$$\begin{aligned}\left(1-G_\lambda\right)^m \varphi_k\left(\frac{t}{n}\right) &= \left(\frac{\mu L^{-2}\left[\left(n\left(1-L\right)\right]^4\right.}{\mu L^{-2}\left[\left(n\left(1-L\right)\right]^4+1\right.}\right)^m \varphi_k\left(\frac{t}{n}\right) \\ &= \left(\frac{\frac{\mu}{\lambda_k^2}}{1+\frac{\mu}{\lambda_k^2}}\right)^m \varphi_k\left(\frac{t}{n}\right)\left\{1+o\left(1\right)\right\} = \left(\frac{\mu}{\mu+\lambda_k^2}\right)^m \varphi_k\left(\frac{t}{n}\right)\left\{1+o\left(1\right)\right\}, \quad (19)\end{aligned}$$

whenever $K_n/n \to 0$, as in (18). Then, for $\lambda = \mu n^4$ with $\mu > 0$ and using (19), we have for the boosted HP filter

$$\begin{aligned}\widehat{f}^{(m)}_{\frac{t}{n},K_n} &:= \left[1-\left(1-G_\lambda\right)^m\right] B^{K_n}\left(r\right) = \sum_{k=1}^{K_n} \sqrt{\lambda_k}\left[1-\left(1-G_\lambda\right)^m\right] \varphi_k\left(\frac{t}{n}\right)\xi_k \\ &= \sum_{k=1}^{K_n} \sqrt{\lambda_k}\left[1-\left(\frac{\mu}{\mu+\lambda_k^2}\right)^m\right] \varphi_k\left(\frac{t}{n}\right)\xi_k + o_{a.s.}(1) \quad (20)\end{aligned}$$

as $n \to \infty$ with $\frac{K_n}{n} \to 0$ as in (19). Next, observe that as $m \to \infty$

$$\sum_{k=1}^{K_n} \sqrt{\lambda_k}\left[1-\left(\frac{\mu}{\mu+\lambda_k^2}\right)^m\right] \varphi_k\left(\frac{t}{n}\right)\xi_k \to_{a.s.} \sum_{k=1}^{K_n} \sqrt{\lambda_k}\varphi_k\left(\frac{t}{n}\right)\xi_k,$$

because, for all $k \leq K_n$, $\left(\frac{\mu}{\mu+\lambda_k^2}\right)^m \leq \left(\frac{\mu}{\mu+\lambda_{K_n}^2}\right)^m = \left(1-\frac{\lambda_{K_n}^2}{\mu+\lambda_{K_n}^2}\right)^m \to 0$ as $m \to \infty$ and $K_n^4/m \to 0$.

We deduce that

$$\begin{aligned}\widehat{f}^{(m)}_{\frac{t}{n},K_n} &= \sum_{k=1}^{K_n} \sqrt{\lambda_k}\left[1-\left(\frac{\mu}{\mu+\lambda_k^2}\right)^m\right] \varphi_k\left(\frac{t}{n}\right)\xi_k + o_{a.s.}(1) \\ &= \sum_{k=1}^{K_n} \sqrt{\lambda_k}\varphi_k\left(\frac{t}{n}\right)\xi_k + o_{a.s.}(1), \quad (21)\end{aligned}$$

as $m, n \to \infty$ with $K_n/n + K_n^4/m \to 0$. It follows that $\widehat{f}^{(m)}_{\frac{t=\lfloor nr \rfloor}{n},K_n} \to_{a.s.} B^{K_n}\left(r\right)$ as $m, n \to \infty$. Moreover, since $B^{K_n}\left(r\right) \to_{a.s.} B(r)$ as $K_n \to \infty$, we deduce that

$$\widehat{f}^{(m)}_{\frac{t=\lfloor nr \rfloor}{n},K_n} \to_{a.s.} B\left(r\right), \quad (22)$$

as $m, K_n, n \to \infty$ with $K_n/n \to 0$, which together ensure that (18) and (19) hold. Then, since $X_n\left(\frac{t}{n}\right)$ is almost surely uniformly well approximated by $B^{K_n}\left(\frac{t}{n}\right)$ for all $t \leq n$ as $n \to \infty$, it follows that $\frac{\widehat{f}^{(m)}_{t=\lfloor nr \rfloor}}{\sqrt{n}} \to_{a.s.} B(r)$, as $m, n \to \infty$. In the original probability space, this means that

$\frac{\widehat{f}^{(m)}_{t=\lfloor nr \rfloor}}{\sqrt{n}} \rightsquigarrow B(r)$, and the stated result follows. □

Proof of Theorem 2. The stochastic trend component x_t^0 of x_t is handled in the same way as the proof of Theorem 1. It remains to show that repeated applications of the HP filter in the boosting algorithm preserve the deterministic polynomial component of the J-th degree in x_t . Let $y_t = \beta_1 \frac{t}{n} + \cdots + \beta_J \left(\frac{t}{n}\right)^J$ denote the polynomial trend so that $x_t = y_t + x_t^0$. The boosting algorithm applies the operator

$$1 - G_\lambda = \frac{\lambda L^{-2} (1-L)^4}{\lambda L^{-2} (1-L)^4 + 1}$$

repeatedly m times. Taking the j-th term of y_t, we note that after m iterations

$$(1 - G_\lambda)^m \left(\frac{t}{n}\right)^j = \frac{\lambda^m L^{-2m} (1-L)^{4m} \left(\frac{t}{n}\right)^j}{\left(\lambda L^{-2} (1-L)^4 + 1\right)^m} = 0 \tag{23}$$

for all $j \leq 4m-1$, since in that event the numerator component of the operator yields

$$\lambda^m L^{-2m} (1-L)^{4m} \left(\frac{t}{n}\right)^j = 0. \tag{24}$$

The polynomial degree J of y_t is finite and so whenever the number of iterations m in the boosted HP algorithm is sufficiently large in the sense that $4m - 1 \geq J$, we have $\left(1 - (1 - G_\lambda)^m\right) y_t = y_t$. The denominator component of the operator in (23) may be handled as in PJ (2015) using Fourier methods. An alternative approach to handling the effect of the denominator in this case is to use an integral version of the operator. In particular,

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

just as in (24) above for all m such that $4m \geq J + 1 \geq j + 1$. Thus, (23) holds and boosting removes the polynomial component of the time series. In effect, boosting the HP algorithm raises its capacity to capture accurately a polynomial trend of any finite order as $m \to \infty$. □

Proof of Asymptotic Approximation (13). We need to show that as $n \to \infty$

$$\operatorname{tr}(B_m) = \operatorname{tr}\left(I_n - (I_n - S(\lambda))^m\right) = n - \operatorname{tr}\left[\sum_{k=1}^{n-2} \left(\frac{4\mu n^4 \left(1 - \cos\left(\frac{k\pi}{n-1}\right)^2\right)}{1 + 4\mu n^4 \left(1 - \cos\left(\frac{k\pi}{n-1}\right)^2\right)}\right)^m\right] \{1 + o(1)\}. \tag{25}$$

Define $d_2' = (1, -2, 1)$ and the $(n-2)\times 1$ unit vectors $\underset{(n-2)\times 1}{e_1} = (1, 0, ..., 0)'$ and $\underset{(n-2)\times 1}{e_n} = (0, 0, ..., 1)'$. We then write the $(n-2) \times n$ second differencing matrix D' as

$$D' = \begin{bmatrix} d_2' & 0 & 0 & \cdots & 0 \\ & d_2' & 0 & \cdots & 0 \\ & & d_2' & \cdots & 0 \\ & & & \ddots & 0 \\ & & & & d_2' \end{bmatrix} = [e_1, M, e_n],$$

where M is the $(n-2) \times (n-2)$ tridiagonal symmetric Toeplitz matrix

$$M = \begin{bmatrix} -2 & 1 & 0 & & & \\ 1 & -2 & 1 & \ddots & & \\ & \ddots & \ddots & \ddots & \ddots & \\ & & 1 & -2 & 1 & 0 \\ & & & 1 & -2 & 1 \\ & & & & 1 & -2 \end{bmatrix}.$$

Use the inverse matrix formula

$$\begin{aligned} S(\lambda) &= I_n - \lambda D \left(I_{n-2} + \lambda D'D\right)^{-1} D' \\ &= I_n - \lambda [e_1, M, e_n]' \left(I_{n-2} + \lambda \left(e_1 e_1' + M^2 + e_n e_n'\right)\right)^{-1} [e_1, M, e_n] \\ &\sim I_n - \lambda [e_1, M, e_n]' \left(I_{n-2} + \lambda M^2\right)^{-1} [e_1, M, e_n], \end{aligned}$$

where the last line follows by

$$I_{n-2} + \lambda D'D = I_{n-2} + \lambda \left(e_1 e_1' + M^2 + e_n e_n'\right) \sim I_{n-2} + \lambda M^2,$$

ignoring the end matrix corrections to the first and last diagonal elements as $n \to \infty$. The $(n-2) \times (n-2)$ central matrix elements of $S(\lambda)$ are given by the symmetric matrix

$$S_{n-2}(\lambda) = I_{n-2} - \lambda M \left(I_{n-2} + \lambda M^2\right)^{-1} M. \tag{26}$$

The eigenvalues of the symmetric tridiagonal Toeplitz matrix M are well known (Noschese, Pasquini, and Reichel, 2013) to be given by $\delta_k = -2\left(1 - \cos\left(\frac{k\pi}{n-1}\right)\right)$, $k = 1, ..., n-2$, and so the eigenvalues of M^2 are $\left\{\delta_k^2\right\}_{k=1}^{n-2}$. It follows that the eigenvalues of $S_{n-2}(\lambda)$ are given by

$$\left\{1 - \frac{\lambda\delta_k^2}{1 + \lambda\delta_k^2}\right\}_{k=1}^{n-2} = \left\{\frac{1}{1 + \lambda\delta_k^2}\right\}_{k=1}^{n-2}.$$

We deduce that

$$\begin{aligned}\operatorname{tr}(B_m) &= n - \operatorname{tr}\left(\left(I_n - S(\lambda)\right)^m\right) \sim n - \operatorname{tr}\left(\left(I_{n-2} - S_{n-2}(\lambda)\right)^m\right) \\ &= n - \sum_{k=1}^{n-2}\left(1 - \frac{1}{1+\lambda\delta_k^2}\right)^m = n - \sum_{k=1}^{n-2}\left(\frac{\lambda\delta_k^2}{1+\lambda\delta_k^2}\right)^m \\ &= n - \sum_{k=1}^{n-2}\left[\frac{4\mu n^4\left(1-\cos\left(\frac{k\pi}{n-1}\right)^2\right)}{1+4\mu n^4\left(1-\cos\left(\frac{k\pi}{n-1}\right)^2\right)}\right]^m,\end{aligned}$$

as required for (25) when $\lambda = \mu n^4$. □

Proof of Theorem 3. We follow the line of argument used in the proof of Theorem 2. The polynomial component $g_n(t)$ is handled in a similar way to the proof of Theorem 2 but with complications arising from the break separating the two segments $t < \tau_0$ and $t > \tau_0$. We analyze these two segments in turn first. Then, following those arguments, we consider limit behavior of the boosted filter at the break point r_0 itself.

(i) Lower Segment

Over the lower segment we have $t = \lfloor nr \rfloor < \tau_0 = \lfloor nr_0 \rfloor$ for $r < r_0$, so that for large enough n we have

$$t + 2m = \lfloor nr \rfloor + 2m < \tau_0 = \lfloor nr_0 \rfloor, \tag{27}$$

which implies that $\mathbf{1}\{t + 2m < \tau_0\} = 1$ since $m = o(n)$. Repeated applications of the HP filter in the boosting algorithm lead to

$$(1 - G_\lambda)^m \left(\frac{t}{n}\right)^j \mathbf{1}\{t < \tau_0\} = \left(\frac{L^{-2}(1-L)^4}{L^{-2}(1-L)^4 + 1/\lambda}\right)^m \left(\frac{t}{n}\right)^j \mathbf{1}\{t < \tau_0\}. \tag{28}$$

Applying the numerator operator gives

$$\begin{aligned}&L^{-2m}(1-L)^{4m}\left(\frac{t}{n}\right)^j \mathbf{1}\{t < \tau_0\} = (1-L)^{4m}\left(\frac{t+2m}{n}\right)^j \mathbf{1}\{t + 2m < \tau_0\} \\ =\ &(1-L)^{4m}\left(\frac{t+2m}{n}\right)^j = 0,\end{aligned}$$

because $\mathbf{1}\{t + 2m - k < \tau_0\} = 1$ for all $k \geq 0$ in view of (27) and because $(1-L)^{4m} t^j = 0$ for all $j \leq J$, which is finite, and $m \to \infty$ which ensures that $4m \geq J + 1$. Next, combining the numerator and denominator operators in (28) we have

$$
\begin{aligned}
&(1-G_\lambda)^m\left(\frac{t}{n}\right)^j = \left(\frac{L^{-2}(1-L)^4}{L^{-2}(1-L)^4+\frac{1}{\lambda}}\right)^m\left(\frac{t}{n}\right)^j \mathbf{1}\{t<\tau_0\}\\
= &\frac{1}{\Gamma(m)}\int_0^\infty s^{m-1}e^{-s}e^{-s\left[L^{-2}(1-L)^4+\frac{1}{\lambda}\right]}ds\,(1-L)^{4m}\left(\frac{t+2m}{n}\right)^j \mathbf{1}\{t+2m<\tau_0\}\\
= &\frac{1}{\Gamma(m)}\sum_{k=0}^\infty\frac{(-1)^k}{k!}\int_0^\infty s^{m+k-1}e^{-s}ds\left[L^{-2}(1-L)^4+\frac{1}{\lambda}\right]^k(1-L)^{4m}\left(\frac{t+2m}{n}\right)^j \mathbf{1}\{t+2m<\tau_0\}\\
= &\sum_{k=0}^\infty\frac{\Gamma(m+k)(-1)^k}{\Gamma(m)k!}\left[L^{-2}(1-L)^4+\frac{1}{\lambda}\right]^k(1-L)^{4m}\left(\frac{t+2m}{n}\right)^j \mathbf{1}\{t+2m<\tau_0\}\\
= &\sum_{k=0}^\infty\frac{(m)_k(-1)^k}{k!}\left[L^{-2}(1-L)^4+\frac{1}{\lambda}\right]^k(1-L)^{4m}\left(\frac{t+2m}{n}\right)^j \mathbf{1}\{t+2m<\tau_0\}. \qquad (29)
\end{aligned}
$$

Observe that if $4m > J+1$

$$
\begin{aligned}
&\left[L^{-2}(1-L)^4+\frac{1}{\lambda}\right]^k(1-L)^{4m}\left(\frac{t+2m}{n}\right)^j \mathbf{1}\{t+2m<\tau_0\}\\
= &\sum_{i=0}^k\binom{k}{i}\frac{1}{\lambda^{k-i}}L^{-2i}(1-L)^{4[m+i]}\left(\frac{t+2m}{n}\right)^j \mathbf{1}\{t+2m<\tau_0\}\\
= &\sum_{i=0}^k\binom{k}{i}\frac{1}{\lambda^{k-i}}(1-L)^{4[m+i]}\left(\frac{t+2[m+i]}{n}\right)^j \mathbf{1}\{t+2[m+i]<\tau_0\}\\
= &\ 0 \qquad (30)
\end{aligned}
$$

for all $j \le J$ and for all $k \le k_n = \lfloor n(r_0-r)/4 \rfloor$ because then for large enough n and with $m = o(n)$ we have

$$
t+2[m+i] \le \lfloor nr \rfloor + 2[m+k] \le \lfloor nr \rfloor + 2m + \lfloor n(r_0-r)/2) \rfloor + 1 < \lfloor nr_0 \rfloor = \tau_0, \qquad (31)
$$

in which case $\mathbf{1}\{t+2[m+i]<\tau_0\} = 1$ for all $i \le k \le k_n$, just as in (27). In this event, it follows that

$$
(1-L)^{4[m+i]}\left(\frac{t+2[m+i]}{n}\right)^j \mathbf{1}\{t+2[m+i]<\tau_0\} = (1-L)^{4[m+i]}\left(\frac{t+2[m+i]}{n}\right)^j = 0, \quad (32)
$$

and then

$$
\left[L^{-2}(1-L)^4+\frac{1}{\lambda}\right]^k(1-L)^{4m}\left(\frac{t+2m}{n}\right)^j \mathbf{1}\{t+2m<\tau_0\} = 0, \qquad (33)
$$

for all $k \le k_n$. Hence, the first k_n terms of the series in (29) are zero for large enough n and so

$$
\begin{aligned}
&\sum_{k=0}^\infty\frac{(m)_k(-1)^k}{k!}\left[L^{-2}(1-L)^4+\frac{1}{\lambda}\right]^k(1-L)^{4m}\left(\frac{t+2m}{n}\right)^j \mathbf{1}\{t+2m<\tau_0\}\\
= &\sum_{k>k_n=\lfloor n(r_0-r)/4 \rfloor}^\infty\frac{(m)_k(-1)^k}{k!}\left[L^{-2}(1-L)^4+\frac{1}{\lambda}\right]^k(1-L)^{4m}\left(\frac{t+2m}{n}\right)^j \mathbf{1}\{t+2m<\tau_0\}
\end{aligned}
$$

Next, for $k > k_n = \lfloor n\left(\frac{r_0-r}{4}\right)\rfloor$, using (32) we have

$$\begin{aligned}
&\left[L^{-2}(1-L)^4+\frac{1}{\lambda}\right]^k (1-L)^{4m}\left(\frac{t+2m}{n}\right)^j \mathbf{1}\{t+2m<\tau_0\}\\
=\;& (1-L)^{4m}\sum_{i=0}^{k}\binom{k}{i}\frac{1}{\lambda^{k-i}}L^{-2i}(1-L)^{4i}\left(\frac{t+2m}{n}\right)^j \mathbf{1}\{t+2m<\tau_0\}\\
=\;& (1-L)^{4m}\sum_{i=k_n+1}^{k}\binom{k}{i}\frac{1}{\lambda^{k-i}}(1-L)^{4i}\left(\frac{t+2[m+i]}{n}\right)^j \mathbf{1}\{t+2[m+i]<\tau_0\}\\
=\;& (1-L)^{4m}\sum_{i=k_n+1}^{k}\binom{k}{i}\frac{1}{\lambda^{k-i}}(1-L)^{4i}\left(\frac{t+2[m+i]}{n}\right)^j \mathbf{1}\{t+2[m+i]<\tau_0\}
\end{aligned}$$

for all $j \le J$. When $i \ge i_n := \lfloor n\left(\frac{r_0-r}{2}\right)\rfloor$ we have

$$t+2[m+i] \ge \lfloor nr\rfloor + 2[m+\lfloor n(r_0-r)/2\rfloor] > \lfloor nr\rfloor + \lfloor n(r_0-r)\rfloor + m > \lfloor nr_0\rfloor = \tau_0,$$

which implies $\mathbf{1}\{t+2[m+i]<\tau_0\}=0$. However, even in this case, application of powers of the differencing operator $(1-L)^{4i}$ leads to adjacent integer values for which $\{t+2[m+i]-(p-1)>\tau_0\}$ and $\{t+2[m+i]-p<\tau_0\}$ for some positive integer p. Hence,

$$\begin{aligned}
&(1-L)^{4i}\left[\left(\frac{t+2[m+i]}{n}\right)^j \mathbf{1}\{t+2[m+i]<\tau_0\}\right]\\
=\;& \sum_{p=0}^{4i}\binom{4i}{p}(-1)^p L^p\left[\left(\frac{t+2[m+i]}{n}\right)^j \mathbf{1}\{t+2[m+i]<\tau_0\}\right]\\
=\;& \sum_{p=0}^{4i}\binom{4i}{p}(-1)^p\left(\frac{t+2[m+i]-p}{n}\right)^j \mathbf{1}\{t+2[m+i]-p<\tau_0\}\\
=\;& \sum_{p=P_{(t)}}^{4i}\binom{4i}{p}(-1)^k\left(\frac{t+2[m+i]-p}{n}\right)^j,
\end{aligned}$$

where $P_{(t)}$ is the smallest p for which $\{0 \le t+2[m+i]-p<\tau_0\}$, in which case $\mathbf{1}\{t+2[m+i]-p<\tau_0\} =$

1 for $p \geq P_{(t)}$. We then deduce that

$$\begin{aligned}
&\sum_{k=0}^{\infty} \frac{(m)_k (-1)^k}{k!} \left[L^{-2} (1-L)^4 + \frac{1}{\lambda} \right]^k (1-L)^{4m} \left(\frac{t+2m}{n} \right)^j \mathbf{1}\{t+2m < \tau_0\} \\
= &\sum_{k>k_n=\lfloor n(r_0-r)/4 \rfloor}^{\infty} \frac{(m)_k (-1)^k}{k!} \left[L^{-2} (1-L)^4 + \frac{1}{\lambda} \right]^k (1-L)^{4m} \left(\frac{t+2m}{n} \right)^j \mathbf{1}\{t+2m < \tau_0\} \\
= &\sum_{k>k_n=\lfloor n(r_0-r)/4 \rfloor}^{\infty} \frac{(m)_k (-1)^k}{k!} (1-L)^{4m} \sum_{i=k_n+1}^{k} \binom{k}{i} \frac{1}{\lambda^{k-i}} (1-L)^{4i} \left(\frac{t+2[m+i]}{n} \right)^j \mathbf{1}\{t+2[m+i] < \tau_0\} \\
= &\sum_{k>k_n=\lfloor n(r_0-r)/4 \rfloor}^{\infty} \frac{(m)_k (-1)^k}{k!} \sum_{i=k_n+1}^{k} \binom{k}{i} \frac{1}{\lambda^{k-i}} (1-L)^{4m} \left[\sum_{p=P_{(t)}}^{4i} \binom{4i}{p} (-1)^k \left(\frac{t+2[m+i]-p}{n} \right)^j \right] \\
= &\ 0,
\end{aligned}$$

because $(1-L)^{4m} (t+2[m+i]-p)^j = 0$ for all $j \leq J$ since $m \to \infty$. This proves that on the lower segment $t = \lfloor nr \rfloor < \tau_0 = \lfloor nr_0 \rfloor$ with $r < r_0$ we have

$$(1-G_\lambda)^m \left(\frac{t}{n} \right)^j \mathbf{1}\{t < \tau_0\} \to 0, \quad \text{as } m, n \to \infty.$$

(ii) Upper Segment

Treatment of the segment $t > \tau_0$ is similarly complicated because application of powers of the differencing operator $(1-L)$ to the segmented polynomial $\left(\frac{t}{n}\right)^j \mathbf{1}\{t \geq \tau_0\}$ involves computations on either side of the break point that occurs at $t = \tau_0$. Proceeding as in (29) we have

$$\begin{aligned}
&(1-G_\lambda)^m \left(\frac{t}{n} \right)^j = \left(\frac{L^{-2}(1-L)^4}{L^{-2}(1-L)^4 + \frac{1}{\lambda}} \right)^m \left(\frac{t}{n} \right)^j \mathbf{1}\{t > \tau_0\} \\
= &\ \frac{1}{\Gamma(m)} \int_0^\infty s^{m-1} e^{-s} e^{-s\left[L^{-2}(1-L)^4 + \frac{1}{\lambda}\right]} ds\, (1-L)^{4m} \left(\frac{t+2m}{n} \right)^j \mathbf{1}\{t+2m > \tau_0\} \\
= &\ \frac{1}{\Gamma(m)} \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \int_0^\infty s^{m+k-1} e^{-s} ds \left[L^{-2}(1-L)^4 + \frac{1}{\lambda} \right]^k (1-L)^{4m} \left(\frac{t+2m}{n} \right)^j \mathbf{1}\{t+2m > \tau_0\} \\
= &\ \sum_{k=0}^{\infty} \frac{(m)_k (-1)^k}{k!} (1-L)^{4m} \left[L^{-2}(1-L)^4 + \frac{1}{\lambda} \right]^k \left(\frac{t+2m}{n} \right)^j \mathbf{1}\{t+2m > \tau_0\}. \qquad (34)
\end{aligned}$$

Next observe that if $4m > J+1$

$$\begin{aligned}
&(1-L)^{4m} \left[L^{-2}(1-L)^4 + \frac{1}{\lambda} \right]^k \left(\frac{t+2m}{n} \right)^j \mathbf{1}\{t+2m > \tau_0\} \\
= &\ \sum_{i=0}^{k} \binom{k}{i} \frac{1}{\lambda^{k-i}} L^{-2i} (1-L)^{4[m+i]} \left(\frac{t+2m}{n} \right)^j \mathbf{1}\{t+2m > \tau_0\} \\
= &\ \sum_{i=0}^{k} \binom{k}{i} \frac{1}{\lambda^{k-i}} (1-L)^{4[m+i]} \left(\frac{t+2[m+i]}{n} \right)^j \mathbf{1}\{t+2[m+i] > \tau_0\} \\
= &\ 0, \qquad (35)
\end{aligned}$$

for all $j \leq J$ and for all $k \leq k_n^\varepsilon = \lfloor n\left(\frac{r-r_0-\varepsilon}{4}\right)\rfloor$ for some small $\varepsilon > 0$ such that $r > r_0 + \varepsilon$. The final line (35) follows because for large enough n and with $m = o(n)$ we have

$$t + 2[m+i] - 4[m+i] \geq \lfloor nr \rfloor + 2m - 4[m + k_n^\varepsilon] = \lfloor nr \rfloor - 2m - 4\left\lfloor n\left(\frac{r-r_0-\varepsilon}{4}\right)\right\rfloor > \lfloor nr_0 \rfloor = \tau_0, \tag{36}$$

so that $\mathbf{1}\{t + 2[m+i] - 4[m+i] > \tau_0\} = 1$ for all $i \leq k \leq k_n^\varepsilon$. It then follows that

$$(1-L)^{4[m+i]}\left(\frac{t+2[m+i]}{n}\right)^j \mathbf{1}\{t + 2[m+i] > \tau_0\} = (1-L)^{4[m+i]}\left(\frac{t+2[m+i]}{n}\right)^j = 0, \tag{37}$$

when $4m \geq J + 1$. We deduce that

$$\begin{aligned} & \sum_{k=0}^{\infty} \frac{(m)_k (-1)^k}{k!} (1-L)^{4m}\left[L^{-2}(1-L)^4 + \frac{1}{\lambda}\right]^k \left(\frac{t+2m}{n}\right)^j \mathbf{1}\{t + 2m > \tau_0\} \\ = & \sum_{k=k_n^\varepsilon}^{\infty} \frac{(m)_k (-1)^k}{k!} (1-L)^{4m}\left[L^{-2}(1-L)^4 + \frac{1}{\lambda}\right]^k \left(\frac{t+2m}{n}\right)^j \mathbf{1}\{t + 2m > \tau_0\}. \end{aligned} \tag{38}$$

Next, for $k > k_n^\varepsilon = \lfloor n\left(\frac{r_0-r-\varepsilon}{4}\right)\rfloor$ and in view of (37), we have

$$\begin{aligned} & \left[L^{-2}(1-L)^4 + \frac{1}{\lambda}\right]^k (1-L)^{4m}\left(\frac{t+2m}{n}\right)^j \mathbf{1}\{t + 2m > \tau_0\} \\ = & \sum_{i=0}^{k} \binom{k}{i} \frac{1}{\lambda^{k-i}} (1-L)^{4[m+i]}\left(\frac{t+2[m+i]}{n}\right)^j \mathbf{1}\{t + 2[m+i] > \tau_0\} \\ = & \sum_{i=k_n^\varepsilon+1}^{k} \binom{k}{i} \frac{1}{\lambda^{k-i}} (1-L)^{4[m+i]}\left(\frac{t+2[m+i]}{n}\right)^j \mathbf{1}\{t + 2[m+i] > \tau_0\}. \end{aligned} \tag{39}$$

When $i \geq i_n^\varepsilon := \lfloor n\left(\frac{r_0-r+\varepsilon}{2}\right)\rfloor$

$$t + 2[m+i] \geq \lfloor nr \rfloor + 2\left[m + \left\lfloor n\left(\frac{r_0-r+\varepsilon}{2}\right)\right\rfloor\right] > \lfloor nr \rfloor + \lfloor n(r_0 - r)\rfloor + m > \lfloor nr_0 \rfloor = \tau_0,$$

which implies that $\mathbf{1}\{t + 2[m+i] > \tau_0\} = 1$. For large k and i in (39) application of the operator $(1-L)^{4[m+i]}$ involves powers of the differencing operator $(1-L)$ that lead to adjacent integer values for which $\{t + 2[m+i] - (p-1) > \tau_0\}$ and $\{t + 2[m+i] - p < \tau_0\}$ for some positive integer p. In

this event,

$$\begin{aligned}
& (1-L)^{4i}\left[\left(\frac{t+2[m+i]}{n}\right)^{j}\mathbf{1}\{t+2[m+i]>\tau_0\}\right] \\
= & \sum_{p=0}^{4i}\binom{4i}{p}(-1)^{p}L^{p}\left[\left(\frac{t+2[m+i]}{n}\right)^{j}\mathbf{1}\{t+2[m+i]>\tau_0\}\right] \\
= & \sum_{p=0}^{4i}\binom{4i}{p}(-1)^{p}\left(\frac{t+2[m+i]-p}{n}\right)^{j}\mathbf{1}\{t+2[m+i]-p>\tau_0\} \\
= & \sum_{p=0}^{P_{(t)}}\binom{4i}{p}(-1)^{k}\left(\frac{t+2[m+i]-p}{n}\right)^{j},
\end{aligned}$$

where $P_{(t)}$ is the largest p for which $\{t+2[m+i]-p>\tau_0\}$, in which case $\mathbf{1}\{t+2[m+i]-p<\tau_0\}=1$ for $p\leq P_{(t)}$ and $\mathbf{1}\{t+2[m+i]-p<\tau_0\}=0$ for $p>P_{(t)}$. We then deduce that for large enough n

$$\begin{aligned}
& \sum_{k=0}^{\infty}\frac{(m)_k(-1)^k}{k!}\left[L^{-2}(1-L)^4+\frac{1}{\lambda}\right]^{k}(1-L)^{4m}\left(\frac{t+2m}{n}\right)^{j}\mathbf{1}\{t+2m>\tau_0\} \\
= & \sum_{k>k_n^{\varepsilon}=\lfloor n(r_0-r)/4\rfloor}^{\infty}\frac{(m)_k(-1)^k}{k!}\left[L^{-2}(1-L)^4+\frac{1}{\lambda}\right]^{k}(1-L)^{4m}\left(\frac{t+2m}{n}\right)^{j}\mathbf{1}\{t+2m>\tau_0\} \\
= & \sum_{k>k_n^{\varepsilon}=\lfloor n(r_0-r)/4\rfloor}^{\infty}\frac{(m)_k(-1)^k}{k!}(1-L)^{4m}\sum_{i=k_n^{\varepsilon}+1}^{k}\binom{k}{i}\frac{1}{\lambda^{k-i}}(1-L)^{4i}\left(\frac{t+2[m+i]}{n}\right)^{j}\mathbf{1}\{t+2[m+i]>\tau_0\} \\
= & \sum_{k>k_n^{\varepsilon}=\lfloor n(r_0-r)/4\rfloor}^{\infty}\frac{(m)_k(-1)^k}{k!}\sum_{i=k_n^{\varepsilon}+1}^{k}\binom{k}{i}\frac{1}{\lambda^{k-i}}(1-L)^{4m}\left[\sum_{p=0}^{P_{(t)}}\binom{4i}{p}(-1)^{k}\left(\frac{t+2[m+i]-p}{n}\right)^{j}\right] \\
= & \ 0,
\end{aligned}$$

because $(1-L)^{4m}(t+2[m+i]-p)^{j}=0$ for all $j\leq J$ since $m\to\infty$. This proves that on the upper segment $t=\lfloor nr\rfloor>\tau_0=\lfloor nr_0\rfloor$ with $r>r_0$ we have

$$(1-G_\lambda)^{m}(t/n)^{j}\mathbf{1}\{t<\tau_0\}\to 0,\quad \text{as } m,n\to\infty.$$

By combining the results for both segments $t<\tau_0$ and $t>\tau_0$, it follows that the boosted filter eventually reproduces accurately the polynomial component $g_n(t)$ for all $t=\lfloor nr\rfloor$ with $r\neq r_0$. The stochastic trend component x_t^0 is treated in the same manner as the proof of Theorem 1, showing that as $m,n\to\infty$ and $\lambda=\mu n^4$ for any fixed $\mu>0$, the boosted filter accurately captures the stochastic trend. It then follows that

$$\frac{\widehat{f}^{(m)}_{\lfloor nr\rfloor}}{\sqrt{n}}\rightsquigarrow g(r)+B(r),\ \text{for all } r\neq r_0,$$

when $\frac{1}{m}+\frac{m}{n}\to 0$ as $n\to\infty$, giving the required result for all $r\neq r_0$.

(iii) Limit theory at the Break Point

From the above analysis, we know that as $m, n \to \infty$ with $\frac{m}{n} \to 0$

$$n^{-1/2} \widehat{f}^{(m)}_{\lfloor nr \rfloor} \rightsquigarrow B_g(r) = g(r) + B(r), \text{ for all } r \neq r_0. \tag{40}$$

In particular,

$$[1 - (1 - G_\lambda)^m] g_n(t) \mathbf{1}\{t = \lfloor nr \rfloor < \lfloor nr_0 \rfloor\} \to g(r) \quad \text{for all } r < r_0,$$

and

$$[1 - (1 - G_\lambda)^m] g_n(t) \mathbf{1}\{t = \lfloor nr \rfloor > \lfloor nr_0 \rfloor\} \to g(r) \quad \text{for all } r > r_0.$$

The limit function $g(r)$ of $g_n(t)$ as $n \to \infty$ has the form

$$g(r) = \begin{cases} \alpha^0 + \beta^0_1 r + ... + \beta^0_J r^J & r < r_0 \\ \alpha^1 + \beta^1_1 r + ... + \beta^1_J r^J & r \geq r_0 \end{cases}$$

so that the limit function B_g in (40) satisfies $\lim_{r \nearrow r_0^-} B_g(r) = B_g\left(r_0^-\right) = g\left(r_0^-\right) + B(r_0)$ because $B(r)$ is continuous and $g(r)$ is continuous for all $r < r_0$ and has finite left limit $g\left(r_0^-\right)$. Similarly, $\lim_{r \searrow r_0^+} B_g(r) = B_g\left(r_0^+\right) = g(r_0) + B(r_0)$ because $g(r)$ is continuous for all $r > r_0$ and has finite right limit $g\left(r_0^+\right) = g(r_0)$. At the break point $\tau_0 = \lfloor nr_0 \rfloor$ itself, we can write the indicator $\mathbf{1}\{t = \tau_0\}$ as the simple symmetric average of the limits of the left and right side indicators

$$\mathbf{1}\{t = \lfloor nr \rfloor = \lfloor nr_0 \rfloor\} = \frac{1}{2}\left[\lim_{r_* \nearrow r_0} \mathbf{1}\{t = \lfloor nr \rfloor \leq \lfloor nr_* \rfloor\} + \lim_{r^* \searrow r_0} \mathbf{1}\{t = \lfloor nr \rfloor \geq \lfloor nr^* \rfloor\}\right].$$

So by virtue of the continuity of the limit function $g(r)$ on the right and left sides of $r = r_0$, the existence of the right and left side limits of $g(r)$ at r_0, and the asymptotic symmetry[15] of the action of the filter about the interior point $\tau_0 = \lfloor nr_0 \rfloor$ when $r_0 \in (0, 1)$ we deduce that

$$n^{-1/2} \widehat{f}^{(m)}_{\lfloor nr_0 \rfloor} \rightsquigarrow B_g(r_0) = \frac{1}{2}\left[g\left(r_0^-\right) + g\left(r_0^+\right)\right] + B(r_0),$$

giving the stated result.

□

[15]The asymptotic symmetry of the filter follows from the asymptotic symmetric Toeplitz representation of the filter given in (26). Importantly, this asymptotic symmetry holds away from the end points and is valid therefore for an interior break point $\tau_0 = \lfloor nr_0 \rfloor$ when $r_0 \in (0, 1)$.