BUSINESS CYCLES, TREND ELIMINATION, AND THE HP FILTER

By

Peter C. B. Phillips and Sainan Jin

June 2015

COWLES FOUNDATION DISCUSSION PAPER NO. 2005

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

http://cowles.econ.yale.edu/

Business Cycles, Trend Elimination, and the HP Filter*

Peter C.B. Phillips†
Yale University, University of Auckland
University of Southampton and Singapore Management University

Sainan Jin
Singapore Management University

June, 2015

Abstract

We analyze trend elimination methods and business cycle estimation by data filtering of the type introduced by Whittaker (1923) and popularized in economics in a particular form by Hodrick and Prescott (1980/1997; HP). A limit theory is developed for the HP filter for various classes of stochastic trend, trend break, and trend stationary data. Properties of the filtered series are shown to depend closely on the choice of the smoothing parameter (λ). For instance, when $\lambda = O(n^4)$ where n is the sample size, and the HP filter is applied to an $I(1)$ process, the filter does not remove the stochastic trend in the limit as $n \to \infty$. Instead, the filter produces a smoothed Gaussian limit process that is differentiable to the 4'th order. The residual 'cyclical' process has the random wandering non-differentiable characteristics of Brownian motion, thereby explaining the frequently observed 'spurious cycle' effect of the HP filter. On the other hand, when $\lambda = o(n)$, the filter reproduces the limit Brownian motion and eliminates the stochastic trend giving a zero 'cyclical' process. Simulations reveal that the $\lambda = O(n^4)$ limit theory provides a good approximation to the actual HP filter for sample sizes common in practical work. When it is used as a trend removal device, the HP filter therefore typically fails to eliminate stochastic trends, contrary to what is now standard belief in applied macroeconomics. The findings are related to recent public debates about the long run effects of the global financial crisis.

Keywords: Detrending, Graduation, Hodrick Prescott filter, Integrated process, Limit theory, Smoothing, Trend break, Whittaker filter.

JEL Classification Number: C32 Time Series Models.

---

*An early draft entitled "Limit Behavior of the Whittaker (HP) Filter" that was circulated in 2002 contained some asymptotics related to the present work but was not completed. The recent Bullard - Krugman (2012) public debate about use of the HP filter to estimate the impact of the GFC on potential output provided an incentive to complete the limit theory, widen its original scope, and provide an econometric perspective on the empirical trend estimates discussed in that debate.

†Corresponding author email: peter.phillips@yale.edu. Phillips thanks the NSF for support under Grant No. SES 12 -85258.

"RBC models can exhibit business cycle dynamics in HP filtered data even if they do not generate business cycle dynamics in pre-filtered data. The combination of a unit root or near unit root in technology and the HP filter is sufficient to generate business cycle dynamics" Cogley and Nasan (1995).

1 Introduction

Whittaker (1923) suggested a method of graduating data (now commonly called data smoothing) that was designed to remove the effects of measurement error and reveal the underlying trend in the data. This work on graduation was preceded by earlier actuarial research, including many studies by DeForrest (1873, 1874, 1876) on interpolative methods based on probabilistic principles.[1] The Whittaker method involved taking a least squares best fit to the data subject to a penalty involving the squared higher order differences of the data. The procedure has many variants depending on the form of the penalty. These were discussed in the book by Whittaker and Robinson (1924) and were studied subsequently by many authors (e.g. Greville, 1957). Whittaker and Robinson (1924) provided a formal justification for their smoothing procedure using Bayesian principles to motivate the penalized least squares procedure. That justification underlies much subsequent work, including the use of smoothness priors in econometrics (Shiller, 1973, 1984) and the spline smoothing methods suggested in Wahba (1978).

The literature is now extensive. Aitken (1925) wrote his doctoral thesis on the subject and provided the first systematic investigation of general numerical procedures. Numerical algorithms for graduating data by these techniques have been used in actuarial work dating back at least to Henderson (1924, 1925, 1938). Camp (1950) gave an overview of the Whittaker-Henderson graduation processes. Schoenberg (1964), Reinsch (1967), and Boneva et.al (1970) developed algorithms from the standpoint of spline fitting, prior to the formal Bayesian approach to spline smoothing that was used in Wahba (1978) which in turn closely echoed the original justification given by Whittaker and Robinson (1924). Recently, Kim et al (2009) discussed some related trend capture methods using modern penalized ℓ_1 estimation where sums of absolute values replace sums of squares in penalizing variations from trend. Diewert and Wales (2006) considered smoothing algorithms based on prior ideas, going back to Sprague (1887), that the second differences of smooth series do not change sign too frequently.

In economics the approach was systematically used by Hodrick and Prescott (1980, 1997), where second order differences were used in the penalty. This version of the method, promoted by Leser (1961) for the purpose of trend construction in economic data, has subsequently become known as the HP filter in economics. In view of the

[1] DeForrest wrote extensively on methods of interpolating and adjusting series using probabilistic principles to equalize the probability of error in the adjusted series. DeForrest cites earlier work by Everest and by Schiaparelli, which he showed may be deduced as special cases of his own methods. Stigler (1978) provides a modern statistical overview of some aspects of DeForrest's work on statistical methods of interpolation.

origins of this approach in the work of Whittaker and the Bayesian probabilistic justification for this smoothing technology given in the treatise by Whittaker and Robinson, we shall use the terminology Whittaker filter in what follows for the general form of this filter. For the last several decades, the HP filter has been used extensively in applied econometric work to detrend data, particularly to assist in the measurement of business cycles. Fig. 1 gives examples of several macroeconomic time series, including real quarterly GDP, real personal consumption expenditures, and industrial production in the US, that are frequently HP filtered in empirical work (e.g. Hansen, 1985; Backus and Kehoe, 1992; Stock and Watson, 1999; Phillips and Sul, 2007). The smoothing parameter used in calculating the HP filter in these illustrations is $\lambda = 1600$, the conventional setting for quarterly data suggested by Hodrick and Prescott (1997).

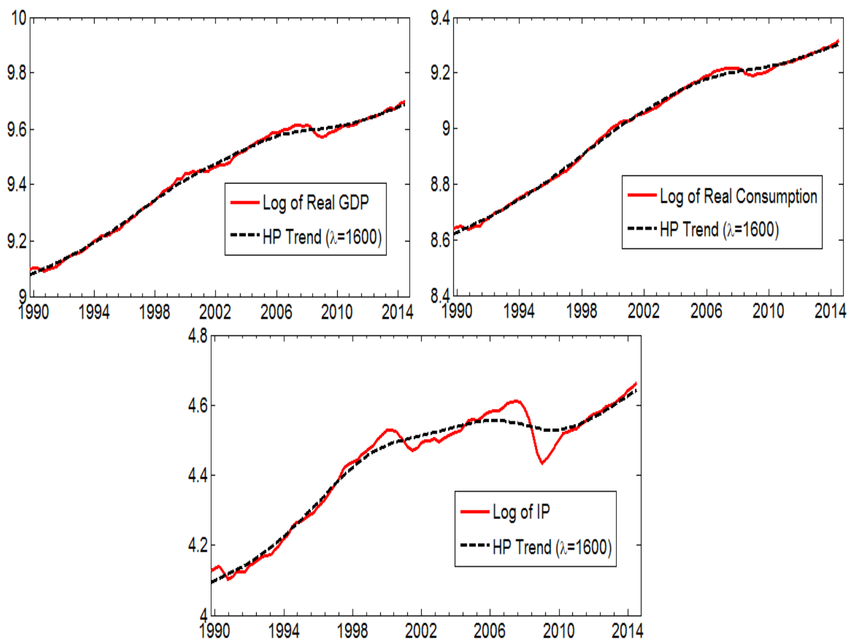
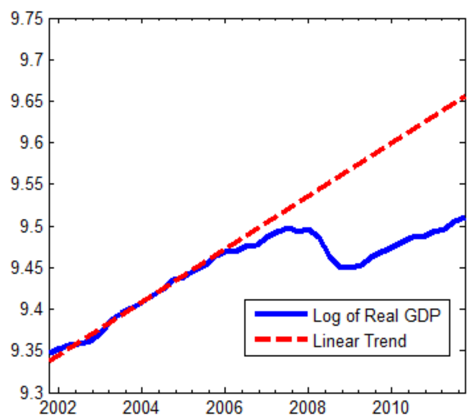


Fig. 1: Time Series plots of quarterly data (1990:I to 2014:IV) for 3 US macro variables accompanied by their HP trends: real GDP, real consumption, and industrial production. All series are seasonally adjusted and in logarithms. Data source: FRED.

Like other trend removal techniques such as trend regression, moving average detrending, and band-pass filtering, the HP filter is often used to produce new time series such as potential GDP and the output gap that are useful in macroeconomic modeling and monetary policy research. This practice has generated enormous discussion in the literature as well as some recent public debate involving James Bullard, President of the St. Louis Federal Reserve, and the economist and New York Times columnist Paul Krugman. Comparing two methods of decomposing US real GDP over 2002:1 to 2012:1, Bullard (2012) argued that detrending via linear time trend regression produces a "large output gap" view of the economy in 2012 because of the large gap between trend and actual GDP. To the extent that trend GDP represented by the trend regression line actually represents potential output, he contends that this view suggests that the housing bubble and ensuing financial crisis "did no lasting damage to the economy" as trend output was largely unaffected and deviations from trend were business cycle effects associated with the great recession. On the other hand, detrending by the HP filter gives a very different picture of the economy, in which Bullard indicates that "relatively slow GDP growth" should be expected following a housing bubble that "probably did some lasting damage to the US economy." Fig. 2 reproduces these two graphics based on our own calculations. The smoothing parameter used for the HP filter is again $\lambda = 1600$. The linear trend is fitted by least squares regression on the data up to 2006:1 combined with a trend projection of that line over the remaining period to 2012:1.



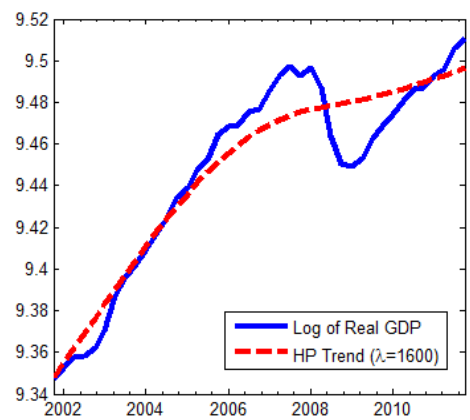


Fig. 2: Did the financial crisis do lasting damage to the US economy? Discovering potential output by decomposing real GDP. The figures shown here reproduce those given in James Bullard (2012) based on our own calculations.

Krugman (2012) challenged the use of the HP filter as a measure of potential output, arguing that "the use of the HP filter presumes that deviations from potential output are relatively short-term, and tend to be corrected fairly quickly," but "... any protracted slump gets interpreted as a decline in potential output". Krugman illustrated the argument with a chart for real US annual GDP over 1919-1939 together with its HP trend, which we reproduce from our own calculations in Fig. 3. The value of the smoothing parameter Krugman used in creating his chart is not stated. But it must be small because the result is a curve that reproduces closely the fine grain shape of the interwar data. (Our calculations in Fig. 3 are based on the choices $\lambda = 1, 2, 6.25, 300$, the setting $\lambda = 6.25$ being the value recommended for annual data by Ravn and Uhlig, 2002.) As we demonstrate below, depending on the value of λ that is chosen (in relation to the sample size), we may expect the detrended (cyclical component of the) series to retain some of the stochastic trend or random wandering characteristics that may be present in the original series. On the other hand, if we use a very large value of λ in relation to the sample size, the fitted trend is now much smoother and the fluctuations about trend appear in the cyclical component as a business cycle effect after trend removal. Quantification of the order of magnitude of λ in relation to the sample size therefore turns out to be of great importance in the interpretation of the results from empirical use of the HP filter.

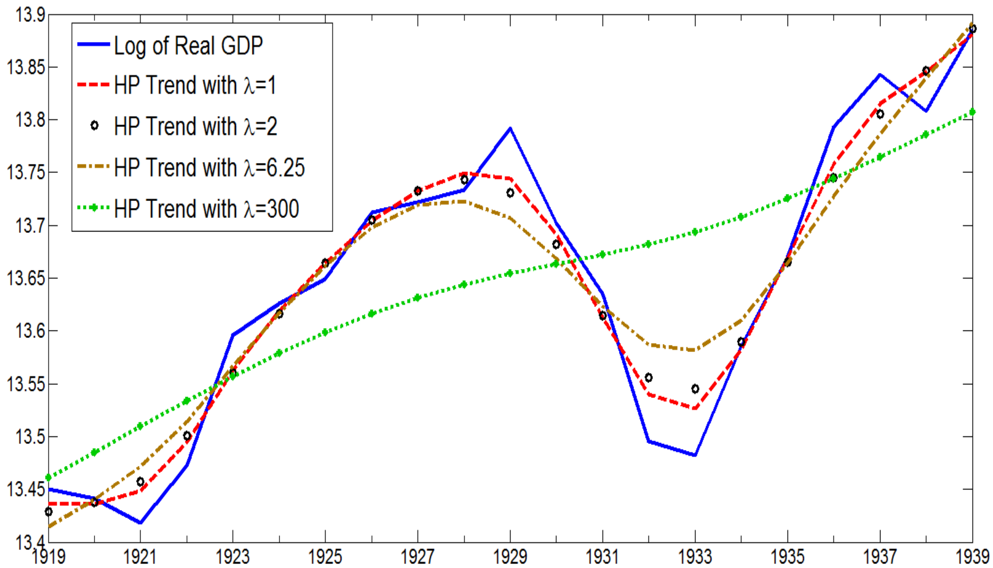


Fig. 3: US GDP and its HP trend during 1919-1939. The graphs shown here reproduce Krugman's (2012) figure based on our own calculations for various choices of λ.

Krugman's view has merit. The HP filter, as Whittaker originally developed it, is a data-smoothing, graduating device. The filter is two-sided, not causal or predictive, and averages data ahead and before each data point.[2] It produces a new series that may be interpreted as a trend only to the extent that it shows the general course of the observed data after graduating out fluctuations. The extent to which such graduation occurs depends on the choice of the smoothing parameter used in the penalty function that penalizes roughness in the observed series. As indicated above, Whittaker and Robinson (1924, pp. 303-306) gave a formal inductive probabilistic argument to justify the precise form of the filter.[3] Their argument relies on finding the 'most probable'[4] values based on a Bayesian principle in which there are prior grounds for believing that these 'most probable values' form a smooth sequence or trend. Such a filtered or smoothed series can therefore provide no guidance about an economy's potential output in the absence of further information and assumptions that describe the smoothness properties of potential output itself and relate these

[2] The filter is necessarily one-sided at the terminal point of the sample and, correspondingly, a version may be produced as a one-sided filter by recursive calculation through the sample observations.

[3] This particular contribution of Whittaker and Robinson (1924) amounts to a modern Bayesian smoothness prior development of the filter. See footnote 8 below for further details.

[4] "The problem is to combine all the materials of judgment – the observed values and the a priori considerations - to obtain the most probable values" (Whittaker and Robinson, 1924).

properties to the design of the filter. Obviously, the HP filter in its usual form cannot do so, because it contains no economic ideas about the nature of potential output in terms of the utilization of an economy's resources, which themselves change over time, whereas the choice parameter λ is fixed.[5],[6]

There is a further issue that complicates the comparison of the two detrending methods shown in Fig. 2. The linear trend regression extrapolates using data up to 2006:1, before the advent of the GFC, from which inference is drawn concerning the extent of the subsequent output gap. On the other hand, the trend function obtained by the HP filter uses all of the data to 2012:1, so that the fitted trend function absorbs the impact of the GFC and the Great Recession, much as is argued in Krugman's comment concerning the use of the HP filter on US real GDP data during the Great Depression. To obtain a fair comparison of the implications of these two alternate methods, it is necessary to use a predictive version of the trend implied by the HP filter based on the same data that is used in the linear trend regression. Such a predictive version of the HP filter can be obtained by using a combination of the HP filter and ARMA forecasting techniques that deliver the required forecasted values that are needed in the HP algorithm. This approach is sometimes used for end-point corrections to the filter and clearly involves mixing two methodologies (e.g. see Duy, 2012, in his comment on Bullard, 2012). Another approach is to use the methods developed in the current paper to obtain a predictive version of the HP filter itself. The theory and numerical algorithm needed to accomplish this predictive version of the HP filter are outside the scope of the present work and will be provided in a later paper.

Because of the commonly given solution form of the filter (see (14) below) it is almost universally assumed that the filter removes unit root nonstationarity in integrated processes up to the 4'th order, following the discussion in King and Rebelo (1993). Nevertheless, practical empirical work and analyses with simulation data (e.g., Cogley and Nasan, 1995) reveal that series often test as having long memory or even a unit root after HP detrending, just as we find in the above illustrations given in Figs. 1 and 2 when we estimate the output gap using a large value of the smoothing parameter λ. To wit, the cyclical residual process of US GDP shown in Fig. 1 after HP filtering tests as unit root nonstationary (with a drift) by a standard Dickey-Fuller (DF) unit root t test with no extra lags giving a p value of 0.20. A similar outcome is obtained with a Phillips-Perron (2008; PP) test. These outcomes are sensitive to the presence of transient dynamics and drift terms in the regression, indicating some fragility in the test results. Table 1 provides unit root and long memory test results for various settings that reveal this fragility.

The final three columns of Table 1 report estimates of the memory parameter

[5] As pointed out below, Leser (1961) and HP (1980) advanced arguments that use of the HP smoother seemed appropriate in terms of long run linear trend behavior for economic aggregates (e.g. for log GDP).

[6] Readers interested in reading more about this public debate on the merits and limitations of the HP filter may consult Tim Duy's (2012) Fed Watch: "Careful with that HP Filter" and the references therein. See also Krugman's (2013) follow-up article on attempts to measure potential output by other methods with specific reference to the debate on European recovery from the recession.

Table 1: Testing HP filtered series for nonstationarity

$\lambda = 1600$	Unit root t tests (ADF and PP)				Exact local Whittle estimates of d		
	p values of ADF tests				$m = n^{0.6}$		
Fig. 1	ADF(0)	ADF(1)	ADFD(0)	ADFD(1)	$\hat{d}$	se	CI
real GDP	0.0052	0.0000	0.2033	0.0389	0.6308	0.1291	[0.3778, 0.8839]
real Cons	0.0097	0.0032	0.3106	0.1239	0.8103	0.1291	[0.5573, 1.0633]
IP	0.0291	0.0000	0.5152	0.0000	0.6670	0.1291	[0.4140, 0.9200]
Fig. 2	0.1819	0.0091	0.8931	0.3148			
real GDP					1.5608	0.1508	[1.2654, 1.8563]
	p values of PP tests				$m = n^{0.7}$		
Fig. 1	ADF(0)	ADF(1)	ADFD(0)	ADFD(1)	$\hat{d}$	se	CI
real GDP	0.0052	0.0031	0.2033	0.1275	1.0913	0.1000	[0.8953, 1.2873]
real Cons	0.0097	0.0054	0.3106	0.2123	1.2370	0.1000	[1.0410, 1.4330]
IP	0.0291	0.0075	0.5152	0.2558	1.3477	0.1000	[1.1517, 1.5437]
Fig. 2							
real GDP	0.1819	0.0919	0.8931	0.7706	1.7605	0.1250	[1.5155, 2.0055]

Notes:
(i) ADF(k) denotes ADF regression with k lagged differences;
(ii) ADFD(k) denotes ADF regression with k lagged differences and a drift.

(d) of the HP detrended GDP time series. The long memory parameter is estimated at $\hat{d} = 1.0913$ with a standard errror of 0.100 using the exact local Whittle (ELW) procedure (Shimotsu and Phillips, 2005) with bandwidth $m = n^{0.7}$. ELW delivers a consistent semiparametric estimator of the long memory parameter of a time series (irrespective of the true value of d) and has an asymptotic $N(d, \frac{1}{4m})$ distribution for all values of d, which enables uniform confidence interval construction. A 95% confidence interval for d is $(0.8953, 1.2873)$ for $m = n^{0.7}$. From these estimates, the hypothesis that there is a unit root ($d = 1$) in the GDP cyclical residual process cannot be rejected. A wide range of other values of the memory parameter are included in the confidence interval but all of these lie in the nonstationary range ($d \geq \frac{1}{2}$). When the bandwidth is set to the lower value $m = n^{0.6}$ (which satisfies Assumption 4 in Shimotsu and Phillips, 2005), the long memory parameter estimate is $\hat{d} = 0.6308$, with a standard errror of 0.129 and 95% confidence interval $(0.3778, 0.8839)$, indicating that the null hypothesis of nonstationary behavior (i.e., $d \geq \frac{1}{2}$) in the cyclical component of GDP cannot be rejected. In this case the confidence interval for the memory parameter includes values in the stationary region ($d < \frac{1}{2}$), consistent with the fragility results found in the unit root tests above. The same finding applies to the other macroeconomic series, including real personal consumption expenditures and industrial production, shown in Figs. 1 and 2.

It is well recognized that the form of the filtered series depends closely on the smoothing parameter (λ) that controls the size of the penalty in the objective funtion. For most macroeconomic applications, this parameter is chosen to be large. For instance, with quarterly data where sample sizes are usually in the region 100 to 300, a typical choice is $\lambda = 1600$. In such applications, we may consider λ to be large relative to n. In such cases, the HP filter produces a smooth series that is usually taken to reflect the underlying trend in the data. For much smaller choices of λ, the filtered series are choppier and follow the original series more closely, as shown in the Fig. 3 illustration. Of course, when $\lambda = 0$, the filtered series is identical to the original series because there is no penalty. When $\lambda \to \infty$ for fixed n, the filter selects a linear trend. When higher order differences are used in the penalty function, the filter selects a higher order time polynomial as the trend when $\lambda \to \infty$, as shown in Phillips (2010).

The intimate dependence of the HP filter on the smoothing parameter λ in relation to the sample size indicates that different forms of limit behavior manifest through the filter as $n \to \infty$. Just as in nonparametric density estimation and regression, we may expect the choice of the tuning parameter to affect the limit theory and limit features such as the asymptotic mean squared error. However, in the case of the HP filter the relationship is largely unexplored and, to the best of our knowledge, there has been no previous study of the asymptotic properties of either the HP or Whittaker filters. This paper therefore seeks to develop a limit theory for these filters that allows for $\lambda \to \infty$ at various rates as $n \to \infty$. We look at stochastic trend, trend break and trend stationary data generating processes. The results show how the limit properties of the filters depend closely on the relative size of λ and n. For macroeconomic applications, our results indicate that the choice $\lambda = O(n^4)$ provides a good approximation to the form of the HP filter as it is presently used in much practical work in economics. The limit function in this case turns out to be a Gaussian stochastic process that is continuously differentiable to the 4'th order when the true limit process (after suitable standardization of the time series) is a Brownian motion with drift. Similar results are shown to apply in the case of data with limit processes that involve Brownian motion with piecewise continuous drift functions. With these conventional settings, therefore, the filter does not remove a stochastic trend but only a smoothed version of a stochastic trend. In effect, and contrary to popular belief in applied macroeconomics, the HP filter does not typically eliminate a time series unit root. This analysis explains the 'spurious cycle' findings in simulation work on the effects of the HP filter, such as that noticed by Cogley and Nasan (1995) in their analysis of artificial data generated by real business cycle models with filtered and unfiltered data.

The plan of the remaining paper is as follows. Section 2 gives the filters and general solution formulae and provides some preliminary analysis and heuristics. Section 3 provides a rigorous development of the limit theory in the leading case where the expansion rates are $\lambda = O(n^4)$ for the HP filter and $\lambda = O(n^{2m})$ for the general Whittaker filter. Section 4 considers similar cases where there are general deterministic drifts and trend breaks in the time series. Section 5 develops asymptotics for faster and slower rates of expansion for λ. When $\lambda = O(n)$ the filters capture both deterministic and stochastic trends in the limit. Some simulations are reported in Section 6. Conclusions and implications are discussed in Section 7. Proofs are given in the Appendix.

2 The Filters and Solution Formulae

The general Whittaker filter decomposes time series data $(x_t : t = 1, .., n)$ into a smooth trend (f_t) and a residual cycle (c_t). The trend f_t is meant to capture the long run growth of x_t, while the residual c_t is often taken to represent a business cycle component (or output gap in the case of real GDP). Writing these components of x_t as

$$x_t = f_t + c_t, \tag{1}$$

the filter computes estimates of f_t and c_t by solving

$$\hat{f}_t = \arg\min_{f_t} \left\{ \sum_{t=1}^{n} (x_t - f_t)^2 + \lambda \sum_{t=m+1}^{n} (\triangle^m f_t)^2 \right\}, \quad \hat{c}_t = x_t - \hat{f}_t, \tag{2}$$

where $\lambda > 0$ is a smoothing parameter and $\triangle^m f_t$ is the m'th difference of f_t for some integer $m \geq 1$, with $\Delta = 1 - L$ and L the lag operator defined by $Lf_t = f_{t-1}$.

The first summation in (2) penalizes a poor fit and the second penalizes lack of smoothness. Whittaker (1923) developed this approach to graduating series using the setting $m = 3$. Aitken (1925, 1926) devised the first numerical algorithm using a Laurent series expansion of the solution function of (2), an advance that enabled practical implementation.[7] Whittaker and Robinson (1924, pp. 304-306) provided a rigorous Bayesian justification for the procedure that led to (2).[8] Their work appears to be the first instance of penalized estimation in statistical theory that is formally based on probabilistic principles, a fact that does not seem so far to have been acknowledged in that literature. This work all concentrated on the case where the penalty term involved squared third differences in the data $(m = 3)$, while acknowledging that more general cases were possible. In what follows, we will therefore reference the general case for arbitrary $m \geq 2$ as the Whittaker filter.

[7] This work, which was contained in Aitken's (1925) doctoral thesis at the University of Edinburgh, was considered so significant an advance, that Aitken was awarded a D.Sc degree in place of a Ph.D (University of Edinburgh Senate Minutes, "Tribute to A. C. Aitken", 19 January 1966.)

[8] Setting $m = 3$, they proposed maximizing the likelihood (or fidelity) of observing the actual observations x_t when the true values f_t were subject to a prior probability, guided by a principle of smoothness and given by the normal law $c_1 e^{-\lambda^2 S}$ for constants $c_1 > 0$ and $\lambda^2 > 0$ with $S = \sum_{t=4}^{n} (\Delta^3 f_t)^2$ a measure of 'roughness'. Fidelity was measured via $F = \sum_{t=1}^{n} h_t^2 (x_t - f_t)^2$, where $h_t > 0$ captured the precision of the t'th observation, and likelihood was represented in terms of the normal law $c_2 \prod_{t=1}^{n} h_t e^{-F}$ for some constant $c_2 > 0$. This principle led to the operational criterion

$$\hat{f}_t = \arg\max_{f_t} \left[c_1 c_2 \prod_{t=1}^{n} h_t e^{-\lambda^2 S - F} \right] = \arg\min_{f_t} \left[F + \lambda^2 S \right]$$

which reduces to

$$\hat{f}_t = \arg\min_{f_t} \left[\sum_{t=1}^{n} (x_t - f_t)^2 + \frac{\lambda^2}{h^2} \sum_{t=4}^{n} (\Delta^3 f_t)^2 \right]$$

in the constant precision case where $h_t = h$ for all t, thereby falling into the class given by (2).

Following early work by Macaulay (1931) and later Leser (1961), Hodrick and Prescott (1997) focussed on the case of second differences ($m = 2$). Leser argued that $m = 2$ was a natural choice for economic time series, giving a family of "quasi-linear trends" approximating the linear trend case for which $\triangle^2 f_t = 0$. Hodrick and Prescott (1997) similarly argued that when decomposing an economic times series like x_t into a growth component measured in logarithms (f_t) and cyclical component (c_t) it is often natural to expect a constant growth rate $\triangle f_t$ in the long run, which in turn implies a linear trend path for f_t. Neither argument is now compelling, particularly in view of the econometric evidence for the presence of stochastic trends and breaks in economic data; and neither addresses the concern evident in Krugman's critique that the prior underlying the usual penalty in the HP filter has little economic content concerning such matters as the smoothness properties of potential output. Nonetheless, the usage $m = 2$ in the filter is near universal in applied econometric work.

2.1 Algebraic and Operator Solutions

Setting $f' = (f_1, ..., f_n)$, $\hat{f} = \left(\hat{f}_1, \hat{f}_2, ..., \hat{f}_n\right)'$ and $x = (x_1, x_2, ..., x_n)'$, the criterion (2) has the matrix form

$$\hat{f} = \arg\min_f \left\{(x - f)'(x - f) + \lambda f' D_m D_m' f\right\}, \tag{3}$$

with solution

$$\hat{f} = \left(I + \lambda D_m D_m'\right)^{-1} x, \tag{4}$$

where D_m' is the rectangular $(n - m) \times n$ Toeplitz matrix

$$D_m' = \begin{bmatrix} d_m' \; 0 \; 0 & \cdots & 0 \; 0 \; 0 \\ 0 \; d_m' \; 0 & \cdots & 0 \; 0 \; 0 \\ 0 \; 0 \; d_m' & \cdots & 0 \; 0 \; 0 \\ \vdots & \ddots & \vdots \\ 0 \; 0 \; 0 & \cdots & d_m' \; 0 \; 0 \\ 0 \; 0 \; 0 & \cdots & 0 \; d_m' \; 0 \\ 0 \; 0 \; 0 & \cdots & 0 \; 0 \; d_m' \end{bmatrix} \tag{5}$$

and d_m' is the $m-$ differencing vector

$$d_m' = \left[\binom{m}{0}, (-1)\binom{m}{1}, ..., (-1)^{m-1}\binom{m}{m-1}, (-1)^m \binom{m}{m}\right].$$

Since $\sum_{j=0}^{m} \binom{m}{j} (-L)^j f_t = \Delta^m f_t$, applying the matrix operator D_m' to $f = (f_1, ..., f_n)$ gives

$$D_m' f = \left[\Delta^m f_{m+1}, \Delta^m f_{m+2}, ..., \Delta^m f_n\right]'. \tag{6}$$

In the HP case $d_2' = (1, -2, 1)$ and we have $D_2' f = \left[\Delta^2 f_3, \Delta^2 f_4, ..., \Delta^2 f_n\right]'$ with

$$D_2' = \begin{bmatrix} d_2' & 0 & \cdots & 0 \\ 0 & d_2' & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_2' \end{bmatrix}.$$

The solution of (3) may be written in several different forms which reveal certain properties of the filtered series $\hat{f}$ and aid in the analysis of its asymptotic behavior. The following result provides an explicit matrix solution that shows the polynomial trend component.

Theorem 1 *For given $n, m,$ and λ, the solution $\hat{f} = (I + \lambda D_m D_m')^{-1} x$ of (3) has the following algebraic form*

$$\hat{f} = R_m \left(R_m' R_m\right)^{-1} R_m' x + D_m \left(D_m' D_m\right)^{-1/2} \left\{I + \lambda D_m' D_m\right\}^{-1} \left(D_m' D_m\right)^{-1/2} D_m' x, \quad (7)$$

where R_m is the $n \times m$ polynomial time trend matrix

$$R_m = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 2 & \cdots & 2^{m-1} \\ 1 & 3 & \cdots & 3^{m-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & n & \cdots & n^{m-1} \end{bmatrix}. \quad (8)$$

Remark 1 Expression (7) decomposes $\hat{f}$ into two components. The first is a polynomial time trend of order $m-1$ whose parameters depend on the least squares regression coefficients $\left(R_m' R_m\right)^{-1} R_m' x$. This time trend is independent of λ so it is present for all values of $\lambda \neq 0$. The second component of (7) is a residual whose importance and magnitude depend critically on the smoothing parameter λ.

Remark 2 For large $\lambda \to \infty$ we can write (7) in series expansion form as follows

$$\begin{aligned} \hat{f} &= R_m \left(R_m' R_m\right)^{-1} R_m' x + \left[D_m \left(D_m' D_m\right)^{-1/2} \left\{I + \lambda D_m' D_m\right\}^{-1} \left(D_m' D_m\right)^{-1/2} D_m'\right] x \\ &= R_m \left(R_m' R_m\right)^{-1} R_m' x + \lambda^{-1} \left[D_m \left(D_m' D_m\right)^{-1} \left\{I + \left(\lambda D_m' D_m\right)^{-1}\right\}^{-1} \left(D_m' D_m\right)^{-1} D_m'\right] x \\ &= R_m \left(R_m' R_m\right)^{-1} R_m' x - \sum_{k=1}^{\infty} (-\lambda)^{-k} D_m \left(D_m' D_m\right)^{-k-1} D_m' x, \quad (9) \end{aligned}$$

the expansion holding for λ large enough to ensure that the latent roots of $\lambda D_m' D_m$ are outside the unit circle. We deduce that for fixed n

$$\hat{f} \to R_m \left(R_m' R_m\right)^{-1} R_m' x = R_m \gamma, \quad \text{as } \lambda \to \infty, \quad (10)$$

where $\gamma = (R'_m R_m)^{-1} R'_m x$ is the least squares regression coefficient of x_t on a polynomial time trend of degree $m-1$. Thus, the general solution $\hat{f}$ tends asymptotically to a trend polynomial of degree $m-1$ as $\lambda \to \infty$ (c.f. Phillips, 2010). In the HP case ($m = 2$), the limit of $\hat{f}$ is a simple linear trend, as is well known. If the data follow a linear trend exactly, then the HP filter reproduces the data since $D'_2 x = 0$ and the projector $P_{R_2} = R_2 (R'_2 R_2)^{-1} R'_2$ preserves x.

Remark 3 The cyclical component of the time series is estimated as the residual $\hat{c} = x - \hat{f}$. As shown in the Appendix, standard projection geometry gives the relationship $R_m (R'_m R_m)^{-1} R'_m = I_n - D_m (D'_m D_m)^{-1} D'_m$, and so

$$\hat{c} = D_m \left(D'_m D_m\right)^{-1} D'_m x - D_m \left(D'_m D_m\right)^{-1/2} \left\{I + \lambda D'_m D_m\right\}^{-1} \left(D'_m D_m\right)^{-1/2} D'_m x,$$

implying that the cyclical component of the Whittaker filter always removes a polynomial time trend of degree $m-1$ from the data.

When $\lambda = \lambda_n \to \infty$ as $n \to \infty$, the asymptotics are much more complex than (10) and depend on the magnitude of the successive terms $\lambda_n^{-k} D_m (D'_m D_m)^{-k-1} D'_m x$ in (9), which in turn depend on the properties of the $n \times n$ matrices $D_m (D'_m D_m)^{-k-1} D'_m$, the stochastic properties of the data x, and the expansion rate of $\lambda_n \to \infty$. These asymptotics are of great interest in practice and have implications for the interpretation of results obtained from HP filtered data, as will become more apparent in what follows.

Tuning parameter choices are well known to be important in nonparametric estimation, affecting bias, variance, and rates of convergence. Similar considerations apply in the present setting where the choice of λ inevitably delimits the performance characteristics of the nonparametric estimate $\hat{f}_t$. As shown below, it is instrumental in determining the capability of the filter to accurately capture trends in the data, particularly stochastic trends, as the sample size grows. The primary concern of the present work is to explore these limits and to examine the relationship between the tuning parameter and the limit form of the filtered series in cases where stochastic trends, deterministic trends, and trend breaks occur in the data.

To start the analysis it is convenient to examine the operator form of the solution of (3). The HP trend solution to (2) when $m = 2$ is frequently written in econometric work (e.g., King and Rebelo, 1993) using operator notation as

$$\hat{f}_t^{HP} = [\lambda L^{-2}(1-L)^4 + 1]^{-1} x_t, \tag{11}$$

where the filter induced by $[\lambda L^{-2}(1-L)^4 + 1]^{-1} = [1 + \lambda(1-L)^2 \left(1 - L^{-1}\right)^2]^{-1}$ is a two-sided moving average of the original time series, as already apparent from (4). The cyclical solution associated with this operator form of the fitted trend $\hat{f}_t^{HP}$ is

$$\hat{c}_t^{HP} = \frac{\lambda L^{-2}(1-L)^4}{\lambda L^{-2}(1-L)^4 + 1} x_t. \tag{12}$$

The numerator of (12) suggests that if x_t is $I(1)$ and satisfies

$$(1-L)x_t = u_t, \tag{13}$$

for some stationary process u_t, then

$$\hat{c}_t^{HP} = \frac{\lambda L^{-2}(1-L)^3}{[\lambda L^{-2}(1-L)^4+1]} u_t \tag{14}$$

so that the fitted cyclical component $\hat{c}_t$ appears to be a stationary process. A similar conclusion applies when x_t is $I(4)$, or $I(4)$ with an accompanying polynomial trend of degree at most four. Hence, as pointed out by King and Rebelo (1993) and as subsequently emphasized in much of the econometric literature, the HP filter apparently renders stationary any time series that is integrated up to 4'th order (or integrated with a 4'th order drift). As discussed below, although the given forms of (11) and (12) are based on an infinite sample, this apparently obvious conclusion that the filter removes unit roots is by no means robust and depends critically on the behavior of the smoothing parameter λ as $n \to \infty$

The solutions (11) and (12) are, in fact, asymptotic approximations because they do not take into account end corrections that manifest in the exact filter solution given by the matrix formula (4). The correct operator form of the solution is given in the next result, which mirrors (4) in operator notation. In what follows, O_ℓ denotes an $\ell \times \ell$ matrix of zeros, O denotes a zero matrix whose dimensions are clear from the context, and e_j denotes the j'th unit vector with unity in the j'th position and zeros elsewhere.

Theorem 2 *For given n, λ, and $m = 2$, the HP filter solution $\hat{f}^{HP}$ satisfies the operator equation $d(L)\hat{f}^{HP} = x$ with matrix operator*

$$\begin{aligned} d(L) &= \begin{bmatrix} d_a(L) & O & O \\ O & \left\{1+\lambda\Delta^2\Delta^{*2}\right\} I_{n-4} & O \\ O & O & d_b(L) \end{bmatrix} \\ &= \left(1+\lambda\Delta^2\Delta^{*2}\right) \mathrm{diag}\left[O_2, I_{n-4}, O_2\right] + EKE', \end{aligned} \tag{15}$$

where $\Delta^* = 1 - L^{-1}$ is the adjoint operator of $\Delta = 1 - L$, $E_a = [e_1, e_2]$, $E_b = [e_{n-1}, e_n]$, $E = [E_a, E_b] = [e_1, e_2, e_{n-1}, e_n]$, $K = diag\,[d_a(L), d_b(L)]$, $d_a(L) = diag\left[1+\lambda\Delta^{*2}, 1+\lambda\Delta^{*2}(-1+2\Delta)\right]$, $d_b(L) = diag\left[1+\lambda\Delta^2(-1+2\Delta^*), 1+\lambda\Delta^2\right]$, and e_j is the j'th unit vector with unity in the j'th position and zeros elsewhere. The operator $d(L)$ may also be written in the form

$$d(L) = \left(1+\lambda\Delta^2\Delta^{*2}\right) I_n + \lambda\Delta^2 EGE' = \left(1+\lambda\Delta^2\Delta^{*2}\right)\left\{I_n + \alpha_\lambda(L)\, EGE'\right\}, \tag{16}$$

where $G = diag\,[A(L), B(L)]$, $A(L) = diag\left[\left(2L^{-1}-1\right), -1\right]$, $B(L) = L^{-2} diag\,[-1, (2L-1)]$, and $\alpha_\lambda(L) = \frac{\lambda\Delta^2}{1+\lambda\Delta^2\Delta^{*2}}$.

Remark 4 End corrections to the filter are contained in the components EKE' and $\lambda\Delta^2 EGE'$ which have rank 4.

Remark 5 The kernel of both operators Δ^2 and Δ^{*2} is the span of the constant and linear trend functions $(1, t)$. The kernel of $\Delta^2\Delta^{*2}$ is the span of the polynomials

$\left(1, t, t^2, t^3\right)$, and so for $\lambda \neq 0$ the identity space of the operator $\left\{1 + \lambda\Delta^2\Delta^{*2}\right\}$ is the span of the polynomials $\left(1, t, t^2, t^3\right)$. Correspondingly, the identity space of the operator $d(L) = \left(1 + \lambda\Delta^2\Delta^{*2}\right) I_n + \lambda\Delta^2 EGE'$ is the span of $(1, t)$. Thus, when $\lambda \neq 0$, the HP filter preserves linear trends, as indicated above from the explicit matrix form of the filter given by (7) when $m = 2$. When $\lambda \to \infty$, the operator $d(L)$ is dominated by $\lambda\Delta^2\Delta^{*2}I_n + \lambda\Delta^2 EGE'$ whose kernel space is the span of $(1, t)$. So the HP filter solution as $\lambda \to \infty$ lies in the intersection of the kernel spaces of $\Delta^2\Delta^{*2}$ and Δ^2, i.e., the span of $(1, t)$.

Remark 6 As $n \to \infty$, the filter $d(L)$ in (16) is dominated by the lead component $\left(1 + \lambda\Delta^2\Delta^{*2}\right) I_n$. As shown in the Appendix, the inverse of the operator $d(L)$ has the explicit form

$$d(L)^{-1} = \left(1 + \lambda\Delta^2\Delta^{*2}\right)^{-1} \operatorname{diag}\left[O_2, I_{n-4}, O_2\right] + \left(1 + \lambda\Delta^2\Delta^{*2}\right)^{-1} \times \left(E_a \left[I_2 + \alpha_\lambda(L) A(L)\right]^{-1} E_a' + E_b \left[I_2 + \alpha_\lambda(L) B(L)\right]^{-1} E_b'\right) \tag{17}$$

whose second component has rank 4 and delivers end corrections to the filter, thereby affecting only the first two and last two entries of $\hat{f}^{HP}$. It follows that typical interior entries $\hat{f}^{HP}_{t=\lfloor nr \rfloor}$ of the solution $\hat{f}^{HP} = d(L)^{-1} x$ are correspondingly dominated as $n \to \infty$ by entries of the lead component $\left(1 + \lambda\Delta^2\Delta^{*2}\right)^{-1} x$, thereby justifying (11) asymptotically.

Remark 7 For the general case (3) with arbitrary $m \geq 2$, more complex calculations related to those leading to (16) show that the Whittaker filter $\hat{f}^W$ satisfies the operator equation $d_m(L) \hat{f}^W = x$, where

$$d_m(L) = (1 + \lambda\Delta^m\Delta^{*m}) \operatorname{diag}\left[O_m, I_{n-2m}, O_m\right] + E_m K_m E_m', \tag{18}$$

with $E_m = [E_{ma}, E_{mb}]$, $E_{ma} = [e_1, .., e_m]$, $E_{mb} = [e_{n-m,}, ..., e_n]$, and diagonal matrix $K_m = \operatorname{diag}\left[A_m(L), B_m(L)\right]$ in which

$$\begin{aligned} A_m(L) &= \operatorname{diag}\left[1 + \lambda(-1)^{-m}\Delta^{*m}, .., 1 + \lambda\left[\Delta^m - (-L)^m\right]\Delta^{*m}\right], \\ B_m(L) &= \operatorname{diag}\left[1 + \lambda\left[\Delta^{*m} - (-L)^{-m}\right]\Delta^m, .., 1 + \lambda\Delta^m\right]. \end{aligned}$$

The specific entries of the diagonal matrices $A_m(L)$ and $B_m(L)$ follow the combinatoric scheme given in the operator system (68) - (69) detailed in the Appendix. The matrices $E_m K_m E_m'$ have fixed rank $2m$ as $n \to \infty$. It follows from (18) that the filter $d_m(L)$ is dominated as $n \to \infty$ by the lead component involving the operator $(1 + \lambda\Delta^m\Delta^{*m})$. The kernel of the operator $\Delta^m\Delta^{*m}$ is the span of the polynomials $\left(1, t, ..., t^{2m-1}\right)$. On the other hand, the elements of the diagonal matrix K_m in (18) are polynomials of the form $1 + a(L)\Delta^m$ where $a(L)$ is a polynomial in L and L^{-1}, so the identity space of the operator K_m is the span of the polynomials $\left(1, t, ..., t^{m-1}\right)$. The Whittaker filter therefore preserves polynomial time trends of degree $m - 1$, as is again evident from (7).

As indicated earlier in the discussion of (14), it is commonly stated in the literature that the cyclical component $\hat{c}_t = x_t - \hat{f}_t$ obtained from the HP filter residual is a stationary process when x_t is a unit root process, which implies that the HP filter is effective in removing a stochastic trend in the data. However, this conclusion does not necessarily follow when the smoothing parameter λ is large. First, as is clear from Remark 2, when $\lambda \to \infty$ the filter only removes a linear (or, in the Whittaker case, a polynomial) trend, which amounts to detrending a unit root process not to the removal of the stochastic trend. It is therefore of considerable interest to determine how the properties of the filter and the induced cyclical component depend on the expansion rate of λ as $n \to \infty$. The specific rate for $\lambda \to \infty$ that ensures removal of a stochastic trend is a natural focus of interest. It is also of interest to learn what expansion rate for λ is required in order to ensure that the filter is capable of more than simple polynomial trend extraction. Finally, in the case of data that involve a stochastic trend and a drift or deterministic trend break, what properties do the filtered data have in the limit as $n \to \infty$? These questions are examined in the following section.

Some immediate heuristics are apparent from (12), which we may write in the form

$$\hat{c}_t = \frac{\Delta^2 \Delta^{*2}}{\Delta^2 \Delta^{*2} + \frac{1}{\lambda}} x_t = \frac{\Delta^2 \Delta^{*2}}{\Delta^2 \Delta^{*2} + o(1)} x_t, \quad \text{as } \lambda \to \infty. \tag{19}$$

The second part of (19) suggests that, if $\lambda \to \infty$ at some suitable rate as $n \to \infty$, the fitted cyclical component may inherit, asymptotically, features similar to those of the original series x_t including its trend or random wandering components. This heuristic reasoning about the asymptotic form of the filter questions the simple and apparently universally accepted conclusion that the HP filter removes stochastic trends and it implies that the apparent trend removal property of (12) may not hold up in large samples. It also corroborates findings from practical work which, as mentioned in the Introduction, often show evidence of stochastic trend persistence or long memory after HP trend removal. This evidence is often interpreted as a 'spurious cycle' outcome of HP smoothing on the residual process (Cogley and Nason, 1995; Cogley, 2006).[9] The limit theory given in the following section explains this spurious cycle in the HP residual, indicates the conditions under which it arises, and gives explicit asymptotic forms to the residual process for data that have stochastic trends and various forms of trend breaks.

The twin issues of whether the trend in the data is removed and whether the induced cycle is spurious are central to much empirical and policy work in economics. They have substantial import for economic management of the business cycle; and they influence the measurement of key economic quantities such as the output gap, as is clear from the Bullard Krugman policy debate over the impact of the global financial crisis on long run potential output of the US economy.

---

[9] Spurious cyclicality may arise with other methods of detrending, such as moving average detrending and the use of band pass filters, as discussed, for example, in Osborn (1995).

3 Limit Theory of the Filters

To develop an asymptotic theory for the filter, we examine the large sample behavior of the operators in (12) and (14). We also need to make precise assumptions about x_t and its limit behavior, so that the trend-capture capability of the filter can be assessed. This section makes rigorous some of the heuristic reasoning of the preceding section concerning the asymptotic behavior of the filter.

Our starting point is to assume that the data x_t have a stochastic trend and to consider the impact of the filter on such a process. Later, we examine cases where the data have deterministic as well as stochastic trend components and a piecewise continuous deterministic drift function is present in the limit process. Suppose that x_t is $I(1)$ as in (13) and u_t is such that a standardized form of x_t satisfies the (weak) functional law (e.g. Phillips and Solo, 1992)

$$X_n(\cdot) = \frac{x_{t=\lfloor n\cdot \rfloor}}{\sqrt{n}} \to_d B(\cdot) = BM(\omega^2), \tag{20}$$

where B is a Brownian motion with (long run) variance ω^2 and $\lfloor \cdot \rfloor$ is the integer floor function. It is convenient in what follows to strengthen (20). Using Lemma 3.1 of Phillips(2007) when u_t has a general linear process (Wold) representation

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

for all $t > 0$, with $\varepsilon_t = iid(0, \sigma_\varepsilon^2)$ and $E\left(|\varepsilon_t|^p\right) < \infty$ for some $p > 2$, it is known that an expanded probability space can be constructed with a Brownian motion $B(\cdot)$ for which uniform convergence holds, viz.,

$$\sup_{0 \le t \le n} \left| \frac{x_t}{\sqrt{n}} - B\left(\frac{t}{n}\right) \right| = o_{a.s.}\left(\frac{1}{n^{1/2-1/p}}\right), \tag{21}$$

so that in this space the functional law convergence (20) takes the strong form

$$\frac{x_{\lfloor nr \rfloor}}{\sqrt{n}} - B(r) = o_{a.s.}(1). \tag{22}$$

In what follows and unless otherwise stated we assume that we are working in this expanded probability space. In the original space the results translate, as usual, into weak convergence mirroring (20).

We make a corresponding normalization assumption on the posited trend process f_t so that the class of allowable (interpolating) functions admits the limiting form of a normalized stochastic trend. Thus, if x_t satisfies (20), we suppose that

$$\frac{f_t}{\sqrt{n}} = F_n\left(\frac{t}{n}\right) \to f(r), \tag{23}$$

where F_n is taken to be a continuous function that interpolates the points $\{f_t/\sqrt{n} : t = 1, ..., n, \}$, and the limit function $f(r) \in C[0,1] \cap QV$ where QV is the class of functions on $[0,1]$ with finite quadratic variation. This assumption allows potentially for Brownian motion limits such as $f(r) = B(r)$. Again, by an appropriate change in the probability space and allowing for stochastic trend processes in the limit, we can interpret the convergence in (23) in the strong form when taken in the same probability space, viz.,

$$\frac{f_{\lfloor nr \rfloor}}{\sqrt{n}} - f(r) = o_{a.s.}(1). \tag{24}$$

Given the standardarization used in (20) and (23), the filtering problem can therefore be formalized so that the HP or Whittaker filter is selected according to the criterion

$$\hat{f}_t = \arg \min_{\frac{f_t}{\sqrt{n}} = F_n\left(\frac{t}{n}\right), F_n \in \mathbb{S}} \left\{ \sum_{t=1}^{n} (x_t - f_t)^2 + \lambda \sum_{t=m+1}^{n} (\triangle^m f_t)^2 \right\}, \tag{25}$$

where $\mathbb{S} = \{F_n \in C[0,1] \cap QV\} \subset \mathbb{L}_2[0,1]$ is a smoothness class that restricts the interpolating function F_n to functions that are continuous and have finite quadratic variation on the interval $[0,1]$. The use of an interpolating function F_n in a certain class such as $\mathbb{S}$ becomes useful as we consider the limit behavior of the filter as $n \to \infty$ and the interval between standardized observations $(1/n)$ shrinks to zero. In some instances, it may be useful to extend this class to admit limit behavior that allows for trend breaks in the limit function $f(r)$ in which case we might use the Skorohod space $D[0,1]$ rather than $C[0,1]$ in the definition of $\mathbb{S}$.

We will see later that the asymptotic solution to the HP filter (2) when (22) holds and $m = 2$ is $f_{HP}(r) = B(r)$ provided λ is finite or passes to infinity slowly enough as $n \to \infty$. In other cases, the limiting trend function $f_{HP}(r)$ takes different forms depending on the expansion rate of λ as $n \to \infty$. In each case, the limiting trend function $f_{HP}(r)$ is stochastic and embodies some stochastic characteristics of the limiting form of the standardized process $n^{-1/2}x_{\lfloor nr \rfloor}$.

To characterize the limiting form of the filter solutions, it is convenient to use a general framework that embodies the limiting stochastic process $B(r)$ as well as other possible stochastic trend processes. To this effect we use the Karhunen-Loève (KL) representation of the limit process in (20) over the interval $[0,1]$, viz.,

$$\begin{aligned} B(r) &= \sqrt{2} \sum_{k=1}^{\infty} \frac{\sin\left[\left(k - \frac{1}{2}\right)\pi r\right]}{\left(k - \frac{1}{2}\right)\pi} \xi_k, \quad \xi_k = iid\ N(0, \omega^2) \\ &= \sum_{k=1}^{\infty} \sqrt{\lambda_k} \varphi_k(r) \xi_k, \end{aligned} \tag{26}$$

(Phillips, 1998; and Phillips & Liao, 2014, for a recent overview), where $\{\varphi_k(r) = \sqrt{2}\sin\left[\left(k - \frac{1}{2}\right)\pi r\right] = \sqrt{2}\sin\left(r/\sqrt{\lambda_k}\right)\}_{k=1}^{\infty}$ is an orthonormal system of eigenfunctions in $L_2[0,1]$ and $\lambda_k = 1/[\left(k - \frac{1}{2}\right)\pi]^2$ are the corresponding eigenvalues. The series (26) is well known to converge almost surely and uniformly in $r \in [0,1]$, which implies that $B(r)$ is arbitrarily well approximated by a finite series $\sum_{k=1}^{K} \sqrt{\lambda_k}\varphi_k(r)\xi_k$ for large enough K. For such cases, it will be convenient to use as the interpolating functions in (25) the specific class $\mathbb{S}_\varphi = \left\{ \sum_{k=1}^{\infty} d_k \varphi_k(r) : \sum_{k=1}^{\infty} d_k^2 < \infty \right\} \subset \mathbb{L}_2[0,1]$ spanned by the ON functions $\{\varphi_k(r)\}_1^{\infty}$.

We next consider some specific expansion rates for λ. The most important turns out to be $\lambda = O\left(n^4\right)$ for the HP filter and $\lambda = O\left(n^{2m}\right)$ for the Whittaker filter. These orders provide critical values that determine whether or not the filters produce a stochastic trend process or a polynomial trend process (i.e., a polynomial with random coefficients) in the limit.

For the HP filter penalty we let $\lambda_{HP} = \mu n^4$ and for the general Whittaker filter set $\lambda_W = \mu n^{2m}$, so that in both cases $\lambda \to \infty$ much faster than n. As the following result shows, the expansion rate is fast enough to ensure that these filters are not consistent for a stochastic trend but they are not so fast as to produce only a simple polynomial time trend limit. Instead, both filters produce limiting Gaussian stochastic processes that embody elements of the stochastic trend (20) that is being modeled. Both of these limiting stochastic processes fall within the usual 'flexible ruler' Bayesian interpretation of the HP and Whittaker filters in the sense that the limit functions are smooth.

Theorem 3 *If x_t satisfies the functional law (22) and $\lambda = \mu n^4$ then the HP filter $\hat{f}^{HP}_{t=\lfloor nr \rfloor}$ has the following limiting form as $n \to \infty$*

$$\frac{\hat{f}^{HP}_{\lfloor nr \rfloor}}{\sqrt{n}} \to_{a.s.} f_{HP}(r) = \sum_{k=1}^{\infty} \frac{\lambda_k^2}{\mu + \lambda_k^2} \sqrt{\lambda_k} \varphi_k(r) \xi_k. \tag{27}$$

When $\lambda = \mu n^{2m}$, the Whittaker filter for general $m \geq 2$ has the corresponding limiting form as $n \to \infty$

$$\frac{\hat{f}^{W}_{\lfloor nr \rfloor}}{\sqrt{n}} \to_{a.s.} f_{W}(r) = \sum_{k=1}^{\infty} \frac{\lambda_k^m}{\mu + \lambda_k^m} \sqrt{\lambda_k} \varphi_k(r) \xi_k. \tag{28}$$

In both cases $\xi_k \sim_{iid} N\left(0, \omega^2\right)$, $\varphi_k(r) = \sqrt{2} \sin\left\{\left(k - \frac{1}{2}\right)\pi r\right\} = \sqrt{2}\sin\left(r/\sqrt{\lambda_k}\right)$, and $\lambda_k = 1/\left\{\left(k - \frac{1}{2}\right)\pi\right\}^2$.

Remark 8 In (27) and (28) the limit processes are random and involve the same component variables ξ_k that appear in the limiting Brownian motion process (26) derived directly from the nonstationary data. Thus, in the case where $\lambda = \mu n^4$ and $n \to \infty$, the HP filtered trend tends to a limiting stochastic process whose components depend on those of the limiting process $B(r)$. Prima facie, this outcome seems different to the case where $\lambda \to \infty$ with fixed n, for which the HP filtered trend is just a simple linear trend. However, even in that case the limiting (as $\lambda \to \infty$) trend process, $R_2 \left(R_2' R_2\right)^{-1} R_2' x$, has random coefficients $\left(R_2' R_2\right)^{-1} R_2' x$. Upon standardization using $F_n = \text{diag}[1, n]$ these coefficients satisfy

$$F_n\left(R_2'R_2\right)^{-1}R_2'\frac{x}{\sqrt{n}}=\left(\frac{1}{n}F_n^{-1}R_2'R_2F_n^{-1}\right)^{-1}\frac{1}{n}F_n^{-1}R_2'\frac{x}{\sqrt{n}}$$
$$\to_{a.s.}\left[\begin{array}{cc}1 & \int_0^1 r\\ \int_0^1 r & \int_0^1 r^2\end{array}\right]^{-1}\left[\begin{array}{c}\int_0^1 B(r)\,dr\\ \int_0^1 rB(r)\,dr\end{array}\right]=:\left[\begin{array}{c}\alpha_{HP}\\ \beta_{HP}\end{array}\right]. \tag{29}$$

Hence, from (9) as $\lambda\to\infty$ for fixed n

$$\frac{\hat{f}^{HP}}{\sqrt{n}}=R_2F_n^{-1}F_n\left(R_2'R_2\right)^{-1}R_2'\frac{x}{\sqrt{n}}+o_p(1). \tag{30}$$

So the leading term of (30) is, in sequential asymptotics[10] as $(n,\lambda)_{\text{seq}}\to\infty$,

$$\frac{\hat{f}^{HP}_{t=\lfloor nr\rfloor}}{\sqrt{n}}\sim\left[R_2F_n^{-1}\right]_{t=nr}F_n\left(R_2'R_2\right)^{-1}R_2'\frac{x}{\sqrt{n}}\to_{a.s.}\alpha_{HP}+\beta_{HP}r, \tag{31}$$

giving the limiting linear trend function $f_{HP}(r)=\alpha_{HP}+\beta_{HP}r$, which has random slope and intercept, both induced by the form of the limiting process $B(r)$. In this case, the limit function $f_{HP}(r)$ carries (smoothed) characteristics of the stochastic trend $B(r)$ only in the two coefficients (α_{HP},β_{HP}). Importantly, the linear trend limit applies when $\lambda\to\infty$ and the specific form of the coefficients (α_{HP},β_{HP}) appearing in (31) holds when $n\to\infty$ subsequently. A similar higher order polynomial limit applies in the case of the Whittaker filter when $(n,\lambda)_{\text{seq}}\to\infty$ - see (63) below.

Remark 9 Functions (27) and (28) provide explicit KL forms for the limit of the trends that are extracted by the HP and Whittaker filters when the original data is $I(1)$ and the tuning parameter $\lambda=\mu n^4$ for the HP filter and $\lambda=\mu n^{2m}$ for the Whittaker filter and constant $\mu>0$. In both cases, it is apparent that the filters do not reproduce the limiting trend process $B(r)$, so the filter does not deliver a consistent estimate of the (stochastic) trend function for these expansion rates of λ. Further, the HP estimate $\hat{c}_t^{HP}$ of the cycle component c_t has the following limiting functional form upon standardization

$$\frac{\hat{c}^{HP}_{\lfloor nr\rfloor}}{\sqrt{n}}=\frac{x_{\lfloor nr\rfloor}}{\sqrt{n}}-\frac{\hat{f}^{HP}_{\lfloor nr\rfloor}}{\sqrt{n}}\to_{a.s.}\sum_{k=1}^{\infty}\left\{\frac{\mu}{\mu+\lambda_k^2}\right\}\sqrt{\lambda_k}\varphi_k(r)\,\xi_k=:c_{HP}(r).$$

This limit function $c_{HP}(r)=B(r)-f_{HP}(r)$ is a stochastic process that is nondifferentiable almost everywhere and inherits the stochastic trend random wandering properties of the limiting Brownian motion process $B(r)$. It is therefore to be expected that for choices of the smoothing parameter that approximate

[10] The notation $(n,\lambda)_{\text{seq}}\to\infty$ signifies that $\lambda\to\infty$ followed by $n\to\infty$ (c.f., Phillips and Moon, 1999).

$\lambda = \mu n^4$ the HP filter fails to remove a stochastic trend and the imputed business cycle estimate $\hat{c}_t^{HP}$ inevitably imports the random wandering character of a stochastic trend, thereby producing 'spurious cycle' phenomena of the type observed in simulations in the past literature.

Remark 10 The explicit forms (27) and (28) enable us to characterise the properties of the limit processes $f_{HP}(r)$ and $f_W(r)$ in relation to the limiting trend function $B(r)$. In particular, $f_{HP}(r)$ in (27) is expressed in terms of the orthonormal basis functions φ_k and the orthonormal Gaussian variates ξ_k. Since $\lambda_k = O\left(1/k^2\right)$, the coefficients in this representation satisfy

$$\frac{\lambda_k^2}{\mu + \lambda_k^2}\sqrt{\lambda_k} = O\left(\frac{1}{k^5}\right),$$

from which we deduce that $f_{HP}(r)$ is a Gaussian stochastic process that is continuously differentiable to the 4'th order. Indeed, its fourth derivative is given by the almost surely convergent series[11]

$$f_{HP}^{(4)}(r) = \sum_{k=1}^{\infty} \frac{\sqrt{\lambda_k}}{\mu + \lambda_k^2}\varphi_k(r)\,\xi_k, \tag{32}$$

which is a non-differentiable Gaussian process similar to Brownian motion for all $\mu \neq 0$. Thus, when $\lambda = \mu n^4$, the trend that is extracted by the HP filter is a very smooth function. In a similar way from its KL representation, it is evident that $f_W(r)$ is a smooth Gaussian process differentiable to order $2m$.

Remark 11 The proof of Theorem 3 shows that for large n the HP trend filter takes the finite series approximate form

$$\frac{\hat{f}_{t,K_n}}{\sqrt{n}} = \sum_{k=1}^{K_n} \frac{\lambda_k^{5/2}}{\mu + \lambda_k^2}\varphi_k\left(\frac{t}{n}\right)\xi_k\left\{1 + o(1)\right\}, \tag{33}$$

where $K_n \to \infty$ as $n \to \infty$ such that $K_n/n \to 0$. Now set $\mu = \mu_K$ in (33) and let $\mu_{K_n} \to 0$ as $K_n \to \infty$. Then, $\frac{\lambda_k^{5/2}}{\mu_K + \lambda_k^2} \to \sqrt{\lambda_k}$ uniformly for $k \leq K_n$ and so

$$\frac{\hat{f}_{t,K_n}}{\sqrt{n}} \to_{a.s} \sum_{k=1}^{\infty} \sqrt{\lambda_k}\varphi_k(r)\,\xi_k = B(r), \tag{34}$$

which suggests that if the tuning parameter $\lambda = \mu_{K_n} n^4 = o\left(n^4\right)$ the approximate HP filter (33) succeeds in capturing the Brownian motion limit process of the stochastic trend as $n \to \infty$.

[11] The series (32) is evidently convergent almost surely by virtue of the L_2 martingale convergence theorem.

The accuracy of the approximation delivered by (33) is illustrated in Fig. 4. In this case the data are generated by taking $n = 100$ equispaced, discrete observations of the Brownian motion (26) calculated using 5000 terms of the series with $\omega^2 = 1$. The data are therefore drawn essentially from a standard Gaussian random walk. Fig. 4 also shows the HP filter computed directly with $\lambda = 1600$ and the limit function approximation $f_{HP}(r)$ computed with $\mu = 0.000016$ so that $\lambda = \mu n^4 = 1600$ using (33) with $K_n = 10$. As is clear from the Figure, the asymptotic form $f_{HP}(r)$ delivers an extremely good approximation to the actual HP filter. The HP filter and its asymptotic approximation both follow the general path of the data but do not reproduce any of its fine grain fluctations with this setting of λ. The only points of deviation appear to be the terminal end points of the series, for which exact end corrections are not included in the asymptotic theory, in contrast to the exact filter solution given by the matrix formula (4). The situation is similar to the empirical example shown in Fig. 3, where for annual real US GDP data the trend extracted by the HP filter with $\lambda = 300$ also follows the general path of the data without capturing all of the fine-grain wandering details.

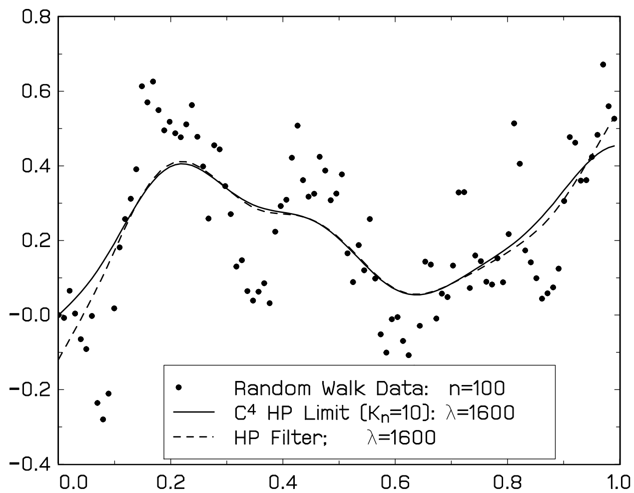


Fig. 4: Random Walk Data, the HP filter, and its C^4 Series Approximation (33) with $K_n = 10$

The asymptotic form of the HP filter (27) when $\lambda = \mu n^4$ turns out to be the solution of the following problem in continuous time

$$\arg\min_f \left\{ \int_0^1 (B(r) - f(r))^2 + \mu \int_0^1 \left[f^{''}(r) \right]^2 dr \right\}. \tag{35}$$

To see this, suppose that the interpolating functions $F_n, f \in C^4$ for all n and $\sup_{r\in[0,1]} \left| F_n^{(4)}(r) - f^{(4)}(r) \right| \to 0$. The normalized first part of (2) has the following limit

$$\frac{1}{n^2} \sum_{t=1}^{n} (x_t - f_t)^2 = \frac{1}{n} \sum_{t=1}^{n} \left(\frac{x_t}{\sqrt{n}} - \frac{f_t}{\sqrt{n}} \right)^2 \to_{a.s} \int_0^1 (B(r) - f(r))^2 dr, \tag{36}$$

by continuous mapping. Next consider

$$\frac{\lambda}{n^2}\sum_{t=1}^{n}(\triangle^2 f_t)^2 = \frac{\lambda}{n}\sum_{t=1}^{n}\left(\triangle^2\frac{f_t}{\sqrt{n}}\right)^2 = \frac{\lambda}{n}\sum_{t=1}^{n}\left(\triangle^2 F_n\left(\frac{t}{n}\right)\right)^2 \tag{37}$$

$$\to \mu\int_0^1 f^{''}(r)^2 dr, \tag{38}$$

again by continuous mapping and the fact that $n^2\left[\triangle^2 F_n(\frac{\lfloor nr\rfloor}{n})\right]$ converges uniformly to $f^{''}(r)$.

From these results we can derive directly the asymptotics of the continuous time HP filter. Suppose f satisfies the initial condition $f(0)=0$ and can be written in terms of the basis functions[12] $\{\varphi_k\}$ as $f(r)=\sum_{k=1}^{\infty}c_k\varphi_k(r)$. Then, since $f\in C^4[0,1]$, the first two derivative series converge pointwise and we have

$$f''(r)=\sum_{k=1}^{\infty}c_k\varphi_k''(r)=\sum_{k=1}^{\infty}\frac{c_k}{\lambda_k}\varphi_k(r).$$

It follows that (35) is equivalent to the following optimization problem with respect to the Fourier coefficients

$$\arg\min_{c_k}\left\{\sum_{k=1}^{\infty}\left(c_k-\sqrt{\lambda_k}\xi_k\right)^2+\mu\sum_{k=1}^{\infty}\left(\frac{c_k}{\lambda_k}\right)^2\right\}. \tag{39}$$

Solving (39) we get

$$c_k=\frac{1}{1+\mu/\lambda_k^2}\sqrt{\lambda_k}\xi_k=\frac{\lambda_k^2}{\mu+\lambda_k^2}\sqrt{\lambda_k}\xi_k,$$

which corresponds precisely to the coefficients that appear in the solution $f_{HP}(r)$ given in (27) above.

In the same way, we can obtain the continuous time Whittaker filter as the solution of

$$\arg\min_{f}\left\{\int_0^1(B(r)-f(r))^2+\mu\int_0^1\left[f^m(r)\right]^2dr\right\}. \tag{40}$$

Assuming that $F_n, f\in C^{2m}$ and $\sup_{r\in[0,1]}\left|F_n^{(2m)}(r)-f^{(2m)}(r)\right|\to 0$, it follows that (40) is equivalent to

$$\arg\min_{c_k}\left\{\sum_{k=1}^{\infty}\left(c_k-\sqrt{\lambda_k}\xi_k\right)^2+\mu\sum_{k=1}^{\infty}\left(\frac{c_k}{\lambda_k^{m/2}}\right)^2\right\},$$

which leads directly to the solution $c_k=\frac{\lambda_k^m}{\mu+\lambda_k^m}\sqrt{\lambda_k}\xi$ and expression (28) for $f_W(r)$.

[12] Since $f(0)=0$ and $f\in C^4[0,1]$, the Fourier series $f(r)=\sum_{k=1}^{\infty}c_k\varphi_k(r)$ is pointwise convergent over the entire interval $[0,1]$.

4 Stochastic Trends with Drift and Breaks

It is often realistic to allow for a limit process that is a stochastic trend with drift, so that in place of (20) we have the weak convergence

$$X_n(\cdot) = \frac{x_{t=\lfloor nr \rfloor}}{\sqrt{n}} \to_d \alpha + \beta r + B(r). \tag{41}$$

A suitable generating mechanism for the discrete time process x_t leading to (41) involves a localized drift function[13], such as $x_t = \alpha_n + \beta_n t + x_t^0$, where x_t^0 is a pure stochastic trend satisfying $X_n^0(r) = n^{-1/2} x^0_{t=\lfloor nr \rfloor} \to_d B(r)$. The accompanying linear trend is sample size dependent with coefficients that satisfy $n^{-1/2}\alpha_n \to \alpha$, and $\sqrt{n}\beta_n \to \beta$. The time trend $\beta_n t \sim \frac{\beta}{\sqrt{n}} t$ then has a local to zero coefficient $\frac{\beta}{\sqrt{n}}$ and the intercept $\alpha_n \sim \sqrt{n}\alpha$ has the same order as the stochastic trend x_t^0, thereby ensuring that (41) holds in the limit as $n \to \infty$. For the general polynomial trend case, we can use the formulation

$$x_t = \alpha_n + \beta_{n,1} t + ... + \beta_{n,J} t^J + x_t^0, \tag{42}$$

with

$$\frac{\alpha_n}{\sqrt{n}} \to \alpha \text{ and } n^{j-\frac{1}{2}} \beta_{n,j} \to \beta_j \text{ for } j = 1, .., J, \tag{43}$$

so that

$$X_n(\cdot) = \frac{x_{t=\lfloor nr \rfloor}}{\sqrt{n}} \to_d \alpha + \beta_1 r + ... + \beta_J r^J + B(r). \tag{44}$$

As before, it is convenient to work in an expanded probability space where (41) and (44) hold *a.s.* using (22). Since polynomial time trends are preserved under the Whittaker filter operation (7) up to degree $J \leq m-1$, the HP and Whittaker filters on x_t will have limit theory comparable to Theorem 3 for the stochastic trend component augmented by a continuous time polynomial trend of corresponding degree. The following result details these limits.

Theorem 4 *If $x_t = \alpha_n + \beta_n t + x_t^0$ where x_t^0 satisfies the functional law (22), the coefficients α_n and β_n satisfy (43), and $\lambda = \mu n^4$ then the HP filter $\hat{f}^{HP}_{t=\lfloor nr \rfloor}$ has the following limiting form as $n \to \infty$*

$$\frac{\hat{f}^{HP}_{\lfloor nr \rfloor}}{\sqrt{n}} \to_{a.s.} f_{HP}(r) = \alpha + \beta r + \sum_{k=1}^{\infty} \frac{\lambda_k^2}{\mu + \lambda_k^2} \sqrt{\lambda_k} \varphi_k(r) \xi_k. \tag{45}$$

If x_t is generated as in (42) with a deterministic trend of degree $J \leq m-1$, then the Whittaker filter with penalty $\lambda = \mu n^{2m}$ has the corresponding limiting form

$$\frac{\hat{f}^{W}_{\lfloor nr \rfloor}}{\sqrt{n}} \to_{a.s.} f_W(r) = \alpha + \beta_1 r + ... \beta_J r^J + \sum_{k=1}^{\infty} \frac{\lambda_k^m}{\mu + \lambda_k^m} \sqrt{\lambda_k} \varphi_k(r) \xi_k. \tag{46}$$

[13] See Phillips, Shi and Yu (2014).

Remark 12 For large n the HP filter of $x_t = \alpha_n + \beta_n t + x_t^0$ takes the finite series approximate form

$$f_{HP}^{K_n}\left(\frac{t}{n}\right) = \frac{\hat{f}_{t,K_n}}{\sqrt{n}} = \alpha + \beta\frac{t}{n} + \sum_{k=1}^{K_n} \frac{\lambda_k^{5/2}}{\mu + \lambda_k^2}\varphi_k\left(\frac{t}{n}\right)\xi_k\left\{1 + o(1)\right\}, \tag{47}$$

where $K_n \to \infty$ as $n \to \infty$ such that $K_n/n \to 0$. As discussed earlier, if we set $\mu = \mu_K$ in (47) and let $\mu_{K_n} \to 0$ as $K_n \to \infty$. Then, analogous to (34) we have

$$\frac{\hat{f}_{t,K_n}}{\sqrt{n}} \to_{a.s} \alpha + \beta r + B(r),$$

suggesting that for smaller tuning parameter rates where $\lambda = \mu_{K_n} n^4 = o\left(n^4\right)$ the HP filter captures the limiting Brownian motion with drift process as $n \to \infty$. Fig. 5 illustrates the HP filter asymptotic approximation (47) to a random walk generated for $n = 100$ with drift using the (limiting) intercept and slope parameter settings $\alpha = 10$ and $\beta = 2$. As in Fig. 4, the HP filter is computed directly with $\lambda = 1600$ and the approximation $f_{HP}^{K_n}(r)$ is computed with $\mu = 0.000016$ so that $\lambda = \mu n^4 = 1600$. Computations are performed using the finite series (47) with $K_n = 10$.

For trends involving stochastic trends with higher polynomial degrees ($K_{HP} \geq 4$ for HP and $K_W \geq 2m$ for Whittaker) the asymptotic forms of the filters project the higher order time polynomials onto lower order polynomials ($J_{HP} = 3, J_W = 2m - 1$) and apply the smoother to the residual process. The same process occurs in the case of data generated with breaking polynomial trends or trends with multiple break points.

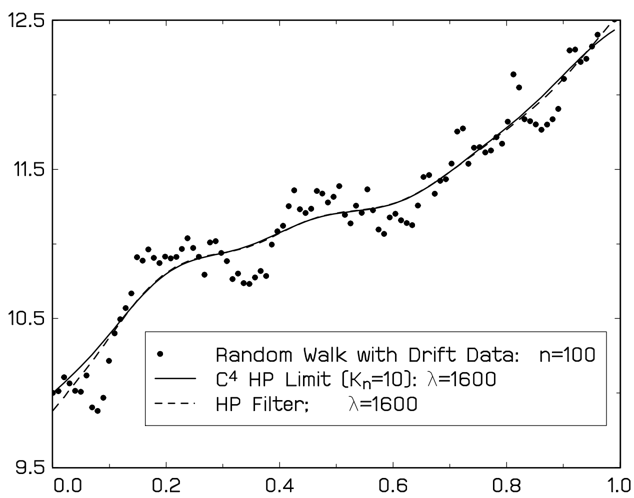


Fig. 5: Random Walk with Drift Data, the HP filter, and its C^4 Approximation (47) with $K_n = 10$ and $\mu = 0.000016$ (giving $\lambda = 1600$).

These cases are covered by the following general result. Suppose that $x_t = g_n(t) + x_t^0$ in which x_t^0 satisfies the functional law (22) and $g_n(t)$ is a piecewise smooth trend function with a finite number of break points and sample size dependent coefficients such that $n^{-1/2} g_n(\lfloor nr \rfloor) \to g(r)$, where $g(r)$ is a piecewise smooth function for $r \in [0,1]$ with convergent Fourier series in $L_2[-\pi,\pi]$. In place of (22) we then have under the same conditions

$$X_n(\cdot) = \frac{x_{t=\lfloor n\cdot \rfloor}}{\sqrt{n}} \to_{a.s.} B(\cdot) + g(\cdot) =: B_g(\cdot). \tag{48}$$

We suppose in what follows that the continuous interpolating function for the deterministic trend component can be written in terms of its Fourier series using the complex exponential basis functions $(2\pi)^{-1/2} e^{ikr}$, so that the interpolating class has the general trigonometric form $\mathbb{S}_\psi = \left\{ \sum_{k=-\infty}^{\infty} c_k \psi_k(r) : \sum_{k=1}^{\infty} c_k^2 < \infty \right\} \subset \mathbb{L}_2[-\pi,\pi]$ with $\psi_k(r) = (2\pi)^{-1/2} e^{ikr}$. The limit function $g(r)$ then has the Fourier series representation

$$g^F(r) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} c_k e^{-ikr} = \frac{c_0}{2\pi} + \frac{1}{\pi} \sum_{k=1}^{\infty} \mathrm{Re}\left[c_k e^{-ikr} \right], \quad r \in [-\pi,\pi], \tag{49}$$

with coefficients $c_k = \int_{-\pi}^{\pi} e^{ikr} g(r)\, dr$. For example, suppose $g_n(t)$ is a trend break polynomial with a single break point at $\tau_0 = \lfloor nr_0 \rfloor$ that takes the form

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

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 $\delta = 0,1$. Then

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

and, in view of the finite number of jump discontinuities in the otherwise smooth function $g(r)$, its Fourier series $g^F(r)$ in (49) converges pointwise over $[0,1]$, although not to $g(r)$ at break points such as r_0, but instead to midpoints of the left and right limits such as $\frac{1}{2}\left\{ g\left(r_0^+\right) + g\left(r_0^-\right) \right\}$. On the other hand, $g^F(r) = g(r)$ for all points of continuity of g and, by standard Fourier analysis (e.g. Tolstov, 1976, pp. 125-129), smooth integral operations on g^F, such as $G^F(r) = \int_0^r g^F(s)\, ds$, have everywhere pointwise convergent (to $G(r) = \int_0^r g(s)\, ds$) Fourier series that are the termwise integrals of the Fourier series of g. The HP and Whittaker filters in this case have the following approximations and limit forms when the trigonometric functions $\left\{ (2\pi)^{-1} e^{-ikr} \right\}$ are used as the basis functions in the Fourier series representation $g^F(r)$ of the interpolating limiting trend function $g(r)$.

Theorem 5 *If $x_t = g_n(t) + x_t^0$ where x_t^0 satisfies the functional law (22) and $g_n(t)$ is a piecewise smooth interpolating function with convergent Fourier series (49),* *then the HP filter* $\hat{f}^{HP}_{t=\lfloor nr\rfloor}$ *with penalty* $\lambda = \mu n^4$ for $\mu \in (0,\infty)$ *has the approximating limit form* $n^{-1/2}\hat{f}^{HP}_{\lfloor nr\rfloor} \rightarrow_{a.s.} f_{HP}(r)$ *as* $n\to\infty$ with

$$f_{HP}(r) = \left\{\frac{c_0}{2\pi} + \frac{1}{\pi}\sum_{k=1}^{\infty}\frac{1}{1+\mu k^4}\operatorname{Re}\left[c_k e^{-ikr}\right]\right\} + \sum_{k=1}^{\infty}\frac{\lambda_k^2}{\mu+\lambda_k^2}\sqrt{\lambda_k}\varphi_k(r)\,\xi_k. \tag{50}$$

The Whittaker filter with penalty $\lambda = \mu n^{2m}$ for $\mu \in (0,\infty)$ *has the approximating limit form* $n^{-1/2}\hat{f}^{W}_{\lfloor nr\rfloor} \rightarrow_{a.s.} f_W(r)$ *as* $n\to\infty$ with

$$f_W(r) = \left\{\frac{c_0}{2\pi} + \frac{1}{\pi}\sum_{k=1}^{\infty}\frac{1}{1+\mu k^{2m}}\operatorname{Re}\left[c_k e^{-ikr}\right]\right\} + \sum_{k=1}^{\infty}\frac{\lambda_k^m}{\mu+\lambda_k^m}\sqrt{\lambda_k}\varphi_k(r)\,\xi_k. \tag{51}$$

Remark 13 The components in braces on the right sides of (50) and (51) are the limiting forms of the HP and Whittaker filters for the breaking trend function $g(r)$. The effect of these filters is to smooth the (possibly discontinuous) limit function $g(r)$ into a smooth curve where breaks are captured by smooth transitions. For the HP case, when trigonometric basis functions are used to represent the trend break function $g_n(t)$, we have the smoothed limit function

$$g^{HP}(r) := \frac{c_0}{2\pi} + \frac{1}{\pi}\sum_{k=1}^{\infty}\frac{1}{1+\mu k^4}\operatorname{Re}\left[c_k e^{-ikr}\right]. \tag{52}$$

This smoothed limit function converges faster to its limit than the original Fourier series $g^F(r) = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty} c_k e^{-ikr}$ in view of the presence of the factor $1/\left(1+\mu k^4\right) = O\left(k^{-4}\right)$ in each term of the series. The extent of smoothing that is involved depends on the magnitude of the parameter μ, with larger μ producing more heavily smoothed versions of the break points in $g(r)$. As is apparent in the examples studied below (Remark 16), small values of μ still produce smoothing but retain greater fidelity to $g(r)$, while smoothing out ripples that occur in finite versions of the Fourier series representation (52).

Remark 14 As shown in the proof of Theorem 4, for large n the HP filter $\hat{f}_t$ of $x_t = g_n(t) + x_t^0$ has the finite series approximate form $\hat{f}_{t,K_n}$ where

$$\frac{\hat{f}_{t,K_n,\mu}}{\sqrt{n}} = g^{HP}_{K_n,\mu}(r) + \sum_{k=1}^{K_n}\frac{\lambda_k^{5/2}}{\mu+\lambda_k^2}\varphi_k\left(\frac{t}{n}\right)\xi_k, \tag{53}$$

with

$$g^{HP}_{K_n,\mu}(r) = \frac{c_0}{2\pi} + \frac{1}{\pi}\sum_{k=1}^{K_n}\frac{1}{1+\mu k^4}\operatorname{Re}\left[c_k e^{-ik\frac{t}{n}}\right].$$

When $K_n\to\infty$ as $n\to\infty$ such that $K_n/n\to 0$, (53) has the limiting form $f_{HP}(r)$ given in (50). Further, if we let $\mu = \mu_K \to 0$ as $K_n\to\infty$ with $K_n/n\to$

0 in (53), we find that $g^{HP}_{K_n,\mu_K}(r) \to \frac{c_0}{2\pi} + \frac{1}{\pi}\sum_{k=1}^{\infty} \mathrm{Re}\left[c_k e^{-ik\frac{t}{n}}\right] = g^F(r)$, which is the same as the deterministic trend break function $g(r)$ except at break points, for which the Fourier series $g^F(r)$ converges but not necessarily to the value of $g(r)$ at the break points. In this event, we have

$$f^{K_n}_{HP}(r) = \frac{\hat{f}_{t=\lfloor nr \rfloor, K_n, \mu_K}}{\sqrt{n}} \to g^F(r) + B(r), \text{ as } n \to \infty,$$

so that the HP filter succeeds in capturing both the continuous part of the stochastic and deterministic trends in the limit. These sequential asymptotics suggest that for tuning parameter expansion rates $\lambda = \mu_n n^4 = o(n^4)$ with suitable $\mu_n \to 0$ as $n \to \infty$, the HP and Whittaker filters will capture the stochastic trend limit function $B(r)$ and the Fourier series form $g^F(r) = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty} c_k e^{-ikr}$ of the trend break. When there are no break points, smooth higher order polynomials are captured exactly in the limit by both filters in this case.

Remark 15 If basis functions other than complex exponentials are used for the interpolating function of the deterministic trend $n^{-1/2}g_n(t = \lfloor nr \rfloor)$ and its limit function $g(r)$, then the HP and Whittaker filters have alternate asymptotic forms in terms of the new basis. In such cases, the smoothness class of interpolating functions $\mathbb{S}_\psi = \left\{\sum_{k=-\infty}^{\infty} c_k \psi_k(r) : \sum_{k=1}^{\infty} c_k^2 < \infty\right\}$ will involve different functions $\{\psi_k(r)\}_1^{\infty}$ from the complex exponentials. For instance, we might use the polynomials $\{1, r, r^2, ...\}$ as a basis or orthogonal versions of them, such as the (shifted) Legendre polynomials $\left\{\tilde{P}_m(r)\right\}_{m=0}^{\infty}$, which are orthogonal over the interval $r \in [0,1]$, where $\tilde{P}_m(r)$ is the (shifted) Legendre polynomial of degree m in r. Then, if the deterministic trend $g_n(t) = \alpha_n + \beta_{n,1}t + ... + \beta_{n,J}t^J$ is itself a high order polynomial with coefficients α_n and β_{nj} satisfying (43), then the limiting trend function has a similar polynomial representation as

$$\lim_{n\to\infty} \frac{1}{\sqrt{n}} g_n\left(\frac{\lfloor nr \rfloor}{n}\right) = g(r) = \alpha + \beta_1 r + ... + \beta_J r^J.$$

In this event, applying the asymptotic form of Whittaker operator $1/\left\{1 + \mu n^{2m}\Delta^m \Delta^{*m}\right\}$ we have

$$\begin{aligned} & \left\{1 + \mu n^{2m}\Delta^m \Delta^{*m}\right\}^{-1}\left[\frac{1}{\sqrt{n}} g_n(t = \lfloor nr \rfloor)\right] \\ = \;& \frac{\alpha_n}{\sqrt{n}} + \frac{\beta_{n,1}}{\sqrt{n}}\lfloor nr \rfloor + ... + \frac{\beta_{n,J}}{\sqrt{n}}\lfloor nr \rfloor^J \\ \to \;& \begin{cases} \alpha + \beta_1 r + ... + \beta_J r^J & \text{for } J < 2m \\ \alpha + \beta_1 r + ... + \beta_J r^J + Q(\mu, r) & \text{for } J \geq 2m \end{cases}, \end{aligned} \tag{54}$$

where $Q(\mu, r)$ is a polynomial in μ of degree $\lfloor J/(2m) \rfloor$ with coefficients involving powers of r such that $\lim_{\mu\to 0} Q(\mu, r) = 0$. The retention of the limit polynomial $g(r)$ in the above expression holds because $\Delta^{mk}\Delta^{*mk}r^J = 0$ for all $k \geq 1$ when $J < 2m$, which explains the first element of (54). When $J \geq 2m$, higher order terms in the expansion produce non zero terms involving powers of μ in addition to the limit polynomial $g(r)$. To illustrate, we evaluate the component $\left\{1 + \mu n^{2m}\Delta^m\Delta^{*m}\right\}^{-1}\left(\frac{t}{n}\right)^j$ for an arbitrary integer j as follows

$$\begin{aligned}
&\left\{1 + \mu n^{2m}\Delta^m\Delta^{*m}\right\}^{-1}\left(\frac{t}{n}\right)^j = \int_0^\infty \exp\left\{-\left[1 + \mu n^{2m}\Delta^m\Delta^{*m}\right]s\right\} ds\left[\left(\frac{t}{n}\right)^j\right] \\
= &\int_0^\infty e^{-s}\sum_{k=0}^\infty \frac{s^k}{k!}\left\{-\mu n^{2m}\Delta^m\Delta^{*m}\right\}^k ds\left[\left(\frac{t}{n}\right)^j\right] \\
= &\sum_{k=0}^\infty\left[\int_0^\infty e^{-s}\frac{s^k}{k!}\left\{-\mu n^{2m}\Delta^m\Delta^{*m}\right\}^k\left[\left(\frac{t}{n}\right)^j\right] ds\mathbf{1}\left\{2mk \leq j\right\}\right] \\
= &\sum_{k=0}^{\lfloor j/(2m)\rfloor}\left\{-\mu n^{2m}\Delta^m\Delta^{*m}\right\}^k\left(\frac{t}{n}\right)^j = \left\{1 - \mu n^{2m}\Delta^m\Delta^{*m} + \mu^2 n^{4m}\Delta^{2m}\Delta^{*2m} + \ldots\right\}\left(\frac{t}{n}\right)^j \\
= &\begin{cases}
\left(\frac{t}{n}\right)^j & \text{for } j < 2m \\
\left(\frac{t}{n}\right)^j + (-1)^{m+1}\mu(2m)! + O\left(n^{-1}\right) & \text{for } j = 2m \\
\left(\frac{t}{n}\right)^j + (-1)^{m+1}\mu(2m+1)!\left(\frac{t}{n}\right) + O\left(n^{-1}\right) & \text{for } j = 2m+1 \\
\left(\frac{t}{n}\right)^j + (-1)^{m+1}\mu\frac{(2m+2)!}{2!}\left(\frac{t}{n}\right)^2 + O\left(n^{-1}\right) & \text{for } j = 2m+2 \\
\vdots & \vdots \\
\left(\frac{t}{n}\right)^j + (-1)^{m+1}\mu\frac{(4m)!}{(2m)!}\left(\frac{t}{n}\right)^{2m} + \mu^2(4m)! + O\left(n^{-1}\right) & \text{for } j = 4m \\
\vdots & \vdots
\end{cases} \\
= &\left(\frac{t}{n}\right)^j + Q_j\left(\mu, \frac{t}{n}\right), \qquad (55)
\end{aligned}$$

where $Q_j\left(\mu, \frac{t}{n}\right)$ is a polynomial in μ of degree $\lfloor j/(2m)\rfloor$ with coefficients involving powers of $\frac{t}{n}$. The explicit form of the coefficients that appear in (55) are obtained by successive differencing. For example, when $k = 1$ and $j = 2m+\ell < 4m$ by successive recursion of the differencing operator $\Delta^m\Delta^{*m} = (-L)^{-m}\Delta^{2m}$ we have

$$\begin{aligned}
\mu n^{2m}\Delta^m\Delta^{*m}\left(\frac{t}{n}\right)^{2m+\ell} &= \mu n^{2m}(-L)^{-m}\Delta^{2m-1}\left\{\left(\frac{t}{n}\right)^{2m+\ell} - \left(\frac{t-1}{n}\right)^{2m+\ell}\right\} \\
&= \mu n^{2m-1}(-L)^{-m}\Delta^{2m-1}\left\{(2m+\ell)\left(\frac{t}{n}\right)^{2m+\ell-1} + O\left(\frac{1}{n}\right)\right\} \\
&= \mu(-L)^{-m}\left\{\frac{(2m+\ell)!}{\ell!}\left(\frac{t}{n}\right)^\ell + O\left(\frac{1}{n}\right)\right\}
\end{aligned}$$

$$= \quad (-1)^m \mu \left\{ \frac{(2m+\ell)!}{\ell!} \left(\frac{t+m}{n} \right)^\ell + O\left(\frac{1}{n}\right) \right\}$$

$$= \quad (-1)^m \mu \left\{ \frac{(2m+\ell)!}{\ell!} \left(\frac{t}{n} \right)^\ell + O\left(\frac{1}{n}\right) \right\},$$

with similar calculations when $k > 1$. The Whittaker filter therefore preserves polynomials of degree $J \leq 2m-1$ asymptotically as $n \to \infty$ when $\lambda = \mu n^4$. When the polynomial has degree $J \geq 2m$, the filter produces additional terms that are contained in $Q_j\left(\mu, \frac{t}{n}\right)$ which all tend to zero as $\mu \to 0$. Hence, just as in the case of trigonometric basis functions, the filter reproduces general deterministic polynomial trends exactly as $n \to \infty$ when $\lambda = \mu n^4$ and $\mu \to 0$. These results all apply to the HP filter with $m = 2$.

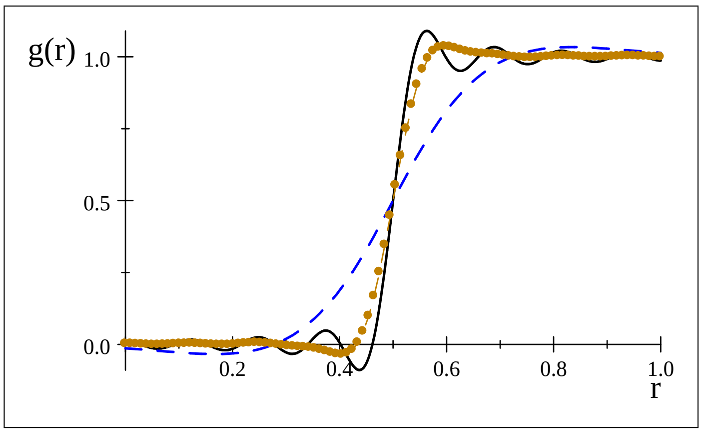


Fig. 6: Fourier series (black solid line:$K = 50$) and HP filter (blue dashed line: $K = 50$, $\mu = 0.0001$; sienna dot-dashed line: $K = 50; \mu = 0.000001$) approximations of a level shift function $g(r)$ with shift at $r_0 = 0.5$.

Remark 16 We illustrate these effects with a linear trend-break function of the type that commonly appears in empirical econometric work. Suppose the data follow a deterministic trend break process with limiting form given by

$$g(r) = (\alpha_1 + \beta_1 r)\, \mathbf{1}\{r < r_0\} + (\alpha_2 + \beta_2 r)\, \mathbf{1}\{r \geq r_0\} \tag{56}$$

for some break point $r_0 \in (0,1)$. The coefficients that appear in the trigonometric Fourier series for $g(r)$ and the limiting HP filter approximation (50) are found, after some calculations that are shown in the Appendix, to be

$$c_0 = \pi(\alpha_1 + \alpha_2) + r_0(\alpha_1 - \alpha_2) + \frac{1}{2}\left(\pi^2 - r_0^2\right)(\beta_2 - \beta_1), \tag{57}$$

and

$$c_k \quad = \quad \frac{e^{ikr_0} - e^{-ik\pi}}{ik}(\alpha_1 - \alpha_2) + \left\{ \frac{e^{ikr_0}}{ik} r_0 - \frac{e^{ikr_0}}{(ik)^2} + \frac{e^{-ik\pi}}{(ik)^2} \right\}(\beta_1 - \beta_2)$$
$$+ \left(\frac{e^{-ik\pi}}{ik}\pi \right)(\beta_1 + \beta_2), \tag{58}$$

for $k \geq 1$. The curves of the Fourier series for $g(r)$ and the limiting HP filter $f_{HP}(r)$ are shown for a level shift in Fig. 6 and for a trend break in Fig. 7. These are computed with finite sums $\sum_{k=1}^{K}$ for large K, replacing the infinite sums in (49) and the term in braces in (50) using the complex exponential basis to construct the interpolating function. In Fig. 6, we consider a level shift function, setting $\alpha_1 = \beta_1 = \beta_2 = 0$ and $\alpha_2 = 1$, giving the simple level shift function $g(r) = \mathbf{1}\{r \geq r_0\}$ in (56), and its HP filter approximations for various μ. In Fig. 7, we set $\alpha_1 = \alpha_2 = 0, \beta_1 = 0.5$, and $\beta_2 = 1$ in (56), giving the linear trend break function $g(r) = 0.5r\mathbf{1}\{r < r_0\} + r\mathbf{1}\{r \geq r_0\}$ with HP filter approximations shown for various values of K.

As is apparent in both Figs. 5 and 6, the smoothing action embodied in the coefficient scale factor $1/\left(1 + \mu k^4\right)$ in (50) helps the HP filter to accelerate the convergence of the Fourier series over a large part of the linear segments of g when μ is small, in addition to smoothing the break discontinuity of g at $r_0 = 0.5$ into a curve that accentuates the continuous transition approximation to the break represented in the finite Fourier series. In this sense the HP filter acts in a manner that resembles a finite trigonometric series approximation while having the twin properties of smoothing out the trigonometric ripples over the linear segments and creating a smooth transition to replace the location shift and trend break in the deterministic function.

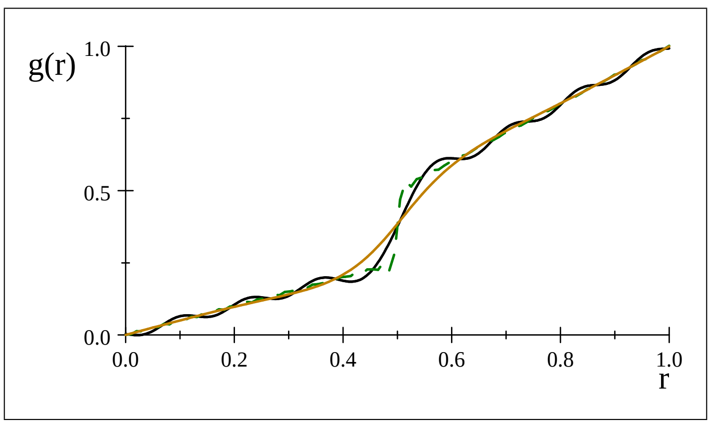


Fig. 7: Fourier series (black solid line: $K = 50$; green dashed line: $K = 250$) and HP filter (sienna solid line: $K = 250, \mu = 0.00001$) approximations of a breaking linear trend function $g(r)$ with break at $r_0 = 0.5$.

Remark 17 Fig. 8 shows equispaced data (with $n = 100$ observations) generated from a Brownian motion with a deterministic location shift corresponding to that of Fig. 6 (viz., $g(r) = \mathbf{1}\{r \geq 0.5\}$), shown against the corresponding HP filtered series and HP limit approximations as $n \to \infty$. As is apparent, the limiting HP filter approximation with $\mu = 0.000016$ (so that $\lambda = \mu n^4 = 1600$) provides a very close approximation the actual HP filter with the usual setting $\lambda = 1600$, whereas when $\mu = 10^{-10}$ (or $\lambda = \mu n^4 = 0.1$) the limiting HP approximation follows the fine grain course of the data in much greater detail, including the sharp level shift at the midpoint ($r_0 = 0.5$)

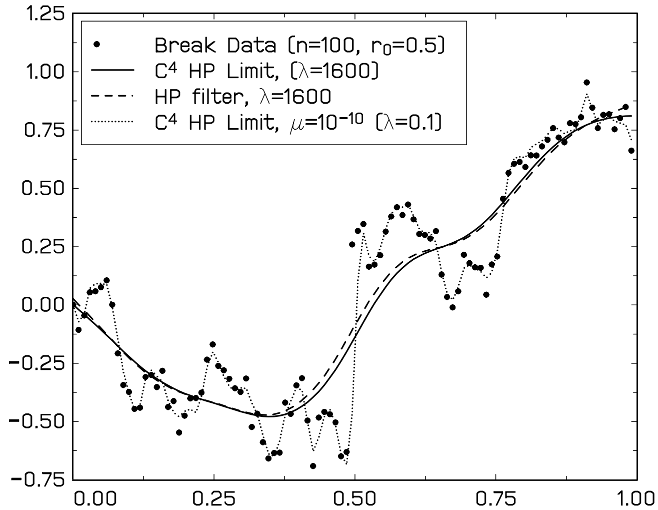


Fig. 8: Random walk with a location shift at $r_0 = 0.5$, shown against the HP filtered series and HP limit approximations ($\mu = 16 \times 10^{-6}$ and $\mu = 10^{-10}$).

Hence, in sequential limits as $\lambda = \mu n^4 \to \infty$ followed by $\mu \to 0$, the HP filter reproduces the correct limit process involving a Brownian motion with a deterministic drift limit $g(r)$ for all r except for break points such as r_0 for which the filter corresponds in the limit to the Fourier series of $g(r)$ given by $g^F(r) = \frac{c_0}{2\pi} + \frac{1}{\pi}\sum_{k=1}^{\infty} \operatorname{Re}\left[c_k e^{-ikr}\right]$.

These results show that, just as in nonparametric function estimation, the path to infinity of the smoothing parameter is important in influencing the asymptotic properties of the HP filter. Moreover, joint limits are not always the same as sequential limits. The obvious example is that limit results for $\lambda \to \infty$ for fixed n followed by $n \to \infty$ are usually very different from results where $\lambda = \mu n^4 \to \infty$ and cases where $(\mu, n)_{\text{seq}} \to \infty$. Furthermore and not unexpectedly, some interpolating functions used in the smoothing class $\mathbb{S}$ for deterministic trends lead to slightly different asymptotics because of their different capacities as approximations to trends of different forms. Thus, time polynomials are better modeled directly in terms of continuous time polynomials than by trigonometric polynomials. In consequence, one requires slightly different divergence rates on the smoothing parameter to achieve the same level of approximation or reproduction of the deterministic trend process in the limit as $n \to \infty$. These differences are reflected in the above results concerning whether we need $\lambda = o\left(n^4\right)$ or $\lambda = O\left(n^4\right)$ rates to embody limiting polynomial time trend solutions exactly in the filtered series. Of course, for breaking trend functions, use of a class $\mathbb{S}$ of continuous interpolating functions will typically lead to continuously differentiable limits that embody the smoothing effects of the HP filter in the continuous limit function, as demonstrated in the examples of a level shift and trend break given above.

5 Limit Theory for Weaker and Stronger Penalties

When x_t is $I(1)$, satisfies (22), and $\lambda = \mu n^4$ the asymptotic form of the trend HP solution is a Gaussian process that is four times continuously differentiable. When the expansion rate of λ as $n \to \infty$ is slower than $O\left(n^4\right)$, the effect of the penalty is weaker and the limit function is not as smooth, at least in this case where the data have a stochastic trend. When the expansion rate of λ exceeds $O\left(n^4\right)$, the penalty is stronger and the limit function is even smoother. These cases are studied next. We concentrate attention here on examining the case where x_t is $I(1)$ and satisfies (22). But closely related results apply in cases where x_t is near integrated (Phillips, 1987) or where the limit process is a continuous stochastic process with deterministic piecewise continuous drift, as will be indicated below.

Slower expansion rates for λ

In the extreme case where λ is fixed as $n \to \infty$ and if x_t and f_t satisfy (22) and (24), the HP optimization problem (2) with $m = 2$ may be written as

$$\begin{aligned} \frac{\hat{f}_t}{\sqrt{n}} &= \arg\min_{f_t/\sqrt{n}} \left\{ \frac{1}{n}\sum_{t=1}^{n}\left(\frac{x_t}{\sqrt{n}} - \frac{f_t}{\sqrt{n}}\right)^2 + \frac{\lambda}{n}\left[\frac{1}{n}\sum_{t=3}^{n}(\triangle^2 f_t)^2\right]\right\} \qquad (59) \\ &= \arg\min_{f_t/\sqrt{n}} \left\{ \frac{1}{n}\sum_{t=1}^{n}\left(\frac{x_t}{\sqrt{n}} - \frac{f_t}{\sqrt{n}}\right)^2 + o\left(\frac{\lambda}{n}\right)\right\}. \end{aligned}$$

It follows that for λ fixed or indeed for any $\lambda = o(n)$, the role of the penalty in the optimization diminishes as $n \to \infty$, leading to the stochastic trend HP solution

$$f_{HP}(r) = \arg\min_f \{ \int_0^1 (B(r) - f(r))^2 \} = B(r).$$

Next suppose, as earlier, that the interpolating fitted function F_n satisfies $F_n(r) \to f(r) \in C[0,1] \cap QV$ to accommodate potential stochastic trend solutions that include nondifferentiable processes like Brownian motion. Define the limiting quadratic variation function of F_n as

$$V_2(r) = \lim_{n\to\infty} \sum_{j=1}^{\lfloor nr \rfloor} \left\{ F_n\left(\frac{j}{n}\right) - F_n\left(\frac{j-1}{n}\right)\right\}^2 = \lim_{n\to\infty} \sum_{j=1}^{\lfloor nr \rfloor} \left\{ f\left(\frac{j}{n}\right) - f\left(\frac{j-1}{n}\right)\right\}^2.$$

In this case the limit behavior (36) for the first component of (59) continues to hold and by the continuity of f we have

$$\sum_{t=1}^{n}\left(F_n\left(\frac{t}{n}\right)-F_n\left(\frac{t-1}{n}\right)\right)^2 \quad \to_{a.s.} V_2(1),$$
$$\sum_{t=1}^{n}\left(F_n\left(\frac{t-1}{n}\right)-F_n\left(\frac{t-2}{n}\right)\right)^2 \quad \to_{a.s.} V_2(1),$$
$$\sum_{t=1}^{n}\left(F_n\left(\frac{t}{n}\right)-F_n\left(\frac{t-1}{n}\right)\right)\left(F_n\left(\frac{t-1}{n}\right)-F_n\left(\frac{t-2}{n}\right)\right) \quad \to_{a.s.} V_2(1).$$

It follows that

$$\begin{aligned}\sum_{t=1}^{n}\left\{\triangle^2 F_n\left(\frac{t}{n}\right)\right\}^2 &= \sum_{t=1}^{n}\left\{\left[F_n\left(\frac{t}{n}\right)-F_n\left(\frac{t-1}{n}\right)\right]-\left[F_n\left(\frac{t-1}{n}\right)-F_n\left(\frac{t-2}{n}\right)\right]\right\}^2 \\ &\to_{a.s.} V_2(1)-2V_2(1)+V_2(1)=0. \end{aligned} \tag{60}$$

Hence, for all $\lambda = \mu n$ with constant $\mu > 0$, we find that the limit of the HP filter is $f_{HP}(r) = B(r)$ as the normalized first element of (59) satisfies

$$\frac{1}{n^2}\sum_{t=1}^{n}(x_t - f_t)^2 \to \int_0^1 (B(r) - f(r))^2 dr \tag{61}$$

and the second element behaves as

$$\frac{\lambda}{n^2}\sum_{t=1}^{n}(\triangle^2 f_t)^2 = \mu\sum_{t=1}^{n}\left(\triangle^2\frac{f_t}{\sqrt{n}}\right)^2 \to 0,$$

in view of (60). So in this case the HP filter trend solution is simply $f_{HP}(r) = B(r)$.

A similar result applies for the Whittaker filter because (61) holds and

$$\frac{\lambda}{n^2}\sum_{t=1}^{n}((\triangle^m f_t)^2 = \mu\sum_{t=1}^{n}\left(\triangle^m\frac{f_t}{\sqrt{n}}\right)^2 \to 0,$$

in the same way as (60). Hence, under the same condition that $\lambda = O(n)$, the Whittaker trend solution is simply $f_W(r) = B(r)$. Stochastic trends are therefore removed by both the HP and Whittaker filters when $\lambda = O(n)$. Similar results clearly hold when the limit process is a continuous stochastic process with determinist drift function as in (48).

Faster expansion rates for λ

From Theorem 2 and (6) the HP filter is the solution of the operator equation

$$d(L)f = \left[\left(1+\lambda\Delta^2\Delta^{*2}\right)I_n + \lambda\Delta^2 EGE'\right]f = x$$

so that for $\lambda = \mu_n n^4$ with $\mu_n \to \infty$ we have

$$\frac{1}{\mu_n} f + n^4 \Delta^2 \Delta^{*2} f + n^4 \Delta^2 EGE' f = \frac{x}{\mu_n}$$

Now suppose that x_t satisfies (22) and correspondingly $f_{t=\lfloor nr \rfloor} = O_{a.s}\left(\sqrt{n}\right)$. Then, if $\frac{\sqrt{n}}{\mu_n} \to 0$, the solution for f satisfies

$$n^4 \Delta^2 \Delta^{*2} f + n^4 \Delta^2 EGE' f = o_{a.s.}(1),$$

so that as $n \to \infty$ in the limit f must lie in the kernel of the operator $n^4 \Delta^2 \Delta^{*2} I_n + n^4 \Delta^2 EGE' = n^4 \Delta^2 \left[\Delta^{*2} I_n + EGE'\right]$ and hence the kernel of the operator Δ^2. That is, if $\lambda = \mu_n n^4$ with $\frac{\sqrt{n}}{\mu_n} \to 0$, f is asymptotically a vector of linear time trends and so

$$\frac{f^{HP}_{t=\lfloor nr \rfloor}}{\sqrt{n}} \to_{a.s.} \alpha_{HP} + \beta_{HP} r, \tag{62}$$

just as in the case where $\lambda \to \infty$ and n is fixed (c.f. Remark 8 and (31)). In a similar way, we find that if $\lambda = \mu_n n^{2m}$ with $\frac{\sqrt{n}}{\mu_n} \to 0$ then

$$\frac{f^{W}_{t=\lfloor nr \rfloor}}{\sqrt{n}} \to_{a.s.} \alpha_W + \beta_{1,W} r + ... + \beta_{m-1,W} r^{m-1}, \tag{63}$$

where the coefficients in the limiting polynomial are given by

$$\begin{bmatrix} \alpha_W \\ \beta_{1,W} \\ \vdots \\ \beta_{m-1,W} \end{bmatrix} = \begin{bmatrix} 1 & \int_0^1 r & \cdots & \int_0^1 r \\ \int_0^1 r & \int_0^1 r^2 & \cdots & \int_0^1 r^2 \\ \vdots & \vdots & \ddots & \vdots \\ \int_0^1 r & \int_0^1 r^2 & \cdots & \int_0^1 r^{2(m-1)} \end{bmatrix}^{-1} \begin{bmatrix} \int_0^1 B(r)\,dr \\ \int_0^1 rB(r)\,dr \\ \vdots \\ \int_0^1 r^{m-1} B(r)\,dr \end{bmatrix}, \tag{64}$$

corresponding to the limiting form of the Whittaker filter when $(n, \lambda)_{\text{seq}} \to \infty$. When the limit process is a continuous stochastic process with piecewise continuous deterministic drift as in (48) rather than a limiting Brownian motion as in (22), (64) continues to hold but with the limit process $B(\cdot)$ replaced by $B_g(\cdot)$.

6 Simulations

We briefly report some simulations that explore the manifestation of unit roots in HP filtered data in finite samples. In particular, let $c^{HP}_{\lambda t} = x_t - \hat{f}^{HP}_{\lambda t}$, where $\hat{f}^{HP}_{\lambda t}$ and $c^{HP}_{\lambda t}$ are the HP fitted trend and cycle for some given λ. We examine evidence for the presence of a unit root in $c^{HP}_{\lambda t}$ in finite samples when the underlying data have a stochastic trend or trend with drift.

As shown in (9) and (10) above, as $\lambda \to \infty$, $\hat{f}^{HP}_{\lambda t} \to (1, t)\left(R_2' R_2\right)^{-1} R_2' x = a + bt$, for some $a = a(X)$ and $b = b(X)$ where $X = (x_t)_{t=1}^n$. This is the case whether or not there is a deterministic trend or trend break in the data. Similarly, from (62), if

$\lambda = \mu_n n^4 \to \infty$ on paths for which $\frac{\sqrt{n}}{\mu_n} \to 0$ as $n \to \infty$, then the standardized fitted trend is again linear and has the form $\frac{f^{HP}_{t=\lfloor nr \rfloor}}{\sqrt{n}} \to_{a.s.} \alpha_{HP} + \beta_{HP} r$, along such joint paths as $(\lambda, n) \to \infty$, which in view of (31) also holds in sequential asymptotics with $(n, \lambda)_{\text{seq}} \to \infty$, and where the coefficients $(\alpha_{HP}, \beta_{HP})$ are given in (29).

Next, suppose that we run a standard UR test from a fitted autoregression with trend on the HP residual series $c^{HP}_{\lambda t}$, viz.,

$$c^{HP}_{\lambda t} = \hat{a} + \hat{b}t + \hat{\theta} c^{HP}_{\lambda t-1} + \hat{u}_t. \tag{65}$$

It is clear that since $\hat{f}^{HP}_{\lambda t} \to a + bt$ we have $c^{HP}_{\lambda t} = x_t - (a(X) + b(X)t)$ as $\lambda \to \infty$, which simply removes a linear trend from x_t irrespective of whether there is a linear trend in the data. Further, suppose that the true model for x_t is a random walk with drift, viz. $x_t = \alpha + \beta t + x^0_t$, where x^0_t is a random walk. Then, for large λ we have

$$c^{HP}_{\lambda t} = x_t - (a(X) + b(X)t) + O\left(\frac{1}{\lambda}\right) = (\alpha - a(X)) + (\beta - b(X))t + X^0_t + O\left(\frac{1}{\lambda}\right).$$

A UR test on $c^{HP}_{\lambda t}$ with fitted trend as in regression (65) is therefore equivalent to a similar UR test on x^0_t for large λ. Hence as $\lambda \to \infty$, a UR test of this type will have rejection probability (of a unit root) equal to the size of the test. This is precisely what the simulations show in Fig. 9. More specifically, the limit theory (62) along the joint path where $\lambda = \mu_n n^4 \to \infty$ with $\frac{\sqrt{n}}{\mu_n} \to 0$, suggests that the UR rejection probability in the fitted cycle $c^{HP}_{\lambda t}$ will eventually equal the size of the test along such joint paths. For each value of n, the empirical rejection rate curves in Fig. 9 are monotically declining as λ increases, just as theory predicts.

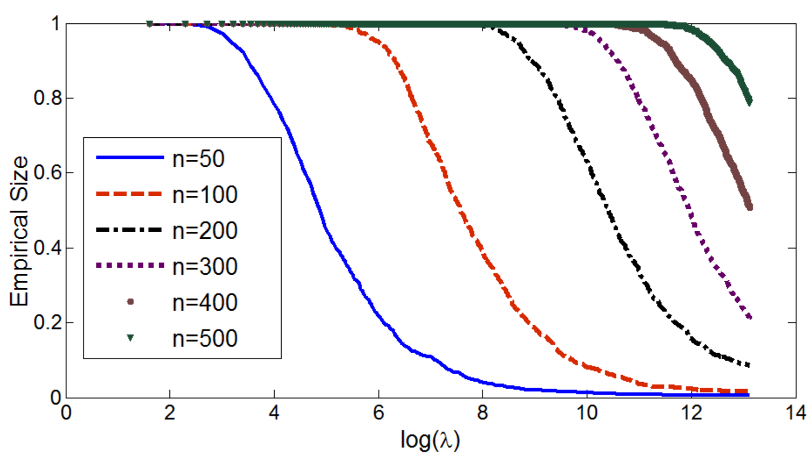


Fig. 9: Empirical size of UR t tests on the residual "cyclical" series $c^{HP}_{\lambda t}$ of a random walk detrended by an HP filter with tuning parameter λ. The nominal size is 0.01. The horizontal scale is measured in logarithms.

7 Conclusion

The Bullard-Krugman debate that was discussed in the Introduction focussed on the measurement of potential output and how this may have been affected by the global financial crisis and the succeeding great recession. That debate focussed attention on the key econometric issue of measuring a latent variable such as potential output that depends critically on the measurement of trend. The debate showed how dramatically economic thinking about the impact of serious shocks like the GFC and major recessions can be influenced by measurement issues. How is it that the measurement of potential output can be so vulnerable to trend elimination methodology?

This paper provides some answers to that important question by analyzing carefully the dependence of trend elimination procedures such as the Whittaker and HP filters. The HP filter is one of the most heavily used econometric methods for measuring business cycles and potential output in empirical research. It is also a smoothing method that belongs to a very general class of nonparametric graduation procedures that depend on a tuning parameter governing the properties of the smoother. As Krugman's position makes clear, long run potential output of an economy can be substantially influenced by great recessions and depressions, which may sufficiently divert resources to impact long run trend components of output. The HP filter has the advantage that, depending on the smoothing parameter (λ) choice, it can encompass long run behavior that encompasses a vast range of possibilities – from a deterministic linear trend, to a smooth Gaussian process, through to stochastic trends and combinations of stochastic trends and deterministic trends that even include trend breaks. However, as our analysis reveals, the processes that lie within the natural capture range of the HP filter depend intimately on the value of the smoothing parameter in relation to the sample size (n). Our results show that a critical expansion rate for λ in terms of n is $O\left(n^4\right)$. Faster rates typically lead to a low order polynomial time trend solution for the HP trend, while slower rates enable the HP trend to capture some features of stochastic as well as deterministic trends and even trend breaks (while still smoothing over the break function).

Like modern nonparametrics, optimal choice of the tuning parameter depends on assumptions about the underlying trend function. If we exclude functions, including stochastic processes, that are differentiable to the 4'th order by insisting on a small smoothness penalty with $\lambda = o\left(n^4\right)$, then the smoother gains the capacity to capture aspects of stochastic trends with random wandering behavior. In that sense, the filter may capture the effects of great recessions and depresssions – as indeed they do in the case of Krugman's Fig. 3 illustration of the use of the HP filter in modeling data around the great depresssion of the 1930s. On the other hand, if we insist on the use of low order polynomial deterministic representation of trend, such as a linear or quadratic time trend to embody long term average growth rates, then the HP filter accommodates such solutions when we insist on a large smoothness penalty by setting $\lambda >> O\left(n^4\right)$ so that $\lambda/n^4 \to \infty$.

It is hoped that this analysis will help to guide empirical work concerned with trend elimination and business cycle research in macroeconomics. It is important, at least, for empirical researchers to be aware that, contrary to current thinking, the

HP filter with a quarterly default setting of $\lambda = 1600$ does not automatically remove unit root stochastic trends in data of sample sizes that commonly arise in practical work. Moreover, it is not so much the value of λ that is important for the material implications of the properties of the filter and induced cycle, but the value of λ in relation to the sample size n.

Economic theory provides empirical researchers with primitive notions about trend that are embodied in steady state growth theories and random wandering processes intended to capture technical change and the operation of efficient markets for foreign exchange, commodities, and stocks. These notions can be used to design smoothing priors, as originally envisaged by Whittaker and Robinson (1924) in their Bayesian formulation of the graduation problem. The modern econometric notion of trend embraces such deterministic and random wandering slowing moving components as well as the potential for intermittent shifts and breaks that lead to more abrupt turning points. While smoothers like the HP filter inevitably 'smooth out' abrupt breaks, it is shown here that they have the capacity to capture most of these different forms of trend. If used with care and with priors that reflect economic thinking about the underlying processes at work in determining latent variables like potential output, our analysis suggests that they may be successfully employed in empirical work to estimate such latent variables in the observed data.

8 Appendix

As in the text, we use the following notation: O_ℓ denotes an $\ell \times \ell$ matrix of zeros, O denotes a zero matrix where the dimensions are clear from the context, and e_j denotes the j'th unit vector with unity in the j'th position and zeros elsewhere.

Proof of Theorem 1 First order conditions and elementary matrix inversion yield

$$\hat{f} = \left(I + \lambda D_m D_m'\right)^{-1} x = \left\{ I - D_m \left[\lambda^{-1} I + D_m' D_m\right]^{-1} D_m' \right\} x,$$

and

$$\begin{aligned} & D_m \left[\lambda^{-1} I + D_m' D_m\right]^{-1} D_m' = D_m \left(D_m' D_m\right)^{-1/2} \left[I + \lambda^{-1} \left(D_m' D_m\right)^{-1}\right]^{-1} \left(D_m' D_m\right)^{-1/2} D_m' \\ = \; & D_m \left(D_m' D_m\right)^{-1} D_m' - D_m \left(D_m' D_m\right)^{-1/2} \left\{ I - \left[I + \lambda^{-1} \left(D_m' D_m\right)^{-1}\right]^{-1} \right\} \left(D_m' D_m\right)^{-1/2} D_m' \\ = \; & D_m \left(D_m' D_m\right)^{-1} D_m' - \left\{ D_m \left(D_m' D_m\right)^{-1/2} \left(I + \lambda D_m' D_m\right)^{-1} \left(D_m' D_m\right)^{-1/2} D_m' \right\}, \end{aligned} \tag{66}$$

since $\left(I + \lambda D_m' D_m\right)^{-1} = I - \left[I + \lambda^{-1} \left(D_m' D_m\right)^{-1}\right]^{-1}$ by direct calculation. Hence

$$\left(I + \lambda D_m D_m'\right)^{-1} = I - D_m \left(D_m' D_m\right)^{-1} D_m' + \left\{ D_m \left(D_m' D_m\right)^{-1/2} \left(I + \lambda D_m' D_m\right)^{-1} \left(D_m' D_m\right)^{-1/2} D_m' \right\}.$$

The stated result follows because $C = \left[D_m \left(D_m' D_m\right)^{-1/2}, R_m \left(R_m' R_m\right)^{-1/2}\right]$ is an orthogonal matrix and $I - D_m \left(D_m' D_m\right)^{-1} D_m' = R_m \left(R_m' R_m\right)^{-1} R_m'$.

Proof of Theorem 2 When n is finite, the following first order conditions hold

$$\begin{aligned}
\frac{\partial}{2\partial f_1}\left\{\sum_{s=1}^{n}(x_s-f_s)^2+\lambda\sum_{s=3}^{n}\left(\Delta^2 f_s\right)^2\right\} &= -(x_1-f_1)+\lambda\left(\Delta^2 f_3\right)=0,\\
\frac{\partial}{2\partial f_2}\left\{\sum_{s=1}^{n}(x_s-f_s)^2+\lambda\sum_{s=3}^{n}\left(\Delta^2 f_s\right)^2\right\} &= -(x_2-f_2)+\lambda\left\{\left(\Delta^2 f_3\right)(-2)+\left(\Delta^2 f_4\right)\right\}=0,\\
\frac{\partial}{2\partial f_t}\left\{\sum_{s=1}^{n}(x_s-f_s)^2+\lambda\sum_{s=3}^{n}\left(\Delta^2 f_s\right)^2\right\} &= -(x_t-f_t)+\lambda\left\{\Delta^2 f_{t+2}-2\Delta^2 f_{t+1}+\Delta^2 f_t\right\}=0\\
\frac{\partial}{2\partial f_{n-1}}\left\{\sum_{s=1}^{n}(x_s-f_s)^2+\lambda\sum_{s=3}^{n}\left(\Delta^2 f_s\right)^2\right\} &= -(x_{n-1}-f_{n-1})+\lambda\left\{\left(\Delta^2 f_{n-1}\right)+\left(\Delta^2 f_n\right)(-2)\right\}=0,\\
\frac{\partial}{2\partial f_n}\left\{\sum_{s=1}^{n}(x_s-f_s)^2+\lambda\sum_{s=3}^{n}\left(\Delta^2 f_s\right)^2\right\} &= -(x_n-f_n)+\lambda\left(\Delta^2 f_n\right)=0.
\end{aligned}$$

In operator form the above equations are

$$\begin{aligned}
\left\{1+\lambda\Delta^2 L^{-2}\right\}f_1 &= x_1,\\
\left\{1+\lambda\Delta^2 L^{-1}\left[-2+L^{-1}\right]\right\}f_2 &= x_2,\\
\left\{1+\lambda\Delta^2\left(1-L^{-1}\right)^2\right\}f_t &= x_t,\ t=3,...,n-2\\
\left\{1+\lambda\Delta^2\left(1-2L^{-1}\right)\right\}f_{n-1} &= x_{n-1}\\
\left\{1+\lambda\Delta^2\right\}f_n &= x_n
\end{aligned}$$

Observe that

$$\begin{aligned}
1+\lambda\Delta^2 L^{-2} &= 1+\lambda\Delta^{*2},\\
1+\lambda\Delta^2 L^{-1}\left(-2+L^{-1}\right) &= 1+\lambda\Delta^{*2}(-1+2\Delta)\\
1+\lambda\Delta^2\left(1-L^{-1}\right)^2 &= 1+\lambda\Delta^2\Delta^{*2}\\
1+\lambda\Delta^2\left(1-2L^{-1}\right) &= 1+\lambda\Delta^2(-1+2\Delta^*)\\
1+\lambda\Delta^2 &= 1+\lambda\Delta^2
\end{aligned}$$

where $\Delta^*=1-L^{-1}$ is the adjoint operator of $\Delta=1-L$ and $\Delta(-L)^{-1}=\Delta^*$ These results are combined in the matrix operator equation

$$d(L)f=x$$

where

$$d(L)=\begin{bmatrix} d_a(L) & O & O\\ O & \left\{1+\lambda\Delta^2\Delta^{*2}\right\}I_{n-4} & O\\ O & O & d_b(L)\end{bmatrix}=\left(1+\lambda\Delta^2\Delta^{*2}\right)\operatorname{diag}\left[O_2,I_{n-4},O_2\right]+EKE'$$

with $E_a=[e_1,e_2]\,,E_b=[e_{n-1},e_n]\,,E=[E_a,E_b]=[e_1,e_2,e_{n-1},e_n]\,,K=\operatorname{diag}[d_a(L)\,,d_b(L)]\,,$
$d_a(L)=\operatorname{diag}\left[1+\lambda\Delta^{*2},1+\lambda\Delta^{*2}(-1+2\Delta)\right]$, and $d_b(L)=\operatorname{diag}\left[1+\lambda\Delta^2(-1+2\Delta^*)\,,1+\lambda\Delta^2\right].$

This gives the first stated result (15). Observe that $d_a(L)$ and $d_b(L)$ can be further decomposed as follows:

$$
\begin{aligned}
d_a(L) &= \operatorname{diag}\left[1+\lambda\Delta^{*2}, 1+\lambda\Delta^{*2}(-1+2\Delta)\right] \\
&= \left(1+\lambda\Delta^2\Delta^{*2}\right) I_2 + \operatorname{diag}\left[\lambda\Delta^{*2}\left(1-\Delta^2\right), \lambda\Delta^{*2}\left\{(-1+2\Delta)-\Delta^2\right\}\right] \\
&= \left(1+\lambda\Delta^2\Delta^{*2}\right) I_2 + \operatorname{diag}\left[\lambda\Delta^{*2}L(2-L), \lambda\Delta^{*2}\left\{-(1-\Delta)^2\right\}\right] \\
&= \left(1+\lambda\Delta^2\Delta^{*2}\right) I_2 + \operatorname{diag}\left[\lambda\Delta^{*2}L(2-L), -\lambda\Delta^{*2}L^2\right] \\
&= \left(1+\lambda\Delta^2\Delta^{*2}\right) I_2 + \operatorname{diag}\left[\lambda\Delta^{*2}L(2-L), -\lambda\Delta^2\right] \\
&=: \left(1+\lambda\Delta^2\Delta^{*2}\right) I_2 + a(L),
\end{aligned}
$$

and

$$
\begin{aligned}
d_b(L) &= \operatorname{diag}\left[1+\lambda\Delta^2(-1+2\Delta^*), 1+\lambda\Delta^2\right] \\
&= \left(1+\lambda\Delta^2\Delta^{*2}\right) I_2 + \operatorname{diag}\left[\lambda\Delta^2\left\{-(1-\Delta^*)^2\right\}, \lambda\Delta^2\left(1-\Delta^{*2}\right)\right] \\
&= \left(1+\lambda\Delta^2\Delta^{*2}\right) I_2 + \operatorname{diag}\left[-\lambda\Delta^{*2}, \lambda\Delta^2 L^{-1}\left(2-L^{-1}\right)\right] \\
&=: \left(1+\lambda\Delta^2\Delta^{*2}\right) I_2 + b(L).
\end{aligned}
$$

Hence, the operator $d(L)$ also has the form

$$
\begin{aligned}
d(L) &= \left\{1+\lambda\Delta^2\Delta^{*2}\right\} I_n + \begin{bmatrix} a(L) & O & O \\ O & O_{n-4} & O \\ O & O & b(L) \end{bmatrix} \\
&= \left\{1+\lambda\Delta^2\Delta^{*2}\right\} I_n + \lambda\Delta^2 \begin{bmatrix} A(L) & O & O \\ O & O_{n-4} & O \\ O & O & B(L) \end{bmatrix}
\end{aligned}
$$

where

$$
\begin{aligned}
a(L) &= \operatorname{diag}\left[\lambda\Delta^{*2}L(2-L), -\lambda\Delta^2\right] = \lambda\Delta^2 \operatorname{diag}\left[\left(2L^{-1}-1\right), -1\right], \\
&=: \lambda\Delta^2 A(L), \quad A(L) = \operatorname{diag}\left[\left(2L^{-1}-1\right), -1\right] \\
b(L) &= \operatorname{diag}\left[-\lambda\Delta^{*2}, \lambda\Delta^2 L^{-1}\left(2-L^{-1}\right)\right] = \lambda\Delta^{*2}\operatorname{diag}\left[-1, (2L-1)\right] \\
&= \lambda\Delta^2 L^{-2}\operatorname{diag}\left[-1, (2L-1)\right] \\
&= \lambda\Delta^2 B(L), \quad B(L) := L^{-2}\operatorname{diag}\left[-1, (2L-1)\right]
\end{aligned}
$$

It follows that we may write the solution as $\hat{f} = d(L)^{-1} x$ with

$$
\begin{aligned}
d(L) &= \left(1+\lambda\Delta^2\Delta^{*2}\right) I_n + \lambda\Delta^2\left\{E_a A(L) E_a' + E_b B(L) E_b'\right\} \\
&= \left(1+\lambda\Delta^2\Delta^{*2}\right)\left(I_n + \frac{\lambda\Delta^2}{1+\lambda\Delta^2\Delta^{*2}} EGE'\right) \\
&= \left(1+\lambda\Delta^2\Delta^{*2}\right)\left\{I_n + \alpha_\lambda(L) EGE'\right\},
\end{aligned}
$$

where $E = [E_a, E_b] = [e_1, e_2, e_{n-1}, e_n]$, $G = \operatorname{diag}[A(L), B(L)]$, and $\alpha_\lambda(L) = \frac{\lambda\Delta^2}{1+\lambda\Delta^2\Delta^{*2}}$, which gives the second result (16).

Proof of (17) The finite dimensional HP filter $\hat{f}$ is the solution of the matrix operator equation

$$\left(1+\lambda\Delta^2\Delta^{*2}\right)\left(I_n+\alpha EGE'\right)\hat{f}=x, \tag{67}$$

where $\alpha=\alpha_\lambda(L), E$, and G are all as given above. Using $(I_n+FF')^{-1}=I_n-F(I+F'F)^{-1}F'$ with $F=\alpha^{1/2}EG^{1/2}$, $\left(I+\frac{1}{\alpha}G^{-1}\right)^{-1}=I-[I+\alpha G]^{-1}$, and noting that $E'E=I_4$, we have

$$\begin{aligned}
&\left[I_n+\alpha EGE'\right]^{-1}=I_n-\alpha EG^{1/2}(I+\alpha G)^{-1}G^{1/2}E'=I_n-E\left(I_4+\frac{1}{\alpha}G^{-1}\right)^{-1}E' \\
=&\ \left(I_n-EE'\right)-E\left\{\left(I_4+\frac{1}{\alpha}G^{-1}\right)^{-1}-I_4\right\}E' \\
=&\ \left(I_n-EE'\right)+E\left(I_4+\alpha G\right)^{-1}E' \\
=&\ \operatorname{diag}\left[O_2,I_{n-4},O_2\right]+E_a\left[I_2+\alpha A(L)\right]^{-1}E_a'+E_b\left[I_2+\alpha B(L)\right]^{-1}E_b',
\end{aligned}$$

Solving (67) we therefore have

$$\begin{aligned}
\hat{f} &= \left(1+\lambda\Delta^2\Delta^{*2}\right)^{-1}\left(I_n+\alpha EGE'\right)^{-1}x \\
&= \left(1+\lambda\Delta^2\Delta^{*2}\right)^{-1}\operatorname{diag}\left[O_2,I_{n-4},O_2\right]x \\
&\quad+\left(1+\lambda\Delta^2\Delta^{*2}\right)^{-1}\left(E_a\left[I_2+\alpha A(L)\right]^{-1}E_a'x+E_b\left[I_2+\alpha B(L)\right]^{-1}E_b'x\right) \\
&= \left(1+\lambda\Delta^2\Delta^{*2}\right)^{-1}\operatorname{diag}\left[O_2,I_{n-4},O_2\right]x \\
&\quad+\left(1+\lambda\Delta^2\Delta^{*2}\right)^{-1}\left(E_a\left[I_2+\alpha_\lambda(L)A(L)\right]^{-1}E_a'x+E_b\left[I_2+\alpha_\lambda(L)B(L)\right]^{-1}E_b'x\right),
\end{aligned}$$

showing that the solution $\left(1+\lambda\Delta^2\Delta^{*2}\right)^{-1}x$ is correct up to the first two and last two elements, which differ via end corrections.

Proof of (18) In the general case where $m\geq 2$, we use the expansion $\Delta^m=(1-L)^m=\sum_{j=0}^m\binom{m}{j}(-L)^{m-j}$ so that

$$\begin{aligned}
\Delta^m f_{m+k} &= \sum_{j=0}^m\binom{m}{j}(-L)^{m-j}f_{m+k}=\sum_{j=0}^m\binom{m}{j}(-1)^{m-j}f_{k+j} \\
\frac{\partial}{\partial f_p}\Delta^m f_{m+k} &= \sum_{j=0}^m\binom{m}{j}(-1)^{m-j}1\{p=j+k\}=\binom{m}{p-k}(-1)^{m+k-p}, \quad \text{for } k\leq p,
\end{aligned}$$

and then the required derivative in the first order conditions is

$$\frac{\partial}{\partial f_p}\left(\Delta^m f_{m+k}\right)^2=\frac{\partial}{\partial f_p}\left(\sum_{j=0}^m\binom{m}{j}(-1)^{m-j}f_{k+j}\right)^2$$

$$= \; 2\left(\sum_{j=0}^{m}\binom{m}{j}(-1)^{m-j}\,f_{k+j}\right)\binom{m}{p-k}(-1)^{m+k-p}$$

$$= \; 2\binom{m}{p-k}(-1)^{m+k-p}\left(\Delta^m f_{m+k}\right), \quad \text{for } k \leq p,$$

which leads to the following explicit forms for the first order conditions:

$$\frac{\partial}{2\partial f_1}\left\{\sum_{s=1}^{n}(x_s - f_s)^2 + \lambda\sum_{s=m+1}^{n}(\Delta^m f_s)^2\right\} = -\left(x_1 - f_1\right) + \lambda\left(\Delta^m f_{m+1}\right)\binom{m}{0}(-1)^m$$

$$= \; -\left(x_1 - f_1\right) + \lambda\left(\Delta^m (-L)^{-m}\right) f_1 = -\left(x_1 - f_1\right) + \lambda\left((-1)^{-m}\,\Delta^{*m}\right) f_1 = 0,$$

$$\frac{\partial}{2\partial f_2}\left\{\sum_{s=1}^{n}(x_s - f_s)^2 + \lambda\sum_{s=m+1}^{n}(\Delta^m f_s)^2\right\}$$

$$= \; -\left(x_2 - f_2\right) + \lambda\Delta^m f_{m+1}\left(\frac{\partial}{\partial f_2}\Delta^m f_{m+1}\right) + \lambda\Delta^m f_{m+2}\left(\frac{\partial}{\partial f_2}\Delta^m f_{m+2}\right)$$

$$= \; -\left(x_2 - f_2\right) + \lambda\left\{\left(\Delta^m f_{m+1}\binom{m}{1}(-1)^{m-1}\right) + \left(\Delta^m f_{m+2}\right)\binom{m}{0}(-1)^m\right\}$$

$$= \; -\left(x_2 - f_2\right) + \lambda\left\{\left(\binom{m}{1}(-1)^{m-1}\right)\Delta^m L^{-(m-1)} f_2 + \binom{m}{0}(-1)^m\,\Delta^m L^{-m} f_2\right\}$$

$$= \; -\left(x_2 - f_2\right) + \lambda\left\{\left(\binom{m}{1}(-1)^{m-1}\right) L\Delta^{*m} f_2 + \binom{m}{0}(-1)^m\,\Delta^{*m} f_2\right\}$$

$$= \; -\left(x_2 - f_2\right) + \lambda\left\{\left(\binom{m}{1}(-1)^{m-1}\right) L + \binom{m}{0}(-1)^m\right\}\Delta^{*m} f_2 = 0,$$

$$\frac{\partial}{2\partial f_m}\left\{\sum_{s=1}^{n}(x_s - f_s)^2 + \lambda\sum_{s=m+1}^{n}(\Delta^m f_s)^2\right\}$$

$$= \; -\left(x_m - f_m\right) + \lambda\sum_{s=m+1}^{n}\frac{\partial}{2\partial f_m}\left(\Delta^m f_s\right)^2$$

$$= \; -\left(x_m - f_m\right) + \lambda\sum_{k=1}^{m}\left(\Delta^m f_{m+k}\right)\frac{\partial}{\partial f_m}\Delta^m f_{m+k}$$

$$= \; -\left(x_m - f_m\right) + \lambda\sum_{k=1}^{m}\left(\Delta^m f_{m+k}\right)\binom{m}{m-k}(-1)^{m+k-m}$$

$$= \; -\left(x_m - f_m\right) + \lambda\sum_{k=1}^{m}\binom{m}{m-k}(-L)^{-k}\left(\Delta^m f_m\right)$$

$$= \quad -x_m + \left[1 + \lambda \sum_{k=1}^{m} \binom{m}{m-k} (-L)^{-k} \Delta^m \right] f_m$$

$$= \quad -x_m + \left[1 + \lambda \left\{ \left(1 - \frac{1}{L}\right)^m - 1 \right\} \Delta^m \right] f_m$$

$$= \quad -x_m + \left[1 + \lambda \left\{ \Delta^m - (-L)^m \right\} \Delta^{*m} \right] f_m,$$

$$\frac{\partial}{2\partial f_t} \left\{ \sum_{s=1}^{n} (x_s - f_s)^2 + \lambda \sum_{s=m+1}^{n} (\Delta^m f_s)^2 \right\} = -(x_t - f_t) + \lambda \sum_{s=m+1}^{n} \frac{\partial}{2\partial f_t} (\Delta^m f_s)^2$$

$$= \quad -(x_t - f_t) + \lambda \sum_{k=1}^{t} \binom{m}{t-k} (-1)^{m+k-t} (\Delta^m f_{m+k}) \quad \text{for } m < t$$

$$= \quad -(x_t - f_t) + \lambda \sum_{q=0}^{m} \binom{m}{q} (-1)^{m-q} (\Delta^m f_{m+t-q}) \quad \text{for } q = t-k = 0, ..., m$$

$$= \quad -(x_t - f_t) + \lambda \sum_{q=0}^{m} \binom{m}{q} (-L)^{-m+q} (\Delta^m f_t) = -x_t + \left[1 + \lambda \sum_{q=0}^{m} \binom{m}{q} (-L)^{-m+q} \Delta^m \right] f_t$$

$$= \quad -x_t + \left[1 + \lambda \left(1 - \frac{1}{L}\right)^m \Delta^m \right] f_t = -x_t + \left[1 + \lambda \Delta^{*m} \Delta^m\right] f_t,$$

$$\frac{\partial}{2\partial f_{n-m+1}} \left\{ \sum_{s=1}^{n} (x_s - f_s)^2 + \lambda \sum_{s=m+1}^{n} (\Delta^m f_s)^2 \right\}$$

$$= \quad -(x_{n-m+1} - f_{n-m+1}) + \lambda \sum_{s=n-m+1}^{n} \frac{\partial}{2\partial f_{n-m+1}} (\Delta^m f_s)^2$$

$$= \quad -(x_{n-m+1} - f_{n-m+1}) + \lambda \sum_{k=1}^{m} (\Delta^m f_{n-m+k}) \frac{\partial}{\partial f_{n-m+1}} \Delta^m f_{n-m+k}$$

$$= \quad -(x_{n-m+1} - f_{n-m+1}) + \lambda \sum_{k=1}^{m} (\Delta^m f_{n-m+k}) \left[\binom{m}{k-1} (-1)^{k-1} \right]$$

$$= \quad -(x_{n-m+1} - f_{n-m+1}) + \lambda \sum_{k=1}^{m} \left[\Delta^m \binom{m}{k-1} (-1)^{k-1} L^{-(k-1)} \right] f_{n-m+1}$$

$$= \quad -(x_{n-m+1} - f_{n-m+1}) + \lambda \sum_{k=1}^{m} \left[\binom{m}{k-1} (-L)^{-(k-1)} \right] \Delta^m f_{n-m+1}$$

$$= \quad -x_{n-m+1} + \left\{ 1 + \lambda \left[\Delta^{*m} - (-L)^{-m} \right] \Delta^m \right\} f_{n-m+1},$$

$$
\begin{aligned}
& \frac{\partial}{2\partial f_{n-1}}\left\{\sum_{s=1}^{n}(x_s-f_s)^2+\lambda\sum_{s=m+1}^{n}(\Delta^m f_s)^2\right\} \\
= & -(x_{n-1}-f_{n-1})+\lambda\sum_{s=n-1}^{n}\frac{\partial}{2\partial f_{n-1}}(\Delta^m f_s)^2 \\
= & -(x_{n-1}-f_{n-1})+\lambda\sum_{k=1}^{2}(\Delta^m f_{n-2+k})\left[\binom{m}{k-1}(-1)^{k-1}\right] \\
= & -(x_{n-1}-f_{n-1})+\lambda\sum_{k=1}^{2}\left(\Delta^m L^{-(k-1)}\right)\left[\binom{m}{k-1}(-1)^{k-1}\right]f_{n-1} \\
= & -(x_{n-1}-f_{n-1})+\lambda\sum_{k=1}^{2}\left[\binom{m}{k-1}(-L)^{-(k-1)}\right]\Delta^m f_{n-1} \\
= & -(x_{n-1}-f_{n-1})+\lambda\left[1+m(-L)^{-1}\right]\Delta^m f_{n-1},
\end{aligned}
$$

and finally

$$
\begin{aligned}
\frac{\partial}{2\partial f_n}\left\{\sum_{s=1}^{n}(x_s-f_s)^2+\lambda\sum_{s=m+1}^{n}(\Delta^m f_s)^2\right\} &= -(x_{n-1}-f_{n-1})+\lambda\frac{\partial}{2\partial f_n}(\Delta^m f_n)^2 \\
&= -(x_{n-1}-f_{n-1})+\lambda(\Delta^m f_n) \\
&= -x_{n-1}-f_{n-1}+\lambda(\Delta^m f_n),
\end{aligned}
$$

giving the full operator system

$$
\begin{aligned}
\left\{1+\lambda(-1)^{-m}\Delta^{*m}\right\}f_1 &= x_1, \qquad (68)\\
\left\{1+\lambda\left[\binom{m}{1}(-1)^{m-1}L+\binom{m}{0}(-1)^m\right]\Delta^{*m}\right\}f_2 &= x_2, \\
&\vdots \\
\{1+\lambda[\Delta^m-(-L)^m]\Delta^{*m}\}f_m &= x_m, \\
[1+\lambda\Delta^{*m}\Delta^m]f_t &= x_t, \quad t=m+1,...,n-m, \\
\left\{1+\lambda\left[\Delta^{*m}-(-L)^{-m}\right]\Delta^m\right\}f_{n-m+1} &= x_{n-m+1}, \\
&\vdots \\
\left\{1+\lambda\left[\binom{m}{1}(-L)^{-1}+\binom{m}{0}\right]\Delta^m\right\}f_{n-1} &= x_{n-1}, \\
\{1+\lambda\Delta^m\}f_n &= x_n. \qquad (69)
\end{aligned}
$$

The system can be written in matrix form as follows:

$$
d_m(L)=(1+\lambda\Delta^m\Delta^{*m})\operatorname{diag}[O_m,I_{n-2m},O_m]+E_mK_mE_m,
$$

with $E_m=[E_{ma},E_{mb}]$, $E_{ma}=[e_1,..,e_m]$, $E_{mb}=[e_{n-m},...,e_n]$, $K_m=\operatorname{diag}[A_m(L),B_m(L)]$ where

$$
\begin{aligned}
A_m(L) &= \operatorname{diag}\left[1+\lambda(-1)^{-m}\Delta^{*m},..,1+\lambda[\Delta^m-(-L)^m]\Delta^{*m}\right], \\
B_m(L) &= \operatorname{diag}\left[1+\lambda\left[\Delta^{*m}-(-L)^{-m}\right]\Delta^m,..,1+\lambda\Delta^m\right],
\end{aligned}
$$

and the remaining entries of the diagonal matrices $A_m(L)$ and $B_m(L)$ follow the combinatoric scheme given in the operator system (68) - (69) above.

Proof of Theorem 3 We use the operator form of the filter (11), which governs its asymptotic behavior as is evident from formulae (15) and (17) showing that the operator takes this form except for the end corrections. Scaling (11) by $\sqrt{n}$ and writing $X_n(r) = n^{-1/2}x_{\lfloor nr\rfloor}$ we have from (21)

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

Since the Karhunen Loève (KL) series representation 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 $\sup_{0\le r\le 1}\left|B^{K_n}(r) - B(r)\right| = o_{a.s.}(1)$. Then

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

if $K_n\to\infty$ as $n\to\infty$. It follows that the HP trend solution has the following approximate form as $n\to\infty$

$$\begin{aligned}\frac{\hat{f}_t}{\sqrt{n}} &= \frac{1}{\lambda L^{-2}(1-L)^4+1}\frac{x_t}{\sqrt{n}} = \frac{1}{\lambda L^{-2}(1-L)^4+1}\left[B^{K_n}\left(\frac{t}{n}\right) + o_{a.s.}(1)\right]\\ &= \sum_{k=1}^{K_n}\sqrt{\lambda_k}\left[\frac{1}{\lambda L^{-2}(1-L)^4+1}\varphi_k\left(\frac{t}{n}\right)\right]\xi_k + o_{a.s.}(1) \qquad (70)\end{aligned}$$

Noting that $\sqrt{\lambda_k} = 1/\left\{\left(k-\frac{1}{2}\right)\pi\right\}$ and $\varphi_k\left(\frac{t}{n}\right) = \sqrt{2}\,\mathrm{Im}\left(e^{\frac{it/n}{\sqrt{\lambda_k}}}\right)$, we have

$$\begin{aligned}n(1-L)\varphi_k\left(\frac{t}{n}\right) &= \sqrt{2}\,\mathrm{Im}\left\{e^{\frac{it/n}{\sqrt{\lambda_k}}}n\left(1-e^{-\frac{i}{n\sqrt{\lambda_k}}}\right)\right\}\\ &= \frac{\sqrt{2}}{\sqrt{\lambda_k}}\mathrm{Im}\left\{e^{\frac{it/n}{\sqrt{\lambda_k}}}\left[\frac{1-\cos\left(\frac{1}{n\sqrt{\lambda_k}}\right)}{\frac{1}{n\sqrt{\lambda_k}}} + i\frac{\sin\left(\frac{1}{n\sqrt{\lambda_k}}\right)}{\frac{1}{n\sqrt{\lambda_k}}}\right]\right\}\\ &= \frac{\sqrt{2}}{\sqrt{\lambda_k}}\mathrm{Im}\left\{e^{\frac{it/n}{\sqrt{\lambda_k}}}\left[O\left(\frac{1}{n\sqrt{\lambda_k}}\right) + i\left(1+O\left(\frac{1}{n^2\lambda_k}\right)\right)\right]\right\}\\ &= \frac{\sqrt{2}}{\sqrt{\lambda_k}}\mathrm{Im}\left\{e^{\frac{it/n}{\sqrt{\lambda_k}}}\left[O\left(\frac{K_n}{n}\right) + i\left(1+O\left(\frac{K_n^2}{n^2}\right)\right)\right]\right\}\\ &= \frac{\sqrt{2}}{\sqrt{\lambda_k}}\left\{\mathrm{Re}\left[e^{\frac{it/n}{\sqrt{\lambda_k}}}\left[1+O\left(\frac{K_n^2}{n^2}\right)\right]\right] + \mathrm{Im}\left[e^{\frac{it/n}{\sqrt{\lambda_k}}}\times O\left(\frac{K_n}{n}\right)\right]\right\},\end{aligned}$$

uniformly for $k\le K_n$ and $t\le n$. Also

$$\begin{aligned}nL^{-1}(1-L)\varphi_k\left(\frac{t}{n}\right) &= \frac{\sqrt{2}}{\sqrt{\lambda_k}}\left\{\mathrm{Re}\left[e^{\frac{it/n}{\sqrt{\lambda_k}}}\left[1+O\left(\frac{K_n^2}{n^2}\right)\right]\right] + \mathrm{Im}\left[e^{\frac{it/n}{\sqrt{\lambda_k}}}\times O\left(\frac{K_n}{n}\right)\right]\right\}\\ &= \frac{\sqrt{2}}{\sqrt{\lambda_k}}\left\{\mathrm{Re}\left[e^{\frac{it/n}{\sqrt{\lambda_k}}}\left[1+O\left(\frac{K_n^2}{n^2}\right)\right]\right] + \mathrm{Im}\left[e^{\frac{it/n}{\sqrt{\lambda_k}}}O\left(\frac{K_n}{n}\right)\right]\right\}.\end{aligned}$$

By repeated argument we find that

$$\begin{aligned}
L^{-2}\left[n\left(1-L\right)\right]^{4}\varphi_{k}\left(\frac{t}{n}\right) &= \sqrt{2}\,\mathrm{Im}\left\{\left(\frac{i}{\sqrt{\lambda_k}}\right)^{4}e^{\frac{it/n}{\sqrt{\lambda_k}}}\left[1+O\left(\frac{K_n^2}{n^2}\right)\right]+\left(\frac{1}{\sqrt{\lambda_k}}\right)^{4}i^{3}e^{\frac{it/n}{\sqrt{\lambda_k}}}\times O\left(\frac{K_n}{n}\right)\right\} \\
&= \frac{\sqrt{2}}{\lambda_k^2}\,\mathrm{Im}\left\{e^{\frac{it/n}{\sqrt{\lambda_k}}}\left[1+i\times O\left(\frac{K_n}{n}\right)+O\left(\frac{K_n^2}{n^2}\right)\right]\right\} \\
&= \frac{\sqrt{2}}{\lambda_k^2}\left\{\mathrm{Im}\left[e^{\frac{it/n}{\sqrt{\lambda_k}}}\left[1+O\left(\frac{K_n^2}{n^2}\right)\right]\right]+\mathrm{Re}\left[e^{\frac{it/n}{\sqrt{\lambda_k}}}\right]\times O\left(\frac{K_n}{n}\right)\right\},
\end{aligned}$$

uniformly for $k \leq K_n$. Thus, the operator $n\left(1-L\right)$ applied to $\varphi_k\left(\frac{t}{n}\right)$ acts asymptotically like the differential operator $D = d/dx$ on $\varphi_k(x)$ and L^{-1} acts asymptotically like the identity. Moreover, well behaved nonlinear functions of $n\left(1-L\right)$ and L^{-1} act asymptotically like the same nonlinear functions of D and the identity. For instance,

$$g\left(n\left(1-L\right),L^{-1}\right)e^{a\frac{t}{n}} = \left[g\left(D,1\right)e^{ax}+o\left(1\right)\right]_{x=\frac{t}{n}} = g\left(a,1\right)e^{a\frac{t}{n}}+o\left(1\right), \tag{71}$$

where $g(D,1) = h\left(D\right)$ is treated as a pseudodifferential operator (e.g., Treves, 1980). A formal justification of (71) uses the Fourier integral representation $h\left(x\right) = \frac{1}{2\pi}\int_{-\infty}^{\infty}e^{ixy}\tilde{h}\left(y\right)dy$ of h in terms of its Fourier transform $\tilde{h}$, so that

$$\begin{aligned}
h\left(D\right)e^{ax} &= \frac{1}{2\pi}\int_{-\infty}^{\infty}e^{iDy}e^{ax}\tilde{h}\left(y\right)dy = \frac{1}{2\pi}\int_{-\infty}^{\infty}e^{a(x+iy)}\tilde{h}\left(y\right)dy \\
&= e^{ax}\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{iay}\tilde{h}\left(y\right)dy = e^{ax}h\left(a\right).
\end{aligned}$$

Now suppose that $\lambda = \mu n^4$ for some $\mu > 0$. The operation in square parentheses in (70) can be evaluated for each term using this argument as follows[14]

$$\begin{aligned}
&\left[\frac{1}{\lambda L^{-2}(1-L)^4+1}\varphi_k\left(\frac{t}{n}\right)\right] = \left[\frac{1}{\mu L^{-2}\left[n(1-L)\right]^4+1}\varphi_k\left(\frac{t}{n}\right)\right] \\
&= \sqrt{2}\,\mathrm{Im}\left[\frac{1}{\mu L^{-2}\left[n(1-L)\right]^4+1}e^{\frac{it/n}{\sqrt{\lambda_k}}}\right] \\
&= \sqrt{2}\,\mathrm{Im}\left[\frac{1}{\mu\left[\frac{i}{\sqrt{\lambda_k}}\right]^4+1}e^{\frac{it/n}{\sqrt{\lambda_k}}}\left\{1+i\times O\left(\frac{K_n}{n}\right)+O\left(\frac{K_n^2}{n^2}\right)\right\}\right]
\end{aligned}$$

[14] In differential form the operator calculation can be performed as

$$\begin{aligned}
&\frac{1}{\mu D^4+1}\varphi_k\left(x\right) = \int_0^{\infty}e^{-\left\{\mu D^4+1\right\}s}\varphi_k\left(x\right)ds = \mathrm{Im}\left[\int_0^{\infty}e^{-s-s\mu D^4}e^{\frac{ix}{\sqrt{\lambda_k}}}ds\right] \\
&= \mathrm{Im}\left[\int_0^{\infty}e^{-s-s\mu\left(i/\sqrt{\lambda_k}\right)^4}e^{\frac{ix}{\sqrt{\lambda_k}}}ds\right] = \mathrm{Im}\left[\frac{1}{1+\mu/\lambda_k^2}e^{\frac{ix}{\sqrt{\lambda_k}}}\right] \\
&= \frac{\lambda_k^2}{\mu+\lambda_k^2}\varphi_k\left(x\right).
\end{aligned}$$

$$= \sqrt{2}\frac{\lambda_k^2}{\mu+\lambda_k^2}\sin\left(\frac{t/n}{\sqrt{\lambda_k}}\right)(1+o(1))$$

$$= \frac{\lambda_k^2}{\mu+\lambda_k^2}\varphi_k\left(\frac{t}{n}\right)(1+o(1)), \tag{72}$$

with the error magnitude holding uniformly for $k \le K_n$ and $K_n/n = o(1)$. Using (72) in (70), we deduce that the asymptotic form of the HP filter can be written as

$$\frac{\hat{f}_{t,K_n}}{\sqrt{n}} = \sum_{k=1}^{K_n}\frac{\lambda_k^{5/2}}{\mu+\lambda_k^2}\varphi_k\left(\frac{t}{n}\right)\xi_k\left\{1+o(1)\right\}. \tag{73}$$

Observe that for $\mu > 0$ the coefficients $\frac{\lambda_k^{5/2}}{\mu+\lambda_k^2} = O\left(k^{-5}\right)$ and so the series $\sum_{k=1}^{\infty}\frac{\lambda_k^{5/2}}{\mu+\lambda_k^2}\varphi_k\left(\frac{t}{n}\right)\xi_k$ converges uniformly and almost surely as $K_n \to \infty$. Hence, when $K_n \to \infty$ as $n \to \infty$ with $\frac{K_n}{n} \to 0$, we have the asymptotic representation of the HP filter trend solution in the case of an $I(1)$ process x_t as

$$\frac{\hat{f}_t}{\sqrt{n}} = \sum_{k=1}^{\infty}\frac{\lambda_k^{5/2}}{\mu+\lambda_k^2}\varphi_k\left(\frac{t}{n}\right)\xi_k + o_{a.s.}(1). \tag{74}$$

The continuous limit form of the HP filter applied to the stochastic trend x_t is therefore

$$f_{HP}(r) = \sum_{k=1}^{\infty}\frac{\lambda_k^{5/2}}{\mu+\lambda_k^2}\varphi_k(r)\,\xi_k,$$

as given in (27).

The corresponding result for the Whittaker filter follows in a similar fashion. In particular, in place of (72) we have the operation

$$\left[\frac{1}{\lambda\Delta^{*m}\Delta^m+1}\varphi_k\left(\frac{t}{n}\right)\right] = \left[\frac{1}{\mu L^{-2m}\left[n(1-L)\right]^{2m}+1}\varphi_k\left(\frac{t}{n}\right)\right]$$

$$= \sqrt{2}\,\mathrm{Im}\left[\frac{1}{\mu L^{-2m}\left[n(1-L)\right]^{2m}+1}e^{\frac{it/n}{\sqrt{\lambda_k}}}\right] = \sqrt{2}\,\mathrm{Im}\left[\frac{1}{\mu\left[\frac{i}{\sqrt{\lambda_k}}\right]^{2m}+1}e^{\frac{it/n}{\sqrt{\lambda_k}}}\left\{1+o(1)\right\}\right]$$

$$= \sqrt{2}\frac{\lambda_k^{2m}}{\mu+\lambda_k^{2m}}\sin\left(\frac{t/n}{\sqrt{\lambda_k}}\right)\left\{1+o(1)\right\} = \frac{\lambda_k^{2m}}{\mu+\lambda_k^{2m}}\varphi_k\left(\frac{t}{n}\right)\left\{1+o(1)\right\}, \tag{75}$$

Thus, in the same way as the HP filter, letting $K_n \to \infty$ such that $\frac{K_n}{n} \to 0$, we find

$$\frac{\hat{f}_t}{\sqrt{n}} = \sqrt{2}\sum_{k=1}^{\infty}\frac{\lambda_k^m}{\mu+\lambda_k^m}\sqrt{\lambda_k}\sin\left(\frac{t/n}{\sqrt{\lambda_k}}\right)\xi_k + o_{a.s.}(1).$$

In continuous form, the limiting trend process is therefore

$$f_W(r) = \sqrt{2}\sum_{k=1}^{\infty}\frac{\lambda_k^m}{\mu+\lambda_k^m}\sqrt{\lambda_k}\sin\left(\frac{r}{\sqrt{\lambda_k}}\right)\xi_k = \sum_{k=1}^{\infty}\frac{\lambda_k^{m+1/2}}{\mu+\lambda_k^m}\varphi_k(r)\,\xi_k, \tag{76}$$

as stated in (28).

Proof of Theorem 4 We again use the operator form of the filter (11) to examine asymptotic behavior. Scaling (11) by $\sqrt{n}$ we can use the earlier result (70) obtained for the component x_t^0 in the proof of Theorem 3. Thus, if $K_n \to \infty$ as $n \to \infty$, we write

$$\begin{aligned}\frac{1}{\lambda L^{-2}(1-L)^4+1}\frac{x_t^0}{\sqrt{n}} &= \sum_{k=1}^{K_n}\sqrt{\lambda_k}\left[\frac{1}{\lambda L^{-2}(1-L)^4+1}\varphi_k\left(\frac{t}{n}\right)\right]\xi_k + o_{a.s.}(1) \\ &= \sum_{k=1}^{K_n}\frac{\lambda_k^{5/2}}{\mu+\lambda_k^2}\varphi_k\left(\frac{t}{n}\right)\xi_k + o_{a.s.}(1) \qquad (77)\end{aligned}$$

Next, using the pseudo-differential integral form

$$\frac{1}{\lambda L^{-2}(1-L)^4+1} = \int_0^\infty \exp\left\{-\left[\lambda L^{-2}(1-L)^4+1\right]s\right\}ds \qquad (78)$$

of the operator $\left\{\lambda L^{-2}(1-L)^4+1\right\}^{-1}$, we find that

$$\begin{aligned}\frac{1}{\lambda L^{-2}(1-L)^4+1}\left\{\alpha+\beta\frac{t}{n}\right\} &= \int_0^\infty \exp\left\{-\left[\lambda L^{-2}(1-L)^4+1\right]s\right\}ds\left[\alpha+\beta\frac{t}{n}\right] \\ &= \int_0^\infty \exp\{-s\}\,ds\left[\alpha+\beta\frac{t}{n}\right] = \alpha+\beta\frac{t}{n}, \qquad (79)\end{aligned}$$

since $\lambda^j L^{-2j}(1-L)^{4j}\left(\alpha+\beta\frac{t}{n}\right)=0$ for all $\lambda>0$ and all $j=1,2,\ldots$. The invariance result (79) also follows immediately from the finite sample representation (7) since the projection operator $R_2\left(R_2'R_2\right)^{-1}R_2'$ is the identity operator on a linear time trend and the differencing operator D_2' eliminates a linear time trend. Note that when $n\to\infty$, the operator $1/(\lambda L^{-2}(1-L)^4+1)$ also preserves polynomials of degree 3 since $\lambda^j L^{-2j}(1-L)^{4j}\left(\frac{t}{n}\right)^3=0$ for all $j\geq 1$. Combining (77) and (79) we have the following approximation to the HP filter

$$\frac{\hat{f}_{t,K_n}^{HP}}{\sqrt{n}} = \alpha+\beta\frac{t}{n}+\sum_{k=1}^{K_n}\frac{\lambda_k^{5/2}}{\mu+\lambda_k^2}\varphi_k\left(\frac{t}{n}\right)\xi_k$$

so that as $K_n\to\infty$, with $K_n/n\to 0$ and $\lambda=\mu n^4$ as $n\to\infty$,

$$\frac{\hat{f}_{t=\lfloor nr\rfloor,K_n}^{HP}}{\sqrt{n}} \to_{a.s.} f_{HP}(r) = \alpha+\beta r+\sum_{k=1}^{\infty}\frac{\lambda_k^{5/2}}{\mu+\lambda_k^2}\varphi_k(r)\,\xi_k,$$

as stated for the HP filter in (45). The same proof applies to the Whittaker filter using the projection invariance of the operator $R_m\left(R_m'R_m\right)^{-1}R_m'$ on polynomial time trends of degree $m-1$. The resulting limiting form of the filter is

$$\frac{\hat{f}_{\lfloor nr\rfloor}^{W}}{\sqrt{n}} \to_{a.s.} f_W(r) = \alpha+\beta_1 r+\ldots\beta_J r^J+\sum_{k=1}^{\infty}\frac{\lambda_k^m}{\mu+\lambda_k^m}\sqrt{\lambda_k}\varphi_k(r)\,\xi_k,$$

as given in (46) for all $J\leq m-1$. In fact, when $n\to\infty$ the dominant asymptotic operator, $1/(\lambda L^{-m}(1-L)^{2m}+1)$, of the Whittaker filter preserves polynomials of degree $2m-1$ since $\lambda^j L^{-mj}(1-L)^{2mj}\left(\frac{t}{n}\right)^J=0$ for all $j\geq 1$ and $J\leq 2m-1$.

Proof of Theorem 5 It is sufficient to work with the case of the Whittaker filter with $m \geq 2$. The data are generated according to $x_t = g_n(t) + x_t^0$ where x_t^0 satisfies the functional law (22) and $g_n(t)$ is a polynomial or a piecewise smooth function with a finite number of break points and sample size dependent coefficients such that $n^{-1/2} g_n(t = \lfloor nr \rfloor) \rightarrow g(r)$, uniformly in $r \in [0,1]$, whose limit $g(r)$ is a piecewise smooth function. Define $\tilde{g}_n(t)$ as the residual function in the relation

$$g_n(t) = \alpha_n + \beta_{n,1} t + \ldots + \beta_{n,2m-1} t^{2m-1} + \tilde{g}_n(t) := p_{n,2m-1}(t) + \tilde{g}_n(t)$$

where $\frac{\alpha_n}{\sqrt{n}} \rightarrow \alpha$ and $n^{j-\frac{1}{2}} \beta_{n,j} \rightarrow \beta_j$ for $j = 1, .., m-1$, and $n^{-1/2} \tilde{g}_n(t = \lfloor nr \rfloor) \rightarrow \tilde{g}(r)$ uniformly in r. Then $g(r)$ and $\tilde{g}(r)$ are piecewise smooth functions on $r \in [0,1]$ that differ by a polynomial of degree $2m-1$. The Fourier series representation of $\tilde{g}(r)$ in terms of the complex orthonormal sequence $\left\{ (2\pi)^{-1/2} e^{ikr} \right\}$ and Fourier coefficients $\tilde{c}_k = \int_{-\pi}^{\pi} e^{ikr} \tilde{g}(r) \, dr$ is

$$\tilde{g}(r) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \tilde{c}_k e^{-ikr} = \frac{\tilde{c}_0}{2\pi} + \frac{1}{\pi} \sum_{k=1}^{\infty} \operatorname{Re}\left[\tilde{c}_k e^{-ikr} \right], \quad r \in [-\pi, \pi], \tag{80}$$

which converges pointwise over $r \in [0,1]$ and converges to $\tilde{g}(r)$ everywhere except at the (finite number of) break points. Integral operators on $\tilde{g}$ such as $\tilde{G}(r) = \int_0^r \tilde{g}(s) \, ds$ are everywhere smooth on $[0,1]$ and therefore have Fourier series that are pointwise convergent to $G(r)$ and that are given by the termwise integrals of the Fourier series of $\tilde{g}$. Define the following function based on K terms of the series in (80)

$$\tilde{g}^K(r) = \frac{\tilde{c}_0}{2\pi} + \frac{1}{\pi} \sum_{k=1}^{K} \operatorname{Re}\left[\tilde{c}_k e^{-ikr} \right], \quad r \in [-\pi, \pi], \tag{81}$$

and the corresponding function for finite samples

$$\frac{1}{\sqrt{n}} \tilde{g}_n^K(r) = \frac{\tilde{c}_{n0}}{2\pi} + \frac{1}{\pi} \sum_{k=1}^{K} \operatorname{Re}\left[\tilde{c}_{nk} e^{-ikr} \right], \tag{82}$$

where $\sup_K \max_{k \leq K} |\tilde{c}_{nk} - \tilde{c}_k| \rightarrow 0$ as $n \rightarrow \infty$. Proceeding as in (75) we have

$$\begin{aligned} \left[\frac{1}{\lambda \Delta^{*m} \Delta^m + 1} e^{-ikt/n} \right] &= \left[\frac{1}{\mu L^{-2m} \left[n(1-L) \right]^{2m} + 1} e_k^{-ikt/n} \right] \\ &= \left[\frac{1}{\mu \left[-ik \right]^{2m} + 1} e_k^{-ikt/n} \left\{ 1 + o(1) \right\} \right], \end{aligned}$$

uniformly in $k \leq K$. Termwise application of the smoothing operator $\left\{ \lambda \Delta^{*m} \Delta^m + 1 \right\}^{-1}$ with $\lambda = \mu n^{2m}$ in the Fourier series (82) leads to

$$\begin{aligned} \frac{1}{\lambda \Delta^{*m} \Delta^m + 1} \left\{ \frac{\tilde{g}_n^K(t)}{\sqrt{n}} \right\} \Bigg|_{t = \lfloor nr \rfloor} &= \frac{1}{\lambda \Delta^{*m} \Delta^m + 1} \left[\frac{\tilde{c}_{n0}}{2\pi} + \frac{1}{\pi} \sum_{k=1}^{K} \operatorname{Re}\left\{ \tilde{c}_{nk} e^{-ik\frac{t}{n}} \right\} \right] \Bigg|_{t = \lfloor nr \rfloor} \\ &= \frac{\tilde{c}_{n0}}{2\pi} + \frac{1}{\pi} \sum_{k=1}^{K} \operatorname{Re}\left\{ \tilde{c}_{nk} \left[\frac{1}{\lambda \Delta^{*m} \Delta^m + 1} e^{-ik\frac{t}{n}} \right]_{t = \lfloor nr \rfloor} \right\} \end{aligned}$$

$$= \frac{\tilde{c}_{n0}}{2\pi} + \frac{1}{\pi}\sum_{k=1}^{K} \mathrm{Re}\left\{ \tilde{c}_{nk} \left[\frac{1}{\mu(-ik)^{2m}+1} e^{-ik\frac{t}{n}} \right]_{t=\lfloor nr \rfloor} \{1+o(1)\} \right\}$$

$$= \frac{\tilde{c}_{n0}}{2\pi} + \frac{1}{\pi}\sum_{k=1}^{K} \frac{1}{\mu k^{2m}+1} \mathrm{Re}\left\{ \tilde{c}_{nk} e^{-ikr} \right\} \{1+o(1)\},$$

which converges as $K, n \to \infty$ with $\frac{K}{n} \to 0$ to

$$\frac{\tilde{c}_0}{2\pi} + \frac{1}{\pi}\sum_{k=1}^{\infty} \frac{1}{\mu k^{2m}+1} \mathrm{Re}\left\{ \tilde{c}_k e^{-ik\frac{t}{n}} \right\}.$$

Then

$$\frac{1}{\lambda \Delta^{*m}\Delta^m + 1}\left\{ \frac{\tilde{g}_n(t)}{\sqrt{n}} \right\}\Bigg|_{t=\lfloor nr \rfloor} \to \frac{\tilde{c}_0}{2\pi} + \frac{1}{\pi}\sum_{k=1}^{\infty} \frac{1}{1+\mu k^{2m}} \mathrm{Re}\left\{ \tilde{c}_k e^{-ikr} \right\}. \tag{83}$$

Analogous to (79) we have

$$\frac{1}{\lambda L^{-m}(1-L)^{2m}+1}\left\{ \alpha + \beta_1 \frac{t}{n} + \ldots + \beta_{2m-1}\left(\frac{t}{n}\right)^{2m-1} \right\}$$

$$= \int_0^{\infty} \exp\left\{ -\left[\lambda L^{-m}(1-L)^{2m}+1 \right] s \right\} ds \left[\alpha + \beta_1 \frac{t}{n} + \ldots + \beta_{2m-1}\left(\frac{t}{n}\right)^{2m-1} \right]$$

$$= \alpha + \beta_1 \frac{t}{n} + \ldots + \beta_{2m-1}\left(\frac{t}{n}\right)^{2m-1}, \tag{84}$$

so that polynomials to degree $J \leq 2m-1$) are preserved under the filter $1/\left(\lambda \Delta^{*m}\Delta^m + 1\right)$. Combining (83) and (84) with the earlier result (76)[15] we have

$$\frac{\hat{f}^W_{\lfloor nr \rfloor}}{\sqrt{n}} = \frac{1}{\lambda \Delta^{*m}\Delta^m + 1}\left\{ \frac{g_n(t) + x_t^0}{\sqrt{n}} \right\}\Bigg|_{t=\lfloor nr \rfloor}$$

$$= \frac{1}{\lambda \Delta^{*m}\Delta^m + 1}\left\{ \frac{p_{n,m-1}(t) + \tilde{g}_n(t) + x_t^0}{\sqrt{n}} \right\}\Bigg|_{t=\lfloor nr \rfloor} \to_{a.s.} f_W(r)$$

where

$$f_W(r) = p_{2m-1}(r) + \left\{ \frac{\tilde{c}_0}{2\pi} + \frac{1}{\pi}\sum_{k=1}^{\infty} \frac{1}{1+\mu k^{2m}} \mathrm{Re}\left\{ \tilde{c}_k e^{-ikr} \right\} \right\} + \sum_{k=1}^{\infty} \frac{\lambda_k^{m+1/2}}{\mu + \lambda_k^m} \varphi_k(r)\,\xi_k. \tag{85}$$

[15] Note that the earlier interpolating class $\mathbb{S}_\varphi = \left\{ \sum_{k=1}^{\infty} d_k \varphi_k(r) : \sum_{k=1}^{\infty} d_k^2 < \infty \right\} \subset \mathbb{L}_2[0,1]$ spanned by the ON functions $\left\{ \varphi_k(r) = \sqrt{2}\sin\left(r/\sqrt{\lambda_k}\right) \right\}_1^{\infty}$ is subsumed within the general trigonometric class $\mathbb{S}_\psi = \left\{ \sum_{k=-\infty}^{\infty} d_k \psi_k(r) : \sum_{k=1}^{\infty} d_k^2 < \infty \right\} \subset \mathbb{L}_2[-\pi,\pi]$ with $\psi_k(r) = \left\{ (2\pi)^{-1/2} e^{ikr} \right\}_{-\infty}^{\infty}$. Hence, an analogous result to (85) is obtained using the basis functions $\psi_k(r) = (2\pi)^{-1/2} e^{ikr}$.

as given in (51). The leading term of (85) is the polynomial $p_{2m-1}(r) = \alpha + \beta_1 r + ... + \beta_{2m-1} r^{2m-1}$, which remains invariant asymptotically under the filter. The term in braces represents the effect of the filter on the residual deterministic drift function $\tilde{g}(r)$, complete with whatever break points occur in $\tilde{g}(r)$. The final component on the right side represents the effect of the filter on the stochastic trend, as studied earlier. The series (85) converges uniformly and almost surely for all $\mu \neq 0$. The result includes polynomials of degree $K \geq 2m$ and trend breaks.

Proof of (57) and (58) We derive the explicit formulae in case of the linear trend break function $g(r) = (\alpha_1 + \beta_1 r) 1\{r < r_0\} + (\alpha_2 + \beta_2 r) 1\{r \geq r_0\}$. The Fourier series of $g(r)$ is

$$g(r) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} c_k e^{-ikr} = \frac{c_0}{2\pi} + \frac{1}{2\pi} \sum_{k=1}^{\infty} \left\{ c_k e^{-ikr} + \bar{c}_k e^{ikr} \right\} = \frac{c_0}{2\pi} + \frac{1}{\pi} \sum_{k=1}^{\infty} \mathrm{Re} \left\{ c_k e^{-ikr} \right\}, \tag{86}$$

with coefficients

$$\begin{aligned} c_0 &= \int_{-\pi}^{r_0} (\alpha_1 + \beta_1 r)\, dr + \int_{r_0}^{\pi} (\alpha_2 + \beta_2 r)\, dr \\ &= \alpha_1(\pi + r_0) + \frac{1}{2}\beta_1(r_0^2 - \pi^2) + \alpha_2(\pi - r_0) + \frac{1}{2}\beta_2(-r_0^2 + \pi^2) \\ &= \pi(\alpha_1 + \alpha_2) + r_0(\alpha_1 - \alpha_2) + \frac{1}{2}\left(\pi^2 - r_0^2\right)(\beta_2 - \beta_1), \end{aligned} \tag{87}$$

and, for $k \geq 1$,

$$\begin{aligned} c_k &= \int_{-\pi}^{\pi} e^{ikr} g(r)\, dr = \int_{-\pi}^{r_0} e^{ikr} (\alpha_1 + \beta_1 r)\, dr + \int_{r_0}^{\pi} e^{ikr} (\alpha_2 + \beta_2 r)\, dr \\ &= \alpha_1 \frac{e^{ikr_0} - e^{-ik\pi}}{ik} + \beta_1 \left\{ \left[\frac{e^{ikr_0}}{ik} r_0 + \frac{e^{-ik\pi}}{ik} \pi \right] - \frac{e^{ikr_0} - e^{-ik\pi}}{(ik)^2} \right\} \\ &\quad + \alpha_2 \frac{e^{ik\pi} - e^{ikr_0}}{ik} + \beta_2 \left\{ \left[\frac{e^{ik\pi}}{ik} \pi - \frac{e^{ikr_0}}{ik} r_0 \right] - \frac{e^{ik\pi} - e^{ikr_0}}{(ik)^2} \right\} \\ &= \frac{e^{ikr_0} - e^{-ik\pi}}{ik} (\alpha_1 - \alpha_2) + \left\{ \frac{e^{ikr_0}}{ik} r_0 - \frac{e^{ikr_0}}{(ik)^2} + \frac{e^{-ik\pi}}{(ik)^2} \right\} (\beta_1 - \beta_2) \\ &\quad + \left[\frac{e^{-ik\pi}}{ik} \pi \right] (\beta_1 + \beta_2), \end{aligned} \tag{88}$$

since

$$\begin{aligned} \int_{-\pi}^{r_0} e^{ikr} (\alpha_1 + \beta_1 r)\, dr &= \alpha_1 \frac{e^{ikr_0} - e^{-ik\pi}}{ik} + \beta_1 \int_{-\pi}^{r_0} e^{ikr} r dr \\ &= \alpha_1 \frac{e^{ikr_0} - e^{-ik\pi}}{ik} + \beta_1 \left[\frac{e^{ikr}}{ik} r \right]_{-\pi}^{r_0} - \beta_1 \int_{-\pi}^{r_0} \frac{e^{ikr}}{ik} dr \\ &= \alpha_1 \frac{e^{ikr_0} - e^{-ik\pi}}{ik} + \beta_1 \left[\frac{e^{ikr_0}}{ik} r_0 + \frac{e^{-ik\pi}}{ik} \pi \right] - \beta_1 \left[\frac{e^{ikr_0}}{(ik)^2} - \frac{e^{-ik\pi}}{(ik)^2} \right], \end{aligned}$$

and

$$\begin{aligned}
\int_{r_0}^{\pi} e^{ikr}\left(\alpha_2+\beta_2 r\right)dr &= \alpha_2\frac{e^{ik\pi}-e^{ikr_0}}{ik}+\beta_2\int_{r_0}^{\pi} e^{ikr}r dr \\
&= \alpha_2\frac{e^{ik\pi}-e^{ikr_0}}{ik}+\beta_2\left[\frac{e^{ikr}}{ik}r\right]_{r_0}^{\pi}-\beta_2\int_{r_0}^{\pi}\frac{e^{ikr}}{ik}dr \\
&= \alpha_2\frac{e^{ik\pi}-e^{ikr_0}}{ik}+\beta_2\left[\frac{e^{ik\pi}}{ik}\pi-\frac{e^{ikr_0}}{ik}r_0\right]-\beta_2\left[\frac{e^{ik\pi}}{(ik)^2}-\frac{e^{ikr_0}}{(ik)^2}\right].
\end{aligned}$$

Then the Fourier series (86) for $g(r)$ has explicit form

$$\begin{aligned}
g(r) &= \frac{(\pi+r_0)\alpha_1+(\pi-r_0)\alpha_2}{2\pi}+\frac{(\beta_2-\beta_1)}{4\pi}\left(\pi^2-r_0^2\right)+\frac{1}{\pi}\sum_{k=1}^{\infty}\mathrm{Re}\left[\left\{\frac{\left(e^{ikr_0}-e^{-ik\pi}\right)(\alpha_1-\alpha_2)}{ik}\right.\right. \\
&\left.\left.+\left\{\left[\frac{e^{ikr_0}}{ik}r_0\right]-\frac{e^{ikr_0}}{(ik)^2}+\frac{e^{-ik\pi}}{(ik)^2}\right\}(\beta_1-\beta_2)+\frac{e^{-ik\pi}}{ik}\pi(\beta_1+\beta_2)\right\}e^{-ikr}\right].
\end{aligned}$$

The special cases shown in Figs. 5 and 6 involve a constant location shift and a trend break shift. For the constant shift case, put $\beta_1=\beta_2=0$ and then

$$g(r)=\frac{\alpha_1+\alpha_2}{2}+\frac{\alpha_1-\alpha_2}{2\pi}r_0+\frac{1}{\pi}\sum_{k=1}^{\infty}\mathrm{Re}\left[\left\{\frac{e^{ikr_0}-e^{-ik\pi}}{ik}\right\}e^{-ikr}\right](\alpha_1-\alpha_2).$$

The corresponding Whittaker filter limit when $\lambda=\mu n^{2m}$ is

$$\begin{aligned}
g_W(r) &= \frac{c_0}{2\pi}+\frac{1}{\pi}\sum_{k=1}^{\infty}\frac{1}{1+\mu k^{2m}}\mathrm{Re}\left\{c_k e^{-ikr}\right\} \\
&= \frac{\alpha_1+\alpha_2}{2}+\frac{\alpha_1-\alpha_2}{2\pi}r_0+\frac{1}{\pi}\sum_{k=1}^{\infty}\frac{1}{1+\mu k^{2m}}\mathrm{Re}\left[\left\{\frac{e^{ikr_0}-e^{-ik\pi}}{ik}\right\}e^{-ikr}\right](\alpha_1-\alpha_2).
\end{aligned}$$

As is evident in Fig. 5 for $m=2$, the HP smoother captures the constant levels better than the finite number of terms of the Fourier series of the function $g(r)$, i.e, the smoother works well in capturing the constant linear levels and the shift is captured as a smooth transition function.

In the trend break case, we work with the function $g(r)=(\alpha_1+\beta_1 r)\mathbf{1}\{r<r_0\}+(\alpha_2+\beta_2 r)\mathbf{1}\{r\geq r_0\}$ and use the Fourier series with coefficients (87) and (88). When $\alpha_1=0,\alpha_2=0,r_0=\frac{1}{2},\beta_1=0.5,\beta_2=1$, we have $g(r)=0.5r\mathbf{1}\{r<r_0\}+r\mathbf{1}\{r\geq r_0\}$ and the Fourier series has the explicit form

$$\frac{\pi}{8}-\frac{1}{32\pi}+\frac{1}{\pi}\sum_{k=1}^{\infty}\mathrm{Re}\left[\left(\left\{\left[\frac{e^{ik/2}}{2ik}\right]-\frac{e^{ik/2}}{(ik)^2}+\frac{e^{-ik\pi}}{(ik)^2}\right\}\left(-\frac{1}{2}\right)+\frac{e^{-ik\pi}}{ik}\frac{3\pi}{2}\right)e^{-ikr}\right],$$

with asymptotic form of the HP filter given by

$$g_{HP}(r)=\frac{\pi}{8}-\frac{1}{32\pi}+\frac{1}{\pi}\sum_{k=1}^{\infty}\frac{1}{1+\mu k^{2m}}\mathrm{Re}\left[\left(\left\{\left[\frac{e^{ik/2}}{2ik}\right]-\frac{e^{ik/2}}{(ik)^2}+\frac{e^{-ik\pi}}{(ik)^2}\right\}\left(-\frac{1}{2}\right)+\frac{e^{-ik\pi}}{ik}\frac{3\pi}{2}\right)e^{-ikr}\right].$$

References

[1] Aitken, A.C. (1925). *The Smoothing of Data.* DSc dissertation, University of Edinburgh.

[2] Aitken, A.C. (1926). On the theory of graduation, *Proceedings of the Royal Society of Edinburgh,* 47, 36–45.

[3] Backus, D. and P. Kehoe (1992). International evidence on the historical properties of business cycles, *American Economic Review* 82, 864-888.

[4] Boneva, L. I., D. Kendall, and I. Stefanov (1970). Spline transformations: three new diagnostic aids for the statistical data analyst, *Journal of the Royal Statistical Society*, Series B 32, 1-71.

[5] Camp, K. (1950). *The Whittaker-Henderson Graduation Processes,* Published by the author, New York.

[6] 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, 253-278.

[7] Cogley, T. (2006). Data Filters, *New Palgrave Dictionary of Economics.*

[8] DeForrest, E. L. (1873). On some methods of interpolation applicable to the graduation of irregular series, such as tables of mortality, etc. *Annual Report of the Board of Regents of the Smithsonian Institution for 1871*, 275-339.

[9] DeForrest, E. L. (1874). Additions to a memoir on methods of interpolation applicable to the graduation of irregular series. *Annual Report of the Board of Regents of the Smithsonian Institution for 1873*, 319-353.

[10] DeForrest, E. L. (1876). *Interpolation and Adjustment of Series.* Tuttle, Morehouse & Taylor, New Haven.

[11] Diewert, E. and T. J. Wales (2006). A 'new' approach to the smoothing problem. Chapter 5, pp. 104-144 in M. T. Belongla and J. M. Binner (Eds.) *Money, Measurement and Computation*, Palgrave MacMillan.

[12] Duy, T. (2012). Tim Duy's Fed watch: careful with that HP filter, http://economistsview.typepad.com/timduy/2012/07/careful-with-that-hp-filter.html

[13] Greville, T. N. E. (1957). On smoothing a finite table: a matrix approach. *SIAM Journal of Applications*, 5, 137-154.

[14] Hansen, G. (1985). Indivisible labor and the business cycle, *Journal of Monetary Economics* 16, 309-327.

[15] Henderson, R. (1924). A new method of graduation, *Transactions of the Actuarial Society of America* 25, 29-40.

[16] Henderson, R. (1925). Further remarks on graduation, *Transactions of the Actuarial Society America*, 26, 52-57.

[17] Henderson, R. (1938). Mathematical theory of graduation, *Actuarial Studies* 4, Actuarial Society of America, New York.

[18] Hodrick, R. J. and E. C. Prescott (1980), Postwar U.S. business cycles: an empirical investigation. Discussion Paper no. 451, Carnegie Mellon University. Published in updated form as Hodrick and Prescott (1997).

[19] Hodrick, R. and E. C. Prescott (1997), Postwar business cycles: an empirical investigation, *Journal of Money, Credit, and Banking*, 29, 1-16.

[20] Kim, S-J, K. Koh, S. Boyd and D. Gorinevsky (2009). ℓ_1 trend filtering. *SIAM Revie*w, 51, 339-360.

[21] King, R.G. and S.T. Rebelo (1993). Low-frequency filtering and real business cycles, *Journal of Economic Dynamics and Control,* 17, 207-231.

[22] Krugman, P. (2012). Filters and full employment (not wonkish, really). New York Times, July 11, 2012. http://krugman.blogs.nytimes.com/2012/07/11/filters-and-full-employment-not-wonkish-really

[23] Krugman, P. (2013).Potential mistakes (wonkish), New York Times, July 6, 2013. http://krugman.blogs.nytimes.com/2013/07/06/potential-mistakes-wonkish/

[24] Leser, C. E. V., 1961. A simple method of trend construction. *Journal of the Royal Statistical Society*, B, 23, 91-107.

[25] Macaulay, F. R. (1931), *The smoothing of time series.* New York: National Bureau of Economic Research.

[26] Osborn, D. R. (1995). Moving average detrending and the analysis of business cycles, *Oxford Bulletin of Economics and Statistics*, 547, 558.

[27] Phillips, P. C. B. (1987). Time series regression with unit roots, *Econometrica* 55, 277-301.

[28] Phillips, P. C. B. (1998). New tools for understanding spurious regressions, *Econometrica*, 66, 1299-1325.

[29] Phillips, P. C. B.(2007). Unit root log periodogram regression. *Journal of Econometrics*, 138, pp.104-124.

[30] Phillips, P. C. B. (2010). Two New Zealand pioneer econometricians, *New Zealand Economic Papers*, 44, 1-26.

[31] Phillips, P. C. B. and Z. Liao. (2014). Series estimation of stochastic processes: recent developments and econometric applications, In J. Racine, L. Su, and A. Ullah (eds), *The Oxford Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics*, Oxford University Press, New York.

[32] Phillips, P. C. B. and H.R. Moon (1999). Linear regression limit theory for nonstationary panel data, *Econometrica,* 67, 1057-1111.

[33] Phillips, P. C. B. and P. Perron (1988). "Testing for a unit root in time series regression," *Biometrika* 75, 335–346.

[34] Phillips, P.C.B., S. Shi and J. Yu (2014). "Specification Sensitivity in Right-Tailed Unit Root Testing for Explosive Behaviour", *Oxford Bulletin of Economics and Statistics*, 76, 315-333.

[35] Phillips, P. C. B. and V. Solo (1992). Asymptotics for linear processes, *Annals of Statistics*, 20, 971–1001.

[36] Phillips, P. C. B. and D. Sul (2007). Transition modeling and econometric convergence tests, *Econometrica*, 75, 1771-1855.

[37] Ravn, M and H. Uhlig (2002). On adjusting the Hodrick-Prescott filter for the frequency of observations, *The Review of Economics and Statistics* 84, 371-375.

[38] Reinsch, H. (1967). Smoothing by spline functions, *Numerische Mathematik* 10, 177-183.

[39] Schoenberg, I. J. (1964). Spline functions and the problem of graduation, *Mathematics* 52, 974-50.

[40] Shiller, R. J. (1973). A distributed lag estimator derived from smoothness priors, *Econometrica*, 41, 775-788.

[41] Shiller, R. J. (1984). Smoothness priors and nonlinear regression, *Journal of the American Statistical Association*, 79, 609-615.

[42] Sprague, T. B. (1887). The graphic method of adjusting mortality tables (with discussion). *Journal of the Institute of Actuaries*, 26, 77-120

[43] Shimotsu K, and P. C. B. Phillips (2005). Exact local Whittle estimation of fractional integration", *Annals of Statistics* 33, 1890- 1933.

[44] Stigler**, S**. M. (1978). Mathematical statistics in the early states**,** *Annals of Statistics,* 6, 239-65.

[45] Stock, J. and M. Watson (1999) Forecasting inflation, *Journal of Monetary Economics* 44, 293-335.

[46] Treves, F. (1980). *Introduction to Pseudodifferential and Fourier Integral Operators, Vols.1 & 2.* New York: Plenum Press.

[47] Wahba, G. (1978). Improper priors, spline smoothing, and the problem of guarding against model errors in regression, *Journal of the Royal Statistical Society*, Series B, 40, 364-372

[48] Whittaker, E. T. (1923). On a new method of graduation. *Proceedings of the Edinburgh Mathematical Association*, 78, 81-89.

[49] Whittaker, E. T. and G. Robinson (1924). *Calculus of Observations.* London: Blackie and Son.