Economics Working Paper Series

2023 - 05

To Boost or Not to Boost? That is the Question

Ye Lu and Adrian Pagan

February 2023

To Boost or Not to Boost? That is the Question*

Ye Lu[†] Adrian Pagan[‡]

February 10, 2023

Abstract

Phillips and Shi (2021) have argued that there may be some leakage from the estimate of the permanent component to what is meant to be the transitory component when one uses the Hodrick-Prescott filter. They argue that this can be eliminated by boosting the filter. We show that there is no leakage from the filter *per se*, so boosting is not needed for that. They also argue that there are DGP's for the components for which the boosted filter tracks these more accurately. We show that there are other plausible DGP's where the boosted filter tracks less accurately, and what is crucial to tracking performance is how important permanent shocks are to growth in the series being filtered. In particular, the DGP's used in Phillips and Shi (2021) have a very high contribution from permanent shocks.

*We would like to thank Sophocles Mavroeidis, Tim Robinson, Daniel Buncic, James Hamilton and James Duffy for the comments on the material in the paper.

[†]School of Economics, University of Sydney. Email: ye.lu1@sydney.edu.au.

[‡]School of Economics, University of Sydney and Center for Applied Macroeconomic Analysis (CAMA), Australian National University. Email: adrian.pagan@sydney.edu.au.

1 Introduction

The Hodrick-Prescott (HP) filter has come in for some bad press in recent times. It is worth looking at some of this criticism and solutions to it in more detail. Some of the criticisms are clearly valid although often ignored. For example, it is a two-sided filter so it needs to be used with care as an independent variable in a regression. An example would be the estimation of an "output gap" from a HP filter on output followed by inputting this into a regression to capture the Phillips curve; see Fukac and Pagan (2010, Section 4). It is a generated regressor that affects the consistency of the regression estimator rather than just the standard errors. A one-sided HP filter obviously avoids some of that criticism, although even a regressor constructed with that may be correlated with the Phillips curve error due to its dependence on current output levels. Others have pointed out that there are issues when the HP filter is used as a stand-alone measure of various concepts such as a potential GDP. Underlying this latter literature is an unobserved components (UC) model of a time series, $y_t = T_t + c_t$, where T_t is a permanent component and c_t a transitory component. Names are often given to c_t, such as "cycle", "gaps", etc., and to T_t, such as "trend". Alternatives to the HP filter for performing a permanent/transitory components decomposition include the Beveridge-Nelson (BN) approach, and there have been suggestions that this may provide a better set of outcomes. The BN approach focuses on producing an estimator $\hat{T}_t = y_t + E_t(\sum_{j=1}^{\infty} \Delta y_{t+j})$. A version that resembles this in some way is proposed by Hamilton (2018) who argue that one should instead use $\hat{T}_t = E_{t-h}(y_t)$, where h is prescribed by the user.

In Section 2 we set out some properties of the HP and related filters. We begin by observing that one cannot recover T_t and c_t from the data on y_t with any value of the key tuning parameter λ. This seems to conflict with some of the conclusions in Phillips and Jin (2021) who examined what happened with the HP filter when λ was allowed to grow with sample size n. They concluded that "It is therefore to be expected that for choices of the smoothing parameter that approximate $\lambda = \mu n^4$, the HP filter fails to remove a stochastic trend and the imputed business cycle estimate $\hat{c}_t$ inevitably imports the random wandering character of a stochastic trend" (Phillips and Jin, 2021, Remark 9, p. 18). In their Table 2, this case is labelled as "inconsistent", in contrast with a fixed λ, where it is termed "consistent". So we need to explore the fact that the latter reference cannot be to the recovery of T_t and c_t. Indeed, we show that the "integration leakage", which seems to be an implication of their result, does not happen, i.e., if there is a permanent component in $\hat{c}_t$ it is sourced from some other feature of the data y_t, and not from an inability to recover T_t. Potential integration leakage for other filters is also discussed. A key requirement for there to be none is that the filter weights should sum to unity, which they do for the HP filter. Section 3 looks at the question of the paper's title. In the event that $\hat{c}_t$ did have a permanent component, Phillips and Shi (2021) proposed to iterate the HP filter (boosting it) so as to remove that. This boosted filter applies the basic HP filter with a fixed λ to $\hat{c}_t$ to produce $\hat{c}_t^{(1)}$, and then applies HP to $\hat{c}_t^{(1)}$ to get $\hat{c}_t^{(2)}$, and so on. There are stopping rules proposed to terminate the iteration. But, if the permanent component is not due to leakage, then we might expect little change in the nature of $\hat{c}_t^{(j)}$ by boosting. As an example of this, Hall and Thomson (2022) show in their Table 7 that there are only relatively minor changes in the persistence of $\hat{c}_t$ when moving from a basic to a boosted HP filter. We also show that this is the case for the industrial production example considered in Phillips and Shi (2021).

However, an interesting feature of the boosted filter which Phillips and Shi (2021) find is that $var(\hat{T}_t^{boost} - T_t)$ is less than $var(\hat{T}_t^{HP} - T_t)$ for a range of specifications of T_t and c_t. That is, the boosted estimator of the permanent component more closely tracks the latent value T_t, even though it can never equal it. This seems to provide a good argument for boosting, as it does not vary any of the parameters assumed by HP but just iterates the filter. One possibility is that the improvement in trend fit is a consequence of a deterioration in that for the cycle. We show that this is not the case, as the variances of the estimators of T_t and c_t around their latent values are identical. Another comes from the fact that the original HP filter was motivated by a components model in which T_t was assumed to be I(2) and c_t to be white noise. Hence the components model implies that y_t is ARIMA$(0, 2, 2)$ with coefficients that depend on λ, which HP set to 1600 for quarterly work. It is true that the derived coefficients fail to match what one would get from fitting an unrestricted ARIMA$(0, 2, 2)$ model to the data, e.g., for quarterly US GDP over 1973/1 to 2014/4:

$$\text{US Data}: \quad \Delta^2 y_t = (1 - .74L - .26L^2)e_t,$$

while the process implied by HP is

$$\text{HP:} \quad \Delta^2 y_t = (1 - 1.77L + .8L^2)e_t.$$

Hamilton (2018) gives examples of this discrepancy for many series. So, perhaps boosting somehow provides a different model for the data, but it is hard to see how this can be. Consequently, if the underlying components model is fixed then one has to explain how $var(\hat{T}_t^{boost} - T_t)$ can become smaller than that from the HP filter. We find that this can occur for some DGP's for T_t and c_t but not others. A reduction seems to be dependent on the permanent shocks having a big effect on Δy_t and $\Delta^2 y_t$. The first might be thought of as at variance with current thinking.

Sections 4 gives some answers to the question posted by the title of this paper.

2 Some Properties of the HP and Related Filters

2.1 Impossibility of trend and cycle recovery

The first property of the HP filter we consider here is dealt with in Pagan and Robinson (2022) and Buncic and Pagan (2022). It is that one can never recover T_t and c_t: since each of the two latent variables are constructed from a single observable using a linear filter, the estimates are in a linear relationship. This failure to recover T_t comes from the fact that $var(\hat{T}_t - T_t)$ is non-zero and does not depend on the sample size n, so asymptotically one cannot recover T_t. In contrast, Phillips and Jin (2021) conclude that, if λ rises slowly enough relative to n, "the filter asymptotically captures the underlying stochastic trend". It is, however, a much more nuanced statement than implying that one is recovering T_t. It actually implies that in the stated circumstances one will be constructing a series $\hat{T}_t$ that follows *the same stochastic process* as T_t, which is different to *recovery*.

To explore this a little further, we consider the case of

$$y_t = T_t + c_t, \quad T_t \sim I(1), \quad c_t \sim I(0), \quad y_t \sim I(1).$$

Here, for any time series (x_t), we say $x_t \sim I(1)$ if $n^{-1/2} x_{\lceil nr \rceil} \to_d \sigma B(r)$ as $n \to \infty$, for some $\sigma > 0$ and the standard Brownian motion $B(\cdot)$ on $[0, 1]$; and we say $x_t \sim I(0)$ if $n^{-1/2} \sum_{t=1}^{\lceil nr \rceil} x_t \to_d \sigma B(r)$ as $n \to \infty$, for some $\sigma > 0$ and standard Brownian motion $B(\cdot)$ on $[0, 1]$. Another way to characterize the difference between $I(1)$ and $I(0)$ processes is through the different stochastic orders of their asymptotically independent periodogram coordinates around frequency zero, i.e. the so-called "low-frequency" coordinates in the sense of Müller and Watson (2008, 2020). The following proposition states such a characterization.

Proposition 1. *Suppose that $(x_t), t = 1, \ldots, n$, is a time series of either $I(1)$ or $I(0)$. We define K periodogram coordinates*[1] $I(\omega_{nk}) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} x_t \sin(\omega_{nk} t)$ *with* $\omega_{nk} = \omega_k / n$ *and* $\omega_k = \left(k - \frac{1}{2}\right)\pi$, *for* $k = 1, \ldots, K$, *where K is a fixed integer. It follows that, as $n \to \infty$, $I(\omega_{n1}), \ldots, I(\omega_{nK})$ are asymptotically independent and for $k = 1, \ldots, K$,*

(i) if $x_t \sim I(1)$, then $n^{-1} I(\omega_{nk}) \to_d N(0, \sigma^2 \omega_k^{-2})$ for some $\sigma^2 > 0$;

(ii) if $x_t \sim I(0)$, then $I(\omega_{nk}) \to_d N(0, \sigma^2)$, for some $\sigma^2 > 0$.

The K periodogram coordinates $I(\omega_{nk})$ correspond to the frequencies that are asymptotically around zero: $\sup_{k=1,\ldots,K} \omega_{nk} \to 0$ as $n \to \infty$. One implication of the above characterization is that for I(1) process, $\sup_{k=1,\ldots,K} I(\omega_{nk}) = O(n)$ and $\sup_{k=1,\ldots,K} I(\omega_{nk}) = O(1)$.

Under the assumptions in Proposition 1, the low-frequency coordinates of y_t and T_t satisfy (i) and those of c_t satisfy (ii) given in the Proposition. What would happen if one applies the two-sided HP filter with smoothing parameter $\lambda > 0$ where λ is allowed to grow with n? It is known that the asymptotic operator solution of the estimated trend and cycle components are $\hat{T}_t = P(L) y_t$ with $P(L) = [1 + \lambda L^{-2}(1-L)^4]^{-1}$ and $\hat{c}_t = [1 - P(L)] y_t$. The frequency responses (i.e. the transfer functions) of these filters are

$$P(\omega) = \frac{1}{1 + 4\lambda(1-\cos\omega)^2} \quad \text{and} \quad 1 - P(\omega) = \frac{4\lambda(1-\cos\omega)^2}{1 + 4\lambda(1-\cos\omega)^2},$$

which are both strictly monotone functions on $[0, \pi]$ taking values between zero and one. The filter gains that determine the spectral properties of $\hat{T}_t$ and $\hat{c}_t$ are simply $|P(\omega)| = P(\omega)$ and $|1 - P(\omega)| = 1 - P(\omega)$.

In light of Proposition 1, we want to understand the behaviour of $|P(\omega_{nk})|$ and $|1 - P(\omega_{nk})|$ at the low frequencies $\omega_{n1}, \ldots, \omega_{nK}$, as $n \to \infty$. To this end, we apply the Taylor expansion of a cosine function, $\cos\omega = \sum_{m=0}^{\infty} (-1)^m \omega^{2m}/(2m)!$, to write

$$|1 - P(\omega)| = \frac{1}{1 + \lambda\omega^4(1 + g(\omega))} \quad \text{and} \quad |1 - P(\omega)| = \frac{\lambda\omega^4(1 + g(\omega))}{1 + \lambda\omega^4(1 + g(\omega))}, \tag{1}$$

where

$$g(\omega) = 8 \sum_{m=1}^{\infty} \frac{(-1)^m (4^{m+1} - 1)}{(2m+4)!} \omega^{2m} = O(\omega), \quad \text{as } \omega \to 0. \tag{2}$$

[1] The usual definition of periodogram coordinates of a time series is the discrete Fourier transform of the time series which is complex-valued involving both cosine and sine components. Here, we use the (Type IV) sine transform to define the periodogram coordinates.

It then follows from (1) and (2) that for each $k = 1, \ldots, K$,

$$|P(\omega_{nk})| = \frac{1}{1 + \lambda n^{-4}\omega_k^4(1 + o(1))} \quad \text{and} \quad |1 - P(\omega_{nk})| = \frac{\lambda n^{-4}\omega_k^4(1 + o(1))}{1 + \lambda n^{-4}\omega_k^4(1 + o(1))},$$

as $n \to \infty$.

Let $I_{\hat{T}}(\omega_{nk})$ and $I_{\hat{c}}(\omega_{nk})$ be, respectively, the low-frequency coordinates of $\hat{T}_t$ and $\hat{c}_t$ around frequency zero. It can then be easily shown that, if λ grows at rate $\lambda = \mu n^4$ for $\mu > 0$, then

$$\begin{aligned} n^{-1} I_{\hat{T}}(\omega_{nk}) &\to_d \sigma N\left(0, \frac{\omega_k^{-2}}{(1 + \mu\omega_k^4)^2}\right), \\ n^{-1} I_{\hat{c}}(\omega_{nk}) &\to_d \sigma N\left(0, \frac{\mu^2\omega_k^6}{(1 + \mu\omega_k^4)^2}\right). \end{aligned}$$

Since $I_{\hat{c}}(\omega_{nk}) = O(n)$, $\hat{c}_t$ must be nonstationary. Moreover, we also note that the asymptotic variance of $I_T(\omega_{nk})$ is different from that of $I_T(\omega_{nk})$, which is stated in Proposition 1 (i). These results are analogous to Theorem 3 in Phillips and Jin (2021, p. 17).

In contrast, if λ rises with n at any power lower than 4, $P(\omega_{nk}) \to 1$ as $n \to \infty$ and

$$n^{-1} I_{\hat{T}}(\omega_{nk}) \to_d \sigma N(0, \omega_k^{-2}),$$

which is the same as the asymptotic distribution of the low-frequency coordinates of the original data. However, just because the low-frequency periodogram coordinates of $\hat{T}_t$ and T_t have the same asymptotic distribution, does not mean that $\hat{T}_t$ recovers T_t. The cycle c_t is even further from being recoverable in that even the asymptotic distributions of the periodogram coordinates $I_{\hat{c}}(\omega_{nk})$ and $I_c(\omega_{nk})$ cannot be matched for any choice of λ.[2]

2.2 Trend-elimination and the row-sum of unity constraint

Phillips and Jin (2021) in their statement (setting $\lambda = \mu n^4$) were concerned about what we call "integration leakage" when $\hat{T}_t$ was not "consistent". We would like to ask the question: does a failure to recover T_t results in $\hat{c}_t$ having a permanent component? Consider how the filter is performed when there are n observations on y_t. As above it is assumed that the series has the structure $y_t = T_t + c_t$, where T_t is the permanent

[2] Since $c_t \sim I(0)$, we have $\sup_{k=1,\ldots,K} I_c(\omega_{nk}) = O(1)$. We only have $\sup_{k=1,\ldots,K} I_{\hat{c}}(\omega_{nk}) = O(1)$ when $\lambda = \mu n^3$, in which case the asymptotic distribution of $I_{\hat{c}}(\omega_{nk})$ is different from that in Proposition 1 (ii). Therefore, the asymptotic distribution of $I_c(\omega_{nk})$ is not recoverable.

component and c_t the transitory component. The objective is to estimate T_t and c_t. To do this with the two-sided HP filter, the two components T_t and c_t are estimated respectively as

$$\hat{T} = Hy \quad \text{and} \quad \hat{c} = (I - H)y$$

where $y, \hat{T}$, and $\hat{c}$ are $n \times 1$ vectors, I is the $n \times n$ identity matrix, and $H = G^{-1}$ where $G = I + \lambda D^\top D$ with D being the $(n-2) \times n$ matrix producing second-order difference. Specifically,

$$G = \begin{bmatrix} 1+\lambda & -2\lambda & \lambda & 0 & 0 & 0 & \ldots \\ -2\lambda & 1+5\lambda & -4\lambda & \lambda & 0 & 0 & \ldots \\ \lambda & -4\lambda & 1+6\lambda & -4\lambda & \lambda & 0 & \ldots \\ 0 & \lambda & -4\lambda & 1+6\lambda & -4\lambda & \lambda & \ldots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix},$$

where the last two rows have the same numbers as the first two but re-arranged.

Letting $h(t, \cdot)$ denote the tth row of H for $t = 1, \ldots, n$, we can write

$$\hat{T}_t = h(t, \cdot)y \quad \text{and} \quad \hat{c}_t = y_t - \hat{T}_t.$$

We define, respectively, the sequential right and left partial row sums of H by

$$b(t, j) = \sum_{k=j}^{n} h(t, k) \quad \text{and} \quad d(t, j) = \sum_{k=1}^{j} h(t, k),$$

for $t, j = 1, \ldots, n$. It is clear that the entire row sum of H for the tth row is $b(t, 1)$ or $d(t, n)$. Then we have the following proposition.

Proposition 2. *Given the partial row sums $b(t, j)$ and $d(t, j)$ in, the estimated transitory component $\hat{c}$ can be written as*

$$\hat{c}_t = [1 - b(t, 1)]y_t + \sum_{j=2}^{n} a(t, j)\Delta y_j, \quad t = 1, \ldots, n,$$

where

$$a(t, j) = \begin{cases} d(t, t+1-j), & 2 \leq j \leq t, \\ b(t, j), & t+1 \leq j \leq n. \end{cases}$$

In the case of the HP filter, it is known that every row of H sums to unity (e.g. De

Jong and Sakarya, 2016), i.e. $b(t,1) = 1$, then it follows from Proposition 2 that

$$\hat{c}_t = \sum_{j=2}^{n} a(t,j)\Delta y_j, \quad t = 1, \ldots, n. \tag{3}$$

Therefore, the estimated c_t inherits the properties of Δy_t rather than y_t. If T_t is $I(1)$ and c_t is $I(0)$ then Δy_t is $I(0)$. Thus what this says is that there is no integration leakage for the HP filter from permanent to transitory component estimates when T_t is $I(1)$.

Now there are other processes that T_t might follow. It could be an integrated process of higher than first order and it could have deterministic terms n^k $(k > 1)$. In that case one needs to look at the values of $\alpha(t,j)$. Sakarya and De Jong (2020, Theorem 3) find that $\hat{c}_t$ is a weakly dependent process if T_t is not of greater integration order than I(4). This continues to hold when there is a linear trend in T_t. The presence of any quadratic and higher order deterministic trends leaves a factor in $\hat{c}_t$ that is non-zero for observations at the beginning and end of the sample (Sakarya and De Jong, 2020, Theorem 5). At the beginning there is a smooth decline to zero and, at the end, a smooth rise. Hence an ADF test will suggest a unit root even though the process is not integrated. So one would want to apply a HP filter to series that have already been adjusted for any polynomial time trend higher than first order when computing $\hat{c}_t$.

To give an example, we take the quarterly log industrial production data y_t from 1919 to 2018 used as the third empirical example by Phillips and Shi (2021). There seems only a linear trend in this series. The regression of Δy_t on Δy_{t-1} and a trend gives the latter a coefficient that is very small and not significantly different from zero (t statistic being $-.7$) while the estimated coefficient on Δy_{t-1} of $-.39$. Consequently, since there is only a linear trend, it should not be surprising that the ADF test on $\hat{c}_t$ yields the value of the test statistic -7.7.

What happens with the other proposed filters? We know that the BN solution is a one-sided filter

$$\hat{T}_t = y_t + \sum_{j=1}^{p} \beta_j \Delta y_{t-j},$$

where p is the order of the autoregression for Δy_t. Clearly the same situation holds. Provided that c_t is transitory so is $\hat{c}_t$. Suppose instead that we used a filter that had the same form as HP but $b(t,1) = \sum_{j=1}^{n} h(1,j) \neq 1$. Then it follows from Proposition 2 that the first term $[1 - b(t,1)]y_t$ is nonzero in the expression of $\hat{c}_t$. Therefore, there is a leakage to the cycle estimate, i.e., the unit root in y_t influences $\hat{c}_t$. So, one needs a filter with the property that the rows of H sum to unity.

Just as for the BN filter, Hamilton (2018)'s filter is a one-sided filter and it involves defining $\hat{c}_t$ as $y_t - E_{t-h}(y_t)$. So we would have the permanent component as $T_t = E_{t-h}(y_t)$ rather than $E_t(y_\infty)$ as in BN. Of course $T_t = \sum_{j=1}^{h} \gamma_j y_{t-j}$ and so

$$\hat{T}_t = \gamma_1 y_{t-1} + \sum_{j=2}^{h} \gamma_j \Delta y_{t-j},$$

meaning that

$$\begin{aligned} \hat{c}_t &= y_t - \gamma_1 y_{t-1} - \sum_{j=2}^{h} \gamma_j \Delta y_{t-j} \\ &= (1-\gamma_1) y_{t-1} + \Delta y_t - \sum_{j=2}^{h} \gamma_j \Delta y_{t-j}. \end{aligned}$$

We are therefore left with an I(1) component unless $\gamma_1 = 1$. In large samples we might expect that the regression which gives an estimate of γ_1 would give such a value when T_t is I(1). Indeed, Phillips and Shi (2021) point out that the estimate of γ_1 with the industrial production data is effectively unity.

3 Boosting the HP Filter

We now turn to looking at how boosting the HP filter can do better at recovering T_t and reducing any serial correlation in c_t. Boosting involves repeatedly using the HP filter on $\hat{c}_t$ to reduce the serial correlation in $\hat{c}_t$. Let us take the basic HP estimate of the transitory component as $\hat{c}_t^{(1)}$ and then iterate this by applying the HP filter to $\hat{c}_t^{(1)}$ to produce $\hat{c}_t^{(2)}$ etc. This can be thought of as computing the Kalman smoother applied to $\hat{c}_t^{(1)}$ and using the UC model

$$\hat{c}_t^{(1)} = \psi_t + c_t,$$

with the assumptions that $\Delta^2 \psi_t$ and c_t are white noise. Because we use the same value of λ as at the first iteration, from (3) it follows that

$$\hat{c}_t^{(2)} = -\sum_{j=2}^{n} \alpha(t,j) \Delta \hat{c}_j^{(1)}.$$

Going back to the the quarterly log industrial production example where we saw earlier that there was no I(1) behaviour in $\hat{c}_t^{(1)}$, now the ADF test statistic for $\hat{c}_t^{(2)}$ on this data is -8.6. There seems little evidence of a unit root in the original HP filtered $\hat{c}_t^{(1)}$ and there is less as one iterates. The stopping rule Phillips and Shi (2021) have used, reported in their Figure 8(b), is based on the value of BIC and not ADF, and it involves 7 iterations. Applying ADF test to $\hat{c}_t^{(7)}$, we find the test statistic being -11. It is interesting that the value of ADF test statistic continues to decline. To understand why we regress $\hat{c}^{(j)}$ on $\hat{c}_{t-1}^{(j)}$ and $\Delta\hat{c}_{t-1}^{(j)}$ and find that the coefficient of $\hat{c}_{t-1}^{(j)}$ is 0.8 for $j = 1$; 0.76 when $j = 2$; and 0.66 when $j = 7$. So there doesn't seem to be any argument for using the boosted filter based on the idea that there will be very strong persistence in $\hat{c}_t$ which boosting is being used to eliminate. The results are very similar to what is in Hall and Thomson (2022). One might add that if one thinks of $\hat{c}_t$ as an output gap why would one want small persistence in it?

Now a feature that Phillips and Shi (2021) point to when boosting is that the variance of the filtered trend estimate around the actual T_t declines with each iteration. Consider a simulation of T_t and c_t that gives some realizations of the latent variables with corresponding realizations for y_t, which we call y_t^R. Then applying HP to y_t^R we have

$$y_t^R = T_t + c_t = \hat{T}_t + \hat{c}_t,$$

producing

$$\xi_{Tt} = -\xi_{ct},$$

where

$$\xi_{Tt} = \hat{T}_t - T_t \quad \text{and} \quad \xi_{ct} = \hat{c}_t - c_t.$$

Consequently, $var(\xi_{Tt}) = var(\xi_{Tt})$ and the variance of the estimated trend is the same as that for the estimated cycle. It is also apparent that ξ_{Tt} and ξ_{ct} are perfectly negatively correlated. These results hold for each boosted estimate of T_t and c_t.

So is it the case that the reduction in $std(\xi_{ct}^{(j)})$ observed from boosting applies generally or is it a result of the particular DGP's for T_t and c_t that Phillips and Shi (2021) use? Ultimately, to judge this we need to have some knowledge of what a reasonable

DGP would be for the components. Take the following one for example:

$$\begin{aligned} y_t &= T_t + c_t \\ \Delta T_t &= \varepsilon_{1t} \\ c_t &= 0.5c_{t-1} + \varepsilon_{2t} + \varepsilon_{2t-1}, \end{aligned}$$

where ε_{jt} are uncorrelated and are independently normally distribution with zero mean and unit variance. The above model has similarities to Clark (1987), and corresponds to DGP3 in Phillips and Shi (2021) when c_t is set to be zero instead. Phillips and Shi (2021) show that for their DGP3, boosting the HP filter results in a very large reduction in the MSE of $(\hat{c}_t - c_t)$, after one boosts a relatively large number of times. On average, the BIC criterion they use as a stopping rule does 9 boosts (shown in their Table 2) and, with our simulations, the MSE reduces from 1.58 to 0.83, which agrees with the figures in their Table 2. Note that with such a DGP where $y_t = T_t$, permanent shocks contribute 100% to the standard deviation of Δy_t. However, when the cycle c_t described above is added to T_t to get a new y_t, this fraction reduces to 75%, while the MSE reduces from 2.17 to 2.03 after the 9th boost. So, with this new DGP for y_t, there is just a 6% reduction in the MSE – much less dramatic.

It seems unlikely that permanent shocks would account for such a high proportion of Δy_t though. Angeletos, Collard and Dellas (2020) argued that TFP shocks were not very important for the standard deviation of US quarterly GDP growth, and Gillman and Pagan (2023) argued for a 20% contribution of the TFP shocks to the standard deviation of Δy_t. To reduce it to that level we need to increase the standard deviation of ε_{2t} to 10, and then we find that the MSE *increases* by 45%, a result that is not attractive for boosting.

Phillips and Shi (2021) are mainly concerned about what happens when there are unaccounted for deterministic trends in T_t, i.e., no adjustment is made to y_t before application of the HP filter. They propose adding on a cubic deterministic trend of the form $.0005t^3$ to DGP3 to get their DGP1. That now results in a large fall in the MSE, as seen in their Figure 3(a). If the shocks for c_t have higher standard deviation than unity (specifically 3.5) we find that the first boost *increases* MSE. Since 3.5 is lower than the value of 10 used earlier, permanent shocks are actually very important to Δy_t in this case.

Mei, Phillips and Shi (2022) construct a DGP for y_t that has T_t as an I(2) process and c_t is

$$c_t = \cos(\pi/10)c_{t-1} - .25c_{t-2} + \gamma\varepsilon_{2t}.$$

They set $\gamma = 5$. This is their DGP1. Here we have less opinions about what the fraction of the standard deviation of $\Delta^2 y_t$ would be due to permanent shocks. Essentially HP set λ based on a judgement about that. When using their components model and $\lambda = 1600$ it translates into a contribution from permanent shocks of around 1% to the standard deviation of $\Delta^2 y_t$. That seems to suggest that permanent shocks are likely to be a minor cause of *fluctuating* growth rates. In the experiment above, setting $\gamma = 5$ gives a contribution of permanent shocks to the standard deviation of $\Delta^2 y_t$ of around 10%, and it results in a 28% reduction in the standard deviation of ξ_{ct} by boosting once. This agrees with their Table 1. Accepting Hodrick and Prescott's reasoning about the relative contribution we would need to set γ to around 98, and not 5, to get that outcome. If one does that then twicing (boosting once) *raises* the standard deviation of ξ_{ct} by 13%. In order to get a reduction in the latter we need to put $\gamma = 28$, and then the contribution of the permanent shock to the standard deviation of $\Delta^2 y_t$ is 1.8%. It is clear that the nature of the DGP is crucial to whether boosting will improve tracking performance and one cannot assume that it will always result in improved tracking.

One interesting question is how different the boosted and basic HP estimates of the cycle are. Figure 1 shows this for the industrial production data over 1919 to 1940. There are differences, and these raise some historical issues. The solid line is the basic HP estimate and it says that the output gap was much larger than one would get from the 7th boost of it. Particularly evident is that the boosted output gap estimates are smaller than the standard HP filter estimates in the first part of the Depression (1930-1933) than in the latter part (1938-1940), which would be counter to narrative evidence.

4 Conclusion

What answers do we give to the question in the paper's title? There are two. First, one does not need to boost to avoid a unit root in $\hat{c}_t$. The possibility of a unit root in $\hat{c}_t$ based on an ADF test is either because c_t has a unit root (a very unusual components model) or one has not adjusted the series y_t for a deterministic trend of higher than second order when estimating c_t. Does boosting eliminate higher order deterministic trends? The answer is in the negative. The impact of such a trend remains in the estimated cycle and does not go to zero. Second, whether you want to boost the HP filter to improve

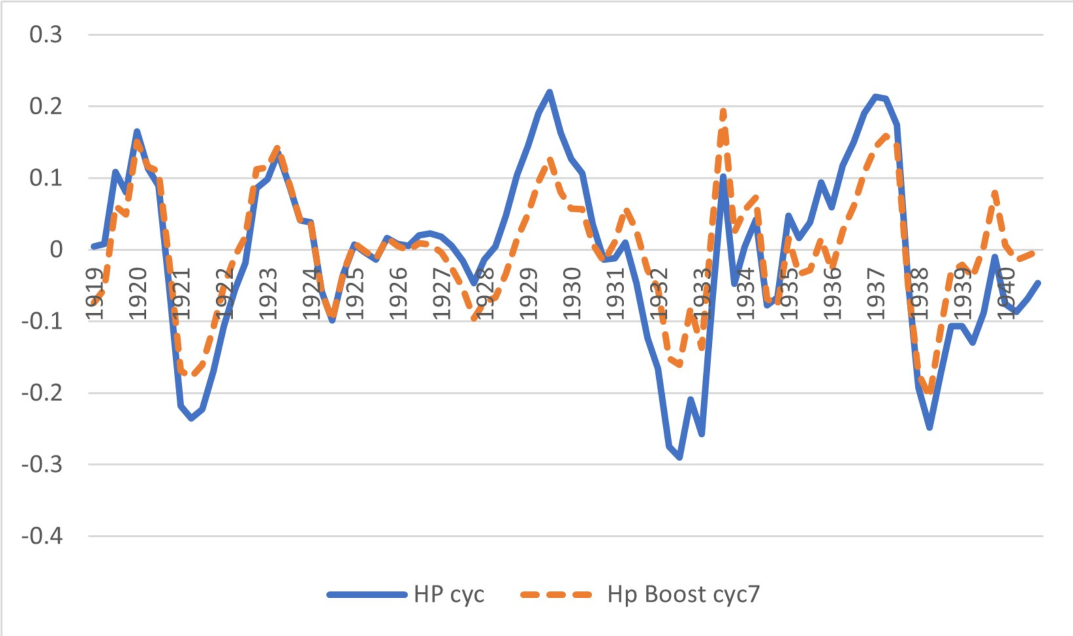

Figure 1: Quarterly US Industrial Production 1919-1940: HP Cycle and Boosted HP Cycle

tracking performance will depend on how important you think the permanent shocks are to Δy_t and $\Delta^2 y_t$. There are plausible specifications of T_t and c_t which result in a poorer tracking performance by boosting. One can never really know about this. One has more confidence that a likely contribution or permanent shocks to Δy_t would result in boosting worsening the tracking of c_t. The situation is much less clear if data follows an $I(2)$ process. It should be noted that in all cases there is a perfect negative correlation between the estimated trend and cycle gaps.

References

Angeletos, G. M., F. Collard and H. Dellas (2020). Business cycle anatomy, *American Economic Review*, 110, 3030-3070.

Buncic, D. and A. Pagan (2022). Discovering stars: Problems in recovering latent variables from Models. *CAMA Working Papers 52/2022*, Australian National University.

Clark, P. K. (1987). The cyclical component of U.S. economic activity, *The Quarterly Journal of Economics*, 102, 797-814.

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

Fukac, M. and A. Pagan (2010). Limited information estimation and evaluation of DSGE models. *Journal of Applied Econometrics*, 25, 55-70.

Gillman, M. and A. Pagan (2023). Investigating cycle anatomy. *CAMA Working Paper 9/2023*, Australian National University.

Hall, V. B. and P. Thomson (2022). A boosted HP filter for business cycle analysis: Evidence from New Zealand's small open economy. *CAMA Working Paper 45/2022*, Australian National University.

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

Harding, D. and A. Pagan (2016). The econometric analysis of recurrent events in macroeconomics and finance. Tinbergen Lectures, Princeton University Press.

Mei, Z., P. C. B. Phillips and Z. Shi, (2022). The boosted HP filter is more general than you might think. Papers 2209.09810, arXiv.org.

Müller, U. K. and M. W. Watson (2008). Testing models of low-frequency variability. *Econometrica*, 76(5), 979–1016.

Müller, U. K. and M. W. Watson (2020). Low-frequency analysis of economic time series. Draft chapter for *Handbook of Econometrics*, Volume 7.

Pagan, A. and T. Robinson (2022). Excess shocks can limit the economic interpretation. *European Economic Review*, 145.

Phillips, P. C. B. and Z. Shi (2021). Boosting: Why you can use the HP filter. *International Economic Review*, 62, 521-570.

Phillips, P. C. B. and S. Jin (2021). Business cycles, trend elimination and the HP filter. *International Economic Review*, 62, 469-520.

Sakarya, N and R. M. De Jong (2020). A property of the Hodrick–Prescott filter and its application. *Econometric Theory*, 36(5), 840-870.

Appendix: Mathematical Proofs

Proof of Proposition 1. For a time series (x_t), we define its partial sum process $S_t = x_1 + \ldots + x_t$ with $S_0 := 0$, and then a sequence of stochastic process $S_n(r) = S_{\lceil nr \rceil}$ for $r \in [0,1]$, where $\lceil a \rceil$ denotes the smallest integer that is greater than $a \in \mathbb{R}$. This sequence of stochastic process, under proper normalization, has different limit depending on the stochastic order of (x_t). In particular, under Part (i) of the proposition where $(x_t) \sim I(1)$, we have

$$n^{-3/2} S_n(r) \to_d \sigma \int_0^r B(s)ds, \quad \text{as } n \to \infty, \tag{4}$$

where $B(\cdot)$ denotes the standard Brownian motion on $[0,1]$. If $(x_t) \sim I(0)$ as assumed in Part (ii) of the proposition, then we have

$$n^{-1/2} S_n(r) \to_d \sigma B(r), \quad \text{as } n \to \infty, \tag{5}$$

where $B(\cdot)$ denotes the standard Brownian motion on $[0,1]$. We combine (4) and (5) to the following result:

$$G_n(r) := n^{-\alpha} S_n(r) \to_d \sigma G(r), \tag{6}$$

where $\alpha = 3/2$ and $G(r) = \int_0^r B(s)ds$ if $x_t \sim I(1)$, and $\alpha = 1/2$ and $G(r) = B(r)$ if $x_t \sim I(0)$.

Now, consider $\varphi_k(r) = \sin(\omega_k r)$, $r \in [0,1]$, with $\omega_k = (k - 1/2)\pi$ for $k = 1, \ldots, K$ where K is a fixed positive integer. That is, $\{\varphi_k(r)\}_{k=1}^k$ are the first K orthonormal basis functions of the Karhunen-Loève representation of the standard Brownian motion on $[0,1]$. Furthermore, we define the first K periodogram coordinates of x_t as

$$I_{nk} = \sum_{t=1}^n \phi_{tk} x_t, \quad k = 1, \ldots, K,$$

where for each $k = 1, \ldots, K$,

$$\phi_{tk} = \frac{1}{\sqrt{n}} \varphi_k(t/n) = \frac{1}{\sqrt{n}} \sin(\omega_{nk} t), \quad \text{with } \omega_{nk} = \omega_k / n,$$

for $t = 1, \ldots, n$. We denote $I(\omega_{nk}) = I_{nk}$ to indicate that they are periodogram coordinates with frequency ω_{nk}.

In light of (6), we write

$$
\begin{aligned}
n^{-\alpha+1/2} I(\omega_{nk}) = n^{-\alpha+1/2} I_{nk} &= n^{-\alpha+1/2} \sum_{t=1}^{n} \phi_{tk}(S_t - S_{t-1}) \\
&= n^{-\alpha+1/2} \left[\phi_{nk} S_n - \phi_{1k} S_0 - \sum_{t=2}^{n} S_{t-1}(\phi_{tk} - \phi_{t-1,k}) \right] \\
&= \sqrt{n}\phi_{nk} G_n(1) - \sum_{t=2}^{n} G_n\left(\frac{t-1}{n}\right) \sqrt{n}(\phi_{tk} - \phi_{t-1,k}) \qquad (7)
\end{aligned}
$$

where $\sqrt{n}\phi_{nk} = \varphi_k(1)$, and

$$\sqrt{n}(\phi_{tk} - \phi_{t-1,k}) = \varphi_k\left(\frac{t}{n}\right) - \varphi_k\left(\frac{t-1}{n}\right) = \varphi_k'\left(\frac{\bar{t}}{n}\right)\frac{1}{n}, \qquad (8)$$

for $\bar{t} \in (t-1, t)$.

It then follows from (6), (7), (8), and continuous mapping that for each $k = 1, \ldots, K$,

$$n^{-\alpha+1/2} I(\omega_{nk}) \to_d Z_k \qquad (9)$$

as $n \to \infty$, where

$$Z_k = \sigma \left[\varphi_k(1) G(1) - \int_0^1 G(r) \varphi_k'(r) dr \right] = \sigma \int_0^1 \varphi_k(r) dG(r) \qquad (10)$$

where the second equality is due to integration by parts and the fact that $G(0) = \int_0^0 B(s)ds = 0$ a.s.

Under Part (i) where $x_t \sim I(1)$, we have $\alpha = 3/2$ and $G(r) = \int_0^r B(s)ds$. Then (9) and (10) yield $n^{-1} I(\omega_{nk}) \to_d Z_k$ where $Z_k = \sigma \int_0^1 \varphi_k(r) B(r) dr$. Therefore, the random vector $Z = (Z_1, \ldots, Z_K)'$ follows the multivariate normal distribution

$$Z =_d \sigma N(0, \Sigma_K)$$

where $\Sigma_K = [\sigma_{jk}]$ is an $K \times K$ covariance matrix with entries

$$\sigma_{jk} = cov(Z_j, Z_k) = \sigma^2 \int_0^1 \int_0^1 \varphi_j(r) \varphi_k(s) \mathcal{C}(r, s) dr ds$$

where $\mathcal{C}(r, s)$ is the covariance function of the standard Brownian motion on $[0, 1]$. Since $\{\varphi_k(r)\}_{k=1}^{k}$ are the first K basis functions of the Karhunen-Loève representation of the standard Brownian motion on $[0,1]$, they are the first K eigenfunctions of $\mathcal{C}(r,s)$ with corresponding eigenvalues ω_k^{-2}. Therefore, we have

$$\sigma_{jk} = \omega_k^{-2}\delta_{jk}, \quad j,k = 1,\ldots,K,$$

where δ_{jk} denotes the Kronecker delta function. We can then deduce that $I(\omega_{nk})$ are asymptotically independent and $n^{-1}I(\omega_{nk}) \to_d N(0,\sigma^2\omega_k^{-2})$, as claimed in Part (i) of the proposition.

Under Part (ii) where $x_t \sim I(0)$, we have $\alpha = 1/2$ and $G(r) = B(r)$. Then (9) and (10) yield $I(\omega_{nk}) \to_d Z_k$ where $Z_k = \sigma \int_0^1 \varphi_k(r)dB(r)$. Therefore, the random vector $Z = (Z_1,\ldots,Z_K)'$ follows the multivariate normal distribution $Z =_d \sigma N(0,\Sigma_K)$ where $\Sigma_K = [\sigma_{jk}]$ is an $K \times K$ covariance matrix with entries

$$\begin{aligned}\sigma_{jk} = cov(Z_j,Z_k) &= \sigma^2 \int_0^1\int_0^1 \varphi_j(r)\varphi_k(s)E[dB(r)dB(s)] \\ &= \sigma^2 \int_0^1 \varphi_j(r)\varphi_k(r)dr = \sigma^2\delta_{jk}, \quad j,k = 1,\ldots,K.\end{aligned}$$

Therefore, we can deduce again that $I(\omega_{nk})$ are asymptotically independent, and $I(\omega_{nk}) \to_d N(0,\sigma^2)$, as claimed in Part (ii) of the proposition. This complete the proof of the proposition. □

Proof of Proposition 2. We start with $t = 1$. By definition of the matrix operation of HP filter, we have

$$\begin{aligned}\hat{T}_1 &= h(1,1)y_1 + h(1,2)y_2 + \ldots + h(1,n)y_n \\ &= h(1,1)y_1 + h(1,2)(y_1 + \Delta y_2) + \ldots + h(1,n)(y_1 + \Delta y_2 + \ldots \Delta y_n) \\ &= b(1,1)y_1 + \sum_{j=2}^{n} b(1,j)\Delta y_j,\end{aligned}$$

and hence $\hat{c}_1 = y_1 - \hat{T}_1 = [1 - b(1,1)]y_1 - \sum_{j=2}^n b(1,j)\Delta y_j$. When $t = 2$, note that

$$\begin{aligned}\hat{T}_2 &= h(2,1)y_1 + h(2,2)y_2 + \ldots + h(2,n)y_n \\ &= h(2,1)(y_2 - \Delta y_2) + h(2,2)y_2 + h(2,3)(y_2 + \Delta y_3) + \ldots + h(2,n)(y_2 + \Delta y_3 + \ldots \Delta y_n) \\ &= b(2,1)y_2 - b(2,1)\Delta y_2 + \sum_{j=3}^{n} b(2,j)\Delta y_j,\end{aligned}$$

from which it follows that $\hat{c}_2 = y_2 - \hat{T}_2 = [1 - b(2,1)]y_2 + d(2,1)\Delta y_2 - \sum_{j=3}^{n} b(2,j)\Delta y_j$. When $t = 3$, we have

$$\begin{aligned}
\hat{T}_3 &= h(3,1)y_1 + h(3,2)y_2 + \ldots + h(3,n)y_n \\
&= h(3,1)(y_3 - \Delta y_3 - \Delta y_2) + h(3,2)(y_3 - \Delta y_2) + \ldots + h(3,n)(y_3 + \Delta y_4 + \ldots \Delta y_n) \\
&= b(3,1)y_3 - (h(3,1) + h(3,2))\Delta y_2 - h(3,1)\Delta y_3 + \sum_{j=4}^{n} b(3,j)\Delta y_j \\
&= a(3,1)y_3 - d(3,2)\Delta y_2 - d(3,1)\Delta y_3 + \sum_{j=4}^{n} a(3,j)\Delta y_j,
\end{aligned}$$

and hence $\hat{c}_3 = y_3 - \hat{T}_3 = [1 - b(3,1)]y_3 + d(3,2)\Delta y_2 + d(3,1)\Delta y_3 - \sum_{j=4}^{n} b(3,j)\Delta y_j$. Analogously, we can easily show that the formula in the proposition holds for all $t = 1, \ldots, n$. $\square$