Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Macroeconomics 27 (2005) 53–67
Journal of MACROECONOMICS
www.elsevier.com/locate/econbase

On suboptimality of the Hodrick–Prescott filter at time series endpoints

Emi Mise [a], Tae-Hwan Kim [b,*], Paul Newbold [c]

[a] Department of Economics, The University of Edinburgh, Edinburgh EH8 9JY, UK
[b] Department of Economics, Yonsei University, Seoul 120-749, Republic of Korea
[c] Department of Economics, University of Nottingham, University Park, Nottingham, NG7 2RD, UK

Received 16 September 2002; accepted 5 September 2003
Available online 20 January 2005

Abstract

The Hodrick–Prescott filter is often applied to economic series as part of the study of business cycles. Its properties have most frequently been explored through the development of essentially asymptotic results which are practically relevant only some distance from series endpoints. Our concern here is with the most recent observations, as policy-makers will often require an assessment of whether, and by how much, an economic variable is "above trend". We show that if such an issue is important, an easily implemented adjustment to the filter is desirable.

JEL classification: E32; C22
Keywords: Hodrick–Prescott filter; Recent cyclical components; Revision error

1. Introduction

The statistical properties of the Hodrick–Prescott (HP) filter proposed by Hodrick and Prescott (1997) have been extensively analysed by, for example, King and

* Corresponding author. Tel.: +82 2 2123 5461; fax: +82 2 313 5331.
E-mail address: tae-hwan.kim@yonsei.ac.kr (T.-H. Kim).

0164-0704/$ - see front matter
doi:10.1016/j.jmacro.2003.09.003

Rebelo (1993) and Ehlgen (1998), from the viewpoint of optimal signal extraction and it is now well known that the HP filter yields an optimal decomposition of a time series into orthogonal components that can be regarded as "trend" and "cycle" if the time series is generated by a particular type of data generating process. This optimality result is based on application of the filter to an infinitely long time series, though for all practical purposes it applies also to the estimation of components at the centre of a moderately long series. However, our concern in this paper is with cyclical components estimation for the most recent time periods, which will be of most interest, for example, to policy makers. Results on HP optimality do not apply here, and indeed the filter is demonstrably suboptimal. This suboptimality has been noted in the literature (see Baxter and King, 1995; Apel et al., 1996; St-Amant and van Norden, 1997). In this paper, we explore the extent of that suboptimality of the HP filter at time series endpoints through extensive Monte Carlo simulations.

The paper is organised as follows. In Section 2, we demonstrate a fact which has not been widely understood in the literature, that the HP filter provides optimal estimators of components that could be viewed as "growth" and "cyclical" for any $I(1)$ or $I(2)$ generating model, whatever value is chosen for the smoothing parameter used in the HP filter. Based on the results in Section 2, we choose some particular $I(1)$ and $I(2)$ models for our analysis in the subsequent sections. In Section 3, we assume that there exists a "true" cyclical component, and that the purpose of the filter is to estimate that component. We further take the standpoint that the true component is the one for which HP yields optimal estimates at the series centre, and go on to assess the quality of the most recent HP figures as estimates of the most recent values of that component. The consequences of HP filtering at time series endpoints can be explored without overt recourse to the concept of "true" components; that is, through the notion of endpoint revisions as in Kaiser and Maravall (1999). Our simulation studies in Section 4 confirm the finding of Kaiser and Maravall (1999) that the use of forecast-augmented series in the HP filter can reduce the revision errors of most recent cyclical components. In addition, we find that the degree of suboptimality of the HP filter and the size of reduction of the revision errors depend on what value is used for the smoothing parameter of the HP filter. Finally, Section 5 concludes.

2. Some technical issues

Given a series of observations y_t $(t = 1, 2, \ldots, T)$ on a time series, the HP filter is an additive decomposition $y_t = y_t^{\mathrm{g}} + y_t^{\mathrm{c}}$ where y_t^{g} is identified as a growth (trend) component and y_t^{c} as a cyclical component. Hodrick and Prescott estimate the growth component as $\hat{y}_t^{\mathrm{g}}$ through solution of the constrained minimisation problem

$$\min_{[y_t^{\mathrm{g}}]_{t=1}^{T}} \sum_{t=1}^{T} (y_t - y_t^{\mathrm{g}})^2 + \lambda \sum_{t=2}^{T-1} \left[(y_{t+1}^{\mathrm{g}} - y_t^{\mathrm{g}}) - (y_t^{\mathrm{g}} - y_{t-1}^{\mathrm{g}}) \right]^2; \quad \lambda > 0, \tag{1}$$

where the parameter λ controls the smoothness of the estimated growth component. Hodrick and Prescott (1997) proposed, on somewhat subjective grounds, a value $\lambda = 1600$ for quarterly data. However, it is desirable to adjust this value when observations of different frequencies are subject to the filter. Backus and Kehoe (1992) suggested an adjustment of the value by multiplying the standard value of 1600 with the square of the frequency of observations relative to quarterly data. For example, the relative frequency is 3 for monthly data and 1/4 for annual data. Hence, the corresponding values of the smoothing parameter is $\lambda = 100$ and 14,400 for annual data and monthly data, respectively. This suggestion has been also used in commercial packages such as EVIEW. We shall use these values throughout the paper. With regard to the choice of the smoothing parameter, it is worth noting that, in research that has gone largely unnoticed in this field, Akaike (1980), while further allowing a seasonal component in the decomposition, proposed precisely the HP approach together with a data-dependent Bayesian procedure for the choice of λ.

Apart from the choice of λ, the structure of the HP filter is identical for all time series. In that sense, one might say that it is not intended to provide "optimal" cyclical component estimates $\hat{y}_t^c = (y_t - \hat{y}_t^g)$ for specific time series. Usually, results on the HP filter in the literature apply to infinitely long series, or in practical terms relate to the midpoints, but not the endpoints, of series of practically interesting length. In subsequent sections we shall be interested in endpoint issues, but here we review the asymptotic optimality results and demonstrate a fact that the HP filter can be regarded as "optimal" at the midpoints for any $I(1)$ or $I(2)$.

King and Rebelo (1993) and Ehlgen (1998) analysed the HP filter in this framework. It can be shown that the estimated cyclical component is provided by the symmetric two-sided filter

$$\hat{y}_t^c = H(L)y_t; \quad H(L) = \frac{(1-L)^2(1-L^{-1})^2}{\lambda^{-1} + (1-L)^2(1-L^{-1})^2}, \tag{2}$$

where L is the lag operator. The HP filter is optimal, in expected squared error sense, for data generating processes of the form

$$\begin{aligned} &(1-L)^2 y_t^g = A(L)\varepsilon_t; \quad y_t^c = A(L)u_t, \\ &A(L) = \sum_{j=0}^{\infty} a_j L^j; \quad \sum_{j=0}^{\infty} a_j^2 < \infty, \end{aligned} \tag{3}$$

where ε_t and u_t are mutually stochastically uncorrelated white noise processes, so that

$$E(\varepsilon_t u_s) = 0; \quad \forall t, s, \tag{4}$$

and where their variance ratio is

$$\lambda = (\sigma_u/\sigma_\varepsilon)^2 \tag{5}$$

where λ is the value of the smoothness parameter used in (1). It is generally agreed that a great many economic time series are integrated of order $d = 1$ or 2. There is some controversy as to which one is typically the more appropriate as shown in

Granger (1997) and Harvey (1997), but there is scant support for higher values. Hence, we shall restrict attention to such generating models with $d = 1$ or 2 for y_t.

Since y_t is the sum of the individual components, it follows from (3) that

$$\begin{aligned}(1-L)^2 y_t &= A(L)\varepsilon_t + (1-L)^2 A(L) u_t = A(L)[\varepsilon_t + (1-L)^2 u_t] \\ &= A(L)(1-\gamma_1 L - \gamma_2 L^2)\eta_t \end{aligned} \tag{6}$$

where η_t is white noise, whose variance along with the parameters γ_i depends through (5) on the smoothness parameter λ. For example, solving the usual autocovariance equalities and imposing the invertibility condition on $1-\gamma_1 L - \gamma_2 L^2$ yields

$$\gamma_1 = 1.56, \quad \gamma_2 = -0.64, \quad \sigma_\eta^2 = 1.57\sigma_u^2, \tag{7}$$

$$\gamma_1 = 1.78, \quad \gamma_2 = -0.80, \quad \sigma_\eta^2 = 1.25\sigma_u^2, \tag{8}$$

$$\gamma_1 = 1.87, \quad \gamma_2 = -0.88, \quad \sigma_\eta^2 = 1.14\sigma_u^2, \tag{9}$$

for annual ($\lambda = 100$), quarterly ($\lambda = 1600$), and monthly ($\lambda = 14{,}400$) frequencies respectively. Consider first the case where y_t is $I(2)$, with stationary autoregressive operator $\phi(L)$ of order p and invertible moving average operator $\theta(L)$ of order q in its generating model. Then in (6) set

$$A(L) = \frac{\theta(L)}{\phi(L)(1-\gamma_1 L - \gamma_2 L^2)} \tag{10}$$

so that

$$\phi(L)(1-L)^2 y_t = \theta(L)\eta_t. \tag{11}$$

Since there is no restriction, other than stationarity and invertibility, on the parameterisations $\phi(L)$ and $\theta(L)$, the implication is that, whatever the choice of λ, an optimal HP decomposition of the form (3) exists. Given σ_η^2 and λ, the variances σ_u^2 and σ_η^2 as well as the coefficients γ_i are determined simultaneously from (5) and (6). It follows from (11), (10), and (3) that in general, if the data generating process for y_t is $ARIMA(p,2,q)$, that for y_t^g is $ARIMA(p+2,2,q)$ and that for y_t^c is stationary $ARMA(p+2,q)$.

Now let y_t be $I(1)$, with stationary autoregressive operator $\phi(L)$ and invertible moving average operator $\theta(L)$, and in (6) set

$$A(L) = \frac{\theta(L)(1-L)}{\phi(L)(1-\gamma_1 L - \gamma_2 L^2)} \tag{12}$$

which is permissible as (3) does not preclude a unit moving average root in $A(L)$. Then

$$\phi(L)(1-L)y_t = \theta(L)\eta_t \tag{13}$$

and it immediately follows that if the process represented by (13) is $ARIMA(p,1,q)$, that for y_t^g is $ARIMA(p+2,1,q)$ and that for y_t^c is stationary $ARMA(p+2,q+1)$ with a unit moving average root. Of course, in such conclusions for either $I(2)$ or

$I(1)$ processes, the possibility exists of pathological cases where cancelling factors in the autoregressive and moving average operators generate lower dimensional generating models for the components series. For example, suppose $\theta(L) = 1 - \gamma_1 L - \gamma_2 L^2$ in (10). Then, the factors in the autoregressive and moving average operators in y_t^g and y_t^c are cancelled, which leads to the case where y_t is more heavily parameterised than the components series (y_t^g and y_t^c). More precisely, in this case, the process of y_t is $ARIMA(p,2,2)$ while those of y_t^g and y_t^c are $ARIMA(p,2,0)$ and $ARIMA(p,0,0)$ respectively.

3. Estimation of recent cyclical components

In this section, our concern is with the most recent HP cyclical component estimates in a series of finite length. Suppose that a time series is generated by the process (3), so that implicitly y_t^c is the cyclical component estimated by the filter. That estimate will be optimal at the centre of a "long" series. However, towards the series endpoints components estimates will in general be inefficient. That conclusion is easily seen in the present context. The filter (2) is symmetric two-sided, but of course such a filter is not directly applicable towards the end-points, and does not of necessity correspond to the solution of (1). However, as follows directly from results of Burman (1980), optimal components estimates follow from augmenting a given series y_t with optimal forecasts (and optimal backcasts if interest is also in the earliest values), and applying the filter to the augmented series. Here, we view the filter as an attempt to estimate the quantity y_t^c of (3), assess the suboptimality of the HP estimates of the most recent time periods, and demonstrate how the inefficiency can be improved by using a forecast-based augmentation to the filter.

Let the time series y_t be generated through (3), so that at the "centre" of a long series the HP filter optimally estimates the cyclical component y_t^c. We shall explore in detail the case where $A(L)$ is a first order autoregressive operator $(1 - aL)^{-1}$, $|a| < 1$, or a first-order moving average operator $(1 - bL)$, $|b| < 1$. As follows from (6) the generating processes for y_t in these cases are, respectively

$$(1 - aL)(1 - L)^2 y_t = (1 - \gamma_1 L - \gamma_2 L^2)\eta_t \tag{14}$$

and

$$(1 - L)^2 y_t = (1 - bL)(1 - \gamma_1 L - \gamma_2 L^2)\eta_t \tag{15}$$

where the γ_i depend on λ of (5). In the simulations that follow, we set $\lambda = 100$, 1600, or 14,400, and, without loss of generality, $\sigma_u^2 = 1$. Also, in these simulations, the white noise processes ε_t and u_t were taken to be independent Gaussian. The usual HP components estimate $\hat{y}_t^c$ then follows directly from y_t ($t = 1,2,\ldots,T$).

Each generated series was augmented by H minimum mean squared error-optimal forecasts, giving series $\tilde{y}_t$ $(t = 1,2,\ldots,T+H)$ where $\tilde{y}_t = y_t$ $(t = 1,2,\ldots,T)$, and the remaining elements of $\tilde{y}_t$ are forecasts based on y_{T-j} $(j = 0,1,2,\ldots)$ and (14) or (15) in the usual way. A detailed discussion on how to generate the forecasts $\tilde{y}_t$ $(t = T+1,\ldots,T+H)$ is provided in Appendix A. Estimation results are more or less invariant to T, provided that the sample size is moderately large, and in our simulations we set its value at 100. The theoretical conclusion on forecast-augmentation strictly requires forecasts infinitely far ahead. However, the weights given by the filter to distant forecasts become negligible. After some experimentation with both real and generated data, we found it sufficient to fix $H = 28$ (corresponding to seven years of quarterly data). We denote by $\widehat{\tilde{y}_t^c}$ the estimated cyclical components obtained by applying the HP filter to $\tilde{y}_t$ $(t = 1, 2, \ldots, T + H)$.

In our experiments, cyclical components are of course known quantities, given by (3) with $A(L) = (1 - aL)^{-1}$ or $A(L) = (1 - bL)$, and in our simulations directly generated from these processes, so it is straightforward to assess the precision of their estimation. We measured this through the standard deviation of estimation error, that is $(y_{T-j}^c - \hat{y}_{T-j}^c)$ for the standard HP filter and $(y_{T-j}^c - \widehat{\tilde{y}_{T-j}^c})$ for the filter applied to the forecast-augmented series, estimated through 5000 replications. These estimates are denoted s and s_f, respectively. The latter, of course, estimates the error standard deviation of optimal estimates of y_t^c of (3). Results for the $AR(1)$ and $MA(1)$ representations of $A(L)$ are given respectively in Tables 1 and 2 for estimation of the cyclical components y_{T-j}^c $(j = 0, 1, 2, 10)$ where Tables 1(a), 2(a) are for $\lambda = 100$, Tables 1(b), 2(b) for $\lambda = 1600$ and Tables 1(c), 2(c) for $\lambda = 14{,}400$. The estimates based on forecast-augmentation are squared error-optimal, and the results of these tables demonstrate the general suboptimality of the HP filter as an estimator of recent cyclical components when that filter is known to provide optimal estimates of such components at a series "centre". The degree of that suboptimality strongly depends on both the values of the model parameters and the values of the smoothing parameter λ. When the typical quarterly value $\lambda = 1600$ is used, the suboptimality of the HP filter is most visible in Table 1(b) for low negative a and in Table 2(b) for high positive b, both of which correspond to substantial negative first autocorrelation in the "true" cyclical component of (3). For the $AR(1)$ case, the suboptimality of the HP filter at the series endpoints becomes more pronounced for the annual frequency ($\lambda = 100$) and less for the monthly frequency ($\lambda = 14{,}400$). On the other hand, when $A(L) = (1 - bL)$, the degree of suboptimality does not change much across various values of λ. As is to be expected, the relative efficiency of the standard HP estimators gradually increases with increasing distance from the series endpoints. By the stage that the observation of interest is ten values from the series end, standard HP is virtually fully efficient in all cases.

The actual values of s are worth investigating. For the most recent observations, these estimation error standard deviations are far from negligible compared to the standard deviation of the cyclical component y_t^c. The standard deviation of y_t^c is $\sqrt{(1 - a^2)^{-1}}$ and $\sqrt{(1 - b^2)}$ for $A(L) = (1 - aL)^{-1}$ and $A(L) = (1 - bL)$, respectively. We tabulate these values for comparison purpose in columns 2–3 (under the headings of DGP1 and DGP2) of Table 5 for the considered values of a and b. The fact that the actual values of s are not proportional to the standard deviation of y_t^c indicates that the behaviour of the trend component y_t^g may play a dominant role in the composite series y_t and may have a larger effect on the size of s than the cyclical component. When $A(L) = (1 - aL)^{-1}$, y_t^g follows an $ARIMA(1,2,0)$

Table 1
Standard deviations of estimators of recent cyclical components in generating processes (3) for which HP is theoretically optimal: (a) $A(L) = (1 - aL)^{-1}$, $\lambda = 100$; (b) $A(L) = (1 - aL)^{-1}$, $\lambda = 1600$ and (c) $A(L) = (1 - aL)^{-1}$, $\lambda = 14{,}400$

Observation	T		$T-1$		$T-2$		$T-10$	
a	s	s_f/s	s	s_f/s	s	s_f/s	s	s_f/s
Panel (a)								
−0.9	0.56	0.64	0.43	0.67	0.33	0.72	0.19	0.99
−0.8	0.46	0.79	0.36	0.81	0.29	0.85	0.20	0.99
−0.7	0.45	0.87	0.36	0.88	0.29	0.91	0.20	0.99
−0.6	0.45	0.92	0.36	0.93	0.29	0.95	0.22	1.00
−0.5	0.45	0.95	0.37	0.95	0.30	0.96	0.23	1.00
−0.4	0.47	0.97	0.38	0.97	0.32	0.98	0.25	1.00
−0.3	0.50	0.98	0.40	0.98	0.34	0.99	0.27	1.00
−0.2	0.53	0.99	0.43	0.99	0.35	1.00	0.29	1.00
−0.1	0.56	1.00	0.45	1.00	0.38	1.00	0.31	1.00
0	0.60	1.00	0.49	1.00	0.42	1.00	0.34	1.00
0.1	0.65	1.00	0.53	1.00	0.45	1.00	0.38	1.00
0.2	0.71	0.99	0.58	0.99	0.50	1.00	0.42	1.00
0.3	0.80	0.99	0.65	0.99	0.56	0.99	0.48	1.00
0.4	0.90	0.97	0.73	0.98	0.63	0.99	0.55	1.00
0.5	1.03	0.95	0.84	0.97	0.72	0.98	0.64	1.00
0.6	1.23	0.92	1.01	0.94	0.87	0.97	0.78	1.00
0.7	1.47	0.90	1.21	0.93	1.07	0.97	1.00	1.00
0.8	1.90	0.86	1.58	0.91	1.41	0.96	1.32	1.00
0.9	2.88	0.80	2.42	0.90	2.18	0.96	2.04	0.99
Panel (b)								
−0.9	0.34	0.74	0.30	0.74	0.27	0.75	0.13	0.97
−0.8	0.30	0.86	0.27	0.87	0.24	0.87	0.14	0.99
−0.7	0.30	0.93	0.27	0.93	0.24	0.93	0.14	1.00
−0.6	0.31	0.95	0.28	0.95	0.25	0.95	0.15	1.00
−0.5	0.32	0.97	0.29	0.97	0.26	0.97	0.16	1.00
−0.4	0.34	0.98	0.30	0.98	0.27	0.98	0.17	1.00
−0.3	0.36	0.99	0.32	0.99	0.29	0.99	0.19	1.00
−0.2	0.38	1.00	0.34	1.00	0.31	1.00	0.20	1.00
−0.1	0.41	1.00	0.37	1.00	0.33	1.00	0.22	1.00
0	0.45	1.00	0.40	1.00	0.36	1.00	0.24	1.00
0.1	0.49	1.00	0.44	1.00	0.40	1.00	0.27	1.00
0.2	0.55	0.99	0.49	0.99	0.44	1.00	0.30	1.00
0.3	0.62	0.99	0.56	0.99	0.50	0.99	0.34	1.00
0.4	0.70	0.98	0.63	0.99	0.56	0.99	0.40	1.00
0.5	0.82	0.97	0.73	0.98	0.66	0.98	0.48	1.00
0.6	1.01	0.95	0.90	0.95	0.82	0.96	0.59	1.00
0.7	1.26	0.93	1.14	0.93	1.03	0.94	0.78	1.00
0.8	1.71	0.89	1.54	0.90	1.40	0.92	1.08	1.00
0.9	2.83	0.81	2.55	0.84	2.33	0.87	1.82	1.00
Panel (c)								
−0.9	0.23	0.82	0.22	0.82	0.20	0.82	0.12	0.89
−0.8	0.22	0.91	0.20	0.92	0.19	0.92	0.12	0.96
−0.7	0.22	0.96	0.21	0.96	0.19	0.96	0.12	0.98
−0.6	0.23	0.97	0.22	0.97	0.20	0.97	0.13	0.98

(*continued on next page*)

Table 1 (*continued*)

Observation	T		$T-1$		$T-2$		$T-10$	
a	s	s_f/s	s	s_f/s	s	s_f/s	s	s_f/s
−0.5	0.24	0.98	0.23	0.98	0.21	0.98	0.14	0.99
−0.4	0.26	0.99	0.24	0.99	0.23	0.99	0.15	1.00
−0.3	0.27	1.00	0.26	1.00	0.24	1.00	0.16	1.00
−0.2	0.30	1.00	0.28	1.00	0.26	1.00	0.17	1.00
−0.1	0.32	1.00	0.30	1.00	0.28	1.00	0.19	1.00
0	0.35	1.00	0.33	1.00	0.31	1.00	0.21	1.00
0.1	0.38	1.00	0.36	1.00	0.34	1.00	0.23	1.00
0.2	0.42	1.00	0.40	1.00	0.37	1.00	0.26	1.00
0.3	0.49	1.00	0.45	1.00	0.43	1.00	0.29	1.00
0.4	0.56	1.00	0.52	1.00	0.49	1.00	0.34	1.00
0.5	0.66	0.99	0.62	0.99	0.58	0.99	0.40	0.99
0.6	0.82	0.97	0.77	0.97	0.72	0.97	0.50	0.99
0.7	1.06	0.95	1.00	0.95	0.94	0.95	0.66	0.98
0.8	1.48	0.92	1.39	0.92	1.31	0.92	0.94	0.98
0.9	2.60	0.85	2.44	0.86	2.29	0.86	1.65	0.97

given by $(1-aL)(1-L)^2 y_t^g = \varepsilon_t$. Hence, as a approaches 1, y_t^g becomes $I(3)$, which can explain why the values of s are so large for high positive a in Table 1(a)–(c). On the other hand, when $A(L) = (1 - bL)$, the trend component follows and $ARIMA(0,2,1)$ process given by $(1-L)^2 y_t^g = (1-bL)\varepsilon_t$. As $b \to 1$, y_t^g becomes $I(1)$ while, as $b \to -1$, y_t^g behaves more like a sum of two perfectly correlated $I(2)$ processes (in fact, in the limit, we have $y_t^g = \sum_{s=1}^{t}\sum_{k=1}^{s}\varepsilon_k + \sum_{s=1}^{t}\sum_{k=1}^{s}\varepsilon_{k-1}$). Hence, it is not surprising to have large estimation error variance for low negative b in Table 2(a)–(c). As Tables 1 and 2 show, in the above mentioned cases, the values of s are disturbingly large. The implication must be that, even if the components decomposition (3), implied by HP filter optimality, is viewed as "reasonable", estimation of such a decomposition can be quite imprecise.

Taken together, the results of Tables 1 and 2 demonstrate that, while suboptimality of the HP estimators at or near the endpoints of a series is a priori obvious, even in cases where HP is theoretically optimal at the series centre, the extent of that suboptimality can be serious. It must be concluded that there is no generating process for which HP yields optimal estimators of the cyclical component at all time periods, though it is clear from the tables that in some cases it comes close to doing so.

4. Revision of most recent cyclical components

In Section 2, we noted that, although it was not specifically developed with that purpose in mind, for any $I(1)$ or $I(2)$ process y_t, the HP estimated cyclical component could be viewed as an optimal estimator, in squared-error-loss sense, of a "true" cyclical component defined in a specific way. In Section 3 we saw that the HP estimators of the most recent values could be far from optimal. In this section we shall examine what is essentially the same issue from a somewhat different perspective, superficially abandoning the notion of a "true" component.

Table 2
Standard deviations of estimators of recent cyclical components in generating processes (3) for which HP is theoretically optimal: (a) $A(L) = (1 - bL)$, $\lambda = 100$; Panel (b) $A(L) = (1 - bL)$, $\lambda = 1600$; and Panel (c) $A(L) = (l - bL)$, $\lambda = 14{,}400$

Observation	T		$T-1$		$T-2$		$T-10$	
b	s	s_f/s	s	s_f/s	s	s_f/s	s	s_f/s
Panel (a)								
−0.9	1.08	0.96	0.88	0.97	0.75	0.98	0.64	1.00
−0.8	1.00	0.96	0.82	0.97	0.51	0.98	0.60	1.00
−0.7	0.97	0.97	0.80	0.97	0.68	0.98	0.57	1.00
−0.6	0.92	0.97	0.75	0.98	0.64	0.99	0.54	1.00
−0.5	0.86	0.98	0.70	0.98	0.60	0.99	0.50	1.00
−0.4	0.80	0.99	0.65	0.99	0.56	0.99	0.47	1.00
−0.3	0.75	0.99	0.62	0.99	0.53	0.99	0.44	1.00
−0.2	0.71	1.00	0.57	1.00	0.48	1.00	0.41	1.00
−0.1	0.65	1.00	0.53	1.00	0.45	1.00	0.37	1.00
0	0.60	1.00	0.49	1.00	0.42	1.00	0.34	1.00
0.1	0.56	1.00	0.45	1.00	0.38	1.00	0.31	1.00
0.2	0.51	0.99	0.41	1.00	0.34	1.00	0.28	1.00
0.3	0.48	0.97	0.38	0.97	0.32	0.98	0.25	1.00
0.4	0.44	0.95	0.35	0.96	0.29	0.97	0.22	1.00
0.5	0.41	0.92	0.32	0.93	0.56	0.94	0.19	1.00
0.6	0.39	0.84	0.31	0.86	0.24	0.89	0.16	0.99
0.7	0.39	0.76	0.30	0.79	0.23	0.83	0.14	0.99
0.8	0.38	0.67	0.29	0.70	0.22	0.76	0.12	0.98
0.9	0.39	0.57	0.29	0.62	0.22	0.69	0.11	0.98
Panel (b)								
−0.9	0.83	0.99	0.74	0.99	0.67	0.99	0.46	1.00
−0.8	0.78	0.99	0.70	0.99	0.63	0.99	0.44	1.00
−0.7	0.75	0.99	0.67	0.99	0.61	0.99	0.41	1.00
−0.6	0.70	0.99	0.63	0.99	0.57	0.99	0.38	1.00
−0.5	0.66	0.99	0.59	0.99	0.53	1.00	0.36	1.00
−0.4	0.61	1.00	0.55	1.00	0.50	1.00	0.34	1.00
−0.3	0.57	1.00	0.52	1.00	0.46	1.00	0.32	1.00
−0.2	0.53	1.00	0.48	1.00	0.43	1.00	0.29	1.00
−0.1	0.49	1.00	0.44	1.00	0.40	1.00	0.27	1.00
0	0.45	1.00	0.41	1.00	0.37	1.00	0.25	1.00
0.1	0.41	1.00	0.37	1.00	0.33	1.00	0.22	1.00
0.2	0.37	1.00	0.33	1.00	0.30	1.00	0.19	1.00
0.3	0.34	0.99	0.30	0.99	0.27	0.99	0.17	1.00
0.4	0.30	0.97	0.27	0.97	0.24	0.97	0.15	1.00
0.5	0.27	0.94	0.24	0.94	0.21	0.94	0.13	1.00
0.6	0.24	0.88	0.22	0.88	0.19	0.89	0.10	0.99
0.7	0.22	0.78	0.20	0.79	0.17	0.80	0.08	0.98
0.8	0.21	0.65	0.19	0.66	0.16	0.68	0.06	0.96
0.9	0.21	0.50	0.18	0.51	0.16	0.53	0.05	0.91
Panel (c)								
−0.9	0.65	0.99	0.61	0.99	0.57	0.99	0.39	1.00
−0.8	0.62	0.99	0.58	0.99	0.55	0.99	0.37	1.00
−0.7	0.59	0.99	0.55	0.99	0.52	0.99	0.34	1.00
−0.6	0.55	1.00	0.52	1.00	0.49	1.00	0.33	1.00

(*continued on next page*)

Table 2 (*continued*)

Observation	T		$T-1$		$T-2$		$T-10$	
a	s	s_f/s	s	s_f/s	s	s_f/s	s	s_f/s
−0.5	0.52	1.00	0.48	1.00	0.45	1.00	0.30	1.00
−0.4	0.48	1.00	0.45	1.00	0.42	1.00	0.29	1.00
−0.3	0.45	1.00	0.42	1.00	0.39	1.00	0.27	1.00
−0.2	0.42	1.00	0.39	1.00	0.37	1.00	0.25	1.00
−0.1	0.38	1.00	0.36	1.00	0.34	1.00	0.22	1.00
0	0.35	1.00	0.33	1.00	0.31	1.00	0.21	1.00
0.1	0.31	1.00	0.29	1.00	0.28	1.00	0.19	1.00
0.2	0.28	1.00	0.26	1.00	0.25	1.00	0.17	1.00
0.3	0.25	0.99	0.24	0.99	0.22	0.99	0.15	0.99
0.4	0.22	0.98	0.21	0.98	0.20	0.98	0.13	0.99
0.5	0.19	0.97	0.18	0.97	0.17	0.96	0.11	0.98
0.6	0.17	0.91	0.16	0.91	0.15	0.91	0.09	0.95
0.7	0.15	0.83	0.14	0.83	0.13	0.86	0.08	0.90
0.8	0.13	0.69	0.12	0.69	0.11	0.69	0.06	0.79
0.9	0.12	0.51	0.12	0.51	0.11	0.51	0.06	0.61

Consider again a time series y_t $(t = 1, 2, \ldots, T)$ and let $\hat{y}_t^c$ $(t = 1, 2, \ldots, T)$ be the HP cyclical component, following in the usual way through (1): specifically, we concentrate on the most recent of these $\hat{y}_T^c$. Suppose now that H time periods have elapsed, so the analyst now has access to y_t $(t = 1, 2, \ldots, T + H)$. The analyst could then pass this entire extended series through the HP filter, obtaining a new estimate $\hat{y}_T^{c*}$ of the cyclical component at time T, revising the original estimate by an amount $(\hat{y}_T^{c*} - \hat{y}_T^c)$. Of course, some revision of this sort would be inevitable, but it seems reasonable to take the view that one would like it to be as small as possible—that is, the standard deviation s of the revision should ideally be no larger than is necessary. The issue of revision size can be directly explored in terms of the generating process for a given series y_t, without recourse to explicit specification of components generating models. We do so here for two types of $I(1)$ processes—the $ARIMA(1,1,0)$ model

$$(1 - \phi L)(1 - L)y_t = \eta_t; \quad |\phi| < 1, \tag{16}$$

and the $ARIMA(0,1,1)$ model

$$(1 - L)y_t = (1 - \theta L)\eta_t; \quad |\theta| < 1. \tag{17}$$

We generated series of T observations from these processes with η_t independent Gaussian with mean 0 and variance $\sigma_\eta^2 = 1$, and applied the HP filter. Generation was continued for H subsequent observations and the filter was also applied to the extended series so that revisions could be calculated: their standard deviations were estimated through 5000 replications. The result is virtually invariant to T, provided that number is moderately large: here we took $T = 100$. The quantity H is chosen sufficiently large for the revision process to "settle down". As in the previous section, we found $H = 28$ to be sufficient.

In fact, the HP filter is easily modified to yield smaller revisions. Define the forecast-augmented series $\tilde{y}_t$ $(t = 1, 2, \ldots, T + H)$ precisely as in the previous section and apply the full HP filter to the complete series $\tilde{y}_t$, taking the estimated cyclical component at time T, $\widehat{\tilde{y}_T^c}$, as an alternative estimator of the time T cyclical component. It should be emphasised that $\widehat{\tilde{y}_T^c}$, depends only on data available at time T—that is, on y_{T-j} ($j \geqslant 0$). For the two models of our study, forecasts can be obtained directly from (16) and (17), employing the same methodology discussed in Appendix A. It is quite clear that such an approach minimises revision standard deviation. The estimated quantity is simply $\hat{y}_T^{c*}$, which is precisely the same linear function of y_t ($t = 1, 2, \ldots, T + H$) as is $\widehat{\tilde{y}_T^c}$ of the forecast augmented series $\tilde{y}_t$ ($t = 1, 2, \ldots, T + H$). But since those forecasts in $\tilde{y}_t$ are minimum mean squared error, so must be $\widehat{\tilde{y}_T^c}$ for the corresponding linear function of y_t ($t = 1, 2, \ldots, T + H$). We estimated in our simulations s_f, the standard deviation of revisions ($\hat{y}_T^{c*} - \widehat{\tilde{y}_T^c}$) of the estimated time T cyclical components when this forecast-augmented approach is used in conjunction with the HP filter, calculating the ratios s_f/s. Simulation results for the generating processes (16) and (17) are given respectively in Tables 3 and 4 for a range of parameter values. When y_t is generated by (16), it can be shown that y_t^c follows an $ARMA$(3,1) process given by $(1 - \phi L)(1 - \gamma_1 L - \gamma_2 L)y_t^c = (1 - L)u_t$. Similarly, when y_t follows (17), the process for y_t^c is $ARMA(2,2)$ given by $(1 - \gamma_1 L - \gamma_2 L)y_t^c = (1 - L)(1 - \theta L)u_t$. The standard deviation of y_t^c in each case can be calculated using the exact formulae for the theoretical autocovariance of general $ARMA(p,q)$ models in McLeod (1975). The standard deviation depends on the smoothing parameter λ through γ_1 and γ_2. The actual values for the considered cases are provided in columns 4–6 (under the heading of DGP3) for the $ARMA(3,1)$ case and 7–9 (under the heading of DGP4) for the $ARMA(2,2)$ case in Table 5. The behaviour of the cyclical

Table 3
Standard deviations of revisions of most recent HP cyclical components when $(1 - \phi L)(1 - L)y_t = \varepsilon_t$

λ	100		1600		14,400	
ϕ	s	s_f/s	s	s_f/s	s	s_f/s
−0.9	0.49	0.79	0.65	0.79	0.82	0.74
−0.8	0.49	0.83	0.67	0.79	0.86	0.74
−0.7	0.50	0.85	0.72	0.79	0.93	0.73
−0.6	0.54	0.84	0.76	0.79	0.98	0.73
−0.5	0.56	0.84	0.81	0.78	1.04	0.73
−0.4	0.58	0.84	0.87	0.78	1.12	0.73
−0.3	0.64	0.84	0.95	0.78	1.20	0.73
−0.2	0.69	0.82	1.01	0.77	1.29	0.72
−0.1	0.74	0.80	1.10	0.76	1.40	0.71
0	0.80	0.79	1.20	0.75	1.53	0.70
0.1	0.88	0.79	1.32	0.75	1.72	0.70
0.2	0.97	0.77	1.49	0.73	1.94	0.69
0.3	1.08	0.74	1.69	0.72	2.21	0.68
0.4	1.24	0.72	1.97	0.69	2.54	0.67
0.5	1.42	0.70	2.35	0.68	3.08	0.65
0.6	1.65	0.67	2.83	0.65	3.84	0.63
0.7	1.92	0.62	3.50	0.61	4.85	0.60
0.8	2.27	0.58	4.69	0.56	6.94	0.55
0.9	2.72	0.52	6.57	0.49	11.03	0.48

Table 4
Standard deviations of revisions of most recent HP cyclical components when $(1 - L)yt = (1 - \theta L)\varepsilon_t$

λ	100		1600		14,400	
ϕ	s	s_f/s	s	s_f/s	s	s_f/s
−0.9	1.47	0.71	2.27	0.71	2.94	0.68
−0.8	1.39	0.71	2.13	0.71	2.78	0.68
−0.7	1.31	0.73	2.05	0.71	2.67	0.68
−0.6	1.28	0.73	1.98	0.72	2.50	0.68
−0.5	1.19	0.74	1.80	0.72	2.33	0.68
−0.4	1.08	0.75	1.69	0.72	2.18	0.69
−0.3	1.04	0.77	1.57	0.75	2.01	0.70
−0.2	0.96	0.77	1.45	0.75	1.85	0.70
−0.1	0.88	0.78	1.32	0.75	1.70	0.70
0	0.80	0.79	1.20	0.75	1.53	0.70
0.1	0.73	0.83	1.08	0.78	1.40	0.72
0.2	0.67	0.83	0.97	0.78	1.25	0.72
0.3	0.60	0.83	0.86	0.79	1.10	0.73
0.4	0.54	0.84	0.75	0.79	0.93	0.74
0.5	0.49	0.84	0.65	0.81	0.80	0.76
0.6	0.44	0.84	0.54	0.84	0.66	0.77
0.7	0.40	0.79	0.44	0.83	0.50	0.80
0.8	0.38	0.72	0.36	0.82	0.37	0.82
0.9	0.37	0.63	0.30	0.72	0.26	0.78

Table 5
Standard deviations of HP cyclical components for various DGP's: $\phi(L)y_t^c = \theta(L)\varepsilon_t$

	DGP1	DGP2	DGP3			DGP4		
			100	1600	14,400	100	1600	14,400
−0.9	2.29	1.35	1.27	1.39	1.55	1.79	2.69	3.64
−0.8	1.67	1.28	1.01	1.17	1.38	1.70	2.55	3.45
−0.7	1.40	1.22	0.92	1.12	1.35	1.61	2.41	3.26
−0.6	1.25	1.17	0.89	1.11	1.38	1.52	2.27	3.07
−0.5	1.15	1.12	0.89	1.14	1.43	1.43	2.14	2.88
−0.4	1.09	1.08	0.90	1.18	1.50	1.35	2.00	2.70
−0.3	1.05	1.04	0.92	1.23	1.58	1.27	1.87	2.51
−0.2	1.02	1.02	0.95	1.30	1.69	1.19	1.74	2.33
−0.1	1.01	1.00	1.00	1.39	1.82	1.12	1.62	2.15
0	1.00	1.00	1.06	1.50	1.98	1.06	1.50	1.98
0.1	1.01	1.00	1.13	1.63	2.17	1.00	1.38	1.81
0.2	1.02	1.02	1.21	1.79	2.41	0.95	1.28	1.64
0.3	1.05	1.04	1.31	1.99	2.71	0.90	1.18	1.49
0.4	1.09	1.08	1.44	2.24	3.10	0.87	1.09	1.34
0.5	1.15	1.12	1.60	2.58	3.63	0.85	1.02	1.21
0.6	1.25	1.17	1.79	3.03	4.38	0.85	0.97	1.10
0.7	1.40	1.22	2.04	3.67	5.52	0.86	0.94	1.02
0.8	1.67	1.28	2.35	4.61	7.42	0.88	0.93	0.97
0.9	2.29	1.35	2.74	6.07	11.13	0.91	0.94	0.96

DGP1: $\phi(L) = 1 - aL$ and $\theta(L) = 1$; DGP2: $\phi(L) = 1$ and $\theta(L) = 1 - bL$; DGP3: $\phi(L) = (1 - \phi L)(1 - \gamma_1 L - \gamma_2 L)$ and $\theta(L) = 1 - L$; and DGP4: $\phi(L) = (1 - \gamma_1 L - \gamma_2 L)$ and $\theta(L) = (1 - L)(1 - \theta L)$.

component y_t^c is highly sensitive to the values of ϕ, θ and λ. As $\phi \to 1$ and $\theta \to -1$, the cyclical component becomes more volatile.

It can be seen from Tables 3 and 4 that, compared with the standard deviation of the cyclical component y_t^c, the revision standard deviations s for the usual HP cyclical component can be very large for all values of λ, particularly when first differences of the series are positively autocorrelated (i.e. ϕ close to 1 and θ close to -1). Given that the behaviour of the cyclical component y_t^c becomes more volatile for such values, it is expected that revision standard deviations increase as $\phi \to 1$ and $\theta \to -1$. It is, however, important to note that the relative size of s (compared to the standard deviation of y_t^c) is also increasing in such cases. This phenomenon can be somewhat mitigated if the filter is applied to the forecast-augmented series, which will generally lead to reductions of at least 20%, and in some cases much more, in these revision standard deviations. When y_t is follows $ARIMA$(1,1,0), larger reductions are generally achieved for higher frequency values of λ regardless of the values of ϕ as shown in Table 3. This observation is also applied to the $ARIMA(0,1,1)$ case (Table 4) except the large positive values ($\theta = 0.7$, 0.8 and 0.9) of the MA coefficient. The results in this section confirm and strengthen the finding of Kaiser and Maravall (1999) by showing the suboptimality of the HP filter at time series endpoints for various data generating processes and for various values of the smoothing parameters.

5. Conclusions

The Hodrick–Prescott filter is often applied to individual economic time series as an initial step in real business cycle analyses. The filter generates cyclical components, which are then subjected to further analysis. Although the view is implicitly taken that actual time series are made up of the sum of growth and cyclical components, little attention is paid to either the structures of or relationship between those components. In particular, the HP filter was not developed to optimally estimate specific unobserved components, but rather is presented as an intuitively plausible transformation. Whether or why this should be so is not our concern.

In Section 2 we note that, whatever the intention, the HP filter does optimally estimate a particular components decomposition, and one might take the view that, inadvertently or otherwise, that is precisely the decomposition that is being estimated when the filter is applied. As we have noted, a number of previous authors have analysed HP from this viewpoint. However, the optimality conclusion strictly applies to infinitely long time series, or from a practical viewpoint to the midpoints of series of typical length. It does not apply at or close to series endpoints. Since the most recent cyclical components might be viewed by practitioners as of most interest, it seems reasonable to analyse the performance of the HP filter here.

At the endpoints the filter is demonstrably suboptimal, and it is easy to construct a modification whose performance is superior from two different, but closely related, perspectives. We examined those in turn in Sections 3 and 4. In Section 3, the "true" cyclical component was taken to be that implied by the optimality results of Section 2, and the estimation of the most recent values of this component was analysed. It might be objected that HP was not explicitly developed as a components estimator, and moreover that the optimal decompositions of Section 2 are not unique, since they impose orthogonality of the trend and cycle, though there is no particular reason to view such a restriction as plausible. In Section 4, we view the filter's output at the series endpoints in terms of revisions—that is, changes to initial components estimates that would inevitably occur as new data became available. It seems reasonable to argue that, on average, the magnitude of such revisions should be as small as possible.

The results of Sections 3 and 4 demonstrate, for specific special model cases, the non-trivial suboptimality of the usual HP filter from both perspectives at series endpoints. Moreover, it is seen that, from each perspective, a simple easily applied remedy generating significant improvements is readily available. The suboptimality of the HP filter at time series endpoints and our forecast-augmentation method can be of great interest to policy makers; especially, central bank economists who like to measure the current output gap as precisely as possible. Research on the practical implementation of our method to estimate the current output gap for the G-7 countries is ongoing.

Appendix A

In this section we explain how to generate the forecasts $\tilde{y}_t$ $(t = T+1, \ldots, T+H)$ used in Section 3. We consider the $AR(1)$ case only since a similar procedure has been used for the $MA(1)$ case. If $A(L) = (1 - aL)^{-1}$, then y_t follows an $ARIMA(1,2,2)$ process:

$$(1 - aL)(1 - L)^2 y_t = (1 - \gamma_1 L - \gamma_2 L)\eta_t,$$

which can be expressed as an $AR(\infty)$ representation as follows:

$$y_t = \sum_{i=1}^{\infty} \pi_i y_{t-i} + \eta_t$$

where $\pi_1 = 2 + a - \gamma_1$, $\pi_2 = -(1 + 2a) + \gamma_1\pi_1 - \gamma_2$, $\pi_3 = a + \gamma_1\pi_2 + \gamma_2\pi_1$ and $\pi_j = \gamma_1\pi_{j-1} + \gamma_2\pi_{j-2}$ for $j \geqslant 4$. The recursive formula for π_j is truncated at T since there are only T observations available in the generation of the forecast $\tilde{y}_{T+1}$. Using the AR(T) approximation, the optimal point forecasts $\tilde{y}_t$ $(t = T+1, \ldots, T+H)$ is given by

$$\tilde{y}_{T+j} = \mathrm{e}_1' A^j Y_T \tag{A.1}$$

where $Y_T = (y_T, y_{T-1}, \ldots, y_2, y_1)'$, $A = \begin{bmatrix} \pi' \\ I_{T-1} 0_{(T-1)\times 1} \end{bmatrix}$, $\pi = (\pi_1, \pi_2, \ldots, \pi_T)'$ and e_1 is a $T \times 1$ vector with one in the first element and zeros elsewhere.

Hence, $\tilde{y}_{T+j}$ is a linear function of Y_T as

$$\tilde{y}_{T+j} = \sum_{i=1}^{T} h_{ji} y_{T-i+1}$$

where h_{ji} determined by (A.1).

References

Akaike, H., 1980. Seasonal adjustment by a Bayesian modeling. Journal of Time Series Analysis 1, 1–13.

Apel, M., Hansen, J., Lindberg, H., 1996. Potential output and output gap. Quarterly Review of the Bank of Sweden 3, 24–35.

Backus, D.K., Kehoe, P.J., 1992. International evidence on the historical properties of business cycles. American Economic Review 82, 864–888.

Baxter, M., King, R.G., 1995. Measuring business cycles: Approximate band-pass filters for economic time series. NBER Working Paper 5022.

Burman, J.P., 1980. Seasonal adjustment by signal extraction. Journal of Royal Statistical Society A 143, 321–337.

Ehlgen, J., 1998. Distortionary effects of the optimal Hodrick–Prescott filter. Economics Letters 61, 345–349.

Granger, C.W.J., 1997. On modelling the long run in applied economics. Economic Journal 107, 169–177.

Harvey, A.C., 1997. Trends, cycles and autoregressions. Economic Journal 107, 192–201.

Hodrick, R.J., Prescott, E.C., 1997. Post-war US business cycles: An empirical investigation. Journal of Money, Credit, and Banking 29, 1–16.

Kaiser, R., Maravall, A., 1999. Estimation of the business cycles: A modified Hodrick–Prescott filter. Spanish Economic Review 1, 175–206.

King, R.G., Rebelo, S.T., 1993. Low frequency filtering and real business cycles. Journal of Economic Dynamics and Control 17, 207–232.

McLeod, I., 1975. Derivation of the theoretical autocovariance function of autoregressive—moving average time series. Applied Statistics 24, 255–256.

St-Amant, P., van Norden, S., 1997. Measurement of the output gap: A discussion of recent research at the Bank of Canada. Technical Report of the Bank of Canada, no. 79.