UNIVERSITY OF NOTTINGHAM

Discussion Papers in Economics

Discussion Paper
No. 03/08

THE HODRICK-PRESCOTT FILTER AT TIME SERIES ENDPOINTS

by Emi Mise, Tae-Hwan Kim and Paul Newbold

May 2003

DP 03/08

ISSN 1360-2438

UNIVERSITY OF NOTTINGHAM

Discussion Papers in Economics

Discussion Paper
No. 03/08

THE HODRICK-PRESCOTT FILTER AT TIME SERIES ENDPOINTS

by Emi Mise, Tae-Hwan Kim and Paul Newbold

Emi Mise is Lecturer, University of Edinburgh, Tae-Hwan Kim is Lecturer and Paul Newbold is Professor, both in the School of Economics, University of Nottingham

May 2003

The Hodrick-Prescott filter at time series endpoints

Emi Mise, Tae-Hwan Kim and Paul Newbold

Department of Economics, University of Nottingham
Nottingham, NG7 2RD, UK

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.

1 Introduction

A common initial step in empirical business cycle analysis is to de-trend a given time series in an attempt to isolate "its cyclical component." The outcome will depend of course on precisely what approach is taken to de-trending. Although there is no uniquely satisfactory answer to this question, a frequently employed strategy is to apply to the raw data a filter proposed by Hodrick and Prescott (1980) in a discussion paper, published many years later essentially unchanged as Hodrick and Prescott (1997). Although their procedure is universally referred to as "the Hodrick-Prescott filter," it should be noted that Akaike (1980) proposed precisely the same approach in the context of seasonal adjustment. The Akaike model incorporates a seasonal component in addition to trend and cyclical components, but his approach to isolating individual components is the same as that of Hodrick and Prescott, whose filter amounts to a special case of Akaike's procedure when seasonality is absent.

The Hodrick-Prescott (HP) filter was not developed to be appropriate, much less optimal, for specific time series generating processes. Rather, apart from the possible choice of a single "smoothness parameter," the same filter is intended to be applied to all series. It is motivated through plausibility rather than optimality considerations. Indeed, even the single free parameter value is generally chosen subjectively, on rather ad hoc grounds. Nevertheless, the properties of the filter have been analysed by, for example King and Rebelo (1993) and Ehglen (1998), from the viewpoint of optimal signal extraction. We discuss this issue in the following section, noting that if a given time series is generated by a process that is integrated of order one or two and is autoregressive-moving average on reduction by differencing to stationarity, HP yields a decomposition that is optimal into orthogonal components that can be regarded as "trend" and "cycle." However, estimated components will not obey the same generating processes as the corresponding "true" components, and a further line of enquiry, as in Harvey and Jaeger (1993) and Cogley and Nason (1995), is to consider the stochastic properties of the estimated components induced by the filter. Also in Section 2 we provide an extended discussion of the application of the HP filter to a random walk.

The optimality result of the previous paragraph 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 the remainder of 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. Sections 3 and 4 of the paper explore the extent of that suboptimality for individual data generating processes from two perspectives that are apparently distinct, but in fact are very closely connected.

In Section 3, we continue with the notion 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 - after all, that concept is not present in the original work of Hodrick and Prescott. We do so in Section 4 through the notion of endpoint revisions. Suppose, having computed an HP decomposition for a given time series, further time elapses and more data accumulate. The filter could be re-applied to the extended data set, which would yield revisions to the original components estimates at the most recent time periods. Although such revisions are necessary (and desirable), one would hope, if current cyclical values are to be taken seriously, that they not be too large. In fact, the HP filter as generally applied leads to revisions with larger standard deviations than necessary, and we analyse and illustrate the extent of this suboptimality. Finally, Section 5 concludes.

2 Some technical issues

Given a series of observations y_t $(t = 1, 2, ..., T)$ on a time series, the HP filter is an additive decomposition

$$y_t = y_t^g + y_t^c$$

where y_t^g is identified as a growth (trend) component and y_t^c as a cyclical component. In much of the business cycle literature, the purpose is to analyse relationships among the cyclical components of given time series. Hodrick and Prescott estimate the growth component as $\hat{y}_t^g$ through solution of the constrained minimisation problem

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

where the parameter λ controls the smoothness of the estimated growth component. Hodrick and Prescott (1980/1997) proposed on somewhat subjective grounds a value $\lambda = 1600$, and we shall follow much applied work that exploits the HP filter in employing that value in the bulk of this paper. However, 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. Nevertheless, the filter that results from the solution to (1) can be viewed in terms of the optimal signal extraction literature pioneered by Wiener (1949) and Whittle (1963) and, crucially in the present context, extended by Bell (1984) to incorporate integrated time series generating processes. Results in that literature generally 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. King and Rebelo (1993) and Ehglen (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

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

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

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

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

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

and where their variance ratio is

$$\lambda = \frac{\sigma_u^2}{\sigma_\varepsilon^2} \tag{5}$$

where λ is the value of the smoothness parameter used in (1). It is difficult to say whether the orthogonality restriction implied by (4) is "reasonable," but clearly permitting an arbitrary correlation structure would lead to lack of identification.

In practice, one directly observes y_t rather than its components, so it is interesting to view the optimality result (3) in terms of the process generating the original series. It is generally agreed that a great many economic time series are integrated of order d, $I(d)$ - that is, require differencing some positive number d of times to induce stationarity. There is some controversy as to whether $d = 1$ or 2 is typically the more appropriate, witness for example the contrasting views of Granger (1997) and Harvey (1997), but there is scant support for higher values. It is standard practice, following Box and Jenkins (1970), to fit to data autoregressive integrated moving average, $ARIMA(p, d, q)$ models and we shall restrict attention to such generating models with $d = 1$ or 2 for y_t.

It should be noted that all such models fit into the framework (3), with $A(L)$ involving a unit moving average root in the case $d = 1$, so that then the growth component y_t^g is also $I(1)$. Since y_t is the sum of the individual components, it follows from (3) that

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

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 yields for $\lambda = 1600$

$$\gamma_1 = 1.77,\ \gamma_2 = -0.80,\ \sigma_\eta^2 = 1.25\sigma_u^2. \quad (7)$$

Consider first the case where y_t is $I(2)$, with stationary autoregressive operator $\phi(L)$ and invertible moving average operator $\theta(L)$ in its generating model. Then in (6) set

$$A(L) = \frac{\theta(L)}{\phi(L)(1-\gamma_1 L - \gamma_2 L^2)} \quad (8)$$

so that

$$\phi(L)(1-L)^2 y_t = \theta(L)\eta_t. \quad (9)$$

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 $\sigma_\eta^2, \sigma_u^2$ is determined through (7), or the corresponding expression for some other λ, and σ_ε^2 through (5). The parameters γ_i of (8) are also functions of λ. It follows from (9), (8), 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

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

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

$$\phi(L)(1-L)y_t = \theta(L)\eta_t \quad (11)$$

and it immediately follows that if the process represented by (11) 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.

We have seen, unsurprisingly since no dimensionality reductions are required, that, although it was not explicitly developed to do so, 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 parameter λ in (1). In that sense these could perhaps be viewed as the "true" components implied by the adoption of HP. After all, the filter estimates such components as accurately as possible in expected squared error sense, and components generated any other way would either be incompatible with the generating process for y_t or more efficiently estimated through some other filter. Nevertheless, as emphasised for example by Ehglen (1998), the stochastic process followed by the component estimate $\hat{y}_t^c$ differs from that followed by the true process y_t^c as a well known consequence of optimal filtering. To illustrate this point, since many economic time series appear to be generated by processes which at least closely resemble random walks, we explore the behaviour of the HP estimated cyclical component $\hat{y}_t^c$ when actual y_t is generated by the process (11) with $\phi(L) = \theta(L) = 1$. Of course, as we have seen, and as follows from (10), the "true" HP cyclical component is then generated by

$$(1 - \gamma_1 L - \gamma_2 L^2) y_t^c = (1 - L) u_t \tag{12}$$

where, for $\lambda = 1600$, the parameters γ_i are given by (7). However, it follows from (2) that

$$\hat{y}_t^c = h(L)\eta_t \quad ; \quad h(L) = \frac{(1-L)(1-L^{-1})^2}{\lambda^{-1} + (1-L)^2(1-L^{-1})^2}.$$

Hence, the autocovariance-generating function of $\hat{y}_t^c$ is

$$g_{\hat{c}}(z) = \sigma_\eta^2 \frac{(1-z)^3(1-z^{-1})^3}{k^2\left[D(z)\right]^2\left[D(z^{-1})\right]^2}$$

where

$$\lambda^{-1} + (1-z)^2(1-z^{-1})^2 = kD(z)D(z^{-1}) \tag{13}$$

and $D(z)$ is a polynomial of degree two in z with $D(0) = 1$. It follows that the generating model for $\hat{y}_t^c$ can be written

$$D^2(L)\hat{y}_t^c = (1-L)^3 v_t \;\; ; \;\; \sigma_v^2 = k^{-2}\sigma_\eta^2 \;\; ; \;\; D(L) = 1 - \gamma_1 L - \gamma_2 L^2 \tag{14}$$

where v_t is white noise. It is permissible to use the same notation γ_i as before since it is straightforward to see from (13) that the algebra that leads to the derivation of those parameters is precisely the same as that which flows from (6). For example, for $\lambda = 1600$, the γ_i values are precisely those given in (7) with $k = 1.25$. This $ARMA(4, 3)$ data generating process, with three unit moving average roots, is, as expected, quite different from the process (12) that generates the "true" y_t^c. The autoregressive operator $D^2(L)$ in (14) has two identical sets of complex conjugate roots, $1.111 \pm 0.125i$ so that $\hat{y}_t^c$ will exhibit an element of damped cyclical behaviour. This phenomenon, which of course is absent in the original random walk series, is in effect spuriously induced by the HP filter.

This result appears rather complex and does not give an easily interpreted impression of how the behaviour of the estimated cyclical component will appear when the HP filter is applied to a random walk series of practically interesting length. To obtain a different perspective, we generated series of $T = 100$ observations from the random walk process (11) with $\phi(L) = \theta(L) = 1$ and η_t normally distributed white noise with zero mean and variance one. The HP filter was applied to each generated series, yielding actual components estimates. To explore the apparent behaviour of the cyclical component, we attempted $ARMA(p, q)$ modelling. All combinations satisfying $p + q$ □ 7 were considered, parameter estimation was through maximisation of the exact Gaussian likelihood, and the order (p, q) was selected through both the SBC and AIC criteria. Table 1 summarises the percentage times models of each entertained order were selected by these criteria. The selected models are generally more lightly parameterised than the $ARMA(4, 3)$ model predicted by the theory. Moreover, this feature is unconnected with the end-effect phenomenon to be discussed in the following two sections (applying the filter to much longer series, and discarding at least 1,000 observations from each end of the series to leave 100 cyclical component estimates unaffected by this phenomenon produced results that did not differ substantially from these of Table 1). From Table 1(a) we see that for the majority of series $\hat{y}_t^c$ is identified by SBC as first order autoregressive. The average value of the autoregressive parameter estimates for these series was 0.67. When an autoregressive component of order at least two was identified, that component almost invariably (on 98% of such occasions) contained complex roots. Also in line with (14), on the great majority of times a model with $q \geq 1$ was identified (88% of such occasions), the fitted model contained at least one unit moving average root. These last two conditional frequencies were repeated for models selected by AIC, though as is standard these models were in the aggregate somewhat more heavily parameterised than the SBC-selected models, the most frequently chosen specification being $ARMA(2, 1)$. One way to summarise these conclusions is that, while the phenomena of complex autoregressive roots and unit moving average roots predicted by the theory could often be detected in practice for series of 100 observations, this is most likely when a model selection criterion, AIC, known to overfit asymptotically, is employed. Under a parsimonious model selection strategy, which is likely to mimic SBC, the filtered series $\hat{y}_t^c$ will often appear to be first order autoregressive with parameter value close to 0.7. This phenomenon of relatively sparse models appearing to provide adequate representations of series of cyclical component estimates stems from the relatively small number of series observations in relation to the number of parameters in the "true model." This is demonstrated in Table 2 which reports results of the same type of simulation experiments, but with 200 replications of series of 500 observations. Even with such a large sample size, the correct $ARMA(4, 3)$ order is selected only on a small minority of occasions, though at least this is the most frequently selected model by AIC. In summary, then, while the theoretical result (14) is correct, that structure will generally not be manifest for series of practically occurring length.

The impact of the features in the estimated cyclical component induced by HP filtering on inference about business cycle "stylized facts" has been discussed by, among others, Harvey and Jaeger (1993) and Cogley and Nason (1995). The former note a problem with a methodology, as employed for example by Canova (1998), that attempts to base business cycle inference on sample cross-correlations at various leads and lags between cyclical component estimates of pairs of series. This approach can generate a "spurious regression" phenomenon in the sense of Granger and Newbold (1974) - that is, the appearance of a strong relation where none exists. Moreover, as a few simulation experiments would demonstrate, a related phenomenon, highlighted by Box and Newbold (1971), is also likely to occur. The sample cross-correlations tend to exhibit a smooth pattern completely unrelated to any real relationship between the variables, and which indeed will materialise when independent $I(1)$ variables are separately subjected to the HP filter to produce estimated cyclical components.

In the following two sections, 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, a conclusion that mirrors that of Wallis (1982) in a study of the linear filter version of the Census $X - 11$ seasonal adjustment procedure. 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.

For example, if the generating process is given by (3), such an approach will yield optimal components estimates for the entire period covered by an observed data set. This approach differs from the usual HP filter, for which for example $\hat{y}_t^c$, following from the solution of (1), will be the same linear function of y_{T-j} $(j \geq 0)$ whatever the true data generating process, whereas optimal forecasts depend on those observations according to that process. We go on to examine the extent of this HP suboptimality from two perspectives. First, in Section 3, we view the filter as an attempt to estimate the quantity y_t^c of (3), and assess the efficiency of the HP estimates of the most recent time periods. Then in Section 4 we abandon the explicit view of a "true" component and ask to what extent the current cyclical component would require revision in a few years time if the filter were reapplied as new data became available - that is, we compare $\hat{y}_T^c$ based on y_{T-j} $(j \geq 0)$ with an estimate based on y_{T+H-j} $(j \geq 0)$ for moderately large H. It is well known that estimates based on forecast-augmented series, as described above, will generate minimum expected squared revisions, and these are compared with revisions that would follow from the standard HP application. The issue is practically relevant, as very often it is the most recent cyclical components that are of greatest interest, since concern might focus on whether, and by how much, an economic variable is currently "above trend."

3 Estimation of recent cyclical components

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

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

and

$$(1 - L)^2 y_t = (1 - bL)(1 - \gamma_1 L - \gamma_2 L^2)\eta_t \tag{16}$$

where the γ_i depend on λ of (5). Specifically, for $\lambda = 1600$ they are given by (7), which also fixes σ_η^2 in terms of σ_u^2, as (5) fixes σ_ε^2. In the simulations that follow, we set $\lambda = 1600$, 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 value $\lambda = 1600$ was fixed in the filter that solves (1). The usual HP components estimate $\hat{y}_t^c$ then follows directly from y_t $(t = 1, 2, ..., T)$.

Each generated series was augmented by H minimum mean squared error-optimal forecasts, giving series $\tilde{y}_t$ $(t = 1, 2, ..., T + H)$ where $\tilde{y}_t = y_t$ $(t = 1, 2, ..., T)$, and the remaining elements of $\tilde{y}_t$ are forecasts based on y_{T-j}

$(j = 0, 1, 2, ...)$ and (15) or (16) in the usual way. Estimation results are more or less invariant to T, provided that sample size is moderately large, and in our simulations we set its value at 80. 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 $\tilde{y}_t$ $(t = 1, 2, ..., T + H)$.

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 10,000 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 3 and 4 for estimation of the cyclical components y_{T-j}^c $(j = 0, 1, 2, 10)$. 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 the values of the model parameters, being most pronounced in Table 3 for low negative a and in Table 4 for high positive b, both of which correspond to substantial negative first autocorrelation in the "true" cyclical component of (3). 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 quite interesting, keeping in mind that $\sigma_u = 1$ is the standard deviation of the white noise generating the quantity of interest y_t^c. For the most recent observations, these estimation error standard deviations are far from negligible, and for some parameter values, notably large positive a in the case of Table 3, 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 3 and 4 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.

To our choice of data generating processes for Tables 3 and 4, it might be objected that the corresponding processes (15) and (16) for y_t are $I(2)$, whereas in practice analysts typically fit $I(1)$ models to actual economic series. However, in a further simulation not reported in detail here, we generated 10,000 replications of series of 100 observations from each of the models of these tables, and applied the usual Dickey-Fuller test to the first differences of these series. Thus, we tested the null hypothesis that the original undifferenced series is $I(2)$ against the alternative that is $I(1)$, choosing by general-to-specific testing the number of lagged changes incorporated in the Dickey-Fuller regressions. In virtually all cases the (true) null hypothesis was rejected at the 5%-level on an overwhelming majority of occasions, the only exception being the $a = 0.9$ case of Table 3, where the rejection rate was still 34%. This finding is a consequence of the fact that the generating processes (15) and (16) with γ_i given by (7) have a moving average root that is close to one - a situation that is well known to generate spurious rejections of the null hypothesis by Dickey-Fuller tests (see, for example, Schwert 1989 and Agiakloglou and Newbold 1992, 1996). The conclusion then is that, even if series were truly generated by such processes, it would be extremely difficult to distinguish such models from $I(1)$ processes for practically commonly occurring sample sizes.

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 then 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.

Consider again a time series y_t $(t = 1, 2, ..., T)$ and let $\hat{y}_t^c$ $(t = 1, 2, ..., T)$ 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 $(t = 1, 2, ..., T + H)$. The analyst could then pass this entire extended series though 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

$$(1-\phi L)(1-L)y_t = \varepsilon_t \quad ; \quad |\phi| < 1 \tag{17}$$

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

$$(1-L)y_t = (1-\theta L)\varepsilon_t \quad ; \quad |\theta| < 1. \tag{18}$$

We generated series of T observations from these processes with ε_t independent Gaussian with mean 0 and variance $\sigma_\varepsilon^2 = 1$, and applied the HP filter. Generation was continued for H subsequent observations and the filter was applied also to the extended series so that revisions could be calculated: their standard deviations were estimated through 10,000 replications. The result is virtually invariant to T, provided that number is moderately large: here we took $T = 80$. 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 $\tilde{y}_t$ $(t = 1, 2, ..., T + H)$ 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} $(j \geq 0)$. For the two models of our study, forecasts can be obtained directly from (17) and (18). It is quite clear that such an approach minimises revision standard deviation. The quantity to be estimated is simply $\hat{y}_T^{c*}$, which is precisely the same linear function of y_t $(t = 1, 2, ..., T + H)$ as is $\widehat{\tilde{y}_T^c}$ of the forecast augmented series $\tilde{y}_t$ $(t = 1, 2, ..., T + H)$. 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 $(t = 1, 2, ..., T + H)$. 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 (17) and (18) are given respectively in Tables 5 and 6 for a range of parameter values.

It can be seen from Tables 5 and 6 that, compared with the standard deviation $\sigma_\varepsilon = 1$ of the white noise generating y_t, the revision standard deviations s for the usual HP cyclical component can be very large, particularly when first differences of the series are positively autocorrelated. This phenomenon can be somewhat mitigated if the filter is applied to the forecast-augmented series, which will lead to reductions of generally at least

20%, and in some cases much more, in these revision standard deviations. It should be emphasised that, although this phenomenon is very closely related to that of the previous section, these results do not presume the estimation by the filter of particular "true" components.

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, precisely that is 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.

References

[1] Akaike, H., 1980, Seasonal adjustment by a Bayesian modeling, *Journal of Time Series Analysis* 1, 1-13.

[2] Agiakloglou, C. and P. Newbold, 1992, Empirical evidence on Dickey-Fuller-type tests, *Journal of Time Series Analysis 13, 471-483.*

[3] Agiakloglou, C. and P. Newbold, 1996, The balance between size and power in Dickey-Fuller tests with data-dependent rules for the choice of truncation lag, *Economics Letters* 52, 229-234.

[4] Bell, W., 1984, Signal extraction for nonstationary time series, *Annals of Statistics* 12, 646-664.

[5] Box, G.E.P. and G.M. Jenkins, 1970, Time series analysis, forecasting, and control (Holden-Day, San Francisco).

[6] Box, G.E.P. and P. Newbold, 1971, Some comments on a paper of Coen, Gomme and Kendall, *Journal of Royal Statistical Society A* 134, 229-240.

[7] Burman, J.P., 1980, Seasonal adjustment by signal extraction, *Journal of Royal Statistical Society A* 143, 321-337.

[8] Canova, F. 1998, Detrending and business cycle facts, *Journal of Monetary Economics* 41, 475-512.

[9] 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.

[10] Ehglen, J., 1998, Distortionary effects of the optimal Hodrick-Prescott filter, *Economics Letters* 61, 345-349.

[11] Granger, C.W.J., 1997, On modelling the long run in applied economics, *Economic Journal* 107, 169-177.

[12] Granger, C.W.J. and P. Newbold, 1974, Spurious regressions in econometrics, *Journal of Econometrics* 2, 111-120.

[13] Harvey, A.C., 1997, Trends, cycles and autoregressions, *Economic Journal* 107, 192-201.

[14] Harvey, A.C. and A. Jaeger, 1993, Detrending, stylized facts and the business cycle, *Journal of Applied Econometrics* 8, 231-247.

[15] Hodrick, R.J. and E.C. Prescott, 1980, Post-war U.S. business cycles: an empirical investigation, Mimeo (Carnegie-Mellon University, Pittsburgh, PA).

[16] Hodrick, R.J. and E.C. Prescott, 1997, Post-war U.S. business cycles: an empirical investigation, *Journal of Money, Credit and Banking* 29, 1-16.

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

[18] Schwert, G.W., 1989, Tests for unit roots: a Monte Carlo investigation, *Journal of Business and Economic Statistics* 7, 147-160.

[19] Wallis, K.F., 1982, Seasonal adjustment and revision of current data: Linear filters for the X-11 method, *Journal of Royal Statistical Society* A 145, 74-85.

[20] Whittle, P., 1963, Prediction and regulation (Van Nostrand, Princeton, NJ).

[21] Wiener, N., 1949, Extrapolation, interpolation and smoothing of stationary time series (Wiley, New York).

Table 1. Percentage times particular $ARMA(p,q)$ models are selected for $\hat{y}_t^c$ from HP filtered random walks ($T = 100$; 1,000 replications)

(a) SBC selection

$q \diagdown p$	0	1	2	3	4	5	6	7
0	0	56.5	4.3	0.3	0.1	0.1	0	0
1	0	1.7	26.1	5.0	0.6	0.3	0.1	.
2	0	0.7	1.9	0.2	0.2	0	.	.
3	0	0.1	1.1	0	0	.	.	.
4	0	0.2	0.5	0	.	.	.	.
5	0	0	0	.	.	.	.	.
6	0	0	.	.	.	.	.	.
7	0	.	.	.	.	.	.	.

(b) AIC selection

$q \diagdown p$	0	1	2	3	4	5	6	7
0	0	7.5	1.8	0.3	0.1	0.1	0	0.4
1	0	0.6	26.5	11.3	7.8	4.3	4.5	.
2	0	0.5	6.3	3.7	2.8	1.4	.	.
3	0	0.5	5.3	1.6	2.6	.	.	.
4	0	0.5	3.3	1.8	.	.	.	.
5	0	0.9	2.1	.	.	.	.	.
6	0	1.5	.	.	.	.	.	.
7	0	.	.	.	.	.	.	.

Table 2. Percentage times particular $ARMA(p,q)$ models are selected for $\hat{y}_t^c$ from HP filtered random walks ($T = 500$; 200 replications)

(a) SBC selection

$q \diagdown p$	0	1	2	3	4	5	6	7
0	0	0.5	0	0	0	0	0	0
1	0	0	22.5	11	8.5	3.5	2	.
2	0	0	13	21	4	0	.	.
3	0	0	7	0	1.5	.	.	.
4	0	0	5.5	0	.	.	.	.
5	0	0	0	.	.	.	.	.
6	0	0	.	.	.	.	.	.
7	0	.	.	.	.	.	.	.

(b) AIC selection

$q \diagdown p$	0	1	2	3	4	5	6	7
0	0	0	0	0	0	0	0	0
1	0	0	5	7	6.5	4.5	13.5	.
2	0	0	6	10	4	7	.	.
3	0	0	5.5	1.5	19	.	.	.
4	0	0	6.5	1	.	.	.	.
5	0	0.5	2.5	.	.	.	.	.
6	0	0	.	.	.	.	.	.
7	0	.	.	.	.	.	.	.

Table 3. Standard deviations of estimators of recent cyclical components in generating processes (3) for which HP is theoretically optimal: $A(L) = (1 - aL)^{-1}$

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
-0.9	0.34	0.73	0.30	0.74	0.27	0.75	0.13	0.98
-0.8	0.30	0.87	0.27	0.87	0.24	0.87	0.14	0.99
-0.7	0.30	0.92	0.27	0.93	0.24	0.93	0.15	0.99
-0.6	0.30	0.96	0.27	0.96	0.25	0.96	0.15	1.00
-0.5	0.32	0.96	0.28	0.97	0.25	0.97	0.16	1.00
-0.4	0.33	0.99	0.30	0.98	0.27	0.98	0.18	1.00
-0.3	0.36	0.98	0.32	0.99	0.28	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	1.00	0.44	1.00	0.30	1.00
0.3	0.61	1.00	0.56	0.99	0.50	0.99	0.34	1.00
0.4	0.71	0.98	0.63	0.98	0.57	0.98	0.40	1.00
0.5	0.84	0.97	0.74	0.98	0.67	0.98	0.48	1.00
0.6	1.00	0.97	0.91	0.96	0.82	0.96	0.59	1.00
0.7	1.28	0.92	1.14	0.93	1.04	0.94	0.76	1.00
0.8	1.73	0.89	1.57	0.89	1.43	0.91	1.09	1.00
0.9	2.81	0.81	2.50	0.85	2.28	0.89	1.86	1.00

Table 4. Standard deviations of estimators of recent cyclical components in generating processes (3) for which HP is theoretically optimal: $A(L) = (1 - bL)$

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
-0.9	0.83	0.97	0.74	0.98	0.67	0.98	0.46	1.00
-0.8	0.79	0.98	0.71	0.98	0.64	0.98	0.44	1.00
-0.7	0.74	0.99	0.67	0.98	0.60	0.98	0.41	1.00
-0.6	0.69	1.00	0.62	0.99	0.56	0.99	0.38	1.00
-0.5	0.65	0.98	0.58	0.99	0.52	0.99	0.36	1.00
-0.4	0.61	1.00	0.55	0.99	0.50	0.99	0.34	1.00
-0.3	0.57	0.99	0.51	1.00	0.46	0.99	0.32	1.00
-0.2	0.53	1.00	0.47	1.00	0.43	1.00	0.29	1.00
-0.1	0.49	0.99	0.43	1.00	0.39	1.00	0.27	1.00
0	0.45	1.00	0.40	1.00	0.36	1.00	0.24	1.00
0.1	0.41	1.00	0.37	1.00	0.33	1.00	0.22	1.00
0.2	0.38	0.98	0.33	0.99	0.30	1.00	0.20	1.00
0.3	0.33	1.00	0.30	0.98	0.27	0.98	0.17	1.00
0.4	0.30	0.95	0.27	0.97	0.24	0.97	0.15	1.00
0.5	0.27	0.92	0.24	0.93	0.22	0.93	0.13	0.99
0.6	0.24	0.88	0.22	0.88	0.19	0.88	0.10	0.99
0.7	0.22	0.77	0.20	0.78	0.17	0.79	0.08	0.98
0.8	0.21	0.66	0.19	0.66	0.16	0.68	0.06	0.97
0.9	0.21	0.46	0.18	0.50	0.16	0.52	0.05	0.91

Table 5. Standard deviations of revisions of most recent HP cyclical components when $(1-\phi L)(1-L)y_t = \varepsilon_t$

ϕ	s	s_f/s
-0.9	0.65	0.79
-0.8	0.68	0.79
-0.7	0.72	0.79
-0.6	0.76	0.78
-0.5	0.80	0.79
-0.4	0.87	0.78
-0.3	0.94	0.78
-0.2	1.00	0.77
-0.1	1.09	0.76
0	1.21	0.75
0.1	1.32	0.75
0.2	1.48	0.73
0.3	1.70	0.72
0.4	1.94	0.70
0.5	2.28	0.68
0.6	2.78	0.66
0.7	3.52	0.62
0.8	4.67	0.56
0.9	6.64	0.49

Table 6. Standard deviations of revisions of most recent HP cyclical components when $(1-L)y_t = (1-\theta L)\varepsilon_t$

θ	s	s_f/s
-0.9	2.27	0.72
-0.8	2.14	0.72
-0.7	2.01	0.71
-0.6	1.92	0.72
-0.5	1.79	0.73
-0.4	1.69	0.72
-0.3	1.56	0.73
-0.2	1.43	0.74
-0.1	1.32	0.74
0	1.21	0.75
0.1	1.08	0.76
0.2	0.98	0.78
0.3	0.86	0.79
0.4	0.74	0.81
0.5	0.64	0.82
0.6	0.54	0.84
0.7	0.44	0.84
0.8	0.36	0.81
0.9	0.30	0.72