The Economic Journal, **109** (*January*), 126–150. © Royal Economic Society 1999. Published by Blackwell Publishers, 108 Cowley Road, Oxford OX4 1JF, UK and 350 Main Street, Malden, MA 02148, USA.

DOES DETRENDING MATTER FOR THE DETERMINATION OF THE REFERENCE CYCLE AND THE SELECTION OF TURNING POINTS?*

Fabio Canova

We examine the sensitivity of turning points classification to detrending and compare the implied cycles to those derived by NBER or DOC researchers. Two dating rules are considered. Turning point dates are broadly insensitive to detrending with one dating rule but not the other and many procedures generate false alarms and miss several common turning points with the latter one. Amplitude and duration properties are also sensitive detrending and dating rules. The Hodrick and Prescott and a frequency domain filters are best in mimicing NBER and DOC cycles, regardless of the dating rule used.

Per questo non abbiamo niente da insegnare: su cio' che piu' somiglia alla nostra esperienza non possiamo influire, in cio' che porta la nostra impronta non sappiamo riconoscerci. Mr. Palomar, Italo Calvino

Is the dating of business cycle turning points sensitive to the choice of detrending? Are the amplitude, duration and persistence characteristics of the resulting reference cycle robust? Is there any detrending method which produces a reference cycle whose turning points match NBER or Department of Commerce (DOC) turning points and replicates features of the US reference cycle?

This paper attempts to shed some light on these three issues. There are several reasons why these questions may be important for business cycle researchers. First, although there is a long history dating business cycle extremes using level data, since Mintz (1969) it has become as common to provide a classification of turning points and business cycle phases using a growth-cycle approach, i.e. using fluctuations around the trend of the series (see e.g. OECD (1986), Zarnowitz (1991*b*) or Niemera (1991)). However, as Zarnowitz (1991*a*) has pointed out, trends vary over time, may interact in a nontrivial way with the cyclical component of the series and be difficult to isolate and measure given the size of the available samples and existing econometric techniques. In standard practice NBER or DOC growth cycles are extracted using elaborate and ad-hoc procedures which are hard to reproduce, involve a substantial amount of judgmental decisions by the researchers and a number of ex-post revisions as more information is obtained over time. It is therefore worthwhile to study, on one hand, whether any available and mechanical detrending procedure can provide a simple rationale for these complicated and subjective approaches and, on the other, whether there is a class of

* I would like to thank two anonymous referees and the editor of this Journal, Danny Quah, Harald Ulhig and the participants of seminars at the Bank of Portugal, Tilburg University and University of Catania for useful comments.

detrending methods which produce reference cycles with 'desirable' properties.

Second, modern business cycle theorists have attempted to build models which reproduced the cyclical features of the data, where cycles are obtained detrending the original series with a variety of methods. Typically, in these models turning points are not as important as in Keynesian trade-cycle theories. However, the recent development of general equilibrium Markov switching models has brought back the attention to turning points as they represent the threshold across which the process changes features. Canova (1998) showed that different trend removal procedures, all of which are reasonable given existing empirical evidence and available econometric tools, induce different properties in the moments of the cyclical component of several real macroeconomic series and different implications for how we perceive the economy to work. It is therefore interesting to check whether the *path* properties of the cycles induced by different detrending methods are also substantially different, thus providing a more complete perspective on the macroeconomic implications of different trend-removal procedures. The ability to reproduce well known characteristics of the US reference cycle can be used as a limited information test to discern among a variety of detrending procedures which *a priori* would be on an equal footing. Third, many researchers have examined the statistical features of the NBER reference cycle over the pre and post WWII period, in particular the amplitude and duration properties (see e.g. Diebold and Rudebush (1990, 1992), Romer (1994) or Watson (1994)) but not much has been done with growth cycles (one exception is Pagan (1997)). It is therefore worthwhile to study whether their statistical features are robust when mechanical detrending procedures are used to construct reference cycles. Fourth, there exists a large branch of the literature which deals with turning point predictions (see e.g. Wecker (1979) or Zellner and Hong (1991)) and the evaluation of the record and the quality of turning point forecasts (see e.g. McNees (1991)). However, the conclusions obtained clearly hinge on having available the 'correct' notion of growth cycle. Therefore, our study may also help researchers studying this problem to select one concept of cycle over another in deciding the validity of various forecasting approaches.

In examining the questions posed in this introduction, the paper focuses on 12 widely used detrending methods (linear and segmented detrending, first order differencing, frequency domain filtering, Hodrick and Prescott filtering, detrending with the Beveridge and Nelson model, with an unobservable components model, with Hamilton's 2-state model, with a one dimensional index model, with Blanchard and Quah's model, with King, Plosser and Rebelo's model and with King, Plosser, Stock and Watson's model). To classify turning points and to construct business cycle phases, I consider two standard mechanical dating rules. The first rule defines a trough as a situation where two consecutive quarter declines in the reference cycle are followed by an increase. Likewise, a peak is defined by two consecutive increases followed by a decline. The second rule selects a quarter as a trough (peak) if there have been at least two consecutive negative (positive) spells in the cycle over a three quarter period. Although the search across detrending methods and dating rules is not exhaustive and more complicated dating rules may generically improve the quality of the outcomes, our work provides a first step in systematically addressing these issues and methodically studying the data.

The results of the paper complement those of Canova (1994). There it was shown that turning point classification is essentially robust to detrending with the first dating rule but not with the other. Here we qualify this statement by showing that with this latter dating rule many standardly reported turning points are missed and many false alarms appear with the majority of the detrending methods. For those turning points which are correctly identified we find that the majority of methods produce dates which slightly lead NBER peaks and troughs and lead DOC peaks but coincide with DOC troughs.

In addition, we demonstrate that the statistical properties of the generated cycles are sensitive to both the detrending procedure and the dating rule. With the first rule, regardless of the detrending procedure used, cycles are slightly asymmetric and the duration of expansions exceeds, on average, the duration of contractions. We also show that there exists only a moderate degree of variability in the duration of each phase and in the amplitude of contractions. Furthermore, we find little evidence that peak dates can be predicted using the information contained in past durations while trough dates are predictable with at least 7 methods. The statistical properties of the various cycles are much more heterogeneous with the second rule. In general, the average duration of contractions exceeds the average duration of expansions and there is a large degree of variability in the duration of each phase and in the amplitude of contractions. Peak dates are more predictable than trough dates but differences across detrending methods are substantial. The only regularity that is robust concerns the persistence of business cycle phases: the amplitude of contractions is in fact uncorrelated with both the duration of contractions and of peak-to-peak cycles, regardless of detrending and dating rule.

The general conclusion seems to be a negative one: statements concerning the location of turning points and the properties of growth cycles are not independent of the statistical assumptions needed to extract trends. While this outcome is somewhat disappointing, our exercise also provides important information on the characteristics of various detrending procedures and on the types of cycles they generate. When we take the ability to reproduce the characteristics of NBER or DOC growth cycles as a limited information test to select a class of detrending procedures over another, the paper indicates that standard methods employed by RBC researchers (the HP filter of Hodrick and Prescott (1980) or the band-pass filters of Baxter and King (1994), here used in frequency domain) are in fact generating growth cycles capturing the essence of what the community perceives as business cycle fluctuations. This obviously does not mean that these filters are appropriate for all purposes, as they may wipe out cycles with important economic features (see Canova (1998)) and may induce spurious patterns in series which do not display any form of classical cyclical fluctuations (see e.g. King and Rebelo (1993) and

Cogley and Nason (1995)). But they appear to provide (i) a solid rationale for the current NBER or DOC practice and (ii) a clear standard to study the path properties of time series generated by dynamic general equilibrium models.

Pagan (1997) has conducted complementary work to that presented here. In a simulation study he tries to assess which data generating process is able to produced duration and symmetry features which reproduce those of classical and growth cycle phases. His result, that a random walk with drift is best, is consistent with our conclusion that the HP and the frequency domain filters are to be preferred in the class of methods used since they are able to extract both deterministic and stochastic trends from the data.

The rest of the paper is organised as follows: the next section describes the various detrending procedures employed in the paper. Section 3 presents the data, the dating rules and the statistics used to characterise the properties of the reference cycle. Section 4 discusses the results. Section 5 concludes describing the implications of the results for current macroeconomic practice.

1. Alternative Detrending Methods

This section briefly describes the 12 procedures used to extract trends from the observable time series. For more details the reader is invited to consult Canova (1995). Throughout the paper I denote the natural logarithm of the time series by y_t, its trend by x_t and its cyclical component by c_t. The methods will be classified according to three characteristics: assumptions on the features of the trend, assumptions on the correlation between x_t and c_t and on whether the methods have statistical or economic justifications. Since only trend and cycle are assumed to exist, all the procedures implicitly assume that y_t has previously been seasonally adjusted and that irregular (high frequency) fluctuations play little role.

Linear detrending (LT) and segmented detrending (SEGM) assume that x_t is a deterministic process which can be approximated by polynomial functions of time and that x_t and c_t are uncorrelated. With segmented detrending we also assume that there is a structural break in x_t at a known time $\bar{t} = 1973,3$.

The basic assumptions of a first order differencing procedure (FOD) are that x_t is a random walk with no drift and c_t is stationary and that the two components are uncorrelated. Similarly, Beveridge and Nelson's (1981) procedure (BN) assumes that y_t has a unit root and that x_t accounts for its nonstationary behaviour. In this procedure x_t is the long run forecast of y_t adjusted for its mean rate of change, so that the trend is the value y_t would have taken if it were on the long–run path. One implication of BN construction is that in this decomposition x_t and c_t are perfectly correlated since they are driven by the same shocks. Since some judgmental decisions need to be made in implementing BN decomposition, I report results for the case where y_t is modelled as an ARIMA(5,1,0), the value of y_t at 1955,2 is used as a initial condition and the quick computational approach of Coddington and Winters (1987) is employed.

The identifying assumptions of the Unobservable Components (UC) procedure are that x_t follows a random walk with drift and that c_t is a stationary finite order AR process. Contrary to a FOD procedure, here x_t and c_t may be correlated (see Watson (1986)). Also in this case some judgmental decisions need to be made: here I report results for the case where c_t is an AR(2), parameters are estimated using moment restrictions as in Carvalho *et al.* (1979) and estimates of the state equations are obtained with the Kalman filter with no smoothing of recursive estimates.

The frequency domain (FREQ) procedure assumes that c_t and x_t are independent, that x_t has most of its power concentrated in a low frequency band of the spectrum and that away from zero the power of the secular component decays very fast (see Sims (1974)). These identification assumptions do not restrict x_t to be either deterministic or stochastic and allow for changes in x_t over time as long as they are not too frequent. Results are presented for the case where c_t includes all the cycles of y_t with length less that 30 quarters. Baxter and King (1994) provide a time dimension version of this filter and study its implication for stylised facts of the business cycle.

Hamilton's (1989) 2-State Markov specification (HAMIL) assumes that although x_t is characterised by a unit root, its innovations are drawn from a binomial distribution. The key identifying assumptions of this procedure are that x_t and c_t are both nonstationary and independent of each other. Results are presented for the case where c_t is an AR(2). Estimates of the free parameters and of the state of the economy are obtained with a EM algorithm and estimates of c_t are obtained recursively (with no smoothing of the recursive estimates).

Detrending using a one dimensional index model (MFREQ) involves the formulation of a multivariate model. In this case I use data on GNP, Consumption, Investment, Real Wage and Capital. The procedures assumes that in the low frequencies of the spectrum of y_t there exists a one dimensional process x_t which is common to all series (see e.g. Stock and Watson (1989)). x_t is characterised by the property that it has all its power at low frequencies and that away from zero it decays very fast. An estimate of c_t is obtained using a multivariate version of the procedure used for the UC model assuming an AR(2) model for each of the six cyclical components.

Contrary to the first eight methods which have been developed in the statistical literature, the next three procedures have economic justifications. King, Plosser and Rebelo (1988) present a neoclassical model of capital accumulation with labour supply choices where there is deterministic labour augmenting technical progress. In their model all endogenous variables have a common deterministic trend (the growth rate of labour augmenting technical progress), fluctuations around this trend are of a transitory nature and independent of the trend. To extract a common deterministic trend I use the six series used for the index model.

King, Plosser, Stock and Watson (1991) propose a version of King, Plosser and Rebelo's (1988) model driven by a nonstationary technological shock. The corresponding statistical common trend representation (see Stock and Watson (1988)) is the multivariate counterpart of the method of Beveridge and Nelson and implies that all the endogenous variables of the model will have a common nonstationary trend (COIN). To produce estimates of c_t for GNP, I use the same six series used for the index model and estimate a vector error correction model with five lags for each variable and one lag of two cointegrating vectors (GNP/consumption, GNP/investment). An estimate of c_t is then obtained summing all the stationary component.

Blanchard and Quah (1989) (BQ) propose a version of Fisher's staggering wage model in which 'demand' shocks have no long run effects on output and unemployment and 'supply' shocks have long run effects on output but not on unemployment. The implied trend-cycle decomposition of GNP has the property that x_t has a unit root, c_t is stationary and the two components are uncorrelated. To estimate c_t, I use the same bivariate VAR specification employed by Blanchard and Quah.

The final procedure, the Hodrick and Prescott (HP) (1980) filter, has two justifications: one intuitive (see Kydland and Prescott (1990) and one statistical (see e.g. Wabha (1980) or Harvey and Jaeger (1993)). In the first case the HP filter is a flexible tool which can accommodate the needs of applied researchers while in the second it is an optimal extractor of a trend which is stochastic but moves smoothly over time and is uncorrelated with the c_t. Smoothness is imposed by assuming that the sum of squares of the second differences of x_t is small. In the RBC literature the free parameter λ, which regulates the extent of the penalty imposed for large fluctuations in x_t, is typically fixed *a priori* to $\lambda = 1600$ for quarterly data. Because Nelson and Plosser (1982) estimated λ to be in the range $[\frac{1}{6}, 1]$ for most of the series they examine, I present results for the standard setting (HP1600) and for a $\lambda = 4$ (HP4), which is closer to Nelson and Plosser's estimates.

Some of the properties of the HP filter when $T \to \infty$ and the penalty function is two-sided have been highlighted by Cogley and Nason (1995) and King and Rebelo (1993). The relationships between the HP and exponential smoothing (ES) filters have been investigated by King and Rebelo (1993).

Before proceeding with the analysis it is useful to stress three important facts which may make the approaches not exactly comparable. First, the information used to compute the trend of the series differs across detrending methods. While most procedures employ information up to the end of the sample, FOD, UC and HAMIL only use the information available at $t-s$ to compute the trend for $t-s+1$. This may generate a more imprecise estimate of the trend and, as a consequence, produce cyclical components which are more erratic than those obtained with other methods. As a consequence, most methods date peaks and troughs having available data for the entire time span, while others 'call them out as they go'. Second, while most methods use maximum likelihood procedures to estimate the parameters, others use only approximate maximum likelihood techniques and with three procedures (FOD, HP and FREQ) no parameter is estimated from the data. Because the sample size is relatively short, this may induce small sample differences in the estimates of the cyclical components. These differences should be kept in mind when comparing turning point dates and the amplitude properties of the estimates of the cyclical component across detrending methods. Third, because the UC model assumes the presence of both an irregular and a cyclical component, care should be exercised in comparing the path properties of c_t (and the record of turning point classification) obtained with UC and other methods since the UC cyclical component is likely to be much smoother than others.

2. The Data, the Dating Rules and Summary Statistics

2.1. *The Data*

The data used in the exercise is taken from the Citibase Tape. The results refer to the logarithm of seasonally adjusted quarterly US series for the period 1955,3–1990,1. For all univariate procedures we use real gross national product in 1982 dollars (Citibase name: GNP82). For multivariate procedures we add to GNP consumption expenditure by domestic residents on nondurables and services (Citibase names: GSC82 + GCN82), fixed investment in plants and equipment plus consumer durables (Citibase names: GINPD82 + GCD82), total number of hours of labour input as reported by establishment survey data (Citibase name: LPMHU), real wage constructed as the ratio of nominal total compensation of nonagricultural employees and the CPI (Citibase names: GCOMP/PUNEW) and a capital stock series constructed using the net capital stock for 1954, the quarterly series for investment and a depreciation rate of 2.5% per quarter. For the BQ decomposition we use, in addition to GNP, the seasonally adjusted unemployment rate for males, age 20 and over, as reported by the US Bureau of Labor Statistics (Table A-39).

2.2. *Determining the Reference Cycle*

The first step in examining the properties of the cycle is to delineate periods of economic expansions and contractions. According to NBER practices, as set out by Burns and Mitchell (1943), this is done by examining the behaviour and the comovements of the cyclical component of a variety of series, and constructing an index of cyclical movements (the reference cycle). From this information a set of reference dates, specifying turning points in aggregate economic activity, are selected and business cycle phases are constructed. This process has the drawback of being time consuming and involving a considerable amount of subjective judgement in selecting reference dates.

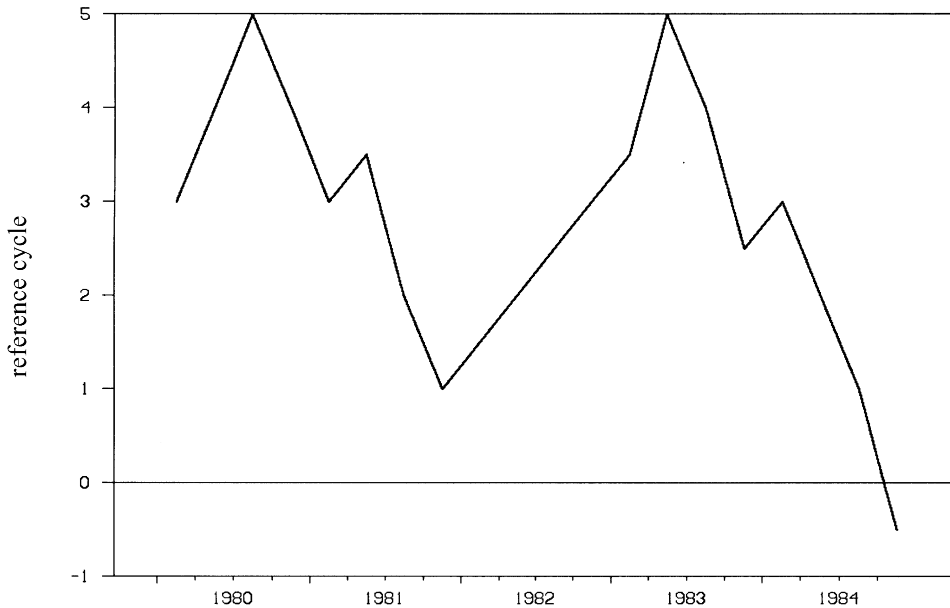
In this paper I depart from the standard Burns and Mitchell approach in several ways. First, as in Simkins (1994) and King and Plosser (1994), instead of constructing an index of cyclical fluctuations, I use the cyclical component of real GNP as a measure of the reference cycle. Although it has been suggested that using the cyclical component of GNP to proxy for the reference cycle fails to capture certain contractions (see e.g. Zarnowitz and Moore (1991)), our choice has the advantage of eliminating judgmental aspects present in the standard procedure and of being easily reproducible. In addition, because a large number of economic variables appear to be procyclical and coincident with GNP, this choice of reference cycle should only minorly distort the dating of turning points even though the amplitude characteristics of the cycle and the severity of contractions may be misrepresented. Finally, the four multivariate procedures do use the information contained in several additional series. Therefore, by comparing the dating record obtained with univariate and multivariate methods we can check whether the information contained in GNP is sufficient to characterise turning points accurately and describe the properties of the reference cycle.

Second, as many have done in this literature, I use mechanical rules to select turning points. However, contrary to e.g. Simkins (1994) or King and Plosser (1994), which use variants of the Bry and Boschen (1971) algorithm, I use two simple and commonly used rules to date turning points. The first classification rule I use is very standard (see e.g. Wecker (1979) or Zellner and Hong (1991)). It defines a trough as a situation where two declines in the cyclical component of GNP are followed by an increase, i.e., at time t, $c_{t+1} > c_t < c_{t-1} < c_{t-2}$. Similarly, a peak is defined as a situation where two consecutive increases in the cyclical component of GNP are followed by a decline, i.e. at time t, $c_{t+1} < c_t > c_{t-1} > c_{t-2}$. The second classification rule is less standard but it has some appealing features (see e.g. Webb (1991) or Pagan (1997)). It selects quarter t as a trough (peak) if there have been at least two consecutive negative (positive) spells in the cyclical component of GNP over a three quarter period, i.e. when $c_t < (>) 0$ and $c_{t-1} < (>) 0$ or when $c_{t+1} < (>) 0$ and $c_t < (>) 0$.

The first classification rule emphasises primarily the duration characteristics of the cycle (no mention of severity is made) and therefore may pick up mild contractions and mild recoveries, while this is not necessarily the case with the second rule, since, e.g., a negative spell in the growth cycle indicates an absolute decline in the level of the series.

On the other hand, the second classification may suffer from amplitude misspecifications if the reference cycle displays multiple sequential peaks (and troughs) in the reference cycle (for an example of this type see Fig. 1, case 2). In general, the first rule may signal the presence of a turning point earlier than the second one. Therefore the two rules balance the scope for an early recognition of the phenomena (at the cost of possible false alarms) vs. its more accurate description (at the cost of a later discovery). Also, it is important to emphasise that we make no adjustments for situations where the reference cycle reaches a plateau around a turning point.

One reason for using these two rules instead of others is that several authors (Wecker, 1979; Webb, 1991) have shown that when they are applied to a standardly constructed (level) cycle they generate turning points which match NBER dates. There are variants and combinations of these two rules which can reduce the frequency of missing signals and discount false alarms (see e.g. Hymans (1973) or Zarnowitz and Moore (1991)) and improve the overall dating recond by using sequential or more flexible approaches (as in Moore and Zarnowitz (1982), Stock and Watson (1990), McNees (1991) and Romer



CASE 1

CASE 2

Fig. 1.

(1994)). However, we restrict attention to these two because they are simple, easily reproducible and provide a useful benchmark to compare the properties of the various reference cycles. These two rules are also preferable to the Bry and Boschen algorithm for our purposes because the latter is sufficiently complicated to render the comparison across detrending methods less transparent. In addition, since the Bry-Boschen algorithm computes turning point dates by detrending the data with a series of MA filters, it is inappropriate to apply it in its original form to detrended data.

As a term of reference in our exercises we use the dating reported by the Center for International Business Cycle Research for the NBER (NBER) and by the Department of Commerce (DOC) (both of which are taken from Niemera (1991)). The 11 series used to construct the NBER cycles are: employees on non-agricultural payrolls, persons engaged in non-agricultural activities, employee-hours in non-agricultural establishment, personal income in constant dollars, total civilian employment, industrial production, manufacturing and trade sales in constant dollars, sales of retails stores in constant dollars, number of unemployed, final sales in constant dollars, GNP in constant dollars. The 11 series used to construct DOC cycles are: average workweek in manufacturing, initial claims for state unemployment insurance, manufacturing new orders for consumers goods and materials in constant dollars, vendors performance, contracts and orders for plants and equipments in constant dollars, building permits, changes in inventories on hands and on order in constant dollars, change in sensitive material prices, stock prices, M2, change in business and consumer credits. The procedure the NBER employs to construct growth cycles is complicated and involves the calculation of the trend of a vector of series by piecewise smooth interpolation of segments of a series obtained by filtering the original data with long term moving averages (see Zarnowitz (1991*a*)). The DOC reference growth cycle, on the other hand, is constructed by detrending the reference index using an exponential smoothing method (Higgins and Poole procedure, see Niemera (1991)). In both cases, turning points and cycle phases are identified using a mixture of mechanical rules and subjective intuition.

2.3. *Summary Statistics*

To analyse the statistical features of the reference cycle and how generated contractions and expansions match up with standardly reported business cycle phases, I compile a number of statistics. To evaluate the dating record of a procedure it is quite common to use statistics based on the differences (in quarters) between the signal and the NBER or DOC turning point. Because such an approach wastes useful information, I employ a different summary statistic based on the timing of the event. I rank the signal as false if a NBER or a DOC turning point does not appear within a ± 3 quarter interval around the signal date and missing if no signal appears within a ± 3 quarter interval around the actual NBER or DOC turning point. The proportion of false alarms and missing signals to the total number of turning points gives an idea of how each detrending procedure trades off the two types of losses. For those turning points which are correctly identified, I also record the proportion of cases where the signal date is leading, coincident or lagging the corresponding NBER or DOC date. This information suggests whether some detrending method generates a systematic bias in recognising standard turning points with one of the two rules.

Together with the dating record of turning points I also present five statistics summarising the statistical properties of each cycle: the average amplitude of contractions, the maximum amplitude and the date at which it occurs, the average duration of expansions and contractions (trough-to-peak (TP) and peak-to-trough (PT) half-cycles) and the percentage of times the economy is expanding. While it is typical to measure the severity of recessions using the distance between the peak and the trough of the cycle, I define severity using the distance of the troughs from the trend line, an approach which is more consistent with the growth-cycle approach adopted in the paper. This measure of the severity is clearly imperfect, but gives a rough idea of how the different detrending methods picture contractions with each of the two dating rules. The percentage of times the economy is in an expansion, on the other hand, is a useful statistic to gauge whether the reference cycle generated by each detrending method and each rule is symmetric or not. We have also computed, the ratio between the percentage of times the economy is above the trend and below the trend. Because the relative ranking across detrending methods and dating rules is unaltered we do not report this statistics here.

To examine whether there is a tendency for contractions and expansions to terminate the longer they have lasted, a question recently investigated using NBER dates by Diebold and Rudebush (1990) and (1992), we computed a nonparametric test for duration dependence of each business cycle phase. This test formally examines whether contractions or expansions have a recurrent and stable (periodic) structure with any detrending method or dating rule, a feature which would facilitate the prediction of turning points. The test, developed by Stephens (1978), is exact even for samples of three durations and incorporates the idea that there is a minimum duration of each phase. In our case the minimum duration (denoted by γ) is two quarters for each phase. This selection is based on the criteria used by NBER researchers in dating contractions and expansions and on the characteristics of the two dating rules we employ. The results we report, however, are not too sensitive to the choice of this parameter within a reasonable range. The statistic used to test for duration dependence is given by:

$$W(t_0 = \gamma) = \frac{(\sum_{i=1}^{N} z_i - \gamma)^2}{N(N+1)[\sum_{i=1}^{N}(z_i - \gamma)^2 - (\sum_{i=1}^{N} z_i - \gamma)^2]} \tag{1}$$

here z_i is the i-th ordered duration for each detrending procedure and each dating rule. The $W(t_0 = \gamma)$ statistic for N durations has an exact small sample distribution which can be recovered from Shapiro and Wilks' (1972) tables using the line corresponding to $N + 1$ durations.

Finally, to study whether there is any systematic relationship between the severity of contractions and either their duration or the duration of full peak-to-peak cycles, we computed Spearman's rank correlation coefficient between the amplitude of contractions and the two types of cycles and tested if they are different. Burns and Mitchell (1943) and Moore (1958) suggested that, because of the way recessions spread in the economy, the association between the severity of the contractions and peak to peak cycles should be stronger than the association between the severity and the duration of contractions. Knowing the severity of a contraction is therefore considered an important ingredient to predict how long it would take the economy to reach another peak level. Also, as emphasised by Romer (1994), the association between the severity and the duration of contractions may indicate how rapidly the effects of a contraction are undone, thereby providing a rough measure of how persistent is this business cycle phase.

3. The results

The results of the investigation appear in Tables 1 and 2. Table 1, Panel *A* reports, for each detrending method, the number of troughs (column 1) and of peaks (column 7) found, the percentage of false alarms and missing signals for troughs (columns 2–3) and for peaks (columns 8–9) and for correct signals if they are leading, coincident or lagging the NBER classification (columns 4–6 and 10–12) for each of the two dating rules. Panel *B* of the table reports the same information when the DOC classification is used as a term of comparison.

Table 2 presents, for each detrending method, the average severity of contractions and its standard deviation (column 1), the maximum amplitude of the contractions and the date at which it occurs (columns 2–3), the percentage of times the economy is in an expansion phase over the sample period (column 4), the average duration of contractions and expansions and their standard deviations (columns 5 and 7), the values of the Stephen's test for duration dependence (columns 6 and 8), the rank correlation coefficient between the amplitude of contractions and the duration of full peak to peak cycles and between the amplitude and the duration of contractions (columns 9 and 10) for each of the two dating rules.

3.1. *Dating Turning Points*

The main features emerging from Table 1 are the sensitivity of turning point classification to detrending and dating rules and, to some extent, the dependence of the results on the reference dating employed. The lack of robustness in the characterisation of business cycle extremes appears in several aspects of the table. First, in agreement with McNees (1991) and Zarnowitz and Moore (1991), the number of complete cycles identified depends on the detrending methods and the dating rule. With the first rule, all methods select at least 8 peaks and 8 troughs (with a maximum of 11), while with the second rule the

Table 1
Panel *A*
Business cycle chronology using NBER dates as reference
Sample 55,3-90,1

	Troughs						Peaks						
Method	Number	% false alarms	% missing signals	Correct LE	Correct CO	Correct LA	Number	% false alarms	% missing signals	Correct LE	Correct CO	Correct LA
					Filter Rule 1							
HP1600	11	45.4	14.2	3	3	0	11	45.4	14.2	5	1	0
HP4	11	54.5	28.4	4	1	0	11	63.6	42.8	4	0	0
FOD	10	90.0	85.7	1	0	0	9	88.8	85.7	0	1	0
BN	9	44.4	28.5	2	1	2	8	50.0	42.8	2	0	2
UC	8	62.5	42.8	2	1	1	9	88.8	83.4	0	0	1
LT	10	50.0	28.5	3	2	0	9	44.4	28.5	3	1	1
SEGM	9	77.7	71.4	1	1	0	8	50.0	42.8	2	1	1
FREQ	10	50.0	28.5	3	2	0	10	50.0	14.2	5	1	0
HAMIL	10	80.0	85.7	2	0	0	9	55.5	42.8	2	2	0
					Filter Rule 2							
MLT	10	60.0	42.8	2	2	0	9	44.4	28.5	3	1	1
MINDEX	8	62.5	57.1	1	1	1	9	66.6	57.1	1	2	0
BQ	8	25.0	14.2	0	4	2	8	37.5	28.5	1	1	3
COIN	9	55.5	42.8	3	1	0	8	50.0	42.8	2	1	1
HP1600	7	42.8	42.8	4	0	0	8	75.0	71.4	2	0	0
HP4	16	62.5	14.2	4	2	0	15	53.3	0.0	5	2	0
FOD	4	50.0	71.4	2	0	0	5	100.0	100.0	0	0	0
BN	3	33.3	71.4	1	1	0	2	100.0	100.0	0	0	0
UC	2	50.0	85.7	0	0	1	3	66.6	85.7	0	0	1
LT	4	75.0	85.7	0	1	0	4	100.0	100.0	0	0	0
SEGM	4	25.0	57.1	3	0	0	5	80.0	85.7	0	1	0
FREQ	8	62.5	57.1	3	0	0	8	50.0	42.8	4	0	0
HAMIL	5	80.0	85.7	0	1	0	4	75.0	85.7	1	0	0
CDT	3	66.6	85.7	1	0	0	3	100.0	100.0	0	0	0
MINDEX	2	0.0	71.4	2	0	0	3	100.0	100.0	0	0	0
BQ	8	62.5	57.1	2	1	0	9	44.4	28.5	3	1	1
COIN	3	33.3	71.4	2	0	0	2	0.0	71.4	2	0	0

Note: With Filter Rule 1 a trough occurs at t if $c_{t+1} > c_t < c_{t-1} < c_{t-2}$ and a peak if $c_{t+1} < c_t > c_{t-1} > c_{t-2}$. With Filter Rule 2 a trough occurs at t if $c_t < 0$ and $c_{t-1} < 0$ or if $c_{t+1} < 0$ and $c_t < 0$ and a peak if $c_t > 0$ and $c_{t-1} > 0$ or if $c_{t+1} > 0$ and $c_t > 0$. A false alarm occurs if there is no turning point within ±3 quarters of the reference date. A missing signal occurs if the method does not signal a turning point within ±3 quarters of the NBER date. In the NBER classification there are 7 troughs and 7 peaks. LE stands for leading, CO for coincident and LA for lagging.

range of extremes identified by the various detrending methods is much larger (between 2 and 15 peaks and 2 and 16 troughs).

Second, the percentage of false alarms varies substantially with the detrending method and the dating rule. For troughs, the percentage of false alarms is between 25 and 90% with the first dating rule and 0 and 80% with the second when the NBER classification is used and between 37.5 and 87.5% with the first dating rule and 0 and 75% with the second when the DOC classification is used. For peaks the heterogeneity is even more evident. When we use the NBER classification as a reference the percentage of false alarms is between

Panel *B*

Business cycle chronology using DOC dates as reference

Sample 55,3-90,1

Method	Troughs Number	% false alarms	% missing signals	Correct LE	CO	LA	Peaks Number	% false alarms	% missing signals	Correct LE	CO	LA
						Filter Rule 1						
HP1600	11	45.4	25.0	2	3	1	11	45.4	0.0	2	2	3
HP4	11	54.5	37.5	2	3	0	11	45.4	14.2	3	1	2
FOD	10	80.0	75.0	1	1	0	9	44.4	28.5	4	1	0
BN	9	55.5	50.0	1	0	3	8	37.5	28.5	1	0	4
UC	8	87.5	85.7	1	0	0	9	77.7	66.6	2	0	0
LT	10	50.0	37.5	1	3	1	9	33.3	14.2	1	2	3
SEGM	9	77.7	75.0	1	1	0	8	25.0	14.2	1	2	3
FREQ	10	50.0	37.5	2	2	1	10	40.0	0.0	2	2	3
HAMIL	10	70.0	62.5	1	2	1	9	33.3	14.2	1	1	4
MLT	10	60.0	50.0	0	3	1	9	33.3	14.2	1	2	3
MINDEX	8	62.5	62.5	0	2	1	9	44.4	28.5	2	1	2
BQ	8	37.5	37.5	0	1	4	8	50.0	42.8	0	2	2
COIN	9	55.5	50.0	0	4	0	8	25.0	14.2	1	2	3
						Filter Rule 2						
HP1600	7	28.5	37.5	5	0	0	8	25.0	14.2	5	1	0
HP4	16	56.2	12.5	5	1	1	15	53.3	0.0	3	3	1
FOD	4	25.0	62.5	3	0	0	5	100.0	100.0	0	0	0
BN	3	33.3	75.0	2	0	0	2	50.0	85.7	1	0	0
UC	2	50.0	87.5	0	0	1	3	66.6	83.4	0	0	1
LT	4	75.0	87.5	1	0	0	4	100.0	100.0	0	0	0
SEGM	4	0.0	50.0	4	0	0	5	40.0	57.1	1	2	0
FREQ	8	25.0	25.0	5	1	0	8	25.0	57.1	5	1	0
HAMIL	5	80.0	87.5	1	0	0	4	75.0	85.7	0	1	0
CDT	3	66.6	87.5	1	0	0	3	100.0	100.0	0	0	0
MINDEX	2	0.0	75.0	2	0	0	3	66.6	85.7	1	0	0
BQ	8	50.0	50.0	1	1	2	9	44.4	28.5	3	0	2
COIN	3	33.3	75.0	2	0	0	2	0.0	71.4	2	0	0

Note: With Filter Rule 1 a trough occurs at t if $c_{t+1} > c_t < c_{t-1} < c_{t-2}$, and a peak if $c_{t+1} < c_t > c_{t-1} > c_{t-2}$. With Filter Rule 2 a trough occurs at t if $c_t < 0$ and $c_{t-1} < 0$ or if $c_{t+1} < 0$ and $c_t < 0$ and a peak if $c_t > 0$ and $c_{t-1} > 0$ or if $c_{t+1} > 0$ and $c_t > 0$. A false alarm occurs if there is no turning point within ±3 quarters of the reference date. A missing signal occurs if the method does not signal a turning point within ±3 quarters of the DOC date. In the DOC classification there are 8 troughs and 7 peaks. LE stands for leading, CO for coincident and LA for lagging.

37.5 and 88.8% with the first dating rule and 0 and 100% with the second dating rule. When we use the DOC classification as a reference the percentage of false alarms is between 25 and 77.7% with the first dating rule and 0 and 100% with the second dating rule.

Third, the percentage of missing signals depends on the detrending method and differs significantly between troughs and peaks. For example, when the NBER classification is used as a benchmark the range of missing troughs is between 14.2 and 85.7% with both dating rules, while the range of missing peaks is between 14.2 and 83.4% with the first rule and between 0 and 100%

Table 2

Statistics of the reference cycle, Sample 55,3-90,1

Method	Amplitude Average	Amplitude Min	Amplitude Date	Expansions %	Expansions TP	Expansions Test	Durations PT	Durations Test	Correlations Amplitude PP	Correlations Amplitude PT
NBER	−2.5	−5.7	75,1	NA	10.80 (8.72)	0.13	7.85 (8.18)	0.22	NA	NA
DOC	NA	NA	NA	NA	7.28 (3.49)	0.21	8.42 (4.25)	0.24	NA	NA
					Filter Rule 1					
HP1600	−0.9 (2.2)	−4.4	82,4	56.9	8.20 (4.68)	0.15	5.18 (1.88)	0.06	0.04	0.15
HP4	−0.6 (0.3)	−1.6	58,1	59.9	7.18 (5.49)	0.07	4.81 (3.18)	0.06	0.02	0.003
FOD	0.004 (0.6)	−1.5	82,1	53.3	6.90 (5.23)	0.08	7.00 (4.27)	0.13	0.06	0.08
BN	−1.6 (5.5)	−9.5	58,3	60.7	8.22 (5.04)	0.14	6.50 (5.58)	0.07(**)	0.24	0.04
UC	0.3 (0.5)	−0.4	76,2	59.1	7.37 (4.98)	0.12	6.00 (4.94)	0.06(*)	0.02	0.05
LT	−1.1 (5.1)	−8.0	82,4	55.3	7.50 (4.88)	0.11	6.55 (5.07)	0.08	0.22	0.01
SEGM	0.2 (2.9)	−5.4	82,4	67.5	10.11 (5.79)	0.18	5.37 (5.95)	0.03(*)	0.36	0.02
FREQ	−0.8 (2.3)	−3.8	82,4	64.4	6.60 (4.14)	0.11	6.36 (6.13)	0.04(*)	0.05	0.04
HAMIL	−1.5 (2.2)	−4.5	59,4	73.5	9.60 (5.48)	0.16	3.88 (4.29)	0.02(*)	0.01	0.06
CDT	−0.1 (4.9)	−9.0	82,4	59.1	8.75 (8.10)	0.09	6.88 (6.75)	0.06(*)	0.38	0.33
MINDEX	−2.4 (6.3)	−10.3	61,3	53.1	9.00 (5.56)	0.08	5.75 (5.67)	0.05(*)	0.09	0.002
BQ	−5.0 (5.4)	−11.3	57,1	65.1	8.00 (5.73)	0.06(*)	6.00 (5.29)	0.14	0.34	0.56
COIN	−2.2 (2.1)	−5.8	60,4	44.0	7.12 (6.74)	0.25	9.25 (6.15)	0.05(*)	0.20	0.21

with the second one. Interestingly, there are two NBER troughs (64,4 and 75,1) and one DOC trough (75,2) which are unrecorded by practically all methods. Note that the 1975 recession was a multiple dips recession in which the growth rate of output was positive during two quarters so that both dating rules find it difficult to appropriately identify the trough date. Notice also that, generally speaking, all methods are worse in dating peaks than troughs with the second rule. This may be due to the fact that peaks appear more as plateau rather than sharp edges and the second rule finds it difficult to clearly pick a turning date in this situation.

Fourth, for those turning points which are identified within the chosen confidence interval, there are differences across detrending methods, types of turning points and, to some extent, benchmark classification. In general, when NBER benchmark is used many detrending methods generate troughs which lead or coincide and peaks which lead the standard classification, regardless of the dating rule employed. The exceptions are BQ and BN detrending which

Table 2
(continued)

	Amplitude			Expansions			Durations		Correlations	
Method	Average	Min	Date	%	TP	Test	PT	Test	Amplitude PP	Amplitude PT
					Filter Rule 2					
HP1600	−0.6 (0.3)	−1.2	62,4	46.4	7.00 (4.00)	0.08	10.50 (7.09)	0.15	0.13	0.62
HP4	−0.6 (0.4)	−1.6	58,1	51.6	4.75 (4.58)	0.02(*)	4.00 (1.67)	0.08	0.009	0.02
FOD	−1.1 (0.5)	−1.5	57,4	12.4	4.50 (3.10)	0.14(*)	23.40 (26.05)	0.27	0.64	0.69
BN	−1.6 (2.2)	−4.2	57,3	35.6	28.66 (18.61)	0.43	20.00 (16.64)	0.30	0.25	0.25
UC	−0.1 (0.1)	−0.2	83,2	23.8	9.66 (9.86)	0.18(*)	36.33 (21.45)	0.48	0.01	0.03
LT	−1.0 (1.0)	−2.5	75,1	54.5	18.40 (21.45)	0.10(**)	15.00 (20.08)	0.10(*)	0.16	0.16
SEGM	−0.9 (0.2)	−1.2	81,4	42.4	13.40 (11.63)	0.24	15.20 (6.09)	0.49	0.16	0.04
FREQ	−0.8 (0.4)	−1.4	60,4	45.5	7.66 (5.36)	0.11	9.00 (4.24)	0.25	0.005	0.03
HAMIL	−0.9 (0.6)	−1.6	80,2	62.1	21.80 (24.89)	0.11(**)	10.00 (6.96)	0.21	0.01	0.04
CDT	−0.6 (0.6)	−1.3	80,2	55.4	19.00 (14.79)	0.20	27.33 (15.88)	0.21	0.001	0.25
MINDEX	−0.9 (0.1)	−1.9	58,1	39.5	36.00 (14.79)	0.33	10.00 (6.24)	0.48	0.02	0.01
BQ	−2.5 (3.1)	−6.1	57,2	48.4	22.00 (20.47)	0.14	19.66 (20.81)	0.26	0.05	0.07
COIN	−0.9 (0.7)	−2.1	70,1	45.6	8.00 (5.29)	0.37	8.11 (5.75)	0.38	0.64	0.18

Note: TP indicates the average duration of trough to peak cycles, PT indicates the average duration of peak to trough cycles. Test reports the value of Stephens' (1978) statistic. Amplitude-PP reports the Spearman's rank correlation coefficient between the amplitude of contractions and the duration of peak to peak cycles. Amplitude-PT reports the Spearman's rank correlation coefficient between the amplitude of contractions and the duration of peak to trough cycles. (**) indicates a statistic with a p-value between 5% and 10%. (*) indicates a statistic with a p-value below 5%. The second row for each method reports standard deviations in parentheses. With Filter Rule 1 a trough occurs at t if $c_{t+1} > c_t < c_{t-1} < c_{t-2}$ and a peak if $c_{t+1} < c_t > c_{t-1} > c_{t-2}$. With Filter Rule 2 a trough occurs at t if $c_t < 0$ and $c_{t-1} < 0$ or if $c_{t+1} < 0$ and $c_t < 0$ and a peak if $c_t > 0$ and $c_{t-1} > 0$ or if $c_{t+1} > 0$ and $c_t > 0$. NBER refers to the NBER chronology reported by the Center for International Business Cycle Research at Columbia University. DOC refers to the Higgins and Poole chronology compiled using the DOC composite index of leading indicators. The features of NBER cycles are from Moore (1983) and Zarnowitz (1991*b*). NA stands for not available.

produce reference cycles whose turning points lag NBER dates both for peaks and troughs. When DOC reference is used, the results are more heterogeneous. With the first dating rule all detrending methods produce troughs which lead or coincide with DOC troughs while with the second dating rule trough dates generally lead DOC troughs. On the other hand, the reported peak dates generally lag standard DOC dates with the first dating rule, but lead with the second. Also with this classification, both the BQ and BN methods produce turning points which tend to lag benchmark dates.

Univariate procedures generally produce trough dates which anticipate benchmark trough dates in several instances, while multivariate procedures select trough dates which, in general, coincide with benchmark dates when the first dating rule is used. This heterogeneity is less evident with the second dating rule but this may be due to the fact that the number of correctly recognised turning points is typically smaller. Overall, with the second dating rule all detrending methods generate turning point dates which lead by about 1-2 quarters, regardless of the benchmark classification used. On average and regardless of the dating rule employed, each method appears to produce smaller discrepancies relative to the DOC classification.

Fifth, although multivariate detrending procedures employ more information to construct the reference cycle than univariate ones, they do not provide a necessarily superior picture in dating business cycle phases. In particular, these methods produce reference cycles whose turning points do not match NBER or DOC dates and for three out of the four detrending methods, the dating performance is definitively inferior relative to the one of univariate procedures with at least one dating rule. While the relevance of this finding clearly depends on the variables included in the econometrician's information set and different information sets may give different conclusions (e.g. including a measure of the slope of the term structure makes the approach look much better), the results suggest that the loss of information incurred in constructing reference cycles using real GNP alone may be small.

In conclusion, the dating of turning points appears to be sensitive to the choice of detrending. Differences emerge in the dates selected, in the number of cycles discovered and in the number of false alarms and missing signals they generate.

3.2. *The Statistical Properties of Growth Cycles*

Next, we study the statistical properties of generated cycles. In particular, we are interested in the amplitude characteristics of contractions and in the duration and persistence of various business cycle phases, as these are the statistics typically employed in the literature to summarise the properties of reference cycles.

3.2.1. Amplitudes

Amplitude measures display significant differences across detrending methods and dating rules. With the first rule the largest average amplitude is -5.0%, which is obtained with BQ, while the others range from -0.1% obtained with MLT to $-$ 2.4% obtained with MINDEX. For two methods (UC and SEGM) the average amplitude of contractions is positive, i.e. on average, contractions were mere slowdowns of economic activity which did not involve crossing below the trend of the real GNP series. With the second rule the average amplitude of contractions is, in general, smaller. The maximum value is -2.5% obtained, once again, with BQ, while the others range from -0.1% obtained with UC to -1.1% obtained with FOD filter.

Additional information on the amplitude of the resulting cycles can be obtained by examining the timing and the severity of the worst contraction. With the first dating rule, the severity of the worst contraction varies substantially with the detrending method, ranging from −0.4% with UC to −11.3% with BQ, with most of the other methods producing a drop of approximately 4.0–5.0% below the trend line. Out of the 13 procedures, 5 picked 1982,4 as the worst time and one 1982,1, while the remaining 7 methods selected dates from 1957 to 1960. Interestingly enough, no method except UC selected a date in the middle of the 1970's as the worst time in the sample. Once again, there is much more homogeneity in the results with the second rule: the range for amplitude of the worst contraction is between −0.2% with UC and −6.1% with BQ, with most other methods producing a maximum fall of 1.5–2.0% below the trend line. This homogeneity however is more the result of the poor dating record of many procedures rather than an intrinsic similarity of the reference cycles generated with this dating rule. This impression is confirmed by the considerable variety of dates picked by each method as the worst contraction date. Four methods selected dates between 1957 and 1958 and two picked 1980,2, but for the rest there appears to be little congruence. Note that LT selected 1975,1 as the worst recession and at this date the cyclical component of GNP was about 2.5% below the trend.

3.2.2. Durations

The average duration of expansions is not sensitive to detrending when the first rule is used: the range is between 6.6 and 10.1 quarters (with a standard deviation of about 5 quarters) and, except marginally for BQ, there is no evidence of duration dependence for this phase. That is, there is no evidence that expansions tend to terminate the longer they have lasted. A somewhat different picture emerges when we look at contractions. In this case the range of average durations is slightly larger, varying from 3.8 to 9.25 quarters, but for 7 out of the 13 methods, the null hypothesis of no duration dependence of contractions is rejected. However, there seems to be no relationship between the average duration of contractions and the rejection of the hypothesis of no duration dependence. Therefore, in agreement with Diebold and Rudebush (1990), the prediction of peak dates is problematic, given the highly irregular nature of expansion phases, but it appears to be easier to predict trough dates. This is generally true regardless of the detrending method.

The average duration of expansions exceeds the average duration of contractions for all reference cycles except those generated with FOD and COIN. Typically, expansions last about 1.5 times longer than contractions. In addition, all reference cycles indicate that the economy is expanding 5–15% more times than contracting. Hence, the cycles obtained are, on average, asymmetric.

With the second rule the features of the durations of business cycle phases strongly depend on detrending. The average duration of expansions ranges from 4.5 quarters with FOD to 36 quarters with MINDEX, while the average duration of contractions ranges from 4 quarters with HP4 to 36.3 quarters with UC. The range of standard deviations is large as well ranging from 4 to 24 quarters for expansions and from 1.6 to 24.1 for contractions. With this second dating rule there is some evidence of duration dependence of expansion cycles for five methods, while contractions display duration dependence only with LT. Moreover, for 7 detrending methods the average duration of contractions exceeds the average duration of expansions and except for HP4, LT, HAMIL and MLT the economy is contracting in more than 50% of the time periods. Finally, there are strong asymmetries in the duration of business cycle phases with FOD and UC detrended data.

3.2.3. Persistence

Burns and Mitchell (1943), Moore (1958) and others have argued that the severity of contractions is an important ingredient to know how long it will take to the economy to recover to its previous peak level. A direct test of their conjecture is impossible within the present context because their analysis did not distinguish the trend from the cycle. As a close substitute, I examine the relationship between various business cycle phases and the severity of contractions. The hypothesis then states that the deeper is the contraction (as measured here by the amplitude of the trough relative to the trend), the longer is the duration of the complete peak to peak cycle. On the other hand, there should be no systematic relationship between the depth of contractions and their duration (see Romer (1994) for an alternative view regarding the relationship between the depth of the contraction and their duration).

Table 2 indicates that the conjecture is not supported in the data even though the results should be interpreted with caution because of the small number of durations available with many procedures, especially with the second rule. In general, although the correlation between the severity of contractions and the duration of full peak to peak cycles appears to be stronger than the correlation between the severity of contractions and their duration for all reference cycles, differences are statistically insignificant. Moreover, in both cases, the rank correlation coefficients are not significantly different from zero and this is true regardless of the detrending method used to construct growth cycles and the dating rule employed to classify turning points.

To summarise, the amplitude and duration properties of the business cycle phases depend, as in the case of turning point classification, on the detrending methods and on the dating rule. However, the persistence properties of contractions and peak-to-peak cycles are robustly unrelated to the severity of contractions.

3.3. *Comparison with Benchmark Growth Cycles*

We next turn to the final question addressed in this paper, i.e. which detrending method reproduces the features of standard growth cycles best regardless of the dating rule employed.

Tables 1 and 2 indicate the HP and the FREQ filters come closest to do the job. In particular, they are the detrending procedures which minimise the unweighted sum of false alarms and missing signals, regardless of the dating rule or the benchmark classification used. These methods are conservative in the sense that the implied reference cycles are sufficiently smooth to avoid the generation of too many false alarms while avoiding missing important signals. As a matter of fact, the HP1600 filter and the FREQ filter capture all DOC peaks with the first dating rule while HP4 captures all NBER and DOC dates with the second dating rule. Hence, if missing a signal is more important than giving a false one, the HP filter should be preferred to the others. On average, the turning points they generate slightly lead NBER turning points and are coincident with DOC turning points.

The similarities between HP1600 and FREQ filters we unveil confirm, on one hand, the low band-pass features of the HP filter highlighted by King and Rebelo (1993) and, on the other, the MA features of the FREQ filter (see also Baxter and King (1994)). However, as pointed out by one of the referees, since the filtering procedures employed by NBER and DOC fall within the class of modified MA filters, it is not completely surprising to find that the HP and FREQ filters are best among the procedures we consider. What is surprising is the fact that univariate, mechanical approaches produce results which are similar to intrinsicaly multivariate and judgemental NBER and DOC approaches.

Among the other methods, the BQ approach does well both in terms of false alarms and missing signals with the first dating rule but is clearly inappropriate with the second dating rule. The Hamilton filter also performs very poorly with the second dating rule where it either misses or incorrectly identifies 13 of the 15 turning points of the sample regardless of the benchmark classification employed. This, however, is not surprising since the method was designed to be optimal with a probabilistic dating rule, i.e. the economy is in a contraction if there is at least 50% probability of being in a low state.

The statistical properties of the various business cycle phases generated with HP and FREQ filters are also broadly consistent with both NBER and DOC cycles. In particular, the HP1600 filter generates cycles which are slightly asymmetric as are the NBER cycles, while the FREQ filter cycle closely replicates the more symmetric pattern of DOC cycles. Moreover, the amplitude characteristics of both benchmark cycles are sufficiently well approximated by the growth cycles generated by these methods.

The worst performers in this comparison with benchmark growth cycles are FOD, LT, SEGM and HAMIL. To investigate why these procedures fail to generate cycles that resemble the ones identified by NBER and DOC researchers I present in Fig. 2 the time paths of the cycles generated by these four detrending methods. Shaded regions represent contractions according to the NBER classification. The reference cycle generated by FOD is very erratic, in many standardly classified contractions it is above the trend and in others it does not conform to the two-quarter-declines-over-three rule. The other three methods produce reference cycles which are visually similar even though the amplitude of the fluctuations differ. Note that all these cycles are persistently

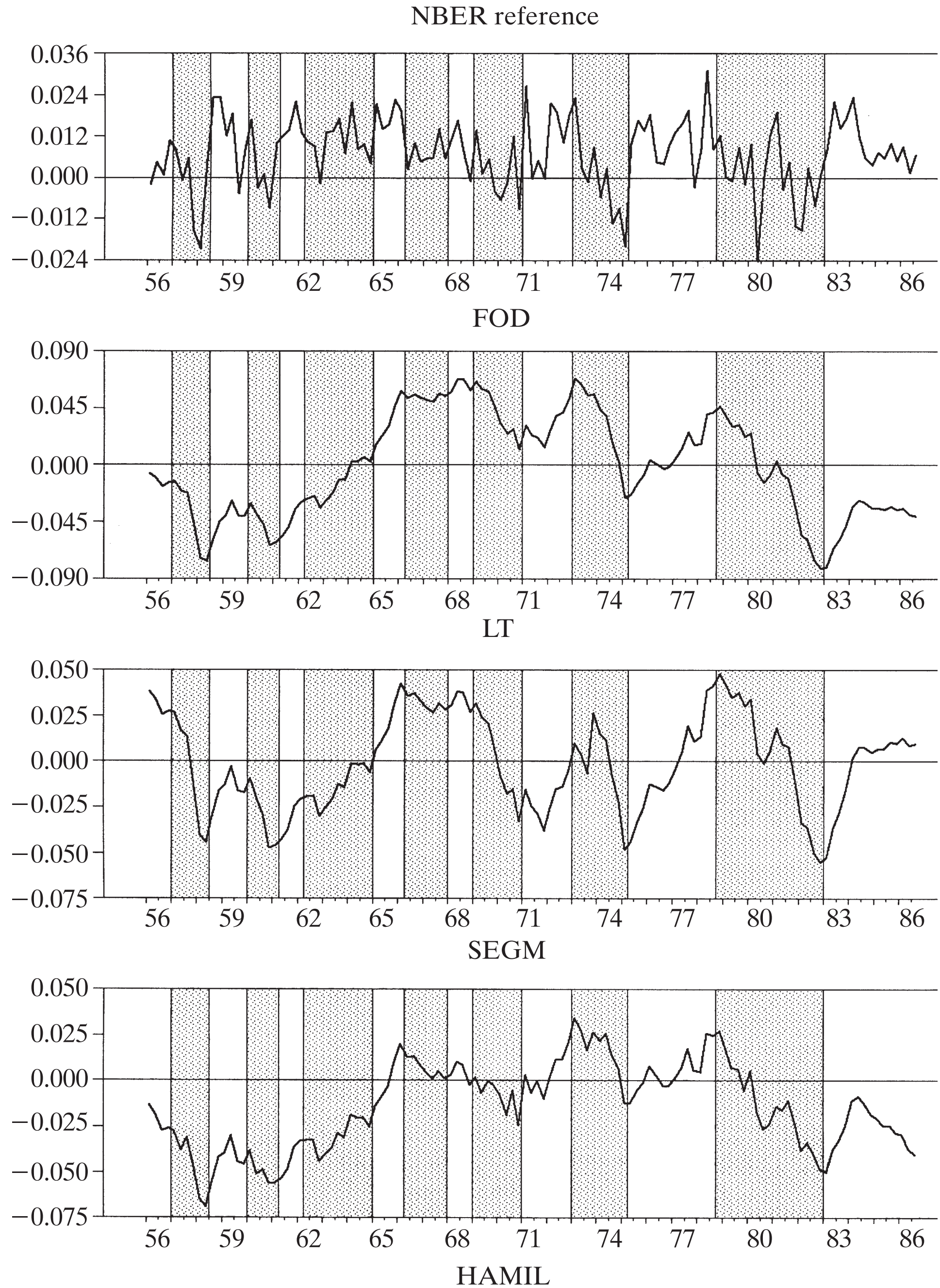


Fig. 2.

on one side of the trend line for long periods of time producing infrequent shifts in the turning point indicator when the second dating rule is used. Furthermore, the average duration of a cycle is 1.5 years with FOD, 7.5 years with HAMIL and more than 8 years with LT and SEGM. Therefore, none of these detrending methods produce cyclical components whose average duration matches the average duration of NBER growth cycles (which have a periodicity of about 4–6 years).

4. Conclusions

This paper examined three questions concerning (i) the sensitivity of turning points classification to different detrending methods and dating rules, (ii) the robustness of the properties of the implied reference cycles and (iii) the ability of different methods to replicate NBER or DOC dating and their business cycles characteristics. We use a variety of detrending methods to separate the trend from the cycle in the data and two different dating rules to select turning points and construct business cycle phases.

Overall, the results indicate that the dating of turning points is sensitive to detrending and dating rules and that both the amplitude and duration properties of the growth cycles generated with alternative detrending methods significantly differ. These results confirm the findings of Canova (1998), who shows that the second moments of the cyclical component of several US real macroeconomic variables are very sensitive to detrending. The sensitivity of outcomes to detrending is easily interpretable since different detrending methods leave cycles of different average duration in the data. What is surprising is that differences in the average duration of cycles are somewhat irrelevant when the first dating rule is used. That is, while the second moment properties of the data vary with detrending, the time paths of the various cyclical components are not too different. These differences are however amplified with the second rule because the crucial factor for dating turning points and selecting business cycle phases is whether the reference cycle is above or below the trend line. In this case, asymmetries emerge because the average spans of time spent above and below the trend line differ across detrending methods.

Is there any sensible way to reduce the range of outcomes by eliminating some detrending methods as ‘unreasonable’? If we take the ability to reproduce a standard turning points classification as a limited information test to select a class of detrending methods, then the results suggest that HP and FREQ filters are those which come closest in reproducing standard dating and business cycle features. Turning points line up in the right way and, regardless of the dating rule, the statistics features of the implied cycle resemble those of NBER or DOC growth cycles. This apparent superiority of this class of low band-pass filters, however, should be weighed against the drawbacks noted by King and Rebelo (1993), Harvey and Jeager (1993), and Cogley and Nason (1995). For a more complete answer on the subject it is therefore necessary to confront the various detrending procedures with alternative and, possibly, more powerful tests.

This paper did not address questions concerning the construction of leading indicators and of useful statistics to evaluate the record and the quality of turning point forecasts. In the literature on the subject (see e.g. Wecker (1979), McNees (1991) or Zellner and Hong (1991)), the results generally hinge on having available a 'correct' reference cycle. Therefore, the results contained in this paper are of interest to researchers engaged in these important activities as they give a rationale for choosing one concept of cycle or one dating rule over another. On the other hand, forecasting exercises comparing both the record and the quality of turning point selections may be a useful class of tests to examine the superiority of one trend specification over another. We plan to conduct these experiments in future research.

Universitat Pompeu Fabra; Universita' di Modena; University of Southampton and CEPR

Date of receipt of first submission: November 1996
Date of receipt of final typescript: June 1998

References

Baxter, M. and King, R. (1994), 'Measuring business cycles: approximate band pass filters for economic time series,' University of Virginia, manuscript.

Beveridge, S. and Nelson, C. (1981), 'A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the ''business cycle'',' *Journal of Monetary Economics*, vol. 7, pp. 151–74.

Blanchard, O. and Quah, D. (1989), 'The dynamic effects of aggregate demand and supply disturbances,' *American Economic Review*, vol. 79, pp. 655–73.

Boldin, M. (1994), 'Dating turning points of the business cycle,' *Journal of Business*, vol. 67, pp. 97–131.

Bry, G. and Boschan, C. (1971), 'Cyclical analysis of time series: selected procedures and computer Programs,' NBER technical working paper 20.

Burns, A. and Mitchell, W. (1943), *Measuring Business Cycles*, New York, N.Y.: NBER.

Canova, F. (1994), 'Detrending and turning points,' *European Economic Review*, vol. 38, pp. 614–23.

Canova, F. (1995), 'Does detrending matter for the determination of the reference cycle and the selection of turning points?,' UPF working paper 115.

Canova, F. (1998), 'Detrending and business cycles facts,' *Journal of Monetary Economics*, vol. 41, pp. 475–512.

Carvalho, J., Grether, D. and Nerlove, M. (1979), *Analysis of Economic Time Series: A Synthesis*, New York, N.Y.: Academic Press.

Coddington, J. and Winters, A. (1987), 'The Beveridge-Nelson decomposition of a time series: a quick computational approach,' *Journal of Monetary Economics*, vol. 19, pp. 125–7.

Cogley, T. and Nason, J. (1995), 'Effects of the HP filter on integrated series,' *Journal of Economic Dynamics and Control*, vol. 19, pp. 253–78.

Diebold, F. and Rudebush, G. (1990), 'The duration dependence of the American business cycle,' *Journal of Political Economy*, vol. 93, pp. 598–616.

Diebold, F. and Rudebush, G. (1992), ' Have postwar economic fluctuations been stabilized,' *American Economic Review*, vol. 82, pp. 994–1005.

Hamilton, J. (1989), 'A new approach to the economic analysis of nonstationary time series and the business cycle,' *Econometrica*, vol. 57, pp. 357–84.

Harvey, A. C. and Jaeger, A. (1993), 'Detrending, stylized facts and the business cycle,' *Journal of Applied Econometrics*, vol. 8, pp. 231–47.

Hymans, S. (1973), 'On the use of leading indicators to predict cyclical turning points,' *Brookings Papers on Economic Activity*, vol. 2, pp. 339–84.

Hodrick, R. and Prescott, E. (1980), 'Post–war U.S. business cycles: an empirical investigation,' Discussion Paper 451, Carnegie Mellon University.

King, R. and Plosser, C. (1994) 'Real business cycles and the test of the Adelmans,' *Journal of Monetary Economics*, vol. 33, pp. 405–38.

King, R., Plosser, C. and Rebelo, S. (1988), 'Production, growth and business cycles: I. The basic neoclassical model,' *Journal of Monetary Economics*, vol. 31(2/3), pp. 195–232.

King R., Plosser, C., Stock, J. and Watson, M. (1991), 'Stochastic trend and economic fluctuations,' *American Economic Review*, vol. 85, pp. 819–40.
King, R. and Rebelo, S. (1993), 'Low filtering and the business cycles,' *Journal of Economic Dynamics and Control*, vol. 17, pp. 207–31.
Kydland, F. and Prescott, E. (1990) 'Business cycles: real facts and a monetary myth,' *Federal Reserve Bank of Minneapolis, Quarterly Review*, (Spring), pp. 3–18.
Lahiri, K. and Moore, G., eds., (1991) *Leading Economic Indicators: New Approaches and Forecasting Record*, Cambridge, Ma:. Cambridge University Press.
McNees, S. (1991) 'Forecasting cyclical turning points: the record in the past three recessions' in K. Lahiri and G. Moore (1991) pp. 151–68.
Mintz, I. (1969) 'Dating postwar business cycles: methods and applications to Western Germany, 1950–1967,' NBER occasional paper 107, New York, NY: NBER.
Moore, G. (1958), 'Measuring recessions,' NBER occasional paper 61.
Moore, G. (1983), *Business Cycles, Inflation and Forecasting*, Cambridge, Ma: Ballinger Publishers for the NBER.
Moore, G. and Zarnowitz, V. (1982), 'Sequential signals of recession and recovery,' *Journal of Business*, vol. 55, pp. 57–85.
Nelson, C. and Plosser, C. (1982), 'Trends and random walks in macroeconomic time series,' *Journal of Monetary Economics*, vol. 10, pp. 139–62.
Niemera, M. (1991), 'An international application of Neftcy's probability approach' in Lahiri and Moore (1991), pp. 91–108.
OECD (1986), *OECD Leading Indicators and Business Cycles in Member Countries, 1960-1985, Sources and Methods*, Working Paper 39.
Pagan, A. (1997), 'Toward an understanding of some business cycle characteristics,' *Australian Economic Review*, vol. 30, pp. 1–15.
Romer, C. (1994) 'Remeasuring business cycles,' *Journal of Economic History*, vol. 54, pp. 573–609.
Shapiro, S. and Wilks, M. (1972), 'An analysis of variance test for the exponential distribution (complete samples),' *Technometrics*, vol. 14, pp. 355–70.
Sims, C. (1974), 'Seasonality in regression', *Journal of the American Statistical Association*, vol. 69, pp. 618–26.
Simkins, S. (1994), 'Do real business cycle models exhibit business cycle behavior?,' *Journal of Monetary Economics*, vol. 33, pp. 573–609.
Stephens, M. A. (1978), 'On the W test for exponentiality with origin known,' *Technometrics*, vol. 20, pp. 33–5.
Stock, J. and Watson, M. (1988), 'Testing for common trends,' *Journal of the American Statistical Association*, vol. 83, pp. 1097–107
Stock, J. and Watson, M. (1989), 'New index of coincident and leading indicators,' *NBER Macroeconomic Annual*, Cambridge, Ma.: MIT Press, pp. 351–93.
Wabha, G. (1980), 'Improper prior, spline smoothing and the problem of guarding against model errors in regression,' *Journal of the Royal Statistical Society*, Series B, vol. 40, pp. 364–72.
Watson, M. (1986), 'Univariate detrending methods with stochastic trends,' *Journal of Monetary Economics*, vol. 18, pp. 49–75.
Watson, M. (1994), 'Business cycle durations and postwar stabilization of the US economy,' *American Economic Review*, vol. 84, pp. 24–46.
Webb, R. (1991), 'On predicting stages of business cycles' in Lahiri and Moore (1991) pp. 109–28.
Wecker, W. (1979), 'Predicting the turning points of a time series,' *Journal of Business*, vol. 52, pp. 35–50.
Zellner, A. and Hong, C. (1991), 'Bayesian methods for forecasting in economic time series: sensitivity of forecasts to asymmetry of loss structures,' in Lahiri and Moore (1991) pp. 129–40.
Zarnowitz, V. (1991*a*) 'What is a business cycle ?,' NBER working paper 3863.
Zarnowitz, V. (1991*b*), *Business Cycles: Theory, History, Indicators and Forecasting*, Chicago, Il.: University of Chicago Press for the NBER.
Zarnowitz, V. and Moore, G. (1991), 'Forecasting recessions under the Gramm-Rudman-Hollings law' in Lahiri and Moore (1991) pp. 257–74.

Table A1

Business cycle chronology, Sample 55,3-90,1

Method	Troughs	Peaks
NBER	58,2; 61,1; 64,4, 67,4, 70,4; 75,1; 82,4	57,1; 60,1; 62,2; 66,2; 69,1; 73,1; 78,4
DOC	58,2; 60,4; 63,1; 67,2; 70,4; 75,2; 82,1; 86,2	59,2; 62,2; 65,4; 67,4; 73,1; 78,2; 84,1
	Filter Rule 1	
HP1600	58,2; 60,4; 62,4; 64,4; 67,2; 70,2; 73,3; 76,3; 79,2; 82,4; 85,2(*)	57,1; 59,2; 62,1; 63,3; 66,1; 68,3; 73,1; 76,1; 77,3; 81,1; 84,2
HP4	58,1; 60,4; 62,4; 64,4; 67,2; 68,4; 73,3; 76,4; 78,1; 82,1; 84,4(*)	57,1; 58,4; 61,4; 63,3; 66,1; 68,2; 70,2; 76,1; 77,3; 81,1; 84,1
FOD	57,2; 62,4; 66,2; 68,4; 72,3; 73,3; 76,3; 79,2; 82,1; 84,4(*)	58,4; 63,3; 67,3; 70,3; 73,1; 75,3; 77,3; 81,1; 83,2
BN	58,3; 61,1; 67,3; 70,3; 73,4; 76,4; 79,3; 83,1 85,1(*)	59,3; 62,4; 68,4; 73,2; 76,2; 77,4; 81,2; 84,3
UC	62,1; 64,1; 66,1; 68,3; 70,1; 76,2; 79,2; 83,2	61,2; 63,3; 64,4; 67,2; 69,3; 74,2; 77,3; 82,2; 85,3(*)
LT	56,3; 58,2; 60,4; 67,2; 70,2; 73,3; 76,3; 79,2; 82,4; 85,2(*)	57,1; 59,2; 62,3; 68,3; 73,1; 76,1; 77,3; 81,1; 84,2
SEGM	56,3; 59,4; 67,2; 68,4; 73,3; 76,3; 79,2; 82,4; 84,4(*)	59,2; 62,3; 68,2; 73,1; 76,1; 77,3; 81,1; 84,2
FREQ	58,1; 60,4; 62,4; 64,4; 67,2; 68,4; 73,3; 76,3; 79,2; 82,4	57,1; 59,2; 62,1; 63,3; 66,1; 68,2; 73,1; 76,1; 77,3; 81,1; 84,2(*)
HAMIL	56,3; 59,4; 62,4; 67,2; 68,4; 73,3; 76,3; 78,1; 82,1; 85,2(*)	59,2; 62,2; 63,3; 68,2; 73,1; 76,1; 77,3; 81,1; 84,2
CDT	56,3; 58,2; 60,4; 67,2; 68,4; 73,3; 76,3; 79,2; 82,4; 85,2(*)	57,1; 59,2; 62,3; 68,2; 73,1; 76,1; 77,3; 81,1; 84,2
MINDEX	58,2; 61,3; 68,4; 73,3; 77,1; 78,1; 79,2; 82,1	57,1; 58,4; 67,3; 73,1; 76,1; 77,3; 78,4; 80,1; 84,2(*)
BQ	57,1; 58,4; 61,2; 64,4; 67,4; 70,4: 78,3; 82,4;	57,3; 59,2; 62,3; 66,3: 69,1: 76,3; 80,3; 84,1(*)
COIN	58,2; 60,4; 67,2; 68,4; 73,3; 76,3; 79,2; 82,1; 85,2(*)	59,2; 62,3; 68,2; 73,1; 76,1; 77,3; 81,1; 84,2
	Filter Rule 2	
HP1600	57,4; 60,3; 62,4; 69,4; 74,3; 80,2 81,4	55,4(*); 59,1; 61,4; 65,2; 72,2; 77,3; 80,4; 84,1(*)
HP4	56,1(*); 58,1; 59,3; 60,4; 62,4; 64,4; 66,4; 70,1; 71,4; 75,1; 76,3; 77,4; 79,1; 80,2; 82,3; 84,4(*)	56,4: 58,4; 60,1; 61,4; 64,1; 65,4; 68,2; 71,1; 73,1; 75,3; 77,2; 78,2; 79,3; 81,1; 84,1
FOD	57,4; 69,4; 74,3; 81,4	56,2; 58,2; 72,1; 75,2; 82,4(*)
BN	57,3(*); 75,1; 80,3(*)	65,2; 77,2
UC	76,1; 83,2	60,4(*); 76,4; 84,3(*)
LT	56,1; 75,1; 76,3; 80,2(*)	55,4(*); 64,1; 76,1; 77,1
SEGM	57,4; 70,1; 74,3; 81,4	55,4(*); 65,1; 73,1; 77,2; 84,1(*)
FREQ	57,4; 60,4; 62,4; 69,4; 74,4; 80,2; 81,4; 86,1(*)	54,4(*); 58,4; 61,4; 65,3; 72,2; 78,2; 80,4; 83,4
HAMIL	56,1(*); 69,4; 75,1; 76,3; 80,2(*)	65,4; 71,1; 76,1; 77,1
CDT	56,3; 74,4; 80,2(*)	55,4(*); 64,1; 77,1
MINDEX	58,1(*); 74,3	55,4(*); 65,2; 81,1(*)
BQ	57,2; 61,1; 63,2; 67,2; 70,2; 74,2; 78,2; 80,1	55,4(*); 60,3; 62,1; 64,3; 68,3; 72,3; 76,2; 78,4; 84,3(*);
COIN	57,3(*); 70,1; 74,1(*)	65,4; 72,4

Note: With Filter Rule 1 a trough occurs at t if $c_{t+1} > c_t < c_{t-1} < c_{t-2}$ and a peak if $c_{t+1} < c_t > c_{t-1} > c_{t-2}$. With Filter Rule 2 a trough occurs at t if $c_t < 0$ and $c_{t-1} < 0$ or if $c_{t+1} < 0$ and $c_t < 0$ and a peak if $c_t < 0$ and $c_{t-1} < 0$ or if $c_{t+1} < 0$ and $c_t < 0$. NBER refers to the NBER chronology reported by the Center for International Business Cycle Research at Columbia University. DOC refers to the Higgings and Poole chronology compiled from the DOC composite index of leading indicators. Both are taken from Niemera (1991) and checked against those reported by Simkins (1994). A * indicates that the previous or the next turning point is censored.