Detrending and business cycle facts: A comment

Craig Burnside*

The World Bank, N11-055, 1818 H Street NW, Washington, DC 20433, USA

Abstract

There is nothing misleading in the fact that different filtering techniques lead to different facts about macroeconomic time series. The fact that economists use a large number of filters to extract the 'cyclical' and 'trend' components of time series simply means that these concepts do not have unique meaning among them. Alternative filters provide different windows through which economists can examine their models and data. It is an open question as to whether some of these windows are more or less interesting to look through. The fact that some economists restrict themselves to a small set of filters is an issue to the extent that they thereby induce a lack of power. Here, I argue that a commonly used method of testing business cycle models induces no such lack of power. © 1998 Elsevier Science B.V. All rights reserved.

JEL classification: B41; C12; C52; E32

Keywords: Filtering; Hypothesis testing; Power

1. Introduction

Canova (1998) has shown that a set of popular filters used in detrending macroeconomic time series have dramatically different properties in the following sense. The second moments of the 'cyclical' components produced by these filters for standard macroeconomic time series such as real GNP, consumption, investment, hours worked, real wages, productivity and the capital stock, are highly sensitive to the choice of filter. The second moments considered include the standard deviation of GNP, the standard deviation of each of the other series relative to GNP, various dynamic cross correlations between pairs of

---

* E-mail: aburnside@worldbank.org

these series, and impulse response functions for each filtered series to an innovation in filtered GNP. Since these moments are commonly the focus of empirical research in the business cycle literature, the information in Canova's paper is of particular interest to researchers in this field.

While the diversity of facts presented in his paper is interesting, Canova (1998) concludes that there is something misleading in the fact that different filtering techniques lead to different facts about macroeconomic time series. In my opinion this is not the case on two grounds. First, he overstates his case, because there are many facts about the business cycle which should be accepted as being robust, especially when detrending and extracting business cycle components of a time series are recognized as being distinct exercises. Second, and more importantly, I will argue that when the facts differ according to the filter, this simply means there are many facts to be explained. Economists will be misled only to the extent that they believe that all filters designed to extract the 'cyclical' and 'trend' components of time series produce the same outcomes. Since, I would argue, it is widely accepted that the concepts of 'trend' and 'cycle' do not have unique meaning among economists, there is not too much to worry about here. However, to the extent that Canova has correctly highlighted a methodological problem, I will argue, in the next section, that he has proposed the wrong solution.

The third section will argue that alternative filters provide different windows through which economists can examine their models and data. It is an open question as to whether some of these windows are more or less interesting to look through. This is a debate that is unlikely to be settled in the near future, and, perhaps, does not need to be. The fact that some economists restrict themselves to a small set of filters is an issue to the extent that they thereby induce a lack of power. In other words, there may be some filters that, presumably due to bad small sample properties, tend to lead to weak evidence against economists' favored models. Or, put differently, there may be filters which tend to make it more difficult to distinguish one business cycle model from another. Since this is an issue of statistical properties of formal hypothesis tests based on filtered data, Canova's analysis sheds no light on this issue. In the third section, I use one example to argue that a commonly used method of testing business cycle models induces no such lack of power.

2. Detrending and filtering

Canova points to two main issues in his paper. First there is the fact that many economists have proposed alternative definitions of the trend in a macroeconomic time series, and that furthermore, each of these definitions of the trend has different implications for the statistical properties of both the trend and the residual, commonly referred to as the cyclical component. Second, he points out

that many economists prefer theoretical explanations of the mechanism driving economic fluctuations, but theory rarely says anything about the exact nature of the trend and its relationship to the cycle.

I will address these issues in two ways. First, are business cycle facts really so sensitive to the choice of detrending filter? Second, even when the facts are sensitive to the detrending method, and theory provides little guidance about how to choose one, should this raise genuine concern among practitioners?

2.1. How robust are the facts?

The facts I will focus on here are ones concerning the second moment properties of the data that are detailed in Canova's first and second tables. To organize my discussion of these facts, I introduce a data set used in Burnside and Eichenbaum (1996b) which includes measures of real output, consumption, and investment per capita, that correspond closely to national accounts measures, and a measure of hours worked per capita, that corresponds to the establishment survey data of the BLS.[1] In Fig. 1 I present estimates of the spectra of four time series, consumption, investment, hours and average productivity compared to the spectrum of output. To have some hope that these estimates have reasonable statistical properties I have linearly detrended the data. We see that each of the spectra displays the typical spectral shape of a deseasonalized macroeconomic time series, with a great deal of power in the low frequencies and a general decline as the frequency increases.

These estimates dramatically illustrate that there are certain facts which are almost certain to be robust to the filtering method, as long as it removes at least as much variation in the series as the linear detrending method does. For example, the spectrum of consumption lies everywhere below that of output. In fact, Canova shows that for all twelve of the filters he considers the variance of consumption is less than that of output, yet he concludes, that 'even within this class of procedures whether consumption is excessively smooth or not depends on detrending'. The only filter for which the relative standard deviation exceeds 0.75 is the one-dimensional index model. The fact that the numerical magnitude of the relative standard deviation ranges from 0.34 to 0.75 for the other filters will not come as a surprise to the reader when we see, in the next section, that the filters weight the various frequencies in very different ways. Another robust

---

[1] The consumption measure is nondurables plus services plus the imputed service flow from durables. The investment measure is private residential and nonresidential fixed investment, plus purchases of consumer durables plus government fixed investment. Output is measured as consumption plus investment, inventory investment, net exports, and government purchases of goods and services excluding investment. The data are described in more detail in Burnside and Eichenbaum (1996b).
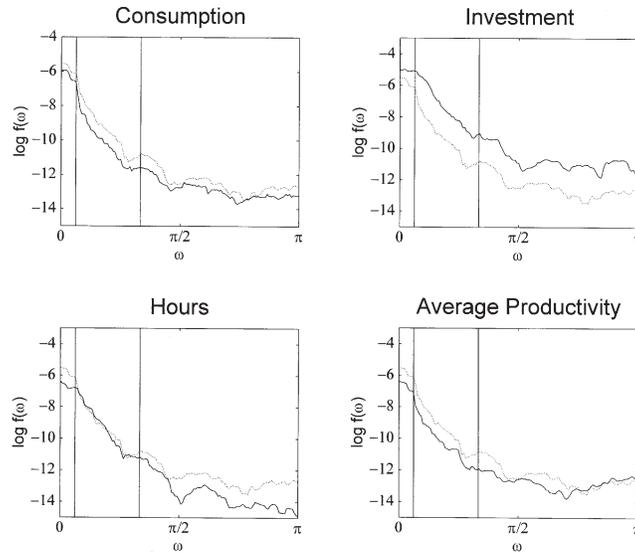
Fig. 1. The spectra of four time series. Each graph illustrates the logarithm of the spectrum of the series indicated (dark line) along with the logarithm of the spectrum of output (dotted line). All series were linearly detrended and are sampled over the interval 1955Q1–1992Q4. The vertical lines delineate the frequencies in the range $[\pi/16, \pi/3]$.

fact is that investment is more volatile than output, since its spectrum lies everywhere above that of output. Again, only the exact magnitude depends on the filter.

While these simple facts are robust across the complete range of frequencies, the facts about hours and average productivity are not. However, it is important at this stage to distinguish between the arguments econometricians have about the right way to detrend, and what I would argue is the general agreement about what the business cycle frequencies are. What I mean by this is that there might be an econometrician who would strongly favor using the first-difference filter to render stationary most standard macroeconomic time series. However, this same econometrician would be unlikely to argue that by so doing he was thereby isolating only the business cycle components of the data. Why? Because, as we shall see in the next section, the first-difference filter puts strong emphasis on high-frequency components.[2] In other words, while the definition of the

---

[2] An econometrician favoring the HP1600 filter would be more likely to make this claim for the following reason. Despite the fact that this filter does not attenuate high frequency components, it puts roughly equal weight on all cycles with a period of less than 6 quarters. Given the typical spectral shape of most time series, and the fact that the filter eliminates low-frequency variation, this means that the business cycle frequencies will be the dominant component of what is left over after filtering.

trend in a time series is not universal, there is more agreement about what business cycles are, and these do not necessarily correspond to the cyclical component defined as the series minus its trend. Canova, himself, argues that business cycle fluctuations have periods of between two and no more than eight years. Sargent (1987), citing the NBER taxonomy, suggests the same.

If we focus on cycles of between 6 and 32 quarters (the band in Figs. 1–3) then we see that hours is about as volatile as output over the cycle, and is less volatile over the other frequencies. Thus, any filter putting most of its weight only on the business cycle frequencies will lead to the conclusion that hours is about as volatile as output. Similarly, average productivity is less variable than output, unless the filter puts strong emphasis on the very highest frequencies. So it seems fair to conclude that productivity is less variable than output over the cycle. In both cases, the behavior of the series within the business cycle frequencies is quite uniform, even if this is not so outside these frequencies.

Now, it is possible that some facts will depend on the particular measures of each time series that we use. For example, the hours series from the household survey may look quite different than the one from the establishment survey. However, this issue, like the one Canova brings up regarding the equivalence of real wages and productivity in real business cycle models, is an issue of measurement, and not one germane to a discussion of robustness to detrending methods.
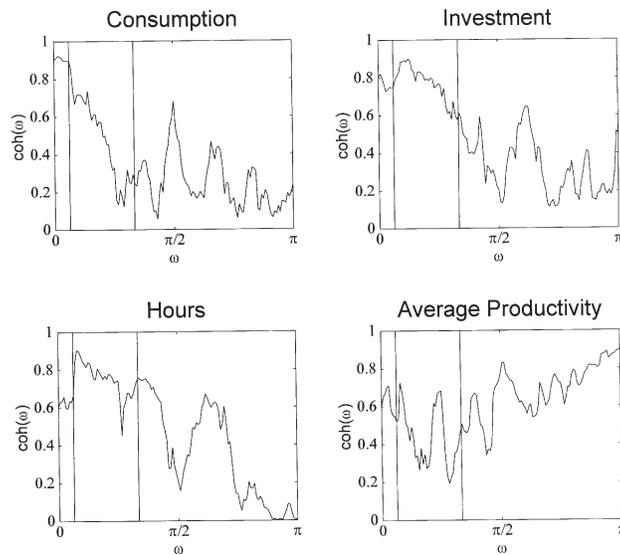

Fig. 2. Coherences of cross spectra with output. Each graph illustrates the coherence of the cross spectrum between the series indicated in the title and output. All series were linearly detrended and are sampled over the interval 1955Q1–1992Q4. The vertical lines delineate the frequencies in the range $[\pi/16, \pi/3]$.
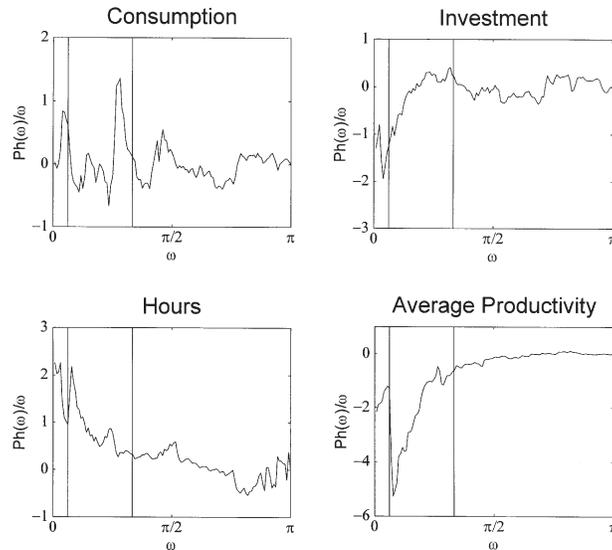
Fig. 3. Phase shifts of cross spectra with output. Each graph illustrates $Ph(\omega)/\omega$, the ratio of the phase of the cross spectrum (between the series indicated in the title and output) to the frequency. All series were linearly detrended and are sampled over the interval 1955Q1–1992Q4. The vertical lines delineate the frequencies in the range $[\pi/16, \pi/3]$.

Regarding the cross correlation patterns, these display more sensitivity to the detrending method. To shed light on this finding I computed estimates of the cross spectra between each of my four variables and output. These estimates are illustrated by the implied measures of coherence in Fig. 2 and phase in Fig. 3. The coherence at frequency ω can be interpreted as the squared correlation between the components of the two series at this frequency. The phase at frequency ω, divided by the frequency, can be interpreted loosely as the number of quarters by which a cycle of frequency ω in output occurs ahead of a similar cycle in each variable.

Fig. 2, which illustrates the coherence, shows that over the entire range of frequencies there is a great deal of variation in the extent to which the time series are correlated with output. For all the variables except average productivity, this correlation is weak at high frequencies and strong at the lowest frequencies. I would argue that this evidence rationalizes the extent of the quantitative variation in the contemporaneous cross correlations presented by Canova. But here, as in Canova's table, the preponderance of the evidence is in favor of the notion that consumption, investment and hours strongly comove with output. And this is certainly true over the business cycle frequencies, if slightly less so for consumption.

Fig. 3, which illustrates the phase of the cross spectrum, shows that outside the business cycle frequencies there is little tendency for any of the variables to lead or lag cycles in output. However, within the business cycle frequencies it would appear that hours is led by output, while investment and to a greater degree average productivity, lead output. Now, if we look over all frequencies, as many of the filters do, it is clear that these lead-lag patterns will depend quite a bit on exactly where the filters put their weight.

To conclude this section, if we accept the notion that arguments about how to detrend are not equivalent to arguments about what the business cycle frequencies are, it is fair to say that there are several regularities concerning business cycles. It remains true that the magnitudes of many second moments are sensitive to the choice of detrending method. In the next section, by focusing on the properties of some of the filters, I will ask whether this sensitivity should be of concern to practitioners.

2.2. Different filters, different outcomes

Canova's paper shows that some second moments, especially, are sensitive to the choice of filtering method. But, as I have alluded to above, this is only natural given that the filters being employed have radically different inherent properties. As long as the practitioner is aware of these inherent properties there is nothing surprising about the fact that the resulting filtered data behave differently.

One way of seeing this is to consider only those univariate detrending methods, catalogued in Canova's paper, which can be represented as either linear or nonlinear filters with coefficients that are independent of the time series in question. These include the first-difference filter, the frequency domain methods and the Hodrick and Prescott (1997) (HP) filter. Consider a raw time series y_t which is covariance-stationary. The spectrum of this time series is defined as

$$F_y(\omega) = \sum_{k=-\infty}^{\infty} c_y(k) e^{-i\omega k}$$

where $c_y(k)$ is the kth autocovariance of y_t. The spectrum describes the variability of the series y_t at different frequencies, ω. Since a cycle of frequency p corresponds to a cycle with period $p = 2\pi/\omega$, it is straightforward to determine from the spectrum of a series whether a significant portion of its variation arises from cycles with interesting periodicity.

Filters can be interpreted as a way of isolating that variation of a time series which occurs within a particular range of frequencies. If we apply a linear filter (FOD, HP1600 or HP4), $B(L)$ to y_t, the resulting filtered series $x_t = B(L)y_t$ will have spectrum $F_x(\omega) = H(\omega)F_y(\omega)$, where $H(\omega) = B(e^{-i\omega})B(e^{i\omega})$ is the transfer
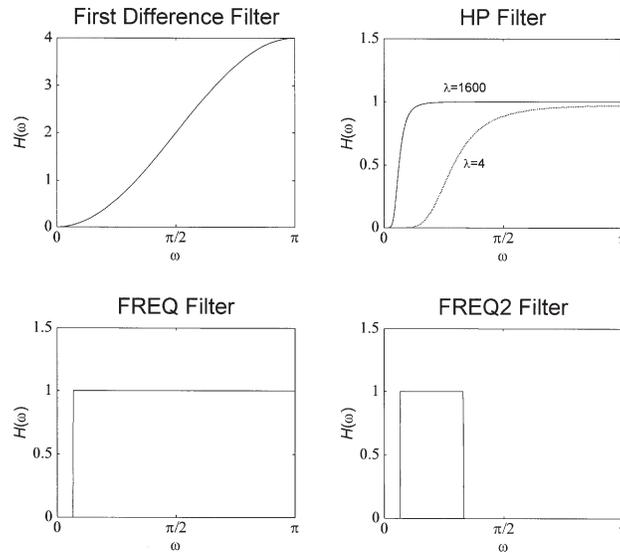
Fig. 4. Transfer functions for five filters. The first-difference filter is $B(L) = 1 - L$. As the sample size grows arbitrarily large the HP filter can be written as $B(L) = C(L)/[1 + C(L)]$ where $C(L) = \lambda(1 - L)^2(1 - L^{-1})^2$. The FREQ filters are defined in the frequency domain, as described in the text.

function associated with $B(L)$. The transfer function indicates the extent to which the filter alters the spectrum of the original time series. If we apply a nonlinear frequency domain filter such as FREQ1 or FREQ2 we obtain the transfer function directly from the definition of the filter, i.e. $H(\omega) = |b(\omega)|^2$ where $b(\omega)$ is the filter defined in the frequency domain.[3]

The transfer functions associated with the five filters mentioned above are illustrated in Fig. 4. These figures assume an arbitrarily large sample size for the HP filter cases, and assume that the frequency domain filters can be implemented in their exact ideal forms. For the first-difference filter the transfer function is

$$H(\omega) = 2 - 2\cos(\omega),$$

for the HP filter it is

$$H(\omega) = \frac{16\lambda^2[1 - \cos(\omega)]^4}{\{1 + 4\lambda[1 - \cos(\omega)]^2\}^2},$$

---

[3] For an application of frequency domain filtering see Baxter and King (1995).

where λ is the parameter of the filter, for the FREQ1 filter it is

$$H(\omega) = \begin{cases} 1 & \text{for } \omega \in [\pi/15, \pi], \\ 0 & \text{otherwise} \end{cases}$$

and for the FREQ2 filter it is

$$H(\omega) = \begin{cases} 1 & \text{for } \omega \in [\pi/15, \pi/3], \\ 0 & \text{otherwise.} \end{cases}$$

Since these are very different functions, it is perhaps not surprising that when they are applied to the same economic time series, the resulting filtered series display very different properties. Although all the filters attenuate the low-frequency movements in a series, only the FREQ2 filter eliminates high-frequency variation, and the first-difference filter magnifies its substantially.

When it comes time to test a model, if the model is a correct description of the data it should be consistent with any fact produced by any filter. Some researchers will focus on some facts because they consider them more interesting or more important. Others may focus on specific facts because they believe they have more information content in a statistical sense. However, any such approach to testing models, i.e. one which uses filtered data, implies that there is a loss of information. What are the consequences of this loss? They are potentially threefold: (i) information interesting to other researchers may be ignored, (ii) because of poor small sample properties, a test based on a single filter may lead to excessive type I errors and (iii) for the same reason, tests based on a single filter may lack the power to distinguish among alternative models. The first problem is familiar, has led to frequent academic debates, and has likely been experienced by most researchers during the editorial process at academic journals. Which filter produces more interesting data is a matter of debate, a debate which is unlikely to be resolved. The second problem could certainly arise, but, given that most models are presumed to be false, it is not the most important. The third problem is a major concern, but Canova's paper sheds no light on this issue. In the next section I try to shed light on this question using a specific example from the real business cycle (RBC) literature.

To conclude this subsection, I would argue that it is hopeless, as Canova does, 'to attempt to identify whether there exists a set of relationships which is invariant to the definition of the cycle employed', and that there is no situation in which 'choosing a standard concept of cycle provides misleading impressions of the comovements of the data'. The first attempt is hopeless when the 'cycle' is defined as the residual from detrending because it would require uniform behavior of each time series over a wide range of frequencies, given the variety of detrending methods considered. Second, we can only be misled by our desire to be, or by not having an understanding of the procedures we employ. That

understanding has been enhanced by Canova's survey because it indicates that there are systematic differences among the popular detrending methods.

2.3. Is there a 'right' way to detrend?

The second issue Canova raises concerns the fact that theory rarely tells us how to detrend. For example, a model may not specifically accommodate a trend. On the other hand, in some cases, theory may act as the justification for a detrending procedure. For example, real business cycle models with exogenous technical progress will typically imply that all real quantities, other than hours worked, will share the same trend, which may be either deterministic or stochastic (see, for example, Singleton, 1988; King et al., 1991). Thus, multivariate methods which extract this common trend from the relevant series are justified by theory. It should be pointed out, however, that all the detrending procedures discussed by Canova could be used in the absence of a theoretical justification simply by justifying their use on statistical grounds.

The question is whether theory needs to be applied in choosing the appropriate filter. Canova argues that it does. The main methodological recommendation in his paper is for a more 'interactive relationship between theory and practice' in which 'theory may indicate which concept of cycle is the object of research and therefore implicitly dictate a class of detrending procedures'. Also 'empirical practice should indicate whether this choice leaves out important features of the data and produces distortions of various kinds'. Since, for the most part, I agree with the second part of this recommendation I will not discuss it further.

I interpret the first part of this recommendation as an argument in favor of the link marked 'A' in Fig. 5. This figure describes an empirical methodology in which the data are prefiltered, and theory, formal or otherwise, is used for two purposes. Through link 'A' it is used to determine which filter is right for the model at hand. Furthermore, using either formal or informal methods, theory is used to describe the behavior data would have if they were generated by a particular model. Canova's argument is that if the filter is chosen without regard to the theory, the filter may obscure the information needed to assess the theory. He uses, as an example, the notion that a labor hoarding model describes the relatively short-term behavior of the economy. He argues that some detrending methods do not allow labor hoarding behavior to be discerned, because they do not focus on cycles of the right periodicity. He suggests that researchers interested in labor hoarding should choose a detrending method that emphasizes cyclical fluctuations of period, say 1 to 3 yr.[4]

---

[4] In his reply to this comment, it would appear that Canova would interpret his paper differently. He suggests that he is not concerned with model evaluation but only with 'collecting a set of interesting statistics'. This seems to abstract from the realities of empirical research, but be that as it may, I am not sure that we should rank facts that are robust over broader ranges of frequencies as more interesting.
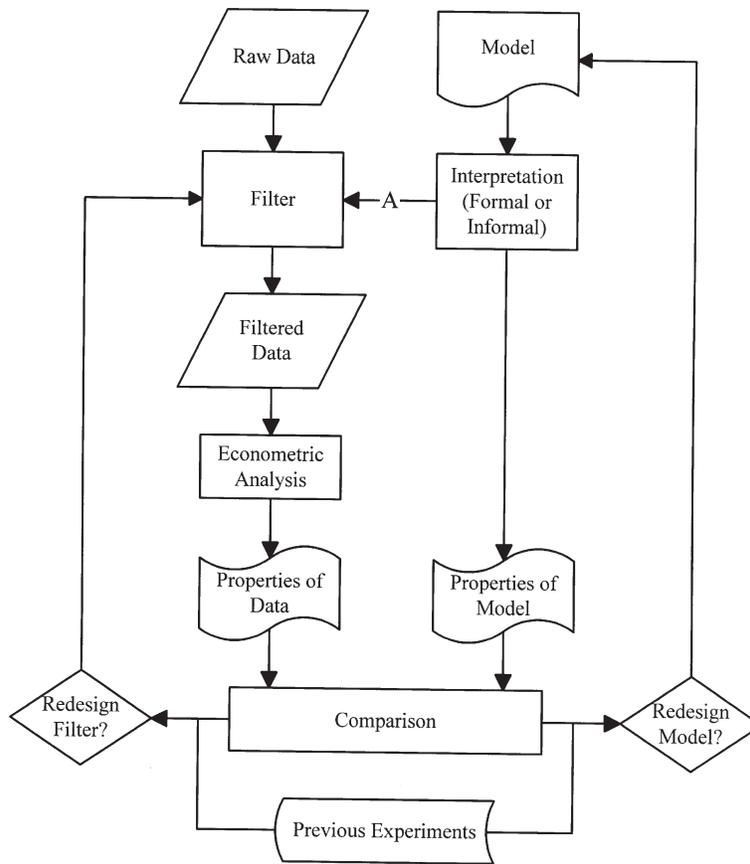
Fig. 5. A flow diagram for an empirical methodology.

There are three reasons to question this recommendation. First, even if we accept the labor hoarding example, rarely will it be the case that a theory describes the behavior of data only at specific frequencies. This simply means that we will rarely be able to implement the recommendation.

Second, it is unnecessary to be so restrictive. Every model should have well-defined implications for data at various frequencies. In the experimental design of Fig. 6, data sampled from an actual economy are treated symmetrically with data from a model economy. These data are filtered, analyzed and compared. If the experiment, along with information from previous experiments suggests that the comparison is a relatively poor one, the researcher has learned that the model economy is a poor description of certain aspects of the actual economy. The researcher can proceed, by using other filters, to learn about other
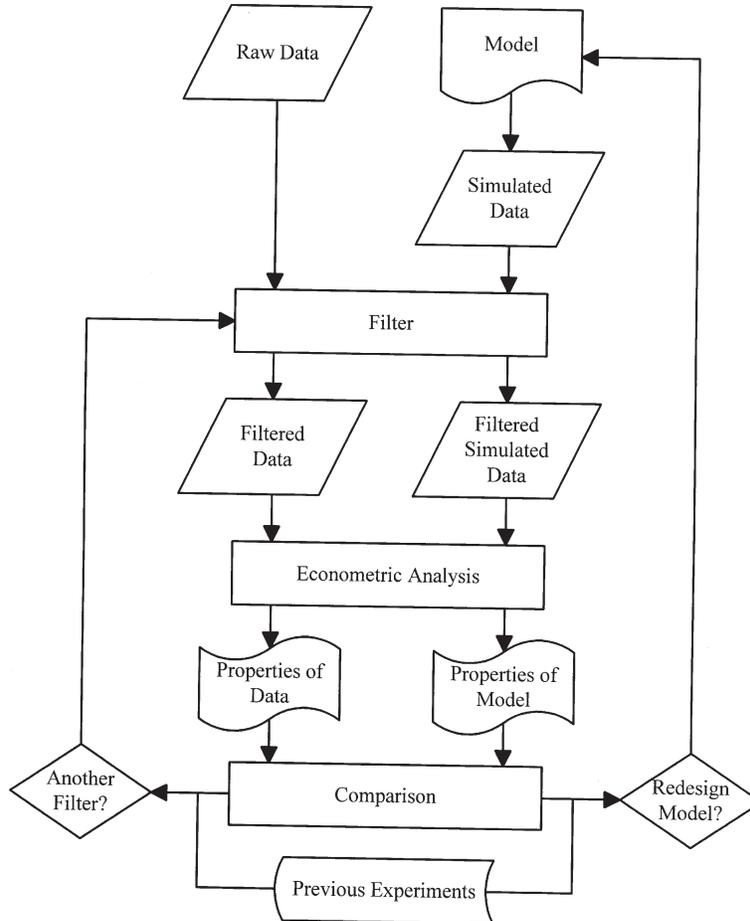
Fig. 6. A Flow diagram for an alternative empirical methodology.

aspects of the model, or he can redesign the model and start over again. If, on the other hand, the comparison is a relatively close one, the researcher has learned that the model economy is a relatively good description of certain aspects of the actual economy. In this case, the researcher will presumably not conclude that his model is true, but will investigate whether the two economies resemble each other along other dimensions. Some of the analysis will presumably make reference to findings using other models. But, the process described in Fig. 6 will only be misleading if the researcher is not thorough enough, and only uses filters which lack the power to (i) distinguish among his models, or (ii) highlight important features of the model or the data. This is the same as my point in the previous section.

Third, the emphasis on the 'A' link in Fig. 5, ignores an inherent weakness in the rest of the experimental design. The main problem with this design is the asymmetry between the analysis of the model and the analysis of the data. To once again take the example of 'hoarding' behavior, it is often argued that such behavior will cause productivity to lead the business cycle. The idea is that unmeasured inputs, such as labor effort, or capital utilization, will be the first to respond to shocks, and that only later, due to adjustment costs, will measured inputs, such as employment, respond. At this level of formalism, however, we know very little about the shocks causing these fluctuations, in terms of their relative importance, persistence, etc. Furthermore, we have not said much about the costs associated with adjusting either the unmeasured or measured inputs. For these reasons, translating the intuitive description of the model into predictions for filtered data is a pointless task.

To take another example, the factor hoarding model I describe in the next section has qualitatively and quantitatively very different implications for the correlation between average productivity and hours worked, depending on whether the data are first-differenced or HP-filtered. A theory in which productivity and hours 'comove' might or might not be consistent with both of these facts, because the statement 'comove' is simply too imprecise to be of use.

Similarly, translating intuition about the workings of a model into the impulse response function corresponding to some kinds of filtered data is impossible. Canova describes the impulse response function of HP-filtered time series to an innovation in HP-filtered output. But such a function has no structural interpretation whatsoever, given that the HP-filter is two-sided and the vector autoregression's further whitening of the filtered data is one-sided. At best, such an impulse response function simply reports a transformation of the partial autocorrelation function of the HP-filtered data. But is there any justification for attempting to discuss whether the patterns displayed by the function are more or less supportive of RBC versus neo-Keynesian theories? I think not.

As a result, a better recommendation would be to follow the formalism of the methodology described by Fig. 6. Here the model and data are treated symmetrically and are subject to formal statistical analysis. A researcher may conclude, after a series of experiments conducted using one filter, that two alternative models are roughly observationally equivalent along the dimensions that have been explored. In the next section I will argue that even using the popular HP filter, and a set of standard statistics, this is an unlikely occurrence. However in such circumstances, a researcher following the methodology of Fig. 3 is unlikely to stop exploring.

3. Exploring the power of business cycle tests

In this section I ask whether tests of business cycle models that are based on a small set of filters induce a lack of power. Specifically, I employ the approach

to estimation and testing employed by Christiano and Eichenbaum (1992), Burnside et al. (1993) and Burnside and Eichenbaum (1996b) which is based on Hansen's (1982) Generalized Method of Moments (GMM). These procedures are straightforward. An exactly identified GMM estimator is used to estimate the parameters of a model. Wald tests are carried out to determine whether the model's predictions for a set of second moments of filtered time series are the same as the corresponding sample moments in US data.

I will not address the properties of two other popular approaches to testing. The first of these, calibration, is similar to the GMM-based approach I describe above. It involves preselection of the model parameters, usually with reference to other empirical evidence, but ignoring the sampling uncertainty inherent to this evidence. Because this uncertainty is ignored, and because moments implied by the model are often compared to sample moments using the so-called 'eyeball' metric, it is less than straightforward to assess the power of calibration-based tests. It is here that Canova's critique may have extra bite. To the extent that the calibrater's 'eyeball' does not adjust its focus to the filter being used, he may be misled by his findings.

Other researchers such as Christiano (1988), Sargent (1989) and McGrattan (1994) have used maximum likelihood methods. Since these methods explicitly consider all aspects of the data, they should not be the subject of scrutiny here.[5]

To address the question of power, I consider two models. The first model (Model *A*) is the model explored by Christiano and Eichenbaum (1992). This is a standard RBC model with the addition of shocks to government spending on goods and services. The second model (Model *B*) is the factor hoarding model explored by Burnside and Eichenbaum (1996b). Model *B* is an extension of Model *A* which allows for cyclical variation in the utilization rate of capital and in the level of unobservable labor effort.

I estimate each of these models using US data from 1955Q1 through 1992Q4 and estimation procedures are described in more detail in an available appendix.

Then I conduct two experiments. In the first experiment, the parameterization of Model *A* implied by my estimates using US data is used as a data generating mechanism to simulate 1000 samples of 152 observations each. For each of these samples I imagine that an artificial econometrician computes a set of sample statistics commonly examined by RBC researchers: the standard deviation of output, σ_y, the standard deviation of consumption relative to that of output, σ_c/σ_y, the standard deviation of investment relative to that of output, σ_i/σ_y, the standard deviation of hours relative to that of output, σ_h/σ_y and the standard deviation of average productivity relative to that of output, σ_{apl}/σ_y. In addition I look at dynamic correlations between average productivity

---

[5] This is true, at least for the time series whose likelihood the researcher has chosen to model. The typical assumption of normality, further rules out the need to consider anything higher than second moments.

and hours as indicators of 'hoarding' behavior, specifically I examine $\rho_{apl,h}^i = \text{Corr}(\text{APL}_t, H_{t+i})$ for $i = -1,0,1$. Finally, to examine the persistence of output, as emphasized by Cogley and Nason (1995) and Burnside and Eichenbaum (1996b), I consider $\rho_y = \text{Corr}(Y_t, Y_{t-1})$. These statistics, any one of which I denote generically as Ψ_1, are computed using both HP-filtered data and first-differenced data. For each of the samples the artificial econometrician also estimates Model A, obtaining parameter estimates denoted $\hat{\theta}_A$. A fully specified version of Model A has implications for each of the above statistics which I denote $\Psi_2^A(\hat{\theta}^A)$.[6] To test Model A, the artificial econometrician can form a Wald statistic based on the difference between Ψ_1 and $\Psi_2^A(\hat{\theta}^A)$ which, under the null that Model A is true, are equal in population. Similarly, the artificial econometrician can test Model B by estimating it, and computing a set of model-based statistics denoted $\Psi_2^B(\hat{\theta}^B)$. By examining the results of these tests across the 1000 samples we can determine the econometrician's probability of rejecting model B (power), which is false, and of rejecting model A (size), which is true. In the second experiment, I simply reverse which of the two models is true.

In his reply, Canova rightly points out that this section ignores the possibility that neither model is the true data generating process (DGP). While the purpose of the exercise is to show that the two models are distinct from one another along various dimensions, it remains an open question as to whether they can be distinguished from other DGPs. It is worth noting, as well, that given the finding of Burnside and Eichenbaum (1996a) that these tests frequently have excessive size, their performance of has been improved. This has been done by using the empirical distribution, across the Monte Carlo draws, of the numerators to form the denominators of the Wald statistics.

The results of the first experiment are reported in Table 1. The table indicates that size is in the neighborhood of what it should be for a test with 5% asymptotic size. Depending on the diagnostic statistic, Model A is rejected in between 4.6% and 11.0% of the samples. On the other hand, there is considerable power to reject Model B. Only $\rho_{apl,h}^{-1}$ for first-differenced data, σ_c/σ_y and σ_h/σ_y for HP filtered data lead to a dramatic lack of power, with probabilities of rejecting Model B of just 5.2%, 6.8% and 2.6%, respectively. For the other statistics power ranges from 17.1% to 100%. Interestingly, in all but one case, $\rho_{apl,h}^{-1}$, there is more power when the data have been first-differenced. But, overall, power is quite good regardless of the filter. The table also indicates the fraction of the experiments in which Ψ_1 was closer to $\Psi_2^A(\hat{\theta}^A)$ than it was to $\Psi_2^B(\hat{\theta}^B)$. This fraction can be thought of as the success rate of a naive econometrician in selecting the correct model. The results indicate that even this naive econometrician will almost always select the right model.

---

[6] The moments commonly reported in RBC studies under the heading 'US data' correspond to Ψ_1. The moments commonly reported under the heading 'Model' correspond to $\Psi_2(\hat{\theta})$.

Table 1
Results of experiments in which Model A is the DGP

	Ψ_1	$\Psi_2^A(\hat{\theta}_A)$	$\Psi_2^B(\hat{\theta}_B)$	Probability of rejecting (%)		Naive (%)
				Model A	Model B	
HP-filtered data						
σ_y	0.0152	0.0153	0.0142	4.8	17.1	60.4
	(0.0017)	(0.0009)	(0.0008)			
σ_c/σ_y	0.463	0.465	0.486	5.8	6.8	68.3
	(0.030)	(0.001)	(0.008)			
σ_i/σ_y	2.40	2.42	2.55	6.5	35.1	84.7
	(0.06)	(0.06)	(0.07)			
σ_h/σ_y	0.629	0.610	0.625	8.5	2.6	43.3
	(0.030)	(0.022)	(0.033)			
σ_{apl}/σ_y	0.463	0.465	0.574	5.8	84.2	96.0
	(0.030)	(0.001)	(0.008)			
$\rho_{apl,h}^{-1}$	0.552	0.609	0.285	7.8	59.8	84.9
	(0.096)	(0.055)	(0.047)			
$\rho_{apl,h}^{0}$	0.672	0.729	0.392	8.5	64.7	89.7
	(0.079)	(0.073)	(0.063)			
$\rho_{apl,h}^{1}$	0.372	0.425	0.796	8.2	98.8	98.2
	(0.105)	(0.048)	(0.064)			
ρ_y	0.698	0.718	0.824	6.7	75.9	90.8
	(0.063)	(0.000)	(0.002)			
First-differenced data						
σ_y	0.0120	0.0119	0.0088	7.1	100.0	100.0
	(0.0007)	(0.0007)	(0.0005)			
σ_c/σ_y	0.449	0.453	0.599	5.8	100.0	100.0
	(0.017)	(0.001)	(0.010)			
σ_i/σ_y	2.41	2.43	2.71	5.5	78.6	96.4
	(0.03)	(0.06)	(0.11)			
σ_h/σ_y	0.633	0.615	0.784	11.0	93.0	99.2
	(0.016)	(0.022)	(0.041)			
σ_{apl}/σ_y	0.449	0.453	0.947	5.8	100.0	100.0
	(0.017)	(0.001)	(0.014)			
$\rho_{apl,h}^{-1}$	0.033	0.030	0.007	4.6	5.2	56.2
	(0.080)	(0.001)	(0.000)			
$\rho_{apl,h}^{0}$	0.702	0.752	-0.345	9.2	100.0	100.0
	(0.040)	(0.074)	(0.022)			
$\rho_{apl,h}^{1}$	-0.074	-0.077	0.856	4.7	100.0	100.0
	(0.076)	(0.004)	(0.036)			
ρ_y	-0.010	-0.007	0.355	4.7	99.6	98.9
	(0.081)	(0.001)	(0.008)			

The table reports the mean and standard deviation (in parentheses) of each statistic across 1000 Monte-Carlo experiments. The probabilities of rejecting are based on statistics compared to the 5% critical value of the chi-squared distribution with one degree of freedom. The numbers reported under the heading 'naive' indicate the fraction of the experiments in which Ψ_1 was closer to $\Psi_2^A(\hat{\theta}^A)$ than it was to $\Psi_2^B(\hat{\theta}^B)$.

Table 2
Results of experiments in which Model B is the DGP

	Ψ_1	$\Psi_2^A(\hat\theta_A)$	$\Psi_2^B(\hat\theta_B)$	Probability of rejecting (%) Model A	Model B	Naive (%)
HP-filtered data						
σ_y	0.0113	0.0135	0.0114	32.0	5.4	82.6
	(0.0014)	(0.0008)	(0.0006)			
σ_c/σ_y	0.481	0.465	0.486	11.3	5.8	54.9
	(0.033)	(0.001)	(0.008)			
σ_i/σ_y	2.52	2.42	2.55	22.9	6.3	64.2
	(0.08)	(0.05)	(0.06)			
σ_h/σ_y	0.655	0.603	0.625	37.1	9.4	78.1
	(0.037)	(0.019)	(0.033)			
σ_{apl}/σ_y	0.576	0.465	0.574	91.9	4.9	93.3
	(0.041)	(0.001)	(0.007)			
$\rho_{apl,h}^{-1}$	0.227	0.628	0.285	94.8	7.6	96.2
	(0.122)	(0.047)	(0.046)			
$\rho_{apl,h}^{0}$	0.323	0.754	0.392	98.0	8.3	98.2
	(0.114)	(0.062)	(0.062)			
$\rho_{apl,h}^{1}$	0.747	0.441	0.797	91.1	9.3	97.1
	(0.063)	(0.041)	(0.063)			
ρ_y	0.810	0.718	0.824	23.4	6.4	84.1
	(0.041)	(0.000)	(0.001)			
First-differenced data						
σ_y	0.0072	0.0105	0.0071	100.0	7.1	100.0
	(0.0005)	(0.0006)	(0.0004)			
σ_c/σ_y	0.590	0.453	0.599	100.0	6.6	99.3
	(0.027)	(0.001)	(0.010)			
σ_i/σ_y	2.70	2.44	2.71	92.5	5.9	91.4
	(0.08)	(0.05)	(0.08)			
σ_h/σ_y	0.817	0.607	0.784	100.0	9.7	99.8
	(0.038)	(0.018)	(0.040)			
σ_{apl}/σ_y	0.944	0.453	0.948	100.0	4.8	100.0
	(0.058)	(0.001)	(0.009)			
$\rho_{apl,h}^{-1}$	0.005	0.031	0.007	7.2	5.0	55.2
	(0.083)	(0.001)	(0.000)			
$\rho_{apl,h}^{0}$	−0.358	0.778	−0.345	100.0	4.7	100.0
	(0.068)	(0.063)	(0.020)			
$\rho_{apl,h}^{1}$	0.829	−0.079	0.856	100.0	10.9	100.0
	(0.023)	(0.004)	(0.036)			
ρ_y	0.351	−0.007	0.355	99.7	4.5	99.4
	(0.066)	(0.000)	(0.005)			

The table reports the mean and standard deviation (in parentheses) of each statistic across 1000 Monte-Carlo experiments. The probabilities of rejecting are based on statistics compared to the 5% critical value of the chi-squared distribution with one degree of freedom. The numbers reported under the heading 'naive' indicate the fraction of the experiments in which Ψ_1 was closer to $\Psi_2^B(\hat\theta^B)$ than it was to $\psi_2^A(\hat\theta^A)$.

Table 2 presents the results of the second experiment in which Model B is the data generating mechanism. The pattern is quite similar to that obtained in the first experiment. In general, tests based on the first-difference filter provide more power, although this is not always the case. Interestingly, the findings in Table 2 directly contradict the claim that the HP filter 'inappropriately characterize[s] [the] phenomenon [of] labor hoarding'. Model B incorporates labor hoarding behavior, but this behavior leads to a distinctive pattern in $\rho_{\mathrm{apl,h}}^{i}$ in both first-differenced and HP-filtered data.

The fact that some of the statistics do not provide any power to reject the false model is not necessarily troubling. The models are not dramatically different in their predictions for some aspects of the data. But, what is clear from the two tables is that an econometrician using a standard method, and the overall evidence, would have little difficulty distinguishing between the two models. Furthermore, the apparent advantage of the first-difference filter in generating power may not be a robust finding. It may have more to do with the specific structure of the data generating mechanisms in my experiments than with something inherent about the filters.

The argument here is not about whether information is lost in the filtering process. It is. In fact, the results indicate that more information about the distinctions between the models is lost when using the HP filter, and the models and moments considered here. However, to reiterate, a researcher following the methodology described in Fig. 6, and using the small set of moments included in my experiments, would have little difficulty identifying the correct model. Had this not been true, the researcher could and should have broadened the set of moments considered in his experiments. An alternative would have been to broaden the set of filters used. Neither of these is obviously preferable, and both, to my mind, are consistent with King and Rebelo's (1993) and Canova's concerns with the properties of the HP filter.

4. Conclusion

To conclude my discussion I refer to the relevant part of Canova's conclusions which I summarize as follows.

1. When data are detrended information is lost. The nature of the information that is lost depends sharply on the filter. In some cases the lost information would have been useful in distinguishing among alternative theories.
2. Using only the HP filter is dangerous.
3. The idea that there exists a single set of stylized facts about business cycles is misleading.
4. Building models which match one set of empirical regularities is a practice than should be reconsidered.

The first of these conclusions is not in dispute. The question is not whether information is lost but whether enough is lost to make the empiricist's job impossible. The analysis in the third section of this paper, I would argue, suggests that there is sufficient information left in the data, and so much structure in the models business cycle researchers use, to quite easily distinguish between competing models.

As for the second point, it seems to me that the exclusive use of any filter is as potentially dangerous as that of any other filter. Any filter has the potential to mask differences between models and data. As it happens, the HP filter does not mask differences between labor hoarding models and their alternatives along the dimensions Canova suggests.

Third, there *is* a single set of stylized facts about business cycles. A statement such as 'output and productivity comove this way over the business cycle' is simply ill-defined if 'business cycle component' refers to the residual from a detrending procedure. The statement has no meaning unless it is made more precise, by specifying how the data have been treated. Once we make more precise statements, we have a unique set of facts. If, on the other hand, by 'business cycles' we mean cycles within the frequency band corresponding to the NBER taxonomy, a fairly large set of robust stylized facts can be constructed.

Fourth, models are compared to a large number of empirical regularities. Most RBC models, for example, are parsimonious, and have many overidentifying restrictions. The fact that there appear to be so many empirical puzzles in the literature casts doubt on the notion that business cycle research is only concerned with fitting a small number of facts. In fact, Canova's paper does a great job of cataloguing these facts.

As for recommendations, the main one Canova gives is for stronger links, where possible, between economic theory and the filtering method. I have argued that this will never be easy and that the main weakness in the empirical methodology he describes is in its lack of statistical formalism. Once a formal statistical methodology is adopted, current theories provide enough strong restrictions to readily distinguish among theories, without too much need to worry about the filter being employed.

Acknowledgements

References

Baxter, M., King, R.G., 1995. Measuring business cycles: approximate band-pass filters for economic time series, NBER Working Paper No. 5022.

Burnside, C., Eichenbaum, M., 1996a. Small sample properties of GMM based Wald tests. Journal of Business and Economic Statistics 14 294–308.

Burnside, C., Eichenbaum, M., 1996b. Factor-hoarding and the propagation of business-cycle shocks. American Economic Review 86 1154–1174.

Burnside, C., Eichenbaum, M., Rebelo, S., 1993. Labor hoarding and the business cycle. Journal of Political Economy 101, 245–273.

Canova, F., 1997. Detrending and business cycle facts. Journal of Monetary Economics 41 (3) (1998) (this issue).

Christiano, L.J., 1988. Why does inventory investment fluctuate so much?. Journal of Monetary Economics 21, 247–280.

Christiano, L.J., Eichenbaum, M.S., 1992. Current real-business-cycle theories and aggregate labor market fluctuations. American Economic Review 82, 430–450.

Cogley, T., Nason, J.M., 1995. Output dynamics in real-business-cycle models. American Economic Review 85, 492–511.

Hansen, L.P., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50, 1029–1054.

Hodrick, R.J., Prescott, E.C., 1997. Postwar U.S. business cycles: an empirical investigation. Journal of Money, Credit and Banking 29, 1–16.

King, R.G., Plosser, C.I., Stock, J.H., Watson, M.W., 1991. Stochastic trends and economic fluctuations. American Economic Review 81, 819–840.

King, R.G., Rebelo, S.T., 1993. Low frequency filtering and real business cycles. Journal of Economic Dynamics and Control 17, 207–231.

McGrattan, E.R., 1994. The macroeconomic effects of distortionary taxation. Journal of Monetary Economics 33, 573–601.

Sargent, T., 1987. Macroeconomic Theory, 2nd ed., Academic Press, San Diego, CA.

Sargent, T., 1989. Two models of measurements and the investment accelerator. Journal of Political Economy 97, 251–287.

Singleton, K.J., 1988. Econometric issues in the analysis of equilibrium business cycle models. Journal of Monetary Economics 21, 361–386.