UNIVARIATE DETRENDING METHOD WITH STOCHASTIC TRENDS

by

Mark W. Watson
Harvard University

Discussion Paper Number 1158
June 1985

Harvard Institute of Economic Research
Harvard University
Cambridge, Massachusetts

I thank Andy Abel and Jim Stock for useful discussions.

Abstract

This paper investigates issues involved in detrending economic time series, when the trend is modelled as a stochastic process. In particular, it considers univariate unobserved component models in which an observed series is additively decomposed into its trend, a random walk with drift, and a residual which follows a stationary stochastic process. Starting from the Wold decomposition for the observed series, three observationally equivalent decompositions are discussed. Optimal detrending methods are derived for the three models. It is shown that the optimal one-sided estimate of the trend is identical for all of the three observationally equivalent models. The paper also discusses the properties of the detrended series, which special attention paid to the use of these series in regression models.

The methods are applied to three macroeconomic time series, GNP, Disposable Income, and Consumption Expenditures for Non-Durables. We find that by parameterising the stochastic processes generating these series in terms of the unobserved components model, sensible estimates of the trend are produced, and long run forecasts superior to those obtained by ARIMA models can be constructed. The detrended disposable income data together wiht the consumption data are used to test one variant of the life cycle hypothesis of consumption.

Key Words

Detrending, Unobserved Component Model, Signal Extraction, Life Cycle Model

1 Introduction

Most macroeconomic time series exhibit a clear tendency to grow over time and can be characterized as "trending." The statistical theory underlying most modern time series analysis relies on the assumption of covariance stationarity, an assumption that is clearly violated by most macroeconomic time series. In applied econometric work it is usually assumed that this statistical theory can be applied to deviations of the observed time series from their trend value. Since it is often the case that these deviations or economic fluctuations are of primary interest, modern time series techniques are often applied to "detrended" economic time series.

Much recent work has been devoted to issues involving trends or long-run components in economic series. Some of this work has devoted itself to the proper characterization of "trends" in economic data. A notable contribution on this topic is the paper by Plosser and Nelson (1981) which considers the use of deterministic and stochastic trends. Other work (e.g. Nelson and Kang (1981), and (1984)) has been concerned with the econometric consequences of misspecification in the model for the trend component. Still more work has addressed the issue of detecting long-run relationships (e.g. Geweke (1983)), or incorporating long run relationships in short run dynamic relationships (e.g. the work on "error correction models" begun in Davidson, et. al. (1978) and the co-integrated processes introduced in Granger(1983) and discussed further in Engle and Granger (1984).)

Research concerning the proper characterization of long run trend behavior in economic time series is important for a variety of reasons. First, the work of Nelson and Kang and others shows that misspecification in the model for the trend can seriously affect the estimated dynamics in an econometric model. Proper estimation of these dynamic relationships is important if they are to be used to test modern theories of economic behavior. These theories put very tight restrictions on the dynamic interrelationships between economic variables. Misspecification of trend components will often lead the analyst to incorrect inferences concerning the validity of these theories. The proper specification of long-run relationships is also critical for long-term forecasting.

This paper proposes a method for removing stochastic trends from economic time series. The method is similar to one originally proposed by Beveridge and Nelson (1981) for ARIMA models, but differs in two important respects. First, the problem is cast in an unobserved components framework, in which the trend is modelled as a random walk with drift, and the remaining "cyclical component " of the series is modelled as a general covariance stationary process. This framework allows us to discuss and construct optimal detrending methods. One of the optimal detrending methods corresponds to the (negative of the) Beveridge and Nelson transitory component. Our empirical results also demonstrate that ARIMA and Unobserved Components formulations which are equivalent in principle, are dramatically different in practice.

The paper is organized as follows. The next section specifies the model for the observed series, the trend component, and the residual "cyclical" component. We begin with a general ARIMA formulation for the observed series and describe how the stochastic process can be "factored" into the two processes for the components. Identification issues are also addressed in this section. Section 3 discusses methods for estimating and eliminating the stochastic trends. The properties of the detrended series are addressed, and special attention is paid to the use of these series in constructing econometric models. The final two sections contain empirical examples. Three postwar US macroeconomic time series, GNP, disposable income, and consumption of nondurables, are analysed. Section 4 looks at each of these series in isolation, and presents and compares various estimates of the detrended series. Section 5 uses the data for nondurable consumption and disposable income to test the "consumption follows a random-walk" implication of one version of the life cycle consumption model. The final section contains a short summary and some concluding remarks.

2 The Model

This section will introduce three models that additively decompose an observed time series, x_t, into a trend and cyclical components. Each of these models will assume, or imply, that the change in x_t is a covariance stationary process. We begin with the Wold representation for Δx_t, which we write as

$$(2.1)\quad \Delta x_t = \delta + \theta^x(B)\, e_t^x$$

where Δ is the first difference operator, $\theta^x(B)$ is a polynomial in the backshift operator B, and e_t^x is white noise. The assumption that x_t is stationary is appropriate for most non-seasonal macro-economic time series. Seasonal series often require a differencing operator of the form $(1-B^s)$ where s is the seasonal span (12 for monthly, 4 for quarterly etc.). Our assumption is not adequate to handle these series, and therefore, we are assuming that the series are either non-seasonal or have been seasonally adjusted.

The representation for the level of the series x_t that we consider in this paper is

$$(2.2) \quad x_t = \tau_t + c_t$$

where

$$(2.3) \quad \tau_t = \delta + \tau_{t-1} + e_t^{\tau} \qquad var(e_t^{\tau}) = \sigma^2_{e\tau}$$

and (c_t, e_t^{τ}) is a jointly stationary process. A variety of specific assumptions about this process will be given below. The component τ_t corresponds to the trend component in the variable x_t; the "detrending" will attempt to eliminate this component. Equation (2.3) represents this trend as a random walk with drift, which can be viewed as a flexible linear trend. The linear trend is a special case of (2.3) and corresponds to the restriction $\sigma^2_{e\tau} = 0$. The forecast function of (2.3) is linear with a constant slope of δ and a level that varies with the realization of e_t^{τ}. More general formulations are certainly possible. Harvey and Todd (1983) consider a model in which the drift term, δ_t, evolves as a random walk. This allows the slope as well as the level of the forecast function of τ_t to vary through time. This formulation implies that x_t must be differenced twice to induce stationarity, and therefore is ruled out by our assumption that Δx_t is stationary.

To complete the specification of the model we must list our assumptions concerning the covariance properties of c_t and the cross-covariances between c_t and e_t^{τ}. We consider three sets of assumptions. The assumptions will differ in the way that c_t and e_{t-k}^{τ} are correlated. We consider each model in turn:

Model 1

The model for c_t is

$$(2.4.1) \quad c_t = \theta^{\tau}(B)e_t^{\tau}$$

where $\theta^{\tau}(B)$ is a one-sided polynomial in the backshift operator. In this model, the innovations in the trend and cyclical components are perfectly correlated. There is a one-to-one correspondence between models of the form (2.1) and models characterized by (2.2), (2.3), and (2.4.1). This implies that the parameters of the unobserved components formulation (2.3) and (2.4.1) are econometrically identifiable. To show this, we show the unique correspondence between the coefficients in the two representations. Equating the two representations for Δx_t we see that

$$\theta^{x}(B)e_t^{x} = [1-(1-B)\theta^{\tau}(B)]e_t^{\tau}$$

so that

$$|\theta^{x}(1)|^2 \sigma^2_{ex} = \sigma^2_{e\tau}$$

The relationship between the remaining coefficients follows directly from equating the coefficients in B in the polynomials

$$[(1-\theta_0^{\tau}) - (\theta_1^{\tau} - \theta_0^{\tau})B - (\theta_2^{\tau} - \theta_1^{\tau})B^2 \cdots]\sigma_{e\tau} = [1-\theta_1^{x}B - \theta_2^{x}B^2 \cdots]\sigma_{ex}$$

The perfect correlation between the innovations in the components is an assumption that some might find objectionable on *a priori* grounds. A model in which the components are completely uncorrelated is given in the next model.

Model 2

In this model the component c_t evolves according to

(2.4.2) $c_t = \theta^{c}(B)e_t^{c}$

Where e_t^{c} and e_{t-k}^{x} are uncorrelated for all k. Like Model 1 the formulation in (2.2), (2.3) and (2.4.2) is econometrically identifiable. To see this, we once again equate the representations for Δx_t corresponding to

(2.1) and (2.2), (2.3), and (2.4.2). This implies

$$(2.5)\ |\theta^{x}(1)|^2\ \sigma^2_{ex} = \sigma^2_{e\tau}$$

The coefficients in $\theta^c(B)$ can be found by forming the factorization of

$$(2.6)\ \theta^{x}(z)\theta^{x}(z^{-1})\sigma^2_{ex} - \sigma^2_{e\tau} = (1-z)(1-z^{-1})\theta^{c}(z)\theta^{c}(z^{-1})\sigma^2_{ec}$$

subject to the usual identifying normalizations (e.g. $\theta_0{}^c = 1$ and the roots of $\theta^c(z)$ are on or outside the unit circle).

Equation (2.6) can be used to show that (2.2)-(2.4.2) place testable restrictions on the Δx_t process given in (2.1). To see this, set $z=e^{-i\omega}$, so that the right-hand side of (2.6) is the spectrum of $(1-B)c_t$. Since the spectrum is non-negative, the left-hand side of (2.6) must be non-negative for all ω. This implies that $\theta^{x}(e^{-i\omega})\theta^{x}(e^{i\omega})\sigma^2_{ex} \geq \sigma^2_{e\tau}$ for all ω, with equality guaranteed at $\omega=0$ by equation (2.5). We then conclude that the spectrum of x_t, $\theta^{x}(e^{-i\omega})\theta^{x}(e^{i\omega})\sigma^2_{ex}$, has a global minimum at $\omega=0$. Only processes with this feature can be represented by Model 2. (This rules out models like an ARIMA(1,1,0) with positive autoregressive coefficient.) The restrictiveness of this assumption will be investigated empirically for three macroeconomic time series in section 4.

Model 3

Model 3 combines the features of Models 1 and 2. The component c_t is represented as

$$(2.4.3)\ c_t = \phi^{c}(B)e_t{}^{c} + \phi^{\tau}(B)e_t{}^{\tau}$$

where $\phi^c(B)$ and $\phi^\tau(B)$ are one-sided polynomials in B. This model can be viewed as a mixture of Models 1 and 2. Since both of those models are individually indentifiable, Model 3 is not. The model can be identified by making some *ad hoc* normalizing assumptions. One normalization would minimize the contribution of $e_t{}^\tau$ to the variance of c_t, and make the model as close to Model 2 as possible. This normalization would produce Model 2 if the spectrum of the change in x_t satisfied the constraint given in equation (2.5). Other normalizations are possible.

In the next section we will discuss "detrending" in the context of the models given above, and discuss the use of detrended data in the construction of econometric models.

3 Estimation Issues

The models presented in the last section suggest that "detrending" should be viewed as a method for estimating and removing the component τ_t from the observed series x_t. If we denote the estimated trend by $\hat{\tau}$, then the detrended series is given by $\hat{c}_t = x_t - \hat{\tau}_t$. Different detrending methods correspond to different methods for estimating τ. A variety of criteria can be used to choose among competing estimation methods. We will consider linear minimum mean square error (lmse) estimators constructed using information sets $\mathcal{I}_h = (x_0, x_1, \cdots, x_h)$. We concentrate on these estimators for a variety of reasons. In addition to the usual reasons, including ease of computation and optimality for quadratic loss, the use of lmse estimators guarantee certain orthogonality properties involving the estimation errors $\tau_t - \hat{\tau}_t = \hat{c}_t - c_t$. These properties play a key role in the formation of instrumental variable estimators that are discussed below. We concentrate on a univariate information set for computational ease. In general, multivariate methods will produce more accurate estimates. The univariate methods considered in this paper can serve as a benchmark to measure the marginal gains from considering more general, multivariate models.

We will discuss detrending in the context of the general model — Model 3. The results for Model 1 can be found by setting $\phi^c(B) = 0$, and $\theta^\tau(B) = \phi^\tau(B)$, and the results for Model 2 can be found by $\phi^\tau(B) = 0$, and $\theta^c(B) = \phi^c(B)$. A convenient starting point for the discussion is the lmse of τ using the information set $(\cdots, x_{-1}, x_0, x_1, \cdots)$. In this case the standard Wiener filter for stationary processes (see e.g. Whittle [1963]) can be extended to this non-stationary case (see Bell[1984]) to yield

$$(3.1)\quad \hat{\tau}_t = V(B)x_t = \Sigma\, v_i x_{t-i}$$

where the coefficients in the two-sided polynomial V(B) can be found from

$$(3.2)\quad V(z) = \sigma^2_{e\tau}\,[1+(1-z^{-1})\phi^{\tau}(z^{-1})]\,[\theta^{x}(z)\theta^{x}(z^{-1})\sigma^2_{ex}]^{-1}$$

We denote $E(w_t \mid \mathcal{I}_h)$ by $w_{t/h}$ for any variable w (where E is used to denote the projection operator). From equation (3.1)

$$(3.3)\quad \tau_{t/h} = E(\hat{\tau}_t \mid \mathcal{I}_h) = \Sigma v_i E(x_{t-i} \mid \mathcal{I}_h)$$

so that the estimates of the trend using the information set $\mathcal{I}_h$ can be formed from the Wiener filter with unknown values of x replaced by forecasts or backcasts constructed from the set $\mathcal{I}_h$.

The form of the filter V(B) given in (3.2) makes it clear that the lmse estimate depends on $\phi^{\tau}(z)$. Different V(B) polynomials will be associated with Models 1, 2, and 3, so that different lmse estimates of c_t will be constructed. Equation (3.2) makes it clear, however, that the difference arises from the way in which future data is used in the construction of $\tau_{t/h}$. All models produce the identical values of $\tau_{t/h}$ for $h \le t$. This is easily demonstrated. From (2.2)

$$x_{t/h} = \tau_{t/h} + c_{t/h}$$

For $h \le t$

$$\tau_{t+k/h} = \tau_{t/h} + k\delta \qquad \text{for } k=1,2, \cdots$$

so that

$$x_{t+k/h} - k\delta = \tau_{t/h} + c_{t+k/h}$$

Since all of the models imply that c_t is stationary (with mean zero),

$$\lim\,(k \to \infty)\; c_{t+k/h} = 0,$$

so that

$$\lim\,(k \to \infty)\,[x_{t+k/h} - k\delta\,] = \tau_{t/h}.$$

This result shows that, in principle, the estimates $c_{t/t}$ (which will be called the filtered estimates) can be formed without access to any specialized software. To calculate the filtered estimates one merely constructs an ARIMA model for x_t and then forecasts the series (less the deterministic increases $k\delta$) into the distant future. This forecast corresponds to the filtered estimate, $\tau_{t/t}$. This estimate of a permanent component was first suggested by Burridge and Nelson(1981) in their permanent/transitory decomposition of economic time series. They define their transitory component as $\tau_{t/t} - x_t = -c_{t/t}$, in the notation above. This discussion shows that their estimate of the permanent component corresponds to an optimal one-sided estimator for the trend in the models under consideration in this paper.

While the optimal filtered estimate $c_{t/t}$ is identified — is not model dependent — its precision is not identified. That is, the mean square of $(c_t - c_{t/t})$ will depend on the assumed model. If Model 1 is used to describe the decomposition of the data then $c_{t/t} = c_t$, so that the mean square error is zero. For the other models, x_t is made up of both noises e^{τ} and e^{c} so that it is impossible to perfectly disentangle τ_t and c_t when only their sum, x_t, is observed. Since the Models 1,2, and 3 are observationally equivalent the mean square of $(c_t - c_{t/t})$ is not identified.

The remainder of this section will focus on the use of estimated values of c_t in linear regression models. Replacing c_t by $c_{t/h}$ in regression models leads to problems similar to those in the classic errors-in-variables model. As the examples below will demonstrate, OLS estimates of regression coefficients will quite often be inconsistent.

We will begin our discussion by writing the orthogonal decomposition of c_t as

$$c_t = c_{t/h} + a_{t/h}.$$

where $a_{t/h}$ is the signal extraction error that arises from the use of information set $\mathcal{I}_h$. Ordinary least square regression estimates rely on sample covariances between observable variables, and the consistency of OLS estimates follow from the consistency of these sample covariances. Consider then, the covariance between an arbitrary variable w and c. From the decomposition of c, we have

$$\mathrm{cov}(w_t c_t) = \mathrm{cov}(w_t c_{t/h}) + \mathrm{cov}(w_t a_{t/h})$$

The $\mathrm{cov}(w_t c_t)$ will be consistently estimated by the sample covariance between w_t and $c_{t/h}$ if $\mathrm{cov}(w_t a_{t/h})=0$. In general this will not be true, so that $\mathrm{cov}(w_t c_t) \neq \mathrm{cov}(w_t c_{t/h})$. Recall, however, that $a_{t/h}$ is a projection error, so that it is uncorrelated with linear combinations of data in $\mathcal{I}_h$. By constructing w_t as a linear combination of the elements in $\mathcal{I}_h$, we can guarantee that $\mathrm{cov}(w_t c_t) = \mathrm{cov}(w_t c_{t/h})$. We will use this fact in construction of instrumental variable estimators. It will be convenient to discuss a variety of estimation issues in the context of some specific models.

The first model has current and lagged values of c_t as independent variables, so that

$$y_t = \Sigma\, c_{t-i}\, \beta_i + z_t'\gamma + \xi_t$$

where z_t is a j vector of observable variables, and we will assume (without loss of generality) that ξ_t is white noise. We can rewrite this model in terms of observables as

$$\begin{aligned} y_t &= \Sigma_{i=1}\, c_{t-i/h}\, \beta_i + z_t'\gamma + \xi_t + \Sigma_{i=1}\, a_{t-i/h}\, \beta_i \\ &= \Sigma_{i=1}\, c_{t-i/h}\, \beta_i + z_t'\gamma + u_t \end{aligned}$$

where u_t is the composite error term $\xi_t + \Sigma_{i=1}\, a_{t-i/h}\, \beta_i$. The unknown parameters $\beta_1, \beta_2, \cdots, \beta_k, \gamma_1, \cdots, \gamma_j$ will be consistently estimated by OLS if the regressors $c_{t-i/h}$ and z_t are uncorrelated with the error term u_t. This may not be true for a variety of reasons.

Two sources of correlation are immediately apparent. First, consider the correlation between $c_{t-i/h}$ and ξ_t. In many models it will be reasonable to assume that ξ_t is uncorrelated with current and lagged values of c_t.

but unreasonable to assume that ξ_t is uncorrelated with future values of c_t. (Correlation between ξ_t and c_t will exist if there is feedback from y_t to c_t.). But when $h>t$, $c_{t-i/h}$ will contain future x_t's, and therefore $c_{t-i/h}$ will contain linear combinations of future c_t's. This may induce a correlation between $c_{t-i/h}$ and ξ_t even though c_{t-i} and ξ_t are uncorrelated. The second source of correlation between the regressors and the disturbance arises from the possible correlation between z_t and $a_{t-i/h}$. These variables will be correlated when z_t contains useful information about c_t not contained in $\mathcal{I}_h$.

These problems can be circumvented by the use of instrumental variables. In particular, the variables in $\mathcal{I}_t$ can be used as instruments to estimate the model. When constructing instrumental variable estimates, it is useful to make use of the fact that, for $h>t$, $c_{t/t} = E(c_{t/h} \mid \mathcal{I}_t)$, so that IV estimates can be formed by regressing y_t on the filtered values, $c_{t/t}$, and $\hat{z}_t$, the fitted values from the regression of z_t onto the set of instruments. Finally, it should be pointed out that the error terms, u_t, may be serially correlated so that standard errors may have to be calculated using the procedures outlined in Cumby, Huizinga, and Obstfeld (1983), or Hayashi and Sims (1983).

The second regression model that we consider is

$$c_t = \Sigma_{i=1}\, c_{t-i}\,\beta_i + z_t'\gamma + \xi_t$$

This model should be interpreted as the true generating equation for c_t, so that ξ_t is uncorrelated with all variables dated t-1 or earlier. Models 1 - 3 described in the last section can be viewed as reduced forms of this model, where the z_t's have been solved out as in Zellner and Palm (1976). Writing the model in terms of observables we have

$$\begin{aligned} c_{t/h} &= \Sigma_{i=1}\, c_{t-i/h}\,\beta_i + z_t'\gamma + \xi_t - a_{t/h} + \Sigma_{i=1}\, a_{t-i/h}\,\beta_i \\ &= \Sigma_{i=1}\, c_{t-i/h}\,\beta_i + z_t'\gamma + u_t \end{aligned}$$

As in the previous model the observed regressors $c_{t-i/h}$ and z_t may be correlated with the error term, u_t, leading to inconsisitency of the OLS estimates. When $h\geq t$, the estimates $c_{t-i/h}$ contain future c_t's and therefore will be correlated with ξ_t. The variables z_t can be viewed as "causes" of c_t and will therefore contain useful information about the c's, which is not contained in the univariate information set $\mathcal{I}_h$. This will induce a correlation between z_t and $a_{t-i/h}$. Instrumental variables can again be used to estimate the model. The data in $\mathcal{I}_{t-1}$ are valid instruments.

In the discussion above we replaced the true values of c_{t-i} by the estimates $c_{t-i/h}$. Since these estimates depend crucially on the model for c_t, a useful alternative is to replace them by the estimates $c_{t-i/t-i}$, which do not depend on the model asumed for c_t. To see the implications of this procedure, rewrite the first regression example as

$$y_t = \Sigma_{i=1}\, c_{t-i/t-i}\, \beta_i + z_t\gamma + \xi_t + \Sigma_{i=1}\, a_{t-i/t-i}\, \beta_i$$

This differs from the formulation above in that each c_{t-i} is estimated using a different information set. Since $\text{cov}(c_{t-i/t-i} a_{t-j/t-j}) \neq 0$ for $i>j$, care must be taken in choosing instruments. Since $\mathcal{I}_t \subset \mathcal{I}_{t-1} \subset \cdots \subset \mathcal{I}_{t-k}$, data in $\mathcal{I}_{t-k}$ are valid instruments, and this set can be used.

Finally, it is important to keep in mind that the inconsistency in OLS estimates will depend on the magnitude of the error in the estimate of c_t. When c_t is estimated very precisely, the inconsistencies from OLS have no practical importance.

4. Univariate Examples

In this section we will analyse three U.S. macroeconomic time series -- real GNP, real disposable income, and real consumption of non-durables. We begin by applying the univariate detrending methods outlined in the last section to the logs of these series, using quarterly data from 1949 through 1984. The estimated univariate models and corresponding trend and cyclical components will be discussed in this section. In the next section we will investigate the relationship between the cyclical components of disposable income and consumption using regression methods.

Two univariate models have been estimated for each series. The first is the usual ARIMA model. The second is an unobserved components (UC) model suggested by the independent trend/cycle decomposition in Model 2. For each time series, we will present and compare the models and their corresponding trend/cycle decomposition. The analysis in the last section indicated that, given the Wold decompositon of the observed series, the optimal one-sided estimates of the components could be formed using any of the observationally equivalent representations of the data. In principle then, it shouldn't matter whether we form the one-sided estimates from the ARIMA model or from the UC model. The results below show that, in practise, it matters a great deal which representation is used. This apparent contradiction arises from the fact the Wold representation for the data is not known. The ARIMA model and the UC model correspond to different parsimonious versions of the Wold representations. Given a finite amount of data it is very difficult to discriminate between these alternative representations for the data sets that we consider.

GNP

The estimated autocorrelations for log GNP suggested that the data were non-stationary (the first estimated autocorrelation was .98). The correlogram for the change in the series suggested that an ARIMA(1,1,0) model was appropriate. The estimated model was:

$$\Delta x_t = \underset{(.001)}{.005} + \underset{(.077)}{.406}\,\Delta x_{t-1}$$

$$SE = .0103 \quad L(1) = .73 \quad L(3) = 6.3 \quad Q(23) = 14.9 \quad \mathcal{L} = 292.07$$

SE is the estimated standard error, L(1) and L(3) are LM statistics for serial correlation in the error term of order 1 and 3 respectively, and Q(df) is the Box-Pierce statistic of the residuals. Under the null hypotheseis of no serial correlation, the LM test statistics are distributed as χ^2_1 and χ^2_3 respectively. The final statistic reported, $\mathcal{L}$, is the value of the log likelihood function.

Interestingly, this estimated model suggests that the spectrum for Δy has a global maximum at the zero frequency, so that decomposition of Δy into an independent random walk and stationary component is not possible. (Recall that this decomposition required that the spectrum had a global minimum at the zero freqency.) This means that the trend/cyclcial decompostion in Model 2 is inappropriate. Neverless, we estimated a model of the form given into Model 2. The results were:

$$x_t = \tau_t + c_t$$

$$\Delta\tau_t = \underset{(.001)}{.008} \qquad (\sigma = .0057)$$

$$c_t = \underset{(.121)}{1.501}\, c_{t-1} - \underset{(.125)}{.577}\, c_{t-2} \qquad (\sigma = .0076)$$

$$SE = .0099 \qquad Q(17) = 10.4 \qquad \mathcal{L} = 294.42$$

The values for σ next to each equation refer to the standard deviation of the disturbance in that equation. The value for SE is the standard error of the innovation in x_t, i.e. the one-step ahead forecast standard error. It is comparable to the SE reported for the ARIMA model.

The unobserved components (UC) model performs slightly better than the ARIMA model in terms of (within sample) one-step ahead forecasting, or equivalently, in terms of the value of the likelihood function. Indeed, both models imply very similar behavior for the short-run behavior of the series. To see this, notice that the UC model implies that

$$(4.1)\quad (1-1.501B - .577B^2)(1-B)x_t = (1-1.501B - .577B^2)(1-B)\, e_t^{\tau} + (1-B)\, e_t^{c}$$

By Granger's Lemma the rhs of (4.1) can be represented as a MA(2), and solving for the implied coefficients yields

$$(1-1.501B - .577B^2)\Delta x_t = (1-1.144B + .189B^2)a_t \qquad (\sigma_a = .0099)$$

The autoregressive representation for the model is

$$(1-1.144B + .189B^2)^{-1}(1-1.501B - .577B^2)\Delta x_t = a_t$$

Carrying out the polynomial division we have, approximately

$$(1-.36B - .05B^2)\,\Delta x_t = a_t$$

which is very close to the estimated ARIMA model.

While the short-run behavior of the UC and ARIMA models are very similar, their long-run properties are quite different. This shows itself in a variety of ways. The spectra implied by the models are quite different at the low frequencies, the models produce markedly different long-run forecasts and give quite different one-sided estimates of the trend and cyclical components. Figure 1 compares the two implied spectra. The ARIMA model implies a spectra with a maximum at the zero frequency, while the UC model implies a model with a minimum at that frequency. Unfortunately, with only 142 data points there are very few periodogram points corresponding to the low frequencies, so that direct frequency domain estimation methods help little in discriminating between the models.

The models can be compared on their ability to forecast over various horizons. In Table 1 we have tabulated the (in-sample) root mean square error (rmse) for both models over various forecast horizons. In addition we show the relative efficiency of the methods. Whereas the models are essentially identical for forecasting one-quarter ahead, they differ markedly in their abilities to forecast over longer horizons. The efficiency of the ARIMA model relative to the UC model falls to 92% at 4 quarters, to 81% at 20 quarters, and to 68% at 40 quarters. This difference in long-run forecasts also implies very different estimates for the one-sided estimates of the trend and cyclical components. In Fgures 2a and 2b we present the actual series and the optimal one-sided estimate of the trend using the two models. The trend in the ARIMA model is very close to the actual series, whereas the trend in the UC model smooths the series considerably. In Figure 2d we compare the (one-sided) estimates of the cyclical component. Here again the differences are substantial. The estimates constructed from the ARIMA model are difficult to interpret, while the estimate corresponding to the UC model corresponds closely to conventional chronologies of postwar cyclical behavior. This correspondence can be seen in figure 2e where we have plotted the UC one-sided estimates, and shade the peak to trough business cycle periods as calculated by the NBER

As we pointed out in the last section, conditional on the estimated model for x_t, the one-sided lmse estimates of c_t are unique, i.e. do not depend on the choice of Model 1, 2, or 3. In addition, if we accept Model 1 as the appropriate decomposition, then conditional on the parameters of the ARIMA model, the one-sided estimates are exact, i.e. they have a root mean square error of zero.If however, we assume that Model 2 or 3 is the correct representation for the cyclical component, then more precise estimates can be constructed. These estimates use future as well as past data to improve the one-sided lmse estimates. Using the estimated UC model, which assumes that Model 2 is the correct representation for c_t, we have constructed the optimal two-sided estimates. The estimate of the trend is plotted in Figure 2c, and Figure 2e compares the optimal one-sided and two-sided estimates of the trend component.

Conditional on the parameter estimates, it is also possible to obtain estimates of the precision of the estimates $c_{t/t}$ and $c_{t/T}$ (the two-sided estimate). Using the estimates from the UC model, the root mean square of $(c_t - c_{t/t})$ = .020. The root mean square of $(c_t - c_{t/T})$ depends on the amount of future data available. For t near the middle of the sample the root mean square of $(c_t - c_{t/T})$ = .017 Both estimates are reasonably precise. Using the orthogonal decomposition of $c_t = c_{t/h} + a_{t/h}$, we see that $var(c_t) = var(c_{t/h}) + var(a_{t/h})$, so that a unit free measure of precision is $R^2(h) = var(c_{t/h})/(var(c_t)$, which shows the proportion of the variance of c_t explained by $c_{t/h}$. For this model $R^2(t)$ = .54 and $R^2(T)$ = .71 (for data near the center of the sample).

Disposable Income

The estimated ARIMA model for disposable income was

$$\underset{}{\Delta x_t} = \underset{(.001)}{.011} - \underset{(.080)}{.210}\, \Delta x_{t-4}$$

SE = .010, L(1) = .01 , L(3) = 1.7 , Q(23) = 22.9 , $\mathcal{L}$ = 297.1

The corresponding UC model was

$$x_t = \tau_t + c_t$$

$$\Delta\tau_t = \underset{(.001)}{.009} \qquad (\sigma = .0057)$$

$$c_t = \underset{(.073)}{1.029}\, c_{t-1} - \underset{(.094)}{.024}\, c_{t-2} + \underset{(.084)}{.051}\, c_{t-3} - \underset{(.058)}{.152}\, c_{t-4} + \underset{(.017)}{.055}\, c_{t-5} \qquad (\sigma = .0076)$$

SE = .009 $\qquad$ Q(14) = 10.4 $\qquad$ $\mathcal{L}$ = 299.6

We can compare the two models for disposable income using the same procedures discussed in the comparison of the models for GNP. First, the UC model produces a lower value of the one-step ahead mean square forecast error, SE. Longer forecast horizons are compared in the second panel of Table 1. Here, the performance of the two models is very similar for 4 to 20 quarters, but markedly different at 40 quarters, where the relative efficiency of the ARIMA forecast falls to 70%. This difference in long-run forecasts also leads to diferent one-sided trend/cycle decompositions for the two models. These are compared in Figure 3a and 3b.

In Figure 3a we compare the optimal one-sided and two-sided estimates of the cycle, where the two-sided estimates are calculated assuming that the UC model is correct. The plots are similar for data after 1954, but differ from 1951 to 1954. The one-sided estimates are rather volatile during this early period, reflecting the small number of data points used in their contruction. For one-sided estimates constructed with a moderately large amount of data the rms of $(c_t - c_{t/t})$ = .019. The rms of $(c_t - c_{t/T})$ is only slightly smaller.[1] The corresponding $R^2(t)$ is .68.

Nondurable Consumption

The ARIMA model for nondurable consumption is a random walk. The estimated model and associated statistics are:

$$Dx_t = \underset{(.0007)}{.0065}$$

SE = .0086, $\qquad$ L(4) = 1.04, $\qquad$ Q(24) = 21.7, $\qquad$ $\mathcal{L}$ = 314.0

FIGURE 3

Trend/Cycle Decomposition for the Log of Disposable Income

Figure 3a

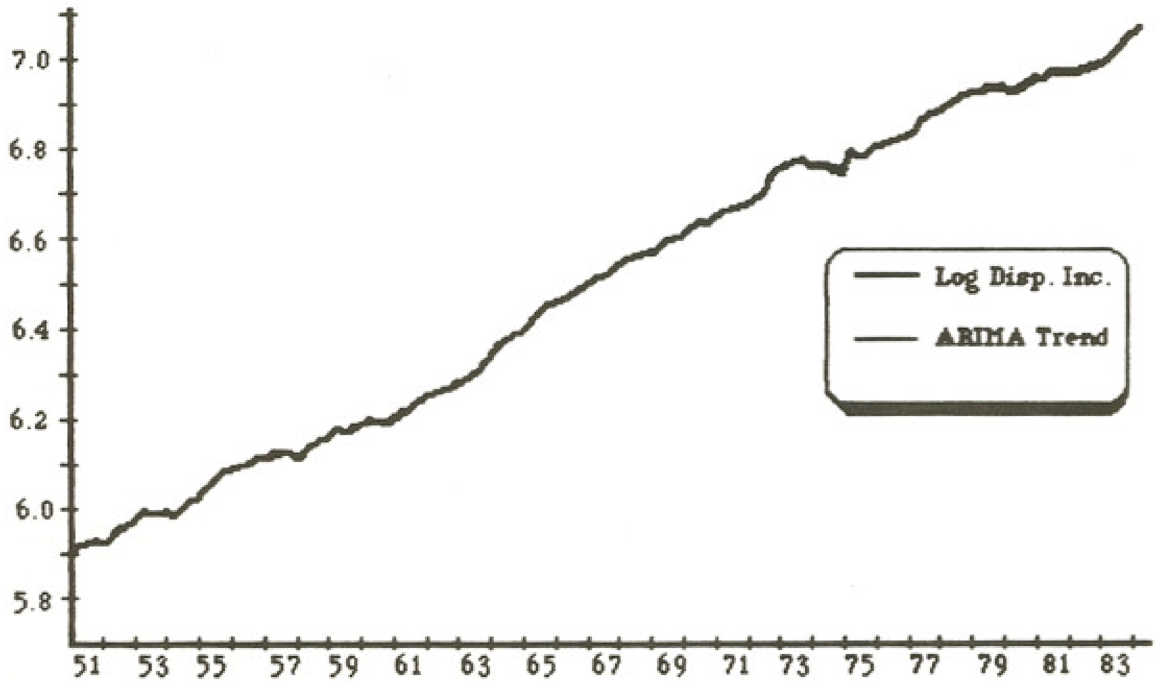


Figure 3b

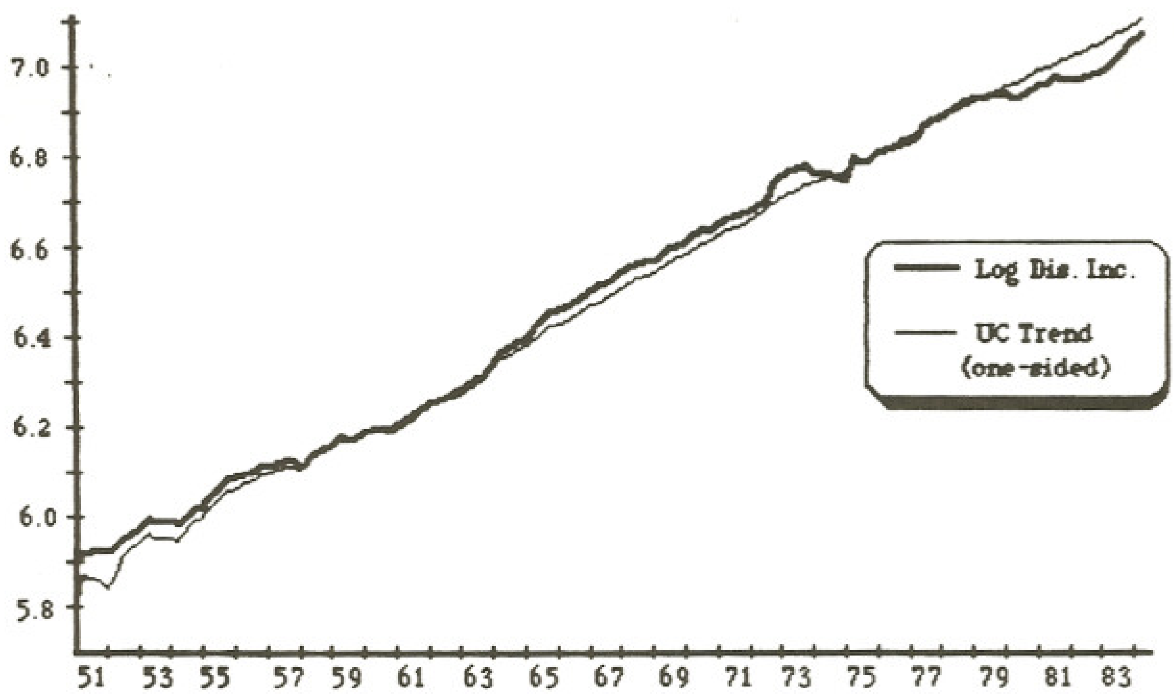


FIGURE 3

Trend/Cycle Decomposition for the Log of Disposable Income

Figure 3c

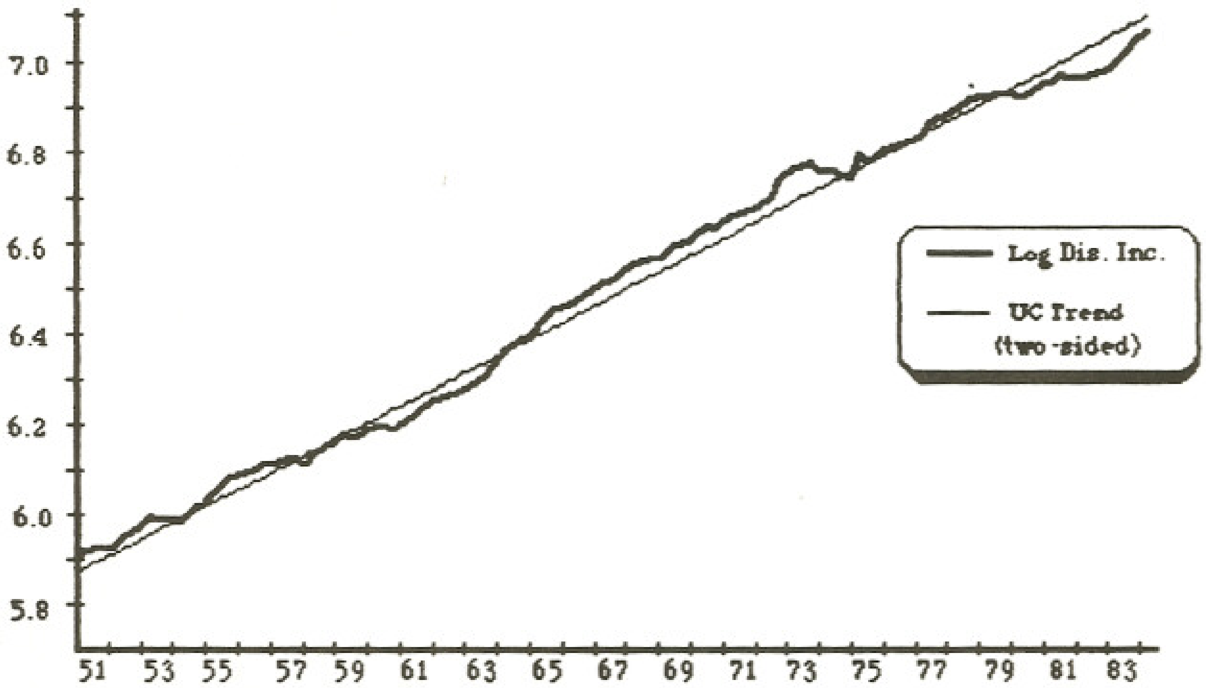


Figure 3d

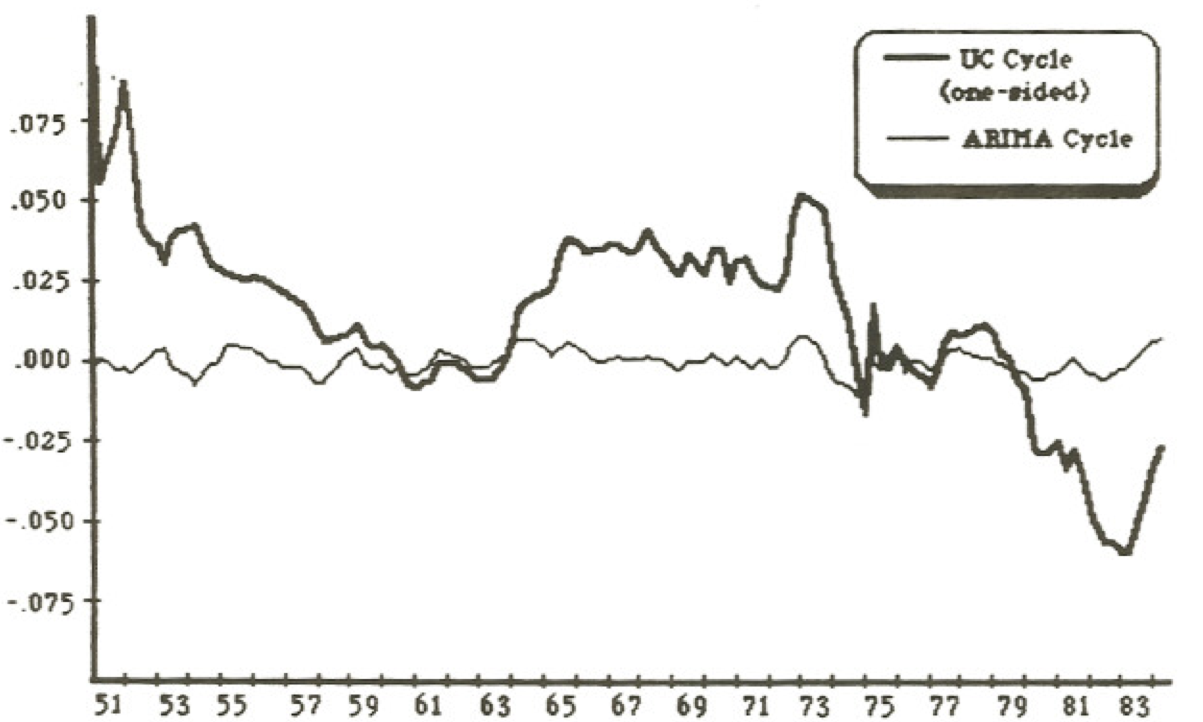
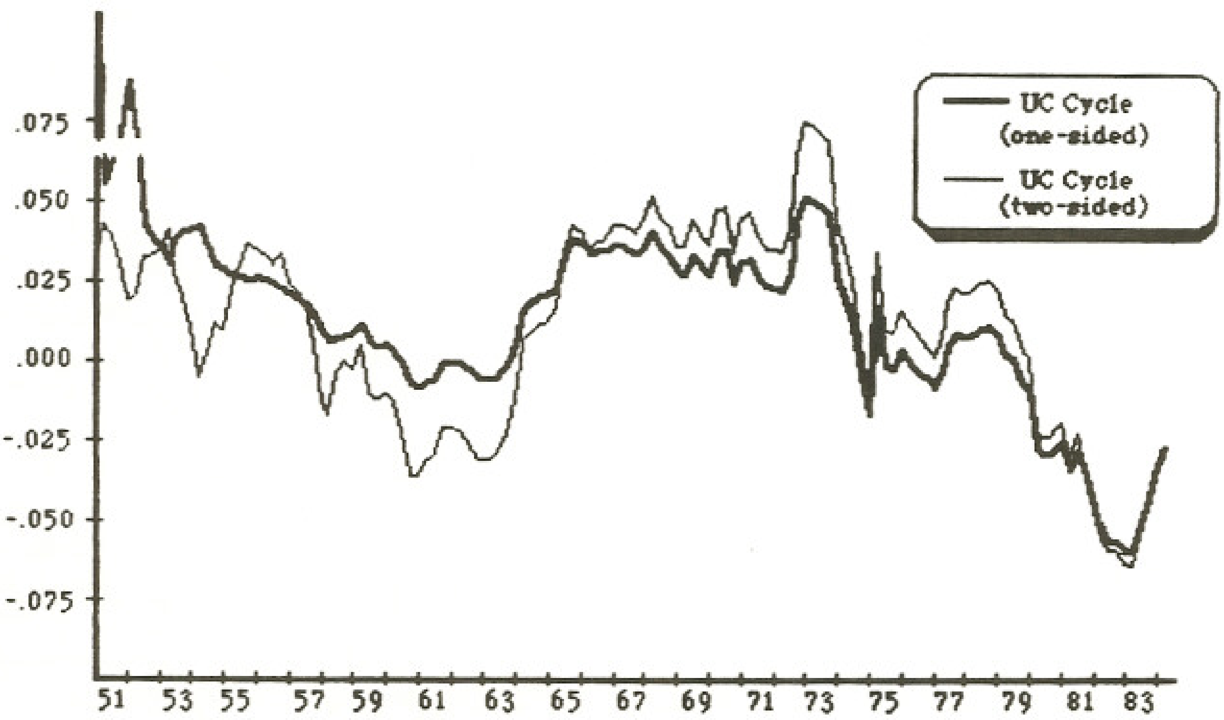


FIGURE 3

Trend/Cycle Decomposition for the Log of Disposable Income

Figure 3e



The estimated UC model was

$$x_t = \tau_t + c_t$$

$$\underset{(.0003)}{\Delta\tau_t = .0067} \qquad (\sigma = .0018)$$

$$c_t = \underset{(.036)}{.940}\, c_{t-1} \qquad (\sigma = .0082)$$

$$SE = .0085, \qquad Q(20) = 14.9, \qquad \mathcal{L} = 314.3$$

The models are clearly very close to one another in terms of their one-step-ahead forecast ability, and likelihood values. If we set the variance of the cyclical component to 0, the UC model implies that x_t is a random walk, so that the random walk model is nested within the UC model. It is therefore possible, in principle, to test the competing models, using a likelihood ratio test. Unfortunately the test is complicated by the fact that the AR coefficient in the model for c_t is not identified under the random walk hypothesis, but it is identified in the more general model. This complicates the distribution of the likelihood ratio statistic; it will not have the usual asymptotic distribution. This problem has been discussed in detail in Watson and Engle (1985) and Davies (1977). They show that the correct (aymptotic) critical value for carrying out the test (using the square root of the likelihood ratio statistic) is bounded below by the critical value for the standard normal distribution. In this example, the square root of the likelihood ratio statistic is .77 which implies a lower bound of the (asymptotic) prob-value of .27. This suggests that the random walk hypothesis cannot be rejected at levels of 27% or less.

The random walk and UC models can also be compared on their longer run forecast abilities. These are shown in the bottom pannel of Table 1. Here again, the estimated UC model dominates. The relative efficiency of the random walk model falls to 95% at 4 quarters, to 81% at 20 quarters, and to 65% at 40 quarters.

The examples in this section tell a consistent story. The short-run forecasting performance of the ARIMA and UC models are very similar. At longer forecast horizons, the performance of the models differ markedly. This difference in the long-run properties of the estimated models, leads to very different estimates of the underlying trend component and cyclical components.

5. Regression Examples

In this section we investigate the relationship between non-durable consumption expenditures and the cyclical component of disposable income. We test the proposition that the change in consumption from period t to t+1 is uncorrelated with the cyclical component of disposable income dated t or earlier. The empirical validity of this proposition, first tested in Hall (1978), has been the subject of ongoing controversy. (See for example, the papers by Flavin (1981), and Mankiw and Shapiro (1984).). We begin by motivating the empirical specification that is used in this paper.

Assume that a consumer is choosing consumption to maximize a time separable utility function, subject to an intertemporal budget constraint, i.e. the consumer solves

$$(5.1)\ \max_{\{C_t\}} E\left(\Sigma_{i=0} (1+\delta)^{-i} u(C_{t+i}) \mid \mathcal{H}_t \right) \qquad \text{subject to} \quad E\left(\Sigma_{i=0} (1+r)^{-i} C_{t+i} \mid \mathcal{H}_t \right) \le W_t$$

where W_t wealth at time t (which includes the expected discounted value of future earning), δ is a time invariant subjective discount factor, $\mathcal{H}_t$ is the information set at time t, and r is the constant one period interest rate The first order conditions for utility maximization imply

$$(5.2)\ E(Z_{t+1} \mid \mathcal{H}_t) = (1+\delta)(1+r)^{-1}$$

where $Z_{t+1} = u'(C_{t+1})/u'(C_t)$ is the marginal rate of substitution for consumption between periods t and t+1. An empirical specification follows from an assumption concerning the functional form of the utility function and the the probabality distribution for Z_t. Here, we follow Hansen and Singleton (1983). Let $z_t = \log Z_t$, and assume that $z_{t+1} \mid \mathcal{H}_t \sim N(\mu_t, \sigma^2)$. The log normality of Z_{t+1} implies that $E(Z_{t+1} \mid \mathcal{H}_t) = \exp(\mu_t + \sigma^2/2)$. But equation (5.2) implies that $E(Z_{t+1} \mid \mathcal{H}_t)$ is constant, so that $\mu_t = \mu$, for all t. If we now assume that u(C) is of the constant relative risk aversion form, so that $u(C) = (1-\beta)^{-1} C^{1-\beta}$, then $z_t = -\beta (c_{t+1} - c_t)$, with $c_t = \log C_t$. When $c_t \in \mathcal{H}_t$, this implies

$$E(c_{t+1}|\mathcal{H}_t) = c_t + \alpha$$

with $\alpha = \beta^{-1}[(\sigma^2/2) + r - \delta]$, so that

$$(5.3) \quad c_{t+1} = c_t + \alpha + e_{t+1}$$

where e_{t+1} is uncorrelated with information available at time t. We will test this proposition by investigating the correlation between Δc_t and lagged values of the cyclical component of disposable income.

Before proceeding to a test of this hypothesis using the data described in the last section, one issue concerning the data should be addressed. When this random walk hypothesis is tested using macro data, the consumption and income figures are usually deflated by population before the analysis begins. This expresses all variables in per-capita terms, so that the data are loosely consistent with a representative consumer notion. We have chosen not to follow this course. While we agree that this transformation is useful in principle, in practice it can lead to serious problems. These problems arise from the errors in the quarterly population series. While the underlying trend in the population estimate is probably close to the trend in the true series, the quarter-to-quarter changes in the estimates series is, most likely, almost entirely noise. Indeed, over 50% of the sample variation in quarterly postwar population growth rates can be attributed to three large outliers in the data series.[2] Since our specification is in quarterly changes, or deviations from a stochastic trends, this increase in the noise to signal ratio will lead to serious inconsistencies. Rather than introduce this additional noisy series into the analysis, we will use the raw log differences of the non-durable consumption data and the stochastically detrended estimates of the log of disposable income.

The analysis of the last section casts some light on the hypothesis embodied in equation (5.3). There we showed that the hypothesis that c_t was a univariate random walk was consistent with the data. We calculated a "t-statistic" associated with this hypothesis that had a value of .77. In this section we ask whether Δc_t can be predicted by linear combinations of disposable income. The models that we will estimate in this section all have the form

$$(5.4)\ \Delta c_t = \alpha + \rho(B) y^c_{t-1} + e_t$$

where y^c_t is the cyclical component in disposable income, and $\rho(B)$ is a one-sided polynomial in B. Under the assumption of the life cycle consumption model outlined above, $\rho(B) = 0$ for any choice of (a one-sided) polynomial. Since y^c_t is not directly observed, the model (5.4) cannot be estimated by OLS. We will estimate the model using various proxies for the unobserved y^c_t data. The sample period is 1954:1 to 1984:2.[3]

The first model that we estimate, regresses Δc on four lags of $y^c_{t/t}$ (i.e. $y^c_{t-1/t-1}, y^c_{t-2/t-2}, \ldots, y^c_{t-4/t-4}$). Since each of these constructed variables is a linear combination of data at time t-1 or earlier, the population OLS coefficients should be zero. The results of this regression are shown in the first column of Table 2. Individually, the coefficients are large and have t-statistics ranging from 1.7 to 2.6. The last entry in the table shows the F-statistic, which tests that all of the coefficients are equal to zero. It takes on the value 2.7, larger than the 5% critical value.

In the next column, the results are presented for the same model, but with $y^c_{t/T}$, the two-sided estimates, used as proxies for y^c_t. The results are simliar. The F-statistic is now 3.8, which is significant at any reasonable level. The results in this column, however, are perfectly consistent with the theory. Recall that the two-sided estimates contain future values of disposable income. Since future values of disposable income are not included in $\mathcal{H}_{t-1}$, Δc_t may be correlated with these variables. (Indeed we would expect innovations in consumption to be correlated with innovations in income.) If we proxy the components y^c_t by the smoothed values, and estimate the coefficients using data in $\mathcal{H}_{t-1}$ as instruments, then the coefficients should not be significantly different from zero. The results of this exercise are shown in the third column. Only the last coefficient is now significant and the F-statistic has fallen to 2.3, significant at the 10% but not the 5% levels.

The results of this section suggest that aggregate postwar US data are not consistent with the life cycle model outlined above.

6. Concluding Remarks

This paper was motivated by the desire for a flexible method to eliminate trends in economic time series. The method that was developed in this paper was predicated on the assumption that deterministic trend models were too rigid and not appropriate for most economic time series. The alternative — modelling economic time series as non-stationary stochastic processes of the ARIMA class — confused long-run and cyclical movements in the series. The useful, fictitious decomposition of a time series into trend and cyclical components could not be used when modelling series as ARIMA processes. The method described in this paper maintains the convenient trend/cycle decomposition, while allowing flexibility in the models for both of the components.

In addition to the general discussion of stochastic detrending offered in the paper, several important empirical results were found. First, while the trend/cycle unobserved components models were shown to be equivalent to ARIMA models, the characteristics of estimated UC and ARIMA models were shown to be quite different for the three economic time series considered. Both models shared the same short run characteristics, but predicted quite different long run behavior of the series. Short run forecasting ability of the models were essentially identical, but the UC models provided uniformly better longer run forecasts. The relative efficiencies for the ARIMA models at a 40 quarter forecast horizon ranged from 65% to 70%.

The paper also discussed the use of stochastically detrended data in the construction of econometric models. Here we demonstrated that care has to be taken to avoid inconsistencies arising from complications similar to errors-in-variables. Our empirical example investigated the relationship between the change in consumption of non-durables and lags of the cyclical component of disposable income. Here we found a significant relationship, which indicates that the simple life cycle model, with its maintained assumptions of constant discount rates, no liquidity constraints, and time separable utility, is not consistent with aggregate postwar US data.

Footnotes

1. Those familiar with recursive signal extraction methods will recognize the $ms(c_t - c_{t/t})$ is a time varying quantity that, for the model under consideration, will converge to a fixed quantity as t grows large. (See Burridge and Wallis(1983).) Since one root of the AR process for c_t is close to the unit circle, and is therefore close to the roots of the AR polynomial for τ_t, the convergence of the $ms(c_t-c_{t/t})$ is very slow. The value of $rms\ (c_t- c_{t/t})$ reported in the text corresponds to the value for 1984:2. The recursive algorithm, the Kalman filter, was initialized with a vague prior, so that rmse's for earlier dates are larger. For example, the value for 1966 is .028. The $rms(c_t - c_{t/T})$ is approximately .019 over the entire sample period.

2. The quarterly population data (published by the Department of Commerce, Bureau of Economic Analysis), are very close to a deterministic trend. The trend is interrupted by three very large quarterly outliers. Over the period 1948:1 to 1980:4 the average quarterly population growth rate was .34%, with a sample standard deviation of .12%. The data show a .85% population growth rate in 59:1, a .79% growth rate in 71:4, and a -.23% growth rate in 72:1. These three data points are responsible for most of the sample variance in postwar population growth rates. They account for 51% of the sample sum of squares . When these data are used to deflate the consumption and income series used in the regressions, the results are dominated by these three data points.

3. We have started the sample period in 1954 to eliminate the observations in which the estimates $y^c_{t/t}$ are very imprecise. See the discussion on footnote 1.

References

Bell, W. (1984), "Signal Extraction for Nonstationary Time Series," **The Annals of Statistics**, 12, pp. 646-684.

Beveridge, S. and Nelson, C.R. (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**, 7, pp. 151-174.

Burridge, P. and K.F. Wallis (1983),"Signal Extraction in Nonstationary Time Series," University of Warwick Working Paper no. 234.

Cumby, R.E., J. Huizinga, and M. Obstfeld (1983), "Two-Step Two-Stage Least Squares Estimation in Models with Rational Expectations," **Journal of Econometrics**, 21, pp. 333-355.

Davidson, J.E.H., David F. Hendry, Frank Srba, and Steven Yeo (1978), "Econometric Modelling of the Aggregate Time-Series Relationship Between Consumer's Expenditure and Income in the United Kingdom," **Economic Journal**, 88, pp. 661-692.

Davies, R. B. (1977), "Hypothesis Testing When a Nuissance Parameter is Present Only Under the Alternative," **Biometrika**, 64, pp. 247-254.

Flavin, M. (1981) "The Adjustment of Consumption to Changing Expectations about Future Income," **Journal of Political Economy**, 89, pp .974-1009.

Geweke, John (1983) "The Superneutrality of Money in the United States: An Interpretation of the Evidence," Carnegie Mellon Discussion Paper, May 1983.

Granger, C.W.J. (1983) "Co-Integrated and Error-Correcting Models," UCSD Discussion Paper 83-13a.

Granger, C.W. J. and Engle, R. F. (1984) "Dynamic Model Specification With Equilibrium Constraints: Co-Integration and Error-Correction," UCSD Discussion Paper August 1984.

Hansen, L. and K. Singleton (1983)," Stochastic Consumption, Risk Aversion, and the Temporal Behavior of Asset Returns," **Journal of Political Economy**, 91, pp. 249-265

Hall, R. E. (1978),"The Stochastic Implications of the Life-Cycle Permanent Income Hypothesis: Theory and Evidence," **Journal of Political Economy**, 86, pp. 971-987.

Harvey, A.C. and P. H. J. Todd (1983), "Forecasting Economic Time Series With Structural and Box-Jenkins Models: A Case Study," (with discussion) **Journal of Business and Economic Statistics**, 1, pp. 299-315.

Hayashi, F. and C. Sims (1983), "Efficient Estimation of Time Series Models with Predetermined, but not Exogenous, Instruments," **Econometrica**,, 51, pp. 783-798.

Mankiw, N.G. and M.D. Shapiro (1984),"Trends, Random Walks, and Tests of the Permanent Income Hypothesis," Cowles Foundation DP # 725.

Nelson, C.R. and Kang, H. (1984) "Pitfalls is the Use of Time as an Explanatory Variable in Regression," **Journal of Business and Economic Statistics**, Vol. 2, pp. 73-82.

__________________ (1981), "Spurious Periodicity in Inappropriately Detrended Time Series," **Econometrica** , 49, pp.741-751.

Plosser, C.I. and Nelson, C. R. (1981) "Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications," **Journal of Monetary Economics** , 10, pp. 139-162.

Watson, M.W. and R. F. Engle (1985), "Testing for Regression Coefficient Stability with a Stationary AR(1) Alternative," forthcoming **Review of Economics and Statistics** , September.

Whittle, P. (1963) **Prediction and Regulation** , London: English Universities Press.

Zellner, A. and F. Palm (1976), "Time-Series Analysis and Simultaneous Equation Econometric Models," **Journal of Econometrics** , 2, pp. 17-54.