THE UNIVERSITY OF CALGARY

PERSISTENCE AND COINTEGRATION IN GDP TIME SERIES

by

GARTH A. RENNE

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE

OF MASTER OF ARTS

DEPARTMENT OF ECONOMICS

CALGARY, ALBERTA

MAY, 1992

ISBN 0-315-79091-1

THE UNIVERSITY OF CALGARY

FACULTY OF GRADUATE STUDIES

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance, a thesis entititled "Persistence and Cointegration in GDP Time Series" submitted by Garth A. Renne, in partial fulfillment of the requirements for the degree of Master of Arts.

Dr. F.J. Atkins, Supervisor
Department of Economics

Dr. A. Serletis
Department of Economics

Gordon Sick

Dr. G. Sick
Faculty of Management

June 1992

ABSTRACT

Two significant extensions of the unit root controversy are the issues of persistence and cointegration in economic time series. Both these areas involve efforts to obtain positive empirical results while dealing directly with the possibility that economic variables may be difference stationary. Whereas the unit root literature threatened to render spurious many commonly accepted economic relationships, the investigation of cointegration is an attempt to re-establish these relationships upon a sounder base of statistical theory. Persistence measurement has the complementary aim of assessing the importance of unit roots to the dynamic behaviour of economic processes. In each case, researchers have applied successively more sophisticated techniques, one of the more recent being spectral analysis. The unfamiliarity of this methodology has made the literature inaccessible to many.

This thesis examines persistence and cointegration in GDP time series, series for which these issues may have important implications for the existence and nature of business cycles, as well as for international linkages which tend to couple together many of the major industrialized economies. Cochrane's variance ratio statistic and Phillips and Ouliaris' cointegration test are utilized for this purpose. Spectral analysis is employed throughout, but is explained largely through the use of straightforward, intuitive arguments and analogies. It is shown that this technique, rather than complicating the analysis, instead enables the results to be interpreted more readily than if time domain methods were used in isolation.

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Dr. F.J. Atkins, for his support and encouragement throughout the development of this thesis; especially appreciated was his willingness to discuss theoretical minutiae only tangentially related to the core subject at hand. I would also like to acknowledge the Canadian Energy Research Institute, and in particular, Michelle Heath, for allowing me to be flexible enough in my work schedule to complete this project.

In addition, I would like to express my sincere thanks to both the Department of Economics and the Faculty of Graduate Studies Scholarship Committee for their financial assistance over the course of my M.A. program.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter One
INTRODUCTION

The measurement of persistence and cointegration of GDP time series relates closely to some of the most fundamental results of macroeconomic and econometric theory. If agents are assumed to behave in an intertemporally optimal manner, then models of the labour market, consumption and savings behavior, as well as more general macroeconomic models can all demonstrate radically different results depending upon the persistence of the exogenous shocks presumed to be driving them. The issue of persistence in time series is also linked to the choice between trend and difference stationary representations of data generating processes, a choice which has significant implications for statistical inference and forecasting, and has even been suggested as a means by which to obtain empirical support for competing models of economic fluctuations. Equally relevant is the observation of interdependencies in the dynamic behavior of GDP time series: evidence of cointegration between these series suggests that while their individual levels may vary extensively, economic forces exist which tend to prevent these series from diverging too much from one another over suitably long time scales.

Given the importance of these topics, it is not surprising that the issues of persistence and cointegration have spawned an extensive body of literature. What is notable about this literature is its comparatively recent development: virtually all of the studies of this type have been published within the last decade. This is due primarily to the fact that the conventional view of economic time series as being stationary about deterministic trends was largely unchallenged until the early 1980s. The nonstationarity exhibited by many economic series, including GDP, was modelled by a deterministic, secular trend; the appropriate response to this nonstationarity was to detrend such series and perform the desired statistical tests on the residuals. The explanation of the cyclical fluctuations about the secular trend--which were associated by theorists with transient disturbances such as monetary shocks--was the main objective, while the trend--thought by most to be attributable to the slow evolution of real factors--was of subsidiary interest.

Theoretical advances in the econometrics of nonstationary time series led to the reconsideration of the traditional interpretation of the dynamics of economic time series. Authors such as Dickey and Fuller (1979) pointed out that secular movements in time series need not be modelled as deterministic trends, but may be better interpreted as being fundamentally stochastic, as is the case for a random walk with drift. Nelson and Plosser (1982) showed that many economic series, including GDP, did not permit the rejection of the null hypothesis that their time series representations contained a unit root. The implications were that these series were difference stationary rather than trend stationary, and that conventional models were mis-specified and therefore likely to give misleading and spurious results.

The concepts of persistence and cointegration are both logical extensions of the unit root literature, the main difference being one of focus. The question of the existence of a unit root is de-emphasized in studies of persistence in favour of the goal of assessing the importance of its contribution to the behavior of the time series in question. The aim is to estimate the proportion of a given shock which can be expected to persist over long time horizons, or, alternatively, to measure the size of the random walk component in the series relative to the process' total variance. Studies of cointegration seek to establish if the first difference filter eliminates all evidence of relationships between series suspected of possessing unit roots, or if the shocks affecting the levels of such series are mutually dependent as a consequence of forces which link the series' dynamic behavior. Both persistence and cointegration presuppose the existence of unit roots: a trend stationary process' repeated return to a trend line implies a persistence of zero, while it is meaningless to consider cointegration of series which do not have to be differenced to make them stationary.

It should be clear from the forgoing that the topics of persistence and cointegration are closely related, and therefore easily studied in conjunction with one another. The most obvious commonality between these attributes of time series is their close association with the "long run." By its very definition, persistence is concerned solely with the properties of stochastic processes over extended periods of time. Similarly, the concept of cointegration of stochastic processes is designed to be compatible with a wide variety of short-run dynamics in the context of an underlying, long run equilibrium relationship (Granger and Engle, 1987). In order to examine these aspects of time series, then, the researcher must have access to two things: long data series (in terms of time span, and not just in terms of the number of observations), and a method by which to focus on long run dynamics within the data so as to preclude short-run information from obscuring long run relationships. This study meets the first condition by utilizing data series consisting of over one hundred annual observations. The second requirement is met through the use of spectral analysis, a relatively straightforward technique for decomposing time series into constituent components, each of which consists of a sinusoidal wave of a specific frequency.

The central aim of this thesis is to apply spectral analysis to the measurement of persistence and comovements in GDP time series, and to demonstrate how this technique leads to intuitive explanations of statistics whose interpretation in the time domain is much more difficult. The following two chapters introduce the theory necessary to accomplish these tasks. Chapter two discusses Cochrane's variance ratio statistic, and shows how Cogley's asymptotically equivalent frequency domain estimator can be obtained from a time series' periodogram. In addition, the variance ratio statistic is related to Campbell and Mankiw's impulse-response measure of persistence; this relationship is shown to be particularly simple when viewed in the frequency domain. Chapter three examines how comovements in time series can be inferred from the coherencies and phase angles of the relevant cross spectra. It also introduces Phillips and Ouliaris' test for cointegration, which is based upon the relative magnitudes of the eigenvalues of the sample spectral density matrix evaluated at frequency zero. In chapter four, these tools are applied to GDP data from nine countries for the period 1871-1984. Specifically, the results obtained by Cochrane and Cogley are extended in two ways: the capability of Cogley's estimator to provide more precise confidence intervals is examined, and Phillips and Ouliaris' cointegration test is applied to pairs of GDP time series, rather than treating the entire set of series as a single system of dynamic processes. Chapter five summarizes the most significant findings from chapter four, and presents some ideas for future research.

Chapter Two

ALTERNATIVE APPROACHES TO PERSISTENCE MEASUREMENT

This chapter is devoted to developing the theoretical background required for interpreting the various statistics used to estimate persistence in time series. Part A gives a brief summary of relevant results from spectral analysis. Parts B and C discuss Cochrane's variance ratio statistic and Campbell and Mankiw's persistence measure, respectively. Part D concludes with a comparison of these two methods, emphasizing their intuitive frequency domain interpretations.

A. Preliminary Theory

In seeking to characterize the serial dependence of a stationary stochastic process (Y_t), the standard approach is to use the autocovariance function, defined as:

$$\begin{aligned} \gamma_k &= E[(Y_t - E(Y_t))(Y_{t-k} - E(Y_t))] \\ &= \mathrm{cov}(Y_t, Y_{t-k}) \end{aligned} \tag{1}$$

This function, which applies to the underlying data generating process, is of course unobservable, and is typically estimated by the following formula, the sample autocovariance function:

$$g_k = \frac{1}{n} \sum_{t=k+1}^{n} (y_t - \bar{y})(y_{t-k} - \bar{y}), \tag{2}$$

where n is the number of observations and $\bar{y}$ is the sample mean.

Equation (2) focuses attention on how each observation is related to the observations which precede it, and so implicitly describes any periodicities which may be present in the data. However, the sample autocovariance function does not make the existence of such cycles readily apparent. Moreover, adjacent values of the sample autocovariance function can be highly correlated, making the information pertaining to the data's cyclical behavior even more difficult to interpret. These drawbacks mean that the autocovariance function is of limited usefullness when one's primary interest is the analysis of periodic components of time series.

Spectral analysis overcomes the disadvantages of the autocovariance function by taking the information which it contains and re-expressing it in a different format. In this approach, stochastic processes are characterized by the Fourier transforms of their autocovariance functions, which, when viewed as functions of an angle w, are known as spectra:

$$f(w) = \sum_{k=-\infty}^{\infty} \gamma_k e^{-iwk} \qquad -\pi < w \leq \pi$$

$$= \gamma_0 + \sum_{k=1}^{\infty} \gamma_k \left(e^{iwk} + e^{-iwk}\right)$$

$$= \gamma_0 + 2\sum_{k=1}^{\infty} \gamma_k \cos(wk) \qquad (3)$$

Expression (3) will be well-defined as long as the underlying stochastic process is both stationary and ergodic.

The sample spectrum, which is referred to as the periodogram, bears the same relationship to expression (3) as the sample autocovariance function does to expression (1):

$$I(w_j) = \sum_{k=-n+1}^{n-1} g_k e^{-ikw_j} \qquad w_j = 2\pi j/n$$

$$j = 0 .. \frac{n}{2} \text{ (n even)}$$

$$j = 0 .. \frac{n}{2} - 1 \text{ (n odd)}$$

$$= g_0 + 2 \sum_{k=1}^{n-1} g_k \cos(kw_j), \tag{4}$$

where n is the number of data points in the series being analysed.

For each angle w_j, the quantity $I(w_j)$ is called the periodogram ordinate at frequency w_j. Its value can be interpreted in several ways. If the data series y_t is assumed to be composed of a number of sine and cosine waves corresponding to the w_js, such that the constituent waves possess the correct random amplitudes so as to make up the original series when they are superimposed, then the ordinates $I(w_j)$ are proportional to the squared amplitudes of the component waves (Engle, 1976). Alternatively, if the data series y_t is regressed on a series of sinusoidal components, each one possessing a frequency w_j, then the values $I(w_j)$ are proportional to the explained sums of squares derived from the regressions (Diggle, 1990). Thus, the periodogram does not contain any information which could not be extracted from the data using more traditional methods of analysis.

It can be shown that the periodogram ordinates corresponding to w_js in the interval $(0,\pi)$ are distributed as:

$$I(w_j) \sim \frac{f(w_j)\,\chi^2}{2}, \tag{5}$$

where the χ^2 term at each frequency is an independently distributed chi-squared variate with two degrees of freedom. Since the expected value of a chi-squared variate is equal to its degrees of freedom, equation (5) shows that each periodogram ordinate is an unbiased estimator of the underlying spectrum.

The most important feature of (4) for the purposes of persistence measurement is the fact that the periodogram partitions the total variation in the series y_t into the contributions of the various frequency components:

$$\sum_{t=1}^{n} y_t^2 = I(0)+2\sum_{j=1}^{m} I\left(\frac{2\pi j}{n}\right) + I(\pi), \tag{6}$$

where m is the largest integer less than n/2 (Diggle, 1990).

Since I(0) is simply n times the square of the sample mean y, and is therefore usually of little interest, data series are generally demeaned before their periodograms are computed. This results in the following relation:

$$\sum_{t=1}^{n} (y_t-\bar{y})^2 = 2\sum_{j=1}^{m} I\left(\frac{2\pi j}{n}\right) + I(\pi) \tag{7}$$

Equivalently,

$$\frac{1}{n} \sum_{t=1}^{n} (y_t-\bar{y})^2 = \frac{1}{n} [(2\sum_{j=1}^{m} I\left(\frac{2\pi j}{n}\right) + I(\pi)] \tag{8}$$

Expression (8) gives the important result that the average value of a (demeaned) series' periodogram ordinates is the series' sample variance g_0. By partitioning the variance of a series into various frequency contributions, the spectrum gives a very general characterization of a stochastic process. This is one of the main advantages of employing spectral analysis in persistence measurements, for it does not require the researcher to assume a particular functional form for the underlying data generating process. Instead, it allows the data to give results without being forced to conform to prior expectations. The flexibility of this non-parametric approach, while desirable, is nevertheless something of a mixed blessing, since one would expect that without the structure of a parametric model, it may be difficult to obtain precise results and narrow confidence limits. This possibility is examined further in chapter four.

The last features of spectral theory which will be required in later sections are the concepts of lag and spectral windows. The formula given in (4) is rarely used in that precise form, but instead is usually modified as follows:

$$I'(w_j)=\lambda_0 g_0 + 2\sum_{k=1}^{n-1}(\lambda_k g_k \cos(kw_j)), \qquad (9)$$

where $\lambda_0=1$, $\lambda_{k+1}<=\lambda_k$, and, typically, $\lambda_k=0$ for $k > K$, $K < n-1$. Collectively, the set of λ_k values are known as a lag window, which can be interpreted as an even function of k which assigns to each autocovariance a weight in the calculation of the periodogram ordinates. When viewed in this way, equation (9) clearly involves taking the Fourier transform of the product of two functions, namely, the autocovariance and lag window functions. By the convolution theorem, which shows that multiplication in the time domain corresponds to convolution in the frequency domain, the windowed periodogram estimate $I'(w_j)$ is equivalent to the convolution of the periodogram estimate derived from (4) and the Fourier transform of the lag window. Thus, utilizing a lag window to calculate $I'(w_j)$ gives the same results as would be obtained by taking a weighted average of periodogram ordinates to derive a smoothed estimate of the spectrum, with the frequency domain weights being given by the Fourier transform of the lag weight function.

The reason for employing a smoothed spectral estimator such as (9) is that the periodogram is an inconsistent estimator of the spectrum: regardless of the length of the data series available, each periodogram ordinate remains a chi-squared variate with two degrees of freedom, as expressed in equation (5). This follows from the fact that each ordinate is independently distributed, so that any additional information obtained by increasing the length of data series available is employed to compute more independent estimates of the underlying spectrum, rather than to decrease the variance of a fixed number of estimates (Engle, 1976). As a result, the unsmoothed periodogram is usually very erratic, with each ordinate's value reflecting sampling errors as much as the underlying spectrum. Utilizing an appropriate spectral window to smooth the sample periodogram allows this sampling variance to be reduced at the expense of the introduction of some bias in the spectral estimator.

B. The Variance Ratio Statistic

The argument which Cochrane (1988) uses to motivate his variance ratio measure of persistence is relatively simple. Assume that a GDP time series y_t, measured in natural logs, is first differenced. The variance of the resulting series, $\sigma^2_{\Delta y_t}$, shows the dispersion of GDP's year-to-year percentage growth about its mean value. Next, suppose that the variance of the GDP series' kth differences--that is, cumulative growth over a k year time horizon--is calculated, and its magnitude compared to $k\sigma^2_{\Delta y_t}$. If the original series was a realization of a trend stationary process, one would expect the series' repeated reversion to a trend line to limit the growth of $\sigma^2(y_t - y_{t-k})$, even if k is increased to arbitrarily large values. The variance of cumulative k year growth will be constrained because years of above-average growth will tend to be followed by lower growth rates as GDP returns to its trend line. If the GDP process was a pure random walk, however, the situation would be quite different: y_t-y_{t-k} would be the sum of k uncorrelated shocks, and so $\sigma^2(y_t - y_{t-k})$ would be expected to increase linearly with the value of k. This line of reasoning led Cochrane to propose the following measure of persistence for a stochastic process:

$$V = \lim_{k\to\infty} \frac{\mathrm{Var}(Y_t - Y_{t-k})}{k\,\mathrm{Var}(Y_t - Y_{t-1})} = \lim_{k\to\infty} V_k \qquad (10)$$

Cochrane proves rigorously that V_k asymptotically approaches zero for any trend stationary process, while for a random walk, V_k's limiting value is unity.

Obtaining an acceptable estimate for V requires that a value of k be chosen, and that estimators for the population moments in (10) be selected. These decisions can be aided by re-writing (9) as:

$$V = \lim_{k\to\infty} \frac{\gamma_0^{\Delta Y_t} + 2\sum_{j=1}^{k-1} \frac{k-j}{k}\, \gamma_j^{\Delta Y_t}}{\gamma_0^{\Delta Y_t}} \tag{11}$$

(See Cochrane (1988) for proof)

The sample version of (11) for a given value of k is:

$$\hat{V}_k = \frac{g_0^{\Delta y_t} + 2\sum_{j=1}^{k-1} \frac{k-j}{k}\, g_j^{\Delta y_t}}{g_0^{\Delta y_t}} \tag{12}$$

This expression shows that the choice of k reduces to the number of autocovariances to be included in the numerator of (12). If too few autocovariances are used, then the trend reversion manifested by negative higher autocorrelations may be obscured, while if too many are used, the resulting estimate may be biased downward, since $\sigma^2(y_t - y_{t-k})$ must approach zero as k approaches the number of data points available. Given these considerations, the following procedure presents itself: calculate a point estimate of V for various values of k, examining the behavior of $\hat{V}_k$ as k is varied. If $\hat{V}_k$ remains relatively constant for k greater than some value k_0, then $\hat{V}_{k0}$ can be selected as an estimate of the underlying parameter V. Campbell and Mankiw (1987) and Cochrane (1988) both use techniques similar to this in order to obtain point estimates of V.

Another approach for estimating V is presented by Cogley (1990). Comparing (12) to (4), it is clear that the numerator of the estimator $\hat{V}_k$ consists of a windowed estimate of the value of the spectrum of ΔY_t at frequency zero. The lag window in this case consists of a "tent-shaped" or triangular function with weights which decline linearly with k, terminating with the (k-1)st autocorrelation. The properties of this window, known as the Bartlett lag window, are well known, and in fact were used by Cochrane (1988) to prove that the distribution of $\hat{V}_k$ is asymptotically normal. Cogley suggests working entirely in the frequency domain, using the Bartlett spectral window to smooth the first few periodogram ordinates as an estimate of the numerator of $\hat{V}_k$. The required spectral window can be obtained by taking the Fourier transform of the "tent-shaped", time domain lag window:

$$
\begin{aligned}
B(w_j) &= 1 + 2 \sum_{l=1}^{K-1} \frac{k-l}{K} \cos(l w_j) \\
&= \frac{1}{k} \left[\frac{\sin\left(\frac{k w_j}{2}\right)}{\sin\left(\frac{w_j}{2}\right)} \right]^2 \qquad (13)
\end{aligned}
$$

(See the appendix for the complete derivation of equation (13)).

The weight function $B(w_j)$ could in principle be calculated for each periodogram ordinate from $2\pi/n$ to π. However, since the function in (13) is concentrated about frequency zero, Cogley recommends that $B(w_j)$ be truncated at a frequency w_m, with a value of m equal to 10 yielding a satisfactory approximation to the entire window when k = 15. Figures 1 and 2 demonstrate this approximation for the case in which k = m = 15 and the data series has 100 observations.

With the values of k and m chosen, the weights $B(w_j)$ can be computed for each ordinate j = 1..m. These weights are then normalized so

Figure 1
The Bartlett Lag Window

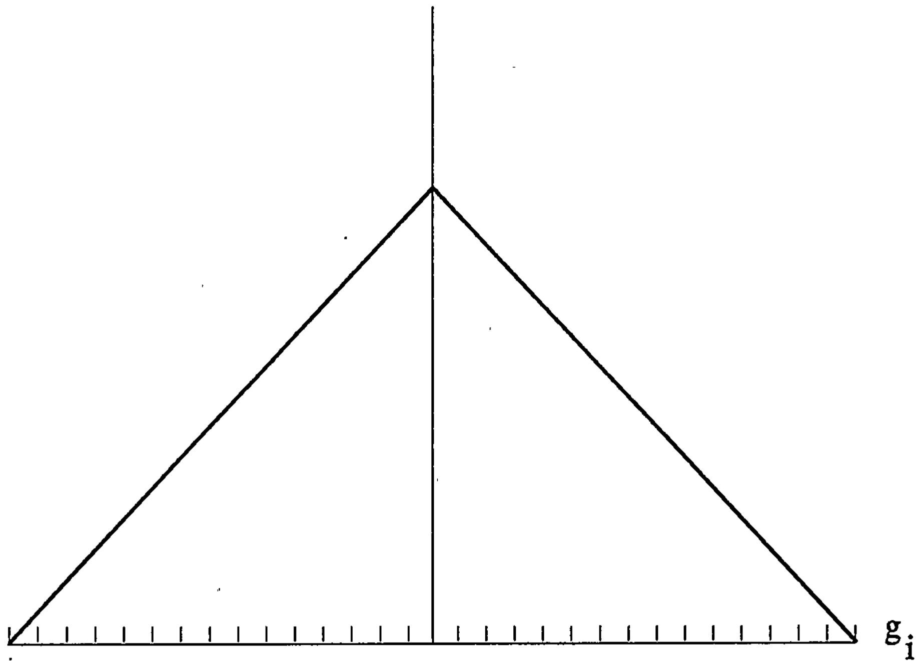
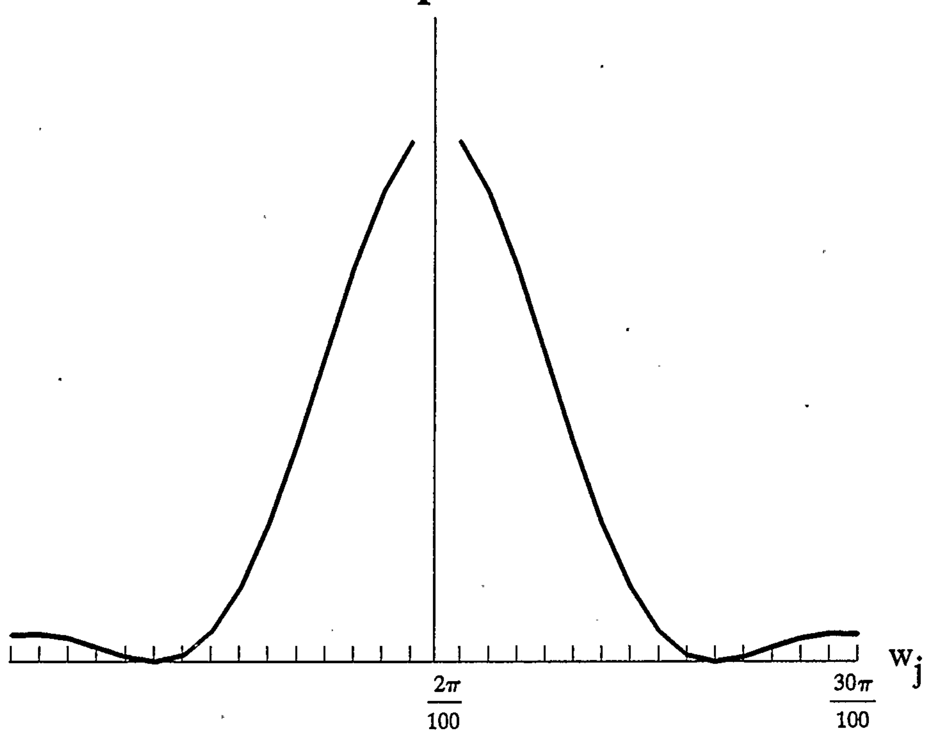


Figure 2
The Bartlett Spectral Window



that they sum to unity, since the objective is to take a weighted average of periodogram ordinates, each of which is an unbiased estimate of a point on the underlying spectrum. After obtaining the resulting windowed estimate for the numerator of (12) using the normalized weights $B_n(w_j)$, computing g_0 allows the frequency domain estimate of V, which Cogley refers to as $\hat{V}_f$, to be calculated.

Although the statistics $\hat{V}_k$ and $\hat{V}_f$ will differ in finite samples, they are asymptotically equivalent, a fact which may seem to cast doubt on the wisdom of incurring the added complication of using the frequency domain to estimate V. However, Cogley shows through Monte Carlo techniques that $\hat{V}_f$ possesses an important advantage over its time domain counterpart: based on the χ^2 distribution of periodogram ordinates, it is possible to construct confidence intervals for V_f which are superior to the normal approximation to the distribution of $\hat{V}_k$. Specifically, Cogley found that the empirical distribution of $\hat{V}_k$ has a lower tail which is too thin, so that confidence intervals based on the normal distribution contain lower values which the Monte Carlo evidence suggests are extremely rare. As a result, Cogley argues that the normal approximation makes it inordinately difficult to reject small values of the variance ratio, which justifies the use of $\hat{V}_f$ in place of $\hat{V}_k$.

The chi-squared approximation to the distribution of the statistic $\hat{V}_f$ can be derived from the distribution of the individual ordinates used to compute the estimator and the asymptotic properties of the periodogram as a whole. The weighted sum of ordinates used to estimate the numerator of the variance ratio is:

$$\hat{V}_f \text{ numerator} = \hat{f}(w_0) = \Sigma \; B_n(w_l) \, I(w_l), \qquad (14)$$

where $w_0 = 0$.

The expected value of this expression is therefore:

$$E(\widehat{f(w_0)}) = \sum_{l=1}^{m} B(w_l)\, f(w_l) \tag{15}$$

Assuming that the spectrum is a smooth function, as n approaches infinity we will have $f(w_l) \approx f(w_0)$. Therefore:

$$E(\widehat{f(w_0)}) \approx f(w_0) \sum_{l=1}^{m} B_n(w_l) = f(w_0) \tag{16}$$

The numerator of $\hat{V}_f$ is thus a consistent estimator of the spectral density of the ΔY_t process at frequency zero. This suggests the following approximate distribution:

$$\widehat{f(w_0)} \sim f(w_0)\, \chi_v^2 / v, \tag{17}$$

where v is known as the equivalent degrees of freedom. Equation (17) implies that:

$$\mathrm{Var}\,(\widehat{f(w_0)}) = \tfrac{2}{\nu}\, [f(w_0)]^2, \tag{18}$$

where use is made of the fact that the variance of a chi-square variate is twice its degrees of freedom. Equations (5) and (14) give the alternative expression:

$$\mathrm{Var}\,(\widehat{f(w_0)}) = \sum_{l=1}^{m} (B_n(w_l))^2\, (f(w_l))^2$$

$$\approx (f(w_0))^2 \sum_{l=1}^{m} (B_n(w_l))^2 \tag{19}$$

Equating (18) and (19) gives the required expression for the equivalent degrees of freedom:

$$v = 2 / \sum_{l=1}^{m} (B(w_l))^2 \tag{20}$$

Equations (17) and (20) allow approximate confidence intervals, based on the chi-squared distribution, to be constructed from point estimates of V_f. By using g_0 as a consistent estimator of γ_0, the denominator of the variance ratio, we obtain:

$$\hat{V}_f \sim \frac{V \chi_v{}^2}{v} \tag{21}$$

Using (21) allows the following 100 (1-a)% confidence interval to be derived:

$$v\hat{V}_f/\chi_{v,1-a/2} \quad <= V < = v\hat{V}_f/\chi^2_{v,a/2} \qquad (22)$$

When using (22), the equivalent degrees of freedom derived from the normalized Bartlett weight function can be rounded to the nearest integer in order to use tabulated χ^2 values.

The procedure outlined above for estimating the variance ratio statistic demonstrates that V admits a very simple interpretation in the frequency domain: it is simply the value of a given ΔY_t process' spectrum at frequency zero divided by the process' total variance. To estimate this value, one simply smooths the first few ordinates of the periodogram and divides by the sample variance, which is equal to the average value of the periodogram's ordinates. The variance ratio statistic is therefore simply a measure of the proportion of the variance in output growth which is accounted for by low-frequency cycles; it is in this sense that it is a measure of the size of the random walk component in GDP. The choice of k in (12), which corresponds via a Fourier transform to the $B_n(w_l)$ in (14), basically defines a choice of "long run" for the sample data in question. It can be shown that as k rises, the weight function $B(w_j)$ becomes more concentrated about frequency zero, so that low-frequency, long-run cycles assume greater and greater importance to the value of $\hat{V}_f$. A large value of k means that many higher-order autocorrelations - and therefore "longer run" data - are employed in the calculation of $\hat{V}_k$, and more weight is assigned to the lowest-frequency cycles which are discernable in the sample series. This fact is made much more obvious when the frequency domain estimator of V is considered.

C. The ARMA Approach

Campbell and Mankiw (1987) have developed another method for measuring persistence in GDP time series, one which involves estimating general ARMA models of GDP growth. This approach assumes that the wide variety of dynamics which can be exhibited by ARMA processes, even those with very few parameters, can adequately mimic the complex movements which characterize series such as GDP. These univariate representations of GDP are not meant to be structural, of course, but are simply approximations to the dynamic behavior of the series in question. The principal advantage of this parametric approach is that once they are estimated, ARMA models enable quantitative persistence estimates to be obtained for various time horizons, including the infinite horizon used to define V.

Campbell and Mankiw assume that the change in the log of GDP can be modelled as a stationary ARMA(p,q) process:

$$\begin{aligned} A(L) \quad \Delta Y_t &= B(L)\,\varepsilon_t \\ \Delta Y_t &= (B(L)/A(L))\,\varepsilon_t \\ \Delta Y_t &= C(L)\,\varepsilon_t, \end{aligned} \tag{23}$$

where A(L) and B(L) are polynomials in the lag operator L of orders p and q, respectively, and ε_t is an uncorrelated error term with variance σ_ε^2. Since ΔY_t is assumed to be stationary, it must be the case that the growth rate of GDP has finite variance. This in turn means that $\lim_{i \to \infty} C_i = 0$, so that innovations in the growth rate in the current period have no effect on the forecast of growth in the distant future.

In this formulation, C_i gives the effect on the growth rate of GDP in period t + i from a shock to ΔY_t in period t. Thus, the cumulative effect on the level of GDP from a current-period innovation over an infinite horizon is:

$$
\begin{aligned}
\text{Effect on GDP level} \quad &= \lim_{k \to \infty} \left[\sum_{i=1}^{k} C_i \right] \\
&= 1 + C_1 + C_2 \ldots \\
&= C\,(1) \qquad (24)
\end{aligned}
$$

Campbell and Mankiw therefore propose that the appropriate measure of persistence in time series is the estimated value of C(1) derived from some suitably parameterized ARMA model, with a given model's acceptability being judged on the basis of the well-known Akaike or Schwarz criterion.

Although it may appear that first-differencing GDP time series presupposes the existence of a unit root and some degree of persistence, this is not the case. To illustrate this fact, assume that GDP, in level form, is itself a stationary ARMA(p,q) process:

$$Y_t = \quad (E(L) / F(L)).\varepsilon_t = D(L)\, \varepsilon_t \qquad (25)$$

Applying the first-difference filter 1-L to (24) yields an ARMA (p,q+1) process:

$$\Delta Y_t = \quad (1\text{-}L)\, D(L)\varepsilon_t = G(L)\, \varepsilon_t \qquad (26)$$

The polynomial in the lag operator G(L) will therefore contain a unit root, and so G(1) will be identically equal to zero. In contrast, if GDP is a pure random walk, then C(L) in (23) will be identically unity, and so will Campbell and Mankiw's proposed persistence measure. Between these extremes, a continuum of possibilities exist, with the precise value of C(1) depending on the correlations between current shocks to output growth and those occurring in future periods. As Campbell and Mankiw observe, it follows that in order to leave open the possibility that the level of GDP is a stationary process, at least one MA parameter must be included in any estimated model along the lines of (23).

D. Comparing the Two Approaches

It might be expected that there exists a close connection between Cochrane's variance ratio and Campbell and Mankiw's measure, since they give identical results if GDP is assumed to be either a trend stationary process or a random walk. However, these two methods are not completely equivalent: in fact, for most stochastic processes, the measures will differ, in some cases substantially. The precise ways in which these two methods for measuring persistence differ from each other can be developed in both the time and frequency domains.

In the frequency domain, the relationship between the two approaches follows from the following fact: if a stochastic process is characterized by the moving average representation (23), then its spectrum is given by the expression:

$$f_{\Delta Y_t}(w) = \| C(e^{-iw})\|^2 \sigma_\varepsilon^2 \qquad (27)$$

In (27), the expression $\|C(e^{-iw})\|^2$ is the squared modulus of the complex number obtained by substituting the value of e^{-iw} in place of L in the polynomial C(L) (Sargent, 1987). This result is just a special case of the convolution theorem, in which the spectrum of the error term, σ_ε^2, is multiplied by the squared modulus of the transfer function $C(e^{-iw})$. Evaluating (27) at frequency zero gives the following result:

$$f_{\Delta Y_t}(0) = \|C(1)\|^2 \ \sigma_\varepsilon^2 \tag{28}$$

Therefore, the spectrum of the differenced GDP process at frequency zero is proportional to the square of Campbell and Mankiw's persistence measure. Dividing (28) by $\sigma_{\Delta Y_t}^2$, the variance of the entire ΔY_t process, one obtains:

$$V = \frac{\|C(1)\|^2 \ \sigma_\varepsilon^2}{\sigma_{\Delta}{}^2{}_{Y_t}} \tag{29}$$

Equation (29) makes very clear why both persistence measures give identical results for trend stationary and random walk processes. Suppose that the GDP process is trend stationary. Then, from the arguments given in the previous section, the polynomial C(L) contains a unit root, and hence C(1) is identically zero. Clearly, then , (29) shows that V must also be zero. On the other hand, if GDP is a random walk, so that the differenced process is composed of pure white noise, then C(1) has the value of unity. Moreover, in this case σ_ε^2 equals $\sigma_{\Delta Y_t}^2$, as the variance of a white noise process is simply the variance of the uncorrelated shocks which drive the system. This means that V will also possess the value of one, a result which is immediately obvious when one considers that the spectrum of a white noise process is simply a constant function of frequency.

Expression (29) also suggests why the two persistence measures may differ. To take a specific example, suppose one begins with a white noise process for the first difference of GDP. Both methods would give a measure of unity for this GDP process' persistence. Now, assume that the process was altered so as to increase the total variance of the differenced process, while leaving the value of C(1) unchanged. If the exogenous shocks driving the system were also kept unaltered, then the value of V for the new process must be lower, while by assumption C(1) will be identical for the two GDP growth processes. The reason for the divergence of the two measures is that the although each shock to the new process has the same long run effect on the level of GDP (i.e., C(1)), the path of dynamic adjustment followed by the two processes differs. Since the denominator of V is the variance of the differenced process, any alteration to the process which leaves the long run impact of shocks unchanged while increasing the short term variability of the series must reduce the value of the variance ratio.

The argument of the preceding paragraph can be put in a more quantitative form, as developed by Campbell and Mankiw (1987). They define a quantity $R^2=1-\sigma_e^2/\sigma_{\Delta Y_t}^2$. This measures the proportion of the variance of the differenced Y_t process which is predictable from the past history of the process. If the differenced process is white noise, and therefore completely unpredictable, then we have $R^2 = 1 - 1 = 0$. For processes which are dominated by the effect of autoregressive or moving average coefficients, R^2 will be correspondingly larger, since much of the future variability in these series is determined by shocks which have already occurred.

Combining this definition with (29) yields:

$$C(1) = \sqrt{\frac{V}{1-R^2}} \qquad (30)$$

The larger is R^2, then, the greater is the difference between the two persistence measures.

Cochrane (1988) has criticized the use of low order ARMA models to measure persistence, claiming that they are likely to give upwardly-biased estimates for typical GDP processes. He bases his argument on the characteristics of the maximum likelihood estimation technique, the method employed by Campbell and Mankiw. Parametric models fitted using maximum likelihood attempt to match the dynamics of the data series over the entire frequency range, including the neighbourhood of frequency zero, the window which is relevant for persistence measurement. More formally, as Christiano and Eichenbaum (1990) show, maximum likelihood seeks (in population) to minimize the average percentage error between the theoretical spectral density of the specified model and the spectral density of the true, underlying data generating process. If the model is specified incorrectly, then maximum likelihood will select parameter values which produce a model with a spectral density which mimics as closely as possible the true spectral density, sacrificing accuracy in some frequency ranges if this allows the estimated model to conform more closely to the underlying process' spectrum over some other, broader intervals.

Maximum likelihood's attempt to imitate the underlying process' dynamics over all frequency ranges is, in most circumstances, a very desirable property. However, in the particular case of persistence measurement, Cochrane argues, it can lead to inaccurate results. The problem arises because of the small width of the frequency interval corresponding to the "long run": as Cochrane observes, the band of frequencies for cycles of twenty years and longer is only $2\pi/20 = \pi/10$ radians wide, while the range for dynamics consisting of cycles of two to four years duration is $\pi/2$ radians wide. Clearly, if a model estimated via maximum likelihood proves incapable of fitting the observed series' dynamics over all frequencies, it will forego accuracy in the narrow "long run" band in order to more closely imitate the series' shorter-run behavior. Christiano and Eichenbaum find evidence of this phenomenon in their Monte Carlo investigations: when an ARMA (2,2) is fitted to data generated by an trend stationary ARMA (3,3), the mis-specified model exaggerates the persistence present in the data, matching the underlying process quite faithfully for high-frequency dynamics and allowing its inability to imitate the underlying model accurately to fall disproportionately on a narrow band about frequency zero. Cochrane claims that this is responsible for the large persistence estimates found by Campbell and Mankiw, and purports to show that due to the very slow trend reversion exhibited by GDP, only very high-order ARMAs could be expected to model the series' long-run behavior.

Cochrane's criticism, while ingenious, is probably of less importance than it may at first appear. If the model being estimated is correctly specified, or if it is sufficiently flexible to fit the underlying process' behavior, then maximum likelihood will give consistent estimates of persistence. Moreover, for virtually any estimation procedure, one can almost invariably propose an underlying process which will lead to misleading results; this does not mean that the procedure should be abandoned unless it can be demonstrated that such results are likely to be the rule rather than the exception. Indeed, Cochrane's own variance ratio statistic can be subjected to a variation of Cochrane's criticism. As Christiano and Eichenbaum demonstrate, if the data generating process is assumed to be their trend stationary ARMA (3,3), and a typical sample size of approximately 100 observations is used, then the variance ratio estimator $\hat{V}_k$ will give upwardly-biased persistence measurements. The reason for this is that Christiano and Eichenbaum's ARMA (3,3), when first differenced, possesses a spectral density which rises very rapidly from the value of zero at frequency zero. Since the periodogram is a consistent estimator of the spectrum, this means that even the ordinates a short distance away from frequency zero employed in calculating the numerator of $\hat{V}_f$ will, on average, have values significantly different from zero. Unless a very large amount of data is available, allowing ordinates very close to frequency zero to be calculated, data generated by the ARMA (3,3) will produce variance ratio estimates which show non-zero persistence. This does not invalidate the variance ratio statistic, but merely underscores the fact that all econometric investigations are potentially vulnerable to problems arising from the unknown nature of the underlying stochastic processes which generate economic series.

Equation (28) shows that both the ARMA and the variance ratio approaches may have difficulty in distinguishing between trend stationary processes and local alternatives which are difference stationary. Since the presence of persistence is governed entirely by the value of C(1), which can take on a continuous range of values from zero upward, it is always true that one can find a difference stationary process which is arbitrarily close to a given trend stationary process in terms of dynamic behavior. This is one of the reasons that the Dickey-Fuller and other unit root tests have been criticized, because for sample sizes typical of those encountered in economics, it is very difficult to distinguish between a process whose first difference possesses a spectral density of zero at frequency zero, and one which contains a very small random walk component. This problem is also present in studies of persistence measurement, although establishing confidence bounds for the random walk component, even if it is small, should be a more tractable task than attempting to classify the underlying process as being either trend stationary or as difference stationary, but with a unit root which has negligible impact on dynamics extending over the length of available data series.

As a conclusion to this chapter, it may be helpful to briefly examine a specific example of a process for which the two persistence measures differ substantially. Consider the following differenced GDP process, from Cochrane (1988):

$$\Delta Y_t = \frac{\varepsilon_t}{1-1.5\,L + 0.95\,L^2} \qquad (31)$$

This process yields a Campbell-Mankiw measure of (1/(1-1.5+0.95)) = 2.22. It is shown by Cochrane that the value of V for this process is 0.2. This large disparity between the two values suggests that R^2 is very large for this process. One would therefore expect the process' short-run dynamics to be very volatile, so that a shock in the current period will give rise to substantial variation in the level of the series, before converging on the long term impact of 2.22 times the value of the original shock. This expectation is borne out by the adjustment process pictured in figure 3, which shows that a unit shock in period one generates a pattern of large cyclical fluctuations in the level form of the series, fluctuations which take a considerable length of time to damp out. The spectrum of the differenced process, displayed in figure 4, shows that the power of the frequencies near zero is comparatively small, thus giving rise to the very low value of V.

This example demonstrates that there is no unique, correct way to characterize the persistence of a stochastic process: while the variance ratio V accurately gauges the small proportion of this process' variance which is accounted for by the lowest frequencies, C(1) provides a measure of the long run effect of current shocks to this system. As long as this fundamental difference between the two methods is kept in mind, there need be no conflict between them, even if they yield seemingly disparate persistence estimates from observed series.

Figure 3
Impulse Response Function For Cochrane's Example

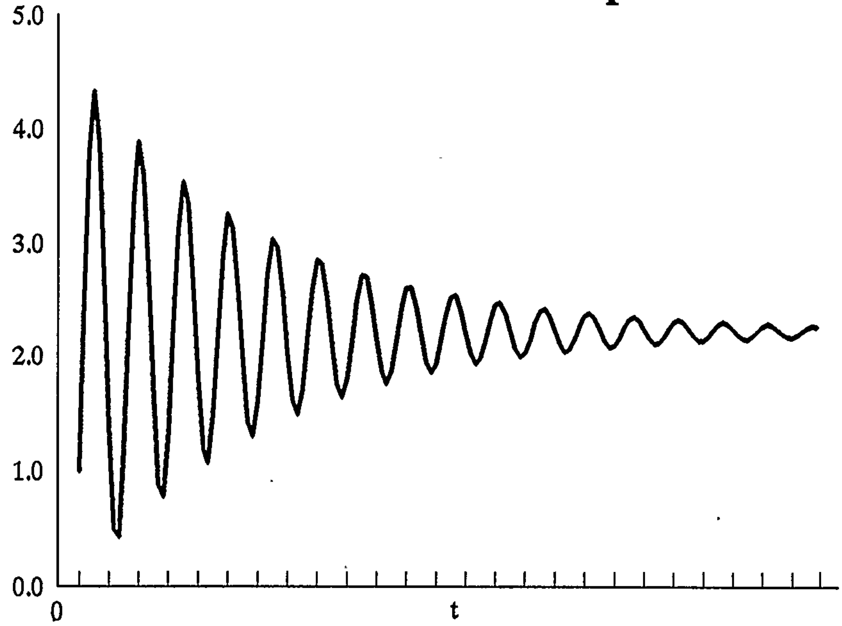
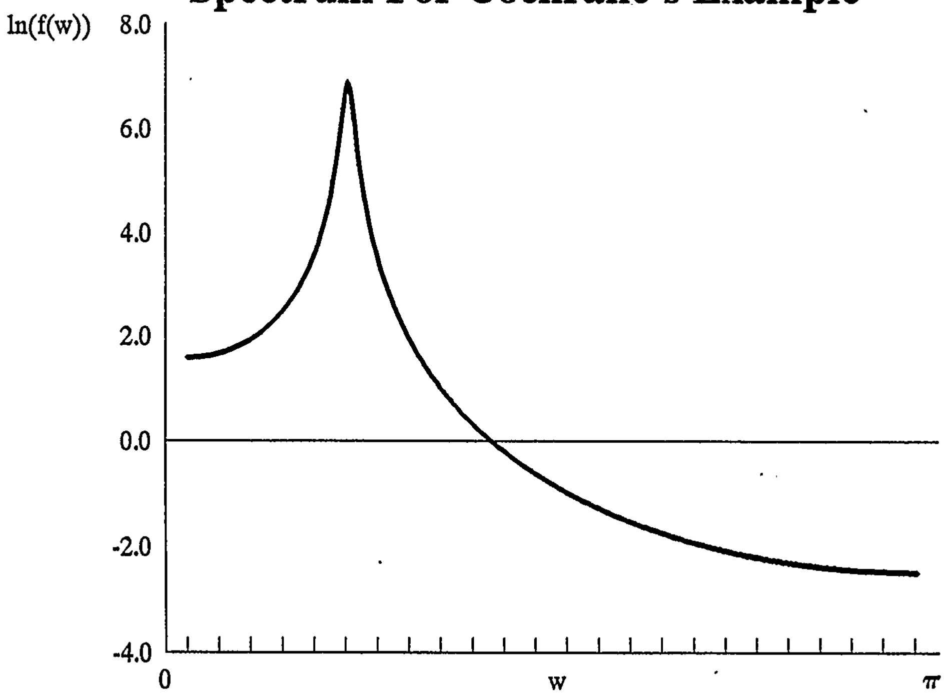


Figure 4
Spectrum For Cochrane's Example



Chapter Three

FREQUENCY DOMAIN ANALYSIS OF COMOVEMENTS AND COINTEGRATION

This chapter develops the basic techniques for examining relationships between time series in the frequency domain. Section A introduces the cross spectrum, along with the two principal means by which it is interpreted: the coherency and phase angle. Section B discusses Phillips and Ouliaris' test for cointegration, which detects relationships between the shocks which drive cointegrated series through the use of the sample spectral density matrix. As was the case in the preceding chapter, emphasis is placed on the relatively intuitive interpretations which can be given to these frequency domain methods.

A. The Cross Spectrum: Estimation and Interpretation

The cross spectrum is simply a generalization of the autospectrum which was introduced in chapter two: it provides a means for characterizing the covariance between specific frequency components of stochastic processes. The cross spectrum of two stationary processes X_t and Y_t is defined as the discrete Fourier transform of the cross-covariance function:

$$f_{xy}(w)= \sum_{k=-\infty}^{\infty} \gamma_{xy}(k)e^{-ikw}, \qquad (32),$$

where $\gamma_{xy}(k)$, the cross-covariance function, is given by:

$$\gamma_{xy}(k) = E[(X_t - E(X_t))(Y_{t-k} - E(Y_t)] \qquad (33)$$

$$= \mathrm{Cov}(X_t, Y_{t-k})$$

Comparing (32) and (3), it is apparent that the spectrum of a stochastic process is simply the cross spectrum of the process with itself.

In the form of (32), the cross spectrum is somewhat difficult to interpret, because unlike the spectrum of a single process, the cross spectrum is generally complex-valued at each frequency w. The reason for this is that whereas the autocovariance function is even, causing the complex sine terms in (3) to cancel one another, in general $\gamma_{xy}(k) <> \gamma_{xy}(-k)$. Evidently, further manipulation of (32) is necessary in order to derive some meaningful expressions.

The usual method for interpreting $f_{xy}(w)$ begins with its decomposition into real and imaginary parts:

$$f_{xy}(w) = \sum_{k=-\infty}^{\infty} \gamma_{xy}(k)\cos(kw) - i\left[\sum_{k=1}^{\infty} (\gamma_{xy}(k) - \gamma_{xy}(-k))\sin(kw)\right]$$
$$= co(w) \qquad -iqu(w) \qquad (34)$$

The function co(w) is called the co-spectrum, while qu(w) is known as the quadrature spectrum. Equation (34) allows the polar representation of the complex-valued expression for $f_{xy}(w)$ to be obtained:

$$f_{xy}(w) = ca(w)\, e^{iph(w)}, \qquad (35)$$

where $ca(w) = \sqrt{co^2(w)+qu^2(w)} = \|f_{xy}(w)\|$ is called the cross-amplitude spectrum and $ph(w) = \arctan(-qu(w)/co(w))$ is the phase spectrum. Expression (35) is just the familiar polar form of the complex number $f_{xy}(w)$ at any given frequency w.

The cross-amplitude spectrum gives a measure of the covariance between specific frequency components of the two processes under consideration, just as the spectrum of a univariate process gives a measure of the variance accounted for by each frequency. Since covariance is not independent of the particular units used to measure each process, the usual practice is to normalize the cross-amplitude spectrum by dividing it by the amplitudes of the X_t and Y_t spectra:

$$coh(w) = \frac{ca(w)}{\sqrt{f_x(w)f_y(w)}} = \frac{\|f_{xy}(w)\|}{\sqrt{\|f_x(w)\|\,\|f_y(w)\|}} \tag{36}$$

Expression (36) gives the coherency between the processes X_t and Y_t at frequency w, which can be interpreted as a correlation coefficient between aligned frequency components. Like a correlation coefficient, the value of coh(w) can range between zero and one. The higher is the coherency, the greater is the tendency for the two processes to move together at the frequency in question. In a similar manner, the squared coherency, $coh^2(w)$, serves the function of an R^2 statistic, showing the proportion of the variance of one process at frequency w which is accounted for by the variance of the other process.

The phase spectrum, ph(w), gives information pertaining to the lag-lead relationship between the processes X_t and Y_t. For example, if two processes are related as follows:

$$Y_t = -BX_{t-l} + \varepsilon_t, \tag{37}$$

it can be shown (Diggle, p. 213) that if B is positive and l is zero, the phase angle spectrum has a constant value of π for all frequencies w. This simply means that reversing the sign on the X_t process produces a new process Y_t which is completely out of phase: at each frequency w, the process Y_t lags X_t by $(ph(w)/w) = (\pi/w)$ time periods. If B is negative and l is non-zero, then $ph(w) = lw$; the Y_t process lags X_t by $(ph(w)/w) = l$ periods at each frequency, and is completely out of phase for cycles of length 2l, corresponding to a frequency of π/l.

The value of the phase angle can range between $-\pi$ and π, representing a lead or lag of up to one-half of a cycle. Unfortunately, there is an inherent ambiguity in the phase angle spectrum: unless sound theoretical reasons can give a priori grounds for expecting one series to have a particular temporal relationship with another, one cannot distinguish between, for example, a lag of one-quarter of a cycle and a lead of three-quarters of a cycle. So while the phase angle allows one to determine the degree of synchronization between series, it is only in rare circumstances that this measure alone will allow definite conclusions to be drawn regarding the temporal order of series' movements.

Estimating the cross-spectrum for the purpose of calculating the coherency and phase spectrum involves the same procedure as in the univariate case. A smoothed cross-spectral estimate can be derived either by substituting sample cross correlations into (34) and using a lag window, or by calculating the cross-periodogram directly from the data series and then smoothing the ordinates (both the real and complex parts) appropriately. Dividing each ordinate of the smoothed cross-periodogram by the square root of the product of the smoothed univariate periodograms, then taking the polar decomposition of the resulting spectral estimates at each frequency, gives the estimated coherencies and phase angles.

Unfortunately, constructing confidence intervals about the point estimates of the coherency and phase angle is not as simple a procedure as the χ^2 approximation used in the univariate case. This is because the variance of these statistics are functions of the true, unobservable population moments. In the case of coherency, the variance of the smoothed estimator $\hat{coh}(w)$ is given by (Diggle, 1990):

$$\mathrm{Var}\,(\hat{coh}(w)) = c_p\,[1-(coh(w))^2]^2, \qquad (38)$$

where c_p is a parameter which depends on the precise weights used in the spectral window. For a rectangular window which takes a simple average of (2p+1) ordinates, c_p is given by 1/(2(2p+1)).

Often, the difficulty in using (38) leads researchers to utilize crude confidence intervals based on Chebychev's inequality and the null hypothesis that the true coherency is zero. However, Jenkins and Watts (1968,ch9) provide a procedure for constructing confidence intervals about any estimated coherency. They note that if the random variable X has variance var(X), then the variance of a function g(X) of that random variable is given by:

$$\mathrm{Var}\,[g(X)] \approx \left(\frac{dg}{dx}\right)_u^2 \mathrm{Var}\,(X), \qquad (39)$$

where the derivative dg/dX is evaluated at the mean, u, of the random variable X. Consider the variance of the transformed coherency estimate arctanh (côh(w)), where arctanh refers to the inverse hyperbolic tangent function. Using (38), (39), the fact that d/dx (arctanh(x)) = $1/(1-x^2)$, and the fact that the smoothed coherency estimator côh(w) is asymptotically unbiased, so E(côh(w)) ≈ coh(w), we see that:

$$\mathrm{Var}\,[\mathrm{arctanh}\,(\hat{\mathrm{coh}}\,(w))] = c_p \frac{[1-(\mathrm{coh}\,(w))^2]^2}{1[1-(\mathrm{coh}\,(w))^2]^2} = c_p \qquad (40)$$

Therefore, the transformed coherency estimator has a variance which is independent of the true population value coh(w).

In order to utilize (40), one must assume a distribution for the transformed point estimate of côh(w). Jenkins and Watts recommend that this distribution be assumed normal, since variance stabilizing transformations such as the one applied above tend to produce probability density functions which are closer to the normal distribution than those of the untransformed variables, (Jenkins and Watts, p77). Also, since the smoothed coherency estimate involves the sum of a number of independent random variables (the periodogram ordinates), the central limit theorem can be invoked to suggest that the assumption of normality may be appropriate. In order to use (40), then, one simply takes the inverse hyperbolic tangent of the point estimate côh(w), then constructs the following 100(1-a)% confidence interval for the transformed variable:

$$\mathrm{arctanh}\,(\hat{\mathrm{coh}}(w)) \pm z_{a/2} \sqrt{c_p} \qquad (41)$$

Notice that the confidence intervals for the transformed variable are symmetric about the point estimate, and that since the function arctanh(x) has range $(-\infty, \infty)$, the upper confidence bound for coh(w) cannot exceed unity. This means that the confidence intervals obtained from the inverse transform of (41) will be narrower, the larger is the point estimate $\hat{coh}(w)$.

In the case of the phase angle, the variance of the smoothed estimator is given by:

$$\mathrm{Var}\,(\mathrm{ph}(w)) = c_p\,[1/(\mathrm{coh}(w))^2-1], \qquad (42)$$

where c_p is as defined above. There is no simple transformation which allows the distribution of ph(w) to be approximated very accurately (Jenkins and Watts, p380). Instead, the usual practice is to substitute the point estimate of coh(w) into (42), then construct extremely approximate 95% confidence intervals by adding and subtracting twice the estimated standard errors from (42) to the point estimate of the phase angle.

The principal advantage which cross-spectral techniques present for the analysis of GDP time series is the same as in the univariate case: this approach allows various frequencies, corresponding to cycles of particular lengths, to be examined individually. This is a very desirable feature, since it is very possible that GDP series may be closely correlated at frequencies associated with business cycles, while over longer periods they may follow approximately independent paths. The opposite may also be true in some cases : while long-run movements in GDP growth may be transmitted from country to country, there may be little reason for business cycles not to be idiosyncratic, particularly for economies with comparatively small foreign trade sectors. Calculating coherencies and phase angles for a broad range of frequencies may allow complicated linkages between countries' GDP growth paths to be characterized more accurately.

B. A Frequency Domain Cointegration Test

Phillips and Ouliaris'(1988) test for cointegration is based on the idea that in any system of cointegrated processes, the shocks which drive the individual series cannot be completely independent, but in the long run must be constrained by the requirement that the series' levels not continually diverge. The same forces which tend to couple the series' movements together over time will impose a relationship upon the shocks driving the system. This dependence should in principle be discernable in the sample variance-covariance matrix of the shocks. Specifically, if some linear relationship exists between any of the system's innovation sequences, then the resulting variance-covariance matrix should have less than full rank.

A complicating factor in devising a test for cointegration from this basic concept is the fact that cointegration is a statistical phenomenon which addresses long run, steady-state behavior of economic variables(Phillips and Ouliaris, 1988). In the short run, cointegrated processes can be expected to exhibit a wide variety of dynamics; these short run dynamics could tend to mask the underlying relationship between the series' levels. This suggests a possible role for spectral analysis: if cointegration manifests itself in the longer-term dynamics of variables, then it should be possible to use the frequency domain to filter out the obscuring high-frequency noise from the system in order to make the long run relationships more conspicuous.

A more formal statement of the consequences of cointegration for the shock sequences driving a stochastic process has been given by Phillips (1986). He proves that a necessary condition for cointegration among processes is that the spectral density matrix of the innovation sequence have less than full rank at the origin. The spectral density matrix contains the univariate spectra along the diagonal, and cross-spectra in the off-diagonal elements. For an arbitrary frequency w, then, the spectral density matrix will not be symmetric: instead, the entries above and below the diagonal will be complex conjugates. However, at frequency zero, the matrix is symmetric, and all of its entries will be real-valued, as can be seen by evaluating (32) at w = 0.

As Cogley (1990) observes, cross-spectra evaluated at frequency zero are proportional to the covariance between the long growth cycles of the series being considered. When combined with the similar meaning of the univariate spectra, this motivates the interpretation of the spectral density matrix as a form of variance-covariance matrix, and provides an intuitive explanation of the reason for the connection between cointegration and this matrix's deficit rank. Another interpretation in the bivariate case can be derived by setting the determinant of the spectral density matrix at frequency zero equal to zero:

$$\begin{vmatrix} f_x(0) & f_{yx}(0) \\ f_{xy}(0) & f_y(0) \end{vmatrix} = 0$$

$$f_x(0)\, f_y(0) - \|f_{xy}(0)\|^2 = 0$$

$$[coh_{xy}(0)]^2 = 1 \qquad (43)$$

Therefore, a necessary condition for the cointegration of two I(1) series is that their innovation sequences be perfectly correlated at frequency zero.

Phillips' deficit rank theorem suggests that a possible test for cointegration would be to estimate the spectral density matrix at frequency zero, then perform some diagnostics to determine if the sample matrix was sufficiently close to being singular. The theory necessary to perform this test is developed by Phillips and Ouliaris (1988). They propose that the eigenvalues of the sample matrix be examined, since the presence of cointegration implies that at least one eigenvalue of the spectral density matrix be zero. In fact, in a multivariate framework, the number of null-valued latent roots is equal to the number of cointegrating vectors, and the corresponding eigenvectors are the cointegrating vectors themselves(Phillips and Ouliaris, 1988). Phillips and Ouliaris recommend the following smoothed estimator for the spectral density at frequency zero:

$$f_{xy}(0) = \frac{1}{k} \sum_{s=1}^{k} \mathrm{Re}(I_{xy}(\tfrac{2\pi s}{n})), \tag{44}$$

where I_{xy} is the cross-periodogram between growth series X_t and Y_t (which, in the case of the diagonal elements, will refer to the same series). Equation (44) is simply a rectangular one-sided spectral window. The reason for the use of a one-sided window is that both the univariate and cross-periodograms possess a form of symmetry about frequency zero: the ordinates at the negative frequencies are the complex conjugates of the corresponding ordinates at the positive frequencies. Of course, the univariate periodograms are entirely real-valued, so ignoring the complex parts of the ordinates has no effect. In contrast, the complex parts of the cross-periodograms would cancel if a two-sided window were used. For simplicity, Phillips and Ouliaris use a one-sided window and ignore the cross-periodograms' imaginary components. They provide theoretical arguments for the use of a rectangular window rather than some alternative weighting scheme.

Once the estimate of the spectral density matrix has been obtained, some suitable confidence interval must be computed, in order to test the null hypothesis of no cointegration. Phillips and Ouliaris develop several alternative procedures, the most appropriate in this case being the following 100(1-a)% upper confidence bound for the ratio of the smallest m+1 latent roots to the sum of all n eigenvalues:

$$\frac{\sum_{j=n-m}^{n}\lambda_j}{\sum_{j=1}^{n}\lambda_j} \leq \frac{\sum_{j=n-m}^{n} l_j}{\sum_{j=1}^{n} l_j} + \frac{z_a B,}{\sqrt{K}}$$

where B
$$= \frac{\sqrt{\left(\sum_{i=n-m}^{n} l_i\right)^2 \left(\sum_{i=1}^{n-m-1} l_i^2\right) + \left(\sum_{i=1}^{n-m-1} l_i\right)^2 \left(\sum_{i=n-m}^{n} l_i^2\right)}}{\left(\sum_{i=1}^{n} l_i\right)^2} \qquad (45)$$

Expression (45) uses the fact that under the null hypothesis Ho:$\lambda_1 > \lambda_2 > \ldots \lambda_n > 0$, where the λ_i's are the eigenvalues of the true, underlying spectral density matrix, the expression $\sqrt{k}(l_i - \lambda_i)/\lambda_i$ is asymptotically normally distributed. If the upper bound in (45) is sufficiently small, one can conclude that the smallest m+1 eigenvalues are negligible, which can be interpreted as evidence of cointegration. Perhaps somewhat arbitrarily, Phillips and Ouliaris suggest the critical value 0.1/n in the case where m=0 and the test focuses on the smallest latent root. If the upper bound is less than this critical value, then one can be approximately 100(1-a)% confident that the smallest eigenvalue of the spectral density matrix is less that 10% of the average of the latent roots, which Phillips and Ouliaris recommend be interpreted as strong evidence of cointegration.

Phillips and Ouliaris also suggest that the above test can be applied in order to support the null hypothesis of no cointegration. In this case, the following lower confidence interval bound is compared to the critical value:

$$\frac{\sum_{j=n-m}^{n} l_j}{\sum_{j=1}^{n} l_j} \frac{-z_a B}{\sqrt{k}} \leq \frac{\sum_{j=n-m}^{n} \lambda_j}{\sum_{j=1}^{n} \lambda_j} \tag{46}$$

If the lower bound exceeds the chosen critical value, then the sample series give considerable support to the null hypothesis, since one can be approximately 100(1-a)% confident that the smallest eigenvalue is larger than 10% of the average of the latent roots.

The idea that cointegration among stochastic processes restricts the dynamics of the processes involved in a manner which affects the rank of relevant matrices has, of course, been exploited by other cointegration tests: for example, it is the motivation behind Johansen's (1988) procedure. In this time-domain test for cointegration, the matrix of interest is derived from the following multivariate model:

$$Y_t = A_1 Y_{t-1} + A_2 Y_{t-2} + \ldots A_k Y_{t-k} + \varepsilon_t \tag{47}$$

This equation can be re-written as (Dickey, Jansen, and Thornton, 1991):

$$\Delta Y_t = \Gamma_1 \Delta Y_{t-1} + \Gamma_2 \Delta Y_{t-2} + \ldots \Gamma_{k-1} \Delta Y_{t-k+1} + \Pi Y_{t-k} + \varepsilon_t \tag{48}$$

where Γ_i=-(I-A_1- ... -A_i) for i=1.. k-1, and Π=- (I-A_1-A_2 ... -A_p) is called the impact matrix. It can be shown that if Π is of full rank, then the vector process Y_t is stationary, and so is every linear combination of its elements. If, however, Π is of deficit rank, then only some linear combinations of Y_t's constituent processes are stationary, namely, those processes formed by the inner product of Y_t with one of the cointegrating vectors of the system. The rank of Π corresponds to the number of linearly independent cointegrating vectors of the Y_t process, so that if Π is the null matrix, every linear combination of Y_t's components will be nonstationary. As in Phillips and Ouliaris' procedure, the restrictions on the process' dynamics caused by cointegration relate closely to the form of linear dependence among the rows of a matrix which characterizes an aspect of the system's dynamic behaviour. Johansen (1988) develops a test to determine if the coefficient matrix Π arising from a particular sample is close (in a statistical sense) to being of a specified rank by determining if the eigenvalues of Π are significantly different from zero (Dickey, Jansen, and Thornton, 1991).

Phillips and Ouliaris' cointegration test is best interpreted as a variation on the commonly used method of principal components analysis, a technique which allows complex systems of related variables to be described in terms of smaller sets of components composed of linear combinations of the original variables. When applied to a stationary vector process X with components $(x_1, x_2 \dots x_n)$, principal components analysis involves a search for linear combinations of the random variables x_i which enable the variance of the entire n-dimensional X process to be accounted for by a smaller number of components. This effectively reduces the dimension of the problem (the number of variables being considered) without sacrificing the information pertaining to the original variables contained in the system's variance-covariance matrix.

If the system X is assumed to possess a variance-covariance matrix denoted by Ω, then the first principal component, $a_1'X$, where a_1 is an n by 1 vector of constants, is determined by solving the following problem:

$$\text{Max } a_1' \Omega a_1 \qquad \text{Subject to: } a_1' a_1 = 1 \tag{49}$$

The motivation behind the formulation of expression (49) lies in the fact that the variance of the new process $a_1' X$ is given by $a_1' \Omega a_1$. This variance is therefore maximized over all n by 1 vectors a_1, with the normalization constraint that the norm of a_1 be unity.

Forming a Lagrangian function from (49) shows that the problem of selecting the principal component $a_1'X$ is equivalent to determining the largest eigenvalue of the variance-covariance matrix Ω:

$$\text{Max } L = a_1' \Omega a_1 - \lambda_1 [a_1' a_1 - 1] \qquad (50)$$

The first-order conditions from 50 give:

$$2\Omega a_1 - 2\lambda_1 a_1 = 0, \quad \text{or:}$$

$$\Omega a_1 = \lambda_1 a_1 \qquad (51)$$

Therefore, a_1 is the eigenvector corresponding to the largest eigenvalue of the matrix Ω. The variance of the first principal component, $a_1' \Omega a_1$, is equal to the eigenvalue λ_1.

Successive components $a_i'X$ are chosen in an analogous manner, with the additional constraint that $a_i' a_j = 0$, so that the a_is are orthogonal to one another. This ensures that the principal components are mutually uncorrelated, since the covariance between two components $a_i'X$ and $a_j'X$ is given by:

$$\begin{aligned} \text{Cov}(a_i' X, a_j' X) &= E(a_i' XX a_j) \\ &= a_i' E(XX) a_j \\ &= a_i' \text{Var}(X) a_j \\ &= a_i' \Omega a_j \\ &= \lambda_i a_i' a_j \\ &= 0 \text{ if } a_i' a_j \text{ is zero.} \end{aligned} \qquad (51)$$

This method of defining principal components ensures that the first principal component $a_1'X$ explains the maximum possible proportion of the system's variance. The second principal component is the normalized linear combination of the original variables having maximum variance out of all linear combinations uncorrelated with the first principal component. Subject to the orthogonality condition, this second component accounts for a maximum of the variance of the system which remains after the removal of the first component's effects. Using the identity $Tr(\Omega) = \Sigma \lambda_i$, it is clear that together, all the principal components partition the total variance of the variables which make up the X process. If $\lambda_k = 0$ for all $k>K$, $K<n$ then the entire system can be expressed in terms of the smaller number (K) of principal components, rather than in terms of the original variables.

The method of principal components has a very simple geometric interpretation. Consider the scatterplot of the original variables x_i. The eigenvector for the first principal component, a_1, defines a new coordinate axis, an axis which will likely be rotated in relation to the axes representing the original variables. The variance of the data along this axis is greater than for any other choice of axis. Similarly, the eigenvector a_2 defines a coordinate axis passing through the origin which is orthogonal to a_1; the data's variation along this second axis is also maximized. Analogous properties are true of the remaining eigenvectors a_1. If some of the λ_is are null-valued, then the scatterplot of the original variables is clustered in such a way that fewer than n axes are required to completely specify the location of any of the observations. To take a simple example, if n is three and the scatterplot clusters into an elliptical pattern, then a_1 would lie upon the ellipse's major axis, a_2 would correspond to the minor axis, and a_3 would be associated with an eigenvalue of zero. The clustering within the data would allow a smaller number of components to account for all of the variance of the original system.

While the preceding discussion may seem rather remote from practical economic applications, the literature does provide an example which is closely related to the ideas of Phillips and Ouliaris: Lucas' "Two Illustrations of the Quantity Theory of Money" (1980). Lucas' approach to testing this proposed long-run, equilibrium relationship involved applying a frequency-domain filter weighted heavily towards the lower frequencies to data series for money supply and price level growth. The effect of this filtering was to produce series with almost all of their variation occurring in the neighborhood of frequency zero. Lucas interpreted this filter as a screening agent which purged the growth series of short-run noise. After filtering, Lucas examined scatterplots of the transformed variables, and noted that the points in the plots tended to cluster about the 45 degree line as the amount of filtering was increased. The implication, Lucas argued, was that there was an underlying linear relationship between the growth rates of money and prices in the long run, a relationship consistent with quantity theory predictions.

The connection between Lucas' work and Phillips and Ouliaris' cointegration test is very straightforward. Consider the variance-covariance matrices for Lucas' filtered money and price level growth series. As the series were filtered more extensively, the increasingly conspicuous linear relationship between them would produce variance-covariance matrices which approached the point of being singular: their smallest eigenvalue would converge to zero. The first principal component of the filtered observations would be (approximately) $1/\sqrt{2}$ (monetary growth + price level growth). This component, corresponding to the eigenvector $(1/\sqrt{2}, 1/\sqrt{2})$, would account for all but a negligible fraction of the variance of the transformed system.

Of course, Phillips and Ouliaris' test does not involve an explicit filter, nor is the variance-covariance matrix calculated. However, the effect of confining attention to the first few periodogram ordinates is analogous to the Lucas filter, and as mentioned previously, the spectral density matrix can be interpreted as a form of variance-covariance matrix. Therefore, the fundamental concepts underlying the two techniques are very similar.

The empirical tools introduced in this chapter provide the means to analyse in very general terms the relationships between GDP time series. With estimates of coherencies and phase spectra, the degree to which movements in one nation's level of economic activity is associated with corresponding fluctuations in others can be assessed. This assessment can distinguish between short-run, business cycle dynamics and low frequency, long run correlations in GDP growth cycles. It can also be applied when constructing parametric models of the comovements in GDP time series (Engle, 1976), although it is not employed for that purpose here. Phillips and Ouliaris' test for cointegration is a natural extension of cross-spectral analysis, providing a more formal method for investigating the possibility that long-term patterns of economic growth are shared by certain countries which may exhibit highly coherent low-frequency economic cycles.

Chapter Four

EMPIRICAL RESULTS

As mentioned in the introduction, the aims of this thesis are to address the following issues:

i) One of Cogley's motivations for introducing his frequency domain estimator of the variance ratio was to enable more reliable confidence intervals to be computed. This leaves open the questions of whether this approach can supply more precise (i.e., narrower) ranges between upper and lower confidence bounds, and if the estimates obtained are sensitive to the choice of the value of the parameter k in equation (12). In order to investigate the effect of varying this parameter, this chapter presents estimates of V_f and the corresponding confidence intervals while k is allowed to take on a large number of values over a fairly broad interval.

ii) Cogley applied Phillips and Ouliaris' cointegration test to a group of nine countries, and obtained evidence that some cointegration was present. In this chapter, this analysis is extended to a pairwise comparison, examining all 36 combinations of the countries in Cogley's data set for possible cointegration. In addition, the relationship between the GDP growth paths of several pairs of countries which were judged to be interesting a priori are studied in more detail by computing the coherencies and phase angles for these combinations of series across the entire spectrum.

A. Data And Software Used

The data employed are the same utilized by Cogley in his development of the frequency domain estimator of the variance ratio. The data set used in the computation of coherencies and phase angles, as well as for the application of Phillips and Ouliaris' cointegration test, consists of real per capita GDP growth for the period 1871 to 1985 for nine countries: Australia, Canada, Denmark, France, Italy, Norway, Sweden, the United Kingdom, and the United States. For the United Kingdom, real annual per capita GDP growth for the period 1831-1985 is used for the calculation of the variance ratio statistic; for all other countries, estimates of the variance ratio are based on data for the period 1871-1985. The author is indebted to Cogley for supplying this data set.

All the periodograms and cross-periodograms necessary for the calculation of the variance ratio statistic and the spectral density matrices were estimated using RATS version 2.0.

B. The Variance Ratio Statistic

In order to investigate the sensitivity of $\hat{V}_f$ to changes in the value of the parameter k in (12), some reasonable interval must be selected over which the parameter will be varied. In chapter two, Cochrane's approach to this problem was outlined. On the basis of this approach, he recommends a value of 15 to 20; these same values are used by Cogley. Clearly, however, there is no single correct choice, with the final selection depending on many factors, including the overall behavior of the estimator for different values of k (which will be sample-specific) and the subjective judgement of how best to resolve the trade-off between greater bias and smaller variance.

The choice of k is representative of the inherent difficulty in measuring persistence as defined by the variance ratio, since the selection of this parameter in essence requires one to fix specifically the ill-defined concept of the "long run." In order to qualify as a "long run" time span, a k year period must be extended enough so that the temporary effects of business cycles begun in the first year of the span will have dissipated after the time period has elapsed (Cochrane, 1988). This will likely lead to long runs which vary both between countries and historical periods. In order to select a range for k, then, one must rely upon some very general assumptions regarding the process which generates fluctuations in economic activity. Time frames of less than ten years in length are almost certainly too brief in duration, given that an average business cycle can last nearly that long from peak to peak. On the other hand, it seems unlikely that typical shocks to the economy continue to cause dynamic adjustments in GDP (as opposed to a permanent but static change in level) past time horizons of 20-25 years. On the basis of these general considerations, the results reported here were obtained by allowing k to vary from 10 to 25, reflecting the implicit belief that in this context, the "long run" lies somewhere within this interval.

The point estimates and 90 percent confidence intervals for the variance ratio using the value m=15 in equation (14) are presented in figures 5 through 13. This format clearly shows the powerful influence which the value of k has on the size of the confidence intervals obtained: in many cases, the intervals for k=25 are almost three times as wide as those for the lowest value of k considered. Even for the value k = 10, however, the intervals are wide in an absolute sense, with the upper confidence bound being more than three times larger than the lower bound. The sizes of the confidence intervals, as indicated by equation (22), are governed by the equivalent degrees of freedom and the magnitude of the point estimates of $\hat{V}_f$. As a result, the narrowest confidence intervals are those for the United States, Canada, and the United Kingdom: the point estimates of $\hat{V}_f$ are lowest for these countries, while the longer data set used for the United Kingdom's estimates gives a larger value of v for each choice of k. These results are consistent with those obtained by Cogley for the values k = 15 and k = 20.

Perhaps the most important feature of the estimates of $\hat{V}_f$ is their stability in response to variation in the parameter k. If the point estimates of $\hat{V}_f$ do not change substantially as k is varied, then one is more justified in using a low value of k in order to generate the narrowest possible confidence

Figure 5

Variance Ratio Estimates - United States

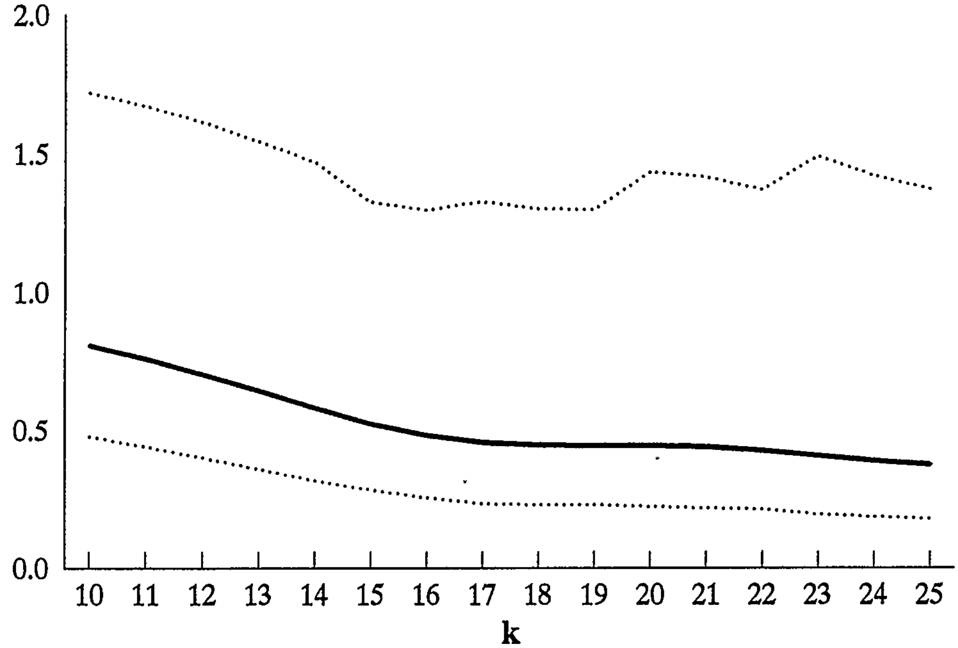
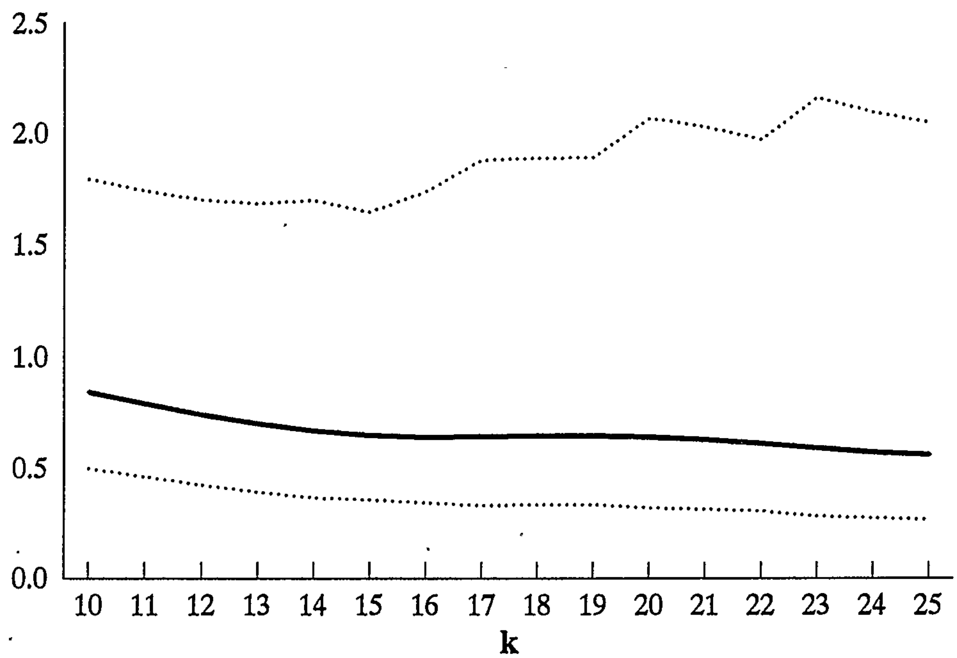


Figure 6

Variance Ratio Estimates - Canada



Point Estimates Confidence Bounds

Figure 7

Variance Ratio Estimates - United Kingdom

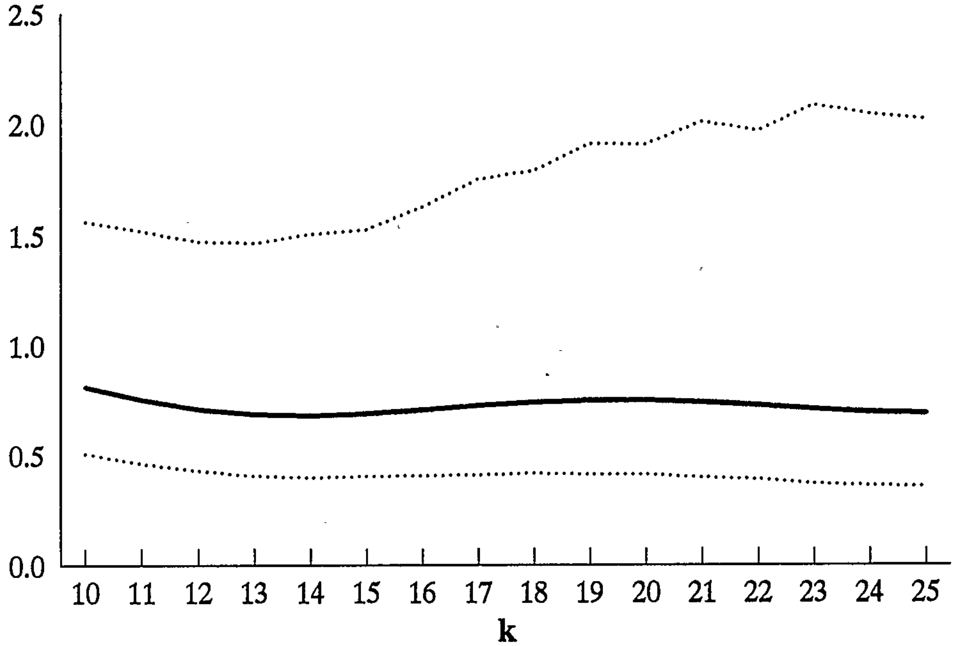


Figure 8

Variance Ratio Estimates - France

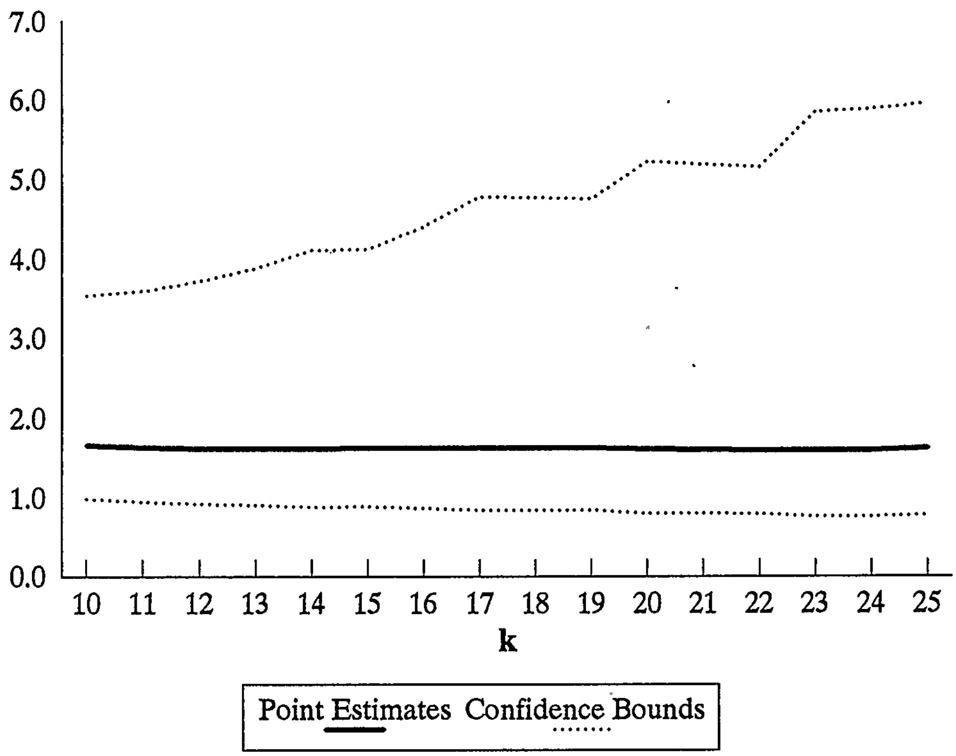


Figure 9

Variance Ratio Estimates - Italy

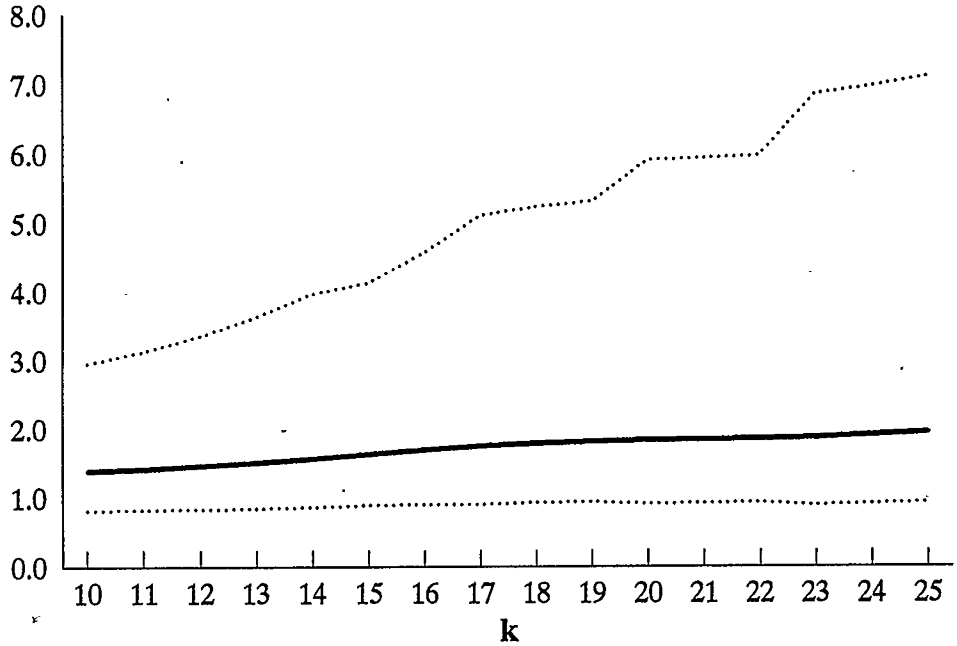
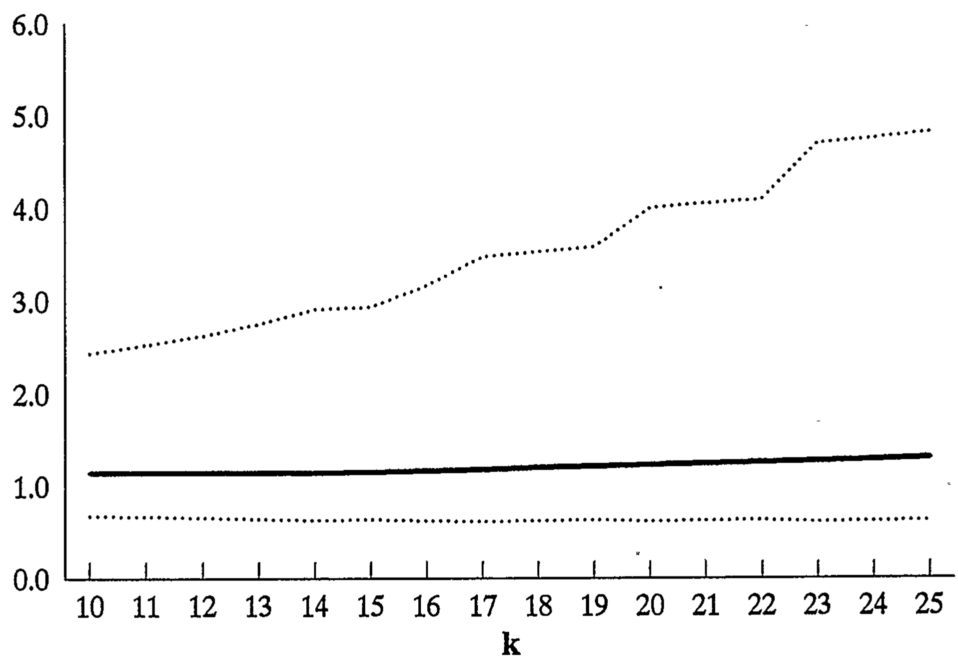


Figure 10

Variance Ratio Estimates - Australia



Point Estimates Confidence Bounds

Figure 11

Variance Ratio Estimates - Sweden

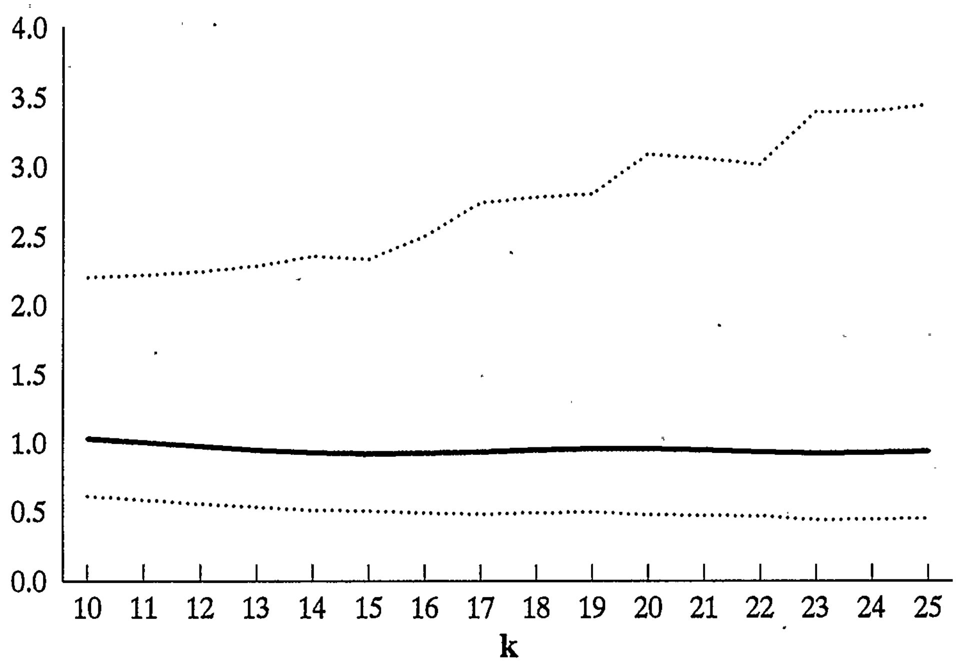


Figure12

Variance Ratio Estimates - Denmark

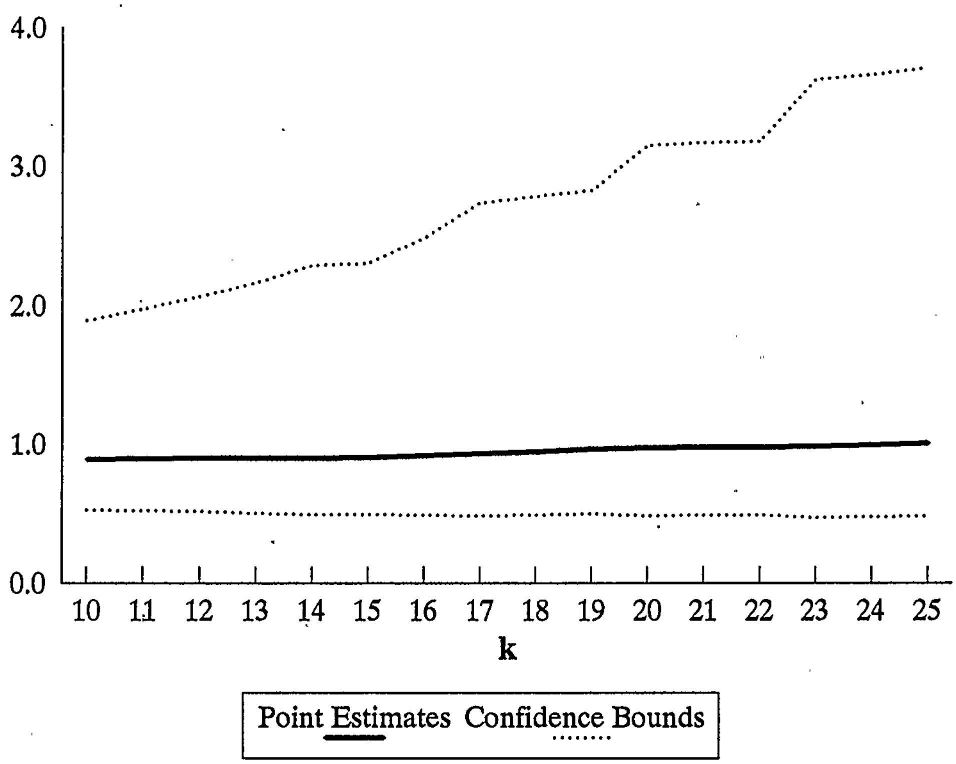
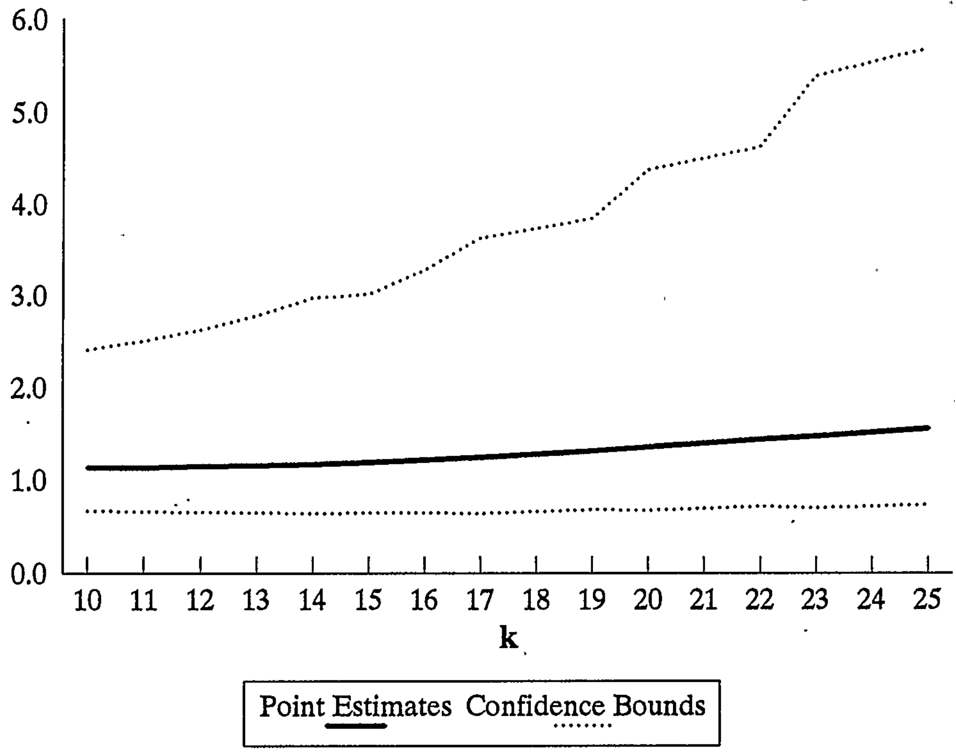


Figure 13
Variance Ratio Estimates - Norway



intervals. If, on the other hand, $\hat{V}_f$ is fairly volatile over the range of k values considered, then it is difficult to defend the use of any particular parameter value, and the broader confidence intervals should be used. Interestingly, perhaps the most variable point estimates of $\hat{V}_f$ (in relative terms) are those for which $\hat{V}_f$ is generally lowest: namely, the estimates for the United States, Canada, and the United Kingdom. In the case of the U.S., $\hat{V}_f$ varies from 0.80 for k = 10 to 0.37 for k = 25; for Canada, the corresponding values are 0.84 and 0.55. The most stable point estimates are those for France and Sweden, which vary over the ranges (1.58,1.65) and (0.92,1.03), respectively. Of course, none of the countries possess point estimates which differ in a statistically significant manner depending on the value of k used, because all the confidence intervals are so broad.

Given the evidence presented in the accompanying figures, it would appear that the size of the random walk component in GDP, or its contribution to the total variation in GDP growth cycles, cannot be determined very precisely for any of the countries examined here. Although the χ^2 approximation to the estimator's distribution may enable one to be more certain that the confidence bounds are accurate from a statistical viewpoint, it can't overcome the fundamental fact that the available data contains very few disjoint "long runs" of reasonable duration. To take but one example, the lowest confidence bound for V in the case of Canada is 0.27, while the highest bound is 2.06. In order to obtain narrower confidence intervals, either a more structured, parametric model would have to be used, or the definition of persistence would have to be abandoned in favor of a more empirically tractable measure.

C. Comovements in GDP Growth Cycles

With data from nine countries available, there are 36 possible combinations of countries for which coherencies and phase spectra could be computed. Of course, the economic relationships between most of these pairs of countries would be of little interest. For this reason, and in order to present the results in a reasonable amount of space, only four pairs of countries are examined in this manner: The United States and Canada, the U.S. and the United Kingdom, the U.K. and Canada, and the U.K. and France. All these countries have been linked by significant trade and/or investment flows over at least some of the time period under consideration, with the connection between the U.S. and Canadian economies being particularly close. One would expect these economic relationships to be manifested in high coherencies, at least over some frequencies.

The coherency and phase angles for the selected pairs of countries are presented in figures 14 to 21. Each figure also includes approximate 95% confidence intervals, computed using the methods outlined in the previous chapter. The confidence intervals for the coherencies have been truncated at zero, while those for the phase angles are not extended past the bounds of $[-\pi,\pi]$. In addition, where there are discontinuities in the estimates of the phase angle (caused by the fact that the arctan function has a range of $(-\pi,\pi)$), the confidence bounds are also interrupted in order to make the presentation easier to interpret. The estimates for both the phase and coherency spectra were computed using rectangular windows of width 11, giving the parameter c_p in (38) and (40) the value 1/22.

Perhaps the most notable feature of the coherencies is that in every case, the highest point estimates are in the low frequency ranges. For each pair of countries except the U.S. and Canada, only in the low frequencies are the point estimates of coherency significantly different from zero over substantial intervals. Except for the U.S. and U.K., the highest coherency estimates are obtained at frequencies between $8/115\pi$ and $18/115\pi$, corresponding to cycle lengths of between approximately 9 and 29 years. This would appear to support the view that growth cycles are more closely related

Figure 14
Coherency Estimates
Canada / United Kingdom

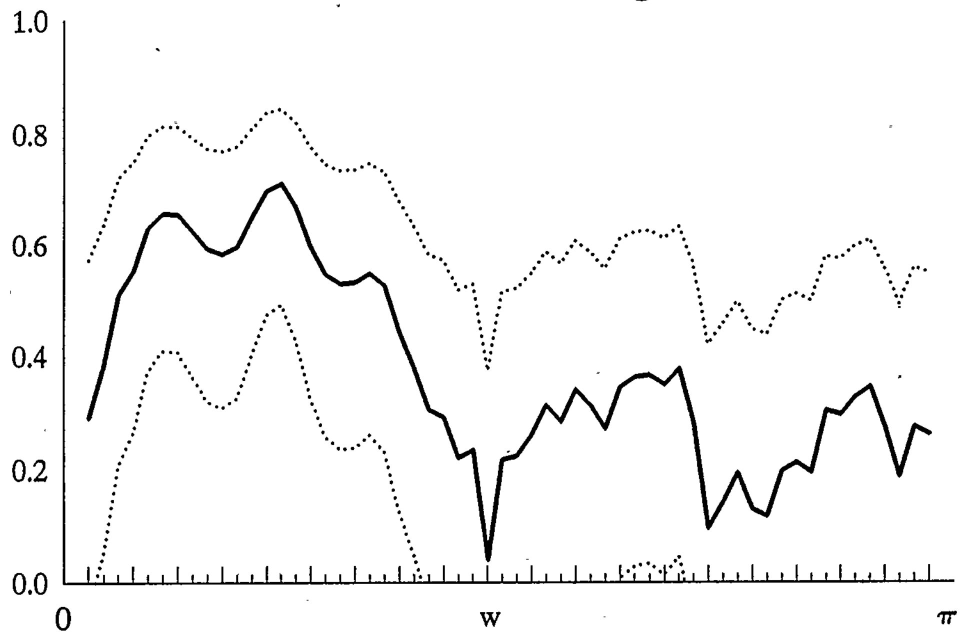
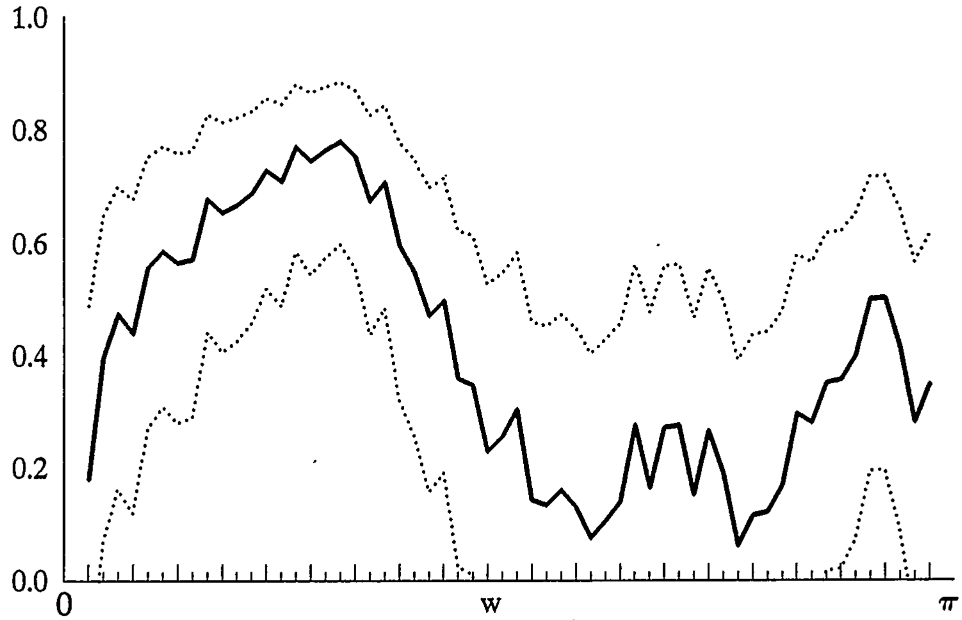


Figure 15
Coherency Estimates
United States / United Kingdom



Point Estimates Confidence Bounds

Figure 16
Coherency Estimates
Canada / United States

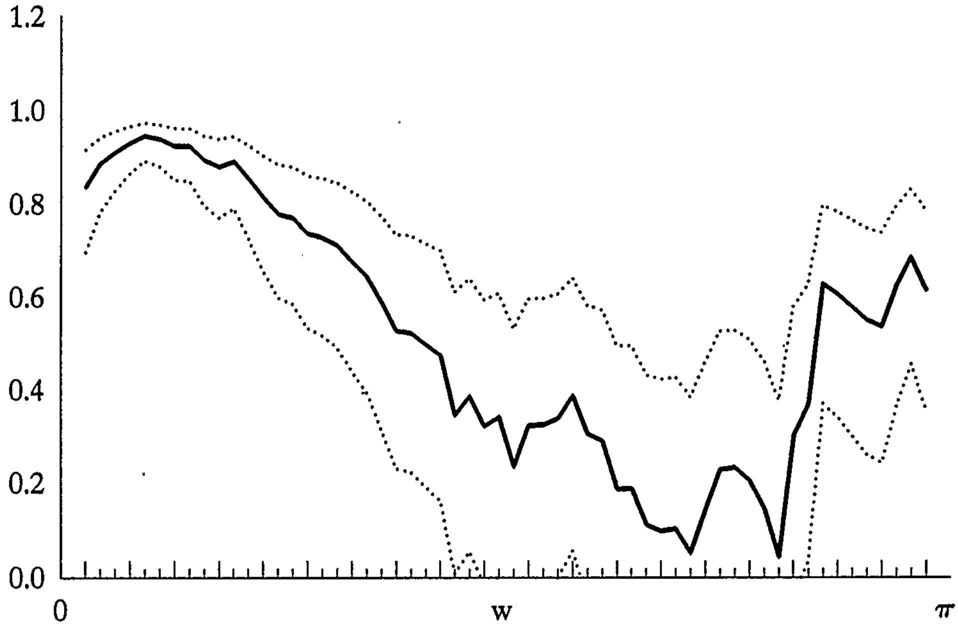


Figure 17
Coherency Estimates
United Kingdom / France

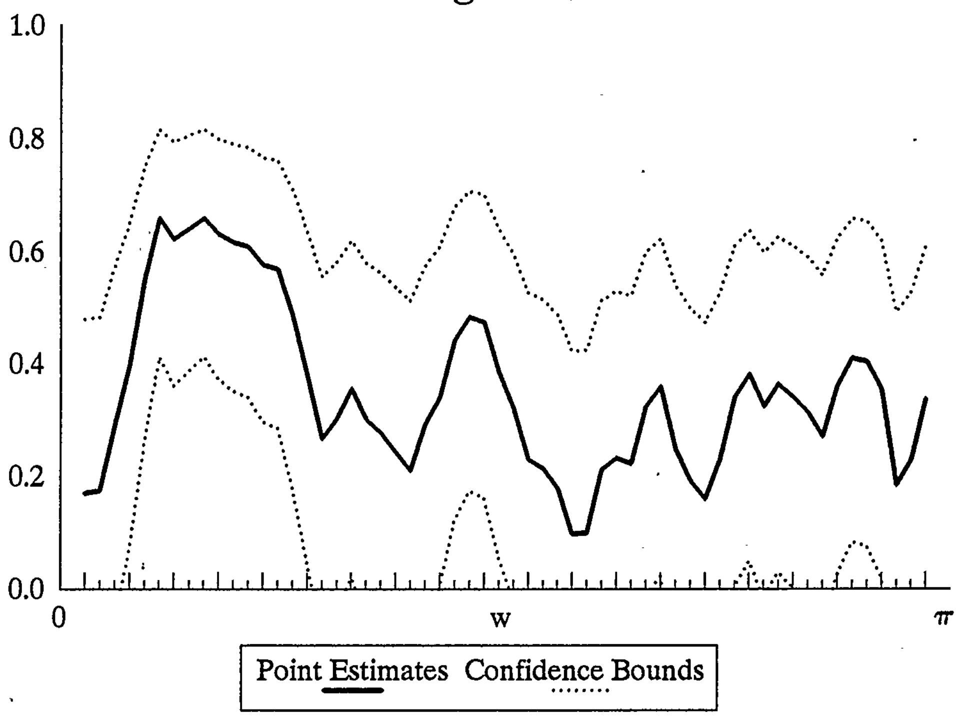


Figure 18
Phase Angle Estimates
Canada / United Kingdom

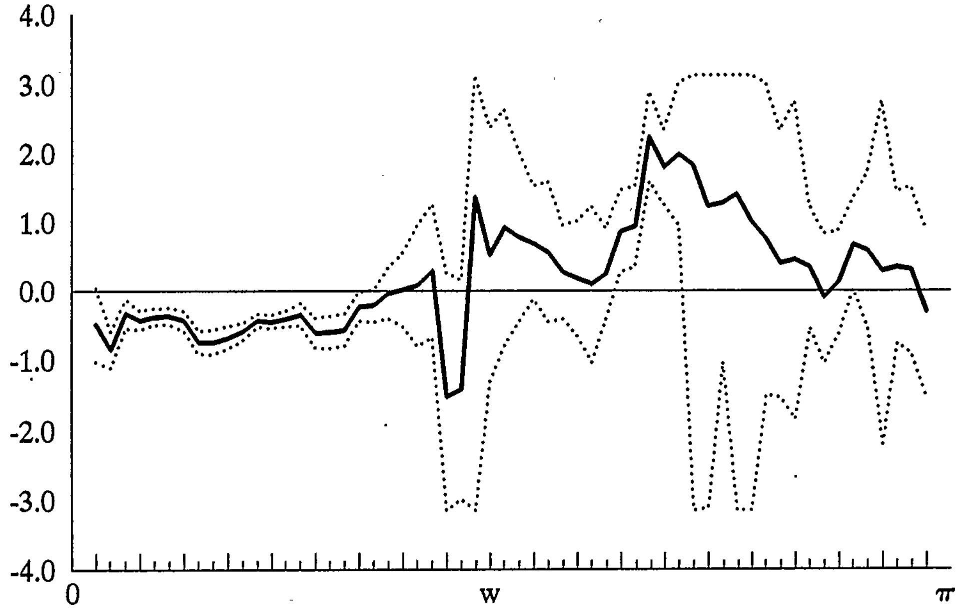


Figure 19
Phase Angle Estimates
United States / United Kingdom

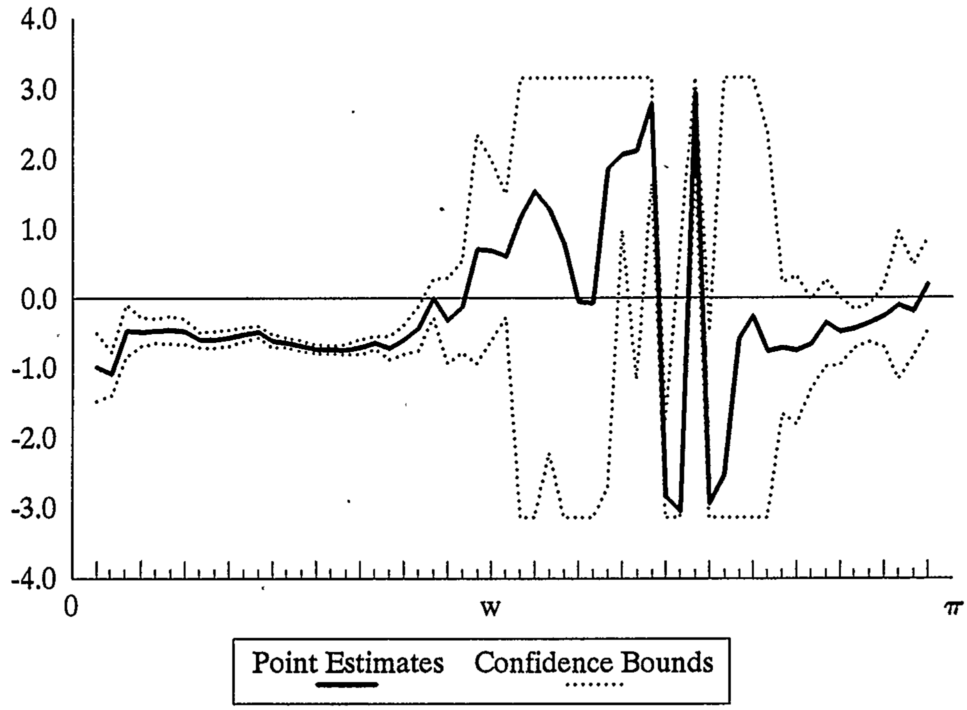


Figure 20
Phase Angle Estimates
Canada / United States

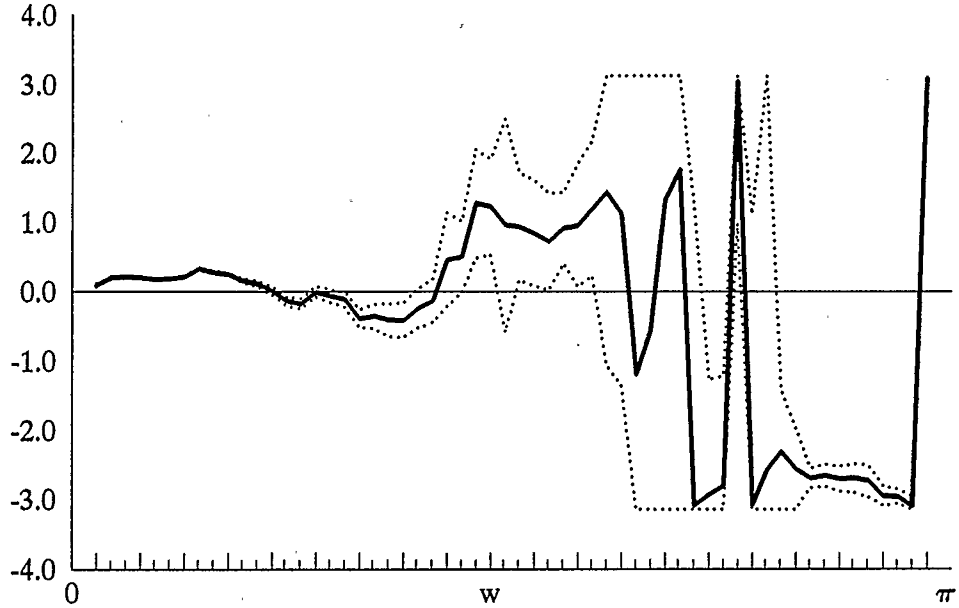
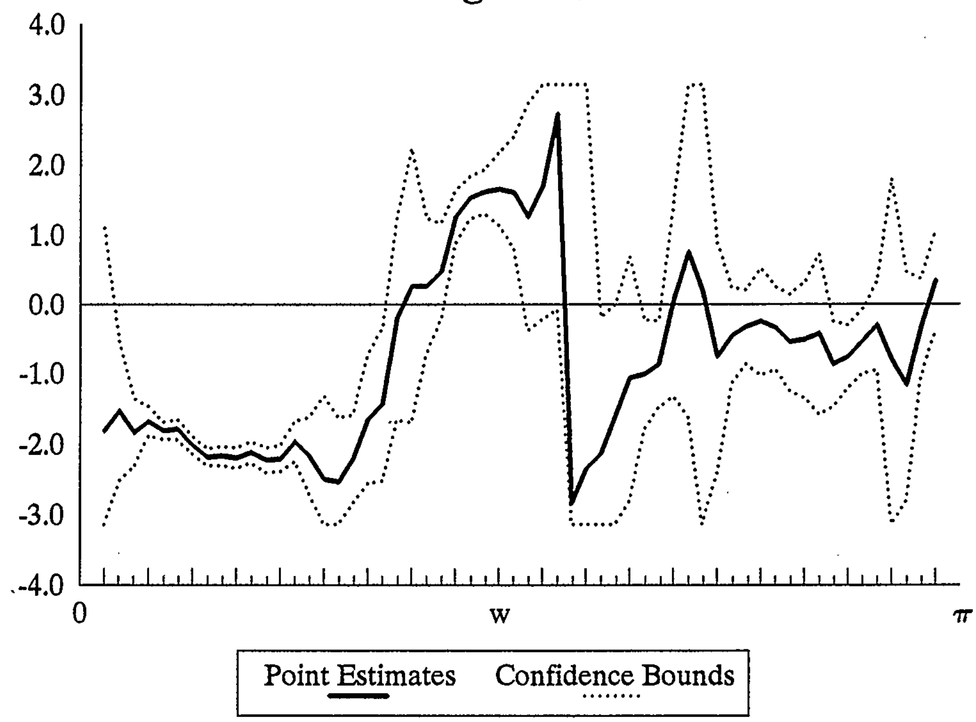


Figure 21
Phase Angle Estimates
United Kingdom / France



over lengthy time periods than for fluctuations occuring over only a few years. A more puzzling aspect of the coherency estimates is that they do not increase monotonically as the frequency of zero is approached: instead, the coherencies in the immediate neighborhood of the zero ordinate are considerably smaller than those between four and eight ordinates away (the U.S.-Canada graph is an exception to this, showing very high coherencies over a very broad band of low frequencies). There is no compelling reason to expect to find this type of result, since cointegration should result in the lowest frequencies being the most closely related.

As was mentioned in chapter three, phase angles are generally very difficult to interpret, even when the suspected dynamic relationship takes a very simple form. It is therefore not at all surprising that little information can be gleaned from the phase angles calculated here. The only feature of interest appears to be that the phase angles for the low frequencies tend to be more stable than those in the higher ranges, suggesting that if there are lag-lead relationships between these pairs of countries, they tend to be more regular and predictable for long growth cycles than for high-frequency fluctuations. This interpretation is, of course, fully consistent with the fact that the coherencies between growth cycles appear to be higher for the lower frequencies.

D. Cointegration Tests

When applying Phillips and Ouliaris' cointegration test to pairs of series, the expressions given in (45) and (46) reduce to:

$$\frac{\lambda_2}{\sum_{i=1}^{2}\lambda_i} \leq \frac{l_2}{\sum_{i=1}^{2} l_i} + \frac{z_a B}{\sqrt{k}}, \text{ and } \frac{l_2}{\sum_{i=1}^{2} l_i} - \frac{z_a B}{\sqrt{k}} \leq \frac{\lambda_2}{\sum_{i=1}^{2}\lambda_i},$$

where $$B = \frac{\sqrt{2\, l_1}\, l_2}{\left(\sum_{i=1}^{2} l_i\right)^2} \qquad (53)$$

These expressions are homogeneous of degree zero in l_1 and l_2, so one can normalize the sample eigenvalues so that $l_1 + l_2 = 1$. This makes very simple the derivation of two "critical" values for the size of the smallest eigenvalue:

$$l_2^{\,crit1} + \frac{z_a B}{\sqrt{k}} = 0.05$$

$$l_2^{\,crit1} + \frac{\sqrt{2 z_a}}{\sqrt{k}}\, l_2^{\,crit1} \left(1 - l_2^{\,crit1}\right) = 0.05$$

$$\frac{\sqrt{2 z_a}}{\sqrt{k}} \left(l_2^{\,crit1}\right)^2 - \left(1 + \frac{\sqrt{2 z_a}}{\sqrt{k}}\right) l_2^{\,crit1} + 0.05 = 0$$

$$l_2^{\,crit2} - \frac{z_a B}{\sqrt{k}} = 0.05$$

$$l_2^{\,crit2} - \frac{\sqrt{2 z_a}}{\sqrt{k}}\, l_2^{\,crit2} \left(1 - l_2^{\,crit2}\right) = 0.05$$

$$\frac{\sqrt{2 z_a}}{\sqrt{k}} \left(l_2^{\,crit1}\right)^2 + \left(1 - \frac{\sqrt{2 z_a}}{\sqrt{k}}\right) l_2^{\,crit2} - 0.05 = 0 \qquad (54)$$

Using the parameter value $k = 5$, the value employed by Cogley and also one of the values used to generate the results reported here, gives the following critical values at the 5 percent level of significance:

$$l_2^{\text{crit1}} = 0.02478$$

$$l_2^{\text{crit2}} = 0.2408 \tag{55}$$

For the value k=10, the corresponding critical values are 0.02515 and 0.2288, respectively. Following Phillips and Ouliaris' suggested procedure, if the smallest (normalized) sample eigenvalue is less than l_2^{crit1}, the data give strong support to the hypothesis of cointegration, while if it is larger than l_2^{crit2}, then there is substantial support for the absence of cointegration.

Table 1 presents the results of Phillips and Ouliaris' cointegration test as applied to the 36 possible pairs of countries which can be formed from the data set. The table is organized as follows: along the diagonal are the estimates of the autospectra at frequency zero; above the diagonal are the cross-spectra; and below the diagonal are reported the smallest normalized eigenvalues of the various sample spectral density matrices. Surprisingly, none of the pairs of data series gives strong evidence in favor of the hypothesis of cointegration: there are no eigenvalues which fall below the lower critical value 0.02478, even for the U.S. and Canada, a case in which the low-frequency coherencies were very large. In 15 of the cases, the smallest eigenvalue is larger than the larger of the two critical values derived above, giving strong support to the hypothesis that no cointegration is present. For the remaining 21 pairs of countries, the smallest eigenvalue lies between the two critical values: for these countries, which include all the pairs examined in more detail above except the combination of the U.S. and U.K., neither the hypothesis of cointegration nor the alternative is supported.

Table 2 presents the same estimates as Table 1 for the value k=10. Once again, none of the eigenvalues are small enough to give substantial support to the hypothesis of cointegration, though for both the United States-Canada and France-Denmark cases, the smallest eigenvalue is

TABLE 1

PHILLIPS AND OULIARIS' COINTEGRATION TEST
(k=5)

	US	CA	UK	FR	IT	AU	SW	DE	NO
US	2.339 x10^{-3}	2.157 x10^{-3}	2.421 x10^{-4}	-7.1616 x10^{-4}	-8.6986 x10^{-4}	1.4202 x10^{-3}	-6.1228 x10^{-5}	3.0196 x10^{-4}	-6.1302 x10^{-4}
CA	*8.0592 x10^{-2}*	2.8382 x10^{-3}	4.3152 x10^{-4}	-3.2104 x10^{-4}	-1.4858 x10^{-4}	1.7738 x10^{-3}	-4.0370 x10^{-4}	2.6140 x10^{-4}	-3.9544 x10^{-4}
UK	*2.3623 x10^{-1}*	*1.8974 x10^{-1}*	7.7136 x10^{-4}	3.7728 x10^{-4}	7.6846 x10^{-4}	5.8827 x10^{-4}	1.5940 x10^{-4}	2.0618 x10^{-4}	2.6672 x10^{-4}
FR	*2.5306 x10^{-1}*	*3.0338 x10^{-1}*	*1.0373 x10^{-1}*	6.4228 x10^{-3}	3.7022 x10^{-3}	5.4988 x10^{-4}	1.4138 x10^{-3}	2.4472 x10^{-3}	1.8795 x10^{-3}
IT	*2.8936 x10^{-1}*	*3.7085 x10^{-1}*	*1.1325 x10^{-1}*	*1.6168 x10^{-1}*	4.7847 x10^{-3}	1.1425 x10^{-3}	1.1290 x10^{-3}	2.0311 x10^{-3}	1.7709 x10^{-3}
AU	*1.8686 x10^{-1}*	*1.4241 x10^{-1}*	*1.8853 x10^{-1}*	*2.47091x10^{-1}*	*2.5316 x10^{-1}*	2.2014 x10^{-3}	3.3022 x10^{-4}	3.8760 x10^{-4}	3.8018 x10^{-4}
SW	*2.3443 x10^{-1}*	*1.8131 x10^{-1}*	*3.9165 x10^{-1}*	*5.4341 x10^{-2}*	*7.7552 x10^{-2}*	*2.2223 x10^{-1}*	7.1934 x10^{-4}	7.9520 x10^{-4}	4.9936 x10^{-4}
DE	*3.3976 x10^{-1}*	*3.0855 x10^{-1}*	*3.3476 x10^{-1}*	*4.4796 x10^{-2}*	*6.0425 x10^{-2}*	*3.3489 x10^{-1}*	*8.4373 x10^{-2}*	1.3302 x10^{-3}	8.6088 x10^{-4}
NO	*2.7498 x10^{-1}*	*2.8659 x10^{-1}*	*3.2097 x10^{-1}*	*8.5684 x10^{-2}*	*8.8225 x10^{-2}*	*3.2718 x10^{-1}*	*2.1340 x10^{-1}*	*1.6901 x10^{-1}*	1.2722 x10^{-3}

Note: The italicized entires are the normalized values of the smallest eigenvalues of the sample spectral density matrices.

TABLE 2

PHILLIPS AND OULIARIS' COINTEGRATION TEST
(k=10)

	US	CA	UK	FR	IT	AU	SW	DE	NO
US	3.8494 x10^{-3}	4.0197 x10^{-3}	1.2282 x10^{-3}	-1.8823 x10^{-3}	1.7237 x10^{-3}	1.5150 x10^{-3}	1.9659 x10^{-4}	-1.3645 x10^{-3}	-1.0370 x10^{-3}
CA	*3.5152 x10^{-2}*	4.8670 x10^{-3}	1.6070 x10^{-3}	-1.8925 x10^{-3}	7.4751 x10^{-4}	1.3972 x10^{-3}	2.1043 x10^{-4}	-9.3225 x10^{-4}	-1.0025 x10^{-3}
UK	*1.7220 x10^{-1}*	*1.2518 x10^{-1}*	1.4200 x10^{-3}	5.1945 x10^{-4}	1.4075 x10^{-3}	6.0893 x10^{-4}	2.6785 x10^{-4}	-2.0820 x10^{-5}	1.1976 x10^{-4}
FR	*2.6287 x10^{-1}*	*3.0847 x10^{-1}*	*1.4552 x10^{-1}*	8.0602 x10^{-3}	3.7748 x10^{-3}	3.2874 x10^{-4}	2.0766 x10^{-3}	2.9871 x10^{-3}	2.9742 x10^{-3}
IT	*3.5524 x10^{-1}*	*3.9127 x10^{-1}*	*1.2923 x10^{-1}*	*2.4603 x10^{-1}*	6.9599 x10^{-3}	7.3523 x10^{-4}	1.1819 x10^{-3}	1.2567 x10^{-3}	1.8152 x10^{-3}
AU	*2.2434 x10^{-1}*	*2.3636 x10^{-1}*	*2.9773 x10^{-1}*	*2.2116 x10^{-1}*	*2.3718 x10^{-1}*	2.3129 x10^{-3}	3.3482 x10^{-4}	-9.0600 x10^{-6}	1.8832 x10^{-4}
SW	*2.1911 x10^{-1}*	*1.8211 x10^{-1}*	*3.7589 x10^{-1}*	*5.7399 x10^{-2}*	*1.0780 x10^{-1}*	*2.9665 x10^{-1}*	1.0980 x10^{-3}	8.7448 x10^{-4}	8.0795 x10^{-4}
DE	*1.5767 x10^{-1}*	*1.9238 x10^{-1}*	*4.9064 x10^{-1}*	*3.1925 x10^{-2}*	*1.4033 x10^{-1}*	*3.8598 x10^{-1}*	*1.5031 x10^{-1}*	1.4541 x10^{-3}	1.1355 x10^{-3}
NO	*2.2185 x10^{-1}*	*2.0606 x10^{-1}*	*4.4852 x10^{-1}*	*4.7323 x10^{-2}*	*1.2387 x10^{-1}*	*3.9998 x10^{-1}*	*1.8772 x10^{-1}*	*1.3015 x10^{-1}*	1.6218 x10^{-3}

Note: The italicized entires are the normalized values of the smallest eigenvalues of the sample spectral density matrices.

relatively close to l_2^{crit2}. For only 22 of the pairs of countries are the eigenvalues smaller than l_2^{crit2}; in 14 of the cases, the eigenvalues exceed this critical value, strongly supporting the hypothesis of no cointegration. It appears that utilizing information corresponding to cycle lengths of 11.5 years and longer (those associated with the first ten periodogram ordinates) seems to have little effect on the evidence in favor of cointegration. However, employing the value k=10 lowers the smallest eigenvalues for three of the four pairs of countries for which coherencies were computed, which accords with the fact that these coherencies tended to rise substantially over the frequency range between the fifth and tenth ordinates.

Overall, there is little evidence that the multivariate cointegration reported by Cogley is expressed in the data's bivariate relationships. This result may be due to a sample size effect: when testing large numbers of series for cointegration simultaneously, the chance of finding statistically significant relationships may be higher than when only two series are examined, just as the R^2 of a regression is always non-decreasing in the number of regressors. Another possibility is that the cointegration between GDP processes is part of a very complex dynamic relationship which is not reducible into simple bivariate models. Using the intuitive interpretation of Phillips and Ouliaris' procedure developed in the previous chapter for the case in which three series are tested, cointegration would be manifested by the clustering of the growth rates about a plane in a three-dimensional space. It is quite possible that a relatively loose pattern of clustering, one which is statistically significant in three dimensions, would not be significant if the points were projected onto any of the planes formed by the coordinate axes, which is essentially what considering only two of the series simultaneously would entail. Of course, both these "explanations" are very speculative, and would definitely bear further examination.

Chapter Five

SUMMARY AND CONCLUSIONS

A. The Measurement of Persistence

It was argued in the introduction that the techniques of spectral analysis were ideally suited to the investigation of persistence in time series, given that the periodogram of a series allows one to separate out from the original data cycles (or "runs") of different lengths. This decomposition of the variance of the series is compatible with a wide range of parametric forms, and does not presuppose any attributes of the underlying data generating process except its stationarity. These same qualities allow spectral analysis to relate in a straightforward manner to the idea of cointegration, which by its very nature is a concept which focuses on the long run properties of time series. Having applied a small number of tests based on the ideas of spectral analysis to GDP time series, the results can be assessed with a view to answering the questions of whether the use of frequency domain techniques allows better and more precise conclusions to be drawn, and what, if any, are the implications of the results for macroeconomic theory.

It was clear from the figures presented in chapter four that while the use of $\hat{V}_f$ in place of $\hat{V}_k$ may have allowed more reliable confidence intervals to be obtained, it did not enable the size of the random walk in output to be estimated with great precision. This was not due to any deficiency in the fundamental techniques used, but the result of the fact that for the length of data series available, only a handful of independent observations of the item of interest -- namely, the "long run" -- are included. Indeed, as Cochrane (1988) has noted, in order to gather enough data for a single observation of the long run as defined by the parameter k in the definition of V, one would have to wait an infinite length of time. Of course, a more tractable version of the long run is utilized in practice, one which focuses on the length of business cycles, and is estimable through the use of the method mentioned in chapter two. However, when coupled with the fact that the variance ratio does not impose any parametric structure on the data, this focus on the long run makes impossible the drawing of any but the most general conclusions about the underlying, population value of the variance ratio, regardless of the procedure used to determine an appropriate value of k.

If the construction of formal confidence intervals for the variance ratio gives little help in forming a conclusion as to the importance of the random walk component in GDP processes, it would appear no definite conclusions can be drawn. However, there are several highly informal, but nonetheless plausible, arguments which suggest that the contribution of random walk components to these series is relatively small. Cochrane (1988) points out that a simple visual inspection of U.S. GDP series reveals that the data is devoid of the low-frequency waves typical of random walk processes; in other words, the "trend" discernable in the data is linear throughout the sample period. This finding is confirmed in the small point estimates for $\hat{V}_f$ derived for the U.S. in the preceding chapter. Figures presented by Cogley (1990) show that with the exceptions of France, Italy, and perhaps Norway, each of the GDP series examined here have appearances similar to the U.S. data, suggesting that for these countries also, the behavior of GDP is more closely analogous to a trend-stationary process than a random walk, at least over the sample periods utilized in this study.

The appearance of time plots of GDP series undoubtedly contributed to the conventional view of GDP, mentioned in the introduction, as a trend stationary process. The fact that this viewpoint prevailed for so long could be construed as evidence that the models derived from the trend-stationary assumption gave acceptable (or at least tolerable) results. The implication is once again that over the time horizons usually considered in economic studies, random walk components do not dominate GDP dynamics. Even now, some ten years after Nelson and Plosser (1982) raised the issue of unit roots in economic series, it remains an open question as to whether they are present at all, or if using trend-stationary models is appropriate. As Perron (1989) found, if the massive shock to GDP of the 1929 Great Crash is viewed as an exogenous event which was entirely separate from the normal process which generates GDP, then the trend-stationary assumption is impossible to reject on statistical grounds. It would be interesting to apply the Peron test to the few countries examined here which gave extremely high point estimates of V. It is quite possible that isolated events such as World War Two, which could cause a significant break or change in GDP's trend, are responsible for the large values of $\hat{V}_f$ obtained for these countries.

The foregoing comments are certainly not powerful enough to motivate the exclusive use of a trend-stationary modelling assumption for GDP processes, since the contrary evidence as presented by Dickey and Fuller (1979) and others is impossible to dismiss. What they do help to establish is that at this point in time, neither the hypothesis of integration nor its trend-stationary alternative is compellingly supported by the data. Given this fact, one cannot determine whether the variance ratio as defined here is zero, or, equivalently, whether GDP exhibits any persistence over infinite time horizons.

This indecision regarding the true value of the variance ratio does have some implications for econometric modelling and macroeconomic theory, though none of them are particularly profound. From a modelling standpoint, the inability to classify the GDP process as being persistent or trend-reverting simply means that the most general parametric models, those which may be trend- or difference stationary, should be used, with the specific sample being considered guiding the choice between the alternatives. Following this line of reasoning, Cochrane(1988) argues that random-walk behavior is not a feature that should be required of macroeconomic models. The results obtained here support this view, since the presence or absence of substantial random walk components depends on the sample under investigation, and is not a "stylized fact" characterizing all GDP processes. For macroeconomic theory, the inability to precisely measure persistence is not as great an obstacle as has been assumed by those who claim that difference-stationary processes characterize real business cycle models, while Keynesian macroeconomic models generate trend-stationary series: as Christiano and Eichenbaum (1990) observe, the distinction between I(0) and I(1) GDP processes does not rule out either of these competing theories.

While the variance ratio measure of persistence may be of limited usefulness given current techniques, there are many aspects of the persistence issue which are both relevant and possibly more likely to yield productive results. One such issue concerns the disparity between the persistence estimates obtained by the ARMA approach and the variance ratio. As was stated in chapter two, these statistics can be expected to differ substantially, depending on the dynamics of the process being considered. However, the fact that they do give such different conclusions for actual series may help characterize the processes which generated them. Cochrane (1988) claims that the ARMA approach gives too much weight to the short-run dynamics of sample series, and so is likely to generate upwardly-biased persistence measures, unlike his own statistic. If this is true, so that the short-term movements in GDP do tend to very persistent (where "persistence" is defined as the presence of some positive autocorrelation in the level of GDP over some horizons, not necessarily infinite), while over the time intervals used in the calculation of $\hat{V}_k$, GDP manifests substantial trend reversion, it may be possible to estimate an interval of time over which shocks to GDP can be expected to dissipate.

B. Comovements and Cointegration

In general, the results obtained here from the use of spectral analysis to examine comovements in GDP time series are more satisfactory than those derived from the estimates of the variance ratio. The high coherencies found between some of the series suggests that low-frequency movements in these countries' GDP's are quite highly correlated, a finding consistent with the presence of cointegration. Unfortunately, Phillips and Ouliaris' test failed to find substantial evidence in favor of this hypothesis, though for many of the pairs of countries, there was no compelling evidence against cointegration. In light of Cogley's results, which suggest that there is cointegration present in the system formed by the nine GDP series considered, the lack of significant pairwise cointegration could be interpreted to mean that the long-term dynamic relationships between these GDP processes are too complex to be reduced to bivariate frameworks. This explanation for the contrast between Cogley's conclusions and the cointegration tests on pairs of GDP series could be investigated further, in part by the methods developed here.

BIBLIOGRAPHY

Campbell, John Y., and Mankiw, N.G., "Are Output Fluctuations Transitory?" Quarterly Journal of Economics 102, (November 1987): 857-880.

Campbell, John Y., and Mankiw, N.G., "Permanent and Transitory Components in Macroeconomic Fluctuations." American Economic Review, Volume 77, No. 2 (May 1987): 111-117.

Christiano, L. and Eichenbaum, M. "Unit Roots in Real GNP: Do We Know and Do We Care?" Carnegie-Rochester Conference Series on Public Policy, Volume 32 (1990): 7-62.

Cochrane, John H., "How Big Is The Random Walk in GNP?" Journal of Political Economy 96 (October 1988): 893-920.

Cogley, T., "International Evidence on the Size of the Random Walk in Output." Journal of Political Economy, Volume 98, No. 3 (1990): 501-518.

Diggle, Peter J. (1990). Time Series: A Biostatistical Introduction. Oxford: Oxford University Press.

Dickey, David A., Jansen, Dennis, and Thornton, Daniel I. "A Primer on Cointegration With an Application to Money and Income". Federal Reserve Bank of St. Louis March-April 1991: 58-78.

Dickey, David A., and Fuller, Wayne A., "Distribution of the Estimators for Autoregressive Time Series With a Unit Root". Journal of the American Statistical Association 74 (1979): 427-431.

Engle, Robert F., "Interpreting Spectral Analyses in Terms of Time Domain Models." Annals of Economic and Social Measurement, 5:1 (1976): 89-109.

Engle, Robert F., and Granger, C.W.J. "Co-integration and Error Correction: Representation, Estimation, and Testing". Econometrica, 55, No. 2 (March, 1987): 251-276.

Jenkins, G.M., and Watts, D.G. (1968). Spectral Analysis and Its Applications. San Francisco: Holden-Day.

Johansen, Soren. "Statistical Analysis of Cointegration Vectors". Journal of Economic Dyanmics and Control 12 (1988): 231-254.

Johansen, Soren and Juselius, Katarina. "Maximum Likelihood Estimation and Inference on Cointegration-With Applications to the Demand For Money." Oxford Bulletin of Economics and Statistics 52,2 (1990): 169-210.

Lucas, Robert E., "Two Illustrations of the Quantity Theory of Money", American Economic Review, Volume 70, No. 5 (December 1980): 1005-1014.

McCallum, Bennett T., "On Low-Frequency Estimates of Long-Run Relationships in Macroeconomics." Journal of Monetary Economics, Volume 14 (1984): 3-14.

Morrison, Donald F., (1990) Multivariate Statistical Methods. New York: McGraw-Hill, Inc.

Muirhead, Robb J., (1982). Aspects of Multivariate Statistical Theory. Toronto, Ontario: John Wiley and Sons, Inc.

Nelson, Charles, R., and Plosser, Charles I., "Trends and Random Walks in Macroeconomic Time Series", Journal of Monetary Economics, Volume 10 (1982): 139-162.

Perron, Pierre. "The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis." Econometrica, Volume 57 (1989): 1361-1401.

Phillips, P.C.B. "Understanding Spurious Regressions in Econometrics." Journal of Econometrics, Volume 33 (1986): 311-340.

Phillips, P.C.B., and Ouliaris, S., "Testing for Cointegration Using Principal Components Methods." Journal of Economic Dynamics and Control 12 (June/September 1988): 205-230.

Raj, Baldev, "Persistence in Output: Further International Evidence From a New Approach", unpublished paper, Wilfred Laurier University, April 1990.

Sargent, Thomas J., (1987). Macroeconomic Theory. San Diego, California: Academic Press, Inc.

Stock, James H., "Unit Roots in Red GNP: Do We Know and Do We Care? A Comment." Carnegie-Rochester Conference Series on Public Policy, Volume 32 (1990): 63-82.

West, Kenneth D., "On The Interpretation of Near Random-Walk Behaviour in GNP". American EconomicReview, Volume 78, No. 1 (March 1988): 202-209.

Whiteman, Charles H., " Lucas on the Quantity Theory: Hypothesis Testing Without Theory". American Economic Review, Volume 74, No. 4 (September 1984): 742-749.

APPENDIX

Derivation of Equation (13):

Note that Cogley's window,

$$\frac{1}{k}\left[\frac{\sin\left(\frac{kw}{2}\right)}{\sin\left(\frac{w}{2}\right)}\right]^2,$$

equals

$$\frac{1-\cos(kw)}{k\,(1-\cos(w))}\quad, \tag{A1}$$

using trigonometric identities.

Beginning with the discrete Fourier transform of the Bartlet lag window at frequency w, we have:

$$\sum_{l=-k+1}^{k-1}\left(\frac{k-|l|}{k}\right)\ e^{-ilw} \tag{A2}$$

$$= \ 1+2\sum_{l=1}^{k-1}\frac{k-l}{k}\cos(lw) \tag{A3}$$

$$= \ \frac{1}{k}\left[k+2\sum_{l=1}^{k-1}(k-l)\cos(lw)\right] \tag{A4}$$

$$= \ \frac{1}{k}[s]. \tag{A5}$$

Thus, s is defined as:

$$\begin{aligned} s = k + 2(k-1)\cos(w) \quad & + 2(k-2)\cos(2w)\ldots \\ & + 2(2)\cos((k-2)w) \\ & + 2(1)\cos((k-1)w) \end{aligned} \tag{A6}$$

$$s = k + (k-1)(e^{iw} + e^{-iw}) + (k-2)(e^{2iw} + e^{-2iw}) \ldots + 2(e^{(k-2)iw} + e^{-(k-2)iw}) + e^{(k-1)iw} + e^{-(k-1)iw} \qquad \text{(A7)}$$

Therefore,

$$\begin{aligned}(e^{iw} + e^{-iw})\, s = k\,(e^{iw} + e^{-iw}) & \\ &+ (k-1)\,[e^{2iw} + e^{-2iw} + e^{0iw} + e^{-0iw}] \\ &+ (k-2)\,[e^{3iw} + e^{-3iw} + e^{iw} + e^{-iw}] \ldots \\ &+ (2)\,[e^{(k-1)iw} + e^{-(k-1)iw} + e^{(k-3)iw} + e^{-(k-3)iw}] \\ &+ [e^{kiw} + e^{-kiw} + e^{(k-2)iw} + e^{-(k-2)iw}] \qquad \text{(A8)}\end{aligned}$$

or:

$$\begin{aligned}(\cos w)\, s = k \cos w \; &+ (k-1)\,[\cos 2w + 1] \\ &+ (k-2)\,[\cos 3w + \cos w] \ldots \\ &+ (2)\,[\cos(k-1)w + \cos(k-3)w] \\ &+ (1)\,[\cos kw + \cos(k-2)w] \qquad \text{(A9)}\end{aligned}$$

Thus,

$$\begin{aligned}(\cos w)\, s = k-1 + (2k-2)\cos w & \\ &+ (2k-4)\cos(2w) + (2k-6)\cos(3w) \ldots \\ &+ (4)\cos(k-2)w + 2\cos(k-1)w + \cos(kw) \qquad \text{(A10)}\end{aligned}$$

Subtracting (A10) from (A6) gives:

$(1-\cos(w)) \quad s = 1-\cos(kw)$, or:

$$s = \frac{1-\cos(kw)}{1-\cos(w)} \qquad \text{(A10)}$$

From (A5), it follows that the discrete Fourier transform of the Bartlett lag window is:

$$\frac{1-\cos(kw)}{k\,(1-\cos(w))},$$

which is just equation (13), as was shown in (A1).