Order Number 9500049

Application of local to unity asymptotic theory to time series regression

Elliott, Graham, Ph.D.

Harvard University, 1994

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

HARVARD UNIVERSITY

THE GRADUATE SCHOOL OF ARTS AND SCIENCES

THESIS ACCEPTANCE CERTIFICATE

The undersigned, appointed by the

Division

Department of Economics

Committee

have examined a thesis entitled

Application of Local to Unity Asymptotic
Theory to Time Series Regression

presented by Graham Elliott

candidate for the degree of Doctor of Philosophy and hereby certify that it is worthy of acceptance.

Signature

Typed name James Stock

Signature

Typed name Gary Chamberlain

Signature

Typed name Takatoshi Ito

Signature

Typed Name Guido Imbens

Date April 27, 1994

Application of Local to Unity Asymptotic Theory to Time Series Regression

A thesis presented

by

Graham Elliott

to

The Graduate School of Arts and Sciences

in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy

in the subject of

Economics

Harvard University
Cambridge, Massachusetts

May 1994

Acknowledgements

Firstly, I would like to thank those who, prior to my visit here, encouraged me in my studies. These include Warwick McKibben, who suggested studying in the US; and Ron Bewley, who helped me see the enjoyment of pursuing an academic career, and my family for supporting me in this. I would also like to thank the Harvard Club of Australia and the Menzies Foundation for subsidizing my first year of study.

In Cambridge, I have learned much from both my colleagues and the faculty. In particular, Antonio Fatas and Judy Hellerstein contributed greatly. I express thanks to Professors Alberto Alesina, Dale Jorgenson, and Mark Watson for their advice and comments. I have benefitted greatly from comments given by Professor Gary Chamberlain. I thank Professor Guido Imbens for the enormous amount of time he has spent in reading and rereading my work, and for the comments he has made, and also Professor Takatoshi Ito, who has given up much time to my work and education. I have been lucky to have had a complete and thorough apprenticeship in the study of time series econometrics from my main advisor, Professor Jim Stock, for which I am greatly appreciative. Without this training, this thesis would not exist.

Lastly, I thank my friends, especially Dagmar, who supported me in my endeavor.

Abstract

In classical hypothesis testing in time series regression, the asymptotic theory employed depends on the stochastic process followed by the regressors. Any approach to inference must either make assumptions on the form of these stochastic processes or use pretests as a selection criteria. This thesis examines this issue when there is serious doubt as to the stochastic properties of the regressor, when shocks to the regressor are persistent. This characterization appears to well reflect most time series data available in economics.

The first chapter provides an overview of the model and the problem for hypothesis testing. When the largest root of the regressor is large, we are unable to decide with data whether or not there is a unit root or a root local to unity. However, the asymptotic distribution employed in second stage hypothesis testing depends on this distinction.

Chapter two examines optimal unit root tests under alternate assumptions that have generally been employed, deriving efficient tests for this case. This chapter shows that the assumptions on the generating process matter in the construction of optimal tests, and provide a new set of tests which can be employed to learn about the largest root in the regressor variable.

In the third chapter, the common practice of conditioning on an exact unit root in the regressor and employing asymptotically efficient cointegrating vector estimation techniques for hypothesis testing is examined. It is shown that even for arbitrarily small deviations

from the assumption of a unit root, this procedure can lead to severe overrejection of the true null hypothesis. The tests can have size up to 1. It is argued in this chapter that unit root pretests cannot overcome this problem.

The final chapter, chapter four, examines hypothesis tests for unbiasedness in the forward exchange rate market. It is shown that the interpretation of tests of various specifications depends on the stochastic process followed by the regressor, as argued above. Potential reasons are given for the rejections of the null hypothesis in the literature are given, including the use of the results in chapter three to show the problems of recent investigations of the null hypothesis using cointegration methods. A new test to distinguish between unbiasedness and static expectations is also introduced.

Contents

Chapter 1: The General Problem in Perspective.

I. Introduction

The general model under investigation in this thesis can be written as a triangular bivariate system of two related time series,

$$\begin{aligned} y_{1t} &= d_{1t} + \alpha y_{1t-1} + v_{1t} \\ y_{2t} &= d_{2t} + \gamma y_{1t-s} + v_{2t} \end{aligned} \qquad (1)$$

where t=1...T, d_{1t} and d_{2t} are deterministic components (d_{1t}=0 will be assumed for the generating process throughout)[1], y_{1t} and y_{2t} are both univariate with k fixed initial values, s may be either zero or 1, $v_t=(v_{1t},v_{2t})'$, and $\Phi(L)v_t=\epsilon_t$ where $\Phi(L)$ has all roots outside the unit circle. In addition, ϵ_t is assumed to be a martingale difference sequence so $E[\epsilon_t]=0$ and $E(\epsilon_t\epsilon_t')=\Sigma$ given information at time t-1 (fourth moments are also assumed to exist). The equations are linked through the correlated residuals. The economic model to be estimated determines s; either a contemporaneous relationship is examined and s=0 or some dynamic ordering is of interest and s=1. Of primary interest is estimation of and inference on the parameter γ when α is close to one. This model is triangular as y_{2t} does not enter the equation for y_{1t}.

This model contains sufficient generality to capture the features of many models of interest

[1] The deterministics here are the relevant alternatives for describing persistance in y_{1t} to stochastic roots, and estimates of α will be inconsistant against such alternatives if d_{1t} is left out of the regression, so it will be kept in the specifications.

in applied time series economics. When $\alpha=1$, the relationship between y_{2t} and y_{1t} is said in the terminology of Engle and Granger (1987) to be a cointegrating relationship, with a cointegrating vector of $(1\ -\gamma)$. When α is not necessarily equal to one, models such as equation (1) are still of great interest as economic theories often yield such relationships, independent of the value of α. If $s=1$, regardless of the value taken by α, we say that y_{1t} Granger causes [in the sense of Granger (1969)] y_{2t}. Such temporal orderings are often derived from theory in economics.

Both of these types of regression models have been widely applied in testing theories in macroeconomics, finance and international finance. Examples of cointegration models estimated in macroeconomics include tests for long run money demand equations [Stock and Watson (1993), Hoffman and Raasche (1991)] and tests of consumption theory [Ogaki (1992)]. In finance, Campbell and Shiller (1987) test the present value model using the cointegration methodology. In international finance the concept and theoretical structure of cointegration includes applications such as testing for long run purchasing power parity [Corbae et al (1992), Choudry et al. (1991)], and long run forward market unbiasedness [Evans and Lewis (1993), Mark et al. (1994)].

Alternately, tests for Granger causality in macroeconomics include tests of Hall's (1978) consumption random walk hypothesis [Hall (1978), Mankiw and Shapiro (1985)]. In finance tests of the unpredictability of stock market returns or returns in other financial markets [Fama (1965) shows that returns should be unpredictable if markets are efficient] revolve around estimating models such as equation (1) above [e.g. Hardevoulis (1990), Hamilton

(1992)]. Similar models obtain in international finance.

A summary of recent applications, including models which are transformations of equation (1), is contained in Appendix 1.

It is well known that the properties of estimates of γ in such models depends on the value taken by α, and on the cross correlation of the errors. If α is fixed and less than one in absolute value, then with the dynamics and simultaneity suitably modelled the t statistic on $\hat{\gamma}$ is asymptotically distributed as a normal with mean zero and variance one. If $\alpha=1$, then the usual limit theory assumptions are violated and the limit distribution for the estimate of γ takes a different form [Stock (1987)]. Further, for α close to one, then the usual limit theorem results appear to break down in practice for empirically relevant samples sizes, in the sense that they do not provide a good approximation to the finite sample distribution.

This thesis is concerned with the case of α unknown but in the region of one. The focus on this case arises from both theoretical and empirical motivations. The theoretical motivation is that it is unusual for an economic model to suggest the exact value of α. This only occurs rarely, and then in very special cases of economic models [e.g. Hall (1978) shows consumption to be a random walk under the dual assumptions of quadratic utility and also that the risk free interest rate and discount rate are constant and equivalent, Fama (1965) derives the efficient markets hypothesis, for which changes in asset prices are not forecastable on the assumption of risk neutral investors]. In many of these models, the assumptions required to obtain a unit root are very restrictive and arguments for their violation are easily made. The empirical motivation is that most macroeconomic, finance and international finance data exhibits strong trending of unknown form[2], such that estimates of $\hat{\alpha}$ typically are large and insignificantly different from one [for macroeconomic data, Nelson and Plosser (1982), for exchange rate data, Meese and Singleton (1982)].

There are two approaches generally taken by researchers in confronting such models when α is not given by theory. The first is to simply remove the problem of dependence by making an assumption about the existence or not of a unit root. In this case, either the potential problem is ignored completely and the normal distribution is employed, or often the 'weight of previous evidence' is that variables are I(1) so the methods of cointegration or non standard asymptotics are used. The alternative procedure, one which is more popular in current literature, is to pretest for a unit root in y_{1t}, and proceed conditional on this result as if it were true. In these situations the researcher fails to reject the existence of a unit root, and proceeds conditional on the existence of a unit root in subsequent tests.

This thesis, and other papers written in conjunction with this thesis, examine a number of questions from a classical viewpoint. How much can we learn from the data about α? What is the effect of pretesting? What is the effect of proceeding on the assumption that α is equal to one when this is not true? How can we conduct inference when α is unknown? Do these results hold any real implications for applied questions in practice? Each of these questions play a role in understanding and evaluating inferences made on models which can

---

[2] This trending could be due to either stochastic trending, i.e. α close to one, or deterministic trending, i.e. $d_{1t} \neq 0$.

be written in the form of equation (1).

The remainder of this chapter discusses how these questions and the results presented in this thesis fit in with both the rest of the literature in econometrics and more specifically with the literature on time series econometrics. The goal is to understand the question at hand, various precedents for solutions, their applicability to this specific problem, and the lessons we may draw for applied research using time series data. In the next section, the asymptotic theory employed to answer the questions raised is motivated. Here it is shown that hypothesis tests depend asymptotically on a nuisance parameter (related to α), where the nuisance parameter summarizes the persistence in y_{1t}. Section 3 examines the question of inference on α, and the information we can expect to receive from tests on this nuisance parameter. This section places chapter 2 of this thesis in perspective. Section 4 examines the testing of hypotheses on γ from a classical (frequentist) testing perspective, reviewing the approach taken in the time series literature and in the econometric literature more generally. Chapter 3 of this thesis examines the most popular approach to hypothesis testing in this framework, that of 'asymptotically efficient cointegrating' estimation tests. The fifth section examines alternative approaches to hypothesis testing in the models considered here: those of nonparametric tests, bootstrapping and of Bayesian approaches. The sixth section sums up.

II Asymptotic Theory for $\hat{\gamma}$

For models such as that given in the previous section, a common statistic employed in undertaking hypothesis tests for theories involving the parameter γ is to examine the t statistic (or pivot) on the estimated value for γ from the OLS regression of the second equation in (1). That is, construct t_γ, which is given by

$$t_\gamma = \frac{(\hat{\gamma} - \gamma_0)}{se(\hat{\gamma})} \tag{2}$$

where γ_0 is the value taken by γ under the null hypothesis being tested.

If $|\alpha| < 1$, then estimation of $\hat{\gamma}$ by ordinary least squares (OLS) and estimating the standard error of the estimate in the usual (robust) way as $\hat{\omega}(\Sigma y_{1t-s}{}^2)^{-1/2}$, where $\hat{\omega}^2 = \hat{S}_{v2t}(0)/2\pi$ [the spectral density of v_{2t} at frequency zero, scaled by 2π][3], yields the result that t_γ has an asymptotic normal distribution with mean zero and variance 1 provided that y_{1t} is uncorrelated with all leads and lags of v_{2t}. If this is not the case, the researcher can use instrumental variables or seemingly unrelated regression techniques.

If instead $\alpha=1$, then t_γ calculated as above has the limit distribution given by

$$t_\gamma \rightarrow \delta\tau_\alpha^d + (1-\delta^2)^{\frac{1}{2}} z \tag{3}$$

where τ_α^d is the distribution of the OLS t statistic testing the hypothesis $\hat{\alpha}=1$ (detrended by the trend specification in d_{2t}) and is a function of standard Brownian motions [see Stock (1994) for details of the statistic τ_α, it was originally derived by Dickey and Fuller (1979), and percentiles of the distribution are given in Fuller (1976)], $\delta=\Omega_{12}/(\Omega_{11}\Omega_{22})^{1/2}$, $\Omega=S_v(0)/2\pi$

[3] In the case of general serial correlation of unknown form as in the model in equation (1), the usual estimator of the variance is inconsistent and must be replaced by an estimate of $S_{v2t}(0)$, see Hansen (1982) or White (1984).

[the spectral density matrix of the residuals of equation 1 at frequency zero, scaled by 2π], and z is a standard normal variable which is asymptotically independent of τ_α^d. Note that here, unlike the previous case, no transformations are required if y_{1t} and v_{2t} are correlated, this simultaneity is subsumed by the nuisance parameter δ.

There are two practical problems facing the applied researcher who wishes to conduct classical inference on $\hat{\gamma}$ when the true value of α is unknown. The first is that, except for the special case of $\delta=0$, it can be seen by directly comparing the two limit distributions that the applicable limit distribution for t_γ depends on the value taken by α. This lack of independence of the limit distribution on the nuisance parameter α presents the chief difficulty in hypothesis testing on γ: the classically constructed t statistic does not have a distribution independent of the parameters of the model. If we consider α as a nuisance parameter, this result shows that t tests on γ depend on this nuisance parameter. In either case, however, estimation of the parameter $\hat{\gamma}$ is consistent for its population value.

The second problem confronting the applied econometrician arises from the different asymptotic behavior of estimates of γ given the size of the largest root in y_{1t}, i.e. α. This is the well known knife edge case of $\hat{\gamma}$ converging at rate T for $\alpha=1$, and at rate $\sqrt{T}$ for $|\alpha|$ fixed and less than one [see Stock (1987)]. The difficulty here is that for values close to one but not exactly equal to one, the asymptotic normal distribution of t_γ (derived under the assumption that α is fixed and less than one in absolute value) does not provide a good approximation to the distribution of t_γ in finite samples of the size usually encountered in practice. This is well documented in practice [Evans and Savin (1981,1984), Ahtola and

Tiao (1984)]. The intuition for this breakdown is straightforward; whilst the asymptotic distributions exhibit this sharp discontinuity at $\alpha=1$, small sample distributions will be continuous in α. This gives the direct implication that for some range over α, asymptotic distributions derived with α fixed will be poor approximations of small sample distributions.

This breakdown can be seen graphically in figure 1. This documents the empirical distribution of t_γ (solid line) along with the standard normal distribution (long dashes)[4]. The particular model estimated here is

$$\begin{aligned} y_{1t} &= 0.95 y_{1t-1} + \epsilon_{1t} \\ y_{2t} &= \gamma y_{1t-1} + \epsilon_{2t} \end{aligned} \tag{4}$$

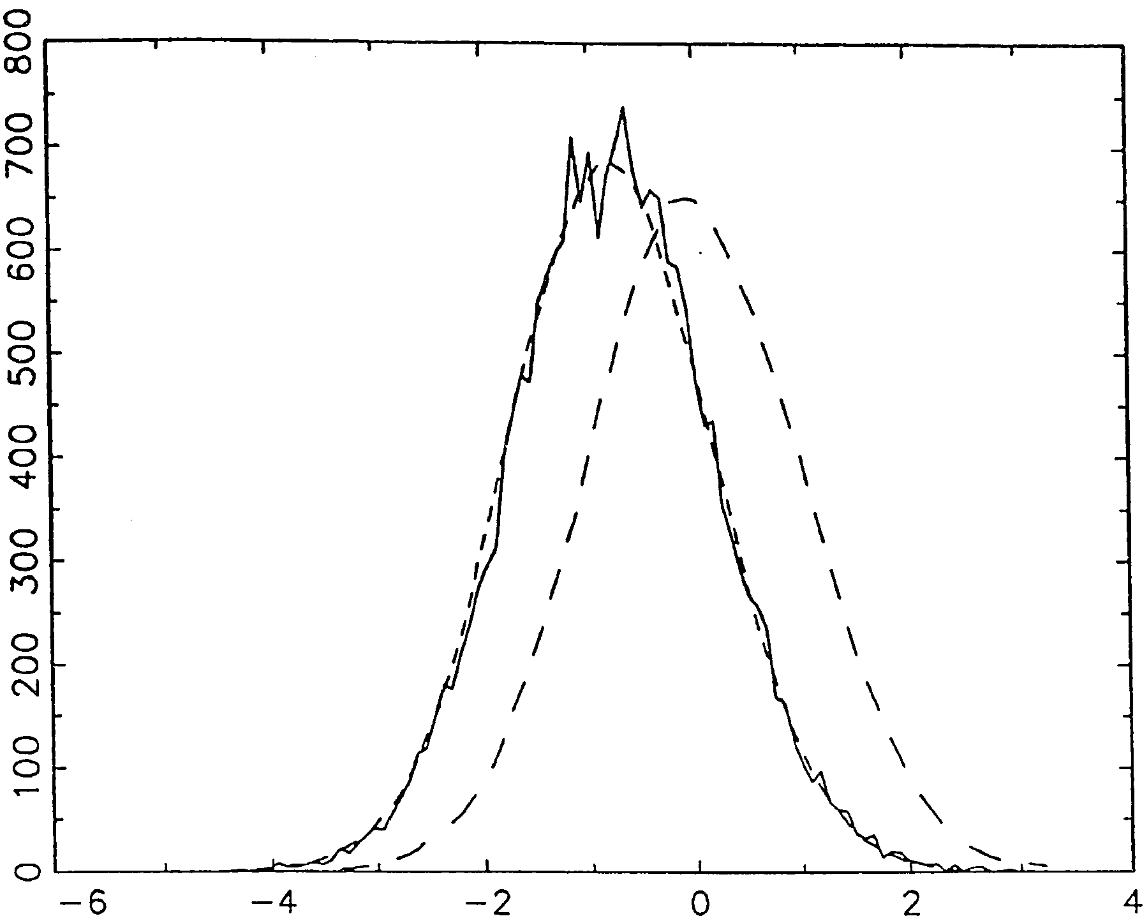
where the residuals v_t are serially uncorrelated with variance covariance matrix Σ, $\Sigma_{12}=0.9$, $\Sigma_{11}=\Sigma_{22}=1$, and $T=100$ (the values for α and T are chosen as empirically relevant possibilities, the shift documented is increasing in Σ_{12}, so this choice highlights the problem).

The most apparent feature of the difference between the distribution with one hundred observations and the normal distribution is that the empirical distribution is shifted significantly to the left. It is also more peaked. This difference between asymptotic theory and the empirical distribution with relevant sample sizes presents a problem for applied researchers when α is unknown, in the sense that even if α could be selected so that it was known that $|\alpha|<1$, the asymptotic normal distribution may not be a useful guide to the distribution of t_γ for the sample size at hand.

---

[4] The empirical distribution is the distribution of estimated t_γ with 100 observations and 20000 Monte Carlo replications. The lack of serial correlation was treated as known.

Figure 1: I(0) and Local to I(1) Distribution Approximations.



Notes: The graph shows the histogram of t statistics testing the true null hypothesis for the model in equation (4) with 100 observations. This is given by the solid line. The long dashed distribution is the N(0,1) asymptotic distribution of t_γ calculated for α fixed and equal to 0.95. The short dashed line is the local to unity asymptotic distribution calculated with c=-5. See the text surrounding equation (4) for the full specification of this model.

An alternate approach to obtaining asymptotic distributions to approximate the distribution of statistics such as t_γ where α is close to but not necessarily equal to one is to consider the distribution obtained when α is a sequence, i.e. $\alpha=1+c/T$, where c is fixed. This is the approach taken by Bobkoski (1983), Cavanagh (1985), Phillips (1987), Chan and Wei (1987) and Chan (1988). In the actual problem, we do not consider α to be converging to one asymptotically, this characterization of α is used only in deriving the asymptotic distribution given a value for α and a fixed sample size.

In this case for α close to one, in the local to unity sense, the limiting distribution of t_γ [given by Elliott and Stock (1992) equation 2.4] is

$$t_\gamma \rightarrow \delta \tau_c^d + (1-\delta^2)^{\frac{1}{2}} z \tag{5}$$

where τ_c^d is the local to unity distribution of the t statistic testing $\hat{\alpha}=\alpha$ with $c=T(\alpha-1)$ and deterministics d_{2t} included in the regression, z is again a normal random variable asymptotically independant of τ_c^d. When $c=0$, this is identical to equation (3) given above, and when c is large and negative, this collapses to the case of $|\alpha|<1$ and fixed [Phillips (1987) shows this for τ_c].

This approximate distribution tends to work well in finite samples. For the example given above in Figure 1, the corresponding value of $c=T(\alpha-1)=100*(0.95-1)=-5$. The asymptotic local to unity distribution for $c=-5$ is also graphed on Figure 1, being the distribution shown by the short dashed line. Clearly, this is an excellent approximation to the estimated empirical distribution, which lies almost entirely on the local to unity asymptotic distribution.

This point is also made in Table 1, which examines the performance of the local to unity and I(0) [i.e. N(0,1)] distributions for a range of values of α between zero and one when 100 observations are available. The model is identical to that in equation (4) above, except that it is examined for a range of values for α and δ. Three panels are presented, for δ = 0.2, 0.5, and 0.8. In each panel the upper and lower 2.5% critical values are employed for each distribution to evaluate the reported empirical upper and lower rejection rates for both the I(0) and local to unity characterizations of the limiting distribution. The chief features of the breakdown of the N(0,1) distribution are seen in each panel, with the extent of the breakdown increasing as δ moves closer to one (from equation (5) we can see that the weight given to the non standard part of the distribution is increasing in δ, there will be no breakdown when δ=0 as the weight is zero here). For values of α close to zero, the normal distribution is a good guide to the limiting behavior of t_γ. For α closer to one, however, use of the asymptotic normal critical values results in overrejection in the lower tail and underrejection in the upper tail. This is to be expected from Figure 1. The extent of this shift in mass is substantial; for δ=0.2, which represents a very mild relationship between the two residuals, in the limit as $\alpha \rightarrow 1$ the standard normal distribution will reject 5% of the time in the lower tail and almost never in the upper tail. For δ=0.8, this lower tail rejection level is 21% (Section 5 in Chapter 3 and the empirical results of Chapter 4 give results on this parameter for applications with real data, this parameter is generally non zero reflecting the general interdependence of economic data).

Table 1:Local to I(1) vs I(0) distribution approximations

	I(0)		loc to unity	
alpha	lower	upper	lower	upper
delta =	0.200			
0.000	0.029	0.026	0.025	0.033
0.100	0.030	0.026	0.025	0.033
0.200	0.029	0.025	0.026	0.033
0.300	0.030	0.024	0.026	0.035
0.400	0.031	0.024	0.028	0.032
0.500	0.031	0.023	0.026	0.035
0.600	0.031	0.023	0.026	0.032
0.700	0.032	0.023	0.027	0.033
0.800	0.033	0.022	0.029	0.030
0.900	0.035	0.020	0.028	0.033
1.000	0.049	0.011	0.025	0.028
delta =	0.500			
0.000	0.028	0.026	0.022	0.035
0.100	0.029	0.024	0.023	0.034
0.200	0.031	0.024	0.023	0.033
0.300	0.033	0.024	0.022	0.032
0.400	0.034	0.022	0.022	0.033
0.500	0.035	0.021	0.025	0.033
0.600	0.037	0.019	0.027	0.033
0.700	0.041	0.017	0.027	0.032
0.800	0.045	0.015	0.029	0.034
0.900	0.050	0.012	0.028	0.032
1.000	0.108	0.003	0.023	0.027
delta =	0.800			
0.000	0.030	0.023	0.016	0.036
0.100	0.030	0.022	0.018	0.035
0.200	0.033	0.022	0.018	0.035
0.300	0.033	0.020	0.019	0.035
0.400	0.036	0.019	0.020	0.034
0.500	0.038	0.017	0.020	0.034
0.600	0.041	0.016	0.022	0.033
0.700	0.046	0.013	0.025	0.031
0.800	0.054	0.010	0.025	0.030
0.900	0.069	0.008	0.030	0.034
1.000	0.208	0.001	0.028	0.027

Notes: Critical values for local to unity results were calculated from Monte Carlo experiments with T=1000 and 5000 replications. The results reported are rejection rates from a Monte Carlo with T=100 and 10000 replications. Critical values for the I(0) case are from standard normal tables.

In contrast, the local to unity distribution has quite good size properties for α close to one. As $\alpha \to 0$, this breaks down as α is quite far from 1. In the lower tail, for large δ, the local to unity asymptotic distribution tends to underreject. However, this underrejection is not really apparent even for $\delta=0.5$. In the upper tail, the tendency is to overreject, a tendency which holds for reasonably large values of α. It is interesting, though, to note that this tendency to over-reject is small.

Recall that the intuition for the breakdown of N(0,1) asymptotics as an approximation of small sample distributions suggested that there would be a range over α where the asymptotic theory approximation would be poor. The local to unity asymptotics make this comment precise. The range of breakdown for α is a region c/T, and this region disappears at rate T. Note also that one alternative approach to using local to unity asymptotics would be to examine small sample distributions directly. This approach is valid but extremely problematic; the small sample distributions would depend on distributional assumptions[5] and would be different for each possible convolution of nuisance parameters in the dynamics of $\Phi(L)$. As yet few attempts to do this have been made. The local to unity approximation, on the other hand, is valid under a wide range of distributional assumptions and nuisance parameters in $\Phi(L)$ are easily handled.

In most economic applications, we expect α to be fairly large. This is follows empirically

---

[5] Guido Imbens has pointed out that the similar asymptotic results for different distributions suggests that the small sample results would also be similar across different distributional assumptions.

from results such as Nelson and Plosser (1982), who fail to reject $\alpha=1$ for many US macroeconomic series, and from Stock (1991a), who inverts the statistic used for the Nelson and Plosser (1982) statistics to obtain confidence intervals on α, and shows that they exclude small values for α. Even the Bayesian results of Dejong and Whiteman (1991) suggest that these roots are fairly large. Results in Elliott, Rothenberg and Stock (1992) show that the tests used in the study by Nelson and Plosser have asymptotic power equal to one (when the data has been detrended) testing the alternative of c=-30 against the null hypothesis of a unit root. This suggests that tests of $\alpha=0.7$ against the null of one with 100 observations would have very high power. This suggests that the relevant area, as regards values of α, to examine is relatively close to one, an area where the local to unity approximation appears to work well. Thus, this thesis concerns itself mostly with roots large and close to one, and employs the approximation of local to unity asymptotic results to examine the large sample behavior of the statistics. This solves to a great extent the second problem confronted by classical researchers[6].

If we accept that the classical finite sample distribution of t_γ is well approximated by the local to unity distribution, then the asymptotic distribution depends on the fixed local to unity parameter c. As before, the limit distribution of the usually estimated t statistic depends on this nuisance parameter, and so estimation of an exact asymptotic distribution requires knowledge of α, or more precisely $T(\alpha-1)$. I have indicated above that this is exactly information which is unknown to the applied researcher, in the sense that economic theory

[6] A third potential problem is that δ is unknown. Lemma 2 of Chapter 3 of this thesis shows that in the range we consider, this nuisance parameter is consistently estimable and so tests invariant to this parameter can be easily constructed.

rarely restricts this parameter to a value.

To see this point a different way, it is perhaps better to consider the approach taken by the cointegration literature for solving for a distribution for γ. Consider the case where $\Phi(L)=1$, i.e. there are no dynamics so that $v_t=\epsilon_t$. Following the algebra in section 3 of Chapter 3 of this thesis (which finds the seemingly unrelated least squares transformation of equation 1), it can be seen that the second equation in (1) can be rewritten as

$$y_{2t} = d_{2t} + \gamma y_{1t} + \varphi(1-\alpha L)y_{1t} + \eta_t^* \quad (6)$$

where $E[\varphi]=\Sigma_{12}\Sigma_{11}^{-1}$ and $E[\eta_t^* \epsilon_{1t}]=0$. This result is shown for the case of $\alpha=1$ in Phillips (1990) and Stock and Watson (1993), where both assume normality of ϵ_t to factor the likelihood and then note that the normality assumption is not required. The cointegration result these authors examined is when $\alpha=1$, this transformation underlies the method of single equation (limited information) cointegration estimation of γ. The purpose of introducing this framework is to examine the estimator of γ and the information required to undertake hypothesis testing on γ independent of knowledge of α.

It is clear that were α known exactly, then the distribution of the t statistic on $\gamma=\gamma_0$, denoted as t_γ^t (where the t superscript here indicates that $\hat{\gamma}$ is estimated from the transformed equation (6) above), has an asymptotic normal distribution. This follows from the orthogonality of the regressors and the residual of equation (6) [see the appendix of Chapter 3 for a proof of this under general conditions]. In the absence of the exactly known value of α, a number of possibilities are available. If we were able to find some way of choosing

α so that it did not affect the limiting distribution of t_γ, then we would be able to construct a test of the null hypothesis with size controlled for unknown α.

The results of Theorem 2 of Chapter 3 in this thesis show that setting $\alpha=1$ in the regression given by equation (6) results in the distribution

$$t_\gamma^t \Rightarrow z - c\delta(1-\delta^2)^{-\frac{1}{2}}\left(\int_0^1 J_c^d(s)^2 ds\right)^{\frac{1}{2}} \tag{7}$$

where δ is as above, $\Sigma_{2.1}=\Sigma_{22}-\Sigma_{11}^{-1}\Sigma_{12}^{2}$, z is a standard normal variable independent of $J_c^d(s)$, and $J_c^d(s)$ is a detrended diffusion (Ornstein Uhlenbeck) process, where $dJ_c^d(s)=cJ_c^d(s)ds+dW^d(s)$, and $W^d(s)$ is a standard Brownian Motion process detrended by d_{2t}. The distribution in (7) depends on the value of $c=T(\alpha-1)$ through two channels; directly as is seen by c entering the equation and indirectly as the diffusion process is indexed by c.

If instead we replace α in equation (6) by $\hat{\alpha}$, the OLS estimate of α, then t_γ^t has the distribution given by

$$t_\gamma^t \Rightarrow z - c\delta(1-\delta^2)^{-\frac{1}{2}}\left(\int J_c^d(s)^2 ds\right)^{\frac{1}{2}} + \delta(1-\delta^2)\,\tau^d_{(\hat{\alpha}=1)} \tag{8}$$

where $\tau_{(\hat{\alpha}-1)}{}^d$ is the distribution of the test statistic testing $\hat{\alpha}=1$ when $\alpha=1+c/T$ (see Phillips (1987) for the derivation of this distribution) detrended by d_{2t}. This distribution also depends on $T(\alpha-1)$.

In each case, we need to know the value of $T(\alpha-1)$ to be able to choose the correct distribution for inference, so neither of these approaches provide useful tests. To do this in the absence of direct knowledge of α, it can be seen that for invariant inference we require that c be consistently estimable. Otherwise, this nuisance parameter affects the distribution of the t statistics for $\gamma=\gamma_0$.

A corrolory to the result that t tests constructed by testing $\hat{\gamma}=\gamma_0$ are not invariant to α is that in those cases where the value of α is given by theory, then the above tests (whether transformed in some way or not) are really testing the joint null that this information on α is correct and the stated null that $\gamma=\gamma_0$. The tests will reject in both the directions of $\alpha\neq\alpha_0$ and $\gamma\neq\gamma_0$ (see Chapter 3 section 7 for discussion of this with reference to cointegration estimation, and Chapter 4 for this in reference to an empirical example). This suggests that if the null hypothesis can be written as a joint test over both α and γ, then inference can proceed. Section 4 of Chapter 4 gives such an example.

As the potential for construction of a test independant of knowledge of α depends on the information we can learn about this nuisance parameter from the theory, we turn to inference over α in the next section.

III Classical Inference on α

Given the dependance of the asymptotic distribution of subsequent hypothesis tests (such as tests on γ in the model here) on the size of the largest root in y_{1t}, there has been substantial interest in tests of $\alpha=1$. A comprehensive and current review of the history and performance of these tests, including a discussion of other reasons why these tests are of interest, is given in Stock (1994). This section will examine first the role tests for a unit root, and more generally inference on α, can take in determining the asymptotic distribution relevant for the second stage. Secondly this section will examine optimal tests for α in the Neyman Pearson sense, and the potential use of such optimal tests in the second stage. Also discussed are possible improvements to these optimal tests by inclusion of covariates [Hansen (1993)][7].

If one were to disregard argument of the previous section, that the 'knife edge' result that the asymptotic distribution of $\hat{\gamma}$ and its associated t statistic depends on only whether $\alpha=1$ or is fixed and less than one in absolute value, then the problem of second stage inference would only require that the first stage consistently select the I(1) ($\alpha=1$) or I(0) ($|\alpha|<1$) models. This type of 'selection' of the correct asymptotics or transformations to obtain second stage asymptotics appears to be behind two-stage testing rules where tests of a unit root leads to use of I(1) asymptotic theory and related transformations of the model if the researcher fails to reject and I(0) asymptotic theory if the researcher rejects the unit root, such as suggested in the 'two step' procedure of Engle and Granger (1987) for cointegration estimation (Appendix 1 shows that such pretesting procedures are the preferred approach in empirical work). In this case, if the pretest is consistent in the sense that the correct distribution is selected asymptotically (and the 'knife-edge' asymptotics were correct), then this strategy would produce consistent tests of hypotheses over γ in the second stage.

[7] The optimal tests are optimal when only the data $\{y_{1t}\}$ is observed. Thus, additional covariates represents extra information outside this framework, and enables potentially greater power in unit root tests.

Elliott and Stock (1992) show that standard tests for a unit root with constant critical values, such as the Dickey and Fuller (1979) τ test, are not consistent pretests due to the asymptotic correlation between the pretest and the second stage test. In order for these tests to consistently classify a series as I(1) or I(0), we require that type I and type II errors in the first stage go to zero. Two methods which achieve this are using standard tests with critical values which vary with the number of observations, or by using a Bayesian classification technique [e.g. Phillips and Ploberger (1991), Stock (1992), Elliott and Stock (1992)], which achieves the same effect.

Whilst these work in the 'knife edge' case where it is presumed that the asymptotic distribution for α fixed and close to one yields normal asymptotics for t_γ, it was shown in the previous section that such asymptotics provide a poor guide to the distribution of t_γ in conventional sample sizes. If one examines the local to unity sequence, then the procedures of the previous paragraph will asymptotically misclassify stationary variables best described by local to unity sequences as being I(1), so the procedures break down [Elliott and Stock (1992), Theorem 3]. Campbell and Perron (1991), in a paper aimed at guiding empirical practice in macroeconomics, argue that this misclassification, or in their terms low power against close alternatives, is potentially an advantage for second stage inference as they suggest using I(1) asymptotics may be better for models close to this model. The smoothness of the small sample distribution may suggest that this is true, but in either case size is not controlled by such 'accidents'. This is made clear in chapter 3, where pretending that close to nonstationary variables are nonstationary leads to potentially very large size distortions in hypothesis tests.

Restricting attention once again to values of α and sample sizes which are well approximated by the local to unity model, it is clear from the previous section that in the absence of knowledge of α, we require consistent estimation of $T(\alpha-1)$. Such consistent estimation would then enable a correction that would enable the removal of the non standard term from equations (7) or (8) above, enabling asymptotic normal inference over the range for α. If this is not available, then $T(\hat{\alpha}-1)$ converging to a distribution would enable the weaker possibility of placing probability statements on c (e.g. confidence intervals) to restrict the range of this nuisance parameter.

When no additional stationary covariates are available [e.g. when γ is unknown in equation (1)], we can write the single equation model for y_{1t} as

$$\begin{aligned} y_{1t} &= d_{1t} + u_{1t} \\ \text{where } u_{1t} &= \alpha u_{1t-1} + v_{1t} \end{aligned} \tag{9}$$

where $\{d_{1t}\}$ are deterministic components and v_{1t} is I(0).

For this model, Dickey and Fuller (1979) show that the OLS estimate of $\hat{\alpha}$ in this autoregression when $\alpha=1$ (when v_{1t} is iid) is consistent for α at the rate T, i.e. $T(\hat{\alpha}-1)$ converges to some distribution. Cavanagh (1985), Phillips (1987), Chan and Wei (1988) extend this result for all $\alpha=1+c/T$, deriving the local to unity distributions for $T(\hat{\alpha}-1)$. Alternative estimators, such as the symmetric least squares estimate of α, also converge at rate T [Dickey, Hasza and Fuller (1984)]. No estimates of $\hat{\alpha}$ converge at the rate required to consistently select the correct local to unity distribution, a rate faster than T for all of the relevant parameter space[8]. Thus, consistent estimation of $T(\hat{\alpha}-1)$ is not available in the classical framework, so this possibility is ruled out[9].

This suggests that the most we can learn from the data about $T(\alpha-1)$ is of the form of asymptotic probability statements on this quantity, i.e. confidence intervals which will contain the true value for α with a known probability. The construction of confidence intervals for this quantity is examined and made operational in Stock (1991a), Andrews (1993) and Kiviet and Phillips (1992). Any test can be used to construct a confidence interval by inverting the test [Kendall and Stuart (1937); see Stock (1991a) for a discussion and application to the Dickey Fuller τ statistic]. Currently, the only results for the general model (general serial correlation) available are for the inversion of the Dickey-Fuller τ test and the Sargan and Bhargava (1983) tests, derived in Stock (1991a). Stock (1991a) shows that the τ test was preferred due to superior small sample performance. This raises the issue of how to select amongst different potential confidence intervals.

To obtain as much information on the range of α as possible from the observed data y_{1t}, we require the construction of the shortest interval possible[10]. When a uniformly most powerful (UMP) unbiased test exists, this can be inverted to yield a uniformly most accurate

---

[8] Hence the non standard distribution in equation (8)

[9] The problem is that c has no real meaning, but is a device for obtaining an approximate limiting distribution, so consistant estimates of this will not be available.

[10] Pratt (1961) discusses optimality concepts for confidence intervals, showing that when there exists some shortest unbiased confidence interval, then this is also optimal from the point of view of minimizing the probability that the confidence interval includes false values [as proposed by Neyman (1937)].

(UMA) confidence interval. Elliott, Rothenberg and Stock (1992) [ERS] show that for the unit root test, no such UMP test exists, thus the UMA interval can not be obtained in this way. This is also the case for null hypotheses which are local to one. ERS does obtain tests which are optimal for a given alternative (a point optimal test) which is shown to be approximately UMP in the sense that it lies almost on top of the power curve for the test of a unit root, not only for the fixed alternative used in it's derivation but also for the sequence of relevant alternatives. Whilst no optimality theory suggests that inverting such a statistic will provide optimal confidence intervals[11], the higher power of the statistics derived in ERS is suggestive of a possible result that confidence intervals constructed by inverting these statistics will be more accurate than inverting tests with lower power (as they would have a higher probability of excluding false values). The construction of such tests is currently under investigation.

In employing the results from Elliott, Rothenberg and Stock (1992), it may be considered to strong to assume that the Eu_{10}^2 is finite. The empirical results in Elliott, Rothenberg and Stock (1992) show that relaxing this assumption to the assumption that u_{10} is drawn from its unconditional distribution under the alternative affects the power of the test in Monte Carlo experiments (this is a well known feature of unit root tests in general - see the discussion in Chapter 2, section 2). The second chapter of this thesis examines this issue and rederives the results of ERS for this case, deriving almost UMP tests for the null of $\alpha=1$. The results of this paper show that the optimal power of such tests is lower than in the conditional case

---

[11] I was unable to find any optimality results for tests and confidence intervals when the test is not invariant to a nuisance parameter.

of ERS (to be expected as less information is being assumed as known). In general, however, the loss from employing these tests instead of those in ERS in the conditional case is large relative to the loss of employing the ERS tests over the Chapter 2 tests when u_{10} is drawn from its unconditional distribution[12]. The relative performance of confidence intervals constructed from inverting these tests, however, remains as yet uninvestigated.

Thus the optimal results show that for a variety of assumptions on the initial condition, the best we can do is place an asymptotic distribution on $T(\hat{\alpha}-1)$, which we denote as $\hat{c}$. Whether or not this is the optimal confidence interval in the sense that it is shortest cannot be derived from theory, and so the best we can do is examine this possibility empirically against other confidence intervals.

A recent paper by Hansen (1993) shows that additional stationary data correlated with v_{1t} in the long run can be employed in tests for a unit root increasing the power of these tests. Whilst the general assumption here is that the model is as in equation (1), extensions of the model in the direction of adding more equations may enable higher power than that obtained by the 'optimal' tests, which are optimal in the absence of such extra information. This research is new and no applications or attempts to invert the statistic for a confidence interval have been undertaken at this time.

---

[12] One could alternatively condition on this initial condition, resulting in the loss of any power that could be derived from observing the first observation. This is at great cost in power in the small samples generally available to researchers.

IV Classical Inference on γ

The result that it is possible to place an asymptotic confidence interval on T(α-1) does not in itself yield a procedure for second stage inference. This section will present some methods for classical inference on γ, taken from the results of Cavanagh, Elliott and Stock (1993) [CES]. The first method discussed does not use data on y_{1t} to help with inference, whilst the others use the results described in the previous section.

Too see how such a confidence interval for α may be useful for second stage inference, consider the sensitivity of inferences using t_γ to values of α (the construction of which is the same for all α). For some α, this test either rejects or it does not. This estimate of t_γ can be examined for a range of plausible values for α, and if the hypothesis is rejected for any of these α, then the null hypothesis is rejected (the rejection of the null hypothesis at some α and not others suggests that any rejection is fragile and depends on knowledge of α). This method constructs a confidence interval invariant to the nuisance parameter by choosing one wide enough to satisfy every possible value for this nuisance parameter[13]. This provides a maximum width confidence interval, as it assumes no knowledge of the nuisance parameter. To achieve a shorter interval, we require some method to limit the range of α. Clearly, confidence intervals on the first stage provide some way to obtain a reduction in the range over α in a strict probability sense.

---

[13] The actual choice of distributions from which the confidence interval is constructed depends on δ, and the estimate of this parameter is affected by α in small samples but not asymptotically.

Ultimately, we wish to place a classical confidence interval on the parameter of interest, γ. The usual frequentist statement for a 100(1-a)% confidence interval for γ can be written

$$P_{\gamma}[\ \underline{\gamma}(y) \leq \gamma \leq \bar{\gamma}(y)\ ] \geq (1-a) \quad \forall \alpha \tag{10}$$

where $\underline{\gamma}(y)$ and $\bar{\gamma}(y)$ are lower and upper bounds for γ as a function of the data y_t, (1-a) is the confidence level, and this probability statement must hold for the entire relevant range of α. To construct such an interval that holds for all α in the relevant range, we require a test for $\gamma=\gamma_0$ which has size not larger than a for all relevant values of the nuisance parameter α. The statistic we will examine here is t_γ, presented in section 2 above.

If we regard the relevant range for α to be such that $-40<c<10$, then percentiles of t_γ can be calculated using Monte Carlo methods. Percentiles of t_γ for a range of values for δ are reported in Figure 3 in CES. As these distributions vary with α, without any extra information on the range for α, then a 90% confidence interval can be constructed by taking the minimum (over c) value of the 5th percentile and the maximum value of the 95th percentile. For any fixed value for α in this range, then the probability statement in equation (10) holds; this is simply the procedure above. The confidence interval for γ can then be calculated in the usual way[14] using these alternative critical values; this is derived in CES and called the sup-bound interval.

This interval is quite conservative, in the sense that for any fixed value of α, the method will

[14] i.e. the confidence interval limits are $[\hat{\gamma} - \bar{d}_a se(\hat{\gamma}), \hat{\gamma} + \underline{d}_a se(\hat{\gamma})]$, where $se(\hat{\gamma})$ is calculated according to the discussion following equation (2), and $\bar{d}_a$ and $\underline{d}_a$ are the upper and lower critical values of the test derived as in the text.

fail to reject estimates of t_γ that would be rejected if α were known. For example, if $T(\alpha-1)=-10$ and $\delta=0.5$ (and d_{2t} is a constant), the (symmetric) sup-bound critical values are $\bar{d}_{0.5}=1.96$ and $\underline{d}_{0.5}=-2.77$, whereas if α were known, then the asymptotic upper and lower critical values from figure 2 are 1.52 and -2.44 respectively. Note, however, that this error from not knowing α is different to those mentioned in previous sections in that size of the test is below stated size rather than above it (this is with probability one, as the α known intervals all lie inside the sup bound interval by construction). Thus, Type I error is controlled to be less than the stated level as is desired in classical inference, and the probability statement given in equation (10) above holds for all α.

This procedure, however, ignores any information that can be obtained from the data on α. The previous section shows that the best we can do is put a confidence interval on this nuisance parameter. CES shows that this information on α can be employed to aid second stage inference using a Bonferroni approach. As discussed in the previous section, tests for $\alpha=1$ can be inverted to form confidence intervals on α. Restricting attention to α inside a first stage confidence bound of level $(1-a_1)$, the second stage critical values can be constructed by examining the percentiles of t_γ that would result in a second stage level $(1-a_2)$ confidence interval. The outer most extreme points of these percentiles give conservative critical values for this restricted range. The Bonferroni inequality then tells us that the size of the Bonferroni test is no greater than $a=a_1+a_2$. Using these critical values results in a classical confidence interval for γ which satisfies equation (10) above, although the results of CES show that this is still a quite conservative test. Results show that whilst this test loses power over the α known case, these power losses are not too extreme. In any case, the α known interval is not feasible.

The actual approach used in CES was to employ the results of Stock (1991) and invert the Dickey-Fuller t statistic to obtain a first stage confidence interval. From the graph of the percentiles of the t_γ statistic in figure 2, it is clear that we wish to limit the range over which α varies as much as possible. The Bonferroni approach allows us to do this directly by increasing a_1, but this comes at a cost of decreasing a_2, thus widening the second stage critical bounds. Results from the previous section suggest that more accurate first stage intervals may be constructed by inverting tests of $\alpha=1$ that are closer to being UMP, or at least more powerful than the Dickey Fuller test.

CES also examines other classical approaches using the joint test over α and γ, although this paper finds that the Bonnferroni and Sup-Bounds tests perform best in terms of power.

The only other regression based approach to this problem is contained in extensions of the fully modified regression approach introduced by Phillips and Hansen (1989). These extensions are found in Kitamura and Phillips (1992) and Phillips (1993a,1993b). These papers extend the procedure to models where the order of the cointegrating space is unknown, without the loss of unbiasedness and chi-square inference. They require the construction of an orthogonal dependant variable vector, such as (1-αL) in equation (6) above, constructed in the method of Phillips and Hansen (1989), i.e. they impose that the largest root of the regressor is one. This term is entered into the regression with a weight estimated non parametrically. In the case of Phillips (1993a), this requires using the first difference of all of the data and relying on the result that if y_{1t} were truly stationary, then its difference is I(-1) and is op(1), and the non parametric correction when the data is stationary disappears asymptotically and so inclusion of this term does not affect the results[15]. When α is considered fixed (whether equal to one or less than one), then this results in chi-square inference - in the method of Phillips and Hansen if the data is I(1) or by usual stationary CLT results if the data is I(0). Thus, this method treats variables with their largest root equal to α where α is close to one as stationary variables. The central result from local to unity asymptotics is that for such values of α, the asymptotic distribution resulting from considering large values of α as fixed is not a good guide to the types of distributions seen with reasonable amounts of data, but instead the distribution resulting from considering c fixed does result in an asymptotic distribution which appears relevant. This suggests that for persistant data, inference using these methods will also result in size distortions. The extent to which this bias appears in practice for these techniques has not yet been investigated.

V Other Approaches to Inference on γ

The above discussion, and the focus of this thesis, has limited itself to classical (frequentist) testing of the hypothesis of interest. This is indeed the approach apparently preferred by the majority of researchers employing time series theory, as is seen by noting that an extremely

---

[15] To achieve this, very specific controls are required to be placed on the speed at which covariances are added in the construction of the non parametric estimates of the spectral density of the residuals at frequency zero.

large proportion of published papers use purely frequentist methods. This is not to say that such methods are thus most relevant for inferences on γ, but it does appear that a complete understanding of the classical properties of such tests in economically relevant theoretical models should be a high priority if the econometrician is to guide empirical practice. Of course, it may be that other methods solve the problems outlined above in a way that is acceptable to researchers, thus making the examination of these alternate methods also extremely interesting. This section examines three such alternative approaches with a view to assessing their applicability to this problem: these are non parametric methods, bootstrapping, and Bayesian methods.

Non-Parametric Approaches

Campbell and Dufour (1991,1993) have examined hypothesis testing on γ in the second equation in (1), particularly as regards orthogonality tests, using Wilcoxen type non parametric (rank and signed-rank) statistics. These statistics are based around quantities such as

$$S_g = \sum_{t=1}^{T} u[(y_{2t} - \gamma_0 y_{1t-1}) y_{1t-1}] \tag{11}$$

where $u[z]=0$ if $z<0$ and one otherwise. If $\gamma=\gamma_0$, then under the extra conditions that y_{1t} and y_{2t} are mean zero and there is no serial correlation in $(y_{2t}-\gamma_0 y_{1t})$ then this has an exact binomial distribution [Campbell and Dufour (1993), Proposition 1]. This distribution is derived and stated in Campbell and Dufour (1993). They go on to examine signed rank tests and other similar quantities as in equation (11). They show that these tests have excellent small sample properties in Monte Carlo experiments with a range of assumptions on the distribution of the residuals v_{2t}.

The major caveats to use of these statistics is that the requirements of mean zero variables and no serial correlation are binding. The theoretical results rely on u[.]=1 having a probability of occurring of one half for each observation, however with serial correlation or non mean zero data this will not be the case. Serial correlation leads to 'runs' of ones and zeros, as quantities such as those in the argument of u[.] in equation (11) stay away from the true zero mean for a number of periods. This probability will also be incorrect if the quantity in the argument does not have exactly mean zero, as would be the case when the data is not mean zero and the true means are unknown.

To get around the problem of serial correlation, the authors propose splitting up the sample. For example, if the residuals are known a priori to follow an MA(1) process, then taking every second observation would result in a serially uncorrelated sample. The problems with this are twofold. Firstly, even in this simple case, half of the observations are lost, which will result in extreme power losses in the types of samples typically found in macroeconomics and finance. Secondly, it is rare to know the order of serial correlation of the residuals, making such a fix unoperational.

This test can be successfully employed in cases where the joint null of $\gamma=\gamma_0$ and no serial correlation is of interest to the applied researcher, as it will have power in both directions. In most macroeconomic and finance applications, however, we are not really interested in the null of no serial correlation. The results of Campbell and Dufour do suggest that non parametric approaches to this problem warrant further investigation - their approach deals successfully with δ non zero and obtains exact finite sample results for a wide range of values of α and distributional assumptions on v_t.

Bootstrap Approaches

There has apparently been no work so far in evaluating the possibility that Efrons' (1979) bootstrap can be successfully applied to this general problem of estimation and testing hypotheses over γ. The discussion here will examine the apparent lack of success of the bootstrap in a simpler problem, that of inference on the first stage estimation of the autoregression, and draw conclusions from this for the problem at hand.

In the special case where $\delta=1$ and no serial correlation in the residuals v_t, (i.e. $v_{1t}=v_{2t}$), then the model in equation (1) with $s=1$ is such that $\gamma=\alpha$ and both equations are identical. In this case, with the additional assumption of $d_{1t}=0$, a number of papers have examined the application of the bootstrap to estimation of α. For fixed $|\alpha|<1$, Bose (1988) shows that the standard bootstrap estimator of $\hat{\alpha}$ is asymptotically valid, in that it replicates the correct asymptotic normal distribution. Raynor (1990) presents Monte Carlo experiments which correspond to these results. Basawa et al (1989) show the same result for fixed $|\alpha|>1$, i.e. the explosive case. Basawa et al (1991) consider the case of $\alpha=1$, and show that the parametric bootstrap distribution (where $\epsilon_{1t}\sim N(0,1)$, and the bootstrap samples for ϵ_t are drawn from a standard normal distribution) is not asymptotically equivalent to the true distribution. They consider the correct asymptotic distribution τ_α, and show that the asymptotic distribution of the t test testing the bootstrap estimate of α is not equivalent to τ_α.

The intuition for this result follows from the results on estimating α in section 3 above. The bootstrap estimates of y_{1t}, denoted by y_{1t}^*, are estimated by cumulating by the equation $y_{1t}^* = \hat{\alpha} y_{1t-1}^* + \epsilon_{1t}^*$, where ϵ_{1t}^* is constructed from draws (with replacement) from the estimated errors of the process. But from the results of the local to unity literature for the distribution of τ_c [Cavanagh (1985), Phillips (1987), Chan and Wei (1987)], we know that the limiting behavior for the t statistic here depends on the actual value $\hat{\alpha}$ used to cumulate the bootstrap residuals to obtain the bootstrap data y_{1t}^*. Thus, the bootstrap considered here cannot replicate the first order asymptotics for τ_α. Basawa et al (1991) indicate that their result holds for all estimators of $\hat{\alpha}$ which converge at rate T (hence this result covers the local to unity case as well).

Ferreti and Romo (1993) present theoretical and empirical results which show that the bootstrap can, however, be employed to test the unit root hypothesis. They make two changes to the bootstrap design considered above. First, upon obtaining ϵ_{1t}^* as above, they demean the residuals. Second, they construct y_{1t}^* under the null hypothesis of $\alpha=1$, i.e. they use the recursion $y_{1t}^* = y_{1t-1}^* + \epsilon_{1t}^*$. From the intuition above, this second design change circumvents at least part of the problems involved with employing the bootstrap. They present Monte Carlo results which indicate that this bootstrap test has similar size properties in finite samples as the Dickey-Fuller test, and possibly enable a slight gain in power over these tests (this may be Monte Carlo error).

Some conclusions for the potential of applying the bootstrap in examining t_γ can be drawn from these results. If the asymptotic theory for $|\alpha|<1$ (or >1) were relevant, then the bootstrap could potentially be applied as it works in the first stage. However, this is the case where it is not needed, as first order asymptotics work well. Potentially, it would allow some finite sample gains, as it does in the first order AR case [Raynor (1990)]. For the problem at hand, however, we are restricted to the range where estimates for α converge at rate T. Hence, if α were known (say we had a null hypothesis for α), then the results of Ferreti and Romo (1993) suggest that the bootstrap may be applicable. Again, if α were known, the first order asymptotic theory for t_γ is known so there is no real need for the bootstrap. In the case of α unknown, it follows fairly directly from the results of Basawa et al (1991) that the bootstrap will not help: bootstrapped data must be cumulated using the estimated value $\hat{\alpha}$ rather than the correct α so the bootstrap would be invalid.

Bayesian Methods

As in the case of the bootstrap approach, the Bayesian methods have apparently not been applied to these particular models, although they have been applied to the first stage (unit root) problem by itself (which as noted above, is a special case of the second stage). Unlike the bootstrap, Bayes methods do, however, have a justified (from the Bayesian point of view) solution to this problem which would at one level simply entail placing a prior distribution over the nuisance parameters of the model (which is to some extent implicitly done in the classical methods by restricting attention to values of α where the local to unity specification applies). In fact, Bayes methods effectively would circumvent the problem by conditioning directly on the y_{1t} sequence that happened to be observed. As is the usual case for the differing approaches of classical and Bayesian methods, the differing techniques reflect differing views on probability and the experiment being undertaken [e.g. see Rothenberg (1983)]. No attempt will be made here to examine these types of arguments, this section will review what is known about Bayesian solutions in the types of models under investigation, the implications results have for empirical modelling, and some conjectures as to how to provide more information on these points. These conjectures are borne out by a small Monte Carlo experiment.

Whilst no literature has directly considered Bayesian inference in models such as on γ in equation (1), two sets of literature are relevant. Firstly, in the case of tests for a unit root, a number of papers have examined Bayesian solutions [Dejong and Whiteman (1991), Phillips (1991), Sims and Uhlig (1991), Uhlig (1992)]. The general results from this literature are that the prior distribution chosen matters asymptotically, that different 'uninformative'[16] priors over α yield different asymptotic results, and that these results are different from those obtained using classical inference. In particular, the unit root hypothesis is rejected far more often in Bayesian analysis.

Phillips (1991) argues that the flat priors usually employed by the Bayesians are not

---

[16] As the prior matters asymptotically for the posterior, different priors attempting to be uninformative over the space for α result in different posterior distributions for α even in large samples, and so are actually not uninformative in practice.

uninformative in autoregresssions and that they bias the results towards rejecting a unit root model. He argues that Jeffreys priors are more relevant, and that these give results closer to the classical result. Kim and Maddala (1991) use Monte Carlo methods to show that the Jeffreys prior gives high weight to roots close to and above unity. Uhlig (1992) shows that to a great extent the differing results over different priors is due to the weight assigned by the prior on explosive roots; if the parameter space is restricted to disallow explosive results then posterior distributions are more similar over different priors. Sims and Uhlig (1991) employ flat priors and show that conditioning on $\hat{\alpha}$, that the marginal distribution for α has a normal distribution, thus p values for the unit root hypothesis will be larger for Bayesian solutions.

Secondly, there has been much work examining Bayesian vector autoregressions (BVAR) of the form $y_t = a(L)y_{t-1} + b(L)x_{t-1}$ for use in forecasting the US macroeconomy [e.g. Litterman (1986)], usually under the mean restrictions (smoothness priors) of $a(1)=1$ and $b(L)=0$. Different mean restrictions would replicate equation (1) with $s=1$, so this literature is related to the question here. The focus of these studies has not been directed at the stochastic properties of the explanatory variables. All efforts regarding the BVAR have focussed on estimation rather than inference over the parameter space, so no lessons are available.

From the unit root results, where the classical results involve non standard distributions whilst the Bayesian ones with uniform priors do not, and the result that t_γ has a non standard distribution driven by the size of α, it is conjectured that the Bayesian and classical results for inference over $\hat{\gamma}$ will also differ. This result would suggest that conclusions drawn depend on the thought experiment being undertaken -a problem for which no ready answer is available when there is no general agreement as to the correct experiment.

The extent of the difference between classical and Bayesian examinations of γ would be expected to be smaller than that found in the examination of a unit root. This follows from the intuition that the unit root case is and extreme bound on the model considered here (where in the unit root case $\delta=1$). A Monte Carlo experiment can be employed to examine the extent of the difference. Here, the model in equation (4) is examined, with the additional assumption of iid normal errors. The model can be rewritten so that the residuals are orthogonal to the regressors so the second equation becomes that in (6). To calculate the posterior distribution for γ, we require prior distributions over the parameters. Here, rewrite the model as

$$\begin{aligned} y_{1t} &= d_1 + \alpha y_{1t-1} + \epsilon_{1t} \\ y_{2t} &= d_2 + \beta_1 y_{1t} + \beta_2 y_{1t-1} + \eta_t \end{aligned} \tag{12}$$

where $\varphi = \beta_1$ and $\gamma = \beta_2 + \varphi\alpha$. Uniform priors were placed over $(\alpha, \beta)'$ [where $\beta=(\beta_1\ \beta_2)'$], with the prior on α bounded between 0.6 and 1.1. The orthogonality of the error terms was treated as known[17].

The posterior distribution for γ can be estimated by Monte Carlo. For any realised dataset y_t, the posterior distribution for α (without the trucation) is $N(\hat{\alpha}, var(\hat{\alpha}))$. The truncation

[17] For the analytic results for the posterior distributions, the variances of the errors were treated as known, although in the simulations they were estimated. This is likely to have only a small effect on the results.

affects this by removing probability mass in this distribution outside of the truncation points. The posterior distibution for ß is N($\hat{\beta}$, var($\hat{\beta}$)), where the two normal distributions are independant as the residuals for each equation are independant. From these distributions, the posterior distribution for γ can be constructed by the formula relating γ to (α ß) and numerical reaslisations of the normal distributions.

The results can be evaluated from a frequentist perspective, or alternatively, can be used to examine frequentist results from a Bayesian perspective. The equal tailed 95% Bayesian confidence interval can be calculated from the simulated posterior distribution. For interpretation from a frequentist perspective, we would desire such an interval to contain the true value for γ for 95% of datasets constructed according to the true model. In 5000 replications, the Bayesian confidence interval calculated as above[18] had a coverage rate of 89%[19].

Alternatively, one could examine the classical confidence intervals from a Bayesian perspective. This would involve examination of the posterior for γ which is covered by the frequentist confidence interval. For each dataset, (here replication in the Monte Carlo), the posterior probability mass contained in the classical confidence interval can be calculated.

---

[18] Roughly 5000 simulations of the normal distributions were employed to compute the posterior distribution for γ.

[19] The bounding of the prior on α had little effect here. For the model examined the coverage rate without bounds on α was 88%. These coverage rates are sensitive to the model examined. For $\alpha=1$ and $\delta=0.99$, a model close to the unit root case, the coverage rate of the Bayesian confidence interval was 69%. In this model, the posterior coverage of the Bonferroni confidence interval was lower in the sense that on more occasions this coverage was less than 95%.

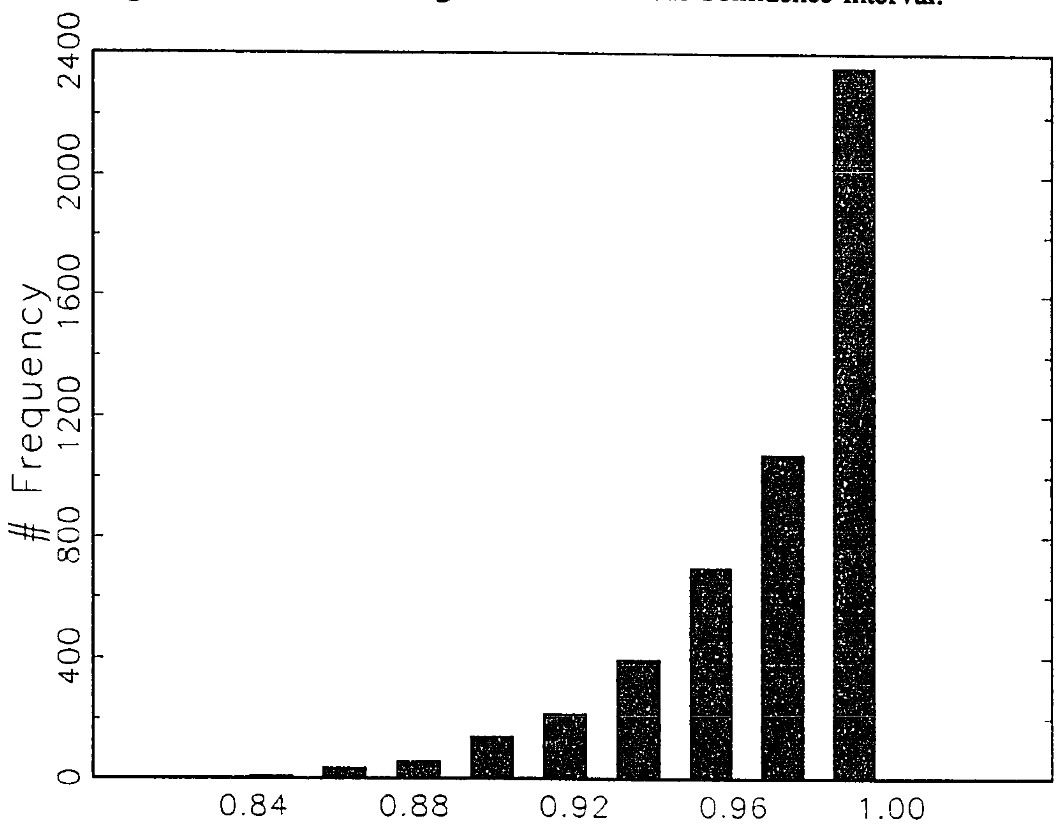
The frequencies of various coverage rates over many datasets can then be reported. Using the Bonferroni confidence interval described in section 4 above[20], this result is reported in Figure 2. Around 80% of the Bonferroni confidence intervals contained over 95% of the posterior probability of γ; almost half of the coverages are very close to one. The smallest posterior coverage of the 5000 replications was over 80%.

The interpretation of these results is unclear. Given the flat prior, there is a difference between Bayesian and frequentist results. For some datasets, different conclusions will be drawn for γ. This was as conjectured above. The coverage rate of the Bayesian confidence interval for γ is distorted less than the Bayesian confidence interval for α. With a time trend included in the specification, Stock (1991b) reports that the Sims and Uhlig (1991) 95% Bayesian confidence interval for α contains the true α 39% of the time when $\alpha=1$ (they employ similar priors as the Monte Carlo experiment above). With the time trend removed, this coverage rate becomes 78%.

---

[20] The confidence interval is calculated with a 1% first stage size and 4% second stage size, yeilding a level 95% confidence interval.

Figure 2: Posterior Coverage of the Bonferroni Confidence Interval.



Notes: The figure presents a histogram of posterior coverage rates for the Bonferroni confidence interval from the Monte Carlo experiment described in Section 5. The vertical axis measures the frequency of various coverage rates. The mid point of each coverage range is given on the horizontal axis.

One possibility for some reconciliation of the results would be to examine what types of priors lead Bayesian results to look like classical results. Sims and Uhlig (1991) invert the classical distribution for $\hat{\alpha}$ to calculate the prior and show that it appears unreasonable from a Bayesian perspective. This approach could also be taken as regards the Bonferroni intervals for $\hat{\gamma}$ derived above. As the coverage rate of the Bayesian confidence interval is not greatly distorted for the model considered here, it is probably the case that the priors required for justifying classical results from a Bayesian veiwpoint may not differ to far from priors considered reasonable by Bayesian analysts. Of course, even if they do differ, this does not motivate frequentists to change behavior as they believe that the wrong thought experiment is being undertaken.

VI Summary

The following chapters presented in this thesis represent part of the work done by the author in attempting to understand the problem at hand, and its implications for empirical work. On their own, each chapter examines a small piece of this puzzle. Other papers written which provide information on the questions raised earlier are Elliott and Stock (1992), Elliott, Rothenberg and Stock (1992), and Cavanagh, Elliott and Stock (1993). Each of these papers is referenced above at the appropriate point.

The following chapter, chapter 2, examines optimal tests for a unit root when the initial condition, usually assumed to be fixed, is instead drawn from its unconditional distribution under the alternative hypothesis. The results of this paper have implications for the above analysis in that they provide new tests for a unit root which work well in certain situations. The theory presented enables comparison with the statistics in ERS and their performance in this alternative case. In addition, the statistics derived have good properties compared to the power bounds, and can be inverted for confidence intervals for the first stage in the Bonferroni method. Also, the null hypothesis of a unit root is interesting in its own right [see Stock (1994) for a discussion of the uses of unit root tests].

Chapter 3 examines the popular cointegration estimator techniques put forward recently and examines their performance in the case where α is unknown, but assumed to be one. Such estimation techniques have recently become extremely popular. The results of this chapter show quite clearly the problem confronting applied researchers testing long run theories, whilst roots are apparently very close to one so normal asymptotics are not good guides for inference, these methods which are asymptotically efficient when $\alpha=1$ can have huge size distortions when α is close to but not exactly one. This result is shown both analytically and with Monte Carlo experiments. Although this chapter does not explicitly examine the role of pretesting for a unit root, we know from the results of Elliott and Stock (1992) that pretests will asymptotically misclassify local to unit roots as unit roots, so pretesting will not rectify this problem.

The fourth chapter examines forward market unbiasedness in the yen/dollar foreign exchange market. Under the null hypothesis of unbiasedness (see the chapter for details), the forward exchange rate should be an unbiased predictor of the future spot rate. Allowing y_{2t} to be the spot rate, and y_{1t-1} to be the forward rate, this suggests that $\gamma=1$ under the null hypothesis.

This has been rejected in recent examinations using the cointegration framework. Chapter 4 shows that these rejections are conditional on the assumption of a unit root in the exchange rate, and the hypothesis cannot be rejected if this is not assumed a priori. Thus, it is shown that the types of theoretical problems discussed in this thesis have real implications, in that they overturn previous results. This chapter also presents other attempts at distinguishing hypotheses, notably distinguishing rational expectations from static expectations, using the local to unity framework and asymptotic theory used in this thesis.

Appendix 1: Applications

It was mentioned above that it is rare for economic theory to suggest a value for α in models such as equation (1). In the absence of these types of theory, the researcher is required to take some sort of stand, explicitly or implicitly, on the stochastic behavior of y_{1t} (i.e. α, which asymptotically dominates this stochastic behavior). This is not just confined to cases such as above, where the model remains the same and different asymptotics are used, but to all applications whether or not the decision is to difference the data to obtain stationarity, or to employ cointegration techniques, or even to undertake permanent/transitory decompositions. To give some idea of the pervasiveness of this decision, examples from recent literature are briefly described.

1. Using Pre-Tests for a Unit Root

In recent work this has been the most popular approach, although often there is no real reason why the researcher is controlling type I error in this way. This approach, when the test fails to reject, leads to either differencing or cointegration analysis. When rejected, normal asymptotics are employed.

Examples in macroeconomics include Stock and Watson (1993), who estimate cointegrating vectors for money demand and Ogaki (1992), who models dissagregated consumption using cointegration methods after pretesting for a unit root. Huag (1991) examines the cointegrating relationship between the government surplus and bonds outstanding using cointegrating techniques after testing for a unit root. Hallman, Porter and Small (1991) use pretests for a unit root in inflation and velocity to suggest the regressing of the change in inflation (the unit root hypothesis was not rejected here) on lags and a constant (proxying for velocity, for which a unit root was rejected). Mehra (1991) estimates a VAR using the price level, productivity adjusted wage and output gap after differencing for stationarity according to unit root pretests. Alogoskoufis and Smith (1991) regress the change in wages on the expected change in prices where this is proxied by lagged changes after ADF pretests suggest that this variable is I(1), and hence can be modelled in differences.

In finance, Hardevoulis (1990) regresses expected returns on the stock market on the lagged dividend price ratio after testing the later for a unit root using ADF tests (even though his stated null is that this variable is I(0)).

In international economics, Clarida (1994) pretests log imports, the log of consumption of domestic goods and the log of the real price of imports for unit roots and upon failure to reject uses them for cointegration analysis. Burda and Gerlach (1992) do the same for the log of real imports, real permanent income, the relative price of consumer non durables and a constructed intertemporal price series, although they do not conduct hypothesis tests on their cointegrating vector. Evans and Lewis (1993), in examining forward market unbiasedness as discussed in section 6 above use previously obtained unit root test results from the literature to justify their cointegrating vector approach. MacDonald and Taylor (1991) use ADF pretests and then undertake hypothesis tests on cointegrating vectors in examining relationships between various interest rates. Choudry, McNown and Wallace

(1991) do the same in testing for long run purchasing power parity (without hypothesis tests on the cointegrating vector). Bohara and Kaempfer (1991) regress a system including real GNP, the average tariff rate and a number of other variables in a differenced VAR after deciding the series were I(1) based on inspection of autocorrelations.

2. Using Asymptotic Normal Distribution

Often, the normal distribution is applied directly with no pretest. This can be due to the belief that the y_{1t} variables are stationary, or that $\delta=0$.

In macroeconomics, Kahn (1992) regresses sales of automobiles on factor prices without pretests, using standard normal asymptotics. He presents some attempts at correcting for simultaneity in the current period.

Examples in finance include Hamilton (1992) who examines the predictability of excess returns in three commodities futures markets around the time of the depression using variables such as the interest rate, lagged spot rate and lagged futures rate. No pretesting is undertaken.

3. Using $\alpha=1$ Nonstandard Distribution

Examples in macroeconomics include Friedman and Kuttner (1992), who look at the relationships between money, income, prices and the interest rate. Cointegration analysis is performed.

Bibliography

Ahtola, J. and G.C.Tiao (1984), "Parameter Inference for a Nearly Nonstationary First Order Auoregressive Model", *Biometrica*, 71, 263-72.

Alogoskoufis, G.S. and R.Smith (1991), "The Phillips Curve, the Persistance of Inflation and the Lucas Critique: Evidence from Exchange-Rate Regimes", *American Economic Review*, 81, 1254-75.

Andrews, D.W.K (1993), "Exactly Median Unbiased Estimation of First Order Autoregressive/Unit Root Model", *Econometrica*, 61, 821-56.

Athreya, K.B. (1987), "Bootstrap of the Mean in the Infinite Variance Case", *Annals of Statistics*, 15, 724-31.

Banerjee, A. and J.Dolado (1987) "Do we Reject Rational Expectations Models too Often? Interpreting Evidence using Nagar Expansions", *Economics Letters*, 32, 19-24.

Basawa, I.V., A.K.Mallik, W.P.McCormick, and R.L.Taylor (1989), "Bootstrapping Explosive Autoregressive Processes", *Annals of Statistics*, 17, 1479-86.

Basawa, I.V., A.K.Mallik, W.P.McCormick, J.H.Reeves, and R.L.Taylor (1991), "Bootstrapping Unstable First-Order Autoregressive Processes", *Annals of Statistics*, 19, 1098-1101.

Bobkoski, M.J. (1983), Hypothesis Testing in Nonstationary Time Series, unpublished Ph.D. thesis, Department of Statistics, University if Wisconsin.

Bohara, A.K. and W.H.Kaempfer (1991), "A Test of Tarriff Endogeneity in the United States", *American Economic Review*, 81, 952-60.

Bose, A. (1988) "Edgeworth Correction by Bootstrap in Autoregressions" *Annals of Statistics*, 16, 1709-22.

Burda, M.C. and Gerlach, S (1992), "Intertemporal Prices and the US Trade Balance", *American Economic Review*, 82, 1234-53.

Campbell, B and J-M Dufour (1991), "Over-Rejections in Rational Expectations Models: A Non Parametric Approach to the Mankiw-Shapiro Problem", *Economics Letters*, 35, 285-90.

Campbell, B and J-M Dufour (1993), "Exact Non-Parametric Orthogonality and Random Walk Tests", C.R.D.E Working Paper #1893, Universite de Montreal, August.

Campbell, J.Y. and P.Perron (1991), "Pitfalls and Opportunities: What Macroeconomists

Should Know about Unit Roots", NBER Macro Annual.

Campbell. J.Y. and R.J.Shiller (1987), "Cointegration and Tests of Present Value Models", *Journal of Political Economy*, 95, 1062-88.

Cavanagh, C. (1985), "Roots Local to Unity", manuscript, Department of Economics, Harvard University.

Cavanagh, C.L., G. Elliott and J.H.Stock (1993), "Inference in Models with Nearly Nonstationary Regressors", manuscript, Department of Economics, Harvard University.

Chan, N.H. (1988), "On Parameter Inference for Nearly Nonstationary Time Series", *Journal of the American Statistical Association*, 83, 857-62.

Chan, N.H. and C.Z.Wei (1987), "Asymptotic Inference for Nearly Nonstationary AR(1) Processes", *Annals of Statistics*, 15, 1050-63.

Choudry, T., R.McNown and M.Wallace (1991), "Purchasing Power Parity and the Canadian Float in the 1950s", *Review of Economics and Statistics*, LXXIII, 558-563.

Clarida, R.H. (1994), "Cointegration, Aggregate Consumption, and the Demand for Imports: A Structural Econometric Investigation", American Economic Review, 84, 298-308.

Corbae, D. K-G Lim, and S.Ouliaris (1992), "On Cointegration and Tests of Forward Market Unbiasedness", *Review of Economics and Statistics*, 728-32.

DeJong, D.N. and C.H.Whiteman (1991), "Reconsidering 'Trends and Random Walks in Macroeconomic Time Series'", *Jounal of Monetary Economics*, 28, 221-54.

Dickey, D.A. and W.A.Fuller (1979), "Distribution of Estimators for Autoregressive Time Series with a Unit Root", *Journal of the American Statistical Association*, 74, 427-431.

Efron, B. (1979), "Bootstrap Methods: Another Look at the Jackknife", *Annals of Statistics*, 7, 1-26.

Elliott, G., T.J.Rothenberg, and J.H.Stock (1992), "Efficient Tests for a Unit Root", NBER Technical Working Paper #130.

Elliott, G. and J.H.Stock (1992), "Inference in Time Series Regression When the Order of Integration of a Regressor is Unknown", NBER Technical Working Paper No. 122, forthcoming Econometric Theory.

Engle, R.F. and C.W.Granger (1987), "Cointegration and Error Correction: Representation, Estimation and Testing", *Econometrica*, 55, 251-76.

Evans, G.B.A. and N.E.Savin (1981), "Testing for Unit Roots: 1", *Econometrica*, 49, 753-79.

Evans, G.B.A. and N.E.Savin (1984), "Testing for Unit Roots: 2", *Econometrica*, 52, 1241-69.

Evans, M.D.D. and K.K.Lewis (1993), "Do Long Term Swings in the Dollar Affect Estimates of the Risk Premia?", Stern School of Business Discussion Paper EC-93-12, New York University, June.

Fama, E.F. (1965) "The Behavior of Stock Market Prices", *Journal of Business*, 38, 34-105.

Ferreti, N. and J.Romo (1993), "Bootstrap Tests for Unit Root AR(1) Models", Unpublished Working Paper 93-22, Universidad Carlos III de Madrid.

Friedman, B.M. and K.N.Kuttner (1992) "Money, Income, Prices and Interest Rates", *American Economic Review*, 82, 472-92.

Fuller, W.A. (1976) *Introduction to Statistical Time Series*, New York: Wiley.

Granger, C.W.J. (1969) "Investigating Causal Relations by Econometric Models or Cross-Spectral Methods", *Econometrica*, 37, 424-38.

Hall A.D., H.M.Anderson and C.W.J.Granger (1992), "A Cointegration Analysis of Treasury Bill Yields", *Review of Economics and Statistics*, 74, 116-26.

Hall, R.E. (1978), "Stochastic Implications of the Life Cycle-Permanent Income Hypothesis: Theory and Evidence", *Journal of Political Economy*, 86, 971-87.

Hallman, J.J., R.D.Porter and D.H.Small (1991), "Is the Price Level Tied to the M2 Monetary Aggregate in the Long Run", *American Economic Review*, 81, 841-58.

Hamilton, J.D. (1992), "Was the Deflation During the Great Depression Anticipated? Evidence from the Commodities Futures Market", *American Economic Review*, 82, 157-58.

Hansen, B.E. (1993), "Rethinking the Univariate Approach to Unit Root Testing: Using Covariates to Increase Power", manuscript, University of Rochester.

Hansen, L.P. (1982) "Large Sample Properties of Generalized Method of Moments Estimators", *Econometrica*, 50, 1029-54.

Hardevoulis, G.A. (1990), "Margin Requirements, Volatility and the Transitory Componant of Stock Prices", *American Economic Review*, 80, 736-62.

Hoffman, D.L. and R.H.Raasche (1991), "Long-run Income and Interest Elasticities of

Money Demand in the United States", *Review of Economics and Statistics*, 73, pp665-74.

Huag, A.A. (1991), "Cointegration and Government Borrowing Constraints: Evidence for the Unitied States", *Journal of Business and Economic Statistics*, 9, 97-101.

Johnson, D. (1993) "Unit Roots, Cointegration and Purchasing Power Parity: Canada and the United States, 1870-1991", in B.O'Reilly and J.Murray (eds) The Exchange Rate and the Economy, forthcoming, Ottowa: Bank of Canada.

Kahn, J.A. (1992) "Why is Production more Volatile than Sales? Theory and Evidence on the Stockout-Avoidance Motive for Inventory-Holding", *Quarterly Journal of Economics*, CVII, 481-510.

Kendall,M.G. and A. Stuart (1967), *The Advanced Theory of Statistics, Volume 2*, second edition, Charles Griffin: London.

Kim, I-M and G.S.Maddala (1991), "Flat Priors vs Ignorance Priors in the Analysis of the AR(1) Model", *Journal of Applied Econometrics*, 6,375-80.

Kiviet, J.F. and G.D.A. Phillips (1992), "Exact Similar Tests for Unit Roots and Cointegration", *Oxford Bulletin of Economics and Statistics*, 54, 349-68.

Kugler, P. and C.Lenz (1993), "Multivariate Cointegration Analysis and the Long Run Validity of PPP", *Review of Economic and Statistics*, 75, pp 180-184.

Litterman, R.B. (1986), "Forecasting with Bayesian Vector Autoregressions - Five Years of Experience", *Journal of Business and Economic Statistics*, 4, 25-38.

MacDonald, R and M.P.Taylor (1991) "Exchange Rates, Policy Convergence, and the European Monetary System", *Review of Economics and Statistics*, LXIII, 553-558.

Mankiw, N.G. and M.D.Shapiro (1985) "Trends, Random Walks, and Tests of the Permanent Income Hypothesis", *Journal of Monetary Economics*, 16,165-74.

Meese, R.A. and K.J.Singleton (1982), "On Unit Roots and the Empirical Modeling of Exchange Rates", *Journal of Finance*, 37, pp 1029-35.

Mehra, Y.P. (1991), "Wage Growth and the Inflation Process: An Empirical Note", *American Economic Review*, 81, 931-37.

Nelson, C.R. and C.I.Plosser (1982), "Trends and random Walks in Macroeconomic Time Series; Some evidence and Implications", *Journal of Monetary Economics*, 10, 139-62.

Neyman, J. (1937), "Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability", Ogaki, M. (1992), "Engels Law and Cointegration", *Journal of Political Economy*, 100, 1027-46.

Pratt, J.W. (1961), "Length of Confidence Intervals", *Journal of the American Statistical Association*, 56, 549-67.

Phillips, P.C.B. (1987), "Towards a Unified Asymptotic Theory for Autoregression", *Biometrica*, 74, 535-47.

Phillips, P.C.B. (1991a), "Optimal Inference in Cointegrated Systems", *Econometrica*, 59, 283-306.

Phillips, P.C.B. (1991b), "To Criticise the Critics: An Objective Bayesian Analysis of Stochastic Trends", *Journal of Applied Econometrics*, 6,333-64.

Phillips, P.C.B. (1993a), "Fully Modified Least Squares and Vector Autoregression", manuscript, Yale University.

Phillips, P.C.B. (1993b), "Robust Nonstationary Regression", manuscript, Yale University.

Raynor, R.K. (1990), "Bootstrapping p Values and Power in the First-Order Autoregression: A Monte Carlo Investigation", *Journal of the American Statistical Association*, 8, pp 251-263.

Rothenberg, T.J. (1983) "The Bayesian Approach and Alternatives in Econometrics - II", in S.E.Fienberg and A.Zellner eds. *Studies in Bayesian Econometrics and Statistics*, North Holland: Amsterdam, pp 55-67.

Sims, C.A. (1991), "Comment by Christopher A. Sims on 'To Criticies the Critics' by Peter C.B. Phillips", *Journal of Applied Econometrics*, 6,423-34.

Sims, C.A. and H.Uhlig (1991), "Understanding Unit Rooters: A Helicopter Tour", *Econometrica*, 59, 1591-99.

Stock, J.H. (1987), "Asymptotic Properties of Least Squares Estimators of Cointegrating Vectors", *Econometrica*, 55, 1035-56.

Stock, J.H. (1991a), "Confidence Intervals for the Largest Autoregressive root in U.S. Macroeconomic Time Series", *Journal of Monetary Economics*, 28, 435-459.

Stock, J.H. (1991b), "Bayesian Approaches to the 'Unit Root' Problem: A Comment", *Journal of Applied Econometrics*, 6, 403-411.

Stock, J.H. (1992), "Deciding between I(0) and I(1)", NBER Technical Working Paper number 121.

Stock, J.H. (1994) "Unit Roots, Structural Breaks and Trends", forthcoming in Engle and McFadden eds. *Handbook of Econometrics*, Volume IV.

Stock, J.H. and M.W.Watson, (1993), "A Simple Estimator for Cointegrating Relationships", *Econometrica*, 61, 783-820.

Uhlig, H. (1992) "What Macroeconomists Should Know About Unit Roots as well: The Bayesian Perspective", forthcoming, *Econometric Theory*.

White, H. (1984) *Asymptotic Theory for Econometricians*. Orlando. Fla: Academic Press.

Chapter 2: Efficient Tests For a Unit Root when the Initial Observation is Drawn From Its Unconditional Distribution

I. Introduction

The asymptotically efficient test for a unit root depends crucially on the assumption regarding the information that can be derived from the starting value for the process. In the case where the starting value is assumed to have the same finite expectation under the null and the alternative, as in equation (1) in Chapter 1 (known as the conditional case), was derived in Elliott, Rothenberg and Stock (1992). In footnote 7 of their paper they mention that their statistic need not be the most efficient test when the initial observation is drawn from its unconditional distribution under the alternative hypothesis. Indeed, in their Monte Carlo evidence, the power of their efficient statistic for this case in finite samples declines dramatically. This point is also shown with Monte Carlo analysis in Pantula, Gonzalez-Farias and Fuller (1992), who show that a number of other statistics have similar power in the unconditional case when there are no nuisance parameters.

The asymptotically most efficient test for a unit root in the unconditional case against some alternative can be derived along the lines followed in Elliott, Rothenberg and Stock (1992). The results of this paper show that in what is usually termed the unconditional case, then the asymptotically most efficient test for a unit root does not depend on nuisance parameters and provides an apparently useful test. In this case, the unconditional power envelope can be derived and used to examine which statistics are asymptotically efficient, and the extent to which other statistics are lacking in this property. The different asymptotic tests derived here suggest alternative tests for inversion for confidence intervals as suggested in Chapter 1.

In this chapter we derive the unconditional power envelope in the first order autoregressive model with normal errors and derive an asymptotically efficient family of tests. We compare this family with other tests in both the unconditional and conditional cases and use this family to suggest two new tests for the null of a unit root. We examine the finite sample power of these tests both when the initial observation is drawn from its unconditional distribution under the alternative, and the usual case where this observation has finite variance. The statistics are compared with others available in the literature for these cases. We also examine the use of these statistics in the extended unconditional case, with the result that the best test depends on which case is believed to be true and the criteria of optimality used.

Section 2 discusses assumptions on the initial condition. Section 3 derives the power envelope for the test of a unit root against a set of stationary alternatives when the initial observation is drawn from the usual unconditional case. The fourth section derives the most efficient test in the simple case, which will be most efficient amongst the class of tests invariant to the trend parameters (constant and linear cases) under this assumption. In the fifth section, power envelopes for various cases are examined using Monte Carlo analysis. Monte Carlo evidence as to the properties of these and other statistics are also presented here. Section six concludes. This chapter only considers the use of y_{1t} from equation (1) in chapter 1 for estimation and inference over α. For the remainder of this chapter, the 1 subscript from y_{1t} will be dropped etc, so $y_t = y_{1t}$ here.

II. The Initial Condition

The time series y_t has the representation

$$y_t = d_t + u_t, \quad t = 1,...,T. \tag{1}$$

where

$$u_t = \alpha u_{t-1} + v_t, \quad t = 2,...,T. \tag{2}$$

the d_t are deterministic (trend) components, $v_t = \epsilon_t$ is a Gaussian process and d_t is known[1].

Under these assumptions and further that $E[u_1]^2$ is finite, Elliott, Rothenberg and Stock (1992) show that no uniformly most powerful test against the relevant alternatives exists and derive the asymptotic power function for the most powerful test against a sequence of stationary alternatives. This maps out a power envelope, being the maximum asymptotic power that classical tests for the null hypothesis of a unit root can achieve. Further, they show that if the trend component is slowly varying, this power envelope is attainable even in the case of the trends being unknown. In the case where this slowly varying condition does not hold, they derive the maximum power envelope for the class of tests invariant to the estimation of this trend.

This assumption on the initial condition was referred to in Elliott, Rothenberg and Stock

[1] The last two restrictions are relaxed for the tests developed later, the assumption of d_t known is relaxed for the derivation of invariant power envelopes.

(1992) and generally in the literature as the 'conditional' case. This terminology is applicable as the power envelope derived under the assumptions of that paper is identical to the power envelope derived from the likelihood conditional on the initial observation (as is the usual use of the term conditional in econometrics) when deterministic terms are known.

It is well known, however, that whilst asymptotic theory used to derive the above results (and thus the limiting representations) does not provide a good approximation when y_0 (or its detrended equivalent) diverges from zero for usually encountered sample sizes [Evans and Savin (1981,1984)]. This has led Perron (1991) and Nabeya and Sorenson (1992) to consider an alternative derivation of the limiting distribution of unit root tests, through using continuous record asymptotics, for which they derive limiting representations of unit root tests which depend on a nuisance parameter summarizing this deviation of the initial value from zero[2]. The addition of the extra nuisance parameter presents the difficulty that this parameter must be estimable for empirical application.

One could alternatively examine where this first shock potentially derives from. One possibility is that this initial value, u_0, is actually the result of some unobserved process that precedes time=0, with the process identical to that which comes after time=0. In this case, if u_t has the same process as in equation (2) above, then

$$u_0 = \sum_{i=0}^{k-1} \alpha^i v_{-i} + \alpha^k u_{-k}$$

[2] Perron (1991) derives the result for iid innovations and no deterministic terms, Nabeya and Sorenson extend the results for the addition of a constant and time trend in the specification of the deterministic terms.

where k is the number of periods back that the process continues. Assume that the variance of v_t is σ^2 and that v_t is serially uncorrelated, and that u_{-k} is fixed. Under the null of $\alpha=1$, the variance of u_0 is equal to $k\sigma^2$, and in the limit as k approaches infinity the variance of u_0 itself converges on infinity. Under the alternative, then u_0 has a variance equal to $\sigma^2 \Sigma\alpha^{2i}$, where the sum is from 1 to k-1. Let this sum be written s(k). In the limit, as $k\rightarrow\infty$, then $\sigma^2 s(k)$ converges to $\sigma^2/(1-\alpha^2)$ for α fixed and less than one. The last limiting result here is known in this literature as the unconditional case (e.g. Dickey, Gonzales-Farias and Pantula (1992)). This will be referred to as the unconditional case in this paper.

Compared to the conditional case, it can be seen that here a stronger assumption is being made on the initial condition. It would thus be expected that by making this assumption, that power of the unit root tests would be improved when this assumption is correct. This is indeed the case when there are no deterministic terms to be estimated. For α close to 1, then under the null (in the limit) the initial value comes from an infinitely wide distribution, and under the alternative it comes from a very wide distribution. Thus, under the null one could choose some very large number, for which the probability that an observed data point is closer to zero is itself close to zero, and so one would conclude that the data is stationary. This is made precise later. This is not the case when deterministic terms are included, as under the null the first observation detrended by d_t will be zero.

As will be seen, different assumptions on the initial observation results in different estimators for the deterministic terms, and thus the limiting representations of the detrended data, and the asymptotic power envelopes for optimal tests for a unit root calculated will differ. This will be true whether or not the problem is treated explicitly, as in this paper, or as in the usual case not examined at all. The advantages of this approach over the continuous record approach are twofold; first, no additional nuisance parameters are introduced, and second; the unconditional case has some intuitive justification.

III. The Asymptotic Power Envelope in the Unconditional Case

This section derives the asymptotically efficient unit root test in the case where the initial condition under the alternative hypothesis is drawn from its unconditional distribution. That is, that when α is not equal to one, then

$$\begin{aligned} u_0 &\sim N(0,\sigma^2 s(k)) \\ \epsilon_t &\sim N(0,\sigma^2), \quad t=2,\ldots,T. \end{aligned} \tag{4}$$

The approach taken here will then be to compute optimal Neyman Pearson tests when under the null hypothesis, the initial observation is drawn from a N(0,kσ^2) distribution, and under the alternative hypothesis the initial observation is drawn from its unconditional distribution as in equation (4).

Consider first the Neyman Pearson test in the case where the deterministic coefficients are known, so that the model depends only on α. The log of the likelihood function under the alternative hypothesis is given by

$$\begin{aligned} \mathcal{L}_A = A &- \frac{T}{2}\ln(\sigma^2) - \frac{1}{2}\ln(s(k)) \\ &- \frac{1}{2s(k)\sigma^2}(y_1^d)^2 - \frac{1}{2\sigma^2}\sum_{t=2}^{T}[(1-\alpha L)y_t^d]^2 \end{aligned} \tag{5}$$

whilst under the null hypothesis, the log of the likelihood is

$$\mathcal{L}_0 = A - \frac{T}{2}\ln(\sigma^2) - \frac{1}{2}\ln k - \frac{1}{2k\sigma^2}(y_1^d)^2 - \frac{1}{2\sigma^2}\sum_{t=2}^{T}(\Delta y_t^d)^2 \qquad (6)$$

In both cases A is a constant.

Solving for the Neyman Pearson most powerful test against some alternative, $\bar{\alpha}<1$, the likelihood ratio test when σ is known is proportional to

$$T(M_T - 1) = \frac{T(\bar{\sigma}^2 - \hat{\sigma}^2)}{\sigma^2} \qquad (7)$$

where

$$T\bar{\sigma}^2 = \frac{(y_1^{+d})^2}{s(k)} + \sum_{t=2}^{T}[(1-\bar{\alpha}L)y_t^{+d}]^2$$
$$T\hat{\sigma}^2 = \frac{(y_1^{+d})^2}{k} + \sum_{t=2}^{T}[\Delta y_t^{+d}]^2 \qquad (8)$$

where these are the variances under the alternative and null hypotheses respectively and the +d superscript denotes that the data has the known trends removed. This statistic is proportional to the negative of the likelihood ratio test (minus a constant) and rejects the null hypothesis for small values of the statistic.

Under the additional (temporary) assumption that the initial condition has k finite under the null hypothesis, and using the local to unity representations for α, i.e. $\alpha=1+c/T$, as derived in Bobkoski (1983), Cavanagh (1985) and Phillips (1987), and noting that some terms are

$o_p(1)$, this can be rewritten to give critical regions of the form

$$\bar{c}^2\sigma^{-2}T^{-2}\sum_{t=2}^{T}u_{t-1}^2 - \bar{c}\sigma^{-2}T^{-1}u_T^2 - \bar{c}\sigma^{-2}T^{-1}u_1^2 < b(\bar{c}) \tag{9}$$

where $\bar{c}=T(\bar{\alpha}-1)$ and s(k) has been replaced by its limit as $k\rightarrow\infty$ as an approximation.

Notice that here, as in the conditional case, the minimal sufficient statistic has dimension greater than that of the unknown parameters (in this case dimension three for c non zero) and depends on the particular alternative $\bar{c}$ chosen, so no UMP statistic exists, even in large samples. If the initial condition were not set so that k is finite under the null, the Neyman Pearson test would be $Op(\sqrt{T})$. Thus, the statistic $\sqrt{T}(M_T-1)$ would be equal to simply $[(k-s(k))/ks(k)]((y_1^d)^2/T)$. The test would place all weight on the first observation asymptotically.

These do not provide any appropriate tests for use in practice, both because of the assumption on the initial condition and because d_t is assumed known. This suggests obtaining the feasible region, restricting attention to tests invariant to the estimation of the parameters describing the deterministic component d_t. In doing this, a result will be that the dependance on k under the null will vanish.

In general, we consider d_t as a polynomial trend, $d_t = \beta' z_t$, where $z_t = (1,t,\ldots,t^k)$. This includes the two leading cases $d_t = \beta_0$, and $d_t = \beta_0 + \beta_1 t$. The most powerful invariant (MPI) test here rejects the null hypothesis if the sum of squared residuals from the GLS regression of y_t on the deterministics is small compared to the sum of squared residuals from the GLS regression under the null. The statistic in this case is given by equation (6) above, where now the variances under the alternative and the null are given by

$$T\tilde{\sigma}^2 = \sum_{t=1}^{T} \tilde{e}_t^2$$
$$T\hat{\sigma}^2 = \sum_{t=2}^{T} \hat{e}_t^2 \qquad (10)$$

where

$$\tilde{e}_t = \tilde{y}_t - \tilde{\beta}\tilde{z}_t$$
$$\hat{e}_t = \hat{y}_t - \hat{\beta}\hat{z}_t \qquad (11)$$

where $\tilde{\beta},\hat{\beta}$ are the GLS estimates of the trend terms under the alternative and null respectively, $\tilde{z}_t = [(1-\bar{\alpha}^2)^{1/2}z_1,(1-\bar{\alpha}L)z_2,\ldots,(1-\bar{\alpha}L)z_T]$, $\tilde{y}_t = [(1-\bar{\alpha}^2)^{1/2}y_1,(1-\bar{\alpha}L)y_2,\ldots,(1-\bar{\alpha}L)y_T]$, $\hat{z}_t = [z_1,\Delta z_2,\ldots,\Delta z_T]$, and $\hat{y}_t = [y_1, \Delta y_2,\ldots,\Delta y_T]$[3].

There are two differences between this formulation and that of equation (8) other than the estimation of the deterministics. The first is that in the estimation of the variance under the null, the initial value does not appear. This is because of the algebra of least squares, $\hat{e}_1$ will always be identically equal to zero. The second difference is that the coefficient on the first observation under the null is set to its limit as $k\to\infty$.

Theorem 1. Suppose that y_t is generated by (1) and (2), the initial conditions are generated by (4) with s(k) replaced by its limit as $k\to\infty$, and d_t is unknown, then under local to unity asymptotics where $\bar{c}=T(\bar{\alpha}-1)$ and $c=T(\alpha-1)$ then as T approaches infinity the most powerful invariant test of $\alpha=1$ vs $\alpha=\bar{\alpha}$ has local asymptotic power function

---

[3] Under the null, a weight of 1/k rather than 1 could be used, however the MLE are invariant to this scaling.

$$\gamma^d(c,\bar{c}) = Pr[\Phi^d(c,\bar{c}) < b^d(\bar{c})]$$

$$\text{where} \quad \Phi^\mu(c,\bar{c}) = \bar{c}^{\,2}\int_0^1 V_c^{\mu 2} - \bar{c}\,[V_c^\mu(1)^2 + V_c^\mu(0)^2] \tag{12}$$

$$\Phi^\tau(c,\bar{c}) = \bar{c}^{\,2}\int_0^1 V_c^{\tau 2} - \bar{c}\,V_c^\tau(1)^2 + \bar{c}\,V_c^\tau(0)^2 + (\beta_1^* - M_c(1))^2$$

and

$$V_c^\mu = M_c - \beta_0^{\mu *}$$
$$V_c^\tau = M_c - \beta_0^{\tau *} - s\beta_1^*$$

where $M_c(s)$, $\beta^{\mu *}{}_0$, $\beta^{\tau *}{}_0$, and $\beta^*{}_1$ are defined in lemma A.2 of the appendix. The μ and τ superscripts refer to the demeaned and detrended cases respectively. The asymptotic power envelope is given by

$$\Gamma^d(c) = \gamma^d(c,c)$$

These rejection regions are feasible rejection regions as they can be attained without knowledge of ß or σ. This shows that there exists a power envelope which does not depend on nuisance parameters. This power envelope differs from that derived by Elliott, Rothenberg and Stock (1992) for the conditional case, thus it is shown that the assumption made on the information contained in the initial condition affects the maximum power of the test. In particular the result that estimation of a slowly varying trend does not affect asymptotic power does not hold for the unconditional case.

In a more general formulation for the unconditional case, with the data being generated by a process with more substantial dynamics, then the efficient test can still be derived for any particular lag order (i.e. v_t is generated by an AR(p) process) but the limiting distribution depends on the choice of the order of the lag polynomial. As we show in the next section that generally dependant data achieve the simple unconditional envelope asymptotically, this suggests that the results provide a minimum bound on the general power envelope and that potentially additional power is attainable if the order of the lag polynomial generating v_t is known and is greater than 1[4].

IV. Efficient Tests for a Unit Root in the Unconditional Case

With the lack of existence of a UMPI test, no test is best over all possible stationary alternatives. This is equivalent to the problem faced in Elliott, Rothenberg and Stock (1992), and the same implementation of point optimal testing [King (1980,1988)] is suggested. The test will be constructed to be asymptotically equivalent to one of the family of optimal tests suggested above, and hence will asymptotically achieve the power envelope at some point.

This section constructs a family to tests which are asymptotically equivalent to the simple unconditional MPI tests derived above when v_t in equation (2) is a general I(0) process. This results in asymptotically efficient tests for the unit root when there is no such dependance, and potentially very close to asymptotically efficient tests when there is some dependance. We assume that y_t is generated by equation (1) and that v_t satisfies

---

[4] In the conditional case of Elliott, Rothenberg and Stock (1992), the power envelopes were the same for general dynamic terms.

Condition A.

(a) $\hat{\gamma}_v(j) \xrightarrow{p} \gamma_v(j)$ for finite fixed j, where $\gamma_v(j) = Ev_tv_{t-j}$ and $\hat{\gamma}_v(j) = T^{-1}\Sigma^{\infty}_{t=j+1}v_tv_{t-j}$.

(b) The partial sum process $\upsilon_T => \omega W$, where W is a standard Brownian Motion on [0,1], and ω is a finite positive constant and $\omega^2 = \Sigma^{\infty}_{j=-\infty}\gamma_v(j)$.

The tests proposed are

$$
\begin{aligned}
Q_T^{\mu} &= (\sum_{t=1}^{T}\tilde{e}_t^2 - \bar{\alpha}\sum_{t=1}^{T}\hat{e}_t^2)/\hat{\omega}^2 \quad \textit{where} \quad z_t=1 \textit{ for } eq(12) \\
Q_t^{\tau} &= (\sum_{t=1}^{T}\tilde{e}_t^2 - \bar{\alpha}\sum_{t=1}^{T}\hat{e}_t^2)/\hat{\omega}^2 \quad \textit{where} \quad z_t=[1,t] \textit{ for } eq(12)
\end{aligned}
\tag{15}
$$

where the residuals are constructed as in equation (11) and the following text. These tests are different from those presented in the previous section in that the statistic is corrected for the long run variance of v_t which is no longer equal to σ^2.

Theorem 2. The asymptotic representations for these statistics are

$$
\begin{aligned}
Q_T^{\mu} &\Rightarrow \Phi^{\mu}(c,\bar{c}) \\
Q_T^{\tau} &\Rightarrow \Phi^{\tau}(c,\bar{c})
\end{aligned}
\tag{16}
$$

Theorem 3. With the added condition that $\omega \to d$, with $0<d<\infty$, under the alternative that α is fixed (and less than one), then these statistics are consistent for this alternative.

These tests have the same limiting distributions as the MPI tests, and so asymptotically achieve the same power envelope at some point $\bar{c}$. For implementation, the researcher must choose the point at which the power curve for the statistic achieves the power envelope.

This is equivalent to choosing the alternative under which to test the data. Popular choices are to choose $\bar{c}$ such that the test is tangent to the power envelope at powers of 50% or 80% [King (1988)]. As will be seen from the power curves in the next section, low power of tests with respect to the asymptotic power envelope is generally more problematic for alternatives around the 50% mark, so the tests here will be chosen to achieve the power curve at this point.

This involves choosing $\bar{c}=-8.3$ (chosen by linear interpolation from estimating the power envelope) in the demeaned case and $\bar{c}=-13.5$ in the detrended case. The assumption on ω^2 under the fixed alternative dictates what estimators of the spectral density at frequency zero can be used, a set which includes those mentioned above.

In Elliott, Rothenberg and Stock (1992), it was found that a useful statistic was the augmented Dickey Fuller τ test where detrending is under the local alternative rather than by OLS as suggested by Dickey and Fuller (1979). This can be examined here. The procedure would be to detrend the data y_t by estimating β, and constructing $y_t^d = y_t - \beta z_t$, and then using this data to test for a unit root as in the usual Dickey Fuller t test but without including any deterministic terms. In this case, the limit distributions of this new statistic, DF-GLS$_u^d$, are given by

$$DF\text{-}GLS_u^d \Rightarrow \left(\int V_c^d\right)^{-\frac{1}{2}} \left[\frac{1}{2}(V_c^d(1)^2 - V_c^d(0)^2 - 1)\right] + c\left(\int V_c^d\right)^{\frac{1}{2}} \tag{17}$$

This statistic will also be evaluated numerically in the following sections.

V. Asymptotic Power Curves and Small Sample Evidence

A. Asymptotic Power Curves and Envelopes.

This section presents Monte Carlo estimates of the power envelopes for the demeaned and detrended MPI tests. We will compare these to the asymptotic power of the P_T tests proposed in the conditional case in Elliott, Rothenberg and Stock (1992). We will also compare the power envelopes of the two tests in the conditional case. This will enable the comparison of differences in asymptotic power under each of the assumptions. The simulations are for T=500, using 5000 Monte Carlo replications. The data are generated according to equations (1) and (2). In the conditional case, the initial observation is drawn from a N(0,1) distribution. In the unconditional case, when $\alpha=1$ the first observation is drawn from a $N(0,k\sigma^2)$ distribution with $k\sigma^2$ set to one (as results are numerically invariant to this parameter), and with $\alpha<1$ the first observation is drawn from a $N(0,1/(1-\alpha^2))$ distribution .

First, we need to evaluate the statistics under the null to determine the critical values of the statistics. Under the null hypothesis of $\alpha=1$, c=0, thus this entails evaluating the limiting representations of the statistics when c=0. This is most easily done to a reasonable level of accuracy by evaluating the distribution using a Monte Carlo experiment as in the previous paragraph, with 1000 observations to ensure convergence to the asymptotic results and 20000 Monte Carlo replications. These asymptotic critical values are reported in Table 1 for both the demeaned and detrended models, for both the Q_T and DF-GLS^u statistics. In both cases, rejections of the null hypothesis in favor of stationary models occurs when the rejections are lower tail.

Table 1: Asymptotic Critical Values

Statistic	Percentiles 0.0100	0.0250	0.0500	0.1000	0.2000
	Demeaned				
Q_T^{μ}	2.1060	2.6419	3.2417	4.1804	5.7711
DF-GLS$^{\mu}$	-3.2710	-2.9775	-2.7211	-2.4440	-2.1283
	Detrended				
Q_T^{τ}	3.5985	4.2468	4.9712	5.9567	7.6511
DF-GLS$^{\tau}$	-3.7800	-3.4851	-3.2399	-2.9796	-2.6570

Notes: Asymptotic percentiles were constructed by Monte Carlo methods using 20000 replications and using 1000 observations, where the model is generated. according to equations (1) and (2) with no deterministic terms and the initial observation equal to zero. The errors were drawn from a N(0,1) distribution.

Figures 1 and 2 compare the power envelopes when the first observation is drawn from its unconditional distribution and its conditional distribution in the demeaned and detrended cases respectively. In each case, the correct MPI test is employed (i.e. $c=\bar{c}$. Figure 1 shows a large decrease in potential power achievable by tests for a unit root when in fact the first observation is drawn from its simple unconditional distribution. An indication of the size of the loss can be seen by looking at the ratio of local alternatives where each power envelope reaches 50% (Pitman efficiency). Reading from the graph, the Pitman efficiency here is approximately 1.36[5]. This means that if the data were actually generated in the unconditional case rather than the conditional case, then when an intercept is included then an extra 36% more observations are required to achieve the same power as if it were truly the conditional case.

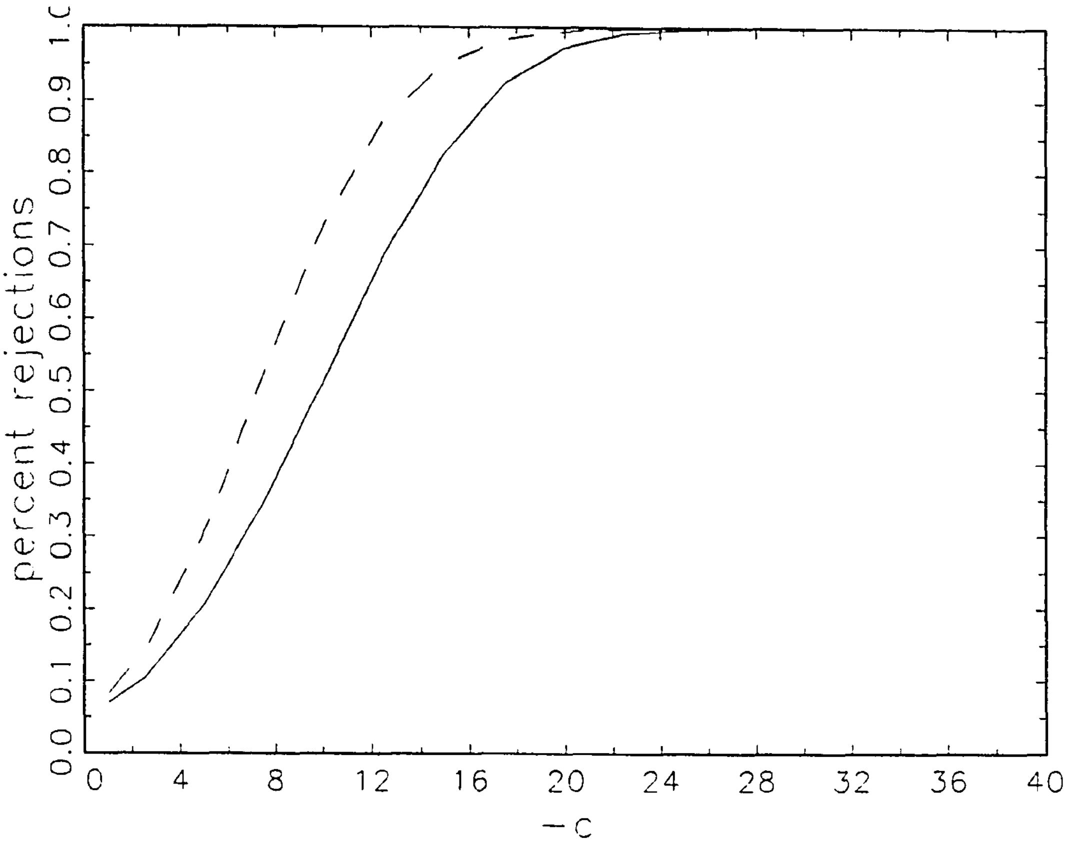
In the conditional case, Elliott, Rothenberg and Stock (1992) show that in the demeaned case, asymptotic power is not decreased when this mean is unknown. This result does not follow in the unconditional case, where lack of knowledge of the constant decreases the power of feasible tests. This appears to be the cause of much of the difference in power between the two cases. This is similar to the well known decrease in power which results from adding an additional deterministic term in such tests as Dickey Fuller tests or other tests where detrending is by methods other than GLS.

When both and intercept and linear trend are estimated, the result is quite different. Whilst the power curve in the unconditional case lies below that for the conditional case in Figure 2, there is no large loss in power such as is seen in Figure 1. The Pitman efficiency loss here is a more modest 1.13%. A possible reason for this feature is from similar reasoning as in the demeaned case. In both the unconditional and conditional cases, estimation of a linear time trend involves a decrease in the feasible rejection region of unit root tests.

---

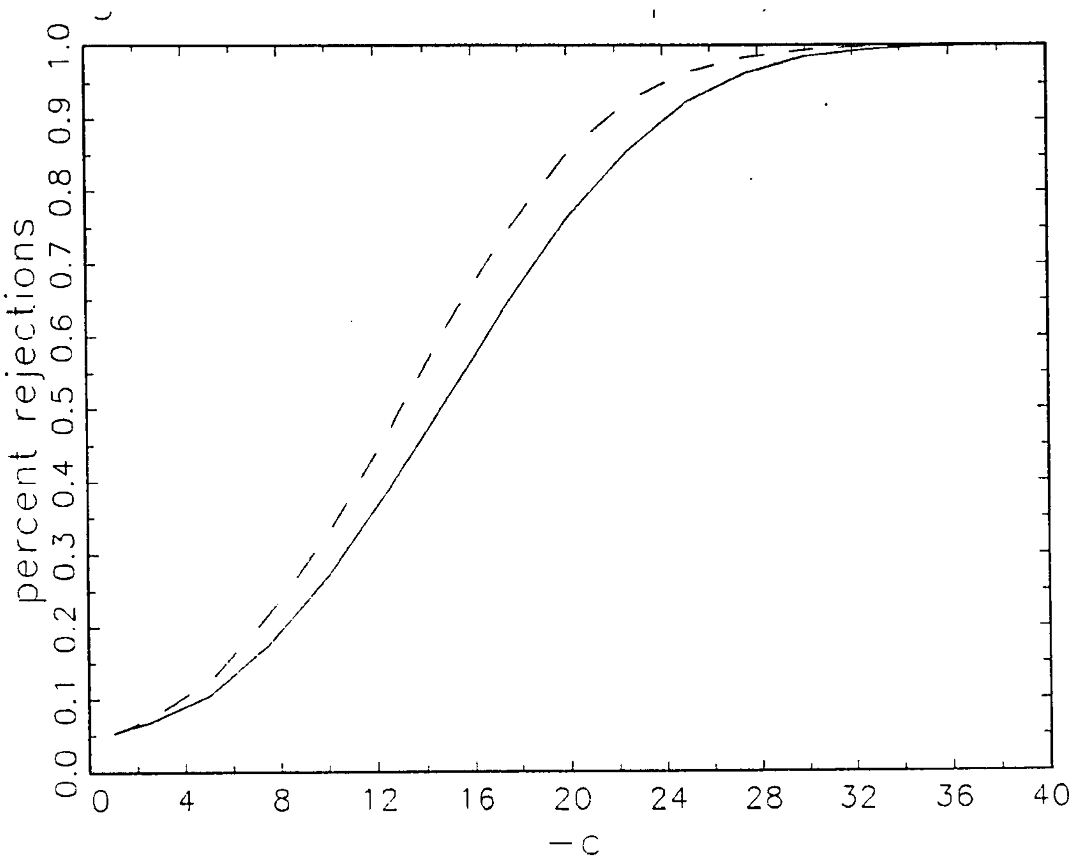
[5] The power curves are estimated for discrete values for $\bar{c}$, the value for $\bar{c}$ at power of 50% is obtained by linear interpolation. The Pitman efficiency is then the ratio of $\bar{c}$'s obtained in this fashion.

Figure 1: Demeaned Power Envelopes



Notes: The figures report asymptotic power functions of the test of the I(1) null when the data is demeaned. The dashed line reports the power envelope from the conditional case, the solid line reports the power envelope from the unconditional case.

Figure 2: Detrended Power Envelopes



Notes: The figures report asymptotic power functions of the test of the I(1) null when the data is detrended with a linear time trend. The dashed line reports the power envelope from the conditional case, the solid line reports the power envelope from the unconditional case.

These results help explain Monte Carlo results in Elliott, Rothenberg and Stock (1992) and other papers. It is found that power of unit root tests decline when the initial observation is drawn from it's unconditional distribution, the loss being more when the data is demeaned than detrended. This is exactly the case for maximal power shown above. The large loss in power in the demeaned case when the assumption is that the first observation is drawn from its unconditional distribution is partly due to the result that in this case, estimation of the mean results in a loss of asymptotic power as the envelope is not similar to the mean known case.

The loss from using the wrong efficient statistic, i.e. the efficient statistic in the conditional case when in reality the data is generated with the first observation from it's unconditional distribution, and vice versa, can be examined asymptotically by generating the asymptotic power curves for these statistics by Monte Carlo experiment. A number of statistics are examined here. The first is the Q_T statistic derived in the previous section. This statistic has power properties almost identical to the unconditional power envelope. The second statistic is the P_T statistic derived by Elliott, Rothenberg and Stock (1992) to be efficient in the conditional case. Also included is the DF-GLSu statistic, which is the equivalent of the augmented Dickey Fuller statistic (Dickey and Fuller (1979)) where instead of detrending by OLS, detrending is done under the local alternative for the same point alternative as the Q_T statistic. The resulting power curves in large samples[6] are plotted in Figures 3 and 4, for the demeaned and detrended cases in the unconditional case. In addition, Pantula,

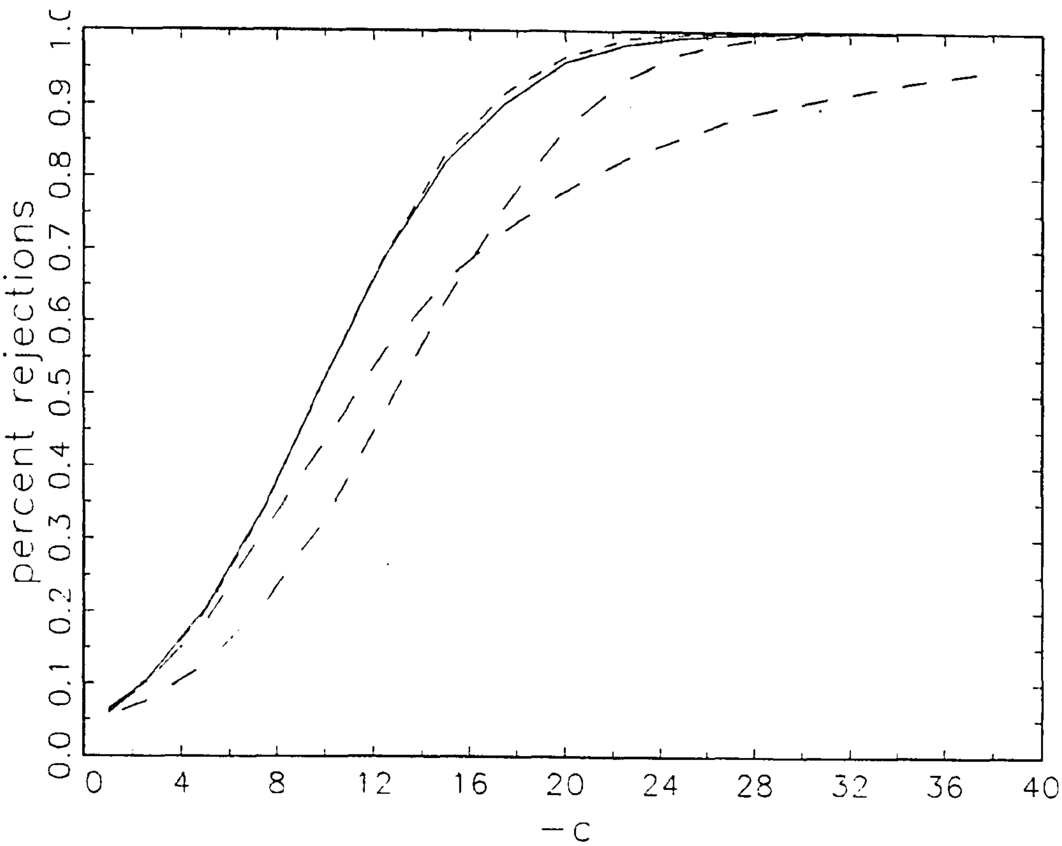
[6] The actual experiment is for T=500, no nuisance parameters are estimated and the number of Monte Carlo replications was 5000.

Gonzales-Farias and Fuller (1992) suggest the use of the symmetric least squares estimator, and a weighted version of this. Each of the two estimators generate a rho-class and tau-class statistic, the former using OLS detrending and the later using GLS detrending under the local alternative. The Sargan-Bhargava statistic (Sargan and Bhargava (1983)), and the Dickey Fuller ρ and τ statistics are also examined.

When the data is demeaned, Figure 3 shows that there can be a reasonably large loss in power from using the 'wrong' conditional MPI statistic (here P_T), but this is mainly for distant alternatives. When the data is detrended, then asymptotically there is very little loss from using the wrong statistic here. This result is shown in Figure 4.

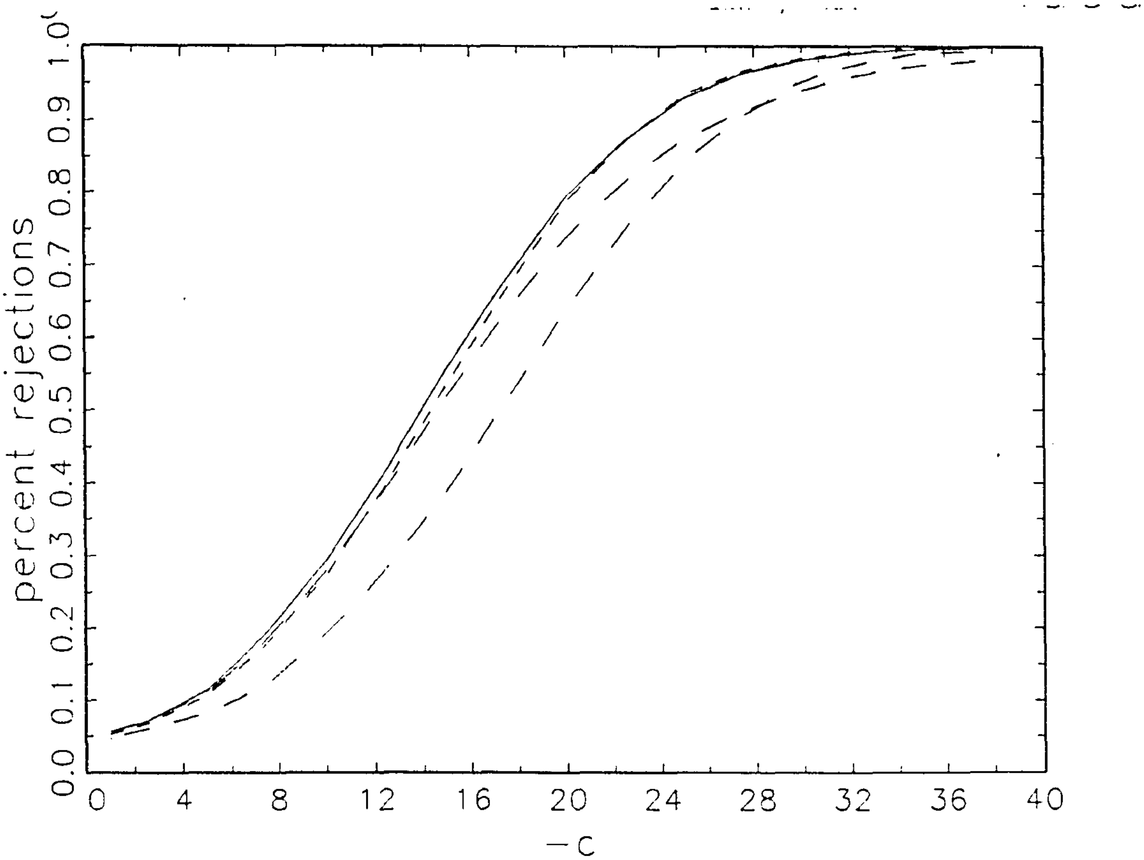
The DF-GLSu statistic tends not to perform as well as the Q_T statistic asymptotically, with a power curve lying below this statistic for most of the relevant range and never rising above the power curve for the Q_T statistic. Even the P_T statistic has better power properties for some alternatives, although for more distant alternatives the DF-GLSu statistic is substantially better. The weighted symmetric least squares τ statistic has power equal, and at times exceeding, the power of the Q_T statistic in the demeaned case but perhaps a little below the Q_T statistic in the detrended case. This statistic is the best performer of the statistics other than Q_T. The DF-GLS statistic has similar large sample power to the P_T statistic in the demeaned case, and a little less power in the detrended case.

Figure 3: Asymptotic Power of Unit Root Tests When Data is Demeaned



Notes: The figures report asymptotic power functions of the test of the I(1) null when the data is demeaned for a variety of test statistics. The upper dashed line reports the power of the WSE test, the upper solid line is for the Q_T statistic, the upper long dashed line is for the P_T and DF-GLS statistics, and the lower long dashed line (lower for small values for c) is the power of the DF τ statistic.

Figure 4: Asymptotic Power of Unit Root Tests When Data is Detrended



Notes: The figures report asymptotic power functions of the test of the I(1) null when the data is detrended for a variety of test statistics. The upper dashed line reports the power of the WSE test, the upper solid line is for the Q_T statistic, the upper long dashed line is for the P_T and DF-GLS statistics, and the lower long dashed line (lower for small values for c) is the power of the DF τ statistic.

The asymptotic results suggest three conclusions. The first is that in the demeaned case, taking a stand as to the distribution of the initial observation will allow the researcher to non trivially increase in the power of the test for a unit root, at the cost of a power loss if the assumption is wrong. This result is similar to one mentioned in Elliott, Rothenberg and Stock (1992), where it was stated that the Gaussian power envelope is the lower bound for power envelopes generated by the class of residuals which satisfy the Lindeburg condition. If one knew the distribution of the residuals satisfies the Lindeburg condition, and that this distribution was non normal, then the researcher could derive statistics which obtain an increase in power over the Gaussian power envelope. A similar result holds here; if one knows that the first observation is generated from it's conditional distribution, then an increase in power over the unconditional power envelope can be achieved. The second conclusion is that this is not as much of a consideration when the data is detrended, use of either statistic works well in large samples.

The third conclusion is that surprisingly, the optimal tests do not have greatly better power than some of the other ad hoc tests, in direct contrast to the conditional case where use of optimal tests results in large gains. This suggests that the method of detrending in this case is not nearly as big an issue here as in the conditional case, although adopting GLS detrending under the local alternative does results in a gain in asymptotic power. This can be seen by noting that the power curve for DF-GLS$^{\mu}$ is always above the power curve for DF-τ. This ranking also holds in the conditional case, so GLS detrending under the local alternative is asymptotically superior to OLS detrending.

B. Small Sample Monte Carlo Evidence.

To examine if the large sample results reported are useful approximations for more relevant sample sizes, we need to examine small samples Monte Carlo results. This is undertaken for a number of different possible models in Tables 2 and 3, for the demeaned and detrended cases respectively. In general, the asymptotic results are borne out by the Monte Carlo evidence, particularly as to the rankings between the statistics. In general, small sample power is less than the asymptotic power.

The second last column of Table 2 gives the small sample results in the unconditional case where the residuals are iid. Against the alternative that α=0.8 (-c=20), the size adjusted power of the Q_T statistic is 0.76, compared to the P_T statistic which has power of 0.73. Other statistics (excepting weighted least squares) have lower power here. It is for these moderate alternative values that the POI tests have the greatest power advantages, as the alternatives are further away the statistics are harder to distinguish. The weighted least squares (WLS) estimator, presented in Pantula et al. (1992) for the no nuisance parameter case, does almost as well as Q_T in the unconditional case when it exploits GLS detrending under the alternative. Appendix 2 of this paper extends the Pantula et al. (1992) WLS estimator (and the SLS estimator [Dickey et al.(1984), Pantula et al (1992)]) to the general I(0) residual case, for use with real data.

TABLE 2: SIZE AND SIZE ADJUSTED POWER : DEMEANED RESULTS

	MA(1), $\theta=$					AR(1), $\phi=$		GARCH MA(1), $\theta=$			Uncond, $\theta=$		
α	-0.8	-0.5	0	0.5	0.8	-0.5	0.5	-0.5	0	0.5	-0.5	0	0.5
Q_T (0.5)													
1.00	0.14	0.11	0.09	0.10	0.43	0.11	0.09	0.12	0.10	0.11	0.11	0.09	0.10
0.95	0.20	0.22	0.22	0.21	0.13	0.21	0.23	0.22	0.22	0.20	0.18	0.18	0.18
0.90	0.40	0.45	0.46	0.43	0.28	0.40	0.46	0.42	0.44	0.42	0.39	0.40	0.40
0.80	0.71	0.78	0.80	0.79	0.65	0.71	0.79	0.75	0.77	0.77	0.73	0.76	0.78
0.70	0.84	0.91	0.92	0.92	0.86	0.85	0.92	0.89	0.90	0.90	0.88	0.91	0.92
DF-GLSu													
1.00	0.08	0.06	0.05	0.07	0.48	0.05	0.05	0.06	0.05	0.08	0.06	0.05	0.07
0.95	0.15	0.15	0.14	0.15	0.14	0.15	0.15	0.15	0.14	0.16	0.14	0.14	0.15
0.90	0.31	0.31	0.32	0.38	0.33	0.28	0.33	0.31	0.32	0.36	0.32	0.32	0.37
0.80	0.64	0.68	0.71	0.82	0.74	0.59	0.75	0.66	0.70	0.79	0.68	0.71	0.82
0.70	0.83	0.88	0.90	0.97	0.91	0.78	0.94	0.86	0.89	0.96	0.87	0.90	0.97
P_T (0.5)													
1.00	0.14	0.11	0.10	0.11	0.42	0.11	0.10	0.13	0.11	0.12	0.11	0.10	0.11
0.95	0.24	0.26	0.26	0.21	0.12	0.26	0.27	0.25	0.24	0.21	0.18	0.18	0.17
0.90	0.46	0.53	0.52	0.41	0.25	0.52	0.56	0.49	0.49	0.40	0.39	0.39	0.38
0.80	0.76	0.83	0.82	0.70	0.56	0.83	0.88	0.80	0.78	0.68	0.67	0.69	0.70
0.70	0.86	0.91	0.90	0.82	0.78	0.93	0.96	0.89	0.88	0.81	0.81	0.82	0.83
DF-GLS													
1.00	0.10	0.08	0.07	0.11	0.45	0.07	0.08	0.09	0.08	0.11	0.08	0.07	0.11
0.95	0.25	0.26	0.26	0.22	0.13	0.26	0.28	0.26	0.25	0.19	0.18	0.17	0.18
0.90	0.52	0.54	0.53	0.43	0.24	0.54	0.62	0.53	0.52	0.39	0.39	0.39	0.39
0.80	0.80	0.85	0.83	0.68	0.37	0.86	0.95	0.83	0.82	0.64	0.67	0.69	0.69
0.70	0.88	0.93	0.90	0.73	0.41	0.95	1.00	0.91	0.89	0.71	0.79	0.80	0.77
DF ρ													
1.00	0.13	0.10	0.08	0.13	0.62	0.09	0.09	0.11	0.10	0.14	0.10	0.08	0.13
0.95	0.17	0.17	0.17	0.17	0.13	0.17	0.17	0.16	0.17	0.17	0.15	0.16	0.15
0.90	0.33	0.35	0.36	0.40	0.31	0.32	0.36	0.34	0.36	0.38	0.35	0.36	0.38
0.80	0.67	0.73	0.76	0.85	0.78	0.64	0.78	0.69	0.74	0.82	0.73	0.76	0.85
0.70	0.85	0.91	0.93	0.98	0.91	0.81	0.95	0.88	0.92	0.97	0.90	0.93	0.98
DF τ													
1.00	0.08	0.06	0.06	0.08	0.46	0.06	0.05	0.07	0.06	0.08	0.06	0.06	0.08
0.95	0.11	0.10	0.10	0.13	0.13	0.10	0.11	0.10	0.10	0.13	0.11	0.10	0.12
0.90	0.23	0.22	0.22	0.31	0.32	0.20	0.25	0.23	0.23	0.29	0.23	0.24	0.32
0.80	0.55	0.56	0.59	0.77	0.79	0.46	0.65	0.55	0.58	0.73	0.57	0.60	0.76
0.70	0.76	0.80	0.84	0.96	0.97	0.67	0.89	0.78	0.82	0.93	0.80	0.83	0.96

TABLE 2: SIZE AND SIZE ADJUSTED POWER : DEMEANED RESULTS (CONT.)

	MA(1), Θ=					AR(1),ϕ=		GARCH MA(1), Θ=			Uncond, Θ=		
α	-0.8	-0.5	0	0.5	0.8	-0.5	0.5	-0.5	0	0.5	-0.5	0	0.5
SB													
1.00	0.15	0.11	0.09	0.09	0.41	0.11	0.08	0.13	0.11	0.10	0.11	0.09	0.09
0.95	0.19	0.19	0.19	0.17	0.13	0.18	0.18	0.18	0.18	0.18	0.17	0.17	0.15
0.90	0.35	0.38	0.39	0.37	0.28	0.34	0.38	0.36	0.38	0.38	0.36	0.37	0.35
0.80	0.66	0.72	0.75	0.75	0.67	0.63	0.71	0.68	0.71	0.74	0.71	0.73	0.74
0.70	0.81	0.88	0.90	0.91	0.88	0.78	0.86	0.85	0.87	0.89	0.87	0.89	0.90
Symmetric LS													
1.00	0.15	0.12	0.09	0.09	0.41	0.11	0.08	0.12	0.11	0.10	0.12	0.09	0.09
0.95	0.18	0.19	0.19	0.17	0.13	0.18	0.18	0.17	0.18	0.18	0.17	0.16	0.15
0.90	0.34	0.37	0.38	0.36	0.28	0.34	0.37	0.35	0.38	0.37	0.36	0.36	0.34
0.80	0.65	0.71	0.74	0.75	0.67	0.62	0.70	0.66	0.71	0.73	0.71	0.73	0.73
0.70	0.81	0.88	0.89	0.91	0.88	0.78	0.86	0.84	0.87	0.89	0.87	0.89	0.90
Weighted Symmetric LS													
1.00	0.15	0.11	0.09	0.10	0.42	0.11	0.08	0.13	0.11	0.11	0.11	0.09	0.10
0.95	0.19	0.20	0.20	0.19	0.13	0.18	0.20	0.19	0.19	0.18	0.17	0.18	0.16
0.90	0.36	0.40	0.42	0.39	0.28	0.35	0.40	0.37	0.39	0.39	0.37	0.39	0.37
0.80	0.67	0.74	0.77	0.77	0.67	0.64	0.73	0.69	0.73	0.74	0.72	0.75	0.76
0.70	0.82	0.89	0.92	0.92	0.88	0.80	0.88	0.86	0.88	0.90	0.88	0.90	0.92

Notes: Each panel gives size and size adjusted power for each statistic over a range of models for v_t from Monte Carlo simulation with one hundred observations and 5000 replications. In each panel, the first row ($\alpha=1$) is size of the statistic. All other rows are size adjusted power, i.e. power adjusted for the possible size distortion for the model. Excepting the GARCH model, the process for v_t can be written $(1-\phi L)v_t=(1-\Theta L)\epsilon_t$. In the unconditional case the same model applies, except that the initial condition u_0 is drawn from its unconditional distribution. In the GARCH model, $v_t = \zeta_t - \Theta\zeta_{t-1}$, $\zeta_t = h_t^{1/2}$, $u_0=0$, and $h_t = 0.65h_{t-1} + 0.25\zeta_{t-1}^2$. The values for the parameters are varied as per the column headings.

As the GLS detrending under the local alternative can be applied to any unit root statistic, we report the DF-GLSu statistic, which is the usual DF-τ statistic with GLS detrending under the (unconditional) alternative presented in the previous section. We also report results for the DF-GLS is the same statistic with GLS detrending under the conditional alternative [from Elliott, Rothenberg and Stock (1992)]. The DF-GLS statistic, as expected from the asymptotic results, does well for closer alternatives but not for more distant alternatives. The DF-GLSu statistic corrects for the problems at further alternatives, but at a cost of lower power for closer alternatives.

Of the other statistics, the symmetric least squares statistic (SLS), the Dickey Fuller ρ statistic and the modified Sargan Bhargava statistic all perform similarly, with power well below the POI tests for moderate alternatives and power similar to Q_T for distant alternatives. The similarity of the SLS and Sargan Bhargava statistics are not surprising, given that their asymptotic distributions are just functions of each other. The Dickey Fuller t statistic has the lowest power against moderate alternatives.

Another asymptotic result which is apparently also true in smaller samples is the tailing off of the power of the P_T statistic as the alternative is far away (as seen for the DF-GLS statistic). In the case of the true $\alpha = 0.7$, Q_T has power 0.91 compared to P_T's power of 0.82. Many other statistics match Q_T for this alternative.

Column 4 of Table 2 presents the same case when the data is generated with the initial condition coming from its conditional distribution. As was the result in Elliott, Rothenberg and Stock (1992), the asymptotically efficient P_T statistic defines the general upper bound for power in this case. For closer alternatives the Q_T statistic has lower power than the optimal P_T statistic and for more distant alternatives actually exceeds this power (although the difference is within Monte Carlo error). Apart from the asymptotically optimal P_T and DF-GLS statistics, Q_T does better than all other statistics (including WLS) in the unconditional case as well for closer alternatives, and there are few differences between statistics at further alternatives. For the case where α=0.90, Q_T has power 0.46 compared to the power of WLS at this same alternative, 0.42.

Table 3 presents results of similar experiments to Table 2 for the detrended case. When the data is generated with as for the unconditional case, almost all statistics have similar power against the alternative that α=0.9 (iid residuals). The exception is the DF τ statistic, which had lower power.

For more distant alternatives, the DF-GLSu statistic and the DF ρ statistics performed the best, with the DF-GLS statistic close behind. This is one of the few cases where the small sample results do not bear out the ordering implied by the asymptotic results. For the case where α=0.7, the DF-GLSu statistic has power of 0.74, the DF-ρ has power 0.73, and the DF-GLS statistic power of 0.71[7].

In the conditional case (iid residuals, column 4), even for moderate alternatives there is a

[7] The standard error for the Monte Carlo estimates is approximately 0.014.

substantial loss in power from using the non asymptotically efficient tests. Against the alternative of $\alpha=0.9$, the DF statistic detrended under the correct conditional alternative has power of 0.23, compared to the power of the DF statistic detrended under the unconditional alternative of 0.21. The differences again, however are modest.

Note that in all cases, the nominal asymptotic and small sample sizes for the tests coincide for the DF-GLS and DF-GLSu statistics, so that large sample critical values can be applied without distortion. These are the only statistics of those with reasonable power that have this property (the DF τ statistic also has very good small sample size properties, but is generally dominated by other statistics on power considerations).

In general, there are two conclusions that can be drawn. Firstly, the asymptotic optimality of the POI statistics appears to carry over to finite samples as well. Secondly, there is a loss from not knowing the true generating process in a given situation, when it is not known if the conditional case or the unconditional case is correct, which of the optimal tests to use. In the demeaned case, the asymptotically optimal statistics of Elliott, Rothenberg and Stock (1992) do far better than other statistics in the conditional case, and do not lose much (except at distant alternatives) in the unconditional case. In the detrended case, there are costs to any strategy. For closer alternatives (usually the ones of interest), the DF-GLS statistic does well in the conditional case and the DF-GLSu statistic does well in the unconditional case.

VI. Conclusion

Potential power that can be achieved by a classical test for a unit root depends on the distribution of the initial observation. Generally, it is assumed that the first observation is from its conditional distribution, a case examined extensively in Elliott, Rothenberg and Stock (1992). This paper extends the method and results of that paper to the case where the initial observation is drawn from it's unconditional distribution. It is well known that a non zero initial condition decreases power of the tests (Evans and Savin (1981,1984), DeJong, Whiteman, Nankervis and Savin (1992)). This is made formal in this paper; the limiting distribution of the Gaussian power envelope is derived for tests when the data is demeaned or detrended and asymptotically efficient tests for a unit root when the data is generated by general heterogenous processes is derived. That these power envelopes and efficient tests differ from those derived in the conditional case means that the researcher must confront this issue in choosing the relevant statistic to employ in practice.

We show that the feasible rejection regions are smaller when the first observation is drawn from it's unconditional distribution than the conditional case. We compare the maximum power obtainable in the unconditional case with that in the conditional case and show that when the data is demeaned, this loss of power is non trivial, although in the detrended case it is less important. As was the case in Elliott, Rothenberg and Stock (1992), the important factor for determining asymptotic power is the method of detrending, which can be applied to any of the statistics in use. The weighted symmetric least squares estimator uses this method of detrending under the null, and has similar power to the Q_T statistic (although not quite as good in the detrended case). The small sample results tend to bear out the asymptotic results.

In terms of which statistic is best suited to testing the null of a unit root in practice, the decision depends on the criterion used and the researchers belief as to the correct assumption on the initial condition. If the criteria is primarily concerned with controlling size in small samples, on must choose between the DF family of tests, choosing the method of detrending. If asymptotic power were the criterion (which could be the case as the small sample results are dependent on the specification of the Monte Carlo experiment), then if the conditional case is believed to be most likely, DF-GLS would be recommended. In the unconditional case, either the Q_T statistic or WLS would be recommended. If one is unsure as to the assumption, the loss from using the DF-GLS statistic in the unconditional case is much smaller than the loss from using the other statistics in the conditional case. If small sample power were the criterion, then the asymptotic results above hold except that the Q_T statistic outperforms the WLS statistic in the conditional case.

Further work could include examining the use of the two sets of asymptotically efficient statistics in tandem. Also, some light on the losses from using either statistic in the 'wrong' situation may be shed from deriving the asymptotic distributions of each statistic in the case that the assumption as to the generation of the initial observation is in error. Other statistics proposed in the literature can be adjusted to use GLS detrending under the local alternative, the small sample properties of these statistics could also be examined in a future work.

In terms of this thesis, the results shed light on the question as to how much information can we extract from the data about the value of α. The choice of optimal tests to invert for a confidence interval follows the arguments over the choice of test for testing the null hypothesis discussed above, as suggested in chapter 1. Clearly, under a range of assumptions we cannot find the value for α from the data, only a range. This chapter shows an alternate test which works well under certain situations and can be utilized in the first stage according to the discussion of the previous chapter.

Appendix 1: Proofs

Two Lemmas are derived before proceeding with the theorems.

Lemma A.1.

The result

$$T(M_T - \bar{\alpha}) = (B_{1T} + \bar{\alpha} B_{2T} + B_{3T})/\hat{\sigma}^2 \tag{A.1}$$

holds exactly where

$$\begin{aligned} B_{1T} &= (1-\bar{\alpha}^2)(y_1^d)^2 \\ B_{2T} &= \sum_{t=2}^{T} (\tilde{\beta} - \hat{\beta})' \Delta z_t \Delta z_t' (\tilde{\beta} - \hat{\beta}) \\ B_{3T} &= (1-\bar{\alpha})^2 \sum_{t=2}^{T} (y_{t-1}^d)^2 + (1-\bar{\alpha})[(y_T^d)^2 - (y_1^d)^2] \end{aligned} \tag{A.2}$$

and $\tilde{\beta}$ and $\hat{\beta}$ are the GLS estimates of β under the alternative and the null respectively, and y_t^d is y_t detrended by GLS under the alternative, i.e. $y_t^d = y_t - \beta z_t$ where $d_t = \beta z_t$ in equation (1).

Proof of Lemma A.1

This result follows from the definitions of $\tilde{\sigma}^2$ and $\hat{\sigma}^2$ in equation (10) and following text. This calculation and result is identical to Lemma A.1 of Elliott, Rothenberg and Stock (1992) except for B_{1T}, which is amended by noting that by the algebra of least squares that $\hat{y}_1^d = 0$. The different coefficient derives from the differing assumption on the initial condition and allowing s(k) to be replaced by its limit as $k \to \infty$ □

Limiting distributions for the deterministic terms under the local alternative are derived in Lemma A.2.

Lemma A.2.

1a) If u_t is given by equation (1), and the distribution of v_t by equation (4) where the limit as $k \to \infty$, then

$$\frac{1}{T^{1/2}}(u_{[Ts]} - u_1) \to \omega W_c(s) + (e^{cs}-1)\xi$$

where $\xi \sim N(0,-\omega^2/2c)$. This defines $\omega\ M_c(s)$. If v_t is iid then the long run standard deviation ω is replaced by σ.

2a) In the demeaned case, for a fixed alternative $\bar{\alpha}$, then $T^{-1/2}(\hat{\beta}_0-\beta_0-u_1) \Rightarrow [-\bar{c}(2-\bar{c})]^{-1}(\bar{c}^2\omega \int M_c - \bar{c}\omega(M_c(1)) = \omega\beta_0^{\mu *}$, where $\omega^2=\gamma_v(0)$. This gives the result that $(y_{[Ts]}^{\mu})/\sqrt{T} \Rightarrow \omega V_c^{\mu}$, where $V_c^{\mu}(s) = M_c(s)-\beta_0^{\mu *}$.

2b) In the detrended case, for a fixed alternative $\bar{\alpha}$, $\Upsilon^{-1}(\hat{\beta}-\beta-Lu_1) \Rightarrow \omega D^{-1}R$, where

$$\Upsilon = \begin{bmatrix} \sqrt{T} & 0 \\ 0 & \sqrt{\frac{1}{T}} \end{bmatrix}$$

$$R = \begin{bmatrix} \bar{c}^2 \int M_c(s)ds - \bar{c}(M_c(1)) \\ (1-\bar{c})M_c(1) + \bar{c}^2 \int sM_c(s)ds \end{bmatrix}$$

and

$$D = \begin{bmatrix} \bar{c}^2 - 2\bar{c} & \bar{c}^2/2 - \bar{c} \\ \bar{c}^2/2 - \bar{c} & 1+\bar{c}^3/2 - \bar{c} \end{bmatrix}$$

where $L_1 = [1\ 0]'$. The first element of $D^{-1}R$ is defined as $\beta_0^{\tau*}$, the second is β_1^*. This gives the result that $(y_{[Ts]}^{\tau})/\sqrt{T} \Rightarrow \omega V_c^{\tau}$, where $V_c^{\tau}(s) = M_c(s)-\beta_0^{\tau*}-s\beta_1^*$.

Proof of Lemma A.2

1a). This result follows from writing u_t as

$$(u_t - u_1) = \sum_{s=1}^{t} \alpha^{t-s} v_s + (\alpha^t - \alpha) u_0 + v_1 \quad \text{(A.4)}$$

The first of these terms is the familiar local to unity characterization in the conditional case, where $\alpha=1+c/T$, and its limit after scaling by $1/\sqrt{T}$ is $\omega W_c(s)$ from the assumption in Condition A(b), the continuous mapping theorem and the functional central limit theorem [Bobkoski (1983), Cavanagh (1985), Phillips (1987), Chan and Wei (1987)].

The second piece comes from the differing assumption on the initial condition. Firstly, $\alpha^{[Ts]}-\alpha \rightarrow e^{cs}-1$. Second, $u_0/\sqrt{T}$ is distributed $N(0,\omega^2/T(1-\alpha^2))$ by the assumption on the initial condition (the variance is the long run variance of v_t for the general MA process). Using $\alpha=1+c/T$, the variance term has denominator $T(1-\alpha^2) = -2c + op(1) \rightarrow -2c$. Finally, v_1 has finite variance so $v_1/\sqrt{T}$ converges in probability to zero by Chebyshev.

2a) Taking the first derivative of the likelihood function equation (4) with respect to β_0 and setting to zero yields the expression

$$\tilde{\beta}_0 = \frac{(1-\bar{\alpha}^2)y_1 + (1-\bar{\alpha})\sum_{t=2}^{T}(1-\bar{\alpha}L)y_t}{(1-\bar{\alpha}^2) + (T-1)(1-\bar{\alpha})^2} \tag{A.4}$$

Multiplying the numerator by the square root of T and the denominator by T, then

$$T^{-\frac{1}{2}}(\tilde{\beta}_0 - \beta_0 - u_1) = \frac{-\bar{c}T^{-\frac{1}{2}}\sum_{t=2}^{T}\Delta u_t + \bar{c}^2 T^{-\frac{3}{2}}\sum_{t=2}^{T}(u_{t-1}-u_1)^2}{-2\bar{c} - \frac{\bar{c}^2}{T} + \frac{T-1}{T}\bar{c}^2} \tag{A.5}$$

This converges to the limit stated noting that $\Sigma_2^T \Delta u_t = u_T - u_1$, using the results of 1a) above, and applying the continuous mapping theorem and functional central limit theorem.

2b) The GLS normal equations can be written

$$[\Upsilon\sum_{t=1}^{T}\tilde{z}_t\tilde{z}_t'\Upsilon]\Upsilon^{-1}(\tilde{\beta} - \beta - Lu_1) = \Upsilon[\sum_{t=1}^{T}\tilde{z}_t\tilde{u}_t' - \sum_{t=1}^{T}\tilde{z}_t\tilde{z}_t'Lu_1] \tag{A.6}$$

Let the right hand side of this equation be $R_T = [R_{1T}\ R_{2T}]'$. After algebraic rearrangement, the first of these terms is given by

$$\sqrt{T}\,R_{1T} = \bar{c}^2 T^{\frac{-3}{2}}\sum_{t=2}^{T}(u_{t-1}-u_1) - \bar{c}T^{\frac{-1}{2}}(u_t-u_1) + op(1)$$

and the second by

$$\frac{1}{\sqrt{T}} R_2 = (1-\bar{c})T^{-\frac{1}{2}}(u_T - u_1) + \bar{c}^2 T^{-\frac{5}{2}} \sum_{t=2}^{T} t(u_{t-1} - u_1) + o_p(1) \qquad \text{(A.8)}$$

Use of the results of the first part of this lemma, the continuous mapping theorem and the functional central limit theorem give the limit results for R_T stated above. The denominator term $\Upsilon \Sigma z z' \Upsilon$ does not involve any stochastic variables. It is straightforward that in the limit this is equal to D.

□

Proof of Theorem 1.

Lemma A.1 shows that $T(M_T - \bar{\alpha})$ can be written as the sum of three pieces. This differs from $T(M_T-1)$ in this problem by the amount $T(1-\bar{\alpha})$, but this extra piece is a constant so can be subsumed into the critical value b(c). Using lemma A.2, the continuous mapping theorem and the FCLT the limiting distributions can be obtained.

In the demeaned case, $\Delta z_t = 0$, so $B_{2T}=0$. $B_{1T} = (-2\bar{c})T^{-1}(y_1 + (\beta_0 - \hat{\beta}_0))^2$, which converges to $(-2\bar{c})V_c^{\mu}(0)^2$ by results of Lemma A.2. Similarly, by results of Lemma A.2 and using the continuous mapping theorem, then

$$B_{3T} \rightarrow \bar{c}^2 \int V_c^{\mu 2} - \bar{c}(V_c^{\mu}(1)^2 - V_c^{\mu}(0)^2)$$

Combining terms gives the result desired scaled by $\omega^2/\hat{\sigma}^2$. Following equation (10), the denominator can be written

$$\hat{\sigma}^2 = \frac{1}{T}\sum_{2}^{T}(\Delta y_t - \hat{\beta}'\Delta z_t)^2 = \frac{1}{T}\sum_{2}^{T}(\Delta u_t)^2$$

where the last equality follows as $\Delta z_t = 0$. This can be further rewritten to obtain

$$\hat{\sigma}^2 = \frac{1}{T}\sum_{2}^{T}(v_t + \frac{c}{T}u_{t-1})^2$$
$$= \frac{1}{T}\sum_{2}^{T}v_t^2 + \frac{2c}{T^2}\sum_{2}^{T}v_t u_{t-1} + \frac{c^2}{T^3}\sum_{2}^{T}u_{t-1}^2$$

The last two of these terms converge to zero. The remaining term converges to $\gamma_v(0)$ by Condition A a). This is equal to ω^2 for v_t iid as in equation (4). Combining results yields $\Phi^{\mu}(c,\bar{c})$.

When a time trend is included, $B_{1T} = (-2\bar{c})T^{-1}(y_1 + (\beta_0 - \beta_0) + (\beta_1 - \beta_1))^2$, which converges to $(-2\bar{c})\omega^2 V_c^{\tau}(0)^2$ by results of Lemma A.2. By results in Lemma A.2 and using the continuous mapping theorem, then

$$B_{3T} \rightarrow \omega^2[\,\bar{c}^2\int V_c^{\tau 2} - \bar{c}(V_c^{\tau}(1)^2 - V_c^{\tau}(0)^2)]$$

For B_{2T}, $\Sigma \Delta z_t \Delta z_t'$ is zero for all but the [2,2] element, which is (T-1). This term can then be written

$$B_{2T} = (T-1)(\tilde{\beta}_1 - \hat{\beta}_1)^2$$
$$= \frac{T-1}{T}[T^{\frac{1}{2}}(\tilde{\beta}_1 - \beta_1) - T^{\frac{1}{2}}(\hat{\beta}_1 - \beta_1)]^2$$

The OLS estimate $\hat{\beta}=(y_T-y_1)/(T-1)$, which after centering and scaling by $\sqrt{T}$ we obtain $(T/(T-1))T^{-1/2}(u_T-u_1)$. This has limit distribution $\omega M_c(1)$, so $B_{2T} \Rightarrow \omega^2[\beta^{\tau*} - M_c(1)]^2$. Combining the terms yields the stated result again multiplied by the ratio $\omega^2/\hat{\sigma}^2$.

Here, again following equation (10),

$$\hat{\sigma}^2 = \frac{1}{T}\sum_2^T (\Delta y_t - \hat{\beta}\Delta z_t)^2$$
$$= \frac{1}{T}\sum_2^T (\Delta u_t)^2 + \frac{2}{T}(u_T-u_1)(\hat{\beta}_1-\beta_1) + (\hat{\beta}_1-\beta_1)^2(T-1)/T$$

The last two of these terms converge in probability to zero. The first of these terms was shown above in the demeaned case to converge to $\gamma_v(0)=\omega^2$ for iid normal residuals yielding the desired result.

The power envelope is obtained by considering the set of tests where $c=\bar{c}$.

□

Proof of Theorem 2.

The proof of this is a straightforward extension of Theorem 1. In Theorem 1 it was shown that $T(M_T-\bar{\alpha})$ converged to $\Phi^d(c,\bar{c})$ multiplied by ω^2/σ^2. Here,

$$Q_T^d = \frac{\gamma_v(0)}{\hat{\omega}^2}\, T(M_T-\alpha)$$

As $T(M_T-\bar{\alpha})$ converges to the limit stated in (16) multiplied by $\omega^2/\gamma_v(0)$, then all we require is that $\hat{\omega}^2$ converges to ω^2 in both the demeaned and detrended cases. This is assured by Condition A b). Estimators of $\hat{\omega}$ include the sums of variances estimators [Newey and West (1989), Andrews (1991)] and the AR estimators [Stock (1988)]. See Stock (1994) for a discussion of the estimators and their small sample properties in the case of unit root testing. For the test of $\alpha=1$, we set $c=0$. □

Proof of Theorem 3.

From Lemma A.1 we see that Q_T^d is the sum of three quadratic terms, divided by a positive constant, and hence is positive. Further, rejections are lower tail. Thus, to show consistency in this case a sufficient condition is that under the fixed alternative, $Q_T^d \xrightarrow{p} 0$ (see Stock (1988)).

In the demeaned case, $(y_{[Ts]}^{\mu})/\sqrt{T} \xrightarrow{p} 0$. This follows from the assumption that under the alternative, u_t have distributions so are bounded; this gives the result $u_t/\sqrt{T} \xrightarrow{p} 0$. This boundedness result can be used to show that $(\hat{\beta}_0-\beta_0-u_1)/\sqrt{T} \xrightarrow{p} 0$. From equation (A.5), this follows if $(u_1/\sqrt{T}, u_T/\sqrt{T}, (\Sigma u_{t-1})/T^{3/2})$ each converge to zero. The first two converge to zero immediately by the boundedness argument. The third result follows similarly, as $(\Sigma u_{t-1})/T$ is bounded. As $(y_{[Ts]}^{\mu})/\sqrt{T}=u_t/\sqrt{T}-(\hat{\beta}_0-\beta_0)/\sqrt{T}$, then $(y_{[Ts]}^{\mu})/\sqrt{T} \xrightarrow{p} 0$ is shown.

From Lemma A.2, $Q_T^{\mu} = (B_{1T}+B_{3T})/\hat{\omega}^2$. $B_{1T} = (-2\bar{c}/T)(y_1^{\mu})^2+op(1)$. This converges to zero by the result in the previous paragraph. The convergence of B_{3T} to 0 also follows directly from the result that $(y_{[Ts]}^{\mu})/\sqrt{T} \xrightarrow{p} 0$. The assumption that $\omega^2 \xrightarrow{p} d$ under the fixed alternative completes the result.

In the detrended case, we can also show that $(y_{[Ts]}^{\tau})/\sqrt{T} \xrightarrow{p} 0$. Consider the equation (A.6) with the same scaling. As before, the denominator term converges to D as it does not involve any stochastic variables. From the results in the demeaned case above, the first row of R, see equation (A.7), converges to zero (it is identical to the numerator of the equation (A.5), which was shown to converge to zero above. Turning to the second row of the right hand side of equation (A.6), given by equation (A.8), the convergence to zero of the first two terms follows from above. To show the result for the last term, note that

$$\frac{1}{T^{5/2}}\sum_{2}^{T} t u_{t-1} = \frac{1}{T^{3/2}}\sum_{2}^{T} \frac{t}{T} u_{t-1} < \frac{1}{T^{3/2}}\sum_{2}^{T} u_{t-1}$$

As shown above, this term converges to zero so $T(\beta-\beta-Lu_1)\xrightarrow{p}0$. This gives the result that $(y_{[Ts]}^{\tau})/\sqrt{T} \xrightarrow{p} 0$. From Lemma A.2, part 2b), we can write $Q_T^{\mu} = (B_{1T}+B_{2T}+B_{3T})/\hat{\omega}^2$, where using the result just shown and analogous arguments as in the demeaned case, both B_{1T} and B_{3T} converge to zero. From equation (A.10), $B_{2T} = (T-1)/T)(\sqrt{T}(\beta_1-\beta_1)-\sqrt{T}(\beta_1-\beta_1))^2$. The convergence of the first of these to zero was shown above. $\sqrt{T}(\beta_1-\beta_1)=(T/(T-1))(u_T/\sqrt{T}-u_1/\sqrt{T})$. This again converges to zero by the boundedness of u_t. Finally, the assumption that $\omega^2 \xrightarrow{p} d$ under the fixed alternative completes the result.

Appendix 2: Generalizing SSE and WSSE

As noted in the text, Pantula, Gonzales-Farias and Fuller (1992) present some estimators which in the case of iid errors appear to have good power properties in the demeaned case. The statistics they present include the symmetric least squares estimator and the weighted symmetric least squares estimator. The forms presented in their paper are not similar under general assumptions on the generating process, so cannot be used in these cases. This appendix derives corrections to these test statistics which enable them to be used under more general assumptions on the innovation process, or more immediately, in the Monte Carlo results presented above.

A. Simple Symmetric Least Squares Estimator.

This estimator, first proposed by Dickey et al. (1984) is given by Pantula et al. (1992) as

$$\alpha_{sym} = \frac{\sum_{t=2}^{T} y_{t-1}^{d} y_t^{d}}{\sum_{t=2}^{T-1} (y_t^d)^2 + \frac{1}{2}(y_1^d)^2 + \frac{1}{2}(y_T^d)^2} \tag{A2.1}$$

This can be rewritten as

$$(\alpha_{sym} - 1) = \frac{-\frac{1}{2}\sum_{t=2}^{T} (\Delta y_t^d)^2}{\sum_{t=2}^{T-1} (y_t^d)^2 + \frac{1}{2}(y_1^d)^2 + \frac{1}{2}(y_T^d)^2} \tag{A2.2}$$

which has the limiting distribution after scaling for the unconditional case of

$$T(\alpha_{sym}-1) \rightarrow \frac{\sigma^2_{\Delta y}}{2\omega^2\int M_c^2} \qquad \text{(A2.3)}$$

Stock (1994) presents the limiting result for this statistic in the conditional case.

As noted by Stock (1994), this statistic has the same limiting distribution (scaled by one half) as the inverse of the Sargan Bhargava statistic [Sargan and Bhargava (1983)][8]. To modify this so that it is invariant, the method employed by Stock (1988) to modify the Sargan Bhargava statistic can be used. This suggests that the modification be to divide by the variance of the change in y_t and multiply by a consistent estimate of the long run variance. This gives the modified statistic

$$MSSE = T(\alpha_{sym}-1)\frac{\hat{\omega}^2}{T^{-1}\sum_{t=2}^{T}\Delta y_t^2} \qquad \text{(A2.4)}$$

which is used in the Monte Carlo simulations.

B. Weighted Symmetric Least Squares Estimators

One of the recommended statistics from Pantula et al is the weighted symmetric least squares estimator which they give as

$$\alpha_{ws} = \frac{\sum_{t=2}^{T} y_{t-1}^d y_t^d}{\sum_{t=2}^{T-1}(y_t^d)^2 + T^{-1}\sum_{t=1}^{T}(y_t^d)^2} \qquad \text{(A2.5)}$$

[8] And thus has equivalent asymptotic power properties to the Sargan Bhargava statistic.

Again rearranging, this results in the test statistic

$$(\alpha_{ws}-1) = \frac{-\sum_{t=2}^{T}(\Delta y_t^d)^2 + \frac{1}{2}[(y_t^d)^2 + (y_1^d)^2] - T^{-1}\sum_{t=1}^{T}(y_t^d)^2}{\sum_{t=2}^{T-1}(y_t^d)^2 + T^{-1}\sum_{t=1}^{T}(y_t^d)^2} \qquad \text{(A2.6)}$$

which has the limiting distribution in the unconditional case

$$T(\alpha_{ws}-1) \rightarrow \frac{\frac{1}{2}(M_c(1)^2 + M_c(0)^2) - \int M_c^2 - \frac{\sigma_{\Delta y}^2}{\omega^2}}{\int M_c^2} \qquad \text{(A2.7)}$$

As this distribution contains the ratio of the variance to the long run variance, it is not invariant when there is general dependance in the errors. Unlike the case above, this is an additive term rather than multiplicative, so the simple correction of above cannot be applied. This is, however, identical to the problem faced by Phillips (1987) with the Dickey Fuller test, so the same solution of adding a variance term and introducing a long run variance term in it's place can be used. This results in the statistic

$$CWSSE = T(\alpha_{ws}-1) + \frac{(T^{-1}\sum_{t=2}^{T}\Delta y_t^2 - \hat{\omega}^2)}{(T^{-2}\sum_{t=2}^{T-1}y_t^2 + T^{-3}\sum_{t=1}^{T}y_t^2)} \qquad \text{(A2.8)}$$

Application of the FCLT and CMT results in the limiting distribution of the statistic of

$$CWSSE \rightarrow \frac{\frac{1}{2}(M_c(1)^2 + M_c(0)^2) - \int M_c^2 - 1}{\int M_c^2} \qquad \text{(A2.9)}$$

Both of the corrected statistics are used in this paper in the Monte Carlos presented in Section 5B. Notice that different methods of detrending other than OLS detrending (used by Pantula et al.) can be used, with limiting distributions adjusted by the correct detrended Brownian Motions. To obtain critical values under various detrending methods, the statistics in equations (A2.4) and (A2.8) were used in Monte Carlos of 10000 replications with $\hat{\omega}$ equal to the theoretical variance of the change in y_t, which was set to one for the experiment.

References

Bobkoski, M.J. (1983), *Hypothesis Testing in Nonstationary Time Series*, unpublished Ph.D thesis, Department of Statistics, University of Wisconsin.

Cavanagh, C. (1985), "Roots local to unity", manuscript, Harvard University.

Chan, N.H. (1988), "On Parameter Inference for Nearly Nonstationary Time Series", *Journal of the American Statistical Association*, 83, 857-62.

Chan, N.H. and C.Z.Wei (1987), "Asymptotic Inference for Nearly Nonstationary AR(1) Processes", *Annals of Statistics*, 15, 1050-63.

DeJong D.N., J.C. Nankervis, N.E.Savin and C.H.Whiteman (1992), "The Power Problems of Unit Root Tests for Time Series with Autoregressive Errors", Forthcoming, *Journal of Econometrics*.

Dickey, D.A., D.P.Hasza, and W.A.Fuller (1984) "Testing for Unit Roots in Seasonal Time Series", *Journal of the American Statistical Association*, 79, 355-67.

Dickey, D.A. and Fuller, W.A. (1979), "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association*, 74, no. 366, 427-431.

Elliott G., T.J.Rothenberg and J.H.Stock (1992), "Efficient Tests for an Autoregressive Unit Root", NBER Technical Working Paper Number 130, December 1992.

Evans, G.B.A and N.E.Savin (1981), "Testing for Unit Roots: 1", *Econometrica*, 49, 753-79.

Evans, G.B.A and N.E.Savin (1984), "Testing for Unit Roots: 2", *Econometrica*, 52, 1241-69.

King, M.L. (1980), "Robust Tests for Spherical Symmetry and their Application to Least Squares Regression", *Annals of Statistics*, 8, 1265-1271.

King, M.L. (1988), "Towards a Theory of Point Optimal Testing", *Econometric Reviews*, 6, 169,218.

Nabeya, S and B.E.Sorenson (1992) "Asymptotic Distribution of the Least Squares Estimators and Test Statistics in the Near Unit Root Model with Non-Zero Initial Value and Local Drift and Trend", manuscript, Brown University.

Pantula, S, G. Gonzales-Farias, and W.A.Fuller (1992), "A Comparison of Unit Root Test Criteria", manuscript, North Carolina State University.

Perron, P. (1991) "A Continuous Time Approximation to the Unstable First Order Autoregressive Process: The Case Without an Intercept", *Econometrica*, 59, 211-36.

Phillips, P.C.B, (1987) "Time Series Regression with a Unit Root", *Econometrica*,

Sargan, J.D. and A. Bhargava (1983), "Testing Residuals from Least Squares Regression for Being Generated by the Gaussian Random Walk," *Econometrica*, 51, 153-174.

Stock, J.H. (1988), "A Class of Tests for Integration", manuscript, Kennedy School of Government, Harvard University.

Stock, J.H. (1994), "Unit Roots and Trend Breaks in Econometrics", forthcoming in Engle,R.F and McFadden eds. The Handbook of Econometrics, vol IV.

Chapter 3: On the Robustness Of Efficient Cointegration Estimators When Regressors Almost Have Unit Roots

I. Introduction

It is often postulated in theoretical economics that there are certain relationships between economic quantities that hold over long periods of time. Engle and Granger (1987), in pathbreaking work, introduced the concept of cointegration and argued that, when individual variables have a unit root[1], cointegration can be an empirically useful method to quantify long run relations. In the following years, a number of techniques emerged which show how to estimate the coefficients of cointegrating vectors efficiently (see Watson (1993) for a thorough overview of these concepts and techniques). The popularly used techniques include the Johansen (1988) maximum likelihood technique, the fully modified procedure of Phillips and Hansen (1989), the dynamic ordinary least squares (DOLS) method of Phillips and Loretan (1991), Saikkonen (1991) and Stock and Watson (1993), and the error correction model (ECM) method of Saikkonen (1992). Each of these methods allow χ^2 inference on the cointegrating parameters, a result enabling the testing of many theories of interest.

This paper examines the robustness of these popular methods for the efficient estimation of cointegrating vectors when there is a stable relationship between the variables but the largest root in the explanatory variables close to but not exactly equal to one, specifically, is local

[1] The inability of tests to reject the presence of a unit root in a number of major macroeconomic time series had been previously shown by Nelson and Plosser (1982).

to unity.

The motivation for examining the robustness of estimators in this direction is that in most empirical applications, the modelling assumption of an exact unit root in the data is usually determined through pre-tests for unit roots in the individual series rather than obtained from the economic theory under investigation, and it is well known that the power of unit root tests can be low for roots close to one (e.g. DeJong and Whiteman (1991), for a survey see Stock (1993)). I show analytically that estimates of the cointegrating vector when there is a large root in the explanatory variable which does not equal one remain consistent but the leading efficient methods of inference on the cointegrating vector are asymptotically distorted and in particular tend to overreject tests of the true null hypothesis. The extent of the overrejection depends on the extent of the departure of the largest root in the regressor from one and on the degree of the simultaneity between the errors of the equation of interest and the explanatory variables. It is shown that for any deviation of the largest root in the explanatory variable from one, then the *size* of the test approaches one as the degree of simultaneity increases. The Monte Carlo evidence shows that this can be important for empirical work. For plausible parameter values, it is shown below that tests with nominal level 5% may actually have asymptotic rejection rates exceeding 50%. It should be emphasized that these distortions remain in arbitrarily large samples and thus are not simply a problem of departures of finite sample distributions from their asymptotic representations.

Each of the available estimates of cointegrating vectors which allow χ^2 inference considered here have this property of overrejecting the true null hypothesis. I examine some frequently used procedures, specifically those of Johansen (1988), the Phillips and Hansen (1989) fully modified procedure, the DOLS procedure (Phillips and Loretan (1991), Saikkonen (1991), Stock and Watson (1993)), and the full system ECM estimator of Saikkonen (1992). The results of this paper will suggest that inference on cointegrating vectors using these methods only be employed when the existence of the unit root in the regressor is derived from the economic theory being tested, or alternatively may be used when the degree of simultaneity between the generating processes of the left and right hand side variables is small.

The next section of the paper discusses the conventional empirical approach to analyzing data which exhibit stochastic trends and the use of the cointegration vector estimators available, motivating the potential for problems to arise. The third section illustrates the results of the paper in a simple bivariate regression where there are no dynamics, which simplifies the analysis to a standard seemingly unrelated regressors problem. This section then provides the major results in a more general situation, where there are dynamics. Section four confirms the results of the asymptotic theory with Monte Carlo evidence, and shows that this asymptotic theory is a useful guide as to the results in finite samples. In section five, empirical examples of the potential problems associated with using standard cointegration techniques for some macroeconomic and finance applications are included. Section six discusses the role of pre-testing for a unit root before applying cointegration methods. This section also shows that Johansen rank tests have power against the alternative that the regressor does not have an exact unit root, selecting a rank greater than the space spanned by the true number of linear relationships between the variables. The final section concludes. All proofs are contained in an appendix.

II. Local to Unit Roots and Cointegration.

This section presents the model discussed in this paper, and the usual approach to inference when data exhibit trends. The model considered here can be written as a triangular representation with stationary vector autoregressive errors,

$$\begin{aligned} y_{1t} &= m_1 + \alpha y_{1t-1} + v_{1t} \\ y_{2t} &= m_2 + \gamma y_{1t} + v_{2t} \end{aligned} \quad (1)$$

where $t=1...T$, m_1 and m_2 are constants ($m_1=0$ will be assumed throughout), y_{1t} and y_{2t} are both univariate with k fixed initial values, $v_t=(v_{1t},v_{2t})'$, and $\Phi(L)v_t=\epsilon_t$ where $\Phi(L)$ has all roots outside the unit circle (this model has an equivalent error correction representation, see Phillips (1991), Watson (1993)). The local to unity parameterization developed in Bobkoski (1983), Cavanagh (1985), Phillips (1987) and Chan and Wei (1987) is employed, so $\alpha=1+c/T$, where c is a fixed (nuisance) parameter. Of interest is estimation and inference on the parameter γ.

It is also assumed that ϵ_t is a martingale difference sequence with respect to $d_{t-1}=\{\epsilon_{t-1},\epsilon_{t-2}...\}$ with four moments, i.e. the following holds;

Condition A.

(i) $E[\epsilon_t|d_{t-1}]=0$.

(ii) $E[\epsilon_t\epsilon_t'|d_{t-1}]=\Sigma$

(iii) $E[\epsilon_{it}^4|d_{t-1}]<\infty$ for $i=1,2$.

In addition, one of two assumptions on $\Phi(L)$ is assumed depending on the estimator studied.

The assumption is either

Condition B1: The order of $\Phi(L)$ is finite and known; or

Condition B2: The order of $d_{21}(L)$ is finite and known[2] where we have

$d_{21}(L)=h_2(L)h_1(L^{-1})'[h_1(L)h_1(L)^{-1}']^{-1}$, and $H(L)=[h_1(L)\ h_2(L)]$ where $H(L)=\Phi(L)^{-1}\Sigma^{1/2}$.

The model considered here is quite simple yet is general enough to include regressions that are estimated in practice. An example is the regression of consumption on income. Shocks to income appear to have lasting effects, so income, here y_{1t}, is modelled as having its largest root close to if not exactly one. The second equation in (1) is simply the linear regression of consumption and income, where the errors in the consumption equation may be correlated with the error driving the variables in the long run, ϵ_{1t}, (so Σ_{12} can be nonzero) and both errors can be serially correlated.

Problems that arise in the estimation of γ are similar to those that arise in the standard SUR problem, where the equations are linked by the off diagonal element of Σ, except that this cross correlation of errors combined with serial correlation results in this linkage between the two regressions occurring at many leads and lags. If $\Sigma_{12}=0$ and there is no serial correlation, then OLS inference in the second equation proceeds as usual, t statistics testing γ have asymptotically normal distributions with the only difference being the faster rate of convergence of $\hat{\gamma}$ (at rate T rather than the usual $\sqrt{T}$).

---

[2] This assumption is not required, but simplifies the proof and accords with the assumption in Stock and Watson (1993).

In the more general case of Σ_{12} nonzero, standard asymptotic theory does not apply, and the estimate of γ converges to a nonstandard limit, as does the corresponding OLS t statistic testing that $\hat{\gamma}=\gamma$. In the special case where $\alpha=1$, i.e. there is an exact unit root in the regressor, then the two variables $\{y_{1t},y_{2t}\}$ are said to be cointegrated (Engle and Granger (1987)). There have been a large number of papers which have examined the estimation of such a system. Early methods such as OLS (proposed by Engle and Granger (1987)) have asymptotic bias of order 1/T and do not allow standard inference (Stock (1987)), but the efficient estimators listed in the introduction yield estimates which are asymptotically unbiased to order 1/T and allow standard normal inference on the parameter of interest. These efficient estimators include the methods of Johansen (1988) (denoted JOH), the Phillips and Hansen (1989) fully modified procedure (PHFM), the DOLS procedures (Phillips and Loretan (1991), Saikkonen (1991), Stock and Watson (1993)), and the method of Saikkonen (1992) (denoted SAIK). Further, Saikkonen (1991) and Phillips (1991) have given asymptotic optimality results for these methods[3] when $\alpha=1$. We will therefore refer to these procedures as the set *S* of asymptotically efficient estimators of cointegrating vectors[4]. These methods are summarized in the notation of this paper in Appendix 1.

These efficient methods have been widely used in the literature. Examples include Ogaki's (1990) examination of Engel Curves, a multitude of papers examining purchasing power

[3] This optimality applies to each of these methods except the IV versions in Phillips and Hansen (1989), when the instrument and right hand side variables are not cointegrated (see Saikkonen (1991)).

[4] This is not a complete list, but it is conjectured that other asymptotically efficient estimators have the same properties.

parity (Fischer and Park (1991), Johansen and Juselius (1992), Johnson (1993), Kugler and Lenz (1993)), Hall et al's (1992) analysis of interest rate yields, and examinations of the long run demand for money in Hoffman and Raasche (1991) and Stock and Watson (1993).

In applications of the cointegration procedure, rather than beginning with a hypothesis that suggests that $\alpha=1$ and employing cointegration, researchers usually resort to employing a pre test to examine whether or not the null of $\alpha=1$ can be rejected. This follows from the stylized facts of Nelson and Plosser (1982), which gave impetus to the study of integrated processes and cointegration in the first place. If the hypothesis of a unit root cannot be rejected, the methods of cointegration are employed. Failure to reject the null hypothesis of a unit root implies that values of α close to one would also not be rejected. Stock (1991) develops a method for placing confidence intervals on α (or correspondingly on c), and applies these to the Nelson and Plosser data set. Typically, these confidence intervals are quite wide. For example, the 90% confidence interval for industrial production in the US with 111 annual observations is $\alpha \in [0.84\ 1.03]$, corresponding to the inclusion of a value of $c=-15$ in the confidence set. The 90% confidence interval for money stock is $\alpha \in [0.69\ 1.03]$, which includes $c=-20$, and that for the Standard and Poors 500 index is $\alpha \in [0.87\ 1.04]$, which contains $c=-10$[5].

The efficient cointegrating methods in set S are derived conditional on the assumption that $\alpha=1$. As the researcher rarely knows the actual true value for α, this conditioning may not be appropriate. Also, it is only in rare cases that a hypothesized value for α comes from

[5] Stock (1991) Table 2.

economic theory. Instead, α is in general an unknown nuisance parameter. This casts doubt on the applicability of available cointegration methods unless $\alpha=1$ is implied by the economic theory under investigation. If this is not implied by the economic theory, then any method of estimation or inference would be required to be robust over a range of sizes of the largest root in the data. It is therefore of interest to examine the properties of the cointegration estimation procedures listed above in the case of α close to but not equal to one.

III. Estimation and roots local to unity

This section derives the limiting representations for the centered and standardized estimates of $\hat{\gamma}$ and the t statistic testing $\hat{\gamma}=\gamma$ when the methods of set S are employed under the generalization that α is local to unity rather than exactly one, giving the main results of the paper. The effect of this generalization is most easily shown in the case where it is known that there is no serial correlation (in equation (1), $\Phi(L)=I$). The results are then generalized to the serially correlated case. For exposition, the estimation method chosen is the DOLS method advanced by Phillips and Loretan (1991), Saikkonen (1991) and Stock and Watson (1993), although as will be shown later each of the estimators in S have the same limiting distributions and thus suffer from the same problems as the procedure analyzed here.

Under the assumption of no dynamics, efficient estimation of γ in the model in equation (1) simplifies to a standard SUR problem, where the two equations are linked only by the contemporaneous correlation between the residuals, Σ_{12}. If α is known, there are no unknown coefficients in the equation for y_{1t} and the SUR estimator of the second equation is given by

$$y_{2t} = m + \gamma y_{1t} + \varphi(1-\alpha L)y_{1t} + \eta_t^* \tag{2}$$

The correlation between y_{1t} and the regression residual has been removed by adding $v_{1t}=(1-\alpha L)y_{1t}$ as a regressor into the regression equation, so η_t^* is orthogonal to both of the right hand side variables in the regression by construction.

Under the assumptions that $\alpha=1$ and no dynamics, the special case of the DOLS estimator when there is no serial correlation is obtained. This involves the estimation of the equation (see Phillips and Hansen (1990) remark (c) p113, Stock and Watson (1993))

$$y_{2t} = m' + \gamma y_{1t} + \varphi(1-L)y_{1t} + \eta_t^{**} \tag{3}$$

which is equivalent to (2) when $\alpha=1$.

The efficient estimator of the cointegrating parameter, γ, when α is assumed to be equal to one is the OLS estimate of $\hat{\gamma}$ from the regression in equation (3). The effect of erroneously assuming α to be equal to 1, and imposing this as such can be seen by rearranging equation (2) so that we have

$$y_{2t} = m + \gamma y_{1t} + \varphi(1-L)y_{1t} - \varphi(\alpha-1)y_{1t-1} + \eta_t^* \tag{4}$$

Comparing equations (3) and (4), we see that the regression in equation (3) omits the term $\varphi(1-\alpha)y_{1t-1}$ from the regression. Thus the effect of the assumption that $\alpha=1$ can be viewed as a classical omitted variable bias problem. This term instead falls into the error term η_t^{**}. As y_t is highly persistent, this omitted variable is positively correlated with the regressor y_{1t}, and hence introduces omitted variable bias into the estimate for $\hat{\gamma}$.

This can be seen by looking at the centered and standardized estimate; we have that

$$\begin{aligned} T(\hat{\gamma}-\gamma) &= \left(\frac{1}{T}\sum \eta_t^{**} y_{1t}^{\mu}\right)\left(\frac{1}{T^2}\sum (y_{1t}^{\mu})^2\right)^{-1} \\ &= \left(\frac{1}{T}\sum \eta_t^{*} y_{1t}^{\mu}\right)\left(\frac{1}{T^2}\sum (y_{1t}^{\mu})^2\right)^{-1} - \varphi T(\alpha-1)\left(\frac{1}{T^2}\sum y_{1t-1}^{\mu} y_{1t}^{\mu}\right)\left(\frac{1}{T^2}\sum (y_{1t}^{\mu})^2\right)^{-1} \\ &= A_{1t} + A_{2t} \end{aligned} \tag{5}$$

Under the assumption that $\alpha=1+c/T$, then $A_{1t} \Rightarrow \Sigma_{11}^{-1/2}\Sigma_{2.1}^{1/2} \int J_c^{\mu}(s)dW_{2.1}(s)(\int J_c^{\mu}(s)^2 ds)^{-1}$, which is the usual result except that local to unity diffusion processes replace the Brownian Motion functionals ($J_c^{\mu}(s)$ is defined after the next equation). However, also note that $A_{2t} \Rightarrow -c\varphi$, introducing a bias of order T^{-1} to the estimate. Combining terms, the asymptotic distribution for the OLS estimator of γ in the misspecified regression (3) is can be written as

$$\begin{aligned} T(\hat{\gamma}-\gamma) &\Rightarrow \Sigma_{11}^{-\frac{1}{2}}\Sigma_{2.1}^{\frac{1}{2}}\int J_c^{\mu} dW_{2.1}\left(\int J_c^{\mu\,2}\right)^{-1} + D \\ \textit{where } D &= -c\delta\Sigma_{11}^{-\frac{1}{2}}\Sigma_{22}^{\frac{1}{2}} \end{aligned} \tag{6}$$

where $\delta=\Sigma_{12}/(\Sigma_{11}\Sigma_{22})^{1/2}$, $\Sigma_{2.1}=\Sigma_{22}-\Sigma_{11}^{-1}\Sigma_{12}^2$, and $J_c^{\mu}(s)$ is a demeaned diffusion (Ornstein Uhlenbeck) process given by $J_c^{\mu}(s)=J_c(s)-\int J_c(r)$, where $dJ_c(s)=cJ_c(s)ds+dW(s)$, and $W(s)$ is a standard Brownian Motion process.

The result in equation (6) shows that the estimates for the cointegrating parameter are biased in small samples when c and δ are nonzero. From the local to unity definition of α, the further is c from zero the less persistent (for negative c) or more explosive (for positive c) is the y_{1t} process. The parameter δ, the correlation coefficient between the two errors in equation (1), gives a standardized measure of the simultaneity between y_{1t} and ϵ_{2t}, the innovation driving the regression of interest. The bias term D/T disappears asymptotically, at the rate T, and is increasing in both c and δ in absolute values. For positive values of δ, a slightly mean reverting process ($\alpha<1$) results in the estimate for γ being overestimated (positive bias). The intuition for the direction of the bias follows directly from the omitted variable bias intuition given above, as y_{1t} and y_{1t-1} are positively correlated and the true value of the coefficient on the omitted term is $\varphi(1-\alpha)$ which is positive for c negative and δ positive[6]. Note that in the true cointegration case of $\alpha=1$, $c=0$ so $D=0$ leaving the estimate asymptotically mean zero mixed normal (Phillips and Hansen (1990), Phillips (1991), Stock and Watson (1993)).

The magnitude of this bias is affected by the variance of the residuals in equation (1). For a given value for δ, the larger is the variance of the shock driving the regressor y_{1t} the smaller is the bias. A larger variance of the residual in the equation for y_{2t}, the larger is the bias.

The term D in equation (6) leads to a departure of the limiting distribution for the t statistic testing $\hat{\gamma}=\gamma$ from a standard normal. From the standard equation for the OLS t statistic on

[6] The population value of φ is $\Sigma_{12}\Sigma_{11}^{-1}$, so it has the same sign as δ, which is given by $\Sigma_{12}\Sigma_{11}^{-1/2}\Sigma_{22}^{-1/2}$.

γ in equation (3), then the t statistic is given by

$$t_{(\hat{\gamma}-\gamma)} = \frac{T(\hat{\gamma} - \gamma)}{\sigma_{\eta}\cdot(\frac{1}{T^2}\sum (y_{1t}^{\mu})^2)^{-1}} + o_p(1) \quad (7)$$

Substituting the two terms from equation (5) for $T(\hat{\gamma}-\gamma)$ and taking the limits as $T\rightarrow\infty$ then we have

$$t_{(\hat{\gamma}-\gamma)} \Rightarrow z - c\delta(1-\delta^2)^{-\frac{1}{2}}(\int_0^1 J_c^{\mu}(s)^2 ds)^{\frac{1}{2}} \quad (8)$$

where z is distributed as a standard normal and z and $(\int J_c^{\mu}(s)^2 ds)^{1/2}$ are independent. In constructing the hypothesis test here, the bias term does not disappear asymptotically as the order of bias T^{-1} is equal to the order of the standard error employed to normalize the estimate. Again, when either there is an exact unit root or $\delta=0$, the nonstandard piece of the distribution drops out of the equation and inference is standard, as per the usual results of cointegration theory. In the general case where neither of these parameters are zero, the distribution of the t statistic testing $\gamma=\gamma_0$ is a mixture of normals with a random mean. In particular, this asymptotic distribution depends on the values of c and δ.

For $\delta>0$, the more mean reverting is the process y_{1t}, or the larger is δ, the larger is the t statistic testing $\hat{\gamma}=\gamma$ when the null is true. This leads to a tendency for the cointegration estimators to over-reject (in the upper tail) the true null hypothesis. For $\delta<0$, the rejections will be lower tail. These results also show that as the absolute value of $\delta\rightarrow1$, the size of the test goes to 1. This is true for all values of c excepting the true cointegrating case when c=0. Even for only very slight departures from c=0, if δ is sufficiently high then size will be far from its nominal value and can be close to 1.

The asymptotic size of the tests for the cointegrating vector can be evaluated more directly by writing the size (or Type 1 error) as

$$Pr[Rej\ H_0 | H_0\ True] = E[\Phi(-z^* - D_2)] + (1 - E[\Phi(z^* - D_2)]) \tag{9}$$

where Φ is the standard normal CDF, $D_2 = -c\delta(1-\delta^2)^{-1/2}(\int J_c^{\mu}(s)^2 ds)^{1/2}$, z^* is the critical value employed for the test, and $E[\cdot]$ is the expectations operator. This shows that as D_2 becomes large, all rejections will shift to one tail. As $|\delta| \to 1$, $D_2 \to \infty$ and so size approaches 1. Of course, the probability that a confidence interval contains γ_0 is simply one minus the size of the test that $\hat{\gamma} = \gamma_0$, so this suggests that coverage rates of confidence intervals constructed using the cointegrating methods will be smaller than one minus nominal size when α diverges from one and $\delta \neq 0$. Thus, for many empirical applications, the probability that the true value of the parameter of interest is contained in the confidence interval is potentially quite small.

With values for $(\int J_c^{\mu}(s)^2 ds)^{1/2}$, the expression for size given in equation (6) can be readily evaluated for various c and δ to examine the asymptotic size distortions from employing estimators in the set *S*. Figure 1 shows the asymptotic size of the cointegration t tests for δ ranging between -1 and 1, where c is equal to -5, -10 and -20 and the above expectation was evaluated with 20 000 Monte Carlo replications and 500 observations[7] (the nominal size

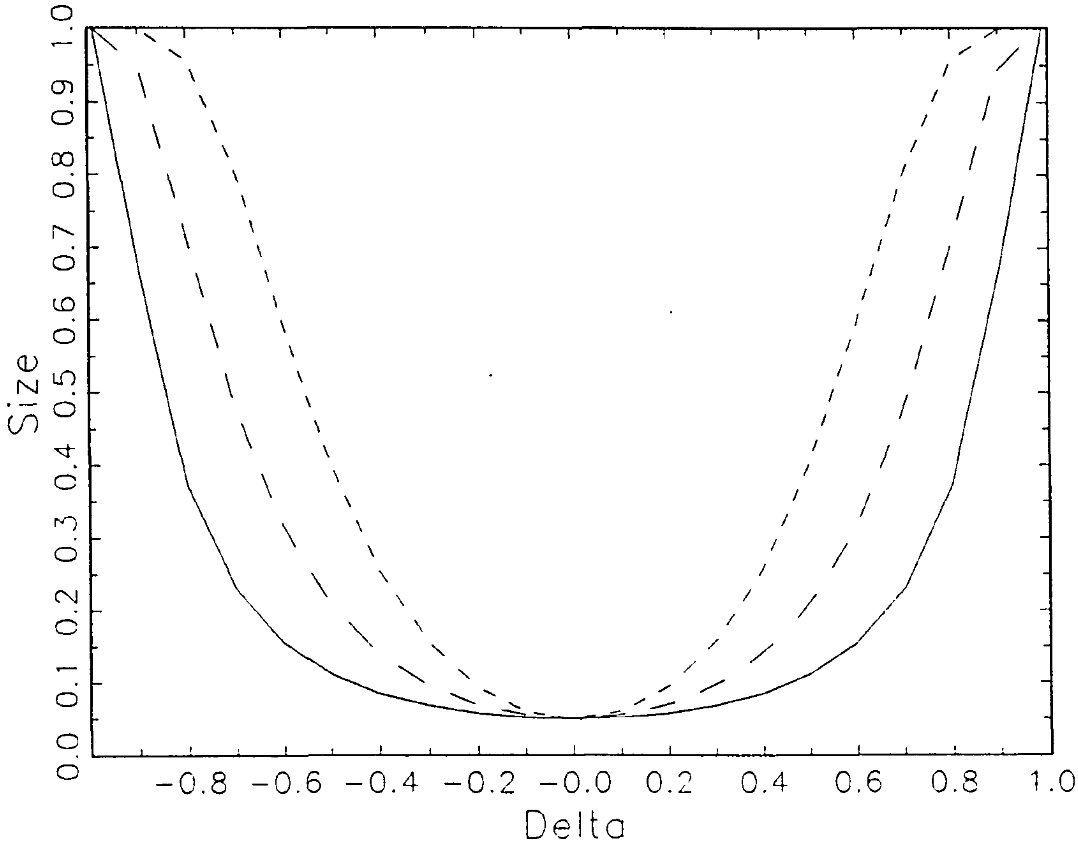
[7] The Brownian motion integral was approximated with a local to unity process driven by a standard normal. For each realization, the initial value is equal to zero. The integral was then calculated and the arithmetic mean of the realizations of the cdf reported.

of the test is 5%). This illustrates the comments made above. The size is increasing in δ at an increasing rate, and goes to 1 as δ gets large.

Interestingly, for mild endogeneity ($|\delta| < 0.5$), asymptotic size does not rise much above 10% for c=-5. But for greater amounts of endogeneity, even for such a slight departure from a unit root asymptotic size is large. Notice that as c gets closer to zero (from -10 to -5), the region over δ for which the size increases steeply is less. This region will get smaller and smaller as $c \to 0$, but the range over δ for which size is large will exist unless c is exactly zero. For lower values of c, the size can be large even for small values of δ.

These results are indicative of the results in the more general situation where there is serial correlation so that $\Phi(L)$ satisfies more general assumptions. Firstly, it will be shown that the effect of assuming $\alpha=1$ on both the bias in the cointegrating vector and inference are the same over the estimators examined here.

Figure 1: Asymptotic Size of Cointegration Hypothesis Tests



Notes: The figure shows asymptotic size of the estimators in S for a range of values for δ. The solid line gives size for $c=-5$, the long dashed line gives size for $c=-10$, and the short dashes gives size for $c=-20$. Size was calculated according to equation (9) and surrounding discussion.

Theorem 1. For the model described in equation (1) and the following paragraph, where the assumptions of Conditions A and B1 are satisfied for each estimator except DOLS, for which Conditions A and B2 hold, then the distribution of $\hat{\gamma}$ estimated using estimators from the set S have an asymptotic distribution given by

$$T(\hat{\gamma}-\gamma) \Rightarrow \Omega_{11}^{\frac{1}{2}}\Omega_{2.1}^{\frac{1}{2}}\int J_c^{\mu}dW_{2.1}^{\mu}\ (\int J_c^{\mu\,2})^{-1} + D \qquad (10)$$
$$\text{where } D = -c\Omega_{11}^{-1}\Omega_{12}$$

where Ω is equal to 2π times the spectral density of v_t at frequency zero.

This is similar to the result derived in equation (4), when there is no serial correlation, with the replacement of the short run variance by its long run counterpart. The estimate of γ converges to its true value at rate T, but suffers a bias of order T^{-1} as a result of the misspecification.

Theorem 2. Under the assumptions of Theorem 1, the asymptotic distribution of the t test for $\hat{\gamma}=\gamma_0$, where γ_0 is the true value for γ, using the set of estimators in S, is given by

$$t_{(\hat{\gamma}=\gamma)} \Rightarrow z - c\delta(1-\delta^2)^{-\frac{1}{2}}(\int_0^1 J_c^{\mu}(s)^2 ds)^{\frac{1}{2}} \qquad (11)$$

where $\delta=-\Omega_{12}/(\Omega_{11}\Omega_{22})^{1/2}$.

These results again are similar to the results in the absence of serial correlation presented in equation (5), and as is standard the only effect of the serial correlation when enough lags are present in the estimation is that variances are replaced by their long run equivalent, the scaled spectral density of the residuals v_t at frequency zero. In the methods based on the triangular representation where non parametric corrections are used (Phillips and Hansen (1989) fully modified OLS), the bias is due to an omitted variable term (as described in the example above). This result will follow for any non parametric correction where v_{1t} is replaced by Δy_{1t} rather than the true quasi difference of y_{1t} in the construction of the orthogonal dependant variable $\hat{y}^*_{2t}$. Thus these results extend to the instrumental variable procedure in Phillips and Hansen (1989)[8], where combined with an estimate of v_{1t}, $v_{3t}=(1-\alpha_3)y_{3t}$, where y_{3t} is the instrument, appears in the construction of $\hat{y}^*_{2t}$, and α_3 is in general unknown[9].

In the DOLS procedure (Phillips and Loretan (1991), Saikkonen (1991), Stock and Watson (1993)), the misspecification is in the lags and leads utilized to orthogonalize the error terms. This misspecification is correlated with the levels variable, rewriting the model in its canonical form shows that the estimate $\hat{\gamma}$ has a population value different from that of the cointegrating parameter γ. The difference is the bias term.

In the full system error correction method of Johansen (1988), the result is also due to the presence of an omitted variable term. In the case of these estimators, the non unit value for α causes the coefficient matrix on the levels term in the ECM to not be of reduced rank. In the Johansen (1988) case, specifying this matrix to be of reduced rank leads to the appearance of the omitted variable. The proof, by way of using an estimator numerically

[8] This procedure is not amongst the asymptotically efficient cointegration vector estimators.

[9] To remain in the class of efficient estimators, y_{3t} and y_{1t} must be cointegrated in the long run so $\alpha_3=\alpha$, which is unknown.

equivalent to the MLE, shows this result for other maximum likelihood estimators as well, thus the estimator in Ahn and Reinsel (1990) has the same properties as stated in the theorems. The result for the Saikkonen (1992) estimator is similar to the DOLS estimator, the misspecification here results in population value for the estimator now includes an extra term, which is the bias in the theorem.

It is conjectured that other efficient estimators of cointegrating vectors not considered here, such as the Park (1992) canonical correlations procedure and the Engle and Yoo (1993) three stage least squares estimator, have the same properties as the estimators in set S.

From the asymptotic results above, one may speculate that a solution would be to estimate α, using this to construct the quasi difference. This will, unfortunately, also result in the loss of χ^2 inference. This can be seen by examining the expression for $T(\hat{\gamma}-\gamma)$ in the DOLS case given in equation (39) in the appendix. The additional bias term involves an expression $T(\alpha-1)$, which in the above analysis is equal to c. If α were replaced by an estimate, say $\hat{\alpha}$, then due to the result that estimates of α are consistent at rate T in the local to unity setting (Bobkoski (1983), Cavanagh (1985), Chan (1988), Phillips (1987)), then $T(\hat{\alpha}-1)$ converges to a nonstandard distribution. This non standard distribution would replace c in the results above, thus even for $\alpha=1$ inference would be non standard. Further, the critical values for this procedure would depend on the true value for c, which of course is unknown. This point was made in this context by Phillips (1991), in his remark (c). See Cavanagh, Elliott and Stock (1993) for further discussion of this point.

IV. Monte Carlo Results

The asymptotic results show that for large numbers of observations, the estimate of the cointegrating vector will still be quite close to its true value (with a bias of order T^{-1}) but for a reasonably large range of δ and c, the tests will badly over-reject the true null hypothesis, even asymptotically. Monte Carlo results can be used to examine the extent of these problems in more usual sample sizes, and give some guidance as to the usefulness of the above asymptotic results in applications.

The model analyzed in the Monte Carlo is a generalization of the familiar triangular representation used by Stock and Watson (1993), Phillips and Loretan (1991), and Hansen and Phillips (1990). The model can be written as

$$\begin{aligned} y_{1t} &= \alpha y_{1t-1} + u_{1t} \\ y_{2t} &= \gamma y_{1t} + u_{2t} \end{aligned} \tag{12}$$

where

$$A(L)u_t = \epsilon_t, \qquad \epsilon_t \ iid\ N(0,\Sigma) \tag{13}$$

$A(L)=I_2-AL$ and $\alpha=1+c/T$.

The estimators examined are Engle and Granger OLS estimator and the estimators in the set S. In all cases, a constant is included to deal with initial observations (and accord with usual practice). The lag lengths included for each case depend on the sample size.

In the simplest case, all elements of A are set to zero so there is no feedback and no short run dynamics, and Σ has diagonal elements equal to one and a non zero off diagonal component, Σ_{12} (which here equals δ due to the normalization of the variances), which is less than or equal to 1 in absolute value.

The implications of Theorem 1 can be seen in Table 1a, which documents the bias in the estimation of $\hat{\gamma}$ for a range of values for δ when c=-5 and -10, and A=0[10]. In this experiment, T=500. From the results of theorem 1, the expected bias in each case can be computed, this is give in the second column E(D). The remaining columns give the average bias calculated from 5000 replications of the MC experiment for each estimator. The expected bias and that of each of the efficient statistics are almost identical, a result to be expected with 500 observations. Only the PHFM estimator has an estimated bias different from the asymptotic bias at three decimal places, this estimator having slightly higher bias than expected from the asymptotic theory in some cases. Here, as expected, with $\delta>0$ and mean reversion the bias is positive and increasing in δ and increasing in $|c|$ as implied by the theoretical results.

[10] Only positive values for δ are reported, as the size results are symmetric in δ (see equation (6)). For $\delta<0$, the asymptotic bias is negative but the same size as for $\delta>0$.

Table 1a: Bias In Cointegration Vectors, T=500

Bias when c=-5.

Σ_{12}	E(D)	DOLS	PHFM	JOH	SAIK	OLS
0	0.000	0.000	0.000	0.000	0.000	0.000
0.1	0.001	0.001	0.001	0.001	0.001	0.003
0.2	0.002	0.002	0.002	0.002	0.002	0.006
0.3	0.003	0.003	0.003	0.003	0.003	0.009
0.4	0.004	0.004	0.004	0.004	0.004	0.012
0.5	0.005	0.005	0.005	0.005	0.005	0.015
0.6	0.006	0.006	0.006	0.006	0.006	0.018
0.7	0.007	0.007	0.008	0.007	0.007	0.022
0.8	0.008	0.008	0.009	0.008	0.008	0.025
0.9	0.009	0.009	0.010	0.009	0.009	0.028

Bias when c=-10.

Σ_{12}	E(D)	DOLS	PHFM	JOH	SAIK	OLS
0	0.000	0.000	0.000	0.000	0.000	0.000
0.1	0.002	0.002	0.002	0.002	0.002	0.005
0.2	0.004	0.004	0.004	0.004	0.004	0.010
0.3	0.006	0.006	0.006	0.006	0.006	0.015
0.4	0.008	0.008	0.009	0.008	0.008	0.020
0.5	0.010	0.010	0.011	0.010	0.010	0.025
0.6	0.012	0.012	0.013	0.012	0.012	0.030
0.7	0.014	0.014	0.015	0.014	0.014	0.035
0.8	0.016	0.016	0.017	0.016	0.016	0.040
0.9	0.018	0.018	0.020	0.018	0.018	0.045

Notes: The first column gives the size of the cross correlation of the residuals. The second column gives the expected bias calculated from the results of equation (6). On each case, the cointegrating vector estimates included 1 lead and lag for DOLS and 1 lag for JOH. The non parametric estimates used in PHFM are calculated using the Bartlett kernel with 3 lags. The results report the average bias in the estimates of γ from 5000 Monte Carlo replications.

Table 1b: Bias In Cointegrating Vectors, T=100.

bias when c=-5.

Σ12	E(D)	DOLS	PHFM	JOH	SAIK	OLS
0	0.000	0.000	0.000	0.000	0.000	0.000
0.1	0.005	0.005	0.006	0.005	0.005	0.014
0.2	0.010	0.010	0.013	0.010	0.010	0.028
0.3	0.015	0.015	0.020	0.014	0.015	0.043
0.4	0.020	0.020	0.026	0.019	0.020	0.057
0.5	0.025	0.025	0.033	0.024	0.025	0.071
0.6	0.030	0.030	0.040	0.029	0.030	0.085
0.7	0.035	0.035	0.046	0.034	0.035	0.100
0.8	0.040	0.040	0.053	0.039	0.040	0.114
0.9	0.045	0.045	0.060	0.045	0.045	0.128

bias when c=-10.

Σ12	E(D)	DOLS	PHFM	JOH	SAIK	OLS
0	0.000	0.000	0.000	0.000	0.000	0.000
0.1	0.010	0.010	0.013	0.009	0.010	0.022
0.2	0.020	0.020	0.025	0.019	0.020	0.045
0.3	0.030	0.030	0.038	0.028	0.030	0.068
0.4	0.040	0.040	0.051	0.038	0.040	0.090
0.5	0.050	0.050	0.064	0.047	0.050	0.113
0.6	0.060	0.060	0.077	0.057	0.060	0.135
0.7	0.070	0.070	0.090	0.068	0.070	0.158
0.8	0.080	0.080	0.103	0.078	0.080	0.181
0.9	0.090	0.090	0.105	0.089	0.090	0.203

Notes: As for Table 1a.

Also included, in the final column, is the bias associated with the straightforward (inefficient) OLS estimation of the cointegrating vector as suggested in Engle and Granger (1987). These biases are of far greater magnitude than those of the other estimates of the cointegration vector. For $c=-5$, the bias in OLS estimates are at least three times the size of the bias of the efficient tests, for $c=-10$ the OLS bias is at least 2.5 times the bias of the efficient tests. The reason for this result can be seen by examining the correctly transformed equation (2) above. In this equation, if α is accurately chosen, then the estimate of γ from (2) will be unbiased. The larger is the 'mistake' in choosing α for the construction of the extra regressor, the larger is the omitted variable bias. Estimation with the Engle and Granger (1987) OLS estimator omits this term altogether. This means that the whole second term on the right hand side of equation (2) is an omitted variable. The efficient class tests S chose $\alpha=1$, which is closer to the true value of α for α local to unity so the bias is smaller.

Table 1b repeats these results for the $T=100$ case. Again, the expected bias in each case is readily calculated and appears to correspond well with the estimated biases for each of the estimation methods (except OLS for the reasons stated above). For both the DOLS and the Johansen estimators, the bias is almost identical to the expected bias, for the PHFM estimator the bias is slightly larger. As implied by the analytic term for the expected bias, these biases are around five times larger than those for $T=500$.

Table 2: Size for Various c

δ=0.5

c	Asy.	DOLS	PHFM	JOH	SAIK
-20.000	0.403	0.352	0.469	0.384	0.421
-10.000	0.208	0.195	0.230	0.209	0.222
-5.000	0.112	0.110	0.121	0.121	0.123
0.000	0.050	0.049	0.047	0.051	0.058
2.000	na	0.204	0.201	0.201	0.223

δ=0.9

c	Asy.	DOLS	PHFM	JOH	SAIK
-20.000	0.999	0.997	1.000	0.997	0.997
-10.000	0.948	0.918	0.967	0.928	0.932
-5.000	0.668	0.633	0.695	0.651	0.661
0.000	0.050	0.049	0.041	0.051	0.058
2.000	na	0.705	0.676	0.709	0.714

Notes: The estimators are constructed as in Table 1a. The column headed Asy. gives the asymptotic size calculated as per equation (9) and the discussion following. The remaining columns are the percentage of rejections for each case where nominal size is 5%. There were 5000 Monte Carlo replications.

Of more interest is inference on the cointegrating vector. Monte Carlo results for the size of the t statistic testing the true null hypothesis in the case T=500 (to approximate the asymptotic results) for $\delta=0.5$ and a range of values for the local to unity parameter c are contained in the first panel of Table 2. The first column of this table gives the value of c, the second the expected asymptotic size of the tests calculated according to equation (6). The remaining columns give the Monte Carlo results for the sizes of each statistic.

In the case where $\alpha=1$ (c=0), cointegration is valid, so the size of the t tests are close to their nominal (asymptotic) size of 5%. Departures from c=0 in either direction result in size increasing; for c=-5 the size of the t statistic testing the true null hypothesis is around 12%, for c=-10 size is around 20%. This means that in these cases the true null hypothesis for the cointegrating parameters will be often rejected when c is not zero, as implied by the second theorem above. These results are common across all of the methods for estimating cointegrating vectors that are considered here.

The increases in empirical size of the t statistics are greater for larger values of δ. In panel B of Table 2, the results are given for $\delta=0.9$. Here, when c=-5, empirical size is around 65%, which is 13 times nominal size. In this case the true null hypothesis will be rejected more often than it is accepted, even for this very slight departure from the correctly specified model. As noted in the introduction, the unit root tests can have very low power against alternatives such as these.

Table 3: Size of Tests of True H0 for various δ.

Panel A : T=500, c=-10.

δ	Asy	DOLS	PH	JOH	SAIK
0.0	0.050	0.050	0.050	0.050	0.050
0.1	0.055	0.053	0.054	0.053	0.053
0.2	0.069	0.064	0.064	0.068	0.070
0.3	0.096	0.091	0.093	0.095	0.098
0.4	0.140	0.134	0.138	0.134	0.137
0.5	0.208	0.203	0.207	0.204	0.210
0.6	0.315	0.309	0.316	0.311	0.317
0.7	0.478	0.467	0.480	0.462	0.473
0.8	0.709	0.706	0.719	0.698	0.707
0.9	0.948	0.948	0.952	0.943	0.946

Panel B : T=100, c=-10.

δ	Asy	DOLS	PH	JOH	SAIK
0.0	0.050	0.050	0.050	0.050	0.050
0.1	0.055	0.052	0.053	0.055	0.058
0.2	0.069	0.067	0.070	0.070	0.077
0.3	0.096	0.099	0.108	0.097	0.112
0.4	0.140	0.149	0.164	0.134	0.158
0.5	0.208	0.218	0.245	0.199	0.229
0.6	0.315	0.335	0.372	0.292	0.336
0.7	0.478	0.491	0.542	0.425	0.479
0.8	0.709	0.718	0.771	0.643	0.707
0.9	0.948	0.947	0.967	0.913	0.942

Notes: Entries are size adjusted rejection rates for each of the estimators over 5000 replications. For each estimator, the absence of serial correlation was treated as known. The nominal size of the tests is 5%.

As can be seen from the results of Table 2, the asymptotic theory gives a relatively good guide as to the size of the tests for 500 observations. To examine this further, Table 3, Panel A, examines size for a range of values for δ when $c=-10$ and $T=500$. In these experiments, the estimators are correctly specified[11]. The computed size here is very close to that expected from the asymptotic theory. Panel B undertakes the same experiment with $T=100$. Here it can be seen that with fewer observations, the size is slightly larger than asymptotic size for lower values of δ and slightly less for larger values of δ. In general, however, the asymptotic theory gives a good guide as to what might be expected in practice.

The asymptotic results also hold in the case where the regressors are serially correlated, some examples of which are given in Table 4. In these experiments, the restriction that A is equal to zero is relaxed and different values for A_{11} and A_{21} are considered. Panel A gives the results for 500 observations when A_{11} is set equal to 0.5 and A_{21} to 0.2. Values of 0, 0.3 and 0.5 are considered for Σ_{12}. These parameterizations yield true values of delta reported in the fourth column. The fifth column reports the asymptotic sizes for the case where $c=-10$, calculated according to equation (9) above. The following columns give the average sizes for each of the tests using the four methods over 5000 replications. For the DOLS, JOH and SAIK procedures, the estimated sizes accord closely with asymptotic theory. The PHFM method has significantly smaller size distortions in these cases.

---

[11] There is no serial correlation, so no lags of dependant variables are used in the parametric methods. For the PHFM method, long run variance covariance matrices are replaced by the standard variance covariance matrix.

Table 4: Size when Regressors are Serially Correlated

A_{11}	A_{12}	Σ_{12}	δ	ASY	DOLS	PHFM	JOH	SAIK
				Panel A: T=500				
0.5	0.2	0	0.196	0.068	0.065	0.053	0.077	0.075
0.5	0.2	0.3	0.464	0.180	0.188	0.098	0.191	0.191
0.5	0.2	0.5	0.629	0.355	0.378	0.173	0.354	0.358
				Panel B: T=500				
-0.5	0.2	0	0.196	0.068	0.064	0.064	0.076	0.079
-0.5	0.2	0.3	0.464	0.180	0.179	0.208	0.196	0.204
-0.5	0.2	0.5	0.629	0.355	0.341	0.421	0.355	0.378
				Panel C: T=100				
0.5	0.2	0	0.196	0.068	0.072	0.087	0.144	0.139
0.5	0.2	0.3	0.464	0.180	0.173	0.141	0.244	0.243
0.5	0.2	0.5	0.629	0.355	0.333	0.227	0.383	0.385
				Panel D: T=100				
-0.5	0.2	0	0.196	0.068	0.073	0.102	0.153	0.180
-0.5	0.2	0.3	0.464	0.180	0.179	0.343	0.260	0.309
-0.5	0.2	0.5	0.629	0.355	0.330	0.654	0.391	0.477

Notes: The model in equations (12) and (13) is estimated with c=-10 and with 3 lags for each procedure excepting the PHFM procedure. The non parametric estimate of the long run variance covariance matrix was computed using the Bartlett kernel with 5 covariances (This was to control size in the c=0 case). The number of Monte Carlo replications is 5000.

The next panel, Panel B, considers the same experiment except that A_{11}=-0.5. The same results hold, although in this case the PHFM procedure now has sizes above those expected by asymptotic theory. This behavior of the PHFM estimator can be traced to the method of estimation of the long run covariance matrix, a result that will not be further examined here.

Panel C and Panel D undertake the same experiments and Panel A and Panel B respectively with 100 observations. This results in size deviating from that of the asymptotic theory in the usual way for time series results, empirical size is generally larger than nominal size when the number of observations is low. The asymptotic theory here still provides a reasonable guide to the types of distortions likely to be found in practice.

V. Applications : Macroeconomic and Financial Regressions

This section presents some empirical cointegrating regressions, estimating the confidence interval on the regressor and the spectral density matrix at frequency zero between the residuals. As was shown in Section 4, the potential for problems to arise is greater the further is α away from 1 and the larger is δ. A confidence interval on α is suggestive of how likely are departures of α from 1 and how large these departures may be. Consistent estimators of δ enable examination of the size of potential effects through this channel. These estimates combined will enable examination of the range of potential asymptotic coverage rates of the confidence intervals constructed around the cointegrating estimates produced by the cointegration techniques in set S. Examples from macroeconomics, international finance and finance are included.

For each of the examples, the 95% confidence interval on the largest root of the regressor is calculated by the method of Stock (1991). With the available number of observations, this translates into a range of possible values for $c=T(\alpha-1)$. We estimate the spectral density matrix of the estimated residuals at frequency zero using an autoregressive (AR) estimator, with the lag length selected by the Bayes Information Criterion. The lag length was truncated so that the number of lags are employed is between some given maximum and zero (the results are robust to increasing the lag length). To compute these estimates, the residuals of the model in equation (1) first must be estimated. This entails the quasi-differencing of the regressor, for which the median unbiased estimate of α is used. The remaining residual is the residual from the estimated cointegrating relationship, where the cointegrating parameters are estimated using the DOLS estimator, and the residuals constructed as $y_{2t}-\hat{\gamma}y_{1t}$. The motivation for using these estimates of γ is that despite the potential biases due to $\alpha\neq1$, these biases were shown in the previous section to disappear at rate T and also be less than the biases from OLS estimation[12].

The range of values for c and the point estimates for δ can be used to examine potential coverage rates of the estimated confidence intervals. As there is a 95% probability that the constructed confidence interval contains the true value for α, the range of potential coverage

---

[12] As discussed after lemma 3 in Appendix 2, the AR estimator of the spectral density matrix of the errors using estimates of the errors converges to its true value despite estimation biases in γ. Small sample properties will, however, be affected.

rates of confidence intervals on γ using the estimators in set S for any nominal size can be computed with a probability of 95%, using the most and least extreme values for α. Typically, $\alpha=1$ is included in the confidence interval so the highest asymptotic coverage rate is equal to the nominal coverage rate. For other values of α, the coverage rate will be less than the nominal coverage rate.

The regression of consumption on income is one potential cointegrating relationship that could be investigated. Using quarterly post-war data[13], the estimated 95% confidence interval on the largest root in the regressor, income, is that $\alpha \in [0.885, 1.03]$. With the sample size employed here, this translates to values for c between -20.3 and 4.8. In addition, three definitions of consumption are employed, these are total consumption, non-durable consumption and consumption of services. For the regression of total consumption on income, the estimate of δ was 0.28. Using this estimate of δ, this translates into coverage rates of between 95% (if $\alpha=1$) to 86% (if c=-20). Similar results under the AR estimator obtain for the regression using non durables consumption in place of total consumption, where the estimate for δ is 0.269. When consumption of services is the dependant variable the estimate for δ is 0.52, which would result in coverage rates potentially as low as 56%[14].

---

[13] Data sources and definitions for this and the other examples are included in Appendix 3.

[14] Some small scale Monte Carlo results suggest that the method employed here underestimate slightly the true value of δ when it is positive. This underestimation increases when the lag length is overspecified. The underestimation is greater for $\alpha<1$, and decreasing in α.

The relationship between the forward and spot rates for exchange rates has generated a large amount of empirical work (see Baille and McMahon (1989)). Recent estimations of this relationship have included regressions of the spot rate on the lagged forward rate, e.g. Evans and Lewis (1992), employing the cointegration methods as both of these series show considerable persistence. Employing monthly data on the spot and forward markets for the YEN/US exchange rate for the period September 1977 to July 1990, the 95% confidence interval on the largest root in the forward rate is $\alpha \in [0.92, 1.03]$, corresponding to values for c between -10 and 5.3. Whilst the potential deviations from $\alpha=1$ here are modest, δ is estimated at 0.778. This suggests that cointegration methods will yield an asymptotic coverage rate of somewhere between the desired 95% rate if α is truly 1 to a very low 40% if c were -10.

A finance regression that is often examined is the regression of the expected returns on the stock market on the lag of the dividend price ratio. Employing 113 annual observations on the S&P index from the dataset used by Shiller (1981), the estimated confidence interval for the largest root in the dividend price ratio includes values for α between 0.825 and 1.03. This corresponds to values for c between -20 and 3.9. The estimated value for δ is around -0.67. This suggests potential coverage rates of the constructed coverage rates as low as 24% (for c=-20). Even if α was equal to its median unbiased estimate (corresponding to a value of c of roughly -10), the coverage rate would be only 58%.

These results show that there is great potential for cointegration methods to give misleading results in real applications. In none of the above examples do we have any theory that the true $\alpha=1$, even in the exchange rate example this hypothesized value for α often only arises due to the assumption that fundamentals driving the exchange rate have a unit root.

VI. Pretests and Rank Tests

The above analysis has been cast when the researcher cannot reject a unit root in standard tests for this phenomena, however no discussion of the effect of pre-tests has been given. In the estimation of cointegrating vectors, two types of pretests are generally undertaken. The first is univariate tests of the null of non stationarity of the vector of y_t variables. The second is the rank test of Johansen (1988), which tests for the null of r cointegrating variables in the system.

In other work (Elliott and Stock (1992), Cavanagh, Elliott and Stock (1993)), the effect of pre test bias in OLS estimation of linear regressions due to the application of conventional unit root tests asymptotically misclassifying local to unity processes as I(1) is studied. Typically, pre-test bias is found to increase the empirical size of the test over the nominal size when data is generated by a local to unity process where $c<0$. In practice, for α sufficiently far from one, the researcher can reject a unit root in the regressor[15], but the range over values of α for which these tests have low power is large. For example, the results of Elliott, Rothenberg and Stock (1992) show that asymptotically, the greatest power we can expect to achieve using classical methods against an alternative such as $c=-5$, is only

[15] See Elliott, Rothenberg and Stock (1992) for asymptotic power of most popular unit root tests.

32% when an intercept is included in the specification of the unit root test. This best possible power drops to only 10% when a time trend must be included. This low power to distinguish between $\alpha=1$ and local alternatives suggests that pre-testing data using univariate unit root tests will not be powerful enough to rule out the possibility that the results in Section three above are relevant in practice.

An alternate pretest often employed in the estimation of cointegrating relationships is the Johansen (1988) likelihood ratio test for the rank of the long run matrix in the error correction model. In the case that all the variables in the system have a unit root, the row rank of this matrix gives the number of long run relationships between the variables. In the appendix it is shown that the error correction model derived from equation (1) can be written $\Delta Y_t = \Psi Y_{t-1} + \pi(L)\Delta Y_{t-1} + \epsilon_t$, where $Y_t=\{y_{1t}\ y_{2t}\}$ and $\Psi=P\Phi(1)P^{-1}M$, and $P^{-1}M$ is given by

$$\begin{bmatrix} (\alpha-1) & 0 \\ \gamma & -1 \end{bmatrix} \tag{14}$$

When $\alpha=1$, $P^{-1}M$ is of reduced rank leading to the usual cointegration result that the long run matrix Ψ is of reduced rank (in equation (1), the row rank would be one as there is one long run relationship between the variables). When α is not equal to one, $P^{-1}M$ is no longer of reduced rank so the coefficient matrix on Y_{t-1} is not of reduced rank. This suggests that rank tests of the form of Johansen (1988) will have power not only in the direction of extra cointegrating vectors but also against the alternative that α is not equal to one, by rejecting the null hypothesis that the rank of Ψ is one against the alternative of two in the model of this paper. In the simple bivariate case analyzed here, the result that there are two cointegrating vectors is analogous the result that the data is stationary. With more variables in the model, this is not true, and the possibility arises that the procedure concludes there are more cointegrating relationships than there are stable relationships between the data is a real one, as the remaining 'cointegrating' vectors are simply the finding of stationary variables in the system.

We do not expect that this type of test will have better power to distinguish between a unit root and a stationary alternative in the regressor, as the rank test method does not include the information known in the univariate test, specifically that the rank of the cointegrating relationships is known. However, the small sample power of this test is examined in a small Monte Carlo experiment in Table 5. In this experiment, the model in equations (12) and (13) is simulated with A=0. For two values of δ, δ=0.5 and 0.9, the percentage number of rejections of the null hypothesis of one cointegrating vector for a range of values for c were computed. For each replication, T=100 and 1000 replications were used to generate the results. As would be expected, size is close to the nominal size of 5% when the rank is truly equal to one (i.e. c=0), and is not affected much by δ. When c departs from zero, the null is rejected more often. The results show that finite sample power against the alternative that c=-5 is only around 12%. Comparing this to the small sample results of univariate tests in Elliott, Rothenberg and Stock (1992), this places the power performance in the direction of stationarity alternatives between the Dickey and Fuller (1979) t test and $T(\rho-1)$ tests. As was shown in that paper, other tests are available which achieve significantly greater power than these tests.

Table 5: Power of Rank Test against $\alpha \neq 1$.

c	$\delta=$ 0.5	0.9
-20	0.751	0.710
-10	0.290	0.285
-5	0.118	0.116
0	0.050	0.051
2	0.094	0.086

Notes: Reported are rejection rates for the null of 1 cointegrating vector when an intercept is included in the Johansen (1988) test. The model generating the data is as for Table 1, with T=100. One lag was employed in estimation. The nominal size is 5% and the number of replications equal to 1000.

VII. Discussion and Conclusion

The clear result of the preceding analysis is that using cointegrating techniques which are conditioned on the premise that the largest root in the explanatory data is equal to one will result in tests of hypotheses with potentially extreme size distortions if the data does not in fact have an exact unit root. From a theoretical perspective, this means that cointegration techniques should not be applied unless the null hypothesis of the economic theory includes the premise that the largest root of the series is indeed equal to one. In practice, the result is not so severe. The theory shows that if the simultaneous equations bias is not so severe, then even with reasonably large deviations of c from zero does not distort the size too much. The results also show that in these misspecified cases, it is still better to use the efficient class estimators examined here rather than OLS. This is especially true for point estimation; point estimates are consistent asymptotically and have smaller biases than OLS.

An alternate way to look at the results of hypothesis tests from the class of estimators S is that the cointegration methods examined here have power in the direction of the largest root diverging from one as well as in the direction of γ diverging from its hypothesized value. Thus, these tests are in fact testing a joint null hypothesis, $\hat{\alpha}=1$ and $\hat{\gamma}=\gamma$. If $\alpha=1$ is part of the null hypothesis being investigated by the researcher, then rejections can be viewed as a rejection of this joint null hypothesis. The same is true of the Johansen (1988) rank test.

The results also suggest that many of the empirical analyses of theories where a unit root is not a part of the null hypothesis may have rejected the hypothesis tested spuriously, where instead the theory is true but the data has its largest root not equal to one. This suggests that the conclusions of such studies be reassessed, especially if the null hypothesis is rejected despite the method of estimation obtaining an estimate of the parameter vector which is plausible from an economic viewpoint.

Fortunately, the size of this simultaneous equations bias is consistently estimable in the local to unity cases, and in fact is generally estimated as a by product of the cointegration estimation procedures. In the Phillips and Hansen (1990) and Saikkonen (1992) methods, the matrix Ω must be estimated, and this contains the relevant information. In the DOLS methods, the estimated value of d(1) plays the same role. Alternatively, the long run variance covariance matrix (spectral density of the residuals at frequency zero) can be estimated in the usual way (Andrews (1991), Andrews and Monahan (1992)). In any application, this matrix can be estimated consistently to give the researcher some idea as to the fragility of any hypothesis rejection using cointegration methods. It is suggested that such estimates be provided along with results of tests and confidence intervals calculated using the methods in set S to enable readers to evaluate the potential for the size problems discussed in this paper.

Finally, it can be seen that there is much room for improvement in the techniques employed to undertake inference with stochastically trending data. The powerful optimality results of Phillips (1991) and Saikkonen (1991) are seen to be relevant only for a smaller class of economic problems, where the null hypothesis of the economic theory holds that $\alpha=1$, rather than the more usual consideration that a unit root cannot be rejected in statistical tests. This suggests that research should proceed towards estimation and hypothesis testing when the size of the largest root in the regressor is not known *a priori*. There are currently two methods in the recent literature which explicitly attempt to deal with the uncertainty over the value of α. The first is an extension of the fully modified approach taken by Phillips and his coauthors. The extension of the fully modified approach are found in Kitamura and Phillips (1992) and Phillips (1993a,1993b). These papers extend the procedure to models where the order of the cointegrating space is unknown, without the loss of unbiasedness and chi-square inference. These methods still require the construction of an orthogonal dependant variable vector, constructed in the method of Phillips and Hansen (1989), i.e. they impose that the largest root of the regressor is one. In the case of Phillips (1993a), this requires using the first difference of all of the data and relying on the result that if y_{1t} were truly stationary, then its difference is I(-1) and is op(1) in the long run, so the non parametric correction when the data is stationary does not affect the results[16]. When α is considered fixed (whether equal to one or less than one), then this results in chi-square inference - in the method of Phillips and Hansen if the data is I(1) or by usual stationary CLT results if the data is I(0). Thus, this method treats variables with their largest root equal to α where α is close to one as stationary variables. The central result from local to unity asymptotics is that for such values of α, the asymptotic distribution resulting from considering large values of α as fixed is not a good guide to the types of distributions seen with reasonable amounts of data, but instead the distribution resulting from considering c fixed does result in an asymptotic distribution which appears relevant. This means that the biases in the parameter vector relating data which appear best described by local to unity generating processes and the asymptotic bias in inference on these parameters will hold for these extensions as well. The extent to which this bias appears in practice has not yet been investigated.

A second set of approaches are followed in Cavanagh, Elliott and Stock (1993). In these methods, classical confidence intervals for γ are constructed taking into account the nuisance parameter c directly. These methods also have not been fully investigated, although preliminary results suggest that losses in power of the tests is not too great when α is unknown, but size is difficult to control (with empirical size tending to be below nominal size). This class of solutions to the problem are currently being pursued by the author.

[16] To achieve this, very specific controls are required to be placed on the speed at which covariances are added in the construction of the non parametric estimates of the spectral density of the residuals at frequency zero.

Appendix 1: Cointegration Estimators

This Appendix presents the estimators in the notation of this paper.

PHFM Method (Phillips and Hansen (1989)).

The PHFM estimate of γ, is given by

$$\hat{\gamma}^{+*} = [\sum_{1}^{T} \hat{y}^{*}_{2t} y_{1t} - T\hat{M}^{*}_{12}][\sum_{1}^{T} y^{2}_{1t}]^{-1} \tag{15}$$

where $\hat{y}^{*}_{2t}=y_{2t}-\hat{\Omega}_{12}\hat{\Omega}^{-1}_{11}\Delta y_{1t}$. The estimates of $\hat{M}_{12}$, $\hat{\Omega}_{11}$ and $\hat{\Omega}_{12}$ are calculated by obtaining consistent estimates of Ω, which is equal to 2π times the spectral density of the residuals of equation (1) at frequency zero. The estimate of ϵ_{1t} is simply Δy_{1t}, and the residual of the cointegrating relationship can be constructed using an OLS estimate of γ. The estimate of Ω based on these residuals can be calculated using the methods of Andrews (1991), Andrews and Monahan (1992), or by using AR estimates as in lemma 3 below. The t statistic is constructed by normalizing $T(\hat{\gamma}^{+*}-\gamma)$ by a consistent estimate of $\hat{\Omega}_{2.1}^{1/2}$ and $(T^{-2}\Sigma y_{1t}^{2})^{-1/2}$.

DOLS method (Phillips and Loretan (1991), Saikkonen (1991), Stock and Watson (1993)). This method estimates γ from the OLS regression

$$y_{2t} = m + d(L)\Delta y_{1t} + \gamma y_{1t} + \eta^{**}_{2t} \tag{16}$$

where m is a constant, d(L) is a two sided polynomial where enough lags have been added (see Stock and Watson (1993) p798 or Saikkonen (1991) for a discussion of this). The t statistic is constructed in the usual way where the standard deviation of the residual is calculated using methods robust to serial correlation.

Saikkonen method Saikkonen (1992).

This method involves the running of the vector error correction model

$$\Delta y_t = \Psi y_{t-1} + \Pi x_t + v_t^* \tag{17}$$

where $x_t = [\Delta y'_{t-1}, \ldots, \Delta y'_{t-k+1}, \Delta y'_{1t-k}]'$ and constructing the estimator of γ as

$$\tilde{\gamma} = -(\tilde{\Psi}_2'\tilde{\Sigma}^{*-1}\tilde{\Psi}_2)^{-1}(\tilde{\Psi}_2'\tilde{\Sigma}^{*-1}\tilde{\Psi}_1) \tag{18}$$

where $\Psi = [\Psi_1\ \Psi_2]$ and $\Sigma^* = E(v^*_t v^*_t{}')$. This residuals variance covariance matrix can be constructed in the usual way. For inference, the variance of the estimator is given by $\Omega_{2.1}(T^{-2}\Sigma y_{1t-1}{}^2)^{-1}$, where $\hat{\Omega}_{2.1}$ is given by the denominator of the estimator in equation (15).

Johansen method (Johansen (1988)).

The Johansen (1988) estimator is an exact MLE derived from the concentrated likelihood given by

$$l(\Sigma^*,\psi) \propto -\frac{T}{2}\ln|\Sigma^*| - \frac{1}{2}\sum_{t=k+1}^{T}(\hat{R}_{0t} - \psi\hat{R}_{kt})\Sigma^{*-1}(\hat{R}_{0t} - \psi\hat{R}_{kt})' \tag{19}$$

where ψ is the long run coefficient matrix containing the parameter of interest, γ, and $\hat{R}_{0t}$ and $\hat{R}_{kt}$ are variables constructed by regressing Δy_t and y_{t-k} on k lags of Δy_t respectively. The method obtains the maximum likelihood estimate of θ where $\Psi = \Psi_2\theta'$, where both Ψ_2 and θ are 2x1 matrices so Ψ is reduced rank. In this bivarate model, $\theta' = [\gamma\ -1]$ and θ is identified up to scale. See Johansen (1988) for details.

Appendix 2: Proofs

To show the results of the paper, first three lemmas are included. The proofs of the two theorems are then derived. The derivations of equations (6) and (8) in the no serial correlation case are special cases of the proofs of the theorems. Finally, equation (9) is derived after the proofs of the theorems.

Lemma 1

This lemma restates a number of limit results for the local to unity regression from Bobkoski (1983), Phillips (1987) and Chan and Wei (1987,1988).

For the model in equation (1) and the conditions in the paragraph following this equation, where Conditions A and B1 hold, then the following convergence results hold:

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[T\lambda]} y_{1t} \Rightarrow \Omega_{11}^{\frac{1}{2}} J_c(\lambda) \tag{20}$$

$$\frac{1}{T}\sum_{t=1}^{T} y_{1t}^2 \Rightarrow \Omega_{11}\int_0^1 J_c(\lambda)^2 ds \tag{21}$$

$$\frac{1}{T^{3/2}}\sum_{t=1}^{T} y_{1t} \Rightarrow \Omega_{11}^{\frac{1}{2}}\int_0^1 J_c(\lambda) ds \tag{22}$$

$$\frac{1}{T}\sum_{t=1}^{T} y_{1t}v_{1t} \Rightarrow \Omega_{11}\int_0^1 J_c(\lambda)dW_1(\lambda) + M_1 \tag{23}$$

where

$$J_c(\lambda) = c\int_0^s e^{c(\lambda-s)} W(s)ds + W(\lambda) \tag{24}$$

and $W_1(\lambda)$ is the limit of $T^{-1/2}\Sigma\epsilon_{1t}$ where the summation is from 1 to $[T\lambda]$ and [.] indicates the greatest lesser integer function. This is known as a diffusion or Ornstein Uhlenbeck process. In addition, M_1 is defined as equal to $\Sigma_i^k {}_{=0}E[v_{10}v_{1i}]$.

The result for the demeaned local to unity representation is

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[T\lambda]}(y_{1t} - \bar{y}_{1t}) \Rightarrow \Omega_{11}^{\frac{1}{2}}(J_c(\lambda) - \int_0^1 J_c(\lambda)d\lambda) \equiv \Omega_{11}^{\frac{1}{2}}J_c^{\mu}(\lambda) \tag{25}$$

where $\bar{y}_{1t}$ is the arithmetic mean of y_{1t}. The other expressions can be similarly demeaned, this being indicated by the μ superscript.

Also,

$$\frac{1}{T}\sum_{t=1}^{T} y_{1t}v_{2t} \Rightarrow \Omega_{11}^{\frac{1}{2}}\Omega_{12}^{\frac{1}{2}}\int_0^1 J_c(\lambda)dW_2(\lambda) + M_{12} \tag{26}$$

where $W_1(\lambda)$ is the limit of $T^{-1/2}\Sigma v_{1t}$ where the summation is from 1 to $[T\lambda]$ and M_{12} is the (1,2) element of $\Sigma_i^k {}_{=0}E[v_0v_k]$.

Writing $v^*_{2t} = v_{2t} - \Omega_{12}\Omega^{-1}_{11}v_{1t}$

$$\frac{1}{T}\sum_{t=1}^{T} y_{1t}v^*_{2t} \Rightarrow \Omega_{11}^{\frac{1}{2}}\Omega_{2.1}^{\frac{1}{2}}\int_0^1 J_c(\lambda)dW_{2.1}(\lambda) + M^*_{12} \tag{27}$$

where $M^*_{12} = M_{12} - \Omega_{12}\Omega^{-1}_{11}M_1$, $\Omega_{2.1}=\Omega_{22}-\Omega^2_{12}\Omega^{-1}_{11}$ and $W_{2.1}(\lambda)$ is a standard Brownian Motion which is independent of $W_1(\lambda)$.

Proof.

The results in equations (20) through (23) are proved in Bobkoski (1983) and Cavanagh (1985) when ϵ_{1t} is iid normal and in Phillips (1987) under more general assumptions than stated here (i.e. ϵ_{1t} is assumed to be strong mixing with 2+d moments (for some d>0) in Phillips (1987), either of the assumptions B1 or B2 satisfy strong mixing and the existence of four moments on ϵ_t are assumed here). Chan and Wei (1987) also show the results of equations (20) through (23) when ϵ_{1t} is serially uncorrelated.

The result in (25) follows directly from (20) and (22). Other demeaned representations follow in the same way. The result in (18) follows from Chan and Wei (1988). The result in (27) follows directly from (23) and (26).

Lemma 2

Some further convergence results are derived from the results in lemma 1. These are that for |k-j| ≥0,

$$\frac{1}{T}\sum_{t=1}^{T} y_t v_t \rightarrow P_1 R + e^{*}\Lambda(0) \qquad (28)$$

$$\frac{1}{T}\sum_{t=j}^{T} y_{t-j} v_t \rightarrow P_1(R - \sum_{i=0}^{j-1}\Lambda(i)) + \Lambda(j) \qquad (29)$$

and

$$\frac{1}{T^2}\sum_{t=j}^{T} y_{t-j} y_t \rightarrow P_1 P_1' \Omega_{11} \int J_c(\lambda)^2 ds \qquad (30)$$

where $R = [R_1\ R_2]$ and R_1 and R_2 are the limit results in equation (23) and (26), $P_1' = [1\ \gamma]$, $\Lambda(j)=E[v_0 v_j]$, $e_1' = [1\ 0]$ and e^* is a 2x2 matrix of zeros except for its (2,2) element, which is equal to one.

Proof: The result in equation (28) follows directly by substituting y_{2t} with $\gamma y_{1t}+v_{2t}$, and using the limit results from equations (23) and (26) and the definition of $\Lambda(j)$. The second result uses the first result and that $y_{1t} = \alpha^j y_{1t-j} + \Sigma_0^{j-1}\alpha^i v_{1t-i}$ to substitute for y_{1t-j}. The result follows from algebra, the limit results in equations (23) and (26) and the definition of $\Lambda(j)$. The final result uses the equation for y_{2t} to substitute for this variable, the substitution in the second result to obtain the expression $T^{-2}\Sigma y_{1t-1}^2$, and also the result in equation (29) to show that cross products are $o_p(1)$.

These two lemmas will be used repeatedly in the proofs that follow.

Lemma 3

For the model in equation (1) and Conditions A and B1, $\hat{\Omega}-\Omega = o_p(1)$, where $\hat{\Omega} = 2\pi\hat{s}_v(0)$, where $\hat{v}_t = [\Delta y_{1t}'\ (y_{2t}-\hat{\gamma}y_{1t})']'$, $\hat{s}_v(0)$ is an estimate of the spectral density of the residuals at frequency zero, and $\hat{\gamma}$ is estimated so that $T(\hat{\gamma}-\gamma)$ is $O_p(1)$.

Proof.

First, notice that $\hat{v}_{1t}=v_{1t} + (\alpha-1)y_{1t-1} = v_{1t} + cT^{-1}y_{1t-1}$. Thus the order of the second term is $\sqrt{T}$ less than the order of v_{1t}. Also, $\hat{v}_{2t}=v_{2t} + (\hat{\gamma}-\gamma)y_{2t-1} = v_{2t} + T(\hat{\gamma}-\gamma)T^{-1}y_{2t-1}$, so again the order of the second term is $\sqrt{T}$ less than the order of ϵ_{2t}. This gives the result that $T^{-1}\Sigma\hat{v}_t\hat{v}_{t-j} = T^{-1}\Sigma v_t v_{t-j} + o_p(\sqrt{T})$ for any j. This can be seen by expanding the expression for $T^{-1}\Sigma\hat{v}_t\hat{v}_{t-j}$ using the expressions above and obtaining, in addition to the term $T^{-1}\Sigma v_t v_{t-j}$, typical terms involving $O_p(1)$ terms multiplied by terms such as $T^{-2}\Sigma v_t y_{t-j-1}'$ and $T^{-3}\Sigma y_{t-1}y_{t-j-1}'$, but following lemma 2 these terms are $o_p(\sqrt{T})$.

To estimate $\hat{\Omega}$, note that $\Omega=\Phi(1)^{-1}\Sigma\Phi(1)^{-1}$, so we require consistent estimates of $\Phi(1)$ and Σ. This can be obtained using the finite VAR regressing $\hat{v}_t$ on k lags, i.e.

$$\begin{aligned}\hat{v}_t &= A(L)\hat{v}_{t-1} + \epsilon_t^s \\ &= Ae_t + \epsilon_t^s\end{aligned} \tag{31}$$

where $e_t = [\hat{v}_{t-1}' \ \ldots \ \hat{v}_{t-k}']'$ and A is the corresponding coefficient matrix. From this regression, we have that $\sqrt{T}(\hat{A}-A) = (T^{-1/2}\Sigma v_t^s e_t')(T^{-1}\Sigma e_t e_t')^{-1}$. The denominator has typical terms equal to

$T^{-1}\Sigma\hat{v}_t\hat{v}_{t-j}$, which were argued above to converge to the same limit as $T^{-1}\Sigma v_t v_{t-j}$ and so the denominator converges to a matrix of variances and covariances. In the numerator, as v_t^s is of the same order as v_t, then $(T^{-1/2}\Sigma\epsilon_t^s\hat{v}_{t-j}'-T^{-1/2}\Sigma\epsilon_t^s v_{t-j}')$ is $o_p(1)$. As ϵ_t^s is a martingale difference sequence, we have that $\sqrt{T}(\hat{A}-A)$ converges to a multivariate normal distribution. Thus $\hat{A} \xrightarrow{p} A$ and $\hat{A}(1) \xrightarrow{p} A(1)$. This results in an estimate $\hat{\Phi}(1)$, which equals I_2-$\hat{A}(1)$, converging to $\Phi(1)$.

We can estimate $\hat{\Sigma} = T^{-1}\Sigma\hat{\epsilon}_t^s\hat{\epsilon}_t^{s\prime}$. The result that $T^{-1}\Sigma\hat{\epsilon}_t^s\hat{\epsilon}_t^{s\prime}-T^{-1}\Sigma\epsilon_t\epsilon_t'$ is $o_p(1)$ follows from the result that $(\hat{A}-A)$ is $o_p(1)$ and the results above on the orders of the estimation errors in $\hat{v}_t$

□

This result is employed twice in the paper. Firstly, this result is used in the proof of theorem 1(a) below. Secondly, noting that $T(\hat{\alpha}^m-1)$ is $0_p(1)$, where $\hat{\alpha}^m$ is the median unbiased estimate of α (see Stock (1991)), then a similar argument to the results above justifies estimates of δ employed in section 5 if we assume that the lag length is finite and less than that employed in each estimation.

The model considered is that of equation (1) where the means are removed by demeaning y_t beforehand. In the following, y_t denotes the demeaned data (the μ superscript is dropped from y_t but not the Brownian motion functionals for notational convenience).

Proof of Theorem 1.

This result will be shown for each of the statistics considered in turn.

a) First, consider the OLS estimator of γ of Phillips and Hansen. Substituting equation (1) into equation (15), we obtain

$$T(\hat{\gamma}^{++} - \gamma) = [\frac{1}{T}\sum_{1}^{T} \hat{v}_{2t}^{**} y_{1t} - \hat{M}_{12}^{*}][\frac{1}{T^2}\sum_{1}^{T} y_{1t}^2]^{-1} \tag{32}$$

Noting that

$$\begin{aligned} \hat{v}_{2t}^{**} &= v_{2t} - \Omega_{12}\Omega_{11}^{-1}\Delta y_{1t} + (\Omega_{12}\Omega_{11}^{-1} - \hat{\Omega}_{12}\hat{\Omega}_{11}^{-1})\Delta y_{1t} \\ &= v_{2t} - \Omega_{12}\Omega_{11}^{-1}v_{1t} - (\alpha-1)\Omega_{12}\Omega_{11}^{-1}y_{1t-1} + (\Omega_{12}\Omega_{11}^{-1} - \hat{\Omega}_{12}\hat{\Omega}_{11}^{-1})\Delta y_{1t} \\ &= v_{2t}^{*} - (\alpha-1)\Omega_{12}\Omega_{11}^{-1}y_{1t-1} + (\Omega_{12}\Omega_{11}^{-1} - \hat{\Omega}_{12}\hat{\Omega}_{11}^{-1})\Delta y_{1t} \end{aligned} \tag{33}$$

then equation (32) can be written

$$\begin{aligned} T(\hat{\gamma}^{**} - \gamma) &= [\frac{1}{T}\sum_{1}^{T} v_{2t}^{*} y_{1t} - M_{12}^{*}][\frac{1}{T^2}\sum_{1}^{T} y_{1t}^{2}]^{-1} \\ &- T(\alpha - 1)\,\Omega_{12}\Omega_{11}^{-1}[\frac{1}{T^2}\sum_{2}^{T} y_{1t-1}y_{1t}][\frac{1}{T^2}\sum_{1}^{T} y_{1t}^{2}]^{-1} \\ &+ (\Omega_{12}\Omega_{11}^{-1} - \hat{\Omega}_{12}\hat{\Omega}_{11}^{-1})[\frac{1}{T}\sum_{2}^{T} \Delta y_{1t} y_{1t}][\frac{1}{T^2}\sum_{1}^{T} y_{1t}^{2}]^{-1} \\ &- (\hat{M}_{12}^{*} - M_{12}^{*})[\frac{1}{T^2}\sum_{1}^{T} y_{1t}^{2}]^{-1} \end{aligned} \tag{34}$$

Note that $T^{-1}\Sigma v^{*}_{2t}y_{1t}$ converges to asymptotically to a mixed normal distribution as v^{*}_{2t} and y_{1t} are asymptotically independent (from lemma 1), and that in the second piece $T^{-2}\Sigma y_{1t-1}y_{1t}$ =

$\alpha T^{-2}\Sigma y^{2}_{1t-1} + T^{-2}\Sigma v_{1t}y_{1t-1} = T^{-2}\Sigma y^{2}_{1t-1} + op(1)$. The last two terms are $o_p(1)$ as $\hat{\Omega}$ is consistently estimable by lemma 3 and

$$\begin{aligned} [\frac{1}{T}\sum \Delta y_{1t} y_{1t}][\frac{1}{T^2}\sum y_{1t}^{2}]^{-1} &= [\frac{1}{T}\sum v_{1t} y_{1t}][\frac{1}{T^2}\sum y_{1t}^{2}] \\ &+ T(\alpha - 1)[\frac{1}{T^2}\sum y_{1t} y_{1t-1}][\frac{1}{T^2}\sum y_{1t}^{2}]^{-1} \\ &= [\frac{1}{T}\sum v_{1t} y_{1t}][\frac{1}{T^2}\sum y_{1t}^{2}]^{-1} \\ &+ c[\frac{1}{T^2}\sum y_{1t-1}^{2}][\frac{1}{T^2}\sum y_{1t}^{2}]^{-1} + o_p(1) \end{aligned} \tag{35}$$

which is bounded, by lemma 1 for the first piece, and the second piece converges to c. Thus the limiting distribution of the estimate is given by the limits of the first two terms in equation (34),

$$T(\hat{\gamma}^{**} - \gamma) \rightarrow \Omega_{11}^{-\frac{1}{2}}\,\Omega_{2.1}^{\frac{1}{2}} \int J_c^{\mu}\, dW_{2.1} \left(\int J_c^{\mu 2}\right)^{-1} + D \qquad \text{where} \quad D = -c\,\Omega_{12}\Omega_{11}^{-1} \tag{36}$$

This is the result stated in the theorem.

b) The DOLS procedure (Phillips and Loretan (1991), Saikkonen (1991) and Stock and Watson (1993)).

To show the result, we can first examine the first examine the correctly specified regression. Following Stock and Watson (1993), we can solve for the efficient estimator for γ in the case where c is known. This is a straightforward extension of their case, where they assume c is known and equal to zero. Assumption A satisfies Stock and Watson's (1993) assumption A, and Condition B2 satisfies their Assumption B.

The errors are given by $v_t=\Phi(L)^{-1}\epsilon_t=\Phi(L)^{-1}\Sigma^{1/2}\Sigma^{-1/2}\epsilon_t=H(L)\xi_t$, where $E[\xi_t\xi_t']=I$. Premultiplying the errors by the lower triangular matrix D(L), where D(L) is in general two sided and D(L) is given by

$$D(L) = \begin{bmatrix} 1 & 0 \\ -d_{21}(L) & 1 \end{bmatrix} \tag{37}$$

where $d_{21}(L)=h_2(L)h_1(L^{-1})'[h_1(L)h_1(L^{-1})']^{-1}$, and $H(L)=[h_1(L)\ h_2(L)]'$ (Stock and Watson (1993)).

Premultipying the model by D(L) results in the equation for y_{2t}

$$y_{2t} = d_{21}(L)(1-\alpha L)y_{1t} + \gamma y_{1t} + \eta_{2t} \tag{38}$$

where by construction the residuals are uncorrelated with all of the right hand side variables of the regression at all leads and lags. The limiting distributions of the statistics in the equation are calculated using methods analogous to Sims, Stock and Watson (1990), as is shown for the $\alpha=1$ case in Stock and Watson (1993).

Writing the equation (38) as $y_{2t}=\delta^{*\prime}Z_t+\eta_t$, where $\delta^*=(d_{21,i}\ \gamma)$ and $Z_t = ((1-\alpha L)y_{1t-i}\ y_{1t})$, for integers i such that $-k\leq i\leq k$, where k is the known upper and finite limits on the order of $d_{21}(L)$, then $\Upsilon(\hat{\delta}^*-\delta^*)=(\Upsilon^{-1}\Sigma Z_tZ_t'\Upsilon^{-1})(\Upsilon^{-1}\Sigma Z_t\eta_t)$ where $\Upsilon=\mathrm{diag}(T^{1/2}I_{k1}\ T)$ and k1 is the number of lags and leads of the quasi differenced term included in the regression. As in the results of Sims, Stock and Watson (1990), the denominator of the expression for $\Upsilon(\hat{\delta}-\delta)$ is asymptotically block diagonal, conformable with the partition for Υ. This can be seen by noting that a typical term outside these blocks is $T^{-3/2}\Sigma y_{1t}(1-\alpha L)y_{1t-j}$, but by the results of lemma 2 this converges to zero. The lag coefficients converge to their true values and are asymptotically mixed normal. Finally, we are left with the expression $T(\hat{\gamma}_c-\gamma)=T(\hat{\delta}^*_3-\delta^*_3)=(T^{-2}\Sigma y_{1t}^2)^{-1}(T^{-1}\Sigma y_{1t}\eta_t)+o_p(1)$, where the c subscript denotes the estimate from the correctly transformed model. As y_{1t} and η_{2t} are orthogonal for all leads and lags, this expression will have an asymptotically mixed normal distribution. From the results of lemma 1, we have

$$T(\hat{\gamma}_c-\gamma) \rightarrow \Omega_{11}^{-\frac{1}{2}}\Omega_{\eta}^{\frac{1}{2}}\int J_c^{\mu}dW_{2.1}^{\mu}\left(\int J_c^{\mu\,2}\right)^{-1} \tag{39}$$

and Ω_η is the long run variance of η_{2t}.

Returning to the estimated equation (16), and writing $\beta'=[d'\ \gamma]$, where d_i are the coefficients of the lag polynomial, and x_t as the regressors in conformable order, then the misspecified model can be written $y_{2t}=\beta'x_t$. The canonical form for this regression can be shown by

$y_{2t}=\beta' x_t=$

$\beta' G^{-1} G x_t = \delta' Z_t$, where Gx_t is given by

$$\begin{array}{cccccccccc|c}
1 & (1-\alpha) & .. & (1-\alpha) & 0 & .. & .. & .. & 0 & 1-\alpha & \Delta y_{1t+k} \\
0 & 1 & (1-\alpha) & (1-\alpha) & 0 & .. & .. & .. & 0 & 1-\alpha & \Delta y_{t+k-1} \\
.. & .. & .. & (1-\alpha) & .. & .. & .. & .. & .. & 1-\alpha & .. \\
0 & .. & 0 & 1 & 0 & .. & .. & .. & 0 & 1-\alpha & \Delta y_{t+1} \\
0 & .. & .. & 0 & \alpha & 0 & .. & .. & 0 & 1-\alpha & \Delta y_{t} \\
0 & .. & .. & 0 & -(1-\alpha) & \alpha & 0 & .. & 0 & 1-\alpha & \Delta y_{t-1} \\
0 & .. & .. & 0 & -(1-\alpha) & -(1-\alpha) & \alpha & 0 & 0 & 1-\alpha & \Delta y_{t-2} \\
0 & .. & .. & 0 & -(1-\alpha) & .. & .. & .. & .. & 1-\alpha & .. \\
0 & .. & .. & 0 & -(1-\alpha) & .. & .. & -(1-\alpha) & \alpha & 1-\alpha & \Delta y_{t-k} \\
0 & .. & .. & 0 & 0 & 0 & 0 & 0 & 0 & 1 & y_t
\end{array} \qquad \textbf{(40)}$$

Upon multiplying this out, we retrieve the canonical regressors in Z_t. Following the results of Sims, Stock and Watson (1990), we can write $\delta^* = (\delta^*_1{}' \ \delta^*_3)$. From the construction of the canonical regression, $\beta = G'\delta^*$. Thus d_i are a linear combination of $d_{21,i}$ and hence are asymptotically normal and converge to their population values at rate $\sqrt{T}$. Further, from this linear relationship we can see that

$$d(1) = \alpha \sum_{i=-k}^{0} d_{21,i} + \sum_{1}^{k} d_{21,i} + (1-\alpha) f[d_{21,i}] \qquad \textbf{(41)}$$

where the function f is linear. Thus, as $T \to \infty$, $\hat{d}(1) \xrightarrow{p} d_{21}(1)$.

Consideration of the final element of β gives the limiting result for the parameter of interest, γ. Using the result that $\beta_3 = \gamma$, then

$$T(\hat{\gamma}_d - \gamma_0) = T(1-\alpha) \sum_{i=-k}^{k} \hat{\delta}_{1,i} + T(\hat{\delta}_3 - \delta_3) \qquad \textbf{(42)}$$

where the d subscript refers to the misspecified estimate.

From the results above for the convergence of δ^*, we have the result

$$T(\hat{\gamma}_d - \gamma_0) \rightarrow \Omega_{11}^{\frac{-1}{2}} \Omega_{2.1}^{\frac{1}{2}} \int J_c^\mu dW_{2.1}^\mu \left(\int J_c^{\mu 2}\right)^{-1} + D \tag{43}$$

where $D=-cd_{21}(1)$. The last two results are to show that $d_{21}(1)=\Omega_{12}\Omega_{11}^{-1}$ and $\Omega_\eta=\Omega_{2.1}$. These are obtained by evaluating the formula for $d_{21}(L)$ given above after equation (37) at L=1. Recall that $\Omega=H(1)H(1)'$. Multiplying out the formula for $d_{21}(1)$ results in $d_{21}(1)=h_2(1)h_1(1)'[h_1(1)h_1(1)']^{-1}=\Omega_{12}\Omega_{11}^{-1}$ which gives the first result. The second result follows from writing $\eta_{2t}=v_{2t}-d_{21}(L)v_{1t}=h_2(L)\xi_t-d_{21}(L)h_1(L)\xi_t=(h_2(L)-d_{21}(L)h_1(L))\xi_t=h^*(L)\xi_t$. Noting that $\Omega_\eta=h^*(1)h^*(1)'$, substituting L=1 and algebra yield $\Omega_\eta=\Omega_{22}-\Omega^2_{12}\Omega^{-1}_{11}=\Omega_{2.1}$.

Note that the removal of the extra assumption that $d_{21}(L)$ has finite order results in extra terms which disappear asymptotically if the order of d(L) in the regression grows at a suitable rate as $T\rightarrow\infty$ (see Saikkonen (1991)).

c) Saikkonen (1992) estimator.

The ECM form in equation (17) is derived from the model in equation (1) by first premultiplying the model in equation (1) by the matrix P, which is given by

$$\begin{bmatrix} 1 & 0 \\ \gamma & 1 \end{bmatrix} \tag{44}$$

This allows us to write the error correction model as $\Delta y_t=My_{t-1}+Pv_t$, where M is given by the lower triangular matrix

$$\begin{bmatrix} (\alpha-1) & 0 \\ \gamma\alpha & -1 \end{bmatrix} \tag{45}$$

As $\Phi(L)v_t=\epsilon_t$, we can write $P\Phi(L)P^{-1}Pv_t=P\epsilon_t$, so the model can be rewritten as

$$\Delta y_t = \Psi y_{t-1} + \Pi x_t + \epsilon_t^* \tag{46}$$

where $\Psi=P\Phi(1)P^{-1}M$ and $\epsilon_t^* = P\epsilon_t$.

Following Saikkonen (1992), let $\underline{\Psi}=\Psi P$, then $\underline{\Psi}=[\underline{\Psi}_1\ \Psi_2]$, and $\underline{\Psi}_1=\Psi_1+\gamma\Psi_2$. The same transformation to the dynamic terms coefficients allows equation (46) to be written

$$\begin{aligned} \Delta y_t &= \Psi PP^{-1}y_{t-1} + \Pi(L)PP^{-1}x_t + \epsilon_t^* \\ &= \underline{\Psi}_1 y_{1t-1} + \Psi_2 v_{2t-1} + \underline{\Pi}_1(L)x_{1t} + \Pi_2(L)x_{2t} + \epsilon_t^* \end{aligned} \tag{47}$$

This final equation yields estimators of the parameters of the model which are infeasible but nonetheless enable calculation of the limiting distributions of the estimators. In the least squares estimation of the final equation in (47), the denominator matrix is block diagonal conformable with the first block being 1x1 in dimension. This can be seen by noting that typical terms off the block diagonal are $T^{-3/2}\Sigma y_{1t-1}v_{2t-1}'$ and $T^{-3/2}\Sigma y_{1t-1}\Delta y_{t-j}'$. The first of these expressions converges to zero by equation (26) from lemma 1. The second term converges to zero in probability, as from above, $\Delta y_t=My_{t-1}+Pv_t$ so $T^{-3/2}\Sigma y_{1t-1}\Delta y_{t-j}' = T^{-3/2}\Sigma y_{1t-1}(My_{t-j-1}+Pv_{t-j})' = T^{-3/2}\Sigma y_{1t-1}y_{t-j-1}'M'+o_p(1) = T^{-3/2}\Sigma y_{1t-1}y_{t-j-1}'\{P^{-1}\}'P'M'+o_p(1) = cT^{-5/2}\Sigma y_{1t-1}y_{1t-j-1}' + c\gamma T^{-5/2}\Sigma y_{1t-1}y_{2t-j-1}' - T^{-3/2}\Sigma y_{1t-1}v_{2t-j-1}$. Each of these three terms converges to zero by lemma 2.

Thus,

$$T(\tilde{\Psi}_1 - \Psi_1) = \frac{\frac{1}{T}\sum \epsilon_t^{*} y_{1t-1}}{\frac{1}{T^2}\sum y_{1t-1}^2} + o_p(1) \tag{48}$$

Using the result that $\epsilon_t^{*} = P\epsilon_t$, and the assumptions on ϵ_t in Condition A give the result that $T(\tilde{\Psi}_1 - \Psi_1) \Rightarrow P\Phi(1)(\int dB_c B_{c1})(\int B_{c1}^2)^{-1}$, where $B_c(r) = \Omega J_c(r)$, and the numeric subscript on B refers to the row number.

The OLS estimates for the remaining coefficients of the model are consistent and converge jointly to their population values at rate $\sqrt{T}$. This gives the result that $\sqrt{T}(\tilde{\Psi}_2 - \Psi_2)$ is $0_p(1)$. The remaining piece needed is that the variance covariance estimator of the estimated residuals, estimated by $T^{-1}\Sigma\tilde{\epsilon}_t^{*}\tilde{\epsilon}_t^{*\prime}$, converges to its population value of $P\Sigma P'$. This follows as all parameter estimates of the regression converge to their population values, so $T^{-1}\Sigma\tilde{\epsilon}_t^{*}\tilde{\epsilon}_t^{*\prime} = T^{-1}\Sigma\epsilon_t^{*}\epsilon_t^{*\prime} + o_p(1) = P(T^{-1}\Sigma\epsilon_t\epsilon_t)P' + o_p(1) \xrightarrow{p} P\Sigma P'$.

The estimator in equation (18) can be written as

$$\begin{aligned} T(\tilde{\gamma} - \gamma) &= -(\tilde{\Psi}_2'\tilde{\Sigma}^{*-1}\tilde{\Psi}_2)^{-1}(\tilde{\Psi}_2'\tilde{\Sigma}^{*-1})T\tilde{\Psi}_1 \\ &= -(\tilde{\Psi}_2'\tilde{\Sigma}^{*-1}\tilde{\Psi}_2)^{-1}(\tilde{\Psi}_2'\tilde{\Sigma}^{*-1})\,T(\tilde{\Psi}_1 - \Psi_1) \\ &\quad - (\tilde{\Psi}_2'\tilde{\Sigma}^{*-1}\tilde{\Psi}_2)^{-1}(\tilde{\Psi}_2'\tilde{\Sigma}^{*-1})T\Psi_1 \\ &= -(\Psi_2'\Sigma^{*-1}\Psi_2)^{-1}(\Psi_2'\Sigma^{*-1}T(\tilde{\Psi}_1 - \Psi_1) \\ &\quad -(\Psi_2'\Sigma^{*-1}\Psi_2)^{-1}(\Psi_2'\Sigma^{*-1}T\Psi_1 + o_p(1) \end{aligned} \tag{49}$$

where in the last expression Ψ_2 and Σ^{*-1} replace their estimated values as these estimates converge to their true population values from above. Noting that $\Psi_2 = \Phi(1)P^{-1}M^{*}e_2$, where

$e_2'=[0\ \ 1]$, $(\Psi'\Sigma^{*-1}\Psi)^{-1}$ is $(e_2'M'P^{-1\prime}\Phi(1)'P'\{P\Sigma P'\}^{-1}P\Phi(1)P^{-1}Me_2)^{-1}=(e_2'M'P^{-1\prime}\Omega^{-1}P^{-1}Me_2)^{-1}+o_p(1)=(e_2'\Omega^{-1}e_2)^{-1}=\Omega_{2.1}^{-1}$ (where we have used the results $\Omega^{-1}=\Phi(1)'\Sigma^{-1}\Phi(1)$ and $P^{-1}Me_2=-e_2$). Similar algebra yields $(\Psi'\Sigma^{*-1}) = -e_2\Phi(1)'\Sigma^{-1}P^{-1}$ and $\underline{\Psi}_1 = P\Phi(1)[(\alpha-1)\ 0]'$. Thus, by straightforward algebra and $c=T(\alpha-1)$ the term $(\Psi'\Sigma^{*-1}\Psi)^{-1}(\Psi'\Sigma^{*-1})T\underline{\Psi}_1 = \Omega_{12}\Omega_{11}^{-1}c$. Substituting the results from this paragraph and the limit distribution for $T(\Psi_1-\underline{\Psi}_1)$ into the last expression in equation (49) yields

$$T(\tilde{\gamma}-\gamma) \rightarrow \int dB_{c2.1}B_{c1}\left(\int B_{c1}^2\right)^{-1} - \Omega_{12}\Omega_{11}^{-1}c \tag{50}$$

which is equivalent to the expression given in the theorem.

d) Johansen (1988) model.

The Johansen (1988) estimator computes the exact FIML estimator, which is identical to the iterated three stage least squares estimator (Theil (1971)). It thus suffices to show that the iterated three stage least squares estimator has the distribution given in the theorem.

The error correction model representation of equation (1) was given above to be equal to $\Delta y_t = \Psi y_{t-1} + \pi x_t + \epsilon_t^*$. This can be rewritten as

$$\begin{aligned}\Delta y_t &= \Psi y_{t-1} + \Psi y_{t-k} - \Psi y_{t-k} + \pi x_t + \epsilon_t^* \\ &= \Psi y_{t-k} + \Psi(y_{t-1} - y_{t-2} + y_{t-2} - \ldots - y_{t-k}) + \pi x_t + \epsilon_t^* \\ &= \Psi y_{t-k} + \pi^* x_t + \epsilon_t^*\end{aligned} \tag{52}$$

which is in the form of Johansen (1988). Comparing this with equation (46) above we have that $\Psi=P\Phi(1)P^{-1}M$, $\epsilon_t^*=P\epsilon_t$ as in (c) above. The only difference between the models is the coefficients on x_t and that x_t is redefined as $[\Delta y'_{t-1}, \ldots, \Delta y'_{t-k}]'$.

The iterated three stage least squares estimator can be computed as follows. First, obtain the consistent estimate of $\tilde{\gamma}$ from this model using the normalization employed in (c). This can then be used to generate estimates of the remaining parameters from the now linear least squares model. The third stage employs the normal equation for γ from the likelihood to obtain a new estimate of $\tilde{\gamma}$, denoted $\tilde{\gamma}_2$. This procedure can be iterated until the parameter estimates converge, when the estimate for γ, given by $\tilde{\gamma}_k$, is numerically equivalent to the MLE for γ. The limit distribution of $\tilde{\gamma}_k$ is equal to that of the theorem.

To obtain the limiting representation of the first stage estimate of γ, we follow exactly the same steps as in (c) above. The model in equation (52) can be rewritten as

$$\begin{aligned} \Delta y_t &= \Psi P P^{-1} y_{t-k} + \pi^*(L) x_t + \epsilon_t^* \\ &= \Psi_1 y_{1t-k} + \Psi_2 \epsilon_{2t-k} + \pi^*(L) x_t + \epsilon_t^* \end{aligned} \tag{53}$$

In the least squares estimation of the final equation in (53), the denominator matrix is again block diagonal conformable with the first block being 1x1 in dimension. The typical terms off the block diagonal now are $T^{-3/2}\Sigma y_{1t-k} v_{2t-1}'$ and $T^{-3/2}\Sigma y_{1t-k}\Delta y_{t-j}'$. The first of these expressions still converges to zero (relate to a result above). The second term also converges to zero in probability, as from above, $\Delta y_t = My_{t-1} + Pv_t$ so $T^{-3/2}\Sigma y_{1t-k}\Delta y_{t-j}' = T^{-3/2}\Sigma y_{1t-k}(My_{t-j-1} + Pv_{t-j})' =$
$T^{-3/2}\Sigma y_{1t-k} y_{t-j-1}'M' + o_p(1) = T^{-3/2}\Sigma y_{1t-k} y_{t-j-1}'\{P^{-1}\}'P'M' + o_p(1) = cT^{-5/2}\Sigma y_{1t-k} y_{1t-j-1}' + c\gamma T^{-5/2}\Sigma y_{1t-k} y_{2t-j-1}' - T^{-3/2}\Sigma y_{1t-k} v_{2t-j-1}$. Again, these terms are $o_p(1)$ by lemma 2.

Thus, we have that

$$T(\tilde{\Psi}_1 - \Psi_1) = \frac{\frac{1}{T}\sum \epsilon_t^* y_{1t-k}}{\frac{1}{T^2}\sum y_{1t-k}^2} + o_p(1) \tag{54}$$

As $T^{-1}\Sigma\epsilon_t y_{1t-1} = T^{-1}\Sigma\epsilon_t y_{1t-1} + o_p(1)$, this converges to $P\Phi(1)(\int dB_c B_{c1})(\int B_{c1}^2)^{-1}$ as in (c).

Also for the same reasons as in (c), both Ψ_2 and $\tilde{\pi}^*$ converge at rate $\sqrt{T}$ to joint normal distributions, and $T^{-1}\Sigma\tilde{\epsilon}_t^*\tilde{\epsilon}_t^{*\prime}$ converges in probability to $P\Sigma P'$. The estimate for $\tilde{\gamma}_1 = -(\Psi_2'\Sigma^{*-1}\Psi_2)^{-1}(\Psi_2'\Sigma^{*-1}\Psi_1)$, which following the steps in (c) using the results here converges to γ (with an $o_p(1)$ bias).

The second stage estimates involve the estimation of the model where ψ is forced to be of reduced rank. The likelihood in this case is proportional to

$$-\frac{T}{2}\ln|\Sigma^*| - \frac{1}{2}\sum(\Delta y_t - \Psi_2\theta' y_{t-k} - \pi x_t)'\Sigma^{*-1}(\Delta y_t - \Psi_2\theta' y_{t-k} - \pi\, x_t) \tag{55}$$

and the second stage equation that is estimated is

$$\Delta y_t = \psi_2\hat{v}_{2t-k} + \pi^* x_t + \epsilon_t^{**} \tag{56}$$

where $\hat{v}_{t-k} = \theta' y_{t-k}$, $\theta' = [1\ -\tilde{\gamma}_1]$, and $\epsilon_t^{**} = \epsilon_t^* + \psi_1 y_{t-k} = \epsilon_t^* + \Phi_1(1)(\alpha-1)y_{1t-k}$ where $\Phi_1(1)$ is the first row of $\Phi(1)$. We can write $X_t = [v_{2t-k}'\ x_t']'$ and Γ as the corresponding coefficient matrix. This enables writing the second stage regression as $\Delta y_t = \Gamma X_t + \epsilon_t^{**}$, so that the centered and scaled estimates of the parameters are given by $\sqrt{T}(\hat{\Gamma}-\Gamma) = (T^{-1/2}\Sigma\epsilon_t^{**}X_t')(T^-$

${}^{1}\Sigma X_tX_t')^{-1}$. The denominator converges in probability to a matrix M_n. This can be shown by noting that there are three types of terms here; $T^{-1}\Sigma x_tx_t'$, $T^{-1}\Sigma\hat{v}_{t-k}\hat{v}_{t-k}'$, and $T^{-1}\Sigma\hat{v}_{t-k}x_t'$. That $T^{-1}\Sigma x_tx_t'$ is $O_p(1)$ follows from noting that a typical element is $T^{-1}\Sigma\Delta y_{t-i}\Delta y_{t-j}'$, and from the model $\Delta y_t = M^* y_t - (\alpha-1)y_{t-1}^* + P\epsilon_t$, where $y_{t-1}^* = \mathrm{diag}(y_{t-1})$. Noting that $M^* y_t = -\alpha e^* v_t$, where e^* is a null matrix with its (2,2) element equal to 1, we have by direct calculation,

$$\frac{1}{T}\sum \Delta y_{t-i}\Delta y_{t-j}' = \frac{1}{T}\sum(-\alpha e^* v_{t-i-1} - T(\alpha-1)y_{t-i-1}^* + Pv_{t-i})(-\alpha e^* v_{t-j-1} - T(\alpha-1)y_{t-j-1} + Pv_{t-j})'$$
$$= \frac{1}{T}\sum(e^* v_{t-i-1}v_{t-j-1}'e^* + Pv_{t-i}v_{t-j}'P' - J^* v_{t-i-1}v_{t-j}'P' - Pv_{t-i}v_{t-j-1}'e^*) + o_p(1) \tag{57}$$

where the terms involving $T(\alpha-1)$ are $o_p(1)$ using results in lemma 2. Each of the remaining terms converge to variances or covariances of ϵ_t, so $T^{-1}\Sigma x_tx_t'$ is $O_p(1)$ and x_t is of order $\sqrt{T}$.

Convergence of $T^{-1}\Sigma\hat{v}_{2t-k}\hat{v}_{2t-k}'$ follows as $T^{-1}\Sigma\hat{v}_{2t-k}\hat{v}_{2t-k}' = T^{-1}\Sigma v_{2t-k}v_{2t-k}' + T^{-1}\Sigma v_{2t-k}y_{t-1}'(\theta'-\theta') + (\theta'-\theta')T^{-1}\Sigma y_{t-k}v_{2t-k}' + (\theta'-\theta')T^{-1}\Sigma y_{t-k}y_{t-k}'(\theta'-\theta')'$. As $T(\theta'-\theta')$ is $O_p(1)$, the final three terms are $o_p(1)$ by lemma 2 so this term converges to Σ_{22}. A typical cross product term is $T^{-1}\Sigma\hat{v}_{t-}$
${}_{k}x_{t-j}' =$
$T^{-1}\Sigma v_{t-k}x_{t-j}' + T(\theta'-\theta')T^{-2}\Sigma y_{t-k}x_{t-j}' = T^{-1}\Sigma v_{t-k}x_{t-j}' + o_p(1)$ as x_t is of order $\sqrt{T}$ from above. As this term is bounded, we have the result for the denominator.

In the numerator, $T^{-1/2}\Sigma\epsilon_t^{**}X_t' = T^{-1/2}\Sigma\epsilon_t^* X_t + \Phi_1(1)T(\alpha-1)T^{-3/2}\Sigma y_{1t-k}X_t' = T^{-1/2}\Sigma\epsilon_t^* X_t' + o_p(1)$. That the final term is $o_p(1)$ can be seen by noting looking at typical terms in X_t. These are v_{2t-k} and Δy_{t-j}. From lemma 2, $T^{-3/2}\Sigma y_{1t-k}v_{2t-k}'$ is $o_p(1)$. Also we have that $T^{-3/2}\Sigma y_{1t-k}\Delta y_{t-j}'$ is $o_p(1)$ by algebra analogous to that above equation (48) and the results of lemma 2. As ϵ_t^*

is a martingale difference sequence, we have that $T^{-1/2}\Sigma\epsilon_t^* v_{2t-k}$ converges to a normal distribution. For the terms $T^{-1/2}\Sigma\epsilon_t^*\Delta y_{t-j}'$, note that following the algebra preceding equation (48) and lemma 2 we have that $T^{1/2}\Sigma\epsilon_t^*\Delta y_{t-j}' = T^{-1/2}\Sigma\epsilon_t^* v_{t-j}'P'$ which converges to a multivariate normal distribution. Thus, we have the result that the second stage estimates $\sqrt{T}(\hat{\Gamma}-\Gamma)$ converges jointly to a normal distribution. For iterated least squares, we require that they are consistent.

In the third stage, the first order condition for γ from the likelihood when reduced rank is imposed is employed to generate the estimate $\tilde{\gamma}_2$. This is given by

$$\Psi_2'\Sigma^{*-1}\sum(\Delta y_t - \Psi_2\theta' y_{t-k} - \pi x_t)' y_{1t-k} = 0 \tag{58}$$

This can be solved for $\tilde{\gamma}_2$ using the second stage estimates derived above. Using the expression in equation (53) to substitute for Δy_t in the third stage estimator results in an expression for $T(\tilde{\gamma}_2-\gamma)$

$$\begin{aligned} T(\tilde{\gamma}_2 - \gamma) = &-(\tilde{\Psi}_2'\tilde{\Sigma}^{*-1}\tilde{\Psi}_2)^{-1}(\tilde{\Psi}_2'\tilde{\Sigma}^{*-1})\frac{1}{T}\sum\epsilon_t^* y_{1t-k}(\frac{1}{T^2}\sum y_{1t-k}^2)^{-1} \\ &-(\tilde{\Psi}_2'\tilde{\Sigma}^{*-1}\tilde{\Psi}_2)^{-1}(\tilde{\Psi}_2'\tilde{\Sigma}^{*-1}\Psi_1) \\ &+(\tilde{\Psi}_2'\tilde{\Sigma}^{*-1}\tilde{\Psi}_2)^{-1}(\tilde{\Psi}_2'\tilde{\Sigma}^{*-1}\sqrt{T}(\tilde{\pi}-\pi)\frac{1}{T^{3/2}}\sum x_t y_{1t-k}(\frac{1}{T^2}\sum y_{1t-k}^2)^{-1} \\ &+(\tilde{\Psi}_2'\tilde{\Sigma}^{*-1}\tilde{\Psi}_2)^{-1}(\tilde{\Psi}_2'\tilde{\Sigma}^{*-1}\sqrt{T}(\tilde{\Psi}_2-\Psi_2)\frac{1}{T^{3/2}}\sum v_{2t-k}y_{1t-k}(\frac{1}{T^2}\sum y_{1t-k}^2)^{-1} \end{aligned} \tag{59}$$

As in (c) above, the consistency of Ψ and Σ^* means that these terms can be replaced by their population values adding an $o_p(1)$ error to the equation. The final two terms in the expression in equation (59) are $o_p(1)$ as each of the pieces is $O_p(1)$ except that $T^{-3/2}\Sigma x_t y_{1t-k}$

and $T^{-3/2}\Sigma v_{2t-k}y_{1t-k}$ are $o_p(1)$ from results in the showing of consistency in the second stage above. The remaining two terms are identical to those of equation (49) in (c) except that y_{1t-1} is replaced by y_{1t-k}. By lemma 2 and that ϵ_t^* is a martingale difference sequence, we have that $T(\tilde{\gamma}_2-\gamma)$ converges to the distribution in the theorem. Note that k-1 successive iterations over stages 2 and 3 yield the estimator $\tilde{\gamma}_k$, which is numerically equivalent to the maximum likelihood estimator and has the distribution of the theorem. Note also that the MLE for $\Omega_{2.1}$ also converges to its true value.

Proof of Theorem 2

a) For the Phillips Hansen estimator, the t statistic testing the true null hypothesis is given by

$$t_{(\hat{\gamma}^{**}-\gamma)} = \frac{T(\hat{\gamma}^{**}-\gamma)}{\hat{\Omega}_{2.1}^{\frac{1}{2}}(\frac{1}{T^2}\sum y_{1t}^2)^{-\frac{1}{2}}} \tag{60}$$

where $\hat{\Omega}_{2.1}$ is a consistent estimator of $\Omega_{2.1}$. By the results of the previous theorem for the numerator in equation (36) and from lemma 1 for the summation in the denominator, then we have the result that

$$t_{(\gamma^{**}-\gamma)} \Rightarrow N(0,1) - D\Omega_{2.1}^{-\frac{1}{2}}\Omega_{11}^{\frac{1}{2}}(\int J_c^{\mu 2})^{\frac{1}{2}} \tag{61}$$

The final term can be shown to be equal to that in equation (11) by dividing the numerator and denominator by $\Omega^{1/2}_{22}$.

b) The same result can be shown in the DOLS case by writing the t statistic testing the true null hypothesis as

$$t_{(\hat{\gamma}_d-\gamma)} = \frac{T(\hat{\gamma}_d-\gamma)}{\hat{\Omega}_\eta^{\frac{1}{2}}(\frac{1}{T^2}\sum y_{1t}^2)^{-\frac{1}{2}}} + op(1) \tag{62}$$

This equation follows from the asymptotic block diagonality of the variance covariance matrix of the regressors. As explained in Stock and Watson (1993), the residuals of the estimated equation (16) (in their case with $\alpha=1$) are serially correlated in the general case, so a consistent estimator of the variance must be employed. In the previous proof it was shown that the misspecified model is simply the correct model reparameterized, so the residuals from the correct and the incorrect specifications are identical. Thus, is a consistent estimator of the residuals for the serially correlated case is employed (i.e. as in Andrews (1991), or an AR estimator as in Stock and Watson (1993)), then $\hat{\Omega}_\eta \rightarrow \Omega_\eta$, which was shown at the end of the proof for theorem 1 to be equal to Ω_{21}. The result follows from using the results of equation (39) for the numerator and the results in the proof for the t statistic in a) above.

c) For the Saikonnen (1992) procedure, employing $\hat{V}_{22}^{-1}$ as the variance estimator in the construction of the t statistic, the result follows directly.

d) Again, the result follows from the definition of the t statistic and the results of Theorem 1, parts (c) and (d).

ADDENDUM - Proof of the results in equations (6), (8) and (9) in the text. The proofs for equations (6) and (8) follow directly from the DOLS cases in the theorem proofs, as they are identical except that there is no serial correlation. In this case, $\Sigma=\Omega$, giving these equations. The result in equation (9) is derived from equation (8). Note that when it is mistakenly believed that $\alpha=1$, then the t statistic for $\hat{\gamma}$ will be compared to the normal distribution, and the null hypothesis rejected if the value for statistic in absolute value is greater than z^*, the normal critical value. Asymptotically, we can compute the expected number of rejections by substituting the asymptotic distribution for the t statistic under the misspecified case, i.e. $N(0,1)+D_2$ from equation (8), and reject the null hypothesis for $E[|N(0,1)+D_2|>z^*]$. Considering the lower tail, this gives the expected number of rejections as $E[\ N(0,1)+D_2<-z^*]$. This can be rewritten as $E[N(0,1)<-z^*-D_2]=E[\Phi(-z^*-D_2)]$, where $\Phi[z]$ is the normal cdf evaluated at z. A similar calculation yields the expected upper tail rejections, and expected size is given by the sum of the expected rejections in each tail.

Appendix 3: Data Sources

This appendix describes the data used in section 5 and its sources.

Consumption Example: All data is quarterly data from the first quarter of 1948 to the first quarter of 1992, and is from the citibase dataset. Income is GDP in 1987 dollars (citibase mnenomic GDPQ). Total consumption is total personal consumption expenditures in 1987 dollars (GCQ), non durables consumption is also in 1987 dollars (GCNQ) and services consumption is total personal consumption expenditures on services in 1987 dollars (GCSQ).

Exchange Rate Example: The exchange rate data is monthly (end month) data on the spot YEN/US dollar rate and the one month forward rate for this currency. The data is from McCallum (1992), Appendix A and runs from September 1977 to July 1990.

Dividend Yield Example: The data is annual observations for ex post real returns and the dividend price ratio from the S&P index over the period 1871-1985. The variables were constructed from dividend and price data from Shiller (1981) as employed in DeJong, Nankervis, Savin and Whiteman (1992).

Bibliography

Ahn, S.K and G.C.Reinsel (1990), "Estimation for Partially Nonstationary Autoregressive Models", *Journal of the American Statistical Association*, 85, pp813-823.

Andrews, D.W.K (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation", *Econometrica*, 59, 817-858.

Andrews, D.W.K and J.C.Monahan (1992) "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator", *Econometrica*, 60, pp953-66.

Baille, R.T and P.C. McMahon (1989), The Foreign Exchange Market: Theory and Econometric Evidence, Cambridge University Press: Cambridge.

Bobkoski, M.J. (1983), Hypothesis Testing in Nonstationary Time Series, unpublished Ph.D. thesis, Department of Statistics, University if Wisconsin.

Cavanagh, C. (1985), "Roots Local to Unity", manuscript, Department of Economics, Harvard University.

Cavanagh, C.L., G. Elliott and J.H.Stock (1993), "Inference in Models with Nearly Nonstationary Regressors", manuscript, Department of Economics, Harvard University.

Chan, N.H. (1988), "On Parameter Inference for Nearly Nonstationary Time Series", *Journal of the American Statistical Association*, 83, pp 857-62.

Chan, N.H. and C.Z.Wei (1987), "Asymptotic Inference for Nearly Nonstationary AR(1) Processes", *Annals of Statistics*, 15, pp1050-63.

Chan, N.H. and C.Z.Wei (1988), "Limiting Distributions of Least Squares Estimates of Unstable Autoregressive Processes", *Annals of Statistics*, 16, pp 367-401.

DeJong, D.N., J.C.Nankervis, N.E.Savin and C.H.Whiteman (1992) "The Power Problems of Unit Root Tests for Time Series with Autoregressive Errors", *Journal of Econometrics*, 53, pp323-43.

DeJong, D.N., C.H.Whiteman (1991), "Reconsidering Trends and Random Walks in Macroeconomic Time Series", *Journal of Monetary Economics*, 28, pp 221-254.

Dickey, D.A. and W.A.Fuller (1979), "Distribution of Estimators for Autoregressive Time Series with a Unit Root", *Journal of the American Statistical Association*, 74, pp427-431.

Elliott, G., T.J.Rothenberg and J.H.Stock (1992), Efficient Tests for an Autoregressive Unit Root", NBER Technical Working Paper No. 130.

Elliott, G. and J.H.Stock (1992), "Inference in Time Series Regression When the Order of Integration of a Regressor is Unknown", NBER Technical Working Paper No. 122.

Engle, R.F. and C.W.Granger (1987), "Cointegration and Error Correction: Representation, Estimation and Testing", *Econometrica*, 55, 251-76.

Engle, R.F. and B.S.Yoo (1993), "Cointegrated Economic Time Series: An Overview with New Results", in R.F.Engle and C.W.J.Granger eds. Long-Run Economic Relations: Readings in Cointegration, Oxford University Press: New York.

Evans, M.D.D. and K.K.Lewis (1993), "Do Long Term Swings in the Dollar Affect Estimates of the Risk Premia?", Stern School of Business Discussion Paper EC-93-12, New York University, June.

Hall A.D., H.M.Anderson and C.W.J.Granger (1992), "A Cointegration Analysis of Treasury Bill Yields", *Review of Economics and Statistics*, 74, pp116-26.

Hansen, B.E. and P.C.B. Phillips (1990), "Estimation and Inference in Models of Cointegration: A Simulation Study", *Advances in Econometrics*, 8, p225-48.

Hoffman, D.L. and R.H.Raasche (1991), "Long-run Income and Interest Elasticities of Money Demand in the United States", *Review of Economics and Statistics*, 73, pp665-74.

Johansen, S. (1988), "Statistical Analysis of Cointegrating Vectors", *Journal of Economic Dynamics and Control*, 12, pp 231-54.

Johansen, S. and K. Juselius (1992), "Testing Structural Hypotheses in a Multivariate Cointegration Analysis of the PPP and UIP of UK", *Journal of Econometrics*, 53, pp 211-44.

Johnson, D. (1993) "Unit Roots, Cointegration and Purchasing Power Parity: Canada and the United States, 1870-1991", in B.O'Reilly and J.Murray (eds) The Exchange Rate and the Economy, forthcoming, Ottowa: Bank of Canada.

Kitamura, Y and P.C.B.Phillips (1992). "Fully Modified IV, GIVE, and GMM Estimation with Possibly Nonstationary Regressors and Instruments", manuscript, Yale University.

Kugler, P. and C.Lenz (1993), "Multivariate Cointegration Analysis and the Long Run Validity of PPP", *Review of Economic and Statistics*, 75, pp 180-184.

Ogaki, M. (1992), "Engels Law and Cointegration", *Journal of Political Economy*, 100, pp1027-46.

McCallum, B. (1992) "A Reconsideration of the Interest Parity Relationship" NBER Discussion Paper No. 4113, July.

Nelson, C.R. and C.I.Plosser (1982), "Trends and random Walks in Macroeconomic Time Series; Some evidence and Implications", *Journal of Monetary Economics*, 10, 139-62.

Park, J.Y. (1992), "Canonical Cointegrating Regressions", *Econometrica*, 60, pp 119-144.

Phillips, P.C.B. (1986), "Towards a Unified Asymptotic Theory for Autoregression", *Biometrica*, 74, 535-47.

Phillips P.C.B. (1991), "Optimal Inference in Cointegrated Systems", *Econometrica*, 59, pp 283-306.

Phillips, P.C.B. (1993a), "Fully Modified Least Squares and Vector Autoregression", manuscript, Yale University.

Phillips, P.C.B. (1993b), "Robust Nonstationary Regression", manuscript, Yale University.

Phillips, P.C.B. and B.E.Hansen (1990), "Statistical Inference in Instrumental Variables Regression with I(1) processes", *Review of Economic Studies*, 57, pp99-125.

Phillips, P.C.B. and M.Loretan (1991), "Estimating Long-Run Economic Equilibria", *Review of Economic Studies*, 59, pp407-36.

Phillips, P.C.B. and J.Y.Park (1988), "Statistical Inference in Regressions with Integrated Processes: Part 1", *Econometric Theory*, 4, 468-497.

Saikkonen, P (1991), "Asymptotically Efficient Estimation of Cointegration Regressions", *Econometric Theory*, 7, 1-21.

Saikkonen, P (1992), "Estimation and Testing of Cointegrated Systems by an Autoregressive Approximation", *Econometric Theory*, 8, pp 1-27.

Shiller, R.J. (1981), "Do Stock Prices Move too Much to be Justified by Subsequent Changes in Dividends", *American Economic Review*, 71, pp 421-36.

Sims, C.A., J.H.Stock and M.W.Watson (1990), "Inference in Linear Time Series Models with Some Unit Roots", *Econometrica*, 58, pp113-44.

Stock, J.H. (1987), "Asymptotic Properties of Least Squares Estimators of Cointegrating Vectors", *Econometrica*, 55, pp 1035-56.

Stock, J.H. (1991), "Confidence Intervals for the Largest Autoregressive root in U.S. Macroeconomic Time Series", *Journal of Monetary Economics*, 28, 435-459.

Stock, J.H. (1993), "Unit Roots, Structural Breaks, and Trends", in McFadden and R.F.Engle eds. Handbook of Econometrics V4, forthcoming.

Stock, J.H. and M.W.Watson (1993), "A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems", *Econometrica*, 61, pp 783-820.

Theil, H (1971), Principles of Econometrics, John Wiley and Sons: New York.

Watson, M.W. (1993), "Vector Autoregressions and Cointegration", in McFadden and R.F.Engle eds. Handbook of Econometrics V4, forthcoming.

Chapter 4: Unbiasedness and Orthogonality Tests in the Forward Exchange Rate Market

I. Introduction

The hypothesis of efficiency in the forward foreign exchange market is a building block in the construction of macroeconomic models of the international economy. Monetary models of exchange rate determination rely on efficiency in this market for interest rates to be internationally determined, a key difference from the portfolio approach [Krueger (1983), p72]. Modern intertemporal rational expectations approaches also rely on this result. Rational expectations itself is taken as a trait of market participants. For these reasons, a large literature has arisen testing efficiency in forward exchange rate markets.

This chapter examines the typical regressions undertaken to test unbiasedness of the forward exchange rate. In particular, the problem of estimation of the confidence interval for the parameter describing the potential bias in the formation of expectations is addressed. It is well known that usual tests for unbiasedness reject when there is a risk premium. This argument is extended here by showing that the sensitivity of parameter estimates and tests depends on the stochastic process followed by the exchange rate, and that negative coefficients may result for small risk premiums. Such negative coefficients are often found in these types of regressions. The most commonly examined tests also magnify potentially economically insignificant deviations from the unbiasedness hypothesis into negative coefficient estimates, providing an alternate explanation of such regression results. Further, it is argued that these parameters, usually given the interpretation of unbiasedness coefficients, are better thought of as error correction coefficients, measuring the correlation between changes in the exchange rate and the distance from equilibrium in the forward market [Hakkio and Rush (1989)].

Instead, regressions which are specified correctly for the relevant parameter space for the unbiasedness coefficient, and are well specified under various plausible stochastic generating processes for the exchange rate, suggest that in the long run, the forward rate is unbiased. This is in direct contrast to recent results [Evans and Lewis (1993), who condition results on an exact unit root in the exchange rate]. Constructed confidence intervals suggest that potential deviation from the null hypothesis do not include differences large enough to be important economically when markets are relatively free of capital controls.

This chapter also examines the possible alternative hypothesis of static expectations. A test of the hypothesis that static expectations are rational is introduced, and this hypothesis is tested with data on forward markets. This test can be used to construct confidence intervals on the unbiasedness parameter under the null hypothesis of static expectations, in order to differentiate between these hypotheses. The ability to distinguish between static expectations and unbiased expectations depends critically on the stochastic process of the spot rate. In particular, when the spot rate follows a random walk, the two hypotheses can never be distinguished (in this case static expectations are unbiased predictors of the future spot rate).

The results suggest that the correct model of expectations is the forward rate plus some noise (potentially a risk premium) correlated with the forward rate. In terms of uncovered interest parity, the result suggests, unlike previous regressions in the literature, that we cannot reject that the hypothesized uncovered interest parity relationship holds with the expected parameters. It is of interest that, in the light of the economically implausible values generated for the coefficient of interest in the unbiasedness regression in the literature, that researchers using the uncovered interest parity relationship in international macroeconomic models have not abandoned their null hypothesis of unity in favor of these implausible estimated values. This paper provides evidence that this is the correct strategy.

The following section examines current practice and the usual results in tests for unbiasedness and orthogonality, and motivates why the hypothesis of biasedness of expectations, and particularly confidence intervals on parameters summarizing the potential extent of unbiasedness, is interesting in its own right. Section 3 explicitly derives the stochastic processes of the variables examined in the regressions of the previous section and examines methods of estimating the unbiasedness parameter and placing confidence intervals on this parameter under various data generating hypotheses for the spot rate. In the fourth section a method is introduced to place confidence intervals on the null hypothesis of static expectations. Section five discusses the results and concludes. Empirical results are presented throughout for data for the YEN/US, Deutschemark/US, Swiss Franc/US and US/British Pound forward exchange rate markets.

II. Current Practice and Usual Results

The hypothesis of rational expectations, combined with both the hypotheses of covered and uncovered interest parity, suggests that the forward exchange rate at any time t for a contract period k should be, on average equal to the actual observed spot rate at time t+k and that any errors are orthogonal to information known at time t. Let s_t denotes the log of the spot exchange rate at time t, $f_{t,k}$ the log of the forward exchange rate at time t for k periods ahead and Ω_t denotes all information known at to the market at time t.

This derivation is standard [e.g. see Baillie and McMahon (1989)]. Covered interest parity (CIP) is given by

$$f_{t,k} - s_t = i_t^* - i_t \tag{1}$$

where i_t is the log of the home interest rate at time t and i^*_t is the log of the foreign interest rate at time t, where the maturities of the interest rates are k periods (to accord with the forward rate maturity). Uncovered interest parity (UIP) can be written as

$$E_t(s_{t+k}) - s_t = i_t^* - i_t \tag{2}$$

and the hypothesis of rational expectations by $s_{t+k}=E_t(s_{t+k})+e_{1t+k}$, where the residual is a martingale difference sequence with respect to all information dated t or before (i.e. Ω_t). Combining these equations results in the expression

$$E_t[s_{t+k} | \Omega_t] = f_{t,k} \tag{3}$$

This expression can be augmented by adding e_{2t} as a risk premium[1] (if e_{2t} is non zero, UIP and CIP cannot simultaneously hold, however CIP can also be adjusted to reflect the risk premium). The result is the second basic equation

$$E_t(s_{t+k}) = \beta f_{t,k} + e_{2t} \qquad (4)$$
$$s_{t+k} = \beta f_{t,k} + e_{2t} + e_{1t+k}$$

Here departures of ß from one indicate biased expectations (so uncovered and covered interest parity cannot both hold) and correlation of e_{1t+k} with time t dated data indicates departures from the assumption of orthogonality. In general, the risk premium e_{2t} is not observed and hence only the composite error term is observed. [See Hodrick (1987) for an overview of asset market derivations of this relationship, Baillie and McMahon (1989) derive this result as above].

This model allows two types of tests to be examined, unbiasedness is examined directly by estimation of ß and orthogonality can be examined by assuming that ß=1, and running the regression

$$(s_{t+k} - f_{t,k}) = \lambda Z_t + w_{t+k} \qquad (5)$$

where Z_t is any information known to market participants at time t [e.g. Hansen and Hodrick (1980), Cumby (1986)]. The researcher can then test for $\lambda=0$. Failure of this test can be interpreted as failure of the orthogonality part of the rational expectations hypothesis in the absence of a risk premium or alternatively evidence of a risk premium, where λZ_t is the

[1] This term also subsumes random errors in expectations, this will be indistinguishable from the risk premium.

estimate of e_{2t}.

Typically, in examining the unbiasedness proposition, rather than estimating β from the relationship in (2) directly, the model is often transformed under the null hypothesis to obtain the regression

$$(s_{t+k} - s_t) = \beta_0 + \beta (f_{t,k} - s_t) + u_t \tag{6}$$

where the joint null hypothesis of unbiasedness and orthogonality becomes $\beta=1$ [and occasionally includes $\beta_0=0$, although this need not be the case as the risk premium e_{2t} may not be mean zero, see Hodrick (1987)]. The reason for undertaking this transformation is usually due to the trending properties of the exchange rate and forward rate, which are usually considered to be non stationary variables [Meese and Singleton (1982) show that the null hypothesis cannot be rejected using the tests of Dickey and Fuller (1979)][2].

The derivation and estimation results of regressions such as (4) and (6) are well documented. In the case of equation (4), see Frenkel (1976,1977,1981), Bilson (1981), Baillie et al (1983) and more recently Corbae et al. (1992), Evans and Lewis (1993) and Mark et al. (1994); for estimates of (6) see Bilson (1981), Fama (1984), Baekart and Hodrick (1993). For the unbiasedness equation (6), typical results suggest that the hypothesis that $\beta=1$ is rejected resoundingly, with t statistics often between -2 and -5. In addition, very implausible values

[2] Pope and Peel (1991) note that the "nonstationarity problem inherent" in testing equations such as (4) will "probably not appertain" in equations such as (6). What they must mean is that the results should be examined conditional on the stochastic property of the spot exchange rate, as in section 3 of this chapter.

for ß are found, in the range of -2 to -5 instead of the null of one. Imposing a value of one and testing (5) is also rejected resoundingly [Hansen and Hodrick (1980), Cumby (1986), Pope and Peel (1991)]. This was surprising given results from testing equation (4), where estimates of ß are usually close to (but less than) one.

The recently constructed dataset of Baekart and Hodrick (1993) can be employed to analyze these regressions, and the extensions that follow[3]. This dataset contains monthly data from 1975 to 1989 for the forward and spot exchange rates between the US and four countries, Japan (YEN), Germany (DM), Switzerland (SF) and the UK (BP). The data is particularly useful not only due to the span of the dataset, but also the effort put in by these authors on ensuring that the spot rate employed accords exactly with the maturity of the futures contract in each month, and also that the effect of transaction costs are removed by using the bid spot price and ask forward price. The data employed are for one month contract duration, so there are no overlapping data concerns.

For all of the regressions estimated in this paper, three sets of results are examined. Firstly, the full sample from 1975 to 1989 is employed. At the beginning of this period, most of the countries employed capital controls of varying degrees, whilst by the end of the period these types of controls had to a great extent been reduced or had disappeared (Ito (1992) for the yen, Eichengreen and Wyplosz (1993) for EMS countries). Typically, these capital controls were phased out gradually. When capital controls are in place, there is no reason

[3] The author is extremely grateful to Professor Hodrick for supplying the data used in this study.

to expect covered interest parity to hold, which would lead us to expect that there is less reason for forward rates to be unbiased predictors of future spot rates. This suggests breaking the sample into (at least) two periods. I have chosen here to break the sample at the end of 1981, so that each regression is estimated over the periods 1975-80:12 and 1981 to 1989. The choice of this date for the yen is straightforward: successive liberalizations in the capital market (with a brief reintroduction in 1979) were mostly finished by this date (the exception is that until 1984, all forward transactions were required to be backed by some actual need for this contract, i.e. an import or export license) with large liberalizations at the end of 1980 (see Ito (1992), Chapter 11). For the EMS countries, the EMS began in 1979, and as this date would result in a very short first sample, extra observations in the early period of the EMS are included in the first sample. Realignments of currencies in the EMS during the second period were commonplace until the last two years of the sample. This break date also accords reasonably closely with that used in Baekart and Hodrick (1993) (who use mid 1980 as a break date)[4].

Results from the commonly estimated regression of equation (6) are reported in Table 1. Looking down the columns for the full sample, we see the usual results from this regression. The estimates of ß are not close to 1, are in fact negative and significantly different from

---

[4] An alternative to specifying a break date from theory would be to test the null hypothesis of no breaks in the data and use any rejection to specify the break date. Recently a large number of papers have examined these types of tests (see Stock (1994) for a review). There are some results in the literature for the types of models considered here. Sephton and Larson (1991) examine rolling cointegrating regressions of equation (4), and find evidence of instability. Chiang (1988) uses the Quandt likelihood ratio test for regressions such as (4) estimated by OLS. Gregory and McCurdy (1986) shows that estimates of ß in equations such as (6) vary over time. None of these papers test the break date formally, in that they do not use methods which account for the search for the break.

one (all t statistics reported are for the null hypothesis of $\beta=1$). This result holds true for all of the currencies examined. To the opposite of intuition from efficient markets theory, for the capital controls sample (1975-81), the estimates of β are closer to zero (or one) and are now insignificantly different from one, whilst in the cleaner less capital controls sample (81-89), these estimates are more extreme than the full sample estimates. In this later sample the null is rejected strongly! The confidence intervals on this parameter in the later period do not include zero yet alone one.

Estimates of the unbiasedness coefficient can also be estimated from the specification in equation (4). OLS results, for all three periods, are reported in Table 2. Under the assumptions of a stochastically trending exchange rate, the t statistics reported are only asymptotically normally distributed if the errors driving the forward rate are uncorrelated with the errors of the equation in (4) at frequency zero. For each of the samples, the estimates of the unbiasedness coefficient β are between 0.9 and 1, and for all cases except the pound in the 1981-89 sample, the confidence interval constructed around the estimate contains the null hypothesis of one. This accords with previous results in the literature. As might be expected from the capital controls story, for each of the currencies (excepting the pound) the estimate for β is smaller (further from one) in the capital controls period, close to 1 for the second period, and somewhere between the two sub sample estimates for the full period. Especially for the first three currencies, the estimates in the second period are very close to one, these regressions suggest that deviations from unbiasedness are probably economically insignificant.

Table 1: Estimates of equation (6) by OLS

	1975-89		1975-1981		1981-1989	
	est	t stat	est	t stat	Est	t stat
YEN						
Constant	-0.010	-3.415	-0.007	-1.984	-0.020	-3.323
ß	-2.149	-4.661	-0.885	-1.499	-4.950	-5.160
CI on ß	-3.472, -0.825		-3.350,1.580		-7.210,-2.690	
DM						
Constant	-0.010	-2.575	-0.003	-0.766	-0.024	-3.811
ß	-2.986	-3.041	-0.588	-0.883	-7.524	-5.018
CI on ß	-5.555,-0.417		-4.111,2.936		-10.854,-4.195	
SF						
Constant	-0.014	-2.904	-0.012	-1.696	-0.020	-2.887
ß	-2.655	-3.427	-1.649	-1.941	-4.686	-4.312
CI on ß	-4.745,-0.565		-4.323,1.025		-7.270,-2.101	
BP						
Constant	0.007	2.146	-0.000	0.060	0.010	4.190
ß	-2.275	-3.865	0.322	-0.483	-5.696	-6.711
CI on ß	-3.937,-0.614		-2.430,3.075		-7.652,-3.741	

Notes: Estimation is for the period given by OLS. The standard errors are corrected using the variance covariance matrix equal to $(X'X)^{-1}\Omega_{x'\epsilon}(X'X)^{-1}$ where $\Omega_{X'\epsilon}$ is estimated using an autoregressive estimator where the lag length for the autoregression is chosen by a Bayesian Information Criterion (BIC). The maximum lag was set to 8. The confidence intervals reported are 95% confidence intervals. The t statistics reported test the null hypothesis of the constant equal to zero and ß=1.

Table 2: Estimates of equation (4) by OLS

	1975-89		1975-1981		1981-1989	
	est	t stat	est	t stat	Est	t stat
YEN						
Constant	0.040	0.784	0.170	0.907	0.010	0.133
ß	0.992	-0.828	0.969	-0.948	0.998	-0.147
CI on ß	0.973-1.011		0.904-1.034		0.969-1.027	
DM						
Constant	0.012	0.919	0.028	1.700	0.005	0.289
ß	0.986	-0.892	0.965	-1.841	0.994	-0.282
CI on ß	0.955-1.017		0.927-1.002		0.953-1.036	
SF						
Constant	0.016	1.411	0.027	1.400	0.009	0.570
ß	0.978	-1.475	0.963	-1.647	0.987	-0.532
CI on ß	0.948-1.007		0.920-1.007		0.939-1.035	
BP						
Constant	-0.017	-1.495	-0.037	-1.585	-0.030	-1.840
ß	0.971	-1.565	0.950	-1.569	0.932	-2.163
CI on ß	0.934-1.007		0.888-1.012		0.870 - 0.994	

Notes: As per Table 1.

If the residuals from equation (4) and the error process driving the futures rate are correlated at frequency zero, OLS is not applicable for inference. Inference in equation (4), conditional on the forward rate having an exact unit root, can proceed using cointegration techniques. In Table 3, results are presented for the same regressions as Table 2 where the

DOLS method [Phillips and Loretan (1991), Saikkonen (1991) and Stock and Watson (1993)] of estimating cointegrating relationships is employed. For the full sample, estimates of ß for all currencies are very close to the null hypothesis of one. Further, the confidence intervals on these estimates are very precise (this drastic decrease in the size of the confidence interval over OLS estimates is a result of the DOLS method using the information contained in the correlation between the error terms of the regression and the process driving the forward rate, giving greater efficiency). In the capital controls sample, the opposite is true. For all currencies excepting the pound, point estimates are below one and very significantly so. In the post capital control sample, estimates are again very close to one (for the yen and pound, this is exact to three decimal places). The largest deviation from one is a very small 0.003. This accords completely with economic theory, when capital controls are in place the forward market is not an unbiased predictor of the future spot rate, in the absence of such controls it is. Even so, the point estimates in the capital control period do not appear to be economically significant from one, and the confidence intervals show that potential deviations from the null hypothesis are small enough that they would not be considered economically significant.

Table 3: Estimates of equation (4) by DOLS

	1975-89		1975-1981		1981-1989	
	est	t stat	est	t stat	Est	t stat
YEN						
Constant	0.005	0.379	0.089	6.846	-0.000	-0.056
ß	0.999	-0.199	0.984	-6.694	1.001	0.455
CI on ß	0.994,1.004		0.980,0.993		0.998,1.004	
DM						
Constant	0.003	2.105	0.011	3.074	0.001	0.785
ß	0.999	-0.483	0.989	-2.440	1.002	1.740
CI on ß	0.995,1.003		0.980,0.998		1.000,1.004	
SF						
Constant	0.006	2.378	0.013	17.814	0.001	1.331
ß	0.998	-0.607	0.988	-11.790	1.003	1.743
CI on ß	0.992,1.004		0.986,0.990		1.000,1.006	
BP						
Constant	-0.002	-0.948	-0.000	-0.025	-0.003	-1.354
ß	1.000	0.024	1.005	0.903	1.000	-0.100
CI on ß	0.991,1.009		0.995,1.015		0.991,1.008	

Notes: The DOLS method for obtaining asymptotically efficient estimates of the cointegrating vector is employed. Eight leads and lags of the change in the regressor are used to orthogonalise the regression, and 3 covariances are employed to estimate the standard errors. The confidence intervals have size 95%. The results are robust to shorter lag length selections.

These results are somewhat similar to previous cointegration results for this regression. Corbae, Lim and Ouliaris (1992) over an earlier time sample, and not breaking the sample due to changes in capital control regimes, reject the null hypothesis for the YEN/US but fail to do so for the other currencies. Evans and Lewis (1993) reject for the YEN/US and DM/US rates. In general, the results from the cointegrating literature are unclear, rejecting the null hypothesis of efficient markets for some datasets and not for others.

It has previously been pointed out that tests of unbiasedness and orthogonality are not mutually exclusive, that the finding that unbiasedness is rejected immediately suggests the rejection of orthogonality, and vice versa, so the tests are indistinguishable. This can be shown as follows. If the result is truly biased but orthogonality holds, we could write the relationship as

$$s_{t+k} = \beta f_{t,k} + u_t, \quad \beta \neq 1, \quad u_t \perp f_{t,k} \tag{7}$$

which could be immediately rearranged to yield a relationship which is unbiased but orthoganality does not now hold, i.e.

$$s_{t+k} = f_{t,k} + \tilde{u}_{t+k}, \quad \tilde{u}_{t+k} = u_{t+k} + (1-\beta)f_{t,k} \tag{8}$$

Here $(s_{t+k} - f_{t,k}) = \tilde{u}_{t+k}$ can be predicted by t dated variables, i.e. $f_{t,k}$. When all variables are stationary, we could not distinguish between these alternative failures of rational expectations. This is true econometrically as estimation techniques yield biased estimates of ß when orthogonality fails and tests such as (5) fail when unbiasedness fails.

But when $f_{t,k}$ is I(1), as cannot be rejected often in unit root tests, such a rewriting would cause the residuals to be non stationary. Thus biasedness now would require that the forward premium, $(s_{t+k}-f_{t,k})$, be non stationary. If there existed some ß not equal to one, but such that the residuals are stationary, then estimates such as equation (4) will estimate this coefficient consistently, thus we would be able to reject unbiasedness. When methods for estimation with non stationary data are employed, the estimate for ß will be consistent even when orthogonality does not hold. Thus with nonstationary data the two hypotheses can be distinguished. Corbae, Lim and Ouliaris (1992) note that this distinction can be made for the correlation between the forward rate and the risk premium.

The hypothesis of unbiasedness is of interest independent of issues of orthogonality. Firstly, through its relation to uncovered interest parity, the coefficient of unbiasedness is here is the same parameter as in the uncovered interest parity relationship, which is a fundamental equation in international macroeconomic models (e.g. see Krueger (1983) for the importance of forward market efficiency in monetary models of the current account). This also gives the reason why a confidence interval on the unbiasedness coefficient is important. A confidence interval enables evaluation of whether or not there exist values for this coefficient consistent with the data which are economically significant. The lack of a plausible estimate for ß has lead the writers of these international macroeconomic models to be reluctant to depart from the null of ß=1. A direct test for the null hypothesis of unbiasedness would therefore allow this strategy to be validated or not, without the complication of requiring orthogonality as well. A second reason for the desire to have a direct test is that the hypothesis of unbiasedness is often an assumption in the target zone literature (see Bartolini and Bodnar (1992)).

The next section focusses on examining the null hypothesis of unbiasedness, particularly the construction of confidence intervals on this parameter. At issue is how the stochastic properties of the spot rate allow us to interpret the regressions presented in the literature and repeated above.

III. Confidence Intervals for the Coefficient of Unbiasedness

The behavior of the estimated coefficients in estimates of equations (4) and (6), and the confidence interval determined by the regressions, depends on the data generating processes for both the spot rate and the forward rate, and the information set.

The assumed data generating process for the spot rate is

$$(1-\alpha L)s_t = C(L)e_{1t} \tag{9}$$

where the roots of C(L) are assumed to lie outside the unit circle (hence the dynamics are stationary and the persistence in s_t is described by α).

Results in the literature suggest that the unit root hypothesis for the spot rate cannot be rejected [Meese and Singleton (1982)]. This suggests that α is close to one, but does not suggest $\alpha=1$ should be imposed, as values for α close to one would also not be rejected. Economic theory also suggests that we may wish to consider these alternatives in our specification. Hodrick (1987) notes that asset pricing models do not necessarily imply that exchange rates follow a random walk, and shows that the time series process followed by the spot rate depends on the time series processes for money supplies and real incomes. Alternatively, even under the assumption of a unit root in the free floating spot rate, Svensson (1992) notes that if there are target zones that the spot rate should display mean reversion. For these reasons it is necessary to consider this possibility, which will be referred to as the local to unity specification, and will be used to model highly persistent data without the extreme assumption of the largest root being equal to unity. We may also wish to allow some dynamic structure in e_{1t}, this is summarized by the lag polynomial C(L). Hakkio (1981) and others have found evidence of dynamics in vector autoregressions of the bivariate system for spot and forward exchange rates. The existence of dynamics is not ruled out on theoretical grounds as there could be a slowly evolving risk premium.

For exposition, notational ease and correspondance with the data employed here, k will be set to one for the remainder of the paper.

From section 2, rational expectations give the result is the second basic equation, equation (4) in the text and repeated here

$$\begin{aligned} E_t(s_{t+1}) &= \beta f_{t,1} + e_{2t} \\ s_{t+1} &= \beta f_{t,1} + e_{2t} + e_{1t+1} \end{aligned} \qquad (10)$$

where departures of ß from one indicate biased expectations (so uncovered and covered interest parity cannot both hold) and correlation of e_{1t+1} with time t dated data indicates departures from the assumption of orthogonality. In general, the risk premium e_{2t} is not observed and hence only the composite error term is observed. Thus all correlations could also be regarded as being a risk premium. The two error terms introduced here are the fundamental errors of the model, where e_{1t} is the error driving the spot rate and e_{2t} is the risk premium. Under the assumption that this is the full system the process for the futures rate can be derived.

The time series process for the futures rate will depend on the time series process for the spot rate for the market and also on the method used to predict the spot rate (i.e. biases, static expectations). Thus, for different alternative hypotheses, different models for the futures rate will arise. The following derivation shows this and shows that under plausible alternative hypotheses, the stochastic process followed by the futures rate is dominated by a largest root equal to the largest root in the spot rate (i.e. α) and that other parameters are only affected in the short run. The short run parameters will be treated as nuisance parameters by the estimation methods.

From equation (9), the rationally expected spot rate in the next period is given by[5]

$$E_t(s_{t+1}) = \alpha s_t + C^*(L)e_{1t} \tag{11}$$

where $C^*_i = C_{i+1}$. Equation (11) can be substituted into equation (10) to obtain

$$\beta f_{t,1} + e_{2t} = \alpha s_t + C^*(L)e_{1t} \tag{12}$$

Subtracting α times the one period lag of equation (12) from equation (12), and substituting from equation (9) for the term $(1-\alpha L)s_t$ yields

---

[5] An implicit assumption here is that the spot rate is completely exogenous. It is taken to depend only on fundamentals, so the operations of the forward market cannot in any way affect the path of the spot rate.

$$f_{t,1} = \alpha f_{t-1,1} + \frac{1}{\beta}[\alpha C(L)e_{1t} + C^{*}(L)(1-\alpha L)e_{1t} - (1-\alpha L)e_{2t}] \tag{13}$$

which gives the generating process of the futures rate in terms of the innovations to the spot rate and the forecast errors/risk premium. Using the result in equation (13) (lagged once) to substitute for αs_{t-1} in equation (9) yields

$$s_t = \beta f_{t-1,1} + e_{1t} + e_{2t-1} \tag{14}$$

where the result that $C(L)e_{1t}-C^{*}(L)e_{1t-1} = e_{1t}$ is used. This is the relationship given in equation (10) above. Collecting equations (13) and (14) gives the system to be estimated. This is

$$\begin{aligned} f_{t,1} &= \alpha f_{t-1,1} + e_{1t}^{*} \\ s_t &= \beta f_{t-1,1} + e_{2t}^{*} \end{aligned} \tag{15}$$

where $e^{*}_{t}=(e_{1t}^{*}\ e_{2t}^{*})'$, $e_{1t}^{*} = [\alpha C(L)e_{1t} + C^{*}(L)(1-\alpha L)e_{1t} - (1-\alpha L)e_{2t}]/\beta$, $e_{2t}^{*} = e_{1t} + e_{2t-1}$, and e^{*}_{t} is in general serially correlated, either through $C(L)\neq 1$ or serial correlation in the risk premium term, and the two error terms are always correlated as they both include e_{1t}.

Now that the model is defined, in the sense that residuals from equations can be related back to shocks with economic meaning, i.e. shocks driving the exchange rate or risk premium, we can evaluated the regressions of the previous section.

Firstly, consider regressions of the form in equation (6). The expected value for the 'unbiasedness' parameter can be evaluated using the results for the stochastic properties of the variables derived above. This will be done conditional on the null hypothesis of rational expectations, i.e. $\beta=1$ and e_{1t} is uncorrelated with all earlier dated variables. The regression to be run in (6) is

$$\Delta s_{t+1} = b(f_{t,1} - s_t) + u_t \tag{16}$$

(for simplicity, no constant is assumed, the result is the same using demeaned data). Now, we can write $\Delta s_{t+1} = \Delta s_t^e + e_{1t+1}$, where $\Delta s_t^e = s_t^e - s_t$ and s_t^e is the one period ahead rationally expected spot rate, and also $(f_{t,1}-s_t) = \Delta s_t^e - e_{2t}$. Now, in population $b = E[\Delta s_{t+1},(f_{t,1}-s_t)]/E[(f_{t,1}-s_t)^2]$. Evaluating this, using the orthogonality assumption yields

$$b = \frac{E[(\Delta s_t^e)^2] - E[\Delta s_t^e e_{2t}]}{E[\Delta s_t^e]^2 + E[e_{2t}^2] - 2E[\Delta s_t^e e_{2t}]} \tag{17}$$

Fama (1984) derives this exact result (their equation 5), and notes that the existence of a risk premium leads to a different value for β. He argues that deviations of β from 1 in such regressions are a "direct measure of the variation of the premium in the forward rate". This is not true for all stochastic processes followed by the spot exchange rate. We can evaluate this expression for different assumptions on α and C(L). If s_t is a martingale, then $\Delta s_t^e=0$ so $b=0$, a direct implication of the impossibility of forecasting the dependant variable, which is the innovation of a martingale. If further, $e_{2t}=0$, this would be indeterminate, as the regressor would be exactly zero under the null hypothesis always. If either α or C(L) is not one, then if $e_{2t}=0$ then $b=1$, as is usually considered the result in applications of this regression. This is an extremely strong assumption, usually at odds with the stories told by researchers using this regression. In the case where $e_{2t}\neq 0$, $b\neq 1$ but instead depends on the correlation between the risk premium and (true) expected depreciation. If the expected depreciation is small, say because the process followed by the exchange rate is close to a martingale (α is close to one), then even for small risk premia, the correlation between the risk premium and the expected depreciation may overwhelm the variance of the expected depreciation, leading to large negative values for E[b] even when the joint null hypothesis of unbiasedness and orthogonality is true. This was noted by Fama (1984) and Hodrick and Srivistava (1986). When we consider only slight deviations from $\alpha=1$, we can see that the expected depreciation can be arbitrarily small, and hence a small risk premia may dominate this and cause negative estimates of ß in regressions such as (6).

The results suggest that the estimated coefficient on the regressor $(f_{t,1}-s_t)$ in (6) should not be interpreted as an unbiasedness coefficient due to either the spot rate following a random walk or the existence of a risk premium. Even if these can be ruled out a priori, the confidence interval on this parameter cannot be interpreted as the confidence interval on the unbiasedness coefficient. For the t statistics on parameters in regressions to be inverted to construct confidence intervals for ß, they must be well specified under both the null and the alternative hypotheses. Taking $\beta\neq1$ as the alternative of interest here, (4) can be used to construct confidence intervals. This is because under the alternative that $\beta\neq1$, the equation is still well specified. Equation (6) is derived under the null of $\beta=1$, and cannot be used for construction of confidence intervals, even if $\beta=1$ under the null hypothesis. Deriving the analog to equation (6) when the alternative may be true gives the result

$$(s_{t+1}-s_t) = \beta_0 + \beta(f_{t,1}-s_t) + (\beta-1)s_t + u_{t+1} \tag{18}$$

This shows that (6) is misspecified under the alternative of biasedness. In order to be able to place a confidence interval on ß, the regression must be well specified for the entire parameter space.

Under this misspecification, small deviations from $\beta=1$ result in the omitted variable s_t, which is potentially correlated with the regressor (this correlation can also be examined using the derivation of the process followed by the forward rate above). This misspecification can also lead to large negative estimates of ß when there are only small deviations from $\beta=1$.

The apparent paradox, that differencing both sides of (4) with s_t causes the population value of the parameter of interest to change, follows by considering the case of cointegration of s_t and f_t, and the interpretation of coefficients on cointegrating variables. That these two variables are cointegrated when unbiasedness holds follows from the result that $f_{t-1}-s_t=e_{2t}$, where the difference between these two variables is stationary. The interpretation for ß here is that it is the coefficient describing the effect of deviations from the equilibrium on Δs_t, as in an error correction mechanism [Davidson et al (1976), Engle and Granger (1987)], and follows directly from the Granger Representation Theorem in Engle and Granger (1987). That this coefficient is the coefficient on an error correction term has been previously pointed out in the context of these regressions [Hakkio and Rush (1989)]. In the random walk case, e_{1t+1} is serially uncorrelated and unforecastable, so the estimated coefficient is zero for the regression. This is true regardless of the true value for ß.

Some examples of negative estimates of ß due to the above reasons are shown in a small Monte Carlo experiment in Table 4. Here, data is generated according to equation (15). The model is such that α is close to or equal to one, there is a large correlation λ between the innovation to the spot rate in the current period and the risk premium, the risk premium is potentially serially correlated (follows an AR(1) with coefficient φ). In each experiment, the results when the true value for $\beta=1$ and 0.98 are presented. In each case, the variance of the innovation to the risk premium has a standard error much smaller than that of an innovation in the spot exchange rate. The results for the estimation of equation (6) are in the final two columns of Table 4.

In the unit root model, when $\beta=1$, the coefficient estimates are close to zero and are negative when the risk premium and the innovation to the spot rate are negatively correlated. If $\beta=0.98$, a small deviation from one, these estimates are negative and large. If the spot exchange rate follows an explosive process, even when $\beta=1$ we are able to obtain large negative coefficients (the size of the coefficient can be arbitrarily changed by changing variances of the innovations). When the exchange rate is slightly mean reverting, negative coefficients are obtained when β deviates from one for the models examined here. The Monte Carlo suggests two results. Firstly, negative estimates from (6) can occur with small risk premia when the markets are efficient. Secondly, these regressions are not very useful in interpreting deviations from unbiasedness. Actually explaining the results in Table 1 above would entail examination of the risk premia in order to simulate a more realistic model [Frankel and Froot (1989) estimate such risk premia]. This cannot be done here as we do not have data on expectations, this is left to further work. The implication drawn is that results from these regressions do not tell us much about unbiasedness.

Table 4: Monte Carlo Results.

Specification			Equation (4)		Equation (6)	
ß	φ	λ	E[ß]	rej rate	E[ß]	rej.
Unit Root Model $\alpha=1$						
1	0	-0.9	1.00	0.07	-0.04	0.12
0.98	0	-0.9	0.98	1.00	-0.93	0.87
1	0.4	-0.9	1.00	0.08	-0.17	0.13
0.98	0.4	-0.9	0.98	1.00	-0.87	0.88
1	0	0.9	1.00	0.05	0.07	0.11
0.98	0	0.9	0.98	1.00	-1.06	0.87
Explosive model $\alpha=1.0056$						
1	0	-0.9	0.99	1.00	-0.66	0.28
0.98	0	-0.9	0.97	1.00	-0.59	0.93
1	0.4	-0.9	0.99	0.998	-0.46	0.26
0.98	0.4	-0.9	0.97	1.00	-0.65	0.94
1	0	0.9	0.99	1.00	-0.61	0.26
0.98	0	0.9	0.97	1.00	-0.65	0.94
Mean Reverting Model $\alpha=0.994$						
1	0	-0.9	1.01	1.00	0.86	0.04
0.98	0	-0.9	0.99	1.00	-1.10	0.71
1	0.4	-0.9	1.01	0.99	0.62	0.04
0.98	0.4	-0.9	0.99	1.00	-1.03	0.72
1	0	0.9	1.01	1.00	0.95	0.05
0.98	0	0.9	0.99	1.00	-1.29	0.65

Notes: The model estimated is that in equation (15) of the text, where $C(L)=1$, $e_{2t}=\varphi e_{2t-1} + e'_{2t}$, $E[e_{1t}e'_{2t}]=\lambda$, $E[e_{1t}^2]=1$ and $E[e_{2t}^2]=0.025$. The estimates reported are mean values of the estimate over 1000 replications with normal errors, the rejection rates are percentage rejections of the null hypothesis that $ß=1$. For the results in equation (6), a robust autocovariance is estimated by the AR method with 3 lags. The estimates of equation (4) were constructed using the DOLS cointegration estimator with 3 leads and lags. The pseudo samples had 179 observations accord with the size of the dataset employed in this chapter.

Under the joint null hypothesis of unbiasedness and orthogonality, along with orthogonality of the regressor and the risk premia, Hodrick (1987) notes that ß can be estimated using (4) by OLS and standard asymptotically normal inference applies. The estimate for ß would be consistent under both the null and the alternative hypothesis, so confidence intervals can be estimated.

In the case that $\alpha=1$, the regression (4) is a cointegrating regression, Stock (1987) shows that estimates from such regressions will be biased of order 1/T if the residuals of equation (15) are correlated at frequency zero, which is true as the innovation in the spot is included in the change in the forward rate. Thus, OLS is not applicable here. In these cases, there are a number of methods which enable asymptotically efficient and unbiased estimation of ß. These include Stock and Watson (1993), Saikkonen (1991,1992), Phillips and Loretan (1991), and Johansen (1988), see Watson (1994) for an overview. It is clear that these methods are applicable from the model derived in equation (15) above. This derivation also describes the expected correlation between the two sets of residuals of the model, as they can be related back ot the specific economic shocks.

Estimation of (4) using the Stock and Watson (1993) method employs a correction in the form of additional variables to control for the lack of orthogonality. This entails estimating the equation

$$s_{t+1} = \beta f_{t,1} + d(L)\Delta f_{t+1,1} + u_t \tag{19}$$

As is usual for the case of cointegration, the faster rate of convergence of ß has the result that this technique has high power in distinguishing deviations of ß from 1. Also, the t statistic for the estimate $\hat{ß}$ has the usual asymptotic normal distribution so confidence intervals can be constructed here for the unbiasedness coefficient. This method has previously been applied to the model in (4) by Evans and Lewis (1993), and other cointegration methods have been applied to this problem by Corbae, Lim and Ouliaris (1992), as discussed above. The results of applying these methods to the problem were presented in the previous section.

The analysis presented above shows that the regressions to be run and their interpretation depend on the stochastic properties of the spot exchange rate; in particular the size of the largest root in the spot rate, the long run correlation between e_{1t}^* and e_{2t}^* (which are the residuals of equation (27), the system being estimated), and the serial correlation in the spot rate. Results are presented in Tables 1 and 2. Using the DF-GLS test of Elliott, Rothenberg and Stock (1992), the null hypothesis of a unit root in the forward rate cannot be rejected for any of the currencies (forward or spot, recall that an implication is that the largest root in the spot and forward markets are the same for all specifications. This is a feature of the data, for each currency the median unbiased estimate of the largest root in the spot rate is equal to that of the forward rate to three decimal places). Further, confidence intervals constructed in the method of Stock (1991) around the estimated largest root in these series using Dickey Fuller (1979) regressions show that the null of a unit root cannot be rejected. Neither, of course, can roots close to unity. The long run correlation between the two residuals of equation (15) is shown to be very large; these values are given for each currency in the row marked δ, where this coefficient is bounded between -1 and 1.

Table 5: Unit Root Tests

	YEN	DM	SF	BP
Spot				
DF-GLS	0.012	-1.184	-0.786	-0.468
DF-GLS$_u$	-1.021	-1.307	-1.603	-1.768
median	1.007	1.005	0.998	0.982
CI	0.971-1.025	0.959-1.024	0.941-1.022	0.925-1.02
Forward				
DF-GLS	0.002	-1.183	-0.786	-0.476
DF-GLS$_u$	-1.035	-1.316	-1.611	-1.768
median	1.007	1.005	0.998	0.982
CI	0.970-1.025	0.958-1.024	0.941-1.022	0.925-1.02
δ	0.982	0.994	0.993	0.861

Notes: The DF-GLS and DF-GLS$_u$ statistics are from Elliott, Rothenberg and Stock (1992) and chapter 2 respectively. The critical values (95% one tailed) are -1.96 and -2.72 respectively. The 95% confidence interval and median unbiased estimates are calculated in the method of Stock (1991). The estimate for $\delta=\Omega_{12}/(\Omega_{11}\Omega_{22})^{1/2}$, where $\Omega=2\pi S_{\epsilon^*}(0)$ and $S_{\epsilon^*}(0)$ is the spectral density of ϵ_t^* at frequency zero where ϵ_t^* are the residuals of equation (15). The median unbiased estimate of α from above and the cointegrating estimate of β from Table 3 are used to identify ϵ_t (see chapter 3 lemma 3 for justification of this estimation in the local to unity case).

Table 6: Stochastic Properties of the Exchange Rate

	1975-89		1975-81		1981-89	
	est	t stat	est	t stat	est	t stat
YEN						
Const	0.231	1.91	0.399	1.97	0.093	0.71
s_{t-1}	-0.004	-1.45	-0.004	-1.04	-0.004	-1.21
Δs_{t-1}	-0.096	-0.97	-0.198	-1.05	-0.006	-0.06
Δs_{t-2}	-0.182	-2.45	-0.418	-2.87		
Δs_{t-3}			-0.163	-1.36		
Deutsch	Mark					
Const	0.013	0.95	-0.027	-0.15	0.153	1.11
s_{t-1}	-0.002	-0.89	-0.002	-0.60	-0.003	-0.77
Δs_{t-1}	-0.129	-1.71	-0.114	-0.93	-0.120	-1.22
SF						
Const	0.045	0.43	-0.064	0.37	0.116	0.87
s_{t-1}	-0.004	-1.17	-0.005	-1.04	-0.003	-0.79
Δs_{t-1}	-0.020	-0.26	0.031	0.26	-0.049	-0.50
BP						
Const	0.124	1.19	0.099	0.62	0.121	0.877
s_{t-1}	0.001	0.49	0.000	0.09	0.002	0.476
Δs_{t-1}	-0.093	-1.24	0.063	0.50	-0.129	-1.32

Notes: The autoregression is $\Delta s_t = \beta_0 + (\alpha\text{-}1)s_{t-1} + a(L)\Delta s_{t-1} + u_t$, where the order of a(L) is selected by a BIC selector starting with a maximum of 8 lags.

Table 6 shows that there is little autocorrelation in the change in the spot exchange rate in any of the subperiods. For the yen, the order of the autoregression changed over subperiods, the results indicating that for the capital controls period there was some autocorrelation (significant second lag), which appears to disappear in the post capital controls period. For none of the other currencies in any period is a lag of the change in the spot rate significant. This suggests that the long run version of unbiasedness has interpretation for the short run as well.

The failure to reject a unit root in the spot rate in Table 5 does not imply that we can condition on α as being known to be equal to one. If we are not able to make the assumption that the spot rate has an exact unit root, the cointegration results above are not valid here, as they are conditional on the fact that this root known to be exactly equal to 1. That this condition matters analytically and empirically has been shown in chapter 3, which shows that for even small deviations from $\alpha=1$, tests based on the cointegration methods can severely overreject the null hypothesis. This result can also be seen from the Monte Carlo results in Table 4. In both the explosive model and mean reverting model, despite small deviations from $\alpha=1$, the cointegrating regression tests rejected the true null hypotheses (rows where $\beta=1$) almost always.

In this case, for α sufficiently close to one, computed confidence intervals for β in (4) can be constructed using the Bonferroni method of Cavanagh, Elliott and Stock (1993)[6] described in chapter 1. The Bonferroni method takes into account a pretest for the size of the largest root in the right hand side variable (here the forward rate), and adjusts the size of the test or confidence interval accordingly. The actual estimation of (4) is directly by OLS, with the OLS estimate being biased but consistent (in the same way as OLS regression

[6] This method applies if α is close to one so that local to unity type asymptotic results better approximate small sample results better than normal distributions.

on cointegration coefficients are biased but consistent). The confidence intervals in this procedure are wider than those calculated when it is assumed there is a unit root (as less information is being assumed) but the method avoids the overrejection of true null hypotheses due to the largest root not being equal to one.

If there is serial correlation in the spot exchange rate (i.e. C(L) not equal to one), then the concept of unbiasedness here has both a short and long run interpretation. For the strict hypothesis of unbiasedness, the coefficients for both the short and long run parts as predictors of the next period spot exchange rate should be equal to one. It could well be the case that in the long run, the forward rate is an unbiased predictor of the future spot rate but in the short run it is not. The coefficient estimated in the Bonferroni method is only consistent for the parameter on the long run component of the forward rate, as the short run dynamics are treated as nuisance parameters in these techniques[7]. This means that there may exist short run biases which will not lead to rejections using the testing procedure employed here. This could be considered unimportant for three reasons. Firstly, the long run components dominate the variance of the next period spot rate, so could be considered much more important. Secondly, these short run components may in any case appear to be insignificant from the results of Table 6. Thirdly, many international macroeconomic models are attempting to capture long run features of the data, thus it is the long run unbiasedness coefficient that is of interest. Alternatively, testing for long run unbiasedness could be considered a weak form test, necessary but not sufficient for unbiasedness of

---

[7] In construction of the t statistic, robust standard errors such as those of Newey and West (1989) or Andrews (1991) should be used to take account of the serial correlation.

expectations when there are short run dynamics.

The confidence intervals from estimating this model and using the Bonferroni method in Cavanagh, Elliott and Stock (1993) for determining the critical values for the t statistic on the unbiasedness parameter is given in Table 7. Note that the point estimates are simply the OLS estimates from estimating equation (4), which are reported in Table 2. The Bonferroni bounds show that the true confidence bounds are shifted (as the confidence interval is around the OLS estimate and not the DOLS estimate) and wider than those when cointegration techniques are used. The reported confidence intervals include the null hypothesis in all cases in all samples except the pound in the post capital controls sample. Despite this lack of rejection, comparing the confidence intervals for the two sub samples shows that in the capital controls period the confidence intervals are much wider and include greater deviations from one than the confidence intervals in the later period. This accords with the economic intuition given.

Table 7: Bonferroni confidence intervals on ß

	1975-89	1975-1981	1981-1989
YEN	0.973-1.024	0.902-1.078	0.969-1.045
DM	0.958-1.038	0.926-1.028	0.952-1.064
SF	0.951-1.028	0.919-1.037	0.938-1.068
BP	0.939-1.033	0.887-1.055	0.842-0.991

Notes: Reported are 95% confidence intervals around the OLS estimates for ß from equation (4) given in Table 2. The confidence intervals are estimated according to the Bonferroni approach of Cavanagh, Elliott and Stock (1993), where the first stage size is 1% and the second stage size is 4% (see Cavanagh, Elliott and Stock (1993) for details).

A caveat to the results is that the Bonferroni method is conservative tends to under reject hypotheses, i.e. at the 5% stated level the true size may be much smaller. Thus, one interpretation could be that the cointegration techniques are invalid as the root is not exactly one, and the results of the Bonferroni methods stand. The conclusion would be that we can not reject the null, either due to it being correct or through lack of power. An alternate interpretation is that the rejections of the DOLS estimator are true and due to its higher power, and not due to biases from the root not equal to unity. The data does not allow us to tell. If one were to believe very strongly that the spot rate contains a unit root, then one could reject the null hypothesis, but this rejection would be due to this belief. The conservative conclusion is that the null hypothesis should not be rejected.

These results give one answer to the quandry that each specification gives wildly different results - the interpretation of the estimates depends on the stochastic process followed by the exchange rate. There appears to be a consensus amongst researchers that the exchange rate follows some process where shocks are persistent, this suggests that the levels equations are more likely to provide useful estimates of the confidence interval on the unbiasedness coefficient. If this persistance is such that the exchange rate follows a process close to a unit root, then this could suggest the reason for the large rejections using equation (6). The results also show that the rejections found with the cointegrating methods, such as in Evans and Lewis (1993) and Corbae et al. (1992) may be due to the critical assumption that the forward rate (and hence spot rate) has an exact unit root.

IV. Static Versus Rational Expectations

The results of section 3 were derived under the assumption that deviations from unbiasedness of expectations were in the direction of ß not equal to one. It is of interest to examine how these results differ if the alternative hypothesis is that market participants have static rather than rational expectations. In this case, the models to be estimated can be specified.

When market participants have static expectations, the futures rate is set according to

$$f_{t,1} = s_t + e_{2t} \tag{20}$$

where e_{2t} is again a random forecast error/risk premium term. This equation can be employed along with equation (9) to derive the generating process for the futures rate. This yields the equation

$$f_{t,1} = \alpha f_{t-1,1} + C(L)e_{1t} + (1-\alpha L)e_{2t} \tag{21}$$

which is the static expectations analog to equation (13) above. The result here can be substituted back into equation (9) to obtain an equation for s_t as a function of the lagged futures rate. This result gives the equation

$$s_t = \alpha f_{t-1,1} + C(L)e_{1t} + \alpha e_{2t} \tag{22}$$

Comparing this to the rational expectations result, we see that under static expectations, $\beta=\alpha$ in equation (4) for this alternative. Note that if $\alpha=1$, then this is optimal (in the absence of dynamics). The static expectations result also results in additional dynamics in the levels unbiasedness regression if there are dynamics in the spot rate. This is a consequence of the suboptimal static expectations having ignored information in the dynamic structure in the setting of the futures rate.

This result can be written as the system

$$\begin{aligned} f_{t,1} &= \alpha f_{t-1,1} + \epsilon_{1t} \\ s_t &= \alpha f_{t-1,1} + \epsilon_{2t} \end{aligned} \tag{23}$$

where $\epsilon_t=(\epsilon_{1t},\epsilon_{2t})$ has a general dynamic structure given by $\epsilon_t=\Phi(L)e_t$ where $\Phi(L)$ is equal to

$$\begin{bmatrix} C(L) & -(1-\alpha L) \\ C(L) & \alpha \end{bmatrix} \tag{24}$$

and has a long run covariance matrix Ω which has off diagonals not equal to zero (hence the two equations are linked by the residuals as in SUR models).

Examining equation (23) above, if $C(L)=1$ and $\alpha=1$ then the null hypotheses of static expectations and rational expectations (unbiasedness) are identical. This is the well known result for random walks (here a martingale) that $f_{t,1}=E_t(s_{t+1})=s_t$. Thus static expectations, merely setting the future price at todays price, is the unbiased strategy.

When we relax the result that the exchange rate has a unit root, these two hypotheses imply different behavior for the forward rate. Now, $f_{t,1}=E_t(s_{t+1})=\alpha s_t \neq s_t$. In the case of short run dynamics, the same result holds true if we consider only the hypothesis of long run unbiasedness. The more mean reverting is the spot rate, i.e. the lower is α, the greater the difference between the two hypotheses. This can be shown graphically in figure 1. This picture shows, in (α,β) space, the two hypotheses. The null hypothesis of unbiasedness is $\beta=1$, shown as the horizontal line. From the results above, $\beta=\alpha$ under the null hypothesis of static expectations. This null hypothesis is shown on figure 1 by the 45° line. Examination of figure 1 shows that the power of statistical tests to distinguish between the hypotheses of static expectations and unbiased expectations depends on the distance α is from one.

This formulation suggests an approach to testing the joint hypothesis that static expectations are unbiased, i.e. H_0: $\alpha=\beta=1$. An F test constructed from OLS estimates can be employed to test this null hypothesis. As the data has a unit root under this null hypothesis, standard chi square critical values will not apply. Critical values for this distribution are tabulated in Cavanagh, Elliott and Stock (1993), Table 4.

Results of these F tests are given in Table 8. In the full sample, no tests reject the null hypothesis. In the capital controls period, tests for the US\YEN and US\SF data reject, whilst in the no capital controls period no test rejects. These tests find that there is some deviation from the joint null in the capital controls period, either in the direction of the largest root in the spot rate not equal to one or in the direction of biased expectations. With the exception of the DM, these results accord with the comparison of the DOLS and Bonferroni results. We can conclude that in the capital controls sample, if $\alpha=1$ then we reject $\beta=1$, alternatively if $\beta=1$ the F tests suggest the rejection of $\alpha=1$. But considering these separately, tests for $\alpha=1$ invariant to β (i.e. the unit root tests) cannot reject and tests of $\beta=1$ invariant to α (the Bonferroni results) cannot reject. Thus, for this sample we have a problem of lack of power.

Table 8: 95% F tests of H_0: $\alpha=\beta=1$ (Static Expectations are unbiased)

	1975-89	1975-1981	1981-1989
YEN	0.482	6.654	1.265
DM	0.440	2.684	1.767
SF	0.699	17.274	1.477
BP	1.300	1.060	0.761

Notes: The statistics reported are F statistics testing the null hypothesis of H_0: $\alpha=\beta=1$ in equation (23). An autoregressive correction estimated with lag length of 8 periods for serial correlation is employed. Asymptotic critical values for the tests are 4.06 for a 90% test and 4.93 for a 95% test.

The general null hypothesis of static expectations can also be examined independently of the hypothesis of unbiased expectations, i.e. testing only $\alpha=\beta$. The confidence interval for the static expections null hypothesis $\alpha=\beta$ could be constructed by inverting a sequence of F tests such as that used above. A size a% test could be undertaken for a sequence of alternative values $\beta^{**}=\alpha=\beta$, and the level (1-a)% confidence interval would be the set of values of β^{**} for which the test of $\alpha=\beta=\beta^{**}$ does not reject. The critical values for the tests depend on the alternative β^{**}, these critical values are readily evaluated using Monte Carlo techniques, Cavanagh, Elliott and Stock (1993) Table 4 report these critical values for selected alternatives.

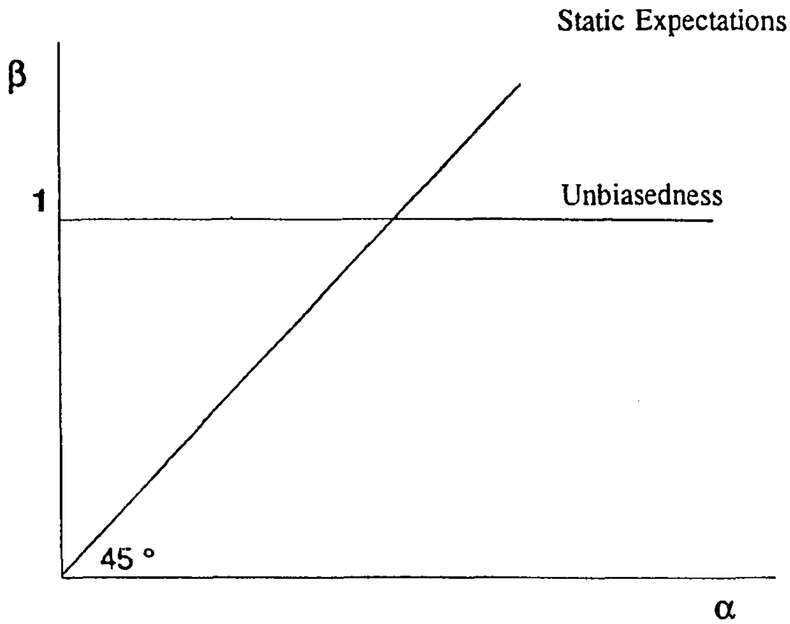


Figure 1. Static Expectations and Unbiasedness

Table 9 reports confidence interval estimates constructed by inverting the F test in this way. These confidence intervals are confidence intervals on the null hypothesis of static expectations. For the full sample, the null hypothesis of unbiased expectations is included in each confidence interval, suggesting that we cannot reject the joint hypothesis of static expectations which are rational. In the capital controls period, the method failed to find any restricted values which were compatible with the data for the YEN and SF, suggesting that static expectations is rejected here. In the case of the DM, the null of unbiasedness is not rejected. For the no capital controls period, the null of static expectations which are unbiased was not rejected for any of the currencies.

Table 9: Confidence Intervals from inversion of F tests of H_0: $\alpha=\beta$ (Static Expectations)

	1975-89	1975-1981	1981-1989
YEN	0.978-1.028	*	0.953-1.019
DM	0.960-1.026	0.967-1.031	0.936-1.019
SF	0.954-1.026	*	0.925-1.025
BP	0.940-1.019	0.875-1.053	0.930-1.040

Notes: The confidence intervals reported are from inverting a sequence of F statistics testing the null hypothesis of H_0: $\alpha=\beta=\beta^{**}$ in equation (23). The values reported are ranges of β^{**} for which the test fails to reject. An autoregressive corrections for serial correlation is employed. Asymptotic critical values for the tests depend on the alternative β^{**} and were computed using Monte Carlo methods. The search procedure was restricted to the interval $-40<c<9.5$, where $\beta^{**}=1+c/T$, and T is the number of observations in the regression. In the cases denoted by *, all tests rejected the null hypothesis. The tests are 5% tests, yielding confidence intervals of level 95%.

V. Discussion and Conclusion

The results above suggests that we cannot reject the unbiasedness hypothesis part of the efficiency hypothesis, at least in the long run, in the absence of extensive capital controls.

Tests neither reject the unit root hypothesis for the exchange rate, the unbiasedness hypothesis invariant to α, or the joint hypothesis of the two. Further, the lack of serial correlation in the spot rate suggests that these long run tests also have short run interpretation.

When there are substantial capital controls, there is clearly more to the analysis of these markets. The empirical evidence suggests that either α or β diverges somewhat from one, suggesting that either the exchange rate does not have a unit root or that the forward market is an inefficient predictor of the future spot rate. Either of these hypotheses are possible - when there is intervention on the part of governments to manage exchange rates, the stochastic properties can change (e.g. a target rate or range can lead to mean reversion, as policies are changed to attempt to stabilize the exchange rate). Alternatively, in periods where capital controls are in place, unbiasedness may fail as the market cannot function freely. Unfortunately, the econometric tests were unable to distinguish between these possibilities for the samples available. The empirical evidence suggests that confidence intervals on the unbiasedness coefficient are wide in this period and include alternatives that are potentially large enough to be economically significant.

The rejections involving estimation of equation (6) suggest very little. The analysis explains why specifications such as (6) reject unbiasedness[8]; they reject due to biasedness, lack of orthogonality or existence of a risk premium. Any of these possibilities are capable of generating the results usually seen in these regressions. The possibility that this result is due

[8] Note that the conclusion here is the opposite of that found by McCallum (1992).

to a risk premium is especially likely if there is close to a unit root in the spot rate, a result found by the confidence intervals on this parameter presented. One cannot tell from these regressions whether the forward rate is unbiased or not.

In terms of the uncovered interest parity hypothesis, arguably the more interesting result for economists, this suggests that the usual coefficient of one on the difference between domestic and foreign interest rates is correct (this difference equals the forward rate by covered interest parity). The analysis also suggests that the risk premia of interest will be correlated with the difference between these two rates. This is not surprising, if these interest rates are subject to government manipulation as they are instruments of monetary policy, then the risk premia is expected to be correlated with their movements.

The tests and models of this paper are weak form [Hodrick (1987)] as the information set is assumed to only include past observations of the spot exchange rate. It would be an improvement theoretically and econometrically to include other information. Another extension that could be undertaken is to examine more fully the role of dynamics in the model. Given that lack of power is apparently a problem in the capital controls sample, perhaps a system modelling approach for this sample could overcome this. Power could be increased through explaining joint correlations which would arise because of the common currency used to measure the exchange rates. It would be desirable to examine estimates of the risk premia directly, which could not be undertaken here as no data on the expectations of market participants was available. This avenue of research is currently being investigated by the author with an alternate dataset.

Bibliography

Andrews, D.W.K (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation", *Econometrica*, 59, 817-858.

Baekart, G. and R.J.Hodrick (1993), "On Biases in the Measurement of Foreign Exchange Risk Premiums", *Journal of International Money and Finance*, 12, pp 115-138.

Baillie, R.T, R.E.Lippens and P.C. McMahon (1983), "Testing Rational Expectations and Efficiency in the Foreign Exchange Market", *Econometrica*, 51, pp 553-63.

Baillie, R.T and P.C. McMahon (1989), The Foreign Exchange Market: Theory and Econometric Evidence, Cambridge University Press: Cambridge.

Bartolini, L., and G.M.Bodnar (1992) "Target Zones and Forward Rates in a Model with Repeated Realignments", *Journal of Monetary Economics*, 30, pp373-408.

Bilson, J.F.O (1981), "The Speculative Efficiency Hypothesis", *Journal of Business*, 54, 435-52.

Cavanagh, C.L., G. Elliott and J.H.Stock (1993), "Inference in Models with Nearly Nonstationary Regressors", manuscript, Department of Economics, Harvard University.

Chiang, T.C. (1988) "The Forward Rate as a Predictor of the Spot Rate - A Stochastic Coefficient Approach", *Journal of Money, Credit and Banking*, 20, 212-232.

Corbae, D. K-G Lim, and S.Ouliaris (1992), "On Cointegration and Tests of Forward Market Unbiasedness", *Review of Economics and Statistics*, pp 728-32.

Davidson, J.E.H, D.F.Hendry, F.Srba and S.Yeo (1978), "Econometric Modelling of the Aggregate Time-Series Relationship between Consumers' Expenditure and Income in the United Kingdom", *Economic Journal*, 88, pp 661-92.

Dickey, D.A. and W.A.Fuller (1979), "Distribution of Estimators for Autoregressive Time Series with a Unit Root", *Journal of the American Statistical Association*, 74, pp427-431.

Eichengreen, B and C.Wyplosz (1993), "The Unstable EMS", *Bookings Papers on Economic Activity*, 1, pp51-124.

Elliott, G., T.J.Rothenberg, and J.H.Stock (1992), "Efficient Tests for a Unit Root", NBER Technical Working Paper #130.

Engle, R.F. and C.W.Granger (1987), "Cointegration and Error Correction: Representation, Estimation and Testing", *Econometrica*, 55, 251-76.

Evans, M.D.D. and K.K.Lewis (1993), "Do Long Term Swings in the Dollar Affect Estimates of the Risk Premia?", Stern School of Business Discussion Paper EC-93-12, New York University, June.

Fama, E.F. (1984) "Forward and Spot Exchange Rates", *Journal of Monetary Economics*, 14, pp 319-38.

Frenkel, J.A. (1976) "A Monetary Approach to the Exchange Rate: Doctrinal Aspects and Empirical Evidence", *Scandinavian Journal of Economics*, 78, 200-24.

Frenkel, J.A. (1977) "The Forward Exchange Rate, Expectations, and the Demand for Money", *American Economic Review*, 67, pp653-670.

Frenkel, J.A. (1981) "The Collapse of PPP in the 1970's", *European Economic Review*, 16, 145-65.

Gregory, A.W. and T.H.McCurdy (1896) "The Unbiasedness Hypothesis in the Forward Foreign Exchange Market", *European Economic Review*, 30, pp 365-81.

Hakkio, C. and M.Rush (1989), "Market Efficiency and Cointegration: An Application to the Sterling and Deutschemark Exchange Rates", *Journal of International Money and Finance*, 8,pp 75-88.

Hodrick, R.J. (1987) The Empirical Evidence on the Efficiency of Forward and Futures Foreign Exchange Markets, Harwood Academic Publishers:Switzerland.

Hodrick, R.J. and S.Srivistava (1986) "The Covariation of Risk Premiums and Expected Futures Spot Rates", Journal of International Money and Finance, 3, 5-30.

Ito, T (1992) The Japanese Economy, MIT Press: Cambridge.

Johansen, S. (1988), "Statistical Analysis of Cointegrating Vectors", *Journal of Economic Dynamics and Control*, 12, pp 231-54.

Krueger, A.0. (1983), Exchange Rate Determination, University of Cambridge Press: Cambridge, U.K.

Mark, N.C., Y.Wu and W.Hai (1993), "Understanding Spot and Forward Exchange Rate Regressions", manuscript, Ohio State University.

McCallum, B. (1992) "A Reconsideration of the Uncovered Interest Parity Relationship", NBER Discussion Paper number 4113, July.

Meese, R.A. and K.J.Singleton (1982), "On Unit Roots and the Empirical Modeling of Exchange Rates", *Journal of Finance*, 37, pp 1029-35.

Newey, W.K. and K.D.West (1987) "A Simple Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix", *Econometrica*, 55, 703-8.

Pope, P.F. and D.A.Peel (1991) "Forward Foreign Exchange Rates and Risk Premia: A Reappraisal", *Journal of International Money and Finance*, 10, 443-56.

Phillips, P.C.B. (1987), "Towards a Unified Asymptotic Theory for Autoregression", *Biometrica*, 74, 535-47.

Phillips, P.C.B. and M.Loretan (1991), "Estimating Long-Run Economic Equilibria", *Review of Economic Studies*, 59, pp407-36.

Saikkonen, P (1991), "Asymptotically Efficient Estimation of Cointegration Regressions", *Econometric Theory*, 7, 1-21.

Sephton, P.S and H.K.Larsen (1991), "Tests of Exchange Market Efficiency: Fragile Evidence from Cointegrating Tests" *Journal of International Money and Finance*, 561-70.

Sims, C.A. (1988) "Bayesian Skepticism on Unit Root Econometrics", *Journal of Economic Dynamics and Control*, 12, 463-74.

Sims, C.A., J.H.Stock and M.W.Watson (1990), "Inference in Linear Time Series Models with Some Unit Roots", *Econometrica*, 58, pp113-44.

Stock, J.H. (1987), "Asymptotic Properties of Least Squares Estimators of Cointegrating Vectors", *Econometrica*, 55, 1035-56.

Stock, J.H. (1991), "Confidence Intervals for the Largest Autoregressive root in U.S. Macroeconomic Time Series", *Journal of Monetary Economics*, 28, 435-459.

Stock, J.H. and M.W.Watson, (1993), "A Simple Estimator for Cointegrating Relationships", *Econometrica*.

Svensson, L.E.O. (1992) "An Interpretation of Recent Research on Exchange Rate Target Zones", Journal of Economic Perspectives, 6, 119-144.

Watson, M.W. (1994), "Vector Autoregressions and Cointegration", in McFadden and R.F.Engle eds. Handbook of Econometrics V4, forthcoming.

Whitt, J.A. Jr. (1992) "Nominal Exchange Rates and Unit Roots: A Reconsideration", *Journal of International Money and Finance*, 11, 539-55.