Board of Governors of the Federal Reserve System

A SIMPLE APPROACH TO ROBUST INFERENCE IN A COINTEGRATING SYSTEM

Jonathan H. Wright

A SIMPLE APPROACH TO ROBUST INFERENCE IN A COINTEGRATING SYSTEM

Jonathan H. Wright*

Abstract: Cointegration requires all the variables in the system to have exact unit roots; accordingly it is conventional for researchers to test for a unit root in each variable prior to a cointegration analysis. Unfortunately, these unit root tests are not powerful. Meanwhile, conventional cointegration methods are not at all robust to slight violations of the requirement that each variable has a unit root. In this paper I show how this difficulty may be circumvented by instrumenting the regressors in the cointegrating regression by deterministic polynomial time trends, or by artificially generated random walks.

KEYWORDS: Cointegration, Local to Unit Roots, Robustness, Instrumental Variables.

1. INTRODUCTION

Cointegration is an enormously important concept in empirical macroeconomics and finance. The idea that there are stable relationships among persistent variables is appealing both theoretically and empirically. The canonical definition of cointegration is that there exists a stationary linear combination of variables that have unit roots. Researchers typically estimate cointegrating regressions by an estimator such as pseudo-Gaussian maximum likelihood in a vector error-correction model (Johansen 1988) or dynamic OLS (Stock and Watson 1993 and others). Assuming that the regressors have exact unit roots, these estimators are asymptotically efficient and asymptotically equivalent and the associated t and F statistics for hypothesis tests on the cointegrating vector have their usual normal and χ^2 null limiting distributions.

But there is an important lack of robustness in these approaches to inference which was first noted by Elliott (1998) and subsequently by Stock and Watson (1996) and Wright (1999). The problem is that if the regressors in the putative cointegrating regression have roots that are slightly below one, tests based on comparing the t and F statistics based on conventional estimators with normal or χ^2 critical values can have size far above the nominal level. Of course, the definition of cointegration requires all the variables to have unit roots and accordingly it is conventional to apply a unit root test to each variable before proceeding to a cointegration analysis. But these unit root tests have poor power in small samples and are not consistent against the alternative that the variable in question

1

has a root that is in a $1/T$ neighborhood of unity, where T is the sample size. The size distortions in the conventional approaches to cointegration analysis are serious even when the variables have roots which are so close to the unit circle that the unit root tests have little power. Elliott derived the nonstandard asymptotic distributions of the conventional t and F statistics for hypothesis tests on the cointegrating vector, when the regressors in the putatively cointegrating regression are in a $1/T$ neighborhood of unity. He showed that severe asymptotic size distortions will result from using the conventional critical values. Allowing the largest autoregressive root of the regressors to depend on the sample size is simply a statistical device which has been found to provide a good approximation to the sampling distribution of estimators and test statistics in small samples with roots close to the unit circle (see e.g. Stock (1995) and the references therein). It also makes concrete the lack of robustness with which this paper is concerned; namely a lack of robustness of conventional cointegration methods to violations of the unit root hypothesis which a pretest will have difficulty in detecting.

The lack of robustness noted by Elliott (1998) is an important practical problem because of the enormity of the potential size distortions and because researchers rarely have reason to be confident that variables have exact unit roots. Indeed, for many economic variables such as interest rates, there are strong theoretical reasons to believe that their evident persistence should be characterized by a root slightly below unity.

This paper proposes an approach to inference in a cointegrating system which is ro-

bust to the regressors having roots slightly below one. The idea involves instrumenting the regressors in the cointegrating regression by deterministic polynomial time trends, or by artificially generated random walks. Although independent of the regressors by construction, these instruments will be highly correlated with the regressors; this is of course a spurious correlation. I show that *in the just identified case*, the F statistic for hypothesis tests on the *entire* cointegrating vector associated with this instrumental variables (IV) estimator has a χ^2 null limiting distribution. This is true regardless of the endogeneity of the cointegrating regressors and regardless of whether the regressors have exact unit roots, or roots local to unity. Confidence sets for the vector of cointegrating coefficients may hence be formed, by appeal to the duality between hypothesis tests and confidence sets, and these confidence sets will not suffer from the lack of robustness noted by Elliott (1998). This approach to inference in a cointegrating system is computationally and conceptually considerably simpler than other available robust alternatives and can in addition be more powerful.

The idea of estimating cointegrating coefficients using deterministic polynomial time trends or artificially generated random walks as instruments was first proposed by Phillips and Hansen (1990), who showed that this estimator was consistent but did not eliminate endogeneity bias. This estimator is not however widely used. The contribution of this paper is to show that the F statistic for hypothesis tests on the entire cointegrating vector associated with this estimator has a standard null limiting distribution that is free of nuisance parameters even when the regressors have roots local to unity, rather than exact unit roots,

3

and so offers a solution to the problem noted by Elliott (1998). This paper is also related to the work of Park (1990), who proposed testing a null of cointegration by testing for correlation between the residuals of the putative cointegrating regression and polynomial time trends. However, this paper is not concerned with testing for cointegration, but rather with inference on the cointegrating coefficients, when the regressors may not have exact unit roots.

The plan of the remainder of this paper is as follows. In section 2 the model and the proposed test are described and the consistency of the test against fixed alternatives is shown. In section 3 its local asymptotic power is compared to that of other tests. Section 4 contains a finite sample Monte-Carlo simulation. An empirical application of the proposed methodology is reported in section 5. Section 6 concludes.

2. THE MODEL AND PROPOSED METHOD

Consider the model

$$y_t = \mu_y + \beta' x_t + u_{1t}$$

$$x_t = \mu_x + v_t, \ v_t = A v_{t-1} + u_{2t}, \ A = I_p + T^{-1}C,$$

where β is a px1 parameter vector and C is a fixed pxp matrix. It is assumed that $u_t = (u_{1t}, u'_{2t})' = D(L)\varepsilon_t$, where $\varepsilon_t = (\varepsilon_{1t}, \varepsilon'_{2t})'$ is a zero-mean i.i.d. random vector, each element of

4

which has 4 finite moments and $D(L)$ is a 1-summable lag-polynomial. Under these assumptions, the functional central limit theorem $(T^{-1/2}\Sigma_{t=1}^{[Tr]}u_{1t}, \ T^{-1/2}\Sigma_{t=1}^{[Tr]}u'_{2t})' \Rightarrow (U_1(r), U_2(r)')'$ is satisfied, where $U_1(r)$ is a scalar Brownian motion and $U_2(r)$ is a p-vector Brownian motion.

If x_t were exactly I(1) ($C = 0$), then this would be the standard model of cointegration in triangular form. As explained in the introduction, I do not wish to impose the exact unit roots assumption and so allow for C to be nonzero. However, x_t is sufficiently persistent that unit root tests will have difficulty in rejecting the exact unit root hypothesis; more precisely, these tests will be inconsistent because A is in a T^{-1} neighborhood of the identity matrix.

The standard methods for analyzing cointegrating systems assume that $C = 0$. Many asymptotically efficient estimates for β have been proposed in this case. Dynamic OLS (Stock and Watson 1993 and others) is one such estimator. Let $\hat{\beta}$ be the dynamic OLS estimate of β, i.e. the estimated coefficient on x_t in an OLS regression of y_t on d_t, x_t, Δx_t and $o(T^{1/3})$ leads and lags thereof. If $C = 0$, $\hat{\beta}$ is superconsistent for β and the associated t and F statistics have their usual normal and χ^2 null limiting distributions. But if $C \neq 0$, while $\hat{\beta}$ is still superconsistent, the t and F statistics do not have these usual limiting distributions (Elliott (1998) and Stock and Watson (1996)). Their limiting distributions instead depend upon C, a nuisance parameter which is not consistently estimable. If the usual normal and χ^2 critical values are used, the t and F tests suffer from serious size distortions. The corresponding confidence sets for the parameter β have coverage that is far below the nominal level.

2.1 *The Proposed Method*

In this paper, I propose an approach to inference in this model which does not require the variables to have exact unit roots. The idea is to estimate β by the just identified instrumental variables estimator, using a p-vector of instruments, z_t, for which I consider two specifications:

Instruments I1: $z_t = (\frac{t}{T}, \frac{t^2}{T^2}, ... \frac{t^p}{T^p})'$, so that $z_{[Tr]} \to \zeta(r) = (r, r^2, ... r^p)'$, as $T \to \infty$.

Instruments I2: z_t is a p-vector random walk with standard normal increments, independent of all the other time series in the system, so that $T^{-1/2} z_{[Tr]} \Rightarrow Z(r)$, as $T \to \infty$ where $Z(r)$ is a standard Brownian motion (independent of $(U_1(r), U_2(r)')')$.

Other deterministic functions of time could be used as instruments (e.g. broken trends), but I limit attention to these two sets of instruments in this paper. Although independent of x_t by construction, these instruments will be highly correlated with x_t; this is of course a spurious correlation. The idea of exploiting this spurious correlation to generate valid instruments was first proposed by Phillips and Hansen (1990) in a model with exact unit roots ($C = 0$). In this paper, I show that the F statistic for hypothesis tests on the entire cointegrating vector associated with this just identified estimator has a standard null limiting distribution that is free of nuisance parameters even when $C \neq 0$, and so offers a solution to the problem noted by Elliott (1998). In the numerical and empirical work in sections 3 and 4, the focus will be on instruments I1, as this appears to give better results.

6

Let $\tilde{\beta}$ denote the IV estimator of β, instrumenting x_t by z_t, i.e.

$$\tilde{\beta} = [\Sigma_{t=1}^{T} z_t^{\mu} x_t^{\mu'}]^{-1} \Sigma_{t=1}^{T} z_t^{\mu} y_t^{\mu}$$

where z_t^{μ}, x_t^{μ} and y_t^{μ} are the deviations of z_t, x_t and y_t from their respective sample means. With instruments I1,

$$T(\tilde{\beta} - \beta) \Rightarrow [\int_0^1 \zeta^{\mu}(r) H_C^{\mu}(r) dr]^{-1} \int_0^1 \zeta^{\mu}(r) dU_1(r) \tag{2.1}$$

while with instruments I2,

$$T(\tilde{\beta} - \beta) \Rightarrow [\int_0^1 Z^{\mu}(r) H_C^{\mu}(r) dr]^{-1} \int_0^1 Z^{\mu}(r) dU_1(r) \tag{2.2}$$

where $\zeta^{\mu}(r) = \zeta(r) - \int_0^1 \zeta(s) ds$, $Z^{\mu}(r) = Z(r) - \int_0^1 Z(s) ds$, $H_C^{\mu}(r) = H_C(r) - \int_0^1 H_C(s) ds$ and $H_C(r)$ is the Ornstein-Uhlenbeck process defined by the stochastic differential equation $dH_C(r) = C H_C(r) dr + dU_2(r)$. Accordingly, $\tilde{\beta}$ is a superconsistent estimator of β, but because the asymptotic distributions in equations 2.1 and 2.2 do not generally have mean zero, $\tilde{\beta}$ has $O(T^{-1})$ asymptotic bias as an estimator of the cointegrating vector (for $C = 0$, this was shown by Phillips and Hansen (1990)). This is also true of the dynamic OLS estimator $\hat{\beta}$ and some numerical simulations indicate that the biases of $\tilde{\beta}$ and $\hat{\beta}$ are of similar magnitudes. Point estimation is accordingly not the motivation for this IV estimator. I do not know of any estimator in this model which is asymptotically unbiased to $O(T^{-1})$, uniformly in C.

The F statistic testing the hypothesis that $\beta = \beta_0$, associated with $\tilde{\beta}$ is

$$F_{IV}(\beta_0) = (\tilde{\beta} - \beta_0)' V (\tilde{\beta} - \beta_0)$$

where $V = \tilde{\sigma}_1^{-2} [(\Sigma_{t=1}^T z_t^\mu x_t^{\mu'})^{-1} \Sigma_{t=1}^T z_t^\mu z_t^{\mu'} (\Sigma_{t=1}^T x_t^\mu z_t^{\mu'})^{-1}]^{-1}$,

$$\tilde{\sigma}_1^2 = \Sigma_{j=-l}^l \ (1 - \tfrac{|j|}{l+1}) \tilde{\gamma}(j),$$

$\tilde{\gamma}(j)$ is the sample autocovariance function of $\tilde{u}_{1t} = y_t^\mu - \tilde{\beta}' x_t^\mu$, and $l = o(T^{1/2})$. If $l = 0$, $\hat{\sigma}_1^2$ is simply the sample variance of $\tilde{u}_{1t}$, but more generally it is the estimator of the scaled zero-frequency spectral density proposed by Newey and West (1987), required because u_{1t} is serially dependent.

The null limiting distribution of $F_{IV}(\beta_0)$ is given in Theorem 1. The proofs of the Theorems are collected in the appendix.

Theorem 1: Under the null hypothesis $\beta = \beta_0$, as $T \to \infty$, $F_{IV}(\beta_0)$ converges to a χ^2 distribution on p degrees of freedom (with either instruments I1 or instruments I2).

Under the fixed alternative $\beta = \beta_1 \neq \beta_0$, F_{IV} diverges to infinity and so the test rejects with probability one, asymptotically.

It is important to note that Theorem 1 would not apply if there were more than p polynomial trends or artificially generated random walks being used as instruments. Instead, the F statistic would have a nonstandard null limiting distribution which would depend on nuisance parameters which are not consistently estimable. This is the reason why I focus

8

exclusively on the just identified case in this paper. Also, even in the just identified case, the null asymptotic distribution of an F or t statistic testing a hypothesis about one or more elements of β, but not the whole vector, is nonstandard (and, again, dependent on nuisance parameters which are not consistently estimable). Only the F statistic testing the hypothesis that the entire parameter vector β takes on a specified value has a χ^2 null limiting distribution, uniformly in C.

2.2 Confidence Sets for β

A researcher can form a confidence set for β with nominal coverage 1-α as $B(\alpha) = \{\beta_0 : F_{IV}(\beta_0) \leq \chi_p^2(1-\alpha)\}$, where $\chi_p^2(1-\alpha)$ is the upper 1-α percentile of the χ^2 distribution on p degrees of freedom. By construction, this confidence set is nonempty and convex[1]. It may be desired to construct a confidence set for one or more elements of β, but not for the whole vector. If β is partitioned as $\beta = (\beta_A', \beta_B')'$, then a confidence set for β_A is also available that has coverage of at least 1-α, asymptotically. This is simply given by $\{\beta_A : (\beta_A', \beta_B')' \in B(\alpha)$ for some $\beta_B\}$. Unlike the confidence set for the whole vector β, this confidence set is asymptotically conservative. It is easy to form a confidence interval for a scalar β_A in this way, without explicitly constructing the confidence set $B(\alpha)$. Partition V

---

[1]It is a maintained assumption throughout this paper that u_{1t} is stationary. If this were not so, then there would be a spurious correlation between z_t and u_{1t}, violating the instrument orthogonality requirement. The relationship between y_t and x_t would a spurious regression and the confidence set for β associated with the IV estimator would be nonempty, but would exclude the true parameter value w.p.1, asymptotically. An important advantage of the approach to robust inference in cointegrating systems proposed by Wright (1999), relative to that in the present paper, is that it does not maintain the stationarity of u_{1t}, but rather involves jointly testing the hypothesis of cointegration and the hypothesis that $\beta = \beta_0$.

and $\tilde{\beta}$ conformably with β as $\begin{pmatrix} V_{AA} & V_{AB} \\ V'_{AB} & V_{BB} \end{pmatrix}$ and $(\tilde{\beta}_A, \tilde{\beta}'_B)'$, respectively. Since $B(\alpha)$ can be written in the quadratic form

$$B(\alpha) = \{(\beta_A, \beta'_B)' : \ (\tilde{\beta}_A - \beta_A)^2 V_{AA} + 2(\tilde{\beta}_A - \beta_A)V_{AB}(\tilde{\beta}_B - \beta_B)$$

$$+ (\tilde{\beta}_B - \beta_B)' V_{BB}(\tilde{\beta}_B - \beta_B) \leq \chi_p^2(1 - \alpha)\},$$

the asymptotically conservative confidence interval for the scalar β_A is simply

$$(\tilde{\beta}_A + \inf_{\beta_B} \frac{-2V_{AB}(\tilde{\beta}_B - \beta_B) - \sqrt{[2V_{AB}(\tilde{\beta}_B - \beta_B)]^2 - 4V_{AA}[(\tilde{\beta}_B - \beta_B)'V_{BB}(\tilde{\beta}_B - \beta_B) - \chi_p^2(1-\alpha)]}}{2V_{AA}},$$

$$\tilde{\beta}_A + \sup_{\beta_B} \frac{-2V_{AB}(\tilde{\beta}_B - \beta_B) + \sqrt{[2V_{AB}(\tilde{\beta}_B - \beta_B)]^2 - 4V_{AA}[(\tilde{\beta}_B - \beta_B)'V_{BB}(\tilde{\beta}_B - \beta_B) - \chi_p^2(1-\alpha)]}}{2V_{AA}}),$$

which can be calculated by straightforward numerical optimization over the p-1 nuisance parameters in β_B. Other techniques are available for testing a hypothesis concerning the entire vector β, as discussed in Wright (1999) and, more briefly, below. In principle, a confidence set for β may be formed by inverting the acceptance region of any one of these tests. These confidence sets may be empty or nonconvex (unlike the confidence set for β proposed in this paper). An asymptotically conservative confidence set for any single element of β may be deduced by projecting out the other parameters. But (especially for large p), this is in practice a highly cumbersome approach involving calculating the underlying test statistic at each point in a p-dimensional grid. Because the method proposed in this paper involves constructing the confidence set for the entire vector β as the standard confidence ellipse associated with the IV estimator, the calculation of an asymptotically conservative

confidence set for a single element of β is greatly simplified. This is a significant practical advantage of the procedure proposed in this paper relative to other available approaches to inference in cointegrating systems which are robust to violations of the assumption of exact unit roots.

2.3 *Extensions to models with trends*

The only deterministic terms in the model as specified in this section are the constants μ_x and μ_y. This is without loss of generality. If μ_y were replaced by a linear time trend (a specification sometimes referred to as stochastic cointegration), then the linear time trend in instruments I1 would have to be replaced by another deterministic function of time, such as $\frac{t^{p+1}}{T^{p+1}}$. Additionally, the demeaned stochastic processes in equations 2.1 and 2.2 would also have to be detrended. But the result in Theorem 1 would be unaffected.

3. ALTERNATIVE METHODS AND LOCAL ASYMPTOTIC POWER

In this section, I consider the local asymptotic power of the proposed test, which I refer to henceforth as IV, and a number of alternatives. In this section, I consider the model with $p=1$, so that x_t is a scalar time series with an autoregressive root $a = 1 + \frac{c}{T}$. The autoregressive root a and local to unity parameter c are written in lower case to emphasize the fact that they are scalars. So the model specifies that

$$y_t = \mu_y + \beta x_t + u_{1t}$$

$$x_t = \mu_x + v_t, \ v_t = a v_{t-1} + u_{2t}, \ a = 1 + \tfrac{c}{T}.$$

Let $\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} = D(1)E(\varepsilon_t \varepsilon_t')D(1)'$ denote 2π times the spectral density of $u_t = (u_{1t}, u_{2t})'$. Let $U_1(r) = \sigma_1 V_1(r)$ and $U_2(r) = \sigma_2 V_2(r)$, where $V_1(r)$ and $V_2(r)$ are standard Brownian motions with correlation $\lambda = \frac{\sigma_{12}}{\sigma_1 \sigma_2}$ (a measure of the long-run endogeneity of the cointegrating regression). Also let $J_c(r)$ denote the scalar Ornstein-Uhlenbeck process defined by the stochastic differential equation $dJ_c(r) = c J_c(r)dr + dV_2(r)$ and note that $T^{-1/2} x_{[Tr]} \Rightarrow \sigma_2 J_c(r)$. Theorem 2 gives the asymptotic distribution of the test statistic $F_{IV}(\beta_0)$ under the sequence of local alternatives $\beta = \beta_0 + b/T$.

Theorem 2: Using instruments I1, under the sequence of local alternatives $\beta = \beta_0 + b/T$, as $T \to \infty$,

$$F_{IV}(\beta_0) \Rightarrow [\tfrac{b\sigma_2}{\sigma_1} \int_0^1 \zeta^\mu(r) J_c^\mu(r)dr + \int_0^1 \zeta^\mu(r)dV_1(r)]^2 / \int_0^1 \zeta^\mu(r)^2 dr$$

where $J_c^\mu(r) = J_c(r) - \int_0^1 J_c(s)ds$. The same result holds with instruments I2, replacing $\zeta^\mu(r)$ by $Z^\mu(r)$.

Wright (1999) lists a number of alternative procedures for testing a hypothesis concerning β and provides expressions for their local asymptotic power. All of these tests have nontrivial local asymptotic power against alternatives in a $1/T$ neighborhood of the hypothesized value of β, so no test has an asymptotic efficiency of zero relative to any other test.

One of these tests is dynamic OLS, which involves a t test of the coefficient on x_t in a regression of y_t on a constant, x_t and Δx_t and $o(T^{1/3})$ leads and lags thereof: I henceforth refer to this as the DOLS test. This does not control asymptotic size, as was first shown by Elliott (1998). Another procedure, proposed in Wright (1999), tests the hypothesis that $\beta = \beta_0$ by testing the implication that $y_t - \beta_0 x_t$ is stationary, using the test of Kwiatkowski et al. (1992). This test, like IV, has asymptotic size equal to the nominal level, uniformly in c. I henceforth refer to this as the KPSS test.

Two other more complicated procedures, described in Wright (1999), are both asymptotically conservative and have similar size and power properties. One is the size-adjusted parametric bootstrap of Stock and Watson (1996). The other involves first constructing S_c, a confidence interval for c, by the method proposed in Stock (1991). Imposing any value of c in S_c, a hypothesis concerning β can then be tested: if any of these tests leads to a rejection, then the null hypothesis is rejected. An upper bound for the asymptotic size of this procedure can be obtained by appeal to the Bonferroni inequality. This procedure was proposed by Cavanagh, Elliott and Stock (1995) in a closely related context. I henceforth refer to this as the BONF test. Generalizations of these two procedures to the case of 2 cointegrating regressors are cumbersome, while they are utterly impractical with more than 2 cointegrating regressors (see Stock and Watson (1996)).

A confidence set for β can be constructed using the size-adjusted parametric bootstrap, BONF, or KPSS tests, by calculating the test statistic at each point in a grid of hypothesized

values and inverting the acceptance region of the test. This first principles construction of a confidence set is circumvented by using the instrumental variables approach proposed in this paper.

The local asymptotic power functions of the IV, DOLS, KPSS and BONF procedures[2] are plotted in Figure 1 for $\sigma_1^2 = \sigma_2^2 = 1$, and various values of c and λ. The IV procedure uses the instruments I1 (results using the instruments I2 are not shown, but give somewhat lower power and are available from the author on request). The nominal size of the test is 5% in all cases[3]. For nonzero λ, the DOLS test has asymptotic size well above 5%, if $c \neq 0$. These massive size distortions mean that the DOLS test cannot be used for reliable inference in the presence of large roots that are not necessarily exact unit roots. The other tests have asymptotic size at or below the nominal level. None of these tests is dominant in terms of local asymptotic power, though the BONF test generally has higher power than the KPSS test, which in turn generally has higher power than IV. The local asymptotic power of all of these procedures falls as c becomes more negative.

---

[2]The local asymptotic power functions for the size-adjusted parametric bootstrap procedure of Stock and Watson (1996) are not shown, but are available in Wright (1999).

[3] Continuous stochastic processes were replaced by discrete approximations of length 500 and 5,000 replications were conducted in each simulation.

4. FINITE SAMPLE RESULTS

While the local asymptotic power of tests is of interest, the finite sample properties of the alternative methods of inference, and especially the finite sample coverage and width of confidence intervals are ultimately of most interest to applied researchers. Accordingly, I also simulated the effective coverage and median width of the alternative confidence intervals for β in the model

$$y_t = \mu_y + \beta x_t + u_{1t}$$

$$x_t = \mu_x + v_t, \; v_t = a v_{t-1} + u_{2t}$$

$$u_{1t} = \phi u_{1t-1} + \theta \varepsilon_{1t-1} + \varepsilon_{1t}, \; u_{2t} = \varepsilon_{2t}$$

where x_t is a scalar, $\mu_y = \mu_x = 0$, $(\varepsilon_{1t}, \; \varepsilon_{2t})'$ is an i.i.d. normal random vector with mean zero and covariance matrix $\begin{pmatrix} (\frac{1-\phi}{1+\theta})^2 & \frac{\lambda(1-\phi)}{1+\theta} \\ \frac{\lambda(1-\phi)}{1+\theta} & 1 \end{pmatrix}$, so that $\Sigma = \begin{pmatrix} 1 & \lambda \\ \lambda & 1 \end{pmatrix}$ and t=1,2,...T. The true value of β is zero. The errors in the cointegrating relationship are allowed to have serial correlation parameterized by ϕ and θ such that $|\phi| < 1$ and $|\theta| < 1$. If $a = 1$, then this is a standard model of cointegration, in which the variables have unit roots, but a may be slightly less than 1. For ϕ close to 1, there is slow mean-reversion to the cointegrating equilibrium.

The effective coverage and median width of the confidence intervals based on the IV, DOLS, KPSS and BONF tests are plotted against λ in Figures 2 and 3, respectively, for a number of values of ϕ, a and θ and a sample size $T=100$. The IV procedure uses the

15

instruments I1. Results using the instruments I2 instead (and for some other parameter configurations) are not shown, but are available from the author on request[4]. All confidence intervals have 95% nominal coverage. For DOLS and IV, the confidence intervals are simply given by the point estimate +/- 1.96 standard errors. For KPSS and BONF, the confidence intervals involve explicitly inverting the acceptance region of the test, over a grid of hypothesized values. The size of the underlying tests can of course be read off from these pictures: the size of the test is one minus the coverage of the confidence interval[5].

The DOLS method has coverage far below the nominal level for $a \neq 1$ and $\lambda \neq 0$. All the other tests control coverage well, except when ϕ is close to 1. The IV and BONF methods yield confidence intervals with similar coverage; both give generally higher coverage than KPSS. The DOLS method gives the confidence intervals with the shortest median width, but this is of little use in the light of the associated low coverage. For $a = 1$, the KPSS and IV methods have very similar median width, both are slightly shorter than the BONF confidence intervals. As a falls, these three confidence intervals all get wider, but this is tendency is most pronounced for the KPSS method and is least pronounced for the BONF method. Overall, there is little difference between BONF and IV in terms of median width. This is also true for KPSS if a is close to 1; otherwise KPSS gives somewhat wider confidence intervals.

---

[4]If the instruments I2 are used instead, the IV confidence intervals have slightly higher effective coverage and slightly higher median width.

[5] In these simulations, the number of leads and lags in the DOLS test, was set to 3 while all Bartlett kernel estimates used 2 autocovariances. 1,000 replications were conducted in each simulation.

In summary, the IV method can give inference that is more accurate than KPSS or BONF, while, in these simulations, it never gives much less accurate inference. Meanwhile IV is the simplest of these three methods, both conceptually and in terms of its practical implementation. Indeed, IV and KPSS are the only computationally feasible robust approaches to inference in the general model of section 2 when there are many regressors in the cointegrating system.

5. MONEY DEMAND: AN EMPIRICAL APPLICATION

A canonical application of cointegration analysis is to the estimation of long-run money demand. If there is a unit root in real output (in logs) and in the interest rate and if there is a stable demand for real money balances (in logs), then the log real money stock, log real output and the interest rate should be cointegrated. The cointegrating coefficients represent the output elasticity and interest semielasticity of long-run money demand. This cointegrating system has been estimated by Hoffman and Rasche (1991), Stock and Watson (1993) and many other authors.

But there is no consensus that income does have an exact unit root and a unit root specification represents an unreasonable population model for the interest rate. Income and interest rates are clearly very persistent, but it seems undesirable to assume that they have exact unit roots and to use methods of inference which are extremely non-robust to a violation of this assumption. Accordingly, I applied the methodology proposed in this paper

17

to form confidence sets for β_y, the output elasticity and β_r, the interest semielasticity in a regression of the log real money stock on log real output and the interest rate. I used a time series of annual data on the money stock, output, prices and interest rates[6] for the years 1900-1998 and also estimated these parameters by dynamic OLS (imposing exact unit roots in the regressors). Figure 4 shows the confidence ellipse for $(\beta_y, \beta_r)'$ using the method proposed in this paper, using instruments I1[7]. The DOLS confidence ellipse for this vector is also shown.

Stock and Watson (1996) and Wright (1999) used the inverse of the acceptance region of the size-adjusted bootstrap and KPSS tests to form confidence sets for $(\beta_y, \beta_r)'$ with annual money demand data. Both those papers obtained confidence sets which were larger than the DOLS confidence ellipses, but not so large as to be uninformative. This is similar to the results which I find, forming the confidence sets as proposed in this paper. The IV confidence ellipse is a little larger than the DOLS confidence ellipse, which does not control coverage in the presence of large but non-unit roots, but is certainly not so large as to be uninformative. Only negative values of the interest semielasticity are included within the

---

[6]The money stock from 1959-1998 is M1, from DATASTREAM (mnemonic USM1). Money stock data from 1900-1914 and 1915-1958 are from *Historical Statistics* (series X267) and Friedman and Schwartz (1970, Table 1), respectively. Real output is measured as the Real Net National Product, from NIPA for 1929-1998, spliced with earlier data obtained from Friedman and Schwartz (1982, Table 4.8). The price series is the net national product deflator, obtained from the same sources as real output. The interest rate is the 6 month commercial paper rate. From 1968-1998, this was obtained from DATASTREAM (mnemonic USCRBYLD); data from 1900-1967 are from Friedman and Schwartz (1982, Table 4.8).

[7]If the artificial random walk instruments (instruments I2) are used instead, the confidence ellipse depends on the seed for the random number generator, but is typically larger than the confidence ellipse using instruments I1. The dependence of the confidence ellipse on the seed is, of course, a practical disadvantage of instruments I2.

confidence sets. Only values of the income elasticity that are over 0.8 are included in the confidence sets, whereas some authors (such as Baba, Hendry and Starr (1992)) have argued that the income elasticity of money demand is around 0.5.

6. CONCLUSION

Standard methods for inference in cointegrating systems require all the variables to have exact unit roots and have been shown by Elliott (1998) to be not at all robust even to slight violations of this condition. In this paper, I have proposed testing a hypothesis concerning the entire cointegrating vector by using the F statistic associated with the just identified instrumental variables estimator, using deterministic polynomial trends or artificially generated random walks as instruments. I have shown that this method is robust to the regressors having roots that are not necessarily exactly equal to one, which are modeled as being local to unity. I have also compared the proposed method of inference with other alternatives which are robust to the regressors not having exact unit roots and found that none of the methods is dominant, but that the proposed method has good coverage and width properties. The approach proposed in this paper is however the simplest of these methods; this and the procedure in Wright (1999) are indeed the only computationally feasible robust approaches to inference in this model when there are many regressors in the cointegrating system.

APPENDIX: PROOF OF THE THEOREMS

I give the proofs of the Theorems for instruments I1 only, without loss of generality.

Proof of Theorem 1: If $\beta = \beta_0$,

$$F_{IV}(\beta_0) = (\tilde{\beta} - \beta_0)'V^{-1}(\tilde{\beta} - \beta_0) = \hat{\sigma}_1^{-2}\Sigma_{t=1}^{T}u_{1t}^{\perp}z_t^{\perp\prime}(\Sigma_{t=1}^{T}z_t^{\perp}z_t^{\perp\prime})^{-1}\Sigma_{t=1}^{T}z_t^{\perp}u_{1t}^{\perp}.$$

Because $\tilde{\beta} - \beta = O_p(T^{-1}), \hat{\sigma}_1^2 \rightarrow_p \sigma_1^2$ and so,

$$F_{IV}(\beta_0) \Rightarrow (\int_0^1 \zeta^{\mu}(r)dV_1(r))'(\int_0^1 \zeta^{\mu}(r)\zeta^{\mu}(r)')^{-1}\int_0^1 \zeta^{\mu}(r)dV_1(r),$$

with instruments I1. Since $\int_0^1 \zeta^{\mu}(r)dV_1(r)\tilde{\ }N(0, \int_0^1 \zeta^{\mu}(r)\zeta^{\mu}(r)')$, it follows that $F_{IV}(\beta_0)$ is asymptotically χ^2 on p degrees of freedom, completing the proof of Theorem 1.

Proof of Theorem 2: $\tilde{\beta} - \beta_0 = \frac{\Sigma_{t=1}^{T}z_t^{\perp}u_{1t}}{\Sigma_{t=1}^{T}z_t^{\perp}x_t^{\perp}} + \frac{b}{T}$, so

$$T\frac{\sigma_2}{\sigma_1}(\tilde{\beta} - \beta_0) = \frac{T^{-1}\sigma_1^{-1}\Sigma_{t=1}^{T}z_t^{\perp}u_{1t}}{T^{-2}\sigma_2^{-1}\Sigma_{t=1}^{T}z_t^{\perp}x_t^{\perp}} + \frac{b\sigma_2}{\sigma_1} \Rightarrow \frac{\int_0^1 \zeta^{\mu}(r)dV_1(r)}{\int_0^1 \zeta^{\mu}(r)J_c^{\mu}(r)dr} + \frac{b\sigma_2}{\sigma_1}.$$

Because $\tilde{\beta} - \beta = O_p(T^{-1}), \hat{\sigma}_1^2 \rightarrow_p \sigma_1^2$ and so,

$$T^2(\frac{\sigma_2}{\sigma_1})^2 V = \hat{\sigma}_1^2(\frac{\sigma_2}{\sigma_1})^2 \frac{T^{-2}\Sigma_{t=1}^{T}z_t^{\perp 2}}{(T^{-2}\Sigma_{t=1}^{T}z_t^{\perp}x_t^{\perp})^2} \Rightarrow \frac{\int_0^1 \zeta^{\mu}(r)^2 dr}{(\int_0^1 \zeta^{\mu}(r)J_c^{\mu}(r)dr)^2}$$

But $F_{IV}(\beta_0) = \frac{(\tilde{\beta} - \beta_0)^2}{V}$, so combining these convergence results completes the proof of the Theorem.

20

REFERENCES

Baba, Y., Hendry, D.F. and Starr, R.M. (1992): The Demand for M1 in the U.S.A., 1960-1988, *Review of Economic Studies*, 59, pp.25-61.

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

Cavanagh, C.L., Elliott, G. and Stock, J.H. (1995): Inference in Models with Nearly Integrated Regressors, *Econometric Theory*, 11, pp.1131-1147.

Dufour, J.M. (1997): Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models, *Econometrica*, 65, pp.1365-1387.

Elliott, G. (1998): On the Robustness of Cointegration Methods when Regressors Almost Have Unit Roots, *Econometrica,* 66, pp.149-158.

Friedman, M. and Schwartz, A.J. (1970): *Monetary Statistics of the United States*, Columbia University Press, New York.

——— (1982): *Monetary Trends in the United States and the United Kingdom*, University of Chicago Press, Chicago.

Hoffman, D. and Rasche, R.H. (1991): Long-Run Income and Interest Elasticities of Money Demand in the United States, *Review of Economics and Statistics*, 73, pp.665-674.

Johansen, S. (1988): Statistical Analysis of Cointegrating Vectors, *Journal of Economic Dynamics and Control*, 12, pp.231-254.

Kwiatkowski, D., Phillips, P.C.B., Schmidt, P. and Shin, Y. (1992): Testing the Null Hypoth-

esis of Stationarity Against the Alternative Hypothesis of a Unit Root: How Sure Are We That Economic Time Series Have a Unit Root?, *Journal of Econometrics*, 54, pp.158-178.

Lucas, R.E. (1988): Money Demand in the United States: A Quantitative Review, *Carnegie-Rochester Series on Public Policy*, 29, pp.137-168.

Newey, W.K. and K.D. West (1987): A Simple Positive Semidefinite Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, *Econometrica*, 55, pp.703-708.

Park, J.Y. (1990): Testing for Unit Roots and Cointegration by Variable Addition, in *Advances in Econometrics: Cointegration, Spurious Regressions and Unit Roots,* Vol 8, eds. T.B. Fomby and G.F. Rhodes, JAI Press, Greenwich.

Phillips, P.C.B. and Hansen, B.E. (1990): Statistical Inference in Instrumental Variables Regression with I(1) Processes, *Review of Economic Studies*, 58, pp.407-436.

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

Stock, J.H. (1995): Cointegration, Long-Run Comovements and Long Horizon Forecasting, in *Advances in Economics and Econometrics: Theory and Applications,* Vol 3, eds. K.F. Wallis and D.M. Kreps, Cambridge University Press, Cambridge.

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

——— (1996): Confidence Intervals in Cointegrating Regressions with Nearly Integrated Regressors, unpublished manuscript, Harvard University.

Wright, J.H. (1999): Confidence Sets for Cointegrating Coefficients Based on Stationarity Tests, *Journal of Business and Economic Statistics*, forthcoming.