Estimating Cointegrating Vectors Using Near Unit Root Variables

Aaron D. Smallwood[1]
University of Oklahoma

AND

Stefan C. Norrbin[2]
Florida State University

This Draft: December 2003

Abstract: This paper argues that the predominant method of estimating equilibrium relationships in macroeconometric models, namely the VECM system of Johansen, is severely flawed if the underlying variables are distributed as near unit root processes. Researchers may apply cointegration techniques to these processes, as the power of rejecting near unit roots using standard unit root tests is extremely low. Using Monte Carlo analysis, we find problematic behavior of cointegration analysis in detecting the true underlying form of the connection between the near unit root processes. Furthermore the connecting vector is imprecisely estimated, resulting in problematic inference for error correction models.

Keywords: near unit root processes; long memory; Johansen cointegration tests.

JEL classifications: C15, C22, C32

[1] Norman, OK 73019. E-mail: Aaron.Smallwood-1@ou.edu.
[2] Tallahassee, FL 32306. E-mail: snorrbin@coss.fsu.edu.

I. Introduction

The seminal article on cointegration by Engle and Granger (1987) resulted in a flurry of re-evaluations of relationships between macroeconomic variables. The most used cointegration methodology has been Johansen's (1988) maximum likelihood procedure.[1] This procedure is appealing to applied researchers as no theoretical structure is necessary. Most applied researchers use unit root tests to test for nonstationarity or assume the underlying distribution to be I(1). However if the underlying variables are nearly unit roots, then unit root tests would tend to incorrectly identify them as unit roots. Emerging research has shown that the cointegration methodology can be inappropriate when the variables are not strictly I(1). Andersson and Gredenhoff (1999) investigate the potential problems with Johansen's technique if unit root variables have a long memory fractional connection, and Gonzalo and Lee (1998) investigate the behavior of cointegration techniques for unrelated long memory variables, finding that the use of the Johansen technique results in spurious findings of equilibria.[2]

We extend the past literature by investigating the performance of cointegration tests under a more general methodology. This paper employs an encompassing approach to long memory variables, namely the GARMA approach, that has a differencing operator equal to $(1-2\eta L+L^2)^d$. The model accommodates unit root, fractional processes, and GARMA processes.[3] In addition, we explicitly analyze the performance of the estimated cointegrating vector using the Johansen procedure and the standard χ^2 inference.

The next section shows that if a researcher has particular interest in a cointegrating relationship they will very rarely detect cointegration with the proper cointegrating vector employing the Johansen techniques. The resulting variance of the estimates of the cointegrating vector in this environment can be excessive. Furthermore, we show that existing tests are likely to find cointegration among certain unrelated long memory processes. We conclude by commenting on the implications for estimating error correction models, and suggest some potential methods for detecting near unit roots.

II. Long Memory and Cointegration

We employ Monte Carlo techniques to assess the applicability of Johansen cointegration techniques in a bivariate system when the original variables are distributed as near unit root processes. The results are based on 5000 simulations for samples of 100, 200, and 500 observations. To conserve space we present only the results based on 200 observations, as this number is representative of the size available to most economic researchers. Following Robinson and Marinucci (2001) we generate a non-stationary long memory series y_t and a stationary residual series z_t. The series x_t is then constructed as $x_t=y_t+z_t$ resulting in a cointegrating vector between x_t and y_t of $[1\ -1]'$.[4] We also consider a specification where z_t follows an AR(1) process with an autoregressive coefficient equal to .7 (see Smallwood and Norrbin, 2003, for details on data generation).

Test and estimation results regarding the cointegrating vector for long memory variables with equilibrium relationships are depicted in Table 1. Throughout, we have used a GARMA specification employing the parameters η and d to represent the different cases. When a fractional model is employed, the value of η is equal to 1. Thus, the differencing parameter is half of its corresponding value in the fractional literature. The unit root process has an η/d pair given by 1.00/0.50. The last column, bottom row of the Table presents the unit root case.

The results show that for the unit root case the correct cointegrating vector is rejected about 6% of the time, slightly above the size for this experiment.[5] The rejection rates are acceptable when the equilibrium residual is an AR(1) process even among near unit root processes. In particular, for AR(1) residuals the rejection rates are close to the size of the test, and a reasonable RMSE of about 2-3% of the true vector.[6]

Ominous results occur, however, when the parent series is not strictly a unit root process and the residual series is not strictly I(0), especially for the case of a fractionally integrated parent series. For example, when the parent series and residual series are fractional processes with η/d pairs given by 1.00/0.45 and 1.00/0.10 respectively, we reject the correct null about the value of the cointegrating vector 87.48% of the time. It is very interesting to note the dramatic disparity between the size of the test relative to the case where the parent series is a unit root process. In particular, it appears clear that even slight deviations from the unit root assumption present serious problems with tests about the cointegrating vector. Interestingly, the results applied to GARMA processes do not present as bleak a picture, although the rejection rates of the true connecting vector are substantially larger than 5% in all cases except when the residual series is an AR(1) process.[7]

Table 1 also reports the median bias and RMSE associated with estimation of the second element of the cointegrating vector. The median bias tends to be relatively small and negative. On the other hand the RMSE can be excessive. Consider the specification in Table 1 where the η/d pair for the parent series is given by .999/.50, while the residual series has a value of d=.4. The RMSE for the second element in this vector is equal 20.0857, implying that the estimated cointegrating vector can be wildly erratic.[8] Such a RMSE translates to over 2000% error in this case, dramatically higher than the 80-90% reported by Anderson and Gredenhoff (1999) for fractional residuals. These results indicate that the use of the MLE approach of Johansen can produce estimates of the cointegrating vector that differ from the true value of the equilibrium vector.[9] Among other things, we feel this may help explain why researchers studying variables that are likely candidates for long memory have found erratic estimates for cointegrating vectors when employing the Johansen procedure.[10]

We also consider the consequences of employing cointegration analysis among unrelated near unit root variables in Table 2. For a given d, as the value of η rises, we see that the rejection rates for the null of no cointegration fall as the process moves parametrically closer to the unit root boundary. Even a process with an η–value slightly less than unity, a process with a very long cycle, leads to a very high rejection rate. Thus a process that closely resembles a unit root would lead the researcher to incorrectly reject in over 55% of the cases for an η/d pair of 0.9995/0.5. In contrast if η goes to unity then the test-statistics falsely indicate cointegration in only 5% of the cases. Also note that the trace-test rejects the null of no cointegration about 10%-30% more often than the eigenvalue-test in all near unit root GARMA cases.

The results of Table 2 show that cointegration tests will incorrectly indicate cointegration when there is none. GARMA processes that are extremely close to unit root processes are more likely to lead to problems than fractional processes. Furthermore a large RMSE results in the cases where the test incorrectly identified a cointegrating vector. Thus the results show that researchers will often incorrectly believe that a cointegrating vector exists, and estimate a vector that is very different from the expected vector.

III. Conclusions and Error Correction Mechanisms

This paper examines the response of testing procedures to near unit roots. Under a very general setting, we show that Johansen's MLE procedure can fail in two ways. First, even when a cointegrating vector is found among near unit root processes having a connection, it is unlikely to find the correct vectors. Secondly if a cointegrating vector does not exist the MLE procedure is likely to find a spurious connection. The implications of the results in this paper on the VECM estimation are important. If the cointegrating vector is misspecified then the VECM is also biased, and incorrect conclusions would be made regarding the short-run response.

The use of MLE methods should be used with caution when variables may be near unit roots. Unexpected cointegrating vectors might indicate a concern. There have been recent breakthroughs in unit root testing that allow one to directly test for a unit root against both long memory and stochastic regime switching processes in the context of a Dickey-Fuller type test (c.f.

Dolado, Gonzalo, and Mayoral, 2002 and Kapetanios, Shin, and Snell, 2003). As these procedures are relatively technical, we offer two potential pre-tests to determine if the problems described in the text are likely a concern. First, the researcher might use the Engle-Granger approach and compare the cointegrating vectors. If there is a substantial difference relative to the Johansen test then this might be evidence of a near unit root. The use of the Engle-Granger test in conjunction with the Johansen procedure has also been advocated by Gonzalo and Lee (1998), and similar to their results, we find that the Engle-Granger procedure is more robust to the near root cases described in the text. Finally, a large difference between the trace and the maximum eigenvalue tests might indicate the existence of near unit roots. In Table 2 the trace test consistently rejected the null more frequently than the maximum eigenvalue test. Because the trace test uses all the eigenvalues, the skewed distribution of the eigenvalues might indicate a potential near unit root case.

FOOTNOTES

1. To date over 1600 citation exists for this article, making it one of the most cited articles in the economic literature.

2. See also Abadir and Taylor (1999) and Elliott (1998).

3. For details on the GARMA model, the interested reader is referred to Chung (1996) and Bierens (2001). Diebold and Inoue (2001) show that the properties of certain regime switching processes are arbitrarily close to long memory. Thus, our results likely extend to a larger class of models, including those that are characterized by regime switching behavior.

4. Following Robinson and Marinucci (2001), we use the property that sum of a non-stationary series and a stationary series produces a non-stationary series by definition.

5. We do not report the rejection rate of the null of cointegration as this is generally consistent with the size of the tests for most cases in this table.

6. Tests on the form of the vector use an LR test and the standard χ^2 inference.

7. The results are conservative in that we considered a number of lags (up to 4) for both the trace and eigenvalue statistic, and report the results for the specification and test that provide the least inferential problems for the Johansen procedure.

8. Similarly to Gonzalo and Lee (1998) we also find that the inferential problem does not improve by a larger sample size.

9. We also conducted analysis using the Engle-Granger test and found that the estimated RMSE was significantly smaller for the same processes as the results reported here.

10. For example, Cheung and Lai (1993) test for a long run connection between the nominal exchange rate, and domestic and foreign prices. They find estimates of the elements of the cointegrating vectors ranging from 0.35 to 25.422 (in absolute value), substantially different from the theoretical expected value of 1.

REFERENCES

Abadir, K, Taylor, M. (1999) On the definitions of (co-)integration, *Journal of Time Series Analysis*, 20, 129-137.

Andersson, M.K, M.P. Gredenhoff, (1999) On the Maximum Likelihood Cointegration Procedure Under a Fractional Equilibrium Error, *Economics Letters*, 143-147.

Bierens, H. (2001) Complex Unit Roots and Business Cycles: Are They Real?, *Econometric Theory* 17, 962-983.

Cheung, Y-W, Lai, K.S. (1993) Long run purchasing power parity during the recent float, *Journal of International Economics*, 34, 181-192.

Chung, C-F. (1996) A generalized fractionally integrated autoregressive moving-average process, *Journal of Time Series Analysis* 17, 111-140.

Diebold, F.X., Inoue, A. (2001) Long memory and regime switching, *Journal of Econometrics* 105, 131-159.

Dolado, J., Gonzalo, J., Mayoral, L. (2002) A fractional Dickey Fuller test for unit roots, *Econometrica* **70**: 1963-2006.

Elliott, G. (1998) On the robustness of cointegration methods when regressors almost have unit roots, *Econometrica* **66**: 149-158.

Engle, R, Granger, C. (1987) Co-Integration and error correction: representation, estimation, and testing, *Econometrica* **55**: 251-76.

Gonzalo, J, Lee, T-H. (1998) Pitfalls in testing for long run relationships, *Journal of Econometrics* **86**: 129-54.

Johansen, S. (1988) Statistical analysis of cointegration vectors, *Journal of EconomicDynamics and Control* **12:** 231-54.

Kapetanios, G., Shin, Y., Snell, A. (2003) Testing for a unit root in the nonlinear STAR framework, *Journal of Econometrics* **112**: 359-379.

Robinson, P, Marinucci, D. (2001) Semiparametric fractional cointegration analysis, *Journal of Econometrics* **105**: 225-247.

Smallwood, A. Norrbin, S. (2003) Long memory processes, cointegration bias, and exchange rate dynamics, *University of Oklahoma Working Papers*: 1-32.

Table 1: **Nonstationary Long Memory Processes With a Stationary Residual**
(size=5%)

	d-value for residual process				
	0.4	**0.3**	**0.2**	**0.1**	**AR 1**
Y-Series Process					
η = 0.995, d = .50					
Rej. Rate α'=[1,-1]	0.3570	0.2828	0.1902	0.1200	0.0456
Median Bias of α'	-0.0596	-0.0113	-0.0022	-0.0005	-0.0006
RMSE of α'	25.5869	0.9461	0.4387	0.0763	0.0268
η = 0.997, d = .50					
Rej. Rate α'=[1,-1]	0.3134	0.2394	0.1994	0.1194	0.0486
Median Bias of α'	-0.0304	-0.0106	-0.0033	-0.0011	-0.0002
RMSE of α'	11.7640	0.2573	0.0356	0.0657	0.0332
η = 0.999, d = .50					
Rej. Rate α'=[1,-1]	0.2940	0.2394	0.1924	0.1316	0.0622
Median Bias of α'	0.0621	0.0127	0.0036	0.0011	0.0000
RMSE of α'	20.0857	1.2579	0.0287	0.0110	0.0175
η = 1.000, d = .40					
Rej. Rate α'=[1,-1]	N/A	N/A	0.8162	0.7910	0.0566
Median Bias of α'	N/A	N/A	-0.0017	0.0015	0.0000
RMSE of α'	N/A	N/A	1.3275	0.0221	0.0314
η = 1.000, d = .45					
Rej. Rate α'=[1,-1]	N/A	N/A	0.8402	0.8748	0.0590
Median Bias of α'	N/A	N/A	0.0013	0.0015	0.0000
RMSE of α'	N/A	N/A	0.5757	0.0113	0.0168
η = 1.000, d = .50					
Rej. Rate α'=[1,-1]	N/A	N/A	0.4932	0.2514	0.0622
Median Bias of α'	N/A	N/A	0.0008	0.0000	0.0003
RMSE of α'	N/A	N/A	0.1609	0.0232	0.0291

Notes: The distribution of the original variable is depicted in each subgroup. The residual series has the same value of η as the original series, while the top row denotes the value of d for the residual series. This d should be multiplied by 2 to be comparable to the d used in the fractional literature. N/A indicates not applicable because these relationships are nonstationary processes.

Table 2: **Long Memory Processes Without an Equilibrium Connection**
(size=5%)

	d-value of Y-series	η-value for Y-series 0.996	0.998	0.9995	1.0000
	0.40				
Rejection rate using Trace test		0.8780	0.6298	0.2986	0.0842
Rejection rate using Max. Eigenvalue		0.5486	0.3846	0.1894	0.0966
Median of α in [1-α]' for null rejected		-0.0002	0.0384	-0.0244	0.1175
RMSE of α in [1-α]' for null rejected		100.5057	71.7410	47.1178	10.1375
	0.45				
Rejection rate using Trace test		0.8656	0.6724	0.4698	0.0538
Rejection rate using Max. Eigenvalue		0.6204	0.4906	0.3810	0.0574
Median of α in [1-α]' for null rejected		0.0205	-0.0240	0.0077	0.0884
RMSE of α in [1-α]' for null rejected		174.0909	67.7188	552.4188	72.3194
	0.50				
Rejection rate using Trace test		0.8278	0.6902	0.5536	0.0546
Rejection rate using Max. Eigenvalue		0.6356	0.5424	0.4708	0.0524
Median of α in [1-α]' for null rejected		0.0099	-0.0199	-0.0063	-0.1676
RMSE of α in [1-α]' for null rejected		35.102	76.1652	23.4341	44.0146

Notes: Two unrelated variables have been generated having a value of η given by the top row of the table and a value of d shown in each subgroup. The d parameter should be multiplied by 2 to be comparable to the d parameter in the fractional literature.