

Testing for multicointegration

Tom Engsted^{a,*}, Jesus Gonzalo^b, Niels Haldrup^c

^a*Department of Finance, Aarhus School of Business, DK-8210 Aarhus V, Denmark*

^b*Department of Statistics and Econometrics, University of Carlos III, Madrid, 28903 Getafe, Spain*

^c*Department of Economics, University of Aarhus, DK-8000 Aarhus C, Denmark*

Received 23 January 1997; received in revised form 29 May 1997; accepted 19 June 1997

Abstract

We suggest how to redefine the multicointegration model of Granger and Lee (1990) in terms of an I(2) system and subsequently propose a one-step procedure for estimation and inference which will have favourable statistical properties compared to the two-step procedure suggested by Granger and Lee. With respect to the single equation residual based cointegration procedure for I(2) systems we tabulate new critical values that are necessary to accommodate the presence of deterministic components. © 1997 Elsevier Science S.A.

Keywords: Cointegration; Multicointegration; I(2) processes

JEL classification: C12; C13; C22

1. Introduction

It is widely recognized that many economic time series are non-stationary and contain stochastic trends such that the notion of cointegration provides the natural starting point for the analysis of such series, see Engle and Granger (1987). Most non-stationary time-series appear to be integrated of order one, I(1), that is, they only need to be differenced once in order to become stationary. However, recent research has shown that some economic variables (e.g. nominal variables like money, prices, wages, and some stock variables) may be better characterized as I(2) series, and various methods for analyzing systems containing I(2) variables have been developed, see e.g. Haldrup (1994), Johansen (1995), Kitamura (1995), Paruolo (1996), and Stock and Watson (1993).

A particular special case where the analysis of I(2) variables becomes relevant follows from the work by Granger and Lee (1989, 1990). They consider the case with two I(1) time series X_t and Y_t which cointegrate such that the linear-combination $Z_t = Y_t - \beta X_t$ is stationary I(0). By definition the cumulated error series $S_t = \sum_{j=1}^t Z_j$ is I(1) and hence it can happen that S_t cointegrates with X_t and/or Y_t , such that, say, $I_t = Y_t - \gamma S_t$ is I(0). This implies that essentially there are two levels of cointegration between just two I(1) time series. Granger and Lee denote this sort of cointegration

*Corresponding author. E-mail: tom@hha.dk

‘multicointegration’, and they demonstrate how such a phenomenon may arise in e.g. a standard cost minimizing framework. The particular economic example they give (and analyze empirically) is the case where Y_t and X_t are production and sales, respectively, such that for $\beta = 1$, Z_t is inventory investment whereby S_t becomes the level of inventories. Another example is provided by Lee (1992) where Y_t is new housing units started, X_t is new housing units completed, Z_t is uncompleted starts, and hence S_t is housing units under construction.

Multicointegration is an important property of the data that needs to be considered empirically. As emphasized by Lee (1992) and Engsted and Johansen (1997) multicointegration will invalidate usual procedures for estimation and testing in cointegrated systems, since the standard error correction model will be misspecified in this case. Naturally this will also have serious consequences for instance in forecasting and hypothesis testing.

I(2) cointegration is relevant for the analysis of multicointegrated time series since implicitly it involves the cumulation of I(1) variables which by definition are I(2). However, the statistical analysis proposed by Granger and Lee for multicointegrated series does not involve variables which are explicitly modelled as being I(2). In fact, they only consider the case where the cointegrating vector at the first level is known in advance, i.e. where $\beta = 1$, and, hence, does not need to be estimated. This implies that the statistical analysis only consists of investigating whether the directly observable I(1) variables X_t , Y_t and S_t are cointegrated. In some cases, however, β is not known beforehand and hence must be estimated. Standard consumption theory, for example, implies that consumption and income cointegrate such that savings are stationary. By cumulating savings we obtain wealth which according to the life-cycle hypothesis should cointegrate with consumption. Hence, there is multicointegration between consumption and income. However, in empirical work one often faces the problem of how to measure durable goods and the associated service flows. Campbell (1987) suggests to let the share of nondurable and service consumption to total consumption be determined as an unknown parameter which can be estimated in a cointegrating regression for income and nondurables and services consumption. Hence, if Y_t denotes income and X_t nondurables consumption, the parameter β in the first level of cointegration is unknown and must be estimated.

For this more general case where β is not known a priori, Granger and Lee propose the following two-step procedure. First, a super-consistent estimate of β is obtained in a regression containing Y_t and X_t . The residuals from this regression, \hat{Z}_t , then provide the estimated Z_t -series. In the second step the cumulated sum of \hat{Z}_t is generated and thus giving the \hat{S}_t series, which subsequently is regressed onto X_t and/or Y_t , resulting in a super-consistent estimate of γ , provided there is multicointegration.

In the present note we show that there are a number of important statistical problems associated with the two-step procedure suggested by Granger and Lee for the case where β has to be estimated, and we propose an alternative one-step procedure with favourable statistical properties. The procedure that we propose simultaneously tests both levels of cointegration by exploiting the fact that multicointegration implies I(2) cointegration in a particular way. The I(2) cointegration analysis of Johansen (1995) may be adopted in this case by considering amongst the model variables the cumulated series, $\sum_{j=1}^t X_j$ and $\sum_{j=1}^t Y_j$. Alternatively the single equation analysis for I(2) systems suggested by Haldrup (1994) may be used. In contrast to the two-step procedure proposed by Granger and Lee, the distributions concerning tests for multicointegration in a one-step procedure will be well-known. Furthermore, provided there is multicointegration, the cointegration parameter at the first level will be estimated at the super-super-consistent rate, $O_p(T^2)$, in the single-step procedure.

Another purpose of this note is to extend existing tables with critical values for the ADF

cointegration test to allow for a trend polynomial in cointegration regressions with both I(1) and I(2) variables. In the present context such a polynomial will arise naturally because if the variables in levels have a non-zero mean or a linear trend (which is most often the case in practice), then, by cumulation, a linear – or even a quadratic – trend will be generated.

2. Estimating and testing multicointegrated systems

Before proceeding it is worthwhile to give a formal definition of multicointegration which extends Granger and Lee (1990). To simplify the arguments we consider the case with a single cointegration vector at each level of cointegration. This will naturally be the case as in Granger and Lee (1989, 1990) where there are only two variables in the system. It is our belief that this is the situation that will most likely occur in practice. However, it is possible, of course, to extend the definition further to the case with multiple multicointegration relations, see Engsted and Johansen (1997).

Definition 1. Multicointegration. Assume that Y_t and the vector time series X_t are cointegrated time series of order $CI(1,1)$ such that $Y_t - \beta X_t = Z_t$ is stationary. X_t is assumed not to have elements that cointegrate. If the integral I(1)-variable $S_t = \Delta^{-1}Z_t = \sum_{j=1}^t Z_j$ cointegrates with X_t (or alternatively, Y_t) such that a parameter vector γ exists whereby $S_t - \gamma X_t$ is also a stationary relation, then Y_t and X_t are said to be multicointegrated. In this case the cointegrating relationship amongst the variables can be written as

$$S_t - \gamma X_t = \Delta^{-1}Y_t - \beta \Delta^{-1}X_t - \gamma X_t \sim I(0) \quad (1)$$

where $\Delta^{-1}Y_t = \sum_{j=1}^t Y_j$ and $\Delta^{-1}X_t = \sum_{j=1}^t X_j$ are now I(2) variables.

Since multicointegration implies cointegration between the cumulated cointegration errors at one level of cointegration with the original variables, the implied consequences w.r.t. estimation and testing are non-trivial. On the face of it there are two procedures to test for multicointegration that seem possible: A two-step procedure and a one-step procedure. Both single equation and multivariate procedures may be considered but our main focus is here on standard single equation residual based procedures.

2.1. A two step procedure

The two step procedure which has been suggested by Granger and Lee (1990) is based on the idea that first cointegration between Y_t and X_t is tested using standard cointegration techniques. If the series are accepted to be cointegrated, i.e. such that $\hat{Z}_t = Y_t - \hat{\beta}X_t$ is stationary, the regression residuals from this first step are cumulated as the new variable $\Delta^{-1}\hat{Z}_t = \sum_{j=1}^t \hat{Z}_j$. In the second step this variable is regressed on X_t and possible deterministics, e.g. an intercept and a time trend, in order to take account of non zero means and drifts of the series. Subsequently the integration order of the regression residuals from this second step regression is tested using the standard procedure. If the errors are I(0) the series are multicointegrated. Although this procedure seems plausible in principle, it appears to be

less attractive in practice. It can be shown¹ that the limiting theory to test for multicointegration is complicated by the fact that the auxiliary regression is based on cumulated regression residuals from another regression. This means that by using regression residuals in the second step the \hat{S}_T variable is bounded by the fact that $\hat{S}_T \equiv 0$ where T is the sample size. Standard methods to test for cointegration therefore become invalidated for this particular type of models, since the asymptotics will be expressed in terms of functionals of a Brownian bridge process rather than a Brownian motion process as is normally the case. Hence the size of standard residual based cointegration tests will be incorrect². Naturally the actual distributions and new critical values can be tabulated to account for the particular distributions. However, as we shall see in the sequel we opt for a one step procedure which may be preferred for several reasons.

2.2. A one step procedure

The basis for this procedure is to consider (1) as the benchmark for the joint estimation of the model parameters. Since the single series may potentially have a drift it is necessary to include deterministic components like a linear and a quadratic trend in the auxiliary integral regression³. Hence the cointegration regression reads

$$\Delta^{-1}Y_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \beta \Delta^{-1}X_t + \gamma X_t + u_t. \quad (2)$$

Alternatively, in place of the non-accumulated X_t series, it is possible to use Y_t as a regressor. The idea is now to test whether the errors u_t from the integral regression (2) follow an I(0) process (the case of multicointegration), an I(1) process (the case of first level cointegration but no multicointegration) and finally the case of an I(2) process where there is no cointegration amongst the variables.

If the variables exhibit multicointegration such that u_t is I(0) then it follows as a special case of Haldrup (1994) (Theorem 1, p. 160) that the least squares estimators $\hat{\gamma} = O_p(T^{-1})$ and $\hat{\beta} = O_p(T^{-2})$. Hence the cointegration parameters corresponding to the I(1) variables can be estimated super consistently whilst parameters associated with the generated I(2) regressors can be estimated super-super consistently. The latter result is especially interesting since the rapid rate of convergence cannot be achieved by using a two stage non-integral regression procedure.

A single equation residual based regression procedure for I(2) cointegration can now be conducted along the lines suggested in Haldrup (1994). The procedure is a straightforward generalisation of the Engle-Granger procedure developed for I(1) variables. Hence least squares regression residuals \hat{u}_t are first constructed from the regression (2) (possibly without a quadratic trend included), and subsequently the integration order can be tested by an augmented Dickey-Fuller t -ratio test, i.e. the t -statistic of ρ_0 from the regression

$$\Delta \hat{u}_t = \rho_0 \hat{u}_{t-1} + \sum_{j=1}^p \rho_j \hat{u}_{t-j} + v_t. \quad (3)$$

¹The technical details and a proof was included in an earlier version of this paper and can be obtained from the authors upon request.

²Note that if the cointegration vector of the first step is given beforehand, no estimation is involved and the second stage of the procedure is perfectly valid. In many cases this is likely to be the case. Our critique only applies when the first step cointegration vector is unknown.

³When a drift is absent in the series it suffices to include a time trend to account for the initial conditions.

The distribution of the Dickey-Fuller t -test is derived and reported in Haldrup (1994) for the situation where at least cointegration at the first level is found. In other words, it is assumed a priori that Y_t and X_t are CI(1,1), or alternatively, $\Delta^{-1}Y_t$ and $\Delta^{-1}X_t$ are CI(2,1). Under the null there is no further cointegration (i.e. no multicointegration in the present set up) and hence u_t is I(1). The distributions are found to depend upon both the number of I(1) regressors in the model, m_1 , and the number of I(2) regressors, m_2 . The critical values for this case are reported in the above reference for the situation where a constant is included in the regression. In Tables 1 and 2 in the present note the

Table 1
Linear trend case

m_1	T	Probability of a smaller value							
		$m_2=1$				$m_2=2$			
		0.01	0.025	0.05	0.10	0.01	0.025	0.05	0.10
0	25	-5.21	-4.72	-4.29	-3.88	-5.81	-5.25	-4.83	-4.41
	50	-4.66	-4.33	-4.01	-3.67	-5.14	-4.77	-4.45	-4.10
	100	-4.55	-4.18	-3.90	-3.59	-4.93	-4.56	-4.31	-3.98
	250	-4.41	-4.08	-3.83	-3.51	-4.81	-4.49	-4.20	-3.91
	500	-4.33	-4.04	-3.78	-3.49	-4.75	-4.42	-4.14	-3.84
1	25	-5.60	-5.10	-4.71	-4.30	-6.24	-5.68	-5.21	-4.80
	50	-5.11	-4.70	-4.42	-4.08	-5.62	-5.22	-4.89	-4.51
	100	-4.85	-4.54	-4.26	-3.94	-5.23	-4.90	-4.62	-4.29
	250	-4.73	-4.43	-4.19	-3.89	-5.11	-4.77	-4.50	-4.20
	500	-4.73	-4.42	-4.15	-3.87	-5.05	-4.74	-4.48	-4.18
2	25	-6.09	-5.57	-5.14	-4.69	-6.70	-6.17	-5.70	-5.22
	50	-5.47	-5.07	-4.74	-4.38	-5.98	-5.53	-5.17	-4.79
	100	-5.21	-4.86	-4.58	-4.26	-5.59	-5.19	-4.93	-4.62
	250	-5.07	-4.79	-4.51	-4.20	-5.35	-5.07	-4.80	-4.51
	500	-5.00	-4.73	-4.48	-4.18	-5.34	-5.02	-4.75	-4.46
3	25	-6.47	-5.95	-5.53	-5.08	-7.19	-6.63	-6.08	-5.89
	50	-5.89	-5.43	-5.13	-4.76	-6.23	-5.81	-5.48	-5.12
	100	-5.52	-5.18	-4.91	-4.59	-5.97	-5.58	-5.25	-4.92
	250	-5.38	-5.05	-4.78	-4.74	-5.69	-5.37	-5.07	-4.80
	500	-5.34	-5.04	-4.78	-4.50	-5.67	-5.33	-5.06	-4.76
4	25	-6.95	-6.37	-5.90	-5.44	-7.61	-6.93	-6.43	-5.91
	50	-6.35	-5.85	-5.47	-5.10	-6.64	-6.18	-5.82	-5.41
	100	-5.86	-5.49	-5.20	-4.89	-6.09	-5.76	-5.50	-5.16
	250	-5.66	-5.35	-5.08	-4.77	-5.95	-5.61	-5.34	-5.04
	500	-5.63	-5.31	-5.06	-4.76	-5.92	-5.56	-5.29	-5.02

Critical values for the cointegration ADF-test allowing for I(2) variables. An intercept plus a trend have been included in the cointegration regression. The indices m_1 and m_2 indicate the number of I(1) and I(2) variables, respectively, on the right hand side of the cointegration regression. The left hand side variable is an I(2) variable. T indicates the sample size.

Note: It is assumed a priori that all the I(2) variables of the model cointegrate into an I(1) relation, so under the null hypothesis residuals are I(1) and under the alternative these are I(0) indicating multicointegration.

The standard errors of the fractiles vary, but generally they lie in the interval {0.01–0.03}. The simulations were based upon 50 000 replications.

Table 2
Quadratic trend case

m_1	T	Probability of a smaller value							
		$m_2 = 1$				$m_2 = 2$			
		0.01	0.025	0.05	0.10	0.01	0.025	0.05	0.10
0	25	-5.77	-5.28	-4.86	-4.43	-6.44	-5.85	-5.42	-4.96
	50	-5.20	-4.81	-4.47	-4.12	-5.61	-5.21	-4.88	-4.52
	100	-4.94	-4.60	-4.32	-4.00	-5.33	-4.97	-4.67	-4.34
	250	-4.77	-4.47	-4.21	-3.92	-5.13	-4.79	-4.52	-4.23
	500	-4.73	-4.43	-4.17	-3.88	-5.07	-4.76	-4.50	-4.21
1	25	-6.21	-5.69	-5.27	-4.83	-6.85	-6.30	-5.82	-5.33
	50	-5.56	-5.16	-4.83	-4.47	-5.99	-5.58	-5.22	-4.86
	100	-5.29	-4.93	-4.64	-4.32	-5.63	-5.27	-4.98	-4.65
	250	-5.11	-4.79	-4.52	-4.23	-5.43	-5.09	-4.84	-4.54
	500	-5.05	-4.75	-4.49	-4.20	-5.35	-5.05	-4.78	-4.49
2	25	-6.66	-6.10	-5.65	-5.20	-7.32	-6.68	-6.21	-5.69
	50	-5.92	-5.50	-5.17	-4.82	-6.35	-5.90	-5.54	-5.16
	100	-5.57	-5.23	-4.95	-4.63	-5.90	-5.54	-5.25	-4.92
	250	-5.42	-5.08	-4.82	-4.52	-5.69	-5.37	-5.10	-4.80
	500	-5.36	-5.04	-4.77	-4.48	-5.61	-5.29	-5.04	-4.76
3	25	-7.12	-6.51	-6.05	-5.55	-7.68	-7.06	-6.55	-6.03
	50	-6.27	-5.85	-5.50	-5.12	-6.63	-6.23	-5.86	-5.46
	100	-5.90	-5.54	-5.25	-4.91	-6.19	-5.85	-5.55	-5.22
	250	-5.71	-5.38	-5.11	-4.81	-5.96	-5.64	-5.37	-5.07
	500	-5.60	-5.30	-5.04	-4.76	-5.85	-5.55	-5.30	-5.02
4	25	-7.61	-6.93	-6.43	-5.91	-8.18	-7.47	-6.93	-6.38
	50	-6.56	-6.15	-5.79	-5.41	-7.00	-6.55	-6.16	-5.76
	100	-6.18	-5.81	-5.52	-5.19	-6.47	-6.10	-5.80	-5.47
	250	-5.96	-5.64	-5.36	-5.05	-6.21	-5.87	-5.60	-5.31
	500	-5.87	-5.57	-5.30	-5.01	-6.12	-5.80	-5.54	-5.26

Critical values for the cointegration ADF-test allowing for I(2) variables. An intercept, a time trend and a quadratic time trend have been included in the cointegration regression. The indices m_1 and m_2 indicate the number of I(1) and I(2) variables, respectively, on the right hand side of the cointegration regression. The left hand side variable is an I(2) variable. T indicates the sample size.

Note: It is assumed a priori that all the I(2) variables of the model cointegrate into an I(1) relation, so under the null hypothesis residuals are I(1) and under the alternative these are I(0) indicating multicointegration.

The standard errors of the fractiles vary, but generally they lie in the interval {0.01–0.03}. The simulations were based upon 50 000 replications.

critical values are extended to the case where, respectively a trend and a trend plus a quadratic trend have been included in the regression as well, i.e. various versions of (2).

Another procedure to adopt is to consider the I(2) procedure for gaussian VAR models advocated by Johansen (1995), where $\sum_{j=1}^t X_j$ and $\sum_{j=1}^t Y_j$ are included amongst the variables of the VAR model. The procedure is straightforward and naturally is more general than the single equation procedure

described above since several multicointegrating relations may be tested for, at least in principle. See also Engsted and Johansen (1997) where a discussion of gaussian VAR models for $I(2)$ variables and the relation to multicointegration is given.

3. Conclusion

Granger and Lee (1990) have suggested how a deeper form of cointegration, multicointegration, may potentially occur for time series integrated of order one when the cumulated errors at one level of cointegration cointegrate with the original variables. Testing for multicointegration is important since the standard error correction model will be misspecified if multicointegration is present and hence will have serious implications for e.g. estimation, inference, and forecasting. Granger and Lee propose a two-stage procedure to test for multicointegration. In this note we suggest how to redefine the multicointegration model of Granger and Lee (1990) in terms of an $I(2)$ system and subsequently propose a one-step procedure for estimation and inference which will have favourable statistical properties compared to the two-step procedure. The problem with the Granger and Lee two step procedure, when the cointegration vector is unknown, is that in their second stage they suggest using cumulated regression errors in a cointegration regression. As we have shown this will invalidate the use of standard tables for cointegration inference since the limiting behaviour of the cumulated series will be a Brownian bridge, rather than a Brownian motion process. The advantage of the one stage procedure that we opt for is that some of the cointegrating parameters, i.e. those associated with generated $I(2)$ variables, can be estimated at the $O_p(T^2)$ consistent rate, – a property that should naturally be exploited in estimation. Both the Johansen multivariate $I(2)$ procedure and a single equation procedure may be adopted (given there is only a single multicointegrating relation). With respect to the single equation residual based cointegration procedure for $I(2)$ systems we tabulate new critical values that are necessary in order to accommodate the presence of deterministic components.

Acknowledgements

We would like to thank Jörg Breitung, Clive Granger, Søren Johansen, Kenneth Kasa and Jesper Lund for helpful comments and suggestions.

References

- Campbell, J.Y., 1987. Does saving anticipate declining labour income? An alternative test of the permanent income hypothesis. *Econometrica* 55, 1249–1273.
- Engle, R.F., Granger, C.W.J., 1987. Co-integration and error correction: Representation, estimation and testing. *Econometrica* 55 (2), 251–276.
- Engsted, T., Johansen, S., 1997. Granger's representation theorem and multicointegration. EUI working paper, ECO No. 97/15, European University Institute, Florence.
- Granger, C.W.J., Lee, T., 1989. Investigation of production, sales and inventory relations using multicointegration and non-symmetric error correction models. *Journal of Applied Econometrics* 4 (supplement), S145–S159.

- Granger, C.W.J., and Lee, T., 1990. Multicointegration. In: Rhodes, G.F., Fomby, T.B. (Eds.), *Advances in Econometrics*, vol. 8. JAI Press, Greenwich, CT, pp. 71–84.
- Haldrup, N., 1994. The asymptotics of single-equation cointegration regressions with I(1) and I(2) variables. *Journal of Econometrics* 63, 153–181.
- Johansen, S., 1995. A statistical analysis of cointegration for I(2) variables. *Econometric Theory* 11, 25–59.
- Kitamura, Y., 1995. Estimation of Cointegrated Systems with I(2) Processes. *Econometric Theory* 11, 1–24.
- Lee, T.-H., 1992. Stock-flow relationships in housing construction. *Oxford Bulletin of Economics and Statistics* 54, 419–430.
- Paruolo, P., 1996. On the determination of integration indices in I(2) systems. *Journal of Econometrics* 72, 313–356.
- Stock, J.H., Watson, M.W., 1993. A simple estimator of cointegrating vectors in higher order integrated systems. *Econometrica* 61, 783–820.