



Cointegration

Cointegration is a statistical property of a collection (X_1, X_2, \dots, X_k) of time series variables. First, all of the series must be integrated of order d . Next, if a linear combination of this collection is integrated of order less than d , then the collection is said to be co-integrated. Formally, if (X, Y, Z) are each integrated of order d , and there exist coefficients a, b, c such that $aX + bY + cZ$ is integrated of order less than d , then X , Y , and Z are cointegrated. Cointegration has become an important property in contemporary time series analysis. Time series often have trends—either deterministic or stochastic. In an influential paper,^[1] Charles Nelson and Charles Plosser (1982) provided statistical evidence that many US macroeconomic time series (like GNP, wages, employment, etc.) have stochastic trends.

Introduction

If two or more series are individually integrated (in the time series sense) but some linear combination of them has a lower order of integration, then the series are said to be cointegrated. A common example is where the individual series are first-order integrated (**$I(1)$**) but some (cointegrating) vector of coefficients exists to form a stationary linear combination of them.

History

The first to introduce and analyse the concept of spurious—or nonsense—regression was Udny Yule in 1926.^[2] Before the 1980s, many economists used linear regressions on non-stationary time series data, which Nobel laureate Clive Granger and Paul Newbold showed to be a dangerous approach that could produce spurious correlation,^[3] since standard detrending techniques can result in data that are still non-stationary.^[4] Granger's 1987 paper with Robert Engle formalized the cointegrating vector approach, and coined the term.^[5]

For integrated **$I(1)$** processes, Granger and Newbold showed that de-trending does not work to eliminate the problem of spurious correlation, and that the superior alternative is to check for co-integration. Two series with **$I(1)$** trends can be co-integrated only if there is a genuine relationship between the two. Thus the standard current methodology for time series regressions is to check all time series involved for integration. If there are **$I(1)$** series on both sides of the regression relationship, then it is possible for regressions to give misleading results.

The possible presence of cointegration must be taken into account when choosing a technique to test hypotheses concerning the relationship between two variables having unit roots (i.e. integrated of at least order one).^[3] The usual procedure for testing hypotheses concerning the relationship between non-stationary variables was to run ordinary least squares (OLS) regressions on data which had been differenced. This method is biased if the non-stationary variables are cointegrated.

For example, regressing the consumption series for any country (e.g. Fiji) against the GNP for a randomly selected dissimilar country (e.g. Afghanistan) might give a high R-squared relationship (suggesting high explanatory power on Fiji's consumption from Afghanistan's GNP). This is called spurious regression:

two integrated $I(1)$ series which are not directly causally related may nonetheless show a significant correlation.

Tests

The six main methods for testing for cointegration are:

Engle–Granger two-step method

If x_t and y_t both have order of integration $d=1$ and are cointegrated, then a linear combination of them must be stationary for some value of β and u_t . In other words:

$$y_t - \beta x_t = u_t$$

where u_t is stationary.

If β is known, we can test u_t for stationarity with an Augmented Dickey–Fuller test or Phillips–Perron test. If β is unknown, we must first estimate it. This is typically done by using ordinary least squares (by regressing y_t on x_t and an intercept). Then, we can run an ADF test on u_t . However, when β is estimated, the critical values of this ADF test are non-standard, and increase in absolute value as more regressors are included.^[6]

If the variables are found to be cointegrated, a second-stage regression is conducted. This is a regression of Δy_t on the lagged regressors, Δx_t and the lagged residuals from the first stage, \hat{u}_{t-1} . The second stage regression is given as: $\Delta y_t = \Delta x_t b + \alpha u_{t-1} + \varepsilon_t$

If the variables are not cointegrated (if we cannot reject the null of no cointegration when testing u_t), then $\alpha = 0$ and we estimate a differences model: $\Delta y_t = \Delta x_t b + \varepsilon_t$

Johansen test

The Johansen test is a test for cointegration that allows for more than one cointegrating relationship, unlike the Engle–Granger method, but this test is subject to asymptotic properties, i.e. large samples. If the sample size is too small then the results will not be reliable and one should use Auto Regressive Distributed Lags (ARDL).^{[7][8]}

Phillips–Ouliaris cointegration test

Peter C. B. Phillips and Sam Ouliaris (1990) show that residual-based unit root tests applied to the estimated cointegrating residuals do not have the usual Dickey–Fuller distributions under the null hypothesis of no-cointegration.^[9] Because of the spurious regression phenomenon under the null hypothesis, the distribution of these tests have asymptotic distributions that depend on (1) the number of deterministic trend terms and (2) the number of variables with which co-integration is being tested. These

distributions are known as Phillips–Ouliaris distributions and critical values have been tabulated. In finite samples, a superior alternative to the use of these asymptotic critical value is to generate critical values from simulations.

Multicointegration

In practice, cointegration is often used for two $I(1)$ series, but it is more generally applicable and can be used for variables integrated of higher order (to detect correlated accelerations or other second-difference effects). **Multicointegration** extends the cointegration technique beyond two variables, and occasionally to variables integrated at different orders.

Variable shifts in long time series

Tests for cointegration assume that the cointegrating vector is constant during the period of study. In reality, it is possible that the long-run relationship between the underlying variables change (shifts in the cointegrating vector can occur). The reason for this might be technological progress, economic crises, changes in the people's preferences and behaviour accordingly, policy or regime alteration, and organizational or institutional developments. This is especially likely to be the case if the sample period is long. To take this issue into account, tests have been introduced for cointegration with one unknown structural break,^[10] and tests for cointegration with two unknown breaks are also available.^[11]

Bayesian inference

Several Bayesian methods have been proposed to compute the posterior distribution of the number of cointegrating relationships and the cointegrating linear combinations.^[12]

See also

- [Error correction model](#)
- [Granger causality](#)
- [Stationary subspace analysis](#)

References

1. Nelson, C.R; Plosser, C.I (1982). "Trends and random walks in macroeconomic time series". *Journal of Monetary Economics*. **10** (2): 139–162. doi:10.1016/0304-3932(82)90012-5 ([http s://doi.org/10.1016%2F0304-3932%2882%2990012-5](http://doi.org/10.1016%2F0304-3932%2882%2990012-5)).
2. Yule, U. (1926). "Why do we sometimes get nonsense-correlations between time series? - A study in sampling and the nature of time series". *Journal of the Royal Statistical Society*. **89** (1): 11–63. doi:10.2307/2341482 (<https://doi.org/10.2307%2F2341482>). JSTOR 2341482 ([ht tps://www.jstor.org/stable/2341482](https://www.jstor.org/stable/2341482)). S2CID 126346450 ([https://api.semanticscholar.org/Cor pusID:126346450](https://api.semanticscholar.org/CorpusID:126346450)).
3. Granger, C.; Newbold, P. (1974). "Spurious Regressions in Econometrics". *Journal of Econometrics*. **2** (2): 111–120. CiteSeerX 10.1.1.353.2946 ([https://citeseerx.ist.psu.edu/view doc/summary?doi=10.1.1.353.2946](https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.353.2946)). doi:10.1016/0304-4076(74)90034-7 (<https://doi.org/10.1016%2F0304-4076%2874%2990034-7>).

4. Granger, Clive (1981). "Some Properties of Time Series Data and Their Use in Econometric Model Specification". *Journal of Econometrics*. **16** (1): 121–130. doi:10.1016/0304-4076(81)90079-8 (<https://doi.org/10.1016%2F0304-4076%2881%2990079-8>).
5. Engle, Robert F.; Granger, Clive W. J. (1987). "Co-integration and error correction: Representation, estimation and testing" (http://pe.cemi.rssi.ru/pe_2015_3_106-135.pdf) (PDF). *Econometrica*. **55** (2): 251–276. doi:10.2307/1913236 (<https://doi.org/10.2307%2F1913236>). JSTOR 1913236 (<https://www.jstor.org/stable/1913236>).
6. https://www.econ.queensu.ca/sites/econ.queensu.ca/files/wpaper/qed_wp_1227.pdf
7. Giles, David (19 June 2013). "ARDL Models - Part II - Bounds Tests" (<http://davegiles.blogspot.com/2013/06/ardl-models-part-ii-bounds-tests.html>). Retrieved 4 August 2014.
8. Pesaran, M.H.; Shin, Y.; Smith, R.J. (2001). "Bounds testing approaches to the analysis of level relationships". *Journal of Applied Econometrics*. **16** (3): 289–326. doi:10.1002/jae.616 (<https://doi.org/10.1002%2Fjae.616>). hdl:10983/25617 (<https://hdl.handle.net/10983%2F25617>).
9. Phillips, P. C. B.; Ouliaris, S. (1990). "Asymptotic Properties of Residual Based Tests for Cointegration" (<https://web.archive.org/web/20210918174602/https://cowles.yale.edu/sites/default/files/files/pub/d08/d0847-r.pdf>) (PDF). *Econometrica*. **58** (1): 165–193. doi:10.2307/2938339 (<https://doi.org/10.2307%2F2938339>). JSTOR 2938339 (<https://www.jstor.org/stable/2938339>). Archived from the original (<http://cowles.yale.edu/sites/default/files/files/pub/d08/d0847-r.pdf>) (PDF) on 2021-09-18. Retrieved 2019-12-14.
10. Gregory, Allan W.; Hansen, Bruce E. (1996). "Residual-based tests for cointegration in models with regime shifts" (https://www.econ.queensu.ca/sites/econ.queensu.ca/files/qed_wp_862.pdf) (PDF). *Journal of Econometrics*. **70** (1): 99–126. doi:10.1016/0304-4076(96)00165-7 (<https://doi.org/10.1016%2F0304-4076%2869%2941685-7>).
11. Hatemi-J, A. (2008). "Tests for cointegration with two unknown regime shifts with an application to financial market integration" (<https://ideas.repec.org/a/spr/empeco/v35y2008i3p497-505.html>). *Empirical Economics*. **35** (3): 497–505. doi:10.1007/s00181-007-0175-9 (<https://doi.org/10.1007%2Fs00181-007-0175-9>). S2CID 153437469 (<https://api.semanticscholar.org/CorpusID:153437469>).
12. Koop, G.; Strachan, R.; van Dijk, H.K.; Villani, M. (January 1, 2006). "Chapter 17: Bayesian Approaches to Cointegration". In Mills, T.C.; Patterson, K. (eds.). *Handbook of Econometrics Vol.1 Econometric Theory* (<https://strathprints.strath.ac.uk/7750/>). Palgrave Macmillan. pp. 871–898. ISBN 978-1-4039-4155-8.

Further reading

- Enders, Walter (2004). "Cointegration and Error-Correction Models" (https://archive.org/details/applieconometr00ende_0). *Applied Econometrics Time Series* (https://archive.org/details/applieconometr00ende_0/page/319) (Second ed.). New York: Wiley. pp. 319–386 (https://archive.org/details/applieconometr00ende_0/page/319). ISBN 978-0-471-23065-6.
- Hayashi, Fumio (2000). *Econometrics* (https://archive.org/details/econometrics00haya_012). Princeton University Press. pp. 623 (https://archive.org/details/econometrics00haya_012/page/n639)–669. ISBN 978-0-691-01018-2.
- Maddala, G. S.; Kim, In-Moo (1998). *Unit Roots, Cointegration, and Structural Change* (<https://books.google.com/books?id=llXBvougICMC&pg=PA155>). Cambridge University Press. pp. 155–248. ISBN 978-0-521-58782-2.
- Murray, Michael P. (1994). "A Drunk and her Dog: An Illustration of Cointegration and Error Correction" (<http://www-stat.wharton.upenn.edu/~steele/Courses/434/434Context/Co-integration/Murray93DrunkAndDog.pdf>) (PDF). *The American Statistician*. **48** (1): 37–39. doi:10.1080/00031305.1994.10476017 (<https://doi.org/10.1080%2F00031305.1994.10476017>). An intuitive introduction to cointegration.

