# Cointegration and Antitrust: A Primer

Jonathan L. Rubin, J.D., Ph.D.[*]
*American Antitrust Institute*

## Introduction

On October 8, 2003, Robert F. Engle and Clive W. J. Granger were awarded the Nobel Prize for their research on the statistical analysis of economic time series. Both made important contributions on their own, but their most influential work by far is contained in a short and elegant paper they published together in 1987.[1] Their paper influenced the way statisticians perform almost all regression analysis.

Their insight, known as cointegration, has been described as a method of uncovering long-run relationships between variables that are concealed by the noise of short-term fluctuations. An engineer might look at this as disentangling the "signal" from the "noise." An economist could consider it a way of distinguishing between a random fluctuation and a correction back to an equilibrium level. A statistician would regard it as a way of doing regression analysis on non-stationary (*i.e.,* stochastically trending) variables that gives statistically valid results.

It is this last (and most technical) aspect of cointegration which accounts for its influence in the econometric world. Cointegration methods will inevitably make their way into the statistical analysis of antitrust issues and, ultimately, into the courtroom. The purpose of this article is to introduce the intuition behind cointegration in the context of antitrust econometrics. The focus will be the multivariate cointegration model pioneered by Johansen.[2]

## The Nature of Time Series and the Problem of Spurious Regression

Econometric studies relevant to antitrust issues are often concerned with time series, *i.e.*, a list of $n$ sequential observations, $X_t = \{x_1, x_2, x_3, ..., x_n\}$ of a particular variable that varies over time. The graph of a typical price series is given in Fig. 1, which shows the price of a particular variety and grade of lumber over a 22-year period.

## Figure 1: Real Price of Lumber, 1975-1996
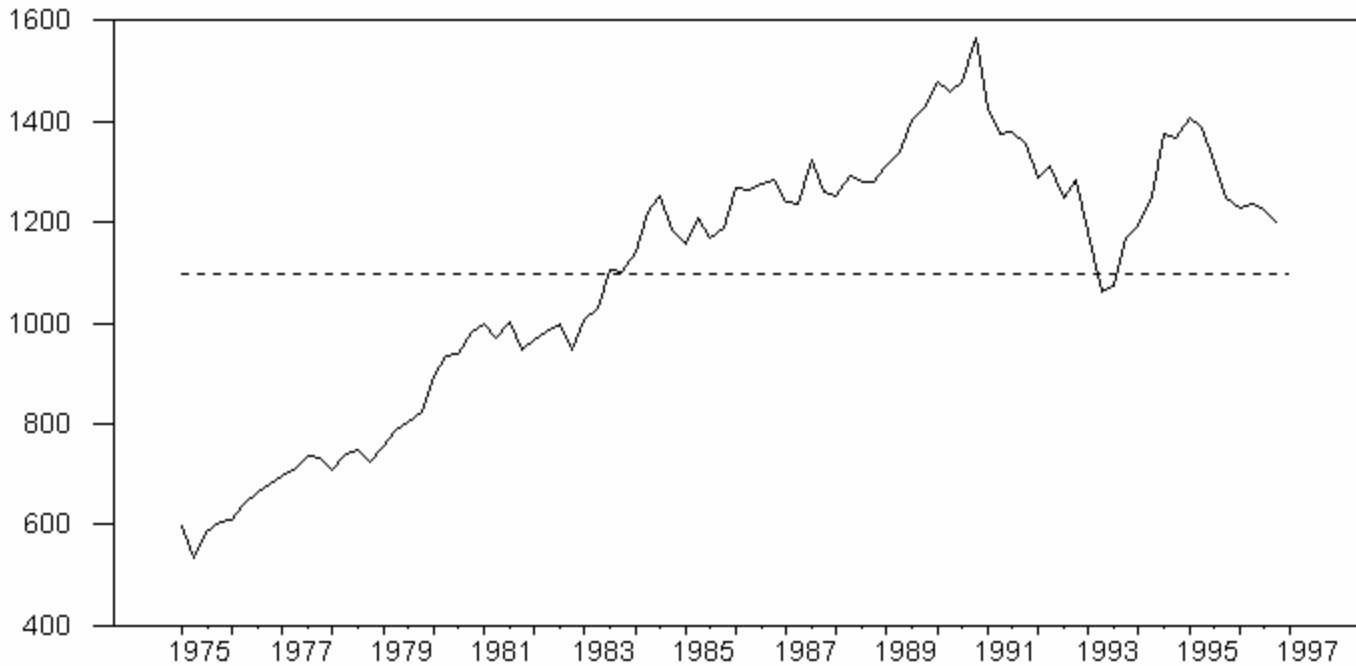


This time series consists of 88 quarterly observations. The mean of the sample (*i.e.*, the average price) is $1,096.75, which is indicated on the graph by the horizontal dotted line. What is most obvious about the data in Fig. 1 is its tendency to move from the lower left of the graph to the upper right, which is typical in any market in which prices tend to increase over time (the prices shown are real, that is, they have been corrected to eliminate the effect of inflation). The significance of this is that the sample mean summarizes the price quite poorly. Except for the periods around 1983 or mid-1993, the statement, "The average price over the sample is $1,096.75" is fairly uninformative, greatly overstating the price in the earlier part of the sample while

understating it in the latter part. A strategy involving waiting a quarter or two for the price to revert to the mean would nearly always fail. Econometricians call this property "non-stationarity," and the price variable in this case is said to be "non-stationary."

An example of a stationary variable would be the time series defined as the *difference* of this price series, telling us to look at the time series consisting of the differences of the prices from one observation to the next. The graph of the lumber prices in differences is shown in Fig. 2.

## Figure 2:  Real Price of Lumber in Differences, 1975-1996
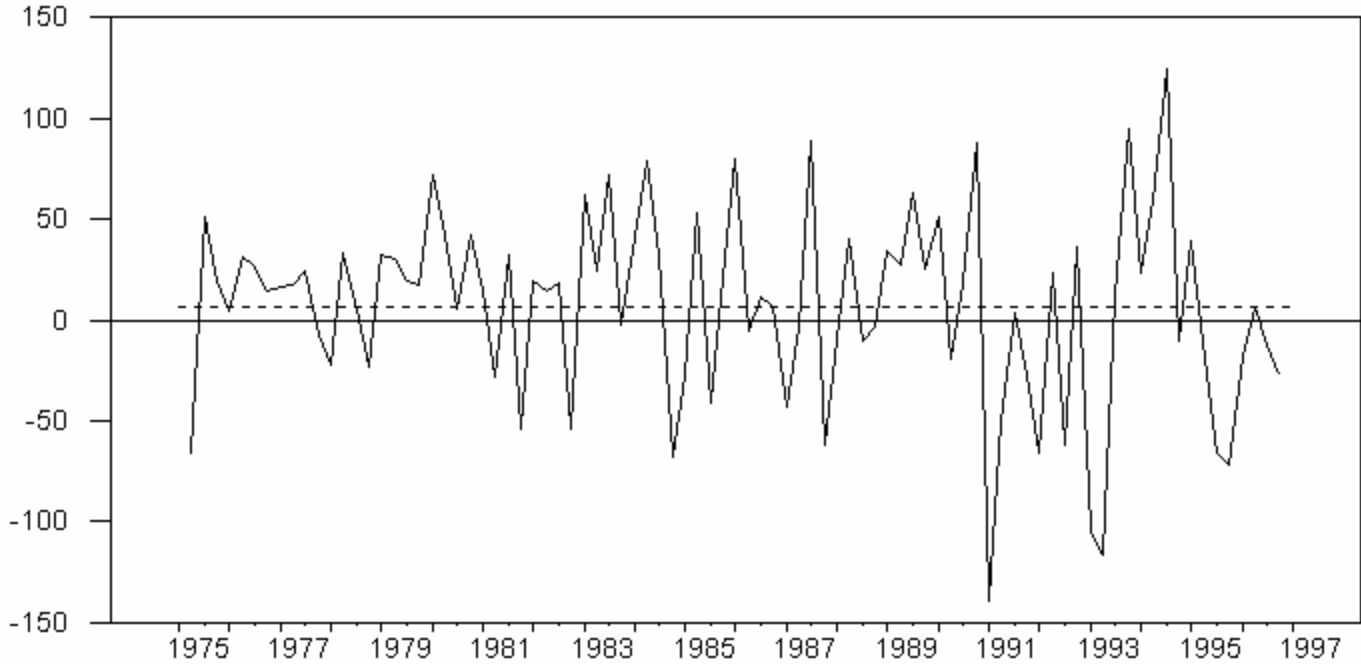


Again, the mean, or average, difference, in this case $6.84, is indicated by the horizontal dotted line.  While the price in levels in Fig. 1 crossed the mean three times, the price in differences crosses the mean 42 times. Clearly, the difference from any given quarter to the next may differ widely from the mean, a quarter or two later the difference reverts to the average.  Such a mean-reverting series is said to be "stationary."

The distinction between stationary and non-stationary time series is important because these data have dramatically different statistical properties.  Standard regression analysis, also known as "ordinary least squares," or "OLS," is said to give the Best, Linear, Unbiased Estimates, *i.e.*, they are

"BLUE," provided that the assumptions underlying the regression model are fulfilled.  Regression estimates that are not BLUE are more likely to be excluded under the Daubert standard, and regression studies involving non-stationary data are not BLUE because they do not fulfill the OLS assumptions.  Econometric studies that do not take non-stationarity into account are flawed, and have little probative value.  It has been shown that regressing two non-stationary series leads to false positives, also known as "spurious regression."[3] Econometric relationships that appear to be statistically significant in the presence of non-stationarity may not, in fact, have any meaningful relationship whatsoever.

## Integration and Cointegration

The technical term for non-stationary time series is that they are "integrated." The cause of such integration can be traced to the accumulation of random influences on the variable. The simplest integrated process is known as a "random walk." The random walk process is said to be integrated of order one because, like the price series in Fig. 1, if it is differenced once it becomes stationary. More generally, if a variable can be made stationary by differencing it *d* times, it is said to be integrated of order *d*. The concept of an integrated time series has not only been extended to higher orders of *d*, but to fractional values of *d* as well.

Ordinarily, the sum of two non-stationary (integrated) time series is also non-stationary. On occasion, however, a unique combination of two integrated time series results in a *stationary* time series, in which case it is said that the data is *cointegrated*.

Intuitively, two series that are cointegrated may be individually non-stationary, but they will not move too far apart over time. A common heuristic example of two cointegrated series is that of a drunk dog-owner walking in a desert. Assuming the owner is drunk enough to have no sense of direction (and does not double-back), his path might resemble a (non-stationary) random walk. His dog's path might also look like a random walk. At any one time they may be close together, and at another time further apart, but over the long run they will move together, and never take off in opposite directions. Before any regression analysis can be considered valid, therefore, the econometrician must be satisfied either that a) the regressors are stationary, or, b) the regressors are cointegrated.

## Multivariate Cointegration

Cointegration theory reaches far beyond explaining, and being able to correct for, spurious regression. It also easily permits a superior approach to multiple regression modeling which virtually eliminates simultaneity bias. Simultaneity bias in regression analysis results when causality runs not only from the explanatory variable to the dependent variable, but "feeds back" from the dependent variable to the explanatory variable, as well. This problem is discussed in a widely available reference on scientific evidence, in which Professor Rubinfeld states,

> The assumption of no feedback is especially important in litigation, because it is possible for the defendant (if responsible, for example, for price-fixing or discrimination) to affect the values of the explanatory variables and thus to bias the usual statistical tests that are used in multiple regression.[4]

The problem is illustrated by supposing that the defendant's expert wants to demonstrate that the price of a product, $P_t$, is determined by three variables: a demand variable, $D_t$, a cost variable, $C_t$, and advertising, $A_t$. Provided that the "no feedback" assumption is fulfilled, the researcher might estimate a multivariate model of the form

$$P_t = \alpha + \beta_1 D_t + \beta_2 C_t + \beta_3 A_t + \varepsilon_t.$$

Setting aside for the time being the spurious regression problem, if the explanatory variables together account for all but residual random variations in the price, and the parameters, $\beta_i$, are all statistically significant, estimations from this model may constitute sound statistical evidence. However, if demand also reacts to price, *i.e.*, $P_t$ also causes variations in $D_t$, then simultaneity bias will invalidate the results.

To remedy this, Dr. Rubinfeld suggests dropping the questionable variable to determine whether its exclusion makes a difference, or expanding the model by adding one or more equations that explain the feedback effect.

The cointegration approach generally solves the problem by expanding the model into a system of equations in which each variable may influence every other variable. The statistical significance of the dependence of each variable on every other variable can then be tested. Instead of the researcher assuming that $P_t$ should be considered the dependent variable and that $D_t$, $C_t$, and $A_t$ should be the explanatory variables, the direction of causality as between each variable can be tested within the model to arrive at a specification that does not suffer from simultaneity bias. This arrangement also permits modeling more complicated dynamics. Thus, a cointegration model can remain agnostic about which of the variables, $P_t$, $D_t$, $C_t$, $A_t$, should be considered dependent or explanatory until there is statistical evidence to support the specification.

## Research and Applications

We have already demonstrated the virtue of the multivariate cointegrated VAR approach in the context of litigation as a diagnostic tool to uncover weakness in regression analyses presented by an opposing expert. The theory offers methods to determine whether variables are integrated of order one or stationary, and thus whether regression results are spurious, whether there is a problem of simultaneity bias, and whether a dynamic rather than a static specification is required.

At the same time, the approach has obvious virtues for *de novo* econometric modeling. Anywhere a linear regression model could

provide probative evidence, the cointegrated VAR model represents a superior approach. But because there are numerous contexts in which the interpretation of a stationary cointegrating process can be theoretically meaningful, cointegration analysis can provide a wealth of other kinds of information to a fact-finder. Of particular interest to the antitrust practitioner is the case in which statistical evidence is needed to determine product or market delineation.

The use of statistical correlation between price series has be justifiably criticized (because of the spurious regression problem, *inter alia*) as a means of determining whether price realizations from potentially substitutable products or from different geographical areas belong to the same or separate markets.[5] But the cointegration paradigm provides a statistically sound basis on which to perform "price tests" for market delineation.

Econometric research has begun to explore this application of cointegration theory,[6] and in recent years the European Commission has relied on cointegration analysis in antitrust and merger cases.[7] Such applications are based on the common-sense notion that non-stationary prices in a single geographical market or for substitutable goods in the same area will not move too far apart over time. In this context the cointegration vector is the stationary process that represents arbitrage and transaction cost differentials between the two products or areas. As Michaels and deVany explain it:

> If two areas are in the same competitive market, their prices will inhabit a band whose width reflects the cost of arbitrage. Those costs include transportation, risk exposure, and information about profitable opportunities. If competition exists, it will quickly bring disparate prices back within their arbitrage limits. If, for

example, bad weather increases price in area *i* while price at area *j* and transmission cost are unchanged, transactions in a competitive market will restore an equilibrium at which the two prices again differ by no more than the arbitrage limits. If the cost of arbitrage varies little over time, two areas are in the same market if the difference between their prices is relatively constant. The statistical technique known as cointegration provides a criterion under which to determine the relative constancy of such a difference. If the prices are not cointegrated, there are no well-defined bounds on the difference between them. If prices in two areas are cointegrated, the areas are in the same economic market. Although the difference between the prices varies with some randomness, there is a high probability that it will remain within arbitrage bounds.[8]

In their study of the BP/Arco merger, Hayes, *et al.* (2001, note 7, *supra*, at 7) cite cointegration studies of the price series of various types of crude oil because, "evidence that prices are cointegrated has been interpreted as an indication that the products in question trade in the same antitrust market. In particular, the fact that the price of [Alaskan North Slope] is cointegrated with world crude prices indicates that ANS trades in a world market for crude oil."[9]

Unlike the U.S. antitrust enforcement agencies and U.S. courts, who may as yet not been called upon to expressly rely on cointegraton analysis as part of their decision-making, the EC has recognized the utility of cointegration methods in the antitrust context. One notable case is the Gencor/Lonrho merger case,[10] in which the advantage of cointegration compared to cross-correlation analysis was emphasized. The case involved the proposed joint

acquisition by Gencor Ltd. of South Africa (a holding company) and Lonrho plc of the U.K. of the Gencor platinum mining and refining assets. One of the issues in the case was whether platinum, gold, silver, rhodium and palladium should be considered separate antitrust product markets. Noting that the prices of these commodities are "highly correlated," the EC recognized that "a high correlation does not in itself imply a causal relationship. Indeed economic price-series data are often non-stationary (*i.e.* trended) and therefore automatically correlated."[11] To reach beyond the limitations of cross-correlations, the Commission undertook a cointegration analysis, which they dubbed "an econometric method which can test whether there is a systematic equilibrium (or long-run) relationship between two or more time-series of data."[12] The Commission concluded:

> The results of the analysis show that the data do not suggest any equilibrium (or long-run) relationship between the respective price levels of platinum, rhodium, palladium, gold and silver, nor of any subset of these metals. This econometric analysis of metal prices indicates that [prices of these metals] tend to vary, over the long run, independently of each other, thus confirming the view that [they] constitute] separate relevant product markets.[13]

Cointegration analysis also played a role in the EC's decision rejecting the acquisition of Lenzing AG, a synthetic fiber manufacturer, by CVC Capital Partners Group Ltd., owner of Acordis, also a manufacturer of synthetic fibers.[14] The relevant market issue depended in part on whether various varieties of fibers belonged to the same product market. While the correlation coefficients were found to be quite high, a statistical test based on the distribution theory that underlies cointegration analysis revealed that the price

gap between "commodity VSF (viscose staple fibers)" and "spun-dyed VSF" was non-stationarity, leading the EC to conclude that these two varieties of fiber do not belong to the same product market.

## An Illustration

Let us conclude with a brief illustrative application of the cointegration approach to delineating the relevant geographical market for a fictional product manufactured by several competitors, each with several facilities throughout the U.S. Assume that the defendant finds it advantageous to claim that the relevant market is the U.S. East Coast, while the plaintiff asserts that the relevant geographic market is narrower, and that Florida and Georgia constitute a separate market. To keep the example simple, we will consider real wholesale prices from suppliers in Boston, Atlanta, and Miami. The data are shown in Fig. 3.

Cross-correlation statistics reveal that the correlation coefficient for the Boston-Miami price pair is 0.41, the Boston-Atlanta pair is 0.55, and the Miami-Atlanta pair is 0.72. While these correlation coefficients tend to suggest that Boston is in a separate market, the results can only be *ad hoc*, and in fact give no reliable guidance to the decision-maker.

## Figure 3: Wholesale Price of Widgets
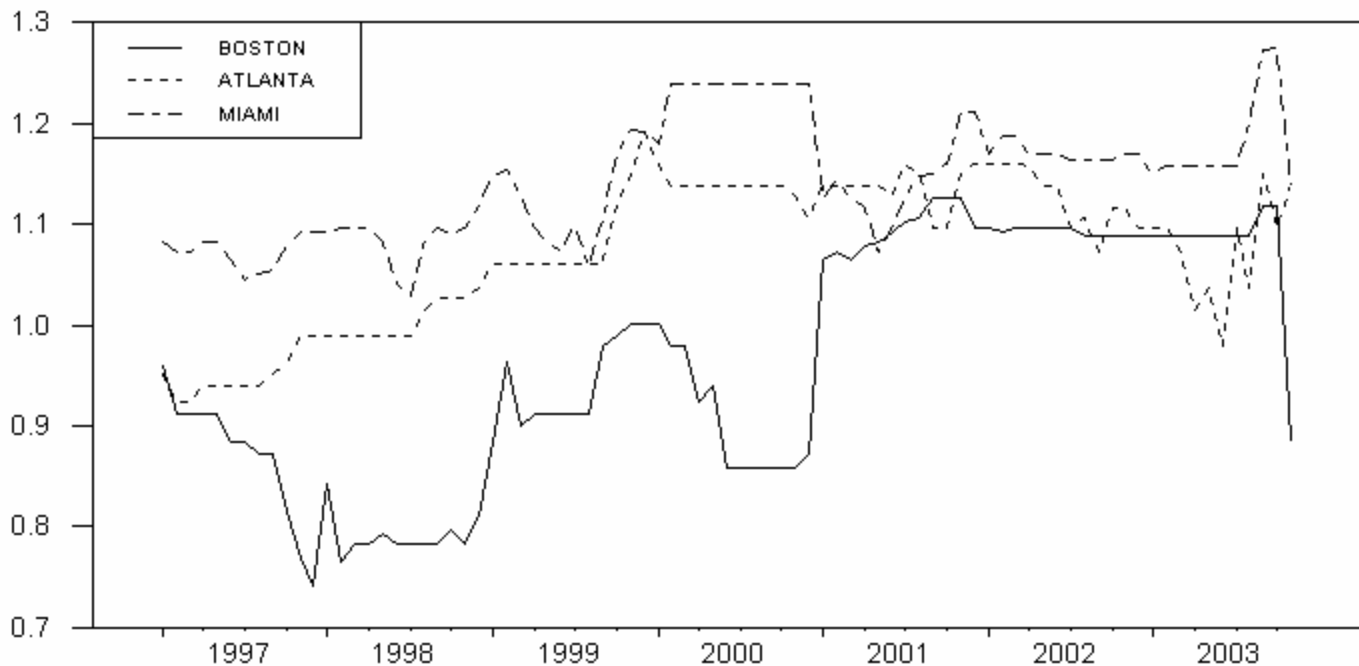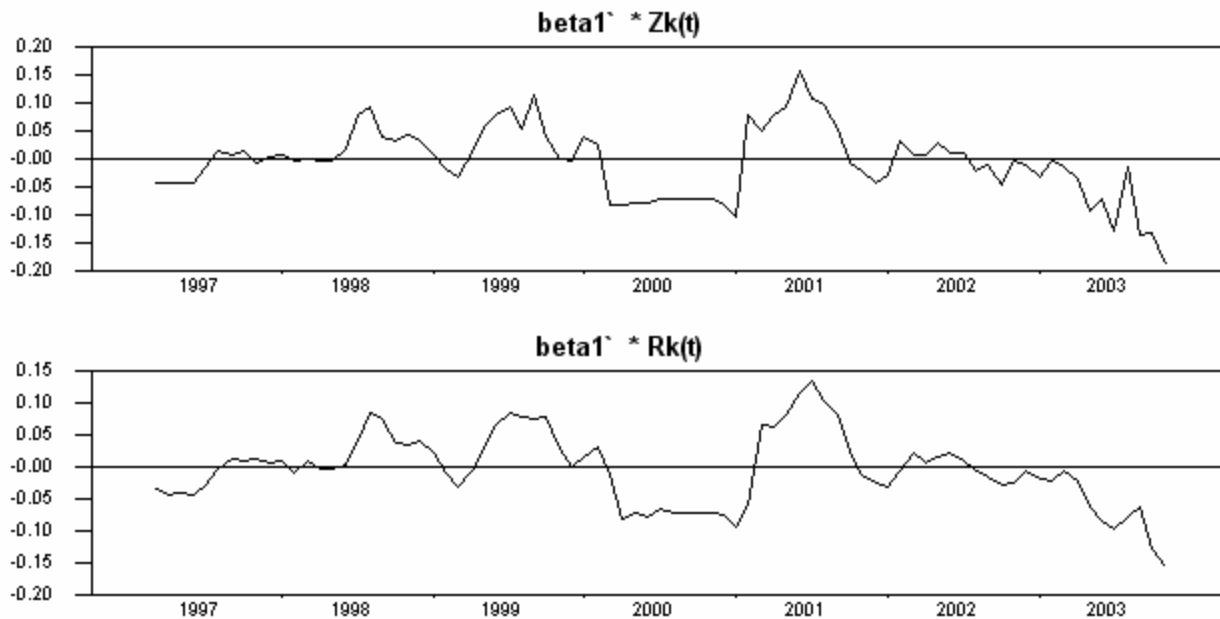Boston, Atlanta and Miami, in Logs

## Figure 4: The Cointegrating Relation



Suppose, however, that a cointegration system is estimated, and that the results indicate that there is a single, unique combination of the three time series that is stationary. We can restrict the model to conform with these finding, estimate the parameters of the cointegrating relation, and determine which variables react to it. Graphs of the cointegrating relation are shown in Fig. 4.

In the upper panel, the cointegrating relation is shown without restrictions on the short-run process. The lower panel shows the cointegrating relation "cleaned" of the influence of the short-run process by regression. Further testing indicates that the Boston prices do not belong in the cointegrating relation, nor do they react to it. Being unable to reject these restrictions establishes the strict exogeneity of the Boston price series. Not only are the Boston prices *weakly* exogenous, in the sense that

they do not enter into the cointegrating relation, but they are also *strongly* exogenous because the variable is not affected by the stationary equilibrium represented by the price gap between prices in Atlanta and Miami, *i.e.*, the cointegrating relation.

Finding a single, stationary cointegrating relation in a three-dimensional system is the same as finding that all three time series are driven by two non-stationary components, or random walks (in the multivariate context, called "common stochastic trends"), because the number of cointegrating relations is the complement of the number of common stochastic trends. The model, therefore, provides strong evidence that the same common trend drives both the Atlanta and Miami price series, while a separate stochastic trend drives prices in Boston. Moreover, since the Boston prices do not enter into the equilibrium represented by the

Atlanta-Miami price gap, and, in turn, the Atlanta-Miami equilibrium relation has no influence on prices in Boston, the model provides strong evidence that suppliers in Atlanta and Miami serve a separate geographical market from the market served by suppliers in Boston.

## Conclusion

The levels of a variable contain its history, while its differences reflect short-term changes. The multivariate cointegrated VAR model is able to examine both types of effects by splitting up the movement of each variable into a long-run and short-term effect, and permitting the researcher to perform separate statistical inference on each. For variables integrated of order one, this is equivalent to splitting the system into non-stationary and stationary components. This approach represents a vast improvement over traditional linear regression analysis, which is invalid in the presence of non-stationarity.

Because regression analysis is such an important econometric tool, and because so many variables are non-stationary, cointegration is likely to find its way into the courtroom soon. Cointegration offers a way of explaining why the regression results presented by the opposing expert are not statistically valid, and a way for your expert to present statistical relationships that can withstand methodological attack. Decision-makers are likely to rely on such statistical methods for help in determining a wide range of antitrust-related issues. Congratulations to Professors Engle and Granger for their tremendous contribution, and for a Nobel Prize well-deserved.

---

[1] Engle, Robert F. and C.W.J. Granger (1987), "Cointegration and error correction: Representation, estimation and testing," Econometrica 55, 251–76.

[2] The "Johansen method" is explored in detail in Johansen, S. (1995), Likelihood-Based Inference in Cointegrated Vector-Autoregressive Models, Oxford University Press, but most econometrics textbooks published since 1997 contain a description of Johansen's approach.

[3] The spurious regression phenomenon was first identified by Yule, G.U. (1926), "Why Do We Sometimes Get Nonsense Correlations Between Time-series?" Journal of the Royal Statistical Society, 89:1–69, and established rigorously by simulation studies by Granger, C.W.J. and P. Newbold (1974), "Spurious Regression in Econometrics," Journal of Econometrics, 2:111–120. Not until more than ten years later, however, was the asymptotic distribution theory applicable to the Granger-Newbold experiments worked out by Phillips, P.C.B. (1986), "Understanding Spurious Regression in Econometrics," Journal of Econometrics, 33:311-340.

[4] Rubinfeld, D.L. (2000), "Reference Guide on Multiple Regression," in Reference Manual on Scientific Evidence, 2nd Ed., Federal Judicial Center, at 196, available at: http://www.fjc.gov/public/ pdf.nsf/lookup/sciman00.pdf/$file/sciman00.pdf

[5] See, e.g., Werden G. and L. Froeb (1993) "Correlation, Causality, and All That Jazz: The Inherent Shortcomings of Price Tests for Antitrust Market Delineation," Review of Industrial Organization, 8: 329–353.

[6] See, e.g., Møllgaard, P. and C.K. Nielsen (2004), "The Competition Law & Economics of Electricity Market Regulation", European Competition Law Review, 25(1): 37-43; la Cour, L.F. and P. Møllgaard, (2002), "Market Domination: Tests Applied to the Danish Cement Industry," European Journal of Law and Economics 14(2), 99-127; Wårell, L. (2002), "Market Integration in the International Coal Industry: An Error Correction Model," Luleå University of Technology Department of Business Administration and Social Science, Division of Economics, available at: http://www.kkv.se/ forskare student/pdf/proj122 2001.pdf; Hayes, J., C. Shapiro, and R.Town (2001), "Market Definition in Crude Oil: Estimating the Effects of the BP/ARCO Merger," Working Paper, available at: http://www.ftc.gov/bc/gasconf/comments2/oilpaperjo hnhayesetal.pdf; Dhar, T.P. and S. Ray (2000), "Understanding Dynamic Retail Competition Through the Analysis of Strategic Price Response Using Time Series Techniques," presented at the American Agricultural Economics Association

Meeting, Tampa, Florida, 2000, available at: http://www.aae.wisc.edu/fsrg/publications/wp2002 7.pdf; Michaels, R.J. and A.S. deVany, (1995) "Market-based rates for interstate gas pipelines: The relevant market and the real market," 16 EnergyL.J. 299–345; and, deVany, A.S., and W.D. Walls, (1993). "Pipeline Access and Market Integration in the Natural Gas Industry: Evidence from Cointegration Tests", The Energy Journal, Vol. 14, No. 4, pp. 1-19.

[7] See, e.g., Commission Decision of 17 October 2001 declaring a concentration to be incompatible with the common market and the functioning of the EEA Agreement (Case No COMP/M.2187 - CVC/Lenzing) (slip opinion available at: http://europa.eu.int/comm/competition/mergers/cases /decisions/m2187_en.pdf); Commission Decision of 24 April 1996 declaring a concentration to be incompatible with the common market and the functioning of the EEA Agreement (Case No IV/M.619 - Gencor/Lonrho) Official Journal L 011 , 14/01/1997 p. 0030 - 0072; and Commission Decision of 31 January 1994 declaring a concentration to be compatible with the common market (Case No IV/M.315-Mannesmann/Vallourec/Ilva) Official Journal L 102 , 21/04/1994 p. 0015 - 0037.

[8] Michaels, R.J. and A.S. deVany (1995), note 6, supra, at 327.

[9] Citing Rodriguez, A.E. and M.D. Williams (1993), "Is the World Oil Market 'One Great Pool'? A Test," Energy Studies Journal, 121–130, for the proposition that cointegration tests "show that a relevant antitrust product market is no narrower than crude oil and the appropriate geographic market is the world."

[10] Note 7, supra.

[11] Id., at 52.

[12] Id., at 53.

[13] Id.

[14] Note 7, supra, at 109–10.