A Primer on Spurious Statistical Significance in Time Series Regressions

Ai Deng[1]

Regression analysis is an important tool in antitrust litigation: it's a formal way to establish an empirical relationship among variables such as prices, quantities, and supply and demand factors. Regression models are commonly used by economic experts to estimate the impact of cartel conduct in price-fixing cases and to investigate competitive effects in merger cases. Proper uses of regression models have been accepted by the courts and have met Daubert standards. But in relying on regressions, economic consultants, as well as attorneys, need to be aware of the possibility of "spurious statistical significance."

Imagine, for instance, that an economic expert decides to use a regression analysis to formalize and test the theory of harm. The expert finds that the regression results in a high R squared (R^2) and produces statistically significant coefficients. In the expert report, the expert explains (1) that a high R^2 shows the model fits the data very well and (2) that the statistically significant coefficients are consistent with a meaningful impact. But then the rebuttal expert report comes back and alleges that the expert's regression produced a false positive—i.e., the coefficient is in fact not significant and the high R^2 is not indicative of a meaningful relationship.[2]

---

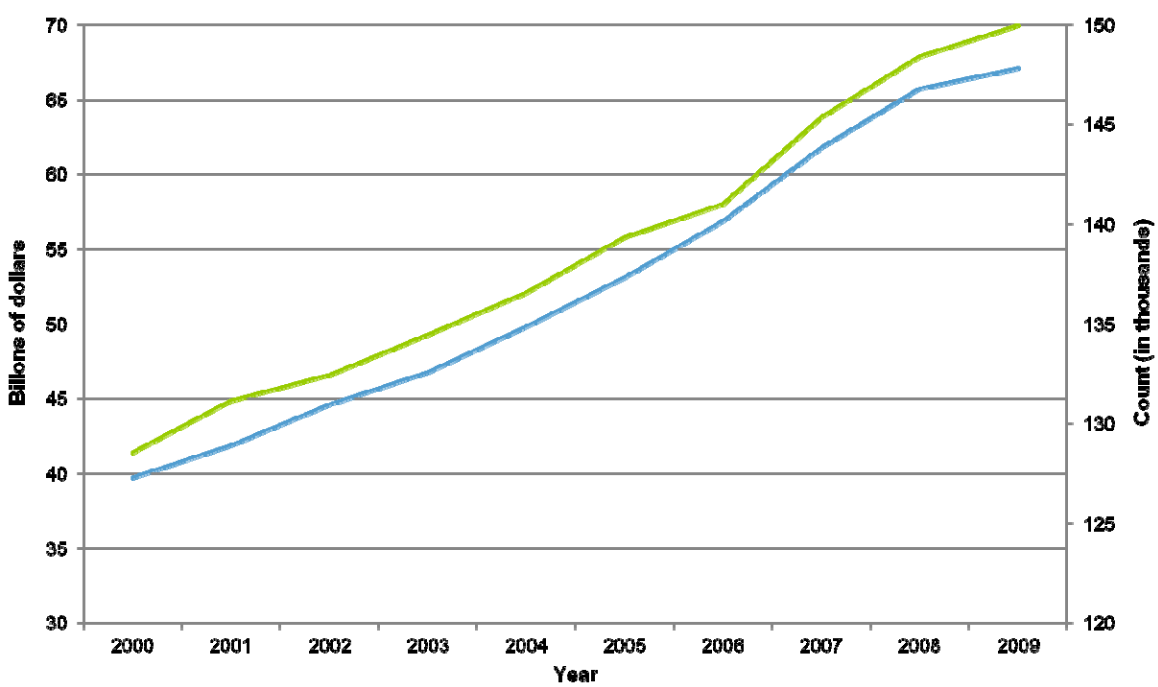
[1] Ai Deng, PhD, is a Principal in the Antitrust and Competition practice and the Energy practice at Bates White Economic Consulting and an adjunct professor of economics at Johns Hopkins University's advanced academic program. He can be reached at ai.deng@bateswhite.com. The author thanks Paul Johnson, Mark Nelson, Jonathan Robell, and Scott Thompson for their comments on early drafts, and Cindy Monroe for excellent editorial assistance. The views expressed in this paper are those of the author and do not necessarily reflect the opinions of Bates White, Johns Hopkins University, or their affiliates.

[2] It is important to keep in mind the fact that statistical significance also depends on the sample size, a factor that an economist might not have full control over in a litigation context. This is an issue we will not address in this article.

In this article, we take a closer look at the problem of false positives. We focus on a common type of economic data that are observed over time, known as *time series*. Such data are routinely processed and analyzed by expert witnesses in litigation matters. Examples include monthly prices and sales of the products in question. In this context, we provide a non-technical discussion about when spurious statistical significance might arise and what measures can be taken to avoid the problem.

Let's now consider a concrete example. Figure 1 shows two variables and their values over time (i.e., a time series). The correlation between the two variables is 0.99 and highly statistically significant.[3] A regression gives similar statistically significant results. These statistical results "confirm" what our eyes see—that the trend lines move in concert over time. It turns out that the blue line shows the amount of money spent on pets in the United

Figure 1: An example of a spurious relationship



---

[3] The correlation or the correlation coefficient between two variables is a number between −1 and 1. If the correlation is 1 (−1), we say that the two variables are perfectly positively (negatively) correlated.

States and that the green line shows the number of lawyers in California.[4] It is therefore highly unlikely that the 99% correlation is "real."

In another example, Professor David Hendry, in his 1980 article "Econometrics: Alchemy or Science," reported on a regression that used a measure of the UK government's money supply and the cumulative rainfall in the United Kingdom. This regression fitted the data quite well and the relationship was highly statistically significant.[5]

The problem is not new. In 1926, Yule, asked "Why Do We Sometimes Get Nonsense Correlations?"[6] Since then, econometricians have come a long way in understanding the problem and in learning how to avoid it. Understanding what produces these unreliable results would allow attorneys and economic experts to (1) ensure that they don't run into the problem themselves and (2) develop appropriate challenges to opposing sides' analysis/arguments.

Intuitively, a statistical regression "looks at" the empirical patterns of how variables move and then infers their relationship. With the two trending variables shown in Figure 1, the regression is "tricked" into believing that there is a true meaningful relationship between the variables. While this example, especially the almost perfect parallel movements, may appear rather contrived, the econometric research has found that spurious statistical

---

[4] This example is taken from http://www.tylervigen.com/view_correlation?id=2956, accessed Nov. 21, 2014. This website, maintained by Tyler Vigen, contains many other examples of potentially spurious correlations. Some of my personal favorites are "US spending on science, space, and technology and Suicides by hanging, strangulation and suffocation" (correlation of 0.99); "number of lawyers in North Carolina and Suicides by hanging, strangulation and suffocation" (correlation of 0.99); and on a sweeter and happier note, "honey producing bee colonies (US) and marriage rate in Vermont" (correlation of 0.94).

[5] David Hendry, "Econometrics: Alchemy or Science," *Economica* 47, no. 188 (1980): 387–406. To be precise, Hendry regressed the money supply on both the cumulative rainfall and the squared cumulative rainfall.

[6] George U. Yule, "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, no. 1 (1926): 1–63.

significance can arise in more subtle ways—as a simulation exercise below will demonstrate, variables certainly do not need to move in parallel as those in our example are made to appear for there to be a risk of spurious statistical significance. Furthermore, extensive research tells us that spurious results such as those in the examples above are *not* mere statistical coincidence.

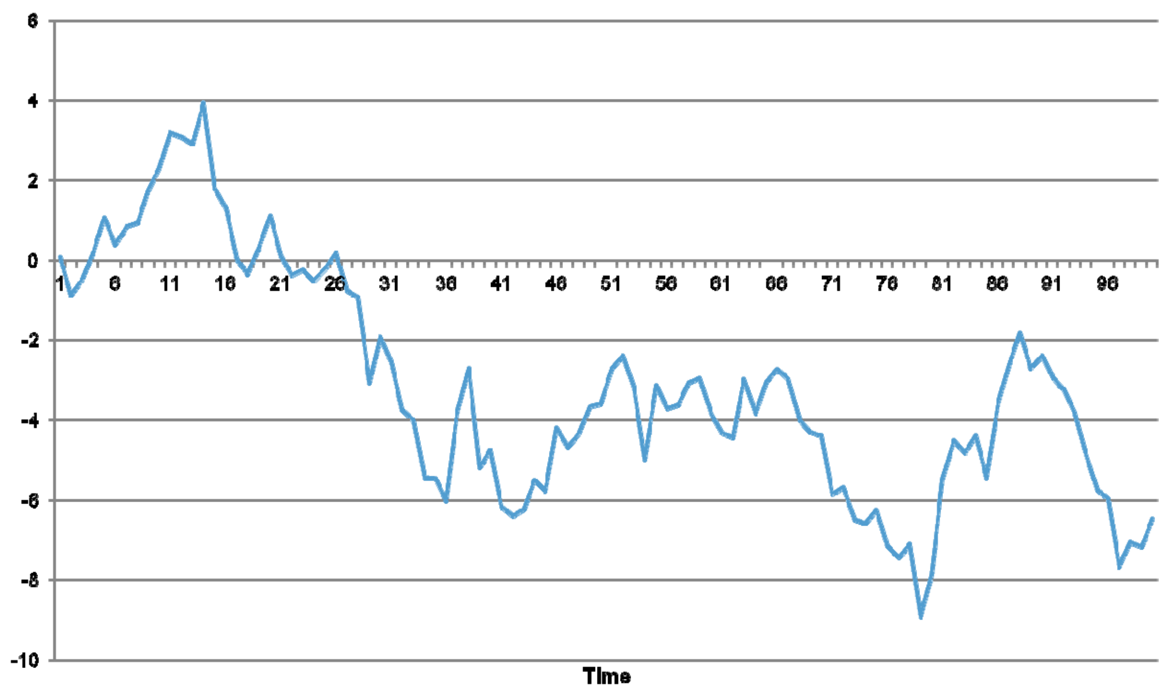
To understand what is going on, it is helpful to know that different types of variables "behave" differently in a statistical analysis including a regression or correlation analysis. Intuitively, trending, slowly moving (nonstationary) variables carry more information content than nontrending (stationary) variables.[7] What do we mean by "more information content"? The following heuristic example will illustrate.

Suppose we want to figure out, by observing the behavior of two colleagues over a number of days, whether those two colleagues are friends. Assume that these two are friends, but that we don't know that. In scenario 1, we see them texting each other, hanging out, and having lunch on many occasions. From those observations we can deduce with a high degree of confidence that they are friends. In scenario 2, we observe that the two have no contact at all (perhaps because they are extremely busy). In the latter case, we can't really know much from what we observe. In scenario 1, there are a lot of activities or "variations" in our observations. These are precisely the types of information that statistical techniques such as correlations or regressions try to exploit. It turns out that nonstationary or trending variables behave similarly to the data in scenario 1. And they typically carry more information than stationary variables.

[7] The technical definitions of stationary and nonstationary variables are related to the constancy and invariance of the mean and (co)variances of the variables. For details, see, for example, James D. Hamilton, *Time Series Analysis* (Princeton, NJ: Princeton University Press, 1994).

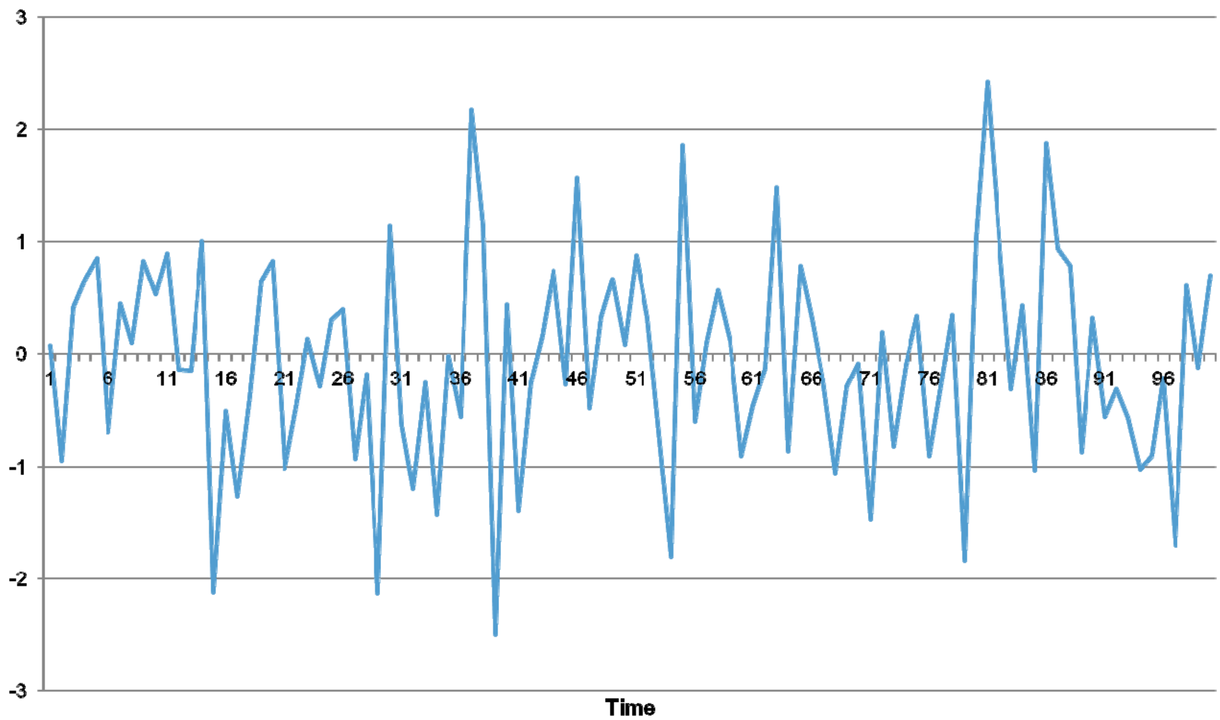
Research has shown that the strong information content in nonstationary variables is a double-edged sword. One type of "nonstationarity" that is well known to be capable of creating problems in regressions is the so-called stochastic trend. Loosely speaking, this is a time trend with variable slopes (i.e., slopes that do not always head in the same direction as a "deterministic" trend would). Such data are also called "integrated" in the econometrics literature. In their 1972 seminal paper, Granger and Newbold found that if two variables both have a stochastic trend, then regressions would (more often than not) indicate a statistically significant relationship between them, even when they are completely

Figure 2: A simulated example of a random walk



independent from each other![8] A classical example of a stochastic trend is the so-called random walk. Figure 2 shows a simulated example. In contrast, Figure 3 shows a simulated example of a particular type of stationary variable: "white noise." The visual difference is striking.[9]

Figure 3: A simulated example of a white noise



To illustrate the type of problem that Granger and Newbold called our attention to, I replicated a small simulation exercise from their study. In this exercise, I used a computer program to generate a large number of independent random walks and tested the statistical significance of the relationship between pairs of these independent time series. Had there been

8 Clive W. J. Granger and Paul Newbold, "Spurious Regression in Econometrics," *Journal of Econometrics* 2 (1974): 111–20. Their study was later extended by other researchers. Phillips provided a mathematical theory to explain these simulation results. Peter C. B. Phillips, "Understanding Spurious Regressions in Econometrics," *Journal of Econometrics* 33 (1986): 311–40.

9 In practice, the difference between stationary and nonstationary time series is often far from clear-cut.

no spurious statistical significance, we should find statistical *insignificant* relationship between the majority of the pairs, simply because they are generated as entirely independent variables. It turned out, however, that out of the 2,000 pairs of *independent* random walks that I generated, the relationship between over 1,500 pairs or 76% are found to be statistically significant by the commonly used two-sided t-test. [10]

So trends, deterministic or stochastic, can do the "trick"; but can other "shapes" of the data also do the "trick?" Yes, they can. Seasonality is one such "trickster." One interesting and well-known example of seasonality producing spuriousness is a regression/correlation analysis on the amount of ice cream sold and the number of deaths caused by drowning.[11] The occurrence of these events in the summer season is the only thing that produces a high correlation and similar patterns in both variables. [12] Another data feature that can result in spurious statistical significance is structural breaks such as level or slope shifts in the data.[13]

Trends, seasonality, and structural breaks are all part of a low-frequency component in that they are all somewhat smooth, slow moving,

---

10 Among those 1,500 pairs, roughly half of them appear to be positively related and the other half negatively related.

11 *See*, *e.g.*, Robert B. Johnson and Larry B. Christensen, *Educational Research: Quantitative*, *Qualitative*, *and Mixed Approaches* (Thousand Oaks, CA: SAGE Publications, 2013), Table 11.2.

12 The issue here is that both variables are affected by a common factor. So in this sense, this example is different from the spurious regression between independent stochastic trends. It is more related to a broader definition of spurious correlation/causation mentioned in the end of the article.

13 Antonio E. Noriega and Daniel Ventosa-Santaulària, "Spurious Regression under Broken Trend Stationarity," *Journal of Time Series Analysis* 27, no. 5 (2006): 671–84.

and/or long lasting. Economists, and attorneys, need to be careful when these components are in the data. [14]

In fact, high correlations between trending variables have been treated cautiously by the courts. An example can be found in Judge Seeborg's recent Order in *In re Optical Disk Drive Antitrust Litigation*. After discussing how the plaintiffs' economic expert offered a correlation analysis to "show that supra-competitive prices paid by Dell and HP as a result of bid-rigging affected prices paid by other purchasers," Judge Seeborg, citing economic experts from both sides, commented: "There appears to be little dispute, however, that strong correlations would arise from the long term price declines and the competitive market forces in any event."[15] The "long term price declines" refer specifically to the downward trend in the optical disk drive prices. What is interesting here is that, unlike our previous heuristic and somewhat extreme examples (money supply vs. cumulative rainfall, or money spent on pets vs. the number of lawyers), the price data being analyzed in this case are at least conceptually related. So the strong statistical correlation might actually reflect a true economic relationship, hence not spurious. But how does one go about trying to resolve this question in practice? The most obvious approach is to examine *additional* economic/econometric evidence to see if they either corroborate or refute the conclusion of a meaningful relationship. For example, economic experts may consider performing a formal cointegration analysis, which will be briefly discussed below, to see if the high correlation is in fact spurious in a statistical sense. Also important are an analysis of the economics of the relevant market and the associated empirical analysis. A detailed discussion is beyond the scope of this article.

Going back to the type of spurious significance problem discussed above, a natural question is how it can be avoided. Earlier economists would simply remove the trend (through a so-called detrending process), and then

---

14 To precisely define trends, however, is actually quite challenging. *See* Halbert White and Clive W. J. Granger, "Consideration of Trends in Time Series," *Journal of Time Series Econometrics* 3, no. 1 (2011): 1–38.

15 Order Denying Motions for Class Certification, *In re Optical Disk Drive Antitrust Litig.*, No. 3:10-md-2143 RS (N.D. Cal. Oct. 3, 2014).

they would analyze these "detrended" data.[16] In other words, nonstationary variables were often detrended to be turned into stationary variables. Usually, proper detrending can take care of the problem, but detrending can also lead to other issues. In fact, the particular issues related to detrending variables with stochastic trends were the basis for the research that garnered Professor Clive Granger (University of California, San Diego) the 2003 Nobel Prize in economics.

Prior to Granger's groundbreaking work, Professor David Hendry commented that a regression between variables with stochastic trends need not produce spurious statistical relationship. Granger, who conducted the simulation study discussed above to illuminate the danger of spurious regression between variables with stochastic trends, set out to prove Hendry wrong; but instead Granger proved Hendry right. That effort led to the Nobel Prize-winning "theory of cointegration."

Cointegration refers to the situation where there is a true relationship between two, or more, variables with stochastic trends.[17] Intuitively, when such variables are cointegrated, the regression/correlation is not spurious in the sense discussed above. An oft-cited heuristic example of a cointegration relationship is that of a drunkard and his leashed dog walking on the street.[18] The drunkard's path may resemble a random walk (hence nonstationary) and so does the path of his dog. But obviously they will not "deviate" too far from each other. One economic example of a plausible cointegrated relationship is

---

[16] Depending on the type of trends, one could detrend data by estimating a regression such as $y_t = \hat{a} + \hat{b}t + \tilde{y}_t$, where $\hat{a}$ and $\hat{b}$ are estimated coefficients and t is the time trend. The detrended data is simply $\tilde{y}_t$. Or in the case of a stochastic trend, take the first difference, i.e., $\Delta y_t = y_t - y_{t-1}$.

[17] In an anecdote, when two young economists, both of whom studied stochastic trends or integrated data, told Granger that they were getting married, Granger said without even thinking: "So then you guys are cointegrated." (Personal communications with Pierre Perron).

[18] Michael Murray, "A Drunk and Her Dog: An Illustration of Cointegration and Error Correction," *American Statistician*, 48, no. 1 (1994): 37–39.

that of aggregate consumption and income.[19] Although both variables tend to increase over time, the pertinent economic theory nevertheless suggests a meaningful underlying relationship between them.

Importantly, Engle and Granger, in their Nobel Prize-winning article published in 1987, proved mathematically that if the variables are actually cointegrated, not only will there be no problem with analyzing the correlation or regression of the variables (without detrending) but doing so will be precisely the *correct* thing to do.[20] Why is that? Because, as explained intuitively above, nonstationary variables carry more information than stationary variables; and as a result their relationships are more accurately estimated by regressions. [21] Detrending would have eliminated the most informative component, i.e., the stochastic trend, from the nonstationary variables. In other words, the very reason that nonstationary variables can cause statistics to go astray is also the very reason that one should not ignore them in a cointegrated regression! But how do we know if the variables are cointegrated in practice? In that same paper, Engle and Granger developed a statistical test to help us answer this question empirically, thus, at least conceptually, giving us a way to properly handle data with stochastic trends.

---

[19] There is evidence that both consumption and income contain a stochastic trend and that they are cointegrated. *See* J. E. H. Davidson , D. F. Hendry, Frank Srba, and Stephen Yeo, "Econometric Modelling of the Aggregate Time-Series Relationship between Consumers' Expenditure and Income in the United Kingdom," *Economic Journal*, 88, no. 352 (1978): 661–92.

[20] Clive W. J. Granger, Robert Engle, "Co-Integration and Error Correction: Representation, Estimation, and Testing," *Econometrica*, 55, no. 2 (1987): 251–76. *See also*, Søren Johansen, "Correlation, Regression, and Cointegration of Nonstationary Economic Time Series," *Bulletin of the ISI LXII* 2007, 2008, 19–26.

[21] In technical terms, it has been shown that the regression estimates between cointegrated nonstationary variables are "super-consistent." A statistical estimate is said to be consistent if it approaches (or "converges") the true (but unknown) value as the sample size gets larger. When econometricians talk about consistency, they often associate it with a "rate," i.e., how fast the convergence is. In regressions of stationary variables, the rate of convergence is usually the square root of the sample size. But if the nonstationary variables are "cointegrated," the rate of convergence turns out to be the sample size, which is a much "faster" rate than in the "stationary" case. Consequently, these estimates are called "super-consistent."

There has been an explosion of academic research on cointegration in the past 30 years. And because many, if not most, economic data (prices, in particular) are nonstationary, cointegration has become an indispensible tool in the economist's toolkit.[22, 23]

To avoid the most basic problem of spurious statistical significance when analyzing time series data, the first line of defense is and should always be the pertinent economic theory. Questions such as does the economic theory support a plausible relationship among these variables should always be asked before any actual regression analysis is undertaken. When the theory is not sufficient or strong enough to convince a careful economist, further diagnostic analysis will be needed. For example, it is often helpful to plot the data to spot trends and to examine the regression residuals (i.e., the variations in the variable of interest that are not explained by the other variables in the regression) for nonstationary behavior. When a formal test is justified and necessary, the economist can apply it to check for evidence of cointegration.

A final caution: while this article focuses on a particular type of statistical illusion, especially with regard to variables with stochastic trends, the word "spurious" as in "spurious regression" and "spurious correlation" is sometimes also used to describe any situation where a false positive is found

---

22 For a nontechnical introduction and a historical perspective of cointegration, see Granger's 2003 Nobel Prize lecture. Clive W. Granger, "Time Series Analysis, Cointegration, and Applications," Nobel Lecture, Dec. 8, 2003, *available at* http://www.nobelprize.org/nobel_prizes/economic-sciences/laureates/2003/granger-lecture.pdf. For more technical details, see, for example, James D. Hamilton, *Time Series Analysis* (Princeton, NJ: Princeton University Press, 1994).

23 With our improved understanding of the related issues, some no longer think that spurious regression is a problem, *as long as we take the effort to properly handle it.* McCallum even goes as far as asking "Is the Spurious Regression Problem Spurious?" *See* Bennett, McCallum, "Is the Spurious Regression Problem Spurious?" *Economics Letters* 107, no. 3 (2010): 321–23. Others have commented that the problem may not appear as easily solvable as McCallum believed. *See* Berenice Martínez-Rivera and Daniel Ventosa-Santaulària, "A Comment on 'Is the Spurious Regression Problem Spurious?'" *Economics Letters* 115, no. 2 (2012): 229–31.

or where there is a more subtle case of "spurious causation." [24] On the topic of causation, it is worth pointing out that similar to correlation, the presence of cointegration *by itself* in general does not imply causation. These issues are beyond the scope of this article. Interested readers can learn more about cointegration by following the references cited in this article.[25]

[24] For example, in the Opinion and Order in the discrimination case *Borden v. Walsh Group*, No. 06 C 4104 (N.D. Ill. Mar. 30, 2012), Judge Lefkow cited the book *The Statistics of Discrimination* to generically define spurious correlation in this way.

[25] The issue discussed in this paper is also distinct from the well-known multiple testing problem in statistics. The multiple testing problem is related to the fact that in the framework of standard (frequentist) hypothesis testing, when a test is applied multiple times (usually over multiple data sets), the (same) null hypothesis may be rejected in some of the applications. But such instances of rejection are not necessarily evidence against the null hypothesis. The interested reader is referred to the vast and still active literature on this important statistical topic for details.