A Drunk and Her Dog: An Illustration of Cointegration and Error Correction

Michael P. Murray

If there exists a stationary linear combination of nonstationary random variables, the variables combined are said to be *cointegrated*. A humorous example of a drunk and her dog illustrates cointegration much as "the drunkard's walk" illustrates random-walk processes. The example makes clear why using first differences of the variables is a mistaken way to look for linear relationships between potentially cointegrated variables. The example also makes clear the link between cointegrated variables and error-correction models.

KEY WORDS: Drunkard's walk; Econometrics; Nonstationary process; Random walk.

1. INTRODUCTION

Teachers of statistics have long used the drunkard's walk to introduce nonstationary processes. Here I adapt the drunkard's walk to clarify the more recently introduced notion of *cointegration* (Granger 1981) and to concretize the link between cointegration and error-correction models (Engle and Granger 1987). The mathematics of cointegration and error correction are sophisticated, but the concepts themselves are simple enough to allow their introduction at elementary levels as a straightforward extension of the drunkard's walk. The link between cointegration and error correction arises naturally from the humorous tale of the drunk and her dog.

2. THE TALE OF THE DRUNK AND HER DOG

The drunk is not the only creature whose behavior follows a random walk. Puppies, too, wander aimlessly when unleashed. Each new scent that crosses the puppy's nose dictates a direction for the pup's next step, with the last scent forgotten as soon as the new one arrives. Thus, the meanderings, x_t and y_t, of both drunks and dogs along the real line can be modeled by the random walk:

$$x_t - x_{t-1} = u_t \tag{1}$$

and

$$y_t - y_{t-1} = w_t, \tag{2}$$

where u_t and w_t are stationary white-noise steps that the woman and dog take each period.

One key trait of random walks is that the most recently observed value of the variable is the best forecaster of future values. If I come out of a bar with a friend who asks me, "Where is that puppy we saw out here earlier?", I am likely to answer, "Well, he was right over there when I went in." We might have the same exchange about a drunk we saw earlier as well.

A second key trait of random walks is that the longer we have been in the bar, the more likely it is that the puppy or the drunk has wandered far from where we last saw them. If my friend and I had been in the bar a long while, I'd say about either the dog or the drunk, "But heaven only knows where they've got to by now." This growing variance in location characterizes the "nonstationarity" of random walks.

But what if the dog belongs to the drunk? The drunk sets out from the bar, about to wander aimlessly in random-walk fashion. But periodically she intones "Oliver, where are you?", and Oliver interrupts his aimless wandering to bark. He hears her; she hears him. He thinks, "Oh, I can't let her get too far off; she'll lock me out." She thinks, "Oh, I can't let him get too far off; he'll wake me up in the middle of the night with his barking." Each assesses how far away the other is and moves to partially close that gap.

Now neither drunk nor dog follows a random walk; each has added what we formally call an *error-correction mechanism* to her or his steps. But if one were to follow either the drunk or her dog, one would still find them wandering seemingly aimlessly in the night; as time goes on, the chance that either will have wandered far from the bar grows. The paths of the drunk and the dog are still nonstationary.

Significantly, despite the nonstationarity of the paths, one might still say, "If you find her, the dog is unlikely to be very far away." If this is right, then the distance between the two paths is stationary, and the walks of the woman and her dog are said to be *cointegrated of order zero*.

To understand the phrase *cointegrated of order zero*, we should first define *integrated series*. Nonstationary series that become stationary when differenced n times are called *integrated of order n*. For a set of series to be cointegrated, each member of the set must be integrated of the same order, n; thus the term *cointegration*. A set of series, all integrated of order n, are said to be cointegrated if and only if some linear combination of the series—with nonzero weights only—is integrated of order less than n. Such a linear combination is called *a cointegrating relationship* (see Engle and Granger [1987] for more details).

Notice that cointegration is a probablistic concept. The dog is not on a leash, which would enforce a fixed distance between the drunk and the dog. The distance between the drunk and the dog is instead a random variable, but a stationary one despite the nonstationarity of the two paths.

More mundanely, we can model the woman's and dog's cointegrated meanderings as

$$x_t - x_{t-1} = u_t + c(y_{t-1} - x_{t-1}) \tag{3}$$

Michael P. Murray is Charles Franklin Phillips Professor of Economics, Bates College, Lewiston, ME 04240. This work was partially supported by the U.S. Department of Housing and Urban Development. The author is grateful to David Aschauer, Rosanne Ducey, Robert Engle, Carl Schwinn, and an anonymous referee for useful comments on a dog-eared earlier draft of this article.

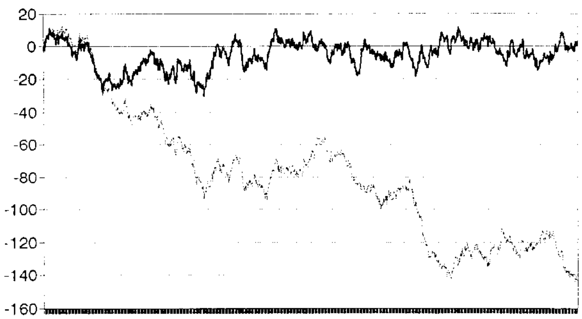

Figure 1. A drunk and two dogs: How close are the dogs to her? — Her dog. --- My dog.

and

$$y_t - y_{t-1} = w_t + d(x_{t-1} - y_{t-1}), \qquad (4)$$

where u_t and w_t are again the stationary white-noise steps of the woman and her dog. The second terms on the right sides are the error-correction terms by which the two wanderers probably stay close together; $(x_{x-1} - y_{t-1})$ is a cointegrating relationship between x and y.

Notice that if the error-correction terms were not stationary, then the steps of the woman and of the dog would also not be stationary, so the two would probably grow far apart over time, despite their efforts to get together. If this were the case, contrary to my expectation when I left the bar, the paths of the woman and of the dog would not be cointegrated of order zero. Engle and Granger (1987) proved, however, that if the drunk and dog follow paths that are both integrated of order one and consistent with the behavior described by Equations (3) and (4), the paths must be cointegrated.

Even with small values of $(c + d)$, the error-correction mechanism has powerful effects on the distance between the drunk and her dog. To illustrate this, Figure 1 shows for a draw of 2000 steps u and w from a standard normal distribution the distances between the drunk and her dog and the drunk and my dog if she and her dog follow Equations (3) and (4) while my dog follows Equation (2); the assumed values of c and d sum to .01. The distance between the drunk and her dog looks quite stationary, whereas the distance between the drunk and my dog looks much more like a random walk. This particular draw of steps is not exceptional in appearance.

In the case of the drunk and her dog, the difference in their locations is stationary. In general, cointegration does not require the difference between the variables to be stationary. Were the drunk and puppy circus performers trying, despite her stupor and his impetuousness, to balance a teeter-totter while otherwise wobbling along the real line, their error-correction device would press them to move in opposite directions as they sought to reestablish balance on the teeter-totter given one another's movements. In this case, the sum of their positions would be a stationary variable with a mean of zero.

Indeed, if the woman and dog were not of the same weight, balancing the teeter-totter would require that the lighter actor be farther away from the fulcrum point. In this case, the stationary variable would be a weighted average of the woman's and dog's positions, not the simple sum of the two. This complexity reflects the fact that two nonstationary variables are cointegrated if there exists any linear combination of the variables that is stationary.

Nonstationary variables pose two threats to conventional regression analysis. The first threat arises with unrelated nonstationary variables that are random walks. Regressing two unrelated random walks against one another results in regression coefficients that are small in relation to their standard errors much less often than the theory of stationary regressors predicts. Consequently, the use of standard distributions when the variables are unrelated and nonstationary will lead to too-frequent rejections of the null hypothesis that there is no relationship between the variables. To guard against such spurious regressions, Granger and Newbold (1974) recommended that regressions among nonstationary variables be conducted as regressions among changes in the variables. In contrast, cointegration analysis avoids such errors by eschewing standard distributions and instead applying the correct distributions.

The second threat arises with truly related nonstationary variables that are integrated of order one. Taking Granger and Newbold's (1974) counsel in this case results in a misspecified regression model. As the story of the drunk and her dog makes clear, if one specifies such a regression model in terms of changes in the variables only, one misses the error-correction mechanism that connects cointegrated variables. Regressions involving the changes of cointegrated variables should also involve the lagged levels of those variables, but with the constraints of the cointegrating relationship imposed. This is a central point of Engle and Granger's (1987) work. (If the data are generated by (3) and (4), then once the cointegrating relationship is accounted for, changes in x and changes in y have no further effect on one another, and lagged changes in x and y have no further effect on current changes. If one omits the cointegrating relationship, however, spurious effects of both these sorts may be observed.)

In addition to illustrating cointegration and error correction, the tale of the drunk and her dog offers a reminder to applied econometricians that the cointegrating relationship is not merely a statistical convenience with no behavioral content. If you estimated by regression that the cointegrating relationship for the drunk and dog were $(x_t - 13y_t)$, I should object that the estimates make no good sense. (Unless, of course, you can tell a different behavioral story than mine, a story that makes the 13 in the cointegrating relationship plausible.) Sims (1980) and others have argued that economic theory tells us too little about the dynamics of economic relationships to impose many useful restrictions on dynamic equations like those describing the steps of the woman or her dog. Arguably, however, economic theory does tell us more about the long-run relationships that hold among economic variables and hence may tell us quite a bit about what we should find in cointegrating relationships. (Hall, Anderson, and Granger (1992), Johansen and Juselius (1990), and Murray (1993) all test such long-run hypotheses about the coefficients in particular cointegating relationships.)

No discussion of regressions involving nonstationary variables would be complete without reminding readers about what I call the *bad companions rule*. When thoughtful parents of a teenager bring their child to the mall for a movie, they are wont to order the child not to leave the mall until the parents return after the movie. The parents' motive is to keep their child from associating with the unruly kids who may go who knows where and do who knows what. Kids who stay close to the mall avoid associating with these bad companions. In econometric analysis, the bad companions rule reappears; stationary variables, like stationary kids, tend not to be associated with those that are nonstationary. Thus, as Stock and Watson (1988) noted, if one regresses a stationary variable against a nonstationary variable, the observed association will in large samples tend to zero as the variation in the stationary variable grows ever smaller in relation to the variation in the nonstationary variable.

[*Received October 1992. Revised March 1993.*]

REFERENCES

Engle, R. F., and Granger, C. W. J. (1987), "Cointegration and Error Correction: Representation, Estimation, and Testing," *Econometrica*, 55, 251–276.

Granger, C. W. J. (1981), "Some Properties of Time Series Data and Their Use in Econometric Model Specification," *Journal of Econometrics*, 16, 121–130.

Granger, C. W. J., and Newbold, P. (1974), "Spurious Regressions in Econometrics," *Journal of Econometrics*, 26, 111–120.

Hall, A. D., Anderson, H., and Granger, C. W. J. (1992), "A Cointegration Analysis of Treasury Bill Yields," *Review of Economics and Statistics*, 74, 116–126.

Johansen, S., and Juselius, K. (1990), "Maximum Likelihood Estimation and Inference on Cointegration—With Applications to the Demand for Money," *Oxford Bulletin of Economics and Statistics*, 52, 169–210.

Murray, M. P. (1993), "Subsidized and Unsubsidized Housing Stocks, 1935–1987: Crowding Out and Cointegration," Dept. of Economics, Bates College, Lewiston, ME.

Sims, C. (1980), "Macroeconomics and Reality," *Econometrica*, 70, 1–47.

Stock, J. H., and Watson, M. W. (1988), "Variable Trends in Economic Time Series," *Journal of Economic Perspectives*, 2, 147–174.