Time Series Analysis, Cointegration, and Applications
(The Nobel Lecture)

Clive W.J. Granger
Department of Economics
University of California, San Diego
La Jolla, CA 92093-0508
USA

Stockholm, Sweden
December 2003

The two prize winners in Economics this year would describe themselves as "Econometricians," so I thought that I should start by explaining that term.

One can begin with the ancient subject of Mathematics which is largely concerned with the discovery of relationships between deterministic variables using a rigorous argument. (A deterministic variable is one whose value is known with certainty.) However, by the middle of the last millennium it became clear that some objects were not deterministic, they had to be described with the use of probabilities, so that Mathematics grew a substantial sub-field known as "Statistics." This later became involved with the analysis of data and a number of methods have been developed for data having what may be called "standard properties."

However, in some areas of application, the data that they generated were found to be not standard, and so special sub-sub-fields needed to be developed. For example, Biology produced Biometrics, Psychology gave us Psychometrics, and Economics produced Econometrics.

There are many types of economic data, but the type considered by Rob Engle and myself is know as time series. Consider the measurement of unemployment rates which is an important measure of the health of the economy. Figures are gathered by a government agency and each month a new number is announced. Next month there will be another value, and so forth. String these value together in a simple graph and you get a "time series."

Rather than show a diagram, I would rather you use internal visualization (I think that you learn more that way). Suppose that you have a loosely strung string of pearls which you throw down, gently, onto a hard table top with the string of pearls roughly stretched out. You will have created a time series with time represented by the distance down the table, the sum of the variable as the distance from the bottom edge of the table to a point, and the set of pearls giving the points in the series. As the placement of one pearl will impact where the next one lies, because they are linked together, this series will appear to be rather smooth, and will not have big fluctuations in value for one term to the next.

Time series can vary in many ways, some are gathered very often, others less frequently. Values for many important financial variables are know not merely daily, but can be found within seconds, if they change, such as highly traded stock prices or exchange rates. These are called "high frequency data" and are the input for Rob Engle's studies.

At the other extreme, some aspects of the overall, or "macro," economy, such as national income, consumption, and investment, may be available only quarterly for many countries, and only annually for others. Population data is also available only annually or less frequently. Many of these series are rather smooth, moving with local trends or with long swings, but the swings are not regular. It is this relative smoothness that makes them unsuitable for analysis with standard statistical procedures, which assumes data to have a property know as "stationarity." Many series in economics, particularly in finance and macro-economics, do not have this property and can be called "integrated" or, sometimes incorrectly, "non-stationary." However, when expressed in terms of changes or rates of returns, these derived series appear closer to being stationary. The string of pearls would be "integrated" as it is a smooth series.

Methods to analyze a single integrated series had been proposed previously by Box and Jenkins (1970) and others, but the joint analysis of pairs, or more, of such series was missing an important feature. It turns out that the difference between a pair of integrated series can be stationary, and this property is known as "cointegration." More formally, two smooth series, properly scaled, can move and turn, but slowly, in similar but not identical fashions, but the distance between them can be stationary.

Suppose that we had two similar chains of pearls and we threw each on the table separately, but for ease of visualization, they do not cross one another. Each would represent smooth series but would follow different shapes and have no relationship. The distances between the two sets of pearls would also give a smooth series if you plotted it.

However, if the pearls were set in small but strong magnets, it is possible that there would be an attraction between the two chains, and that they would have similar, but not identical, smooth shapes. In that case, the distance between the two sets of pearls would give a stationary series and this would give an example of cointegration.

For cointegration, a pair of integrated, or smooth series, must have the property that a linear combination of them is stationary. Most pairs of integrated series will not have the property, so that cointegration should be considered as a surprise when it occurs. In practice, many pairs of macro-economic series seem to have the property, as is suggested by economic theory.

Once we know that a pair of variables has the cointegration property it follows that they have a number of other interesting or useful properties. For example, they must both be cointegrated with the same hidden common factor. Further, they can be considered to be generated by what is know as the "error-correction model," in which the change of one of the series is explained in terms of the lag of the difference between the series, possibly after scaling, and lags of the differences of each series. The other series will be represented by a similar dynamic equation. Data generated by such a model is sure to be cointegrated. The error-correction model has been particularly important in making the idea of cointegration practically useful. It was invented by the well-known econometrician Dennis Sargan, who took some famous equations from the theory of economic growth and made them stochastic.

The early development of the cointegration idea was helped greatly by colleagues and friends in the Scandanavian countries, including Søren Johansen and Katerina Juselius in Copenhagen who developed and applied sophisticated testing procedures, Svend Hylleberg in Århus who extended the theory to seasonal data, and Eilev Jansen and his colleagues at the Bank of Norway, who successfully applied it to a large econometric model of Norway. To complete the set, Timo Teräsvirta, who is from Finland but now lives in Stockholm, helped develop models that were useful in nonlinear formulations of cointegration. I am delighted that they are all here as my guests.

The modern macro economy is large, diffuse, and difficult to define, measure, and control. Economists attempt to build models that will approximate it, that will have similar major properties so that one can conduct simple experiments on them, such as determining the impacts of alternative policies or the long-run implications of some new institution. Economic theorists do this using constants suggested by the theory, whereas the econometrician builds empirical models using what is hopefully relevant data and which captures the main properties of the economy in the past. All models simply assume that the model is correct and extrapolate from there, but hopefully with an indication of uncertainty around future values.

Error-correction models have been a popular form of macro model is recent years, and cointegration is a common element. Applications have been considered using almost all major variables including investment, taxes, consumption, employment, interest rates, government expenditure, and so forth.

It is these types of equations that central banks, the Federal Reserve Bank, and various model builders have found useful for policy simulations and considerations.

A potentially useful property of forecasts based cointegration is that when extended some way ahead, the forecasts of the two series will form a constant ratio, as is expected by some asymptotic economic theory. It is this asymptotic result that makes this class of models of some interest to economic theorists who are concerned with "equilibrium." Whether the form of equilibria suggested by the models is related to that discussed by the theorists is unclear.

These ideas and models are fairly easily extended to many variables. Once the idea of cointegration (a name that was my own invention, incidentally) became public it was quickly picked up and used by many other econometricians and applied economists. There are now numerous citations and applications based on it. Rob Engle and I quickly realized that the concept of cointegration and its extensions could be used to explain and remove a variety of difficulties that we had observed in our own research and that of others. It seemed to be the missing piece in our approach to modeling groups of series.

An example is a problem known as "spurious regressions." It had been observed, by Paul Newbold and myself in a small simulation published in 1974, that if two independent integrated series were used in a regression, one chosen as the "dependent variable" and the other the "explanatory variable," the standard regression computer package would very often appear to "find" a relationship whereas in fact there was none. That is, standard statistical methods would find a "spurious regression." This observation lead to a great deal of reevaluation of empirical work, particularly in macro-economics, to see if apparent relationships were correct or not. Many editors had to look again at their list of accepted papers. Putting the analysis in the form of an error-correction model resolves many of the difficulties found with spurious regression.

I am often asked how the idea of cointegration came about; was it the result of logical deduction or a flash of inspiration? In fact, it was rather more prosaic. A colleague, David Hendry, stated that the difference between a pair of integrated series could be stationary. My response was that it could be proved the he was wrong, but in attempting to do so, I showed that he was correct, and generalized it to cointegration, and proved the consequences such as the error-correction representation. I do not always agree with the philosopher Karl Popper, but in his book "The Logic of Scientific Discovery," according to Hacohen (2000), page 244, Popper believed the "discovery was not a matter of logic" but rather the application of methodology, which fits the discovery of cointegraion. This insight intrigues me partly because the book appeared almost exactly at the time of my birth, in September 1934. At this same time Popper was debating Heisenberg on the relevance of probability theory in physics. It happens to be the case that echoes of that debate still persists, but relating to economics. My position is that it is clear that we can best describe many of the movements of economic variables, and the resulting data, using probabilistic concepts. I should also point out that 1934 was also the year the J.M. Keynes finished the first draft of "The General Theory of Employment, Interest, and Money" although it was very many years before I became aware of his book.

As an aside, I wrote this lecture whilst visiting the Department of Economics of the University of Canterbury in New Zealand, where Karl Popper also spent some years after World War II.

Before considering the usefulness of the new methods of analysis, I would like to take a personal detour. This Prize has climaxed my year which started with me being named a Distinguished Fellow of the American Economic Association. Previously in my career, I have been Chair of two economics departments, yet I have received very little formal training in economics. One third of my first year as an undergraduate at the University of Nottingham was in economics, with introductions to micro and in national accounts, and that was it. Whatever other knowledge I have, it has come from living amongst economists for about forty years, by osmosis, attending seminars, having discussions with them, and general reading. My question is: does this say something about me, or something about the field of economics? I think it is true to say that I am not the first Nobel Prize winner in economics to have little formal training in economics. I wonder if economics has less basic core material than is necessary for mathematics, physics, or chemistry, say. Economics does have a multitude of different aspects, applications, and viewpoints which has to each form their own basis, at least in practice. Economic theory does seem to maintain common concepts and features but these may be quite simplistic and are not necessarily realistic.

Possibly because it is not tied down by too many central concepts but certainly because economics involves a myriad of topics, both theoretical and applied, it is a hot-bed of new ideas, concepts, approaches, and models. The availability of more powerful computing at low cost has just increased this activity even more.

In my reading I came across a statement (unfortunately I have forgotten who the author was) noting that "economics is a decision science, concerned with decision matters, such as consumers, employers, investors, and policy makers, in various forms of government, institutions, and corporations." I fully accept this viewpoint as it follows that the "purpose of economics is to help decision makers make better decisions." That statement is useful because it gives us a foundation with which we can compare and evaluate specific pieces of economic analysis. We can ask "how will a decision maker find this result useful?"

As I stated before, the main uses for the economic techniques that I helped develop, such as cointegration, was to build statistical models linking major economic variables that both fit the available data better and agree with the preconceptions of the model constructors about what the construction should look like.

There are a number of stages to the forecasting process; getting the central forecast and then uncertainty bounds around it to give some idea of the risks involved in using this forecast. Finally, previous forecasts have to be gathered and evaluated. Hopefully any tendencies, trends, or swings in the errors can be detected so that one can learn and produce better forecasts in the future. The process of forecast evaluation, plus the use of combinations of forecasts from different series, is an on-going research project at the University of California, San Diego.

Forecasts do not just come from time series, but also from panels of data, which can be thought of as a group of series of a similar nature measured from different sources. An example would be monthly inflation rates from each of the Scandinavian countries. Once one is clear what is the purpose of the analysis, suitable techniques can be formulated.

I have recently been involved in such a project where the purpose is to study the future of the Amazon rainforest in Brazil. This forest covers an area larger than all the countries in the European Union put together, but it is being cut down quite rapidly. I was one of five authors who produce a report (Anderson, et al, Cambridge University Press (2003)) which include a model that could forecast the decline of the forest under various policy scenarios. The forest is not being cut down for its timber, but to get at the land that the timber stands upon to produce food. Unfortunately, unlike European ex-forest land, its useful life span is often rather short, often becoming "fallow" within five years of being forested.

The advantage of being an academic econometrician is the possibility of working on data from many areas. I have run pricing experiments in real super markets, I have analyzed data from stock markets, commodity prices – particularly god and silver prices, interest rates, considered electricity demand in small regions, the female labor force participation, river flooding, and even sun spots. All data present their own unique problems and I continue to find data analysis fascinating, particularly in economics.

An earlier concept that I was concerned with was that of causality. As a post-doctoral student in Princeton in 1959-1960, working with Professor John Tukey and Oskar Morgenstern, I was involved with studying something called the "cross-spectrum," which I will not attempt to explain. Essentially one has a pair of inter-related time series and one would like to know if there are a pair of simple relations, first from the variable *X* explaining *Y* and then from the variable *Y* explaining *X*. I was having difficulty seeing how to approach this question when I met Dennis Gabor who later won the Nobel Prize in Physics in 1971. He told me to read a paper by the eminent mathematician Norbert Weiner which contained a definition that I might want to consider. It was essentially this definition somewhat refined and rounded out that I discussed, together with proposed tests in the mid 1960's. The statement about causality has just two components:

1. The cause occurs before the effect; and
2. The cause contains information about the effect that that is unique, and is in now other variable.

A **consequence** of these statements is that the causal variable can help forecast the effect variable after other data has first been used. Unfortunately, many users concentrated on this forecasting implication rather than on the original definition.

At that time, I had little idea that so many people had very fixed ideas about causation, but nay did agree that my definition was not "true causation" in their eyes, it was only "Granger causation." I would ask for a definition of true causation, but no one would reply. However, my definition was pragmatic and any applied researcher with two or more time series could apply it, so I got plenty of citations. Naturally, many ridiculous papers appeared.

When the idea of cointegration was developed, over a decade later, it became clear immediately that if a pair of series was cointegrated then at least one of them must cause the other. There seems to be no special reason whey there two quite different concepts should be related; it is just the way that the mathematics turned out.

As a brief aside for those of you with more technical training, what I have been telling you about so far has mostly been concepts using linear models. Everything can be generalized to the nonlinear situation and recently efforts have been pushing into using similar concepts in conditional distributions, which is a very general form. It appears that causality will play a basic role in the generalization of the error-correction model, but that is still a work-in-progress.

I am not sure if the empirical studies on causation have proved to be so useful, although the debate relating to money supply and prices was interesting. The concept does help with the formulation of dynamic models in more useful ways.

I started this lecture talking about econometrics. We econometricians love numbers, but let me end with a few. The first two Nobel Prizes in Economics were to econometricians, Ragnar Frisch and Jan Tinbergen, for which we are very proud. In all there are now eight of us with the Prize, representing 15% of the economics winners. However, in the current millennium, we represent about 44% of the winners, which I vew as a healthy local trend.

Over my career and before today, I have met twenty-one Nobel Laureates: one in Physics (Dennis Gabor, 1970), on in Peace (Phillip Noel Baker, 1959), plus 18 Prize winners in Economics. Without exception I have found them to be both very fine scholars and having excellent personalities, willing to help a younger, inexperienced worker when seeking their advice or meeting them socially. I hope that I am able to live up to their very high standard.

References

M.H. Hacohen: "Karl Popper: The Formative Years, 1902-1945." Cambridge University Press, 2000.

K. Popper: "The Logic of Scientific Discovery." Hutchinson: London, 1959.

C. Box and G. Jenkins: "Time Series Analysis, Forecasting, and Control." Holden Day: San Francisco, 1970.

C.W.J. Granger, Lykke Andersen, Eustaquio Reis, Diana Weinhold, and Sven Wunder: "The Dynamics of Deforestation and Economic Growth in the Brazilian Amazon." Cambridge University Press, 2003.