September 11, 2004

# Long Memory of Statistical Time Series Modeling

by Emanuel Parzen

Texas A&M University
NBER/NSF Time Series Conference
September 17, 2004

The aim of this paper is (1) improvement of teaching, practice, and history of statistical science and statistical time series analysis, and (2) celebration of Robert Engle and Clive Granger, and their 2003 Economics Nobel Prize. Many concepts discussed are summarized by the key words.

KEY WORDS: Robert Engle, Clive Granger, simple statistical methods, invention, innovation, answer machines, R. A. Fisher fame and mathematical statisticians, static data, dynamic data, ergodic data, enumerative studies, analytic studies, Deming, what is time series analysis good for, spectral analysis, filters, exponential model, reproducing kernel Hilbert spaces, learning from data, pattern inference, Akaike information criterion, model selection, Tukey, quantile summary of data and populations, quantile unification of confidence intervals, forecasting, prediction, innovations, Box Jenkins ARIMA, Kalman filter, state space and unobserved components, unit roots, cointegration, ARCH, nonlinear models, nonparametric models, future of time series analysis in a world of change, Leontieff.

## 1. A Talk of Celebration with Few Formulas and Much Philosophy

I would like to welcome you with a greeting in Texan: “Shalom all you’all.” In Texan “you’all” means you singular and “all you’all” means you plural.

There are many reasons to be happy that the 2003 Economics Nobel Prize was won by our friends Clive Granger (whose 70th birthday on September 4, 2004 we should note) and Robert Engle. Their prize brings honor and recognition to Time Series Econometrics and also to the discipline of Statistical Time Series Analysis, and the annual NBER/NSF seminars on time series analysis.

We have the pleasure of celebrating, learning, and expressing our admiration for, the hard work and accomplishments of Granger and Engel. I would like to add the perspective of remembering the long memory of time series analysis.

Time series analysis can claim to be 342 years old. Time series econometrics is about 40 years old. Our understanding of the past and future of time series analysis is helped by the 1997 book by Judy Klein *“Statistical Visions in Time, History of Time Series Analysis 1662-1938,”* Cambridge University Press. She teaches us that:

Many methods associated with analysis of stationary time series were invented by manipulators of time series data, merchants, financiers, speculators. These methods include harmonic analysis, first differences and logarithms of data, autocorrelation, autoregression, regression on functions of time, moving averages, and decomposition of time series.

The challenge of time series analysis is comprehending processes of change using analytical tools designed for making comparisons in history where history was frozen.

## 2. Invention, Innovation, Answer Machines

Statistical methods (and also economic models according to Granger 1998 Marshall Lectures) are useful to the extent that their output is useful to the ultimate consumers: managers (who make decisions) and researchers (learning from nature its science). John Tukey (Granger's mentor in spectral analysis) recommended methods with simple theory because he defined the true power of a method to be "simple enough to be applied frequently." I agree with this definition of power of a method but argue that to be applicable a method needs to be simple not in its theory but in the presentation and interpretation of its numerical and graphical outputs.

I believe that the concept of "*answer machine*" is required to explain what statisticians do. Mathematics finds a definite answer to a problem; statistics provide answer machines which are formulas that can be adapted to compute and compare answers under varying assumptions that are acceptable to applied researchers. My current research is about answer machines for confidence intervals.

I propose the problem of computing the variance of grouped data as an illustration of my proposal that statistics cannot provide answers, only answer machines. The calculation of variance has several answers which depend on assumptions such as: values in an interval are at the center; the values are uniform on the interval; the values have a beta distribution on the interval.

The history of statistics is that new methods are "*invented*" by pioneers motivated by applications. "*Innovation*" by statisticians improves methods to make them widely and rigorously applicable. Statisticians innovate (improve) statistical methods using mathematical models and optimization. Statistical science does not make methods; it makes them better.

We should be concerned about whether the future of statistical science will continue to play its role of elegantly abstracting methods from their practical roots so that they can be learned by researchers in many fields (and thus provide technology transfer and links between all sciences).

I question whether the methods invented by R. A. Fisher would have become famous without the innovations of American and Scandinavian statisticians who mathematized and popularized Fisher's "cook book" exposition.

Cointegration provides examples of the interdisciplinary importance of statistics and the distinction between invention and innovation. Cointegration invented for economic modeling can be applied to weather forecasting. Cointegration may have been invented by several researchers but it was clearly innovated by Clive Granger.

I would modify Stigler's famous theorem about names of methods to say that those credited with an idea or method may not have been the inventors but they were innovators.

## 3. Static versus Dynamic Data, Enumerative versus Analytic Studies

The concepts of *sample* and *process* are central to understanding the challenge of applying statistics to economics. Introductory courses in statistics are teaching only half the story. They usually define the goal of statistical methods to be learning from a sample about a population, and emphasize the importance of defining the sampling frame and taking a random sample. They do not warn that the data we are observing may be a process and not a sample.

We need to popularize the ideas of Ed Deming in his 1975 paper "On Probability as a Basis for Action," *The American Statistician*. Deming distinguishes between enumerative studies (where we can imagine sampling from a population or sampling frame) and analytic studies (where data is collected from process of measurement or experimentation such as arises in quality control of a production process). Deming's abstract eloquently states his plea for greater care in the practice and teaching of statistical theory (see section 14).

I propose that we teach about *static data* and *dynamic data*. Static data $X_1, \cdots, X_n$ can be regarded as a sample from a population in which the order of observation is not important. We study the order statistics $X(1;n) \leq \cdots \leq X(n;n)$ of values sorted in increasing order which leads to quantile data analysis (median, quartiles, box plot). If order of observations matters we call $X_1, \cdots, X_n$ a process or dynamic data.

November 2004 issue of *Statistical Science* on non-parametric statistics will contain my review paper "Quantile Probability and Statistical Data Modeling."

It is inspiring to read the 1846 words of Adolphe Quetelet (the teacher of Florence Nightingale), quoted by Klein (p. 259). "Statistics then supposes a state to be for an instant stationary, so that the elements attached to its existence may be enumerated; whilst political history follows it in its march, and verifies all the phenomena it presents. The one science is to the other what, in a different order of things, statics are to dynamics, what rest is to motion. Generally, statistics relate to the present, leaving the past to history, and the future to politics." (Adolphe Quetelet [1846] 1849, 176.)

We define "*Ergodic*" data intuitively to be data for which time averages equal ensemble averages.

Static data statistical methods extend to ergodic data such as stationary time series which are no memory (identically distributed observations which are independent) or short memory (moderately dependent identically distributed process).

Evolutionary (long memory) time series are processes. The challenge of statistical methods for dynamic data has been successfully developed (compare Kalman filter, the pioneering work of Grace Wahba on nonparametric functional data analysis, and the enormous popularity of reproducing kernel Hilbert spaces as a universal method for learning from data). The challenge of applying statistics to dynamic data has been accomplished by applying a deep (rather than routine) understanding of how parametric and non-parametric methods for static data work. Complicated theory can provide outputs simple enough for ultimate consumers. "Everything should be as simple as possible, but not simpler" (to quote Albert Einstein).

## 4. What is Time Series Analysis Good for?

Statistics graduate students should be prepared to answer questions asked by recruiters such as “What is time series analysis good for?” I would welcome your written answers. My answer: (1) in 1900 when researchers computed cross-correlations between economic variables they obtained high spurious correlations which led them to discover high autocorrelation; (2) a major application is forecasting of future sales (under the assumption that the future is like the past) in order to compare them to actual sales (and determine credit of a new advertising campaign); (3) monitoring patients in an intensive care unit you are recording time series of medical conditions which you want to analyze for early warning of fatal conditions; (4) monitoring quality on a production line avoid being fooled into detecting a change which is not significant if account is taken that the data is not independent but can be modeled as a stationary time series. Moral: Statistics students should learn these ideas early rather than late or never.

## 5. Spectral Analysis, Filters

Statistical time series analysts who apply frequency domain methods (as well as time domain methods) have more reason to appreciate the accomplishments of Granger and Engle which include published applications of frequency domain methods:

**Engle**

1973 Band Spectrum Regression
*International Economic Review*

1976 Interpreting spectral analysis in terms of time domain models
*Annals of Economic and Social Measurement*

**Granger**

1969 Investigating casual relations by econometric models and cross spectral analysis
*Econometrica*

**Engle and Granger**

1983 Applications of spectral analysis in econometrics. In Krishniah & Brillinger (eds.), *The Handbook of Statistics*, vol. 3, *Time Series and the Frequency Domain*, Amsterdam: North-Holland.

Econometricians have told me that they do not want to invest their time in learning about frequency domain time series analysis and spectral analysis because they can solve their problems in the time domain. I argue that economists should be educated about the broad scientific impact of the theory and practice of statistical spectral analysis in the study of processes. They should learn concepts of linear time invariant filter and Fast Fourier Transform as the fast algorithm to compute sample correlations.

Any real world system can be regarded as a *filter* (black box) which links inputs and outputs. To understand operations on a time series like first difference or moving average it helps to view your new time series as the output of a “filter” whose input is the old time series.

If you love dynamic data you should appreciate spectral analysis which provides a universal nonparametric statistical method to summarize a time series $Y(t), t = 1, \cdots, n$.

The dichotomy between static data and dynamic data manifests itself in the two ways available to develop theory of spectral analysis as ensemble averages and time averages. We mentioned these two ways when we defined ergodic data.

The covariance function $R(h)$ of a time series can be defined by an ensemble average

$$R(h) = E[Y(t)\, Y(t+h)]$$

or a time average

$$R(h) = \lim_{n \to a} \frac{1}{n} \sum_{t=1}^{n-h} Y(t)\, Y(t+h).$$

A controversy which seems to be still alive is whether the definition of $R(h)$ should divide by $n$ or $n-h$ (misled by the Holy Grail of being unbiased).

Important tools are spectral distribution and log spectral density. The Fourier transform of the log spectral density, called "cepstral" coefficient, can be used to learn time domain models (such as AR or MA) from first estimating the log spectral density (see 2004 thesis of my Ph.D. student Scott Holan, "*Time Series Exponential Models: Theory and Methods*").

We can make a long list of the contributions of spectral analysis to innovations of statistical methods that are broadly practiced.

(1) John Tukey's "exploratory data analysis"
(2) Hirotugu Akaike's information criterion AIC (celebrated in 2003 in Japan at a conference on "Science of Modeling").
(3) Statistical inference by Hilbert space methods, probability priors on function space, Reproducing Kernel Hilbert Space methods of learning from data.

To maintain our long memory of time series analysis I include a list of diverse approaches to spectral density estimation.

(1) Direct: kernel smoothing of sample spectral density and spectral distribution
(2) Indirect: kernel smoothing of sample correlations
(3) Direct: kernel smooth of log of narrow-band smoothed sample spectrum
(4) Indirect: kernel smoothing of sample cepstral correlations
(5) AR model selection, Yule Walker estimates
(6) AR model selection, Burg estimates
(7) EXP spectral estimators (Exponential model for spectral density)
(8) Wavelet estimators of log spectral density; Percival and Walden (2000) present an excellent discussion of wavelet methods for estimating log spectral density.
(9) FEXP spectral estimators (Fractional exponential long range model); Velasco and Robinson (2000) provide an elegant treatment of the role of tapering and FEXP models.

## 6. Forecasting, Prediction Theory, Box Jenkins Models

The art and science of forecasting is an industry to which Prof. Granger has made award winning contributions (his 1993 paper "Forecasting Stock Prices: Lessons for Forecasters," *International Journal of Forecasting*).

In the (Hilbert space) theory of prediction of stationary time series an important concept is "innovations", the prediction errors $Y^n(t)$ of a one-step ahead infinite memory predictor $Y^m(t)$ of a time series $Y(t)$. A time series model is a filter whose input is the original time series and whose output is the innovations, a time series with no memory or white noise.

For long memory (evolutionary) time series we can in practice find several model representations (filters) whose innovations have the same properties (such as mean square error). To select a model for a time series, and formulas for predictions for several steps ahead, one approach is to find a filter (or several filters in series) whose input is the observed time series and whose output is white noise innovations. George Box and Gwilym Jenkins in their 1970 influential book "*Time Series Analysis, Forecasting and Control*" pioneered the ARIMA approach to time series model identification which taught us to think about data modeling, that statistical science is NOT just about estimation, testing, or even data analysis. Data modeling is done in systematic way by an iterated series of steps which I formulate as SIEVE: *sample* statistics, *identify* parametric model, *estimate* parameters and orders, *validate* goodness of fit, and *estimation* non-parametric.

Interval forecasts and density forecasts are being increasingly used in practical real time forecasting. I love the title of the 1989 paper by Granger, White, and Kamstra: "Interval Forecasting: An Analysis Based on ARCH-Quantile Estimators," *Journal of Econometrics*.

Multi-step forecasting of long memory time series is a very active research field. We are enthusiastic about FEXP (fractional exponential) models, the long memory generalization of the EXP (exponential) model proposed by Bloomfield (1973) but only popular (cited) in recent years. Hurvich (2002) develops FEXP forecasting methodology.

On the topic of selecting models I cannot resist quoting my own philosophy about the nature of error terms $e(t)$.

"None of the previous work should be construed as a demonstration of the inevitability of MA disturbances in econometric models. As Parzen's proverb reminds us the disturbance term is essentially man-made, and it is up to man to decide if some of his creations are more reasonable than others." Nicholls, Pagan, and Terrell (1975) p. 117.

"God made X (the data), man made all the rest (especially, the error term)." Quoted in Nicholls et al (1975). We add that in many instances humans, not God, construct the data. (Carter and Zellner report on the ARAR Error Model).

## 7. Model Selection, Akaike Information Criterion

Criteria for choosing model orders (the number of parameters in a time domain model of a time series) have a history of 30 to 40 years, and have had an enormous impact on the broad discipline of statistics. The most prominent innovation has been Akaike's Information Criterion, which in 2003 celebrated its 30$^{th}$ anniversary. Of historical interest is Parzen's CAT criterion for AR model selection.

Goodness of fit theory (or lack of fit theory) and spectral distribution tests for white noise benefit from model selection criteria. The idea that "a model fits if it predicts" can be attained by testing zero values for components orthogonal to the components used to estimate parameters. The theory and practice of model selection and model fit is still one of the outstanding problems of statistical science.

## 8. State Space, Unobserved Components

Every statistician should learn about theory and practice of (1) static data modeling of a process $Y(t)$ by model

$$Y(t) = m + s \;\; e(t)$$

where $e(t)$ is Normal (0,1) white noise and $\mu$ and $s$ are parameters to be estimated, and (2) dynamic data modeling by Kalman filter model

$$Y(t) = m(t) + s \;\; e(t)$$

$$m(t) = m(t-1) + s_1 \; e_1(t)$$

where $e(t), e_1(t)$ are Normal (0,1) white noise.

Reconciling the static and dynamic models is easier if one adopts a Bayesian interpretation of the parameter $\mu$ and defines the problem of statistical inference as computing the posterior quantile function of $\mu$ given the data. My recent research is concerned with showing that quantile functions provide unified summaries of data (sample distributions), population probability distributions, p-values of parameters, and posterior distributions of parameters.

An excellent review of theory and practice (in honor of the outstanding career of Jim Durbin) is available in the 2004 book *State Space and Unobserved Component Models*, edited by Andrew Harvey, Jan Koopman, and Neil Shephard, Cambridge University Press. The book concludes with a paper "Fitting Genes to the Human Genome with Hidden Markov Models" by Richard Durbin, FRS, son of Jim Durbin. The work of son and father, in biology and economics, are linked by applying statistical methods for dynamic data analysis.

## 9. Unit Roots, Cointegration

Time series econometrics has added to the general theory of statistical time series analysis the theory of univariate unit roots (which determine if a stochastic trend is present in a time series) and the theory of cointegration (which investigates empirically long run relationships among a number of variables). These theories have had an enormous influence.

## 10. ARCH

Classical time series models provide models for the conditional mean of $Y(t)$ given past values. ARCH was innovated by Engle to model time varying conditional variances and volatility (empirically observed to have highs and lows). Its generalizations have been very influential in practice. We should remember Engle's pioneering papers:

1982 Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. *Econometrica*.

1983 With D. Kraft. Multiperiod forecast error variances of inflation estimated from ARCH models. In A. Zellner (ed.), *Applied Time Series Analysis of Economic Data.* Washington, D.C.: Bureau of the Census.

**11. Nonlinear Models, Nonparametric Models**

Models which relax the assumptions of the classical Gaussian linear stationary models for time series are clearly important. I would appreciate opinions on how successful they have been in applications to economics.

**12. Pattern Inference and the Future of Time Series Analysis**

Time series analysis, signal processing engineering, and communication theory have modern extensions in pattern and image analysis. Research on general pattern theory by Ulf Grenander stimulated Donald and Stuart Geman in the early 1980's to forge a link between digital image analysis and Monte Carlo Markov Chain methods for statistical inference.

A unified approach to modern statistical inference can be developed which practices "Data Analysis = Data Modeling = Data Simulation." The long history of time series analysis can be expected to continue in which its future shape will reflect the many exciting developments that we can expect to happen in modern computer intensive statistical practice and theory.

Research topics certain to be important are vector (multiple) time series forecasting and principal components. Important research is being done in Texas by Mel Hinich, Buddy Gray, and Willa Chen. A concise 3 page survey of time series analysis is the 2000 paper of Victor Solo, "The End of Time Series," *Journal of the American Statistical Association*. A review of the impact of modern statistical time series analysis should mention the influential research of David Brillinger.

**13. How Statistics Revolutionized Science in the 20th Century**

A major theme of this paper has been the impact on the broad discipline of statistics of dynamic data, ergodic data, and statistical time series analysis. A charming and informative book on the history, philosophy, and timeline of statistics is 2001 book of David Salsburg "*The Lady Tasting Tea, How Statistics Revolutionized Statistics in the Twentieth Century,*" Freeman. In his conclusion Salsburg mentions the implications of lack of references to innovative mathematical statisticians in many re-inventions by applied researchers, and discusses examples of controversy about how to apply probability to real life if we adopt the sample space axiomatic Kolmogorov definition of probability. Salsburg asks if these problems make probability irrelevant to research on a world of change.

I am optimistic that statisticians are solving the problem of modeling change and that we are successfully adapting static data statistical methods to dynamic data in a way that is objective by practical standards. One key is for us to provide answer machines, not answers. To cope with the fact that we are uncertain about assumptions we should compute and compare comprehensive answers under diverse assumptions.

Salsburg's book has many wonderful stories about the careers of statisticians. I reprint for economists a story Salsburg tells about Wassily Leontief (page 177). I first tell an amusing story about Leontief. He went to a cemetery office in Connecticut to buy a cemetery plot. When asked his profession by the office manager, he replied "An economist, and a very good one." The manager replied that the cemetery had a famous economist Schumpeter, and the plot next to him was for sale by his family. Leontief, buying the plot, noted "That Schumpeter, he always was a speculator."

"I was on a committee with Jerry Cornfield, convened in 1973 as part of a set of hearings before a Congressional committee. During a break in our work, Cornfield was called to the phone. It was Wassily Leontief, an economist at Columbia University, calling to say that he had just been awarded the Nobel Prize in economics and wanted to thank Cornfield for the role Jerry had played in the late 1940s when Leontief had come to the Bureau of Labor Statistics for help.

Leontief believed that the economy could be broken down into sectors, like farming, steel manufacturing, retailings, and so forth. Each sector uses material and services from the other sector to produce material or a service, which it supplies to those other sectors. This interrelationship can be described in the form of a mathematical matrix. It is often called an "input-output analysis." When he first began investigating this model at the end of World War II, Leontief went to the Bureau of Labor Statistics to help gather the data he needed. To assist him, the bureau assigned a young analyst who was working there at the time, Jerome Cornfield.

Leontief could break the economy down into a few broad sectors, such as putting all manufacturing into one sector, or he could subdivide the sectors into more specific ones. The mathematical theory of input-output analysis requires that the matrix that describes the economy have a unique inverse. That meant that the matrix, once assembled, had to be subjected to a mathematical procedure called "inverting the matrix." At that time, before the widespread availability of computers, inverting a matrix was a difficult and tedious procedure on a calculator. When I was in graduate school, each of us had to invert a matrix – I suspect as a kind of rite of passage "for the good of our souls." I remember trying to invert a 5 x 5 matrix and taking several days, most of which I spent locating my mistakes and redoing what I had done wrong.

Leontief's initial set of sectors led to a 12 x 12 matrix, and Jerry Cornfield proceeded to invert that 12 x 12 matrix to see if there was a unique solution. It took him about a week, and the end result was the conclusion that the number of sectors had to be expanded. So, with trepidation, Cornfield and Leontief began subdividing the sectors until they ended with the simplest matrix they thought would be feasible, a 24 x 24 matrix. They both knew this was beyond the capacity of a single human being. Cornfield estimated that it would take him several hundred years of seven-day work-weeks to invert a 24 x 24 matrix.

During World War II, Harvard University had developed one of the first, very primitive computers. It used mechanical relay switches and would often jam. There was no longer any war work for it, and Harvard was looking for applications for its monstrous machine. Cornfield and Leontief decided to send their 24 x 24 matrix to Harvard where its Mark I computer would go through the tedious calculations and compute the inverse. When they sought to pay for this project, the process was stopped by the accounting office of the Bureau of Labor Statistics. The government has a policy at that time; it would pay for goods but not for services. The theory was that the government had all kinds of experts working for it. If something had to be done, there should be someone in government who could do it.

They explained to the government accountant that, while this was theoretically something that a person could do, no one person would be able to live long enough to do it. The accountant was sympathetic, but he could not see a way around the regulation. Cornfield then made a suggestion. As a result, the bureau issued a purchase order for capital goods. What capital goods? The invoice called for the bureau to purchase from Harvard "one matrix, inverted.""

**14. On Probability as a Basis for Action, W. Edwards Deming**

"Abstract:

The aim of the author is improvement of statistical practice. The author distinguishes between enumerative studies and analytic studies. An enumerative study has for its aim an estimate of the number of units of a frame that belong to a specified class. An analytic study has for its aim a basis for action on the cause-system or the process, in order to improve product of the future. A fair price to pay for an inventory is an example of an enumerative study. Tests of varieties of wheat, insecticides, drugs, manufacturing processes, are examples of analytic studies: the choice of variety or treatment will affect the future out-turn of wheat, future patients, future product. Techniques and methods of inference that are applicable to enumerative studies lead to faulty design and faulty inference for analytic problems.

It is possible, in an enumerative problem, to reduce errors of sampling to any specified level. In contrast, in an analytic problem, it is impossible to compute the risk of making a wrong decision. The author provides a number of examples, and pleads for greater care in the writing and teaching of statistical theory and inference."