of Demand," *Journal of Marketing Research*, 35, 384–389.
Allenby, G. M., and Ginter, J. L. (1995), "Using Extremes to Design Products and Segment Markets," *Journal of Marketing Research*, 32, 392–403.
Allenby, G. M., and Lenk, P. J. (1994), "Modeling Household Purchase Behavior With Logistic Normal Regression," *Journal of the American Statistical Association*, 89, 1218–1231.
Allenby, G. M., Leone, R. P., and Jen, L. (1999), "A Dynamic Model of Purchase Timing With Application to Direct Marketing," *Journal of the American Statistical Association*, 94, 365–374.
Allenby, G. M., and Rossi, P. E. (1999), "Marketing Models of Consumer Heterogeneity," *Journal of Econometrics*, 89, 57–78.
Bagozzi, R. P., Yi, Y., and Nassen, K. (1999), "Representation of Measurement Error in Marketing Variables: Review of Approaches and Extension to Three-Facet Designs," *Journal of Econometrics*, 89, 393–421.
Boatwright, P., and Rossi, P. E. (1999), "Estimating Price Elasticities With Theory-Based Priors," *Journal of Marketing Research*,
Bradlow, E. T., and Schmittlein, D. C. (1999), "The Little Engines That Could: Modeling the Performance of World Wide Web Search Engines," *Marketing Science*, forthcoming.
Bradlow, E. T., and Zaslavsky, A. M. (1999), "A Hierarchical Latent Variable Model for Ordinal Data From a Customer Satisfaction Survey With 'No Answer' Response," *Journal of the American Statistical Association*, 94, 43–52.
Chipman, H., George, E., and McCulloch, R. (1998), "Bayesian CART Model Search," *Journal of the American Statistical Association*, 93, 935–960.
DeSarbo, W. S., Munrai, A. K., and Munrai, L. A. (1994), "Latent Class Multi-Dimensional Scaling: A Review of Recent Developments in the Marketing and Psychometric Literature," in *Advanced Methods for Marketing Research*, ed. R. P. Bagozzi, Cambridge, U.K.: Blackwell.
Erdem, T., and Keane, M. P. (1996), "Decision-Making Under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets," *Marketing Science*, 15, 1–20.
Gelfand, A. E., and Smith, A. F. M. (1990), "Sampling-Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, 85, 398–409.
George, E., and McCulloch, R. (1995), "Variable Selection via Gibbs Sampling," *Journal of the American Statistical Association*, 88, 881–889.
Geweke, J. (1989), "Simulation Estimation Methods for Limited Dependent Variable Models," in *Handbook of Statistics*, Vol. 11, eds. G. S. Maddala, C. R. Rao, and H. D. Vinod, Amsterdam: North-Holland.
Gonul, F., and Srinivasan, K. (1996), "Estimating the Impact of Consumer Expectations of Coupons on Purchase Behavior: A Dynamic Structural Model," *Marketing Science*, 15, 262–279.
Greenleaf, E. (1992), "Improving Ratings Scale Measures by Detecting and Correcting Bias Components in Some Response Styles," *Journal of Marketing Research*, 29, 176–188.
Kamakura, W., and Russell, G. (1989), "A Probabilistic Choice Model for Market Segmentation and Elasticity Structure," *Journal of Marketing Research*, 26, 379–90.
Lenk, P., DeSarbo, W., Green, P., and Young, M. (1996), "Hierarchical Bayes Conjoint Analysis: Recovery of Partworth Heterogeneity from Reduced Experimental Designs," *Marketing Science*, 15, 173–191.
Mehta, N., Rajiv, S., and Srinivasan, K. (1999), "Active Versus Passive Loyalty: A Structural Model of Consideration Set Formation," working paper, Carnegie-Mellon University.
Montgomery, A. L. (1997), "Creating Micro-Marketing Pricing Strategies Using Supermarket Scanner Data," *Marketing Science*, 16, 315–337.
Neelemegham, R., and Chintagunta, P. (1999), "A Bayesian Model to Forecast New Product Performance in Domestic and International Markets," *Marketing Science*, 18, 115–136.
Ripley, B. D. (1996), *Pattern Recognition and Neural Networks*, Cambridge, U.K.: Cambridge University Press.
Rossi, P. E., Gilula, Z., and Allenby, G. M. (1999), "Overcoming Scale Usage Heterogeneity: A Bayesian Hierarchical Approach," working paper, University of Chicago, Graduate School of Business.
Rossi, P. E., McCulloch, R. E., and Allenby, G. M. (1996), "The Value of Purchase History Data in Target Marketing," *Marketing Science*, 15, 321–340.

Time Series and Forecasting: Brief History and Future Research

Ruey S. TSAY

1. A BRIEF HISTORY

Statistical analysis of time series data started a long time ago (Yule 1927), and forecasting has an even longer history. Objectives of the two studies may differ in some situations, but forecasting is often the goal of a time series analysis. In this article I focus on time series analysis with an understanding that the theory and methods considered are foundations and tools important in forecasting.

Applications played a key role in the development of time series methodology. In business and economics, time series analysis is used, among other purposes, (a) to study the dynamic structure of a process, (b) to investigate the dynamic relationship between variables, (c) to perform seasonal adjustment of economic data such as the gross domestic product and unemployment rate, (d) to improve regression analysis when the errors are serially correlated, and (e) to produce point and interval forecasts for both level and volatility series.

To facilitate discussion, I denote a time series at time index t by z_t and let Ψ_{t-1} be the information set available at time $t-1$. It is often assumed, but not necessarily so, that Ψ_{t-1} is the σ field generated by the past values of z_t. A model for z_t can be written as

$$z_t = f(\Psi_{t-1}) + a_t \tag{1}$$

where a_t is a sequence of iid random variables with mean 0 and finite variance σ_a^2. It is evident from the equation that a_t is the one-step-ahead forecast error of z_t at time origin $t-1$ and, hence it is often referred to as the innovation or shock of the series at time t. The history of time series analysis is concerned with the evolution of the function $f(\Psi_{t-1})$ and the innovation $\{a_t\}$.

Ruey S. Tsay is H. G. B. Alexander Professor of Econometrics and Statistics, Graduate School of Business, University of Chicago, IL 60637 (E-mail: *ruey.tsay@gsb.uchicago.edu*). This work was supported by the National Science Foundation, the Chiang Ching-Kuo Foundation, and the Graduate School of Business, University of Chicago.

Journal of the American Statistical Association
June 2000, Vol. 95, No. 450, Vignettes

The series z_t is said to be (weakly) stationary if its first two moments are time-invariant under translation. That is, the expectation $E(z_t) = \mu$ is a constant, and the lag covariance function $\gamma_l = \text{cov}(z_t, z_{t-l})$ is a function of l only. The autocorrelation function of a stationary z_t is simply $\rho_l = \gamma_l/\gamma_0$. For a linear series, understanding the behavior of ρ_l is the key to time series analysis.

1.1 Unification in Research

The publication of *Time Series Analysis: Forecasting and Control* by Box and Jenkins in 1970 was an important milestone for time series analysis. It provided a systematic approach that enables practitioners to apply time series methods in forecasting. It popularized the autoregressive integrated moving average (ARIMA) model by using an iterative modeling procedure consisting of identification, estimation, and model checking. In the framework of (1), an ARIMA(p, d, q) model assumes the form

$$w_t = (1 - B)^d z_t,$$

$$f(\Psi_{t-1}) = c + \sum_{i=1}^{p} \phi_i w_{t-i} - \sum_{j=1}^{q} \theta_j a_{t-j}, \qquad (2)$$

where $p, d,$ and q are nonnegative integers; c is a constant; and B is the backshift operator such that $Bz_t = z_{t-1}$. The series w_t is referred to as the dth differenced series of z_t. Using polynomials, one can write the ARIMA model in a compact form, $\phi(B)(1 - B)^d z_t = c + \theta(B)a_t$, where $\phi(B) = 1 - \sum_{i=1}^{p} \phi_i B^i$ and $\theta(B) = 1 - \sum_{i=1}^{q} \theta_i B^i$ are two polynomials in B. The two polynomials $\phi(B)$ and $\theta(B)$ have no common factors, and their 0's are assumed to be outside the unit circle. In practice, it is common to assume further that a_t is Gaussian. The foregoing assumptions imply that z_t is stationary if $d = 0$. When $d \neq 0$, z_t is said to contain a unit root or to be unit-root nonstationary.

Once an ARIMA model is built and judged to be adequate, forecasts of future values are simply the conditional expectations of the model if one uses the minimum mean squared error as the criterion. The success of ARIMA models generated substantial research in time series analysis. However, the history of time series analysis was not as smooth as one might think. To begin with, time series analysis was originally divided into *frequency domain* and *time domain* approaches. Proponents of the two approaches did not necessarily see eye to eye, and there were heated debates and criticisms between the two schools. The time domain approach uses autocorrelation function ρ_l of the data and parametric models, such as the ARIMA models, to describe the dynamic dependence of the series (see Box, Jenkins, and Reinsel 1994 and references therein). The frequency domain approach focuses on spectral analysis or power distribution over frequency to study theory and applications of time series analysis. A power spectrum of a stationary z_t is the Fourier transform of the autocorrelation function ρ_l (see Brillinger 1975 and Priestley 1981 and references therein). Cooley and Tukey (1965) made an important advance in frequency-domain analysis by making spectral estimation efficient.

Times change, and it is fair to say that the sharp separation between the two approaches is gone. Now, the objective of an analysis and experience of the analyst are the determining factors between which approach to use. Likewise, the difference between Bayesian and non-Bayesian time series analyses is also diminishing. There remain some differences between Bayesian and non-Bayesians in our profession, but the issue has been shifted those of practicality rather than philosophy (see Kitagawa and Gersch 1996 and West and Harrison 1989 and the references therein for Bayesian time series analysis.) Durbin and Koopman (2000) provided both classical and Bayesian perspectives in time series analysis.

1.2 Technical Developments

The advances in computing facilities and methods have profound impacts on time series analysis. Within the so called "traditional analysis" (i.e., linear Gaussian processes with parametric models) there are many important developments. Outlier analysis and detecting structural breaks have become an integral part of model diagnostics (see, e.g., Chang, Tiao, and Chen 1988 for outlier detection and Martin and Yohai 1986 for influential functionals). Outlier analysis in time series is concerned with aberrant observations in z_t and a_t and the changes in the mean of z_t and the variance of a_t. Many model selection criteria have been proposed to help model selection (e.g., Akaike 1974, Hannan 1980), and some important advances in pattern identification methods have also been developed; for example, the R- and S-array of Gray, Kelley, and McIntire (1978) and the extended autocorrelation function of Tsay and Tiao (1984) that is capable of handling both stationary and unit-root nonstationary series. Indeed, there have many developments in ARMA model identification; (see Choi 1992). The exact likelihood method now becomes the standard method of estimation (Ansley 1979; Hillmer and Tiao 1979). The foregoing developments are not in isolation with other developments in the area, and their impacts are not limited to linear Gaussian time series models.

Generally speaking, two important technical advances in the recent history of time series analysis have generated much interest on the topic. The first advance is the use of state-space parameterization and Kalman filtering (Harrison and Stevens 1976; Kalman and Bucy 1961 and the references therein). This happened largely in the 1980s, as evidenced by the explosion in the papers published in statistical journals that have "state-space" or "Kalman filter" in their titles. A simple state-space model for z_t can be written as

$$z_t = \mathbf{H}\mathbf{S}_t, \qquad \mathbf{S}_{t+1} = \mathbf{F}\mathbf{S}_t + \mathbf{R}_t, \qquad (3)$$

where $\mathbf{S}_t$ is the state vector at time t, $\mathbf{H}$ is a row vector relating the observation z_t to the state vector, $\mathbf{F}$ is a state transition matrix, and $\mathbf{R}_t$ is a state innovation with mean 0 and a fixed nonnegative definite covariance matrix. The first equation in (3) is called the observation equation; the second, the state equation. An ARIMA model in (2) can be put into a state-space model in (3). Similarly, a state-space model in (3) implies an ARIMA model for z_t. However, the correspondence between state-space models in (3) and ARIMA models in (2) is not one-to-one; (see Aoki 1987).

The original purpose of introducing Kalman filter into time series analysis was mainly to evaluate efficiently the exact Gaussian likelihood function of a model and to handle missing observations (Jones 1980). The exact likelihood function of a model can be written as a product of consecutive conditional distributions, $p(z_1, \ldots, z_n) = \prod_{i=1}^{n} p(z_i|\Psi_{i-1})$, and the Kalman filter provides closed-form formulas for the evolutions of the conditional mean and conditional variance of z_t. The usefulness of the technique was extended beyond estimation, however. It led to developments of new methods for signal extraction (Kohn and Ansley 1989; Wecker and Ansley 1983), for smoothing and seasonal adjustment (Kitagawa and Gersch 1996), and for renewal interest in structural models (Harvey 1989), among many others.

The second technical advance in recent time series analysis is the use of Markov chain Monte Carlo (MCMC) methods, especially Gibbs sampling (Gelfand and Smith 1990), and the idea of data augmentation (Tanner and Wong 1987). The applicability of MCMC methods to time series analysis is widespread, and indeed the technique has also led to various new developments in time series analysis. (See, e.g., Carlin, Polson, and Stoffer 1992 for nonnormal and nonlinear state-space modeling and McCulloch and Tsay 1993 for inference and prediction of autoregressive models with random mean and variance shifts, including using explanatory variables to estimate transition probabilities in mean and variance.) The MCMC methodology also led to increasing use of simulation methods in time series analysis, especially in tackling complicated problems that were impossible to handle a few years ago.

1.3 Methodological Developments

The past several decades also brought many important advances in time series methodology.

Nonlinearity and Nonnormality. Theory and methods have been developed for many nonlinear and non-Gaussian models, marking a generalization of the functional form $f(\Psi_{t-1})$ in (1). In the statistical literature, Tong (1990) provided an excellent summary of recent developments in nonlinear models, such as bilinear models and threshold autoregressive (TAR) models. For a bilinear model, $f(\Psi_{t-1})$ involves some cross-product terms between z_{t-i} and a_{t-j}, where $i, j > 0$. A TAR model, on the other hand, uses a piecewise linear model for $f(\Psi_{t-1})$ over a threshold space. For example, the model

$$f(\Psi_{t-1}) = \begin{cases} \phi_1 z_{t-1} & \text{if } z_{t-1} \geq r \\ \phi_2 z_{t-1} & \text{if } z_{t-1} < r, \end{cases}$$

where $\phi_1 \neq \phi_2$, is called a two-regime TAR model of order 1 with threshold variable z_{t-1} and threshold r. This simple TAR model is piecewise linear in the space of z_{t-1}, not in time. It can produce several nonlinear characteristics commonly observed in practice, such as asymmetry between increasing and declining patterns in a periodic time series. Many tests now are available to detect nonlinearity in a series, and some studies show that nonlinear models can indeed improve forecasting in certain situations (see, e.g., Montgomery, Zarnowitz, Tsay, and Tiao 1998).

In the econometric literature, some important developments in nonlinear models have also emerged. The advance, however, focuses on the evolution of the conditional variance of a_t in (1). Let $h_t = E(a_t^2|\Psi_{t-1})$ be the conditional variance of a_t given Ψ_{t-1}. This quantity is referred to as the volatility of a_t in econometrics and finance. The autoregressive conditional heteroscedastic (ARCH) model of Engle (1982) assumes that h_t is a positive deterministic function of Ψ_{t-1}. For the simplest ARCH(1) model, h_t becomes $h_t = \alpha_0 + \alpha_1 a_{t-1}^2$, where $\alpha > 0$ and $1 > \alpha_1 \geq 0$. A special feature of the ARCH model is that under certain conditions the innovation a_t is heavy-tailed distributed and hence is unconditionally non-Gaussian even though it is conditionally Gaussian. Another feature of ARCH models is volatility clustering. Because volatility plays an important role in options pricing, the ARCH model has attracted much attention. Its also leads to the introduction of generalized ARCH (GARCH) models and stochastic volatility models. In general, the ARCH-family models use a positive deterministic (typically quadratic) equation to govern the evolution of h_t over time, whereas stochastic volatility models allow h_t to have its own innovational series. For example, an GARCH(1, 1) model assumes the form

$$h_t = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 h_{t-1}, \tag{4}$$

where $\alpha_0 > 0, \alpha_1 \geq 0, \beta_1 \geq 0$ and $1 > \alpha_1 + \beta_1$. A simple stochastic volatility model may assume the form $\ln(h_t) = \alpha_0 + \ln(h_{t-1}) + v_t$, where v_t is a white noise sequence. The added innovation v_t increases the flexibility of the model as well as complications. A GARCH model has an ARMA-type representation, so that many of its properties are similar to those of ARMA models. For example, let $\eta_t = a_t^2 - h_t$, which is the squared innovation without its conditional expectation. Then the GARCH(1, 1) model in (4) can be written as

$$a_t^2 = \alpha_0 + (\alpha_1 + \beta_1) a_{t-1}^2 + \eta_t - \beta_1 \eta_{t-1}, \tag{5}$$

where it is easy to see that η_t is a martingale difference. The stationarity condition $\alpha_1 + \beta_1 < 1$ of a GARCH(1, 1) becomes apparent from (5). The econometric literature also concerns the evolution of $f(\Psi_{t-1})$. The Markov switching model of Hamilton (1989) uses a state variable to govern the choice of a linear model for $f(\Psi_{t-1})$. The state variable then evolves over time based on a transition probability matrix.

Multivariate Process. Methods for analyzing multivariate series have been developed, especially in structural specification of a vector system. The usefulness and need of considering jointly several related time series were recognized a long time ago (see Quenouille 1957). However, multivariate analysis is often confined to vector autoregressive (VAR) models. One can think of many reasons for this lack of progress, but two reasons stand out:

a. The generalization of univariate ARMA models to vector ARMA models encounters the problem of identifiability. As a simple illustration, the following bivariate AR(1) and MA(1) models are identical:

$$\begin{bmatrix} z_{1t} \\ z_{2t} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} z_{1,t-1} \\ z_{2,t-1} \end{bmatrix} + \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix},$$

$$\begin{bmatrix} z_{1t} \\ z_{2t} \end{bmatrix} = \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix} - \begin{bmatrix} 0 & 0 \\ -2 & 0 \end{bmatrix} \begin{bmatrix} a_{1,t-1} \\ a_{2,t-1} \end{bmatrix}.$$

where z_{1t} and z_{2t} are two time series and $a_t = (a_{1t}, a_{2t})'$ is a sequence of independent bivariate Gaussian random vectors with mean 0 and a constant positive-definite covariance matrix. From the model, $z_{1t} = a_{1t}$, so that $z_{1,t-1} = a_{1,t-1}$. The equality of the two models then results from $z_{2t} = 2z_{1,t-1} + a_{2t} = 2a_{1,t-1} + a_{2t}$. Such exchangeable forms between AR and MA models cannot occur in the univariate case.

b. Multivariate models are much harder to estimate and to understand, and there is a propensity to use perceived simpler models.

These difficulties have been largely overcome. The identifiability problem can be solved by using either the Kronecker indices (Akaike 1976) or the scalar component model (SCM) of Tiao and Tsay (1989). The relationship between Kronecker indices and orders of SCM was given by Tsay (1991). Both the Kronecker index and SCM can easily identify the existence of exchangeable models; in fact, both methods identify the underlying structure of a linear system.

A related development in multivariate time series analysis is the cointegration of Engle and Granger (1987). Roughly speaking, cointegration means that a linear combination of marginally unit-root nonstationary series becomes a stationary series. It has become popular in econometrics because cointegration is often thought of as the existence of some long-term relationship between variables. In the statistical literature, the idea of a linear combination of unit-root nonstationary series becoming stationary was studied by Box and Tiao (1977). Associated with cointegration is the development of various test statistics to test for the number of cointegrations in a linear system. Despite the huge literature on cointegration, its practical importance is yet to be judged; (see Tiao, Tsay, and Wang 1993). This is due primarily to the fact that cointegration is a long-term concept that overlooks the practical effects of scaling factors of marginal series.

Long-Range Dependence. To model the fact that sample autocorrelation functions in some real-world time series tend to decay much slower than those of an ARMA model, Granger and Joyeux (1980) and Hosking (1981) introduced the idea of fractional difference to handle what is called long-range dependence in a series. This led to the development of autoregressive fractionally integrated moving average (ARFIMA) (p, d, q), model where d is a positive but noninteger real number. The long-range dependence occurs in many fields; (see Beran 1994). In finance, it was found that the squared or absolute returns of an asset typically exhibit long-range dependence (e.g. Ding, Granger, and Engle 1993).

Finally, models for time series that assume only discrete values are also available; see a recent review article by Berchtold and Raftery (1999), which contains many references. For forecasting, there is a trend toward using the predictive distribution rather the mean squared error to evaluate the accuracy of forecasts.

1.4 Theoretical Developments

The most recent widespread theoretical development in time series analysis is the theory of unit-root. In its simplest case, the theory is concerned with asymptotic properties of statistics of a random walk process, $z_t = z_{t-1} + a_t$, where $z_0 = 0$ and a_t is a martingale difference satisfying $E(a_t|\Psi_{t-1}) = 0$ and $E(|a_t|^{2+\delta}) < \infty$, where δ is a positive real number. Consider, for example, the ordinary least squares estimate $\hat{\phi} = (\sum_{i=1}^{n} z_{t-1}^2)^{-1}(\sum_{i=1}^{n} z_t z_{t-1})$ of the autoregression $z_t = \phi z_{t-1} + \varepsilon_t$, where ε_t denotes the error term. Unlike in the stationary case, here the limiting distribution of $\hat{\phi}$ is a functional of a standard Brownian motion. (See Chan and Wei 1988 for a comprehensive treatment of least squares estimates with various characteristic roots on the unit circle and Phillips 1987 for a nice treatment of time series regression with a unit-root.) This newly established asymptotic result is interesting in several ways. First, it is a case in which the convergence rate of $\hat{\phi}$ is n^{-1} not the usual $n^{-.5}$, where n is the sample size. Second, the mathematics involved are elegant. Third, the nonstandard limiting distribution is by itself exciting, opening a new set of testing problems that are of interest to many econometricians and statisticians. In particular, the unit-root test is to consider the null hypothesis $H_0 : \phi = 1$ versus the alternative hypothesis $H_a : \phi \neq 1$ (see Dickey and Fuller 1979).

The unit-root problem has attracted much attention because (a) it provides a formal test to determine the order of differencing in using ARIMA models, (b) it opens an area of testing in which the proper test statistic depends on the existence of a nonzero constant term c in (1) and the multiplicity of the unit root (i.e., d) in (2), and (c) the other AR and MA parameters, if they exist, become nuisance parameters that disappear asymptotically, but might have serious implications in finite-sample cases. A multivariate extension of Chan and Wei (1988) was given by Tsay and Tiao (1990) and that of the unit-root test is the so-called cointegration test. The unit-root problem has also been extended to the MA case; (see Davis and Dunsmuir 1996).

Another area of theoretical development is the theory of nonlinear models, especially the issue of geometrical ergodicity and stationarity of nonlinear processes. For the simple TAR(1) model, Chen and Tsay (1991) and Petruccelli and Woolford (1984) provided some interesting results. In general, useful theorems for studying stability of a nonlinear process were given by Meyn and Tweedie (1993) and Tong (1990, appendices).

The theory of long-range dependent processes and that of processes with heavy-tailed distributions also mark substantial progress; see Beran (1994), Resnick (1997), Samorodnitsky and Taqqu (1994), and the references therein.

2. FUTURE RESEARCH

An important driving force of future research in time series analysis is the advance in high-volume data acquisition. Consider, for instance, transaction-by-transaction data common in financial markets, or communications networks, and in e-commerce on the internet. These data are available now and must be processed properly and efficiently in a globally competitive business environment. But the special features of the data, such as large sample sizes, heavy tails, unequally-spaced observations, and mixtures of multivariate discrete and continuous variables, can easily render existing methods inadequate. Analysis of these types of data will certainly influence the directions of future time series research.

In my personal opinion, the following topics will attract substantial interest of time series researchers. First, the use of multivariate models either in a vector ARMA framework or a state-space form will increase, partly because of the need to study the dynamic relationships between variables and partly because of the advances in computing facilities. Part of the work here will be to identify common characteristics between the marginal series. Second, theory and applications of nonlinear and non-Gaussian models will continue to flourish. The development here is likely to move closely with nonparametric methods or computing intensive methods. Within the parametric framework, we can expect to see new models that use different equations to govern the evolutions of lower-order conditional moments. Third, heavy-tail modeling and extreme value analysis will become a necessity in some areas of application, such as telecommunication and high-frequency financial data analysis. Fourth, there will be a trend to study not only the usual time series data, but also the time duration between observations. In other words, times of occurrence will play an increasingly important role in time series analysis and forecasting. This line of research also marks a marriage between theories of time series analysis and point processes and between researchers in econometrics and statistics. Fifth, methods will be developed to efficiently and effectively process large-scale datasets.

In terms of technical developments, we will continue to see mixtures of Bayesian and Non-Bayesian analyses. The ideas of data augmentation and MCMC methods are here to stay and will flourish. Data mining will become part of time series data analysis, and we need to make good use of it.

REFERENCES

Akaike, H. (1974), "A New Look at the Statistical Model Identification," *IEEE Transactions on Automatic Control*, AC-19, 716–723.

——— (1976), "Canonical Correlation Analysis of Time Series and the Use of an Information Criterion," in *Systems Identification: Advances and Case Studies*, eds. R. K. Methra and D. G. Lainiotis, New York: Academic Press, pp. 27–99.

Ansley, C. F. (1979), "An Algorithm for the Exact Likelihood of a Mixed Autoregressive Moving Average Process," *Biometrika*, 66, 59–65.

Aoki, M. (1987), *State-Space Modeling of Time Series*, New York: Springer-Verlag.

Beran, J. (1994), *Statistics for Long-Memory Processes*, New York: Chapman and Hall.

Berchtold, A., and Raftery, A. E. (1999), "The Mixture Transition Distribution Model for High-Order Markov Chains and Non-Gaussian Time Series," Technical Report 360, University of Washington, Dept. of Statistics.

Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (1994), *Time Series Analysis: Forecasting and Control* (3rd ed.), Englewood Cliffs, NJ: Prentice-Hall.

Box, G. E. P., and Tiao, G. C. (1977), "A Canonical Analysis of Multiple Time Series," *Biometrika*, 64, 355–365.

Brillinger, D. R. (1975), *Time Series: Data Analysis and Theory*, New York: Holt, Rhinehart and Winston.

Carlin, B. P., Polson, N. G., and Stoffer, D. S. (1992), "A Monte Carlo Approach to Nonnormal and Nonlinear State-Space Modeling," *Journal of the American Statistical Association*, 87, 493–500.

Chan, N. H., and Wei, C. Z. (1988), "Limiting Distributions of Least Squares Estimates of Unstable Autoregressive Processes," *The Annals of Statistics*, 16, 367–401.

Chang, I., Tiao, G. C., and Chen, C. (1988), "Estimation of Time Series Parameters in the Presence of Outliers," *Technometrics*, 30, 193–204.

Chen, R., and Tsay, R. S. (1991), "On the Ergodicity of TAR(1) Processes," *Annals of Applied Probability*, 1, 613–634.

Choi, B. (1992), *ARMA Model Identification*, New York: Springer-Verlag.

Cooley, J. W., and Tukey, J. W. (1965), "An Algorithm for the Machine Calculation of Complex Fourier Series," *Mathematics of Computation*, 19, 297–301.

Davis, R., and Dunsmuir, W. (1996), "Maximum Likelihood Estimation of MA(1) Processes With a Root on or Near the Unit Circle," *Econometric Theory*, 12, 1–29.

Dickey, D. A., and Fuller, W. A. (1979), "Distribution of the Estimates for Autoregressive Time Series With a Unit Root," *Journal of the American Statistical Association*, 427–431.

Ding, Z., Granger, C. W. J., and Engle, R. F. (1993), "A Long Memory Property of Stock Returns and a New Model," *Journal of Empirical Finance*, 1, 83–106.

Durbin, J., and Koopman, S. J. (2000), "Time Series Analysis of Non-Gaussian Observations Based on State-Space Models From Both Classical and Bayesian Perspectives" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 62, 3–56.

Engle, R. F. (1982), "Autoregressive Conditional Heteroscedasticity With Estimates of the Variance of United Kingdom Inflations," *Econometrica*, 50, 987–1007.

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

Gelfand, A. E., and Smith, A. F. M. (1990), "Sampling-Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, 85, 398–409.

Granger, C. W. J., and Joyeux, R. (1980), "An Introduction to Long-Memory Time Series Models and Fractional Differencing," *Journal of Time Series Analysis*, 1, 15–29.

Gray, H. L., Kelley, G. D., and McIntire, D. D. (1978), "A New Approach to ARMA Modeling," *Communications in Statistics*, Part B, 7, 1–77.

Hamilton, J. D. (1989), "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57, 357–384.

Hannan, E. J. (1980), "The Estimation of the Order of an ARMA Process," *The Annals of Statistics*, 8, 1071–1081.

Harrison, P. J., and Stevens, C. F. (1976), "Bayesian Forecasting" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 38, 205–247.

Harvey, A. C. (1989), *Forecasting, Structural Time Series Models and the Kalman Filter*, Victoria, Australia: Cambridge University Press.

Hillmer, S. C., and Tiao, G. C. (1979), "Likelihood Function of Stationary Multiple Autoregressive Moving Average Models," *Journal of the American Statistical Association*, 74, 652–660.

Hosking, J. R. M. (1981), "Fractional Differencing," *Biometrika*, 68, 165–176.

Jones, R. H. (1980), "Maximum Likelihood Fitting of ARMA Models to Time Series With Missing Observations," *Technometrics*, 22, 389–395.

Kalman, R. E., and Bucy, R. S. (1961), "New Results in Linear Filtering and Prediction Theory," *Transactions of ASME, Journal of Basic Engineering*, 83D, 95–108.

Kitagawa, G., and Gersch, W. (1996), *Smoothness Priors Analysis of Time Series*, New York: Springer-Verlag.
Kohn, R., and Ansley, C. F. (1989), "A Fast Algorithm for Signal Extraction Influence and Cross-Validating in State-Space Models," *Biometrika*, 76, 65–79.
Martin, R. D., and Yohai, V. J. (1986), "Influence Functionals for Time Series," *The Annals of Statistics*, 14, 781–818.
McCulloch, R. E., and Tsay, R. S. (1993), "Bayesian Inference and Prediction for Mean and Variance Shifts in Autoregressive Time Series," *Journal of the American Statistical Association*, 88, 968–978.
Meyn, S., and Tweedie, R. L. (1993), *Markov Chains and Stochastic Stability*, New York: Springer-Verlag.
Montgomery, A. L., Zarnowitz, V., Tsay, R. S., and Tiao, G. C. (1998), "Forecasting the U.S. Unemployment Rate," *Journal of the American Statistical Association*, 93, 478–493.
Petruccelli, J., and Woolford, S. W. (1984), "A Threshold AR(1) Model," *Journal of Applied Probability*, 21, 270–286.
Phillips, P. C. B. (1987), "Time Series Regression With a Unit Root," *Econometrica*, 55, 277–301.
Priestley, M. B. (1981), *Spectral Analysis and Time Series*, Vols. I and II, London: Academic Press.
Quenouille, M. H. (1957), *The Analysis of Multiple Time Series*, London: Griffin.
Resnick, S. I. (1997), "Heavy Tail Modeling and Teletraffic Data," (with discussion), *The Annals of Statistics*, 25, 1805–1869.
Samorodnitsky, G., and Taqqu (1994), *Stable Non-Gaussian Random Processes*, New York: Chapman and Hall.
Tanner, M., and Wong, W. H. (1987), "The Calculation of Posterior Distributions," (with discussion), *Journal of the American Statistical Association*, 82, 528–550.
Tiao, G. C., and Tsay, R. S. (1989), "Model Specification in Multivariate Time Series," (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 51, 157–213.
Tiao, G. C., Tsay, R. S., and Wang, T. (1993), "Usefulness of Linear Transformation in Multivariate Time Series Analysis," *Empirical Economics*, 18, 567–593.
Tong, H. (1990), *Non-Linear Time Series: A Dynamical System Approach*, Oxford, U.K.: Oxford University Press.
Tsay, R. S. (1991), "Two Canonical Forms for Vector ARMA Processes," *Statistica Sinica*, 1, 247–269.
Tsay, R. S., and Tiao, G. C. (1984), "Consistent Estimates of Autoregressive Parameters and Extended Sample Autocorrelation Function for Stationary and Nonstationary ARMA Models," *Journal of the American Statistical Association*, 79, 84–96.
——— (1990), "Asymptotic Properties of Multivariate Nonstationary Processes With Applications to Autoregressions," *The Annals of Statistics*, 18, 220–250.
Wecker, W. E., and Ansley, C. F. (1983), "The Signal Extraction Approach to Nonlinear Regression and Spline Smoothing," *Journal of the American Statistical Association*, 78, 81–89.
West, M., and Harrison, P. J. (1989), *Bayesian Forecasting and Dynamic Models*, Berlin: Springer-Verlag.
Yule, G. U. (1927), "On a Method of Investigating Periodicities in Disturbed Series With Special Reference to Wölfer's Sunspot Numbers," *Philosophical Transactions of the Royal Society London*, Ser. A, 226, 267–298.

Contingency Tables and Log-Linear Models: Basic Results and New Developments

Stephen E. FIENBERG

1. HISTORICAL REMARKS ON CONTINGENCY TABLE ANALYSIS

Contingency table analysis is rooted in the turn-of-the-century work of Karl Pearson and George Udny Yule, who introduced the cross-product, or odds ratio, as a formal statistical tool. The subsequent contributions by R. A. Fisher linked their methods to basic statistical methodology and theory, but it was not until 1935 that Maurice Bartlett, as a result of a suggestion by Fisher, utilized Yule's cross-product ratio to define the notion of second-order interaction in a $2 \times 2 \times 2$ table and to develop an appropriate test for the absence of such an interaction (Bartlett 1935). The multivariate generalizations of Bartlett's work, beginning with a 1956 article by Roy and Kastenbaum, form the basis of the log-linear model approach to contingency tables, which is largely the focus of this vignette. Key articles in the 1960s by M. W. Birch (1963), Yvonne Bishop (1975), John Darroch (1962), I. J. Good (1963), Leo Goodman (1963), and Robin Plackett (1974), plus the availability of high-speed computers, led to an integrated theory and methodology for the analysis of contingency tables based on log-linear models, culminating in a series of books published in the 1970s. (Historical references can be found in various sources including Bishop, Fienberg, and Holland 1975, Carriquiry and Fienberg 1998, Fienberg 1980, and Haberman 1974.)

The next section outlines some of the basic results on likelihood estimation for log-linear models used to describe interactions in contingency tables, as the theory emerged by the early 1970s. I then briefly describe some of the major advances of the next three decades related to log-linear models. There is now an extensive literature on other classes of models and other methods of estimation, especially Bayesian, but I treat these only in passing, not because they are unimportant, but rather because they draw on similar foundations. Finally, I outline some important open research problems.

Many statisticians view the theory and methods of log-linear models for contingency tables as a special case of either exponential family theory or generalized linear models (GLMs) (Christensen 1996; McCullagh and Nelder 1989). It is true that computer programs for GLM often provide

Stephen E. Fienberg is Maurice Falk University Professor of Statistics and Social Science, Department of Statistics and Center for Automated Learning and Discovery, Carnegie Mellon University, Pittsburgh, PA 15213 (E-mail: *fienberg@stat.cmu.edu*). This work was supported by National Science Foundation grant no. REC-9720374.

Journal of the American Statistical Association
June 2000, Vol. 95, No. 450, Vignettes