Working Paper Series

WORKING PAPER NO. 866

SOME RECENT DEVELOPMENTS IN BAYESIAN STATISTICS AND ECONOMETRICS

by

Arnold Zellner

DEPARTMENT OF AGRICULTURE AND RESOURCE ECONOMICS AND POLICY
BERKELEY
CALIFORNIA AGRICULTURAL EXPERIMENT STATION
University of California

DEPARTMENT OF AGRICULTURAL AND RESOURCE ECONOMICS AND POLICY
DIVISION OF AGRICULTURE AND NATURAL RESOURCES
UNIVERSITY OF CALIFORNIA AT BERKELEY

WORKING PAPER NO. 866

SOME RECENT DEVELOPMENTS IN BAYESIAN STATISTICS AND ECONOMETRICS

by

Arnold Zellner

California Agricultural Experiment Station
Giannini Foundation of Agricultural Economics
September 1998

Some Recent Developments in Bayesian Statistics and Econometrics

by

Arnold Zellner
(U.of Chicago, visiting the U. of the Orange Free State)

Summary of presentation to Maxent '98 Meeting, Max Planck Institute for Plasma Physics, Garching b. Munich, Germany, July 27-31, 1998, and to be published in the Conference volume honoring Edwin T. Jaynes.

September, 1998

established various prizes, including the Savage Thesis Award, the Mitchell Award, etc. and published important Bayesian volumes that have had a great impact. Further, the number of published Bayesian research papers has increased amazingly over the years and some current and past editors of leading journals are Bayesians. Given these developments, it appears that we are witnessing the beginning of a Bayesian Era in statistics and econometrics, a major event in the history of science; see, e.g. Zellner (1988a).

I suggest that this upsurge in Bayesian research, applications and publications has occurred because there is a growing awareness of the impressive foundations of Bayesian analysis and, most importantly, that Bayesian analysis has produced many solutions to theoretical and practical problems that are as good as or better than those produced by other approaches. While a review of all areas of application is impractical, attention is focussed below on some recent developments in one important area of application, namely Bayesian forecasting. Then some uses of information theory in relation to Bayesian analysis will be presented. See, e.g the review articles on information theory and Bayesian analysis by Soofi (1996,1997), a collection of 48 papers on recent Bayesian research in Berry, Chaloner and Geweke (1996), the information theory text by Cover and Thomas (1991), applications of maxent to econometric problems in Fomby and Hill (1997) and Golan, Judge and Miller (1996) and articles on Bayesian analysis and information theory in Zellner (1997). In these works, there are many references to the pioneering research of Edwin T. Jaynes who has made many fundamental contributions to Bayesian and maximum entropy methodology and applications; see Jaynes (1974,1983,1984). A final section provides a brief summary and comments on prospects for the future.

II. Bayesian Modeling and Forecasting

The problem of model formulation is present in all fields of science. How to obtain models that explain the past well and predict accurately is a major issue. Prior views play an important role in model formulation in both Bayesian and non-Bayesian approaches. Generally, some, including Jeffreys, Jaynes, and many others, start with a sophisticatedly simple model and complicate it if necessary. However others, recommend starting with a complicated, general predictive densities to make point and turning point forecasts.

Hong (1989) made draws from the posterior density of the parameters of (1) to investigate dynamic properties of the models for each of the 18 countries in our sample. That is, he made draws from the marginal trivariate Student-t posterior density for the three autoregressive parameters and for each draw computed the roots of the third order process. He found that about 85% of the draws led to two complex conjugate roots and one real root; that is, the probability is about 0.85 that the process has an oscillatory component. From the drawn values, he computed posterior densities for the periods and amplitudes of the oscillatory components for each country's model and found amplitudes less than one and periods of about 4 to 5 years. . In addition, the implications of a uniform prior density for the three AR parameters for the properties of the roots of the process were investigated and found to be reasonable for the problem at hand. These results and the exact finite sample posterior densities for coefficients of our models and their predictive densities are very useful. In particular, means of one year ahead predictive densities that are optimal point forecasts relative to quadratic loss were employed and yielded forecasts that are better in terms of root mean squared error than those of random walk, AR(3), Nelson-Plosser and other benchmark models. The simple addition of leading indicator variables to our AR (3) model led to improved point forecasts and turning point forecasts.

Another important improvement in forecasting performance was provided by use of various forms of Bayes-Stein shrinkage. That is, the country coefficient vectors were assumed drawn from a normal density with a fixed, unknown mean vector. Using this assumption, and generalizations of it, Bayesian shrinkage forecasts were computed and generally were much better than non-shrinkage point and turning point forecasts as reported in papers in Zellner (1997) and also by Putnam and Quintana (1995), Quintana et al (1995), LeSage (1996) and others. In addition to shrinkage, we also introduced various time series processes for the random coefficient vectors in each of our country models, as have Putnam and Quintana (1995), Quintana et al (1995) and others. In the simplest case, we allowed each country's coefficient vector to follow a vector random walk, that is, $\beta_{it} = \beta_{it-1} + e_{it}$. Another model involved the assumption

Using these procedures, about 70 percent of 158 turning points were forecasted correctly in early work and in more recent work, about 70 percent of 211 were correctly forecasted.Here too, Bayesian shrinkage or pooling of information across equations for different countries produced improved turning point forecasts which are quite a bit better than those of naïve models, e.g. a coin flipper or an eternal optimist who always forecasts NDT at the top and UT at the bottom, etc.; see papers in Zellner (1997) and Zellner and Min (1998). In Zellner, Tobias and Ryu (1997) forecasts were defined to be of three types, namely minor DT, major DT and NDT at the top and minor UT, major UT and NUT at the bottom. Computing predictive probabilities for these events from predictive densities and using a 3x3 loss structure, it was possible to derive the minimum expected loss optimal turning point forecast, say minor DT.

That new Bayesian methods have proven fruitful in our and others' forecasting work is a major new development. Further, after having found an equation such as (1), or variants of it, that have been shown to work fairly well in forecasting, there is a challenge to subject matter theorists to explain this result. In this connection, Hong (1989) and Min (1992) have shown how equation (1) can be derived from forms of some macroeconomic theoretical models. In Zellner and Palm (1974,1976), Palm (1983) and Zellner (1994), the algebraic derivations of transfer function models, like that in (1), from general multivariate time series models are described and illustrated. In our work, we use information in specific simple relations like (1) that have been shown to work reasonably well in practice to formulate theoretical models that imply relations like (1) for these variables. Thus in line with the a priori advice of Jeffreys (1957, 1967), Jaynes (1980) and many others, we start simply and complicate if we have to do so in contrast to others who start with a general model and try to simplify it by testing downward. The latter approach involves a prior view that contrasts sharply with that of Jeffreys, Jaynes and many others and has yet to produce macroeconomic models that explain the past and forecast well. For further discussion of these simplicity/complexity issues and Jeffreys' measure of complexity, see Zellner (1998b). The role of prior information in model construction, e.g. the Jeffreys-Wrinch simplicity postulate that involves associating higher prior probabilities with sophisticatedly maximization…is thoroughly established…This makes it seem scandalous that the exact relation of entropy to the other principles of probability theory is still rather obscure and confused. But now we see that there is, after all, a close connection between entropy and Bayes's theorem. Having seen this start, other such connections may be found, leading to a more unified theory of inference in general. Thus in my opinion, Zellner's work is probably not the end of an old story, but the beginning of a new one."

As part of the "new story," Zellner (1991) has considered the prior and sample information inputs to be of differing quality in deriving an optimal information processing rule.
Also, one can weight the prior information and the sample information differently in deriving an optimal IPR which results in having the optimal output density function for the parameters being proportional to the prior raised to the power w times the likelihood function raised to the power 1-w, with $0<w<1$. Then there are dynamic information processing problems to solve that involve generation and processing of sequential information given various costs of acquiring new information and of adjusting to it. As Jaynes rightly pointed out, there are indeed many new stories to be developed. See also, Hill (1988), Cover and Thomas (1991), Bernardo and Smith (1994), Soofi (1996, 1997) and Zellner (1997) for recent considerations of information processing and new and old measures of the information provided by an experiment. The new definition in Zellner (1997, p.149), the information in the joint density of the observations and parameters minus the information in the prior is operational and more comprehensive than the old definition.

Another area of great activity has been research on what to do when the form of the likelihood is unknown. Of course, maxent has been and is being employed to produce models for observations in many areas, as Jaynes (1988) and Zellner (1991) have pointed out. Recently, Stutzer (1996) has shown how maxent can be used to produce "risk neutral" densities in financial economics while Golan, Judge and Miller (1996), Soofi (1996) and Zellner (1997, p. 100ff.) have shown how to produce many statistical and econometric models by maxent. These maxent if the measurements have been made properly with no left out variables, outliers, etc., the assumption that $E\bar{u}|D = 0$ can be made which implies $E\theta|D = \bar{y}$ and the maxent proper post data density for the parameter, subject to this mean constraint is the exponential density, $f(\theta|D) = (1/\bar{y})\exp\{-\theta/\bar{y}\}$. This is a Bayesian method of moments (BMOM) post data density based on one moment condition. See, e.g., Zellner (1994a, 1997, 1997a), Green and Strawderman (1996), Tobias and Zellner (1997) for applications of the approach, using additional moment restrictions on the realized errors, to a broad range of univariate and multivariate regression and other models. With additional restrictions on future error terms'moments, maxent predictive densities have been derived and used in Bayes' factors to compare alternative models in Tobias and Zellner (1997). In Golan, Judge and Miller (1996, 1997), the parameter and realized error term spaces are discretized and moments of realized errors and parameters are obtained and used in a maxent framework to get proper post data densities for parameters and for future observations. A number of applications are presented. Whether it is better to proceed in terms of a few parameters' moments as in the BMOM approach or many parameters associated with the discretization of the parameter space in empirical likelihood, nonparametric, and the Golan, Judge, Miller (1996) generalized entropy approaches is an issue that deserves attention. As Jeffreys and many others have recognized, introducing many parameters may lead to a good fit but will not generally lead to good predictions. It is possible to use Bayes' factors, see, e.g. Kass and Raftery (1995) and Tobias and Zellner (1997), and other model selection procedures and actual performance in prediction and forecasting to attempt to resolve this important range of issues.

IV Summary and Conclusions

As is evident from what has been presented above, information theory has and is playing an important role in Bayesian analysis in statistics and econometrics in producing prior densities, models for observations, information processing rules and post data densities for parameters and future observations when the form of the likelihood function is not known, etc.. As pointed out

Hill, B.M. (1988), "Comment," The American Statistician, 42, No. 4, Nov., 281-282.

Hong, C. (1989), "Forecasting Real Output Growth Rates and Cyclical Properties of Models: A Bayesian Approach," Ph.D. thesis, Dept. of Economics, U. of Chicago.

Jaynes, E.T. (1974), "Probability Theory with Applications in Science and Engineering: A Series of Informal Lectures," Dept. of Physics, Washington U., St. Louis, MO.

Jaynes, E.T. (1983), "Papers on Probability, Statistics and Statistical Physics," R.D. Rosenkrantz, ed., D. Reidl: Dordrecht, Netherlands.

Jaynes, E.T. (1984), "The Intuitive Inadequacy of Classical Statistics," Epistemologia VII (Special Issue on Probability, Statistics and Inductive Logic), 43-74.

Jaynes, E.T. (1985), "Highly Informative Priors, " in J.M. Bernardo, et al. (eds.), Bayesian Statistics 2, Amsterdam: North-Holland, 329-352.

Jaynes, E.T. (1988), "Comment," The American Statistician, 42, No. 4, Nov., 280-281.

Jeffreys, H. (1957), Scientific Inference, Cambridge:Cambridge U. Press.

Jeffreys, H. (1967), Theory of Probability, London: Oxford U. Press.

Kass, R.E. and A.E. Raftery (1995), "Bayes Factors," J. of the American Statistical Association, 90, 773-795.

Kass, R.E. and L. Wasserman (1996), "Formal Rules for Selecting Prior Distributions: A Review and Annotated Bibliography," J. of the American Statistical Association, 91, 1343-1370, corrections, 1998.

Keuzenkamp, H.A. and M. McAleer, "Simplicity, Scientific Inference and Econometric Modelling," The Economic Journal, 105, (January), 1-21.

Kuezenkamp, H.A. and M. McAleer, "The Complexity of Simplicity," Mathematics and Computers in Simulation," 43, 553-561.

LeSage, J.P.(1996), "A Comparison of Techniques for Forecasting Turning Points in Regional Employment Activity," Ch. 4, 41-51, in Berry, D.A., K.M. Chaloner, and J.K.Geweke, (1996).

Min, C. (1992), "Economic Analysis and Forecasting of International Growth Rates Using Bayesian Techniques," Ph.D. thesis, Dept. of Economics, U. of Chicago.

Schick (eds.), Maximum Entropy and Bayesian Methods, Kluwer Academic Publishers, 17-31.

Zellner, A. (1994), "Time Series Analysis, Forecasting and Econometric Modelling: The Structural Econometric Modelling, Time Series Analysis (SEMTSA) Approach," J. of Forecasting, 13, 215-233.

Zellner, A. (1994a), "Bayesian Method of Moments (BMOM) Analyis of Mean and Regression Models," presented at World Meeting of the International Society for Bayesian Analysis, Alicante, Spain, 1994, and in Zellner (1997). 291-304.

Zellner, A. (1997), Bayesian Analysis in Econometrics and Statistics: The Zellner View and Papers, Cheltenham, UK: Edward Elgar Publ. Ltd

Zellner, A. (1997a), "The Bayesian Method of Moments (BMOM): Theory and Applications," in Fomby, T.B. and R.C. Hill (1997), 85-105.

Zellner, A. (1998), "Keep It Sophisticatedly Simple," ms., to be published in H. Kuezenkampf, M. McAleer, and A. Zellner (eds.), Simplicity, Inference and Econometric Modelling, Cambridge U. Press, a collection of papers presented at the Tilburg Conference on Simplicity.

Zellner, A., J.L. Tobias and H.K. Ryu (1998), " Bayesian Method of Moments Analysis of Time Series Models with an Application to Forecasting Turning Points in Output Growth Rates," presented at World Meeting of the International Society for Bayesian Analysis, Istanbul, 1997, and to be published in Estadistica, 1999.

Zellner, A. and C.Min (1998), "Forecasting Turning Points in Countries' Output Growth Rates: A Response to Milton Friedman," ms., to be published in J. of Econometrics.

Zellner, A. and F.C. Palm (1974), "Time Series Analysis and Simultaneous Equation Econometric Models," J. of Econometrics, 2, 17-54.

Zellner, A. and F.C. Palm (1976), Time Series Analysis of Structural Monetary Models of the U.S. Economy," Sankhya 37, Series C, 12-56.