Working Paper Series

WORKING PAPER NO. 867

DISCUSSION OF PAPERS PRESENTED AT 1999 ASSA MEETING IN NEW YORK BY (1) FOSTER AND WHITEMAN, (2) GOLAN, MORETTI AND PERLOFF, AND (3) LAFRANCE

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. 867

DISCUSSION OF PAPERS PRESENTED AT 1999 ASSA MEETING IN NEW YORK BY (1) FOSTER AND WHITEMAN, (2) GOLAN, MORETTI AND PERLOFF, AND (3) LAFRANCE

by

Arnold Zellner

California Agricultural Experiment Station
Giannini Foundation of Agricultural Economics
January 1999

Discussion of Papers
Presented at 1999 ASSA Meeting in New York by
(1) Foster and Whiteman, (2) Golan, Moretti and Perloff and (3) LaFrance

by
Arnold Zellner
University of Chicago
arnold.zellner@gsb.uchicago.edu
http://gsbwww.uchicago.edu/fac/arnold.zellner/index.html

The three papers presented at this session involve interesting and useful applications of the maxent and Bayesian approaches to the analysis of important problems. FW use Bayesian and maxent techniques to analyze an option pricing problem, GMP employ maxent in the analysis of a job choice problem involving the use of a sample selection model and L analyzes an important food nutrition problem using a Bayesian method of moments maxent procedure. The papers illustrate the fact that Bayesian and maxent procedures are useful in these and other applications. Further, GMP and L have compared alternative maxent and other approaches in their papers in a scientific effort to appraise their relative merits. The authors are to be congratulated for providing these concrete applications of Bayesian and maxent procedures that will hopefully reduce the number of vacuous philosophical debates about the appropriateness of alternative approaches. (See Maasoumi (1990), Soofi (1996) and Zellner (1991) for accounts of the roles of information theory and Bayesian analysis in economics, econometrics and statistics and Davis (1941) for early, significant contributions to the uses of entropy in economics and econometrics.)

In the FW paper, "An Application of Bayesian Options Pricing to the Soybean Market," a predictive density for unknown future spot and futures prices is derived from a vector autoregressive process using a relatively uniformative prior density for its many (81) parameters and adjusted by use of maxent procedures to yield a risk neutral distribution that is used to price options. By use of Bayesian methods, the authors are successful in dealing with estimation risk. They conclude:

> "We use simple examples to show that relative to the standard Black (1976) model, as well as a non-parametric procedure advocated by Stutzer (1996) a procedure that makes use of numerical Bayes techniques to develop an underlying predictive density holds significant promise. That these techniques work well for complicated time series models...and without informative prior information is particularly encouraging and suggests that additional efforts to tune the model and to employ non-sample information will be fruitful."

They remark that in subsequent work larger samples and richer specifications of their time series models will be employed "to document performance over a longer sample, across expiration months, across puts and calls, and across strike prices."

My comments on this useful paper will hopefully be helpful to the authors in their future work. First, as shown in Tables 3 and 4, adding the constraints incorporating current market information from the options market (at-the-money option prices) improves predictive performance substantially. However, there is a striking difference in predictive performance between Table 3 and Table 4 that deserves some comment. In Table 3, relating to prediction of January prices, the medians of the percentage errors for three methods range from 12 to 35 % in absolute value while in Table 4, relating to the prediction of March prices, the medians are equal to 1% in absolute value for the three methods. In addition, for all the methods considered, there are a number of predictions that are far from accurate. Are there explanations for these large errors? Finally, if one were to use these predictions for speculation, what would be the resulting rates of return?

Second, it may be that multivariate transfer functions with relevant lagged input variables, time-varying parameters and shrinkage techniques such as used by Putnam and Quintana (1993) and in our past work on forecasting...see papers in Zellner (1997), will probably work better than VAR (very awful regression) models containing many parameters and implying complicated processes for individual variables. In addition, a uniform prior on the VAR parameters can be very informative about the properties of the roots of the process and other functions of the parameters.

Third, the authors successfully deal with estimation risk by use of Bayesian methods as has been done in earlier Bayesian portfolio analysis; see, e.g., Bawa, Brown and Klein (1979) and Putnam and Quintana (1993). However, it may also be worthwhile to consider model uncertainty and dispersion of prior beliefs as well as estimation risk. That is, it is possible to compute posterior probabilities for alternative models and/or priors and use them to combine alternative views and models as has been done in the forecasting area for many years. Whether such combination of expert opinion and allowance for model uncertainty by "averaging over models and experts" will provide better predictions is the key issue. Many times in forecasting, as demonstrated by the Blue Chip Company's forecast performance, averaging experts' forecasts provides better predictive performance. Averaging over expert opinions and over models seems possible and may lead to better predictions of option prices. Also, model uncertainty and dispersion of prior beliefs may be important in the definition of risk neutral distributions.

The GMP paper, "An Informational Based Sample-Selection Estimation Model of Agricultural Workers' Choice Between Piece-Rate and Hourly Work," presents a "new generalized maximum entropy (GME) approach to estimation of sample-selection models with small data sets, such as are found in many empirical economic analyses." This is indeed an important topic since most non-Bayesian methods for analyzing this class of models have just a large sample justification. The authors state that, "For small samples, the GME approach produces more stable estimates and has smaller mean square error measures than other well-known estimators such as ordinary least squares, Heckman's two-step method, full-information maximum likelihood, and Ahn and Powell's method." These conclusions are based on an analysis of a sample of data relating to choices made by male and female agricultural workers with respect to working in piece-rate or time-rate jobs.

As regards the model specification, if the time-rate and piece-rate sectors are distinct, one might expect to have different variables affecting wage rates in the two sectors whereas the authors' model specifies the same X matrix of input variables for the two sectors. Further material describing the two sectors would be helpful.

The GME framework is clearly explained, especially the reparameterization of the problem from the betas to the p's, a "set of proper probabilities, defined over the supports." The same sort of transformation or reparameterization is done with respect to the epsilons, the realized errors in terms of a set of proper probabilities, the q's. Since the p's and q's are new hyperparameters, it is particularly important to know how many are actually used and how the number and spacing of intervals were determined. For example, in constructing histograms, one has to decide on the number of intervals to employ given the range and in this regard Sturges' Rule and work by Hartigan have provided useful procedures for determination of cell size for histograms. It is possible that this work can be extended to the multidimensional "histograms" that the GME approach utilizes. Further, if sample sizes are small and many p's and q's are introduced, it is possible that a problem of "over-fitting" may be encountered. That is, the fit may be very good but out of sample prediction may not be very good. See, e.g., Adkins (1997), who performed Monte Carlo experiments to appraise the performance of GME techniques relative to that of maximum likelihood logit and probit techniques in estimation (within sample) and prediction (out-of-sample) using a binary choice model. In his conclusions he remarks,

> "Unfortunately, the in-sample and out-of-sample predictive performance of the GME estimator as specified in this paper leaves something to be desired. As the signal- to-noise ratio gets small..., the GME is a good choice. Unfortunately, as this value increases, its performance relative to other estimators diminishes. A clear cut recommendation is not possible, although the performance of the GME is better the larger the number of betas." (p.195)

In view of Adkin's results, it would appear useful for the authors to study out-of-sample properties of their procedures. In this comparison, they can also include traditional Bayesian (TB), procedures, Bayesian method of moments (BMOM) and other procedures and use their associated predictive densities to compute Bayes's factors, as has been done in Tobias and Zellner (1997). As is well known, Bayes'factors incorporate a penalty for additional parameters that could be important in comparing GME with other procedures that use fewer parameters.

Within-sample results are interesting but without thorough consideration of the extra information provided by truncating the parameter and realized error terms' distributions and the effects of introducing many, many parameters, it is difficult to understand statements such as, "...the GME coefficients tend to have much smaller asymptotic standard errors than the other estimates." The within-sample measures of prediction are similarly hard to interpret since with enough parameters we can get a perfect within-sample fit. However, such a fitted model will usually do poorly in out-of-sample predictive performance, as many have noted. That the Heckman maximum likelihood procedure in some cases "...either fails to converge or its estimated correlation coefficient lies outside the [-1,1] in finite samples" is an important finding. Here, we have a procedure that is asymptotically justified, and as is well known, such properties say little about finite sample properties in many cases; see, e.g. Phillips (1983) and Zellner (1998). It would be valuable to use MCMC methods along with data augmentation procedures to implement TB and BMOM methods for further comparison with the results so far obtained. In implementing these approaches, diffuse and informative prior densities can be employed that match to some extent the extra information provided by truncation of the parameter and realized error spaces in the GME approach. Further, as remarked above, Bayes' factors can be computed using predictive densities provided by alternative approaches.

As regards diagnostic checks of the adequacy of the authors' model, very few have been performed. One can use the realized error term analysis described in Chaloner and Brant (1988), and Albert and Chib (1993) to do outlier analysis, etc. Obviously, if the model is misspecified, many conclusions may be affected. In the one check on the appropriateness of including certain variables in the X and C matrices, a chi-squared test is performed at the five percent level with the conclusion that the nine variables' coefficients are equal to zero with no mention of the power of the test. Much more diagnostic checking of the model would be desirable, say using cross entropy measures which, as is well known, are directly related to expected log posterior odds, see, e.g. Good (1950), Kullback (1959) and Tobias and Zellner (1997).

In summary, this is a path-breaking paper showing how GME methods can be employed to analyze an important class of models. The remarks above are intended to help the authors improve and extend certain aspects of their very valuable and interesting research.

The third paper, "Inferring the Nutrient Content of Food with Prior Information," is by Jeffrey T. LaFrance. As he states, "Using unpublished documents from the HNIS, estimates of the percentages of seventeen nutrients supplied by twenty-one foods were compiled for the period 1952-1983. The Bayesian Method of Moments is applied to this data set to obtain a proper prior for the purpose of drawing year-to-year inferences about the data base."

He starts with the assumption that, "...we have a stable, theoretically consistent reduced form empirical model of the demand for foods..." While this is a good starting point for the analyses that the author carries out, the question arises as to whether it is possible to derive a theoretical demand system in which consumers are assumed to take account of nutritional content of foods in making their budget-constrained, utilitymaximizing choices. If no such system exists, perhaps we should get some economic theorists busy to produce one, e.g. by maximizing U(x) subject to a budget constraint and $Ax \geq c$, where c is a vector of minimal requirements of each nutrient and A is the author's "nutrient content matrix."

The author introduces the "nutrient content matrix," denoted by A. Probably as a first approximation, he assumes the elements of A to be constant over the decades, something like a no technical change assumption. The data plots in his Fig. 1 indicate some trend and other variation in certain of the cal/lb. measures for different foods. Perhaps a generalization to permit the elements of A to have possible trends would be interesting, even though, as the author points out, the trends are not very pronounced.

As regards methodological approaches, he provides a description of how the GME approach can be employed to solve his problem, namely, how to make inferences, e.g. obtain estimates of the elements of the matrix A that reflect prior information obtained independently of the current inference problem. After demonstrating how a GME solution can be obtained, he raises fundamental issues about the prior inputs to the GME procedure, namely the ranges of parameters and, the number, N, of subintervals to employ, noting that each choice, "...generates a different solution for the probability weights and therefore for the elements...", that is the elements of A. His suggested solution for this problem "... is to let N go to infinity and use a continuous density function for both the prior and the posterior." He then provides continuous GME solutions and raises the question as to what are appropriate choices for a pre-data prior distribution, a post-data posterior distribution, which becomes the pre-forecast distribution? He indicates that minimization of the K-L cross entropy relative to a uniform prior provides a solution equivalent to the GME solution. Given this result, he decides to pursue a BMOM approach, see, e.g., Zellner (1997a) to yield a post data density for the rows of A. He shows how such a post data density can be employed in the present problem to provide a solution to his problem and comments, "We end up with a very simple least squares rule as the solution to what started out as a difficult and highly ill posed inference problem. I find this quite delightful!"

In his closing paragraphs, LaFrance points to other properties of his solution and compares it to those provided by classical and traditional Bayesian approaches. Having these properties stated is indeed very valuable and the fact that the BMOM solution is rather simple is indeed important. However, we are still faced with the problem of how many moments to use in the BMOM approach? How does it compare with solutions provided by TB and GME approaches? As is obvious, deductive logic will not answer these questions entirely. What are needed, as stated above, are applications of good model selection techniques, e.g. Bayesian posterior odds that can be employed to choose among or combine alternative models.

The authors of these three stimulating papers are to be congratulated for their contributions which have helped considerably to enhance understanding of new, alternative approaches and to provide impressive empirical results.

References

Adkins, L. "A Monte Carlo Study of a Generalized Maximum Entropy Estimator of the Binary Choice Model." Fomby, T.B. and R.C. Hill, eds., *Advances in Econometrics, Vol.*

12, Applying Maximum Entropy to Econometric Problems, Greenwich, Ct & London, UK,: Jai Press(1997): 183-197.

Albert J.H. and S. Chib. "Bayesian Analysis of Binary and Polychotomous Response Data." *J. Am. Stat.Assoc.* 88 (1993): 669-679.

Bawa, V.S., S.J. Brown and R.W. Klein, eds., *Estimation Risk and Optimal Portfolio Choice*, Amsterdam: North-Holland Publishing Co., 1979.

Chaloner, K.M.and R. Brant. "A Bayesian Approach to Outlier Detection and Residual Analysis." *Biometrika*. 75 (1988): 651-659.

Davis, H.T. *The Theory of Econometrics.* Bloomington,IN: Principia Press, 1941.

Good, I.J. *Probability and the Weighting of the Evidence.* New York: Hafner Publ. Co., 1950.

Kullback, S. *Information Theory and Statistics.* New York: J. Wiley & Sons, Inc., 1959.

Maasoumi, E. "Information Theory." J. Eatwell, M. Millgate and P. Newman, eds.,*The New Palgrave Econometrics*. New York: W.W. Norton & Co.1990: 101-112.

Phillips, P.C.B. "Exact Small Sample Theory in the Simultaneous Equations Model." Z. Griliches and M.D. Intriligator, eds., *Handbook of Econometrics*. Vol. 1. Amsterdam: North-Holland Pub. Co., 1983: Ch 8, 449-516.

Putnam, B. H, and J.M. Quintana."Bayesian Global Portfolio Management," *American Statistical Association's Proc.of the Sect. on Bayesian Statistical Science*. 1993: 110-118.

Soofi, E.S. "Information Theory and Bayesian Statistics." Berry, D.A., K.M. Chaloner and J.K.Geweke,eds. Bayesian Analysis in Statistics and Econometrics. New York: John Wiley & Sons, Inc., 1996: 179-190.

Tobias, J.L. and A. Zellner. "Further Results on Bayesian Method of Moments Analysis of the Mutiple Regression Model." H.G.B. Alexander Research Foundation, U.of Chicago.1997. Paper presented to Econometric Society Meeting at Caltech and at the ISBA World Meeting, Instanbul, Turkey.

Zellner, A. "Bayesian Methods and Entropy in Economics and Econometrics." W.T. Grandy, Jr. and J.H. Schick, eds., *Maximum Entropy and Bayesian Methods*. Kluwer Academic Pub., 1991: 17-31.

__________. Bayesian *Analysis in Econometrics and Statistics: The Zellner View and Papers*. Cheltenham, UK and Lyme, US: Edward Elgar Publishing Co., Ltd. 1997.

__________. "The Bayesian Method of Moments (BMOM): Theory and Applications." Fomby, T.B and R.C. Hill, eds., *Advances in Econometrics. Vol. 12, Applying Maximum Entropy to Econometric Problems.* Greenwich, CT & London, UK: Jai Press, 1997a: 85-105.

__________. "The Finite Sample Properties of Simultaneous Equations' Estimates and Estimators: Bayesian and Non-Bayesian Approaches." L.R. Klein,ed. Studies in Econometrics in Honor of Carl F. Christ, J. of Econometrics Annals Issue (March-April, 1998): 185-212 and in Zellner (1997).