

378.794  
G43455  
WP=867

*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

Walter Library  
Dept. of Applied Economics  
University of Minnesota  
1994 Buford Ave - 232 ClaOff  
St. Paul MN 55108-6040 USA

DEPARTMENT OF AGRICULTURE AND  
RESOURCE ECONOMICS AND POLICY

BERKELEY

CALIFORNIA AGRICULTURAL EXPERIMENT STATION

*University of California*

378,794  
G43455  
WP-867

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](mailto: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 uninformative 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