Econometric and Statistical Data Mining,
Prediction and Policy-Making
By

Arnold Zellner, University of Chicago*

Abstract

How to formulate models that work well in explanation, prediction and policy-making is a central problem in all fields of science. In this presentation, I shall explain the strategy, our Structural Econometric Modeling, Times Series Analysis (SEMTSA) approach that my colleagues and I have employed in our efforts to produce a macroeconomic model that works well in point and turning point forecasting, explanation and policy-making. Data relating to 18 industrialized countries over the years, taken from the IMF-IFS data base have been employed in estimation and forecasting tests of our models using fixed and time varying parameter models, Bayesian posterior odds, model combining or averaging, shrinkage, and Bayesian method of moments procedures. Building on this past work, in recent research economic theory and data for 11 sectors of the U.S. economy have been employed to produce models for each sector. The use of sector data and models to forecast individual sectors' output growth rates and from them growth rates of total U.S. output will be compared to use of aggregate data and models to forecast growth rates of total U.S. output. As will be seen, IT PAYS TO DISAGGREGATE in this instance. Last, a description of some steps underway to improve and complete our Marshallian Macroeconomic Model of an economy will be described.

I. Introduction

When Ham Bozdogan invited me to present a lecture at the U. of Tennessee, C Warren Neel Conference on Statistical Data Mining and Knowledge Discovery, he mentioned that he had attended my presentation of our modeling and forecasting work at the June 2000 International Society for Bayesian Analysis (ISBA, website: www. Bayesian.org) and Eurostat meeting in Crete and that he thought it was a good example of data mining. This was news to me since I did not have a good definition of data mining and was unsure about how it represented what we were doing in our work with a large data set relating to 18 industrialized countries over a period of about 50 years. After attending this U. of Tennessee Conference, I believe that I have a better understanding of data mining and believe that it is useful to relate it to general views of scientific methodology and some past work in statistics and economics, which I shall do in Section II. I provide a brief overview of scientific methodology and its objectives and indicate where, in my opinion, data mining fits into the overall process and what some of its unique features are.

---

* H.G.B. Alexander Distinguished Service Professor Emeritus of Economics and Statistics. Research financed by Alexander Endowment Fund, CDC Management Fund and the National Science Foundation.

Then in Section III, I shall describe what Ham called our past data mining efforts with special emphasis on methodological tools that may be useful in other data mining efforts. These include the Structural Econometric Modeling, Time Series Analysis (SEMTSA) approach put forward by F.C. Palm and myself, Zellner and Palm (1974,1975) and pursued further by many over the years; see papers by leading workers in Zellner and Palm (2003). In this approach, the value of keeping it sophisticatedly simple (KISS) is emphasized for "obvious" reasons…see e.g. Zellner, Kuezenkamp and McAleer (2001) for further consideration of simplicity and complexity issues. In addition to the overall SEMTSA approach, there is a brief discussion of new results in information processing, including a new derivation of the Bayesian learning model, Bayes' Theorem and generalizations of it; see, Soofi (2000) and Zellner (1997, 2002). One special learning approach is the Bayesian method of moments (BMOM) that permits researchers to derive post data moments and post data densities for parameters and future observations without use of sampling assumptions for the given data, likelihood functions and prior densities. Since, as I learned at this Conference, there is often great uncertainty regarding the forms of likelihood functions and prior densities in data mining, an approach, such as the BMOM approach, that permits individuals to perform inverse inference without specifying the form of the likelihood and using a prior density, will probably be useful. Examples will be provided to illustrate general principles. For more on the BMOM and applications of it see, e.g. Green and Strawderman (1996), Mittelhammer et al. (2000), van der Merwe et al. (1998,2001), Zellner (1994,1997a, b, 2002) and Zellner and Tobias (2001).

In Section IV, some details regarding the models and methods employed in analyzing, modeling and forecasting data for 18 industrialized countries will be described. Also, considerations regarding aggregation/disaggregation issues will be briefly discussed along with an example involving alternative methods for forecasting the median growth rate of 18 countries' real output growth rates, that illustrates that in this instance, it pays to disaggregate; see Zellner and Tobias (2000). Further, disaggregation by industrial sectors using Marshallian models for each sector will be described and some results, taken from Zellner and Chen (2001), of the use of sector models in forecasting aggregate outcomes will be provided and compared to those derived from models implemented with aggregate data. Last, in Section V considerations regarding the form of a complete Marshallian Macroeconomic Model will be presented along with some thoughts regarding future research will be presented.

II. Brief Comments on Scientific Method and Data Mining

As discussed in Jeffreys (1937,1998) and Zellner (1984,1996,1997), it is the case that scientific work in all areas involves induction, deduction and reduction. Induction involves (a) measurement and description and (b) use of given models or subject matter generalizations to explain features of past data, predict as yet unobserved data and to help solve private and public decision problems. As regards reduction, in Zellner (1996, p. 5) it is pointed out that the famous physicist C.S. Pierce commented that "reduction suggests that something may be; that is, it involves studying facts and devising theories to explain them. For Pierce and others the link of reduction with the unusual fact is emphasized." And of course, deduction, that is logical or mathematical proof or disproof plays a role in induction and reduction but is inadequate alone to serve the needs of scientists. As emphasized by Jeffreys (1937,1998), scientists need to be able to make statements less extreme than "proof," "disproof" and "I don't know." They generally make statements such as "A probably causes B" or "Theory I is probably better than Theory II" or "The parameter's value probably lies between 0.40 and 0.50." Fortunately, Bayesian methods are available to quantify such statements, to update such probabilities as more data become available and to provide predictions that can be checked with additional data; for a survey of Bayesian analysis with many references, some to downloadable, free software for performing Bayesian calculations, see Berger (2000). That Bayesian methods permit updating of probabilities representing degrees of confidence in models is a very important capability. In this connection, note that causality has been defined as predictability according to a law or set of laws, that is a dependable, subject matter model or set of models that have performed well in past work and thus have high probabilities associated with them. See, e.g., Feigl (1953), Aigner and Zellner (1988), and Zellner (1984), for further discussion of the concept of causality, empirical tests of causality and its relation to applied data analysis and modeling.

There is no doubt but that data miners are involved in inductive, deductive and reductive scientific work involving measurement, description, modeling, explanation, prediction and decision-making. What appears to be rather unique about the current data mining area are the large samples of data utilized and the powerful computers and algorithms that are available for analyzing them. However, as has been recognized by many, there are the old issues of data quality, appropriate measurement procedures and good statistical descriptive measures that are obviously very important. Further, given a tentative model for a very large data set, say a multivariate regression or time series model, the use of appropriate estimation, testing, model evaluation and other inference procedures is critical. For example, use of inappropriate testing and model evaluation methods, e.g. the 5% level of significance in Neyman-Pearson test procedures, can lead to costly errors when sample sizes are very large; see, e.g. Zellner (1996, p 302, ff.) and references cited therein for further discussion of this important problem. Further, the accept-reject framework often used is many times not as good as the use of techniques that permit a set of models to be identified and to use them in model averaging to obtain estimates and predictions. Further, in making decisions or producing forecasts, it has been emphasized that model uncertainty as well as other types of uncertainty should taken into account, as for example in formal model combining and averaging procedures (See Zellner (1997, Part V), for procedures for model-combining and applications to forecasting problems that appear readily applicable to modeling and forecasting in some data mining problems.)

As regards reduction, that is studying data and devising models to explain them, there is the issue of simplicity versus complexity. Some advise researchers to start with a sophisticatedly simple model and complicate it if necessary. Note that a sophisticatedly simple modeling approach involves taking account of what is already known about a problem in order to avoid making stupid errors. In industry, there is the saying, "Keep it simple stupid" (KISS). Since some simple models are stupid, I changed the interpretation of KISS to "Keep it sophisticatedly simple," which is in line with Einstein's advice to make models as simple as possible, but no simpler. On the other hand there are those who advocate starting with a large, complicated, "encompassing" model and testing downward to discover a useful model. Of course, it is impossible to prove deductively whether it is better to KISS or to use an "encompassing" approach. However, it is relevant to note that there are many, many complicated encompassing models and if one chooses the wrong one, all that follows will be unsatisfactory. While there are many successful sophisticatedly simple models in the sciences, e.g. s= $1/2gt^2$, F =ma, PV=RT, E= mc^2, etc., it is difficult to find one successful large, complicated model in any area of science. For additional consideration of these simplicity/complexity issues and some measures of the complexity of models, see Zellner, Kuezenkamp and McAleer (2001).

Since, as mentioned above, recognition of unusual facts is many times very important in leading to new insights and models, it occurred to me that there must be ways of producing unusual or surprising facts. After some thought, I developed a list of eight such procedures, see Zellner (1984, pp. 9-10), many of them quite obvious, namely, study unusual groups, e.g. the very poor or the very rich, unusual historical periods, e.g. hyperinflations, or great depressions, produce extreme data that are inconsistent with current models' predictions, etc. Also, with large data sets, there is a tendency to throw away "outlying" observations rather than try to explain why they are "outlying." As is well known, "outlying " points may be evidence that assumed functional forms for relations, e.g., linear forms, and/or assumed distributions for error terms are incorrect. Thus "outlying" or "unusual" data points deserve much attention and explanation, if possible. For brevity, I shall not review analyses that have been adversely affected by deleting rather than explaining, outlying data points.

When the available data, say time series data, include many, many variables, as in data relating to the world's numerous economies, and include many unusual features, e.g. similar, but not exactly the same upward and downward movements in growth rates of output and other variables, etc., there is the problem of how to formulate explanatory models that explain past variation and are helpful in predicting new data and in policy-making. Many researchers and governmental units worldwide have approached this problem by building complicated, multi-equation, linear and non-linear, stochastic models, so-called structural econometric models. Unfortunately, not many of these models, if any, have performed satisfactorily in explanation and prediction. Others, have resorted to large statistical models, usually vector autoregressive (VAR) models in attempts to get good forecasting models. These attempts, too, have not been successful in the sense that their point and turning point forecasts have not been very good. Faced with these problems and approaches, years ago, Franz Palm and I put forward a combined structural econometric modeling, time series analysis (SEMTSA) approach that attempted to integrate the best features of earlier approaches. See Zellner and Palm (1974, 1975, 2003) and Zellner (1984, 1997) for descriptions of the approach and applications of it to a variety of problems. The approach involves starting simply with models for individual variables, determining forms for them that reflect past subject matter information, fit the data reasonable well and forecast satisfactorily. Then these tested components are combined to form a multivariate model and its performance in explaining variation in past data and in forecasting future data is studied and continually improved. Throughout the process, subject matter theory is used in attempts to "rationalize" or "explain" why the empirical models work in practice and in an effort to produce a "structural" or "causal" model rather than just an empirical, statistical forecasting model that does not explain outcomes or possible causal relations very well. See below for further discussion of the SEMTSA approach's properties and some experiences in applying it in our efforts to produce a structural macroeconomic model that works well in explaining past data, predicting new data and in policy-making.

III. The Structural Econometric Modeling, Time Series Analysis (SEMTSA) Approach

In the mid-twentieth century, impressive theoretical and empirical developments occurred in the statistical time series analysis area in the work of Quenouille (1957), Box and Jenkins (1970), and others. The univariate and multivariate autoregressive, moving average (ARMA) models employed in this statistical time series analysis approach, seemed very different from the causal, dynamic, structural econometric models formulated and estimated by the Nobel Prize winners, Tinbergen, Klein, and Modigliani and many others and used by many governmental units world-wide. Since many were confused about the relationship of these two classes of models, in Zellner and Palm (1974), we clarified the relationship and pointed to some unusual features of multivariate statistical time series and structural econometric models. This led to a strategy of model-building that we called the SEMTSA approach and first applied in a paper, Zellner and Palm (1975) to evaluate a monetary model of the U.S. economy and utilized in other works included in Zellner and Palm (2003).

3.1 The SEMTSA Approach

The initial problems considered in the SEMTSA approach were (1) the implications of the basic multiple time series or multivariate ARMA model put forward by Quenouille (1957) and others for properties of processes for individual variables and (2) its relationship to dynamic structural econometric models. In this connection, Quenouille's and others' multiple time series model, that includes a vector autoregression as a special case, for an mx1 vector of variables at time t, z(t) is:

(1) $H(L)z(t) = F(L)e(t)$ t = 1,2,…, T

where $H(L) = I + H_1L + ... + H_pL^p$ and

$F(L) = F_o + F_1 L + ... + F_q L^q$ are matrix lag operators with L the lag operator such that, $L^r z(t) = z(t-r)$, the H's and F's given matrices and e(t) is a zero mean white noise error vector with covariance matrix I.

Is the model in (1) a good one for modeling and forecasting? As Palm and I suggested, a relevant issue is what are the implied processes for individual variables in the vector z(t). If H(L) is invertible, it is straightforward to derive the marginal processes for individual elements of z(t). In general these turn out to be very high order autoregressive-moving average processes, not at all like the models for individual variables identified by many time series workers. Thus the general model in (1) has implications that are not in agreement with empirical findings indicating a need for modifications of it, say imposing restrictions, or introducing non -linearities, etc.

As mentioned above, the relation of (1) to dynamic linear structural econometric models (SEMs), discussed in most past and current econometrics textbooks, was unclear to many. Here we pointed out that if the vector of variables, z(t) is partitioned into a sub vector of "endogenous" variables, y(t) and a sub vector of "exogenous" variables, x(t), that is $z(t)' = [y(t)', x(t)']$, the relation would be clear. Note that endogenous variables are variables whose variation is to be explained by the model whereas exogenous variables, e.g. weather variables, etc. are input variables to the model. The assumption that the sub vector x(t) is a vector of exogenous variables leads to a sub matrix of H(L) being identically zero and F(L) being block diagonal, very important restrictions on the general model in (1). When these restrictions are imposed, the system becomes:

(2a) $H_{11}(L)y(t) + H_{12}(L)x(t) = F_{11}(L)e_1(t)$

(2b) $H_{22}(L)x(t) = F_{22}(L)e_2(t)$

where $H_{11}(L), H_{12}(L)$ and $H_{22}(L)$ are sub matrices of $H(L)$ and the assumption that x(t) is exogenous implies that the sub-matrix $H_{21}(L) \equiv 0$. Also, $F_{11}(L)$ and $F_{22}(L)$ are sub matrices of $F(L)$ with the assumption that x(t) is exogenous implying that the off diagonal matrices $F_{12}(L) \equiv 0$ and $F_{21}(L) \equiv 0$.

Equation system (2a) is in the form of a linear dynamic SEM while (2b) is the system assumed to generate the exogenous or input variables. Thus there is compatibility between the multiple time series model in (1) and the linear dynamic SEM. Further, as emphasized in our SEMTSA approach, it is important to check the implications of (2a) for the forms of the processes for individual variables in y(t). Given that the matrix lag operator $H_{11}(L)$ is invertible, that is $H_{11}^{-1} = H_{11}^{a} / |H_{11}|$ where the dependence on L has been suppressed, H_{11}^{a} is the adjoint matrix and $|H_{11}|$ is the determinant of H_{11}, a polynomial in L, it is direct to solve for the transfer function system associated with (2a), namely:

(3) $|H_{11}|y(t) = -H_{11}{}^{a}H_{12}x(t) + H_{11}{}^{a}F_{11}e_1(t)$

Note that since the same lag operator $|H_{11}|$ hits all the elements of y(t) and if there is no canceling of common roots, this implies that the autoregressive part of the transfer functions for the elements of y(t) should be IDENTICAL, a restriction that can be and has been tested in applied work. Further, restrictions on the coefficients in (2a) imply testable restrictions on the parameters of the transfer functions in (3) that can be and have been tested. Last, empirical procedures for obtaining or identifying forms of transfer functions from information in data are available in the literature and have been applied to determine whether the forms of the transfer function equations derived analytically in (3) are the same as those determined or identified empirically. See Zellner and Palm (1975, 2003) for examples in which these procedures are applied to a variety of problems and revealed that the information in the data was in conflict with the implications of models in many cases.

In our initial empirical work to formulate good models to explain and predict macroeconomic variables, in Garcia-Ferrer et al. (1987), we decided to start with analysis of models for a key macroeconomic variable, the rate of growth of real output, as measured by real GDP, of an economy. After determining that simple time series models, e.g. autoregressive models did not work well in forecasting and with the form of (3) in mind, we formulated the following model for the output growth rate of economy i in year t, denoted by the scalar $y_i(t)$,

(4a) $$\begin{aligned} y_i(t) &= b_{oi} + b_{1i}y_i(t-1) + b_{2i}y_i(t-2) + b_{3i}y_i(t-3) + \\ &+ b_{4i}M_i(t-1) + b_{5i}SR_i(t-1) + b_{6i}SR(t-2) + b_{7i}WSR(t-1) + e_i(t) \end{aligned}$$

or

(4b) $y_i(t) = x_i(t)'b_i + e_i(t)$

t= 1,2,…,T and i = 1,2,…, N

where M = growth rate of real money, SR = growth rate of real stock prices and WSR = the median of the countries' growth rates of real stock prices. The form of (4) was rationalized as follows: (a) The AR(3) assumption allowed for the possibility of having complex roots leading to oscillatory movements (see Hong (1989) for supporting empirical evidence). (b) Burns and Mitchell (1946) had established that money and stock prices lead in the business cycle using pre World War II data going back to the 19th century for France, Germany, U.K., and U.S. (c) The variable WSR was included to model common shocks hitting all economies. The model in (4) was named an autoregressive, leading indicator (ARLI) model and implemented with data for 9 and then 18 industrialized countries in point and turning point forecasting experiments with encouraging results.

Later (4) was modified to include a current measure of world output, w(t), the median of 18 countries' growth rates in year t, and an ARLI model for w(t) was introduced, namely,

(5a) $y_i(t) = g_i w(t) + x_i(t)'b_i + e_i(t)$

and

(5b) $$w(t) = a_o + a_1 w(t-1) + a_2 w(t-2) + a_3 w(t-3) + a_4 MM(t-1) + a_5 MSR(t-1) + a_6 MSR(t-2) + u(t)$$

where MM and MSR are the median growth rates of real money and real stock prices, respectively, and u is an error term. We called this two equation model the ARLI/WI model and showed in Zellner and Hong (1989) that it performed somewhat better in forecasting experiments with data for 18 countries than the one equation ARLI model, shown above in (4), using various Stein-like shrinkage estimation and forecasting techniques. Also, the RMSE's of forecast compared favorably with those associated with large, structural OECD forecasting models. In addition, variants of the ARLI and ARLI/WI models performed well in forecasting turning points in countries' growth rates with 70% or more of 211 turning points correctly forecasted for 18 countries with the "downturn" and "no downturn" forecasts being somewhat better than the "upturn" and "no upturn" forecasts; see Zellner and Min (1999) for these and additional results with references to earlier analyses. And very important, the ARLI/WI model was shown to be compatible with various macroeconomic theoretical models, e.g., a Hicksian IS/LM model in Hong (1989), a generalized real business cycle models in Min (1992) and an aggregate supply and demand model in Zellner (1999) . This is a fundamental aspect of the SEMTSA approach, namely a fruitful interaction between data analysis and subject matter theoretical models. Some additional variants of these models will be discussed in Sect. IV.

3.2 Statistical Inference Procedures

With respect to inference procedures for the ARLI and ARLI/WI models, mentioned above, various procedures were employed including least squares (LS), maximum likelihood (ML), traditional Bayesian(TB) and the recently developed Bayesian method of moments (BMOM). For these time series models, while LS and ML procedures readily yielded point estimates, they did not provide, among other things, finite sample standard errors, confidence and predictive intervals, probabilities relating to various models' forms and probability statements regarding possible downturns or upturns in countries' growth rates. In contrast, TB procedures readily provided finite sample results including intervals for parameters and future observations, shrinkage estimates and forecasts, probabilities associated with alternative model forms, etc. given the traditional inputs to TB analyses, namely, a prior density, a likelihood function and Bayes' theorem. As is well known, Bayes' theorem yields the important general result that a posterior density for the parameters is proportional to the prior density for the parameters times the likelihood function, with the factor of proportionality being a normalizing constant that can be evaluated analytically or numerically.

Thus given the inputs, a prior density and a likelihood function, it is direct to obtain a posterior density for the parameters and also a predictive density for future observations. In an effort to relax the need for such inputs, the Bayesian method of moments (BMOM) was introduced in Zellner (1994); see also Zellner (1997) and the references cited in Section I for further results and applications. The BMOM approach permits investigators to obtain post data moments and densities for the parameters and future observations without using prior densities, likelihood functions and Bayes' theorem and may be useful to data miners in performing statistical inference.

To illustrate the BMOM approach, consider time to failure given observations that are assumed to satisfy, $y(i) = q + u(i)$, i= 1,2,…,n where q is a parameter of interest with an unknown value. On summing both sides of this relation and dividing by n, we have the given mean of the observations $\bar{y} = q + \bar{u}$ and on taking a subjective expectation of both sides, $\bar{y} = Eq + E\bar{u}$, where it is to be recognized that $\bar{y}$ has a given observed value, e.g. 3.2, whereas q and $\bar{u}$ have unobserved values that are considered subjectively random. Now if we assume that there are no outliers or left out variables and the form of the above relationship is appropriate, we can assume that the subjective mean of $\bar{u}$ is zero, that is, $E\bar{u} = 0$ that implies $Eq = \bar{y}$. Thus we have a post data mean for q without introducing a prior density and a likelihood function. Also, as is well known, the mean is an optimal estimate relative to a quadratic loss function. Further, the maximum entropy or most spread out proper density for q with given mean = $\bar{y}$, is easily derived and well known to be the exponential density, that is, $g(q|D) = (1/\bar{y})\exp(-q/\bar{y})$ with $0 < q < \infty$ and where D denotes the given observations and background information. This post data density can be employed to make inverse probability statements regarding possible values of q, e.g. $\Pr\{1.2 < q < 2.3|D\}$, the objective of Bayes' (1763) original paper and thus the name Bayesian method of moments. This is one example of the BMOM.

To apply the BMOM to equation (4) above, express the equation in standard regression form, $y = Xb + e$ where y is a vector of given observations, X a matrix of observations on the input variables, b a coefficient vector and e a vector of realized error terms. The elements of b and of e are treated as parameters with unknown values, as previously done in the TB literature in Chaloner and Brant (1988) and Zellner (1975). Assuming that X is of full column rank, we can write: $b \equiv (X'X)^{-1}X'y = Eb + (X'X)^{-1}X'Ee$, where E represents the subjective expectation operator. Assuming that the functional form of the relation is satisfactory, i.e. no left out variables, and no outliers, no errors in the variables, etc., we can make Assumption I: $X'Ee = 0$ which implies from the relation above that $Eb = b = (X'X)^{-1}X'y$, that is the post data mean for b is the least squares estimate. Further, we have $y = XEb + Ee$ and thus Assumption I implies that y is the sum of two orthogonal vectors, XEb and Ee. In addition, $Ee = y - XEb = \hat{e}$, the least squares residual vector. On making a further assumption regarding the post data covariance matrix of the realized error vector, see, e.g. Zellner (1997, p. 291ff) and Zellner and Tobias (2001), the second moment matrix of b is shown to be: $E(b - Eb)(b - Eb)' = (X'X)^{-1}s^2$ with $s^2 = \hat{e}'\hat{e}/v$ and $v = T - k$, the "degrees of freedom" parameter, where T is the number of observations and k is the number of elements of b. Note that all these post data moments have been derived without using a likelihood function and a prior density. However, more prior information can be introduced via use of a conceptual sample and in other ways, as shown in Zellner et al (1999), and other papers cited above.

With the moments of b as given in the previous paragraph, the least informative, maxent density for b is a multivariate normal density with mean b, the least squares estimate of b, and covariance matrix $(X'X)^{-1}s^2$. This density can be employed to make probability statements regarding possible values of the elements of b and functions of them quite readily. Also, for future, as yet unobserved values of y, assumed to satisfy, $y_f = X_f b + e_f$, given X_f, if we make assumptions regarding the moments of e_f, the moments of y_f can be easily derived. For example, if it is assumed that $Ee_f = 0$, then $Ey_f = X_f Eb = X_f b$, the least squares forecast vector. With a further assumption regarding the second moment of e_f the second moments of the elements of y_f are available and these moments can be used as side conditions in deriving a proper maxent predictive density for y_f. For some computed examples involving these and other assumptions with applications to forecasting turning points, see Zellner et al (1999).

In addition to providing BMOM post data and predictive moments and densities, information theory has been utilized in Zellner (1988, 1997, 2002) to produce learning models, including Bayes' Theorem, in a very direct approach. Namely information measures associated with inputs and outputs to an information processing problem are considered. In the TB approach there are two inputs, a prior and a likelihood function, and two outputs, a posterior density for the parameters, and a marginal density for the observations. On forming the criterion functional, information out minus information in and minimizing it with respect to the form of the output density for the parameters, subject to its being a proper density, the solution is to take the posterior density for the parameters equal the prior density times the likelihood function divided by the marginal density of the observations, that is the result is Bayes' theorem and when this is done, the information in = the information out and thus this information processing rule is 100% efficient. See comments on these results in Soofi (2000) and by Jaynes, Hill, Kullback and Bernardo in Zellner (1997).

As pointed out in Zellner (1997,2002), one can input just a likelihood function, and no prior, as R. A. Fisher wanted to do in his fiducial approach, and solve for an optimal form for the output density for the parameters that is proportional to the likelihood function. Or one can assign given weights to the input prior and sample information and solve for an optimal form of the output density for the parameters. Also, when the input information is just in the form of given moments for parameters, the optimal form for the output density is the BMOM solution. Dynamic problems in which the output of one period is the input to the next period, along with additional sample information, have been formulated and solved using dynamic programming techniques. Again when the traditional inputs are employed, a prior and a likelihood function, it is optimal to update densities using Bayes' theorem. Having various information processing or learning models "on the shelf" to be used when appropriate is important just as is the case with various static, dynamic and other models in engineering, physics, economics, business and other fields

Having presented some discussion of general methods, attention in the next section will be focused on these and other methods used in our empirical, data mining work.

IV. Methods Employed in Data Analysis, Modeling and Forecasting

In our work since the mid 1980's, we have used data from the International Monetary Fund's International Financial Statistics database at the U. of Chicago, a very large database with data on many variables for over 100 countries' economies. Since measures of output may contain systematic biases, etc., we thought it wise to log variables and take their first differences to help get rid of proportional measurement biases. Further, these growth rate measures are of great interest. However, we were not too sanguine about inducing covariance stationarity by first differencing given the many different kinds of shocks hitting economies, e.g. wars, strikes, financial crises, droughts, policy blunders, etc.

Very important in our work and in presentations was and is the graphical presentation of our data using boxplot techniques to display general properties of the economic fluctuations for the 18 countries in our sample and for 11 sectors of the U.S. economy in later work. As we observed, there are somewhat regular fluctuations present in the three variables with the money and stock price variables showing a tendency to lead, as recognized earlier by Burns and Mitchell (1946) in their monumental "data mining" study, Measuring Business Cycles in which they used data for France, England, Germany and the U.S. going back to the 19^{th} century to characterize properties of economic fluctuations in many variables.

From our plots of the data, we noted that there appear to be some outlying points. As yet, we have not formally tested for outliers nor utilized procedures for allowing for outliers in estimation and forecasting. Rather than face possible charges of "rigging the data" or "massaging the data" by eliminating outliers, etc., we decided to use all the data that relate to periods in which unusual events occurred in many countries, e.g. the Korean, Vietnamese and Gulf wars, energy crises of the 1970s, institution of wage and price controls, some abrupt changes in monetary and fiscal policies, etc. No dummy variables or other devices were used to deal with these unusual historical events. It was assumed and hoped that our lagged stock market and monetary variables would reflect the impacts of such events on economies.

With the data described, it was thought important to carry along some "benchmark" models in forecasting. In earlier studies, random walk models and simple time series models, e.g. third order autoregressive models, AR(3) models, or Box-Jenkins ARIMA models had been found to forecast better than some large scale models; see, e.g. the discussion of such studies by Christ, Cooper, Nelson, Plosser and others in Zellner (1984, 1997a). However, we generally found that such benchmark models, as well as other models, tended to make large errors in the vicinity of turning points in economic activity. Thus we thought it very important not just to study point forecasts but also to develop methods for making good turning point forecasts.

With respect to point forecasting, In Zellner and Hong (1989), we estimated the fixed parameter ARLI and ARLI/WI models, discussed above in connection with (4)-(5), using data, 1954-1973, for first 9 and then 18 industrialized countries and made one-year ahead point forecasts with parameter estimates updated year by year for the period 1974-1984. In this work, our ARLI and ARLI/WI models' performance was much better than that of various benchmark models and competitive in terms of root mean squared errors (RMSEs) to that of some large-scale OECD models. Also, the ARLI/WI model performed better than the ARLI model. Similar results were found to hold in later studies involving extended forecast periods, e.g. 1974-1997. See, e.g., papers in Sect. IV of Zellner (1997) and Zellner and Palm (2003) for some of these results.

The methodological tools utilized in this forecasting work included the following:

(1) Finite sample posterior densities for parameters of relations were employed in estimation of parameters, analyses of the realized error terms and study of the properties of dynamic relations, such as (4)-(5) above. For example, in Hong (1989) using data for each of 18 countries, draws were made from the trivariate Student t marginal posterior density of the autoregressive parameters of the ARLI model in (4) and for each draw, the three roots of the process were computed. It was found that in a large fraction of the draws, about .85, that there were two complex roots and one real root associated with the model. Also, the computed posterior densities for the periods of the cycle associated with the complex roots were found to be centered at about 3 to 5 years and those for the amplitudes were centered at values less than one for both the real and complex roots. Note that in this work, there was no need to rely on asymptotic approximate results, usually employed in non-Bayesian work with time series models.

(2) Means of predictive densities for future observations were employed as point forecasts that are optimal relative to a quadratic predictive loss function. With use of diffuse priors and usual likelihood functions or by use of BMOM predictive means, these forecasts were identical to least squares forecasts when fixed parameter models were employed but not in cases involving use of random parameter models and shrinkage assumptions. Recall that a usual predictive density for a future vector of observations, y_f is given by $h(y_f|D) = \int q(y_f|q)g(q|D)dq$, where D denotes the given data and $g(q|D)$ is the posterior density for the vector of parameters q .

(3) Predictive densities were employed to compute optimal turning point forecasts. For example, given a definition of a turning point, e.g. given two previous growth rates below the third, that is y(T-2), y(T-1)< y(T), where T denotes the current period, if y(T+1) <y(T), this is defined to be a downturn (DT), while if y(T+1) is not below y(T), this outcome is defined to be no downturn (NDT). Given such a sequence of outcomes up to period T and a predictive density for y(T+1), it is direct to compute the probability that y(T+1) is less than y(T), that is the probability of a downturn, P, and that of no downturn, 1-P. Then on considering a two by two loss structure associated the acts, forecast DT and forecast NDT and the possible outcomes, namely DT or NDT, it is direct to derive the forecast that minimizes expected loss. For example, if the 2x2 loss structure is symmetric, it is optimal for forecast DT when P>1/2 and NDT when P<1/2. Similar analysis applies to forecasting upturns and no upturns. Using such simple turning point forecasting procedures with a number of alternative models, about 70 per cent or more of 211 turning points for 18 countries were forecasted correctly. Also, such turning point forecasts were better than those provided by a number of benchmark procedures, e.g. coin tossing, or systematic optimistic or systematic pessimistic forecasts, etc. Such fine performance in turning point forecasting was a pleasant surprise! See Zellner and Min (1999) for results using many variants of the models in (4) and (5).

(4) Posterior odds were employed in variable selection, model comparison and model combining. As regards model selection, several researchers using our data and their model identification procedures determined forms for our ARLI model, given in (4) above, that differed from ours in terms of lag structures, etc. One group reported mo dels that forecasted much, much better than our models, so much better that I suspected something must wrong. Indeed it turned out that they had used all the data to fit their model and then used a portion of the sample for forecasting experiments. When this error was corrected, the improved performance of their model disappeared. On the other hand several others using their model identification procedures did produce models that are competitive with ours. This prompted our use of posterior odds to compare various variants of our model. We employed 8 input variables, the 7 shown in connection with (4) and an eighth, the money variable lagged two periods for a total of 8 possible input variables and thus $2^8 = 256$ possible

linear models that included our model and those of our two "competitors." Fortunately, our model compared favorably with the 255 alternative models, including those of our competitors that also performed reasonably well. For methods and specific results, see Zellner and Min (1993) and Zellner (1997). These methods are applicable to many other model selection problems and involve a correction for the possibility that on considering so many different models, one may have fit the data well just by chance, a "model selection" effect that has been discussed by Jeffreys (1998, p. 253 and p. 278) and in our paper.

Also, posterior odds have been very useful in comparing fixed parameter and time varying parameter models, showing a slight preference for time varying parameters. As regards time varying parameters, they have been widely employed in engineering state space modeling. In our context, there are many reasons why parameters, e.g. those in (4) above, might be time varying, including aggregation effects, changes in technology and/or tastes and preferences, effects of changes in economic policies, i.e. Lucas effects, etc. Our earliest work involved use of the assumption that the coefficient vector in (4) followed a vector random walk, $b_i(t) = b_i(t-1) + v_i(t)$ with $v_i(t)$ a white noise error vector and of Bayesian recursive methods to update parameter and predictive densities. Later, other variants were employed, namely

(i) $b_i(t) = q + u_i(t)$,

(ii) $b_i(t) = q(t) + u_i(t)$ and $q(t) = q(t-1) + h(t)$, etc.

Here we are introducing dynamic Stein-like assumptions that the individual country coefficient vectors are distributed about a mean vector that may be a constant, as in (i) or may be generated by a random vector random walk process, as in (ii). These and other models both for our ARLI and ARLI/WI models were implemented with data and found to perform better than models that did not incorporate Stein-like shrinkage effects, both for fixed and time varying parameter versions. Posterior odds for such time varying parameter models versus fixed parameter models were derived and computed. They indicated some support for the use of time varying parameters and much support for the use of shrinkage assumptions. For example, using shrinkage, the median RMSE of forecast for the 18 countries annual forecasts, 1974-87, was 1.74 percentage points while without shrinkage it was 2.37 percentage points. For more on these empirical results, derivations of posterior odds and use of them in combining models, see Min and Zellner (1993) and Zellner (1997). In this work involving combining models and their forecasts, comparisons were made with results of non-Bayesian combining methods of Bates, Granger and others. In these calculations, the Bayesian combining techniques performed slightly better than non-Bayesian combining techniques but did not result in much improvement relative to uncombined forecasts.

(5) As mentioned earlier, our empirical forecasting models were shown to be compatible with certain theoretical macroeconomic models. While this compatibility with economic theoretical models was satisfying, there was still the need to improve the explanatory and forecasting performance of our

models and to include additional variables to be forecast. How this was to be done and what was done is the subject of the next section.

IV. Disaggregation and the Marshallian Macroeconomic Model

In connection with the need to improve and extend our models, it was thought advisable to disaggregate the output variable since with disaggregated data there would be more observations and information that might improve forecasts and provide better explanations of the variation in total output. Further, in Zellner and Tobias (2000) the results of an experiment were published to indicate that disaggregation can be effective in improving forecasting precision. In particular, (a) equation (5a), was employed to forecast future values of w(t), median growth rate of 18 countries. These results were compared to those provided by two other procedures, namely, (b) use of (5a) and (5b) to forecast individual countries' growth rates and use of the median of the 18 forecasts as a forecast and (c) use of just equation (4) to forecast individual countries' growth rates and use of the median of these forecasts as a forecast of the median growth rate. These calculations indicated that method (b) that involves disaggregation was much better than methods (a) and (c) indicating that some forms of disaggregation can lead to improved forecasts. These empirical results suggested that improved forecasts of future, aggregate output of a country might be obtained by summing forecasts of its components, as shown theoretically in papers by de Alba and Zellner (1991) , Lütkepohl (1986) and others. The main issue was how to disaggregate total output in a meaningful, fruitful and effective manner.

One morning while shaving, it occurred to me that it might be worthwhile to disaggregate by industrial sector, e.g. agriculture, mining, construction, retail trade, etc. using traditional Marshallian demand, supply and firm entry/exit relations for each sector. While demand and supply relations had appeared in many earlier macroeconomic models, few, if any, included entry and exit relations. For example, in real business cycle models, there is the representative firm and one wonders what would happen should this firm shut down.

Fortunately, in earlier work, Veloce and Zellner (1984), we had derived and implemented a model of the Canadian furniture industry to show the importance of taking account of entry and exit of firms and the number of firms in operation in explaining variations in furniture supply. When the variable, number of firms in operation was omitted from the supply equation, as in many previous studies, very unsatisfactory results were obtained whereas with the inclusion of this variable, more sensible estimation results were obtained. In Zellner (2001), a slightly revised version of the original Marshallian sector model was derived that involved competitive conditions with N profit maximizing firms, each with a Cobb-Douglas production function including neutral and factor augmenting technical change. Assuming that firms maximize profits, the firm's supply function was derived and multiplied by N(t), the number of firms in operation a time t and the real price of output, p(t) to obtain a supply relation for industry real sales, S(t) = p(t)N(t)q(t) which when differentiated with respect to time, led to the following dynamic supply relation:

(6) (1/S) dS/dt = (1/N)dN/dt + $h(1/p)dp/dt$ + a SUPPLY

where h is a parameter and a denotes a linear function of exogenous variables shifting supply, e.g. rates of growth of the real wage rate, the real price of capital etc.

Further, with a "log-log" demand function for the sector's output, Q, multiplied by the real price of output, real sales demanded is S = pQ and log S is a linear function of the logs of real price, real income, real money balances and other variables. On differentiating log S with respect to t, the following relation is obtained:

(7) (1/S)dS/dt = v(1/p)dp/dt + b DEMAND

where v is a parameter and b = a linear function of the rates of change of variables shifting the demand function, e.g. real income, real money, etc.

As regards entry and exit of firms, we assume:

(8) (1/N)dN/dt = c (S – F) ENTRY-EXIT

where c and F are positive parameters, the latter associated with a cost of entry. Also, in this model, S is proportional to real industry profits.

Thus we have three equations for the three variables, p(t) price, S(t) real sales and N(t) number of firms in operation. On substituting from (8) in (6) and then eliminating the variable (1/p)dp/dt from the remaining two equations, the final, implied equation for S(t) is:

(9) (1/S)dS/dt = r – S/F + g

where r is a parameter and g is a linear function of the rates of change of the input variables shifting the supply and demand relations, e.g. the real wage rate, the real price of capital services, real income, real money balances, etc. Note too that for g= constant or g = 0, (9) is in the form of the logistic differential equation with solution the logistic function that has been often used to model many industries' output. Showing that such empirical logistic functions are compatible with a traditional Marshallian model of a competitive industry is indeed satisfying. See Zellner (2001) for more information regarding the derivation of this model.

With the above results available, in Zellner and Chen (2001), discrete versions of (9) were formulated and fitted using data for 11 sectors of the U.S. economy. Before turning to particular models, note that it is well known that a discrete version of the homogenous part of (9) is in the form of a chaotic model that can have quite unusual outputs; see, e.g. Kahn (1990, Ch. 16) and Koop, Peaaran and Potter (1996) for plots of the outputs of such chaotic models. This raises the question as to whether the world is better modeled using (1) a model using continuous time, (2) a model based on discrete time or (3) a mixed continuous-discrete time model , as for example used in modeling certain biological populations as mentioned in Zellner (2002) with reference to Cunningham (1958). While it would be interesting in future work to entertain a mixed model and to compute posterior odds on it versus a discrete time model, in recent work, Zellner and Chen (2001), we started with discrete time models, one of which is

$$
\begin{aligned}
(10)\quad (1-L)\log S_t &= a_o + a_1 S_{t-1} + a_2 S_{t-2} + a_3 S_{t-3} + b_1 (1-L)\log Y_t \\
&+ b_2 (1-L)\log M_{t-1} + b_3 (1-L)\log w_t + b_4 ((1-L)\log SR_{t-1} + u_t
\end{aligned}
$$

where S denotes real sector sales, Y real income, M, real money, w real wage rate, SR real stock price index and u an error term.

It is seen that the rate of change and levels of S enter the equation, a so-called "cointegration effect" that follows from the economic theoretical model presented above. The lags in the S variable were introduced to represent possible lags in the entry equation and in the demand equations. With parameters allowed to vary over sectors, the relation in (10) was fitted to 11 sectors' data for the U.S. economy, agriculture, mining, construction, etc. using data, from the early 1950s to 1979 to fit the models and then the relations were employed to forecast one year ahead with estimates updated year by year, 1980 to 1997. With 11 sector forecasts available each year, they were used to produce forecasts of the rate of change of aggregate real output (GDP) year by year. Such forecasts were compared to forecasts obtained from aggregate annual data using (i) an AR(3) model, (ii) an ARLI model and (iii) an aggregate one sector Marshallian model in the form of (10) The benchmark AR(3) model missed all the turning points and yielded a mean absolute error (MAE) of forecast = 1.71 percentage points and root mean squared error (RMSE) = 2.32 percentage points. The aggregate Marshallian model's MAE = 1.48 and RMSE = 1.72 percentage points are much lower than those associated with the AR(3) and ARLI models imp lemented with aggregate data.

When a time series AR(3) model was implemented using least squares methods for each sector and forecasts were aggregated to form forecasts of the rates of growth of aggregate real GDP, there was not much improvement, namely a MAE = 1.65 and RMSE = 2.26 percentage points. In contrast, the ARLI model and the Marshallian equation in (10), implemented with the disaggregated sector data and using least squares estimation and forecasting methods produced much better annual forecasts of rates of growth of aggregate GDP with forecast MAE = 1.25 and RMSE = 1.47 percentage points for the Marshallian model in (10) and MAE =1.32 and RMSE = 1.62 percentage points for the ARLI model. Thus this set of experiments indicates that IT PAYS TO DISAGREGATE for two of the three models considered above.

In other experiments we employed various estimation and forecasting methods that yielded some improved forecasts relative to those provided by least squares. For example, using a seemingly unrelated regression approach for joint estimation of the 11 sector relations in the form of (10), taking account of differing error term variances and their correlations across equations, and one year ahead forecasting with the 11 sectors' data, the aggregate growth rate forecasts had MAE = 1.17 and RMSE = 1.40 percentage points. Also,, many other estimation and forecasting techniques employed yielded MAEs ranging from 1.2 to 1.4 and RMSEs ranging from 1.4 to 1.6 percentage points thus indicating that it was probably the added information provided by disaggregation that produced improved forecasting performance. See Zellner and Chen (2001) for detailed presentation of the data and forecasting results. Also, it should be noted that forecasts for the highly variable agricultural, mining , construction and durable goods industrial sectors were not very accurate and need to be improved, e.g. by introduction of weather variables for the agricultural sector, etc. With such improvements, there will be improvements not only in sector forecasts but also probably in aggregate GDP forecasts. Also, perhaps fitting all three relations, shown in (6)-(8) above for each sector may lead to improved sector and aggregate forecasts.

Note that a MAE of about 1.2 percentage points for forecasting annual real GDP for the U.S., as obtained in the above forecasting experiments, compares favorably with MAEs reported in Zarnowitz (1986, Table 1, p. 23). As he reports, for the periods, 1969-1976 and 1977-84, the MAE = 1.2 percentage points for the one year ahead forecasts of the rates of growth of real GNP made by the U.S. Council of Economic Advisors. Of course, "on line" forecasters have to cope with preliminary data problems not present in our forecasting experiments. However such on line forecasters typically use a lot of judgmental, outside information to adjust their models' forecasts and sometimes combine models' and others' forecasts in efforts to get improved forecasts that we did not do.

In summary, the above results based on our Marshallian sector models are encouraging and further work will be done to get improvements. Along these lines, it will be useful to add relations to get a closed model of the economy, our Marshallian Macroeconomic Model. The first steps in this direction are described in the next section.

V. A Complete Marshallian Macroeconomic Model

Above, supply, demand and entry relations were formulated for each sector. Attention was focused on the final product market in which each sectors' producers are assumed to sell. To close the model, there is a need to add factor markets, that is labor, capital and money markets as well as an international sector. The roles of intermediate goods and inventories need attention. And the operations of federal, state and local governments have to be incorporated in the model. Consideration of the birth of new sectors and the deaths of old sectors is needed as well as allowance for regulated and imperfectly competitive sectors. Improvements in entry and investment equations are possible. Fortunately, there is much valuable research on many of these topics in the literature, which can be incorporated in our model and hopefully, improve it. However, what is initially needed is a "bare bones" complete model, that is a "Model T" that works reasonably well.

In our approach to this last problem, we have started analyzing a one sector closed economy MMM model by adding labor, capital and money markets to the product market discussed above. From initial analyses, currently underway with G. Israilevich, when such additions are made, the model is mathematically tractable and leads to final equations for the variables of the system that are in forms similar to that presented in (9) above. However, when we go to a two sector model, there are interesting interactions between sectors produced by dependencies in demand and supply relations that affect properties of solutions. Elements of stability and instability are encountered for certain values of strategic parameters. Every effort is being made to keep the model in a form so that mathematical analyses of its properties are tractable. If not, computer simulation techniques will be employed to help determine the properties of the overall model including its responses to changes in policy variables, e.g. money, tax rates, etc. as has been done in much past work, including Zellner and Peck (1973). And of course, additional forecasting experiments will be carried forward using as much new data as possible.

It was a pleasure having an opportunity to discuss our work and results with many at this Conference. I hope that the above account of our modeling experiences using large data sets will be of interest and value to many data miners and that our future work will benefit from research that you have done and are currently carrying forward. Hopefully, we shall all strike gold.

References

Aigner, D. and Zellner, A. (eds.) (1988), Causality, Annals Issue of the Journal of Econometrics, 39, 234pp.

Berger, J.O. (2000) , "Bayesian Analysis: A Look at Today and Thoughts of Tomorrow," J. of the Am. Stat. Assoc., 95, Dec., 1269-1276.

Box, G. and Jenkins, G. (1970), Time Series Analysis: Forecasting and Control, San Francisco: Holden-Day.

Burns, A. and Mitchell, W. (1946), Measuring Business Cycles, New York: Columbia U. Press & National Bureau of Economic Research.

Cunningham, R. (1958), Introduction to Non-Linear Analysis, New York: McGraw-Hill.

De Alba, E. and Zellner, A.. (1991), "Aggregation, Disaggregation, Predictive Precision and Modeling, ms., 7 pp.

Feigl, H. (1953), "Notes on Causality," in Feigl, H. and Brodeck, M. (eds.), Readings in the Philosophy of Science, New York: Appleton-Century-Crofts, 408-418.

Garcia-Ferrer, A., Highfield, R., Palm, F. and Zellner, A.(1987), "Macro-economic Forecasting Using Pooled International Data," J. of Business and Economic Statistics, 5, 53-67.

Green, E. and Strawderman, W. (1996), "A Bayesian Growth and Yield Model for Slash Pine Plantations," J. of Applied Statistics, 23, 285-299.

Hong, C. (1989), Forecasting Real Output Growth Rates and Cyclical Properties of Models, Ph.D. Thesis, Dept. of Economics, U. of Chicago.

Jeffreys, H. (1973), Scientific Inference, Cambridge: Cambridge U. Press, 3rd ed.
Jeffreys, H. (1998), Theory of Probability, Oxford Classic Series, Oxford: Oxford U. Press, reprint of 3rd revised edition.
Koop, G., Pesaran, M. and Potter, S. (1996), Impulse Response Analysis in Nonlinear Multivariate Models," J. of Econometrics, 74, 119-147.
Lütkepohl, H. (1986), "Comparisons of Predictors for Temporally and Contemporaneously Aggregated Time Series,"
International J. of Forecasting, 2, 461-475.
Kahn, P. (1990), Mathematical Models for Scientists and Engineers, New York: Wiley.
Min, C. (1992), Economic Analysis and Forecasting of International Growth Rates, Ph.D. Thesis, Dept. of Economics, U. of Chicago.
Min, C. and Zellner, A. (1993), "Bayesian and Non-Bayesian Methods for Combining Models and Forecasts with Applications to Forecasting International Growth Rates," J. of Econometrics,56, 89-118, reprinted in Zellner (1997a).
Mittelhammer, R., Judge, G., and Miller, D. (2001) Econometric Foundations, Cambridge: Cambridge U. Press.
Quenouille, M. (1957) The Analysis of Multiple Time Series, New York: Hafner Publishing Co.
Soofi, E. (1996), "Information Theory and Bayesian Statistics," in Berry, D., Chaloner, K. and Geweke, J., Bayesian Analysis in Statistics and Econometrics, New York: Wiley, 179-189.
Soofi, E. (2000), "Information Theory and Statistics," J. Am. Stat. Assoc., 95, Dec., 1349-1353.
Van der Merwe, A., Pretorius, A., Hugo, J. and Zellner, A. (2001), "Traditional Bayes and the Bayesian Method of Moments Analysis for the Mixed Linear Model with an Application to Animal Breeding," South African Statistical Journal, 35, 19-68.
Veloce, W. and Zellner, A. (1985), "Entry and Empirical Demand and Supply Analysis for Competitive Industries," J. of Econometrics, 30, 459-471.
Zarnowitz, V. (1986), "The Record and Improvability of Economic Forecasting," Economic Forecasts, 3, 22-31.
Zellner, A. (1984), Basic Issues in Econometrics, Chicago: U. of Chicago Press.
________ (1994a), "Time Series Analysis, Forecasting and Econometric Modeling: The Structural Econometric Modeling,
Time Series Analysis Approach," invited paper with discussion in J. of Forecasting, 13, 215-233.
__________(1994b), " Bayesian Method of Moments/Instrumental Variable Analysis of Mean and Regression Models," in Lee, J., Johnson, W. and Zellner, A. (eds.), Modelling and Prediction, New York: Springer, 61-76, reprinted in Zellner (1997a).
_________ (1996), An Introduction to Bayesian Inference in Econometrics, Wiley Classics Library, New York: Wiley.
_________ (1997a), Bayesian Analysis in Econometrics and Statistics: The Zellner View and Papers, invited contribution to
M. Perlman and M. Blaugh (eds.) Economists of the Twentieth

Century Series, Cheltenham, UK: Edward Elgar Publ. Ltd.
_________ (1997b), "The Bayesian Method of Moments: Theory and Applications," in Fomby, T. and Hill, R. (eds.), Advances in Econometrics, 12, 85-105, Greenwich, CT: JAI Press.
_________(1999), "Bayesian and Non-Bayesian Approaches to Scientific Modeling and Inference in Economics and Econometrics," invited paper for Ajou U. Research Conf. Honoring Prof. Tong Hun Lee, Korea J. of Money and Finance, 5, 11-56.
_________(2001a), "Remarks on a 'Critique' of the Bayesian Method of Moments (BMOM)," J. of Applied Statistics 28, No. 6, 775-778.
________ (2001b), "The Marshallian Macroeconomic Model" in T. Negishi et al (eds.), Economic Theory, Dynamics and Markets, Boston/Dordrecht: Kluwer Acad. Publishers, 19-29.
________ (2002), "Information Processing and Bayesian Analysis," J. of Econometrics, 107, 41-50.
________ and Chen, B. (2001), "Bayesian Modeling of Economies and Data Requirements," 5, 673-700.
________ and Hong, C. (1989), "Forecasting International Growth Rates Using Bayesian Shrinkage and Other Procedures," J. of Econometrics, 40, 183-202.
________, Kuezenkamp, H. and McAleer, M. (2001), "Simplicity, Inference and Modelling," Cambridge: Cambridge University Press.
________ and Min, C (1999), "Forecasting Turning Points in Countries' Output Growth Rates: A Response to Milton Friedman," J. of Econometrics, 88, 203-206.
________ and Palm, F. (1974), "Time Series Analysis and Simultaneous Equation Econometric Models," J. of Econometrics, 2, 17-54, reprinted in Zellner and Palm (2003).
_______ and Palm, F. (1975), "Time Series Analysis of Structural Monetary Models of the U.S. Economy," Sankya, Series C, 37, 12-56, reprinted in Zellner and Palm (2003)
________ and Palm, F. (eds.), (2003), The Structural Econometric Modeling, Time Series Analysis Approach, to be published by Cambridge U. Press.
________ and Peck, S. (1973), "Simulation Experiments with a Quarterly Model of the U.S. Economy," in Powell, A. and Williams, R. (eds.), Econometric Studies of Macro and Monetary Relations, Amsterdam: North-Holland, 149-168, reprinted in Zellner (1984).
________, Tobias, J. and Ryu, H. (1999a), Bayesian Method of Moments Analysis of Time Series Models with an Application to Forecasting Turning Points in Output Growth Rates," Estadistica, 49-51, pp. 3-63, with invited discussion.
________, Tobias, J. and Ryu, H. (1999b), "Bayesian Method of Moments (BMOM) Analysis of Parametric and Semi-parametric Regression Models," South African Statistical Journal, 31, 41-69.
________ and Tobias, J. (2001), "Further Results on Bayesian Method of Moments Analysis of the Multiple Regression Model," International Economic Review, 42, No. 1, 121-140.