Information Processing and Bayesian Analysis

Arnold Zellner[*]

Graduate School of Business, University. of Chicago, Chicago, IL 60637, USA

Abstract

Science involves learning from data. Herein this process of learning or information processing is considered within the context of optimal information processing, as in Zellner (1988, 1991, 1997). Information criterion functionals are formulated and optimized to provide optimal information processing rules, one of which is Bayes' theorem. By varying the inputs and using alternative side conditions, various optimal information processing rules are derived and evaluated. Generally output information = input information for these rules and thus they are 100% efficient learning rules. When different weights or costs are associated with alternative inputs, "anchoring" like effects, much emphasized in the psychological literature are the results of optimal information processing procedures. Further, dynamic information processing results are reviewed and extensions noted. Last, some implications of the information processing approach for learning from data will be discussed.

JEL classification: C0, C1

Keywords: Optimal information processing; Bayes' Theorem; Learning theory

1. Introduction

In past work, Zellner (1988, 1991, 1997) scientific inference that involves learning from data and experience has been viewed as information processing. Input information contained in a likelihood function and in a prior density has been considered in relation to the information in an output density for parameters and a marginal density for the observations. Since we wish to have the output information as close as possible to the input information, a criterion functional, namely output information minus input information, was minimized with respect to the choice of an output density for the parameters and it was found that the optimal information processing rule is Bayes' theorem. Also, when Bayes' theorem is employed, output information = input

---

[*] Tel.: +1-773-702-7145; fax: +1-773-702-0458.
E-mail address: arnold.zellner@gsb.uchicago.edu

Research financed by the National Science Foundation and by income from the H.G.B. Alexander Endowment Fund, Graduate School of Business, University of Chicago.

information and thus the process is 100% efficient in the sense that no information is lost in the information processing procedure, an information conservation property of Bayes' theorem. In this past work, entropic measures of information have been employed and Jaynes (1988) commented as follows on the above approach and results: ". . . entropy has been a recognized part of probability theory since the work of Shannon 40 years ago, and the usefulness of entropy maximization as a tool in generating probability distributions is thoroughly established. . . This makes it seem scandalous that the exact relation of entropy to the other principles of probability is still rather obscure and confused. But now we see that there is, after all, a close connection between entropy and Bayes' theorem. Having seen this start, other connections may be found, leading to a more unified theory of inference. Thus in my view Zellner's work is probably not the end of an old story, but the beginning of a new one." (pp. 280-281) See also Hill (1988), Soofi (1996, 1997) and Bernardo and Smith (1994) for further comments on optimal information processing and Bayes' theorem.

In the current paper, the information processing approach will be reviewed and extended. For example, information processing when a single input, say a likelihood function is employed will be analyzed. Note that Fisher in his fiducial inference approach was interested in just inputting a likelihood function and not a prior density in his effort to perform inverse inference, that is to obtain post data densities for parameters of likelihood functions. Further, we shall consider information processing when the input information is a weighted average of the information in a prior density and the information in a likelihood function. It will be seen that when the weights are far from being equal, "anchoring" effects are observed that are optimal in an information processing sense. Also, it is possible to get optimal information processing solutions when the input information is just moment side conditions on parameters, as in the Bayesian method of moments approach described and applied in Green and Strawderman (1996), LaFrance (1998), Tobias and Zellner (1999), Zellner (1994, 1996, 1997), and Zellner, Ryu and Tobias (1999). From results so obtained, it is possible to see how much information is lost relative to the case in which the input information utilized is that in a prior density and a likelihood function. Note that Hill (1988, p. 281) points to "size-based sampling" as a procedure, among others, ". . . where it is routine for some statisticians to violate the likelihood principle, and thus also the conservation property, without ever changing the model." Finally, the case of initial ignorance, that is little or no initial information will be considered and related to earlier work on rational behavior under ignorance by Arrow and Hurwicz (1972), Cohen and Jaffray (1980), Hogarth and Einhorn (1992) and others.

Above, we have considered "static" information processing. The dynamic formulation of information processing put forward by a former graduate student in notes for my course will be reviewed and extensions noted. He showed that traditional Bayesian temporal updating of posterior densities to incorporate new sample information is optimal from an information processing point of view using a dynamic programming approach. However, with costs of change and other dynamic side conditions introduced, traditional Bayesian updating will not be optimal. As in many fields, solutions to dynamic problems often differ from solutions to static problems.

The plan of the paper is as follows. In Section 2, the information processing approach is presented and used to provide results using various informational inputs and side conditions. Relative measures of output information are derived and compared. Then in Section 3, dynamic formulation is reviewed and extended. Last, in Section 4, a summary of results, comments on their implications and some suggestions for future research are presented.

2. The Information Processing Approach to Learning from Data

2.1 Bayes Theorem as a learning model

Learning from data and experience is an important activity in science. Some learn informally while others use Bayes' theorem as a learning model. The Bayesian learning model utilizes two informational inputs, a prior density and a likelihood function and two output densities, a post data density for the parameters and a marginal density for the observations. In past work, the following information criterion functional was employed in analyzing information processing:

$$\Delta\left[g(q \left| D \right. \right] = OutputInfo - InputInfo \qquad (1)$$

where $g(q \left| D \right.)$ is a proper probability density function for $q \subset \Theta,$ a vector of parameters given D, the observed data, y, and prior information I, that is D= (y, I). Since we wish to have the output information as close as possible to the input information, the functional in (1) was minimized with respect to the choice of g, subject to its being a proper density and the result obtained is Bayes' theorem, namely:

$$g^{*}(q \left| D \right.) = p(q \left| I \right.) l(q \left| D \right.) / h(y \left| I \right.) \qquad (2)$$

where $p(q \left| I \right.) = p$ is the prior density, $l(q \left| D \right.) = l$ is the likelihood function and $h(y \left| I \right.) = h$ is the marginal density of the observations. Explicitly, the criterion functional that was minimized to obtain the result in (2) is:

$$\Delta(g) = \int g \ln g dq + \int g \ln h dq - \int g \ln p dq - \int g \ln l dq \qquad (3.a)$$

$$= \int g \ln\{g / [p l / h]\} dq \qquad (3.b)$$

In (3a) the information in a density is represented by its expected log height relative to uniform measure. On combining the terms in (3a), (3b) is obtained that is the Jeffreys-Kullback-Leibler (JKL) distance measure between g and g* given in (2). Since the JKL measure is known to be non-negative, taking g = g* in (3b) provides a minimal value such that input information = output information. Thus the Bayesian information processing rule (IPR) in (2) is 100% efficient. No information is lost or added when (2) is employed, an information conservation property.

Further, it should be recognized that the above optimization problem can be solved subject to moment side conditions, that is, $\int q^{i} g dq = m_{i}, i = 0,1,2,...,m$ by use of Lagrange multipliers to obtain $g \propto pl\exp\{\sum c_i q^i\}$, where the c's are Lagrange multipliers. Also differential equation side conditions can be employed, e.g. $d\ln g/dq = q(q)/z(q)$, resulting in the optimal solution, $g \propto p\,l\exp\{c[q(q)+z'(q)]\}$, where c is a Lagrange multiplier. If q and z are given particular forms, the side condition implies that g is a member of the Pearson class of densities; see, e.g., Jeffreys (1998) for a thorough discussion of the differential equation that yields the Pearson system of probability density functions. Further, the following "anchoring" side conditions can be used, $\int g\log(g/p_a)dq = a$ and/or $\int g\log(g/l)dq = b$, with given a and b denoting the distances between g and p and g and l, respectively. Thus information in a variety of side conditions can be introduced, as in general maxent problems, see e.g., Golan, Judge and Miller (1996) to obtain information processing rules that incorporate such information.

2.2 Quality of prior and sample information

In past work, Zellner (1988a, 1991, 1997, p. 171), it was recognized that the quality of the input prior and sample information may vary and thus quality adjusted priors and likelihood functions were introduced, namely, $p_a = q_1(p)$ and $l_a = q_2(l)$ and employed in place of p and l in the criterion functional in (3a). On minimizing it with respect to the choice of g subject to its being proper, the result is:

$$g_a \propto p_a l_a = q_1(p)q_2(l) \qquad (4)$$

For example, in connection with (4), the following may be assumed, namely, $q_1(p) \propto p^a$ and $q_2(l) \propto l^b$ with a and b having values in the closed interval 0 to 1. Note that raising a density to a fractional power will in many cases result in its being more spread out. If a = 0, indicating that the prior information is of very low quality, g is taken proportional to the likelihood function raised to the power b, i.e. $g_a \propto l^b$. If the sample information is considered of very low quality, b = 0, and $g_a \propto p^a$. To illustrate, in the case in which we have n independent observations, each with mean q and variance=1 and employ a normal prior with mean $\bar{q}$ and variance = 1/h, the quality adjusted output density is:
$g_a \propto \exp\{-ah(q-\bar{q})^2/2\}\exp\{-b\Sigma(y_i-q)^2/2\} \propto \exp\{-h'(q-q_m)^2/2\}$,
with $h' = ah+bn$ and the post data mean = $q_m = (ah\bar{q}+bn\bar{y})/(ah+bn)$, where $\bar{y}$ is the sample mean. It is seen that the values of a and b affect the precision and location of the post data density. With a = b = 1, the standard results are obtained.

The above example illustrates the effects of introducing quality considerations in making inferences about a mean. It is the case that the above adjustments for quality lead to sample and prior information being down weighted and to a posterior density for the mean with greater dispersion relative to the situation in which prior and sample information are of very high quality, i.e. a = b =1.

When we have a form for g, such as that given in (4), that differs from g* in (2), we can compute the information loss using the JKL information measure, $\int g^* \log(g^* / g_a) dq \succ 0$ or $E \log g^* > E \log g_a$, that is, the expected log height of g^* is greater than expected log height of g_a for $g^* \neq g_a$. Such calculations can reveal the loss of information associated with the use of low quality inputs or the use of just one input, say a likelihood function, rather than the usual two inputs, namely prior and likelihood functions. Further, if two forms of a prior density are under consideration to be combined with a given likelihood function, the information in their associated post data densities can be computed using the JKL information measure.

2.3 Weighting the prior and sample information

In some circumstances, it may be appropriate to weight the prior information and the sample information differently using non-negative weights, w_1 and w_2 that can reflect the costs of departing from a uniform prior density and a uniform density for the observations given the parameters. Then the criterion functional in (3) becomes:

$$\Delta(g) = \int g \log g dq + \int g \log h dq - w_1 \int g \log p \, dq - w_2 \int g \log l dq \qquad (5)$$

And minimization with respect to the choice of g subject to its being proper, yields:

$$g^{**} \propto p^{w_1} l^{w_2} \qquad (6)$$

The result in (6) is a modified version of (2) and similar in form to (4).

As regards (6), if $w_1 = 0$ and $w_2 = 1$, we are inputting just the information in the likelihood function and not inputting the information in the prior, just as Fisher wished to do in his fiducial inference approach and we obtain an optimal post data density for the parameters that is proportional to the given likelihood function. Also, this results when we input a uniform prior for the parameters along with the information in the likelihood function. If we just input a prior density and no likelihood function, then the optimal post data density for the parameters is the prior density. If the prior is diffuse, the output density will be diffuse in line with the precept, nothing in nothing out. Then too, if one is “deeply attached” to a prior, he/she can choose the value of w_1 accordingly, an example of an “anchoring” effect. Also, the JKL measure as well as the measure of the information provided by an experiment, Zellner (1997, pp. 149-153) can be employed to compare the information content of various solutions.

2.4 Rational behavior under ignorance

Arrow and Hurwicz (1972) and Cohen and Jaffray (1980), provide results on “rational behavior under ignorance” that usually involve going to extremes. Cohen and Jaffray (1980) write: “The axioms of Arrow and Hurwicz imply that the decision maker’s preferences can take into account only the outcomes yielded by the acts, and in fact, among these outcomes, only the extremal values." (p. 1289) and “ A property worthy of note exhibited by these criteria [for rational behavior under ignorance] is that, in a first-order approximation, they depend on the sole comparison between the extremal values of acts. . . Our results concern only complete ignorance. They nonetheless succeed in casting a doubt on the existence of personal probabilities in the case in which the decision maker possesses little information on events [that is, is ignorant]"(p. 1296). Thus it appears that given these authors' definition of ignorance, rational behavior under ignorance will involve "going to extremes" and can not involve the use of personal probabilities to maximize expected utility or minimize expected loss in solving problems "rationally."

The above authors are concerned with "personal" probabilities and do not consider Jeffreys' (1998) well known use of unbounded measures in his representation of ignorance about parameters' values. For example, if we use an improper prior for a location parameter $\theta, p(\theta) \propto const., -\infty < \theta < \infty$, we have Pr (a<$\theta$<b) = 0 and Pr (c<$\theta$<d)=0, and the ratio of these probabilities is indeterminate, which for Jeffreys (1998) is a representation of ignorance. Also, see Rényi (1970) for an axiom system for probability theory that accommodates such unbounded measures. Assigning values of 0 to the above probabilities is indeed rather extreme behavior as is also the indeterminate odds statement. Note too that with Jeffreys' improper priors, it is not possible to minimize, say, the expectation of a quadratic loss function since the needed moments don't exist. As regards information processing, if a Jeffreys' improper prior is the only input information and we minimize the associated JKL information measure, the optimal value of g is the improper prior. That is "no information in," "no information out". On the other hand, if we are not totally ignorant, e.g., we might know the mean of a non-negative parameter, say $E\theta = m$. Then using this side condition, the optimal value for $g = (1/m)\exp\{-\theta/m\}$. Such side conditions arise when $y_i = \theta + e_i$, i =1,2,..., n, where the y's are given observations, say times to failure, and the e's are realized error terms. On summing both sides, dividing by n and taking the expectation of both sides, $\bar{y} = E\theta + E\bar{e}$ where E is a subjective expectation operator. If there are no outliers or missing variables and the functional form of the relation is satisfactory, it can be assumed that $E\bar{e} = 0$ and then $E\theta = \bar{y}$, a moment side condition and the density that minimizes the JKL information measure relative to uniform measure or maximizes the entropy is the exponential density, $g = (1/\bar{y})\exp\{-\theta/\bar{y}\}$, a Bayesian Method of Moments (BMOM) result; see the references given at the end of the paper for applications of the BMOM approach to many problems that can be given this information processing interpretation.

Above the exponential density was obtained as the output of an information processing problem given just the information stated above. It is obvious that if there is ignorance about outliers being present in the sample of data, the mean may take on an extreme, unusual value. Thus in this and many other problems, ignorance about properties of the data may, as is well known, result in "optimal" mean values that assume "ignorant" extreme values. See Izan (1980) for analysis of past empirical studies that neglected to take account of outliers. The same can be said in cases in which one "ignorantly" uses models in which there are left out variables or other problems with an assumed functional form for a relation. For analysis of such an example, see Zellner and Moulton (1985) wherein it was pointed out that use of c = a + by + u, is an inappropriate relation to employ for permanent consumption, c and permanent income y in order to test the hypothesis that a = 0. Note that 0<c<y, and thus given y, a and b, u is not

normally distributed. Further, if under the alternative hypothesis, a>0, then at very low levels of y,
$Ec|y,a,b$ can be larger than y, which is an “ignorant” extreme implication. By using the functional form, suggested in Zellner and Sankar (1967), log [z/(1+z)] = a’+ b’ log y + e, with z = c/y, the above difficulties are avoided. Thus, as is well known, ignorance about the inputs to information processing can result in absurd, extreme, “optimal” outputs.

2.5 Some comments on static information processing

To guard against such extreme, absurd outputs, diagnostic checking and predictive testing of models, using new data, have long been recommended. In diagnostic checking, many times some alternative model, as mentioned in the previous paragraph is available.
Then the information in the alternative models produced by alternative optimal information processing rules can be compared and evaluated. In prediction, new data become available and rather than static information processing, we have to consider temporal, sequential, dynamic information processing procedures.

3. Dynamic Information Processing

In this section, we shall briefly review the analysis of dynamic information processing brought to my attention by a student in my course in the early 1990s[1] and then consider additional problems. In the formulation of dynamic information processing, the problem is to choose an optimal sequence of functions, $g(q\,|D_t, t) \equiv g(q,t)$, where $D_t = (y_1, y_2, ..., y_t)$, the available data at time t, that minimizes the following criterion functional:

$$\sum_{t=1}^{t} \int \Pi(q,t) \log\left[\Pi(q,t)\, p(t) / \Pi(q,t-1)\, f(q,t)\right] dq \qquad (7)$$

subject to: $\int \Pi(q,t)\,dq = 1$, for all t and $\Pi(q,0)$ a given prior at t=0 and where

$p(t) \equiv p(D_t|t)$, $f(q,t) \equiv f(D_t|q,t)$ and $\Pi(q,t) \equiv \Pi(q\,|D_t,t)$, the post data density for q at time t given data up to time t, denoted by D_t.

The objective function in (7) is a dynamic version of that used in Zellner (1988) and the optimization problem is a dynamic programming problem. The optimal first and subsequent periods’ solutions are obtained by backward induction and use of Bellman’s principle of optimality. Surprisingly, the optimal sequential solution is the usual Bayesian updating procedure. That is, for period t, t = 1,2, T, the solution is:

---

[1] The student did not provide a reference for the results reported below. If anyone has this information, please send it to the Editor and author.

$$\Pi^*(q,t) = \Pi^*(q\ ,t-1)\ f(q,\ t)\ /\ p(t) \tag{8}$$

an optimal information processing rule relative to the objective function in (7). Note that (8) is the usual updating rule using Bayes' theorem that permits investigators to combine information from more than one sample in a logical, optimal fashion. Further, period-by-period the value function, evaluated using the optimal solution in (8) is equal to zero. That is, period-by-period, input information = output information and thus the dynamic information processing procedure is 100% efficient.

As in the static case, it is possible to allow for differing quality of the prior and sample information or to weight the sample and prior information differently and solve for the optimal sequential solution using the dynamic programming approach. Solutions will be similar to those obtained in the static case. Further in the case that the information is valued, as in (7) above, in the dynamic formulation discounting can be introduced to value net information obtained in the future. That is, a discount factor, say $b(t) = 1/(1+r)^t$, where r is the discount rate, would be used in the summation in (7) and the modified criterion functional can be employed to obtain sequential solutions. Thus, many problems can be recast in a dynamic programming framework and solutions obtained using standard dynamic programming techniques.

In Hogarth and Einhorn (1992), their "belief adjustment model" is explained and shown to be useful in interpreting experimental data relating to belief updating. Their model, a "partial adjustment" model, given in their equation (1), makes the change in beliefs, $S_t - S_{t-1}$ equal to an adjustment "weight" w_t, $0 \le w_t \le 1$, times the difference between the subjective evaluation of current information, $s(x_t)$ minus a "reference point", R , "against which the impact of the t'th piece of evidence is evaluated." (p.8). That is, their belief adjustment model is given by:

$$S_t - S_{t-1} = w_t[s(x_t) - R] \tag{9}$$

To derive a "partial adjustment" equation in the information processing framework that can be compared to the Hogarth-Einhorn equation the following dynamic loss function, incorporating a "target" function and a cost of changing beliefs is minimized by choice of the form of $g_t(q|D) \equiv g_t$,

$$L_t = (g_t - g_t*)^2 + c_t(g_t - g_{t-1})^2 \tag{10}$$

where g_t* is target function given in (2), which is an optimal solution given no costs of change. Of course, other target functions can be utilized. On minimizing the integral of (10) with respect to choice of g_t, the solution is:

$$g_t^o = (g_t^* + c_t g_{t-1})/(1+c_t) \tag{11a}$$

or

$$g_t^o - g_{t-1} = a_t(g_t * - g_{t-1}) \tag{11b}$$

with the “adjustment parameter” $a_t = 1/(1 + c_t)$. While the solution in (11) is a one period solution, it is possible to get solutions to the T period problem, given initial conditions, a dynamic quadratic programming problem. Further, (10) is just one possible form of the loss function. One could replace it with a JKL distance loss function and obtain the associated solution. That the solution in (11b) is similar in form to the Hogarth-Einhorn equation is noteworthy.

4. Summary and Conclusions

In this paper, it has been shown that learning or information processing rules can be derived analytically by optimizing information criterion functionals. Bayes’ theorem or information processing rule was shown to be an optimal rule both in static and dynamic learning contexts. On varying the conditions surrounding the learning process, e.g., allowing for cost of changing beliefs or weighting prior and sample information differently, new learning models were derived, one of them close in form to the “belief adjustment” model of Hogarth and Einhorn (1992). That solutions to more general learning problems differ in form from Bayes’ theorem or information processing rule is not an indication of irrationality or ignorance. Rather they reflect differing assumptions regarding the information processing problem to be solved. Such differences arise in many fields, e.g. in economics when the forms of the utility functions or profit functions are changed. The empirical task is to determine which rules work well in solving specific information processing problems. Future work will involve solving specific information processing problems using alternative assumptions and using data to evaluate predictions yielded by alternative information processing rules, the usual procedure for evaluating alternative procedures and models. It is hoped that the results in this paper will be useful to those who wish to pursue further work on information processing rules.

References

Arrow, K. and Hurwicz, L. (1972), “An Optimality Criterion for Decision Making Under Ignorance,” in Carter, C. and Ford, J., eds., Uncertainty and Expectations in Economics, Oxford: Basil Blackwell & Mott, Ltd.

Bernardo, J. and Smith, A. (1994), Bayesian Theory, New York: Wiley.

Berry, D., Chaloner, K. and Geweke, J. (1996), Bayesian Analysis in Statistics and Econometrics: Essays in Honor of Arnold Zellner, New York: Wiley.

Cohen, M. and Jaffray, J. (1980), “Rational Behavior Under Complete Ignorance,” Econometrica, 48, 1281-1300.

Golan, A., Judge, G. and Miller, D. (1996), Maximum Entropy Econometrics: Robust Estimation with Limited Data, New York: Wiley.

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

Guszcza, J. (1992), Student note for Bayesian Econometrics doctoral course, Grad. School of Business and Dept. of Economics, U. of Chicago.

Hill, B. (1988), "Discussion," American Statistician, 42, No. 4, 281-282.

Hogarth, R. and Einhorn, H. (1992), "Order Effects in Belief Updating: The Belief-Adjustment Model," Cognitive Psychology, 24, 1-55.

Izan, H. (1980), "To Pool or Not to Pool? A Re-examination of Tobin's Demand Problem," J. of Econometrics, 13, August.

Jaynes, E. (1988), "Discussion," American Statistician, 42, No. 4, 280-281.

Jeffreys, H. (1998), Theory of Probability, reprinted 3rd ed., 1st ed., 1939, in Oxford Classic Texts, Oxford: Oxford U. Press.

La France, J. (1999), "Inferring the Nutrient Content of Food with Prior Information," American J. of Agricultural Economics, 81, August, 728-734.

Renyi, A. (1970), Foundations of Probability, San Francisco: Holden-Day.

Soofi, E. (1996), "Information Theory and Bayesian Statistics," in Berry,D.,Chaloner, K. and Geweke, J., eds. (1996), 179-189.

______ (1997), "Information Theoretic Regression Methods," Advances in Econometrics, Vol. 12, Applying Maximum Entropy to Econometric Problems,
Fomby T. and Hill, R., editors, Greenwich, CT: Jai Press, 25-83.

Tobias, J. (1999), Three Essays on Bayesian Inference in Econometrics with an Application to Estimating the Returns to Schooling Quality, Ph.D. thesis, Dept. of Economics, U. of Chicago.

______ and Zellner, A. (1999), "Further Results on the Bayesian Method of Moments Analysis of the Multiple Regression Model," to appear in International Economic Review.

Zellner, A. (1988) "Optimal Information Processing and Bayes' Theorem," American Statistician, 42, No. 4, 278-284 (with invited discussion).

________ (1991), "Bayesian Methods and Entropy in Economics and Econometrics," in Grandy, W. and Schick, L. (eds.), Maximum Entropy and Bayesian Methods, Netherlands: Kluwer Publ., 17-31.

________ (1994), “Bayesian Method of Moments (BMOM) Analysis of Mean and Regression Models,” in Lee, J., Johnson, W. and Zellner, A. (eds.) Prediction and Modelling Honoring Seymour Geisser, New York: Springer Verlag, 61-74 and in Zellner (1997), 291-304.

________ (1996), An Introduction to Bayesian Inference in Econometrics, Wiley Classics Library, reprint of 1971 ed., New York: Wiley.

________ (1997), Bayesian Analysis in Econometrics and Statistics: The Zellner View and Papers,” in Economists of the 20th Century Series, Cheltenham, UK: Edgar Elgar Publ.

________ (1997a), “The Bayesian Method of Moments (BMOM): Theory and Applications,” Advances in Econometrics, Vol. 12, Applying Maximum Entropy to Econometric Problems, Fomby, T. and Hill, R., eds., Greenwich, CT: Jai Press, 85-105.

Zellner, A. and Moulton, B. (1985), “Bayesian Regression Diagnostics with Applications to International Consumption and Income Data,” J. of Econometrics, 53. 187-211.

_____ and Sankar, U. (1967), “ On Errors in the Variables,” ms, Social Systems Research Institute, U. of Wisconsin, Madison, 45pp.

_____, Tobias, J. and Ryu, H. (1999), “Bayesian Method of Moments Analysis of Times Series Models with an Application to Forecasting Turning Points in Output Growth Rates,” to appear in Estadistica with invited discussion.