



AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search

<http://ageconsearch.umn.edu>

aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

378.794
G43455
WP-865

Working Paper Series

WORKING PAPER NO. 865

KEEP IT SOPHISTICATEDLY SIMPLE

by

Arnold Zellner

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

DEPARTMENT OF AGRICULTURE AND
RESOURCE ECONOMICS AND POLICY

BERKELEY

CALIFORNIA AGRICULTURAL EXPERIMENT STATION

University of California

378,794
G434SS
WP-865

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

WORKING PAPER NO. 865

KEEP IT SOPHISTICATEDLY SIMPLE

by

Arnold Zellner

California Agricultural Experiment Station
Giannini Foundation of Agricultural Economics
January 1999

Keep It Sophisticatedly Simple

by

Arnold Zellner*
U. of Chicago

1. Introduction

Some years ago, I came upon the phrase used in industry, "Keep It Simple Stupid," that is KISS and thought about it in relation to scientific model-building. Since some simple models are stupid, I decided to reinterpret KISS to mean "Keep It Sophisticatedly Simple." In any event, KISS is very popular in many scientific and non-scientific areas. For example, the slogan of the Honda Motor Company is, "We make it simple." The Dutch Schipol airport in its advertising claims that, "It excels because it is simple and convenient." And it is well known that Einstein advised in connection with theorizing in the natural sciences, "Make it as simple as possible but no simpler." Also, the famous physicist Jaynes (1985, p.344) wrote, "We keep our model as simple as possible so as not to obscure the point to be made and also to heed Arnold Zellner's wise advice about 'sophisticatedly simple' models."

Many, including myself, have for long advocated that workers in econometrics and statistics follow the advice of natural scientists and others to keep analyses and models sophisticatedly simple. In addition, I have pointed out that there are many important, sophisticatedly simple models and methods that work well in practice, that is in explanation

and prediction, namely $s = \frac{1}{2} gt^2$, $E = mc^2$, $PV = RT$, maxent, etc. in the physical sciences and the laws of demand and supply, the Fisher equation, no arbitrage conditions, Marshall's competitive industry model, Friedman's and Becker's consumer models, the method of least squares, maximum likelihood techniques, Bayesian analysis, etc. in economics, econometrics and statistics. See Zellner (1997) for further discussion of these topics.

Further, for many years, I have challenged many audiences to give me an example of a large, complicated model in any field that works well in explaining past data and experience and in predicting new data. As yet, I have not heard of a single one.

Certainly, the many large-scale, complicated macroeconomic models of national economies, involving hundreds of nonlinear stochastic difference equations have not been very successful in explanation and prediction. See e.g., Christ (1951), Friedman (1951), Cooper (1972), Nelson (1972), Nelson and Plosser (1982), Meese and Rogoff (1983), McNees (1986), Garcia-Ferrer, et. al (1987) and Smyth (1983) for evidence on the forecasting performance of a sample of complicated macroeconomic models. In general, these studies found their forecasting performance to be unsatisfactory and in a number of instances no better or worse than that of random walk models and other simple, univariate time series models. As many have noted, if a large scale model using many variables and data on them along with much background subject matter information and theory can not perform better in prediction than a simple random walk model, the large, complicated model is probably defective. Rather than complicated, large Stanley steamers that often break down and sometimes explode, we need dependable Model T or A Fords which can be improved with additional work.

Further, Adelman and Adelman (1959), Hickman (1972), Zellner and Peck (1973) and others have reported the results of simulation experiments with several, leading complicated macroeconometric models that cast doubt on their value. Indeed, with some models that have hundreds of nonlinear stochastic difference equations, it is very difficult to show that they have a unique solution and their dynamic properties are hard to establish. If questions about these basic issues can not be answered, then, in my opinion, these models should be labelled, "UNSAFE FOR USE." Recently, after much unsatisfactory performance, U.S. Federal Reserve officials decided to scrap the large scale, complicated Federal Reserve-MIT-PENN model and commence construction of a new model. Similarly, the Federal Reserve Bank of Minneapolis is revising its complicated vector autoregressive (VAR, i.e., Very Awful Regression) models after their poor performance in forecasting turning points, etc. Also, on a visit to the Federal Reserve Bank of New Zealand a few years ago, I was told by researchers there that they have stopped using big, complicated models and are developing simple aggregate demand and supply models in attempts to understand their economy and to make good forecasts.

When James Tobin, a Nobel Prize winner in economics, was on the U.S. Council of Economic Advisors some years ago, I asked him if he used complicated macroeconometric models in his work for the Council. He responded that he didn't because he and others could not understand the workings and output of such models and thus did not have much confidence in them. When I asked what he actually did use to analyze problems, he remarked that he used a simple multiplier-accelerator model on the back of an envelope and even though he didn't have too much confidence in it, he said at least he understood what he was doing. In addition, in personal correspondence with Jan Tinbergen I learned that he too

favored starting analyses with sophisticatedly simple models. The same can be said for other Nobel Prize winners, Gary Becker, Milton Friedman, Robert Lucas, Merton Miller, Robert Solow, George Stigler, James Tobin and many others. That this is the case with respect to these leading researchers and many others contrasts markedly with the statement in Keuzenkamp and McAleer (1995, p.17), "In econometrics, Zellner is a rare exception in support of the simplicity postulate; Zellner...opposes a 'top-down' (general to specific) approach." On the contrary, there are many in econometrics and economics who oppose a 'top down' approach and like to KISS.

In addition to many other scientists, Sir Harold Jeffreys advocated the use of simple models in his books, *Theory of Probability* (1939, 1948, 1961, 1967, 1988) and *Scientific Inference* (1931, 1957). With respect to the former book, the eminent statistician I. J. Good (1980, p. 32) has written, "In summary, Jeffreys's pioneering work . . . has had a large permanent influence on statistical logic and techniques. In my review (Good, 1962b), I said that Jeffreys's book on probability 'is of greater importance for the philosophy of science and obviously of greater practical importance, than nearly all the books on probability written by professional philosophers lumped together.'"

With this in the way of an introduction, I shall now provide an overview of the remaining sections of this paper. In Section 2, a brief review of the thoughts of Sir Harold Jeffreys, formerly of Cambridge University, on simplicity and its role in science will be provided.¹ Then his numerical measure of the complexity of differential equation systems will be reviewed and discussed. In Section 3, Jeffreys's measure will be applied to several

central time series and other models that are often used in econometrics and statistics. Some specific procedures for extending Jeffreys's measure of complexity to apply to difference equation systems, distributions of random structural and measurement errors will be provided employing engineers's state space modeling concepts. The discussion and methods of Sections 2 and 3 are considered in relation to some structural econometric and time series modelling procedures that have been employed in practice in Section 4. Examples will be provided to illustrate general points. Last, in Section 5 some conclusions and topics that deserve further work are presented. }

II. Jeffreys's Views on Simplicity and His Measure of Model Complexity

In Section 1 a number of prominent, important simple models and methods from several sciences were mentioned. They all have the property that they are widely used in analyzing many problems and yield useful results. I believe that such impressive behavior of relatively simple models has been appreciated by many and thus the widespread belief in the efficacy of sophisticatedly simple models. On this issue, Jeffreys (1961, pp. 4-5) has written,

"It is asserted, for instance, that choice of the simplest law [or model] is purely a matter of economy of description or thought, and has nothing to do with any reason for believing the law [or model] . . . I say on the contrary, the simplest law is chosen because it is the most likely to give correct predictions; that the choice is based on a

¹ In Zellner (1980) a volume dedicated to Jeffreys, leading statisticians and econometricians summarize his contributions to the philosophy of science, statistical science and Bayesian analysis.

reasonable degree of belief; and the fact that deductive logic provides no explanation of the choice of the simplest law [or model] is an absolute proof that deductive logic is grossly inadequate to cover scientific and practical requirements.”

Further, he points out that “. . . the tendency to claim that scientific method can be reduced in some way to deductive logic is the most fundamental fallacy of all: it can be done only by rejecting its chief feature, induction.” (p. 2) Jeffreys defines induction as generalization from past data and experience to explain past experience and data and predict future experience and data. In his book, *Theory of Probability*, he develops and successfully applies a system of inductive logic that serves the needs of scientists and applied workers in all fields.

In Jeffreys’s discussion of models, he explicitly points out that including too many terms in a relation can improve fit but “the conclusion is that including too many terms will lose accuracy in prediction instead of gaining it.” (1961, p. 46) He goes on to explain,

“All we have to say is that the simpler laws [or models] have the greater prior probabilities. This is what Wrinch and I called the *simplicity postulate*.” (p. 47)

He remarks that he requires that laws or models be put in an order of decreasing prior probability that represents a degree of confidence in a particular law or model. As he points out,

“To make the order definite, however, requires a numerical rule for assessing the complexity of a law. In the case of laws expressible by differential equations this is easy. We could define the complexity of a differential equation, cleared of roots and fractions, by the sum of the order, the degree, and the absolute values of the coefficients.” (1961, p. 47)

He also mentions that, “All the laws of classical physics are in fact expressible by differential equations, and those of quantum physics are derived from them by various systematic modifications.” (1961, p. 47) It is also the case that that many economic and econometric models are in differential or difference equation form.

Thus in terms of explaining a particular phenomenon, there are, in Jeffreys’s view, an infinity of possible models. He indicates how to produce an ordering of models with respect to degree of complexity. Then he suggests assigning probabilities to each with simpler models given higher probabilities in a convergent sequence that is assumed to sum to one. The probabilities so assigned to models can be employed to form prior odds and used to compute posterior odds in evaluating alternative models with data. As many have noted, such model evaluation procedures incorporate “penalties for complexity,” see, e.g. Zellner and Min (1993, p. 396).

With regard to other uses of Jeffreys’s complexity measure, if simpler models are given higher prior probabilities of being adequate, then in the choice between a simple model, say M1 and a complicated model, say M2, there is a preference for M1, a priori. Also, if the construction of M1 is less costly than that of M2, the preference for M1 is probably strengthened. Further, if M1 is chosen and found to be inadequate, many believe that it is

easier to determine the causes of the inadequacy and to remedy them in the case of a simple model than in the case of a complicated model. These are considerations that seem important in the choice between M1 and M2 and Jeffreys's prior probabilities associated with simple and complicated models play an important role in this and other model choice problems.

To illustrate his measure of complexity, Jeffreys (1961, pp. 48-49) considers "... laws of the following form, $y = ax^n$ where n is restricted to be a particular integer. Cleared of the parameter a this gives the differential equation $x \frac{dy}{dx} = ny$ the complexity of which is $n + 4$." The parameter a is cleared from the equation by differentiating $\log y = \log a + n \log x$ with respect to x which yields the equation presented above, a first order differential equation with the sum of the absolute values of its coefficients equal to $n + 1$ and its degree equal to 2 and thus of complexity $n + 4$. Note that the order of a differential equation is the order of the highest derivative appearing in it; see, e.g., Bakker (1966, p. 101) or texts on differential equations. As regards the degree of the equation, here Jeffreys considers it to be 2, the degree of the product $x \frac{dy}{dx}$, a convention that is widely used even though the mathematical definition of the degree of a differential equation is given in Bakker (1966, p. 94) as "The degree of the highest order derivative." According to this definition, the above equation is of degree 1 and it would be equal to 1 for any value of q in the following differential equation,

$x^q \frac{dy}{dx} = ny$ whereas Jeffreys's value of the degree of this equation is $q + 1$, a measure that reflects the added complexity vis a vis, e.g. a value of $q = 0$. Given these considerations,

we shall employ Jeffreys's convention in appraising the degree of differential and difference equations in what follows.

Jeffreys also considers the relation $y = ax^n$ in the case in which the value of n is "wholly arbitrary," that is not necessarily equal to an integer. From $\log y = \log a + n \log x$, $x d^2 y / dx^2 = (n/x) dy / dx$ and when $n = (x/y) dy / dx$ is substituted in this last relation, we obtain Jeffreys's equation, $xy d^2 y / dx^2 + y dy / dx = x dy / dx$ which is of order 2, degree 3 and has sum of the absolute values of the coefficients equal to 3 and thus its complexity measure is 8. Note that the terms $x(dy/dx)^2$ and $xy d^2 y / dx^2$ are each of degree 3, the degree of the equation.

Jeffreys also illustrates his complexity measure as follows. For the law, $s = a$, where s is distance and a is a constant, he states that "it would be written as $ds / dt = 0$ [with t denoting time], with complexity $1 + 1 + 1 = 3$. $s = a + ut + \frac{1}{2} gt^2$ would become $d^2 s / dt^2 = 0$ with complexity $2 + 1 + 1 = 4$ and so on." (p. 47) Note that in this last case, he employs the homogeneous, second order differential equation for s . Solution of this homogeneous equation yields the complementary function $a + ut$, which when added to a particular solution of the nonhomogeneous equation, e.g. $\frac{1}{2} gt^2$, provides a general solution. Further for the nonhomogeneous equation, $d^2 s / dt^2 = g$ or $d^2 y / dt^2 = 1$, where $y = s/g$, the complexity is $2 + 1 + 2 = 5$, larger than for the homogeneous equation considered above. Below, Jeffreys's complexity measure will be discussed further and used in conjunction with laws or models expressed in discrete time by use of difference equations with additive stochastic errors.

It should be emphasized that Jeffreys regards his rule as a first rough approximation to measuring complexity. If we write his rule as follows:

$$\begin{aligned} C &= \text{Order} + \text{Degree} + \text{Sum of absolute values of} && (1) \\ & && \text{normalized coefficients} \\ &= O + D + S \end{aligned}$$

where “normalized coefficients” means that “differential equations must be cleared of factors common to all terms” (p. 48) since otherwise “multiplying the whole differential equation by an integer would apparently increase the complexity, though the result is exactly equivalent to the original equation.” (p. 48) Very importantly, Jeffreys does NOT just count the number of parameters in a model. Rather he adds the absolute values of the parameters which of course can be expressed as their number, N , times their average absolute value. Thus a parameter that has a large absolute value contributes more to Jeffreys’ measure of complexity, C , than one that has a small absolute value. To repeat, just counting the number of parameters to represent complexity is not in accord with Jeffreys’s approach.

As regards the common practice of “counting parameters” to measure complexity, Keuzenkamp and McAleer (1997, p. 554) are very critical of this practice and write, “Much of the literature in the philosophy of science and in econometrics deals with a popular definition of simplicity based on counting the number of parameters (or variables) or a model....For example, Popper...identifies simplicity with the paucity of parameters.... In practice, Rissanen’s Minimum Description Length criterion is similar to defining a measure of simplicity by counting parameters. The paucity of parameters measure of simplicity does

not seem satisfactory, in general, as several simple examples will demonstrate.” We shall consider some of these simple examples below and show that Jeffreys’s measure is useful in distinguishing between or among models that have the same number of parameters with respect to their relative complexity. As indicated above, just counting the parameters does not take account of their magnitudes nor of other properties of the models in which they are embedded, e.g. order and degree of differential or difference equations. However, as pointed out above, the number of parameters appearing in a model is one ingredient of Jeffreys’s measure of complexity.

To illustrate Jeffreys’s rule further, consider a model that is often used to represent the growth of human and other populations, the differential equation with solution the logistic growth curve, namely,

$$dN/dt = r N (1 - N/K) \tag{2a}$$

or

$$d(N/K)/dt = r (N/K)(1 - N/K) \tag{2b}$$

where $N = N(t)$, population at time t , and r and K are parameters, the latter the limiting, equilibrium population. (2b) is a first order differential equation of degree two with the sum of the parameters’s absolute values $1 + 2r$. Thus Jeffreys’s complexity measure is

$1 + 2 + 1 + 2r = 4 + 2r$. A discrete time version of (2b) is

$$N(t)/K - N(t-1)/K = r [N(t-1)/K][1 - N(t-1)/K] \tag{3}$$

which is a first order difference equation of second degree and thus the Jeffreys's measure of complexity is $4 + 2r$. If the discrete time model that best approximates the continuous time model in (2) were to be considered, it would be more complicated than the model given in (3). Open problems are whether the differential equation is more complicated than the difference equation in this and other cases and how to measure the complexity of a mixed difference-differential equation model, say the model in (2) with a finite gestation lag introduced that can yield a solution with damped oscillatory solutions whereas (2) does not have oscillatory solutions.

Another simple example is the Harrod-Domar growth model for total real income, Y , namely $S = sY$, the savings equation and $I = v dY/dt$, the investment equation and $I = S$, the equilibrium condition. Then we obtain $v dY/dt = sY$ or $dz/dt = 1$, with $z = (v/s)\log Y$. Thus we have the complexity $C = 1 + 1 + 2 = 4$. If the savings function is elaborated to permit the parameter s to depend on Y - see Zellner and Moulton (1985) for some empirical evidence on this point, so that $S/Y \rightarrow 0$ as $Y \rightarrow 0$, and $S/Y \rightarrow 1$ as $Y \rightarrow \infty$, let us see how this alters the complexity of this model. Let $s = Y^\lambda / (Y^\lambda + A)$ with $A, \lambda > 0$, then $v(Y^\lambda + A)dY/dt = Y^{\lambda+1}$, or $r = d \log Y / dt = Y^\lambda / v(A + Y^\lambda)$ is the equilibrium condition, a nonlinear first order differential equation that is more complicated than that above.

Of course Jeffreys's rule in (1) is in a simple linear form with all terms receiving the same weight. Many other forms of (1) can be contemplated but true to his philosophy, Jeffreys starts with the simplest form and points to the need to investigate the extent to which it performs satisfactorily and would modify it if necessary to produce improved performance in measuring complexity. For example, one could consider various monotonic functions of C , say $f(C)$ as a measure of complexity rather than C itself. Or one might

contemplate that C , the measure of complexity is related to order (O), degree (D) and sum of normalized coefficients, S , by $C = h(O,D,S)$ with h being a homogenous function of a given degree or perhaps a nonhomogenous function. See Zellner and Ryu (1998) for some differential equations defining production functions's forms that can be ordered with respect to complexity that may be helpful in characterizing complexity functions. As with the forms of utility functions, loss functions, production functions, etc., choice of an appropriate functional form is an important issue that is usually approached by starting with simple forms and seeing how well they work in applications and modifying them if necessary. In this connection, see Tinbergen's discussion of the forms of social welfare functions for use in policy-making, Arrow's, and Solow's work on the CES production function, Friedman's, Becker's, Tobin's and Modigliani's consumer models and Lucas's rational expectations model to appreciate how these leading researchers in economics developed sophisticatedly simple models, much in the spirit of the approach described by Jeffreys. Namely, start simply and complicate the model only if necessary. Indeed, learning how and why a simple model performs poorly is important information in attempts to improve it.

Last, as is obvious, sophisticated simplicity is a relative term. It has to be considered in connection with our current state of knowledge. A random walk model may be a sophisticatedly simple model when little knowledge is available about the subject under consideration. However, in areas in which there is much subject matter information, a random walk model is probably a stupid model. And other simple models that are known to be at variance with our knowledge, are also stupid. A sophisticatedly simple model will not be in conflict with what is known and provides more in the way of understanding than currently available models. In this connection, Jeffreys's (1967, p. 50) comments on the

Schrödinger partial differential wave equation in physics are illuminating, "One reviewer of Scientific Inference argued that on my theory Schrödinger's equation would have so high a complexity, and therefore so low an initial probability, that it could never have acquired a high probability. This argument overlooks the greater part of the book, which shows how by consideration of different types of data in turn and corresponding modification of the laws Schrödinger's equation is actually reached. Besides those mentioned as stages in the approximation hosts of others have been excluded as having negligible posterior probabilities, and it may well be true that Schrödinger's equation is the simplest of the survivors."

With this said about Jeffreys's views and his measure of complexity, the issue of how his measure works in connection with widely used time series models in econometrics and statistics will now be considered.

III. Measures of Complexity for Time Series Models

Herein, we apply Jeffreys's measure of complexity to some time series models encountered in econometrics and statistics. Then we shall consider measures of complexity for the distributions of error terms and relate these measures to our past work on producing forecasting and structural models for 18 countries' output growth rates; see Garcia-Ferrer, et al (1986), Zellner and Hong (1989, 1991), Hong (1989). Zellner, Hong and Gulati (1991). Zellner, Hong and Min (1991), Min (1992) and the review paper summarizing work in these and other papers, Zellner (1994).

Consider the following deterministic autoregressive models of order 1 and 2, denoted by AR(1) and AR(2), namely, $y(t) = a y(t-1)$ and $y(t) = a y(t-1) + b y(t-2)$. Jeffreys's measure

of complexity for the AR(1) and AR(2) are $O = 1$, $D = 1$ and sum of the absolute values of the coefficients, $1 + |a|$, that is $3 + |a|$ and $4 + |a| + |b|$, respectively and thus the AR(2) is more complex than the AR(1) given that the coefficient of $y(t-1)$ is the same in the two models. If it is not, then it is possible that the AR(2) model can have a complexity measure that is smaller in value than that for the AR(1) model. An example is $y(t) = b y(t-2)$ which may be less complex than the above AR(1) model given that a is much larger than b in absolute value. Further, this last AR(2) containing just one parameter b is simpler than an AR(2) model containing two parameters, a and b . Also in terms of an AR(1) model, a model with the absolute value of a larger than one, namely an explosive model is more complicated than an AR(1) model with the absolute value of a less than one, that is a non-explosive process. Similar considerations apply to AR(2) models in that explosive models will tend to be more complicated than non-explosive models. See Zellner (1996, p. 196) for a plot of regions of the parameter space associated with explosive and non-explosive oscillatory and non-oscillatory solutions of an AR(2), process. And if we consider an m 'th order AR process, it will usually be more complex in terms of Jeffreys's measure than lower order processes. In summary, it is clearly the case that Jeffreys's complexity measure can be applied to deterministic AR processes without difficulty and appears to yield sensible results.

We now add an error term to our AR process, e.g., $y(t) = a y(t-1) + u(t)$. Then to evaluate Jeffreys's complexity measure, we can consider the homogenous part of the AR process, namely $y(t) = a y(t-1)$ and proceed as in the previous paragraph to evaluate his measure of complexity. Note however that if $u(t)$ is not white noise, e.g., it might be generated by an invertible AR(1) process, $u(t) = c u(t-1) + e(t)$ or $u(t) = e(t)/[1 - cL]$, where

L is the lag operator such that $Lu(t) = u(t-1)$, we have, $[1 - aL][1 - cL]y(t) = e(t)$. It is seen that the homogenous part of this last equation is a second order AR with complexity $3 + |a + c| + |ac|$. This AR process with an AR error term has been called an ARAR model in Carter and Zellner (1996). On the other hand, the error term $u(t)$ might be generated by an invertible moving average process of order one, say $u(t) = e(t) - q e(t-1) = [1 - qL] e(t)$, where $e(t)$ is white noise and it is assumed that $1 - qL$ is invertible, that is $-1 < q < 1$. Then from $[1 - aL]y(t) = [1 - qL] e(t)$, $\{[1 - aL]/[1 - qL]\}y(t) = e(t)$. The lefthand side of this last equation, the homogeneous difference equation, is an INFINITE order AR process, as is well known, and thus extremely complicated according to Jeffreys's complexity measure, much more complicated than the ARAR model considered above that has a finite second order homogeneous difference equation even though the two equations have the same number of parameters, as pointed out by Keuzenkamp and McAleer (1997). Researchers have recognized complications of working with moving average error terms, e.g. multimodal likelihood functions, "piling up" phenomena, etc. In Carter and Zellner (1996) many comparisons of ARAR and ARMA models are made in univariate and multivariate cases and the general result is that ARAR models are simpler to interpret and implement. Model selection techniques and data are employed to compare ARAR and ARMA models to determine which is better supported by the information in the data, see, e.g., Zellner and Geisel (1970).

Note that in the above stationary ARAR model, $(1 - aL)(1 - cL)y(t) = e(t)$ the sum of the absolute values of the parameters is $1 + |a + c| + |ac|$, where $-1 < a, c < 1$. In an unrestricted second order AR(2), the sum of the absolute values of the unrestricted parameters can

obviously be larger than that for the above ARAR model which is then less complex and has higher prior probability than an unrestricted AR(2) process.

Jeffreys's measures of complexity can be applied to higher order AR, ARAR and ARMA models. It was mentioned that the AR representation of invertible ARMA models is an INFINITE autoregression whereas that of the ARAR model is a FINITE autoregression. Thus Jeffreys's complexity measure has a greater value for the ARMA model than for the ARAR model since the order of the former is infinite, unless truncated in which case the order will generally be very high. For a particular transfer function, $a(L)y(t) = b(L)x(t)$, an analysis similar to that above indicates that those with MA(q) error terms are in general more complicated than those with AR(q) error terms. And both of these models will be simpler than the above transfer function with ARMA (p, q) error terms.

As regards the densities for the error terms introduced above, it is well known that many probability density functions can be produced as solutions of differential equations. For example the solution to $d \log f / dx = -1$ is the exponential density, $f(x) = \exp(-x)$, to $d \log f / dx = -x$, the normal density, $f(x) = [1 / \sqrt{2\pi}] \exp\{-x^2 / 2\}$, to $d \log f(x) / dx = -(1 + ax + bx^2 + cx^3)$, the exponential quartic density. Further, the Pearson system of densities can be obtained as solutions to the following differential equation; see, e.g. Jeffreys (1967, p.74),

$$d \log f / dx = -(x - a) / (b_0 + b_1x + b_2x^2) \quad (4)$$

Clearly, Jeffreys's measure of complexity can be evaluated for these differential equations and leads to the conclusion that for the uniform density, $d \log f / dx = 0$ is the simplest, with C

= 3, while C = 4 and 5 for the exponential and normal densities. Also, it is the case that maxent densities that arise from maximizing entropy, $-\int f(x)\log f(x)dx$, here defined relative to uniform measure, with respect to choice of $f(x)$ subject to moment side conditions,

$$\int x^i f(x)dx = \mu_i, i = 0,1,\dots,m, \mu \text{ 's having known values, have the form}$$

$f(x) = \exp\{-(\lambda_0 + \lambda_1 x + \lambda_2 x^2 + \dots + \lambda_m x^m)\}$ where the λ 's are Lagrange multipliers. On logging this last expression and differentiating by x , an explicit differential equation involving $d\log f/dx$ and powers of x is obtained and its complexity can be measured using Jeffreys's complexity measure. Note that the maxent solution may be a density for the subject matter part of a model or for the error term.

Thus adding error terms to models involves increasing their complexity. If the differential equation for the error term density has complexity C' , this can be added to the complexity for the model's differential equation, C , to obtain an overall measure $C + C'$. As noted above, many other possible ways of combining C and C' can be considered.

While many model builders just add random error terms to their models, state space engineers, statisticians and econometricians add measurement equations, deterministic or stochastic, to their state equation models. The measurement equations link the variables of the model, the so-called state variables to the measurements or measured variables. To illustrate, consider Friedman's (1957) model for the state variables, permanent consumption C_p and permanent income Y_p , $C_p = kY_p$, with k assumed constant. This is a state equation with no error term. One could write $C_p = kY_p + u^*$, where u^* is a state equation error term. The complexity of this relation and of its error term density can be evaluated as shown

above. But this model is not sophisticatedly simple since C_p and Y_p are not observable.

Friedman

presented the following measurement equations to make his model operational, $C = C_p + u$ and $Y = Y_p + v$, where C and Y are measured real consumption and real income and u and v are the measurement equation error terms that Friedman calls temporary consumption and temporary income, respectively. Additional assumptions about the properties of u and v have to be introduced to make the model operational, say u and v are independent normal variables with zero means and finite variances. These assumptions have to be adequate to identify the key parameter k and other parameters of the model. If they are not, the model is not sophisticatedly simple. Note that in this "errors in the variables" model with no state space error term, having the intercept parameter in the state equation be equal to zero, a key economic property of the Friedman model, is adequate to identify the model's parameters; see, e.g. Zellner (1996, p. 128ff) for discussion of this identification problem.

The assumed measurement equations of the Friedman model are relatively simple and have been made more general in applications by introduction of proxy or instrumental variables to represent Y_p , or by introduction of adaptive or rational expectations models as well as seasonal time series models to provide time series proxies for Y_p . If the measurements come from survey data, then the measurement equations may have to be elaborated to take account of systematic reporting errors, the design of the survey, and other features of the process generating the data, e.g. non-random attrition, missing observations, etc., etc. It is clear that if there are no good simple measurement models relating C_p and Y_p to the measurements, the overall model would not be operational and hence only a theoretical model, perhaps empirically useful in the future. One strength of the Friedman

model is that it is readily implementable with available data and makes predictions, many of them verified with data..

One other consideration that must be taken into account is the constancy of the parameters of models. In Friedman's model the parameter k is sometimes assumed constant, particularly in cointegration analyses. However, Friedman's theory indicates that k is a function of the interest rate, ratio of nonhuman wealth to total wealth, family size, tastes and preferences, etc. Perhaps the parameter k above is a time varying parameter that should be modeled, as in state space models with time varying parameters. If a time series model for k is added to the model, that indeed would add to the complexity of the model but in certain circumstances might improve its performance. For example, in explaining Kuznets's finding of a relatively constant savings rate for the U.S. since the turn of the century until after World War II, Friedman did not just point to the constancy of the parameter k . He pointed out that two main offsetting influences kept S/Y relatively, constant, namely, the fall in average family size and the decline in the proportion of farmers and other entrepreneurs in the labor force; see Friedman (1957). In addition, aggregation effects, Lucas effects, changes in tastes and technology and other factors can cause parameters' s values to change through time.

At this point, we recognize that in modeling in the sciences, there are at least six components, (1) state equations, (2) state equations's parameters, (3) state equations's errors, (4) measurement equations, (5) measurement equations's parameters and (6) measurement equations's errors. As indicated above, Jeffreys's measure of complexity is operational in connection with (1)-(6) insofar as components are in differential or difference equation form. Adding these measures of complexity will provide a first approximation to

the overall complexity. If a weighted average or some other combination of the above components of complexity can be shown superior to the simple sum, then of course it should be employed.

We now turn to review some work in modeling to show the roles played by the above concepts in actual analyses of macroeconomic data relating to eighteen industrialized countries's economies.

IV. Structural and Time Series Modeling Procedures and Results

In Section I, some remarks were made about certain macroeconomic modeling procedures. Some years ago, Franz Palm and I (Zellner and Palm, 1974) investigated the relation between multivariate time series(TS) models and dynamic structural econometric models (SEM) in an effort to produce improvements in modeling strategy. We began by considering Quenouille's (1957) multivariate ARMA model,

$$H(L)z(t) = F(L)e(t) \quad (5)$$

where $H(L)$ and $F(L)$ are invertible matrix polynomial lag operators, $z(t)$ is an $m \times 1$ vector of variables, e.g. output, prices, consumption, etc. In period t , and $e(t)$ is an $m \times 1$ vector of zero mean white noise errors, $Ee(t) = 0$ and $Ee(t)e(t)' = I_m$ for all t . If $F(L) = F_0$, an $m \times m$ non-singular matrix, the process is a VAR. Also, Quenouille's system can incorporate cointegration relations along with other relations. Now Palm and I asked, if we start with this complicated MVARMA or VAR process, what are its implications and are they sophisticatedly simple? One easy operation is to solve the m -equation system to determine the implied processes on individual variables. By multiplying both sides by the inverse of

$H(L) = \text{adjoint matrix of } H(L) / \text{determinant of } H(L) = H^*/H_d$, we have $z(t) = (H^*/H_d)F(L)e(t)$ or $H_d z(t) = H^*F(L)e(t)$. For an individual component of $z(t)$, say $z_i(t)$, the implied process is $H_d z_i(t) = b_i' e(t)$, where b_i is the i 'th row of $H^*F(L)$. For many values of m , say $m = 5$ or $m=10$, the determinant H_d will usually be an extremely high degree polynomial in L giving rise to an extremely high order AR on $z_i(t)$. Similarly, the elements of b_i' will be high degree polynomials in L giving rise to a high order MA process for the error term in this "final equation.": Thus the unrestricted general linear MVARMA model or VAR model implies very, very complicated, according to Jeffreys's measure, ARMA models for individual variables. Since such highly complicated models have not generally been identified using annual data, the complicated implications of these general models do not square with the data.

To simplify somewhat the complicated MVARMA model, econometricians generally assume that some of the variables in the $m \times 1$ vector $z(t)$ are endogenous, denoted by $y(t)$ and the remainder are exogenous, denoted by $x(t)$, or $z(t)' = (y(t)', x(t)')$ and with a corresponding partitioning of $H(L)$ and $F(L)$, the system becomes:

$$H_{11}(L)y(t) + H_{12}(L)x(t) = F_{11}(L)e_1(t) \quad (6)$$

and

$$H_{22}(L)x(t) = F_{22}(L)e_2(t) \quad (7)$$

where (6) is the dynamic linear structural equation system and (7) is the MVARMA process for the exogenous variables, $x(t)$ implied by the overall ARMA process in (5). The

assumption that $x(t)$ is exogenous imposes the restrictions that a submatrix of $H(L)$, $H_{21}(L)$ and two submatrices of $F(L)$, $F_{12}(L)$ and $F_{21}(L)$, a simplifying assumption that reduces the number of unknown parameters considerably. However, without further assumptions, the parameters in the structural equation model in (6) are not identified. Also, when Palm and I solved for the marginal processes for individual variables, they were still very complicated ARMA processes. Also, (6) can be solved for the transfer equations associated with the dynamic SEM. These relate single endogenous variables to current and lagged values of the exogenous variables and have the general form, obtained by multiplying both sides of (6) by the inverse of $H_{11}(L)$ and solving for $y_i(t)$, the i 'th endogenous variable,

$$H_{11}(L)y_i(t) = b(L)x(t) + c(L)e_1(t) \quad (8)$$

where $b(L)$ and $c(L)$ are vectors the elements of which are polynomials in the lag operator L

See Zellner and Palm (1975) for the transfer functions associated with several variants of Friedman's monetary macroeconomic model, Hong (1986) for those associated with several variants of a Keynesian IS-LM model and Min (1992) for those associated with several generalized real business cycle models.

Since transfer functions represent the variation of a single variable that is related to past values of itself and just possibly current and lagged values of exogenous variables, they are somewhat simpler than reduced form equations and also equations of VAR models. Thus it is a possible, simple starting point for model construction. That is, empirically determine the forms of (8) for each endogenous variable and then try to determine the form

of the SEM system in (6) that algebraically yields the empirically determined transfer functions. See the references in the previous paragraph for some results.

One important endogenous variable is the rate of growth of an economy's total output as measured by real GDP. Since there may be biases in measuring real GDP, in our empirical work, Garcia-Ferrer, et al., (1987) and Zellner (1994), we logged measured real GDP and first differenced log GDP to get the rate of growth of GDP. The logging and differencing was thought to help remove certain types of systematic biases in measured real GDP, a

simple but important reason for doing it, and also because there is great interest in the rate of growth of real GDP, denoted by y_{it} for the i 'th country in the t 'th year. After finding that a simple AR(3) model for this variable did not work because it missed in forecasting turning points, we added several leading indicator variables to the model, namely the lagged rates of growth of real money and real stock prices, variables that Burns and Mitchell empirically discovered tended to lead in the business cycles for France, Germany, the UK and the U.S. using pre World War II data. In addition, we added the lagged annual median growth rate of real stock prices for the countries in our sample to the equation, a proxy for the world rate of return. This common effect variable took most of the contemporaneous correlation out of our country error terms which very importantly reduced the number of non-zero parameters in our 18x18 covariance matrix for the 18 countries in our sample by making it diagonal. The autoregressive model containing lagged leading indicator variables is called an ARLI model. In forecasting experiments with the ARLI model, it was found to be superior to various random walk models, AR(3) models, the famous Nelson-Plosser ARIMA model, and

to several OECD macroeconometric models for various countries. The root mean squared error of forecast for our out of sample forecasts were employed in these computations. Further, it was found that Stein shrinkage, that involves assumptions regarding the distribution of countries' coefficient vectors, namely that they are concentrated about some central value, to be estimated, improved forecasts considerably. Here by making the individual country parameters random and time-varying and integrating them out to express our models in terms of relatively few hyperparameters, we reduced the complexity of our models and obtained improved results. See also the paper by Putnam and Quintana (1995) who review forecasting in the financial economics area and comment favorably on some of our procedures that they have found quite useful.

Not only were the annual point forecasting results produced by our methods and models encouraging but also their turning point forecasts are surprisingly good, namely about 70 per cent or more of 158 turning points, 1974-86, correctly forecasted; see Zellner, Hong and Min (1991) for the new Bayesian decision theoretic turning point forecasting methods employed and results. In recent work, Zellner and Min (1998), the data were extended to include 192 turning points and the former favorable results were still encountered. In this work, the models's turning point forecasts were compared with various naive modelers' forecasts, e.g., a coin flipper, an eternal optimist, who always forecasts no downturn and upturn, an eternal pessimist, who always forecasts downturn and no upturn, and a deterministic four year cycle forecaster, who always forecasts downturn at the top and upturn at the bottom.

Thus these relatively simple variants of a basic ARLI model work reasonably well but as with almost all models still need improvement. Again a simple solution is being investigated, namely disaggregation by industrial sector, namely agriculture, construction, manufacturing, wholesale trade, retail trade, etc. As an engineer might put it, instead of viewing the economy as one big oscillator, we view the economy as made up of coupled sectoral oscillators. For each sector, sophisticatedly simple and useful Marshallian demand, supply and entry (DSE) models, have been formulated and will be implemented to forecast sectoral outputs. Then the sectoral output forecasts will be aggregated to obtain a forecast for aggregate output that can be compared with the forecasts obtained from various models for aggregate output. Since use of the sectoral data involves increasing the sample size and introducing specific, important sectoral variables and linkages it is expected on theoretical and intuitive grounds that improved forecasts will be obtained. See consideration of these issues in the discussion of Zellner (1994). All of this contrasts markedly with a strategy of starting with an "encompassing" or general MVARMA model or a VAR model and testing downward. As Keuzenkampf and McAleer (1995, p.17) rightly stated, I and many others do not favor such a strategy for the simple reason that in decades during which it has been used, it has not as yet produced macroeconomic models that perform well in explanation and prediction. In my opinion, it is much more fruitful to KISS and complicate only if necessary. Needless to say, this conclusion owes much to the wisdom and analysis of Harold Jeffreys and many other leading researchers mentioned earlier.

V. Summary and Conclusions

In this paper, I have attempted to summarize Jeffreys's views on simplicity and to review and extend his measure of complexity. On the latter topic, Jeffreys's measure of complexity has been extended to apply to difference equation models and differential equations defining probability density functions. Thus it becomes possible to measure the complexity of state equations, measurement equations and the density functions for their error terms. Bringing in the complexity of probability density functions and of measurement equations is important in terms of appraising the overall complexity of a model that purports to explain the past and predict the future.

Further, I have emphasized the virtues of sophisticatedly simple models and pointed out that simple models that contain logical errors, are at variance with known facts, etc. are simply stupid. A sophisticatedly simple model takes appropriate account of the techniques and knowledge present in a field and is logically sound. It is suggested that it is more worthwhile to develop sophisticatedly simple models than to develop complex models. When a sophisticatedly simple model does not work well, some added, well-chosen complexity may improve the situation. For example, disaggregating total output by sector and modeling coupled sectors is probably a better strategy for understanding and predicting total output than by modeling total output in an aggregate model. The former strategy represents a complication of the latter that will probably be justified in terms of superior performance in explanation and prediction. In such a case, added sophisticated complexity, not too much, may be justified.

There is much more work to be done on the appropriate ways to combine the elements of Jeffreys's measure of complexity and of the broadened measures presented in

this paper. More analysis of the sequential decision problem of model choice and subsequent improvement using appropriate prior probabilities and criterion functions would be welcome. And last, not much has been said about the important issue of how new, sophisticatedly simple models are conceived and developed, the objective of reductive inference that I have briefly discussed in Zellner (1996, p. 5ff.). How important, new sophisticatedly simple models that work well in explanation and prediction are produced is a topic that deserves much attention.

List of References

I. Adelman and F. L. Adelman, 'The Dynamic Properties of the Klein- Goldberger Model,' *Econometrica* 27 (1959), 569-625.

C.C.T. Baker, *Dictionary of Mathematics* (New York: Hart Publishing Co., 1966)

R.A.L. Carter and A. Zellner, 'The ARAR Model for Time Series,' *Proceedings of the American Statistical Association's Section on Bayesian Statistical Science* (1996), 226-231.

- C.F. Christ, 'A Test of an Econometric Model for the United States, 1921-1947' in Conference on Business Cycles (New York: National Bureau of Economic Research, Inc., 1951), 35-107.
- C.F. Christ, 'Judging the Performance of Econometric Models of the U.S. Economy,' *International Economic Review* 16 (1975) 54-74.
- R.L. Cooper, 'The Predictive Performance of Quarterly Econometric Models of the U.S. Economy,' in B.G. Hickman (ed.), *Econometric Models of Cyclical Behavior*, Vol. 2 (New York: Columbia U. Press, 1972), 813-926.
- Friedman, 'Comment on 'A Test of an Econometric Model for the United States, 1921-47' by Carl Christ,' in Conference on Business Cycles (New York: National Bureau of Economic Research, Inc., 1951), 107-114
- M. Friedman, *A Theory of the Consumption Function* (Princeton: Princeton U. Press, 1957)
- A. Garcia-Ferrer, R.A. Highfield, F.C. Palm and A. Zellner, 'Macroeconomic Forecasting Using Pooled International Data,' *Journal of Business and Economic Statistics* 5 (1987), 53-67.
- B.G. Hickman (ed.), *Econometric Models of Cyclical Behavior* (New York) Columbia U. Press, 1972)
- C. Hong, 'Forecasting Real Output Growth Rates and Cyclical Properties of Models: A Bayesian Approach,' Ph.D. thesis, Dept. of Economics, U. of Chicago (1989).

- E.T. Jaynes, 'Highly Informative Priors,' in J.M. Bernardo, et al. (eds.), *Bayesian Statistics 2* (Amsterdam: North-Holland Publ. Co., 1985), 329-352.
- H. Jeffreys, *Theory of Probability* (Oxford: Oxford U. Press, 1939, 1948, 1961, 1967, 1988).
- Jeffreys, *Scientific Inference* (Cambridge: Cambridge U. Press, 1931, 1957).
- H.A. Keuzenkamp and M. McAleer, 'Simplicity, Scientific Inference and Econometric Modelling,' *Economic Journal* 105 (1995), 1-21.
- H.A. Keuzenkamp and M. McAleer, 'The Complexity of Simplicity,' *Mathematics and Computers in Simulation*, 43 (1997), 553-561.
- R.B. Litterman, 'Forecasting with Bayesian Vector Autoregressions: Five Years of Experience,' *J. of Business and Economic Statistics*, 4 (1986), 25-38 (with discussion).
- S.K. McNees, 'Forecasting Accuracy of Alternative Techniques: A Comparison of U.S. Macroeconomic Forecasts,' *J. of Business and Economic Statistics*, 4 (1986), 5-23 (with discussion).
- R. Meese and K. Rogoff, 'The Out of Sample Failure of Empirical Exchange Rate Models: Sampling Error or Misspecification? In J. A. Frenkel (ed.) *Exchange Rates and International Economics*, (Chicago: U. of Chicago Press, 1983).
- Min, 'Economic Analysis and Forecasting of International Growth Rates Using Bayesian Techniques,' Ph.D. thesis, Dept. of Economics, U. of Chicago (1992).

C. Min, 'Bayesian and Non-Bayesian Methods for Combining Models and Forecasts with Applications to Forecasting International Growth Rates,' *J. of Econometrics*, 56 (1993), 89-118.

C.R. Nelson, 'The Predictive Performance of the FRB-MIT-PENN Model of the U.S. Economy,' *American Economic Review* 62 (1972), 902-917.

C.R. Nelson and C. I. Plosser, 'Trends and Random Walks in Macroeconomic Time Series,' *J. of Monetary Economics*, 10 (1982), 139-162.

B. Putnam and J.M. Quintana, 'The Evolution of Bayesian Forecasting Models,' in B. Putnam (ed.), *Applying Quantitative Discipline to Asset Allocation* (London: Euromoney Publications, 1995).

D.J. Smyth, 'Short-run Macroeconomic Forecasting: The OECD Performance,' *J. of Forecasting*, 2 (1983), 37-49.

A. Zellner (ed.), *Bayesian Analysis in Econometrics and Statistics: Essays in Honor of Sir Harold Jeffreys*, (Amsterdam: North-Holland, 1980).

A. Zellner, 'Time-series Analysis, Forecasting and Econometric Modelling: the Structural Econometric Modelling, Time-series Analysis (SEMTSA) Approach,' *J. of Forecasting*, 13(1994), 215-233.

A. Zellner, *An Introduction to Bayesian Inference in Econometrics* (New York: Wiley Classics Series, 1996, reprint of 1971 ed.)

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

A. Zellner and M.S. Geisel, 'Analysis of Distributed Lag Models with Applications to Consumption Function Estimation,' *Econometrica* 38 (1970), 865-888.

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

A. Zellner and C. Hong, 'Forecasting International Growth Rates Using Bayesian Shrinkage and Other Procedures,' *J. of Econometrics, Annals*, P. Schmidt (ed.) *Issues in Econometric Forecasting*, 40 (1989), 183-202.

A. Zellner and C. Hong, 'Bayesian Methods for Forecasting Turning Points in Economic Time Series: Sensitivity of Forecasts to Asymmetry of Loss Structures,' in K. Lahiri and G. Moore (ed.), *Leading Indicators: New Approaches and Forecasting Records* (Cambridge: Cambridge U. Press, 1991), 129-140.

Zellner and B.R. Moulton, 'Bayesian Regression Diagnostics with Applications to International Consumption and Income Data,' *J. of Econometrics*, (1985), 187-211.

Zellner and S. Peck, 'Simulation Experiments with a Quarterly Macroeconometric Model of the U.S. Economy,' in A.A. Powell and R.A. Williams (eds.), *Econometric Studies of Macro and Monetary Relations*, (Amsterdam: North-Holland, 1973), 149-168.

A. Zellner C. Hong and G. M. Gulati, 'Turning Points in Economic Time Series, Loss Structures and Bayesian Forecasting,' in S. Geisser, J. Hodges, S.J. Press and A. Zellner

(eds.), *Bayesian and Likelihood Methods in Statistics and Econometrics: Essays in Honor of George A. Barnard* (Amsterdam: North-Holland), 371-393.

Zellner, C. Hong and C. Min, 'Forecasting Turning Points in International Output Growth Rates Using Bayesian Exponentially Weighted Autoregression, Time-Varying Parameter, and Pooling Techniques,' *J. of Econometrics*, 49 (1991), 275-304.

Zellner and H. Ryu, 'Alternative Functional Forms for Production, Cost and Returns to Scale Functions,' *J. of Applied Econometrics*, 13 (1998), 101-127.

Zellner and C. Min, 'Forecasting Turning Points in Output Growth Rates: A Response to Milton Friedman,' manuscript (1998) to appear in *J. of Econometrics*.

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