Volume Title: Evaluation of Econometric Models

Volume Author/Editor: Jan Kmenta and James B. Ramsey, eds.

Volume Publisher: Academic Press

Volume ISBN: 978-0-12-416550-2

Volume URL: http://www.nber.org/books/kmen80-1

Publication Date: 1980

Chapter Title: Problems and Issues in Evaluating Econometric Models

Chapter Author: Jan Kmenta, James B. Ramsey

Chapter URL: http://www.nber.org/chapters/c11691

Chapter pages in book: (p. 1 - 11)

Problems and Issues in Evaluating Econometric Models

JAMES B. RAMSEY

DEPARTMENT OF ECONOMICS
NEW YORK UNIVERSITY
NEW YORK, NEW YORK

and

JAN KMENTA

DEPARTMENT OF ECONOMICS
UNIVERSITY OF MICHIGAN
ANN ARBOR, MICHIGAN

As in most scientific disciplines there is in economics a considerable gap between econometric theory and practice. While the actual practice of applied economists is only occasionally up to the "best available" procedures as determined by econometric theorists, it is also true that econometric theory frequently ignores the hard and often most relevant procedural problems faced by applied economists or theorists wishing to test their hypotheses.

The former statement is commonplace, while the latter may be more novel. With respect to the former, the applied economist needs to learn how more recent but, paradoxically, often simpler procedures can be implemented, and what are the potential benefits and costs of such implementation. Correspondingly, the econometric theorist needs to keep in mind the needs of the applied economist and the economic theorist. Both are interested in effective inference; the theorist wishes to test hypotheses about economic behavior, while the applied economist wishes to weigh the effects of alternative policies. However, difficulties occur for the users of econometric procedures in that frequently the econometric prescriptions are inappropriate for the situation in hand.

ISBN 0-12-416550-8

Inferences based on normal distributions are of little use where the underlying distributions are clearly not normal. Parametric inference is of little use where the economist's theoretical knowledge is limited to directions of change. Classical and even Bayesian inferential procedures are of little help where samples from a population are limited in size, unique, and nonrepeatable except at very high cost. Procedures for which the results are sensitive to extensive and detailed specifications of the maintained hypothesis are of little use where that knowledge does not exist.

These, then, are the main issues which provided the initial motivation for the Conference on Econometric Methodology held in Ann Arbor in June 1977 and for this book of conference proceedings. While there exists a vast potential for econometric research into specific topics generated by the inferential issues mentioned above, the editors and organizers of the conference decided to concentrate on certain major themes.

The first theme involves the degree of detail and precision with which a model is specified; at the extremes we might characterize the distinction as one between formal and informal models, although a more accurate distinction is between parametric and nonparametric specification. For example, compare the statements "the expected value of quantity demanded decreases with increases in price" (nonparametric) with "the expected value of quantity demanded is given by the function $\gamma_0 e^{-\gamma_1 p}$ where p is real price and γ_0 and γ_1 are unknown but estimable parameters." Recently, more attention has been paid to nonparametric (at least less parameter-specific) models in recognition of the fact that rejecting economic hypotheses couched in terms of a specific model may merely be a rejection of the modeler's detailed parametric specification and not the economic hypothesis of interest.

The second and related theme has to do with the robustness of a model's inferences to errors in the specification of the model. The claim is frequently made that since all models are approximations, then all models are in error. While we need not take such an absolute position, it is clearly farsighted for us to recognize that our models may be in error, at least to a small extent, so that we should seek inferential procedures which are not sensitive to the more likely errors. For example, we may suspect that the disturbances are not normally distributed, though they are symmetrically distributed about zero without very fat tails. A corollary notion is to consider in terms of regression analysis those sets of observations which have the greater relative impact on the inferences. This provides a springboard for a detailed analysis of regression results in looking for evidence of model specification errors.

Often, the sensitivity of a model to errors in its specification can be related to the level of aggregation being used. A more informative approach to this aspect of the problem is to consider the extent to which micro (or individual) behavioral coefficients can be inferred from estimates of coefficients in macro (or group) relationships.

Further, one must not forget the straight forward, but crucially important, procedure of checking one's model for specification errors before attempting to use the statistical results. Specification error analysis is the more important, the more parameter-specific (and hence usually the more sensitive to specification errors) the model is.

This leads to a third theme which involves formal methods for the comparison of models. Econometric models are now and have been for some time sufficiently complex in structure as to require the development of new methods and criteria for choice between alternative models of a specified economic situation.

A fourth theme which has only recently been the subject of intense examination is the appropriate role of time series analytical methods in econometric models. The earlier and overstated dichotomy was between sophisticated, but purely statistical, data analysis with no economic theory content on one side and theoretically specified models analyzed with little attention to the possibility of a complex time series structure in the stochastic elements on the other side. Very recently, attempts have been made to reconcile the conflicts between the time series approach and theoretical modeling.

A fifth and last theme concerns the potential benefits and costs of using experimental data to test economic hypotheses. From one perspective, the development of an experimental methodology and its use in economics can serve as a substitute for further attempts to refine, improve, and expand methods for extracting information from historical data. From another perspective, experimental data may enable us to examine behavioral relationships not directly observable from historical data no matter how ingenious the inferential methods.

The organizers of the conference envisaged two sets of people who would be interested in the outcome and to whom this book is addressed. First, it was hoped that the economic theorist and the applied economist would benefit in that they would be exposed to a discussion of the issues mentioned above and would perceive that the implementation of the suggested econometric procedures is not an insurmountable task as is unfortunately all too frequently the case. Further, and perhaps more importantly at this time, the users of econometrics would learn from this book that the conventional methods of analysis must be used with much greater caution and more concern about the correct interpretation of one's statistics than has been the general case to date.

Second, the sponsors of the conference also hoped that those interested in econometric theory would be stimulated by the discussion to explore new and improved techniques in a variety of inferential situations, to recognize that analysis of variance and regression are not to be regarded as inevitable panaceas to all statistical problems and that frequently we wish to learn in situations in which the available *a priori* knowledge is scant.

1. A Topical Overview

The book has five sections; each deals in turn with one of the major issues raised in this introduction. The objective of this section is to review very briefly the included papers, topic by topic, in order to demonstrate the relationship between the papers and the contribution of each to the specific aims of the Conference.

1.1. Informal Models

Part I of the book contains four main papers plus discussion, each of which deals with an aspect of the role of informal or parameter-free models. The basic situation is one in which researcher's knowledge is both incomplete and imprecise. For example, one might postulate that the conditional mean of one variable is positively related to another but not be able to make a more precise statement than that. In the past a typical response to such a situation has been to postulate a *specific* model of positive association, say a simple linear regression model, and proceed as if the assumptions involved in that parameterization of the problem were known to hold. The general type of response being recommended in this book through the contributed papers is to recognize explicitly one's inherent ignorance of the situation and proceed accordingly.

The authors focus on three alternative and largely complementary procedures. Mayer, in discussing exploratory data analysis techniques, relies least on *a priori* theoretical specifications of a model and most on systematic examinations of the data for potential regularities or observed relationships between the variables. In one sense the output of exploratory data analysis is a series of potential general statements about an economic situation which would be subject to testing in a more formal sense with other data.

The next two papers by Wold and Bookstein, respectively, are also in the "let the data tell their own story" approach, although more reliance is placed in these papers on *a priori* specifications of a linear structure. The procedure devised by Wold is called partial least squares. The method is used to obtain estimates of linear relationships between unobserved "latent variables" through indices of the latent variables created by weighted sums of observed variables. Bookstein's paper presents a geometric interpretation of the Wold procedure. The underlying maintained hypothesis is that there exist linear relationships between various weighted sums of the observed variables. The objective of the procedure is to find which weighted sums are related with what weights. In essence, the Wold procedure tries to let the data determine the index weights and the pattern of nonzero regression coefficients between the latent variables which are represented by the indices. Once again, one should interpret the results as "regularities" in the data which imply theoretical hypotheses to be tested with other data sources.

Hildebrand, Laing, and Rosenthal examine in their paper some solutions to inferential problems when the conditional predictions are in terms of sets. Using one of the authors' examples, game-theoretic predictions are most often in terms of specifying equilibrium sets, wherein changes in conditioning events lead to different equilibrium sets.

The common feature in the motivations underlying these various articles is a desire to lower the *a priori* informational content required by the statistical analysis—both in terms of the economic theory and in terms of the distributions of the random variables included in the model. The objective is to include in the maintained hypothesis only those specification statements about which the researcher is convinced that they are true. Thus this approach to econometric inference is to recognize that the known specification elements are meager, that assuming greater knowledge than one has is inferentially dangerous, and that relatively unstructured analysis of the data will be more productive than more parametrically complex work in generating testable hypotheses.

1.2. Specification Errors and Sensitivity Analysis

The first paper in this part, by Kelejian, is concerned with spelling out the conditions under which a macro-relationship can be said to exist when stable micro-relationships are known to hold. The inferential problem tangentially addressed in this paper is to consider the conditions under which useful inferences about behavioral parameters can be made using aggregated data even when one knows the form of the specific functional relationships between micro-variables. Kelejian also examines the circumstances under which estimates using aggregated variables can be used to make useful inferences about the parameters of mean micro-relationships, assuming that the parameters of individual micro-relationships differ randomly and independently from the mean relationship.

Theoretical research of this type is of great use in specifying clearly the theoretical limits of inference from the available data no matter what estimation procedure is used. Thus the Kelejian work enables one to evaluate the potential information which could be gained from observed data even under ideal sampling procedures. The point of this discussion is that there is no use considering estimates or hypotheses tests about a parameter until one knows that the parameter is relevant. As a most severe case, if one's analysis leads to the conclusion that no stable relationship between aggregate data exists, then any statistical inference is vitiated immediately. Less severe cases may be summarized by saying that inference about certain effects can be made but not about others.

The next paper, by Welsch, is concerned with a different aspect of the limitations imposed by a model and the data on the ability to make inferences. The Welsch paper addresses the issue of informal checks of the assumptions in the maintained hypothesis. An item of particular interest in this paper is the procedure for analyzing the sensitivity of the regression results to various observations, both singly and in groups, a procedure known as regression diagnostics. A related matter is Welsch's suggestion to use robust and bounded influence estimation procedures which provide estimates which are less sensitive to "outliers" than the standard ordinary least squares procedure. There has been some controversy over the use of robust estimators as estimators in their own right. However, if the object of one's analysis is a careful examination of residuals in order to ascertain the possible presence of specification errors, the use of robust procedures is particularly useful in that the effect of outliers in the distribution of the residuals is emphasized relative to the ordinary least squares procedure.

The Dent and Geweke paper is also concerned with checking the maintained hypothesis, but the procedures are much more formal and more specialized than those of Welsch. Dent and Geweke are interested in the problem of testing exogeneity and overidentifying restriction specifications in the context of what the authors call a complete dynamic simultaneous equation model.

The final main paper in this part, Hill's, takes a slightly different viewpoint to the effects of possible model misspecification. The Hill thesis is in essence that one should choose one's estimating procedure in order to have the least sensitivity to those aspects of the model specification in which one has least confidence. Hill explores in some detail within the context of simple linear models the sensitivity of alternative estimators to deviations from the maintained hypothesis, especially with respect to assumptions about the functional forms of the underlying distributions.

All four main papers in this part are complementary and illustrate the breadth of the problems to be faced in the analysis of the interaction between model specification and inferential procedures. At the most fundamental level one can question whether or not an estimable stable relationship between observable variables can usefully be said to exist. However, given that some relationship is theoretically posited to hold, the next step is to examine whether or not the maintained hypothesis is as assumed. Alternatively, one might consider choosing an estimation procedure which is insensitive to one's doubts about the model specification.

1.3. Formal Decision Rules for Comparing Models

There are only two main papers in Part III on formal decision rules for comparing models. The Chow paper is essentially concerned with the problem of how one characterizes or "describes the essential distinguishing features" of a complex dynamic multivariate model—this being the first necessary step in comparing models. As is well known, especially by those who have tried, the initial task of merely understanding a large dynamic model is not easy. But if such models are to be compared and evaluated, one must be able to understand them, be able to separate the essential from the inessential features, and have some notion of how comparisons with respect to one aspect of two models can be traded off against another. For a simple example, if model A provides more accurate forecasts of unemployment than does model B but less accurate forecasts of inventory changes, one may need to weigh the relative advantages of models A and B with respect to these two characteristics.

Chow's analysis provides an initial approach to these problems through the use of optimal control techniques. The basic idea is intuitively clear. The author recommends specifying a loss function which will enable one to transform the model's multivariate endogenous output values into a simple measure of "closeness" of model paths to prescribed paths, as determined by the initial values of the "control" variables. The fundamental aspect of the comparison is in terms of the relative sensitivity of alternative models to changes in initial control variable values as measured by changes in the specified loss function.

The Kadane and Dickey paper takes a very different approach to model comparison. The basic methodology is Bayesian; the context is linear regression. The first result is that choosing between two linear models on the basis of the posterior odds ratio is not entirely useful. Further, that if the Occam's razor concept is to be followed in Bayesian procedures for model selection, the researcher needs to specify a "utility function" with respect to which there are positive rewards for simplicity. In the practical trade-off between parameter parsimony and forecasting accuracy there is no alternative to specifying one's preference function over the space of estimators.

1.4. Time Series Analysis in Econometrics

The fourth part deals with the role of time series in estimating econometric models. The first paper by Howrey is a careful review of the role of time series analysis and an evaluation of the relative merits of the time and frequency domain approaches. Until recently, the comparison between "econometric" and time series approaches to problems was one wherein the "econometric" approach was economic theory intensive but unsophisticated with respect to the spectral structure of the random variables, whereas the time series approach was sophisticated with respect to the postulation of complex ARMA processes but virtually ignored economic theory. This simple dichotomy is now fast disappearing.

However, in one respect (as pointed out frequently by Granger) a pure time series approach can legitimately be taken with respect to the residuals from a regression analysis, since by design and intent of the estimation the regression residuals summarize the data information after the available economic information has been used. What is left "ought to be" pure white noise or at least an ARMA process independent of any regressor time paths. The suggestion of some stochastic process in the residuals can be interpreted as indicating the presence of specification error in the original model.

Howrey shows in his paper how time series techniques, especially spectral methods, can be used to aid the analysis of theoretically specified time series models. Examples are the estimation of covariance indices for Aitken procedures, Granger type tests for exogeneity ("causality" in Granger's and Sim's terms), and identification of the appropriate lag structure in the specification of dynamic models.

Engle's paper which deals with a Lagrange multiplier test in the context of time series models using spectral techniques is a particularly fitting example of Howrey's general thesis. From one point of view the Engle paper provides an extension of specification error tests of the Ramsey type to the time series (nonspherical disturbance term covariance matrix) type.

The Hatanaka and Wallace paper in its theme reverts to the notion discussed earlier of robustness, or its antithesis, sensitivity of estimators. The authors show in their paper that, in a distributed lag model suffering the nearly inevitable ill-conditioning of the regressor matrix, the estimation of low order moments of the lag distribution can be made much more precise than that of the original parameters of the model. This paper reminds us of an important but often neglected fact, namely, that reparameterization of a regression problem will frequently lead to an estimation problem which can be solved more effectively than can the original.

1.5. Experimentation and Tests of Economic Hypotheses

The last part of the book consists of two main papers plus discussion, both of which illustrate the role and potential role of experimentation in the development of economic theory. While only the most immediate and obvious experiments have been carried out so far, it is clear that the papers by Smith and by Kagel and Battalio mark a most significant contribution to useful econometric procedures. Until the work cited by these authors, the traditional view was that economics was inherently a nonexperimental science; we now see that this is no longer a relevant statement.

The chief difficulty, still not quite successfully resolved, is not the actual *performance* of some experiment but concerns the *inferences* which can legitimately be made. This is a universal problem but one of particular force in biological and even more so behavioral experiments. The issue is simply to what extent and under what circumstances can one infer general principles of behavior (human or otherwise) from experimental results. The problems are numerous and for the most part obvious; essentially how can one assume that the experimental situation does not itself affect behavior. Human subjects recognize the experiment as a game and have a tendency to play games with the game or, most often, to behave in a way which they feel the experimenter would like to observe. Animal subjects sometimes are also as "obliging" as human subjects.

Nevertheless, if experimentation can be successfully used at least for generating testable hypotheses, then our methodological tool kit will have been expanded significantly. If so, traditional econometric concerns will become of less relevance, and new challenges will arise. There will be thought given to checking the assumption that the experimental context did not influence the outcome; sample design will become of central importance; experimental design and control will become a new topic in econometrics, and new opportunities for choosing between alternative economic models will be introduced.

2. Research Recommendations

The brief comments in this section are meant to provide useful insights and suggestions for the practitioner and stimulating ideas for the econometric theorist. At the outset it is to be noted that a central and fundamental concept underlying the discussion throughout the entire book is the vital importance of the maintained hypothesis. Various papers have been concerned with numerous ways of reducing the specificity of the maintained hypothesis, others with checking its incorporated assumptions, still others with offseting by choice of procedure one's doubts about it, and finally there was the suggestion that through experimentation one can be much more confident about the specification of the maintained hypothesis.

2.1. Recommendations for Practitioners

The first important notion is that where possible one should try to tailor one's model specifications to the level of information one has; the "sin" of overspecifying and overparameterizing models should be minimized. Procedures now exist (and new ones are being developed) which can facilitate inferences from parametrically parsimonious models. Generally one should proceed by easy stages, not try to enter a marathon race before learning how to run. At the lowest information level, one may well be best advised to "explore the data" and recognize that with little prior information the task is more one of generating hypotheses than of testing them.

The second useful notion is that whatever the maintained hypothesis, it is seldom specified with complete confidence. Hence, before proceeding to draw inferences from the specified models with an easy conscience, one must check out the maintained hypothesis as carefully as possible. At the beginning one can use the intuitively appealing, robust, distribution free methods of Welsch and others to gain some feel for model sensitivity, data structure, and possible model misspecifications. Later one can advance to more formal and powerful specification error testing procedures of the Ramsey, Engle, Dent and Geweke type.

In the process of specifying the models two further issues are to be kept in mind. First, if using aggregate data, one ought to question the extent to which microtheory indicates that a stable macrorelationship is observable and under what circumstances. Second, before beginning estimation one should consider how to reparameterize the models in order to gain more useful information more efficiently from the observed data.

With time series data, one should not be reluctant to use time series procedures in cases where the precise stochastic structure of the model and lag lengths are unknown. In addition, spectral methods may at times provide more efficient solutions to "traditional" econometric problems.

Finally, one should keep in mind the feasibility of obtaining empirical evidence on our hypotheses by controlled experiments.

2.2. Recommendations for Econometric Theorists

Only a smattering of what the editors feel are some of the more fundamental or productive suggestions can be made, for in fact the possibilities are far too numerous to mention.

The analysis of models lacking detailed structure and parsimoniousness in parameters needs much more work; the surface of this topic has only been scratched. Exploratory data analysis is really an art, Wold's soft modeling is basically linear, and the Hildebrand, Laing, Rosenthal approach deals only with set predictions for discrete variables. What is really needed is some notion of a process of learning which uses models with little structure and a procedure for developing from such models others that have greater structure and are more parameter intensive.

With respect to the formulation of macro models (or at least models with highly aggregated data), more research needs to be done to spell out the stochastic framework within which stable macro relationships can be achieved. Also needed are procedures to check whether or not the necessary preconditions do in fact hold. An early example of an attempt at this approach together with a careful statement of the preconditions needed can be found in Ramsey (1972).

The development of interest in specification error tests since the first few early articles in the late 1960s and early 1970s is gratifying, but much remains to be done. At the moment, the set of specification error tests is a ragbag of miscellaneous procedures. The chief difficulty occurs with the presence of more than one error and with the resulting problem of how to isolate and identify the separate effects. Further, many of the more powerful tests are themselves subject to the same criticism as are the models, namely that they often rely for their power on aspects of the maintained hypothesis about which little is known for sure. In short, one needs specification error tests which are general, robust, and suitable for use with parameter parsimonious models. Further, despite the progress made to date much remains to be done in terms of devising tests in the context of the "complete dynamic simultaneous equation model." A more specific suggestion is to reevaluate the role of robust estimation in terms of its contribution to the power of specification error tests.

In this context the development of experimental economic studies will lead to a concern for experimental designs which will either prove robust to errors in the maintained hypothesis or which will facilitate the testing of the assumptions contained in the maintained hypothesis.

Currently, time series analysis treads an uneasy path between formal classical inferential procedures and the intuitive procedures of exploratory data analysis. This ambivalence in approach needs to be reconciled, most likely by the development of more formal model evaluation and comparison procedures in the context of complex multivariate time series models.

REFERENCE

Ramsey, J. B. Limiting functional forms for market demand curves. *Econometrica*, 1972, 40(March), 327–341.