

Preface

Econometric theory, as presented in textbooks and the econometric literature generally, is a somewhat disparate collection of findings, rather than the well integrated organized whole it might seem at first sight. Its essential nature is to be a set of demonstrated results that increase over time, each logically based upon a specific set of axioms or assumptions, but rather than becoming collectively complete, these inevitably remain a fragmentary body of knowledge. The practice of econometric theory consists of selecting from and applying this literature, as well as simultaneously evaluating it, so as to test its applicability and range, and support its further advance.

Today this practice is closely associated with the creation, development, and use of computer software and “econometric software” is the operational expression for this theory. Originally, the development of this software focused on the implementation of a progressively enlarging set of estimators, but now its best expression involves the attempt to provide not only the means to estimate parameters in different ways but also to test each of the underlying assumptions in the most meaningful way. The argument that might be made to buttress these assertions begins from the observation that the range of estimators that have been discovered by econometricians is reasonably extensive and that some of these are particular to a specific context. However, the most generally applied estimator is Ordinary Least Squares and, except in those situations where there is a priori knowledge of its unsuitability, its role is to be the starting point. To the extent that, either in practice or in principle, OLS plays this role, the consequence is then to give a particular importance to the set of supplementary evaluative tests that are applied in conjunction with it.

Following from these considerations, this monograph presents, classifies, and documents the particular diagnostic tests associated with Ordinary Least Squares regression that are provided by the existing econometric software packages, with the goal of supplying reliable and useful information to three categories of people: econometric software developers, econometric theorists, and more generally those economists who use this software. Towards this end, it attempts to both discover and evaluate the present state of the art. The research behind it has been undertaken in the form of an interactive survey, conducted not as an external examination but instead with the active assistance and collaboration of the econometricians who have created, designed and developed these packages. Furthermore, this investigation has

been embarked upon with certain specific fact-finding intentions. One of these is to provide a generally useful set of benchmark values of these diagnostic statistics. Another is to determine how the individual tests have been implemented, assessing both their degree of commonality and, wherever there are differences, why these have occurred. However, this study should also be viewed in a broader context: it is one of an ongoing series of contributions collaboratively produced by a number of people since about 1995, who together attempt to consider and assess the various computational aspects of the modern applied economics research environment. The common perspective is that the existing software packages collectively define the operational state of the art of econometrics, and at least certain aspects of the general applied economic research environment, and that therefore it is vitally important for both econometricians and applied economists to understand the specific characteristics of the research facilities available to and actually used by economists.

The statement of intent just provided clearly serves as an abstract. However, it is possible that the scope of the investigation may not yet be wholly self-evident. On the face of it, the idea of identifying, classifying, and presenting a particular set of diagnostic tests seems simple enough, but any evaluation of the state of the art must also compare what is with what could be, inevitably spawning a number of questions, among them being: how and why were the particular tests implemented by developers originally chosen? More specifically, what constitutes the most appropriate set of diagnostic tests, both individually and collectively? Is this the set actually offered by the existing packages, perhaps as a result of some type of invisible hand selection process? Can a uniquely appropriate set actually be determined? In addition, certain more general questions may also ultimately need to be addressed, among them: why limit the present study to Ordinarily Least Squares?

What is an econometric software package anyway? The definition of econometric software is actually a nice question and to answer it raises more questions, the first of which can be expressed as one of nature versus nurture; that is, should this answer be approached by first addressing the nature of econometrics – or is its content simply the evolutionary result of its nurture? The content of econometrics is of course a question that has been addressed repeatedly since the 1930s, with as yet no definitive conclusion reached. However, assuming for argument's sake that its characteristics are capable of being pinned down one might ask, can the nature of econometric software actually be determined on this basis? Or to put this last question in slightly different words, is it reasonable to expect that “econometrics” is so well defined in its subject matter that a single econometric software package could in principle serve as an operative expression of it? Might it instead be more appropriate to ask, to what degree are the existing packages an operative reflection of particular topic areas of the published econometrics literature? Given this more restricted interpretation, is it then possible that the match between the literature and the software offerings is less than exact, not only in subject matter but also in terms of the specific characteristics of the algorithmic implementation?

These are all questions that finally need to be considered, but at the outset rather than attempt to answer them directly, particularly in the order asked, it might be better to approach the matter in a somewhat crabwise fashion, beginning by asking

first, what is software? Fortunately, the word “software,” generally referring to the set of instructions that cause an electronic computer to operate in a particular way, is nowadays a concept that is familiar to anyone who has used such a device, particularly during the past approximately thirty years, since the introduction of the microcomputer. “Hardware,” in contrast, is of course the physical substance of a computer, without the electricity. These machines are now ubiquitous and their use ordinarily requires conscious interaction: from time to time, a user must “boot” and, occasionally, “restart” the computer in a rather direct, hands-on manner, as well as both “install” and “execute” software packages. Once the operating system and so-called “applications” packages are in execution, they cause the machine to perform its useful work. The idea of “software” therefore requires little explanation as a general concept, but there is nonetheless an evident degree of ambiguity inherent in the compound phrase “econometric software.” Specifically, the definition of this software might be considered from any of several perspectives: as software created *for* economists and econometricians or *by* them, or that takes as its subject matter the set of econometric techniques, or, possibly, simply as being that software that economists or econometricians happen to choose to use professionally.

However, most fundamentally the critical issue is how this software shapes the economist or econometrician’s interaction with his or her research materials and the specific way in which theories and hypotheses are then confronted by empirical evidence. Notice that the just mentioned alternative perspectives are each potentially important to a particular evolutionary consideration of the computational methods employed by economists. For example, “happening to choose,” as a possible historical explanation, would appear to imply at least a degree of exogenous influence by some other discipline on the creation of this software, since the software’s prior existence is thereby indicated. Furthermore, if this *ex post* choice is *commonly* made, it also implies at least a touch of disciplinary instability; that is, if economists persistently choose to use “other” software, rather than what has been developed endogenously, the extreme implication ultimately might be either the ongoing or a one-time reorientation of economics, if not econometrics, because of the software that happens to be used. After all, the tools employed can affect the specific work done and hence the results obtained. Therefore, in the end, the evolutionary outcome could finally be research that is done in conformity with and is shaped by some “other” tradition, then affecting for better or worse the progress of econometrics.

In contrast, the several questions raised earlier about the inherent nature of econometric software obviously take as given a presumed global stability, implicitly interpreting such software as being created in conformity with the existing econometrics literature, with the techniques offered originally established by that literature and then subsequently affecting its ongoing development. However, when such questions are considered in this endogenous context, casting the definition in terms of either *who* creates the software or *for whom*, such considerations in each case still potentially raise questions concerning the likely evolutionary path of economics and econometrics. *Who* creates the software, for instance, can result in unexpected effects if the creators bring to the task an imperfect understanding of

the current practice of econometrics. *For whom*, in contrast, suggests the existence of a strongly held, or at least essentially self-conscious, concept of the particular practices of economists and econometricians. It is possible of course to view at least certain of these various potentialities as tenuous, but in the general scheme of things it is important to give each of them some consideration, for none can be simply rejected out of hand as being preposterous.

The fundamental question that is addressed by this study is the present state of the art and an important consideration is how this came about. Rejecting both chance and instability as likely possibilities, the selection of software for the present study was made on the basis of the stated intentions of the developers of the programs chosen, which initially required some investigation to identify the population of candidate packages. It was first necessary to examine the relevant economics and statistical literature, including the various lists of software that have been advertised or posted on pertinent Internet websites during the past five to ten years as being “of interest” to economists, particularly sites that aim to identify “resources for economists,” such as www.rfe.org. Once a unified population list had been thus created, the developers or vendors of these packages, as relevant, were contacted and each simply asked if the software they offered was *intended* to be “econometric software.” This “sampling” method is obviously somewhat informal and of course left open the possibility that a given supplier might make an unsustainable claim, or that an appropriate package might remain undiscovered, but actually each of these potential problems were easy to minimize. To minimize wrongful inclusion, each selected developer needed to provide information and participate actively, to continue as a member of the set. To minimize wrongful exclusion, the list of selected packages has been advertised widely among economists since early 2003. Initially, this list was circulated to discover possible additional candidate software packages. Later, in 2004, a compendium of these packages, complete with developer-provided descriptions, was published as both a chapter of a book and a paper in a special issue of the *Journal of Economic and Social Measurement*. Simultaneously, a fairly detailed study was made of the various design characteristics of the included packages and the findings also published in the book and special issue. This process has provided both publicity and a clear statement of the properties of the included packages, as well as contact information for the future. Notwithstanding any classification problems that might be associated with this approach, its evident virtue is that it permits inferences to be drawn about the characteristics of the included packages from their designers’ and developers’ stated intentions. Furthermore, it is reasonable to expect, over time, that packages that are developer-identified as econometric are more likely to characterize and reflect econometric practice than those intentionally developed for some other purpose.

An associated consideration was the choice of data to be used to make comparisons among the different software packages. The principal data set chosen was selected well in advance. It consists of a previously published set of observations that have the virtue of being both widely available and long-known to econometricians, thus placing little burden on each software developer to acquire and use. After all, it was not obvious in advance exactly what might be discovered, whether for

instance the conformity between packages would be sufficiently great as to make the comparisons to all intents and purposes a non-event. In practice, as a general choice, this data set proved to be quite satisfactory as a means of demonstrating a variety of interesting results, results that *intentionally* do not depend upon any type of stress testing or the particular (read unusual) characteristics of the data employed. It has long been known, from numerical accuracy tests developed by Wilkinson, among others, designed specifically to test for specific types of computational errors and problems, that it is often possible – with some ease – to expose particular shortcomings. In other contexts, such testing might be desirable, even as a supplement to the present study. But, in this study, the first question to be considered was: what are the differences, if any, under normal, even benign conditions? It would always be possible, later, to stress test, but there is actually much to be learned in the absence of this type of, potentially hostile, accuracy testing. Recall from the discussion above that the results reported in this volume are the result of active (and cordial) cooperation among econometric software developers, so that it was critical, from the beginning, to proceed in a way that would provide a beneficial joint learning experience. Not only was the purpose of this study to determine what values and test statistics each econometric software package might produce, but why.

As the Chinese say, each journey begins with a single step; this study constitutes that first step. However, as the comparisons began to be made, it became clear that in certain instances the original data set chosen could not produce fully informative results. Therefore, it proved useful to employ an alternative, easy to obtain, set of observations on US GDP, which specifically have been used to illustrate aspects of the computation of specific unit root test statistics. But in all cases, with the sole exception of certain historical displays replicated to illustrate former ways of presenting information, the principal numeric results shown in this volume are generated using one or the other of these data sets and the observations themselves are included in its appendix, as well as still being available from the original published sources in hard copy, if not machine-readable form.

Furthermore, to insure the possibility of replication, each of these values has been independently computed by a minimum of two econometric software packages and usually by more than two. In particular, in the case of the diagnostic tests performed in common by a number of packages, most if not all of these have been tested for agreement. Whenever differences have been discovered, a concerted effort has been made to determine why. In certain cases this evaluation has led to recoding on the part of individual developers, although the details of such changes have not been examined here: the purpose of the present study is not to track and report on the historical changes in the individual packages, but rather simply to display reproducible results produced by the existing set of econometric software packages. An objective of this study is to provide benchmarks. In the future, anyone who wishes to examine the numerical characteristics of a given package will be able to assess them in the light of these results.

A central evaluative finding of this study is that in many cases the numbers that have been produced by the surveyed packages differ in various important ways. However, it must also be added quickly that there were actually only a few

differences discovered in those instances in which developers *intentionally implemented the same formulae*. Differences found almost always occurred because of differences in the ways in which individual developers independently implemented the tests, sometimes reflecting the particular identifying names used for test statistics but in other cases the implementation of variations on known formulae. Furthermore, as is discussed later, program developers in some instances have intentionally implemented a particular form of a diagnostic test, knowing in advance that they were producing a variant. Sometimes these computational differences reflect the developer's firmly held belief that this variant has better diagnostic properties. In other cases, the differences occurred inadvertently. Ideally, a benefit of studies such as the present one is that inadvertent differences will be minimized in the future.

In addition to numeric differences, when comparing one package pair-wise against another, it was also found that packages commonly differ in their range of offered diagnostic tests. There are a number of test statistics that are, or optionally can be, displayed by all or nearly all packages, but there are also a number of others provided only by one or two packages. To a degree, the observed differences characterize the recent evolution of econometric software, for during the past ten to fifteen years there has been an increasing tendency for econometric software packages to display more and more test statistics, reflecting among other things the peer review process of economics and econometrics journals, but also the more forensic methodology of modern econometrics. However, as just indicated, the particular choice of the diagnostic statistics generated and displayed has often been made by econometric software designers both competitively and independently, with the result that there are now noticeable differences between packages — at least when they are evaluated, snapshot fashion, as of a particular date. These differences may not persist over time, once publicized, but they are documented here by a set of tables that identify the test statistics offered by each of the individual packages. In contrast, the numeric results displayed are more restricted: as mentioned earlier, only values *independently* reported by two or more packages are displayed. The italics reflect that individual packages are not always developmentally independent of each other.

This display choice is of course open to criticism. The range of statistics reported by the packages collectively, as well as the fact that a particular statistic is generated uniquely by a given package, are each findings that are of course quite relevant to the present study in its attempt to discover and evaluate the state of the art. Therefore, this information is provided. However, whenever a statistic is computed uniquely or agreement between two or more independent packages could not be confirmed, the consequence was to preclude its use as a benchmark value.

Evidently, this study is restricted in scope. Ideally, it might be desirable to describe each of the existing econometric software packages individually, explaining both their present characteristics and the reason for those characteristics. There is a story to be told in each case, but it is a story that can only be told by the individual developers. Similarly, as an ideal, it might be desirable to establish a full set of benchmark values for each of the statistics reported by each of the packages. However, this goal also needs to be pursued separately later. The particular aims, form, organization, and even the conclusions of this study reflect the newness of

this type of investigation. It appears to be the first of its type, so that the best that can be expected is for it to provide a limited result. On the positive side, this limited result can act as an arbitrating ploy. One of the possible consequent effects may be a greater commonality of pertinent results in the future, for inevitably any study such as this one affects the design of software packages. It makes software developers and users each more aware of the facilities provided by the software packages included in the study. It exposes the ways in which a given package differs from other packages, which in itself is a spur to developers. The developers of individual packages inevitably view certain other packages as prime competitors, so that in this case especially, the information provided by a study such as this one is, at the minimum, market informing.

More generally, the intention of this study is not to judge any particular package relative to another, or to expose the computational defects of individual packages. No attempt has been made to identify a particular set of numeric results with any given package, except in the case of certain pertinent examples, which are provided simply to illustrate the differences that can occur between packages. This study does not evaluate individual packages, but instead the econometric software package as a classification. To contrast and compare individual econometric software packages more specifically might provide useful information for users of those packages, but this course of action would distract attention from the essential findings of both commonalities and reported differences. In the final analysis, the essential purpose of this study is to play the role of a mutual learning experience for all the participating developers. Ideally, one of the effects will be to establish a collective understanding of the present state of the art.

Of course, the present investigation is also limited by its focus on a single estimation technique. The choice to consider only diagnostic tests that are directly associated with the Ordinary Least Squares parameter estimation technique can be seen as motivated by several considerations. The first is that it can be argued that one of the principal distinguishing characteristics of any econometric software package is its inclusion of this technique among the offerings, notwithstanding that it may be the inclusion of other techniques that distinguishes such packages from statistical and other types. It is the most basic of all the econometric parameter estimation techniques, as well as the one most commonly used. However, in this choice too, practicality has been allowed to dictate. Historically, there have been surprisingly few attempts to consider either the design of econometric software packages or the way in which econometric techniques have been implemented in these packages, including their numerical accuracy and the other computational characteristics of the results produced, so that it has seemed most sensible to start with what is ostensibly the base case and to work outwards from there.

Even as a first examination, rather than a complete examination, the present study could not have been produced without the active assistance, advice and support of econometric software developers individually and collectively. I am grateful to my fellow econometricians who are software developers, as well as to others who have provided information, comments and suggestions. I appreciate in particular the assistance and advice of Jerry Adams, Irma Adelman, Micah Altman,

Richard Anderson, Terry Barker, David Belsley, Herman Bierens, Jon Breslaw, Guiseppe Bruno, Ian Cahill, Paul Calvert, Allin Cottrell, Clint Cummins, Kenneth Berk, James Davidson, Tom Doan, Jurgen Doornik, Mark Eisner, Ray Fair, Arthur Goldberger, Richard Goldstein, Clive Granger, William Greene, Bronwyn Hall, Robert Hall, Stephen Hall, Tim Harrison, David Hendry, Peter Hollinger, Charles C. Holt, Lawrence Klein, Robert Lacey, Cynthia Latta, Edward Leamer, James LeSage, David Lilien, James MacKinnon, Keith May, Michael McCracken, Bruce McCullough, Marc Nerlove, Ray O'Brien, William Peterson, Peter Phillips, Richard Pierse, Robert Pindyck, Brian Poi, Ross Preston, Duo Qin, David Reilly, Colin Rose, Ronald Schoenberg, Lucy Slater, Houston Stokes, Daniel Suits, Robert Summers, William Teeters, Gareth Thomas, Kenneth White, Mike Wickens, Vince Wiggins, Mark Wiley, Clifford Wymer, Achim Zeileis, and Arnold Zellner for their patient willingness to assist me. I am also indebted to Arthur Goldberger, Bronwyn Hall, David Hendry, and Peter Hollinger for kindly making available to me both early manuals and particular examples of regression displays that have been used in the preparation of this monograph. I express my especial debt to Houston Stokes for his constant willingness to provide me repeatedly with code, documentation, and analytical assistance during the process of determining the characteristics of the various computations made. I am wholly responsible for all errors of fact and omission, and for all opinions stated that are not identified as quotations.

Newagen, Maine
12 January 2007

Charles G. Renfro

The Practice of Econometric Theory
An Examination of the Characteristics of Econometric
Computation

Renfro, C.G.

2009, XVI, 311 p. 28 illus., Hardcover

ISBN: 978-3-540-75570-8