Teaching Statistical Methods to Undergraduate Economics Students

By WILLIAM E. BECKER*

Little published work has appeared on the teaching of statistics in economics. This lack of attention is surprising since John Seigfried and James Wilkinson (1982) found that nearly 85 percent of the 546 schools surveyed required at least one course in statistics for an undergraduate major in economics, and approximately 50 percent of the departments of economics offered their own course. After introductory economics, only intermediate macroeconomics and money and banking have higher enrollments than statistics. This paper focuses on the goals and objectives of the introductory economics statistics course. It also gives attention to the more advanced econometrics courses undergraduates may elect to take.[1] It emphasizes the quantitative skills all undergraduate majors in economics should possess irrespective of their career plans.

As pointed out by W. Lee Hansen (1986), it is much easier to talk about what should be covered in a course than to define the student competencies to be gained. The latter requires the definition of student activities while the former treats the student as a passive recipient. If students are to do more than regurgitate definitions, duplicate proofs, or perform repetitive computations, then what is expected from them must be specified. Whether an instructor "covers" something or whether students can actually do something with what is "covered" are two different issues. This paper focuses on what students should be expected to do, not on what should be covered in economics statistics courses.

I. Applying Statistical Methods to Economic Data

Eric Sowey defines econometrics as the study of "...applying statistical methods to economic data" (1983, p. 257).[2] An economics department's rationale for offering an introductory statistics course or more advanced econometrics courses, as opposed to having them offered by a department of statistics or mathematics, rests on an argument that there is something unique about applying statistical techniques to economic data.[3]

*Indiana University, Bloomington, IN 47405. I am indebted to Suzanne Becker, Rendigs Fels, W. Lee Hansen, Donald Harnett, Thomas Johnson, Peter Kennedy, Catherine Melfi, and Robert Sandy for helpful comments on earlier drafts of this article.

[1]According to Seigfried and Wilkinson, the undergraduate econometrics courses had the seventeenth highest enrollment among the 62 courses listed. They were taught by 46.5 percent of the 512 departments for which information was available, but only 5.9 percent of the entire sample of 546 departments required an econometrics course for undergraduate majors in economics.

[2]Sowey defines econometrics as "the discipline in which one studies theoretical and practical aspects of applying statistical methods to economic data for the purpose of testing economic theories (represented by carefully specified models) and of forecasting and controlling the future path of economic variables" (p. 257). The words "practical aspects of applying" are downplayed by Sowey; he presents a highbrow view of what econometricians should teach, with the teaching of applications left to others.

[3]Donald Waldman and I (1987) prepared a graphical interpretation of probit analysis where the dependent variable was unmeasurable utility and only one of two outcomes was observable for each sample individual's choice. A statistician suggested that we redo the example using a "continuous real-world measure" as the dependent variable. He suggested that we first fit a least squares regression to this continuous measure and then split the data to form the dichotomous dependent variable for the probit model. Students would gain insight in comparing the results. While a statistician untrained in economics might think that students will find such comparisons enlightening, the unmeasurable nature of utility and the modeling of discrete choices is lost and thus has little relevance in a statistics course offered by an economics department.

Yet the word "applying" gets lost in many discussions about teaching courses labeled "economics statistics" and "econometrics." It is the application of statistical measures and statistical inference in an economic analysis that must be emphasized in an economics department's offerings.

To learn to apply the tools of statistics to economic data, there is little justification for examples involving the drawing of balls from urns, coin and dice tricks, or examples from the natural sciences. Students need to be involved in working exercises, considering case studies and solving problems which reflect what economists do. This implies replacing the urns with a preselection pool from which individuals are hired; replacing the coins and dice with surveys in which individuals face multiple choice responses, and replacing examples from genetics with examples from quality control.

Analyzing economic issues and drawing conclusions based on economic data may be the best, if not the only, way to sharpen the ability of students to "emphasize the interpretation, the limitations and the significance for economics of statistical techniques," which according to Sowey (p. 282) is a goal of "nonspecialist courses" in econometrics. Jacques Drèze (1983), commenting on Sowey's suggestion that applications should take a back seat to theory when class time is short in "specialist courses," stated that practical illustrations and examples should be introduced all along in the teaching of economics statistics. Both basic and more advanced principles can be put to practical use through examples drawn from the academic journals and popular press. The short or mini case study approach introduced into the teaching of economics by Rendigs Fels (1974) can be easily and fruitfully modified for teaching economic statistics. Short case studies enable instructors to demonstrate a specific form of statistical analysis while giving students an opportunity to observe how concepts and theory are used to examine particular problems or issues.[4] When confronted with a similar problem, it seems more likely that students can make the transfer than if the specific form of analysis had been presented without the case application.

There are those who argue that even undergraduate students should be able to work with large, real-world data sets on which a myriad of analytical techniques can be performed. Some introductory textbooks include a large data set which is used in each chapter. The mini case study approach of Fels, however, suggests that before students are confronted with compound problems and large data sets, they should be exposed to real or simulated situations that involve a limited number of concepts and easily managed data sets.

In using large data sets, it is difficult if not impossible for beginning students to see how an array of observations and a specific method of analysis can illuminate a question or problem. For example, in a histogram of 1,000 observations, students may not be able to appreciate the influence of extreme values on different measures of central tendency; with only a few well-chosen observations, the effect of extreme values is made apparent. More important, use of one large data set to show multiple statistical techniques may give the impression that a good

[4]As an example of a short case, consider the following from myself and Donald Harnett:

"You are the state commissioner of insurance. An irate consumer group wants you to initiate a costly review process that it hopes will result in the fining of an insurance company for deceptive advertising. The insurance company is advertising that the average processing time to settle a household damage claim fully is only 8 days. The consumer group claims that the majority of such claims are not settled in even 13 days. Thus, the group asks 'How can the insurance company be guilty of anything short of fraud?' What action will you take before responding?" [1987, p. 81]

This case calls attention to the difference between the mean and median, their relationship to skewness, and the need for measures of dispersion. In the case approach, statistical concepts are taught only if they can be used in a problem-solving situation. This means that the first question asked by the instructor in constructing a syllabus is not what topics should be taught, but rather what type of problems and questions should the students be able to analyze and answer. Once the problem and questions are specified the statistical concepts, methods, and techniques to teach will be apparent.

analysis involves using all the statistical techniques available without regard to their appropriateness. Students should come to realize that the problem or question dictates the data needed and the analytical techniques employed, and not the other way around. Fels demonstrated that many short cases can help students recognize this point. Starting with a limited number of statistical concepts and analytical techniques, students learn how to select the concepts and techniques appropriate for solving different classes of problems. Since the emphasis is on problem solving, students quickly realize that memorization of a lot of definitions, the ability to plug numbers into formulas, and agility with a hand calculator are not critical in applying statistical methods in economics. To foster this realization, instructors must ensure that their presentations and instructional materials aim at problem solving and do not reflect only a preconceived notion about what topics must be covered.

II. Statistical Analysis Requires Computer Software

To apply statistical methods to economic data, students need computer programs for handling computations. What they must do with a program (arithmetic, transformations, graphics, search procedures, etc.) and how sophisticated the program must be (both in terms of programming knowledge and statistical routines) depends on the course. But if econometrics is viewed as an area of applied statistics, then all undergraduate econometrics courses starting with the introductory course should use computer programs.

Surveys completed by textbook publishers and academics alike suggest that instructors are already using computer programs such as MINITAB, SAS, and SPSS in statistics courses. How these programs are used remains in doubt. For example, in the E. L. Rose, J. A. Machak, and W. A. Spivey (1986) survey of business programs, 99 percent of the responding schools said they covered simple regression in the introductory statistics course and 88.8 percent considered multiple regression. Yet, 48.7 percent of the responding schools stated that they do not use the computer for regression with cross-sectional data and 57.3 percent do not use it for regression with time-series data. About 17 percent do not allocate any class time to using computers in statistics, and 68.4 percent allocate only 1 to 4 hours during the term for this purpose.

The first course in statistics should require students to load and run a menu-driven statistics package (for example, MICROSTAT on a CPM or DOS-based microcomputer, IDA on a VAX minicomputer, or STATPAK on an IBM mainframe). Menu-driven programs are advantageous because they do not require extensive class time for students to learn to operate the program. For example, a *PC Magazine* review of MICROSTAT stated that "someone completely unfamiliar with MICROSTAT could follow the clear menus and obtain useful output" (Marvin Bryan, 1986, p. 214). Unlike the detailed instructions required to run a regression on command-driven programs, menu-driven or interactive programs prompt the user for specification details (for example: What is the dependent variable? How many explanatory variables will be used? Identify the explanatory variables? etc.). Of course, to follow these menus, students must understand what they are trying to do, and once the output is obtained, they must be able to explain its meaning. Only then is the output useful. Unless students attempt statistical work on a computer, however, they will never appreciate the potential of statistical methods in economic analysis.

By the end of the first course in statistics, students should be able to use a menu-driven statistics package to enter and retrieve data, perform data transformations, calculate descriptive statistics, generate frequency distributions, determine probabilities with the binomial, hypergeometric, poisson, normal, t, F, and *Chi*-square distributions, and fit least squares regressions. Only a few hours of class time are typically required to help students gain the computer and program expertise required to perform these functions.

Time devoted to computer training also can be used to teach concepts which previously required student knowledge beyond high school algebra. For example, during the computer training, students can learn how the computer generates distribution functions without first learning how to integrate density functions. Students can learn how the regression algorithm is based on the idea of minimizing the sum of squared errors. Without ever taking a derivative, they can develop an understanding of least squares through discussions of computer-generated scatterplots.

Class time need not and should not be devoted to teaching computer use outside the context of statistics. Just as students should learn that the problem or issue under consideration gives rise to the data and statistical techniques employed, they should come to realize that the data and statistical techniques suggest the computer program to be used and not the other way around. This later student recognition may not be achievable in the first statistics course, but is achievable by the end of the more advanced econometrics course.

Unlike microcomputer software, use of statistical packages on minicomputers and mainframes has disadvantages for beginning students. Instructors must invest time to make arrangements to obtain and maintain student accounts, which typically expire after the course is over, prohibiting continued use by students. Moreover, unless future employers possess these programs and hardware, students will be unable to demonstrate their practical knowledge. Microcomputer-based programs overcome these disadvantages. Through the use of site licenses, student versions, education discounts, or public domain software, students can purchase copies of statistical packages at little more than the cost of a disk. Bundled textbook, workbook, and statistical packages are now selling for under $55.

The ease of menu-driven programs in generating descriptive statistics, calculating probabilities from specific distributions, and performing other routine computations assures their future use by students who continue their education in econometrics and those who terminate their training after the first course in statistics. Continuing students will need to have their library of computer programs expanded. Powerful and yet relatively easy-to-use programs, such as LIMDEP, RATS, and GAUSS, make it unnecessary for these students to learn structured programming languages to do matrix manipulations, perform maximum-likelihood estimation, test linear restrictions, or design their own statistics.

III. Outdated Computational Procedures

The time required to help students learn how to use menu-driven statistics programs can be more than offset by savings in instructional time devoted to teaching computational algorithms. With programs to determine probabilities based on appropriate distributions, beginning students no longer need the ability to approximate the binomial with the normal, use the binomial to approximate the hypergeometric or use the z when the t distribution is appropriate. Interpolating between values in z or t tables is a skill that students no longer need.

Students should no longer be expected to demonstrate their proficiency at plugging numbers into computational formulas. Instead, they should be called upon to demonstrate the principles involved in a computational procedure. Such learning does not imply that students know all the alternative computational versions of a formula. For example, students should know that the variance is the average squared deviation of observations around their mean. This definition of the variance implies that students calculate a finite population variance from the formula[5] $\sigma^2 = \Sigma(f_i/N)(x_i - \mu)^2$. To understand the definition of the variance,

[5]This is not to argue that the calculation of sample variances should be deemphasized, or that the calculation of the variance of a continuous random variable has no place in the introductory course. Until students understand degrees of freedom, however, calculations of sample variances need not lead to student understanding that the variance is nothing more than the average of the squared deviations of observations around their mean. Similarly, beginning students have difficulty grasping the meaning of the variance if they are first confronted with an integral which involves the product of a density function and the squared deviations of the continuous random variable around its mean.

however, students need not demonstrate proficiency with all the algebraically equivalent forms of this equation.

At more advanced levels of econometrics, computer programs make matrix manipulation relatively easy; this implies no need to emphasize hand calculator procedures. For example, in the estimation of regression coefficients, students no longer need to see or work with the transformation matrix

$$A=\begin{bmatrix}1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1\end{bmatrix}-\frac{1}{n}\begin{bmatrix}1 & 1 & \dots & 1\\ 1 & 1 & \dots & 1\\ \dots & \dots & \dots & \dots\\ 1 & 1 & \dots & 1\end{bmatrix}$$

which is used to convert the n observations on each variable (as contained in the X matrix and y vector) to deviations from sample means. A computational procedure that first requires converting data to deviations via this A matrix and then calculating slope coefficients in a second step and the intercept in a third is of no value (as a learning experience or for practical purposes) to students who have access to computer programs for matrix operations. The A matrix has nothing to do with understanding regression coefficient estimation, coefficient interpretation, and the role of coefficients in prediction. This understanding comes from working directly with the products of the X matrix, the $(X'X)^{-1}$ matrix and the $X'y$ vector. Diagrams showing how these matrices project observations into predictions and errors provide the intuition for least squares regression analysis.

Elimination of coefficient estimation based on the A matrix would remove 3 pages from the 39 pages in Chapter 5 of J. Johnston's (1984) econometrics text. Similar savings would be realized in other books where the k-variable linear model is introduced in matrix form. Eliminating discussions of all such outdated computational procedures frees class time to help students gain a better grasp of when and why particular statistical techniques may be appropriate, and makes more time available for the instructor to help students formulate correct interpretations of their results. It also provides class time to introduce the more contemporary modeling and estimation techniques associated with problems involving time-series, discrete choice, and qualitative data.

Modern computer statistical packages make it possible for students to use certain maximum-likelihood techniques which a few years ago could be performed and understood only by Ph.D. econometricians. Using computer printouts that show the iterative values tried by the program, students with an understanding of least squares regression can learn probit, logit, and Tobit modeling without first taking four semesters of calculus. For example, using a probit specification, Waldman and I demonstrate with a three-dimensional diagram the maximum-likelihood method. We show the line that best describes an unobservable but normally distributed variable y^* given the relative frequencies of observations that are known to be above (or below) a threshold value of y^*, at each value of an independent variable. This development and interpretation of the line's properties require an understanding of density, mass, and the maximizing principle, but not the actual calculation of integrals or derivatives.

In conclusion, the availability of easy-to-use and powerful microcomputer programs puts sophisticated statistical techniques within the reach of undergraduates. Practical applications should never again take a back seat to theory in the introductory statistics course or the more advanced econometrics courses. No longer should students walk away from these courses saying that econometrics is "a marvelous array of pretend-tools which would perform wonders if ever a set of facts should turn up in the right form" (G. D. N. Worswick, 1972, p. 79). Our job as teachers is to assist undergraduate students in gaining an understanding of operational tools and techniques while they learn to appreciate the limitations of the procedures and their legitimate use in statistical inference.

REFERENCES

Becker, William E. and Harnett, Donald L., *Business and Economics Statistics With Computer Applications*, Reading: Addison Wesley, 1987.

______ **and Waldman, Donald M.,** "The Probit Model," in William Becker and William Walstad, eds., *Econometric Modeling in Economic Education Research*, Boston: Kluwer-Nijhoff, 1987, 135–40.

Bryan, Marvin, "Business Forecasting," *PC Magazine*, August 1986, *5*, 211–34.

Drèze, Jacques H., "Nonspecialist Teaching of Econometrics: A Personal Comment and Personalistic Lament," *Econometric Reviews*, No. 2, 1983, *2*, 291–99.

Fels, Rendigs, "Developing Independent Problem-Solving Skills in Economics," *American Economic Review Proceedings*, May 1974, *64*, 403–07.

______ **and Uhler, Robert G.,** *Casebook of Economic Problems and Policies*, St. Paul: West Publishing, 1st ed., 1974 (5th ed. edited by Fels and Stephen Buckles, 1981).

Hansen, W. Lee, "What Knowledge is Most Worth Knowing For Economics Majors?," *American Economic Review Proceedings*, May 1986, *76*, 149–52.

Johnston, J., *Econometric Methods*, New York: McGraw Hill, 3rd ed., 1984.

Rose, E. L., Machak, J. A. and Spivey, W. A., "A Survey of the Teaching of Statistics in Business Schools," unpublished paper, Graduate School of Business Administration, University of Michigan, October 1986.

Seigfried, John J. and Wilkinson, James T., "The Economics Curriculum in the United States," *American Economic Review Proceedings*, May 1982, *72*, 125–38.

Sowey, Eric R., "University Teaching of Econometrics: A Personal View," *Econometric Reviews*, No. 2, 1983, *2*, 255–89.

Worswick, G. D. N., "Is Progress in Economic Science Possible?," *Economic Journal*, March 1972, *82*, 73–86.