2016 HES PRESIDENTIAL ADDRESS: STATISTICAL INFERENCE IN ECONOMICS, 1920–1965: CHANGES IN MEANING AND PRACTICE

BY

JEFF BIDDLE

I review changes over time in the meaning that economists in the US attributed to the phrase "statistical inference," as well as changes in how inference was conducted. Prior to WWII, leading statistical economists rejected probability theory as a source of measures and procedures to be used in statistical inference. Trygve Haavelmo and the early Cowles Commission econometricians developed an approach to statistical inference based on probability theory, but the arguments they offered in defense of this approach were not always responsive to the concerns of earlier empirical economists that the data available to economists did not satisfy the assumptions required for such an approach. Despite this, after a period of about twenty-five years, a consensus developed that methods of inference derived from probability theory were an almost essential part of empirical research in economics. I conclude with some speculation on possible reasons for this transformation in thinking about statistical inference.

I. INTRODUCTION

"Statistical inference," as I use the phrase in this paper, is the process of drawing conclusions from samples of statistical data about things that are not fully described or recorded in those samples.[1] This broad definition is consistent with several narrower

---

Jeff Biddle, Michigan State University. This article was the HES Presidential Address given in Durham, North Carolina, June 19, 2016.
[1]Statistical inference is one approach to the broader problem of "generalizing" scientific knowledge, which, at the point of its generation, is necessarily quite "specific" in a number of ways. See Morgan (2014).

understandings of the phrase, and indeed, different economists from the period I survey attached different meanings to "statistical inference"; thus my reference to "changes in meaning" in the title of the paper. More prosaically, I think of statistical inference in economics as the methods used by economists to answer certain questions: How do we judge the extent to which the data we have support or refute a proposition about the world in general? How should we use data to estimate the economic quantities in which we are interested? And how do we assess the reliability of those estimates once we have made them?

I will also make occasional reference to "statistical theory," by which I mean the study of the properties of statistical measures calculated using sample data—e.g., the sample mean or a sample regression coefficient—as measures of characteristics *of that sample*. Statistical theory in this sense is distinct from both statistical inference and probability theory, although it has often been used in conjunction with probability theory in the process of statistical inference. This occurs, for example, when a researcher makes assumptions about the distributions of random variables in an unknown population and about a process by which samples are drawn from that population, then derives the properties of statistical measures calculated from the samples as estimators of the parameters of the assumed distributions. One purpose of this paper is to explore changes over time in economists' views of the role that probability theory should play in statistical inference.

There is currently a broad consensus among economists on what statistical inference means in economics. Today's graduate student learns that "statistical inference" is a set of procedures for analyzing data. One starts by making assumptions about the joint distribution of a number of random variables in a population of interest; the joint distribution is characterized by fixed parameters that embody important but unknown facts about some phenomenon of interest. Typically these assumptions are made to facilitate the use of one of a set of canonical "statistical models" with well-understood properties, such as the linear regression model.

The researcher is also assumed to have a random sample of observations of many of the relevant variables. At this point, statistical inference becomes a matter of applying formulas to this sample data. These formulas produce estimates of the parameters, and also measures of the reliability of those estimates—"standard errors" and so forth. The formulas have been derived from probability theory, and there is a set of standardized procedures—also justified by probability theory—for using the estimates to test hypotheses about the parameters.

To judge from econometrics textbooks published since the 1970s, "statistical inference in economics" means no more and no less than the application of these procedures. William Greene, in his popular graduate-level econometrics textbook, calls this approach "the classical theory of inference," and correctly notes that "the overwhelming majority of empirical study in economics is done in the classical framework" (Greene 2000, p. 97).[2]

But it has not always been thus. Prior to WWII, very few empirical economists in the US made any use of these tools of statistical inference when drawing conclusions

---

[2]Some modern books, including Greene (2000), briefly discuss an alternative Bayesian approach to inference, an approach likewise based in probability theory.

from statistical data, even though discussions of the foundational principles of this "classical theory of inference" could be found in standard textbooks on economic statistics in the 1920s, and important contributions to the classical theory were being made and disseminated to interested economists throughout the 1930s. I will begin by describing the assumptions and arguments upon which leading empirical economists of this period based their rejection of the classical theory of statistical inference: that is, the inferential procedures derived from, and justified by, probability theory.[3] In doing so, I will point out a similarity between the attitudes of the American empirical economists of the 1920s and 1930s towards statistical inference and those expressed by John Maynard Keynes in his *Treatise on Probability*. I argue that this similarity is partly due to influence, and describe a means by which Keynes's views came to shape the thought and practice of empirical economists in the US. I then turn to Trygve Haavelmo and the early Cowles Commission econometricians, and discuss how their arguments in favor of applying the classical theory of inference to economic data did and did not respond to the concerns of those empirical economists of the 1920s and 1930s who had rejected probability theory. Haavelmo and the Cowles Commission are widely seen as having launched a revolution in the practice of econometrics, and they did, but I do not think it is appreciated how long it took for their revolutionary approach to statistical inference to become the standard practice of the profession. I argue that statistical inference without probability was an important part of empirical economics at least into the 1960s. In a final section, I offer a summary characterization of the changes in the meaning and practice of statistical inference in economics from the pre-WWII decades to the last decades of the twentieth century, and speculate about the causes of these changes. I should also note that this essay focuses mainly on the ideas and activities of economists working in the United States; I have not ascertained whether the ideas and practices of empirical economists working outside the US developed in the same way.

II. STATISTICAL INFERENCE WITHOUT PROBABILITY: THE 1920S AND 1930S

Keynes's Treatise on Probability

Early in his career, Keynes had a significant interest in probability and statistics, leading to his 1921 *Treatise on Probability*.[4] In the part of the *Treatise* devoted to statistical inference, Keynes made the distinction between the descriptive function of the theory of statistics, which involved devising ways of representing and summarizing large amounts of data, and the inductive function, which "seeks to extend its descriptions of certain characteristics of observed events to the corresponding characteristics of other events that have not been observed." This inductive function of statistics he called the "theory of statistical inference," and he noted that it was currently

---

[3]What I report here is broadly consistent with Morgan's (1990, pp. 230–238) seminal account of the arguments underlying the rejection of probability theory during this period.
[4]Aldrich (2008) described the development of Keynes's ideas on probability over the period from 1908 to 1921.

understood to be "closely bound up with the theory of probability." The theory of probability provided the basis for the calculations of probable errors for statistical measures like the mean and the correlation coefficient, and it justified the use of these probable errors in making inferences about unobserved phenomena on the basis of descriptive statistics calculated for a sample (Keynes 1921, p. 371).

Keynes then spent several chapters explaining why this approach to statistical inference was an unsound basis for drawing conclusions from statistical data. As he stated at one point, "To apply these methods to material, unanalyzed in respect to the circumstances of its origin, and without reference to our general body of knowledge, merely on the basis of arithmetic and those characteristics of our material with which the methods of descriptive statistics are competent to deal, can only lead to error and delusion" (Keynes 1921, p. 384).

The quote hints at one of Keynes's central themes: that induction from statistical material, like all induction, required arguments by analogy. In the case of statistical induction, the key analogy was between the circumstances surrounding the generation of the data used to calculate descriptive statistics and the circumstances surrounding the phenomena about which one wished to draw conclusions. Assessing the extent of resemblance between these two sets of circumstances required knowledge beyond the sample, and the techniques of statistical inference based in probability theory offered no means for taking such knowledge into account.

Keynes accompanied his criticism with an outline of a more satisfactory approach to statistical induction, but it was a sketchy outline. At times, he seemed to be rejecting any role for probability theory in a separate "logic of statistical inference," and arguing that the process of drawing inferences from statistical data should essentially follow the logic of ordinary induction. Other times, he seemed to be pointing towards a new theory of statistical inference that synthesized his own theory of induction by analogy with probability-based statistical procedures based on the work of the German statistician Wilhelm Lexis. For my purposes, however, what Keynes's contemporaries believed him to be saying is more important than his intended message.

Keynes, Warren Persons, and the Statistics Texts of the 1920s

John Aldrich (2008) tells us that almost all the leading statistical theorists of the 1920s rejected Keynes's arguments concerning statistical inference, and Mary Morgan (1990) observes that the *Treatise* has seldom been cited by econometricians. As both Aldrich and Morgan note, however, one prominent American economist wholeheartedly embraced Keynes's message, and that was Warren Persons.

Warren Persons was a highly original and skilled economic statistician who made important contributions to the analysis of time series in the period just prior to the 1920s.[5] These contributions earned him the presidency of the American Statistical Association (ASA) in 1923, and in his Presidential Address, he spoke at length on statistical inference as it related to economic forecasting. One of Persons's messages was that "statistical probabilities provide no aid in arriving at a statistical inference." In defending that proposition, Persons borrowed many of Keynes's arguments, and quoted liberally from Keynes's *Treatise*.

---

[5]For more on Persons's contributions, see Morgan (1990, pp. 56–63).

Persons asserted that the time series data used in forecasting did not meet the conditions required for use of the theory of probability. Probability-based inference required a random sample of independent observations drawn from a well-defined population, but a time series could not be regarded as a random sample from a population in any but an "unreal, hypothetical sense," and the individual elements of a time series were not independent of one another. Further, the logic behind using inferential procedures based in probability theory for forecasting assumed that the forecasting period was essentially another random draw from the same population as the sample used to calculate the forecast. Yet, the forecaster typically knew enough about the forecasting period to know ways in which it differed from the period covered by his sample, information that the probability-based measures could not incorporate.

Persons also borrowed some of Keynes's arguments regarding the use of analogy in statistical induction: statistical results from time series data provided a stronger basis for inference if they were stable when calculated for different subperiods of the data, if they were consistent with results found in other periods and under different circumstances, and if the results "agree with, are supported by, or can be set in the framework of, related knowledge of a statistical or non-statistical nature" (Persons 1924, pp. 4–5).

As noted above, others have pointed out Persons's embrace of Keynes's views on probability and statistical inference. What I would like to suggest is that Keynes's arguments, as interpreted by Persons, helped to provide the intellectual justification for the attitude, widely held by leading empirical economists for over two decades, that "statistical inference" could and should be conducted without the apparatus of the classical theory of inference.

During the 1920s, there was a significant change in the way that economic researchers presented statistical data. In particular, there was a sizable jump in the percentage of empirical papers presenting summary statistics like means and correlation coefficients in addition to tables of numbers. This change coincided with a growing attention to training in statistical methods in American economics departments (Biddle 1999). Accompanying this statistical revolution was the publication, in 1924 and 1925, of first editions or new editions of four textbooks on statistical methods for economists. For the next several years, at least until the 1940s, the overwhelming majority of advanced economics students in the US were introduced to the subject by one of these books. And each one of them reflected Persons's view of Keynes's position on statistical inference.

The most prominent of the four was Frederick Mills's *Statistical Methods* (Mills 1924). Mills's views on statistical inference are worth looking at not only because of the popularity of his textbook, but also because from the mid-1920s to the late 1930s, Mills was essentially the in-house authority on statistical theory at the National Bureau of Economic Research, which sponsored a non-trivial fraction of the empirical research being done in the US prior to WWII.

Mills's text discussed the classical theory of statistical inference in a chapter entitled "Statistical Induction and the Problem of Sampling." He adopted Keynes's distinction between the descriptive and inductive functions of statistics, and, like Persons, regarded statistical inference as a synonym for statistical induction. Mills described the meaning of a representative sample, derived formulas for the probable errors of common statistical measures, and explained how those probable errors could measure the reliability of a sample statistic as an estimate of a characteristic of the population

from which the sample was drawn. But he then strongly cautioned his readers against using these measures, arguing that the circumstances that justified their use were "rarely, if ever," met in economic data, and that they could not account for the sampling biases and measurement errors that represented the most serious obstacles to reliable statistical inference. He recommended that instead of inferential techniques based on probability theory, the statistician use "actual statistical tests of stability," such as the study of successive samples and comparisons of descriptive statistics across subsamples of the population—the same sorts of inferential procedures recommended by Keynes and Persons.

The general message was the same, but the debt to Keynes and Persons more explicit, in Edmund Day's 1924 textbook, *Statistical Analysis*. Day asserted that the theory of probability was hardly applicable to most economic data, and that inferences should be based on "non-statistical tests of reasonableness" and the logic of analogy. In support of this, Day reproduced several paragraphs from Persons's ASA Presidential Address. William Leonard Crum, senior author of *An Introduction to the Methods of Economic Statistics* (Crum and Patton 1925), worked closely with Persons, and taught economic statistics at Harvard in the 1920s and 1930s. In a review of Day's book, Crum endorsed Day's discussion of inference (Crum 1926, p. 545), while, in his own text, he barely mentioned probability theory as an inferential tool. Finally, a second edition of Horace Secrist's *An Introduction to Statistical Methods* text appeared in 1925. He included new material on probable error measures, but argued, with the help of a quote from Persons, that these measures were largely inapplicable to economic data.

Statistical Inference in Agricultural Economics

Another source of economists' views on statistical inference in the 1920s and 1930s is the literature of agricultural economics. As is well known, agricultural economists were at the forefront of the statistical revolution of the 1920s, and a good indication of the opinions of the period's leading agricultural economists on a broad range of topics related to empirical research is provided by a two-volume report, "Research Method and Procedure in Agricultural Economics," sponsored by the Social Science Research Council and published in 1928. The report was assembled by four prominent agricultural economists, based on contributions from a number of active researchers.[6]

The editors' stance on statistical inference in economics was made clear early in the section of the report devoted to "statistical method": probability theory had little to contribute to statistical inference in economics, further development of the theory was not likely to soon change this situation, and the overreliance on mathematical formulas in analyzing statistical data was a mistake made by men of limited training and understanding (Advisory Committee 1928, p. 38).

The message that the data of agricultural economics did not meet the assumptions required for probability-based inference, and that statistical inference required both

---

[6]The editors were John Black, Edwin Griswold Nourse, Lewis Cecil Gray, and Howard R. Tolley. Although the view of leading agricultural economists on statistical inference, as reflected in the report, parallels in many ways the Keynes via Persons view described above, I have no evidence that these writers based their views on a reading of either Keynes or Persons.

a priori analysis and empirical knowledge beyond that provided by the sample, was repeated by other contributors to the report. The section on sampling explained that most samples in agricultural economics did not meet the definition of random samples, a stance that is noteworthy, given that agricultural economists were more likely than empirical researchers in any other field of economics to have had a hand in the collection of the samples with which they worked. The section of the report specifically devoted to statistical inference was written largely by Elmer J. Working, the economist sometimes credited with producing the first complete and clear analysis of the identification problem.[7] Working observed that statistical inference was sometimes narrowly defined as the process of generalizing from a sample to the universe from which the sample was drawn, with terms like 'statistical induction' or 'statistical logic' applied to the problem of generalizing from a sample to some more or less similar universe. Working saw little of value in this distinction, arguing that the universe of which an economist's sample was truly representative was seldom the universe about which the economist wished to draw inferences. For this reason, the classical formulas for probable errors were of limited usefulness.

How, then, could an economist determine whether a correlation found in a sample was likely to provide a reliable basis for inference about the relationship between variables in that universe that really interested him? Like Keynes and Persons, Working argued that the tests of stability created by Wilhelm Lexis had a role to play—the relationship should at least be stable within the sample if it was to be trusted as an indication of what might be true beyond the sample. But one could have even more faith in a generalization based on a sample correlation if that correlation represented a true cause-and-effect relationship, and much knowledge of which correlations represented causal relationships could be provided by economic theory.

Producing such sample correlations was a tricky business, however, as the causal systems underlying economic phenomena involved many factors. Ideally, the economist would know this system of causal relations, and his sample would allow him to use multiple correlation analysis to account for all the relevant variables in the system. But these ideal conditions were almost never realized. Thus, an important part of statistical inference was being able to judge both how close one was to this ideal and the effects of deviations from the ideal.

Working provided a concrete example of this approach to inference by analyzing a multiple regression model explaining the price of butter. Combining theoretical considerations with a test of within-sample stability, he concluded that the sample estimates did not represent stable cause-and-effect relationships. He cautioned, however, that this sort of inferential process was difficult to standardize. Each case presented unique problems that would tax "to the utmost" the mental powers of the economist.[8]

In a 1930 paper, Working made a number of these points again, and added an additional theme: his concern with a growing tendency among agricultural economists to substitute refined statistical methods for sound inferential procedures. Working endorsed the use of the most up-to-date multiple regression techniques in analyzing

---

[7]See Morgan (1990, pp. 170ff.) for a discussion of this claim.
[8]Working's section on inference is found on pages 270ff. in Advisory Committee (1928). The quote is on page 281.

sample data, but believed that often those using refined techniques tended to be less careful and thorough in making inferences from their data, relying solely on statistical analysis rather than logical and theoretical analysis. Working suggested that this was because the greater precision and sophistication of the methods engendered greater confidence in the measurements they produced, while at the same time making it more difficult to understand exactly what those measures represented. That some of the refined statistical techniques that Working saw being substituted for "reasoned inference" were the tools of the classical theory of inference is clear from his comment that "the fact that the theory of sampling gives no valid ground for expecting statistical relationships of the past to hold true in the future has been pointed out repeatedly. Nevertheless, it appears that there continues to be widespread misunderstanding and disregard of this fact" (Working 1930, esp. pp. 122–125; the quote is from p. 124).[9]

The 1930s

The 1930s saw further developments of the classical theory of inference, and the discussion of these developments in the journals read by economists (Biddle 1990). Morgan (1990) identifies a handful of influential economists in the 1930s who were enthusiastic about applying the classical methods of inference to economic data, including Tjalling Koopmans, whose dissertation reinterpreted Ragnar Frisch's confluence analysis in a way that made it amenable to the application of the classical theory of inference, and Harold Hotelling, who, throughout the 1930s, taught mathematical statistics to economics PhD students at Columbia. She also notes that this sort of work had little noticeable impact on the literature in the 1930s.

John Aldrich identifies Henry Schultz as one American economist who was willing to use probability-based inference in the 1930s, but this observation requires some qualification. Schultz, who worked primarily on the estimation of demand curves, was recognized as an authority on the application of regression techniques to time series data, and was the leading economic statistician at the University of Chicago from the late 1920s until his death in 1938. Although Schultz understood the classical theory of inference, measures derived from the theory played only a minor role in his attempts to draw conclusions from his estimates. His main strategy was to estimate several specifications of a demand curve, using different methods to control for the trend, different functional forms, and so forth. Then, using a variety of economic and statistical considerations, he would distill from these results a best estimate of the true demand curve.[10]

In the 1930s, Schultz began a project of statistically testing the restrictions implied by the neoclassical theory of the consumer, and he used the standard errors of the coefficients of his estimated demand curves as a metric for determining whether the relationships between the coefficients were close enough to those implied by the theory to count as a confirmation of the theory (Schultz 1938, chs. XVIII, XIX). And, in the 1938 book summarizing his life's work, he included an explanation of the fundamentals of the classical theory of hypothesis testing (Schultz 1938, pp. 732–734).

---

[9]Working continued to make these arguments throughout the 1930s. See Working (1932, 1939).
[10]See, e.g., Schultz (1925, pp. 606, 627–628).

However, the same 1938 book includes a discussion of standard errors and significance testing that reveals serious reservations about the usefulness of the classical theory of inference in the analysis of time series data. Schultz argued that time series could not be considered as samples of a definable universe in any reasonable sense. Also, there were typically changing patterns of dependence between successive observations in a time series. So, standard errors derived from time series, Schultz wrote, "do NOT [and the emphasis is his] tell us the probability that the true value of a given parameter will be found between such and such limit." But that, of course, was precisely the question that the classical theory of inference had been developed to answer. Instead, Schultz said, standard errors calculated for coefficients of a time series regression should be regarded as only one measure of goodness of fit (Schultz 1938, pp. 214–215).

In 1938, Frederick Mills issued a revised edition of his textbook (Mills 1938). He noted in the preface that in the years since his last book, statistics had experienced "ferment and creative activity similar to that associated with Karl Pearson and his associates," including the development of improved methods for testing hypotheses and a "sounder foundation" for statistical inference. As a result, the material on inferential methods derived from probability theory had been expanded (Mills 1938, pp. vii–viii).

Although Mills's revised chapter, "Statistical Induction and the Problem of Sampling," now emphasized the importance of the classical measures of reliability, he did not tone down his warnings about the use of these measures to any significant extent.[11] Still, students using Mills's textbook were shown how to conduct Ronald Fisher's analysis of variance, calculate standard errors for regression coefficients, and conduct T-tests when dealing with small samples. Mills himself used classical inferential methods in a 1926 paper that proposed and tested a hypothesis regarding the cause of a secular change in the duration of business cycles, although he was careful to note that he was testing only the likelihood that observed changes in cycle duration were due to "sampling fluctuation alone" (Mills 1926).

Both Mills and Schultz were aware of, and concerned about, the stringent data requirements for the valid application of the classical inferential measures. However, they were also committed to making economics more like the natural sciences, in which theoretical hypotheses were tested empirically. One can conjecture that their enthusiasm for the potential of the classical theory of inference to provide economists with a more precise way of determining the extent to which statistical evidence supported theoretical assertions occasionally pushed them to "damn the torpedoes" and go full speed ahead with the classical inferential methods.

One goal of this section has been to establish that the era of "statistical inference without probability" preceding WWII was not rooted in an unreflective rejection of probability theory, and neither was it the result of a naiveté about the problems created for empirical research by the inexact nature of economic theory. Instead, leading empirical economists articulated a sophisticated, reasonable case against drawing conclusions

---

[11]Mills now made his claim that the conditions assumed in deriving the classical measures were "rarely" rather than "rarely, if ever," met in economic data, and that "some degree of departure" from the ideal conditions for the use of the measures did not rob them of value. However, there was new material warning about the dangers of using the measures with time series data (Mills 1938, pp. 485–487).

from economic data using inferential procedures based on probability theory. But this case did not carry the day, and statistical inference in economics came to mean the use of the classical theory of inference. I now take up that part of the story.

III. HAAVELMO, COWLES, AND THE ARGUMENT FOR USING PROBABILITY THEORY

There is wide agreement that the work of Trygve Haavelmo and the Cowles Commission econometricians in the 1940s and early 1950s was crucial to the eventual acceptance by economists of the classical theory of statistical inference.[12] The Cowles group's justification of the application of probability theory to economic data is found in Haavelmo's (1944) essay, "The Probability Approach in Econometrics."[13] In the introduction, Haavelmo acknowledged, and promised to refute, the prevalent opinion that applying probability models was a "crime in economic research" and "a violation of the very nature of economic data" (Haavelmo 1944, p. iii). At the center of his refutation was an ingenious reconceptualization of the inferential problem facing economists.

The leading statistical economists of the 1920s discussed the classical theory of inference in terms of using sample data to draw conclusions about a universe or population from which the sample had been drawn, and they were dubious about attempts to fit time series data into this "sample and universe" conceptual scheme. When it came to time series data, there did not seem to be any good answer to the question "What is the universe from which this is a sample?" Haavelmo deemed this way of thinking about the classical theory to be too narrow. In its place, he proposed the idea of a time series as a set of observations brought about by a mechanism capable of generating an infinity of observations. The mechanism could be characterized by a probability law, and the task of statistical inference was to learn the characteristics of this probability law (Haavelmo 1944, pp. iii, 48). This provided, I think, a compelling answer to the objection that a time series was not representative of any definable universe. However, I am going to argue that neither Haavelmo's essay nor the empirical methods used by the Cowles econometricians provided convincing answers for other major elements of the pre-existing case against applying the classical theory of inference to economic data.

As is well known, the Cowles econometricians typically assumed that the probability law governing the mechanism that produced their data could be represented as a set of linear equations, each one with a "systematic" part describing relationships between variables observed in the data, and an unobserved disturbance term. It was also necessary to make assumptions about the joint distribution of the disturbance

---

[12]When I refer to the Cowles econometricians, I mean both those econometricians formally affiliated with the Cowles Commission in the 1940s and 1950s, and those who were clearly working on the research program laid out in Haavelmo's (1944) essay and in the Cowles econometric monographs.

[13]As Hoover (2014) points out, the Cowles Commission econometricians took the correctness of Haavelmo's arguments on this score as a general background assumption of their entire endeavor, and had little to say about them.

terms and the observable variables.[14] One drew on economic theory and institutional knowledge to specify the systemic part of the assumed probability law. Ideally, the distributional assumptions regarding the relationships between the observable and unobservable variables also rested on theory and prior knowledge, but often those assumptions were made primarily to facilitate the application of the classical inferential procedures.

The Cowles approach to statistical inference involved two distinguishable aspects. One was the unbiased estimation of the parameters of the systematic part of the assumed probability law—'the structural parameters,' as they came to be called. But estimation of structural parameters produced, as an important by-product, estimates of the unobserved disturbances. These estimated disturbances were needed to estimate the variances of the structural parameter estimates, which were in turn necessary for hypothesis testing and the construction of confidence intervals.

I detail these two aspects because each required a different type of assumption to succeed. The proof that a proposed estimator provided unbiased estimates of the structural parameters required the assumption that the unobserved disturbances were uncorrelated with at least a subset of the variables observed in the data. The proof that an estimated variance covariance matrix of the parameter estimates was unbiased required an assumption that the unobserved disturbances were independently and identically distributed, or, alternatively, that any deviations from independent and identical distribution could be expressed in a relatively simple mathematical model.

Finally, an assumption underlying it all was that the econometrician's data could be thought of as a random sample—either a random sample of an identifiable population of interest, or, in terms of Haavelmo's reconceptualization, a random draw from the infinity of sample points that could be generated by the economic mechanism in which the economist was interested.

Consider now the relationship between these assumptions necessary to the Cowles program of statistical inference and the pre-existing arguments against the application of the classical theory of inference to economic data. Note first that Elmer Working's attitude towards the role of theory in empirical research was similar to that of Haavelmo and Koopmans. All agreed that the goal of empirical analysis in economics should be to measure those correlations in the data that represent true cause-and-effect relationships, that economic theory helped to identify which partial correlations represented such causal relationships, and that the systems of causal relationships underlying economic data were complex and involved many variables. But Working emphasized that the available data were unlikely to include the full set of variables involved in the relevant causal system, and that those unobserved variables were probably correlated with the observed variables included in the regression, thus biasing the resulting parameter estimates. The classical inferential methods employed by the Cowles

---

[14]As Arthur Goldberger's influential econometrics text explained: "The models of econometric theory will have three basic components: a specification of the process by which certain observed 'independent variables' are generated, a specification of the process by which unobserved 'disturbances' are generated, and a specification of the relationship connecting these to observed 'dependent variables.' Taken together these components provide a model of the generation of economic observations, and implicitly, at least, they define a statistical population. A given body of economic observations may then be viewed as a sample from the population" (Goldberger 1963, p. 4).

Commission did not solve this problem. Instead, application of those methods required an assumption that the problem did not exist.[15] The Cowles econometricians were willing to make this assumption. Working was not. Instead, he argued that thinking about the consequences of the failure of this assumption was an important part of statistical inference.

Another common pre-war objection to the use of probability models with economic data was that successive observations in a time series were not independent. The Cowles econometricians offered a strategy for solving this problem, which was to propose and estimate a parametric model of the dependence between observations, then use the parameter estimates to create a series of estimated disturbances that were not serially correlated. This was an ingenious strategy, but there is reason to believe that the statistical economists of the 1920s and 1930s would not have been satisfied with it. For example, in discussing the problem of serial correlation, Henry Schultz (1938, p. 215) commented that successive items in a time series "generally occur in sequences of rises and falls which do not repeat one another exactly." In other words, the dependence between observations could not be captured by a parsimoniously parameterized model.

As discussed above, the leading statistical economists of the 1920s and 1930s were also unwilling to assume that any sample they might have was representative of the universe in which they were interested. This was particularly true of time series, and Haavelmo's proposal to think of time series as a random selection of the output of a stable mechanism did not really address one of their concerns: that the structure of the "mechanism" could not be expected to remain stable for long periods of time. As Schultz pithily put it, "'the universe' of our time series does not 'stay put'" (Schultz 1938, p. 215). And, the belief that samples were unlikely to be representative of the universe in which the economists had an interest applied to cross-section data as well. The Cowles econometricians offered little to assuage these concerns.

It is not my purpose to argue that the economists who rejected the classical theory of inference had better arguments than did the Cowles econometricians, or had a better approach to analyzing economic data, given the nature of those data, the analytical tools available, and the potential for further development of those tools. I offer this account of the differences between the Cowles econometricians and the previously dominant professional opinion on appropriate methods of statistical inference as an example of a phenomenon that is not uncommon in the history of economics. Revolutions in economics, or "turns," to use a currently more popular term, typically involve new concepts and methods. But they also often involve a willingness to employ assumptions considered by most economists at the time to be inadvisable, a willingness that arises because the assumptions allow progress to be made with the new concepts and methods.[16] Obviously, in the decades after

---

[15]Arguably, Haavelmo was much more sensitive to this potential problem than the Cowles econometricians who followed him. See Boumans (2015, ch. 4).

[16]As an early commentator on Cowles econometrics noted, "anyone using these techniques must find himself appealing at every stage less to what theory is saying to him than to what solvability requirements demand of him. Certain it is that the empirical work of this school yields numerous instances in which open questions of economics are resolved in a way that saves a mathematical theorem. Still, there are doubtless many who will be prepared to make the assumptions required by this theory on pragmatic grounds" (Hastay 1951, p. 389).

Haavelmo's essay on the probability approach, there was a significant change in the list of assumptions about economic data that economists were routinely willing to make in order to facilitate empirical research. Explaining the reasons for this shift is, I think, a worthwhile task for historians of economics, and at the end of this paper I offer some conjectures that I hope will contribute to that task. However, I now turn to a question of timing: How long did it take for the widespread belief that methods associated with the classical theory of inference were unimportant, if not inappropriate, tools for statistical inference in economics to be replaced by a consensus that the methods associated with the classical theory of inference were not only appropriate, but were the primary methods for designing estimation procedures, assessing the reliability of estimates, and testing hypotheses? The evidence I present below suggests that this process took twenty years or more.

IV. STATISTICAL INFERENCE WITH AND WITHOUT PROBABILITY: PEDAGOGY AND PRACTICE 1945–1965

The Early Textbooks on "Econometrics"

Prior to the early 1950s, there were, at least in the opinion of those enthusiastic about the Cowles econometric program, no adequate textbooks teaching the new econometrics. No one was teaching the material in any way to graduate students at Harvard or Yale; interested Harvard students studied Cowles monographs on their own.[17] The first two textbooks teaching the new econometrics, including the use of techniques of inference based on the classical theory, were Gerhard Tintner's *Econometrics,* published in 1952, and Lawrence Klein's *Textbook of Econometrics*, which appeared the following year. Both Klein and Tintner had been affiliated with the Cowles Commission in the 1940s.

As I have noted, prominent empirical economists of the 1920s and 1930s considered and rejected the idea that the phrase "statistical inference" referred specifically to the use of inferential methods derived from probability theory, with some other phrase, like "statistical induction," to be applied to other approaches to drawing conclusions from statistical data. However, it is clear from the first pages of "The Probability Approach" (1944) that Haavelmo accepted the now-standard identification of the phrase "statistical inference" with applications of probability theory, and both Tintner and Klein did likewise in their books.

Klein's section on statistical inference—which he defined as "a method of inferring population characteristics on the basis of samples of information"—was a review of the basics of the classical theory of inference. The concept of the unbiased, minimum variance estimator was offered as an ideal. The construction of confidence intervals using the sample variance of an estimator was explained, as were the fundamentals of Neyman–Pearson hypothesis testing. In subsequent chapters, readers were shown how to apply these tools when estimating linear regression and simultaneous equations models.

---

[17]On the lack of good textbooks on the new econometrics prior to 1952, see, e.g., Carter (1953) or Solow (1953). The observation about Harvard and Yale is from Shiller and Tobin (1999). Klein reports that Cowles essays were also circulated among students at MIT in the 1940s (Mariano 1987).

The reliability of forecasts from time series models was discussed almost entirely in terms of sampling error and the variance of the regression residual, although it was allowed that "structural change" might cause a forecast to be inaccurate, and that "a well rounded econometric approach to forecasting" would take outside information into account (Klein 1953b, pp. 50–56, 252–264). Tintner also taught students how to apply the classical inferential tools in the context of regression analysis, although, somewhat surprisingly, the estimation of a simultaneous equations system was not discussed.

After a lull of about a decade, there appeared another set of econometrics texts destined for wide adoption, books by Arthur Goldberger (1963), Jack Johnston (1963), and Edmond Malinvaud (1966). In these books, which in many ways would set the pattern for subsequent twentieth-century graduate econometrics texts, statistical inference in economics meant an application of the classical theory of inference.[18]

Both Malinvaud and Goldberger presented Haavelmo's conceptualization of the sample-population relationship underlying statistical inference in terms of learning about the laws governing a stochastic process that generated the sample data. The econometrician's stochastic model, which was derived from economic theory, reflected the basic form taken by these general laws. Malinvaud explained that statistical inference "proper" never questioned this model, but that in practice, economic theory lacked the precision to provide the econometrician with a framework he could rely upon—more than one model framework might be plausible. Typically, there was no "strictly inductive procedure" for choosing among models in such cases. It was unfortunate but true that the choice would have to be based on "an analysis in which our empirical knowledge will be less rigorously taken into account": for example, applying the inferential tools to many different data sets to eliminate frameworks that did not seem to agree with the facts (Malinvaud 1966, pp. 64–65, 135).

Like Malinvaud, Goldberger argued that statistical inference proper began after the model was specified; it was only during model specification that a priori assumptions or information beyond the sample played a role in empirical analysis. Note the contrast between this conception of how theory and background information should be integrated into empirical analysis and Elmer Working's view of the process of statistical inference. Like Goldberger, Working believed researchers should rely on theory in determining the relationships to be estimated, but Working saw a further role for theory and background knowledge after estimation, as an aid to interpreting the results and to diagnosing whether the estimates actually reflected stable cause-and-effect relationships.

Graduate Instruction in Econometrics

I believe that the early 1960s, in part due to the appearance of these textbooks, mark something of a turning point in the pedagogy of statistics/econometrics, after which graduate students in economics would routinely be taught to understand statistical estimation and inference in terms of the classical theory, whether in the context of a Cowles-style presentation of simultaneity, identification, and so forth, or simply the

---

[18]Goldberger (1963, p. 125); Malinvaud (1966, p. 67).

more prosaic instruction in testing regression coefficients for "statistical significance" and in constructing confidence intervals.[19]

In the 1950s, however, the amount of systematic instruction in the application of the classical theory of inference available to economics graduate students depended on where they were being trained. For example, Frederick Mills was still teaching statistics to the Columbia graduate students in the 1950s, and it was not until 1962 that a departmental committee at Columbia recommended that econometrics be offered as a field for graduate students (Rutherford 2004).

At the University of Wisconsin, graduate students in the 1940s were taught statistics by professors from the university's School of Commerce in a fashion that Milton Friedman described as "inadequate" in an evaluation requested by a Wisconsin faculty member (Lampman 1993, p. 118). In 1947, the department hired Martin Bronfenbrenner, a student of Henry Schultz, with the expectation that he would teach statistics to graduate students. Bronfenbrenner appears to have taught the course only on an irregular basis during the 1950s, and a look at his own research would suggest that he offered students instruction in the classical inferential methods as one of many aids to statistical inference, probably not the most important.[20] The picture at Wisconsin changed dramatically with the arrival of Guy Orcutt in 1958 and Arthur Goldberger in 1959, both strong advocates of the use of classical theory of inference (Lampman 1993, pp. 131, 171, 229). In the mid-1950s, before he came to Wisconsin, Goldberger was teaching econometrics to graduate students at Stanford, and Goldberger himself was trained in the early 1950s at the University of Michigan by Lawrence Klein (Lodewijks 2005; Kiefer 1989).

At the University of Chicago, home of the Cowles Commission until 1955, the situation was complicated. Students could learn Cowles-style econometrics from teachers like Henri Thiel, Haavelmo, and Zvi Griliches. But these methods were eschewed for actual empirical work. Among Griliches's students in the early 1960s was future econometrician Gangadharrao S. Maddala, who considered writing a dissertation in econometric theory before opting for a more empirically oriented topic. Maddala reports that neither he nor any other student in his cohort who was writing an empirical dissertation used anything more complicated than ordinary least squares regression (Krueger and Taylor 2000; Lahiri 1999).

Statistical Inference without Probability in the 1950s and 1960s: Some Important Examples

Another type of evidence on the timing of the emergence of a consensus behind the classical theory of inference is negative in nature: that is, examples of important

---

[19]By the 1960s, the question of whether simultaneous equations techniques produced more reliable estimates than equation-by-equation application of ordinary least squares was a topic of debate (e.g., Christ 1960). However, most empirical economists who turned away from the "simultaneous equations" part of the Cowles econometric program remained committed to the classical theory of inference.

[20]On the sporadic nature of Bronfenbrenner's teaching of econometrics, see Lampman (1993, pp. 96, 131). Regarding what Bronfenbrenner likely taught about statistical inference, in a survey of the failure of postwar economic forecasts, he commented, "If time and staff are available, econometric analysis, with consumption functions playing a prominent role, is worth attempting, if only as a rough check on the consistency of whatever other estimates are made" Bronfenbrenner (1948, pp. 319–320).

empirical research studies of the 1950s and early 1960s that did not involve the classical tools of inference in any important way. The researchers involved were assessing the reliability of estimates, were drawing conclusions from data, were asserting and testing hypotheses, but they were not using the tools of the classical theory of inference to do so.

In 1957, Milton Friedman published his theory of the consumption function. Friedman certainly understood statistical theory and probability theory as well as anyone in the profession in the 1950s, and he used statistical theory to derive testable hypotheses from his economic model. But one will search his book almost in vain for applications of the classical methods of inference.[21] Six years later, Friedman and Anna Schwartz published their *Monetary History of the United States*, a work packed with graphs and tables of statistical data, and generalizations based on that data. But the book contains no classical hypothesis tests, no confidence intervals, and no reports of statistical significance.

A colleague of Friedman's at Chicago, and a student of Henry Schultz, H. Gregg Lewis was a member of the research staff of the Cowles Commission in the early 1940s. During the 1950s, Lewis worked with a number of students writing dissertations on the impact of labor unions on workers' wages, measuring what came to be known as the 'union/non-union wage gap.' In the late 1950s, he set out to reconcile the wage-gap estimates from these and other studies. The resulting book, *Unionism and Relative Wages in the United States*, appeared in 1963. Using basic neoclassical reasoning and the statistical theory of regression, Lewis attempted to determine when differences in measured wage gaps were due to biases related to differences of data and method, and when they could be linked to economic factors. Lewis made little use of standard errors or significance testing in drawing his conclusions. For example, he would sometimes report a range for the likely value of the wage gap in a particular time period or labor market, but these ranges were not the confidence intervals of classical inference theory. Instead, they were Lewis's judgments of the likely size of biases due to inadequacies of the data. Sampling error, the source of imprecision in estimation with which classical inferential methods were designed to deal, was hardly mentioned.

The human capital research program was also born at the University of Chicago in the late 1950s and early 1960s, with the key contributions being made by Theodore Schultz, Gary Becker, and Jacob Mincer. Theodore Schultz was the veteran of the three, and although he had written one of the early papers in the economics literature explaining Fisher-style hypothesis testing (Schultz 1933), he had since lost confidence in more complex empirical methods, and used cross-tabulations and estimation techniques borrowed from national income accounting to measure key human capital concepts and to establish general patterns. The young Gary Becker's 1965 book, *Human Capital,* contained a fair amount of statistical data, some meant to test hypotheses derived from the book's theoretical models. But Becker did not use the tools of classical inference theory. The empirical innovator of the three was Jacob Mincer, who, among other things, developed the "earnings regression" approach to measuring the

---

[21]Teira (2007) covers Friedman's familiarity with probability theory and classical methods of inference, and offers an explanation of his non-use of these methods.

rate of return to education. Mincer used probability theory as a modeling tool in his work on human capital theory. But the tools of classical inference theory played little role in his early empirical work on human capital, or, for that matter, in his 1974 magnum opus, *Schooling, Experience, and Earnings* (Mincer 1958, 1962, 1974).

Statistical inference without probability was not just a Chicago phenomenon in this period. Economists trained in the final decades of the twentieth century are familiar with Okun's Law, an empirical generalization concerning the relationship between changes in unemployment and changes in the GNP. Arthur Okun, a policy-oriented macroeconomist, first presented the law and the empirical evidence behind it in 1962, and it was quickly accepted as a basic macroeconomic fact. The important point for my purpose, however, is the nature of the evidence he presented: results from three alternative specifications of a regression involving unemployment and GNP. There were no confidence intervals reported for the regression coefficient(s) that embodied the law, and neither was their "statistical significance" mentioned.[22]

Then there is Simon Kuznets. In his 1949 Presidential Address to the American Statistical Association, Kuznets emphasized that the data used by social scientists were generated by social processes rather than by controlled experiments, and he discussed which methods were and were not well suited to the analysis of such data. And clearly on Kuznets's "not well suited" list were the classical theory of inference and the Cowles Commission style of econometrics.

For example, Kuznets argued that properties of measurement errors in social data differed across data sets and across variables within data sets, so they could not be dealt with by any "broad and standard scheme." At a minimum, one needed particular knowledge of the process behind each measurement. Further, recently developed inferential methods derived from probability theory presupposed that the analyst understood, and could control for, systematic patterns in the data. Yet, it was only through the analysis of data that we could gain knowledge of these patterns, or develop hypotheses regarding them. It was tempting, but inappropriate, to assume such knowledge at the beginning of the analysis in order to employ "the elegant apparatus developed for an entirely different situation" (Kuznets 1950, p. 8). Kuznets said the same thing more pointedly a year later, arguing that statistical analysis in economics should be "divorced from the rather mechanistic approach that employs mathematical models as hasty simulacra of controlled experimentation (which they are not)" (Kuznets 1951, p. 277).

At this time, Kuznets was beginning a major empirical research program on the causes of economic growth. Over the next two decades, he would analyze a great deal of statistical data, proposing hypotheses and drawing conclusions with varying degrees of confidence, including his conjecture of an inverted u-shaped relationship between a nation's level of economic development and its level of income inequality, which became known as the "Kuznets curve." And his analysis did not involve, and neither did his conclusions rest on, classical hypothesis tests.

My final example of an important empirical research program that made no use of the classical theory of inference is Wassily Leontief's input-output analysis. During the 1940s and 1950s, there was great enthusiasm for input-output analysis among young, mathematically oriented economists who were rising to positions of prominence in the

---

[22]A look at Okun (1957) shows that this was not atypical of his empirical work.

profession, partly because Leontief's work provided a template for using theoretical models expressed as linear simultaneous equation systems as tools for empirical analysis and forecasting.[23] But, as he made clear in the early 1950s, Leontief differed from many of his admirers on one point: Cowles-style econometrics and the classical theory of inference were not an important part of his vision of the best way forward for an empirically based economics.

To implement his model empirically, Leontief first needed to assign empirical values to its parameters, the input-output coefficients. He described his approach to this estimation task as "direct observation," contrasting it with what he called the "indirect statistical inference" employed by "the modern school of statistical econometricians" (Leontief 1952, pp. 2–3, 7). He argued that the versions of the classical inferential techniques developed by the econometricians simply could not do the tasks being asked of them. In the search for the large samples required to make their methods valid, they had been driven towards what he called the "treacherous shoals of time series analysis," where, in order to get a time series long enough to permit valid inference in the presence of autocorrelation, they were forced to assume "invariance in relationships which actually do change . . . over time." The work of the modern econometricians was original and imaginative, but not the best use of the profession's resources (Leontief 1952, pp. 2–3).

Leontief used his empirically specified model to estimate unknown price and quantity relationships and to forecast the effects of hypothetical economic changes, so he still faced the inferential problem of assessing the reliability of such estimates and forecasts, or, more correctly, of the empirical model that had produced them. He and his followers developed various techniques for doing so, but none were based in probability theory (e.g., Leontief 1951, pp. 216–218; Holzman in Leontief 1953).

Cowles econometricians predictably took issue with Leontief's rejection of "indirect inference." Both Lawrence Klein (1953a) and Leonid Hurwicz (1955) reviewed the collection of input-output studies that included Leontief's essay quoted above, and both found much to praise. However, Klein warned that without the use of probability theory to provide "scientific bases for separating good results from bad, … the path of empirical research in input-output study was dangerously close to the realm of arbitrary personal judgment" (Klein 1953a, p. 262). Hurwicz also worried that, lacking the discipline imposed by the classical methods, the inferences drawn by the input-output researchers were prone to be influenced by the "author's state of mind," and he hoped for research that would reveal complementarities between Leontief's empirical methods and those of the econometricians (Hurwicz 1955, pp. 635–636).

As it happens, input-output analysis has remained a form of empirical research with little connection to the classical theory of inference, although it is no longer a research tool employed at the frontiers of the profession (Biddle and Hamermesh 2016). However, if you look at the post-1970 empirical research that built on the various other examples of inference without probability that I have offered, you will find classical hypothesis tests and indications of statistical significance taking a prominent place in the discussion of the empirical results.

---

[23]See, for example, the introduction and first part of Cowles Monograph 13 (Koopmans 1951).

V. DISCUSSION AND SPECULATION

By the 1970s, there was a broad consensus in the profession that inferential methods justified by probability theory—methods of producing estimates, of assessing the reliability of those estimates, and of testing hypotheses—were not only applicable to economic data, but were a necessary part of almost any attempt to generalize on the basis of economic data. I am far from being able to explain why this consensus, which ran counter to the strongly held opinions of the top empirical economists of thirty years earlier, developed when it did. But I will offer a few conjectures, which I hope seem worthy of some further thought and investigation. In doing so, I am going to make use of the concept of mechanical objectivity, introduced by Lorraine Daston and Peter Galison (1992) and fruitfully applied to the history of quantification in the social sciences by Theodore Porter in his book *Trust in Numbers* (1995).

This paper has been concerned with beliefs and practices of economists who wanted to use samples of statistical data as a basis for drawing conclusions about what was true, or probably true, in the world beyond the sample. In this setting, "mechanical objectivity" means employing a set of explicit and detailed rules and procedures to produce conclusions that are objective in the sense that if many different people took the same statistical information, and followed the same rules, they would come to exactly the same conclusions. The trustworthiness of the conclusion depends on the quality of the method. The classical theory of inference is a prime example of this sort of mechanical objectivity.

Porter contrasts mechanical objectivity with an objectivity based on the "expert judgment" of those who analyze data. Expertise is acquired through a sanctioned training process, enhanced by experience, and displayed through a record of work meeting the approval of other experts. One's faith in the analyst's conclusions depends on one's assessment of the quality of his disciplinary expertise and his commitment to the ideal of scientific objectivity. Elmer Working's method of determining whether measured correlations represented true cause-and-effect relationships involved a good amount of expert judgment. So, too, did Gregg Lewis's adjustments of the various estimates of the union/non-union wage gap, in light of problems with the data and peculiarities of the times and markets from which they came. Keynes and Persons pushed for a definition of statistical inference that incorporated space for the exercise of expert judgment; what Arthur Goldberger and Lawrence Klein referred to as 'statistical inference' had no explicit place for expert judgment.

Speaking in these terms, I would say that in the 1920s and 1930s, empirical economists explicitly acknowledged the need for expert judgment in making statistical inferences. At the same time, mechanical objectivity was valued—there are many examples of economists of that period employing rule-oriented, replicable procedures for drawing conclusions from economic data. The rejection of the classical theory of inference during this period was simply a rejection of one particular means for achieving mechanical objectivity. By the 1970s, however, this one type of mechanical objectivity had become an almost required part of the process of drawing conclusions from economic data, and was taught to every economics graduate student.

Porter emphasizes the tension between the desire for mechanically objective methods and the belief in the importance of expert judgment in interpreting statistical evidence. This tension can certainly be seen in economists' writings on statistical

inference throughout the twentieth century. However, it would be wrong to characterize what happened to statistical inference between the 1940s and the 1970s as a displacement of procedures requiring expert judgment by mechanically objective procedures. In the econometric textbooks published after 1960, explicit instruction on statistical inference was largely limited to instruction in the mechanically objective procedures of the classical theory of inference. It was understood, however, that expert judgment was still an important part of empirical economic analysis, particularly in the specification of the models to be estimated. But the disciplinary knowledge needed for this task was to be taught in other classes, using other textbooks.

And in practice, even after the statistical model had been chosen, the estimates and standard errors calculated, and the hypothesis tests conducted, there was still room to exercise a fair amount of judgment before drawing conclusions from the statistical results. Indeed, as Marcel Boumans (2015, pp. 84–85) emphasizes, no procedure for drawing conclusions from data, no matter how algorithmic or rule bound, can dispense entirely with the need for expert judgment. This fact, though largely unacknowledged in the post-1960s econometrics textbooks, would not be denied or decried by empirical economists of the 1970s or today.

This does not mean, however, that the widespread embrace of the classical theory of inference was simply a change in rhetoric. When application of classical inferential procedures became a necessary part of economists' analyses of statistical data, the results of applying those procedures came to act as constraints on the set of claims that a researcher could credibly make to his peers on the basis of that data. For example, if a regression analysis of sample data yielded a large and positive partial correlation, but the correlation was not "statistically significant," it would simply not be accepted as evidence that the "population" correlation was positive. If estimation of a statistical model produced a significant estimate of a relationship between two variables, but a statistical test led to rejection of an assumption required for the model to produce unbiased estimates, the evidence of a relationship would be heavily discounted.[24]

So, as we consider the emergence of the post-1970s consensus on how to draw conclusions from samples of statistical data, there are arguably two things to be explained. First, how did it come about that using a mechanically objective procedure to generalize on the basis of statistical measures went from being a choice determined by the preferences of the analyst to a professional requirement, one that had real consequences for what economists would and would not assert on the basis of a body of statistical evidence? Second, why was it the classical theory of inference that became the required form of mechanical objectivity?

With respect to the first question, Daston and Galison (1982), in discussing the shift to mechanical objectivity, emphasize concerns within the community of scientists

---

[24]One could say that the standard errors and hypothesis tests of classical theory came to provide threshold conditions for using sample data to make an argument about phenomena beyond the sample. For example, if an estimated sample relationship was not statistically significant, it could not be used to make a claim about the relationship outside the sample. If the sample relationship was significant, a threshold was cleared, and the relationship could be offered as possible evidence of the population relationship of interest. However, it could still be argued that the evidence was not credible, due, for example, to the failure of the data to satisfy the assumptions required for the estimation procedure to produce consistent estimates.

regarding their own abilities to produce unbiased interpretations of empirical data. Porter makes the point that an embrace of mechanical objectivity has often been the response of a weak discipline to external pressures, as when outside audiences whom the discipline wishes to influence doubt the discipline's claims to expertise and objectivity (Porter 1995, p. ix). Use of a mechanically objective procedure, particularly one endorsed by other more respected disciplines, can help overcome such doubts.

It does not seem to me that a narrative emphasizing either of these factors will fit the facts surrounding economists' embrace of the classical theory of inference. As I have mentioned, since the 1920s, individual empirical economists have been attracted to mechanically objective procedures, and this is in part because they offered insurance against biased interpretations due to the analyst's "personal equation." However, the intensity of concern over such bias differed from economist to economist, and I see no reason to think that the prevalence of this concern increased significantly in the 1950s and 1960s. Neither am I aware of important external patrons of economics who might have developed a concern, during the late 1950s and early 1960s, that the empirically based claims of economists could not be trusted if they were not produced using methods approved by the wider scientific community, such as classical hypothesis tests. Indeed, the conventional wisdom is that the public prestige of economists was at a high point in the early 1960s, just when the use of classical inference methods was becoming *de rigueur* in empirical research.

Perhaps searching for an explanation that focuses on the classical theory of inference as a means of achieving mechanical objectivity emphasizes the wrong characteristic of that theory. In contrast to earlier forms of mechanical objectivity used by economists, such as standardized methods of time series decomposition employed since the 1920s, the classical theory of inference is derived from, and justified by, a body of formal mathematics with impeccable credentials: modern probability theory. During a period when the value placed on mathematical expression in economics was increasing, it may have been this feature of the classical theory of inference that increased its perceived value enough to overwhelm long-standing concerns that it was not applicable to economic data. In other words, maybe the chief causes of the profession's embrace of the classical theory of inference are those that drove the broader mathematization of economics, and one should simply look to the literature that explores possible explanations for that phenomenon rather than seeking a special explanation of the embrace of the classical theory of inference.

I would suggest one more factor that might have made the classical theory of inference more attractive to economists in the 1950s and 1960s: the changing needs of pedagogy in graduate economics programs. As I have just argued, since the 1920s, economists have employed both judgment based on expertise and mechanically objective data-processing procedures when generalizing from economic data. One important difference between these two modes of analysis is how they are taught and learned. The classical theory of inference as used by economists can be taught to many students simultaneously as a set of rules and procedures, recorded in a textbook and applicable to "data" in general. This is in contrast to the judgment-based reasoning that combines knowledge of statistical methods with knowledge of the circumstances under which the particular data being analyzed were generated. This form of reasoning is harder to teach in a classroom or codify in a textbook, and is probably best taught using an

apprenticeship model, such as that which ideally exists when an aspiring economist writes a thesis under the supervision of an experienced empirical researcher.[25]

During the 1950s and 1960s, the ratio of PhD candidates to senior faculty in PhD-granting programs was increasing rapidly.[26] One consequence of this, I suspect, was that experienced empirical economists had less time to devote to providing each interested student with individualized feedback on his attempts to analyze data, so that relatively more of a student's training in empirical economics came in an econometrics classroom, using a book that taught statistical inference as the application of classical inference procedures. As training in empirical economics came more and more to be classroom training, competence in empirical economics came more and more to mean mastery of the mechanically objective techniques taught in the econometrics classroom, a competence displayed to others by application of those techniques. Less time in the training process being spent on judgment-based procedures for interpreting statistical results meant fewer researchers using such procedures, or looking for them when evaluating the work of others.

This process, if indeed it happened, would not explain why the classical theory of inference was the particular mechanically objective method that came to dominate classroom training in econometrics; for that, I would again point to the classical theory's link to a general and mathematically formalistic theory. But it does help to explain why the application of mechanically objective procedures came to be regarded as a necessary means of determining the reliability of a set of statistical measures and the extent to which they provided evidence for assertions about reality. This conjecture fits in with a larger possibility that I believe is worth further exploration: that is, that the changing nature of graduate education in economics might sometimes be a cause as well as a consequence of changing research practices in economics.[27]

I will close with an appeal for future research. I hope that this paper has conveyed a sense of the wide variety of approaches employed by economists of the twentieth century to extract knowledge from statistical data—and I do not mean here just the variety of statistical methods, but the variety of general strategies for combining economic theorizing, logical reasoning, and the analysis of statistical information to justify claims about the nature of economic phenomena. I believe there is something to be gained from increasing our understanding of this variety, through both case studies of the research practices of various individuals and groups, and work that

---

[25]Porter (1995, esp. pp. 7 and 200) also notes this contrast between mechanically objective procedures and procedures resting on disciplinary expertise.

[26]According to NSF data (available at http://www.nsf.gov/statistics/nsf06319/pdf/tabs1.pdf; accessed 21 December 2016), the annual number of economics PhD degrees earned per year in the US was relatively stable during the 1930s and 1940s at about 100 to 120. In the period from 1950 to 1955, this increased to about 300 per year; during the period from 1955 to 1960, to 331 per year; during the period from 1960 to 1964, to 432 per year; and during the period from 1965 to 1970, to 685 per year. Tabulation of the annual reports of "Degrees Conferred" published in the *American Economic Review* shows a smaller, but still substantial, increase in the annual rate from the early 1950s to the early 1960s (about 30%). Under any reasonable assumption about the processes determining faculty sizes at economics departments with PhD programs, this growth in PhD degrees conferred implies a large increase in the student-to-faculty ratio in those programs.

[27]Kaiser (2004) points to the postwar explosion of the student-teacher ratio in US physics graduate programs as a factor influencing research practices in physics.

aims at synthesis, perhaps identifying a few basic patterns underlying the variety. There is, of course, already much good work of this sort, but there remains much empirical research from the twentieth century that is worth a closer look. It is worth a closer look because of the light it might shed on normative questions of methodology, and for how it could inform our attempts, as historians of economics, to understand the changes over time in the sort of evidence and arguments that economists find convincing.

REFERENCES

Advisory Committee on Economic and Social Research in Agriculture of the Social Science Research Council. 1928. *Research Method and Procedure in Agricultural Economics*. Two volumes. New York: Social Science Research Council.

Aldrich, John. 2008. "Keynes Among the Statisticians." *History of Political Economy* 40 (2): 265–316.

Biddle, Jeff E. "Statistical Economics, 1900–1950." *History of Political Economy* 31 (4): 607–652.

Biddle, Jeff E., and Daniel Hamermesh. 2016. "Theory and Measurement: Emergence, Consolidation, and Erosion of a Consensus." Unpublished manuscript.

Boumans, Marcel. 2015. *Science Outside the Laboratory*. Oxford: Oxford University Press.

Bronfenbrenner, Martin. 1948. "The Consumption Function Controversy." *Southern Economic Journal* 14 (3): 304–320.

Carter, Charles Frederick. 1953. "Review of *Econometrics* by Jan Tinbergen; *Econometrics* by Gerhard Tintner." *The Economic Journal* 63 (249): 118–120.

Christ, Carl. 1960. "A Symposium on Simultaneous Equation Estimation: Simultaneous Equation Estimation: Any Verdict Yet?" *Econometrica* 28 (4): 835–845.

Crum, William L. 1926. "Recent Texts on the Methods of Economic Statistics." *The Quarterly Journal of Economics* 40 (3): 530–546.

Crum, William, and Alson C. Patton. 1925. *An Introduction to the Methods of Economic Statistics*. Chicago: A. W. Shaw.

Daston, Lorraine J., and Peter Galison. 1992. "The Image of Objectivity." Special Issue, "Seeing Science." *Representations* 40: 81–128.

Day, Edmund. 1925. *Statistical Analysis*. New York: Macmillan.

Friedman, Milton. 1957. *A Theory of the Consumption Function*. New York: NBER.

Friedman, Milton, and Anna J. Schwartz. 1963. *A Monetary History of the United States, 1867–1960*. New York: NBER.

Goldberger, Arthur. 1963. *Econometric Theory*. New York: John Wiley and Sons.

Greene, William H. 2000. *Econometric Analysis*. Upper Saddle River, NJ: Prentice Hall.

Haavelmo, Trygve. 1944. "The Probability Approach in Econometrics." *Econometrica* 12 (Supplement): iii–115.

Hastay, Millard. 1951. "Review of *Statistical Inference in Dynamic Economic Models*, ed. by J. Marshcak and T. Koopmans." *Journal of the American Statistical Association* 46 (255): 388–390.

Hoover, Kevin. 2014. "On the Reception of Haavelmo's Econometric Thought." *Journal of the History of Economic Thought* 38 (1): 45–66.

Hurwicz, Leonid. 1955. "Input-output Analysis and Economic Structure." *The American Economic Review* 45 (4): 626–636.

Johnston, Jack. 1963. *Econometric Methods*. New York: McGraw Hill.

Kaiser, David. 2004. "The Postwar Suburbanization of American Physics." *American Quarterly* 56 (4): 851–888.

Keynes, John Maynard. 1921. *A Treatise on Probability*. London: MacMillan and Co.

Kiefer, Nicholas. 1989. "The ET Interview: Arthur Goldberger." *Econometric Theory* 5: 133–160.

Klein, Lawrence R. 1953a. "Review of *Studies in the Structure of the American Economy." Journal of Political Economy* 61 (3): 260–262.

———. 1953b. *A Textbook of Econometrics*. Evanston, IL: Row, Peterson and Co.

Koopmans, Tjalling C., ed. 1950. *Statistical Inference in Dynamic Economic Models*. New York: John Wiley and Sons.

———. 1952. *Activity Analysis of Production and Allocation*. Cowles Commission Research in Economics, Monograph 13. New York: John Wiley and Sons.

Krueger, Alan B., and Timothy Taylor. 2000. "An Interview with Zvi Griliches." *Journal of Economic Perspectives* 14 (2): 171–189.

Kuznets, Simon. 1950. "Conditions of Statistical Research." *Journal of the American Statistical Association* 45 (249): 1–14.

———. 1951. "Statistical Trends and Historical Changes." *The Economic History Review* 3 (3): 265–278.

———. 1955. "Economic Growth and Income Inequality." *The American Economic Review* 45 (1): 1–28.

Lahiri, Kajal. 1999. "The ET Interview: G. S. Maddala." *Econometric Theory* 15 (5): 753–776.

Lampman, Robert J., ed. 1993. *Economists at Wisconsin, 1892–1992*. Madison, WI: Board of Regents of the University of Wisconsin System.

Leontief, Wassily. 1951. *The Strucuture of American Economy 1919–1939: An Empirical Application of Equilibrium Analysis*. Second edition, enlarged. New York: Oxford University Press.

———. 1952. "Some Basic Problems of Stuctural Analysis." *The Review of Economics and Statistics* 34 (1): 1–9.

Leontief, Wassily, ed. 1953. *Studies in the Structure of the American Economy*. New York: Oxford University Press.

Lewis, H. Gregg. 1963. *Unionism and Relative Wages in the United States: An Empirical Inquiry*. Chicago: University of Chicago Press.

Lodewijks, Jon. 2005. "The ET Interview: Professor Jan Kmenta." *Econometric Theory* 21 (3): 621–645.

Malinvaud, Edmond. 1966. *Statistical Methods of Econometrics*. Amsterdam: North Holland.

Mariano, Roberto S. 1987. "The ET Interview: Lawrence Klein." *Econometric Theory* 3: 409–460.

Mills, Frederick. 1924. *Statistical Methods Applied to Economics and Business*. First edition. New York: Henry Holt & Co.

———. 1926. "An Hypothesis Concerning the Duration of Business Cycles." *Journal of the American Statistical Association* 21 (156): 447–457.

———. 1938. *Statistical Methods Applied to Economics and Business*. Revised edition. New York: Henry Holt & Co.

Mincer, Jacob. 1958. "Investment in Human Capital and Personal Income Distribution." *Journal of Political Economy* 66 (4): 281–302.

———. 1962. "On-the-Job Training: Costs, Returns, and Some Implications." *Journal of Political Economy* 70 (5): 50–79.

———. 1974. *Schooling, Experience, and Earnings*. New York: NBER.

Morgan, Mary. 1990. *The History of Econometric Ideas*. Cambridge, UK: Cambridge University Press.

———. 2014. "Resituating Knowledge: Generic Strategies and Case Studies." *Philosophy of Science* 81 (December): 1012–1024.

Okun, Arthur M. 1957. "The Value of Anticipations Data in Forecasting National Product." Cowles Foundation Discussion Paper No. 40. http://cowles.yale.edu/sites/default/files/files/pub/d00/d0040.pdf. Accessed 21 December 2016.

———. 1963. "Potential GNP: Its Measurement and Significance." Cowles Foundation Paper No. 190. Reprinted from the 1962 Proceedings of the Business and Economics Statistics Section of the American Statistical Association. http://cowles.econ.yale.edu/P/cp/py1963.htm. Accessed 21 December 2016.

Persons, Warren M. 1924. "Some Fundamental Concepts of Statistics." *Journal of the American Statistical Association* 19 (145): 1–8.

Porter, Theodore. 1995. *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton, NJ: Princeton University Press.

Rutherford, Malcolm. 2004. "Institutional Economics at Columbia University." *History of Political Economy* 36 (1): 31–78.

Schultz, Henry. 1925. "The Statistical Law of Demand as Illustrated by the Demand for Sugar." *Journal of Political Economy* 33 (6): 577–631.

———. 1938. *The Theory and Measurement of Demand*. Chicago: University of Chicago Press.

Schultz, Theodore W. 1933. "Testing the Significance of Mean Values Drawn from Stratified Samples." *Journal of Farm Economics* 15 (3): 452–475.

Secrist, Horace. 1925. *An Introduction to Statistical Methods*. Revised edition. New York: Macmillan.

Shiller, Robert J., and James Tobin. 1999. "The ET Interview: Professor James Tobin." *Econometric Theory* 15 (6): 867–900.

Teira, David. 2007. "Milton Friedman, the Statistical Methodologist." *History of Political Economy* 39 (3): 511–528.

Working, Elmer J. 1930. "Evaluation of Methods Used in Commodity Price Forecasting." *Journal of Farm Economics* 12 (1): 119–133.

———. 1932. "Indications of Changes in the Demand for Agricultural Products." *Journal of Farm Economics* 14 (2): 239–256.

———. 1939. "Graphic Method in Price Analysis." *Journal of Farm Economics* 21 (1): 337–345.