Statistical Inference in Economics, 1920-1965: Changes in Meaning and Practice*

Jeff E. Biddle
Dept. of Economic
Michigan State University
July 2016

This paper reviews 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. Haavelmo and the econometricians associated with the early Cowles Commission developed an approach to statistical inference based on concepts and measures derived from 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 25 years, a consensus developed that methods of inference derived from probability theory were an almost essential part of empirical research in economics. I close the paper with some speculation on possible reasons for this transformation in thinking about statistical inference.

*This paper is the basis for my Presidential Address to the History of Economics Society, delivered in June of 2016. In preparing that address and this paper, I benefitted from the helpful comments of Maurice Boumans, Dan Hirschman, Kevin Hoover, Mary Morgan, and Tom Stapleford.

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] As will become clear in what follows, during the period I survey different economists have had different understandings of the meaning of "statistical inference"; thus my reference to "changes in meaning" in the title of the paper. However, all fit within the broader usage that I have adopted. Also, I distinguish statistical inference from what I will call statistical theory, by which I mean the study of the properties of various summary statistical measures calculated with sample data – e.g., the sample mean or a sample regression coefficient – as measures of the characteristics *of that sample*. Statistical theory in this sense is often used in conjunction with, but is distinct from, a branch of probability theory, one that involves making assumptions about the distributions of random variables in an unknown population and about the properties of samples drawn from those populations, then deriving the properties of statistical measures calculated from the samples as estimators of the parameters of those assumed distributions. The changing view of the role that probability theory should play in the process of statistical inference in economics is one of the major themes of the paper.

More prosaically, I think of statistical inference in economics as the attempt to answer certain questions: How do we judge whether the data we have support or refute a proposition about the world in general? How should we use data to estimate the economic quantities we are interested in? And how do we assess the reliability of those estimates once we have made them? Over the course of the twentieth century, economists occasionally made explicit their views on how best to answer these questions, but more often their views were revealed by their research

[1] One can think of statistical inference as 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).

practices, and in the textbooks they wrote and used to teach those practices to aspiring economists.

There is currently a broad consensus among economists, dating to the last decades of the 20th century, on what statistical inference means. Today's graduate student comes to understand "statistical inference" as a set of procedures for analyzing statistical data. One starts by making assumptions about the joint distribution of a number of random variables of interest 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 the assumptions are patterned on one of a set of canonical "statistical models" with well understood properties, such as the linear regression model.

It is also assumed that the economist has 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 information. A formula produces estimates of the parameters, and other formulas produce measures of the reliability of those estimates -- "standard errors" and so forth. These formulas have been derived from probability theory, and there is a set of standardized procedures – also derived from probability theory -- for using the estimates produced by the formulas to test hypotheses about the parameters.

To judge from most econometrics textbooks published since the 1970s, "statistical inference in economics" means no more and no less than the application of this set of procedures. William Greene, in his popular graduate-level econometrics textbook, refers to this approach as "the classical theory of inference", and points out that, "the overwhelming majority of empirical study in economics has been done in the classical framework", a fact which is both the reason for and perhaps a consequence of the complete identification of statistical inference with this “classical theory of inference” in other modern econometrics texts (Greene 2000, p. 97). [2]

But it has not always been thus. Despite the fact that discussions of the foundational principles of Greene’s classical theory of inference could be found in standard textbooks on economic statistics in the 1920s, and despite the fact that important contributions to the classical theory were being made and disseminated to interested economists throughout the 1930s, prior to WWII very few empirical economists in the US made any use of these tools of statistical inference when drawing conclusions from statistical data. 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 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 will then turn to Haavelmo and the early Cowles Commission econometricians and discuss how their arguments in favor of applying the classical theory of statistical inference to economic data did and did not respond to the concerns of those statistical economists of the 1920s and 30s 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

[2] In a few, however, one can find some brief introductory phrases from which it can be discerned that what is about to be explained is one of possibly several approaches to or theories of statistical inference. Green, for example, spends a few pages discussing an alternative Bayesian theory of inference, which is also based on probability theory.
[3] Although I look at some different authors, what I have found is largely consistent with Morgan’s (1990, pp. 230-38) seminal account of the arguments underlying the rejection of probability theory during this period.

profession. I try to shed some light on the timing of this transformation in thinking about statistical inference. In particular, I argue that statistical inference without probability was an important part of empirical economics at least into the mid 1960s. In a final section I offer a some conjectures about the reason for these changes in the meaning and practice of statistical inference in economics from the pre-WWII decades to the last decades of the 20th century. 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.[4]

I. Statistical Inference Without Probability: the 1920s and 1930s

A. Keynes's *Treatise on Probability*

Early in his career, John Maynard Keynes had a significant research interest in probability and statistics. His fellowship dissertation, the final version of which was submitted in 1908, was on "The Principles of Probability". But his interests gradually turned more towards monetary theory, and his last substantial work on probability and statistics was his 1921 *Treatise on Probability*, a heavily revised version of the fellowship dissertation. One of the book's five parts, making up about a quarter of the total content, was entitled "The Foundations of Statistical Inference" (Aldrich 2008).

Keynes began his discussion of statistical inference by making 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

[4] Morgan (1997) provides an count of how leading British economists approached the problem of statistical inference in the decades just prior to the period I examine.

other events that have not been observed". This part of statistics he called the theory of statistical inference, and he noted that it was currently understood to be "closely bound up with the theory of probability"(Keynes 1921, p. 327). The theory of probability provided the basis for the calculation of probable errors for statistical measures like the mean and the correlation coefficient, and it justified the use of probable errors in making inferences about unobserved phenomena on the basis of descriptive statistics calculated for a sample.

Keynes spent several chapters explaining why this approach to statistical inference was actually 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 outlines. At times he seems 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 seems 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 statisticians Wilhelm Lexis. For my purposes, however, what Keynes's contemporaries believed him to be saying is more important than his intended message.

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

John Aldrich (2008) tells us that almost all of 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 and statistician wholeheartedly embraced Keynes's message, and that was Warren Persons.

Persons was a highly original and skilled economic statistician, with a solid command of statistical theory as it existed in the early decades of the 20th century. In the late nineteen teens, while developing a method forecasting business conditions using time series data, Persons made important and lasting contributions to economic statistics. These included his model of economic time series as the resultant of the actions of four distinct components, that is, the trend, seasonal, cyclical, and irregular components, along with the statistical techniques he developed for isolating and analyzing the cyclical component of time series.[5]

Persons's contributions earned him election as President of the American Statistical Association in 1923, and his Presidential Address, in which he spoke at length on "the nature of statistical inference" as it related to forecasting, revealed the influence of Keynes's treatise. Persons asserted that the methods of reasoning required to make an inference about the future based on statistical results derived from past data, and to judge the quality of that inference, were

[5] Person's forecasting method was embodied in the Harvard Business Barometer. For more on Persons, his methods, and the Harvard Barometer, see Friedman (2009) and Morgan (1990, pp. 56-63).

simply those of ordinary induction. He considered the argument that the mathematical theory of probability gave the statistician and the economic forecaster tools for drawing conclusions from statistical data beyond those traditionally associated with the logic of induction. But Persons rejected this view as "wholly untenable". The time series data used in forecasting did not meet the conditions required for the use of the mathematical theory of probability as an aid to inference. Probability-based inference required a random sample of independent observations drawn from a well-defined population, but a given time series from a past period could not be regarded as a random sample in any but an "unreal, hypothetical sense", and the individual elements of a time series were not independent of one another. Even if a time series were regarded as a random sample of independent observations, the logic behind using inferential procedures and measures derived from probability theory for forecasting assumed that the forecasting period was essentially another random draw from the sample used to calculate the forecast. Yet the forecaster typically had enough information about the forecasting period to know ways in which it would differ from a randomly chosen period within the sample, information that the probability-based measures could not incorporate (Persons 1924, quotes from p. 6).

Persons let his audience know that this thesis that "statistical probabilities provide no aid in arriving at a statistical inference" had already been developed "with skill and success" by Keynes in his *Treatise on Probability*, and he quoted Keynes at length. In describing the proper way of using statistical data from the past to make inferences about the future, Persons also borrowed some of Keynes arguments regarding the use of analogy in statistical induction: statistical results from time series data provided a stronger basis for inference if the results were stable when calculated for different sub-periods 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 I noted above, Persons' embrace of Keynes's views on probability and statistical inference have been noted before.[6] What I would like to suggest is that the arguments of Keynes, 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 empirical economic researchers presented statistical data. While the share of journal articles using statistical data in some way remained roughly constant each decade from 1900 to 1950, in the 1920s there was a sizable jump in the percentage of authors of empirical papers who calculated summary statistics like means and correlation coefficients rather than simply presenting tables of numbers. This change coincided with a growing attention to training in statistical methods in US economics departments at both the undergraduate and graduate level (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 would be introduced to the subject by one of these books. And each one of them reflected Persons' view of Keynes's position on statistical inference.

The most prominent of the four was Frederick Mills' "Statistical Methods" (Mills 19. Mills' views on statistical inference are worth looking at not only because of the popularity of his

[6] There is reason to believe that Keynes changed Persons' mind regarding the applicability of probability theory, given Person's earlier endorsement of the use of the probable error as a measure of whether a sample correlation represented a true relationship (Persons 1919, pp. 125-127). This was noted by Cox (1925).

textbook and the high esteem in which it was held, but also because from the mid-twenties to the late thirties, 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.[7]

Mills discussed the classical theory of statistical inference late in his text 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 the formulas for the probable errors of various common statistical measures, and explained how those standard errors could act as measures of 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 which 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 using inferential techniques based on probability theory, the statistician use "actual statistical tests of stability", such as the study of successive samples and the comparison 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 much more explicit, in Edmund Day's 1924 textbook *Statistical Analysis*. In discussing the interpretation of statistical results, Day asserted that the theory of probability was hardly applicable to most

[7] An indication of the widespread circulation of Mills book, and the high esteem in which it was held, can be found in reviews of the revised (1938) edition, such as Brown (1939) and Yntema (1939).

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 close to two pages of text from Persons' ASA presidential address. Day's chapter on inference was noteworthy in that he went beyond platitudes about the need for statisticians to use judgment and to consider all relevant statistical and non-statistical factors: He offered students a systematic method of approaching statistical inference, in the form of a checklist of seven considerations to be kept in mind when attempting to generalize from statistical results.

An Introduction to the Methods of Economic Statistics, by Crum and Patton, appeared in 1925. Senior author W.L Crum taught courses in economic statistics at Harvard in the 1920s and 1930s, and worked closely with Warren Persons. His book barely mentioned probability theory as an inferential tool, and that alone is important, but Crum's opinion on the matter was conveyed in his review of Day's textbook, which praised the chapter I have just described on the interpretation of results. Finally, a second edition of Horace Secrist's Statistics text came out in 1925, and included new material both deriving and explaining probable error measures. Secrist also argued, with the help of a quote from Persons, that these measures were largely inapplicable to economic data.

C. 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. They had access to a large and growing body of agricultural statistics, monetary support from the government, and a mandate to produce useful knowledge, a mandate interpreted in a way that favored empirical research. We have a good indication of the opinions of the period's leading agricultural economists on a broad range of topics related to empirical research, thanks to a two volume report on "Research Method and Procedure in Agricultural Economics" sponsored by the Social Science Research Council and published in 1928. The report was put together by a subcommittee of four prominent agricultural economists, based on contributions from a number of active researchers.[8]

The subcommittee's stance on statistical inference in economics was made clear early in the 300 page 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 over-reliance on mathematical formulas in analyzing statistical data was a mistake made by men of limited training and understanding.[9]

The message that the data of agricultural economics did not meet the assumptions required for probability-based inference, and that statistical inference required both 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 fill the requirements of random sampling, a stance which is noteworthy, given that agricultural economists were much 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

[8] The subcommittee members were John Black, E.G. Nourse, L.C. Gray, and H. 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.

[9] "The economic statisticians, however, generally take the position that the mathematics of sampling and error and inference thus far developed, which holds rigorously only for pure chance and simple samples of entirely unrelated events, is inadequate for the needs of economic phenomena, and that there is little prospect of mathematical analysis soon being developed that will be adequate. Once the assumptions of pure chance are violated, inference has to proceed along other lines than those based on simple mathematical probability." (Advisory Committee, 1928, p.38)
"To the extent that statistics makes use of the mathematical formulae of pure chance, it must accept all their assumptions. How the logic of mathematics starts with assumptions and introduces more at every step, is not generally realized by the ordinary manipulator of formulae. This is why men trained to a limited extent in mathematics with no understanding of the philosophy of it rely so implicitly on formulae when they enter the field of statistics." (idem.)

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.[10] 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, especially since the universe of which an economist's sample was truly representative was seldom the universe about which the economist wished to draw inferences. Since the universe from which the sample was drawn was not likely the universe about which the economist wished to draw conclusions, 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 which really interested him? Like Keynes and Persons, Working argued that the tests of stability created by Lexis had a role to play – the relationship should at least be stable within the sample if it was to be trusted to 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. Under ideal conditions, the economist would know this system of causal relations. He would have a sample that would allow him to use multiple correlation analysis to account for all the relevant variables in the system, and this

[10] See Morgan (1990, pp. 170ff) for a discussion of this claim.

would require that all of the relevant variables actually be observed and be varying in his sample. The closer the economist could come to this ideal, the more reliable his inferences from sample correlations would be. An important part of statistical inference, then, was being able to judge both how close one was to this ideal, and the effects of deviations from the ideal.

By way of providing a general framework for tackling this task, Working offered a non-technical description of how to assess the bias introduced into a multiple regression coefficient by the absence from the analysis of a relevant variable. He also presented a concrete example involving a multiple regression model explaining the price of butter, in which he combined theoretical considerations with a test of within-sample stability to conclude that the estimates of sample relationships did not represent stable cause-and-effect relationships. He cautioned, however, that was no general procedure or technique to this inferential process. Each case presented unique problems that would tax "to the utmost" the mental powers of the economist.[11]

In a 1930 paper, Working made a number of these points again, this time with special reference to the problem of forecasting agricultural commodity prices. Working added an additional theme to this paper, however: his concern with a growing tendency among agricultural economists to substitute refined statistical methods for the sound inferential procedures he was recommending. Working endorsed the use of the most up-to-date multiple regression techniques in analyzing sample data, but believed that too often, those using refined techniques tended to be less careful and thorough in making inferences about the future from past 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 it the same time making it more difficult to understand exactly what those measures represented. That some of the refined statistical techniques that

[11] Working's section on inference is found on pp. 270ff in Advisory Committee (1928). The quote is on p. 281.

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."[12]

D. The 1930s

Efforts to disseminate both "refined statistical techniques" and the classical methods of statistical inference picked up in the early thirties. The inferential methods recently developed by R.A. Fisher, Jerzy Neyman, and Egon Pearson were discussed in survey articles in the *Journal of the American Statistical Association* and *Econometrica*; at this time, a substantial majority of the members of the American Statistical Association were economists (Biddle 1990). In 1930, Mordecai Ezekiel published *Methods of Correlation Analysis*, which came to be used as an advanced textbook and a handbook for empirical economists wishing to employ the new methods. Ezekiel's book included a chapter intended to bring to the attention of US economists what he called "the new methods of judging the reliability of conclusions" associated with Fisher and others. In discussing these methods, he explained the idea of sampling error, and the problem of using information from a sample to draw conclusions about other samples from the same universe. The caveats that accompanied the discussion were relatively mild: The new methods were based on assumptions, and it was the statistician's duty to know if those assumptions were reasonable; the new methods only dealt with the problem of sampling error, and provided no

[12] Working 1930, esp. pp. 122-125. The quote comes from pg. 124. Keynes (1921, p. 329) expresses the same concern. Working continued to argue against the unreflecting application of probability-based methods of statistical inference throughout the 1930s, although he took to using the phrase "statistical inference" to refer only to the application of those methods in drawing conclusions from statistical data. See Working (1932, 1937, 1939).

information about the effect on the reliability of one's conclusions of bias introduced by the sampling process (Ezekiel 1930, pp 13ff).

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 Koopmans, whose dissertation reinterpreted 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 Ph.D. students at Columbia.[13] She also notes that this sort of work had little noticeable impact on the literature in the 1930s.

Aldrich (2008) points to Henry Schultz as one US economist who was willing to use probability theory, and that is true, but this statement requires some qualification. From the early 1920s to his untimely death in 1938, Henry Schultz worked on the empirical estimation of demand curves. He was recognized as an authority on the application of regression techniques to time series data, and was probably the most influential economic statistician at the University of Chicago from the late 1920s until his death, training graduate students and advising colleagues, such as Paul Douglas, who wished to use more advanced statistical techniques on their empirical work (Douglas 1939).

From the beginning of his career, Schultz was familiar with how probability theory could be used to justify using the standard errors of descriptive statistics as tools for inference, and in his work during the 1920s estimating demand curves, he routinely reported the standard errors of correlation coefficients and of estimated regressions. But these measures played only a minor role in Schultz's attempts to draw conclusions about the true elasticity of demand from his sample estimates. His main strategy was to estimate a wide variety of specifications of the

[13] See also Brown and Sturges (1937, pp. 699-706) for an defense of the usefulness of the classical theory of inference aimed at agricultural economists.

demand curve, using different methods to control for the trend, different functional forms, and so forth. Citing a variety of economic and statistical considerations, he would choose some of the resulting elasticity estimates as more reliable than others, and from among those, distill a best estimate of the true elasticity.[14]

In the 1930s Schultz adopted the practice of reporting standard errors of his regression coefficients, and he often used the concept of statistical significance in assessing what one could safely conclude from an estimated coefficient (e.g., Schultz 1933). He also began a project of using estimated demand curves to test the restrictions placed on price and income elasticities by the neoclassical theory of the consumer. Although he did not conduct formal Fisher-style statistical tests of the restrictions, he did use the standard errors of the coefficients as a metric for determining whether the relationships between various estimated coefficients were close enough to those implied by theory to count as a confirmation of the theory (Schultz 1938, Chapts XVIII, XIX). And his massive 1938 book summarizing his life's work included an appendix explaining the fundamentals of the classical theory of hypothesis testing, as explained to him by Jerzy Neyman (Schultz 1938, pp. 732-734).

However, one finds in the same 1938 book a long discussion of standard errors and significance testing, clearly written in the late thirties, that reveals serious reservations about the usefulness of the classical theory of inference in the analysis of time series data. Although he worked exclusively with time series data, Schultz argued that time series could not be considered as samples for 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) allow us to answer the question "what is the probability that the true value of a given parameter will be found between

[14] See, e.g., Schultz (1925, pp. 606, 627-628).

such and such limit". But that, of course, was precisely the question that the classical theory of inference was typically used to answer. Instead, Schultz said, standard errors calculated for coefficients of a time series regression should be regarded only as another measure of goodness of fit, though useful in that role (Schultz 1938, pp. 214-215). As one looks at the body of Schultz's work, one sees an eclectic approach to drawing inferences from sample results, involving statistical measures of goodness of fit, considerations from economic theory, considerations of coefficient stability across differing model specifications when theory did not speak clearly, and, occasionally, tests of significance.

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 testing hypotheses and a "sounder foundation" for statistical inference. As a result, in the new book the material on inferential methods derived from probability theory had been expanded from two to six chapters (Mills 1938, pp. vii-viii).

Although Mills's revised chapter on "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. [15] Still, students using Mills's textbook could now learn how to conduct Fisher's analysis of variance, calculate standard errors for regression coefficients and conduct T-tests when confronted with small sample. It should also be noted that Mills himself had used classical inferential reasoning in a 1926 paper that proposed and tested a hypothesis regarding the cause of a secular change in the duration of business

[15] The claim that the conditions assumed in deriving the classical measures were "rarely, if ever" met in economic data became simply "rarely", and Mills also now taught 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).

cycles, although he was careful to note that he was only making a statement about the likelihood that observed changes in cycle duration were due to “sampling fluctuation alone.” (Mills 1926).

Both Mills and Shultz were intellectually aware of and concerned about the stringent data requirements for the valid application of the classical inferential measures. However, they were also strongly committed to making economics more like the natural sciences, in which theoretical hypotheses were put to empirical test. 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, nor was it the result of a naiveté about the problems created for empirical research by the inexact nature of economic theory. Leading empirical economists articulated a sophisticated, reasonable case against drawing conclusions 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 turn to that part of the story.

II. Haavelmo, Cowles, and the Argument for Using Probability Theory

Most everyone agrees that the work of Trygve Haavelmo and the Cowles Commission econometricians in the 1940s and early 1950s was crucial to the eventual widespread acceptance classical theory of statistical inference.[16] The Cowles group’s justification of the application of

[16] In what follows, when I refer to the Cowles econometricians, I mean not only those econometricians formally affiliated in some way with the Cowles commission in the 1940s and 1950s, but also those, like Henri Thiel, who

probability theory to economic data is found in Haavelmo's (1944) essay on the Probability Approach in Econometrics.[17] In the introduction Haavelmo acknowledged, and promised to refute, the prevalent opinion among empirical economists 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 that the classical theory of statistical inference was designed to solve.

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 were dubious about attempts to fit time series data into this concrete "sample and universe" conceptual scheme. When it came to time series data, there did not seem to be any reasonable answer to the question "what is the universe from which this is a sample?" Haavelmo explicitly identified this way of thinking about probability theory as "too narrow." In its place, he proposed the idea of the time series as a set of observations generated by a mechanism, one that was capable of generating an infinity of observations. The mechanism could be characterized by a probability law, and the task of statistical inference was to discover that probability law (Haavelmo 1944, pp. iii, 48). This provided, I think, a very compelling answer to the objection that a time series was not representative of any definable universe. Those who had trouble constructing a plausible explanation of the universe of which a time series was a sample had less trouble imagining the economy as a mechanism (or, to use a term quickly adopted by the Cowles econometricians, a "structure"), one governed by stable laws, with the

---

were clearly working on the research program laid out in Haavelmo's (1944) essay on the Probability Approach and in the Cowles econometric monographs.

[17] As Kevin Hoover (2014) has pointed 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.

data generated by the economic structure being a source of evidence regarding those laws. However, I am going to argue that neither Haavelmo's essay nor the empirical methods used by the Cowles Commission econometricians provided convincing answers for some of the other major elements of the preexisting case against applying the classical theory of inference to economic data.

As is well known, the Cowles econometricians operated with the assumption that the probability law governing the mechanism that produced their data could be represented as a set of linear equations, each with a "systematic" part, describing relationships between the variables observed in their data, and an unobserved disturbance term. It was also necessary to make assumptions about the joint distribution of the disturbance terms and the observable variables.[18] 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 the main motivation behind those distributional assumptions was that they were necessary for the valid application of classical inferential procedure.

The Cowles approach to statistical inference involved two distinguishable aspects. One was the unbiased estimation of the coefficients, or parameters, of the systematic part of the assumed probability law – the structural parameters, as they came to be called. The development of new estimators – that is, alternatives to the traditional least squares algorithm -- that could produce unbiased estimates of these parameters when the structural equations formed a

[18] As students were told in Arthur Goldberger's influential econometrics text (to be discussed below): "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." (Goldberger1963, p. 4).

simultaneous system was of course one of the best known contributions of the Cowles econometric program. But the estimation of structural parameters produced, as an important by-product, estimates of the unobserved disturbances. These estimated disturbances were needed to estimate the variances and covariances of the estimated structural parameters, which were in turn necessary for classical hypothesis testing and the construction of confidence intervals for the structural parameters.

I detail these two aspects because each required a different type of assumption to succeed. The proof that any 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.[19] The proof that an estimated variance covariance matrix of the parameter estimates was unbiased – whether the structural parameter estimates themselves were unbiased or not -- 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 parsimonious mathematical model, as when it was assumed that serial correlation in the unobserved disturbances could be properly represented with a linear equation involving one or two parameters.

Finally, an assumption underlying the entire inferential procedure 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.

[19] The important question of "identification" hinged on the number and pattern of assumptions one made about non-correlation between observable variables and unobserved disturbances.

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.

It is worth pointing out that Elmer Working's attitude towards the role of theory in empirical research has important similarities to that of Haavelmo and Koopmans: The goal of empirical analysis in economics should be to measure those correlations in the data that represented true cause and effect relationships. Economic theory was an important aid in identifying which partial correlations represented such causal relationships. The systems of causal relationships underlying economic data were typically complex, involving many variables, and it was necessary for the researcher, aided by theory, to consider the complete system of causal relationships when attempting to estimate any of them. One consequence of taking this view, a consequence recognized by Working and a central point of emphasis for Cowles, was that the causal system generating economic data would involve simultaneous relationships, and that the partial correlations produced by ordinary least squares estimation would not properly measure those relationships. Another consequence of this view, the one emphasized by Working, was that the available data were unlikely to include the full set of variables involved in the relevant causal system. If those unobserved variables were correlated with the observed variables included in the regression analysis, the resulting parameter estimates would be biased. The classical inferential methods employed by the Cowles commission did not solve this problem. Instead, application of those methods required an assumption that the problem did not exist. [20] The Cowles econometricians were willing to assume routinely that unobserved variables of their system were uncorrelated with the observed. Working was not

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

willing to make this assumption, instead counseling his readers that they be attentive to the consequences of the likely failure of this assumption when drawing inferences from statistical results.[21]

One objection to the use of probability models with economic data that Haavelmo explicitly mentioned as arising from an overly narrow view of probability theory was the contention that successive observations in a time series were not independent. The Cowles econometricians did offer a strategy for solving this problem, which was to propose a parametric statistical model of the dependence between observations, estimate the parameters of that model, and then use the parameter estimates to create a series of estimated disturbances that were not serially correlated. This was an original and ingenious path of research to pursue, and I have seen no evidence that anything like it was ever considered by the statistical economists of the 1920s and 1930s whom I have mentioned so far. But there is reason to believe that at least some of them would not have been satisfied with it. Consider what Henry Schultz (1938, p. 215) said about the problem of serial correlation: "nor are the successive items in the series independent of one another. They generally occur in sequences of rises and falls which do not repeat one another exactly, and which are often masked by many irregularities." In other words, the dependence between observations could not be captured by a parsimoniously parameterized model. Schultz would probably not have been willing to make the simplifying assumptions regarding the nature of serial correlation that the Cowles econometricians needed to make in order to employ the inferential methods that were a crucial element of their econometric research program.

[21] Working listed four conditions "in order that the regression may without a doubt express the true cause-and-effect relationships": the system of cause and effect relationships could be expressed in a way that was "linear in the parameters", there was no perfect collinearity between the variables in the causal system, and, more importantly, all relevant variables could be included in the analysis and were observed to actually vary in the data. These four conditions, Working thought, "probably never exist in actual study of economic data." (Advisory Committee 1928, pp. 278-79)

As I have noted, 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 they cared about. 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). Working commented that there was nothing in the theory of sampling that warranted our saying that "the conditions of covariance obtaining in the sample (would) hold true at any time in the future" (Advisory Committee 1928, p. 275) . As I have already noted, Persons went further, arguing that treating a time series as a sample from which a future observation would be a random draw was not only inaccurate but ignored useful information about unusual circumstances surrounding various observations in the series, and the unusual circumstances likely to surround the future observations about which one wished to draw conclusions (Persons 1924, p. 7). 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 to little assuage these concerns except the hope that it would be possible to specify the equations describing the systematic part of the mechanism of interest in a way that captured the impact of factors that made for structural change in the case of time series, or factors that led cross section samples to be systematically different from the universe of interest.

It is not my purpose to argue that the economists who rejected the classical theory of inference had better arguments than 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 only wish to 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 analytical methods. But they also often involve a willingness to employ assumptions considered by most economists at the time to be too unrealistic, a willingness that arises because the assumptions allow progress to be made with the new concepts and methods.[22] Obviously, in the decades after Haavelmo's essay on the probability approach, there was a significant change in the list of assumptions about economic data that empirical 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 I offer some conjectures that I hope will contribute to this effort at the end of this paper. However, I will 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 economic to be replaced by a consensus the methods associated with the classical theory of inference were not only appropriate, but were the primary and essential 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 to run its course.

[22] 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)
.

III. Statistical Inference with and without probability: Pedagogy and Practice, 1945-1965.

A. 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. In 1950, the Cowles Commission had published its monograph no. 10, a collection of papers from a Conference on "Statistical Inference in Dynamic Economic Models" (Koopmans 1950). The papers dealt with all aspects of the Cowles Econometric program, and during the early 1950s, statistician George Kuznets used monograph 10 to teach graduate students in agricultural economics at the University of California. But as a textbook, the monograph was quite wanting: Lawrence Klein described it as "rigorous in an overbearing pedantic manner". [23]

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.[24]

As I have noted, prominent empirical economists of the twenties and thirties considered and rejected the idea that the phrase "statistical inference" referred specifically to the use of

[23] On the lack of good textbooks on the new econometrics prior to 1952, see Carter (1953), Solow (1953), Bronfenbrenner (1953), and Arrow (1954). The observation about Harvard and Yale is from Shiller and Tobin (1999). Klein reports that Cowles essays were also circulated among interested parties at MIT in the 1940s (Mariano 1987). G. Kuznets' use of monograph 10 is reported by Z. Griliches in Krueger and Taylor (2000). The Klein quote is from Klein (1954, p. 343). In 1953, the Cowles Commission published monograph 14, *Studies in Econometric Method* (Hood and Koopmans 1953), which was intended as a more accessible exposition of the ideas and methods found in monograph 10.

[24] An English translation of Tinbergen's text *Econometrics* appeared in 1951, but the Dutch version had been written in 1941, and did not reflect the innovative ideas and methods introduced by Haavelmo and Cowles.

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” 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. 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 1952, pp. 50-56, 252-264).

Tintner also taught students how to apply the tools of the classical theory of inference in the context of regression analysis, although, somewhat surprisingly, the estimation of simultaneous equations system was not discussed. Another surprise was Tintner’s rejection, in principle, of the frequentist theory of probability that was at the root of the classical theory of inference. He argued that Keynes’s concept of probability as the degree of belief in a theory on the basis of a body of empirical evidence was a better foundation for inference in econometrics. But, he allowed, very little progress had been made in building a formal theory of inference on this foundation, so for the time being, it was necessary to use the classical theory, although "it was not very satisfactory from a philosophical and epistemological point of view (Tintner 1952, p. 17)".

The appearance of these textbooks was greeted enthusiastically by some reviewers, but others sounded the old concerns about the applicability of the classical theory of inference to economic data. C. F Carter, after acknowledging the need for the modern econometrics texts, worried that books like those by Tintner and Tinbergen conveyed a false sense of the precision provided by methods like significance testing, given that the methods would often be used on "data to which they were not quite suited (Carter 1953, p. 120)." In a well-informed and somewhat positive review of Cowles monograph 10, Millard Hastay of the National Bureau noted that the Cowles econometricians worked with methods that were technically correct only for "linear systems, serially independent disturbances, error-free observations, and samples of a size not generally obtainable in economic time series today." Assumptions were made for the sake of validating the methods, not because of their accordance with reality. "Painfully little" was known about the reliability of the methods if the assumptions did not hold, though perhaps experimentation with them could remedy that situation. It would be unfair to suggest that the authors did not understand these difficulties, Hastay observed, "but it seems to this reviewer that they have not given adequate weight to their reservations in tempering the promise of success held out for their method (Hastay 1951, p. 389)."

After a lull of about a decade, there appeared another set of econometrics texts destined for wide adoption, books by Goldberger (1963), Johnston (1963), and Malinvaud (1966). In these books, which in many ways would set the pattern for subsequent 20th century graduate econometrics texts, statistical inference in economics meant an application of the classical theory of inference.[25]

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 between 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 spent a chapter illustrating this basic reality of applied econometrics by examining various models of the consumption function (Malinvaud 64-65, 135, chapt. 4).

Malinvaud also reminded readers that the inferential methods he was presenting depended on assumptions such as normality of and independence between certain variables, assumptions often chosen for convenience. This made it important to investigate theoretically the impact on the properties of one's estimators and tests of changes in the assumptions of the model, and to seek more robust estimators and tests.

Like Malinvaud, Goldberger argued that statistical inference proper began only after the model was specified; it was only during model specification that *a priori* assumptions or

[25] "The discussion follows the classical theory of inference, distinguishing among point estimation, interval estimation, and hypothesis testing" Goldberger (1963, p. 125). See also Malinvaud (1966, p. 67).

information beyond the sample played a role in empirical analysis. Goldberger mentioned the Bayesian approach to statistical inference, which offered a probability-theoretical framework for combining a priori information with sample information, but argued that the classical theory of inference was sufficiently flexible to allow prior beliefs or information about the distributions of the model parameters to be both incorporated into the estimation procedure and tested (Goldberger 1963, p. 142, sec. 5.6). Note the contrast between Goldberger's explanation of how theory and background information should be integrated into empirical analysis and Elmer Working's description and demonstration of the process of statistical inference in the Social Science Research Council volume discussed earlier. Like Goldberger, Working believed researchers should rely on theory in determining the regression relationships to be estimated, but Working's exposition showed that he also 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.

B. 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 a more prosaic instruction in constructing confidence intervals and testing ordinary least squares regression coefficients for "statistical significance".[26] In the 1950s, however, the amount of systematic instruction in the classical theory of inference available to interested economics graduate students, not to say average economics graduate students, depended on where they were being trained. At least that is the impression I have gotten from reminiscences of leading econometricians, and histories of some of the larger US graduate programs in the 1950s. 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 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.[27] The picture at Wisconsin changed dramatically with the arrival of Guy Orcutt in 1958 and Arthur Goldberger in

---

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

[27] On the sporadic nature of Bronfenbrenner's teaching of econometrics, see Lampman (1993, p. 131, p. 96). In 1948-49, advanced statistics courses were still being taught by the Commerce faculty members whose courses Friedman had deemed inadequate (Lampman 1993, p. 88). Regarding what Bronfenbrenner likely taught about statistical inference, in a survey of the failure of post-war economic forecasts, he commented that "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). Bronfenbrenner (1953) praises the work and methods of the NBER, whose researchers practiced inference without probability.

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). During the early 1950s, Cowles alumnus Carl Christ was teaching econometrics at Johns Hopkins (Ghysels 1993), and at Iowa State University, future econometricians like George Judge could take classes in econometrics from Gerhard Tintner and Cowles associate Leonid Hurwicz (Bera 2013).

At the University of Chicago, home of the Cowles Commission until 1955, the situation was complicated. Zvi Griliches came to Chicago in 1954, having already been exposed to Cowles-style econometrics at the University of California, and received further training from Henri Theil and Haavelmo, among others. In 1957 he began teaching econometrics at Chicago. Among Griliches's students in the early 1960s was future econometrician G. S. Maddala, who considered writing a dissertation in econometric theory before opting for a more empirically oriented topic. But Maddala reports that econometrics at Chicago was very "low key:" Neither Maddala nor any other student in his cohort who was writing an empirical dissertation used anything more complicated than ordinary least squares regression. High tech methods were eschewed for actual empirical work, in part due to the influence of Milton Friedman. (Krueger and Taylor 2000, Lahiri 1999)

C. **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 empirical research studies of the 1950 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: hypotheses about the relationships between estimates of the marginal propensity to consume for different groups and from different types of data. But one will search his book almost in vain for applications of the classical methods of inference.[28] 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, as well as numerous generalizations based on that data. But the book contains no classical hypothesis tests, no confidence intervals, no reports of statistical significance or insignificance, and only a handful of regressions.

H. Gregg Lewis was a colleague of Friedman's at Chicago. Lewis, who received his Ph.D. from Chicago, joined the faculty in 1939 to replace Henry Schultz as a teacher of graduate level statistics, and was on 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 in various industries, measuring what came to be known as the

[28] Exceptions are a few references to the statistical significance of particular estimates (Friedman 1957, pp. 141, 146, and 175). 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.

union-non-union wage gap, and in the late 1950s he set out to reconcile the wage gap estimates from these studies and a number of others.

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 the method of measurement, and when they could be plausibly linked to economic factors. While Lewis reported standard errors for some of the regression coefficients he discussed, he made little if any use of them 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 judgments made by Lewis, based on statistical theory, of the likely size of biases due to inadequacies of the data. These sorts of biases were the main focus of Lewis's efforts to assess the accuracy and reliability of wage gap estimates. 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 Viner. 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 tables, cross tabulations, and estimation techniques borrowed from national income accounting to measure key human capital concepts and to establish general patterns. Gary Becker was still less than ten years beyond his Ph.D when he published *Human Capital*, a statement of the results of seven years of research (Becker 1965). Although the book contained a fair amount of statistical data, some meant to test hypotheses derived from the book's theoretical models, Becker did not employ the tools of classical inference theory. The empirical innovator among the three was Jacob Mincer, who among other things, developed the "earnings regression" approach to measuring the rate of return to human capital. Mincer, who like Becker was just beginning his career in the late 1950s, had used statistical and probability theory as modeling tools in his dissertation, and continued to do so in his work on human capital. But the tools of classical inference theory played little role in his early articles on human capital theory (Mincer 1958, 1965), nor for that matter in his 1974 magnum opus, *Schooling, Experience, and Earnings*.[29]

But statistical inference without probability was not just a Chicago phenomenon in this period. Any economist trained in the last three decades of the 20th century was taught Okun's Law, an empirical generalization positing a stable relationship between changes in the unemployment rate and changes in the GNP. Arthur Okun, a policy-oriented macroeconomist at Yale, first presented the law and the empirical evidence behind it in 1962, in the context of a discussion of the measurement of potential GNP, and it was quickly accepted as a basic macroeconomic reality. The important point for my purpose, however, is the nature of the evidence presented: results from three alternative specifications of a regression involving

[29] In Mincer (1962, p. 71), the concept of statistical significance is used in a way not strictly consistent with the classical theory: "However, in terms of the investment hypothesis . . . the positive sign at X2 is puzzling. Could it possibly reverse if the analysis were expanded to include such variables as urbanization, unionization, race, marital status? Such an expansion, if feasible, would be desirable. I experimented with inclusion of two easily accessible variables: X5, percentage of males older than fifty-five, and X6, percentage of non-whites in an occupation. Neither was statistically significant. Their inclusion did not increase the correlation coefficient, nor did it affect the coefficient of X2. The inclusion of the racial variable X6, however, lowered the coefficient of X1 and weakened its reliability." In Mincer (1974, p. 91), a mention of statistical significance is accompanied by this comment: "All the estimated coefficients shown in Table 5.1 are highly significant in a sampling sense: the coefficients are many times larger than their standard errors. This is due to the very large sample size, though size alone is not a sufficient condition for statistical significance."

unemployment and GNP. There were no confidence intervals reported for regression coefficient(s) that embodied the law, nor was their "statistical significance" mentioned.[30]

Then there is Simon Kuznets. In December of 1949, Kuznets was the President of the American Statistical Association. A substantial part of his presidential address was devoted to discussing the full implications of the "trite" observation that the data used by social scientists were generated not by controlled experiments, but through social processes. He emphasized that the interests of those who collected these data differed from those of social scientists, so their views of what to collect, how to collect it, and how much to collect led to a supply of data that fell short of any reasonable standard for scientific analysis. The uncontrolled respondents upon whose reports the data were based gave inaccurate reports, sometimes on purpose; the concepts employed in collecting the data did not correspond to the theoretical concepts of researchers, and so on. But Kuznets's long catalogue of data problems was not a counsel of despair, rather it was a prelude to a discussion of what methods were and were not well suited to the analysis of social data. And clearly on the "not" list were the classical theory of inference and the Cowles Commission style of econometrics.

For example, Kuznets argued that the nature of the measurement errors in data, and their relationships to the true values of the variables being measured, varied within and across data sets, so that they could not be dealt with by any "broad and standard scheme." At a minimum, particular knowledge of the character of bias or ignorance associated with each measurement was needed. Adding to the difficulty created by these errors was the essential complexity and

[30] A look at Okun (1957) shows that this was not atypical of his empirical work. A comment in Okun (1961) reveals an attitude that the classical theory of inference may be useful but is by no means essential in empirical economics: after describing the selection of preferred regression specifications through a process of eliminating variables with insignificant coefficients, he notes that "The mining of the data through experimentation detracts from the statistical evaluation of the results. This is no catastrophe, however, since the study is designed to develop optimal point estimates rather than tests hypotheses."

changeability of social phenomena. Both traditional and newer inferential methods derived from probability theory presupposed that the analyst understood, and could control for, systematic patterns in the data that changed over time. Yet it was only through the analysis of data that we could gain the knowledge necessary to control for these patterns, or develop testable 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 the important activity of statistical analysis 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).

Lest we be tempted to consider these the curmudgeonly comments of a member of some old guard, we should remember that Kuznets was 49 years old at the time, with the presidency of the American Economic Association still five years in his future, and at the beginning of a major and influential empirical research program on the causes of economic growth. Over the next two decades, Kuznets assembled and analyzed a great deal of statistical data, proposing hypothesis 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, that is, the “Kuznets curve” (Kuznets 1955). And his analysis did not involve, nor did his conclusions rest, on classical hypothesis tests.

To provide another example of what statistical inference without probability could look like, I will give a brief account of how Kuznets approached one particular inferential task associated with his project creating historical national income estimates for the United States (Kuznets et al., 1946). Kuznets and his assistants on the project, Lillian Epstein and Elizabeth

Jenks, developed strategies for dealing with a number of problems and gaps in the available statistical material. For example, an estimate of the total income produced by an industry might be based on data pertaining to only part of that industry. Or an estimate pertaining to a particular year might have to be interpolated from data pertaining to nearby years. And reporting errors were an endemic concern. In the face of these data difficulties, Kuznets believed it important to report margins of error for his estimates.

If it were possible to make specific assumptions concerning the distributions of the errors in the data, Kuznets commented, one could assign margins of error to the estimates using "the full armory of weapons of statistical analysis of sampling errors and limits of inference" (Kuznets et al. 1946, p. 535). But Kuznets believed that the nature of his statistical material made that approach inadvisable, so he developed an alternative procedure for assessing the reliability of his estimates that was in principle transparent and replicable. Kuznets commented that the "logic involved was clear enough for any student to follow," and the specific knowledge required was no greater than what any "intelligent reader" could gather from the extensive descriptions of data and methods included in the text (Kuznets et al. 1946, p. 535).

Kuznets had produced, for each year, estimates of 13 measures pertaining to each of 40 industries. Working independently, he, Jenks, and Epstein placed each of the 520 annual estimates into one of four "maximum probable error" categories, each representing a range, such as "maximum error between 10 and 15 percent". They also assigned error margins to certain larger aggregates, sometimes directly, and sometimes by calculating weighted averages of the errors margins assigned to the subcomponents that together comprised the aggregate (subcomponents were assigned the midpoint value of their error category to make the calculation of averages possible). The assigned error margins of each of the three researchers were reported separately, with no attempt to reconcile differences, as Kuznets believed that the presentation of three separate, uninfluenced opinions would be more “objective”.[31]

Error margins were first assigned using only information on the data and methods used to create the estimates. Then two additional checks of reliability were applied. First, the team compared their estimates, where possible, to those constructed by other researchers. Second, they developed a test of their procedure for interpolating or extrapolating missing data values. The procedure itself was based on the assumption that each quantity that was not measured in every year maintained a fairly stable ratio to another quantity that was observed annually. This assumed fixed ratio could be estimated using data from years in which both quantities were observed. Then, if the desired quantity was not observed in a given year, it could be estimated by multiplying the ratio by that year’s value of the annually observed quantity. The test of this procedure was to use it to estimate the intermittently observed quantities for years in which they actually had been observed, then compare the estimates to the actual values. As a result of these two reliability checks, Kuznets decided to increase the error margin of every estimate by 50%.[32]

My final example of statistical inference without probability is Wassily Leontief’s input-output analysis. During the 1940s and 1950s, there was tremendous enthusiasm for input-output analysis among young, mathematically oriented economists who were rising to positions of prominence in the profession.[33] Linear simultaneous equations models such as those developed by Leontief could be used to represent a variety of economic processes, and provided a fruitful starting point for the rigorous mathematical derivations of a bounty of theorems regarding the behavior of those processes. And, Leontief’s work provided a template for the use of such

[31] The procedure for assigning error margins is described in detail in Kuznets (1946, pp. 501-538)
[32] This was done by changing the four “maximum error range” categories. “3 to 7%” with a midpoint of 5% became “5 to 10%” with a midpoint of 7.5%; “8 to 12%” with a midpoint of 10% became “11 to 20%”, with a midpoint or 15%, and so on (Kuznets et al. 1946, p. 502n1; the interpolation tests are described on pp. 486-500).
[33] See, for example, the introduction and first part of Cowles monograph 13 (Koopmans, 1951).

models as tools of empirical analysis that had the potential to aid greatly in the tasks of economic forecasting and control, tasks on which many of the mathematically oriented economists of that period placed a very high priority.

But 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 economic science. Leontief, like the Cowles econometricians, began with a model expressed as a system of linear equations, and faced the task of assigning empirically derived values to the parameters of the model's equations. The parameters of Leontief's model were the input-output coefficients, each of which represented the quantity of one industry's output required as an input into the production of a unit of another industry's output. Leontief estimated these "production function" parameters by calculating the inverse ratio of the net value of one industry's output to the value of a second industry's output sold to the first industry, using information from the national income accounts, censuses of manufacturing and agriculture, and so on. However, Leontief hoped that input-output coefficients created using his simple ratio method would over time be replaced by numbers based on careful engineering studies of individual industries. Leontief described this general approach to estimating model parameters as a method of "direct observation", as opposed to the "indirect statistical inference" employed by "the modern school of statistical econometricians" (Leontief 1950, pp. 2-3, 7).[34]

The basic methods of statistical inference were important tools of empirical analysis, Leontief allowed, but the refined versions of these techniques developed by the econometricians

[34] "The empirical components which are to be included in the theoretically prefabricated analytical schemes can be arrived at either through direct observation, or through the use of more or less intricate methods of indirect statistical inference. The production function of some particular industry can, for example, be obtained directly through the collection of relevant technical, engineering information or it can be determined indirectly by way of a rather intricate interpretation of supply and demand reactions in the industry." (Leontief 1952, p. 3)

to deal with the inadequacies of existing economic data simply could not do the tasks being asked of them. Further, in the search for the large samples required to make their methods valid, econometricians had been driven towards the "treacherous shoals of time series analysis", where they were forced into the "fatal choice" between too-short series of highly autocorrelated data and longer series which required the assumption of "invariance in relationships which actually do change and even lose their identity over time." The work of the modern econometricians was original and imaginative, but not the best use of the profession's resources (Leontief 1952, p. 2-3).

Assigning values to the model parameters was only a preliminary step in input-output analysis, preparatory to using the empirically specified model as a tool for estimating unknown price and quantity relationships, or forecasting the effects of hypothetical economic changes. But how to assess the reliability of the estimates and forecasts, or more correctly, of the empirical model that had produced them? Leontief and his disciples developed procedures for "testing" their models, procedures that represent another version of inference without probability in the 1950s and early 1960s. For example, one of Leontief's tests of his input-output model for 1939 was to use it along with the output of final goods in 1929 to "predict" the distribution of output across all industries in 1929, then compare this prediction to the actual 1929 data, and to predictions based on "conventional statistical methods".[35] Another approach was to assess the sensitivity of the forecasts produced by an input-output model to alternative specification decisions, such as the decision of how to group industries in the necessary process of aggregation (Holzman in Leontief, ed., 1953).

Cowles econometricians predictably took issue with Leontief's rejection of "indirect inference." Both Lawrence Klein (1953) and Leonid Hurwicz (1955) reviewed the collection of

[35] Leontief (1951), pp. 216-218.

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." He would later suggest an approach to putting input-output analysis into a framework that would allow the application of classical inferential techniques (Klein 1962). Leonid 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 inference without probability, although it is no longer a research tool employed at the frontiers of the profession (Biddle and Hamermesh 2016). However, it is safe to say that if you look at most of the post-1970 empirical research building on or otherwise reacting to the various other examples of statistical inference without probability that I have offered, you will find classical hypothesis tests, discussion and indications of statistical significance, and so on, taking a prominent place in the discussion of the empirical results.

IV. Discussion and Speculation

By the 1970s, there was a very 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 30 years earlier, developed when it did. But I will, in closing, 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 Daston and Galison (1992) in their writings on the history of scientific objectivity, and fruitfully applied to the history of quantification in the social sciences by Theodore Porter in his book *Trust in Numbers*.

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 sample data. The analyst's expertise is acquired through a training process sanctioned by a scientific discipline, as well as through experience making similar decisions using similar data subject to the surveillance of other experts. One's faith in the analyst's conclusions depends largely on one's assessment of the quality of his disciplinary expertise, but also of his commitment to the ideal of scientific objectivity. Elmer Working, in describing how to determine whether measured correlations represented true cause and effect relationships, was describing a process that depended heavily on expert judgment. Likewise, expert judgment played a big part in Gregg Lewis's assessments of how regression estimates of the union-non-union wage gap from various studies should be adjusted 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, the importance and propriety of applying expert judgment in the process of statistical inference was explicitly acknowledged by empirical economists. At the same time, mechanical objectivity was valued– it is easy to find examples of empirical economists 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 technology for achieving mechanical objectivity. In the post-1970s consensus regarding statistical inference in economics, however, application of this one particular form of mechanical objectivity became an almost required part of the process of drawing conclusions from economic data, taught in a standardized way to every economics graduate student.

In his case studies of the history of the pursuit of objectivity in the analysis of social and economic data, Porter emphasizes the tension between the desire for mechanically objective methods and the belief in the importance of expert judgment in arriving at and communicating statistical results. This tension can certainly be seen in economists' writings on statistical inference throughout the 20th 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. As I have pointed out, in the 1920s and 1930s there was disagreement over whether the phrase “statistical inference” should be applied to all aspects of the process of drawing conclusions based on statistical data, or whether it meant only the use of formulas derived from probability theory to create estimates from statistical data, measure the reliability of those estimates, and use those estimates to test hypotheses. The econometrics textbooks published after 1960 explicitly or implicitly accepted this second, narrower, definition, and their 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 economic models to be estimated. But the disciplinary knowledge needed for this tasks was to be taught in other classes, using other textbooks.

And something else was and is still left largely unspoken in the textbook descriptions of procedures for statistical inference: Even after the statistical model has been chosen, the estimates and standard errors calculated, and the hypothesis tests conducted, there is room for an empirical economist to exercise a fair amount of judgment, based on his specialized knowledge, before drawing conclusions from the statistical results. Indeed, no procedure for drawing conclusions from data, no matter how algorithmic or rule bound, can dispense entirely with the need for expert judgment (Boumans 2015, esp. pp. 84-85). And few empirical economists would deny that the interpretation of statistical results often involves a good deal of expert judgment. Empirical economists have not foresworn the need for expert judgment in the interpretation of statistical data.

This does not mean, however, that the widespread embrace of the classical theory of inference, leading to the near ubiquity of tests and measures associated with that theory in the empirical economics literature since the 1970s, was simply a change in style or 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 based 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.[36] Further, based on my reading of the empirical literature of the post-1970 in economics, I would say that once an author had justified an empirical model with a theoretical argument, a presentation and discussion of the coefficient estimates and significance tests associated with one or two versions of the empirical model was often a sufficient amount of "interpretation of results" to satisfy journal editors and referees.[37]

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 which had real consequences for what economists would and

[36] One could say that the standard errors and hypothesis test 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, one could not use it 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 was still possible to argue 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. These arguments tended to draw heavily on expert disciplinary knowledge.

[37] See Biddle and Hamermesh (2016) and Singleton and Panhans (2016) for more about the use of theoretical models in empirical microeconomic research in the 1970s and 1980s.

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, in discussing the shift to mechanical objectivity Daston and Galison (1992) emphasize concerns within the community of scientists regarding their own abilities to produce reliable and unbiased interpretations of empirical data. Porter makes the point that a discipline's embrace of mechanical objectivity has often been a response to pressures from outside the discipline, as when outside audiences whom the discipline wish to influence doubt the discipline's claims to expertise and its practitioners' abilities to render objective judgments (Porter 1995, p. ix). If the discipline's empirically-based knowledge claims are produced by a mechanically objective procedure, particularly a general purpose procedure endorsed by other, more respected disciplines, those claims are less suspect.

On initial reflection, I have not found an obvious candidate explanation for the widespread acceptance by economists of the classical theory of inference that fits in with either of these general patterns. As I have noted, since the 1920s individual empirical economists have been attracted to mechanically objective procedures because of the insurance they offered against interpretations that were influenced by the analyst's "personal equation". However, the intensity of such a concern, and thus the attractiveness of mechanical objectivity, will differ from economist to economist, and I see no reason to think that the prevalence of this concern began to increase significantly in the late 1950s and 1960s in a way that would have led to the widespread conclusions that the standardized procedures of classical inference theory were a necessary part of empirical economic analysis. Nor am I aware of important potential patrons of the economics profession 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 not produced using methods approved by the wider scientific community, such a classical hypothesis tests. Indeed, the conventional wisdom is that the public prestige of economists was at a high point in the early 1960s, the time when the use of classical inference methods was becoming *de rigueur* in the empirical research published in the leading journals.

Perhaps searching for an explanation that focuses on the classical theory of inference's status as a means of achieving mechanical objectivity emphasizes the wrong characteristic of that theory. In contrast to earlier forms of mechanical objectivity found in economics, such as the methods of time series decomposition taught in almost every textbook on economics statistics published between the wars, the classical theory of inference is a form of mechanical objectivity that 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 formalism and mathematical expression in economics was increasing, it may be this feature of the classical theory of inference that increased its perceived value enough to overwhelm longstanding concerns about its inapplicability to economic data. In other words, perhaps the chief causes of the profession's embrace of the classical theory of inference are those which drove the broader mathematization of economics, and one should simply look to the large and interesting literature that explores and debates possible explanations for that phenomenon rather than seeking a special explanation of the embrace of the classical theory of statistical 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 reasoning is how they are taught and learned. Like mechanically objective procedures in general, the classical theory of inference as used by economists can be taught as set of rules and procedures, recorded in a textbook and applicable to "data" in general. It can be taught to a dozen or more students simultaneously, without any of them necessarily having to apply the procedures to a particular sample of data. This is in contrast to the judgment-based reasoning that has always been a part of empirical economics, reasoning that combines knowledge of statistical methods with knowledge of the processes that generated the particular data being analyzed and the circumstances under which they were generated. This form of reasoning is hard 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 works as a research assistant for or writes a thesis under the supervision of an experienced empirical economist.[38]

During the 1950s and 1960s, the ratio of Ph.D. candidates to senior faculty in Ph.D. granting programs was increasing rapidly.[39] I conjecture that a consequence of this 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 would come in an econometrics classroom, using a book that taught, and defined statistical inference as consisting of, applications of classical inference procedures. As training in empirical economics came more and more to be classroom

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

[39] According to data compiled by the NSF (available at http://www.nsf.gov/statistics/nsf06319/pdf/tabs1.pdf) the annual number of economics P.D degrees earned per year in the US was relatively stable during the 1930s and 1940s at about 100 to 120. In the 1950-55 period, this increased to about 300 per year; during the 1955-60 period to 331 per year; during the 1960-64 period 432 per year, and during the 1965-1970 period 685 per year. Tabulation of the annual report s of "Degrees Conferred" published in the *American Economic Review* also shows a substantial increase in the annual rate from the early 1950s to the early 1960s (about 30%), although this is smaller than the increase indicated by the NSF data from the same two periods. Still, under any reasonable assumption about the processes determining faculty sizes at economics departments with Ph.D. programs, this growth in the number of Ph.D.s' conferred implied a large increase in the student to faculty ratio in those programs.

training, competence in empirical economics came more and more to mean mastery of the mechanically objective techniques taught in the econometrics classroom using a small number of similar textbooks, a competence displayed to others by application of those procedures. Less time in the training process 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 help 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, along with the rising prestige in the profession of mathematical modes of analysis and expression. But it does help to explain why the application of one particular set of mechanically objective procedures came to be regarded as a necessary and sometimes the only 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 worthy of further exploration, that is, that the changing nature of graduate education in economics might be a cause as well as a consequence of changing research practices in economics.[40]

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 20th 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.

---

[40] 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.

I believe there is something to be gained from increasing our understanding of this variety, both through case studies of the actual research practices of various individuals and groups, and work that 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 a lot of empirical research that went on in the 20th 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 considered convincing by economists.

References

Advisory Committee on Economic and Social Research in Agriculture of the Social Science Research Council. *Research Method and Procedure in Agricultural Economics*. 2 Vols. New York: Social Science Research Council, 1928.

Aldrich, John. Keynes Among the Statisticians. *History of Political Economy*, Vol. 40, No. 2 (Summer 2008), pp. 265-316.

Arrow, Kenneth J. Review of *A Textbook of Econometrics. Journal of the American Statistical Association*, Vol. 49, No. 266 (Jun., 1954), pp. 393

Bera, Anil. The ET Interview: George Judge. *Econometric Theory*, Vol 29, No. 1(February 2013) , pp 153-186

Biddle, Jeff E. Statistical Economics, 1900-1950. *History of Political Economy*, Vol. 31, No. 4 (Winter 1999), pp. 607-652.

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

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

Braun, E. W. and Sturges, Alexander. Price Analysis as a Guide in Marketing Control. Journal of Farm Economics, Vol. 19, No. 3 (Aug. 1937) pp. 691-706.

Bronfenbrenner, Martin. Review of Tintner's *Econometrics*. *The Annals of the American Academy of Political and Social Science*, Vol. 285 (Jan., 1953), pp. 181-182

Bronfenbrenner, Martin. The Consumption Function Controversy. *Southern Economic Journal*, Vol. 14, No. 3 (Jan., 1948), pp. 304-320

Brown, Frederick. Review of Mills, F.C. *Statistical Methods Applied to Business and Economics*, revised edition. *Economica*, New Series, Vol. 6, No. 24 (Nov., 1939), pp. 468-469

Carter, C. F. Review of *Econometrics* by Jan Tinbergen; *Econometrics* by Gerhard Tintner. *The Economic Journal*, Vol. 63, No. 249 (Mar., 1953), pp. 118-120

Christ, Carl. A Symposium on Simultaneous Equation Estimation: Simultaneous Equation Estimation: Any Verdict Yet? *Econometrica*, Vol. 28, No. 4 (Oct., 1960), pp. 835-845

Cox, Garfield. Review of *The Problem of Business Forecasting*. Edited by W. M. Persons, W. T. Foster, and A. J. Hettinger, Jr. *Journal of Political Economy*, Vol. 33, No. 3 (Jun., 1925), p. 365

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

Daston, L.J. and Galison, Peter. The Image of Objectivity. *Representations,* Special Issue: Seeing Science (Autumn 1992), pp. 81-128.

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

Douglas, Paul H.. "Henry Schultz as Colleague". *Econometrica* 7.2 (1939): 104–106.

Ezekiel, Mordecai. *Methods of Correlation Analysis*. New York: John Wiley and Sons, 1930.

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

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

Friedman, Walter. The Harvard Economic Service and the Problems of Forecasting. *History of Political Economy* Vol. 41, no. 1 (Spring 2009), pp. 57-88.

Ghysels, Eric. The ET Interview: Marc Nerlove. *Econometric Theory*, Vol. 9 (1993) 117-143.

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

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

Haavelmo, Trygve. The Probability Approach in Econometrics. *Econometrica*, Vol. 12, Supplement (Jul., 1944), pp. iii-115

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

Hoover, Kevin. On the Reception of Haavelmo's Econometric Thought. Journal of the History of Economic Thought, vol. 38 no. 1 (March 2014): 45-66.

Hurwicz, Leonid. Input-output Analysis and Economic Structure. *The American Economic Review* 45.4 (1955): 626–636.

Kaiser, David. The Postwar Suburbanization of American Physics. *American Quarterly*, Vol. 56, No. 4 (Dec., 2004): 851-888

Keynes, J.M. *A Treatise on Probability*. London: MacMillan and Co., 1921

Kiefer, Nicholas. The ET Interview: Arthur Goldberger. Econometric Theory, Vol. 5 (1989): 133-160.

Klein, L. R. Review of *Studies in the Structure of the American Economy*. *Journal of Political Economy*, 61.3, (1953): 260–262

Koopmans, T. C., editor. Activity Analysis of Production and Allocation. Cowles Commission Research in Economics, Monograph 13. New York: Wiley & Sons, 1952.

Koopmans, T. C., editor. *Statistical Inference in Dynamic Economic Models*. New York: Wiley and Sons, 1950.

Krueger, Alan B., and Timothy Taylor. An Interview with Zvi Griliches. *Journal of Economic Perspectives*, 14.2 (2000): 171-189.

Kuznets, Simon. "Economic Growth and Income Inequality". *The American Economic Review* 45.1 (1955): 1–28.

Kuznets, Simon. "Statistical Trends and Historical Changes". *The Economic History Review* 3.3 (1951): 265–278.

Kuznets, Simon. "Conditions of Statistical Research". *Journal of the American Statistical Association* 45.249 (1950): 1–14.

Kuznets, Simon, assisted by Lillian Epstein and Elizabeth Jenks. National Product Since 1869. Vol. 2. New York: NBER, 1946.

Lahiri, K. The ET Interview: G.S. Maddala. Volume 15, no. 5 (October 1999); 753-776

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

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

Leontief, Wassily. "Some Basic Problems of Stuctural Analysis". *The Review of Economics and Statistics* 34.1 (1952): 1–9.

Leontief, Wassily. *The Strucuture of American Economy 1919-1939: An Empirical Application of Equilibrium Analysis*. 2nd edition, enlarged. New York, Oxford University Press, 1951.

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

Lodewijks, Jon. The ET Interview: Professor Jan Kmenta. *Econometric Theory* Vol. 21, No. 3 (2005),:621-645

Mariano, Roberto S. The ET Interview: Lawrence Klein. *Econometric Theory*, Vol. 3 (1987): 409-460.

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

Mills, Frederick. An Hypothesis Concerning the Duration of Business Cycles. *Journal of the American Statistical Association*, Vol. 21, No. 156 (Dec., 1926), pp. 447-457.

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

Mincer, Jacob. "Investment in Human Capital and Personal Income Distribution". *Journal of Political Economy* 66.4 (1958): 281–302.

Mincer, Jacob. "On-the-job Training: Costs, Returns, and Some Implications". *Journal of Political Economy* 70.5 (1962): 50–79.

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

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

Morgan, Mary. Searching for Causal Relations in Economic Statistics: Reflections from History. Pp. 47-80 in *Causality in Crisis? Statistical Methods and the Search for Causal Knowledge in the Social Sciences*., ed. by V. McKim and S. Turner. South Bend, IN: University of Notre Dame Press, 1997.

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

Okun, Arthur. 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) 1963. (accessed at http://cowles.econ.yale.edu/P/cp/py1963.htm)

Okun, Arthur M. "Monetary Policy, Debt Management and Interest Rates: A Quantitative Appraisal" Cowles Foundation Discussion Paper No. 125 (June 1961) (accessed at http://cowles.yale.edu/sites/default/files/files/pub/d01/d0125.pdf)

Okun, Arthur M. "The Value of Anticipations Data in Forecasting National Product" Cowles Foundation Discussion Paper No. 40, (October 1957). (accessed at http://cowles.yale.edu/sites/default/files/files/pub/d00/d0040.pdf)

Persons, Warren M. Some Fundamental Concepts of Statistics. *Journal of the American Statistical Association*, Vol. 19, No. 145 (Mar., 1924), pp. 1-8

Persons, Warren M. “II. The Method Used”. *The Review of Economics and Statistics* 1.2 (1919): 117–139.

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

Rutherford, Malcolm. Institutional Economics at Columbia University. *History of Political Economy*, Duke University Press, vol. 36 (Spring 2004): 31-78, Spring.

Schultz, Henry. *The Theory and Measurement of Demand*. Chicago: University of Chicago Press, 1938.

Schultz, Henry. “A Comparison of Elasticities of Demand Obtained by Different Methods”. *Econometrica* 1.3 (1933): 274–308.

Schultz, Henry. “The Statistical Law of Demand as Illustrated by the Demand for Sugar”. *Journal of Political Economy* 33.6 (1925): 577–631.

Schultz, Theodore W.. “Testing the Significance of Mean Values Drawn from Stratified Samples”. *Journal of Farm Economics* 15.3 (1933): 452–475

Secrist, Horace. *An Introduction to Statistical Methods*, revised edition. New York: MacMillan, 1925.

Shiller, Robert J., and James Tobin. “The ET Interview: Professor James Tobin”. *Econometric Theory* 15.6 (1999): 867–900.

Teira, David. Milton Friedman, the Statistical Methodologist. *History of Political Economy*, vol. 39 (Fall 2007); pp. 511-528.

Working, Elmer J. Evaluation of Methods Used in Commodity Price Forecasting. *Journal of Farm Economics*, Vol. 12, No. 1 (Jan., 1930) pp. 119-133.

Working, Elmer J. Indications of Changes in the Demand for Agricultural Products. *Journal of Farm Economics*, Vol. 14, No. 2 (Apr., 1932), pp. 239-256

Working, Elmer J. New Indices of Agricultural Supplies and Carry-Over. *The Review of Economics and Statistics*, Vol. 19, No. 3 (Aug., 1937), pp. 144-153

Working, Elmer J. Graphic Method in Price Analysis. *Journal of Farm Economics*, Vol. 21, No. 1, Proceedings Number (Feb., 1939), pp. 337-345

Yntema, Theodore O. Review of Mills, F.C. *Statistical Methods Applied to Business and Economics*, revised edition. *Journal of Political Economy*, Vol. 47, No. 3 (Jun., 1939), pp. 449-450