Book Review Column

Reflections on The History of Econometrics Induced

Robert L. Basmann
State University of New York
at Binghamton.

Nancy J. Wulwick
Old Dominion University

1. Issues in Historiography

The history of science and the history of economics since their recent inception as academic disciplines have remained largely distinct fields of investigation. This has been partly for institutional reasons. Most historians of science have been historians or philosophers by occupation, while historians of economics usually have been economists. As a result, the histories of science and economics often have revealed different methodological concerns. In addition, the subject matter of the two disciplines has differed. The history of science has dealt with mathematics, technology, the natural sciences, the behavioral sciences, anthropology and sociology, yet rarely with political science or economics.

In the past, most historians of economics too took little interest in the history of other sciences. This situation markedly changed by the 1970s, when many economists viewed their discipline as in a state of crisis. They saw several conflicting theories in a position to dominate the discipline, with econometrics unable to choose between these theories. Consequently, historians of economics turned to the history of science for insights into how to treat the issues posed by the crisis. Scientific methodology and the historiography of science became important topics of the history of economics. The recent appearance of the first two books on the subject of the history of econometrics, *The History of Econometric Ideas* by M. S. Morgan and *A History of Econometrics* by R. J. Epstein reveals this concern with the methodology and historiography of science.

Morgan, a member of an economic history department, covers the history of econometric ideas from the late nineteenth century to 1944. She explains that econometrics originated as a result of the adaptation by economists of statistical tools and concepts borrowed from biometrics and astronomy. Her book divides the history of these borrowings into two major periods. During the period up to 1944, economists treated economic variables as essentially following deterministic laws presented by economic theory. This treatment acted as an obstacle to the use of statistical inference in the comparison of theories. In 1944 Haavelmo, presented a coherent framework in which to treat economic laws as probabilistic. Morgan sees this event as a "probabilistic revolution", which laid the basis for econometrics.

R. J. Epstein, an economist, covers the history of econometrics from 1944 to the present, with an emphasis on the period after 1944. He deals with the issues that have arisen during the recent debates among econometricians concerning the scientific basis of current methodology. Epstein views the modern history of econometrics as falling into two main periods, marked by the dominance of different "research programs" in conventional econometric practice. The research program of structural estimation intended econometrics to serve a role in economics that was analogous to that of the controlled experiment in physics. That research program was abandoned in the 1970s by numerous economists in favor of vector autoregressions, which lack a basis in economic theory and serve the purpose of forecasting.

The periodisation of the history of econometrics suggested by these two books is based on concepts current in the history of economics and science. A history of econometrics written thirty years ago most likely
would have taken a different perspective. Few historians or methodologists of science at that time thought in terms of revolutions or competing research programs. The main theme of the historiography of mature science was rather one of continuous progress. In fact, neither of our authors explicitly discuss progress.

Let us say that a science that successfully solves its conceptual and empirical problems is progressive. It would be possible in principle to write of econometrics as a progressive discipline. For a start, think of the evolution in econometrics of mathematical tools, data bases, computing facility, techniques of estimation, modelling, diagnostic tests, and diversity of uses. No wonder that econometric articles by the late 1950s composed a large part of the growing economics literature. Yet, despite the many specific successes of econometrics, our authors deliver the impression that econometrics has not achieved general progress. Epstein concludes that the discipline so far has failed to solve its original scientific problem, how to choose amongst plausible competing theories. Morgan's history ends too early to make a decisive assessment of econometrics, but she broaches the conclusion that econometrics so far has not pursued rigorously the "adequate experimental method" provided by Haavelmo (1944) to develop and test economic theories in a probability framework. A purpose of these histories (more explicitly so with Epstein than Morgan) is to help solve these conceptual deficiencies of econometrics. Both books, far from being counsels of despair, are ultimately confident of the capacity of econometrics to serve as the observational arm of economics.

Morgan's book treats four main subjects, business cycle models (1875-1938); demand analysis (1914-1934); the first macroeconomic models (of Tinbergen, 1935-1939); and formal econometric models (1935-1944). Given this choice of subjects, the primary organization of the study is based on the definition of problems in econometrics, rather than on time or historical people. Indeed, one chapter of the text takes the form of an imaginary dialogue that represents the arguments (about errors-in-variables and equations) without assigning any of the viewpoints (outside the footnotes) to historical people. The book relies almost entirely on published primary sources, in the form of journal articles, books, reviews, and official reports. The author primarily confines the narrative to what historians call internal factors, or intellectual events and influences within economics. In reconstructing the intellectual arguments, the narrative avoids a retrospective, or teleological approach to history: the author gives attention to "very poor work" and methods that are outdated and excludes crude comparisons of ideas that occurred during and after the time period under study. The narrative is oriented towards econometric problems as treated within each historical period.

The problem-oriented method implicitly adopted by Morgan is a familiar one in the history of science. The method includes four steps: (1) The historian defines a problem that the science raised at a particular period. (2) She examines the literature and artifacts of the time to ascertain the methods available to solve this problem. (3) Given the resources available to science, the historian must explain why scientists ever thought that their proposed solution of the problem was an adequate one. (4) Finally, the historian may assess the adequacy of the historical solution in light of current knowledge. The exposition in The History of Econometric Ideas of the notion of the business cycle as kindled by shocks, in the form sunspots or weather changes, offers an excellent example of how to understand in the context of the events of the late nineteenth century what economists since have mocked. We find that in discussing a number of other statistical notions, The History of Econometric Ideas is deficient in fulfilling the second and third tasks of the historian of science which requires that each attempt at problem-solving appear in its appropriate historical context.

Epstein discusses early demand analysis, the theory of structural system estimation presented by the Cowles Commission in the 1940s, the development of alternative schools of estimation in Britain and Sweden, the problem of exogeneity, and the use of vector autoregression. The focus of interest of the discussions lies on the development and the conceptual problems of estimation of equations in structural systems. According to this method of estimation, statistical inference is determined within the framework of a probabilistic model of the phenomenon under study and the generation of the data. As obvious from the richness of the narrative of the debates about structural estimation, the volume relies on a wide range of published sources and several archival sources. The book accurately reports technical arguments
and offers a lively auxiliary text from which to learn about modern econometric arguments at the graduate level.

In contrast to Morgan, Epstein suggests that external factors affected the history of structural estimation, in regard to both the directions taken and the pace of problem-solving by econometricians. For example, Epstein states that during World War II the Cowles Foundation supported econometric theory to help policymakers predict the structural changes that followed from economic policy interventions, such as wartime price controls. Researchers at the Cowles Commission broached the difficult problems posed by the concept of structural change, including the role of exogenous variables and the presence of relations invariant to policy interventions. Few resources were available to solve such problems during the liberal, Cold War period, when structural estimation was required to predict the effects of Keynesian stabilization policies on an unchanging structure. "The "new economics", Epstein comments, "was well established in the thinking of a great many political leaders and their academic counselors." They believed that the implementation of Keynesian fiscal tools would maintain steady prosperity. The leading researchers at the Cowles Commission, "Koopmans, Klein, and Marschak were anxious to take part in making this dream a reality" (pp. 87, 92, 187-188). It was not until the criticism of stabilization policy by new classical economics in the 1970s that econometric theorists reconsidered the research problems involved in the treatment of structural change following interventions into the economic system. Despite the apparent importance of finance, political ideology or institutional affiliation at several points in the narrative, A History of Econometrics lacks any explicit statement about the role of external factors on research.

No history is objective, in the sense of being free of the interpretation of the author. The barest narrative which only compiles recorded facts exercises judgement by excluding facts that are arguably relevant to the topic at hand. Even the historian's selection of topics upon which to write is disputable. However, the viability of certain topics is hard to dispute. For instance, no one actually would challenge the creative historian who chose to write another book on the topic of the Second World War. Similarly, the topics of the two histories under review have a timeless dimension that invites further publications: The major topic of Morgan's history is the "data-theory gap", that of Epstein's narrative, the problem of "multiple hypotheses", or "model-choice" - two themes that are inescapable in the primary literature of econometrics of their periods. The rest of this review critically discusses the problems of the data-theory gap and model-choice in light of the interpretations by Morgan and Epstein. We hope that the readers of Methodus will read both their informative and provocative books, having been convinced of the reasonableness of our remarks.

2. The Data-Theory Gap

Imagine that you are an econometrician looking at the statistical scattergraphs shown in Figures 1 and 2, which show the percentage changes and the level of price and quantity of pig iron for the period 1871-1912, according to data which H. L. Moore (1914) used in one of the first statistical studies in economics. Do you see a pattern in the scattergraphs? You may notice that the figure 1 shows the rudimentary pattern of a sloping band. Probably you will find it useful to describe the pattern of scatterpoints approximately by fitting some simple mathematical expression to the scattergraph. You now want to interpret this mathematical description of the scattergraph in terms of some real world activity. You may be reminded that abstract theories of consumer and factor demand exist, with their idealized concepts of commodities, prices, quantities demanded, and their internal definitions of the demand function. You may wonder, would it be valid to say that the mathematical description of the scattergraph is an approximation of the idealized demand curve of abstract demand theory? To answer this question you find that you must answer many other questions first. For instance, do the averaged prices and the annual quantities of pig iron fulfill the implicit definitions of price and quantity demanded within the formal theory? This question is typical of problems that lie at one edge of the data-theory gap. At the other edge of the gap are problems relating to the abstract theory itself. Solving these problems requires answers to many questions. An abstract theory defines several concepts that lack an observational counterpart. You will find that some versions of abstract demand theory incorporate special assumptions, and upon reflection, suspect that these are ad hoc assumptions. Which version of an abstract theory
should you select as corresponding to your fitted curve? Should you select a version incorporating Walras’s assumption that the marginal utility of a commodity depends only the amount of that commodity? Asking questions like the above reminds you that there is no such thing as the data-theory gap, but there are many individual data-theory gaps to be bridged in economics.

One thing all data-theory gaps have in common is this: If you want to bridge one of them securely you can’t slight work on either abutment. Slighting of work on the abstract theory end of the interpretative bridge usually involves a failure to get a clear distinction between what the theory actually does imply, and what it may seem to imply but really doesn’t. Mistaken but fashionable beliefs that a theory implies certain empirical conclusions it really cannot rigorously imply have been greater obstacles to general progress in economic science than the more readily detectable mistakes in handling observational data. There are many reasonable approaches to bridging a data-theory gap. Some start from the data end, some from the theory end; some concentrate on the statistical tools needed in the process. We shall emphasize the theory-end.

The issue of the relation between data and theory preoccupied the economists who pioneered the use of statistics to analyze markets in the interwar period. H. L. Moore’s (1914) study included one of the first of these statistical applications. It is Morgan’s interpretation that the treatment of demand and supply by in Moore’s book exposed key problems in the treatment of the data-theory gap. The attempt by statistical economists in the interwar period to solve these problems started the serious analysis of simultaneous equations in econometrics. We shall focus on Morgan’s discussion of Moore (1914) as the basis for a critical commentary of her historical treatment.4

Economic Cycles: Their Law and Cause (1914) studied the markets for several representative crops and pig iron, or crude iron cast in blocks or pigs, which was an unusually good barometer of the trade cycle (Moore, p. 105; Mitchell 1913, p. 199). Moore’s problem was to derive the demand curve for these commodities from statistics. He estimated the relation between price \( p \) and quantity \( q \), having expressed the data in percentage terms,

\[
P = (p_t - p_{t-1})/p_{t-1}
\]

\[
Q = (q_t - q_{t-1})/q_{t-1}
\]

The use of percentages was designed to eliminate the effects of increasing population and price-level changes and facilitate the estimate of price elasticity of demand (pp. 69, 83, 105). Moore estimated the linear relation between \( P \) and \( Q \) using the method of least squares. Given Moore’s data, the sources of which he carefully recorded, we replicated his estimates for pig iron:

\[
(1) \ P = -4.571 + 0.5228 \ Q, \ R^2 = 0.289071, \quad (0.110) (0.000) \quad , \quad DW \ stat= 2.126527
\]

Our estimates of the constant and the slope of the line and the coefficient of determination are virtually the same as those found by Moore (p. 114). The important point to note is that the estimate of the slope in the case of pig iron (in contrast to crops) was positive. Moore referred to equation (1) as "a new type of demand curve", which did not suffer from the statistical limitations of Marshall’s demand curve (Mirowski, 1990). “There have been grave doubts”, Moore thought “as to whether the practical difficulty of deriving the [Marshall’s] demand curve would ever be overcome” (p. 65). Unfortunately, Moore used the terms ‘demand’ in ways that for many economists elicited meanings that Moore probably never intended, yet were entrenched in Marshall’s theory of demand.

As Moore explained, his statistical demand equations for hay and other crops

“unlike the classical theory of demand which was limited to the simple enunciation of this one characteristic, ceteris paribus, ... apply to the average changes that society is actually undergoing. They summarize the changes in prices that are to be expected from changes in the supply of the commodity, thus enabling one to predict the probable variation in price that will follow upon an assigned variation in the amount of the commodity” (p. 77).

His account of the place of the statistical upward-sloping “demand” curve for pig iron in the trade cycle went as follows: A rise in crop yields prompts increases in trade and the demand for crops and factor inputs including pig iron, which causes a general rise in input prices, including that of pig iron. The opposite changes occurred given a fall in crop yields (p. 28). Hence the price of producer goods varied positively with the “increase and decrease of ‘ultimate demand’ which lies back of the flow and ebb of trade” (pp. 105, 111-2). The relative changes in
the price and the quantity of pig iron were correlated because both variables depended on the crop cycle (pp. 114, 116). No wonder Moore referred to his curve as a "new type of demand curve"!

The treatment of demand and supply in Economic Cycles elicited critical comments by several economists of the period, comments that Morgan surveys. R. A. Leffeldt (1915) agreed with Moore that Moore's equations that described the relations between "demand and price ... are not 'demand curves' as defined by the theoretical economist," because all "other things" were not held constant. However, he criticized Moore for calling curve for pig iron (our equation 1) "a new type of demand curve, sloping the opposite way to the usual kind! But the curve (for pig iron) is not a demand curve at all, but much more nearly a supply curve". Leffeldt hypothesized how a shifting demand curve for pig iron confronting a fixed supply curve could have traced out an approximate upward-sloping supply curve in Moore's data (Leffeldt, p. 411; Morgan, p. 167). P. G. Wright's review (1915) of Economic Cycles, which devoted relatively little space to the issue of Moore's upward-sloping demand curve for pig iron, asserted Marshall's law of demand based on diminishing marginal utility and hypothesized that Moore's statistical curve was the effect of rapid, continuous increases in the demand for pig iron in conditions of constant supply (Wright, pp. 632, 638-9; Morgan, p. 167). In an essay on statistical demand curves, E. J. Working (1927) corrected Wright, explaining that Moore's curve for pig iron resulted from the shifting back and forth of the relation between the relative change in price to the relative change in quantity demanded, given comparative stability on the supply side of the market (pp. 214, 222). In his obituary of Moore, Stigler (1962) accepted the critical reviews of Leffeldt and Wright, adding that although Moore's demand curve for pig iron "was permissible for a predictive cycle theory, it did not yield the demand functions of economic theory" (pp. 361, 367). In sum, as Epstein remarked, "it has become part of the folklore of econometrics that Moore did not know that a shifting demand curve could trace out a static supply curve" (pp. 17-8).

A History of Econometric Ideas accepts the folklore about Moore. According to Morgan, Moore's empirical work produced a relationship which he interpreted as a positive demand curve for pig iron. This of course contradicted the negative relationship stipulated by standard economic theory, and he used this opportunity to launch a strident attack on conventional economic method and demand theory (p. 28).

She continues,

Moore's approach was a mixture. At its worst, it involved both the unthinking application of theory to data and the adoption of empirically derived relationships without reference to theory (p. 141).

Does Morgan want to give the impression that Moore's application of theory to data was unusually unthinking for his time (and ours)? Does she want her readers to believe that critics of Moore's positively sloped "demand curve" for pig iron were significantly less "unthinking" in their applications of theory? It is important to consider now in what respects Moore's approach in Economic Cycles may have been a mixture, what he actually mixed in his approach, to what theory he should have, but did not, refer his empirically derived relationships, and especially what was "unthinking" about his application of theory to data.

In challenging the characterization of Moore's approach as "unthinking" we are not concerned chiefly to defend Moore against an undeserved charge. Of greater importance is that the incident of the finding of an upward-sloping "demand curve" for pig iron exemplifies an important historical fact concerning the way in which many economists and econometricians have viewed the relation between economic theory and econometric analysis: In making critical evaluations of empirical econometric studies, many economists have felt free to use economic theory in a question-begging spirit. They, and other critics, assume as already demonstrated what has not been adequately tested before, and what econometricians propose to test. This question-begging spirit is tolerated by the majority of economists. Unfortunately, Morgan is blind to its practice by the critics of Moore (Wulwick, 1992).

What was the state of theory as it pertained to the "law of demand" in the English-speaking world around the time when Moore was deriving his famous upward-sloping demand curve for pig iron and the critical reviews of Economic Cycles were written? At the time, Marshall maintained that the demand curve for each product must have a negative slope, given three assumptions,
an additive utility function according to which the utility of a commodity is a function only of the quantity of the commodity, diminishing marginal utility (two assumptions of Walras) and constant marginal utility of money income (Silberberg and Walker, 1984). Of course, Marshall recognized that positively sloped demand curves existed, but he assumed that the case of the Giffen good, as he called it, was rare. While the Marshallian utility function remained popular amongst most economists around 1944, mathematical economists preferred the theory of demand based on the convex indifference map introduced by Edgeworth (1881). Work by V. Pareto in Italian in the 1890s (cited in Stigler 1950, p.136) and his Manuel in Italian (1906, or the French translation 1909, pp. 583-5), then by Zawadski (1914) in French and E. Slutsky (1915) in Italian argued that in the general case the demand for an inferior good may be upward-sloping. The theory of utility and demand essentially reached its present general form in the Economic Journal of 1913, at the hands of W. E. Johnson (Allen 1936, p. 120). From Johnson English-speaking economists could have learned that the concept of the convex indifference map and linear budget line implied this much and no more about the own-price slopes of demand functions for consumer and producer goods: In a system of $N$ demand functions, no more than $N-1$ of them can have positive slopes with respect to own-prices at any given set of prices and total expenditure, $(p_1, p_2, ..., p_N, E)$. Economics could offer no theoretical support for the generalization that all or most demand functions are downward-sloping with respect to own-prices at all sets of prices and total expenditures $(p_1, p_2, ..., p_N, E)$. In short, the neoclassical theory of Moore’s day showed that without assumptions designed to imply that all goods have negatively-sloped demand curves, the theory of demand provided essentially no support for the generalization known as the “law of demand”. We lack any evidence that Moore’s decision to refrain from using the popular, Marshallian “law of demand” was motivated in any way by the flux in demand theory. He certainly was aware (apparently in contrast to his critics) that there was a problem concerning the formal derivation of Marshallian demand curves. The important point is that to demean Moore as Morgan has because his statistical derivation of an upward-sloping demand curve for pig iron gave him a chance to attack demand theory makes no sense in light of the fact that the only rigorous theory of demand in Moore’s day permitted upward-sloping demand curves. From the point of view of historiography, two problems arise here: Morgan has neglected the task of the historian to explain why Moore himself thought that he adequately explained the statistical demand curve in the context of what he saw as the dynamics of the business cycle. She also has ignored the fact that one actually could explain Moore’s demand curve for pig iron given the contemporary neoclassical theory of demand.

It appears to have been even more widely believed in Moore’s time than now that scientific theories could be tested piecemeal. The physicist and historian, Pierre Duham (1906) stressed in La théorie physique, son objet et sa structure the great difficulty of meaningfully testing merely a part of a “theory. Some scientists, Duham wrote, generally think that each one of the hypotheses of a theory can be taken in isolation, checked by experiment, and when many tests have established its reality, given a definite place in the theory. Duham argued that in any science, when the data seem to disconfirm an hypothesis, the whole theory of which that hypothesis is a part is disconfirmed. While the finding that data and an hypothesis disagree calls for a modification of some hypotheses constituting a theory, this finding cannot single out any one hypothesis – in particular, not the one apparently “tested”, as necessarily requiring modification. From this perspective, it would have appeared impossible for Moore to have treated the statistical upward sloping for pig iron in light of the theory of demand without looking at the whole system of demand curves. Economists in 1914, however, lacked the understanding of how to do this.

Neither Stigler nor Morgan asked whether economists in 1914 should have been surprised on empirical grounds to encounter a positively-sloping demand curve for pig iron (Stigler 1954, p.229-231; 1962). For such surprise to have been warranted, there would have to have been already a considerable amount of experience with econometric testing and estimation of such demand system. There would have had to be available a lot of quantitative and institutional knowledge of the pig iron market and markets closely related to it. Those conditions were not met in 1914. Morgan criticizes Moore for viewing data in a way that contradicted economic theory; she then criticizes Keynes for viewing data according to the dictates of economic theory.
J. M. Keynes, according to Morgan, took the view that "if the [statistical] results are not in accordance with [one's own] theoretical preconceptions, then blame the method [econometrics] and the data but not the theory" (p.124).

That view of Keynes was similar to the one stated by T. Haavelmo (1943, p.13), though Morgan's tone is rather more strident. To confirm that view, Morgan refers to two commentaries by Keynes - his 1939 book review of the first Keynesian econometric model of the business cycle, by J. Tinbergen and his 1940 reply to Tinbergen. Her conclusion appears implausible once one reads those commentaries in light of a fuller range of primary sources, including Keynes's correspondence and *The Treatise on Probability*. Our discussion of Keynes's views of econometrics, like other such discussions, uses those sources.

The League of Nations had asked Keynes to review a draft of the Tinbergen's *Statistical Testing of Business Cycles, Volume One* prior to publication. Keynes, who was recovering from a heart attack, took the review to be "severe holiday task". One reason for Keynes's dismay was, as he wrote to the League, "my lack of familiarity with the matter". He suggested that prior to publication the League send the draft to "someone more competent in these matters than I am" (CW p. 289). Even at the end of the published book review, Keynes stated that "I hope that I have not done injustice to a brave pioneer effort". Many of the points in his published review appear in the form of questions. In light of the fact that Keynes's health prevented him from attending the League of Nations with Harrod and other economists for two days of questioning of Tinbergen, it is possible that the questions in Keynes's book review were not all rhetorical (Bateman 1987, pp.109-10). Keynes's questions (which Morgan calls "invalid" and "quibbles") touched upon many of the difficulties that modern econometricians have experienced when working with the linear regression model, including simultaneity bias, omitted variable bias, multicollinearity, how to model expectations and shocks, the treatment of unobserved variables, spurious correlation from the use of proxy variables, how to measure variables, misspecified lags and trends, incorrect prefiltering of data, and statistical vs economic significance. Evidently, Keynes believed that multiple correlation was an extremely difficult project. Indeed, he did not mince his words about the effect of a failure to prepare the required tasks adequately, referring to the dangers of dealing in "black magic" and making a "brand of statistical alchemy" still not "ripe to become a branch of science" (CW p. 320). At the same time, Keynes was in a position to see that with improved computing machines and better economic statistics, many more economists besides Tinbergen (whom he likened to "Newton, Boyle and Locke [who] all played with alchemy") would be engaged in multiple correlation analysis (CW pp. 317, 320).

The "central question" for Keynes concerned "the logic of applying the method of multiple correlation to unanalyzed economic material" for the purposes of statistical inference (CW p. 286). Apparently unaware of recent developments in statistical theory, Keynes (1939) referred the reader to the discussion of statistical inference in *The Treatise on Probability*. As *The Treatise* explained, the statistical analysis of natural and social data potentially served two purposes, description and induction. In light of the statistical theory of the late nineteenth century, Keynes argued that "single correlation" of natural or social variables required a demonstration that the observed relative frequencies of the variables behaved like the results of urn drawings (1921, p. 426). This meant that the observed relative frequencies had to be stable for any very long series of data and the same (outside of random influences) across any subseries of that data. According to Keynes, data series from biometrics, chemistry and atomic physics fulfilled the condition of stability for the purposes of induction. In contrast, as the *General Theory* (1936) implied, Keynes thought that some frequencies in macroeconomics were "prima facie extremely unpromising material" for the purposes of statistical induction. While Keynes accepted the use of multiple correlation to reveal the cost curve of a firm or the time lags involved in the multiplier, when it came to the case of the causes of investment. Tinbergen's investigation of the stability of the estimated coefficients, limited as it was by amount of available data, failed to persuade Keynes of the usefulness of the method of correlation. The central message here – indeed, in Keynes's whole discussion of the use of econometrics – was that economists should test a model before they accept the estimates of its parameters. This prescription, as the next section of our review shows, was familiar to econometricians.
Keynes was concerned that econometricians would not arrive at their results objectively. Comparing econometricians to the seventy translators of the Septuagint who according to legend though shut up in separate rooms gave identical translations, he asked, "[w]ould the same miracle be vouchsafed if seventy multiple correlators were shut up with the same statistical material? Any anyhow, I suppose, if each had a different economist perched on his a priori, that would make a difference to the outcome" (CW pp. 319-20). We now consider the problem of conflicting outcomes due to the choice of different estimating models.

3. The Problem of Model Choice

Model choice is the current name for the what econometricians in the early days of structural estimation called multiple hypotheses. J. Marschak, as research director of the Cowles Commission in the 1940s, viewed the problem of multiple hypotheses as "the still remaining core of current criticisms against the application of statistics to economics" (Epstein, p. 106). M. Friedman, who attended many Cowles Commission seminars during 1946-1948, persistently asked his audience, "[h]ow does one choose a model, given that numerous possible models exist for the same period?". Marschak apparently answered "that the true model would be revealed by the accumulation of more data because the coefficients of the false models must asymptotically approach zero" (Epstein, p. 107). As we shall see, Marschak underestimated the difficulty of the task at hand.

Epstein’s narrative reveals the problem of multiple hypotheses arising in three main forms (pp. 53, 104, 107). The most fundamental is the problem of observational equivalence, because it is inherent in the theoretical concept of data generated by a market that clears demand and supply. According to Haavelmo (1944), observational equivalence is a three term relation involving two different hypotheses and one set of shared variables. The hypotheses are observationally equivalent, or mathematically equivalent with respect to the set of variables, if, and only if, they both assign identical hypothetical probabilities to every event describable by those variables. Observationally equivalent hypotheses are "... indistinguishable from the point of view of observations" (p. 74). Consequently, it is impossible for statistical inference to decide between observationally equivalent hypotheses.

The second problem of multiple hypotheses arises when several different hypotheses that are not observationally equivalent receive approximately the same support from a given batch of relevant data when appropriate tests are performed and published. Empirical data are capable of deciding between the two hypotheses but the given batch fails to do so.

Finally, the problem of multiple hypotheses arises whenever economists disagree about different models without having inquired whether the models are observationally equivalent or having performed the joint tests of significance of the set of restriction on the models. Epstein’s narrative leaves the reader with the impression that the failure to test has been common in econometrics (pp. 4, 104-6, 224).

We shall focus on the second type of problem of model choice, which involves the testing of multiple hypotheses in a probabilistic framework.

The "probabilistic revolution", according to Morgan, "occurred in econometrics with the publication of Trygve Haavelmo’s The Probability Approach in Econometrics in 1944" (p. 229). Haavelmo’s probability model came from R. A. Fisher’s (1915) model of a hypothetical infinite population. According to the famous statistician, P. C. Mahalanobis, R. A. Fisher introduced for the first time the brilliant technique of representing a sample of size n by a point in n-dimensions. Such representation has proved extremely useful in subsequent work not only in the theory of distribution, but also in other fields of statistical theory such as the work of J. Neyman and E. S. Pearson (Mahalanobis 1938, p. 268).

Haavelmo adopted probability theory and R. A. Fisher’s likelihood concept as a result of his familiarity with the work of Jerzy Neyman (Morgan, p. 242). In a series of articles published between 1928 and 1938, Neyman and E. S. Pearson extended R. A. Fisher’s definition of likelihood of a simple statistical hypothesis to composite statistical hypotheses, or classes of simple hypotheses.

The "anti-probability lobby", as Morgan calls it, argued that statistical theory was inapplicable to economics because of "[t]he problems of non-independendence of successive observations and non-homogeneous time periods" (Morgan, p. 243). Armed with R. A. Fisher’s concept of a hypothetically infinite population and the
representation of a sample of size \( n \) as a point in \( n \)-dimensions, Haavelmo sought to counter the prevailing view among economists that

"[p]robability schemes ... apply only to those series of observations where each observation may be considered as an independent drawing from one and the same 'population'."

Haavelmo then repeated R. A. Fisher's argument (though without citing Fisher) that it is not necessary that the observations should be independent and that they should all follow the same one-dimensional probability law. It is sufficient to assume that the whole set of, say \( n \), observations may be considered as one observation of \( n \) variables (or a "sample point") following an \( n \)-dimensional joint probability law, the "existence" of which may be purely hypothetical (Morgan, p. 243).

The *Probability Approach in Econometrics* in Chapter 4 included a concise exposition of the Neyman-Pearson theory of testing statistical hypotheses and estimating parameters. The book discussed the questions, "[d]oes this scheme [the Neyman-Pearson theory of tests of hypotheses] represent a useful instrument by which to deal with the problem of verifying economic theories? Can it help us to understand better the nature of these problems, and to reach practical solutions of them?" (p. 68). Haavelmo's answers to those questions were clearly affirmative. Yet, as Morgan states in the concluding sentence to her book, "by the 1950s the founding ideal of econometrics, the union of mathematical and statistical economics into a truly synthetic economics, had collapsed" (p. 264, 261).

We shall consider why the collapse occurred so quickly.

Undoubtedly, statistical inference in simultaneous structural systems is a difficult task. Under Gauss-Markov conditions, the distribution of the regressors depends neither upon their unknown coefficients nor the variances and covariances of the disturbances. This ideal condition disappears with the introduction of just one lag of the dependent variable. Then each coefficient and the asymptotic standard deviation of its maximum likelihood estimator vary together, and the latter varies with other coefficients as well. The complexity is compounded when the regression equation is a reduced-form equation derived from a system of structural equations, where reduced-form coefficients depend on at least one coefficient in each of several structural equations. In these cases, the *statistical intuition* acquired through considerable experience with regression under the Gauss-Markov condition is of no avail without long and tedious mathematical analysis.

The early "revolutionaries" of econometrics, as Epstein suggests, underestimated the magnitude of the difficult task ahead of them and did not organize to get it done (pp. 62, 105, 108, 110-4, 127-9). Realizing the goals of the "probabilistic revolution" and the Cowles Commission program required the work of many hands over a long period of time. The project needed motivated and trained people to develop a broad and deep infrastructure of theorems and further discoveries. Unfortunately, the Cowles Commission pioneers turned from these essential technical and organizational tasks too soon to formal economic theory and forecasting.

The models of L. Klein (1950), which have had great influence on the teaching of econometrics in the postwar period, provide a major example of the failure of econometricians to pursue the "probabilistic revolution" to fruition. According to Morgan, following Haavelmo's work in the 1940s, econometricians believed that economic theories should be explicitly formulated as statistical hypotheses, so that methods of inference based on probability theory could be used to measure and test the relationships. Klein was in the process of testing a cycle (or macroeconomic) model for the Cowles group, using *their newly developed methods of estimation and inference* (p. 56; emphases added).

We shall argue that Klein misused the Cowles Commission's "newly developed methods of estimation and inference" in his *Economic Fluctuations in the United States 1921-1941*, an argument for which Epstein lends support (Klein 1950, Epstein, pp. 104-6, 111, 115-8). In so far as statistical testing and estimation of econometric models is concerned, there is no continuity of thought or practice with the probability approach of Haavelmo. Klein, in practice, if not in profession, was the first influential defector from the "probabilistic revolution".

Klein was a member of the Cowles Commission during 1944-1947, when the new methods were developed, including T. W. Anderson's development of the test statistic for
identifying restrictions (Epstein, p. 100). This statistic is used to evaluate the observed significance level for a null hypothesis pertaining to the set of restrictions. The observed significance level for a test of identifying restrictions is the chance of observing a value of the test statistic at least as large as observed, if the hypothesis is true. Klein failed to report the observed level of significance of the set of economic restrictions in Model I of Economic Fluctuations in the United States 1921-1941. Basmann and Hwang (1990) calculated the observed level of significance to be
\[ P = 29 \times 10^{-11}, \]
from the data used to estimate the model by limited information maximum likelihood methods that Klein said he used. This means that a population of which Klein's Model I is true would produce a test statistic as large as 94.98 by chance about 3 times in one trillion. Such a minute observed significance level would ordinarily motivate strong reluctance to accept the economic theory incorporated in Model I.

C. Christ (1951) presented various tests of a revised version of Klein's Model III. Christ's out-of-sample tests of Model III detected serious shortcomings of that model.

Epstein reports that these results "were especially discouraging in that no structural changes were presumed to have taken place and the within sample fits were extremely close" (p. 111). However, close within-sample fits do not imply that tests of hypothetical restrictions imposed on the fitting will not produce observed significance levels that are infinitesimal like that for Model I above. It is not possible to compute a system test (that is, a test of all restrictions simultaneously) in order to ascertain the observed significance level for Klein's Model III. The restrictions are shielded from tests because the number of predetermined variables exceeds the sample size. This means that a considerable number of reduced-form parameters have to be set equal to zero before estimation can be performed. In effect this amounts to an ad hoc null hypothesis, whose observed significance level in a larger sample might itself be infinitesimal to begin with. Hence, Christ tested restrictions on his revised version of Klein's Model III equation-by-equation, using T. W. Anderson's identification test mentioned above (Christ 1951, Table 4, p. 82). Unfortunately, Christ calculated the test statistics incorrectly, and greatly overstated the observed significance levels. Discussants of the monograph, among them M. Friedman, L. R. Klein, and J. Tinbergen, failed to notice Christ's mistake, or they failed to call it to Christ's attention. The corrected values, as compared to Christ's estimates (in parentheses) are: investment, 0.066 (1.0); inventory, 0.0178 (2.0); production, 0.00025 (3.4); private wages and salaries, 0.000015 (4.0); employment, 0.000041 (4.2); wage rate, 0.00021 (5.1); consumption, 0.0017 (6.2); owned housing, 48 x 10^{-11} (7.0); rent, 0.000188 (9.0); rental housing, 0.20 (10) (Christ, Table 2, pp. 72-8).

With two exceptions, the observed significance levels are clear signals to re-think the underlying theory and assumptions of Model III before using it for any purpose.

A hallmark of the "Cowles Commission method" was this methodological maxim, (i) it is natural to use the test of the set of restrictions on a given equation before proceeding to estimate its coefficients, which is a subsequent step in computational procedure. It is likewise natural to abandon without further computation a set of restrictions strongly rejected by the test (Koopmans and Hood 1953, pp. 184-5).

It is safe to say that this maxim would preclude the making of out-of-sample predictions or forecasts using reduced-forms calculated from structural estimates of Klein's Model I, III or any other model with a tiny level of observed significance.

Clearly, the replications of the original "experiments" of Christ, Klein and Moore provide a vital source from which to gain a greater understanding of the purpose, the options and the limitations of newly discovered economic constructs and econometric methods. We hope to see more histories of econometrics, with a sharper focus on statistical theory and a "hands-on" approach, now that A History of Econometrics and The History of Econometric Ideas so solidly have laid the necessary groundwork.

Notes
1. The authors thank Thomas Mayer for his helpful comments.
2. The chapter on the probabilistic revolution appeared also in Morgan (1987).
4. According to the later the Slutsky-Hicks tradition, all income-compensated demand function must be negatively sloped with respect to own-price (Hicks 1939, Value and
5. Mirowski (1990) thought that Moore had gone "sour" on neoclassical theory generally.

6. Morgan briefly discussed Moore's theoretical rationale for his upward sloping demand curve for pig iron on pp. 28, 166.


10. Letter from Keynes to Tyler, 23 August 1938, CW p. 285

11. p. 318.

12. From Keynes, CW, 1939, 1940. Part of this list appears in Hendry 1980, p. 296; Bateman 1987, p. 108.

13. Keynes (1939) cites work he did thirty years ago (CW p. 135). *The Treatise on Probability* was written during 1908-1912 and published in 1921. Part V is on statistical inference.


15. 1921, p. 428; Letter from Keynes to Harrod, 16 July 1938, CW p. 229.


18. Basmann and Hwang, p. 139. In contrast to the significance level stated in Basmann and Hwang (1988), the significance level in this paper pertains to model I with the trend left in. Wenti Wang did the reestimate. (Likelihood ratio test statistic = 84.19. There are 16 restrictions. Observed significance level is calculated from the asymptotic distribution of the test statistic, i.e., chi-squared with 16 degrees of freedom.)


20. He used logarithms to the base 10 where he should have used natural logarithms. This resulted in division of the correct values to test criteria by 2.302585, causing considerable exaggeration of observed significance levels.

21. It remains, as Epstein remarked, "one of the mysteries of the evolution of econometrics that they (the tests of identifying restrictions) are seldom reported in practice" (p. 100).

References


Book Review


