THEORY AND MEASUREMENT[*]:
EMERGENCE, CONSOLIDATION AND EROSION OF A CONSENSUS

Jeff E. Biddle and Daniel S. Hamermesh[**]

[*]This was the motto of the original Cowles Commission.
https://en.wikipedia.org/wiki/Cowles_Foundation

[**]Professor, Michigan State University; professor, Royal Holloway, University of London, Sue Killam professor emeritus, University of Texas at Austin and research associate, IZA and NBER.
We thank Robert Goldfarb and Thomas Wiseman for helpful suggestions, and participants in the 2016 HOPE conference on "Becoming Applied" for very useful discussions.

I. Background and Introduction

A number of economists have recently argued that significant changes have taken place in the conduct of empirical microeconomic research. Joshua Angrist and Jorn-Steffen Pischke (2010) wrote of a "credibility revolution" in empirical economics, driven by "a focus on the quality of empirical research designs." They were particularly encouraged by the growing use of what were being called "quasi-experimental" designs. This research did not involve new or complex statistical estimators, but instead used applications of standard least-squares regression such as "regression discontinuity" or "difference in difference" methods. A "research design" was a proposal for applying one of these flexible approaches to observational data, along with an argument that the circumstances that had generated these data would allow "credible identification" of a "causal effect." Central to the credibility revolution was the growing ability of empirical economists to recognize and properly analyze situations in which "human institutions or the forces of nature" had created "informative natural or quasi-experiments".

Many economists, however, see in Angrist and Pischke's credibility revolution not only a change in the nature of the data and methods used in empirical microeconomic research, but also a change in the role of economic theory. Keane (2010, p. 54), in a comment on the Angrist-Pischke manifesto, objected to their "notion that empirical work can exist independently from, or prior to, economic theory".[1] Others see the change reaching beyond the way that research is conducted to the way that future microeconomic researchers are being trained. Deaton (2009) questioned the value of the new "design-based" approach to empirical microeconomic research, but also lamented, "… the wholesale abandonment in American graduate schools of price theory",

[1]Cf. Angrist and Pischke's comment that, "Economic theory helps us understand the picture that emerges from a constellation of empirical findings, but does not help us paint the picture (p. 23)."

commenting that "empiricist and theorist seem further apart now than at any time in the last quarter century'.

We seek to determine whether there has indeed been a changed role of theory in empirical microeconomic research, both in the selection of questions to be explored and in the choice of statistical methods and data used to answer them. We do this by analyzing large samples of empirical microeconomic articles in the top scholarly journals in economics from three different periods: 1951-55, 1973-77, and 2007-08.

Two hypotheses guide this analysis. The first follows from Backhouse and Cherrier's (2014) sketch of the emergence, by the 1970s, of a broad consensus regarding the nature of microeconomic theory and how it should be used in empirical research. Building on their observations, we locate the seminal ideas underlying this consensus in three distinct approaches to empirical microeconomics in the early post-war period: the Cowles Commission approach to "econometric" research; the style of empirical microeconomic research developed at the University of Chicago beginning in the 1950s, and an approach to empirical research approach exemplified by Leontief's input-output analysis.

Each of these approaches embodied the idea that microeconomic theory, defined as the logical analysis of the consequences of optimizing behavior on the part of individual economic agents, was "prior to" and necessary for empirical analysis. This idea was reflected by the prominence typically given to formal theoretical models in empirical studies. The model would be presented and analyzed in advance of descriptions of the empirical techniques and results, with these latter tightly linked to the theoretical model. The beliefs about proper research practice embodied in these approaches exerted a strong influence on the conduct of empirical research and the training of researchers in the four decades following World War II, and our empirical measures

are designed to detect important manifestations of them in published research. We expect the influence of these ideas to be noticeable in the empirical literature of the 1950s and dominant in the empirical literature of the 1970s.

Our second hypothesis is that the share of articles with these characteristics was noticeably lower in the early 2000s than in the 1970s, due to the emergence of the new approach to empirical research that attributed less importance to formal theoretical models of optimization and equilibrium as vehicles for defining specific research questions and hypotheses. Instead, the posing of questions to be answered or hypotheses to be tested is more likely to occur through informal discussions of earlier models, perhaps mixed with the conjectures of other social scientists or commonsense opinions of participants in policy debates. The Angrist-Pischke article and the studies it cites as early milestones in their credibility revolution serve as exemplars of this approach, which we will call the "experimentalist paradigm." [2]

II. Measurement with Theory: The Emergence of a Post-War Consensus in Empirical Microeconomic Research

By the 1970s a broad consensus had developed regarding proper research practices in empirical microeconomics, one which displayed the influence of three distinct ways of conducting empirical economic research that emerged in the 1940s and 1950s. We label these approaches the Cowles Commission approach, as exemplified by Haavelmo's (1944) treatise on the probability approach in economics and Koopmans' contributions to the "measurement without theory" debate (Koopmans 1947, Vining and Koopmans 1949); the "Chicago" approach, which among other things was seen as an operationalization of Friedman's (1953) methodological ideas; and "Leontief's approach", the pioneering application of which was Leontief's input-output analysis.

[2]The label is also used by Angrist and Pischke and seems apt, given the liberal use of the language of controlled experimentation by the followers of this approach when describing their non-experimental research, e.g., frequent references to treatment effects, treatment and control groups, placebo tests, and so forth.

Proponents of the three approaches differed in their guidelines about empirical strategy, and the post-1970 consensus displayed a corresponding heterogeneity about these matters. But the three approaches shared a similar set of ideas about the meaning of "microeconomic theory" and the role of theory in research, and the consensus embodied this similarity.

This view can be summarized by the following propositions: 1) Theory consists of models of individual optimizing agents and/or equilibrium in market-like interactions between optimizing agents, preferably expressed in mathematical form. 2) Such models, i.e. "theory," should be used to identify and define the quantities and relationships to be measured by empirical methods, as discovering and measuring relationships between economic phenomena is of primary importance for empirical research. 3) Theoretical models can be analyzed to produce hypotheses that can be tested empirically. Generating and testing such new hypotheses is the central activity of economic science. 4) Theory provides significant guidance to empirical design, often providing important information about how to measure key concepts, the appropriate sort of data and estimation technique to use, etc.

In short, theory, as defined in our first proposition, identifies what empirical researchers should be looking for and points out what methods and data to use. Economic science progresses as theory is used to indicate new relationships to search for, and as empirical research determines which of those relationships actually exist. In these senses, theory precedes empirical analysis in economics.

Most of our propositions can be found in well-known explications and demonstrations of each of these three approaches. Haavelmo (1944) stated that even the simplest empirical tasks are impossible without a prior theoretical framework. Theory was to be used to specify the probabilistic model upon which the empirical analysis would be based. Haavelmo did not

explicitly equate "economic theory" with the theory of individual optimization, but he frequently identified theory with statements about the behavior of individuals and firms (pp. 8, 21-22, 28, 51-52); and the bulk of the economic examples he used came from neoclassical theory – demand functions, indifference surfaces, utility functions, etc. (pp. 6, 18, 22, 27, 33).

The Cowles commitment to the concept of theory stated in proposition (1) was more strongly expressed in Koopmans's contributions to the "Measurement without Theory" debate (Koopmans 1947, p. 166; Vining and Koopmans 1949, p. 80), in which he emphasized the essential role of such theory in both the identification of questions to be answered through empirical research and the choice of methods for answering them. The belief that theory should be expressed mathematically, as a system of simultaneous stochastic equations embodying hypothesized causal relationships between economic phenomena, was a defining feature of the Cowles approach. So too was the commitment to the development of elaborate estimation techniques to quantify the "parameters" that represented those causal relationships.[3]

Milton Friedman's essay on the methodology of positive economics also placed theory as we define it – "relative price theory . . . which reached almost its present form in Marshall's *Principles*" – at the center of empirical microeconomic research. There was "enormous room for extending the scope and improving the accuracy" of that theory. Such theoretical development was the "ultimate goal" of positive economics, which was to be achieved using the theory to develop "predictions about phenomena not yet observed" and testing those predictions against "factual evidence." (Friedman, 1953, pp. 3-4, 5, 24, 26).

[3] The Cowles Commission econometrics program is often associated with macroeconomic research, but from the beginning the Cowles pioneers applied their methods to microeconomic problems, such as demand systems and production functions.

Friedman's methodological ideas took concrete form in the dissertations and subsequent papers of a generation of empirical microeconomists trained at Chicago in the 1950s and 1960s. These studies typically began with a mathematical model of individual optimization subject to constraints and/or an equilibrium model of the interaction of optimizing agents. The model's ability to generate new hypotheses often arose from its representation of optimizing behavior in some human activity that had not previously been so represented (e.g., educational attainment, crime, divorce, or suicide), or its redefinition of the constraints faced by agents to include an aspect of the activity that had been assumed away in previous models. It was common to show that the new model predicted the existence of empirical relationships not previously well established, which the author would list as the implied testable hypotheses. A statistical analysis would follow, with descriptions of why it provided credible tests of the hypotheses.[4]

An important distinction between what we are calling the Cowles and Chicago approaches lay in how the theoretical model was brought to the data. The Cowles approach involved a theoretical model consisting of a system of equations that could be estimated directly by linear regression or maximum likelihood techniques. In its pure form this Cowles-inspired style of empirical microeconomic research came to be known colloquially as the "structural" approach. In practice it typically generated two models – a set of equations that embodied the theoretical model, and a derived empirical model ("estimating equations") that were approximations to the theoretical model. The parameters of the estimating equations represented important theoretical relationships, and the theoretical model might impose limits on the possible values of certain parameters, thus creating opportunities for testing theory using probability-based inferential techniques.

[4]Archetypical examples of this "model – listing of hypotheses – description of test procedures" approach to research from Chicago-trained economists include Oi (1962) and Rosen (1968, 1969).

Many of the Chicago-affiliated researchers working in the 1950, 1960s and 1970s practiced a less formalistic style of empirical microeconomic research. A mathematically expressed theoretical model would still drive the analysis, but the goal would not be to estimate directly the equations constituting the model.[5] As with the Cowles-style empirical research, and despite the emphasis placed by Friedman (1953) on the importance of falsification testing, much of this empirical literature in was devoted not to testing model predictions that could potentially be refuted, but instead to measuring the magnitudes of important model concepts that were assumed to exist, such as the rate of return to education or the union/non-union wage gap.

Another common purpose of empirical research in both the Cowles and Chicago styles could be termed reinterpretation, with empirical patterns and correlations in the data pertinent to some class of social activity being "explained" or reinterpreted in terms of the theoretical model. Often there would be no possibility that the empirical analysis would identify a pattern in the data that would falsify the model; rather, different patterns would support different interpretations of the data in terms of the model.[6]

These descriptions of the "structural" and "Chicago" styles of empirical analysis are of course ideal types. What is important is that two distinguishable styles of combining economic theory with regression-based statistical methods were acceptable to the editors who selected papers

[5]What we call the Chicago approach would sometimes be referred to colloquially as the "reduced from" approach. This reduced-form style of combining theory and empirical analysis also characterized the structure-conduct-performance (SCP) school of empirical Industrial Organization that originated at Harvard in the 1950s. In SCP research, models of profit-maximizing firms were used to derive hypotheses about partial correlations between measures of market structure and measures of industry performance, hypotheses that were then tested using regression analysis. An alternative, "structural" approach to empirical IO emerged in the 1980s. See Breshnahan (1989).

[6]For example, Lewis's (1956) reinterpretation of trends in labor force participation in terms of the neoclassical model of labor supply, or Becker's (1965, Ch. VII) interpretation of empirical age-earnings profiles in terms of the human capital model.

for publication in the top economics journals, and they were widely adopted by empirical microeconomists with no explicit connection to either Chicago or the Cowles Commission.

Wassily Leontief, in his 1941 *Structure of American Economy*, introduced a quite different way of combining theory and data. Like the Cowles econometricians and the Chicago microeconomists, Leontief gave theory a privileged place in empirical analysis, that of determining "what factual data are to be secured and how they are to be used within the framework of a particular analytical scheme." He described his input-output model as "an attempt to apply the economic theory of general equilibrium." Leontief might appear to have offered an alternative to neoclassical models, given his assumption of fixed factor proportions in production and fixed budget shares for goods consumed by households, as well as his statement that his model was a "formal rejection of marginal productivity theory.",At least early on, however, he regarded his model as an approximation to a neoclassical economy in which cost-minimizing firms varied factor proportions in response to changing input prices, and consumer demand was responsive to changes in output prices. Approximation was necessary due to the nature of the data available, and the adequacy of the approximation was an empirical question.[7]

It is not Leontief's theoretical model itself that left the more significant mark on the post-1970 consensus, but its empirical implementation. He described his general approach to estimating model parameters, e.g., input output coefficients, 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). In Leontief's approach, assigning values to the model parameters through various means preceded the use of the empirically specified model as a tool for estimating unknown price and quantity relationships or forecasting the effects of

[7]See especially Leontief (1950, p. 201), but also pp. 40, 42, 152, 203-204, 214-216, and Leontief (1952, p. 8).

hypothetical economic changes. In the 1950s and 1960s, this general approach came to be used in conjunction with other types of theoretical models, including "spatial equilibrium" models (e.g., Fox, 1953), and linear programming or activity analysis models (e.g., Hildreth, 1955). By the 1980s the approach was being used with computational general equilibrium models for such tasks as estimating the economy-wide impact of major changes in tax policy (see Ballard and Johnson, this volume).[8]

In our analysis of empirical microeconomic articles we look for easily identifiable markers of the post-1970 consensus view of the role of theory: Does the article estimate a relationship between social or economic variables? Does it include a formal theoretical model? Is that model presented in mathematical form? We also look for evidence of the distinct empirical approaches we have described as coexisting within the consensus: Does the author test the model, in the sense of describing results of empirical procedures that would be inconsistent with the model? Does the author attempt to estimate directly the model's parameters, as called for in the Cowles-inspired structural approach? As noted in the introduction, we believe that the markers consistent with the consensus view of the role of theory will be present in the articles from the 1950s, but much more prevalent in those from the 1970s.

Our hypothesis about the more recent emergence of a new approach to empirical microeconomics implies that many articles from the early 2000s will lack certain characteristics associated with the post-1970 consensus. As the experimentalist paradigm emphasizes the more accurate measurement of causal relationships, we do not expect a decrease in the share of articles devoted to measuring relationships between variables. We do, however, expect a decline in the share of articles that develop explicit theoretical models to identify questions to be answered or to

[8]Johansen's growth model, discussed in ---- (this volume), also employed what we call Leontief's approach.

design the empirical methods used to answer those questions. This would also entail fewer articles employing the "structural" approach to estimation.

III. Measuring the Role of Theory in Empirical Microeconomics

A. Sample and Measurement Criteria

Any attempt to quantify secular changes in the role of economic theory in published empirical microeconomic research must first address certain procedural questions: What kind of theoretical discussion should be counted as informing applied work? Should one count only formal theory or also looser, theory-based discussions? On what set of applied work should we focus a formal evaluation? In what follows we describe the specific choices we have made that implicitly answer these questions.

We create samples of articles in empirical microeconomics from 1951-55, 1973-77 and 2007-08, restricting the samples to articles in the so-called "Top 5" journals, the *American Economic Review* (*AER*), *Econometrica* (*ETRCA*), *Journal of Political Economy* (*JPE*), *Quarterly Journal of Economics* (*QJE)* and *Review of Economic Studies* (*RESTUD*).[9] We include only regular articles, not reviews, Presidential or Nobel Prize addresses, comments or replies. Those from the first two periods must have been at least five pages long to be included in the samples; but to reflect the profession's increased logorrhea those in the last period must have been at least ten pages.

We exclude articles that use only macroeconomic data. We include articles that test microeconomic ideas that are fundamental to macroeconomics using micro data (e.g., examining expectations formation using data on individuals' expectations) and empirical studies in finance

Formatted: Indent: First line: 0.5"

[9]The second period includes as a subset empirical articles from the two years used by Hamermesh (2017), while the third period includes all the empirical micro articles in his 2007-08 sample.

that examination patterns of stock prices over time. The requirement is not that the sample be cross-section or panel data, but rather that the test be essentially based formally or even very informally on microeconomic ideas and that it be examined using data at the micro level. Thus an examination of the aggregate consumption function using aggregate time series would not be included in our sample, but a study of life-cycle patterns of consumption using the PSID is included.

Table 1 presents the percentage distributions by journal of the articles in our samples. The small percentages in the *RESTUD* reflect its historical concentration on theory and methodology, while the declining representation in *ETRCA* reflects its increasing turn away from empirical work. The percentage of articles from the *JPE* declines because that journal has not expanded the number of articles published in each issue or the number of issues published per year.

We developed a coding instrument designed to categorize consistently each of the 512 studies in the sample. Given changes in research styles and the need to have a consistent set of coding instructions, the protocols cannot fit each era perfectly. We believe, however, that despite their generality they do allow us to infer differences in research styles in applied microeconomics over this 57-year period.

Table 2 describes the five variables coded for each of the articles in the samples and reproduces the coding instructions. We pre-tested this scheme by taking one article from each of the three samples and attempting to code each of the five variables. With the exception of one variable on one of the three articles, our separate coding of these three articles matched perfectly. Encouraged by this agreement, we then proceeded independently to code all 512 articles. The degree of agreement was less than in the pre-test. For that reason, all of the analyses will be based

on each author's separate coding; and we will only conclude that a particular difference in research styles exists across the three periods if it exists in both authors' coding.

V2 is the focus of much of the discussion, as it categorizes articles by the extent to which they are based in theory. Much of our empirical analysis focuses on a binary variable indicating that the article has some theoretical basis (V2 = 2, 3 or 4) or has essentially none (V2 = 0 or 1). The descriptions in Table 2 were developed before looking at the articles and reflect characteristics common to the articles from the post-1960 period. We found, like Backhouse (1998, pp. 89-90), that the journal literature from before 1960 does not always fit well into categories developed in light of the research practices that had come to dominate empirical microeconomics by the 1970s. Many articles were essentially descriptive accounts of institutions, regulations, etc., with no implicit or explicit use of theory as we define it, nor any statistical testing. Both authors coded these types of articles, which had essentially disappeared from the top journals by the 1970s, as V2=0.

The category V2 = 1 captures articles that aim to measure something that past economists have identified as important for economic theory and policy, such as the rate of return to education or the extent of wage discrimination against some group, but that do not develop or analyze a theoretical model. Rather, the relationship to be measured is discussed in a way that assumes broad agreement among readers about its definition. The article may include a *statistical* model that illustrates why previous studies have produced biased measures of the relationship, and why the author's empirical approach is less likely to produce a biased measure. These arguments about the presence or absence of bias are not, however, derived from nor based in an economic theoretical model but instead come in the form of plausible assertions or brief references to other studies.

This type of article is distinct from another type which also has the measurement of an economic relationship as a goal, but presents an explicit theoretical model. The model may provide a more precise definition of the relationship to be measured or a new or more complete understanding of the behavior underlying it, explain why previous attempts at measurement were potentially flawed, justify the use of a specific statistical technique, or demonstrate the suitability of the data being used. Such an article is coded $V2 \geq 2$, denoting that theory is present and is used.[10]

The sample for the early 2000s includes a number of articles using experimental data, many from the field of behavioral economics. Because empirical research using experimental data typically shares the characteristics that we have identified as markers of the post-1970 consensus, these articles did not create the need for special categories in any of the variables. In most such articles a mathematical model of individual behavior or market equilibrium is analyzed to produce testable hypotheses or identify parameters or relationships of interest. The methods used to analyze individual choices or the equilibria arising from the model are the same as those found in traditional microeconomic models. Such articles were coded as $V2 = 4$. If a behavioral or experimental economics paper did not present or reference a specific theoretical model, but only mentioned a general concept such as "hyperbolic discounting" or "prospect theory", it was coded as $V2 < 2$

The variable V1 identifies papers that attempted to measure a relationship between variables, as the post-1970 consensus considered identification and measurement of relationships between economic phenomena the key to building an empirically-based economic science. V3 indicates whether the author of the study claimed to be conducting falsification testing, and V4 tries to identify papers adopting the "structural" approach to estimation. V5 was intended to

[10]This distinction between two types of measurement articles is discussed more fully in Panhans and Singleton's case study of the empirical literature on the rate of return to education (this volume).

capture papers that, while not using theory to identify a relationship, did use a theoretical model in the design and justification of the author's estimation procedure.

Both authors/evaluators assigned the same value for V1 in 93% of cases, and gave the same rating to 59% of the articles when applying the five-point scale of V2. On the dichotomized version of V2 (separating articles that made some use of theory from those that did not) there was agreement in 84% of cases. There was less inter-rater agreement on V3--only 58% of those articles that both had coded V2 $\geq$ 2. Agreement regarding the presence of structural estimation (V4) was somewhat higher, 76% on those articles that both authors assigned V2 $\geq$ 2.

B. Estimates

Table 3 presents for each period the percentage of empirical articles attempting to measure one or more relationships between variables, as judged by each author/evaluator. The statistics are consistent with our description of the emergence of the idea that discovery and measurement of relationships between economic phenomena is the ultimate purpose of empirical research in economics. The top panel of Table 3 looks at the percentage of articles by time period assigned V1 = 1. In the 1950s sample, both evaluators judged that a substantial share (though not a majority) of empirical articles dealing with microeconomic topics, while presenting numbers or describing means and perhaps further distributional information for variables considered in isolation, were not concerned with measuring relationships.[11] In the samples from the 1970s and 2000s, however, almost every empirical article included an attempt to measure relationships between variables.

Next we present the distributions of V2 by time period. For both evaluators, the percentage with V2 = 0 falls from the 1950s to the 1970s, then rises again in the recent period. The rise is less striking in Evaluator A's ratings, but the differences across periods are statistically significant for

[11]This agrees with the earlier surveys of the empirical journal literature that found that statistical techniques for measuring relationships were not commonly used in the 1950s (Backhouse 1998).

both sets of ratings. The share of articles with V2 ≥ 2 rises significantly between the early 1950s and the early 1970s, and then is significantly lower again in the 2000s.. The typical article categorized as V2 = 0 in the 2000s sample is consistent with the experimentalist paradigm discussed above. The main focus of the article is on convincing the reader that the relationship of interest has been "credibly identified" by the author's "empirical strategy". The relationship being measured is not, however, rooted in a previously specified economic model.

We refined the analysis of the categorizations of V2 using probit models describing the indicator based on V2 ≥ 2 that included controls for period and journal. All the differences between periods reported in both panels of the table remained statistically significant. The probits also indicated that articles coded as V2 < 2 in the 2000s were more prevalent in the *QJE* and the *JPE* than in *RESTUD* or *ECTRA*, with the *AER* in between.

Considering the percentage of articles coded as V2 = 4, both evaluators saw a greater use of mathematical models in the 1970s than in the 1950s or in the 2000s. The drop from the 1970s sample to the 2000s sample is smaller in Evaluator B's coding, and the t-statistic testing the difference between the two periods is only 1.47. Considering only those articles in which theory was viewed as playing an important role (V2 ≥ 2), the probability that the model would be expressed in mathematical form increased from the 1950s to the 1970s, while the decline in the use of mathematical models from the 1970s to the 2000s was largely due to a decline in the number of articles in which economic theory, however presented, played a substantive role.

The bottom parts of Table 3 present results on V3–V5, identifying the influence of the methodological ideas that coexisted within the post-1970 consensus. V3 = 1 if the evaluator saw in the article, for which V2 ≥ 2, a description of a way that the empirical results could falsify the model. V4 was the "structural estimation" variable, coded V4 = 1 if the empirical procedure was

designed to estimate directly one or more parameters of the theoretical model. There was only a low level of agreement between the evaluators on V3, so there is little we can say with confidence on the issue this variable was designed to reflect. The picture is clearer, however, on V4: Both evaluators see the use of structural approaches to empirical microeconomics increasing from the 1950s to the 1970s, then returning by the 2000s to close to the 1950s level. The decrease from the 1970s sample to the 2000s sample occurs even among papers that otherwise fit the post-1970 consensus. It is possible that this is another consequence of the "credibility revolution," as those who have embraced quasi-experimental approaches to empirical microeconomic research, some of whom served as editors of top economics journals in the early 2000s, are particularly skeptical towards using structural estimation techniques (see, e.g., Angrist and Pischke 2010, pp. 20-22).

The paucity of articles for which V5 = 1 in the 1970s and 2000s samples masks a change between the two periods in the perceived role of theory in empirical research. V5 was intended to identify instances in which theory informed the design of the statistical aspects of the empirical project, for example, to identify exclusion restrictions in a simultaneous-equations model. This is a role for theory distinct from its use in identifying relationships to be measured or developing hypotheses to be tested. In the 1970s formal theory was often used in this way, but almost exclusively in articles in which the authors also used formal or informal theoretical models to define or identify the relationships to be measured, so V5 was not coded for these articles. In the 2000s sample, as noted, there are a number of papers for which V2 < 2, and for all of them both evaluators set V5 = 0, indicating an attitude among adherents to the experimentalist paradigm that economic theory is not an important tool for designing an "empirical strategy."

Evaluator A also tabulated papers that employed Leontief's approach. Eight of the 26 papers that used a mathematical model in the 1950s sample employed this approach. In the 1970s

sample only 10 out of 138 did so; and in the sample from the 2000s only 3 out of 110. Leontief's non-econometric approach to estimating simultaneous equations models has essentially disappeared from microeconomic research published in top journals.

C. Discussion

The results in Table 3 support our two hypotheses regarding the emergence, consolidation and subsequent erosion of a consensus concerning the role of theory in the conduct of empirical microeconomic research. The research approaches that characterized what we have termed the post-1970 consensus, with empirical analysis organized around a formally explicated theoretical model, are still evident in the articles published in top economics journals in early 2000s. They share space, however, with a significant number of articles in which formal economic models play little or no role, almost all of which employ the empirical methods of the experimentalist paradigm. The evidence suggests that research employing the structural approach to empirical microeconomics has been disproportionately affected in this crowding-out process.

To examine how the market for ideas has treated empirical articles classified by their basis in economic theory, we collected data from the Web of Science (WoS) on the citations through December 2014 received by each article in our 2007-08 sample. The estimates show that articles for which V2 was coded as 0 or 1 were at least as influential as those articles coded as $V2 \geq 2$. The results are not altered qualitatively if we use citations in Google Scholar instead of the WoS.

The rhetoric of economists associated with the experimentalist paradigm has centered on establishing higher standards for what counts as good empirical analysis and has not involved questioning the value of traditional microeconomic theory as a tool for empirical economic research. Indeed, some of the leading examples of the quasi-experimental approach involve creative uses of theoretical modeling. But in research, as in other endeavors, increased attention to

one task leads to decreased attention to others; the same is true in graduate training, where more time spent teaching students to identify and exploit the situations that might represent "natural experiments" means less time spent learning to specify and manipulate theoretical models. This may be the main mechanism behind the diminution of formal theoretical modeling in the literature of the experimentalist paradigm.

In the conclusion of his treatise on the probability approach, Haavelmo (1944, p. 114) explicitly recognized the "tremendous amount of work" that would be involved in conducting research along the lines he had laid out. The period when the post-1970 consensus dominated empirical microeconomics was punctuated by critiques from within, citing how routine research practice had not lived up to the methodological standards Haavelmo set. Leamer's (1983) well-known contribution to this critical literature was the jumping off point for Angrist and Pischke (2010, p.12), and they in turn conjured a picture of empirical research in the 1980s and 1990s as a degraded Cowlesian program, in which it was acceptable to "mechanically invoke a simultaneous equations framework, labeling some variables endogenous and others exogenous, without substantially justifying the exclusion restrictions ...." As the credibility revolution proceeds, it may involve a similar gap between the demanding methodological standards set by the leaders and the ordinary research practice of the rank and file.[12]

The results presented thus far suggest at least a partial erosion of the post-1970 consensus on empirical methodology; but because we wished to obtain information on a reasonably long period of citations to recent articles, we restricted the third period to articles published in 2007-08.

[12]Leamer (2010, p. 33) in responding to Angrist and Pischke, raised this possibility when he expressed the concern that despite Angrist and Pischke's own understanding of the need for careful thought when assessing empirical results, "their students and their students' students may come to think that it is enough to wave a clove of garlic and chant "randomization" to solve all our problems, just as an earlier cohort of econometricians have acted as if it were enough to chant 'instrumental variable'."

One wonders, therefore, whether the trend away from theory in empirical work has continued to the present. To assuage wonderment we collected data on all 132 empirical studies published in the same five journals in 2015, of which 63 were published in the *AER*, a concentration explained at least in part by its current annual run of 11 regular issues.[13] For each of these articles the evaluators again coded V2.

The correlation of the authors' coding was 0.81, and the correlation for the collapsed indicator, V2 ≥ 2, was 0.82, both slightly higher than for the 2007-08 sample. The indicator for Evaluator A's coding V2 ≥ 2 averaged 0.62 (s.e. = 0.04), while that for Evaluator B averaged 0.64 (s.e. = 0.04). These are lower still than those in Table 3, but qualitatively quite similar. A fair conclusion is that the change documented between the 1970s and 2007-08 has persisted.

IV. The Treatment of Economic Theory in the Labor Market

The evidence just presented, along with Hamermesh's (2013) finding of a sharp diminution in the amount of purely theoretical research in the top journals, is arguably consistent with Backhouse and Cherrier's (2014) conjecture that there has been a change in the status accorded economic theory by the economics profession. Put differently, as compared to the 1970s, empirical microeconomists are paying less attention to theory in recent years. At the same time, there has been no diminution in the use of mathematical modeling in the purely theoretical microeconomic literature, and arguably the average complexity of the mathematics used has increased. So implicitly, theorists are "talking amongst themselves" more than before. With this change in the focus of the profession, an interesting question is how the market for economists has responded

[13]Because final issues of *Econometrica* and the *Journal of Political Economy* were not available at the time we collected these data, we used the final issues from 2014.

and, in particular, how the determinants of salaries of academic economists who differ by specialty have changed as the nature of the profession has changed.

To examine this question we use several sets of data on the academic-year salaries of economists, data originally assembled for other purposes. Hamermesh *et al* (1982) and Hamermesh (1989) collected data on 100 full professors of economics at six major public universities for the academic years 1979-80 and 1985-86.[14] In addition to salary, the data included information on each scholar's Ph.D. year, the number of citations received in the previous five years (as tabulated in the Social Science Citation Index) and the person's current or prior status as an academic administrator. Hamermesh and Pfann (2012) collected the same information for 525 full professors at 41 public universities, including all six universities from the earlier study, for the 2007-2008 academic year, except that the data covered each person's total lifetime citations in the WoS.

For each of these data sets we designated an indicator variable, Theorist, describing those whose primary (usually only) work was in microeconomic theory. Because the designation as "theorist" is arbitrary, a theorist colleague was consulted to create an alternative designation in the 2007-08 data that was used as a check on the other indicator. In the earlier sample one-eighth were classified as theorists. The fraction in those schools in 2007-08, as classified by the (non-theorist) authors, was more than double, while the theorist's classification generated about the same fraction as in the earlier sample. In percentage terms there were fewer theorists in the entire 2007-08 sample

[14]The institutions are the University of Illinois—Urbana-Champaign, the University of Michigan, University of Maryland, University of Minnesota, University of Wisconsin—Madison, and Michigan State University.

than in the six schools that were present in both samples, perhaps a reflection of the generally higher quality rankings of those six in the entire sample.[15]

Table 4 presents estimates of the determinants of the logarithm of nine-month salaries in the six schools that are included in both samples. The left-hand panel presents estimates for the longitudinal (1979-80 and 1985-86) sample of 100 economists, while the right-hand panel lists estimates for the same schools for 2007-08. For the latter sample we present results using both classifications of the sample members as being theorists. In all cases we present estimates with no controls, with the vector of control variables (econometrician, citations, experience and administrator experience), and with these and institution fixed effects (since theorists may be sorted across schools that differ in average compensation).

In the earlier sample theorists receive a pay premium of roughly ten percent, an estimate that is robust to the inclusion of either the control variables and/or school fixed effects. At least in this sample, at a time when theory appears to have been crucial to the conduct of empirical research, theorists commanded a pay premium. Looking at the results for 2007-08, the conclusion depends on whether we use the authors' broad definition or the narrower classification provided by a theorist. Assuming that the latter is more appropriate (and the large increase in the fraction that we classified as theorists between the two periods suggests the narrower definition may be better), the evidence suggests that there was some diminution of this premium, but that it was still present. While none of the estimates based on the later sample is statistically significant by conventional standards, what we view as the best estimate, shown in the final column, does have a t-statistic exceeding one.

[15]All six are among the Top 30 schools ranked in Hamermesh (2017); only five of the remaining 35 schools in his sample are ranked at least this high.

The sample of six schools is quite narrow and was dictated by the difficulty of obtaining salary data in the early 1980s. Table 5 shows the results of estimating these earnings equations on the much broader sample of 41 schools in 2007-08, a sample that we view as being representative of the upper echelon of public higher education in economics in the United States. Again, the results using the narrower definition of "theorist" are more precise and probably more credible. While the best estimates, controlling for school fixed effects and other measures of the scholars' job experience and professional impact (shown in the final column of the table) are not statistically significant, the point estimate of 6 percent differs little from that shown for 1979-85 in Table 4. It suggests that there still exists a pay premium for theorists, a premium that is larger if one ignores the fact that theoretical work is relatively less-cited today than in the 1970s (see Hamermesh, 2017).

One might wonder why, in the face of a decline in the role of theory informing applied economics, the pay premium for theorists has failed to disappear. One possible explanation is that employers have moved up the demand curve for theorists, but this is inconsistent with the data: We noted above that in the six institutions on which we have information in 1979-85 and 2007-08, the fraction of theorists (broadly classified) has not changed. Beyond this, many explanations are possible, with our favorite being the idea that theory is a merit good, with theorists receiving pay premia simply for being theorists.

V. Conclusions and Implications

Our survey of the content of economic journal articles confirms our hypotheses about the rise and subsequent decline of a consensus about how theory should be used in empirical microeconomic research. The early 1950s were a transitional period. About half the articles from that time reflect ideas about economic theory and norms governing empirical microeconomic

research before WWII; and although many of those articles were intellectually based in neoclassical economic theory, they did not develop new theory nor did they involve explicit estimation of theoretically-based concepts and parameters. However, an almost equal number embodied newer ideas about the meaning of economic theory and its appropriate role in empirical research. These new ideas contributed to and were bolstered by a rapidly growing emphasis on mathematical theory in graduate economic education, and they came to form the basis of the new consensus, which is reflected in the articles from the 1970s. This consensus view held that empirical microeconomic research projects should be organized around an explicitly articulated theoretical model, and should involve the measurement of economic parameters and/or the testing of hypotheses derived from that model. This period lasted 20 to 25 years; if we had taken our sample of articles from any interval between the late 1960s and the early 1990s, we would likely have seen summary statistics for our variables that were similar to those from the 1970s.

The articles from the 2000s show empirical microeconomics again to be in a transitional period with respect to the role of theory. The share of articles presenting new theoretical models or developing new hypotheses has fallen, and the use of mathematics to express economic theory is also less common than in the 1970s. The research approaches of the post-1970 consensus have by no means disappeared, but a significant number of articles from the 2000s reflect the influence of a new "experimentalist paradigm". In these articles, emphasis is placed on developing a "credible" estimate of a causal relationship between some shocking variable and some outcome of interest, with little or no effort given to linking the relationship being estimated to a formally explicated theoretical model. A sizable proportion of empirical microeconomists have switched from concern about explicitly basing their empirical project in economic theory to, as the London School of Economics motto states, *rerum causas cognoscere*. Here, however, the "*causas*" are

superficially causal relationships rather than any underlying behavior that generates the estimated impacts.

We also examine how the changing academic labor market has rewarded theorists compared to applied economists. We conduct such an examination using longitudinal data on 6 large public universities in 1979-85 and 2007-08, and on another 35 schools in 2007-08. The evidence suggests that theorists earned roughly 10 percent premium pay over applied economists with the same academic experience and the same professional recognition, measured by citations in scholarly journals, during the early period. This premium had eroded only slightly if at all by the early 2000s.

In the early 1970s the senior author overheard one of his much-senior applied-economist colleagues, who had received his graduate training in the late 1920s, lamenting to another senior colleague, "When is the mathematical stuff in economics going to end?" Our evidence suggests that the importance of the "mathematical stuff" has diminished in today's applied economic research. It is much less important today than it was 40 years ago to have one's applied research grounded in economic theory and more important that one can demonstrate an explicit causal relation between two measures that may or may not reflect economic behavior. Modeling the behavior itself is no longer an essential part of many empirical articles in top journals.

Applied economic research is not a monolith; nonetheless, in broad outlines it can be characterized by changing styles and emphases over the past 70 years. The aspects of applied economic research that we examined – ideas, rhetoric, and practices related to the meaning of economic theory and the role of theory in empirical research – enjoyed a relatively long period of stability, but are now in flux. The experimentalist paradigm may come to be the new consensus approach to empirical microeconomic research, or it may instead fade in influence, as did

Leontief's "direct observation" approach. Either way, it is unlikely that empirical economics will revert to the paradigm of the post-1970 consensus—intellectual history is not cyclical. Rather, the desires of today's more senior scholars who, like the very senior applied economists of the early 1970s, deplore much current research, will be realized—but undoubtedly not in the ways they might imagine or even desire.

REFERENCES

Angrist, Joshua and Jörn-Steffen Pischke. "The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics." *Journal of Economic Perspectives*, 24(2010): 3-30.

Backhouse, Roger. The Transformation of Economics 1920-1960, Viewed through a Survey of Journal Articles. Pp. 85-107 in *From Interwar Pluralism to Postwar Neoclassicism* (Annual supplement to vol. 30, *History of Political Economy*), ed. Mary S. Morgan and Malcolm Rutherford. Durham: Duke University Press, 1998.

Backhouse, Roger and Béatrice Cherrier, "Becoming Applied: The Transformation of Economics after 1970," Unpublished Paper, Birmingham University, 2014.

Becker, Gary. *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education.* New York: NBER, 1964.

Bresnahan, Timothy. Empirical Studies of Industries with Market Power. Pp. 1011-57 in *Handbook of Industrial Organization*, vol. 2, ed. Richard Schmalensee and Robert Willig. Amsterdam: North-Holland, 1989.

Deaton, Angus. Instruments of Development: Randomization in the Tropics and the Search for the Elusive Keys to Economic Development. January, 2009 (accessed 11/13.2015, at https://www.princeton.edu/~deaton/downloads/Instruments_of_Development.pdf)

Friedman, Milton. The Methodology of Positive Economics. In *Essays in Positive Economics*. Chicago: University of Chicago Press, 1953, 3-43.

Fox, Karl A. A Spatial Equilibrium Model of the Livestock-Feed Economy in the United States. *Econometrica*, 21 (Oct. 1953): 547-566

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

Hamermesh, Daniel. Why Do Fixed-Effects Models Perform So Poorly? The Case of Academic Salaries. *Southern Economic Journal*, 56 (July 1989): 39-45.

Hamermesh, Daniel. Six Decades of Top Economic Publishing: Who and How. *Journal of Economic Literature*, 51 (March 2013): 162-72.

Hamermesh, Daniel. Citations in Economics: Measurement, Uses and Impacts. *Journal of Economic Literature*, 55, forthcoming 2017.

Hamermesh, Daniel, George Johnson and Burton Weisbrod. "Scholarship, Citations and Salaries: Economic Rewards in Economics" *Southern Economic Journal*, 49 (Oct. 1982): 472-81.

Formatted: Font: Italic

Hamermesh, Daniel and Gerard Pfann. Reputation and Earnings: The Roles of Quality and Quantity in Academe. *Economic Inquiry*, 50 (2012): 1-16.

Hildreth, Clifford. Economic Implications of Some Cotton Fertilizer Experiments. *Econometrica*, 23 (Jan. 1955): 88-98.

Keane, Michael. A Structural Perspective on the Experimentalist School. *Journal of Economic Perspectives*, 24 (Spring 2010): 47-58.

Koopmans, Tjalling. Measurement without Theory. *Review of Economics and Statistics*, 29 (Aug. 1947): 161-72

Leamer, Edward. Let's Take the Con out of Econometrics. *American Economic Review*, 73 (Mar., 1983): 31-43.

Leamer, Edward. Tantalus on the Road to Asymptopia. *Journal of Economic Perspectives*, 24 (Spring 2010): 31-46

Leontief, Wassily. *The Structure of American Economy, 1919-1939;An Empirical Application of Equilibrium Analysis*. 2nd Edition. New York: Oxford University Press, 1950.

Leontief, Wassily. Some Basic Problems of Structural Analysis. *Review of Economics and Statistics*, 34 (Feb., 1952):. 1-9.

Lewis, H. Gregg. Hours of Work and Hours of Leisure. In *Proceedings* of the *Ninth Annual Meeting, Industrial Relations Research Association* (1956): 192-206.

Oi, Walter. Labor as a Quasi-Fixed Factor. *Journal of Political Economy*, 70 (December 1962): 538-555

Panhans, Matthew and John Singleton, "The Credibility Revolution: or, How Empirical Economics Jettisoned Most of Mas-Collel and Econometrics," Unpublished Paper, Duke University, November 2014.

Rosen, Sherwin. Short-Run Employment Variation on Class-I Railroads in the U.S., 1947-1963. *Econometrica*, 36 (Jul. - Oct. 1968): 511-529.

Rosen, Sherwin. On the Interindustry Wage and Hours Structure. *Journal of Political Economy*, 77 (Mar. - Apr. 1969): 249-273.

Vining, Rutledge and Tjalling Koopmans. Koopmans on the Choice of Variables to be Studied and the Methods of Measurement (with Reply). *Review of Economics and Statistics*, 31 (May 1949): 77-91.

Table 1. Percent Distributions of Samples of Applied Economics Articles, 1950s, 1970s, 2000s

	Journal					
Period	*AER*	*ETRCA*	*JPE*	*QJE*	*RESTUD*	N =
1951-55	24	19	28	21	8	105
1973-77	32	14	37	14	3	195
2007-08	37	9	14	29	11	212

Table 2. Evaluation Instrument for Empirical Microeconomic Articles

Variable	Description
1	1 if there is something bivariate in it, estimated via regression, purposeful comparisons of means across subgroups, or time series graph, if learning the relationship of a variable with time is a stated purpose of the article. Not univariate graphs, or tables of means and standard deviations without meaningful subgroup comparisons; 0 otherwise.
2	0 Pure policy evaluation—no link to theory. 1 No theory but based on theory. This would includes ROR on education and gender wage examples. 2 Cites one or more theoretical models created by other researchers. 3 A logical elaboration of a theoretical model, but no math. 4 Presents a mathematical model (at least one equation) in which an estimable relationship plays a role.
3	1 if the empirical work is presented as a test of the model, in that the authors describe empirical results that would be inconsistent with the model; 0 if not. Skip if V2<2.
4	1 if an explicit parameter in a formal model is estimated; 0 if not. Skip if V2<2.
5	1 if an explicit discussion of some theoretical basis for the particular estimation procedure chosen is given; 0 if not.. Skip if V2>1.

Table 3. Means, their Standard Errors, and Inter-evaluator Correlations of V1-V5, by Time Period

	Time Period:					
	1951-55		**1973-77**		**2007-08**	
Evaluator:	**A**	**B**	**A**	**B**	**A**	**B**
Variable:						
V1 = 1	0.58	0.70	0.92	1	1	1
	(0.05)	(0.05)	(0.02)	(0)	(0)	(0)
Correlation		0.70		1		1
V2 (Percent Distributions)						
0	39	40	7	15	18	40
1	18	23	8	8	15	2
2	10	14	6	24	10	15
3	9	7	8	4	5	1
4	24	16	71	49	52	42
Inter-rater Correlation		0.67		0.49		0.79
V2 ≥2:	0.43	0.37	0.85	0.77	0.67	0.58
	(0.05)	(0.05)	(0.03)	(0.03)	(0.03)	(0.03)
Correlation		0.57		0.46		0.74
V3 = 1	0.65	0.59	0.46	0.84	0.68	0.84
if V2≥2	(0.07)	(0.08)	(0.04)	(0.03)	(0.04)	(0.03)
Correlation		0.41		0.08		0.10
V4 = 1	0.29	0.33	0.41	0.47	0.29	0.24
if V2≥2	(0.07)	(0.08)	(0.04)	(0.04)	(0.04)	(0.04)
Correlation		0.85		0.46		0.41
V5 = 1	0	0	0	0.07	0	0.15
if V2<2	(0)	(0)	(0)	(0.05)	(0)	(0.04)
N=	105		195		212	

Table 4. Determinants of (Logarithm) of Academic-Year Salary, Six Public Universities, 1979-80, 1985-86 and 2007-08*

	1979-80, 1985-86 (N=100)				2007-08 (N=107)			
Ind. Var.								
Theorist (broad definition)	0.097 (0.068)	0.082 (0.070)	0.117 (0.058)	0.028 (0.060)	-0.011 (0.051)	-0.002 (0.050)		
Theorist (narrow definition)							0.029 (0.070)	0.086 (0.069)
Control variables**	---------	x	x		x	x	x	x
School fixed effects (6)	----------	--------	x	-------	--------	x	-----	x
R^2	0.642	0.761	0.820	0.037	0.353	0.434	0.354	0.445

*Standard errors are in parentheses below the parameter estimates here and in Table 5. In the 1979-80, 1985-86 sample they are clustered on individuals. Each equation here and in Table 5 also contains an indicator for theoretical econometricians, and the equations in the first panel here include an indicator for year.

**Quadratics in post-Ph.D. experience and five years of citations, an indicator of prior/current administrator status.

Table 5. Determinants of (Logarithm) of Academic-Year Salary, 525 Faculty at 41 Public Universities, 2007-08

Ind. Var.	Theorist (broad definition)			Theorist (narrow definition)		
Theorist	0.105	0.030	0.013	0.198	0.093	0.060
	(0.032)	(0.028)	(0.025)	(0.045)	(0.041)	(0.036)
Control variables*	-------	-------	x	-------	------	x
School fixed effects	-------	x	x	-------	x	x
R^2	0.029	0.363	0.513	0.043	0.368	0.515

*Quadratics in post-Ph.D. experience and lifetime citations, an indicator of prior/current administrator status.