Southern Economic Journal 2005, 71(3), 556–565

Theory versus Application: Does Complexity Crowd Out Evidence?

Philip R. P. Coelho* and James E. McClure†

Donald F. Gordon hypothesized that mathematical complexity in economics is inversely related to operationalism. Here we (i) operationalize Gordon's hypothesis, (ii) test for the significance of trends in complexity for the *American Economic Review*, *Economic Journal*, *Journal of Political Economy*, and *Quarterly Journal of Economics*, and (iii) test Gordon's hypothesis by conducting analyses of the contents of articles from the *American Economic Review* (*AER*), as well as the contents of articles citing the *AER* articles. The results do not refute the hypothesis that complexity crowds out operationalism in economics. Additionally, the presence of significant, positive trends toward complexity suggests that the magnitude of the crowding out is on the rise in these journals.

JEL Classification: A11

1. Introduction

Winnowing theories by appeals to evidence is a practice that dates to the beginnings of modern science. Here we test the hypothesis of Donald F. Gordon (1955) that complex mathematical statements are less operational than other economic statements. Operationalism means that non–self-referential evidence has a dominant role in the assessment of theories. Mathematical "proofs" of lemmas and theorems are self-referential and are generally nonoperational;[1] the "proven" theorems may or may not be operational.

Gordon, echoing concerns raised in 1920 by Alfred Marshall (1964)[2] about the use of mathematics in economics, argued that

> ... the essential point is the difference between theories using a large number of functions and those using one or two, since formal and mathematical reasoning is normally required when the number of relationships simultaneously being considered becomes large. As we have seen, even though each may be quite plausible, a combination of very many will rarely be so. Consequently, it happens that the cases in which formal and mathematical reasoning is most likely to be required are precisely the cases in which, for other reasons, the validity of any conclusions is likely to be

* Department of Economics, Ball State University, Muncie, IN 47304, USA; E-mail 00prcoelho@bsu.edu; corresponding author.

† Department of Economics, Ball State University, Muncie, IN 47304, USA; E-mail jmcclure@bsu.edu.

For comments, suggestions, and assistance we express our appreciation to the following people: Moheb Ghali, Tung Lui, Frank Machovec, Eric Munshower, Robert Ohsfeldt, Gary Santoni, Lee Spector, John Umbeck, the editor, and two anonymous referees of this journal. All remaining errors are our responsibility.

Received March 2004; accepted July 2004.

[1] All that mathematical "proofs" show is that the symbolic language is internally consistent.

[2] Samuelson (1952, p. 57) stated that Marshall's disdain for "long chains of logical reasoning" was because "Marshall treated such chains as if their truth content was subject to radioactive decay and leakage—at the end of n propositions only half truth was left, at the end of a chain of $2n$ propositions, only half of half the truth remained, and so forth in a geometric multiplier series converging to zero truth."

> conjectural. It is frustrating but nevertheless true that, where mathematics is most likely to be useful, the theory is least likely to be valid, while, where the theory is most likely to be true, complex deduction is generally not needed (p. 160).

Gordon argued that the realms of mathematics and real-world economic behavior are not identical. Operational propositions are less likely to arise from protracted mathematical formalism. He used an example of a theory relating three variables x, y, and z to illustrate:

> Again, the relationship between x and y may be stable long enough for a shift along that function but not stable long enough for a shift along that function plus a subsequent shift along another [z] (p. 155).

Expressed more formally, let the relationship between x and y be expressed as $y = f(x)$, and that between y and z be $z = g(y)$. Substituting $f(x)$ into the second expression, a composite function, $z = g[f(x)]$, is obtained. Differentiating the composite yields

$$dz/dx = (dz/dy)\times(dy/dx). \tag{1}$$

Equation 1 expresses the impact of a change in x upon z as an *indirect effect*; a change in x leads to a change in y, and the change in y *then* leads to a change in z. Mathematical conventions assume that the indirectness of the effect of x upon z is irrelevant. It is irrelevant because units of measure such as historic time do not exist in pure mathematics. But if a mathematical technique is used to represent a real-world situation in which it takes time for a change in *one variable to affect another*, then the functional relationship may be devoid of practical application. Gordon emphasized that economic phenomena are time dependent; the more functions that were linked in a theory, the more likely it is that the passage of time will materially affect the relationships in ways that are inherently unpredictable. Gordon saw the timelessness implicit in mathematics as an impediment to operationalizing complex relationships between and among variables in economic models.[3]

Our analysis of the Gordon hypothesis extends the literature that has brought the content of published journal articles and citation data to bear on issues in the history of economic thought. In a classic article, George J. Stigler (1969, p. 229–30) concluded

> Economics ... has a useful past, a past that is useful in dealing with the future. Many useful commodities and services are not produced in society because they are worth less than they cost: it remains the unfulfilled task of the historians of economics to show that their subject is worth its cost.

Since Stigler, the literature has provided practical reasons for studying the history of economics. In an analysis of the citations of "great" economists, Gary Anderson, David Levy, and Robert Tollison (1989, p. 182) showed that although "a considerable number" of the listed economists had little connection to the "living" literature, "a fair number of pre-twentieth century economists have impressive citation counts ... What Ricardo, Marx, and Smith, et al. may not have been able to solve may be what is most important about their work for contemporary economists." In another article, David Laband and Robert Tollison (2000) quantified aspects of intellectual collaboration in economics; one example was a positive relationship between the probability of coauthorship and the frequency of "equations, tables, figures, and appendices" (p. 641). Finally, Laband, Tollison, and Karahan (2002) conducted a content analyses for *The American Economic Review* publications that produced insights into editorial quality control, the decline of commentary, and rent seeking by authors.

---

[3] Without reference to Gordon, Leontief (1971) reasoned analogously: "Uncritical enthusiasm for mathematical formulation tends often to conceal the ephemeral substantive content of the argument behind the formidable front of algebraic signs" (pp. 1–2).

2. Evidence on Gordon's Hypothesis

The Gordon hypothesis is that complex mathematical statements are less likely to be operational relative to other economic statements: We offer evidence on this proposition.[4] We use data from the *JSTOR* (*Journal Storage*) archive for 1963 through 1996.[5] We collected data on four general interest economic journals in the archive: *The American Economic Review* (*AER*), *The Economic Journal* (*EJ*), *The Journal of Political Economy* (*JPE*), and *The Quarterly Journal of Economics* (*QJE*), as well as *The Journal of Economic History* (*JEH*). We included the *JEH* because it is empirically oriented; we wanted to observe how a journal that emphasizes real-world applications compared to the general interest journals. The *Social Sciences Citation Index* provided data for our citation analysis. We used EViews3 and EViews5 software.

Trends in Theoretical Complexity

We examine the trends of mathematical complexity in the literature to assess the importance of the Gordon hypothesis. If trends in the publication of mathematically complex papers are constant or declining, then the Gordon hypothesis is relatively less important than if the trends turned out to be increasing. Toward an assessment of trends in complexity, we conducted an annual full-text search in our sample journals for either of the terms "multiple equilibria," or "lemma." Articles found to contain either (or both) of these terms were viewed as being more mathematically complex than those that did not contain them.

The terms "lemma" and "multiple equilibria" were selected as proxies because they are indicative of mathematical complexity, and because their initial usages in the journals in *JSTOR* were in 1910 and 1934, respectively. These terms were used infrequently prior to 1963, but they were in usage.[6] We did not "clean" the data; articles that contained "lemma" and/or "multiple equilibria" that did not contain complex models were not excluded.[7]

The procedures for organizing data were as follows: let $LM_{i,t}$ denote the *JSTOR* count of articles containing the terms "lemma" and/or "multiple equilibria" for journal i in year t (for example, $LM_{AER,1963}$ would represent the *JSTOR* count of articles in the 1963 *American Economic Review* containing "lemma" and/or "multiple equilibria").[8] To correct for changes in the numbers of articles published, we divided $LM_{i,t}$ by the total number of articles published by journal i in year t, denoted as $TOTAL_{i,t}$. These totals were found from the *JSTOR* count of the number of articles containing four

---

[4] We are *not* testing whether economics is becoming more or less empirical; a preliminary investigation suggests that the journal literature is becoming more empirical. What we are doing is examining the empirical content of mathematically complex publications in economics. If the literature is becoming more empirical *and* if mathematically complex articles are defying this trend, then evidence supporting the Gordon hypothesis is enhanced.

[5] JSTOR provides electronic copies of journals; it is text searchable on a variety of levels.

[6] From 1910 to the beginning of our sample period in 1963 the terms "lemma" and "multiple equilibria" appeared 33 times in the four general interest journals (including *The Papers and Proceedings of the AER*) in *JSTOR*. The same search criteria of JSTOR for the years 1963 to 1996 yielded a count of 853. The year 1996 was the terminal year because the *EJ* and *JEH* were covered only through 1996 in the version of *JSTOR* available to us.

[7] This is because (i) we want our procedures to be easily replicated, and (ii) biases that enter into our sorting process invariably produce more statistical "noise"; the introduction of this "noise" makes the attainment of statistical significance more difficult. Any bias introduced by not "cleaning" the data is a bias against accepting Gordon's hypothesis.

[8] There were no double countings; articles containing both terms were counted once. Articles in the *AER Papers and Proceedings* were excluded because the selection criteria for these differ from the criteria for the *AER* (and also the other journals).

Table 1. Number of Articles (LM) and Percentage of Articles (PCTLM) Containing the Terms "Lemma" and/or "Multiple Equilibria" by Journal and by Year (Percentages Were Rounded to the Nearest Percent)

	AER		*EJ*		*JPE*		*QJE*		*JEH*	
Year	LM	PCTLM	LM	PCTLM	LM	PCTLM	LM	PCTLM	LM	PCTLM
1963	0	0	0	0	0	0	0	0	0	0
1964	1	1	0	0	0	0	0	0	1	0
1965	1	2	0	0	0	0	2	4	0	0
1966	2	3	0	0	1	2	6	11	0	0
1967	2	2	1	2	2	2	1	2	0	0
1968	1	1	0	0	0	0	2	5	0	0
1969	2	2	2	5	2	3	5	10	0	0
1970	4	3	0	0	0	0	2	3	0	0
1971	4	3	0	0	2	2	3	6	0	0
1972	2	1	0	0	1	1	2	3	0	0
1973	4	3	0	0	5	4	5	10	1	0
1974	3	2	0	0	2	2	1	2	1	2
1975	4	4	0	0	7	9	1	1	0	0
1976	6	5	3	5	3	3	5	9	0	0
1977	3	3	2	5	3	4	5	10	0	0
1978	5	5	1	3	3	4	2	4	2	4
1979	10	8	4	7	7	8	4	8	0	0
1980	7	6	1	2	1	1	9	9	1	0
1981	9	8	1	1	2	3	5	11	1	2
1982	6	5	1	2	6	8	4	9	1	0
1983	13	10	3	5	9	16	10	18	1	0
1984	5	4	4	6	10	17	6	12	0	0
1985	14	11	7	9	7	10	11	16	1	0
1986	6	5	7	10	11	16	8	16	1	2
1987	12	12	6	9	12	18	14	29	2	2
1988	13	13	7	12	11	17	9	18	0	0
1989	11	9	9	14	12	17	12	27	1	0
1990	19	18	9	10	11	17	11	20	0	0
1991	24	21	10	10	9	16	10	17	2	0
1992	18	18	10	11	13	25	10	18	2	2
1993	14	15	7	7	5	9	15	32	0	0
1994	16	17	10	11	8	16	6	14	0	0
1995	17	19	10	11	9	18	8	20	2	0
1996	24	30	8	8	12	26	11	27	1	0

commonly used words that *JSTOR* would search: "because," "which," "first," and/or "then."[9] The percentage of articles in each journal i and for each year t containing "lemma" and/or "multiple equilibria" is denoted $\text{PCTLM}_{i,t}$.[10]

Table 1 presents the data for LM and PCTLM for each journal from 1963 to 1996; it suggests that (i) our measure of complexity is increasing in the general interest journals (*AER*, *EJ*, *JPE*, and *QJE*) and (ii) there is no tendency toward increasing usage of these terms in the *JEH*. To formally test for the presence and significance of trends we conducted augmented Dickey-Fuller unit root tests.

[9] *JSTOR* will not count terms such as "the" or "and." Other common terms were tried, but none were as inclusive; our goal was to get as close an approximation to the total number of publications as possible given our resource constraints.

[10] The percentages were calculated by the following: $\text{PCTLM}_{i,t} = (\text{LM}_{i,t}/\text{TOTAL}_{i,t}) \times 100$.

Table 2. Augmented Dickey-Fuller Unit Root Tests for the PCTLM Time Series Data Found in Table 1

	Level		First Difference	
Journal	Intercept	Intercept and Trend	No Constant	With Constant
$AER_{\text{Full Sample}}$	0.77	−2.82	−7.72*	−8.19*
$AER_{1963-1981}$	0.92	−3.37**	−4.49*	−5.34*
$AER_{1982-1996}$	0.02	−3.56**	−5.18*	−5.55*
$EJ_{\text{Full Sample}}$	−1.79	−3.53*	−8.25*	−8.26*
$JPE_{\text{Full Sample}}$	−0.15	−4.72*	−8.69*	−5.82*
$QJE_{\text{Full Sample}}$	−0.31	−5.86*	−2.55*	−2.90**

The unit root tests were not conducted for the *JEH* because there were insufficient nonzero observations to make the results meaningful.

* Significant at the 5% level.

** Significant at the 10% level.

Table 2 shows the unit root test results for the levels and first differences of the PCTLM time series for the *AER, EJ, JPE,* and *QJE.*[11] The statistics in the table's second column indicate that the PCTLM series for $AER_{1963-1981}$, $AER_{1982-1996}$, $EJ_{\text{Full Sample}}$, $JPE_{\text{Full Sample}}$, and $QJE_{\text{Full Sample}}$ possess stationary and significant time trends. This implies that for each series the existence of a unit root must be rejected. Although the statistics found in the row labeled $AER_{\text{Full Sample}}$ may appear to suggest the existence a unit root, the results just mentioned for the PCTLM series for the $AER_{1963-1981}$ and $AER_{1982-1996}$ indicate that there is actually a structural break in the PCTLM series for the $AER_{\text{Full Sample}}$ that occurs in 1981.[12]

To measure the magnitudes of the trends in the PCTLM time series, for each journal i we independently estimated the linear trend equation:

$$\text{PCTLM}_{i,t} = K_i + \beta_i \text{ TREND}_i. \tag{2}$$

Here, TREND_i and K_i represent the time trend and constant for journal i. Table 3 summarizes the results of estimating Equation 2 for each journal. The coefficient estimates for TREND_i are all positive, and the p values indicate they are significant at the 1% level. The estimate for PCTLM in the $AER_{\text{Full Sample}}$ is subject to specification error. The significance of the trend for PCTLM for $AER_{\text{Full Sample}}$ was *not* established by the unit root test in Table 2. The estimates for PCTLM for the *JEH* are also unreliable as indicated by the adjusted R^2 statistic shown in Table 3.

The TREND estimates that are both reliable and significant in Table 3 are those for the PCTLM series for $AER_{1963-1981}$, $AER_{1982-1996}$, and the full samples for the *QJE, JPE,* and the *EJ*. The estimated coefficients for the TREND variable range from 0.32 for the $AER_{1963-1981}$ to 1.34 for the $AER_{1982-1996}$. This means that, on average, every 10 years the percentage of articles containing "lemma" and/or "multiple equilibrium" rose by 3.2% for the $AER_{1963-1981}$ series and 13.4% for the $AER_{1982-1996}$ series.[13]

[11] There were insufficient data for meaningful testing for unit roots to be conducted on the PCTLM time series for the *JEH*.

[12] Other evidence confirming the significance of this structural break is presented later in Endnote 14. It is worth noting that 1981 is the year when the editorship of the *AER* changed from George H. Borts (who had been editor since 1969) to Robert W. Clower (who was replaced in 1985 by Orley Ashenfelter who continued until 2001).

[13] The estimates and standard errors for the PCTLM series for $AER_{1963-1981}$ and $AER_{1982-1996}$ indicate the independence of the trend estimates. The 95% confidence intervals are nonoverlapping (the intervals were established by adding and subtracting two times the respective standard errors to the respective coefficient estimates). The interval for the $AER_{1963-1981}$ series is 0.32 ± 0.12 and the interval for the $AER_{1982-1996}$ series is 1.34 ± 0.46.

Table 3. Least Squares Estimates of $PCTLM_{i,t} = K_i + \beta_i\ TREND_i$

	Intercept (K_i)			Trend$_i$			
Journal	Coefficient Estimate	Standard Error	Probability	Coefficient Estimate	Standard Error	Probability	Adjusted R^2
$AER_{\text{Full Sample}}$	−2.50	1.22	0.05	0.63	0.06	0.00*	0.74
$AER_{1963-1981}$	0.37	0.59	0.54	0.32	0.06	0.00*	0.63
$AER_{1982-1996}$	−20.99	6.81	0.01	1.34	0.23	0.00*	0.69
$EJ_{\text{Full Sample}}$	−1.56	0.81	0.06	0.89	0.04	0.00*	0.72
$JPE_{\text{Full Sample}}$	−2.94	1.28	0.03	0.70	0.07	0.00*	0.77
$QJE_{\text{Full Sample}}$	−0.08	1.62	0.96	0.72	0.08	0.00*	0.69
$JEH_{\text{Full Sample}}$	0.17	0.30	0.58	0.01	0.02	0.00*	−0.01

* Significant at the 1% level.

Tests of the Gordon Hypothesis

We test the Gordon hypothesis by (i) comparing the contents of more complex articles to the contents of a random sample of articles and (ii) comparing the contents of articles that cite more complex articles to the contents of a random sample of articles.[14]

An Empirical Analysis of Articles' Contents

We compared the contents of complex articles with the contents of less mathematically complex articles. We conducted a content analysis of a subsample of the 1963 to 1996 period, and limited (due to the costs of scrutinizing each article's contents) our analysis to the *AER*. The operational content of the articles using the terms "lemma" and "multiple equilibria" in the *AER* for the years 1975, 1980, 1985, 1990, and 1995 was compared to the operational content of a random sample of the *AER* articles from the same years not containing the terms.

The pages of each article were inspected. Articles containing casual empiricism and/or references to "stylized facts" were counted as nonoperational articles. Similarly, articles that presented self-referential simulations were designated nonoperational. But articles containing data from surveys and/or experiments were counted as operational.

The construction of the random sample followed standard statistical procedures. For each of the years a list of all *AER* articles was created, we removed from this list any citations that were on the list of articles containing "lemma" and/or "multiple equilibria." We excluded any citations to *The Papers and Proceedings of the AER*. Finally, we excluded from both the random sample and from the population all citations that had the terms "comment," "reply," and/or "rejoinder" in their titles. These procedures allowed a comparison between the operational content of original articles containing the terms "lemma" and/or "multiple equilibrium" to original articles not containing the terms.

[14] Repeating two previous points (i) the trend estimates can be used as a gauge of the importance of Gordon's hypothesis, but they are not a test of it, and (ii) again, our tests of the Gordon hypothesis are *not* trying to assess the extent to which empirical analysis is occurring in economics. Again, Gordon's hypothesis says nothing about the trends in economics toward, or away from, empiricism; his hypothesis only states that increased mathematical complexity in economic research reduces the probability of it being assessed empirically.

Table 4. Ratios of Numbers of *AER* Articles Containing Data to Total Numbers of *AER* Articles, by Year and by Source (*AER* Population Containing the Terms "Lemma" and/or "Multiple Equilibria" versus *AER* Random Sample)

Year	Ratios of Numbers of *AER* Articles Containing Data to Total Number of *AER* Articles[a]	Ratio of Number of *AER* Articles Containing Data to Total Number of *AER* Articles[b]
1975	1/4 (25)	1/10 (10)
1980	1/7 (14)	3/10 (30)
1985	2/13 (15)	4/10 (40)
1990	3/19 (16)	4/10 (40)
1995	3/15 (20)	7/10 (70)

[a] Source: *AER* articles containing the terms "Lemma" and/or "Multiple Equilibria" (%).
[b] Source: Random sample of *AER* articles (%).

In the five sample years there were a total of 58 *AER* articles containing either "lemma" and/or "multiple equilibria" (excluding comment articles and articles in *The Papers and Proceedings*).[15] Of the 58 articles, 10 contained nonreferential analysis of statistical data (in other words, approximately 18% of the total articles included analysis of data). The distribution of these articles over time and the presence of data are presented in Table 4.

The random sample had a total of 50 articles from JSTOR. To select 10 articles for each sample year, we employed a table of random digits and chose 10 articles from the *AER* for each sample year whose edited JSTOR rank corresponded to the random digits.[16] The distribution of the articles in the random sample and its characteristics are also in Table 4. In the random sample of 50 articles, 38% had data.

To test the Gordon hypothesis, we estimated Equation 3 by the binary probit method to assess the impact of the appearance of the selected terms (lemma and/or multiple equilibria) on the probability that the article contained data:

$$\mathrm{DAT} = c + \alpha_1 \mathrm{LM} + \alpha_2 \mathrm{YR80} + \alpha_3 \mathrm{YR85} + \alpha_4 \mathrm{YR90} + \alpha_5 \mathrm{YR95}, \tag{3}$$

where (i) DAT equals one for articles with data analysis, and zero otherwise; (ii) LM equals one for articles containing the term "lemma" and/or "multiple equilibria," and zero otherwise; (iii) YR80, YR85, YR90, and YR95 are dummies for 1980, 1985, 1990, and 1995; and (iv) c is a constant.

Table 5 displays the results. The coefficient on the LM variable is negative and is significant at the 1% level. This means that the presence of the terms "lemma" and/or "multiple equilibria" in an article has a negative impact on the probability that the article has any empirical content. These results are consistent with the hypothesis that theoretical complexity reduces operationalism.

A Content Analysis of Citations

A question remains: What is the relative contribution made by complex mathematical models to future operational economic analyses? We addressed this by comparing the operational content of

[15] The original sample had 61 entries containing the terms; two of these had "comment," "reply," or "rejoinder" in their titles and were eliminated. Also removed was a presidential address (Amartya Sen 1995). Consequently, the data in the years 1975, 1980, 1985, 1990, and 1995 in the comparison sample we obtained for the *AER* are not strictly comparable to the data that were used to construct Tables 1, 2, and 3. Because we were interested in the overall *trends* in Tables 1, 2, and 3, we included all citations.

[16] The table of random digits was taken from Morris H. DeGroot (1975).

Table 5. Impact of Mathematical Complexity on the Probability of an Article Having Operational Content

Variable	Coefficient	Standard Error	*z* Statistic	Probability
Constant (c)	−0.845	0.415	−2.035	0.42*
LM	−0.798	0.281	−2.838	0.005**
YR80	0.403	0.535	0.753	0.451
YR85	0.606	0.507	1.197	0.231
YR90	0.619	0.494	1.250	0.211
YR95	1.117	0.501	2.228	0.026*

Dependent variable, DAT; method, binary probit; 107 in the sample (adjusted for endpoints); 29 DAT = 0 observations; 78 DAT = 1 observations; mean DAT = 0.271; SD DAT = 0.447; SE of regression = 0.425; Akaike information criterion = 1.167; sum squared residuals = 18.169; Schwarz criterion = 1.316; log likelihood = −56.383; Hannan-Quinn criterion = 1.227; restricted log likelihood = −62.518; average log likelihood = −0.527; LR statistic (5 df) = 12.270; McFadden R^2 = 0.098; probability (LR statistic) = 0.031.

* Significant at the 5% level.

** Significant at the 1% level.

articles that cite articles containing complex mathematics to the operational content of articles that cite articles that are less complex. To this end, we undertook a content analysis of the citations of the articles from the two data sets from the *AER* on which we had performed content analysis. We examined the contents of all articles in *JSTOR* that cited articles in the two data sets. The analysis was for the five years following publication. For 1975 we searched the *Social Science Citation Index* for the years 1976 through 1980; for the articles from 1980, we looked at the citing articles between 1981 and 1985; and so on for the articles in 1985, 1990, and 1995. Each base article had its own list of citing publications from the *JSTOR* archives for the five-year period following publication.[17]

Table 6 lists the ratios of citations with operational content (data analysis) to the total numbers of *JSTOR* citations found in the *Social Science Citation Index* for 1975–1980, 1980–1985, 1985–1990, 1990–1995, and 1995–2000.[18] In all intervals, except that from 1980 to 1985, the ratios of citations tracing to the random sample exceeded those tracing to the *AER* population of articles containing "lemma" and/or "multiple equilibria."

To test Gordon's hypothesis, we used the binary probit method to assess the impact of the appearance of the selected terms in the source articles in the citation period on the probability that citations contained data in the following:

$$\text{CDAT} = c + \alpha_1\text{LMNSRC} + \alpha_2\text{YR80} + \alpha_3\text{YR85} + \alpha_4\text{YR90} + \alpha_5\text{YR95}, \tag{4}$$

where (i) CDAT equals one for articles having data analysis, and zero otherwise; (ii) LMNSRC equals one if the citing *JSTOR* article cites an *AER* source article containing "lemma" and/or "multiple equilibria," and equals zero otherwise; (iii) YR80, YR85, YR90, and YR95 are dummies variables for 1980, 1985, 1990, and 1995;[19] and (iv) c is the constant term.

The results of the probit estimation are in Table 7. The estimated LMNSCR coefficient is negative

[17] Citations include notes, comments, replies, and rejoinders as well as articles in the *AER Papers and Proceedings*. The search was limited to economics and finance; we limited our citation analysis to those that came up in the search using the author list as it occurred in *JSTOR*. Any listing that incorrectly cited the ordering of an article's authors was not included in our analysis.

[18] Only the Social Science Citation Index listings that were in JSTOR were included.

[19] Estimations (unreported) showed that the significance of the year dummies varied with the choice of the base year, but the significance of LMSRC was invariant to changes in the base year.

Table 6. Ratios of *JSTOR* Citations Having Operational Content to Total *JSTOR* Citations, by Year and by Citation Source (*AER* Articles with "Lemma" and/or "Multiple Equilibria" versus Random Sample of *AER* Articles)

Years	Ratios of *JSTOR* Citations Containing Data to Total Number of JSTOR Citations*	Ratios of *JSTOR* Citations Containing Data to Total Number of JSTOR Citations†
1976–1980	3/20 (15)	4/16 (25)
1981–1985	6/11 (55)	1/7 (14)
1986–1990	15/56 (27)	8/13 (62)
1991–1995	17/69 (25)	13/32 (41)
1996–2000	5/27 (19)	38/51 (75)

* Source: *AER* articles with "Lemma" and/or "Multiple Equilibira" in 1975, 1980, 1985, 1990, and 1995 (%).
† Source: Random sample of *AER* articles in 1975, 1980, 1985, 1990, and 1995 (%).

and significant at the 1% level. This can be interpreted as meaning that the presence of the term "lemma" and/or "multiple equilibria" in the source article has a negative impact on the probability of a citation containing any empirical analysis. These results are consistent with the hypothesis that theoretical complexity reduces the operationalism.

3. Summary and Conclusions

The assumption that resources are scarce relative to human wants is used in economics to generate operational statements about how things in the world behave. Empirical evidence and statistical analyses allow us to (i) cull theories whose predictions are inconsistent with observational reality and (ii) provide circumstances in which theories are applicable.

A tradeoff between operationalism and the mathematical complexity of economic theories was suggested by Alfred Marshall, directly hypothesized by Donald F. Gordon, and restated by Leontief. This article tested Gordon's hypothesis. Over the period of the study, analyses of the contents of

Table 7. Impact of Theoretical Complexity on the Probability of Citations Having Operational Content

Variable	Coefficient	Standard Error	*z* Statistic	Probability
Constant (*c*)	−0.515	0.256	−2.013	0.044**
LMNSRC	−0.695	0.165	−4.200	0.000***
YR80	0.660	0.383	1.723	0.085*
YR85	0.636	0.297	2.143	0.032**
YR90	0.436	0.280	1.559	0.119
YR95	0.902	0.287	3.147	0.002***

Dependent variable: CDAT; method, binary probit; 301 sample observations; 191 CDAT = 0 observations; 110 CDAT = 1 observation; Mean CDAT = 0.364; SD CDAT = 0.482; SE of regression = 0.455; Akaike information criterion = 1.229; sum squared residuals = 61.045; Schwarz criterion = 1.303; log likelihood = −179.038; Hannan-Quinn criterion = 1.259; restricted log likelihood = −197.603; average log likelihood = −0.595; LR statistic (5 df) = 37.130; McFadden $R^2 = 0.094$; probability (LR statistic) = 5.64×10^{-7}.

* Significant at the 10% level.
** Significant at the 5% level.
*** Significant at the 1% level.

complex mathematical articles and of the contents of the articles that cited the complex articles failed to refute the hypothesized tradeoff. Mathematically complex articles were less operational and were less likely to be cited in articles containing operational statements. Nevertheless, editors appear to have become consistently more likely to publish complex theorizing as shown by the presence of significant and positive trends toward increasing mathematical complexity in the time series data for general interest journals between 1963 and 1996.[20] In contrast, the empirically oriented *JEH* has shown no trend toward increasing complexity.

References

Anderson, Gary M., David M. Levy, and Robert D. Tollison. 1989. The half-life of dead economists. *Canadian Journal of Economics* XXII:174–83.

DeGroot, Morris H. 1975. *Probability and statistics*. Reading, MA: Addison-Wesley.

Gordon, Donald F. 1955. Operational propositions in economic theory. *Journal of Political Economy* 63:150–61.

JSTOR. 2000–2002. http://www.jstor.org.

Laband, David N., and Robert D. Tollison. 2000. Intellectual collaboration. *Journal of Political Economy* 108:632–62.

Laband, David N., Robert D. Tollison, and Gokhan Karahan. 2002. Quality control in economics. *Kyklos* 55:315–33.

Leontief, Wassily. 1971. Theoretical assumptions and nonobservable facts. *American Economic Review* 61:1–7.

Marshall, Alfred. 1964. *Principles of economics*. 8th edition. London: MacMillan & Co.

Samuelson, Paul A. 1952. Economic theory and mathematics—an appraisal. *American Economic Review* 42:56–66.

Sen, Amartya. 1995. Rationality and social choice. *American Economic Review* 85:1–24.

Social Science Citation Index. 1963–1996. Philadelphia, PA: Thompson Scientific.

Stigler, George J. 1969. Does economics have a useful past? *History of Political Economy* 1:217–30.

[20] This is not a condemnation of nonoperational theories. Such theories *may* generate operational statements in the future; there is a *potential* payoff. But economic analysis requires an assessment not only of the probabilities and magnitudes of potential benefits, but also the costs. The publication of nonoperational theories entails sacrificing the net benefits that forgone operationalized analyses would have generated.