ECONOMICS IN PRACTICE

The Market for Lemmas: Evidence that Complex Models Rarely Operate in Our World

PHILIP R.P. COELHO AND JAMES E. McCLURE

ABSTRACT

Many major economic journals publish models that can neither generate operational statements nor be challenged by evidence. Authors sometimes motivate these enterprises by allusions to “stylized facts.” Often, it is only in concluding remarks that authors provide vague directions about how “future research” might allow their results to operate in the realm of evidence. Coelho and McClure (2005, 562-564) present evidence that in the American Economic Review 1963 through 1996 “[m]athematically complex articles were less operational and were less likely to be cited in articles containing operational statements.” Empirical research suggests that the probability of subsequent articles appearing with refuting data, or any data, is substantially lower than in less mathematically complex articles. Another reason to doubt that mathematically complex models can generate operational research is that their assumptions are often complicated, substantively obscure, and unworldly.

Economics uses evidence to assess theories. Theories that do not provide evidentiary or testable propositions at reasonable costs are usually disregarded. The ability to formalize refutable statements and find evidence for or against them is operationalism. Statements that cannot be operationalized are what Wolfgang Pauli called “not even wrong.” Refutations instruct us on what is wrong; non-operational statements lack this virtue.

1 Miller College of Business, Ball State University. Muncie, Indiana 47306.
2 Klein and Romero (2007) develop several necessary conditions for when a “model” qualifies as a “theory”; examining the main articles in the 2004 volume of the Journal of Economic Theory they provide evidence for the “paucity of theory” in JET.
3 See Daniel M Hausman (1989, 120-121) on “unrealistic” assumptions, and on “tractability” see Frank Hindriks (2005).
4 The phrase not even wrong is being used in the debate over string theory in physics. Peter Woit has au-
Operationalism is not the only way to assess theories. Several authors (such as Coase 1982, 26f; Gibbard and Varian 1978, 669; Hausman 1992; Sugden 2000, 131) allow for a richer and more complex framework, wherein theories and evidence are mutually formulated, and encompassing formulations then judged by broad sensibilities. Voices critical of the emphasis on statistical operationalism often point out that the aspiration of statistical testing is not always attainable, and some conjectures are too important to neglect for that reason. Other forms of empirical argument must sometimes be sought (see Coase 1975, 58).

Here we focus on operationalism, the “stronger” empirical standard, for several reasons. First, operationalism is primarily what economists have in mind when they speak of “empirical work.” Second, weaker standards of evidence are more difficult to replicate and assess. Non-statistical approaches are generally more impressionistic, the details involved in replication typically leave room for great ambiguity. In our view, the profession’s emphasis on statistical operationalism as the marker of “empirical work” may be overdone, but not inordinately. Rather, our complaint is that empirical argumentation all too frequently is neither undertaken nor considered. Further, the ideas represented by complex models typically lack any apparent significance that might legitimately exempt them from operationalist demands.

Deploying evidence is a judgmental endeavor that depends upon such things as the particular application of the theory, the costs of establishing background conditions, and data acquisition. After establishing the conditions, judgments about “size” and “fit” can be considered; this is close to the concerns of D. N. McCloskey’s (1983).

Lemmas are formulated in proofs so complex that it is useful to divide the task into intermediate steps (Lemma 1, Lemma 2, . . . et cetera), like a stopover. We use the presence of the term “lemma(s)” as an indicator of mathematical complexity. The term “lemma(s)” has become increasingly frequent in the journal literature.
Figure 1 (above) presents evidence on the trend in the usage of lemmas in some top journals in economics. The vertical axis of Figure 1 represents the numbers of articles found per decade in a full-text search of the JSTOR data base that contain either the term lemma or lemmas in the American Economic Review, Economic Journal, Journal of Political Economy and Quarterly Journal of Economics. The appearance of the word “lemma(s)” was rare in the first six decades of the twentieth century, but during the last four decades it became increasingly frequent.

Radioactive Decay in Long Chains

Alfred Marshall addressed “long trains of deductive reasoning”:

It is obvious that there is no room in economics for long trains of

7 The year 2000 was chosen as the last year of consideration because at the time of our investigation JSTOR did not provide data for all of the four journals considered beyond that year. The first appearance of lemma(s) in any of these journals was in Edgeworth’s (1910) article in Ej. The numeric results shown in Figure 1 are: 1900-1910 (one article); 1911-1920 (two articles); 1921-1930 (zero); 1931-1940 (zero); 1941-1950 (two); 1951-1960 (zero); 1961-1970 (22); 1971-1980 (98); 1981-1990 (245); 1991-2000 (353). Data for 2001 became available for all four journals subsequent to the creation of this figure. In the year from 2000 to 2001 the JSTOR count of articles in the AER, EJ, JPE, and QJE containing the term “lemma” or “lemmas” was 83. The accelerating expansion of the “market for lemmas” in the final four decades of the 20th century continued in the first year of the 21st.

8 More inclusive measures of “mathematical complexity” could be presented, but looking at lemma(s) is good enough to illustrate the trend towards publication of articles of increasing mathematical complexity. For a more formal investigation with a more inclusive measure, see Coelho and McClure (2005, 560-561). In the future if the use of the term lemma is stigmatized, and authors of lengthy proofs avoid the stigma by calling intermediate steps in lengthy proofs something else (e.g., step 1, step 2, ...), then the value of this proxy for mathematical complexity would be reduced. Given the trends in the usage of the term lemma that we document, there is no evidence that its usage is being curtailed currently.
deductive reasoning; no economist, not even Ricardo, attempted them. It may indeed appear at first sight that the contrary is suggested by the frequent use of mathematical formulae in economic studies. But on investigation it will be found that this suggestion is illusory, except perhaps when a pure mathematician uses economic hypotheses for the purpose of mathematical diversions; for then his concern is to show the potentialities of mathematical methods on the supposition that material appropriate to their use had been supplied by economic study. He takes no technical responsibility for the material, and is often unaware how inadequate the material is to bear the strains of his powerful machinery. But a training in mathematics is helpful by giving command over a marvelously terse and exact language for expressing clearly some general relations and some short processes of economic reasoning; which can indeed be expressed in ordinary language, but not with equal sharpness of outline. And, what is of far greater importance, experience in handling physical problems by mathematical methods gives a grasp, that cannot be obtained equally well in any other way, of the mutual interaction of economic changes. (Marshall 1920, 644, emphasis added)

Paul Samuelson notes that both Alfred Marshall and John Stuart Mill spoke “of the dangers involved in long chains of logical reasoning;” and he explains that:

Marshall treated such chains as if their truth content was subject to radioactive decay and leakage—at the end of \( n \) propositions only half the truth was left, at the end of a chain of \( 2^n \) propositions, only half of half the truth remained, and so forth in a geometric multiplier series converging to zero truth. (Samuelson 1952, 57, emphasis added)

Subsequently, Donald F. Gordon (1955, 160) said: “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.” Using an example of a theory relating three distinct variables \( x, y, \) and \( z \), Gordon reasoned: “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 \( [z] \)” (155).

Problems arise if \textit{ceteris paribus} breaks down. As the length of a mathematical chain in an economic theory increased, Gordon suggested, it would become increasingly likely that the passage of time would in unpredictable ways compromise
the assumed stability of the chain. The timelessness implicit in multiple mathematical linkages was seen by Gordon as an obstacle to operationalizing complex mathematical theories about economic phenomena.9

The contrast between economic analytics (or “pure theory”) and statistical/econometric analytics10 is informative; in statistical analytics (such as in the development or refinement of a statistical test) assumptions are not affected by the passage of time. A century from now, the calculation of a Chi-squared statistic will require the same mathematical steps that are used today. In contrast, the subjects of economic theories are affected by the passage of time; a century from now the income elasticity of the demand for gasoline will have changed.

Table 1 provides evidence bearing upon the proposition that mathematical complexity in economic analytics tends not to be operationalized. It lists all articles in the 1980 volumes of the Journal of Economic Theory that contained 5 or more lemmas. The columns list how many lemmas each article had, how many citations each article had up to the June 2006, how many of the citing articles had empirical data, how many of the citing articles empirically tested a proposition of the original article, and how many citing articles refuted a proposition of the originating article.11

The 12 articles with five or more lemma generated 237 citations to them in the following (approximately) quarter century. Nine of the 237 citing articles contained empirical data, two had empirical data that had something to do with the propositions of the original article, and none had a definitive test leading to an acceptance or rejection of a proposition of the original article.12 In short, the

---

9. Wassily Leontief (1971, 1-2) echoed Gordon’s concerns about the timelessness implicit in mathematics: “Uncritical enthusiasm for mathematical formulation tends often to conceal the ephemeral substantive content of the argument behind the formidable front of algebraic signs.”

10. We use the term “economic analytics” to include both models and theories. We understand the distinction between “models” and “theories” and that not all models are theories, and, in conjunction, what we term “statistical/econometric analytics” would usually be called “statistical/econometric theory.” Again, the term “theory” may not always be appropriate because much work is often solely analytic refinements and explorations, as opposed to an explanation.

11. The ISI Web of Science was used to identify citations. Our search of this database occurred during the first two weeks of June of 2006. After citations were identified, each citing article was individually inspected to see whether it: (a) contained empirics, (b) attempted a direct assessment of any of the authors’ theoretical propositions, and (c) contained empirical assessments that accepted or rejected any proposition of the authors. Citing articles containing only casual empiricism were not counted as containing empirics, nor did the presence of simulations qualify them as containing empirics. However, citing articles containing data from surveys and/or experiments did qualify them as containing empirics.

12. For the originating (1980) articles listed in Table 1, the average number of lemmas per article is (79/12) or 6.58. For comparison purposes we counted the numbers of articles in JET in 2005 having 5 or more lemmas (there were 21 such article), and we counted the numbers of lemmas in these articles (there were 165 lemmas). In the 2005 set of articles the lemmas per article was 7.86. Comparing JET
12 originating articles have to date defined no operational propositions.\textsuperscript{13}

Determining whether an article’s propositions have been operational-
ized requires a fair amount of labor. The five-lemma threshold is the only sample we investigated. Twelve articles are not a large sample, but the results regarding those 12 are suggestive.

**The Most Cited Top-Journal Articles Rarely Contain Lemmas**

Kim, Morse, and Zingales (2006) have compiled a comprehensive list of articles published between 1970 and 2002 in 41 prominent journals in economics (and econometrics) that generated at least 500 citations to them as of June 2006. We accept this list as a proxy for what economists regard as best practice. We recognize that citation is an imperfect proxy for “best practice” or “what mattered most.” Citation counts are open to numerous criticisms (Klein and Chiang 2004, 137-39 summarize the concerns). Still, citation counts are widely considered the “gold standard” in assessing the impact of an article. Despite our misgivings we use the list developed by Kim, Morse, and Zingales to examine the impact of mathematization on article quality.

Using the their data we took all the articles that were published in the four top general-interest journals *AER, EJ, JPE*, and *QJE* that had 500 or more citations, and examined each to count the number of lemmas that appear in the articles. Table 2 summarizes the findings:

<table>
<thead>
<tr>
<th>Journal</th>
<th>Total Number of articles*</th>
<th>Number of articles that created at least one lemma</th>
</tr>
</thead>
<tbody>
<tr>
<td>AER</td>
<td>18</td>
<td>0</td>
</tr>
<tr>
<td>EJ</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>JPE</td>
<td>26</td>
<td>0</td>
</tr>
<tr>
<td>QJE</td>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td>TOTALS</td>
<td>59</td>
<td>1</td>
</tr>
</tbody>
</table>

*Data extracted from Table 2 of: Kim, Morse, and Zingales (2006, 15)

Of the fifty-nine articles in *AER, EJ, JPE*, and *QJE* that have been cited more than 500 times, only one article contained an author-written lemmas. Cho and Kreps (1987) used two lemmas in their publication in the *QJE.*
recognize that it takes decades to accumulate 500 citations and the lemma trend only really started in the 1970s. Still, the results suggest that mathematical complexity has almost never been professionally rewarded with super-high citations and publication in the top general-interest journals.

**Lemma Usage in the Most Cited: Economic vs. Statistical/Econometric Analytics**

Here we compare lemma usage in two types of articles: economic versus statistical/econometric analytics. We make this comparison because the Gordon hypothesis argues that there is an inverse correlation between mathematical complexity and operationalism in economic analytics. Coelho & McClure (2005, 562-564) provide empirical evidence supporting the Gordon hypothesis. In contrast, there is no reason to expect that the Gordon hypothesis applies to statistical/econometric analytics. We hypothesize that the citation “payoff” to complex mathematics in economic analytics will be smaller than the payoff in statistical/econometric analytics.

Again using the list of articles with 500+ citations, we added four more journals: two top model-building journals, *Review of Economic Studies* and *Journal of Economic Theory*, and two top statistical/econometric journals, *Journal of the American Statistical Association* and *Econometrica*. We examined all the articles on the Kim, Morse, Zingales list from the following 8 journals: *AER*, *Econometrica*, *EJ*, *JASA*, *JET*, *JPE*, *QJE*, and *ReStud*. This produced a list of 108 articles in all. Each article was examined individually to determine whether it: (1) it contained at least one author-written lemma; and (2) was devoted to economic analytics or to statistical/econometric analytics.

The scorings of the 108 are shown in the Excel file linked from Appendix 1, at the end of this paper, and results are displayed in the Table 3. Of the total 108 articles considered, 21 percent had at least one author-created lemma. Contrasting articles concerned with economic versus statistical/econometric analytics, we find that the percentage containing at least one lemma is 11 percent for the former versus 52 percent for the latter.

The frequencies are consistent with our hypothesis: The citation “payoff” to mathematical complexity in economic articles is smaller than in statistical/econometric articles. To assess whether the difference in lemma usage among most widely cited articles is statistically significant depending upon article purpose, a Chi-squared test was conducted using the data in Table 3. The Chi-squared statistic is 20.1. This leads to a rejection of the null hypothesis (that the frequency of lemma usage in the most-cited articles in statistical/econometric analytics is the same as the lemma frequency in articles in economics) at the 1% level.
Table 3: Contingency Table: Most Widely Cited Articles in Alternative Types of Articles by Lemma Usage

<table>
<thead>
<tr>
<th>Article Purpose</th>
<th>Articles with No Lemmas</th>
<th>Articles with at Least One Lemma</th>
<th>TOTAL</th>
<th>Percentage with at Least One Lemma</th>
</tr>
</thead>
<tbody>
<tr>
<td>Economic Analytics</td>
<td>72</td>
<td>9</td>
<td>81</td>
<td>11%</td>
</tr>
<tr>
<td>Statistical/Econometric Analytics</td>
<td>13</td>
<td>14</td>
<td>27</td>
<td>52%</td>
</tr>
<tr>
<td>TOTAL</td>
<td>85</td>
<td>23</td>
<td>108</td>
<td>21%</td>
</tr>
</tbody>
</table>

Again, the results of Table 3 are not meant as a direct test of the Gordon hypothesis, instead these results provide insights into the social returns to the usage of mathematical complexity in economic versus statistical/econometric analytics. Among the sample, the apparent return to complexity is significantly lower in economics, both statistically and quantitatively. From the perspective of operationalism, this makes intuitive sense: widely cited statistical/econometric analytics generally supply directly or contribute indirectly to econometric tests and techniques for the manipulation of data. These articles are widely cited because what they supply is useful for examining data in articles that are operationalizing theories.

Concluding Remarks

This paper is not a general criticism of the usage of mathematics in economics; it is instead about the displacement of operationalism as the core pursuit of economics by the pursuit of mathematical elegance and generality.\(^\text{15}\) Pauli’s indictment “not even wrong” says “even” because non-operational models are worse than wrong whenever they draw resources away from the creation and examination of operational propositions or fail to provide any information, insights, or hypotheses about observational reality.

Alfred Marshall (1920, 1) stated that economics is: “a study of mankind in the ordinary business of life.” This is in contrast to the mathematical ideal of gen-

\(^\text{15}\) In their 1986 analysis of the efficient quantity of mathematics in economics, Grubel and Boland argued: “Our study has one clear-cut conclusion: The editors of economics journals should reduce the space devoted to mathematically oriented material” (439).
erality, elegance, and “pure” theory unblemished by the pursuit of worldly considerations. If we are dealing with the “ordinary business of life” we are unlikely to encounter either absolute “Truth” or the elegance that is sought by purists. In the Marshallian tradition the best we can hope for are conditional statements that are dependent upon time and a host of other circumstances; here the use of mathematics will be tempered by measurements, operationalism, experience, history, and all the nuances that are relevant to the purposes at hand.

During the last century, economists have discussed the implications of mathematically complexity in economic theory. In 1920, Alfred Marshall stated that it was “obvious” that there was “no room in economics for long trains of deductive reasoning.” What was obvious to Marshall was not obvious to the economics profession writ large. In the mid-twentieth century the increasing mathematical complexity of economics led Donald Gordon (1955, 161) to speculate that concerns for operationalism in economics implied that “the practice of proliferating and manipulating functions has gone to somewhat incautious limits.”

The evidence here indicates that mathematical complexity in economics has expanded exponentially beyond the levels that Gordon decried as “incautious.” Mathematical complexity has commanded more resources in economics, yet the additional complexity has generated little in the way of operational propositions. Concerns for operationalism, measurement, empiricism, statistical testing, and history are the focus of an economics discipline that attempts to explain phenomena that exist in the world that real people inhabit.

**APPENDIX**

Using the list of 500+ citations articles found in Kim, Morse, Zingales (2006; Table 2), we examined the 108 articles published in *American Economic Review, Econometrica, Economic Journal, Journal of the American Statistical Association, Journal of Economic Theory, Journal of Political Economy, Quarterly Journal of Economics, and Review of Economic Studies* to determine whether it: (1) it contained at least one author-written lemma; and (2) was devoted to economic analytics or to statistical/econometric analytics. The results are summarized in our Table 3 above. The Excel file linked here contains the details. [Link](#).

---

16 In explaining *why* mathematical complexity spreads, Gordon Tullock (2005, 47) reasoned that it spreads in fields where opportunities for original research are limited relative to the number of people in the field. “One symptom of the existence of this condition is the development of very complex methods. Calculus will be used where simple arithmetic would do, and topology will be introduced in place of plane geometry. In many fields of social science these symptoms have appeared.”
REFERENCES


**Articles Listed in Table 1**


**About the Authors**

**Philip R. P. Coelho** is a Professor of Economics in the Miller College of Business at Ball State University. He has published in the *American Economic Review, Journal of Economic History, Southern Economic Journal, Economic Inquiry,* and other journals. His current research interests are the economic effects of morbid diseases, the biological basis of economic behavior, and ethics and behavior in business. His email is 00prcoelho@bsu.edu.

**James E. McClure** (Ph.D., Purdue University, 1983) is a Professor of Economics in the Miller College of Business at Ball State University. His publications appear in journals such as *Economic Inquiry, Journal of Economic Behavior and Organization,* and *Southern Economic Journal.* His primary research focus is applied microeconomics. His email is jmcclure@bsu.edu.

Go to January 2008 Table of Contents with links to articles