June 28, 2018

The strategies of economic research

An empirical study

Martin Paldam[1]

Abstract: The paper analyzes the structure of strategies of economic research on a sample of all regular research papers in 10 general interest journals every 5th year from 1997 to 2017. It is 3,415 papers, with an annual upward trend of 3.3%. I have classified the papers into eight categories: The fraction in theory and empirics are almost equal large. Most empiric papers use the classic strategy, which derives an operational model from theory and run regressions. Several trends are highly significant - notably two main ones: The fraction of theoretical papers has fallen by 26 pp (percentage points), while the fraction of papers using the classic strategy has increased by 15 pp. Many papers using the classic strategy have been analyzed using meta-analysis, which show that the typical paper exaggerate the results reported substantially. There is no reason to believe that other strategies have smaller problems.

Keywords: Research strategies, survey of journals
Jel classification: A14, B41

[1] Department of Economics and Business, Aarhus University, Fuglesangs Allé 4, DK-8210 Aarhus V, Denmark. E-mail: mpaldam@econ.au.dk, URL: htpp://www.martin.paldam.dk.

1. Introduction

This paper studies the research strategies used in economics on a sample of 3,415 papers. It is all regular papers published in years 1997, 2002, 2007 and 2017 in 10 *general interest* journals, listed in section 3. That is journals that publish (at least some) papers where a one-page executive summary may interest people outside economic research. The analysis builds on my belief that truth exists, but it is difficult to find, and *all* eight strategies have serious problems. This means that only when many studies using different methods reach a joint finding can we safely believe that it is true.

As shown in Table 1, the annual number of papers in our ten journals has increased 1.9 times or by 3.3% per year. I have classified the 3,415 papers in eight categories: (C1) Economic theory, (C2) Statistical methods, (C3) Surveys, (C4) Experiments, (C5) Event studies, (C6) Descriptive, (C7) Classic empirics, and (C8) New empirics. Section 2 briefly describes the eight strategies, and mentions their main problem(s). Hereby I do not imply that all – or even most – papers have the said problem, but we rarely know how serious the problem is when we read a paper.

Table 1. The 3,415 papers

Year	Papers	Fraction	Annual increase		
1997	464	14.4	From	To	In %
2002	518	15.3	1997	2002	2.2
2007	661	18.3	2002	2007	1.2
2012	881	26.0	2007	2012	5.9
2017	891	25.9	2012	2017	0.2
Sum	3,415	100	1997	2017	3.3

More than 90% of the papers are easy to classify, but there is a stochastic element involved in the classification of the rest. Thus, the study has some – hopefully random – measurement error. The Appendix gives the full counts per category, journal and year. It also reports shares by category and the change in the shares. By looking at data over two decades, I study how economic research develops.

The researchers in economics has three main layers: The A-level of about 500 tenured researchers working at the top-ten universities. They publish in the top 10 journals that bring less than 1,000 papers per year. They mainly cite each other, but they greatly influence other researchers.[2] The B-level consists of about 15-20,000 researchers who work in the other 4-500 research universities and publish in the next level of about 150 journals. The C-level includes more than 100,000 academic economists at institutes of higher learning who mainly teach, though some also manage to publish.[3]

The 150 B-level journals (also) try hard to have a serious refereeing process. If our selection is representative these journals have increased the number of papers published from about 7,500 in 1997 to about 14,000 papers in 2017. Thus, the B-level dominates our science, and the 10 journals chosen are all from that level. For the five years covered by our sample, it is about 6%, but only 5 of the 20 years are included. Thus, our sample covers about 2% of all papers published in the period, but it is a lager fraction of the papers in general interest journals.

By a crude estimate, the 150 journals have published a bit more than 200,000 papers in the 20 years. It is impossible for anyone to read more than a small fraction of this flood of papers. Consequently, researchers compete for space in journals and for attention from the readers, as measured in the form of citations. To be cited a paper has to hold a clear message. Obviously, there is an element of sales promotion in many papers.[4] This causes *exaggeration,* which is a joint problem for all eight strategies. This is in accordance with economic theory, which predicts that researchers will report results that are too good; see Paldam (2018).

In the empirical part of economics, methods have been developed to summarize results quantitatively, notably results reached from regressions. Section 4.5 discusses these meta-methods. It confirms that papers tend to exaggerate the results reached.

A 'perfect' study is very demanding, and requires far too much space to report, especially if the paper looks for usable results. Thus, each paper is just one look at an aspect of the problem analyzed. This has caused some proliferation of empirical strategies, but the classic strategy still dominates, and the domination has actually grown in the last 20 years.

2. They are surrounded by a large, ever-changing 'rim' of PhD students, non-tenured staff, visitors and research assistants, giving a large production and builds links to other institutions, notably at the B-level.
3. The World of Learning organization report on 36,000 universities, colleges and other institutes of tertiary education and research. Although many of the C-level institutions are quite modest, others are large. If half has a program of economics, with a staff of at least five economists, the total staff is 100,000.
4. The publication process has many market-like properties, and I shall refer to is as a market, even when it is not a perfect market.

2. The eight categories

Table 2 lists the categories and main numbers discussed in the rest of the paper. The first three strategies, (C1) to (C3), are theory methods discussed in section 2.1. The next two, (C4) and (C5), are experimental methods covered by section 2.2. Finally, section 2.3 covers (C6) to (C8), which are the methods using statistical inference.

Table 2a. All 3,415 papers. Fractions in percent

Group		Type of paper	Fraction	Groups
	(C1)	Economic theory	45.2	
Theory	(C2)	Statistical technique, including forecasting	2.5	49.6
	(C3)	Surveys, including meta-studies	2.0	
Experimental	(C4)	Experiments in laboratories	5.7	6.4
	(C5)	Events, including real life experiments	0.7	
	(C6)	Descriptive, deductions from data	10.7	
Data inference	(C7)	Classic empirical studies	28.5	43.7
	(C8)	New empirical techniques	4.5	

Table 2b. The change in the fractions from 1997 to 2017 in percentage points

Group		Type of paper	Fraction	Groups
	(C1)	Economic theory	-25.9	
Theory	(C2)	Statistical technique, including forecasting	2.2	-24.7
	(C3)	Surveys, including meta-studies	-1.0	
Experimental	(C4)	Experiments in laboratories	7.7	9.0
	(C5)	Events, including real life experiments	1.3	
	(C6)	Descriptive, deductions from data	2.4	
Data inference	(C7)	Classic empirical studies	15.0	15.8
	(C8)	New empirical techniques	-1.7	

Note: Section 3.4 tests if the pattern observed in Table 2b is statistically significant. The Appendix reports the full data.

2.1 Theory: (C1) to (C3)

(C1) ***Economic theory***. It is papers, where the main content is the development of a theoretical model. In a few papers, the analysis is verbal, but more than 95% rely on mathematics, though the technical level differs. Theory papers often start by a descriptive introduction giving the stylized fact the model explains, but the bulk of the paper is the formal analysis, building a model and deriving proofs of some propositions from the model. The working of the model often used illustrated by simulations, which include a calibration that differs greatly by the efforts made to reach realism. Often the simulations are in lieu of an analytical solution or illustrations suggesting magnitudes of the results reached.

The ideal theory paper presents a (simple) new model that recasts the way we look at something important. Such papers are rare and obtain large numbers of citations. Most theoretical papers present variants of known models and obtain few citations.

Theoretical papers suffer from the problem known as *T-hacking,*[5] where the able author by a careful selection of assumptions can tailor the theory to give the results desired. Thus, the proofs made from the model may represent the ability and preferences of the researcher rather than the properties of the economy.

(C2) ***Statistical method***. Papers reporting new estimators and tests are published in a handful of specialized journals in econometrics and mathematical statistics – such journals are not included. In our general interest journals, some papers compare estimators on actual data sets. If the demonstration of a methodological improvement is the main feature of the paper, it belongs to (C2), but if the economic interpretation is the main point of the paper, it belongs to (C7) or (C8).[6]

Some papers, including a special issue of Empirical Economics (vol. 53-1), deal with forecasting models. Such models normally have a weak relation to economic theory. They are sometimes justified precisely because of their eclectic nature. They are classified as either (C2) or (C6), depending upon the focus. It appears that different methods works better on different data sets, and perhaps a trade-off exists between the user friendliness of the model and the improvement reached.

(C3) ***Surveys***. When the literature in a certain field becomes substantial, it normally presents a motely picture with an amazing variation, especially when different schools exist in the field. Thus, a survey is needed, and our sample contain 68 survey articles. They are of two types, where the second type is still rare:

(C3.1) ***Assessed surveys***. Here, the author reads the papers and assesses what the most reliable results are. Such assessments demand judgement that is often quite difficult to distinguish from priors, even for the author of the survey.

(C3.2) ***Meta-studies***. They are qualitative surveys of estimates of parameters claimed to be the same. Over the two decades from 1997 to 2017, about 500 meta-studies have been

5. The concept of *T-hacking* is used for the tailoring of theory to fit priors or interests of the author. T-hacking is closely related to data-mining and overfitting discussed under (C7), the classic strategy.

6. In dubious cases, I have used the conclusion of the paper to assess its main purpose. In addition, it has played a role if the dataset used to illustrate the method seems suspiciously exotic.

made in economics. Our sample includes five, which is 0.15%.[7] Meta-Analysis has two levels: The basic level collects and codes the estimates and studies their distribution. This is a rather objective exercise where results replicate rather well. The second level analyzes the variation between the results. This is less objective. The papers analyzed by meta-studies are empirical studies using method (C7), though a few uses estimates from (C6) and (C8).

2.2 Experimental methods: (C4) and (C5)

Experiments are of three distinct types, where the last two are rare, so they are lumped together. They are taking place in real life.

(C4) (Lab) ***Experiments***. The sample had 1.9% papers using this method in 1997, and it has expanded to 9.7% in 2017. It is a technique that is much easier to apply to micro- than to macroeconomics, so it has spread unequally in the 10 journals, and many experiments are reported in a couple of special journals that are not included in our sample.

Most of these experiments take place in a laboratory, where the subjects communicate with a computer, giving a controlled, but artificial, environment.[8] A number of subjects are told a (more or less abstract) story and payed to react in either of a number of possible ways. A great deal of ingenuity has gone into the construction of such experiments and in the methods used to analyze the results. Lab experiments do allow studies of behavior that are hard to analyze in any other way, and they frequently show sides of human behavior that are difficult to rationalize by economic theory. It appears that such demonstration is a strong argument for the publication of a study.

However, everything is artificial – even the payment. In some cases, the stories told are so elaborate and abstract that framing must be a substantial risk;[9] see Levitt and List (2007) for a lucid summary, and Bergh and Wichardt (2018) for a striking example. In addition, experiments cost money, which limits the number of subjects. It is also worth pointing to the difference between expressive and real behavior. It is typically much cheaper for the subject to 'express' nice behavior in a lab than to be nice in the real world.

(C5) ***Event studies*** are studies of real world experiments. They are of two types:

(C5.1) ***Field experiments*** analyze cases where some people get a certain treatment and

---

7. The 500 meta-studies are 0.25% of the 200,000 papers, so the five meta-studies are representative. They are not enough to form a separate category.
8. Some experiments are more informal, by using different classrooms or in rare cases phone interviews. I have even seen a couple of studies where it was unclear how the experiment was done.
9. If the issue has a low saliency for the subjects, the answers in polls and presumably in experiment depends upon the formulations of the story told. The word 'framing' is used to cover the deliberate use of the formulation-dependency to reach results desired by the researcher.

others do not. The 'gold standard' for such experiments is double blind random sampling, where everything (but the result!) is pre-announced; see Christensen and Miguel (2018). Experiments with humans require permission from the relevant authorities, and the experiment takes time too. In the process, things may happen that compromise the strict rules of the standard.[10] Controlled experiments are expensive, as they require a team of researchers. Our sample of papers contains no study that fulfills the gold standard requirements, but there are a few less stringent studies of real life experiments.

(C5.2) ***Natural experiments*** take advantage of a discontinuity in the environment, i.e. the period before and after an (unpredicted) change of a law, an earthquake, etc. Methods have been developed to find the effect of the discontinuity. Often such studies look like (C7) classic studies with many controls that may or may not belong. Thus, the problems discussed under (C7) will also apply.

2.3 Empirical methods: (C6) to (C8)

The remaining methods are studies making inference from 'real' data, which are data samples where the researcher has no (or little) control of the data generating process, but the researcher does select the sample.

(C6) ***Descriptive studies*** are deductive. The researcher describes the data aiming at finding structures that tells a story, which can be interpreted. The findings may call for a formal test. If one clean test follows from the description,[11] the paper is classified under (C6). If a more elaborate regression analysis is used, it is classified as (C7). Descriptive studies often contain a great deal of theory.

Some descriptive studies present a new data set developed by the author to analyze a debated issue. In these cases, it is often possible to make a clean test, so to the extent that biases sneak in, they are hidden in the details of the assessments made when the data are compiled.

(C7) ***Classic empirics*** has three steps: (i) It starts by a theory that is developed into an operational model. (ii) Then it presents the data set, and (iii) finally it runs regressions.

The significance levels of the t-ratios on the coefficient estimated assume that the regression is the first meeting of the estimation model and the data. We all know that this is rarely the case. In practice, the classic method is often just a presentation technique. Its great

10. Justman (2018) studies the well-known STAR experiment in education and show how such problems arise. New medical drugs have to go through a number of independent trials, and a meta-study of these trials. Big efforts are often made to reach the gold standard, but still the meta-study regularly shows biases.
11. By a *clean test,* I mean a test that contains no control variables that are not part of the model.

virtue is that it can be used to analyze real problems outside academia. But the relevance comes with a price:

In order to reach general (ceteris paribus) results, the model has to contain control variables for the special conditions affecting the data sample. These variables do not follow from the theory, but they may be found by experiments. Thus, this strategy is rather malleable, making it susceptible to the preferences and interests of researchers and sponsors. This means that some papers using the classic technique are not what they pretend, as already pointed out by Leamer (1983); see also Paldam (2018) for new references and theory. The fact that data mining is tempting suggest that it is often possible to reach smashing results,[12] making the paper fun to read, which may be precisely why it is cited.

(C8) ***New empirical techniques***. Partly as a reaction to the problems of (C7) the last 3-4 decades has seen a whole set of new empirical techniques. They include co-integration tests and different types of VARs, but there are also studies using Bayesian techniques, Kalman Filters, causality tests, etc. I have found 162 (or 4.7%) papers using these techniques. The fraction was highest in 1997. Since then it varied, but with no trend.

I think that the main reason for the lack of success for the new empirics is that it is quite bulky to report a careful set of co-integration tests or VARs, and they often show results that are far from useful in the sense that they only appeal to insiders. With some introduction and discussion, there is not much space left in the article. Therefore, we are dealing with a cookbook that makes for a rather dull dish, which is difficult to sell in the market.

Note the contrast between (C7) and (C8): (C7) makes it possible to write papers that are too good, while (C8) makes them too dull. This contributes to explain why (C7) is getting (even) more popular and the lack of success of (C8), but then, it is arguable that it is more dangerous to act on exaggerated results than on results that are weak.

12. Data mining occurs when the published result is chosen from many made. It may lead to overfitting as shown by the simulation study Paldam (2016).

3. The ten journals

The 10 journals chosen are: (J1) Can [Canadian Journal of Economics], (J2) Emp [Empirical Economics], (J3) EER [European Economic Review], (J4) EJPE [European Journal of Political Economy], (J5) JEBO [Journal of Economic Behavior & Organization], (J6) Inter [Journal of International Economics], (J7) Macro [Journal of Macroeconomics] (J8) Kyklos, (J9) PuCh [Public Choice], (J10) SJE [Scandinavian Journal of Economics].

Section 3.1 discusses the choice of journals, while section 3.2 considers how journals use to deal with the pressure for publication. Section 3.3 shows the marked difference in publication profile of the journals, and section 3.4 tests if the trends in strategies are significant.

3.1 The selection of journals

I have used four selection criteria:

(i) Top ten journals are excluded.

(ii) Journals should be *general interest* journals – methodological journals are excluded.

(iii) Papers are in English, but the Canadian Journal include one paper in French.

(iv) The journals should be open to researchers from all countries, so that the majority of the authors are from outside the country of the journal.[13]

Table 3. The 10 journals covered

Journal		Volumes published					Papers published						Growth
Code	Name	1997	2002	2007	2012	2017	1997	2002	2007	2012	2017	All	% p.a.
(J1)	Can	30	35	40	45	50	68	43	55	66	46	278	-1.9
(J2)	Emp	22	27	32-43	42-43	52-53	33	36	48	104	139	360	7.5
(J3)	EER	41	46	51	56	91-100	56	91	89	106	140	482	4.7
(J4)	EJPE	13	18	23	28	46-50	42	40	68	47	49	246	0.8
(J5)	JEBO	32	47-49	62-64	82-84	133-144	41	85	101	207	229	663	9.0
(J6)	Inter	42	56-58	71-73	86-88	104-109	45	59	66	87	93	350	3.7
(J7)	Macro	19	24	29	34	51-54	44	25	51	79	65	264	2.0
(J8)	Kyklos	50	55	60	65	70	21	22	30	29	24	126	0.7
(J9)	PuCh	90-93	110-113	130-133	150-153	170-173	83	87	114	99	67	450	-1.1
(J10)	SJE	99	104	109	114	119	31	30	39	57	39	196	1.2
	All	-	-	-	-	-	464	518	661	881	891	3,415	3.3

Note. Growth is the average annual growth from 1997 to 2017 in the number of papers published.

13. This means that open journals from small countries such as Canadian Journal of Economics, Kyklos and Scandinavian Journal are included, while the good regional journals from the USA are excluded and so are the main German and French journals.

Papers included are regular research articles. Consequently, I exclude short notes to other papers and book reviews, [14] except for a few article-long discussions of controversial books.

The reason to exclude methodological journals is that methods are not interesting to outsiders. But they are developed to be used in general interest journals. From studies of citations, we know that useful methodological papers are highly cited. If they remains unused we presume that it is because they are useless, though, of course, there may be a long lag.

3.2 Creating space in journals

As mentioned in the introduction, the annual production has now reached about 1,000 papers in top journals and about 14,000 papers in the group of good journals.[15] This production has roughly doubled the last twenty years. The hard-working researcher will read perhaps 100 papers a year, which is 0.5% of the mass papers published.

We know that the publication record of researchers has become increasingly important for their career. In addition, the demand for economists increases, so the number of teachers and researchers at universities and research institutions increases as well. This has greatly increased the demand for space in journals, and we have seen that the number of papers published has increased by 3.3 % per year. Journals have used two methods to create space:

Book reviews have dropped to less than 1/3. Perhaps it also indicate that economists read fewer books than they used to. Journals have increasingly come to use smaller fonts and larger pages, allowing more words per page. The journals from North-Holland Elsevier that have managed to cram about two old pages into one new one. This makes it easier to publish papers, while they become harder to read.

Many journals have changed their numbering system for the annual issues, making it less transparent how much they publish. Only three – Canadian Economic Journal, Kyklos and Scandinavian Journal of Economics – have kept the schedule of publishing one volume of four issues per year. It gives about 40 papers per year. Public Choice has a (fairly) consistent system with four volumes of two (double) issues per year – this gives about 100 papers. The remaining journals have changed their numbering system and increased the number of papers published per year – often dramatically.

From these arguments follow that the wave of publications is not created by the demand

14. Thus, from Vol 41.3/5 of the European Economic Association the three survey papers have been included, while the 53 short proceeding papers have not been included. In addition, I have excluded special issues on the history of a journal or important researchers.

15. In addition, many more modest journal exist, and they seems to proliferate, notably because it is increasingly easy to publish on the net.

for reading material, but by the demand of researchers for outlets for their work. Consequently, the study confirms the old observation by C.L. Temple (1918, p 242): “ … as the world gets older the more people are inclined to write but the less they are inclined to read.”

3.3 *How different are the journals?*

The appendix reports the counts for each year and journal of the research strategies. From these counts, a set of χ^2-scores for the three groups of strategies have been calculated. They are reported in Table 4. It gives the χ^2-test comparing the profile of each journal to the one of the other nine journals taken to be the theoretical distribution.

Table 4. The methodological profile of the journals (χ^2-scores)

Journal Code	Name	(C1) to (C3) Theory	(C4) & (C5) Experiment	(C6) to (C8) Empirical	Sum $\chi^2(3)$-test	P-value (%)
(J1)	Can	7.4(+)	15.3(−)	1.7(−)	24.4	0.00
(J2)	Emp	47.4(−)	16.0(−)	89.5(+)	152.9	0.00
(J3)	EER	17.8(+)	0.3(−)	16.5(−)	34.4	0.00
(J4)	EJPE	0.1(+)	11.2(−)	1.0(+)	12.2	0.31
(J5)	JEBO	1.6(−)	1357.7(+)	41.1(−)	1404.4	0.00
(J6)	Inter	2.4(+)	24.8(−)	0.1(+)	27.3	0.00
(J7)	Macro	0.1(+)	18.2(−)	1.7(+)	20.0	0.01
(J8)	Kyklos	20.1(−)	3.3(−)	31.2(+)	54.6	0.00
(J9)	PuCh	0.0(+)	11.7(−)	2.2(+)	13.9	0.14
(J10)	SJE	10.5(+)	1.8(−)	8.2(−)	20.4	0.01

Note: The χ^2-scores are calculated relative to all other journals. The sign (+) or (−) indicate if the journal has too many or too few papers relatively in this category. The p-values for the $\chi^2(3)$-test always reject that the journal has the same methodological profile as the other nine journals.

The test rejects that the distribution is the same for any of the journals. The closest to the average is the European Journal of Political Economy and Public Choice. The two most deviating scores are for the most micro-oriented journal JEBO that brings ‘too’ many experimental papers, and, of course, Empirical Economics brings ‘too’ many empirical papers.

3.4 *Trends in the use of the strategies*

Table 2b already gave an impression of the main trends in the strategies preferred by economists. I now test if these impressions are statistically significant. The tests have to be tailored to disregard three differences between the journals: (i) their methodological profiles; (ii) the number of papers they publish; and (iii) the trends in these numbers. Table 5 reports a set of distribution free tests, which overcome these differences.

Table 5. Trend-scores and tests for the strategies across the 10 journals

Journal Code	Journal Name	(C1) Theory	(C2) Stat met	(C3) Survey	(C4) Exp.	(C5) Event	(C6) Descript.	(C7) Classic	(C8) New stats
(J1)	Can	[1, 9, 0]	[6, 3, 1]	[6, 3, 1]	[3, 1, 6]	[3, 1, 6]	[6, 4, 0]	[8, 2, 0]	[5, 4, 1]
(J2)	Emp	[2, 8, 0]	[6, 4, 0]	[0, 7, 3]	[0, 4, 6]	[3, 4, 3]	[6, 4, 0]	[8, 2, 0]	[4, 6, 0]
(J3)	EER	[3, 7, 0]	[4, 0, 6]	[1, 9, 0]	[9, 1, 0]	[3, 1, 6]	[7, 3, 0]	[8, 2, 0]	[3, 7, 0]
(J4)	EJPE	[1, 9, 0]	[0, 0,10]	[1, 9, 0]	[4, 0, 6]	[4 ,0 ,6]	[4, 6, 0]	[9, 1, 0]	[8, 1, 0]
(J5)	JEBO	[2, 8, 0]	[6, 1 , 3]	[6, 3, 1]	[7, 3, 0]	[6, 1, 3]	[4, 6, 0]	[8, 2, 0]	[2, 4, 3]
(J6)	Inter	[1, 9, 0]	[0, 0,10]	[0, 0,10]	[0, 0,10]	[0, 0,10]	[8, 2, 0]	[8, 2, 0]	[4, 6, 0]
(J7)	Macro	[6, 4, 0]	[5, 5, 0]	[7, 2, 1]	[0, 0,10]	[0, 0,10]	[9, 1, 0]	[3, 7, 0]	[1, 9, 0]
(J8)	Kyklos	[2, 8, 0]	[0, 0,10]	[2, 2, 6]	[2, 7, 1]	[0, 0,10]	[4, 6, 0]	[9, 1, 0]	[2, 2, 6]
(J9)	PuCh	[3, 7, 0]	[4, 3, 3]	[6, 3, 1]	[4, 3, 3]	[0, 0,10]	[5, 5, 0]	[6, 4, 0]	[6, 3, 1]
(J10)	SJE	[1, 9, 0]	[4, 0, 6]	[6, 3, 1]	[1, 3, 6]	[3, 1, 6]	[6, 4, 0]	[6, 4, 0]	[6, 1, 1]
All 100 per col.		[22,78,0]	[35,16,49]	[35,41,24]	[30,22,48]	[22, 8,70]	[59,41,0]	[73,27,0]	[42,43,13]
Binominal test		**0.00%**	**1.10%**	56%	33%	**1.61%**	8.86%	**0.00%**	100%

Note: The three trend scores in each $[I_1,I_2,I_3]$-bracket, are a Kendall-count over all 10 combinations of years. I_1 counts how often the share goes up. I_2 counts when the share goes down, and I_3 counts the number of ties. Most ties occur when there is no observations either year. Thus, $I_1 + I_2 + I_3 = 10$. The tests are two-sided binominal tests disregarding the zeroes. The test results in bold are significant at the 5% level.

The tests are done on the shares of each research strategy for each journal. As the data cover five years, it gives 10 pairs of years to compare.[16] The three trend-scores in the []-brackets count how often the shares go up, down or stay the same in the 10 cases. This is the count done for a Kendall rank correlation comparing the five shares with a positive trend (such as 1, 2, 3, 4 and 5). The first set of trend scores for (C1) and (J1) is [1, 9, 0]. It means that 1 of the 10 share-pairs increases, while 9 decreases and no ties are found. The two sided binominal test is 2%, so it is unlikely to happen. A great majority of the journals in the (C1)-column have a majority of falling shares. The important point is that the counts in one column can be added – as is done in the all-row – This gives a powerful trend test that disregard differences between journals and the number of papers published.

Four of the trend-tests are significant: It is the fall in theoretical papers and the rise in papers using the classical strategy, but there is also a rise in the share of stat method and event studies. It is surprising that there is no trend in the number of experimental studies, see however Table A2.

16. The ten pairs are: (1997, 2002), (1997, 2007), (1997, 2012), (1997, 2017), (2002, 2007), (2002, 2012), (2002, 2017), (2007, 2012), (2007, 2017), and (2012, 2017).

4. Discussion

The discussion assumes that the development in the strategies pursued by researchers in economics is a reaction to the demand and supply forces on the market for economic papers. Since a time span of 20 years is considered, economic theory predicts that the demand factors come to dominate. However, it is a problem if it is the demand for economic research or the demand by researchers for outlets for their research.

The shares add to 100, so the decline of one strategy means that the others rise. Section 4.1 looks at the biggest change – the reduction in theory papers. Section 4.2 discusses the rise in two new categories. Section 4.3 considers the large increase, in the classic strategy, while section 4.4 looks at what we know about that strategy from meta-analysis. Finally, section 4.5 turns to the slow penetration of the market by authors from the Far East.

4.1 The decline of theory: economics suffers from theory fatigue

The papers in economic theory have dropped from 59.5% to 33.6% – this is the largest change for any of the eight strategies. It is also highly significant in the trend test.

Most theory papers start from the standard model, and argue that a well-known conclusion reached from the model hinges upon a debatable assumption – if it changes, the conclusion changes. Such papers are not very exciting, but they are useful. From a literature on one main model, the profession learns it strengths and weaknesses. It appears that no generally accepted method exists to summarize this knowledge in a systematic way, though many thoughtful summaries have appeared.

I think that there is a deeper problem explaining theory fatigue. It is that many theoretical papers are quite unconvincing. Granted that the calculations are done right, believability hinges on the realism of the assumptions at the start, and of the results presented at the end. In order for a model to convince, it should (at least) demonstrate the realism either of the assumptions or of the outcome.[17] If both ends appear to hang in the air, it becomes a game giving little new knowledge about the world, however skillfully played.

The theory fatigue has caused a demand for simulations demonstrating that the models can mimic something in the world. Finn Kydland and Edward C. Prescott pioneered calibration methods (see their 1991). Sometimes the calibration is carefully done, but it often appears like a numerical solution of a model that is too complex to allow a mathematical solution.

17. The methodological point goes back to the large discussion generated by Friedman (1953).

4.2 Two examples of waves: One still rising and another is fizzling out

When a new method of gaining insights in the economy first appears, it is surrounded by doubts, but it also promises a high marginal productivity of knowledge. After some time the doubts subside, and many researchers enter the field. After some time this will cause the marginal productivity of the method to fall and the method becomes less interesting. The eight strategies include two newer ones: Lab Experiments and New Stats.[18]

It is not surprising that papers with lab experiments are increasing, though it did take a long time: The first papers developing the technique (Vernon Smith 1962) were from the 1960s, and Charles Plott organized the first experimental lab 10 years later. This created a new standard for experiments, but required an investment in a lab and some staff. Labs became more common in the 1990s as pc's got cheaper and IT-programs were written to handle experiments, but only 1.9% of the papers in the 10 journals reported lab experiments as late as in 1997. This has now increased to 9.7%, so the wave is still rising. The trend in experiments is concentrated in a few journals, so the trend test in Table 5 is insignificant.

In addition to the rising share of lab experiment papers in some journals, a journal of experimental economics was started in 1998, where it published three issues with 281 pages. In 2017 it had reached four issues with 1,006 pages,[19] which is an annual increase of 6.5%.

Compared with the success of experimental economics, the (more motely) category of New Stats has had a more modest success, as the fraction of papers in the 5 years are 5.8, 5.2, 3.5, 5.4, and 4.2, which has no trend.

Here, some investment is also needed, but it is mainly investments in human capital.[20] Some of the papers using the classic methodology do contain one table with Dickey-Fuller tests or some eigenvalues of the data matrix, but they are normally peripheral to the analysis. Also, a couple of papers use Kalman filters and a dozen papers using Bayesian VARs. However, it is clear that the New Stats has not really made much headway into the general interest journals.

4.3 The steady rise of the classic method

The typical classic paper provides estimates of a key effect that decision-makers outside academia want to know. This makes the paper policy relevant right from the start, and it is easy

18. The key inventor of each method was awarded a Nobel prize: Vernon Smith (in 2002) for Experimental Economics and Clive Granger (in 2003) for New Stats. The seminal papers were written in the 1960s and 1970s.
19. In this journal, the layout of articles has remained unchanged, so the number of pages is a good measure.
20. Johansen and Juselius (1990) introduced a second wave of new stat. About 2/3 of the paper is packed with mathematical statistics, while 1/3 tries to demonstrate the usefulness of the new tools on Danish and Finnish monetary data, but the paper reaches no policy conclusions that I could find. It is possible to see the paper as an effort in foundation building, which will make later papers more solid.

to write a one page executive summary to the said decision-makers.

The three-step convention (see section 2.4) is often followed rather loosely. It is common that the paper is justified by some stylized facts. If the theory is a variant of some model developed elsewhere, the theory-part may be rather short. The estimation model is nearly always much simpler than the theory. Thus, while the theory can be used to derive the model, the reverse does not apply. Sometimes, the model seems to follow straight from common sense, and if the link from the theory to the model is thin, it begs the question: Is the theory really necessary? In such cases, it is hard to be convinced that the tests 'confirm' the theory, but then, of course, tests can only say that the data does not reject the theory.

The classical method is often only a presentation devise. Think of a researcher who has reached a nice result through a long and tortuous path, including some failed attempts to find a nice publishable result. It is not possible to describe that path within the severely limited space of an article. In addition, such a presentation would be rather dull to read, and none of us likes to talk about wasted efforts that in hindsight seem a bit silly. Here the classic method becomes a convenient presentation device.

All datasets presumably contain some general and some special information, where the latter depends on the special circumstances prevailing when the data were compiled. Thus, the regression should be controlled for these circumstances in order to reach the general result. Such ceteris paribus controls are not part of the theory, so they have to be added. This requires judgement and often some experiments. It is difficult for the reader to know if only one set of controls were tried, or if a search aimed at producing certain results has been carried out.

Many papers using the classic strategy present some estimates using new stats, such as estimators controlling for simultaneity/causality.[21] Other papers use probit/logit and various regressors on panel data that control for residual correlation, etc. It is quite common to throw in some bits of exotic statistics technique to demonstrate the ability of the researcher as does the theory section. This presumably helps to generate credibility.

4.4 Knowledge about classic papers reached from meta-studies

A meta-study in economics analyze a set of *K* (primary) papers, with *N* estimates, that claim to

21. Such estimators are: (i) two-stage instrument regressions. They use instruments that frequently seem rather arbitrary, and often they are just listed in a note to a table. (ii) Arellano-Bond-Blundell GMM estimators that self-generate a large number of instruments that stay within the black box of the method. In addition (iii) Granger-type causality tests are used to compare the conditional lag structure of two time series. They have the problem that they often reach results that are rather fragile – and cumbersome to report – when more variables are included.

be of the same effect.[22] The estimates are normally from studies using the classic strategy. At least 7,000 papers have been coded and analyzed.[23] The meta-methods have now bloomed into meta-meta studies covering thousands of papers; see Ioannidis *et al.* (2017) and Doucouliagos *et al.* (2018). Three general results stand out:

(M1) The range of the estimates is typically amazingly large, given the high t-ratios reported. This suggests that t-ratios are less than they are supposed to be.

(M2) Publication bias (exaggerations) are quite common. As a crude rule of thumb, exaggeration is by a factor of two.

(M3) The meta average estimated from all K studies normally converges, and for $K > 30$, the meta average has normally stabilized to a well-defined value.

Individual studies using the classic method often look better than they are, and thus they are more uncertain than they appear, but we may think of the value of convergence for large Ks as the truth. The exaggeration is largest in the beginning of a new literature, but gradually it becomes smaller. Thus, the classic method does generate truth when the effect searched for has been studied from many sides. The word *research* does mean that the search has to be repeated! It is highly risky to trust a few papers only.

Meta-analysis has found other results such as: (i) Results in top-journals do not stand out. It is necessary to look at many journals, as many papers on the same effect are needed. (ii) Little of the large variation between results is due to the choice of estimators.

A similar development should occur also for experimental economics. Experiments fall in families: A large number covers the prisoners dilemma games, but there are also many studies of dictator games, auction games, etc. Surveys summarizing what we have learned about these games seems highly needed. Assessed summaries of old experiments are common, notably in introductions to papers reporting new ones. It should be possible to extract the knowledge reached by sets of related lab experiments in an objective way, by some sort of meta-technique, but this has barely started. The first pioneering meta-studies of lab experiments do find the usual wide variation of results from seemingly closely related experiments.[24]

---

22. A brief intro to meta-analysis is Paldam (2015), while a fine textbook is Stanley and Doucouliagos (2012).
23. Table 2 show that 28.5% of the papers use the classical method. 200,000 papers have been published since 1997. Assume that 75% are in general interest journals. This gives $0.285 \ x \ 0.75 \ x \ 200'000 \approx 40{,}000$ classic papers. If all papers covered by meta-analysis are from this group, the share of papers coded is $100 \ x \ 7{,}000/40{,}000 \approx 17\%$. Some of the papers covered are from other journals (or working papers) and from before 1997, but still we assess that 10% of all papers that might have been subjected to meta-analysis have actually been. This is a substantial sample but we do not know how representative it is.
24. See e.g., Engel (2011) and Fiala and Suetens (2017).

4.5 A note about the entry of papers from East Asia

Research institutions are increasingly integrated, both through the market for papers and through the many international exchange programs. Researchers are somewhat footloose. The authors from East Asia (Chinese-Japanese-Korean area) are rather easy to identify from their names and the home pages of the authors, though some may be from the East Asian diaspora in the West notably North America.

Table 6 shows the rate of penetration of East Asians into the market. There is no clear trend. Most of the co-authored papers appear to reflect co-operation created by the international PhD-programs. Thus, it appears that such programs work.

Table 6. Rate of penetration for East Asian authors

	1997	2002	2007	2012	2017	Growth p.a.
East Asian author	8.6	6.4	4.1	5.0	8.0	1.8%
East Asian co-author	5.6	4.2	5.7	7.6	8.9	1.9%
Sum	14.2	10.6	9.8	12.6	16.8	1.8%

Table 7. The strategies preferred by authors from East Asia

	(C1) to (C3) Theory	(C4) & (C5) Experiment	(C6) to (C8) Empirical	Sum $\chi^2(3)$-test	P-value (%)
East Asian author	12.1(+)	4.7(−)	7.8(−)	24.6	**0.00**
East Asian co-author	0.4(+)	1.4(+)	3.0(−)	3.0	15.7

Note: see Table 4.

Table 7 shows that papers with East Asian co-authors do not differ from all other papers, but papers with only East Asian authors have a significantly higher preference for theory.

A rather similar process is happening for researchers from Latin America and the former East Bloc (of the USSR and its satellite states), while the share of authors from the Indian sub-continent increased somewhat earlier.

5. Conclusions

The study presents evidence that over the last 20 years economic research has moved away from theory towards empirical work using the classic strategy.

From the eighties onward there has been a steady stream of papers pointing out that the classic strategy suffers from excess flexibility. It does deliver relevant results, but they tend to be too good. While, increasingly, we know the size of the problems of the classic strategy, systematic knowledge about the problems of the other strategies are missing. It is possible that the problems are smaller, but we do not know.

However, it is clear that obtaining solid knowledge about the size of important effects requires a great deal of papers analyzing many aspects of the effect, and a carful quantitative survey. It is a well-known principle in the harder sciences that results need repeated independent replication to be truly trustworthy. In economics, this is only accepted in principle.

The classic method of empirical research is gradually winning, and this is a fine development: It does give answers to important policy questions. These answers are highly variable and often exaggerated, but through the efforts of many competing researchers, solid knowledge will gradually emerge.

References:

Bergh, A., Wichardt, P.C., 2018. Mine, ours or yours? Unintended framing effects in dictator games. INF Working Papers, No 1205. Research Institute of Industrial Economics, Stockholm

Christensen, G., Miguel, E., 2016. Transparency, reproducibility, and the credibility of economics research. P.t. NBER WP 22989. Forthcoming in the *Journal of Economic Literature*

Doucouliagos, H., Paldam, M., Stanley, T.D., 2018. Skating on thin evidence: Implications for public policy. *European Journal of Political Economy* online first: https://doi.org/10.1016/j.ejpoleco.2018.03.004

Engel, C., 2011. Dictator games: A meta study. *Experimental Economics* 14, 583-610

Fiala, L., Suentes, S., 2017. Transparency and cooperation in repeated dilemma games: A meta study. *Experimental Economics* 20, 755-771

Friedman, M., 1953. *Essays in Positive Economics*, 1953. University of Chicago Press, Chicago

Ioannidis, J.P.A., Stanley, T.D., Doucouliagos, H(C). 2017. The power of bias in economics research. *The Economic Journal* 127, F236–F265

Johansen, S., Juselius, K., 1990. Maximum likelihood estimation and inference on cointegration – with application to the demand for money. *Oxford Bulletin of Economics and Statistics* 52, 169-210

Justman, M., 2018. Randomized controlled trials informing public policy: Lessons from the project STAR and class size reduction. *European Journal of Political Economy*, https://doi.org/10.1016/j.ejpoleco.-2018.04.005

Kydland, F., Prescott, E.C., 1991. The econometrics of the general equilibrium approach to business cycles. In Hylleberg, S., Paldam, M., eds. 1991. New approaches to empirical macroeconomics. Special issue of the *Scandinavian Journal of Economics* 93(2), 161-78

Leamer, E.E., 1983. Let's take the con out of econometrics. *American Economic Review* 73, 31-43

Levitt, S.D., List, J.A., 2007. On the generalizability of lab behaviour to the field. *Canadian Journal of Economics* 40(2), 347-70

Paldam M., 2016. Simulating an empirical paper by the rational economist. *Empirical Economics* 50, 1383-1407

Paldam, M., 2015. Meta-analysis in a nutshell: Techniques and general findings. *Economics. The Open-Access, Open-Assessment e-journal* 9, April 14th 2015

Paldam, M., 2018. A model of the representative economist, as researcher and policy advisor. *European Journal of Political Economy* online first: https://doi.org/10.1016/j.ejpoleco.2018.03.005

Smith, V., 1962. An experimental study of competitive market behavior. *The Journal of Political Economy* 70, 111-37

Stanley, T.D., Doucouliagos, H., 2012. *Meta-Regression Analysis in Economics and Business*. Routledge, Abingdon

Temple, C.L., 1918. *Native Races and Their Rulers; Sketches and Studies of Official Life and Administrative Problems in Nigeria*. Argus, Cape Town

Appendix: Two tables

Table A1. The counts for each of the 10 journals

			Number of papers	(C1) Theory	(C2) Stat theory	(C3) Surveys meta	(C4) Experi-ments	(C5) Event studies	(C6) Descrip tive	(C7) Classic empiric	(C8) New empiric
A: Year 1997	J1	Can	68	47		2			10	8	1
	J2	Emp	33	11	5			1	7	3	6
	J3	EER	56	34		3			4	12	3
	J4	EJPE	42	29		2			5	6	
	J5	JEBO	41	26			7		3	5	
	J6	Inter	45	35					1	7	2
	J7	Macro	44	18					1	10	15
	J8	Kyklos	21	10			1		4	6	
	J9	PuCh	83	40		7	1	1	8	26	
	J10	SJE	31	26		1				4	
B: Year 2002	J1	Can	43	27	1				5	7	3
	J2	Emp	36	1	14	1			4	7	9
	J3	EER	91	63		4	3		4	17	
	J4	EJPE	40	27		2			2	9	
	J5	JEBO	85	52		3	14		10	5	1
	J6	Inter	59	40					4	9	6
	J7	Macro	25	8	2	1				6	8
	J8	Kyklos	22	6			1		2	13	
	J9	PuCh	87	39	2		1		14	31	
	J10	SJE	30	18			2			10	
C: Year 2007	J1	Can	55	26		4			6	17	2
	J2	Emp	48	4	8				3	23	10
	J3	EER	89	55		2	1		8	20	3
	J4	EJPE	68	36		2			9	20	1
	J5	JEBO	101	73			10	3	3	12	
	J6	Inter	66	39					4	21	2
	J7	Macro	51	30	1				6	10	4
	J8	Kyklos	30	2		1			6	20	1
	J9	PuCh	114	53			4		19	38	
	J10	SJE	39	29	1			1	2	6	
D: Year 2012	J1	Can	66	33	1		1	1	8	21	1
	J2	Emp	104	8	16				17	38	25
	J3	EER	106	56			7	1	7	33	2
	J4	EJPE	47	12		1			2	31	1
	J5	JEBO	207	75	2	9	50		17	52	2
	J6	Inter	87	36					17	33	1
	J7	Macro	79	32	2	3			12	14	16
	J8	Kyklos	29	8					2	19	
	J9	PuCh	99	47			2		2	48	
	J10	SJE	57	32			2	1	22		
E: Year 2017	J1	Can	46	20	1	5			9	9	2
	J2	Emp	139	1	25			4	30	60	19
	J3	EER	140	75	1	1	16		13	32	2
	J4	EJPE	49	14			2	1	4	27	1
	J5	JEBO	229	66	1	3	63	9	11	76	
	J6	Inter	93	42					10	33	8
	J7	Macro	65	28	1	9			10	13	4
	J8	Kyklos	24	1			1		3	19	
	J9	PuCh	67	33		1	3		10	20	
	J10	SJE	39	19		1	1	1	4	12	1

Table A2. Counts, shares and changes for all ten journals

	Number	(C1)	(C2)	(C3)	(C4)	(C5)	(C6)	(C7)	(C8)
Year		I: Sum of counts							
1997	464	276	5	15	9	2	43	87	27
2002	518	281	19	11	21	0	45	114	27
2007	661	347	10	9	15	4	66	187	23
2012	881	339	21	13	62	3	106	289	48
2017	891	299	29	20	86	15	104	301	37
All years	3415	1542	84	68	193	24	364	978	162
Year		II: Average fraction in per cent							
1997	100	59.5	1.1	3.2	1.9	0.4	9.3	18.8	5.8
2002	100	54.2	3.7	2.1	4.1	-	8.7	22.0	5.2
2007	100	52.5	1.5	1.4	2.3	0.6	10.0	28.3	3.5
2012	100	38.5	2.4	1.5	7.0	0.3	12.0	32.8	5.4
2017	100	33.6	3.3	2.2	9.7	1.7	11.7	33.8	4.2
All years	100	45.2	2.5	2.0	5.7	0.7	10.7	28.6	4.7
Trends-scores		[0,10,0]	[7, 3, 0]	[4, 6, 0]	[9, 1, 0]	[5, 5, 0]	[8, 2, 0]	[10, 0, 0]	[3, 7, 0]
Binominal test		**0.13**	34	37	**2.1**	100	11	**0.13**	34
From	To	III: Change of fraction in percentage points							
1997	2002	-5.2	2.6	-1.1	2.1	-0.4	-0.6	3.3	-0.6
2002	2007	-1.8	-2.2	-0.8	-1.8	0.6	1.3	6.3	-1.7
2007	2012	-14.0	0.9	0.1	4.8	-0.3	2.0	4.5	2.0
2012	2017	-4.9	0.9	0.8	2.6	1.3	-0.4	1.0	-1.3
1997	2017	-25.9	2.2	-1.0	7.7	1.3	2.4	15.0	-1.7

Note: The Trend-scores are calculated as in Table 5. Compared to the results in Table 5 the results are similar, but the power is less than before. However, note that the results in Column (C4) dealing with experiments is stronger in Table A2. This has to do with the way missing observations is treated in the test.