Research Article

Martin Paldam*[#]

Methods Used in Economic Research: An Empirical Study of Trends and Levels

Abstract: The methods used in economic research are analyzed on a sample of all 3,415 regular research papers published in 10 general interest journals every 5th year from 1997 to 2017. The papers are classified into three main groups by method: theory, experiments, and empirics. The theory and empirics groups are almost equally large. Most empiric papers use the classical method, which derives an operational model from theory and runs regressions. The number of papers published increases by 3.3% p.a. Two trends are highly significant: The fraction of theoretical papers has fallen by 26 pp (percentage points), while the fraction of papers using the classical method has increased by 15 pp. Economic theory predicts that such papers exaggerate, and the papers that have been analyzed by meta-analysis confirm the prediction. It is discussed if other methods have smaller problems.

1 Introduction

This paper studies the pattern in the research methods in economics by a sample of 3,415 regular papers published in the years 1997, 2002, 2007, 2012, and 2017 in 10 journals. The analysis builds on the beliefs that truth exists, but it is difficult to find, and that *all* the methods listed in the next paragraph have problems as discussed in Sections 2 and 4. Hereby I do not imply that all – or even most – papers have these problems, but we rarely know how serious it is when we read a paper. A key aspect of the problem is that 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. Only when many studies using different methods reach a joint finding, we can trust that it is true.

Section 2 discusses the classification of papers by method into three main categories: (M1) *Theory*, with three subgroups: (M1.1) economic theory, (M1.2) statistical methods, and (M1.3) surveys. (M2) *Experiments*, with two subgroups: (M2.1) lab experiments and (M2.2) natural experiments. (M3) *Empirics*, with three subgroups: (M3.1) descriptive, (M3.2) classical empirics, and (M3.3) newer empirics. More than 90% of the papers are easy to classify, but a stochastic element enters in the classification of the rest. Thus, the study has some – hopefully random – measurement errors.

Section 3 discusses the sample of journals chosen. The choice has been limited by the following main criteria: It should be good journals below the top ten A-journals, i.e., my article covers B-journals, which are the journals where most research economists publish. It should be general interest journals, and the journals should be so different that it is likely that patterns that generalize across these journals apply to more (most?) journals. The Appendix gives some crude counts of researchers, departments, and journals. It assesses that there are about 150 B-level journals, but less than half meet the criteria, so I have selected about 15% of the possible ones. This is the most problematic element in the study. If the reader accepts my choice, the paper tells an interesting story about economic research.

All B-level journals try hard to have a serious refereeing process. If our selection is representative, the 150 journals have increased the annual number of papers published from about 7,500 in 1997 to about 14,000 papers in 2017, giving about 200,000 papers for the period. Thus, the B-level dominates our science. Our sample is about 6% for the years covered, but less than 2% of all papers published in B-journals in the period. However, it is a larger fraction of the papers in general interest journals.

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. It should be

---

[#] Home page: http://www.martin.paldam.dk

* Corresponding author: Martin Paldam, Department of Economics and Business, Aarhus University, Fuglesangs Allé 4, DK-8210 Aarhus V, Denmark, e-mail: mpaldam@econ.au.dk

uncontroversial that papers that hold a clear message are easier to publish and get more citations. Thus, an element of sales promotion may enter papers in the form of *exaggeration*, which is a joint problem for all eight methods. This is in accordance with economic theory that predicts that rational researchers report exaggerated results; see Paldam (2016, 2018). For empirical papers, meta-methods exist to summarize the results from many papers, notably papers using regressions. Section 4.4 reports that meta-studies find that exaggeration is common.

The empirical literature surveying the use of research methods is quite small, as I have found two articles only: Hamermesh (2013) covers 748 articles in 6 years a decade apart studies in three A-journals using a slightly different classification of methods,[1] while my study covers B-journals. Angrist, Azoulay, Ellison, Hill, and Lu (2017) use a machine-learning classification of 134,000 papers in 80 journals to look at the three main methods. My study subdivide the three categories into eight. The machine-learning algorithm is only sketched, so the paper is difficult to replicate, but it is surely a major effort. A key result in both articles is the strong decrease of theory in economic publications. This finding is confirmed, and it is shown that the corresponding increase in empirical articles is concentrated on the classical method.

I have tried to explain what I have done, so that everything is easy to replicate, in full or for one journal or one year. The coding of each article is available at least for the next five years. I should add that I have been in economic research for half a century. Some of the assessments in the paper will reflect my observations/experience during this period (indicated as my assessments). This especially applies to the judgements expressed in Section 4.

2 The eight categories

Table 1 reports that the annual number of papers in the ten journals has increased 1.9 times, or by 3.3% per year. The Appendix gives the full counts per category, journal, and year. By looking at data over two decades, I study how economic research develops. The increase in the production of papers is caused by two factors: The increase in the number of researchers. The increasing importance of publications for the careers of researchers.

Table 1: The 3,415 papers

Year	Papers	Fraction	Annual increase		
			From	To	In%
1997	464	13.6	1997	2002	2.2
2002	518	15.2	2002	2007	4.0
2007	661	19.4	2007	2012	4.6
2012	881	25.8	2012	2017	0.2
2017	891	26.1			
Sum	3,415	100	1997	2017	3.3

2.1 (M1) Theory: subgroups (M1.1) to (M1.3)

Table 2 lists the groups and main numbers discussed in the rest of the paper. Section 2.1 discusses (M1) theory. Section 2.2 covers (M2) experimental methods, while Section 2.3 looks at (M3) empirical methods using statistical inference from data.

2.1.1 (M1.1) Economic theory

Papers are where the main content is the development of a theoretical model. 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.

In a few papers, the analysis is verbal, but more than 95% rely on mathematics, though the technical level differs. Theory papers may 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. It is often demonstrated how the model works by a set of simulations, including a calibration made to look realistic. However, the calibrations differ greatly by the efforts made to reach realism. Often, the simulations are in lieu of an analytical solution or just an illustration suggesting the magnitudes of the results reached.

Theoretical papers suffer from the problem known as *T-hacking*,[2] 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

---

[1] Hamermesh also analyzes the pattern of author ages and co-authors. Card and DellaVigna (2013) further study the sociology of publications in the top journals.

[2] 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 in Sections 4.3 and 4.4 on the classical method.

Table 2: The 3,415 papers – fractions in percent

	Three main groups	Fraction		Eight subgroups	Fraction
(M1)	Theory	49.6	(M1.1)	Economic theory	45.2
			(M1.2)	Statistical technique, incl. forecasting	2.5
			(M1.3)	Surveys, incl. meta-studies	2.0
(M2)	Experimental	6.4	(M2.1)	Experiments in laboratories	5.7
			(M2.2)	Events, incl. real life experiments	0.7
(M3)	Data inference	43.7	(M3.1)	Descriptive, deductions from data	10.7
			(M3.2)	Classical empirical studies	28.5
			(M3.3)	Newer techniques	4.5

represent the ability and preferences of the researcher rather than the properties of the economy.

2.1.2 (M1.2) 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 (M1.2), but if the economic interpretation is the main point of the paper, it belongs to (M3.2) or (M3.3).[3]

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 (M1.2) or (M3.1), depending upon the focus. It appears that different methods work better on different data sets, and perhaps a trade-off exists between the user-friendliness of the model and the improvement reached.

2.1.3 (M1.3) Surveys

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

2.1.3.1 (M1.3.1) Assessed surveys

Here, the author reads the papers and assesses what the most reliable results are. Such assessments require judgement that is often quite difficult to distinguish from priors, even for the author of the survey.

2.1.3.2 (M1.3.2) Meta-studies

They are quantitative surveys of estimates of parameters claimed to be the same. Over the two decades from 1997 to 2017, about 500 meta-studies have been made in economics. Our sample includes five, which is 0.15%.[4] 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 seem to replicate rather well.[5] The second level analyzes the variation between the results. This is less objective. The papers analyzed by meta-studies are empirical studies using method (M3.2), though a few use estimates from (M3.1) and (M3.3).

2.2 (M2) Experimental methods: subgroups (M2.1) and (M2.2)

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

---

3 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.

4 The 500 meta-studies are 0.25% of the 200,000 papers, so the five meta-studies are representative. They are not enough to have their own group. A similar number of meta-studies have been made in business economics.

5 All that can differ is coding errors, and when they are random they hardly matter. However, I know of three meta-study replications only.

Table 3: The change of the fractions from 1997 to 2017 in percentage points

	Three main groups	Change		Eight subgroups	Change
(M1)	Theory	−24.7	(M1.1)	Economic theory	−25.9
			(M1.2)	Statistical technique, incl. forecasting	2.2
			(M1.3)	Surveys, incl. meta-studies	−1.0
(M2)	Experimental	9.0	(M2.1)	Experiments in laboratories	7.7
			(M2.2)	Events, incl. real life experiments	1.3
(M3)	Data inference	15.8	(M3.1)	Descriptive, deductions from data	2.4
			(M3.2)	Classical empirical studies	15.0
			(M3.3)	Newer techniques	−1.7

Note: Section 3.4 tests if the pattern observed in Table 3 is statistically significant. The Appendix reports the full data.

2.2.1 (M2.1) 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.[6] A number of subjects are told a (more or less abstract) story and paid 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;[7] 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.

(M2.2) **Event studies** are studies of real world experiments. They are of two types:

(M2.2.1) **Field experiments** analyze cases where some people get a certain treatment and others do not. The "gold standard" for such experiments is double blind random sampling, where everything (but the result!) is preannounced; 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.[8] 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.

(M2.2.2) **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 (M3.2) classical studies with many controls that may or may not belong. Thus, the problems discussed under (M3.2) will also apply.

2.3 (M3) Empirical methods: subgroups (M3.1) to (M3.3)

The remaining methods are studies making inference from "real" data, which are data samples where the

---

6 Some experiments are more informal, they use classrooms or in rare cases phone interviews. I have even seen a couple of studies where it was unclear how the experiment was done.

7 If the issue has a low saliency for the subjects, the answers in polls and presumably in experiments depend 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.

8 Justman (2018) studies the well-known STAR experiment in education and shows 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.

researcher chooses the sample, but has no control over the data generating process.

(M3.1) **_Descriptive studies_** are deductive. The researcher describes the data aiming at finding structures that tell a story, which can be interpreted. The findings may call for a formal test. If one clean test follows from the description,[9] the paper is classified under (M3.1). If a more elaborate regression analysis is used, it is classified as (M3.2). 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.

(M3.2) **_Classical empirics_** has three steps: It starts by a theory, which is developed into an operational model. Then it presents the data set, and 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; see also point (m1) in Section 4.4. In practice, the classical method is often just a presentation technique. The great virtue of the method is that it can be applied to real problems outside academia. The relevance comes with a price: The method is quite flexible as many choices have to be made, and they often give different results. Preferences and interests, as discussed in Sections 4.3 and 4.4 below, notably as point (m2), may affect these choices.

(M3.3) **_Newer empirics_**. Partly as a reaction to the problems of (M3.2), the last 3–4 decades have seen a whole set of newer empirical techniques.[10] They include different types of VARs, Bayesian techniques, causality/co-integration tests, Kalman Filters, hazard functions, etc. I have found 162 (or 4.7%) papers where these techniques are the main ones used. The fraction was highest in 1997. Since then it has 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 are unclear and difficult to interpret. With some introduction and discussion, there is not much space left in the article. Therefore, we are dealing with a cookbook that makes for rather dull dishes, which are difficult to sell in the market.

Note the contrast between (M3.2) and (M3.3): (M3.2) makes it possible to write papers that are too good, while (M3.3) often makes them too dull. This contributes to explain why (M3.2) is getting (even) more popular and the lack of success of (M3.3), but then, it is arguable that it is more dangerous to act on exaggerated results than on results that are weak.

3 The 10 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], and (J10) SJE [Scandinavian Journal of Economics].

Section 3.1 discusses the choice of journals, while Section 3.2 considers how journals 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 methods are significant.

3.1 The selection of journals

As mentioned in the introduction, I assess that there is about 150 B-level journals, which are good journals below the top ten, and thus, the journals where most research economists publish. I have picked 10 of these. The choice has been somewhat reduced by three criteria:

(i) They should be general interest journals – methodological journals are excluded. By general interest, I mean that they bring papers where an executive summary may interest policymakers and people in general. (ii) They should be journals in English (the Canadian Journal includes one paper in French), which are open to researchers from all countries, so that the majority of the authors are from outside the country

---

9 By a _clean test_, I mean a test that contains no control variables that are not part of the model.

10 Thus, we distinguish between regression techniques (including extensions) and techniques that claim to be better. Some of the newer techniques are from the 1970 and thus not really new. Sometimes these techniques can be presented as an extra column in a table or a paragraph, but they tend to dominate the paper.

Table 4: The 10 journals covered

Journal		Volume number					Regular research 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	42–3	52–3	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	47–9	62–4	82–4	133–44	41	85	101	207	229	663	9.0
(J6)	Inter	42	56–8	71–3	86–8	104–9	45	59	66	87	93	350	3.7
(J7)	Macro	19	24	29	34	51–4	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	90–3	110–3	130–3	150–3	170–3	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.

of the journal.[11] (iii) They should be sufficiently different so that it is likely that patterns, which apply to these journals, tell a believable story about economic research. Note that (i) and (iii) require some compromises, as is evident in the choice of (J2), (J6), (J7), and (J8) (Table 4).

Methodological journals are excluded, as they are not interesting to outsiders. However, new methods are developed to be used in general interest journals. From studies of citations, we know that useful methodological papers are highly cited. If they remain unused, we presume that it is because they are useless, though, of course, there may be a long lag.

The choice of journals may contain some subjectivity, but I think that they are sufficiently diverse so that patterns that generalize across these journals will also generalize across a broader range of good journals.

The papers included are the regular research articles. Consequently, I exclude short notes to other papers and book reviews,[12] except for a few article-long discussions of controversial books.

3.2 Creating space in journals

As mentioned in the introduction, the annual production of research papers in economics has now reached about 1,000 papers in top journals, and about 14,000 papers in the group of good journals.[13] The production has grown with 3.3% per year, and thus it has doubled the last twenty years. The hard-working researcher will read less than 100 papers a year. I know of no signs that this number is increasing. Thus, the upward trend in publication must be due to the large increase in the importance of publications for the careers of researchers, which has greatly increased the production of papers. There has also been a large increase in the number of researches, but as citations are increasingly skewed toward the top journals (see Heckman & Moktan, 2018), it has not increased demand for papers correspondingly. The pressures from the supply side have caused journals to look for ways to create space.

Book reviews have dropped to less than 1/3. Perhaps, it also indicates that economists read fewer books than they used to. Journals have increasingly come to use smaller fonts and larger pages, allowing more words

---

11 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.
12 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.

13 Card and DellaVigna (2013) show that submissions to top-journals have doubled, while the number of articles has stayed fairly constant, so rejection rates have doubled. Thus, the pressures on the B-level journals has increased. Some of the journals analyzed have published the number of submissions and the rejection rate, both seem to have increased substantially. In addition, many more modest journals exist, and they seem to proliferate, notably because it is increasingly easy to publish on the net.

Table 5: The methodological profile of the journals – χ^2-scores for main groups

Journal		(M1)	(M2)	(M3)	Sum	P-value
Code	Name	Theory	Experiment	Empirical	$\chi^2(3)$-test	(%)
(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 (−) indicates if the journal has too many or too few papers relatively in the 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.

per page. The journals from North-Holland Elsevier have managed to cram almost two old pages into one new one.[14] 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.

Thus, I assess the wave of publications is caused by the increased supply of papers and not to the demand for reading material. Consequently, the study confirms and updates the observation by 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 methods. From these counts, a set of χ^2-scores is calculated for the three main groups

of methods – they are reported in Table 5. 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.

The test rejects that the distribution is the same as the average for any of the journals. The closest to the average is the EJPE and Public Choice. The two most deviating scores are for the most micro-oriented journal JEBO, which brings many experimental papers, and of course, Empirical Economics, which brings many empirical papers.

3.4 Trends in the use of the methods

Table 3 already gave an impression of the main trends in the methods 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: their methodological profiles, the number of papers they publish, and the trend in the number. Table 6 reports a set of distribution free tests, which overcome these differences. The tests are done on the shares of each research method for each journal. As the data cover five years, it gives 10 pairs of years to compare.[15] 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

---

14 During the work on this article I have had the opportunity to discuss with four editors, and in the period covered I have been a co-editor myself. Thus, I have seen these processes at work.

15 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).

Table 6: Trend-scores and tests for the eight subgroups of methods across the 10 journals

Journal		(M1.1)	(M1.2)	(M1.3)	(M2.1)	(M2.2)	(M3.1)	(M3.2)	(M3.3)
Code	Name	Theory	Stat met	Survey	Exp.	Event	Descript.	Classical	Newer
(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 are 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.

comparing the five shares with a positive trend (such as 1, 2, 3, 4, and 5).

The first set of trend-scores for (M1.1) and (J1) is [1, 9, 0]. It means that 1 of the 10 share-pairs increases, while nine decrease and no ties are found. The two-sided binominal test is 2%, so it is unlikely to happen. Nine of the ten journals in the (M1.1)-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 disregards differences between journals and the number of papers published. (Table A1)

Four of the trend-tests are significant: The fall in theoretical papers and the rise in classical papers. 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, but see Table A2 (in Appendix).

4 An attempt to interpret the pattern found

The development in the methods pursued by researchers in economics is a reaction to the demand and supply forces on the market for economic papers. As already argued, it seems that a key factor is the increasing production of papers.

The shares add to 100, so the decline of one method 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 classical method, while Section 4.4 looks at what we know about that method from meta-analysis.

4.1 The decline of theory: economics suffers from theory fatigue[16]

The papers in economic theory have dropped from 59.5 to 33.6% – this is the largest change for any of the eight subgroups.[17] It is highly significant in the trend test. I attribute this drop to theory fatigue.

As mentioned in Section 2.1, the ideal theory paper presents a (simple) new model that recasts the way we look at something important. However, most theory papers are less exciting: They start from the standard model and argue that a well-known conclusion reached from the model hinges upon a debatable assumption – if it changes, so does the conclusion. Such papers are useful. From a literature on one main model, the profession learns its 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.

---

16 A contributing factor may be the shift in the relative costs making it relatively cheaper to do empirical papers.
17 Hamermesh (2013) and Angrist et al. (2017) found the same fall in the share of theory-papers, using different samples.

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 of either the assumptions or the outcome.[18] 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. Kydland and Prescott pioneered calibration methods (see their 1991). Calibrations may be carefully done, but it often appears like a numerical solution of a model that is too complex to allow an analytical solution.

4.2 Two examples of waves: one that is still rising and another that 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. Gradually the doubts subside, and many researchers enter the field. After some time this will cause the marginal productivity of the method to fall, and it becomes less interesting. The eight methods include two newer ones: Lab experiments and newer stats.[19]

It is not surprising that papers with lab experiments are increasing, though it did take a long time: The seminal paper presenting the technique was Smith (1962), but only a handful of papers are from the 1960s. 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 PCs got cheaper and software was developed to handle experiments, but only 1.9% of the papers in the 10 journals reported lab experiments 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 6 is insignificant, but it is significant in the Appendix Table A2, where it is done on the sum of articles irrespective of the journal.

In addition to the rising share of lab experiment papers in some journals, the journal Experimental Economics was started in 1998, where it published 281 pages in three issues. In 2017, it had reached 1,006 pages in four issues,[20] which is an annual increase of 6.5%.

Compared with the success of experimental economics, the motley category of newer empirics 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. Newer stats also require investment, but mainly in human capital.[21] Some of the papers using the classical methodology contain a table with Dickey-Fuller tests or some eigenvalues of the data matrix, but they are normally peripheral to the analysis. A couple of papers use Kalman filters, and a dozen papers use Bayesian VARs. However, it is clear that the newer empirics have made little headway into our sample of general interest journals.

4.3 The steady rise of the classical method: flexibility rewarded

The typical classical 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 in many cases, it is possible to write a one page executive summary to the said decision-makers.

The three-step convention (see Section 2.3) is often followed rather loosely. The estimation model is nearly always much simpler than the theory. Thus, while the model can be derived from a theory, 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

---

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

19 The key inventor of each method was awarded a Nobel prize: Vernon Smith (in 2002) for Experimental Economics and Clive Granger (in 2003) for various Newer methods. The seminal papers were written in the 1960s and 1970s.

20 In this journal, the layout of articles has remained unchanged, so the number of pages is a good measure.

21 Johansen and Juselius (1990) introduced a second wave of co-integration techniques. 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. The paper reaches no policy conclusions that I could find, but it is possible to see the paper as an effort in foundation building, which will make later papers more solid.

the tests "confirm" the theory, but then, of course, tests only say that the data do not reject the theory.

The classical method is often only a presentation devise. Think of a researcher who has reached a nice publishable result through a long and tortuous path, including some failed attempts to find such results. 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 classical method becomes a convenient presentation device.

The biggest source of variation in the results is the choice of control/modifier variables. All datasets presumably contain some general and some special information, where the latter depends on the circumstances prevailing when the data were compiled. 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 many possible controls may be added. The ones chosen for publication often appear to be the ones delivering the "right" results by the priors of the researcher. The justification for their inclusion is often thin, and if two-stage regressions are used, the first stage instruments often have an even thinner justification.

Thus, the classical method is rather malleable to the preferences and interests of researchers and sponsors. This means that some papers using the classical 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 suggests that it is often possible to reach smashing results, making the paper nice to read. This may be precisely why it is cited.

Many papers using the classical method throw in some bits of exotic statistics technique to demonstrate the robustness of the result and the ability of the researcher. This presumably helps to generate credibility.

4.4 Knowledge about classical papers reached from meta-studies

A meta-study in economics analyzes a set of K (primary) papers, with estimates that claim to be of the same effect.[22] The estimates are normally from studies using the classical method. At least 20,000 papers have been coded and analyzed.[23] The meta-methods have now bloomed into meta-meta-studies covering thousands of papers; see Ioannidis, Stanley, and Doucouliagos (2017) covering 159 meta-studies of 6,700 primary papers and Doucouliagos, Paldam, and Stanley (2018) covering 101 meta-studies of 4,300 primary papers. Three general results stand out:

(m1) The range of the estimates is typically amazingly large, given the high t-ratios reported. This confirms that t-ratios are problematic as claimed in Section 2.3.

(m2) Publication biases (exaggerations) are common, i.e., meta-analyses routinely reject the null hypothesis of no publication bias. My own crude rule of thumb is that exaggeration is by a factor two – the two meta–meta studies cited give some support to this rule.[24]

(m3) The meta-average estimated from all K studies normally converges, and for $K > 30$, the meta-average normally stabilizes to a well-defined value, see Doucouliagos et al. (2018).

Individual studies using the classical 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 Ns (number of observations) as the truth. The exaggeration is largest in the beginning of a new literature, but gradually it becomes smaller. Thus, the classical 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.

---

22 Stanley and Doucouliagos (2012) is a fine textbook of meta-analytical method in economics, while Paldam (2015) is a brief intro.

23 Table 2 shows 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 $0.285 \times 0.75 \times 200,000 \approx 40,000$ classical papers. Most papers covered by meta-analysis are from this group. Thus, a crude estimate is that the share of papers coded is $100 \times 7,000/40,000 \approx 17\%$. The papers covered are from a wide range of journals (or working papers) and some are from before 1997, but still I 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 The argument assumes that theory predicts that the effect studied is positive, such as one. Exaggeration therefore means that the published result is significantly larger than one, such as two. The (rare) case where theory predicts that the effect is zero is not discussed at present.

Meta-analysis has found other results such as: 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. 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 cover prisoner's dilemma games, but there are also many studies of dictator games, auction games, etc. Surveys summarizing what we have learned about these games seem 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 a quantitative 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.[25] A recent large-scale replicability study by Camerer et al. (2018) finds that published experiments in the high quality journal Nature and Science exaggerate by a factor two just like regression studies using the classical method.

5 Conclusion

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

From the eighties onward, there has been a steady stream of papers pointing out that the classical method suffers from excess flexibility. It does deliver relevant results, but they tend to be too good.[26] While, increasingly, we know the size of the problems of the classical method, systematic knowledge about the problems of the other methods is weaker. It is possible that the problems are smaller, but we do not know.

Therefore, it is clear that obtaining solid knowledge about the size of an important effect requires a great deal of papers analyzing many aspects of the effect and a careful 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 classical 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

Angrist, J., Azoulay, P., Ellison, G., Hill, R., & Lu, S. F. (2017). Economic research evolves: Fields and styles. *American Economic Review* (*Papers & Proceedings*), *107*, 293–297.

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 Econ, Stockholm*. München: CESifo.

Brodeur, A., Cook, N., & Heyes, A. (2020). Methods matter: p-Hacking and publication bias in causal analysis in economics. *American Economic Review*, *110*(11), 3634–3660.

Camerer, C. F., Dreber, A., Holzmaster, F., Ho, T.-H., Huber, J., Johannesson, M., … Wu, H. (27 August 2018). *Nature Human Behaviour*. https://www.nature.com/articles/M2.11562-018-0399-z

Card, D., & DellaVigna, A. (2013). Nine facts about top journals in economics. *Journal of Economic Literature*, *51*, 144–161

Christensen, G., & Miguel, E. (2018). Transparency, reproducibility, and the credibility of economics research. *Journal of Economic Literature*, *56*, 920–980

Doucouliagos, H., Paldam, M., & Stanley, T. D. (2018). Skating on thin evidence: Implications for public policy. *European Journal of Political Economy*, *54*, 16–25

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*. Chicago: University of Chicago Press.

Hamermesh, D. (2013). Six decades of top economics publishing: Who and how? *Journal of Economic Literature*, *51*, 162–172

Heckman, J. J., & Moktan, S. (2018). Publishing and promotion in economics: The tyranny of the top five. *Journal of Economic Literature*, *51*, 419–470

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

---

25 See e.g., Engel (2011) and Fiala and Suentes (2017).
26 That is, too large and too significant, see Brodeur, Cook, and Heyes (2020).

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*, *54*, 167–174

Kydland, F., & Prescott, E. C. (1991). The econometrics of the general equilibrium approach to business cycles. *Scandinavian Journal of Economics*, *93*, 161–178

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*, 347–370

Paldam, M. (April 14th 2015). Meta-analysis in a nutshell: Techniques and general findings. *Economics. The Open-Access, Open-Assessment E-Journal*, *9*, 1–4

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

Paldam, M. (2018). A model of the representative economist, as researcher and policy advisor. *European Journal of Political Economy*, *54*, 6–15

Smith, V. (1962). An experimental study of competitive market behavior. *Journal of Political Economy*, *70*, 111–137

Stanley, T. D., & Doucouliagos, H. (2012). *Meta-regression analysis in economics and business*. Abingdon: Routledge.

Temple, C. L. (1918). *Native races and their rulers; sketches and studies of official life and administrative problems in Nigeria*. Cape Town: Argus

Appendix: Two tables and some assessments of the size of the profession

The text needs some numbers to assess the representativity of the results reached. These numbers just need to be orders of magnitude. I use the standard three-level classification in A, B, and C of researchers, departments, and journals. The connections between the three categories are dynamic and rely on complex sorting mechanisms. In an international setting, it matters that researchers have preferences for countries, notably their own. The relation between the three categories has a stochastic element.

The World of Learning organization reports on 36,000 universities, colleges, and other institutes of tertiary education and research. Many of these institutions are mainly engaged in undergraduate teaching, and some are quite modest. If half of these institutions have a program in economics, with a staff of at least five, the total stock of academic economists is 100,000, of which most are at the C-level.

The A-level of about 500 tenured researchers working at the top ten universities (mainly) publishes in the top 10 journals that bring less than 1,000 papers per year;[27] see Heckman and Moktan (2020). They (mainly) cite each other, but they greatly influence other researchers.[28] The B-level consists of about 15–20,000 researchers who work at 4–500 research universities, with graduate programs and ambitions to publish. They (mainly) publish in the next level of about 150 journals.[29] In addition, there are at least another 1,000 institutions that strive to move up in the hierarchy.

---

27 Roughly, they publish half of their papers in the top-ten journals, and this fills half of these journals.
28 They are surrounded by a large, ever changing "rim" of PhD students, non-tenured staff, visitors and research assistants, which gives a large scientific production and builds links to other institutions, notably at the B-level.
29 As the promotion systems at universities in many countries use journal classifications, many lists of A, B and C journals exist, where the A list is 5–10. The lists are not the same, but they overlap. I hope that the number 150 corresponds to the number of general interest B-journals that the country of the reader recognizes.

Table A1: The counts for each of the 10 journals

		Main group	(M1)			(M2)		(M3)		
		Subgroup	(M1.1)	(M1.2)	(M1.3)	(M2.1)	(M2.2)	(M3.1)	(M3.2)	(M3.3)
		Number papers	Theory	Stat. theory	Surveys meta	Experiments	Event studies	Descriptive	Classical empiric	Newer empiric
A:	(J1)	Can	68	47		2			10	8	1
Year	(J2)	Emp	33	11	5			1	7	3	6
1997	(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:	(J1)	Can	43	27	1				5	7	3
Year	(J2)	Emp	36	1	14	1			4	7	9
2002	(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:	(J1)	Can	55	26		4			6	17	2
Year	(J2)	Emp	48	4	8				3	23	10
2007	(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:	(J1)	Can	66	33	1		1	1	8	21	1
Year	(J2)	Emp	104	8	16				17	38	25
2012	(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:	(J1)	Can	46	20	1	5			9	9	2
Year	(J2)	Emp	139	1	25			4	30	60	19
2017	(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 for subgroups

	Number	(M1.1)	(M1.2)	(M1.3)	(M2.1)	(M2.2)	(M3.1)	(M3.2)	(M3.3)
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	3,415	1,542	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 6. Compared to the results in Table 6, the results are similar, but the power is less than before. However, note that the results in Column (M2.1) dealing with experiments are stronger in Table A2. This has to do with the way missing observations are treated in the test.