November 16-2022

Research methods in economics

Martin Paldam
Department of Economics, University of Aarhus, Denmark
E-mail: mpaldam@econ.au.dk and home page: http://www.martin.paldam.dk

Abstract
Economic journals publish 15,000 papers annually. The share of empirics is rising. Journals demand short and clear papers, so they often concentrate on one aspect of the problem, while other papers analyze other aspects. Three main methods are: (i) Theory needs few clear assumptions. As many assumptions are possible the results vary. (ii) Experimental papers report lab-experiments, where the behavior of people are studied under controlled conditions. This may cause expressive behavior, where people behave better than they actually do. (iii) Classical empiric papers, where available data are subjected to regression analysis. This suits problems where decision makers want answers. However, it is tempting to run many regressions and pick the best. It may not replicate. The diversity of papers on the same problem means that it is analyzed from many sides, as it should. But to look at one papers only may be misleading.

Keywords:
Strict vs partial replication
Theoretical model variants
Expressive lab-experiments
Data mining and regressions

1. Introduction

The economic literature is large and diverse. The top-ten A-journals publish about 1,000 papers a year. The next level of about 150 B-journals publish 14,000 papers. These numbers grow by app 3% per year. This is an annual production of 15,000 papers, but there are probably a further 500 more modest journals publishing at least as many papers. The organization World of Learning reports on 36,000 universities, colleges and other institutions of tertiary education and research. If half of these institutions have a staff of at least five economists, there is a stock of academic economists of about 100,000 persons.

The 15,000 papers are the way economic researchers communicate. Papers are on average 20 pages long in the printed version. Nobody can read or even look at 15,000 papers. We all prefer papers that give clear results, and we remember papers with a simple core. The profession discriminates against long papers with complex and weak results.

To make a perfect research paper in economics is almost impossible. Over three decades Berndt P. Stigum tried to develop the conditions for the perfect economic paper combining theory and empirics. His description grew to 1,000 pages; see Stigum (2015). He also published an easy reader's versions of his work, but he did not manage to get below 300 pages. A paper that meets his demands would surely exceed 200 pages, and thus be unacceptable in any journal.

For the researcher it is essential to learn how to cut out one piece of a larger research project and make it fit into a neat package of 20 printed pages. It is important to keep in mind that referees, editors, and readers want clarity. Thus, readers of economic journals should keep in mind that a paper is a carefully limited package dealing with one aspect of the problem analyzed. It is likely to be a bit too clear. That is, results tend to be exaggerated.

2. All economic papers: Some statistics about the methods used

The many papers of the economic literature are rather different. Only a few papers have tried to survey the methods used. Fortunately, they gave almost the same results. Hamermesh (2013) covers 748 articles six years a decade apart in three A-journals. Angrist *et al*. (2017) use a machine-learning classification of 134,000 papers in 80 journals to look at the three main methods: theory, empirics, and econometrics.

Table 1. The methods used in a sample of papers

Main group	All papers	Trend 1997	Trend 2017	Change	Largest sub-group in 2017
1. Theory	49.7%	63.8%	39.1%	-24.7%	Economic theory 33.6%
2. Experimental	6.5%	2.3%	11.3%	9.0%	Lab experiments 9.7%
3. Data inference	43.8%	33.9%	49.6%	15.7%	Classical method 33.8%

Count based on a sample of 3,415 papers in ten journals, every fifth year from 1997 to 2017. Econometric theory is counted as theory, while applied econometrics is counted as data inference. Source: Paldam (2021).

Table 1 looks at a somewhat different methodological classification of 3,415 papers. Most papers use one main method, but about 5% of the papers have been difficult to classify. The sample of papers was chosen to be representative. It covered all annual issues for ten journals in the B-group, five years apart from 1997 to 2017. Table 1 reports the classification reached. To the extent it overlaps with Angrist *et al.* (op cit.), the results are the same.

The two decades covered have seen large changes. (a) A large fall in (1) the share of theoretical papers, and (b) a corresponding large increase in (2) and (3) the share of empirical papers. The largest and fastest growing part of the empirical papers uses the well-known conventions of the ***classical method***; see Table 2.

Table 2. The four steps of the classical method, and two temptations invalidating tests

Step 1	Survey of previous literature, pointing to the innovation in the paper
Step 2	Derivation of model with operationalization to allow estimation
Step 3	Presentation of the data selected for the project
Step 4	Regressions confirming the model. Rejections may occur, but they are rare
	Ad hoc control variables are often added to give variants of the model
Temptation 1	Research loops: Empirical results lead to revision of model, so that it is confirmed
Temptation 2	Data mining: Many models are run - the researcher chooses the best, suppresses the rest

The fact that few economic papers reject their model suggests that many researchers give in to the two temptations also mentioned in Table 2. Another sign is that t-ratios cluster just above two; see Brodeur *et al*. (2016). It certainly means that empirical models urgently need independent replication. That is, replication of the same model on another data set and by another researcher. We shall return to the problems of the classical method below. Due to these problems, many post-classical empirical techniques have been developed. They are used in less than 5% of the papers, and there is no clear trend.

3. The need for replication – meta-analysis to the rescue

All sciences know that empirical results need repeated, independent replication before they can be trusted. Due to the two temptations, economics studies certainly need replication. Economists also know that replication studies are difficult to publish. Only 0.1% of all economics papers are replicated (often with embarrassing results); see Müller-Langer *et al.* (2019). A large literature points to the replication crisis in economics and the other social sciences. Hence, there is a problem.

Assume that we look at an effect, β, where a basic theoretical model explains β qualitatively. Typically, the theory predicts the sign on β. Paper A provides a nice clear analysis, and a fine estimate of β, with the right sign. There is no paper with an independent replication of A, but if β is important, there is likely to be a β-literature with many other estimates of β. The β-literature uses variants of the model, different data samples, and several estimators. Maybe there are 75 papers with, on average, 10 estimates $b \approx \beta$ in each. Thus, we have 750 estimates of β. Instead of strict replication, the profession has a ***swarm of partial replications***.

Meta-analysis is a technique to analyze one such swarm and to study if it gives a result β^* that can be considered the true value. Thus, meta-studies aim at replacing replication. Table 3 surveys how a meta-study is done.

Table 3. The four to five steps of a meta-analysis

Step 1	All studies with estimates b of β are collected
Step 2	The estimates and the way they are reached are coded using the same scale
Step 3	The distribution of the estimates b is analyzed. It often has asymmetries that suggest bias
Step 4	Calculation of the mean, $\bar{b}$, and the meta-average, b_M, that corrects the mean for the said bias
Step 5	It is analyzed why the estimates differ, and the mean may be further corrected

Steps (1) – (4) are robust in the sense that if the same β-literature is analyzed by two independent meta-studies they should get virtually the same result. Step (5) is much less robust and may even allow priors to sneak back; see Paldam (2022).

Both step (1) and especially (2) are very labor intensive. If the models used to estimate β are different, it may be necessary to make different meta-studies, but in practice the operational models are normally much less different than the theoretical models. Thus, they are often easy to code by just adding a binary dummy for the theory.

About 1,000 meta-studies have been done in economics. Each of these studies deal with

the literature that claims to estimate the same β. The studies show that most β-literatures have an ***amazingly large variation***. It is quite common to find highly significant results that differ by a factor two or even three. In addition, the distribution of the estimates often suggests ***publication bias***, in the form of exaggeration.

The variation and exaggeration point to the ***flexibility*** of economics. A research paper of 20 pages requires many choices. Researchers are humans, and thus they (we) have priors – due to preferences and interests – that inevitably influence the choices. When a meta-analyst can identify that a fraction of the researchers of a certain β-literature has a particular prior, a binary dummy can be coded. Such prior variables normally become significant with the expected sign. But the common goal of all researchers of writing clear papers with the "right" signs (as discussed in a moment) may be enough to create bias.

My own rule of thumb, when I read a paper in a field I do not know, is to expect an exaggeration by a factor two, in accordance with the typical result of meta-studies; see Ioannidis *et al*. (2017), and Doucouliagos *et al.* (2018).

The following three sections discuss the three methods from Table 1, stressing both what the method can do and the main problems of the method.

4. Methods and their problems (1): Theory

Table 1 shows that theory papers that used to dominate in economics are a rapidly falling share. The profession suffers from theory fatigue. In addition, empirical techniques have seen a large technical progress in computers and the software used by economists. The progress has caused large cost reductions in empirical work. That is, relative prices have changed dramatically to the disadvantage of theory.

A theoretical paper starts by the definition of a set of formal assumptions, including relations about behavior. Then math is used to allow the researcher to derive the consequences of the behavior assumed. The assumptions are typically highly stylized such as: let us assume the economy produces one good, and has one agent who lives two periods. Everybody knows that such assumptions are a huge simplification, but they concentrate attention on the mechanism studied, and make the model tractable. Simulations showing that the model can produce something that looks real are often added to theoretical models. This is not classified as empirics in the survey used for Table 1.

Some theory papers start from a well-known puzzle. By a clever twist, the tools of

economics are brought bear on the problem, and the puzzle is solved. Such papers are rare, and highly cited. After a while, we all know the twist, and the paper has changed the way we think, which is the purpose of theory!

Most theory papers are much less exiting, as they are variants. They start from a well-known model and show that the solution is different if one of the assumptions are changed, e.g., instead of one good, the economy has two, or the same solution can be reached in a more general way if more advanced math is used, e.g., to handle n goods and k periods. This is useful, as it helps us understand what a certain model can do and not do. Such papers are difficult to summarize, and the marginal insight reached soon dwindles. Thus, they become dull to read and contribute to theory fatigue.

Theoretical papers suffer from the problem of *t-hacking*, where the able theoretician chooses assumptions that – within a range – produce the results desired. If the range is small, the theoretical result is robust, but often it is not so small.

Theory only provides qualitative information about β – typically only the expected sign. This gives rise to the concept of the "right" sign, which causes researchers to censure results with the "wrong" sign. Thus, the results come to confirm the theory, and the mean is exaggerated. This may be revealed by meta-analysis that shows the truncation of the distribution.

The policy using β needs knowledge about its size. This requires empirical estimation. One way to obtain knowledge is by experiments.

4. Methods and their problems (2): Experiments

While the fraction of theory has dropped, experimental economics has increased from 2% to 11% in the last 20 years; see Table 1. Experiments in economics are of three types, where only one is important in practice.

Real life experiments are done by giving a certain treatment to some people and compare the treated with a similar group of non-treated. To make the groups similar, they are often selected by a lottery, and the non-treated are given a placebo treatment that looks like the real thing. As in medicine, the non-treated are likely to experience placebo effects. Thus, the *gold standard* is the double-blind experiment where neither the ones giving the treatment nor the ones treated know who are treated and who are given the placebo treatment. This is easy to do with a pill, where a perfect look alike can be made, but the double-blind design is quite difficult to practice in economic experiments. Such experiments are rare, as they are expensive,

require a major team, and take a great deal of time, see Christensen and Miguel (2018). There were precisely zero in the sample of papers covered by Table 1.

However, there were a couple of ***natural experiments*** where an important shift happened unexpectedly, so people's behavior before and after could be compared.

Experiments nearly always mean ***lab experiments***. Almost 10% of the papers published dealt with such experiments. It is a method where manuals and special software are available, and many labs have been established. A lab consists of some cells that contain a controlled computer. It only contains the software for the experiment, and it can only communicate with the central computer of the experiment. The cost of a new experiment is a change of the program and a modest payment to the players, who are normally students at the university having the lab. The key idea is to let people play a game that can be solved analytically, in the highly controlled environment. When, e.g., 100 people play the same game, one can study the outcome. It often deviates non-randomly from the analytical solution. The advantage of lab-experiments is that it allows the researcher to study behavior that is difficult to study in the real world, where it is hard to create a controlled environment. However, it is an artificial situation, which may not generalize to real life.

It is likely that the experiment is affected by ***expressive behavior***. When people play for moderate payments, and they know that what they do is carefully recorded, they may act as they would like to do more than what they would do in a real situation. Thus, we expect the experimental studies show that people are too nice.

Experimental papers may have publication bias just like other empirical papers. The prior of the researcher is to find that people behave differently from the prediction of economic theory, so that something interesting is found. Thus, we expect that published experiments show that people differ too much from the predictions of economic theory. Recently, a large-scale replication study of experimental papers in the social sciences found a publication bias exaggerating results by a factor 2; see Camerer *et al.* (2016) for the part of the project dealing with economics.

While experimental economics is a semi-empirical technique, the big chunk of empirics attempts to infer people's behavior from the huge amount of data produced by the economy. The data are collected and distributed by statistical agencies, with the explicit purpose of helping all of us make decisions.

5. Methods and their problems (3): Statistical inference

As mentioned, economic research is increasingly empirical and uses the classical method of Table 2. In practice, the key method is to run regressions on existing data, where the data-generating process is outside the control of the researcher. Though, of course, the data have to be selected. I know of cases where researchers searched high and low to find a dataset showing what they wanted to find. The toolbox of economists is full of regression tools. When all is done, the paper is presented as if the classical method is used.

As an aside, it would be interesting to know how many regressions the 100,000 academic economists run every day. It could easily run into 25,000, so it is truly staggering to think of how many regressions the profession runs every year.

It is well known – at least since Leamer (1983) – that the classical method is ***flexible*** and contains a great deal of make-believe. Many researchers see the flexibility as a main virtue of the method, as it allows them to write fine clear papers. But the method generates the problem of publication bias.

The key problem is that it is normally possible to make many variants of the model, especially by adding ad hoc control variables. Computers are fast, and the econometric packages used by researchers are very user-friendly. Once the data is in the computer, it is easy to run hundreds of regressions. Then the rational researcher will surely choose the best results, and in a good many cases this makes it possible to reach clear results, and by some research loops it can be written up to look very convincing!

Results presented as if the researcher has followed the classical method may be influenced by both loops and mining, making the tests reported much weaker than they look. It also predicts that many β-literatures are biased, especially if they deal with effects where many researchers have similar priors.

This happens when the effect deals with something good we all want and when a big institution, notably a public spending program, exists in the field trying to do the good. Such programs typically want results that show that the programs work. Thus, results that are too big are better than result that are too small. Big institutions typically finance research in its field and have ways to reward loyal researchers.

The post-classical methods are often cumbersome and difficult to report, and they tend to reach unclear results. A handful of tables with co-integration tests, or five pages showing the paths generated from VARs, are valuable for future research, but not very exciting to read.

Especially when they provide unclear end weak results. In addition, they are difficult to interpret. Thus, the post-classical methods often produce weak and complex results, while the classical method can be made to produce clear results. The trends reported show that the profession tends to prefer results that are too good rather than results that are too dull, irrespective of how true they are.

6. Conclusion

This survey started by noting that communication in economics is by journal articles of about 20 pages. In addition, it noted that few problems in economics can be fully analyzed in one such paper. Consequently, most important problems are analyzed by many papers.

Thus, we are faced with a swarm of partial results, where researchers have tried to catch the effect (β) in many ways. ***This is surely great***. We want problems to be analyzed in many ways, but it also means that it is dangerous to act on the results in one paper. The truth about β is likely to be hidden within the swarm. It also shows how much serious papers differ. This is often an eye-opener. Estimates in fine papers published in perfectly decent journals often deviate substantially, such as by a factor 2 or 3.

It is a major project to make a full meta-study. Consequently, people use different short cuts. Some just look at the paper in the journal with the highest rank, as they argue that this is the estimate with the highest quality. Meta-studies have often analyzed the relation between the estimated β's and the rank of the journal. The result typically finds no connection. The paper in the top journal is often an early paper that pioneers new methods, and they tend to vary more; see Doucouliagos et al (2018). Paper number 50 in a middle-ranking journal may be more reliable.

Another method is to look at a handful of studies, calculate the average, and adjust for the typical publication bias, which is likely to be by a factor two. This is surely a crude method, but it is better than to use the average result.

At the end, I should say that the above is not meant as a critique of economics, but as a warning against illusions. Many decisions are economic in nature, so economic analysis is highly needed. But all users of such analysis should note that one paper is one aspect only, and that we economists are human and have priors and interests as everybody else.

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

Brodeur, A., Lé, M., Sangnier, M., Zylberberg, Y., 2016. Star Wars: The empirics strike back. *American Economic Journal: Applied Economics* 8(1), 1-32

Camerer, C. F., Dreber, A., Holzmaster, F., Ho, T.-H., Huber, J., Johannesson, M., … Wu, H., 2016. Evaluating replicability of laboratory experiments in economics. *SCIENCE* 351, Issue 6280, 1433-36

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

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

Ioannidis, J.P.A., Stanley, T.D., Doucouliagos, H., 2017. The power of bias in economics research. *Economic Journal* 127, 236-65

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

Mueller-Langer, F., Fecher, B., Harhoff, D., Wagner, G.G., 2019. Replication studies in economics - How many and which papers are chosen for replication, and why? *Research Policy* 48(1), 62-83

Paldam, M., 2018. A model of the representative economist, as researcher and policy advisor. *European Journal of Political Economy* 54, 1-15

Paldam, M., 2021. Methods used in economic research: An empirical study of trends and levels. *Economics* 15, 28-42

Paldam, M., 2023. Meta-mining: The political economy of meta-analysis. *Kyklos* 76, 125-140

Stigum, B.P., 2015. *Toward a Formal Science of Economics: The Axiomatic Method in Economics and Econometrics.* (4 versions) MIT press, Cambridge MA