EJW
ECON JOURNAL WATCH
Scholarly Comments on Academic Economics

Econ Journal Watch 10(2)
May 2013: 136-156

Regression Costs Fall, Mining Ratios Rise, Publication Bias Looms, and Techniques Get Fancier: Reflections on Some Trends in Empirical Macroeconomics

Martin Paldam[1]

Link to Abstract

The following deals with an 'iceberg' property of empirical economic research: Only a small fraction of the statistical estimates done are actually published, so most of the berg is hidden below the water. That part is the dangerous part. This paper suggests that it is growing, and that this is a problem.

The problem can be analyzed by considering the body of research that attempts to determine the size of some important parameter β. The research output has two dimensions: The visible dimension of the reported results and the invisible dimension of unreported results. The total number of regressions run is given by the product NJ, where $J - 1$ is the average number of unreported estimates behind each of the N reported estimates. The paper speculates about the invisible dimension using economic theory and some general results from the recent wave of meta-studies, which analyze the visible dimension.

The key observation that has prompted this paper is that the cost of running a regression has fallen rapidly: Data is becoming increasingly available, and our computers and econometric packages are becoming faster and more user-friendly. The researcher, therefore, will run more regressions and get more estimates of β. For this simple reason it may be suspected that the number of regressions made per regression published—J, which can be called the *mining ratio*—is rising. I will argue

1. Aarhus University, 8210 Aarhus V, Denmark. I am grateful for comments from Chris Doucouliagos, Erich Gundlach, and Sarah Necker, and for all comments when this paper was presented at the 2012 MAER-Net Colloquium, September, Perth, Western Australia and at the European Public Choice meeting in Zürich in April 2013. Also, I want to thank the referees for their feedback.

that this is likely to lead to a rise in publication bias, which is the excess size and significance of the published estimates, as compared to the true values.

A looming increase in publication bias should be counteracted.[2] Indeed we see another trend today, namely a rapid rise in the number and sophistication of econometric techniques that do increase regression costs. Meta-studies typically show that little of the variation between empirical results is due to variation in econometric techniques, which suggests that new techniques have a small marginal productivity. It is tempting to see the trend toward econometric sophistication as part of a response to the falling cost of regression, but there are no doubt other reasons to demand that front-line econometrics is used.

This paper discusses problems that are present in all empirical economic research. In micro the dataset is normally an important part of the experiment. The sample is often large, independent, and expensive, and thus replications are independent. Below we concentrate on macro investigations that use smaller publicly available datasets. An experiment is basically one regression, so meta-analysis is done by MRAs, meta-regression analyses. Since the data often overlap between studies, independent replication is difficult.

It is common in such literatures that NJ, the number of regressions done to determine the size of a certain β, exceeds the number of available data observations, which may be 50 annual observations, one for each of only 50 countries.[3] That is, the 'mining collective' of researchers analyzing β has, in a manner of speaking, exhausted the degrees of freedom many times.[4]

Data mining has long been discussed by applied economists; see, e.g., the essays in Hylleberg and Paldam (1991). D. F. Hendry (2001) has recommended that data be broken into two, where you mine the first to your heart's content and use the second to validate the best model found; the problem is that you may mine the second part too. E. E. Leamer (1983 and later) has a series of papers on the "con" element in econometrics caused by data mining, arguing for (rather cumbersome) Bayesian methods. In the field of growth regressions a large discussion has taken place, which is summarized in chapter 12 of Barro and Sala-i-Martin (2004).

This paper has some relation to the discussion started by John Ioannidis (2005) dealing with medical research, where meta-analysis originated. That discus-

2. Hopefully journal editors and other gatekeepers can do something, but perhaps they are part of the problem.

3. The reader may think that β is the slope of the IS curve, or that of the LM curve, Phillips curve, consumption function, etc. Or maybe β is the effects of education, democracy, or development aid on economic growth.

4. The smallest of the literatures mentioned in note 3 is the one on aid effectiveness. It uses cross-country data averaged over four to ten years, where about 1,000 observations are available. Here about 1,500 regressions have been published, and if $J \approx 25$, it means that the degrees of freedom have been used more than 30 times.

sion deals with problems that are fairly close to those in microeconomics. Even with independent data samples compiled to estimate a certain effect, one can still vary the explanatory models substantially.

In economics the seminal paper was T. D. Stanley and Stephen Jarrell (1989), along with the collection of papers in Roberts and Stanley (2005). However, the state-of-the-art technique was only developed in Stanley (2008), which greatly helped in pushing the ongoing wave of meta-studies. Chris Doucouliagos and I have discussed some of the problems considered here, but only based on the development aid effectiveness literature (Doucouliagos and Paldam 2009). The present paper is an attempt to show that economic principles provide some insights into the mechanisms underlying the problems of data mining and publication bias, and it speaks to how much these problems can be reduced by fancier econometrics.

The theory of the rising mining ratio

The analysis below considers a literature aimed at estimating some parameter β by regression experiments. The studies of β have two dimensions: A visible dimension, comprising the N-set of regressions that are reported in the literature; and an almost entirely invisible dimension that comprises the J-set of regressions run as part of the research processes behind each of the N published results. For ease of presentation I make two assumptions: (i) Each study reports one estimate of β, reached by a regression. (ii) At any point in time J_t is the same for all researchers, and J_t moves smoothly over time. The mean J over time in the β-literature is $\underline{J}$. The full body of regressions to determine β is $N\underline{J}$.

In the 150 papers of the aid effectiveness literature, the number of published regressions per paper has gradually increased from about five to about ten during the last 40 years (see Doucouliagos and Paldam 2011). In this regard the aid effectiveness literature appears to be typical. It is important that the individual author has little influence on that number, which is largely determined by editors and referees. Given that the number is (almost) exogenous, it might as well be assumed that it is one, and so assumption (i) is that each paper in the β-literature reports one estimate. This will save one index on the variables below.

The mining ratio J differs from one paper to the next. I assume that if, by a miracle, all J_{it} were revealed for one β-literature, and plotted over t, a well determined moving average (kernel) would appear in these data. The analysis deals with papers having J's equal to that moving average; in this sense we discuss the J of the typical paper. Assumption (ii) is that J is the same for all papers written at time t and that J_t has a smooth trend over time. This saves another index on the variables below.

The rest of this section of the paper deals with the *J*-dimension of the *β*-literature, that is, with the *J* regressions run as part of the research process behind each published result. The profession agrees that *J* is private information for the individual researcher. We cannot peer into the private space of the individual researcher, so *J* is unknown. I have applied introspection and informally polled friends and colleagues, and my guess is that, in 2013, *J* is well above 25. It should be added that there are search strategies that make it difficult to define what counts as a regression privately undertaken. A simple example is that econometric packages typically provide estimates with a handful of diagnostic tests pointing to the nature of variables that can increase the fit. So perhaps the true increase in *J* is larger than a simple count will reveal, since packages are doing some of the experimentation for us.

Thus, *J* is the unobserved private choice of each researcher. Economic intuition suggests that he will go on regressing as long as he feels that $MC(J) < MB(J)$, where $MC(J)$ is the marginal costs and $MB(J)$ is the marginal benefits. The *J* chosen is hence the solution to:

$$MB_t(J_t) = MC_t(J_t), \text{ where } t \text{ is an index for time that is used to discuss } J = J_t \quad (1)$$

This equation is meant to be specific to the individual, even to the individual investigation, but I economize on notation. The equation expresses the equi-marginal principle so common in economic analysis. As far as I know it has not been clearly stated in any analysis of *J*.

The total costs of macroeconomic research are high: Macroeconomic data are expensive to compile, econometric packages are costly to write, and human capital of researchers in the field is expensive to produce and acquire. The individual paper carries only a tiny fraction of these costs. Most macro data are free to use. At the operative margin, the human capital of researchers is a sunk cost. Econometric packages are installed on the computers of researchers, who know how to use them!

The costs for the author of the empirical part of a paper arise from the opportunity costs of the following steps in the work: (i) The data set has to be collected from the relevant sources, (ii) it should be organized in the computer in the form needed by the econometric package; then (iii) the necessary commands have to be given to the computer to run the regression; finally, (iv) the results should be studied. The time necessary to implement each of the four steps keeps falling.

When I started in research more than 40 years ago, data had to be copied from tables in books found in libraries. Computers were mainframes, and the data had to be punched onto cards that were entered into the computer by a special staff. Once read into the computer the cards could be stored on magnetic tape. You had access to the mainframe a couple of times a day only, and then your tape had to be mounted. The early statistical programs were clumsy, poorly documented, and not very user-friendly. But they were a huge advance from the pre-computer technology. Lots of orders were necessary for running a regression, and if you got something just a tiny bit wrong, as frequently happened, half a day was wasted. Results were provided on big sheets of paper with green lines.

Today, data libraries are virtual in the form of webpages hosted on computers around the world. Most of the relevant pages post data in a format that is user-friendly, so that data can be downloaded straight into a program, such as Excel, that all researchers have on their own personal computers. To combine data from different sources and get them into the form necessary for the econometric package still needs some work, but the work typically only takes a couple of days. Personal computers now have about the same brainpower as the typical mainframe 40 years ago. The companies producing econometric programs are impelled by the competition to make large efforts to make the packages user-friendly.

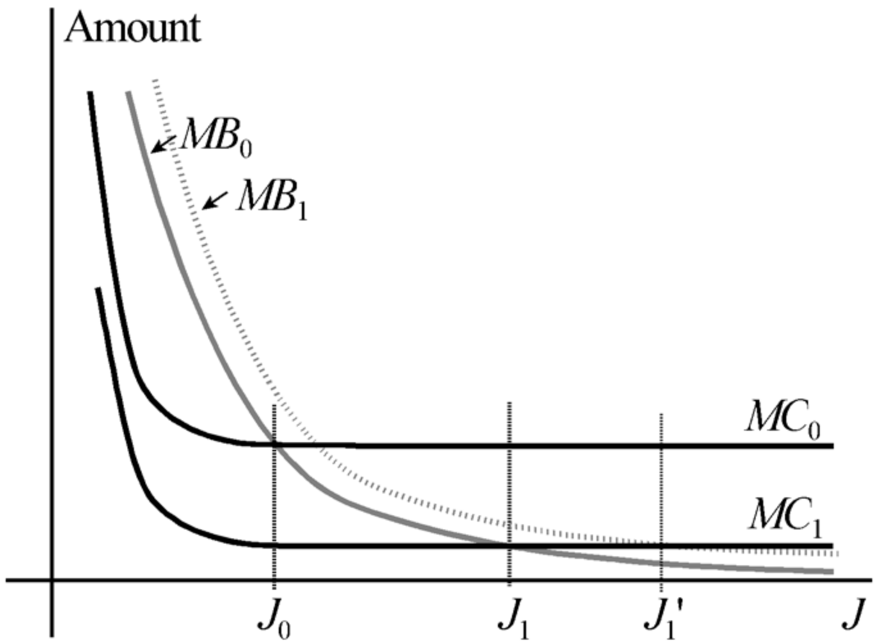
Therefore, it is a strong fact that the *MC* curve is shifting downward over time. In the 40 years the cost of the marginal regression has probably dropped between 100 and 200 times—say 150 times. A fall by 150 times over 40 years is 3.5 times per decade or 13.5% per year. The drop is probably approximately log-linear. Technological jumps have occurred, but regression costs do not fall neatly in steps. Think of the change from mainframes to PCs: productivity gains were small in the beginning because PCs were slow and few programs were available, but the productivity gains have kept growing.

The form of the *MC* curve

The first regression still needs some days of work, and maybe the next few regressions also need new data to be entered into the program, but then it will quickly fall to almost nothing—just a few keystrokes. Also, the output may just be a screen to be looked at for a moment, and then you rush to the next screen till something looks good. Thus, the *MC* curve flattens out. This is drawn in Figure 1.

For $t = 0$ the *MC* curve is MC_0. For $t = 1$, one decade later, the curve is MC_1. The two curves are drawn to reflect the crude orders of magnitude mentioned. That is, the shift in the *MC* curve is almost three-quarters of the way down toward the horizontal axis. It was argued that the cost per regression was falling along a log-linear path. As the picture is the same (after a multiplicative transformation of the axes) next decade, this also implies that the *J* increases log-linearly.

Figure 1. The mining ratio (of the average researcher): A function of benefits and costs



Note: The initial situation (t = 0) and the end situation (t = 1), are indicated with a subscript. They differ by one decade. The curves are drawn as explained in the text.

The reader may wonder, as I do, how long the downward shifts in the $MC(J)$ curve can continue. The next step, already underway, will make it so data available on the Internet are downloadable directly into the main econometric packages. It appears that the Stata package is winning market dominance, and some of the main data providers, including the World Values Survey and the World Development Indicators (World Bank), are already providing data in Stata format. Stata's data handling ability increases, and, of course, computers keep improving. There is space for a continuation of the downward shift in the $MC(J)$ curve.

The form of the *MB* curve

Seen from the point of view of the individual author, the marginal benefit of one extra regression depends (though not entirely, surely) on the increase in his chance of publication and his expected benefits from publishing. The main factor determining the shifts of the $MB_t(J_t)$ curve over time is probably the increasing use of publication information (about citations and impact factors of journals) on the academic job markets. The increasing use of such information in the job market leads me to suggest that the *MB* curve shifts a bit upward over time. The MB_1 curve in Figure 1 shows the effect of an upward shift in the *MB* curve—it further increases the equilibrium value for J from J_1 to J_1'. The MB_1 curve is dotted to indicate that it is difficult to assess the size of the shift. But how much *MB* rises is secondary; the story is mainly about the fall in *MC*.

The solution to equation (1) is the intersection of the two curves. As drawn, the intersection J_0 moves to J_1, or probably to J_1'. Already, J_1 is twice as large as J_0, and J_1' is even further from J_0. I have assumed that the two situations are one decade apart. Hence, the graph suggests that J increases 3.5 times each decade. Obviously this estimate is very rough. But I think that it is hard to draw the two curves in a way that is both reasonable and gives only a small rise. I proceed on the assumption that the mining ratio is rapidly rising and that the rise is log-linear.

Findings from meta-analysis

This section switches to the visible dimension of the β-literature, i.e., it turns to the N-set of reported estimates, b_i, of β. They have standard errors s_i, t-ratios $t_i = b_i/s_i$, and precisions $p_i = 1/s_i$.

Meta-analysis was developed in medicine where it is a standard technique. More than 100,000 meta-studies have been made in medicine. A handful of text-books have appeared—the main one is Hunter and Schmidt (2004). A medical experiment is costly, and hence it is reported. Also, it typically uses data that are unique to the experiment, but meta-studies still point to important problems (see Ioannidis 2005).

In microeconomics an experiment often involves independent data collection and a set of calculations, which might be a regression, but also they may not be, as is often the case in medicine. It is possible to vary the model used, but the data are to provide an answer to a question. Hence, the model variation is not so large. Still a meta-analysis of the replicability of the result is necessary. The situation in microeconomics is fairly close to the situation in medicine.

In macroeconomics the data are often available, so an experiment is a regression. And the meta-analysis is done by regressions on regression coefficients (MRA). As discussed a regression is so cheap that many of the ones made remain unreported.[5] Macro data are limited and many studies are therefore done on over-

5. In the analysis of macro models all relations need to include controls so that the ceteris paribus assumption holds. If, e.g., the researcher considers 25 potential controls and a tidy model contains fewer than six controls, then there are about 70,000 possible model variants to try.

lapping data, so the meta-technique has had to be adjusted to be equally useful. The adjustment took some time, as mentioned. The first textbook on meta-analysis in economics is Stanley and Doucouliagos (2012), and recently a set of the guidelines has been published; see Stanley et al. (2013). Before 2005 only a handful of meta-studies had been done in economics, but since 2008 the number has increased to about 500. As of 2012, only one study (Doucouliagos and Stanley 2012) has looked across many meta-studies. The meta-studies have coded approximately 20,000 primary papers. About half of the meta-studies are in macroeconomics.

At the basic level, meta-analysis is very robust, so that if two teams make an independent meta-study of the β-literature, they reach virtually the same result.[6] It consists of four steps: (1) The literature is collected; (2) it is coded; (3) the distribution of the estimates is analyzed by the funnel graph (see Figure 2); and (4) the FAT-PET is estimated, where FAT stands for Funnel Asymmetry Test and PET stands for Precision Estimate Test.

If the estimates are weighted with their precisions, this will move the mean towards β. This is the idea behind the FAT-PET (from Stanley 2008). It is the regression:

$b_i = \beta_M + \beta_F/p_i + u_i$, where β_M is the PET, β_F is the FAT, as explained in the next paragraph, and u_i is noise. It is a hyperbola converging to β_M for $p \rightarrow \infty$, as drawn in Figure 2b. (2)

The FAT is a Funnel Asymmetry Test, so that if $\beta_F \neq 0$, by a standard t-test the funnel is asymmetric. The literature referred to shows that it is a rather powerful test. In Figure 2a, β_F is zero. The PET is the Precision Estimate Test, β_M. It is the PET meta-estimate of β. A standard t-test tells if it differs from zero. In Figure 2b the β_F/p_i term converges to zero and $\beta_M = \beta$. The literature shows that in the face of publication bias, β_M is a fine estimate of β. Hence, the publication bias is estimated as:

$PB = (\underline{b} - \beta) \approx (\underline{b} - \beta_M)$. Note that the publication bias is $\underline{b}$ if $\beta = 0$. (3)

Figure 2 may help the reader to build intuition about the meta-analysis, and Table 1 is meant to provide some help to keep track of the notation. In Figure 2, each funnel presents the distribution of the N estimates b_i as a scatter over their precision p_i (on the vertical axis). In the case analyzed, economic theory predicts that $\beta > 0$, so estimates where $b < 0$ have the 'wrong' sign.

6. Both at step 1 and especially at step 2 it is difficult to fully escape mistakes, but experience shows that a small number of stochastic errors matter very little; see Doucouliagos and Paldam (2008), Mekasha and Tarp (2013), and Doucouliagos and Paldam (2013).

Figure 2a is the ideal funnel where all estimates are published. The least precise results scatter most, and as precision increases the results move closer. This gives a lean, symmetrical funnel, so the FAT-PET is vertical. The FAT $\beta_F \approx 0$, and the PET meta-average $\beta_M \approx \underline{b} \approx \beta$.

Figure 2. Funnel diagrams and the FAT-PET estimate

Figure 2a. The ideal funnel   Figure 2b. A more realistic case

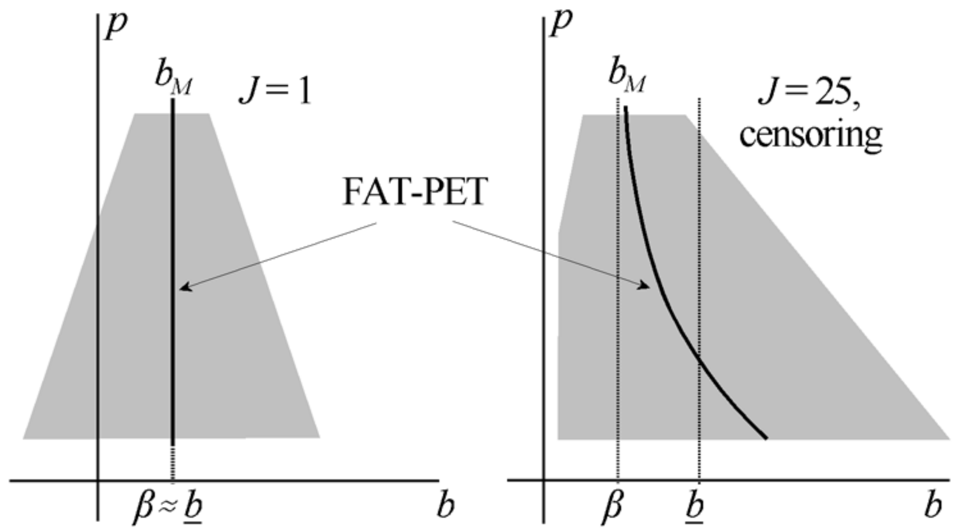


TABLE 1. Summary of meta-analysis terms

Term	Definition considering the β-literature	Explanation
Funnel	(b_i, p_i) scatter plot, where $i = 1, \ldots, N$	Distribution of the N-set of estimates
Funnel width	The coefficient of variation of the N-set	Standard deviation divided by mean
Ideal funnel	The model is true and all estimates are published	Funnel is lean and symmetric
Empirical funnel	The reported estimates in the β-literature	Funnel is wide and often asymmetric
Arithmetic mean	Disregard funnel asymmetry, $\underline{b}$, may be weighted	Best average for ideal funnel, where $\underline{b} \approx \beta_M$
PET meta-average	Adjusts mean for funnel asymmetry, β_M	The best average if publication bias
FAT-PET MRA	$b_i = \beta_M + \beta_F/p_i + u_i$	MRA, Meta Regression Analysis
	FAT, $\beta_F \neq 0$, Funnel Asymmetry Test	FAT-PET is joint estimate of FAT and PET
	PET, $\beta_M \neq 0$, Precision Estimate Test	Assumes the asymmetry is publication bias
Publication bias	$PB = \underline{b} - \beta \approx \underline{b} - \beta_M$, when $\beta \approx \beta_M$	The publication bias is the mean of the selection bias in the individual studies

Figure 2b is a more realistic funnel where each researcher runs 25 regressions and selects one for publication. Here the choice set is wider. The situation permits the researcher to select estimates with the correct sign and high significance. The figure is a sketch of the effect of such selections. Thus, the funnel will be wider, especially at the base. As it is censored at zero it looks asymmetric. Hence, the FAT-PET is hyperbolic, so the FAT $\beta_F < 0$ and $\underline{b} > \beta$, but the PET meta-average is still $\beta_M \approx \beta$.

The excess width and frequent asymmetry of the *N*-set

The width of a funnel is measured by the coefficient of variation of the *N*-set. Most estimates in the typical *N*-set have t-ratios above 2, so precision is high. Thus, one should expect funnels to be lean. Ideal funnels that occur in simulations, where $J = 1$, are indeed lean—they are also symmetrical. Ideal funnels have another property that will be used below: The distribution of the *N*-set is normal. We can only be sure that simulated funnels are ideal, and the normality property holds rather well for simulated funnels. In simulation experiments the variation is generated by the data, and most of the variance of empirical funnels is probably from model variation. But by the central limit theorem, the normality assumption still holds for variation generated from many sources. Below it is assumed that ideal funnels are normally distributed.

When empirical funnels are compared with the lean and symmetric ideal funnels, two major findings emerge: empirical funnels always have *excess width* and they are often *asymmetric*. Doucouliagos and Stanley (2012) compare a whole set of meta-studies showing these properties. Especially when the size of β is debated, the width becomes large. In the case where $\beta = 1$, the typical width is between 2 and 5, even when the typical t-ratio reported is above 2.

This paper requires that we guess at the coefficient of variation of the typical *J*-set. It is easy to give reasons to expect that the *N*-set varies more. The *N*-set is calculated from many data sets, while the typical *J*-set is estimated from one data set. The *N*-set is often made by researchers from different schools who want to find different results, while the *J*-set is made by one researcher. But good reasons also exist to expect that the *J*-set varies more. For instance, many researchers select the reported estimates (the b_p's) after some averaging. Second, the profession often decides that a certain sign on β is the right one, as discussed in the next section. The likely result is some censoring, reducing the variation of the *N*-set. All said, it appears reasonable to expect that the two variances are roughly the same size.

The asymmetry indicates a problem that calls for an explanation. In principle, an asymmetry may have many explanations.[7] From the way the asymmetry looks

7. Callot and Paldam (2011) study how funnels can become asymmetric. It appears that the funnels are robust to most standard econometric problems such as misspecification of the function, non-normal residuals, etc. To generate a significant asymmetry such problems have to be so large that it is unlikely that they are undetected. Also, funnels do become asymmetric due to randomly omitted control variables, but this asymmetry rarely looks like censoring.

one can often explain what is going on. By far the most common explanation is publication bias that looks as in Figure 2b.

From asymmetry to publication bias

Publication bias follows from the iceberg property of empirical research: Only $1/J$ of the results are reported, while the remaining $(J-1)/J$ estimates remain unpublished and invisible. A simple question is: Is the visible part representative of the whole berg? The answer is obvious: The incentives facing the researcher will surely make him select results that are *better* than the average: That is, they are polished so that they have relatively high t-ratios, and they might be censored to conform to the priors of the author.[8] The problem is less if many separate authors have different priors.

The profession, however, often develops a dominating body of thinking, or beliefs, taken to be *the* theory in the field. Such thinking produces a main prior which typically suggests a 'correct' sign on β. Authors, colleagues, editors and other gatekeepers may all have that prior. Estimates with the 'wrong' sign tell everybody that the research is suspect! Funnels often have appearance of having been censored accordingly. If the theory is right (as it surely is!), estimates will converge to the true β and thus have the correct sign as data samples increase. But results from small data samples should have a large variation and contain a good deal of estimates with the wrong sign. It is common to find that such estimates have suspiciously few results with the wrong sign (as in Figure 2b); see Stanley and Doucouliagos (2012, chapter 4). The asymmetry is often visible to the naked eye, and the FAT is a powerful test of asymmetry.

When the funnel has this asymmetry, the mean $\underline{b}$ is biased by the missing values, and the bias is in the direction of the prior, so that $\underline{b}$ is an exaggerated estimate. The PET estimate of the meta-average, β_M, corrects the mean for the publication bias. The publication bias found by formula (3) has been estimated in many meta-studies. It does have a large variance across the bodies of economic literature analyzed (see Doucouliagos and Stanley 2012), but a crude average seems to be about 2. Consequently, publication bias is not a negligible problem. The analysis offered above argues that the publication bias was smaller in the past and will be larger in the future because of the steady rise in J, unless something happens to counteract the trend.

8. Two additional bias-generators are: (i) In sponsored research, sponsor interests produce priors. (ii) An author with prior results in a field has a confirmation prior, so that relation (2) has to be estimated using clustered standard errors, with author clusters. Clustered standard errors are also used when more results are reported in the same paper.

Once a body of literature is coded, the meta-analyst can ask various questions by adding an extra variable in equation (2). Some examples are: Q1: Are results published in better journals different? Here the impact factor of each journal is added. Q2: Are the results in the A-journal different? Here an A-journal dummy is added. Q3: Are results in papers by female researchers different? Here a gender dummy is added. Such questions are common in meta-studies, and the answers for different β-literatures are not always the same. However, the answers to the three questions mentioned seem to generalize. They are: A1: No; A2: Frequently; and A3: No.

One important question to ask is how independent the data are across the studies, so it is important that variables for the data samples used are coded.

The small effect of estimators

One question that has been frequently asked deals with the effect of estimators. The author has seen most of the approximately 250 meta-studies in macroeconomics (and is co-author of 10) and has yet to see one that has found a substantial effect of any particular estimator.

A literature often unfolds as follows: Old studies in the literature use the old M_0-estimator, and many of the new studies add a new M_1-estimator, or bring only M_1-estimates. When the M_1-estimator is introduced into the literature, the initial paper shows that the results are indeed different, and this is announced as a methodological breakthrough. If the new estimator is presented as a minor honing of the coefficient, it is unlikely to be easy to publish, but journals do like major methodological breakthroughs. Many meta-analysts have learned to view such sales talk with a dose of skepticism.

Let us imagine a typical example: The β-literature has 500 estimates of β. The range of estimates is from −0.55 to +0.95. The two main averages are $\underline{b} = 0.40$ and $b_M = 0.20$, so the publication bias is about 2 as it is on Figure 2b. The breakthrough paper showed that while the old M_0-estimate is 0.30 the new M_1-estimate is 0.70, which is significantly larger. When coding the subsequent literature you soon find one paper where the M_0-estimate is 0.50, supplemented with an M_1-estimate of 0.45. In the next study the M_0-estimate is −0.25 and the M_1-estimate is −0.40. When the M_1-estimator has been used sufficiently often, the meta-analyst can ask if it matters by adding an M_1 dummy in regression (2). As far as I know the effect of estimators is normally small and often insignificant—it certainly is in the studies I have participated in[9]—and it *always* explains only a minor part of the variation.

9. Doucouliagos and Paldam (2011) study the effect of adjusting the aid effectiveness effect for simultaneity by this technique. Out of 1,000 estimates, 200 are adjusted. The test shows that they do not differ.

Personally, these findings from meta-analyses tally well with my experience in doing primary studies: New estimators are made to *hone* coefficients. It assumes that the researcher has a good estimate already and that it is important to chip away small to moderately sized biases. But the message from the amazing width of funnels found in the typical meta-analysis is that our knowledge is far from the honing stage in most research fields.

An assessment of the publication bias as a function of the mining ratio *J*

The perspective now switches from the N-set back to the J-set of estimates. They are b_j, where $j = 1, \ldots, J$, from which the researcher selects one b_p for publication. This makes b_p one of the b_i's. If it is assumed that the J-set is normally distributed, analytical results can be derived from any well defined selection rule. If the selection rule is simple, so is the result. If the size of the coefficient of variation, μ, for the J-set, was known, the orders of magnitude could be calculated. Below the guesstimate (from above) of $\mu \approx 2$ is used. When it is assumed that $\beta = 1$, it follows that $\mu = \sigma/\beta = \sigma \approx 2$, where σ is the standard deviation of the J-set.

The distribution of the *J*-set and the selection bias

The J-set is the set of regressions the researcher thinks contains the true estimate of β. Hence, before any of the estimates are selected the J-set is an ideal funnel, where the best estimate of β is $\beta \approx \underline{b}$. I have argued above that it is reasonable to assume that ideal funnels are normally distributed, and that it appears likely that the researcher selects a b_p that is better than $\underline{b}$. Hence, there is a selection bias:

$$SB = b_p - \beta \approx b_p - \underline{b}. \text{ If } \beta = 0, \text{ the selection bias is } b_p. \tag{4}$$

The selection bias is the 'micro' version, for the individual estimate, of the 'macro' publication bias of the whole literature. If everybody uses the same selection rule and J is constant, $PB = SB$. In practice PB is the mean of the SBs of the N individual studies.

For ease of presentation the J-set is sorted with b_1 as the smallest and b_J as the largest. From the normality assumption it follows that the J-set has the distribution:

$$b_j \approx \beta + \sigma\Phi((j - 0.5)/J), \text{ for } j = 1, \ldots, J, \text{ where } \Phi \text{ is the inverse of the standard normal cumulative distribution function.} \tag{5}$$

This shows that b_j differs from β with $\sigma\Phi((j-0.5)/J)$, which is the selection bias:

$$SB = \sigma\Phi((j-0.5)/J, \text{ if the researcher selects } b_j. \quad (6)$$

Imagine that the selected regression is chosen by a well defined *selection rule*. It is possible to think of many selection rules. I shall only discuss three rules that are tractable and known to occur occasionally. This means that some average of the three may approach the actual situation, as discussed at the end of this section.

Three selection rules

Selection Rule 1

Here the researcher wants to make a completely unbiased choice. As already explained this means that he uses the following rule:

$$b_p = \underline{b} = b_{J/2} \approx \beta \text{ so that (6) yields } SB = 0, \text{ and if all researchers use Rule 1 also } PB = 0. \quad (7)$$

With this rule the increase in J only means that the standard deviation of the N-set falls, and it does so by the square root of J. This makes the coefficient of variation of the N-set fall by the same. The N-set funnel is ideal, and as J goes up the funnel should become leaner rather fast. This does not seem to happen, so even if some authors use Rule 1, it cannot be the dominating rule in practice.

Selection Rule 2

The researcher wants to make a good but moderate choice, so he selects b_p at a fractile $F > 0.5$, such as $F = 0.75$. It is better than β, but it might still be defensible. This will happen if the researcher finds reasons to discard half of the regressions and then selects the regression in the middle of the remaining 'good' half. Then the other regressions in the good half can be used to demonstrate the robustness of the selected one.

$$b_p \approx \beta + \sigma\Phi(F) \quad (8)$$

$F = 0.75$ yields $b_p \approx \beta + \sigma\Phi(0.75) = \beta + \sigma 0.674$, and thus, (6) gives $SB = \sigma 0.674$. If $\sigma = 2$, the bias is 1.35, so that if $\beta = 1$, $b_p = 2.35$.

This selection bias is independent of J. If everybody uses this selection rule it also produces a constant publication bias that is independent of J. The falling cost of regression has no consequences for the publication bias. The increase in J only means that the precision increases. In the language of meta-analysis, the funnel is still symmetric—provided *everybody* uses the same rule. In passing it should be noted that in this 'sensible' case, publication bias occurs with a symmetric funnel, and the FAT-PET does not detect and correct the bias.

Selection Rule 3

The researcher selects the largest estimated coefficient as the best regression.

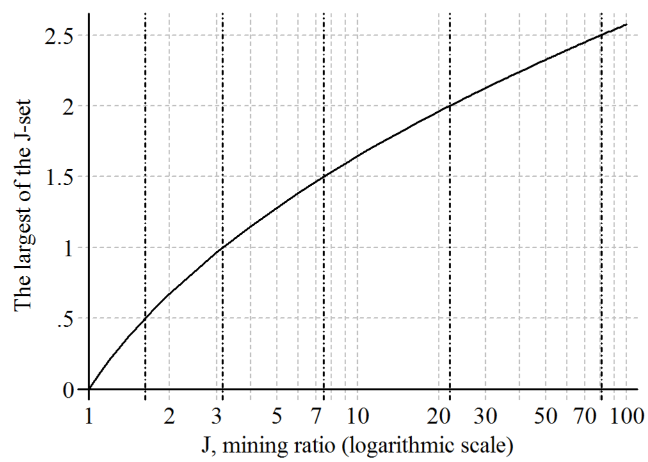
$$b_p = b_J \approx \beta + \sigma\Phi((J - 0.5)/J) = \beta + \sigma\Phi(1 - 0.5/J) \quad (9)$$

Figure 3 shows (9) for $\beta = 0$ and $\sigma = 1$. Here (6) becomes: $SB = \sigma\Phi(J - 0.5/J)$. For $\sigma = 1$ it is also the line drawn on Figure 3.

This expression depends upon J and as will be shown the dependence is quite strong. To select the largest b in the J-set is an extreme choice, but the qualitative results reported still stand as long as the selection depends on J.

Figure 3 gives the normalized path of b_J, for $\beta = 0$ and $\sigma = 1$. All other paths are reached by shifting the vertical axis to start in β and multiplying the curve by σ. To accommodate the conclusion from Figure 1 that the rise in J is log-linear, the J-axis on Figure 3 is logarithmic.

Figure 3. The path of b_J calculated from equation (9) for $\beta = 0$ and $\sigma = 1$



Note: The broken vertical lines are the intersection points for b_J = 0.5, 1, …, 2.5 reported in column (3) of Table 2.

The b_J curve rises monotonically, but we are only interested in the integer values for J, as reported in column (2) of Table 2. Column (1) gives 7 values for b_J. J

is an integer and column (2) gives the closest matching J for each of the values of b_J, while (3) gives the value to three or four significant digits. Our standard assumption is $\sigma = 2$ as shown in column (6) in Table 2, but it is also shown what happens when $\sigma = 1$ and 3.

TABLE 2. The intersection points for b_J

(1)	(2)	(3)	(4)	(5)	(6)	(7)
Calculations from eq. (9) for $\beta = 0$ and $\sigma = 1$				Selection bias from eq. (6)		
b_J	Closest J	Exact J	Ln exact J	$\sigma = 1$	$\sigma = 2$	$\sigma = 3$
0	1	1	0	0	0	0
0.5	2	1.62	0.48	0.5	1	1.5
1	3	3.15	1.15	1	2	3
1.5	7	7.48	2.01	1.5	3	4.5
2	22	21.98	3.09	2	4	6
2.5	81	80.52	4.39	2.5	5	7.5
3	370	370.4	5.91	3	6	9

The assessment that most researchers make at least 25 regressions for each one published means that the selection bias is on the order of 2 to 6. This appears high, as the typical publication bias is only around two, but Rule 3 is extreme.

How bad is the rise in J really?

The β-literature consists of N estimates. The N estimates are likely to use a range of research strategies, corresponding to selection rules such as the ones discussed. Imagine that the researchers chose among the three rules so that the publication bias, PB, is a weighted sum of the three biases:

$$PB = 0w_1 + 1.36w_2 + 4w_3 = 1.36w_2 + 4w_3, \text{ where the } w\text{'s are weights.} \quad (10)$$

If the weights are equal, $w_1 = w_2 = w_3 = 1/3$, the weighted sum (10) is $PB = (1.36 + 4)/3 = 1.8$, which is much like the average publication bias. This allows a crude estimate, for if J rises by 3.5 times in the next decade to about 80, this increases the bias to $(1.36 + 5)/3 = 2.1$. After another decade it is 2.5. These orders are very uncertain, but they do suggest that there is a growing problem.

Is fancier econometrics a device to reduce J?

I think the research community will agree that the rapid rise in J poses a problem. I have often heard senior members in the research community express misgivings about the ease with which the coefficients of the size desired can be generated by standard regression techniques, and about the inflated size of t-ratios. We all know that we are in a game that is too easy to play. One reaction is to demand graphical presentation of the data to show that the effect examined is visible to the naked eye; some demand robustness; others turn to theory and demand a rigorous derivation from first principles; and still others demand fancy estimators.

The core insight is that a new empirical result can be trusted only after enough independent replication. This is precisely where meta-studies come in. They become increasingly important—for with a high J it becomes easier to estimate something that looks like a replication on new data. In order to be credible, replications using new data should be made with *precisely* the same model, and be done by another researcher. In Paldam (2012) I discuss a case where a new and very nice theoretical model was empirically confirmed on a data set that proved to be an outlier. The first paper was so convincing that it managed to establish a widespread prior about the right signs.[10] The signs survived seven independent replications, by researchers who managed to find models that looked almost like the first model and generated the right signs, before it broke down after eight replications. The lesson is that, as J rises, the number of independent replications necessary for a result to be credible rises.

However, it is my impression is that the most widespread reaction to the problem is to increase the costs of regressions by demanding that techniques be fancier, so that the costs of regressions rise and J falls. Due to the rapid development of estimators, the old method M_0 soon changes to $M_1, M_2, \ldots, M_t$. Maybe the key reason to demand the latest M_t-estimator be used is not the small improvement it gives to the estimate, but the increase in cost caused by the use of the new estimator. The costs are high if the M_t-estimator is not (yet) included in the standard econometric package available to the researcher. He will then have to make an effort to master the new technique and to get it to run on his PC.

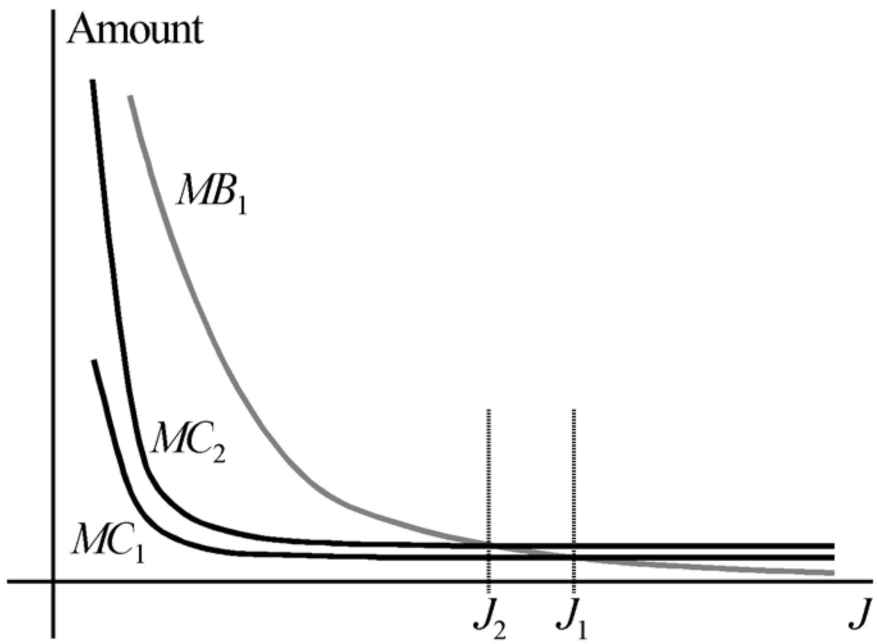
Figure 4 shows what happens when editors and referees demand that a new estimator is used. Here MC_1 is the costs of using the same estimator as before. It is the same as on Figure 1, though the vertical axis is compressed a bit. MC_2 is the costs of using the new estimator. The demand for the new estimates increases the initial costs of running regressions; but once you get the M_t-estimator to run on your PC, the marginal costs soon decrease almost to normal. So the MC_2 curve is above the MC_1 curve, but mostly in the beginning, i.e., J_2 is smaller than J_1, but not much smaller.

If the new estimator has a honing effect only, the pattern found by the two estimators is much the same, and the researcher will typically do the J experiments

10. It is also possible that the authors of the first paper were referees to the most of the following papers.

with the old estimator and then just re-estimate the selected estimates with the new one. Thus, the researcher may just make a data set ready and seek the assistance of an econometrician who knows the new estimator.

Figure 4. The effect of demanding that a new estimator is used



Note: MC_1 and MB_1 are as on Figure 1. The new estimator changes MC_1 to MC_2.

The market for econometric packages is a worldwide competitive market where the leadership edge can be lost quite quickly. Hence, if the companies producing these programs discover that there is a demand for the M_r-estimator, it is soon included in the package, or somebody writes an add-on program that you can use to run M_r-estimates in your favorite package. I have recently changed from Stata 10 to Stata 12. The inclusion of new estimators is quite impressive, and StataCorp (located in Texas) supports the community of Stata users by providing an exchange of add-on codes, which allows access to a world of estimators.

My own experiences are that when a new estimator is introduced in a literature it takes perhaps a year before the profession demands that it be used in the relevant papers.[11] And then it takes perhaps another half year before it is included in the standard econometric packages. There is a window of opportunity where the

11. It is not enough that a certain estimator is presented in a paper about econometric theory. However, once somebody starts to use it, it quickly becomes the new standard.

M_r-estimator can be used to reduce J, but it is not a large window. Once the M_r-estimator is included, the MC_2 curve falls again.

But there is an effect till the window closes. Thus the development of the M_r-estimator has served a useful purpose. The small (and falling) size of the window is somewhat countered by the increase in the production of new and still fancier estimators. It is necessary for the producers of econometric packages to be behind if publication bias is to be kept at bay.

It is arguable that the theory in the previous section—though it explains something—does not explain enough. It seems to me that the outpouring of new estimators is bigger than necessary to reduce J. An additional theory of why ever-fancier techniques are developed comes from Mike Felgenhauer and Elisabeth Schulte (2011). The key idea is that to increase the credibility of papers, authors have to put something at stake. Well known authors can use their reputation, but young authors have not accumulated enough reputation, so they must do something else: They show that they can jump through high econometric hoops. Perhaps the jumping through the high hoops is a signal of their ability, their value to the research community, and their dedication. Since it takes a lot of effort and time to learn these jumps, they have put all that work at stake.

Conclusions

This paper deals with the mining ratio J, which is the number of regressions made for each published. By necessity the mining ratio J reflects activities that are private information to the individual authors. But as the costs of regressions keep decreasing, the mining ratio must increase. I have suggested that the incentives on authors cause authors to select the best of their regressions for publication. Hence, there is a selection bias, and with a larger set of regressions to select from the selection bias will rise.

The average selection bias is the publication bias. Meta-analysis normally finds that funnels are much wider than expected from the t-ratios reported. Often they are also asymmetric in ways suggesting publication bias. It follows that publication bias is common in economics. The increase in the bias depends on the rules by which researchers select which regressions to publish.

Some of the excess variation is due to the introduction over time of new estimators. Yet another observation from meta-analysis is that changing estimators explain little of the variation in the typical economic literature. It appears that the efforts made to improve estimators seem to be excessive relative to the benefits, at least in macroeconomics.

All this poses something of a paradox and leads me to suggest that perhaps one reason to welcome ever fancier estimators is the effect they may have in slowing down the alarming rise in the mining ratio.

References

Barro, Robert J., and Xavier Sala-i-Martin. 2004. *Economic Growth*. 2nd ed. Cambridge, Mass.: MIT Press.

Callot, Laurent, and Martin Paldam. 2011. The Problem of Natural Funnel Asymmetries: A Simulation Analysis of Meta-Analysis in Macroeconomics. *Research Synthesis Methods* 2: 84-102.

Doucouliagos, Hristos, and Martin Paldam. 2008. Aid Effectiveness on Growth: A Meta Study. *European Journal of Political Economy* 24: 1-24.

Doucouliagos, Hristos, and Martin Paldam. 2009. The Aid Effectiveness Literature: The Sad Results of 40 Years of Research. *Journal of Economic Surveys* 23: 433-461.

Doucouliagos, Hristos, and Martin Paldam. 2011. The Ineffectiveness of Development Aid on Growth: An Update. *European Journal of Political Economy* 27: 399-404.

Doucouliagos, Hristos, and Martin Paldam. 2013. The Robust Result in Meta-Analysis of Aid Effectiveness: A Response to Mekasha and Tarp. *Journal of Development Studies* 49: 584-587.

Doucouliagos, Hristos, and T. D. Stanley. 2012. Are All Economic Facts Greatly Exaggerated? Theory Competition and Selectivity. *Journal of Economic Surveys* 27: 316-339.

Felgenhauer, Mike, and Elisabeth Schulte. 2011. Strategic Private Experimentation. Working paper. **Link**

Hendry, D. F. 2001. *Econometrics: Alchemy or Science? Essays in Econometric Methodology*. Oxford: Oxford University Press.

Hunter, John E., and Frank L. Schmidt. 2004. *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*. 2nd ed. London: Sage.

Hylleberg, Svend, and Martin Paldam, eds. 1991. *New Approaches to Empirical Macroeconomics*. Oxford: Blackwell.

Ioannidis, John P. A. 2005. Why Most Published Research Findings Are False. *PLOS Medicine* 2(8): 696-701. **Link**

Leamer, E. E. 1983. Let's Take the Con Out of Econometrics. *American Economic Review* 73: 31-43.

Mekasha, Tseday Jemaneh, and Finn Tarp. 2013. Aid and Growth: What Meta-Analysis Reveals. *Journal of Development Studies* 49: 564-583.

Paldam, Martin. 2012. An Essay About Publication Bias and Strikes. Conference paper. Link

Roberts, Colin J., and T. D. Stanley, eds. 2005. *Meta-Regression Analysis: Issues of Publication Bias in Economics*. Oxford: Wiley-Blackwell.

Stanley, T. D. 2008. Meta-Regression Methods for Detecting and Estimating Empirical Effect in the Presence of Publication Bias. *Oxford Bulletin of Economics and Statistics* 70: 103-127.

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

Stanley, T. D., Hristos Doucouliagos, Margaret Giles, Jost H. Heckemeyer, Robert J. Johnston, Patrice Laroche, Jon P. Nelson, Martin Paldam, Jacques Poot, Geoff Pugh, Randall S. Rosenberger, and Katja Rost. 2013. Meta-Analysis of Economics Research Reporting Guidelines. *Journal of Economic Surveys* 27: 390-394.

Stanley, T. D., and Stephen B. Jarrell. 1989. Meta-Regression Analysis: A Quantitative Method of Literature Surveys. *Journal of Economic Surveys* 3: 161-170.

About the Author

Martin Paldam is professor (emeritus) of Economics at Aarhus University, Denmark. He is the author of about 200 published papers and seven books (link), mostly within public choice, development and meta-analysis. He has participated in all of the annual workshops of the Meta-Analysis of Economics Research Network (MAER-Net). He is international research fellow at the Kiel Institute for World Economy, and has been president of the European Public Choice Society, honorary professor at Deakin University in Melbourne, Australia, and co-editor at the *Scandinavian Journal of Economics* and the *European Journal of Political Economy*. His email address is mpaldam@econ.au.dk.

Go to archive of Economics in Practice section
Go to May 2013 issue

Discuss this article at Journaltalk:
http://journaltalk.net/articles/5796