5/4-2015

Simulating an empirical paper by the rational economist

Martin Paldam

Published in *Empirical Economics* 50:1383–1407 (2016)

Abstract: Papers in economics often try to find the 'best' estimate of a parameter. If researchers behave as predicted by economic theory, the research process can be modeled and simulated. The 'best' estimate is selected from *J* regression experiments by a selection rule, *SR*, which gives the researcher's preferences for the fit (significance) and size of the estimated parameter. Eight *J*s and five *SR*s are considered. To find a stable pattern for all 40 cases, 70 million regressions are simulated. The key results are: (1) All rational *SR*s cause the selected estimate to be substantially biased in the direction of the priors of the researcher. (2) All such rules give almost the same bias. It can be assessed from a set of estimates, and (3) the standard PET estimate of the meta-average reduces the bias by more than 90%.

Keywords: Meta-analysis, selection of regressions, publication bias
JEL: B4, C2

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

1. Introduction: Research as a process of search and choice

Think of an important parameter β that is taken to be the effect of x on y. Many researchers have tried to find the 'best' estimate of β. This has led to the β-literature. The typical paper in this literature is cast in the form of the standard research strategy with three stages:[1]

I. It starts from an intuition that leads to a theory, which is often a twist of a previous theory. It provides a qualitative prediction about the parameter. However, the decision-makers using β want a quantitative estimate. It is reached by the next two stages.

II. The theory is operationalized into an estimation model. It is rare that only one estimation model follows from a theory. In addition to the term of interest, βx, the model contains other variables: Some follow more or less closely from the theory, while others are *ceteris paribus* controls, meant to control for sample heterogeneity. Thus, the β-literature can be seen as attempts to catch the same β.

III. The estimation stage has two steps: A *search* among model variants, where each gives an estimate. The marginal cost of a regression is next to nothing, so many are run. The search is guided by the fit (t-ratio) and size of the estimated parameter and leads to the *choice* of the main result to report. It confirms the prediction and thus the theory.[2]

The paper models this process of search and choice by economic theory and simulates the outcome. A number of experiments and polls have demonstrated that economists are relatively rational.[3] If economic theory applies to people in general, it applies even more to economists. To model and simulate the process within the format of a paper has required a great deal of simplification, but I hope that the reader will agree that I catch the basic process.

Stage I: The theory is a data generating process (DGP), and the qualitative prediction is that $\beta > 0$.[4] In the DGP the true value is $\beta = 1$, so the prediction is true. Stage II is an estimation model (EM). In the simulations the DGP/EM-pair produces estimates.

Stage III has two steps: The search step consists of the production of a set of J estimates.[5] They are characterized in the two dimensions of fit and size, which gives a neat graphical representation. The choice step is simulated as a standard textbook decision using the production possibility frontier (PPF) of the J estimates and the indifference curves (IC) of

[1] The stages may represent the way the work was done, but loops from III to I or II are probably common.
[2] Fanelli (2010) examines the fraction of papers confirming the hypothesis proposed in the paper in 20 different sciences. In economics 87% of papers confirm the theory tested.
[3] The classical paper is Marwell and Ames (1981), see also Carter and Iron (1991) and Kirchgässner (2005). The theory of the rational economist is developed in Paldam (2015).
[4] The analysis is robust to the prediction $\beta < 0$. The case that $\beta = 0$ is simpler, but disregarded.
[5] The paper takes J as given. Paldam (2013a) model J from the costs and benefits of the marginal regression.

the researcher. In the size-fit diagram the PPF is the rim of the *J* estimates. The solution is the published estimate.

Parameter heterogeneity is assumed to be accounted for by the ceteris paribus controls that differ between studies – the simulations have no parameter heterogeneity and hence no ceteris paribus controls.[6] I simplify further by using a DGP that just contains the β-term, while everything else is hidden in a big noise term. Also, I assume that the influence of referees and editors works through author decisions. Hence, it can be internalized in his decision.

The standard research strategy is popular as it has two properties that are strongly desired by researchers: It is doable, and it leads to publications. However, it has frequently been criticized as too prone to moral hazard.[7] It will be shown that it does lead to publication bias, defined as a systematic difference between the true value of β and the reported estimates. Therefore, it has often been recommended that results are tested by robustness experiments and out-of-sample predictions. However, nothing prevents an author from including such tests in the search.[8] Hence, the validity of results ultimately rests on independent replications.

The most systematic replication analysis is the technique of meta-analysis. In economics approximately 750 meta-studies have been done. Each of these covers the literature that reports estimates of one parameter. This is about 50 papers in averages, so about 40,000 economic papers have been coded. This effort tells a lot about the pattern of results in empirical economics. The distribution of the results in a literature is often skewed in ways that suggest systematic publication biases. To correct for such asymmetries, the analysis estimates a meta-average.[9] It is often only half of the mean.

The selected estimates are a function of *J* and *SR*. The pattern in the estimates is analyzed by meta-analysis. It is also used to calibrate the simulations, so that they look like in a typical meta-study. As β is known, the true publication bias can be calculated – it confirms that the estimated meta-average is much closer to the true value than is the mean.

Section 2 presents the DGP/EM techniques for the simulated β-literature and the basic meta-technique. Section 3 discusses the five *SR*s used. They are chosen to span as much of the possibility space as possible. Section 4 reports the results from 70 million simulated regressions. Section 5 compares the results from the selection rules, and section 6 concludes.

[6] The ceteris paribus controls are imperfect. This adds to the noise-term in the simulations. The noise term is simulated as white noise. Hence everything is estimated by OLS.
[7] The literature on this problem is surveyed in Paldam (2013a) and Ioannidis and Doucouliagos (2013). Broader aspects are covered in Young *et al*. (2008).
[8] The average paper includes a dozen robustness checks and sometimes out-of-sample predictions.
[9] The meta-average is an estimated coefficient in a meta-regression analysis (MRA) aiming at estimating β.

2. Empirical and simulated β-literatures

Section 2.1 introduces the terminology of the paper. Sections 2.2 and 2.3 presents the techniques used to analyze a β-literature. Section 2.4 shows how such a literature can be simulated. Section 2.5 considers the two main priors assumed in the paper.

2.1 Terminology: Empirical and simulated N-sets

The paper studies a generic β-literature taken to be a set of N estimates, b_i, of the same parameter β. The full N-set is:

(1) $n_i = (b_i, s_i, t_i, p_i)$, where $i = 1, \ldots, N$, where, s_i, is the standard error, $t_i = b_i/s_i$, is the t-ratio, $p_i = 1/s_i$ is the precision. This means that $t_i = b_i\, p_i$, so that $p_i = t_i/b_i$.

An empirical meta-study collects the N-set from a literature search that tries to find all published estimates, b_i, pertaining to be of the same β.[10] The research process for these papers is unknown. A simulated meta-study uses a generated N-set. Here the research process is known. In my simulations each n_i is selected, by a selection rule, SR, from a set of J_i regressions done. Thus, the total number of regressions made to produce the β-literature is:

(2) $N\,\overline{J} = \sum_{i=1}^{N} J_i$, where $\overline{J}$ is the average number of estimates for each published.

N of these estimates are published. In empirical meta-studies the remaining $N(\overline{J} - 1)$ estimates are private information of the authors, and this extends to the Js as well. In simulations everything is known. The unweighted arithmetic means of the N-set and the J_i-set are $\overline{b}$ and $\overline{b}_{Ji}$ respectively.

When the simulated N-set is analyzed by the tools described in the next section, the mean is supplemented with estimated meta-averages, M. Practitioners of meta-analysis claim that these meta-averages are (much) closer to the true value β than the mean. As simulation results accumulate, the claim has been substantiated, even when the meta-averages are rarely

[10] For ease of presentation it is assumed that each study brings only one estimate. The average study actually reports about 10. The standard way to handle this is to use clustered standard errors, where each paper is a cluster. We dispense with this complication.

perfect estimates.[11] The results below add further evidence. The paper studies the expected/equilibrium values with a *:

(3) $\bar{b}^* = \bar{b}^*(J, SR)$ mean.

(4) $M^* = M^*(J, SR)$ meta-average that should be a good' estimate of β.

(5) $PB_T{}^* = \bar{b}^* - \beta$ true publication bias.

(6) $PB^* \approx \bar{b}^* - M^*$ estimated publication bias.

If M works as we want, (6) is a good estimates of (5). The pattern in (3) to (6) is simulated for a range of Js and SRs. As the DGP has a large variation to mimic empirical funnels, many simulated N-sets are needed to get close to the expected values. It will come as no surprise that a *publication bias* (5) and (6) appear for most values of J and SR, but it is not trivial how this bias looks over the range of Js and SRs. This is shown in Tables 2 to 6 below.

2.2 The funnel – looking at the distribution of the N-set[12]

The funnel is the (p_i, b_i)-scatter. It displays the distribution of the N-set in a particularly revealing way as is illustrated by Figures 5 to 8 below.[13] Funnels have a broad base for low precision and a narrow top for high precision. The simplest measure of the width of the funnel is the standard deviation, v, of the N-set. It will be used below.

In the *ideal* funnel b_i is independent of p_i, so the funnel is symmetric, and all averages are good estimates of β. This is the case when all estimates are published, as in the case below when $J = 1$. In this case the funnel width corresponds to the level of t-ratios of the estimates. The FAT, funnel asymmetry test, from Egger *et al.* (1997) tests if b_i depends on p_i. It is the first term in equations (7) and (8) in Table 1. Asymmetries cause the averages to differ, so the asymmetry needs an explanation to assess which average to prefer.

I have looked at many empirical funnels. My impression is that they have three

[11] See Stanley (2008), Callot and Paldam (2011), Paldam (2013b), Stanley and Doucouliagos (2014), and Reed *et al.* (2014).

[12] This paper uses only level one of the meta-analyses. Much more knowledge can be reached by perusing the recent textbook Stanley and Doucouliagos 2012 and the guidelines Stanley *et al.* 2013. A meta-study demands that the estimation models differ only in ways that can be coded. Although the theories in the β-literature often differ substantially between papers, the reported estimation models are normally fairly similar at the formal level. Few papers are so different that they cannot be included in meta-studies.

[13] One of reasons that researchers ought to make meta-studies in their field is that it is very thought provoking to look at the funnels. Stanley and Doucouliagos (2010) present many funnels and discuss their interpretation.

properties: (p1) About 2/3 are asymmetric. (p2) The estimated *b*'s are normally rather significant. (p3). The funnel is often amazingly wide considering the *t*-ratios.

Below it is shown what the rational behavior of economists does to funnels. The analysis starts from an ideal funnel where $J = 1$. As J increases rational selection causes three things to happen: (1) The funnel becomes increasingly asymmetric. (2) The average *t*-ratio increases. (3) The width of funnel remains as high as in the ideal funnel or even higher. It corresponds to my impressions of the properties of empiric funnels.

Academic economists tend to believe that the asymmetry is due to differences in the quality of the studies. Meta-analysts have often tried to analyze if this is the case by including a variable for the impact factor of the publication outlet. Also, dummies for techniques and time trends are often included in meta-studies. The effects of such variables tend to be small and they are often insignificant. This suggests that it does not invalidate the analysis below that it uses simple models and estimators in the DGP/EM-pair.

2.3 Two generations of meta-averages trying to correct the mean for funnel asymmetries

In a number of papers T.D. Stanley (notably Stanley, 2001) argued that the most common asymmetry is due to censoring of 'unreasonable' estimates. This cuts/reduces tails of the distribution and makes funnels leaner. Imagine that the left half is censored. That is a strong asymmetry. However, the average *t*-ratios stay roughly the same, while the funnel width falls to half. Some funnels look censored, but by no means all and few are lean.

Table 1. The definitions of the FAT-PET MRA and the PEESE MRA

		(a)	(b)	(c)
Equation	Name	Basic formulation	Curve in funnel	Estimation formulation
(7)	FAT-PET	$b_i = \beta_F s_i + \beta_M + u_{i1}$	$p_i = \beta_F/(b_i - \beta_M)$	$t_i = \beta_F + \beta_M p_i + v_{i1}$
(8)	PEESE	$b_i = \beta_F s_i^2 + \beta_P + u_{i2}$	$p_i = (\beta_F/(b_i - \beta_p))^{0.5}$	$t_i = \beta_F s_i + \beta_P p_i + v_{i2}$
Terms		FAT, β_F	PET, β_M	PEESE, β_P

Note: The three models in each row are equivalent. The *u*'s and *v*'s are disturbance terms. The abbreviations are: MRA, meta-regression analysis; FAT, funnel asymmetry test; PET, precision estimate test; and PEESE, precision-effect estimate with standard errors. Note that formulations (8c) used in the estimations have no constant. Stanley (2008) uses β_0 and β_1 for β_F and β_M. I find the second terminology easier to remember. The FAT-PET has largely replaced the MST (meta-significance test) proposed by Card and Krueger (1995) to deal with polishing, where researchers select by fit. In an earlier version of this paper the MST was run in parallel with the other MRAs, and I can confirm that the FAT-PET is better.

In 2008 Stanley developed the FAT-PET MRA, where the PET corrects the mean for the bias detected by the FAT. It was developed to handle censoring. The PET is the second term of equation (7) in Table 1. Later he developed the PEESE MRA of equation (8) in the table (see Stanley and Doucouliagos 2014). The logic of both estimated meta-averages is easiest to see from (7a) and (8a) that show how the estimates converge to β_M or β_P when s_i goes to zero. The difference is the speed of convergence as p rises. The FAT-PET and the PEESE curves often look similar when the estimates are depicted in the funnel.

It is important to note that Stanley considers the PET a test for $\beta \neq 0$ only. This is why he developed the PEESE. Stanley and Doucouliagos (2013, 2014) provide arguments from statistical theory and simulations to demonstrate these points in the case of censoring. I analyze the biases produced by rationality – they are different.

In empirical meta-studies the PET and the PEESE are widely used as an estimate of β, even when little is known about the research process behind the estimates.[14] Thus, historically PET and the PEESE are two generations of an estimator of β. A great deal is known about these estimates when the bias is due to censoring. The evidence cited above suggests that economists are quite rational. Therefore, it is important to see what happens, under the assumption of rationality.

Below it is shown that the PEESE MRA works poorly in this case, but fortunately the FAT-PET MRA work rather well also when it comes to biases due to rationality – those who have published meta-analyses using the PET will be happy to see the results.

2.4 A simulated β-literature: The DGP/EM-pair and the 'Chinese boxes' of Figure 1

To make the analysis tractable, it is kept simple: The DGP/EM pair is the same; the DGP has no constant; β = 1; and all control variables are treated as parts of the stochastic noise. Consequently, the noise term, εt, is chosen to be quite large.

(9a) DGP: $y_t = \beta\, x_t + \varepsilon_t$, where, $x_t = N(0, \sigma_x^2)$ and $\varepsilon_t = N(0, \sigma_\varepsilon^2)$.

(9b) EM: $y_t = b\, x_t + u_t$, estimated by OLS.

The three parameters in the DGP are: $\beta = 1$, $\sigma_x^2 = 2$ and $\sigma_\varepsilon^2 = 10$.[15] Section 4.7 discusses the (small) effect on the analysis when the three parameters are changed. The DGP is used to

[14] Also, by the author in his work with Chris Doucouliagos from 2008 onward.
[15] This gives rather high t-ratios, so that the estimates stabilize reasonably fast.

produce samples of m observations on which the EM is estimated. Thus, the DGP/EM-pair is the *production function* for the results in the simulated β-literature.

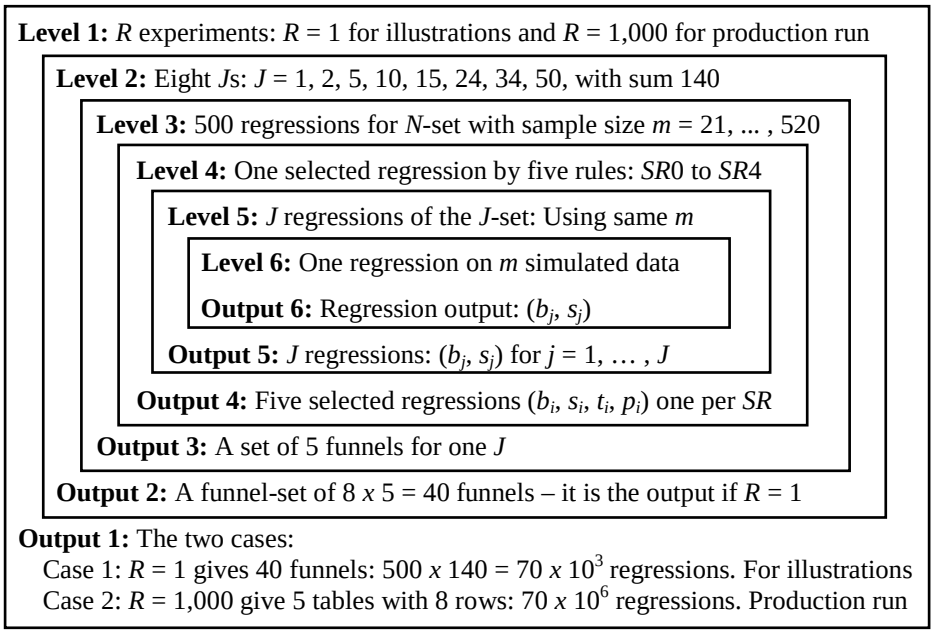


Fig. 1. The six nested levels in the simulation set-up

Note: Simulation variables: R, numbers of experiments; J, regressions searched; N, regressions selected for each funnel; m, sample size; SR, selection rule. Regression output variables: b_i, estimate of β; s_i, standard error; t_i, t-ratio; p_i, precision.

Figure 1 shows that the simulations have six nested levels: Level 1 is the R experiments. One experiment, R = 1, gives a set of 5 x 8 = 40 funnels, which are for 5 SRs and 8 Js. Single experiments are used for calibration and illustrations, such as Figures 5 to 8. A production run is for R = 1,000. It gives a set of five tables with seven rows as reported in Tables 2 to 6. The five tables thus report averages of 1,000 simulations of the funnel-set. Level 2 is the 8 values of J = 1, 2, 5, 10, 15, 23, 34, 50, with the sum 140. Level 3 is the N = 500 selected estimates for each funnel, calculated with different sample sizes, from m = 21 to 520.

Level 4 is five selection rules $SR0$, … , $SR4$, which are defined and discussed in section 3. They are used on each of the J-sets to select one observation for a funnel. Level 5 is the J-set of regressions made for the J chosen at level 2. Note that the five SRs use the same J-set. Finally at level 6 each regression is made by the DGP/EM-pair of equations (9).

Due to the nesting, the numbers of regressions quickly multiply into high values: R = 1 requires 140 x 500 = 70 x 10^3 simulated regressions to generate one funnel set. R = 1,000

require 70 x 10^6 simulated regressions to generate a set of 5 tables.[16]

Production runs have been made for R = 1, 100, and 1,000. The pattern in the results is analyzed as done in Figures 10 to 14 below. The curves have the same basic structure for all three Rs, but for R = 10 the pattern has a visible stochastic element. For R = 1,000 most curves have stabilized into a smooth pattern, but Figure 10 for the PET-bias has an enlarged vertical axis, and a small stochastic variation remains. However, the path of the effects is very clear.

The simulated researcher is naïve and up to no mischief: His theory (DGP) and model (EM) are simple, and the only choice made is formalized by *SR*. The priors of real humans affect more choices in research: Weakly justified controls are in and out of models or appear as instruments in two-stage estimates; data-sets are chosen, outliers are deleted, etc. Consequently, I expect the simulated biases to be too small.

2.5 The two main priors: For fit and size

The β-research community includes referees and editors. Such communities often have a couple of main priors that affect the average paper:

The first main prior is close to human nature. It is for clarity. We do not like to reach wishy-washy results when we have worked for long on a project, and neither referees nor editors like papers with such results. Consequently, authors polish results to make them look better. Thus, the first main prior is for *fit*, i.e., for large t-ratios. Section 3.3 considers the case where this prior is the only one. It is the first extreme selection rule *SR*1.

The second main prior comes from theoretical beliefs, moral/political opinions and sponsor interests. It leads to priors about the size of the estimate. Below it is assumed that the relevant mixture of these priors aggregate to one main prior,[17] which is for *size*, i.e., for large estimates. Section 3.4 turns to the case where this prior is the only one. It is the second extreme selection rule *SR*2.

I am fully aware that few researchers use a selection rule so extreme as *SR*1 or *SR*2. However, they define the outer limits to the selections a rational researcher will make: She will select an estimate that is smaller than the largest only if the selected estimate is more significant. And she will select an estimate that is less significant than the most significant only if the selected estimate is larger. Thus, the two extreme selection rules *SR*1 and *SR*2,

[16] A production run for R = 1,000 took a (moderately fast) pc running 24 hours a day 6½ days.
[17] If the priors predict different outcomes, it may result in funnels with more than one top. This is also the case if β has several distinct values. Funnels with more than one top require a meta-analysis at level two.

limit the range of rational choices.

3. A model of the rational economist, with five selection rules

The rational economist behaves as predicted by economic theory. It predicts exactly how he chooses the best estimate. Section 3.1 presents the model in a fit-size diagram. His indifference curves give a selection rule, *SR*. Section 3.2 sets the baseline, *SR*0. Sections 3.3 and 3.4 present the two extreme selections, *SR*1 and *SR*2. Section 3.5 discusses if they give different results in practice. Section 3.6 deals with a reasonable selection *SR*3. Finally, section 3.7 turns to satisficing behavior.

3.1 The size-fit diagram of Figure 2 and the PPF-IC analysis

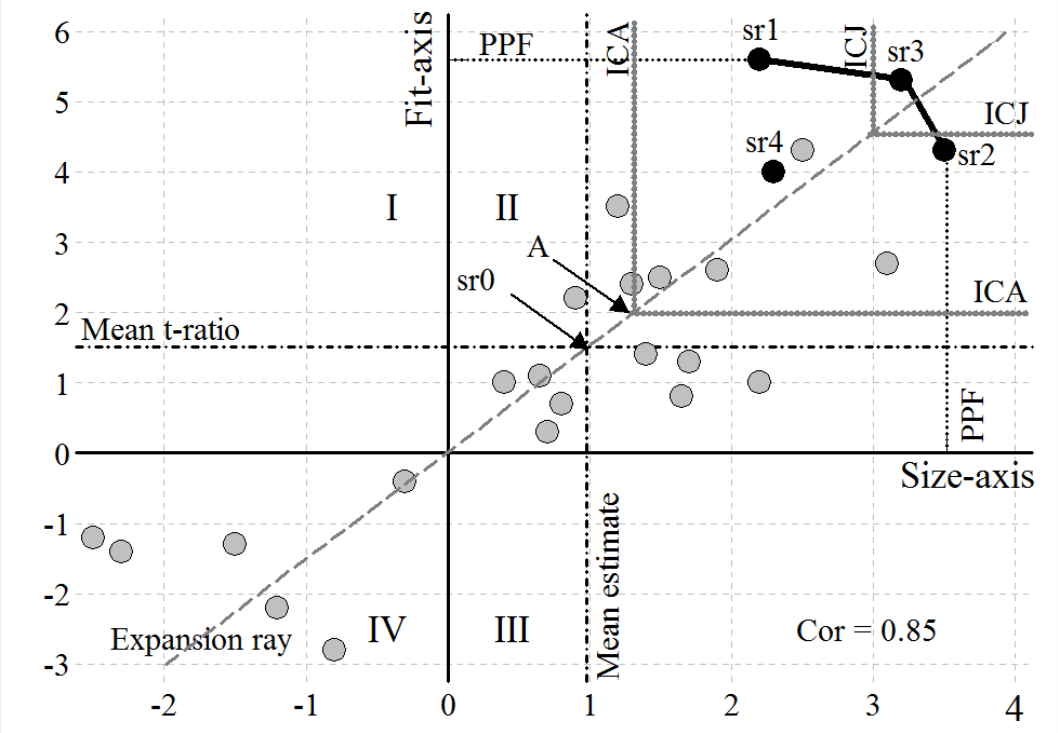


Fig. 2. A typical example of $J = 25$ regression estimates of β

Note: 'Size' is the estimated coefficient, 'fit' is its *t*-ratio. Drawn for $\beta = 1$, $\bar{b} = 0.98 \approx \beta$, while $\bar{t} = 1.51$, as indicated by the two 'mean' lines. It follows that $\bar{s} = 0.65$ and $\bar{p} = 1.54$. If the data sample is $m = 40$, the *t*-ratio should exceed 2 at the 5% level of significance. Figure 2 is calibrated so that it is reasonable to assume that it can generate the empirical example of Figure 3.

Figure 2 is an example of $J = 25$ estimates, shown as a (b, t)-scatter, to illustrate the 5

selection rules. Each selection rule *SR*, selects a sr-point on the figure. The size and fit divide the plane in 4 quadrants. The priors of the researcher are for positive *b*'s, so points in quadrant IV have the 'wrong' sign. Also, quadrant I and III are empty by definition, so the correlation between size and fit will normally be positive.

A ray is a straight line through origo (0,0). All rays with a positive slope have the property that when a point on the ray moves 'out' into quadrant II, both the fit and the size increase. Thus, it represents 'better' results by both preferences.

The choice can be formulated in the language of economic theory.[18] The *J* estimates are the output of the DGP/EM production function. Thus, the outer rim of the *J* points is the PPF-curve, as indicated on Figure 2. All interior points are worse (closer to origo) than some points on the PPF. The horizontal and the vertical sections of the PPF are dotted as they cannot be touched by a reasonable IC, as the reader will know. Thus the only part of the PPF that is relevant for the rational choice is the bold convex curve from sr1 to sr2.The ICs of the researcher are for fit and size. The optimal choice of the researcher is the kink-point on the PPF that touches his utmost (and hence best) IC. The kink-points are stochastic, but for many simulations of *J* points the average PPF-curve will surely be nicely rounded, and it moves out for *J* rising. Once the ICs are chosen, an expansion path should follow in the usual way. The simulations assume that the expected expansion path is a ray.

The theory behind Figure 2 leads to the important (if trivial) ***first key result***: When the estimates are distributed randomly around the true value, an optimal selection based on size and fit must inevitably lead to an exaggerated estimate. This confirms the empirical result of the first meta-meta-study (Doucouliagos and Stanley 2013). A typical size of the exaggeration is 2, so that $\overline{b} \approx 2\beta$, as in the example used for Figure 3 below. However, the quantitative results for different selection rules are still interesting, and it is not trivial if β can be retrieved from a set of biased ones by a meta-average such as the PET or the PEESE.

3.2 SR0, the baseline, where the mean is selected. It is point sr0 on Figure 2

Imagine a researcher who has no priors at all, or more likely, a researcher who manages to suppress his priors. He reports the median of the *J* regressions and some measure of their variance. For a symmetrical distribution the expected median is the average.

[18] The reader will recall the terminology from the introduction. The DGP/EM is the data generating process and the estimating model. The PPF is the production possibility frontier and the IC is the indifference curve.

*SR*0: Select the average $\bar{b}_J$. It also shows what the researcher should expect to find in the next regression. The results from *SR*0 are reported in Table 2 below.

The standard deviation of each $\bar{b}_J$ in the *N*-set falls by the square root to the number of observations. Thus, *SR*0 should cause the funnel width, measured by *v*, the standard deviation, of the *N*-set to fall by $1/\sqrt{J}$, as *J* grows.[19] The central result for an unbiased researcher is point sr0 on Figure 2. In the simulations it is close to the true value of β by design. The choice by *SR*0 for $J > 1$ is a point inside the PPF-curve. Thus, it is not a rational choice.

3.3 SR1, selection by the fit of the estimate. It is point sr1 on Figure 2

The first of the two extreme *SR*s is due to the first main prior for fit. It has often been noted that our profession is greatly concerned about statistical significance, even at the expense of economic significance.[20] Imagine a researcher who has found a theoretically satisfactory model giving an estimate *b*, which he believes to be a good estimate of β, but where *b* is insignificant. Polishing means that he searches for a model variant close to the good model that increases the fit of *b*. That is, he makes the *J* experiments to find an estimate with a good *t*-ratio. The selection rule in the polishing case is thus by the *t*-ratio and independent of the size of *b.*

*SR*1: Select the *b* with the highest *t*-ratio. Here the indifference curves are *horizontal*. The sr1-point is larger than 1; see Table 3 below.

3.4 SR2, selection by the size of the estimate.[21] It is point sr2 on Figure 2

The second extreme *SR* is due to the second main prior for size. If the profession believes that $\beta > 0$, most researchers will discriminate against negative values, and to be on the safe side

[19] The points of the funnel are calculated from samples with different length, so the simple $1/\sqrt{J}$ rule is an approximation only. Table 2b shows that it is a rather good approximation.

[20] It is not uncommon to read papers where the statistical fit of estimates is stressed, but it is left to the reader to find out what the size of the estimates means. D.N. McCloskey has argued that the preference for statistical over economic significance is harmful; see McCloskey (1998) and Ziliak and McCloskey (2008). Below we analyze this question as the difference between the choices *SR*1 and *SR*2.

[21] This *SR* is easy to solve analytically, as done in Paldam (2013a). The solution assumes that the *J*-set is normally distributed, and uses the inverse of the cumulative normal distribution to calculate the expected value of the largest of *J* observations.

they may select large positive values. *SR*2 takes this idea to the extreme:

*SR*2: Select the largest *b* in the *J*-set. Here the indifference curves are *vertical*. The sr2-point is larger than 1; see Table 4 below.

As mentioned in section 2.5 it is unlikely that either of the two extreme selection rules is used in practice, but they define the end-points for any optimal selection based on fit and size. This is why only the section of the PPF between sr1 and sr2 is a solid line. Figure 2 has only one point on this section, so it becomes the one and only *reasonable* selection. Section 3.6 consider the case where there are more reasonable observations to choose from.

3.5 Is selection by fit and size the same?

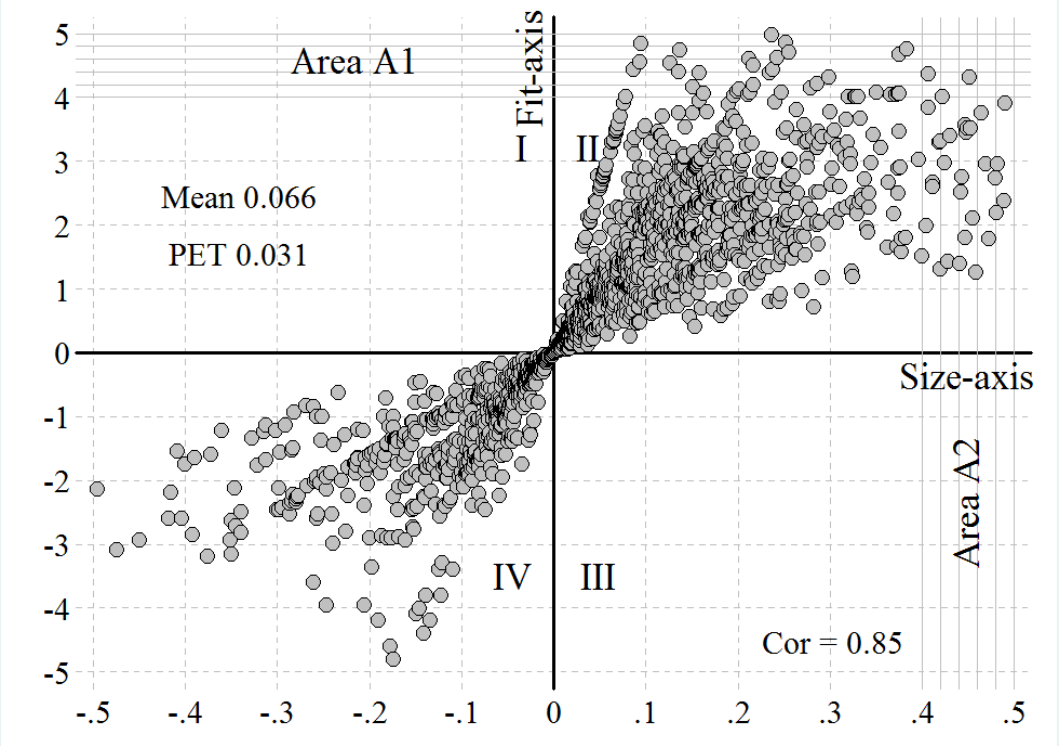


Fig. 3. The scatter of the *t*-ratios over the estimates in the aid effectiveness literature

Note: The 1,777 points depicted are available in stata format from the URL: http://www.martin.paldam.dk/Papers/Meta-AEL/Aidongrowth.dta. 91 observations are outside the frames of the graph. The figure uses the same format as Figure 2, but the estimates are converted to partial correlations. The data are analyzed in Doucouliagos and Paldam (2008 to 2015).

Selection by fit is often seen as fairly innocent, while selection by size is assumed to

give a substantial bias. The stylized example on Figure 2 shows that *SR*1 and *SR*2 give results that differ, but not much compared to the full range of estimates. Also, it is easy to construct examples where *SR*1 and *SR*2 select the same point. This is likely to happen for small *J*s.

Figure 3 shows the example of a (b_i, t_i)-scatter of the estimates in a literature. The correlation cor(b_i, t_i) = 0.85 (as on Figure 2), so the two variables are strongly related. Contrary evidence appears from the two areas A1 and A2: Area A1 contains the observations with the highest *t*-ratios (between 4 and 5). It is shaded with horizontal lines. It is likely that these 43 observations are selected by fit (*SR*1). Area A2 contains the observations with the highest coefficients (between 0.4 and 0.5). It is shaded with vertical lines. It is likely that these 32 observations are selected by size (*SR*2). Only 3 observations are in the checkered area where the A1 and A2 areas overlap. Thus, the empirical example provides contradictory evidence. Fortunately, the simulations in section 4 clear up the contradiction by showing that *SR*1 and *SR*2 give surprisingly similar results.

3.6 SR3, the selection of the best mix of fit and size. It is point sr3 on Figure 2

I think that researchers look for the estimate that is best by some mixture of size and fit. We want *SR*3 to choose a point on the relevant part of the PPF, if such points exist. To make the results tractable, two simplifying assumptions are made:

(a1) The expansion path is a ray that intersects the relevant part of PPF. The slope of this ray represents the researcher's trade-off between fit and size.

(a2) To make the indifference curve chooses a point between sr1 and sr2 it is taken to be vertical above the expansion path and horizontal to the right of this path. Two such indifference curves, ICA and ICJ are shown on Figure 2.

Everything follows once the expansion ray is chosen. It is chosen as the ray through the baseline point sr0 ($\overline{b}_J$, $\overline{t}_J$). It represents the realistic expectation the researcher tries to improve upon. The expansion ray has the formula:

(10) $t = (\overline{t}_J / \overline{b}_J)\, b = (1/\overline{s}_J)\, b = \overline{p}_J\, b$, so a point on this ray is $(b, t) = (b, \overline{p}_J\, b) = (\overline{s}_J\, t, t)$.

The expected path is $t = \overline{p}_J^* b$, but in practice both $\overline{t}_J$ and $\overline{b}_J$ are stochastic. The ICA-curve has a kink in *A* discussed in section 3.7. ICJ contains one point only. It is calculated as follows: First all *J points* in the *J*-set are converted to *raypoints* on the expansion ray.

(11a) $h(b, t) = b - \overline{s}_J\ t$ the horizontal distance to the expansion ray, at point ($\overline{s}_J\ t$, t).

(11b) $v(b, t) = t - \overline{p}_J\ b$ the vertical distance to the expansion ray, at point (b, $\overline{p}_J\ b$).

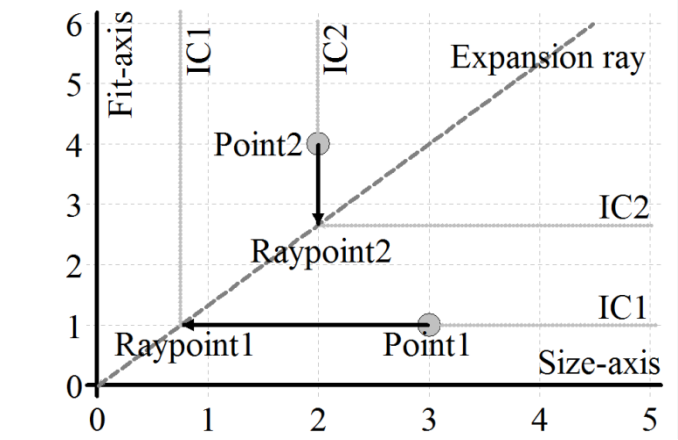


Fig. 4. The conversion of points to raypoints
Note: Drawn as a part of Figure 2. The ray has the equation (10). Point1 is (3, 1) so that the horizontal distance to the ray is negative and raypoint1 is (1(3/4), 1) = (0.75, 1). Point2 is (2, 4,) so that the horizontal distance to ray is positive and Raypoint2 is (2, 2(4/3)) = (2, 2.67). Point2 is preferred to Point1 as Raypoint2 is further from origo than Raypoint1. Also, IC2 is better than IC1.

If (b, t) is not on the expansion ray, two possible raypoints occur. Choose the one closest to origo. If (b, t) is below the ray as point1 on Figure 4, it is converted to raypoint1 ($\overline{s}_J\ t$, t) by (11a). If (b, t) is above the ray as point2, it is converted to raypoint2 (b, $\overline{p}_J\ b$) by (11b). Thus, the *J*-set is converted to raypoints. *SR*3 chooses the utmost. This corresponds to:

*SR*3: Select the last point inside the outmost IC that still contains a point. The sr3-point is larger than 1; see Table 5 below.

3.7 SR4 selects the first satisfactory estimate. It is point sr4 on Figure 2

Until now it has been assumed that *J* is exogenous. *SR*4 is the *SR* of a researcher in a hurry. She stops the search when a satisfactory result is reached, even if *J* is not reached. That is, she

selects the first point with a raypoint that exceeds the aspiration level *A*. [22]

*SR*4: Select the first acceptable result where the raypoint exceeds *A*. On Figure 2 eight points are within the ICA. By chance *SR*4 is the one reached first. All these points are larger than 1; see Table 6 below.

In the simulations we set A so low that the results for *SR*4 really deviate. Note that it will be by chance only if *SR*4 reaches the PPF. The example eight satisfactory points of which three are on the rim, so the chance of hitting one of these is 3/8 = 0.375.

4. The results from 70 million simulated regressions

Section 4.1 explains the format of the five tables reporting the results. Each of the following five sections covers one selection rule. First some typical funnels generated by the rule are shown, and then the table of the average results for the 1,000 funnels is reported. Section 4.7 explains what happens when the parameters of the simulations are changed. For $J = 1$ all selections are the same, as only one estimate can be selected. Hence, there are only 36 different funnels.

4.1 The format of Tables 2 to 6 reporting the average results

Each row is for one *J*. As just mentioned the first row for $J = 1$ is always the same. When *J* increases the tables diverge. Row nine is an average, giving a (crude) estimate of the outcome, when authors use different *J*s.

Column (1) gives the *J*-value; column (2) reports the mean of the 1,000 means, $\overline{\overline{b}} \approx \overline{b}^*$; (3) gives $\overline{\overline{t}} \approx \overline{t}^*$, which is the average of the average *t*-ratios, and (4) holds the average width (standard deviation), $\overline{v} \approx v^*$, of the funnel.

The FAT-PET MRA is reported in columns (5) to (8); (5) is the average estimated FAT, $\overline{b}_F \approx \beta_F$, while (6) counts how often the FAT rejects symmetry; (7) is the average estimated PET meta-average, $\overline{b}_M \approx \beta_M$, while (8) counts how often b_M differs from 1, so that

[22] It is chosen as the first point on the expansion ray from section 3.6, where *t* exceeds 2. Here *b* is $2\,\overline{s}_J$. In the example of Figure 2, this *b*-value is 1.34. If no satisfying point is reached before *J*, use *SR*3.

the PET fails to find the true value. The PEESE MRA is reported in column (9) and (10). Column (9) reports the average meta-average, $\overline{b}_P \approx \beta_P$, while (10) counts how often b_S differs from 1, so that the PEESE fails to find the true value. The FAT-term from the PEESE is not reported as it is almost the same as the one reported in (5) and (6). The three count-columns (6), (8) and (10) use the 5% level of significance in the tests.

Only *SR*1 is related to censoring, but it is actually rather different. Therefore, it is unknown how the PET and the PEESE reacts. However, both MRAs have been applied on empirical funnels generated by research processes that are likely to be affected by rationality.

4.2 SR0: the baseline, where the selection is unbiased and the funnels are symmetric

Two funnels are shown. Figure 5a is the ideal funnel for J = 1, where the average t-ratio is 3.1. The figure is symmetrical around 1 and as lean as predicted by the t-ratios of the estimates (see Callot and Paldam, 2011). This figure is the same for all five *SR*s.

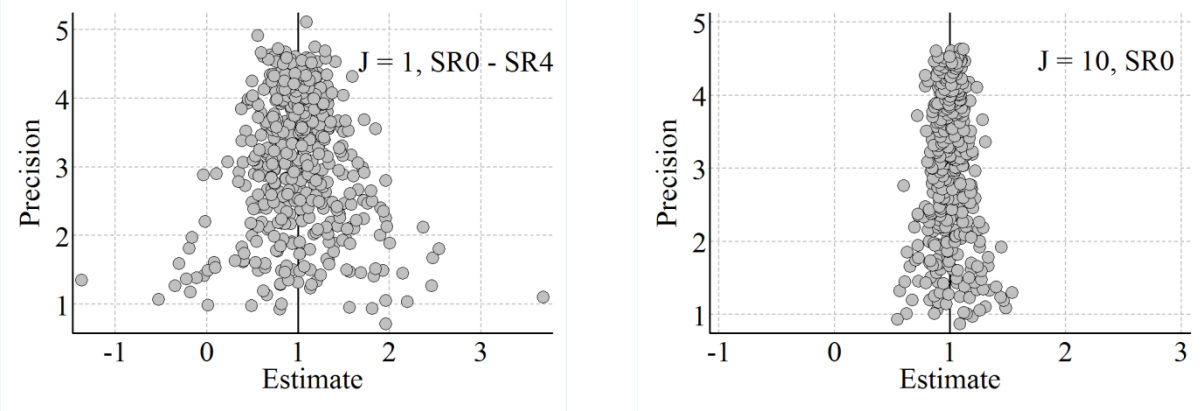


Fig. 5. *SR*0: The ideal funnel: Selecting the average

Fig. 5a. For J = 1. Table 2 to 6 row (1)

Fig. 5b. For J = 10. Table 2 row (4)

Table 2 reports the average result for the 1,000 funnels: The mean, $\overline{\overline{b}}$, the PET, $\overline{b}_M$, and the PEESE, $\overline{b}_P$, are all close to the expected value 1 as they should. The tests in columns (5), (7), and (9) reject the true results in 5% of the cases as they should. Also, the FAT finds no funnel asymmetry. In this case there is no bias – it is reassuring that none is found

Table 2. Selection rule *SR*0, the ideal selection

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
		Descriptive statistics			FAT-PET FAT asym. test		PET meta-avr.		PEESE Meta-average	
Row	J	$\overline{\overline{b}}$	$\overline{\overline{t}}$	$\overline{v}$	$\overline{b}_F$	Not 0	$\overline{b}_M$	Not 1	$\overline{b}_P$	Not 1
(1)	1	1.001	3.145	0.411	-0.004	5.8	1.002	6.2	1.001	5.8
(2)	2	1.000	3.147	0.289	0.000	4.6	1.000	4.8	1.000	6.1
(3)	5	1.000	3.144	0.183	0.004	6.7	0.999	5.6	0.999	5.2
(4)	10	1.000	3.145	0.130	-0.001	4.2	1.000	4.9	1.000	4.7
(5)	15	1.000	3.145	0.106	0.000	4.8	1.000	5.0	1.000	4.5
(6)	23	1.000	3.146	0.085	-0.001	4.7	1.000	4.8	1.000	5.3
(7)	34	1.000	3.146	0.070	0.000	6.6	1.000	6.5	1.000	5.3
(8)	50	1.000	3.145	0.058	0.000	4.3	1.000	3.9	1.000	4.5
(9)	Avr.	1.000	3.145	0.167	0.000	5.2	1.000	5.2	1.000	5.2

Note: Figure 5 shows funnel from rows (1) and (4). The table is explained in section 4.1.

Table 2b. The fall in v from column (3) in Table 2 – compared with $1/\sqrt{J}$

J	1	2	5	10	15	23	34	50
Col (3) rel	1	0.705	0.445	0.316	0.257	0.208	0.171	0.141
$1/\sqrt{J}$	1	0.707	0.447	0.316	0.258	0.209	0.171	0.141
Difference in%		0	0.3	0.5	-0.1	0.3	0.3	0.2

Note: 'Col (3) rel' is the data in column (3) of Table 2 divided by 0.41, the estimate for $J = 1$.

The t-ratio $\overline{\overline{t}}$ stays constant. Table 2b shows that the averaging for $J > 1$ reduces the width of the funnel by $1/\sqrt{J}$ as predicted. This is also shown by the funnel in Figure 5b. As mentioned in section 2.2, most meta-studies show that empirical funnels are rather wide relative to the *t*-ratios of the estimates. Thus, researchers often fail to control their priors.

4.3 SR1, selection by fit

The first extreme rule selects by fit only. The results are reported in Table 3. Columns (2) and (3) show what happens to $\overline{\overline{b}}$ and $\overline{\overline{t}}$ as J rises. Obviously $\overline{\overline{t}}$ rises, but so does $\overline{\overline{b}}$. Recall from (1) that $p = t/b$. As $\overline{\overline{t}}$ and $\overline{\overline{b}}$ rise roughly proportionally, the p's stay fairly constant.

This is nicely illustrated on left hand panel of Figure 6. For J =1 the funnel is the same as for all the *SR*s, but the three funnels of the left hand panel of Figure 6 show that when J rises the funnel moves to the right and change form. It becomes more and more asymmetric,

and few negative values appear from $J = 5$. Also, the funnel becomes more sausage-like. Note that even when the *t*-ratio goes up the standard deviation of the funnel stays almost constant. This is precisely the three properties we wanted to see in section 2.2.

Empirical funnels are a mixture of funnels with different *SR*s and *J*s. Therefore, they do not look as sausage-like as the extreme funnels generated. However, they often have tails at the bottom, as the simulated ones.

Table 3. Selection rule *SR*1, the largest t-ratio

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
		Descriptive statistics			FAT-PET				PEESE	
					FAT asym. test		PET meta-avr.		Meta-average	
Row	J	$\overline{\overline{b}}$	$\overline{\overline{t}}$	$\overline{v}$	$\overline{b}_F$	Not 0	$\overline{b}_M$	Not 1	$\overline{b}_P$	Not 1
(1)	1	1.001	3.145	0.411	-0.004	5.8	1.002	6.2	1.001	5.8
(2)	2	1.205	3.726	0.352	0.586	99.3	0.991	5.6	1.103	99.9
(3)	5	1.416	4.344	0.342	1.218	100.0	0.978	10.1	1.208	100.0
(4)	10	1.545	4.739	0.359	1.612	100.0	0.971	17.1	1.272	100.0
(5)	15	1.612	4.949	0.371	1.825	100.0	0.966	24.2	1.304	100.0
(6)	23	1.676	5.154	0.387	2.034	100.0	0.962	33.2	1.335	100.0
(7)	34	1.731	5.331	0.401	2.215	100.0	0.957	42.9	1.361	100.0
(8)	50	1.782	5.497	0.416	2.390	100.0	0.951	56.0	1.384	100.0
(9)	Avr.	1.496	4.611	0.380	1.485	-	0.972	24.4	1.246	88.2

Note: Left hand panel of Figure 6 shows funnel from rows (3), (5) and (8). The table is explained in section 4.1.

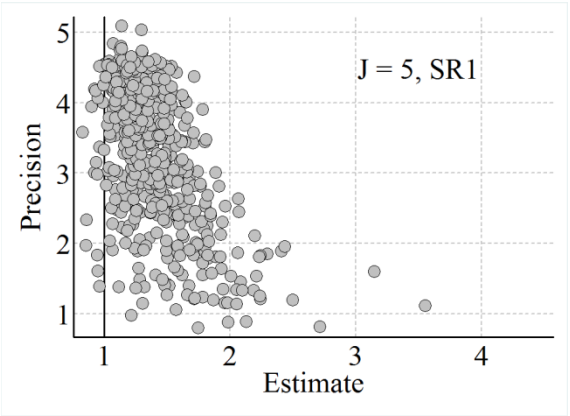


Fig. 6.1. *SR*1 for $J = 5$. Table 3 row (3)

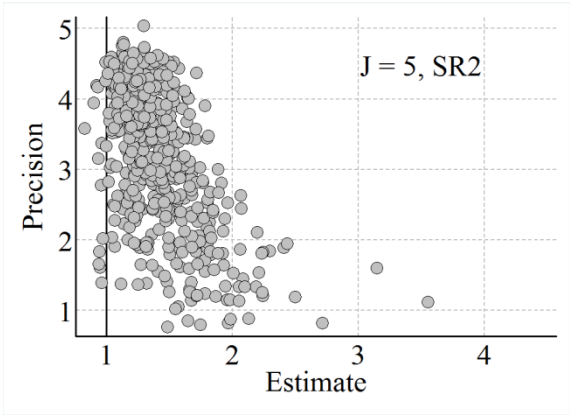


Fig. 6.2. *SR*2 for $J = 5$. Table 4 row (3)

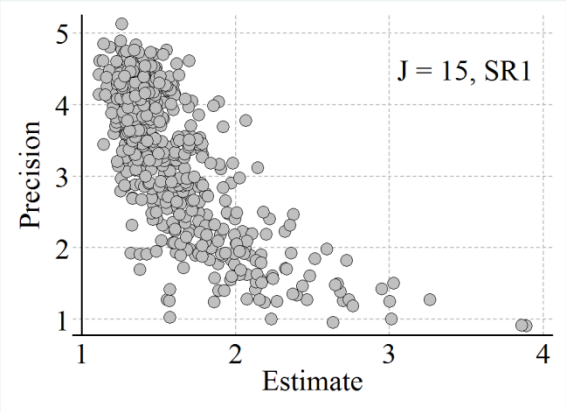


Fig. 6.3. *SR*1 for *J* = 15. Table 3 row (5)

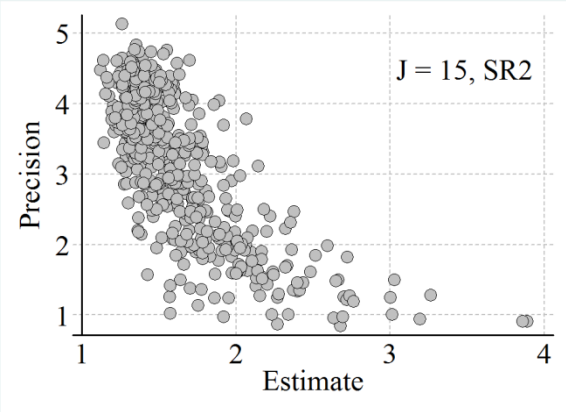


Fig. 6.4. *SR*2 for *J* = 15. Table 4 row (5)

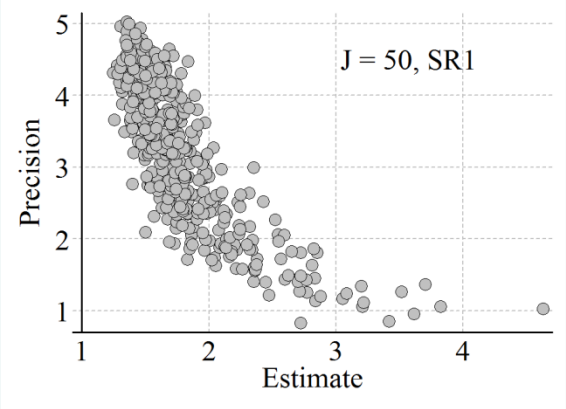


Fig. 6.5. *SR*1 for *J* = 50. Table 3 row (8)

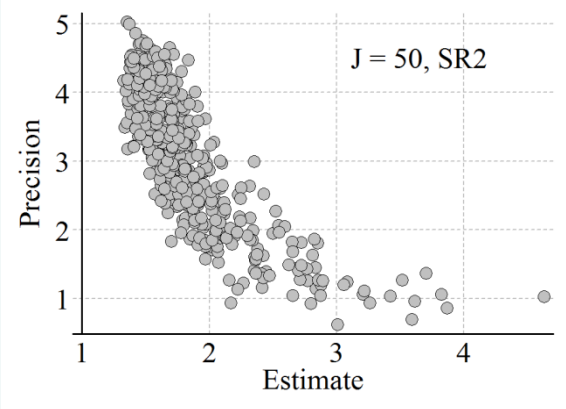


Fig. 6.6. *SR*2 for *J* = 50. Table 4 row (8)

Fig. 6. *SR*1 and *SR*2: Comparing funnels selected by the two extreme *SR*s

The increasing asymmetry of the funnel as *J* rises causes the FAT to rejects symmetry more and more decisively as *J* increases. The two meta-averages were made to handle censoring, but the PET in column (7) manages to stay within 5% of the true value. The PEESE in column (9) is in the right direction, but remains at about halfway between the mean and the true value. When the two curves are included in the funnels the PEESE-curves become vertical too soon.

4.4 SR2, selection by size

*SR*2 is the second extreme rule – it selects by size only. The funnels shown are reported as the right hand panel of Figure 6, while Table 4 reports the simulation results. A comparison of the 3 x 2 graphs and columns (2) in tables 3 and 4 gives the ***second key results*** in the paper: The

biases produced by fit and size are similar.

Table 4. Selection rule *SR*2, the largest estimate

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
		Descriptive Statistics			FAT-PET				PEESE	
					FAT asym. test		PET meta-avr.		Meta-average	
Row	*J*	$\overline{\overline{b}}$	$\overline{\overline{t}}$	$\overline{v}$	$\overline{b}_F$	Not 0	$\overline{b}_M$	Not 1	$\overline{b}_P$	Not 1
(1)	1	1.001	3.145	0.411	-0.004	5.8	1.002	6.2	1.001	5.8
(2)	2	1.209	3.714	0.352	0.566	99.1	1.001	5.0	1.110	100.0
(3)	5	1.431	4.306	0.345	1.157	100.0	1.003	4.2	1.228	100.0
(4)	10	1.571	4.675	0.368	1.509	100.0	1.011	7.0	1.306	100.0
(5)	15	1.645	4.868	0.386	1.693	100.0	1.015	9.1	1.348	100.0
(6)	23	1.717	5.054	0.408	1.870	100.0	1.019	13.6	1.389	100.0
(7)	34	1.780	5.214	0.430	2.023	100.0	1.023	18.4	1.424	100.0
(8)	50	1.839	5.361	0.453	2.162	100.0	1.027	26.7	1.458	100.0
(9)	Avr.	1.524	4.542	0.394	1.372	-	1.013	11.3	1.283	88.2

Note: Right hand panel of Figure 6 shows funnel from rows (3), (5) and (8). The table is explained in section 4.1.

Obviously the choice by size *SR*2 gives a larger bias in the mean than does *SR*1, but it is only by 8% even for $J = 50$. Also, *SR*2 causes the *t*-ratio to increase a little less than *SR*1. Consequently the 3 pairs of graphs are amazingly similar as well.

Like *SR*1 also *SR*2 causes *t*-ratios to rise, funnels to become more asymmetric, and the funnel width to stay approximately constant. The PET still works rather well. While the PET was marginally smaller than the true value when *SR*1 was used, it is now marginally larger. The PEESE is still about halfway between the mean and the true value.

4.5 SR3, the realistic selection by both fit and size

*SR*3 is the realistic case, where the researcher looks at both fit and size. As any combined *SR*3 is between *SR*1 and *SR*2, it is no wonder that the two pictures in Figure 7 look much like the corresponding pictures on Figure 6. Table 5 shows that all results are close to the average for *SR*1 and *SR*2, which have the opposite signs. Hence, the results from the PET are even better for *SR*3 than for *SR*1 and *SR*2.

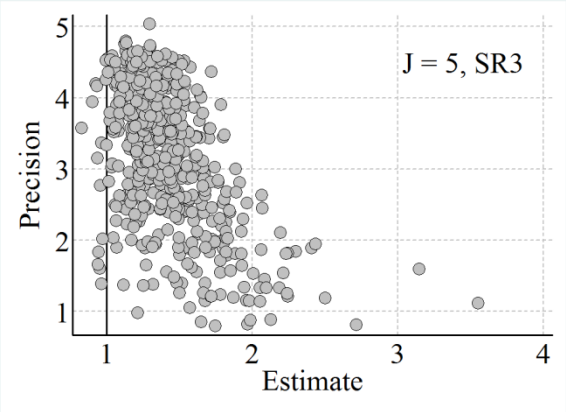


Fig. 7a. For J = 5. Table 5 row (3)

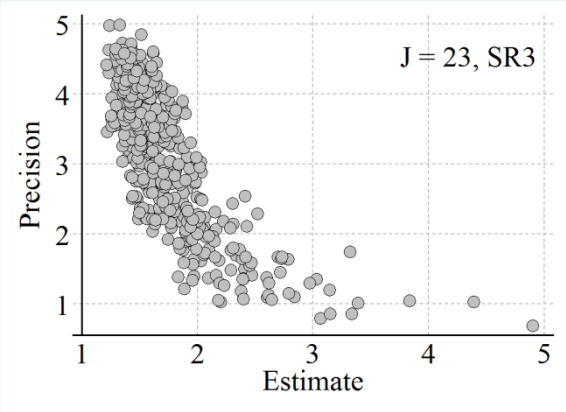


Fig. 7b. For J = 23. Table 5 row (6)

Fig. 7. *SR*3: The combined funnel: Selecting the optimal combination of fit and size

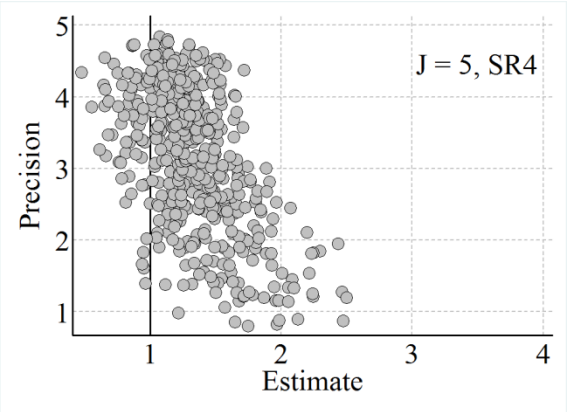
Table 5. Selection rule *SR*3, the combined selection

	(1)		(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
		Descriptive			FAT-PET				PEESE	
		Statistics			FAT asym. test		PET meta-avr.		Meta-average	
Row	J	$\overline{\overline{b}}$	$\overline{\overline{t}}$	$\overline{v}$	$\overline{b}_F$	Not 0	$\overline{b}_M$	Not 1	$\overline{b}_P$	Not 1
(1)	1	1.001	3.145	0.411	-0.004	5.8	1.002	6.2	1.001	5.8
(2)	2	1.208	3.723	0.352	0.580	99.4	0.996	4.7	1.107	99.9
(3)	5	1.427	4.332	0.344	1.196	100.0	0.990	5.6	1.219	100.0
(4)	10	1.563	4.716	0.364	1.576	100.0	0.990	6.3	1.290	100.0
(5)	15	1.635	4.919	0.379	1.777	100.0	0.989	7.5	1.327	100.0
(6)	23	1.705	5.116	0.399	1.976	100.0	0.988	6.7	1.363	100.0
(7)	34	1.765	5.286	0.418	2.148	100.0	0.987	10.1	1.394	100.0
(8)	50	1.821	5.444	0.438	2.309	100.0	0.986	11.0	1.422	100.0
(9)	Avr.	1.515	4.585	0.388	1.445	-	0.991	7.3	1.265	88.2

Note: Figure 7 shows funnel from rows (2) and (5). The table is explained in section 4.1.

4.6 SR4, the satisficing choice and a summary

The results of applying *SR*4 are much as the three previous *SR*s for small *J*s, as expected, but as *J* rises more choices differ as this *SR* may stop selecting well before it reaches *J*. This is easy to see on the two funnels of Figure 8, which looks as they combine funnels with different *J*s. It means that the publication bias increases less when *J* rises as seen in column (2) of Table 6. Here the PET and the PEESE are less different. They still adjust the average to gets closer to 1 than the mean, but the PET is not as efficient as for the three previous *SR*s.



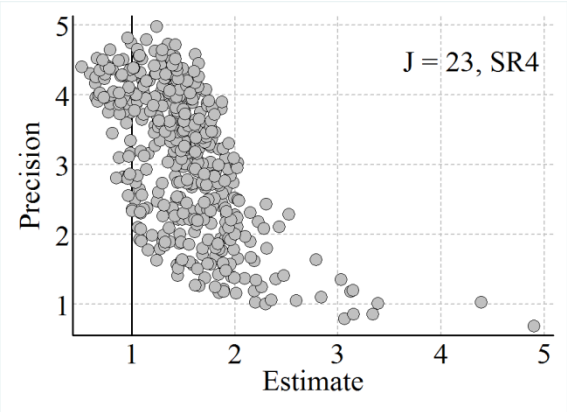


Fig. 8. *SR*4: The satisficing funnel

Fig. 8a. For $J \leq 5$. Table 6 row (3)

Fig. 8b. For $J \leq 23$. Table 6 row (6)

Table 6. *SR*4: The satisficing *SR*, select first acceptable result

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
		Descriptive Statistics			FAT-PET FAT asym. test		FAT-PET PET meta-avr.		PEESE Meta-average	
Row	J	$\overline{\overline{b}}$	$\overline{\overline{t}}$	$\overline{v}$	$\overline{b_F}$	Not 0	$\overline{b_M}$	Not 1	$\overline{b_P}$	Not 1
(1)	1	1.001	3.145	0.411	-0.004	5.8	1.002	6.2	1.001	5.8
(2)	2	1.176	3.602	0.358	0.710	100.0	0.917	55.7	1.056	89.2
(3)	5	1.341	4.035	0.373	1.347	100.0	0.851	98.2	1.112	100.0
(4)	10	1.430	4.284	0.405	1.636	100.0	0.837	98.3	1.153	100.0
(5)	15	1.475	4.412	0.423	1.742	100.0	0.844	97.6	1.181	100.0
(6)	23	1.514	4.530	0.440	1.835	100.0	0.851	93.9	1.208	100.0
(7)	34	1.548	4.633	0.457	1.907	100.0	0.860	86.4	1.232	100.0
(8)	50	1.580	4.728	0.473	1.970	100.0	0.871	75.4	1.256	100.0
(9)	Avr.	1.383	4.171	0.418	1.393	88.2	0.879	76.5	1.150	86.9

Note: Figure 8 shows one funnel from rows (3) and (6). The table is explained in section 4.1.

The five tables of results cover 40 combinations of *J* and *SR.* The five cases (for *J* =1) are the same, so only 36 cases are different. Eight of these give unbiased estimates. One is the case for *J* =1. The seven remaining unbiased cases are for *SR*0. In all the 32 remaining cases, where a bias is possible, it does appear. The four rational *SR*s all have the property that when *J* rises they causes funnel to be increasingly asymmetric, and to have steadily rising *t*-ratios, while the funnels remains as wide as the corresponding ideal (for *J* = 1) or even wider.

4.7 Varying the parameters and missing aspects

The setup of the analysis contains few parameters that can be varied: J, SR, m, β, σ_x, and σ_ε. The experiments reported cover, m, J, and SR, which are taken to cover realistic ranges. If β is changed, all that happens is a linear shift along the horizontal axis. Thus, only the effects of two standard deviations, σ_x and σ_ε, remain to be analyzed. They have been submitted to a set of experiments (for $R = 10$) keeping $\sigma_\varepsilon^2 = 10$ constant, while $\sigma_x^2 = 1$, 1.5, 2 (used as the main case), 3 and 4. The estimate of $\overline{\overline{b}}$, the FAT and the PET remained virtually unchanged. However, $\overline{\overline{t}}$ varies proportionally with σ_x^2, for all values of J.

Thus, the results are rather robust within the frames used, but they are restrictive in some respects. The most problematic is that the experiments deal with data variation, not model variation. If model variation is added the funnel width will increase.

5. The pattern in the average results

Each of the five tables of results reports a set of experiments with one selection rule. They are compared across rules in this section. First, section 5.1 compares the publication bias for the five *SR*s, and then section 5.2 discusses what happens in the realistic case where authors use different *J*s and *SR*s. Section 5.3 turns to the PET and the PEESE. Finally section 5.4 considers the FAT and the standard deviation.

5.1 The first key result: A substantial publication bias

Columns (2) in the five tables give the results for the average mean. If $\beta = 1$ is deducted from the estimated means it gives the true publication bias shown on Figure 9. The curves are so smooth that it is easy to interpolate to all intermediate values of J and to extrapolate to higher values of J as well. Thus, the analysis is rather robust to a wide range of *J*s and *SR*s.

No bias is a horizontal line at zero. The *SR*0-line is close to this, but the other four *SR*s, which are for rational researchers, all show a substantial positive bias for $J > 1$.[23] The bias is already about 20% for $J = 2$ and then it rises to about 80% for *SR*1-3 and $J = 50$.

[23] The four curves bend almost as logarithmic curves, but if the vertical axis of Figure 9 is logarithmic, the curves still bend a little, see Paldam (2013a) for *SR*2. As the bend is strongest for small *J*, the simulations for J = 3, 4 and 7 have also been run. They are used in the drawings.

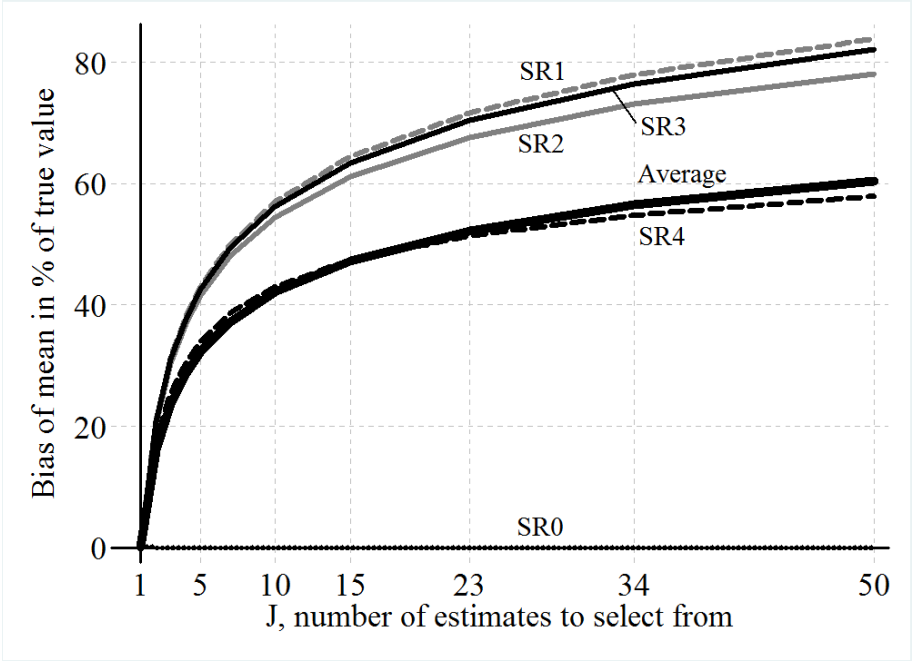


Fig. 9. The paths of the publication bias for the mean in % of $\beta = 1$

Note: The curves are $100(\overline{\overline{b}} - \beta)/\beta = 100(\overline{\overline{b}} - 1)$ from columns (2) in Tables 2 to 6.

The introduction made the assumption that economic theory predicted that $\beta > 0$, and β was accordingly chosen at 1. This gave a bias so that the average result was well above 1. Thus, the simulations show *exaggeration*: Results are more in accordance with author priors and interests than they should be. This has consequences such as: Theory is confirmed too often, sponsors can be rewarded with results they like, etc.[24]

5.2 The second key result: The bias is robust as long as authors are rational

Rational authors look for the optimal combination of the fit and size of the estimate. They produce a bias somewhere between the *SR*1 and *SR*2 lines on Figure 9. As the two lines are so close, it hardly matters at all what the weights are. Consequently, it does not matter if different authors put different weight on fit and size.

However, *SR*0 and *SR*4 are different. It has been argued that *SR*0 must be rare, but maybe *SR*4 is more common. However as long as the two deviating *SR*s are used with a reasonable constant frequency, they have a constant effect.

[24] A search in February 2015 for 'sponsor bias' in Google Scholar gave 136,000 hits. The broader concept of 'publication bias' gave 2.99 million hits. Many of these are to empirical studies, notably in medicine.

From interviews and introspection I believe that the researchers of a typical economics paper use some mixture of the five selection rules. Often they are even mixed in the same paper, and many researchers find it difficult to fully explain the choices made. Perhaps the average of the five *SR*s (shown) may be a realistic guess of the publication bias?

Also, authors use different values of *J*. I assess that the realistic range is between 20 and 40. The average curve on Figure 9 is fairly flat in that range, giving an average publication bias of about 55%. From the meta-studies made so far in economics, this appears on the low side. But then section 2.4 argues that the simulations catch some of the bias only.

The mixture of *J*s in the typical β-literature has another consequence. It makes the form of the funnels somewhat softer. So it is no wonder that the sausage-formed funnels seen on Figure 6, for high values of *J*, do not occur in practice.

5.3 The third key result: The PET reduces the bias by more than 90%

Columns (7) of the five tables show the PET meta-averages. Figure 10 uses an enlarged vertical axis, so there is a bit random variation left. However the paths of the expected estimates are very clear. They show that the PET is much closer to one (β) than is the mean, even in the case of *SR*4. However, the PET is not a perfect estimate of β.

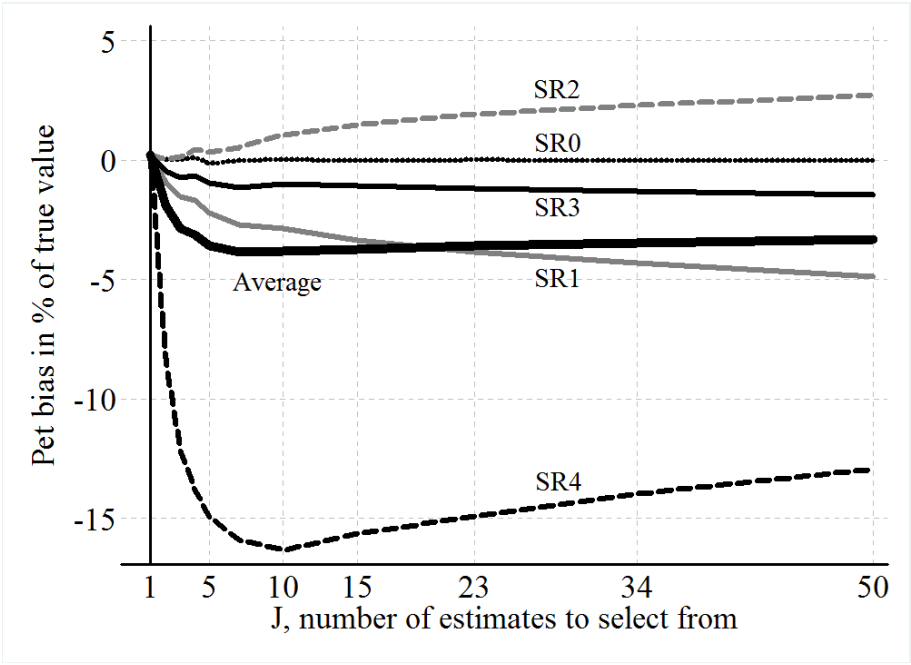


Fig. 10. The paths of the bias of the PET in % of β

Note: The curves are $100(\overline{b}_M - \beta)/\beta = 100(\overline{b}_M - 1)$ from columns (7) in Tables 2 to 6.

*SR*0 has no selection bias, and the *SR*0-curve is horizontal at zero as it should be. In the three cases *SR*1 to *SR*3, the PET bias is rather close to zero. For *SR*4 the PET bias is about 15%. Still, the average result is within 5% of the true value for all *J*s examined. Consequently, the bias of the PET is less than one tenth of the bias of the mean.

The similar graph for the PEESE is much as an average between Figure 9 and 10. For biases due to rationality the PET works much better than the PEESE. For the meta-practitioner it is a problem that empirical funnels are reached by researchers using unknown research strategies, which are almost certainly different. We know from the work referred by T.D. Stanley (*op. cit.*) that the PEESE is a bit better in the case of censoring. Imagine that *c*% researchers censor, while (1 – *c*)% are rational then even for quite large values of *c* such as 75% the PET is likely to get closer to the true value than the PEESE.

5.4 The FAT and the standard deviation

The FAT is known as a fine test. Figure 11 compares the five sets of FAT-results. It should reject asymmetry for $J = 1$ and for all estimates using *SR*0, and detect asymmetry in all other cases. This is precisely what it does. It is interesting that the results are rather similar for all four optimizing *SR*s, and that the highest test-values are for *SR*1.

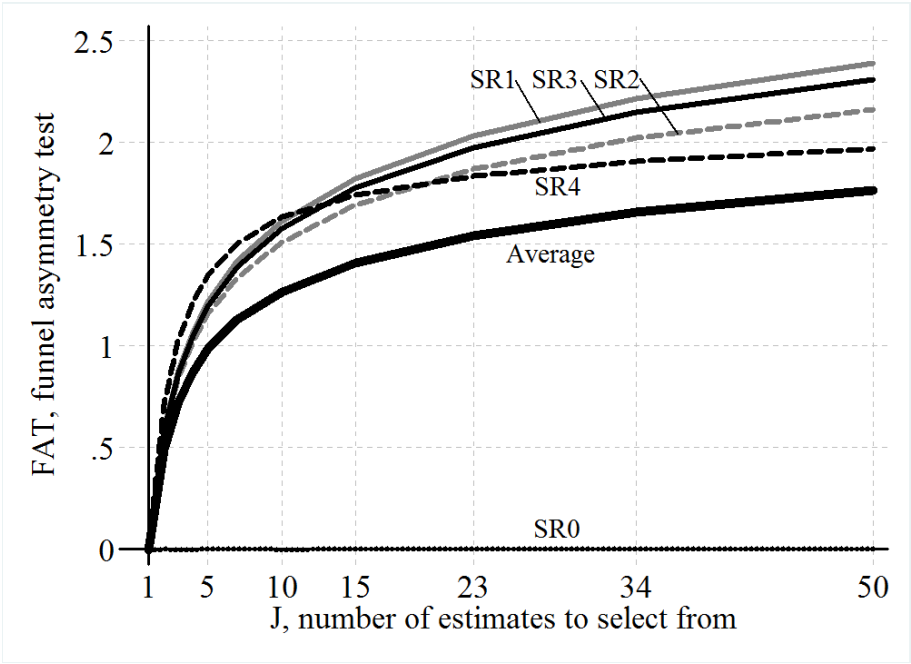


Fig. 11. The FAT

Note: The curves are $\overline{b_F}$ from columns (5) in Tables 2 to 6.

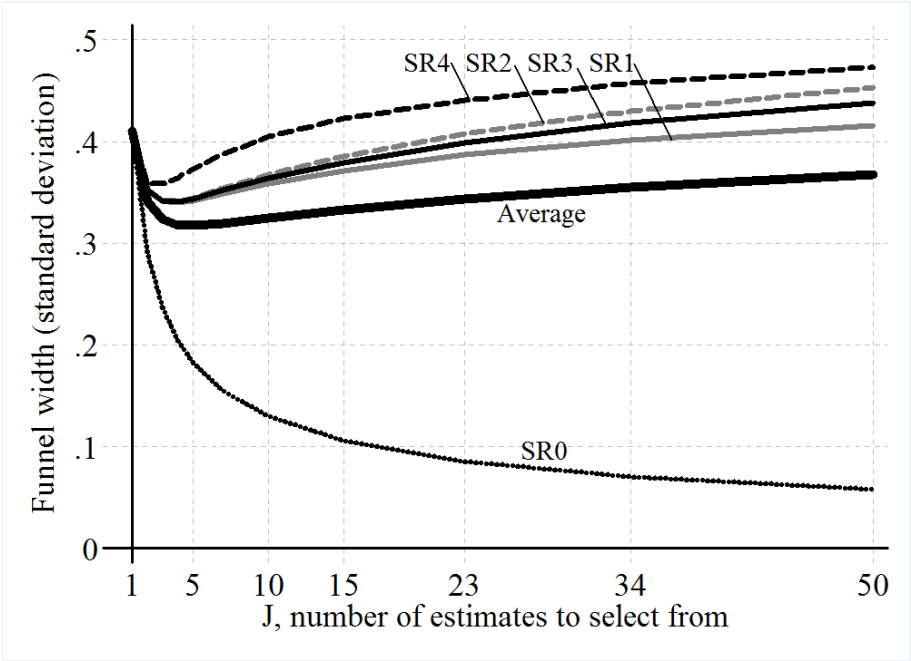


Fig. 12. The effect of the selection rules on ν, the funnel width

Note: The curves are $\overline{\nu}$ from columns (4) in Tables 2 to 6.

One of the most puzzling observations from empirical meta-studies is the amazing widths of funnels. It is analyzed by the variable, ν, in column (3) of the tables. Figure 12 shows the ν-lines for all the SRs. All curves start at 0.412 for J = 1, and then they fall, but only the SR0-curve keeps falling. The other four SRs all turn to rise and become higher than at the start. It happens at the same time as the level of significance rises, and the funnel becomes more asymmetrical. This corresponds to my assessment of the properties of the typical empirical funnel in section 2.2. It is certainly more realistic than the path for SR0.

6. Conclusions on the rational researcher

Economic theory assumes that humans are rational and try to make the best decisions. When the assumption of rationality is applied to economists it allows us to make predictions about a researcher, who try to find the 'best' estimate of a certain parameter.

The paper models the search and selection process for published estimates. Once the data are in the computer the costs of producing estimates, are next to nothing. Thus, most researchers make more estimates than they can possibly publish. The rational choice follows from the familiar framework of a production possibility frontier and a tangential indifference curve. This process is simulated as a function of the number of estimates per published one,

and a set of the most likely selection rules, based on the fit and size of the estimate. Three key results are reached:

The first key result is sad but unsurprising: When the number of regressions per published estimate is larger than one, all rational selection rules produce a bias, which is often substantial. The bias is in direction of the prior for size of the researcher. It is enough, if she prefers a positive result, to get a positive publication bias.

The second key result is more surprising: As long as the selection is based on the fit and size of the estimate, the bias is almost the same irrespective of the weigh the researcher places on the fit and the size. It follows that even if different researchers use different weights it does not matter very much for the bias generated.

The third key result is the good news: When the set of estimates of the same parameter is treated by the tools of meta-analysis it allows us to see if a bias occurs, and reduce the bias by more than 90%.

Acknowledgements: The simulation program is written by Jan Ditzen, who is a great expert in stata simulations. The paper has been presented at a seminar at Deakin University (Melbourne), and at the Danish society for Econometrics Meeting at Sandbjerg Manor. Also, it has been presented as a keynote at the European Public Choice Meeting in Cambridge (UK) and at the Meta-Analysis of Economics Research Network Colloquium in Athens (Greece). I am grateful to the referees and to all who have commented, notably to Chris Doucouliagos, Doug Hibbs, Robert Reed, Jörg Schäpfer and in particular Tom Stanley.

Replication: The stata do file doing the simulations is posted on the URL: http://www.martin.paldam.dk/Simulations.php.

References:

Brodeur A, Lé M, Sangnier M, Zylberberg Y (2013) Star Wars: the Empirics Strike Back. IZA DP No. 7268, Bonn, Germany

Callot L, Paldam M (2011) Natural funnel asymmetries. A simulation analysis of the three basic tools of meta-analysis. Research Synthesis Methods 2: 84–102

Card D, Krueger AB (1995) Time-series minimum-wage studies: a meta-analysis. Am Econ Rev 85: 238-43

Carter JR, Irons MD (1991) Are economists different, and if so why? J of Econ Perspectives 5: 171-77

Doucouliagos H, Paldam M (2008) Aid effectiveness on growth. A meta study. Eur J Pol Econ 24: 1-24

Doucouliagos H, Paldam M (2015) Finally a breakthrough? The recent rise in the size of the estimates of aid effectiveness. For Arvin, M., ed., Handbook on the Economics of Foreign Aid. Edward Elgar

Doucouliagos H, Stanley T.D (2013) Are All Economic Facts Greatly Exaggerated? Theory Competition and Selectivity. J Econ Surv 27: 316-39

Egger M, Smith GD, Schneider M, Minder, C (1997) Bias in meta-analysis detected by a simple, graphical test, British Medical J 316: 629-34

Fanelli D (2010) ‘‘Positive’’ Results Increase Down the Hierarchy of the Sciences. PLoS ONE 5(4): 1-10

Ioannidis J, Doucouliagos C (H) (2013) What's to know about the credibility of empirical economics? J Econ Surv 27: 997-1004

Kirchgässner G, (2005) (Why) are economists different? Eur J Pol Econ 21, 543-62

Marwell G, Ames RE (1981) Economists fee fide, does anyone else?: Experiments on the provision of public goods, IV. J of Public Econ 15, 295-310

McCloskey DN 2nd ed. (1998) The Rhetoric of Economics. University of Wisconsin Press, Madison

Paldam M (2013a) Regression Costs Fall, Mining Ratio Rise, Publication Bias Looms, and Techniques get Fancier. Reflections on some trends in Empirical Macroeconomics. Econ J Watch 10, 136-56

Paldam M (2013b) How do partly omitted control variables influence the averages used in meta-analysis in economics? Economics Working Papers, Aarhus University 2013-22

Paldam M (2015) The rational economist in research: A model. Posted at http://www.martin.paldam.dk/Meta-method.php

Reed WR, Florax RJGM, Poot JA (2014) A Monte Carlo analysis of alternative meta-analysis in the presence of publication bias. Economics e-journal, forthcoming

Stanley TD (2001) Wheat from Chaff: Meta-Analysis as Quantitative Literature Review. The J of Econ Perspectives 15, 131-150

Stanley TD (2008) Meta-regression methods for detecting and estimating empirical effect in the presence of publication bias. Oxford Bulletin of Econ and Stat 70, 103-27

Stanley TD, Doucouliagos H (2010). Picture this: A simple graph that reveals much ado about research. J of Econ Surv 24: 170–91

Stanley TD, Doucouliagos, H (2012) Meta-Regression Analysis in Economics and Business. Routledge, Abingdon (UK)

Stanley TD, Doucouliagos H (2014) Meta-regression approximation to reduce publication selection bias Research Synthesis Methods 5: 60-78

Stanley TD, Doucouliagos H, Giles M, Heckmeyer JH, Johnson RJ, Laroche P, Nelson JP, Paldam M, Poot J, Pugh G, Rosenberger RS, Rost K (2013) Reporting Guidelines for Meta-Regression Analysis in Economics. J of Econ Surv 27: 390-4

Young NS, Ioannidis JPA, Al-Ubaydli O (2008) Why current publication Practices may distort science. PLoS Med 5(10): 1418-22

Ziliak ST, McCloskey DN (2008) The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice, and Lives. University of Michigan Press, Ann Arbor