22/9-2015

Biases in papers by rational economists

The case of clustering and parameter heterogeneity

Martin Paldam, Department of Economics and Business, Aarhus University, Denmark[1]

Abstract: The paper considers the literature on an important parameter. The literature can be modeled and simulated if researchers are rational, so that they behave as predicted by economic theory. Previous work has analyzed the parameter homogeneity case. Heterogeneity means that the variation within papers is smaller than between papers. Three key results generalize from the previous case: (1) Rationality biases the published estimates substantially in the direction of the priors of the researcher. (2) The bias is robust to all rational selection rules the researcher may use. (3) The PET estimate of the meta-average reduces the bias by more than 90%. The present case gives two additional results (4) It does not matter for the bias whether 1 or 10 estimates are published. What matters is the number of estimates for each published. (5) Parameter heterogeneity makes funnels more top-heavy, and increases the variation somewhat.

Keywords: Meta-analysis, selection of regressions, publication bias

JEL: B4, C2

Acknowledgements:
The simulation program is written by Jan Ditzen. It is available at the URL: http://martin.paldam.dk/Simulations-2.php. The theory of the rational economist is given in Paldam (2015c). The model originated in Paldam (2013) that gives a theory of the number of estimates per published result, *J*, based on the costs and benefits of the marginal regression. *J* is assumed to be exogenous in the present paper. Paldam (2015b) gives a detailed presentation of the simulation framework used at present. Presented at the MAER-net Colloquium in Prague, September 2015, I am grateful for the comments.

---

1. Street address: Fuglesangs Allé 4, Building 2632 (L), DK-8210 Aarhus V. E-mail: mpaldam@econ.au.dk.
URL: http://www.martin.paldam.dk.

1. Modeling the rational economist

This paper simulates the research of economists, who behave as predicted by economic theory. In the case analyzed they try to find the best estimate of a parameter β, where 'best' is defined in the two dimensions of size and fit (t-ratio) that all studies report. The parameter is important so *M* researchers have produced a β-literature with *N* estimates. For ease of presentation it is assumed that each of the *M* researchers publish one paper with 10 estimates. They show that the result is robust.

It is assumed that economics has a basic theory about β. However, the theory is qualitative and allows many variants of the estimation model. Also, different data samples and estimators can be used. Thus, estimates of β have a substantial ***PPS***, production possibility set. The ***PPF***, production possibility frontier, is the part of the rim of the *PPS* where production is effective, so that an increase in one of the two dimensions causes a fall of production in the other.

Each author searches only some of the *PPS* as he looks at a subset of models and one data sample, but his *PPS* is still sizable, and it has a *PPF*. Economic theory says that the estimate chosen by a rational researcher is the one where his utmost ***IC***, indifference curve, touches the *PPF*. Thus, papers differ both by the *PPF* and the *IC*s of the author.

The *M* papers are presented in the usual way: (i) First a theory is developed. It is normally the basic theory with a twist. (ii) It is turned into an estimation model. (iii) A search is made of versions of the model on a data sample. (iv) The ten best estimates are selected for publication.

This process is modeled and simulated as follows: (i) The theory is a data generating process, ***DGP***; (ii) The estimating model, ***EM***, is the same as the *DGP*; (iii) The DGP/EM-pair generates estimates. For each of the 10 published results, *J* estimates are searched by the researcher. Their efficient rim is the production possibility frontier, ***PPF***; (iv) The researcher's selection rule, ***SR***, represents his indifference curves, ***IC***, which are a function of the fit and size of the estimates; (iv) The *PPF* and *IC*s give one published estimate.

The solution is a function of five variables: They are: *J*, *SR* and three variations (σ_β, σ_x, σ_ε) in the stochastic terms generating the data sample. Each choice of these five variables gives a solution. Even when a very parsimonious DGP is chosen, it still has five variables to vary in order to map the pattern in the solutions. Many simulations of each case are necessary for the pattern to stabilize on the equilibrium/expected results. Fortunately, the results have a simple pattern, which is easy to interpolate. Eight values are used for *J* and three for the *SR*. For the three variations a central case is chosen. It is supplemented by a low and a high variant.

The simulated β-literature is analyzed by the tools of meta-analysis, developed precisely for that purpose. A publication bias is defined as a systematic difference between the published estimates of β and the true value. It is due to the rational behavior of researchers, so it is a *rationality bias*. The meta-analysis estimates two meta regression analyses, ***MRA***s, to retrieve the true value. The simulations show that one of these MRAs (the FAT-PET) finds and corrects most of the rationality bias. The other (PEESE) works less well.

This set-up was already used in Paldam (2015b), which studied the pattern in the selected result for eight values of *J* and five values of the selection rule, *SR*. The analysis used many simplifications. Two of the most problematic were the assumptions of independent estimates and parameter homogeneity, modeled by assuming that each paper publishes one independent estimate of $\beta = 1$.

The present paper relaxes these assumptions. The estimates are clustered in papers with 10 estimates in each. An exogenous variation is added to β, so that $\beta = N(1, \sigma_\beta^2)$ where each paper uses one draw from the β-distribution.[2] Thus, papers differ and the variation of the estimates is smaller within papers than between papers.

However, the analysis still confirms the three main results of the previous study: (r1) All rational selection rules produce a substantial publication bias as soon as $J > 1$. (r2) The range of rational selections gives much the same bias. It is always in the direction of the main priors of the researchers. (r3) The PET meta-average corrects more than 90% of the bias.

In addition two new results appear. (r4) It does not matter for the bias if one or ten estimates are published – what matters is the number of estimates *J* per published one. (r5) The extra variation when $\sigma_\beta \neq 0$ makes the funnel more top heavy and causes the meta-analysis to reject that $\beta = 1$ in more cases. Section 2 explains how the simulations are set up. Section 3 looks at the rational researcher. Section 4 gives detailed results in the central case. The results are compared in section 5. Section 6 analyzes a low and a high variant relative to the central case for the three variances. Finally, section 7 concludes.

2. The set-up means that if σ_β is set at zero, β is the same for all estimates both within and between papers, i.e., there is no paper-structure in the simulations, and everything is as in Paldam (2015b), which is thus the limiting case for $\sigma_\beta \rightarrow 0$.

2. The set-up of the simulation experiments

The reader should note three numbers:

N = 500 is the number of 'published' estimates of β in the simulated β-literature. The *N*-set comes from *M* = 50 'papers' with *K* = 10 estimates in each.

J = 1, 2, 5, 10, 15, 23, 34 and 50 (where $\sum J = 140$) are the number of estimates made for each 'published'. In fact, three *N*-sets are selected by the selection rule *SR*0, *SR*1 and *SR*2.[3]

R = 1,000 or 10,000 is the number of experiments made for a production run. Each such run produces 8·3 =24 *N*-sets, which is one for each value of *J* and *SR*.

Section 2.1 describes the simulations. Section 2.2 looks at the results retained for each *N*-set. *R* is crucial for the computer time necessary to reach the three tables. Section 2.3 discusses how large *R* needs to be. Finally, section 2.4 is an introduction to the meta-analysis used to study the *N*-set.

2.1 The DGP/EM set-up and the five variables: J, SR, σ_β^2, σ_x^2 and σ_ε^2

The DGP/EM pair is the same, and it has only a β-term:

(1a) DGP: $y_t = \beta\, x_t + \varepsilon_t$, where $\beta = N(1, \sigma_\beta^2)$, $x_t = N(0, \sigma_x^2)$ and $\varepsilon_t = N(0, \sigma_\varepsilon^2)$.

(1b) EM: $y_t = b\, x_t + u_t$, estimated by OLS.

The three standard deviations in the DGP are: σ_β, σ_x and σ_ε. One value of β is generated for each paper, while one value of σ_x and σ_ε are generated for each observation in the data sample used to estimate $b \approx \beta$. The central case is $(\sigma_\beta, \sigma_x, \sigma_\varepsilon) = (0.3, 2, 10)$. It is the same values for σ_x and σ_ε used in (Paldam 2015), where the two variations were chosen to generate realistically looking distributions (funnels) of the estimates. Section 6 gives the experiments with variations around these values.

In the DGP/EM all control variables are treated as parts of the noise terms. Therefore, σ_ε is quite large. The expected value of β is 1, thus, the theory is true, even when β varies between papers**.** The new parameters in this paper are *K* = 10 and σ_β. To see the effect of σ_β, it is also made rather large, $\sigma_\beta = 0.3$, relative to the expected mean $\beta = 1$. It gives a 95% confidence interval of (0.4 to 1.6) for the βs. In addition to the choice of the three standard deviations, two more choices are made, *SR* that is discussed in section 3.1 and *J* defined above.

3. *J* is taken to be exogenous to the variables discussed as it is determined by the marginal costs and benefits of making regressions. The marginal costs of a regression are very small and the benefits substantial, so many are made.

One *N-set* mimics a β-literature of 500 'published' estimates, where each published estimate is selected from J estimates made. For each estimate a new data sample is made, where each observation uses a new draw of $N(0, \sigma_x^2)$ and $N(0, \sigma_\varepsilon^2)$.

For each paper β is drawn once from $N(1, \sigma_\beta^2)$. This creates a cluster-dependency within the *N*-set, where the 10 estimates in a paper is a *cluster*. In empirical bodies of literature each paper typically works with the same data sample, and a 'family' of models, which are a sub-set of the larger class of β-models. Also, the author has the same priors throughout the paper. Thus the 500 estimates within each *N*-set have 50 values of β that are distributed around 1, with a standard deviation of σ_β. The clustering in papers gives the *N*-set a panel structure, with less variation within the papers than between the papers. Various ways have been developed to handle this, but the results are not very sensitive to these methods, so I just stick to a simple fixed-effect framework.

2.2 R experiments: Increasing R till the pattern in the results becomes smooth

One simulation experiment covers 24 = 3·8 *N*-sets, which is one for each of the three *SR*s, and each of the eight values of J. It means that $\sum J = 140$ regressions are made. Thus, one experiment requires $500 \cdot 140 = 7 \cdot 10^4$ simulated regressions.

The key result is the pattern in the expected/equilibrium results to which the average results converge. *R* experiments are made to study the pattern in the following six result-variables: b, t, v, β_F, β_M and β_P – defined in Table 1 – when the five variables J, *SR*, σ_β, σ_x and σ_ε vary.

The experiments generate averages of averages for b and t that soon become rather stable, but it turns out to be more difficult to get stable averages for the meta-averages, notably the PET, β_M. It is close to 1 so it gives a small range of variation. This allows an enlargement of the range, so that even a small variation becomes visible, see Figure 6 below.

Below I run J = 1, 2 and 5 with R = 10,000. It gives $(1 + 2 + 5) \cdot 500 \cdot 10{,}000 = 4 \cdot 10^7$ regressions and J = 10, 15, 23, 34 and 50 with R = 1,000. It gives $(10 + 15 + 23 + 34 + 50) \cdot 500 \cdot 1{,}000 = 6.6 \cdot 10^7$ regressions, so the sum is $1.06 \cdot 10^8$ regressions. The variability experiments described in section 6 are run for R = 500. They add $140 \cdot 6 \cdot 500 \cdot 500 = 2.1 \cdot 10^8$ regressions, so everything sums to $3.16 \cdot 10^8$ regressions. This takes about one month of computer time.[4]

2.3 The meta-analysis: Analyzing N-sets of 500 selected estimates

The technique of meta-analysis is developed precisely to analyze *N*-sets. The technique is

4. I have used 2 fast pcs running for 3 weeks, which include the calibration of the models.

thoroughly covered by the recent textbook Stanley and Doucouliagos (2012).[5] About 750 empirical meta-studies have been made in economics.

A meta-study covers an *N*-set of empirical estimates that pertains to be of the same parameter, even when we suspect that there is some parameter heterogeneity in practice. They are from a set of papers that are so similar as regards the estimating models that the differences can be coded.[6] The average meta-study covers about 50 studies, where each reports about 10 estimates. Thus, these numbers are used in the simulations, so the typical empirical *N*-set can be used to calibrate the simulations, as has been done.

The data for the analysis is the *N*-set: (b_i, t_i, s_i, p_i).The b_i's are rescaled to the same scale. As $t_i = b_i/s_i$ is a ratio, with no unit of measurement, s_i and $p_i = 1/s_i$ are automatically rescaled when the *b*'s are. When the literature has been collected and coded, it is analyzed in two ways:

(i) Graphically by displaying the distribution of the *N*-set as the (p_i, b_i)-scatter that is termed the funnel. It should have a broad base for low precision and a narrow top for high precision. Three specimens are shown in Figures 1 and 2.

Table 1. The 5 variables analyzed: *J*, *SR*, σ_β, σ_x and σ_ε

Variable	Explanation (possibilities)	One published	One *N*-set	*R* experiments
Mechanics of search and selection of the publishes estimates , see section 3.1				
Regressions	Number made and searched	*J*	*J*·500	*R*·140·500
Numbers searched	*J* = 1, 2, 5, 10, 15, 23, 34 or 50	one chosen	one	all 8
Selection rule	*SR* = *SR*0, *SR*1 or *SR*2	one chosen	one	all 3
Published	Each *J* and *SR* selects one	1	500	*R*·24·500
Descriptive analysis of the estimates				
Estimate	Coefficient and t-ratio	(b, t)		
Mean	Unweighted arithmetic		$\bar{b}$	$\bar{\bar{b}}$
Mean *t*-ratio			$\bar{t}$	$\bar{\bar{t}}$
Meta-analysis of each *N*-sets, see section 2.4				
Funnel	The distribution of the *N*-set		Graph	
Std of *N*-set	Funnel width		v	$\bar{v}$
The FAT	Funnel asymmetry test		β_F	$\bar{\beta}_F$
The PET	Meta average to catch β		β_M	$\bar{\beta}_M$
The PEESE	Meta average to catch β		β_P	$\bar{\beta}_P$

Note: The search mechanism has two parameters: *J* the number of estimates made per published one and *SR* the selection rule that picks one. Note that the sum $\sum J = 140$. A paper uses the same data sample. The selection process is run *10* times, so 10·*J* regressions are made and 10 estimates are published.

5. A short introduction to the technique is found in Paldam (2015a).
6. Estimating models are much more similar than theoretical models, so typically only a few papers have to be left out.

(ii) Statistically, by estimating the statistics listed in Table 1. The special meta-estimates are two, MRAs, which are regression on regression estimates: (1) the ***FAT-PET*** and (2) the ***PEESE***:

(1) $b_i = \beta_M + \beta_F s_i$, where β_F is the FAT and β_M is the PET.[7]

(2) $b_i = \beta_P + \beta_F s_i^2$, where β_F is an alternative FAT, and β_P is the PEESE.[8]

The H0 of no asymmetry is $\beta_F = 0$. In that case $\beta_M \approx \beta_P \approx \bar{b}$. If $\beta_F \neq 0$, $\bar{b}$ differs from both β_M and β_P, which are the values that the two MRAs converge to when s_i goes to zero.

Obviously β_M and β_P should be good estimates of the true value of β. This is the way they have been treated by the meta-community: They are taken to be two estimates of the *meta-average*. The difference between (1) and (2) is the speed of convergence. Stanley (2008) considers the case where the funnel is fully or partly censored and shows that the FAT-PET works rather well to catch β. Later Stanley introduced the PEESE and Stanley and Doucouliagos (2014) showed that the PEESE is even closer to β than the FAT in that case.

As mentioned, rational researchers create biases too. *Rationality biases* are different from censoring biases.[9] In Paldam (2015b) the FAT-PET MRA did catch most of these biases, while the PEESE did less well, though better than the mean. It will be shown that this result generalizes also in the present case. In about 2/3 of empirical meta-studies the FAT detects asymmetry, and it is often obvious from a visual inspection of the funnel. If the study covers, e.g., 50 primary studies, it is unlikely that the research process has been the same in all studies, and we rarely have more than a vague knowledge of the research processes leading to the asymmetry, anyhow. The present study examines the possibility that researchers are rational.

Thus, we have to admit that the conditions for applying the MRAs of meta-analysis are unlikely to hold strictly. However, till now all simulations where the true value of β is known have shown that the PET is (much) closer to β than is the mean. It is important to study the properties of these estimators in a range of circumstances to look for cases where they are safe to use. This is what is done in the rest of the paper.

7. The FAT-PET MRA is from Stanley (2008). The acronym is for Funnel Asymmetry Test and Precision Estimate Test and Meta Regression Analyses

8. The PEESE MRA is from Stanley and Doucouliagos (2007 and 2014). The acronym is for Precision-Effect Estimate with Standard Errors.

9. It is not irrational to discard results that must be *wrong*, but it is not at the same level of rationality as a full optimization as discussed in the rest of the paper.

3 The choices of the researcher

Section 3.1 makes the claim that rational researchers reach much the same solutions. Section 3.2 considers the choice of *J*, while the next sections turn to *SR*: Section 3.2 discusses the altruistic choice of truth only, while section 3.3 turns to the range of rational *SR*s. Section 3.4 looks at the ideal funnel and empirical funnels.

3.1 The claim that different researchers make similar choices

It is a widespread belief that if a group of, say, 50 independent researchers with different utility functions and interests study the same problem, truth will result. Meta-analysis contradicts that view by showing that most empirical literatures have biases. This must mean that a mechanism exists to make indifference curves and the choices they generate more alike than this belief assumes. I discuss two such mechanisms:

(Mec 1) Researchers have to 'sell' their projects to sponsors and their papers to journals. Therefore, they will have to fulfil the expectations of sponsors, referees and editors. The rational author tries to internalize the preferences of these players. Thus, the market homogenizes the preferences of the rational researchers.

(Mec 2) Estimates published in economics are (nearly) always given by a size and a fit, i.e., a *t*-ratio or a standard error. This must mean that the size and the fit are seen as the two key dimensions in the evaluation of the results, i.e., in the utility functions of authors.[10] Rational researchers are likely to have different trade-offs between the fit and size of estimates. A key finding below – demonstrated in sections 4.4, 4.5 and discussed in 5.1 and 5.2 – is that all rational choices give amazingly similar results.

The two mechanisms are independent of each other. This suggests that the combined effect is quite strong.

3.2 The choice of J

The choice of the researcher is modeled by two variables: The number *J* of regressions per published one and the selection rule SR.

The rational choice of *J* is where the marginal cost of an extra regression equals the margi-

10. Other variables may enter also, but the paper simplifies and assumes that the two variables are the only variables in the utility function. This is a main reason why the DGP/EM is so simple.

nal benefit.[11] On the marginal costs side it is clear that once the data is in the computer, the marginal costs of another regression is next to nothing. On the marginal benefit side, it is also clear that fine empirical results increase the publication chance for the paper. As publications are crucial for the career of the researcher, the value is quite high. It is easy to reach intersection points such as J = 200. The only way to get lower values is to assume that the marginal benefits fall rapidly after some point such as J = 15. And as 10 estimates have to be chosen, it is possible that J per estimate stays under 50.

Below I use J = 1, 2, 5, 10, 15, 23 34 and 50. I am confident that the upper end of this scale is more realistic. Fortunately, the results are rather stable for high values of such as J = 23 to 50. However, they move a lot for low values, so here the *J*s are closer together.

It is assumed that the decision about J is already made when I turn to the indifference curves of the researchers. They are modeled by the variable, *SR*, which is the rule that selects one 'published' result from the J estimates. Three SRs are considered: One altruistic and two rational.

3.3 The baseline of an altruistic researcher, who looks for truth only

The choice of a researcher who has a prior for truth and no other prior will chose *SR*0:

*SR*0: The researcher selects the best estimate of the expected value of β. As the J estimates are the best ones the researcher has thought of, *SR*0 is the mean $\bar{b}_J$ of the J estimates. It is also the expected value if the researcher decides to run one more regression.

If truth is revealed in the long run, it may be rational – with that time horizon. In the meantime the researcher stands out as a guy who finds smaller and less significant estimates than other researchers who have estimated β and who consequently are likely referees. They will not like the paper of the *SR*0-researcher, and mot editors will their referees. Sponsors, who want large coefficients, will look for more rational researchers. Hence, the truth seeker will disappoint both university appointment committees, who want publications, and administrators, who want sponsor money they can tax. This will have a negative impact on his career, so truth seeking is indeed altruistic.

Economists acknowledge the existence of altruism, but normally it is found to be a minor factor in the behavior of people, and a dozen studies of the behavior of economists, see Kirchgässner (2005) for a survey, find that economists behave more in accordance with economics than other people. Also, Table 2 in section 4.2 shows that *SR*0 leads to results that are at variance

11. A crude attempt to evaluate the marginal costs and benefits are made in Paldam (2015c).

with the empirics as discussed.

3.4 *The two ends of the range of rational selections*

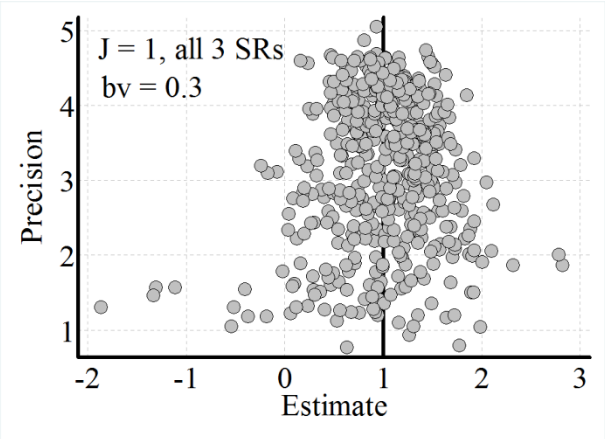
The other two SRs are extreme versions of the rational ones that follow from two main priors of the researchers:

*SR*1: The first main prior is for *fit*. Researchers want the fit of the preferred estimate to be high, so the selection rule *SR*1 maximizes the *t*-ratio.

*SR*2: The second prior is for *size*. Economic theory, moral/political beliefs and sponsor interests are taken to aggregate into a prior for size, so *SR*2 maximizes the size of the estimate.

Most researchers want a compromise between the two *extreme SR*s. They are willing to trade some size for more fit and vice versa. So the two *SR*s give the limits of the rational choices. A key result in Paldam (2015b) is that the two *SR*s give similar results for the relevant statistics. This result generalizes below. The rationality bias is robust to all rational selection rules.

Figure 1. The ideal funnel, with J = 1, in the central case



Note: One of the 10,000 funnels in row (1) of Tables 2 to 4, which is the same.

3.5 *How should the simulated funnels look to be realistic?*

In simulations it is easy to generate ideal funnels where no censoring takes place. This happens when all estimates are published, i.e. for J = 1. An ideal funnel is symmetric and as wide as can be predicted from the average t-ratio of the estimates. Figure 1 is the ideal funnel in the central case. Rows (1) in Tables 2-4 are the statistics for 10,000 repetitions of this funnel. Hence, row (1) is the same in the three tables, but as J rises the tables diverge.

The funnel looks as the corresponding funnel in Paldam (2015b, Figure 5a), but the parameter heterogeneity makes the funnel a bit more top-heavy, and the extra variation of β also gives a little more variation. I have looked at many empirical funnels. My impression is that they have four properties:

(p1) The estimated b's are normally rather significant, i.e., t-ratios are high.

(p2) The funnel is often amazingly wide given the level of precision. One explanation of the width is the variability of β as modeled. We measure the funnel width as the standard deviation, v, of the N-set.

(p3) About 2/3 of all empirical funnels are asymmetric – this caused the various averages to differ. If the estimates deviate due to random noise – including the noise in β – the funnel should be symmetrical, and in that case 'all' averages are the same. Consequently, an explanation of the asymmetry is needed.

(p4) They normally have a narrower top than Figure 1. Thus, it appears that the choice of σ_β is a rather large one.

Note that (p2) and (p3) are combined in practice – empirical funnels are often wide and asymmetrical at the same time. The analysis below shows that the rational behavior or researchers generate funnel asymmetries at the same time as funnels keep their width.

4. Results: The central case

In the central case $(\sigma_\beta^2, \sigma_x^2, \sigma_\varepsilon^2) = (0.3, 2, 10)$ as already mentioned. Section 4.1 explains the format of the tables in the next 3 sections that cover the three *SR*s.

4.1 The format of Tables 2 to 4 reporting the average results

The expected/equilibrium values of the statistics reported are provided with a star ‘*’. Due to the many replications of the *N*-set, the averages are close to the expected values.

The tables have nine rows: The first eight rows are for one *J*. The first row for $J = 1$ is always the same, as only one estimate can be selected when $J = 1$. When *J* increases, the three tables diverge. Row nine is an average giving a (crude) estimate of the outcome when authors use different *J*s.

The tables have ten columns: Column (1) is the *J*-value; column (2) is the mean, $\bar{\bar{b}} \approx b^*$; and column (3) is $\bar{\bar{t}} \approx t^*$. These means are averages of averages (i.e. over all $500 \cdot R$ estimates) so they are very stable. The remaining statistics are averages of the *R* estimates only: (4) holds the average width that is also rather stable, $\bar{v} \approx v^*$, of the funnel.

The FAT-PET MRA is reported in columns (5) to (8). Here (5) is the average estimated FAT, $\bar{b}_F \approx \beta_F$, while (6) counts how often the FAT rejects symmetry; (7) is the average estimated PET meta-average, $\bar{b}_M \approx \beta_M$, while (8) counts how often b_M differs from 1, so that the PET fails to find the true value. While $\bar{b}_F$ is so stable that we can trust that $\bar{b}_F = b_F^*$, the PET still has some variation, as discussed in section 5.4.

The PEESE MRA is reported in columns (9) and (10). Column (9) gives the average meta-average, $\bar{b}_P \approx \beta_P$, 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 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 react. However, both MRAs have been applied on empirical funnels generated by research processes that are likely to be (strongly) affected by rationality.

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

The baseline is the *SR* of the researcher with a prior for truth only. The results are reported in Table 2. Column (5) shows that the average funnel in the table is symmetrical, so that all three averages $\bar{\bar{b}}$, $\bar{b}_M$ and $\bar{b}_P$ should be 1, as is actually the case. Note also that the average t-ratios stay constant at 3.15. When *SR*1 and *SR*2 are used $\bar{\bar{t}}$ rises with *J*.

The counts in columns (6), (8) and (10) are all between 0.45 and 0.75, thus the FAT often rejects symmetry and the FAT and the PET fail to find the true value. This is due to the variation in *β*. When σ_ε goes to zero, the three rejection rates fall to 0.05 as shown in Paldam (2015b). I return to this point in section 5.4.

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

	(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*	$\bar{\bar{b}}$	$\bar{\bar{t}}$	$\bar{v}$	$\bar{b}_F$	Not 0	$\bar{b}_M$	Not 1	$\bar{b}_P$	Not 1
(1)	1	1.001	3.152	0.504	0.001	0.729	1.001	0.631	1.001	0.574
(2)	2	1.000	3.151	0.414	0.002	0.667	1.000	0.573	1.000	0.522
(3)	5	1.000	3.150	0.348	-0.004	0.616	1.001	0.531	1.001	0.483
(4)	10	0.998	3.146	0.322	-0.012	0.604	1.003	0.517	1.000	0.454
(5)	15	0.999	3.148	0.315	-0.014	0.601	1.004	0.528	1.001	0.495
(6)	23	1.000	3.151	0.307	-0.005	0.574	1.002	0.483	1.001	0.443
(7)	34	1.000	3.149	0.305	0.002	0.588	0.999	0.499	1.000	0.450
(8)	50	1.001	3.153	0.303	-0.001	0.556	1.001	0.451	1.001	0.405
(9)	Avr.	1.000	3.150	0.352	-0.004	-	1.001	-	1.001	-

Note: *R* = 10,000 in rows (1) to (3) and *R* = 1,000 in rows (4) to (8).

Column (4) reports the funnel width. It is a function of the variation in σ_ε and σ_β, and the selection rule:

(3) $v = v(J, \sigma_\varepsilon, \sigma_\beta, SR_i)$, where *i* = 0, 1 and 2

It is not easy to solve (3) analytically for *SR*1 and *SR*2, but it can be solved for *SR*0. Paldam (2015b) considers the case $\sigma_\beta = 0$, and shows that the width $v(J) = v(1)/\sqrt{J}$. This gives a rather fast convergence for *J* rising. When $\sigma_\beta > 0$, it enters as a minimum, to which the funnel width converges.

Thus, I expect that $v(J)$ converges from $v(1)$ to σ_β, as indeed it does in column (4) of Table 2. See also section 6.2.

Thus, the selection rule *SR*0 causes the funnel width to fall as J increases. For small values of σ_β, it even gets close to zero. This is not what we observe – refer to point (p2) in section 3.2. I conclude that few economists are as altruistic as assumed by *SR*0.

4.3 SR1, selection by fit only

Selection by t gives the results reported in Table 3. The average b in column (2) and t in column (3) both increase by J, but their ratio stays constant. The funnel width, v, stays almost constant.

The FAT increases from 0 to more than two. Thus, the funnel moves to the right and becomes more and more skew, while keeping its width. The change in the funnel as J raises is illustrated by comparing Figure 1 (for $J = 1$) to Figure 2a (for $J = 10$) and to Figure 2b (for $J = 50$). The parameter heterogeneity changes the funnels a little. It becomes a bit more top-heavy and the tail to the right becomes a bit less pronounced than in the case of $\sigma_\beta = 0$.

The key observation is that the PET works remarkably well as reported in column (7): It is always within 5% of the true value of β. This is much better than the PEESE that is 'only' half of the way between the mean and the true value. Once again it should be noted that both meta-averages reject the true value surprisingly often.

Table 3. Selection rule *SR*1, the best fit

	(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	$\bar{\bar{b}}$	$\bar{\bar{t}}$	$\bar{v}$	$\bar{b}_F$	Not 0	$\bar{b}_M$	Not 1	$\bar{b}_P$	Not 1
(1)	1	1.001	3.152	0.504	0.001	0.729	1.001	0.631	1.001	0.574
(2)	2	1.204	3.730	0.458	0.572	0.295	0.996	0.599	1.105	0.202
(3)	5	1.414	4.351	0.446	1.173	0.013	0.993	0.573	1.213	0.009
(4)	10	1.541	4.741	0.457	1.553	0.002	0.989	0.567	1.277	0.001
(5)	15	1.609	4.953	0.466	1.754	0.000	0.989	0.577	1.312	0.000
(6)	23	1.674	5.163	0.476	1.965	0.000	0.985	0.522	1.343	0.000
(7)	34	1.729	5.338	0.489	2.152	0.000	0.978	0.541	1.368	0.000
(8)	50	1.780	5.509	0.500	2.312	0.000	0.978	0.514	1.393	0.000
(9)	Avr.	1.494	4.617	0.475	1.435	-	0.988	-	1.251	-

Note: Generated for the same regressions as Table 2.

Figure 2a. Funnel SR1, with J = 10

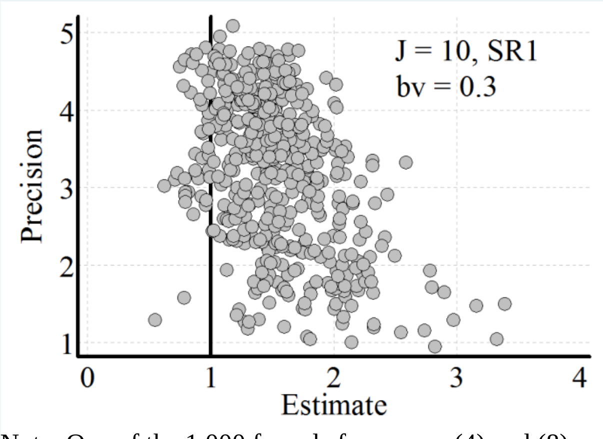
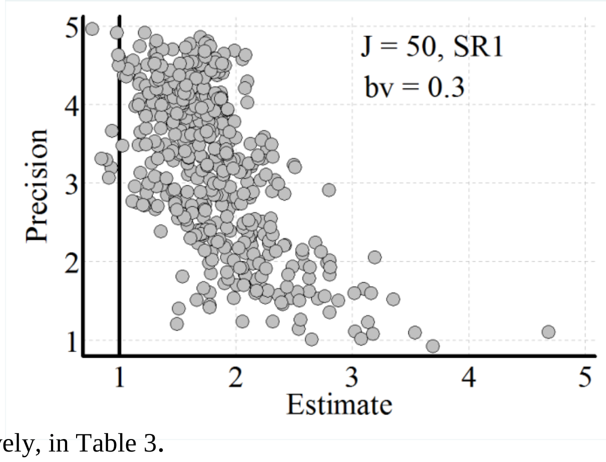


Figure 2b. Funnel SR1, with J = 50



Note: One of the 1,000 funnels from rows (4) and (8), respectively, in Table 3.

4.4 *SR2, selection by size*

The second selection rule is by size alone. The results are reported in Table 4. They are almost the same as Table 3. As the selection is by size, the average mean raises a little more when J rises, and conversely the average *t*-ratio rises a little less. The effect of changing from *SR*1 to *SR*2 is surprisingly small. This will be further analyzed in section 5. The key result is, once again, that the PET works well. It is much better than the PEESE. The funnels from *SR*2 look so similar to the ones from *SR*1 that they are not worth showing.

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

	(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	$\bar{\bar{b}}$	$\bar{\bar{t}}$	$\bar{v}$	$\bar{b}_F$	Not 0	$\bar{b}_M$	Not 1	$\bar{b}_P$	Not 1
(1)	1	1.001	3.152	0.504	0.001	0.729	1.001	0.631	1.001	0.574
(2)	2	1.209	3.717	0.459	0.569	0.289	1.000	0.600	1.109	0.186
(3)	5	1.429	4.311	0.452	1.148	0.012	1.006	0.572	1.229	0.005
(4)	10	1.567	4.675	0.468	1.498	0.001	1.013	0.561	1.305	0.001
(5)	15	1.642	4.870	0.483	1.680	0.000	1.019	0.565	1.348	0.000
(6)	23	1.716	5.060	0.501	1.867	0.000	1.021	0.543	1.389	0.000
(7)	34	1.778	5.217	0.520	2.030	0.000	1.021	0.539	1.422	0.000
(8)	50	1.838	5.369	0.539	2.162	0.000	1.028	0.492	1.457	0.000
(9)	Avr.	1.522	4.546	0.491	1.370	-	1.013	-	1.282	-

Note: Generated for the same regressions as Tables 2 and 3.

By comparing Tables 3 and 4 it appears that the two selection rules, *SR*1 and *SR*2, produce results that are quite similar. In practice the reasonable researcher uses some compromise between *SR*1 and *SR*2, which is a weighted sum of the two choices. The fact that they are so similar means that it matters little if researchers use different weights, as further discussed in section 5. This is the mechanism (Mec 2) already announced in section 3.1.

5. Comparing results

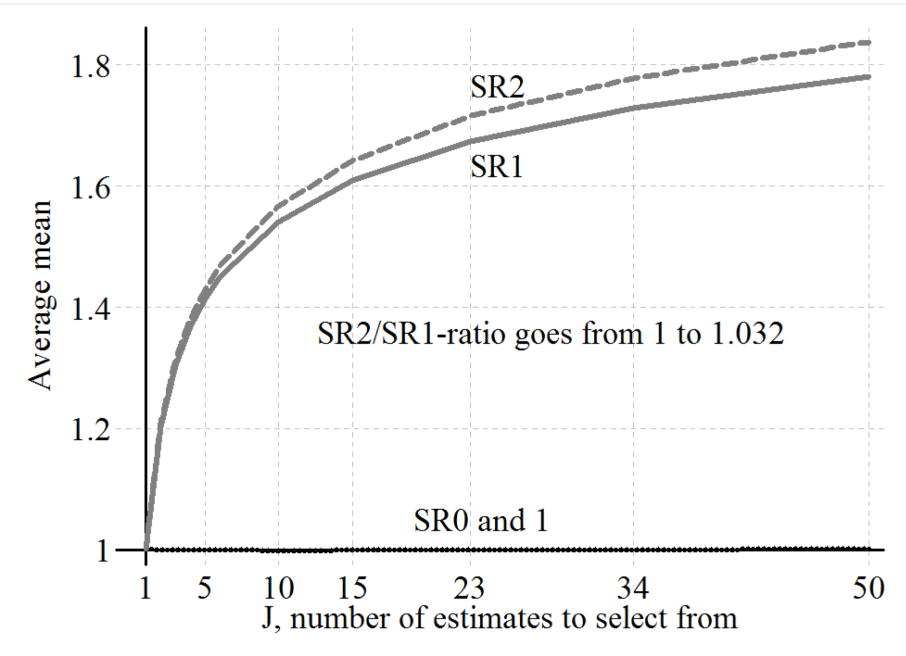
Section 5.1 compares the bias in the mean – that is the publication bias – while section 5.2 looks at the average estimate of the t-ratio. The bias in PET meta-average is covered in section 5.3. Section 5.4 looks at the width of funnels. Note that the figures contain simulations for J = 3, 4 and 6 not reported in the Tables.

5.1 The extreme choices: A preference for size only and for fit only

Columns (2) in Tables 2 to 4 give the estimates of the mean as a function of J. The three columns are drawn as curves on Figure 3. Thanks to the large number of replications the curves are smooth. Thus I conclude that they are close to the expected values b^*. When the true value $\beta = 1$ is subtracted a precise estimate of the publication bias appears.

The *SR*0-curve is close to the horizontal axis at 1, indicating no publication bias as expected. But, the two extreme rational selection rules, *SR*1 and *SR*2, give bias as soon as $J \geq 2$. As expected it is always positive. It confirms the well-known results that the bias is in the direction of the prior. The bias reaches 40 % already for J = 5, and then it gradually goes to 80%. As *SR*2 chooses the largest estimate, the *SR*2 curve is at top, but the gap between the *SR*1 and *SR*2 is small.

Figure 3. The paths of the publication bias for the mean

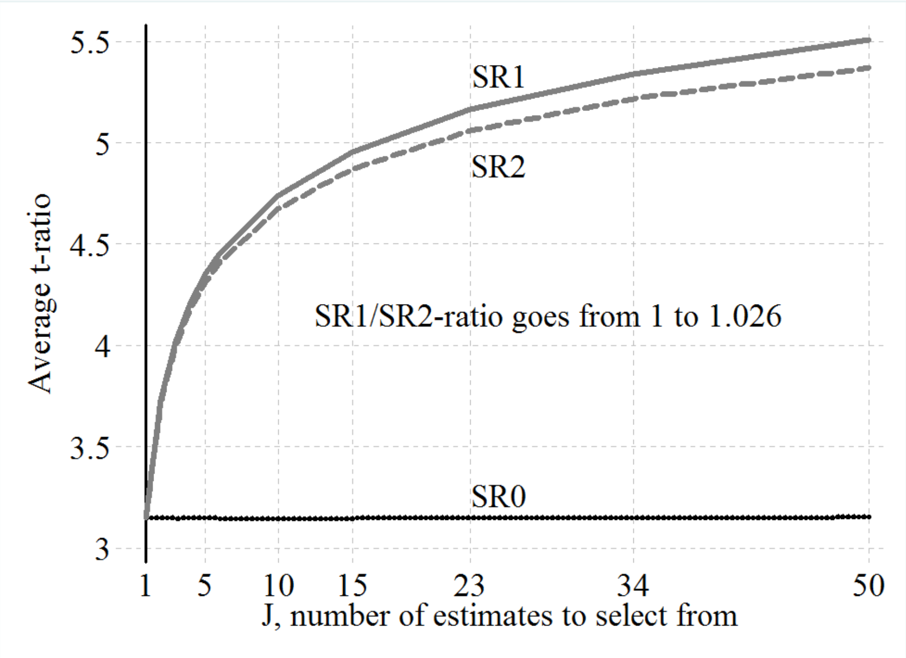


Note: Drawn for R = 1,000, except for low values $J \leq 6$, where R = 10,000.

5.2 *The t-ratio*

Columns (3) in Tables 2 to 4 give the estimates of the *t*-ratio as a function of *J*. It is drawn on Figure 3. The curves look like the ones for the mean; however, now the three curves start at 3.152.[12] The curve for SR0 remains constant, while the curves for SR1 and SR2 rise. But compared to Figure 2 the order of the two curves is reversed. Now the SR1-curve is at the top, as expected. Note that the gap between the two curves remains narrow. Any realistic trade-off between the fit and the size causes an *SR*-line in the narrow gap between the two extremes. Thus, it hardly matters if the researcher optimizes the fit or the size of the estimate.

Figure 4. The paths of the t-ratio



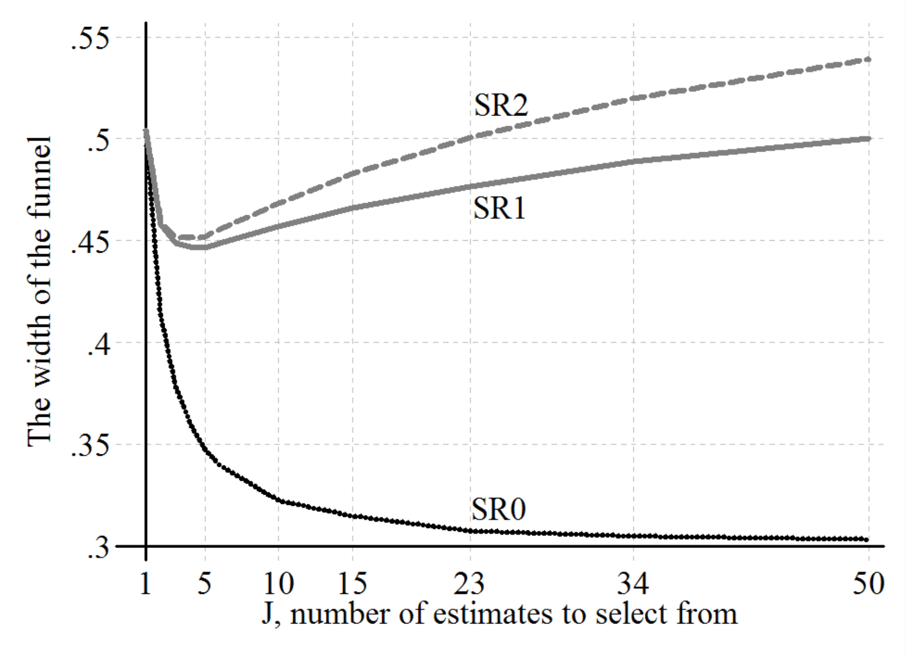
Note: Drawn for R = 1,000, except for low values $J \leq 6$, where R = 10,000.

5.3 *The width of the funnel*

Columns (4) in Tables 2 to 4 give the estimates of the funnel width calculated as the STD if the *N*-set is a function of *J*. For *SR*0 the funnel width converges to 0.3, which is the standard deviation of the βs. This is, of course, precisely as section 3.2 argued that it should. The path of convergence is $1/\sqrt{J}$.

12. The starting value of 3.152 looks deceptively like the square root of σ_ε = 10; but it is a more complex expression, and in section 5.3 it is 5.265 in the low case σ_ε = 6, and 2.255 in the high case σ_ε = 14. Thus, the higher the noise term, the lower is the average t- ratio, precisely as it should.

Figure 5. The width of the funnel



Note: Drawn for R = 1,000, except for low values $J \leq 6$, where R = 10,000.

For both *SR*1 and *SR*2 $\overline{V}$ stays remarkably stable. For *SR*2 the curve for $\overline{V}$ is always slightly above the one for *SR*1. For a realistic *SR* between *SR*1 and *SR*2, we expect that the funnel width curve is between the ones for *SR*1 and *SR*2. Thus, we know that when *J* rises all rational *SR*s cause the funnel to be skewer, but not leaner as I claimed is realistic in section 3.2.

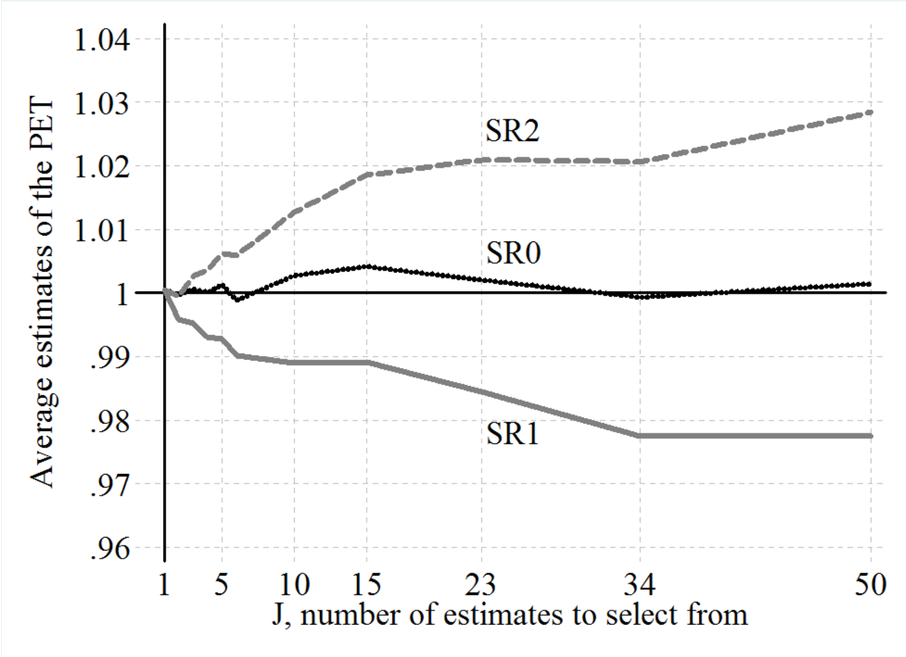
5.4 The bias of the PET

Columns (7) in the three tables show the PET-estimate of β. Here the estimates cover 'only' *R* replications, so the estimates are not as precise as the ones for the mean. Also, the range on the vertical axis is 1/10 of the range on Figure 3. Thus, we finally see curves that are not perfectly smooth. However, the paths of the curves on Figure 4 are still rather clear.

The most amazing finding is that the three PET-curves stay within 1 $\pm$ 0.03 of the true value of β = 1. The PET was not designed to catch the biases in the case where the bias is due to the rationality of researchers, but the PET still does a fine job. However, it is not perfect.

The PET does actually reject that β = 1 in about half of the cases. See columns (8) in the three tables. For σ_β = 0 the rejection rate is about 20%, but for σ_β = 0.3 the rejection rate increases to about 50%. So, I conclude that PET does get close to the true value of β, but in many cases it finds something that is not precisely true.

Figure 6a. The paths of the bias of the PET



Another interesting observation from Figure 6a is that when some researchers put more weight on the fit and others look more at the size, the PET bias curve will be between the SR1 and the SR2 curves. The SR1-curve is below 1 and the SR2-curve is above 1 by almost the same magnitude. Therefore, the bias of the average will be very close to 1, and, thus, almost without bias.

6. Experiments with the three variations using *SR*1

Six experiments – listed in Table 6 – are made with low and high cases for the three variation parameters, σ_β, σ_x and σ_ε. The results are reported in Tables 7 to 9 in sections 6.1 to 6.3, respectively.

Table 6. The three cases for three standard deviations

	(1)	(2)	(3) Present paper	(4)	(5)	(6)	(7) Previous paper, Paldam (2015b)	(8)	(9)
	Stochastic term	Choice	Draw	Low	Central	High	Draw	Central	Experiments
(1)	$\beta = N(1, \sigma_\beta^2)$	σ_β	One per paper	0.15	0.30	0.45	None	Zero	None
(2)	$x_t = N(0, \sigma_x^2)$	σ_x	One per observation	1	2	3	As now	2	1, 1.5, 3, 4
(3)	$\varepsilon_t = N(0, \sigma_\varepsilon^2)$	σ_ε	One per observation	6	10	14	As now	10	None

The experiments have generated 6 sets of tables corresponding to Tables 2 to 4. As the pattern in the estimates is unsurprising, the reporting is condensed into three tables. They cover the main results for *SR*1 only. Column (1) reports *J* as before. Then follow 3 sections with 3 columns each: The low case, the central case and the high case. The central case reported in columns (5), (6) and (7) is the same as in Table 3.

Table 7. The experiments with σ_β^2 the variation of β

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
		Low case $\sigma_\beta = 0.15$			Central case $\sigma_\beta = 0.3$			High case $\sigma_\beta = 0.45$		
Row	J	$\bar{\bar{b}}$	$\bar{b}_M$	$\bar{V}$	$\bar{\bar{b}}$	$\bar{b}_M$	$\bar{V}$	$\bar{\bar{b}}$	$\bar{b}_M$	$\bar{V}$
(1)	1	1.000	1.008	0.434	1.001	1.001	0.504	1.003	1.020	0.603
(2)	2	1.203	0.978	0.382	1.204	0.996	0.458	1.202	0.969	0.566
(3)	5	1.415	0.986	0.369	1.414	0.993	0.446	1.415	1.016	0.551
(4)	10	1.544	0.971	0.384	1.541	0.989	0.457	1.544	0.996	0.559
(5)	15	1.609	0.972	0.396	1.609	0.989	0.466	1.606	1.012	0.57
(6)	23	1.674	0.968	0.409	1.674	0.985	0.476	1.671	1.011	0.574
(7)	34	1.727	0.964	0.423	1.729	0.978	0.489	1.721	1.014	0.579
(8)	50	1.779	0.956	0.434	1.780	0.978	0.500	1.775	1.004	0.588
(9)	Avr.	1.494	0.975	0.404	1.494	0.988	0.475	1.492	1.005	0.574

Note: The central case is from Table 4. The low and the high case is run for $R = 500$.

6.2 *The experiments with σ_β the variation in β*

Table 7 shows that σ_β has no effect on the publication bias. This was expected from the above analysis, but the confirmation is still important.

It is also reassuring to see that the change on σ_β has only a small effect on the PET meta average. The value $\sigma_\beta = 0.3$ was chosen to be large, so $\sigma_\beta = 0.45$ is very large. But the PET still stays within 2% of the true value, for all *J*s. However, the value of σ_β does influence the funnel width, $\overline{v}$, which rises with σ_β as it should.

6.3 *The experiments with σ_x^2, the variation term for the explanatory variable*

Table 8 shows that the higher the variation term of the explanatory variable, the lower is the bias. However, the PET-meta average catches the true value rather well in any case. In all three cases the funnel width remains rather constant when *J* changes, though its size depends upon σ_x. This result should be seen in connection with the results in Table 9, where the noise term in the estimating equation σ_ε varies.

Table 8. The experiments with σ_x, the variation of x

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
		Low case $\sigma_x = 1$			Central case $\sigma_x = 2$			High case $\sigma_x = 3$		
Row	*J*	$\overline{\overline{b}}$	$\overline{b}_M$	$\overline{v}$	$\overline{\overline{b}}$	$\overline{b}_M$	$\overline{v}$	$\overline{\overline{b}}$	$\overline{b}_M$	$\overline{v}$
(1)	1	1.000	1.015	0.867	1.001	1.001	0.504	1.002	1.014	0.402
(2)	2	1.413	0.969	0.763	1.204	0.996	0.458	1.132	0.975	0.377
(3)	5	1.846	0.995	0.737	1.414	0.993	0.446	1.269	1.003	0.366
(4)	10	2.114	0.969	0.771	1.541	0.989	0.457	1.351	0.988	0.371
(5)	15	2.250	0.974	0.799	1.609	0.989	0.466	1.391	0.997	0.378
(6)	23	2.386	0.967	0.828	1.674	0.985	0.476	1.432	0.997	0.379
(7)	34	2.497	0.960	0.858	1.729	0.978	0.489	1.464	0.998	0.382
(8)	50	2.606	0.946	0.886	1.780	0.978	0.500	1.498	0.991	0.387
(9)	Avr.	2.014	0.974	0.814	1.494	0.988	0.475	1.317	0.995	0.380

Note: The central case is from Table 4. The low and the high case is run for *R* = 500.

6.4 *The experiments with σ_ε, the noise term in the estimating equation*

Table 9 shows the reverse pattern of the one in Table 8. The higher the variation term of the explanatory variable, the higher is the bias. However, the PET-meta average still catches the true value rather well.

Table 9. The experiments with σ_ε^2, the variation in the model residuals

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
		Low case $\sigma_\varepsilon = 6$			Central case $\sigma_\varepsilon = 10$			High case $\sigma_\varepsilon = 14$		
Row	J	$\bar{\bar{b}}$	$\bar{b}_M$	$\bar{V}$	$\bar{\bar{b}}$	$\bar{b}_M$	$\bar{V}$	$\bar{\bar{b}}$	$\bar{b}_M$	$\bar{V}$
(1)	1	1.002	1.013	0.384	1.001	1.001	0.504	1.001	1.015	0.643
(2)	2	1.117	0.975	0.363	1.204	0.996	0.458	1.287	0.970	0.575
(3)	5	1.239	1.004	0.353	1.414	0.993	0.446	1.588	0.996	0.556
(4)	10	1.312	0.990	0.356	1.541	0.989	0.457	1.773	0.974	0.577
(5)	15	1.347	1.000	0.361	1.609	0.989	0.466	1.866	0.980	0.595
(6)	23	1.383	1.000	0.362	1.674	0.985	0.476	1.959	0.976	0.611
(7)	34	1.411	1.002	0.364	1.729	0.978	0.489	2.035	0.972	0.629
(8)	50	1.442	0.995	0.368	1.780	0.978	0.500	2.110	0.961	0.647
(9)	Avr.	1.282	0.997	0.364	1.494	0.988	0.475	1.702	0.980	0.604

Note: The central case is from Table 4. The low and the high case is run for $R = 500$.

The funnel width, $\bar{V}$, falls when σ_ε rises. This might appear counterintuitive at a first look. However, it tallies well with the fact that the bias falls. Thus, Tables 8 and 9 show that it is the relation between σ_x and σ_ε that matters: If σ_x is small and σ_ε is large, it will result in wide funnels and vice versa. As empirical funnels are surprisingly wide, the choice of a much smaller variation σ_x than σ_ε is realistic.

The three sets of experiments with the variation variables σ_β, σ_x and σ_ε show a rather simple pattern that is easy to interpolate and extrapolate to the range of interest. The three key observations are that when $J > 1$: (i) The rational researcher always produces a bias. (ii) When the researchers want positive estimates the bias is always positive. (ii) The PET always greatly reduces that bias.

7. Conclusions: Rational researchers produce publication bias

This paper simulates the empirical literature about an important parameter, β, which is characterized by its size and fit (t-ratio). The analysis assumes that researchers behave rationally as predicted by economic theory. The key variables that need to be modelled are J, the number of estimates made per estimate published, and SR, the selection rule used to choose the published estimate. For both choices economics make a clear prediction.

J will be chosen where the marginal cost of an extra regression equals the marginal benefit. At low values of J the marginal costs are much lower than the marginal benefits. Consequently, authors typically have to choose the regressions published from the much larger set of regressions produced. The optimal choice is where the author's production possibility frontier touches his utmost indifference curve.

It is assumed that the β-literature consists of N = 500 estimates that are clustered in papers that report 10 estimates each. The 10 estimates are reported to show the robustness of the main result. The parameter, β, has a central value of 1, but it is allowed to vary randomly between papers, so the literature has parameter heterogeneity. It means that the estimates vary less within papers than between papers.

A previous study assumed that the 500 estimates are independent. It reached three key results: (r1) When J is larger than one, all rational selection rules produce a bias, which is often substantial. The bias is in the direction of the priors of the researcher. (r2) The bias is almost the same irrespective of the weight the researcher places on fit and size. (r3) 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 the PET estimate reduces the bias by more than 90%. These results generalize to parameter heterogeneity. Two additional results have been found: The reason given by authors for publishing 10 estimates is that it shows the robustness of the main result. (r4) This paper shows that the bias survives as long as the number of estimates per published one is larger than 1. What matters for the bias is the number of estimates made per published one, not the number of estimates published.

(r5) Parameter heterogeneity makes it more likely that the PET rejects the true value of β = 1, but the rejections are to either side with almost the same probability, so the PET is still a fine estimate of the true value.

References:

Kirchgässner, G., 2005. (Why) are economists different? *European Journal of Political Economy* 21, 543-62

Paldam, M., 2013. Regression costs fall, mining ratio rise, publication bias looms, and techniques get fancier. Reflections on some trends in empirical macroeconomics. *Econ Journal Watch* 10 (2): 136-56.

Paldam, M., 2015a. Meta-analysis in a nutshell: Techniques and general findings. *Economics E-Journal* 9, 2-15

Paldam, M., 2015b. Simulating an empirical paper by the rational economist. *Empirical Economics* online first. DOI: 10.1007/s001881-015-0971-6

Paldam, M., 2015c. A model of the research of the rational economist. URL: http://www.martin.paldam.dk/Papers/Meta-method/Rational-economist-model.pdf

Stanley, T.D., 2008. Meta-regression methods for detecting and estimating empirical effects in the presence of publication selection. *Oxford Bulletin of Economics and Statistics* 70(1), 103–127

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

Stanley, T.D., Doucouliagos, C(H). 2014. Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods 5*, 60-78

Stanley, T.D., Doucouliagos, C(H)., 2007. Identifying and correcting publication selection bias in the efficiency-wage literature: Heckman meta-regression. School Working Paper, Economics Series 2007–11, Deakin University.