Aarhus, 23/1-2014

How do partly omitted control variables influence the averages used in meta-analysis in economics?

Martin Paldam [1]

Abstract:

Meta regression analysis is used to extract the best average from a set of *N* primary studies of one economic parameter. Three averages of the *N*-set are discussed: The mean, the PET meta-average and the augmented meta-average. They are affected by control variables that are used in some of the primary studies. They are the POCs, partly omitted controls, of the meta-study. Some POCs are ceteris paribus controls chosen to make results from different data samples comparable. They should differ. Others are model variables. They may be true and should always be included, while others are false and should always be excluded, if only we knew. If POCs are systematically included for their effect on the estimate of the parameter, it gives publication bias. It is corrected by the meta-average. If a POC is randomly included, it gives a bias, which is corrected by the augmented meta-average. With many POCs very many augmentations are possible. The mean of all augmented meta-averages is also the mean of the *N*-set. If it has a publication bias so do the average augmented meta-averages.

Keywords: Meta-analysis, omitted variables, meta-average

JEL: B4, C9

Acknowledgements

The simulations in sections 4 and 5 were done by Jan Ditzen and Orimar Sauri. I am grateful for their care. I have also benefitted from many discussions with Chris Doucouliagos and Tom Stanley. The paper has been presented at the Danish Econometric Society meeting at Sandbjerg, Denmark, in May 2013, and at the MAER-Net (URL in references) conference in Greenwich, September 2013.

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

1. Introduction: Three problems and a deep issue

The following deals with the meta-analysis of a generic β-literature. It consists of M papers that report N estimates of something that pertains to be the same parameter β. I assume that β is the effect of x on y. Each of the N estimates is from a linear regression, with y as the dependent variable. It contains the term of interest, βx, and a set of K controls.[2]

The meta-analysis considers three averages of the N-set: The (arithmetic) mean $\underline{b}$, the PET meta-average β_M, and the augmented meta-average β_A that easily becomes a large family. The ***first problem*** analyzed in the paper is the strengths and weaknesses of the three averages.

The β-literature contains a total of L controls, where $L > K$. Once L is large the K controls can be chosen in many ways giving a wide range of estimates. Some researchers seem to consider the L-set as a trove of possible choices allowing them to find the right result. The L controls are the POCs, partly omitted controls, in the meta-study. While the term of interest has to be included in all estimates of the β-literature the POCs differ. The ***second problem*** is the effect of the choice of POCs on the three averages.

The choice of POCs has a stochastic element. It follows that POCs may be true, so that they should be included, or false, so that they should be excluded. It appears that false POCs are likely to be common. Till now meta-analysis has disregarded false POCs. The ***third problem*** is how false POCs influence the meta-analysis.

The three problems are aspects of the ***deep issue*** of comparability: The estimates in the β-literature are done on different data samples. The meta-analysis compares these estimates. Four observations support comparability: (i) It makes sense if, and only if, the estimates fulfill the ***cp*** (ceteris paribus) condition, and most estimating models contain *cp-controls*. (ii) Many papers claim that they derive the estimation model from economic theory. (iii) Researchers often take some POCs to belong in the model, so that they are *model variables*, which should be included in all estimates even if they generate an insignificant coefficient. They may even be derived from the theory. (vi) Also, most studies have literature reviews where estimates are compared – authors must think that this makes sense.

Points (i) and (iii) remain implicit in many papers, and it is rare that a paper explicitly distinguishes between model variables and cp-controls. They are normally lumped together as controls – thrown in with little justification. The choice of POCs is a gray area in economics.

2. Example: β may be the effect of a tax rate, x, on the rate of unemployment, y. Economists know that the sign on β is positive. However, when a government wants to change the tax rate it is not enough to know the sign on β, and a large literature, such as 400 papers, studies the size of the effect, β. Meta-analysis claims that we may learn more about the size of β from a study of the 400 papers than from adding paper number 401.

A POC is missing in a model in four cases:

(Ca1) It is a cp-control which should not be included due to the data analyzed.[3]

(Ca2) It was tried, but it did not work.[4] Thus, it should be treated as zero.

(Ca3) The author did not try this control. Thus, in principle it is an omitted variable.

(Ca4) It is a false variable that the author omitted as he should. Thus, some other papers contain a false variable.

The reader should keep the four cases in mind. The augmented meta-average adjusts for case (Ca3), but it is biased in the other cases. Below an alternative augmentation method is presented that adjusts for case (Ca4), but it is biased in the other cases.

It is not the job of meta-analysis to judge the literature. The job is to analyze what the literature has found as objectively as possible. This gives two levels in the meta-analysis.[5] Level one bypasses the gray area of the POCs. Here a high level of objectivity can be reached, in the sense that two meta-analysts doing the same study independently, will reach much the same result. Level two concentrates on the POCs. Here the attempt to handle things objectively has to grapple with two 'impossibly' large numbers: The number of model variants reached by a selection of POCs, and the number of possible augmentations.

The two levels of meta-analysis are discussed in section 2 and section 3 respectively. Section 4 reports a set of simulations of level one of the meta-analysis, while section 5 simulates level two. Section 6 concludes. All definitions and variables are listed in the Appendix for easy reference. Efforts have been made to make this paper accessible to non-economists. They should note the Appendix defining the terms.

The paper reports results from more than half of the simulations made.[6] The pattern in the remaining simulations is documented in Paldam (2013b) available from the home page of the project: http://www.martin.paldam.dk/Meta-Method.php.

---

3. A typical cp-control is a regional dummy that should not be included if the data does not cover the region.
4. Due to the pressures of space such variables are often left out. Maybe, it is mentioned (often in a footnote) that they were tried, but gave no result. However, the coding for the meta-study rarely covers such remarks.
5. The state-of-arts in meta-analysis is covered in the recent textbook Stanley and Doucouliagos (2012). The reader should also consult Stanley *et al.* (2013) for a set of guidelines.
6. The simulation experiments done for the paper can be measured in regressions: A total of about 21 million have been run. Tables 3 to 6 cover about 12 mill of these regressions, while the eight tables in Paldam (2013b) document the pattern in the remaining 8.5 mill.

2. Level one: The funnel and the first two averages $\underline{b}$ and β_M

Section 2.1 introduces the two levels of meta-analysis, and section 2.2 presents the recipe for level one of the meta-analysis. Section 2.3 discusses the two averages, $\underline{b}$ and β_M, while section 2.4 deals with the hidden dimension in research, the *J* regressions made for each selected publication. Section 2.5 looks at selection in practice.

The data for level one is the *N*-set of estimates, b_i, including standard errors, s_i, from which precisions, $p_i = 1/s_i$, and *t*-ratios, $t_i = b_i/s_i$ are calculated. Also, a *t*-variable for the time of publication should be coded, so that b_i-set can be sorted into a b_t-set.

2.1 The two levels of meta-analysis and main priors in the literature

Meta-analysis has two levels: ***Level one*** is discussed in the present section. It estimates the first two averages $\underline{b}$ and β_M, which correct the mean, $\underline{b}$, for publication bias. It disregards the POCs. This is right if all POCs are cp-controls. I show that it also makes sense if there are enough POCs. Level one has a clear recipe that produces robust results.

Level two is discussed in section 3 and later. It has two aims: Firstly, it studies the effect of the POCs on the width of the distribution of the *N*-set. Secondly, it estimates the third estimate, β_A, which corrects the meta-average for POC-bias by the augmentation method, explained in section 3.4. It makes in case (Ca3). Level two is still rather unsettled, and the results are much less robust than the ones at level one.

Most *meta-studies* show that the economic profession has a ***main prior*** about the parameter of interest β. It is typically a sign prior, such as $\beta > 0$. Priors come from economic theory, common political/moral views, and the interests of dominating sponsors. Such priors cause results to be *exaggerated*. Also, science in general has a prior for clarity, which is enforced by the publication process. It leads to polishing of results, so that t-ratios rise.

2.2 The recipe and the problem of the big gap

Level one has five steps (s1) to (s5):

(s1) A search should be made for the complete β-literature.[7]

(s2) The (b_i, s_i, p_i, t_i)-data from the β-literature should be coded into a worksheet. In

7. Researchers outside meta-analysis often argue that the meta-study should be restricted to new papers, papers in top journals, or papers using some new estimation technique. Meta-analysts prefer to include everything and code variables for such restrictions. This allows the analyst to ask if they really matter.

principle, the coding can be done in one way only.[8]

(s3) The *funnel* is the (p_i, b_i)-scatter. It displays the distribution of the N-set. Low precision estimates scatter most, so the base of the funnel is wide. As precision increases it narrows. The form of the funnel provides important qualitative information.

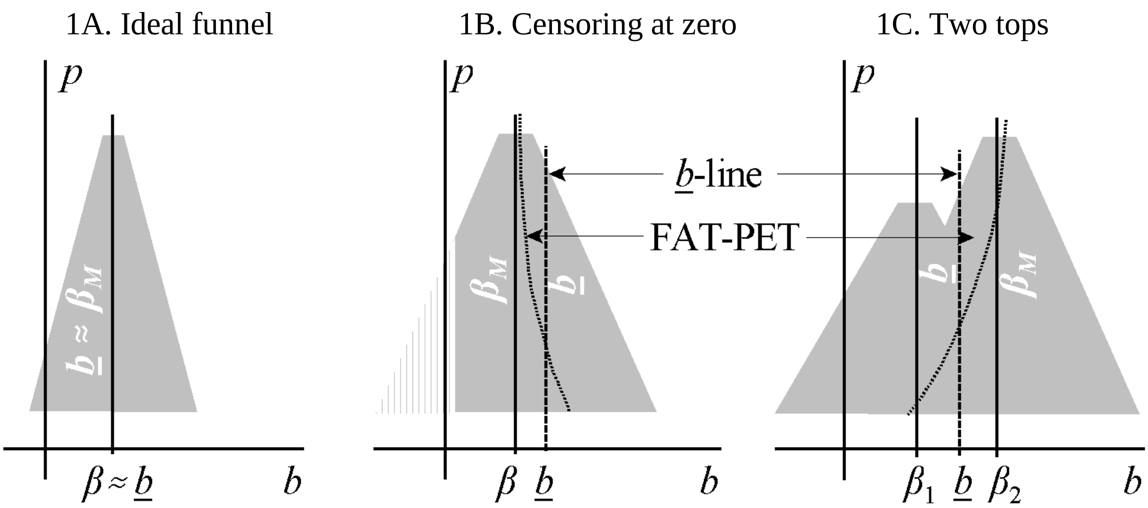
(s4)[9] The *path* in the b_t-set should be studied to identify *time dependencies*, as trends and structural shifts, indicating breakthroughs in models or estimating techniques.

(s5) Finally, the result ($\underline{b}$, β_F, β_M) is calculated as explained in section 2.3.

The recipe is clear and the steps are well defined. Thus, they can be done in one way only, and the results are robust, see Doucouliagos and Paldam (2013a) for a case study.

Statistical theory about regression coefficients and many simulation experiments (see e.g. section 4.2 below) show how funnels should look: They have two properties: They are symmetric and as lean as suggested by the t-ratios of the estimates, as sketched on Figure 1A. The main problem at level one is the big gap between the way funnels should look and how they actually look.

Figure 1. Funnels with typical forms affected by POCs



Note: Each figure shows three curves: The vertical axis of symmetry through β, the vertical $\underline{b}$-line, and the FAT-PET curve that converges to β_M when p rises. On Figure A all three curves are the same, so that $\beta \approx \underline{b} \approx \beta_M$. On Figures B and C the three curves differ, and on Figure C β even becomes two lines. Simulated versions of Figures 1A and 1C will appear as Figures 3 and 4 respectively, see also Paldam (2013b) for many examples.

8. The bibliography and the coded worksheet should be published on the net. Some random search and coding errors will occur at (s1) and (s2). When N is large a small number of moderately sized stochastic errors matter little. Large coding errors are easy to discover in (s3).

9. While the other steps are now standard (s4) is often forgotten, see section 3.1.4 in Stanley and Doucouliagos (2012). Time dependencies are discussed in section 3.3. If no time dependencies are identified old estimates are as good as new ones.

At present about 500 meta-studies have been made in economics, and most have published the funnel, so it is well known how empirical funnels look: They are amazingly wide and most are asymmetric as sketched on Figures 1B and 1C. At level one it is possible to deal with the width coming from trends and structural shifts, but they are normally a small fraction only of the excess width. Thus, it is not trivial to identify the true value of β.

The paper considers two types of funnel asymmetries: Censoring (Figure 1B) and multi-tops (Figure 1C). The PET meta-average corrects for censoring as discussed in the rest of the present section. Censoring is caused by a systematic selection of the POCs. If all models used the same controls there would be no asymmetry, as shown in section 5.

2.3 The Mean, $\underline{b}$, the FAT, β_F, and the PET meta-average, β_M

The simplest way to summarize the literature is to calculate the (arithmetic) mean:

(1) $\underline{b} = \sum_{i=1}^{N} b_i / N$, which might be weighted $\underline{b}^w = \sum_{i=1}^{N} (w_i b_i) / \sum_{i=1}^{N} w_i$, with weights w_i

The FAT, β_F, is the funnel asymmetry test, and the PET meta-average, β_M. They are jointly calculated by the FAT-PET MRA: [10]

(2a) $b_i = \beta_M + \beta_F s_i + u_i = b_i = \beta_M + \beta_F/p_i + u_i \rightarrow \beta_M$, when precision $p \rightarrow \infty$. It becomes

(2b) $t_i = \beta_M p_i + \beta_F + v_i$, after division by s_i.[11]

The idea of the FAT is that if a funnel is symmetric, the average at each level of precision is the same, so that the average is independent of s_i. The FAT is a powerful test of asymmetry. It detects both censoring and multi-tops. If the funnel is symmetric, the FAT $\beta_F = 0$, so that $\underline{b} = \beta_M$. Iff $\beta_F \neq 0$, it means that $b_M \neq \underline{b}$. β_M is the censoring-corrected estimate of the mean. Section 4 studies an ideal funnel that becomes increasingly censored.

Relation (2) shows that β_M is the limit at the top of the funnel as p goes to infinity. It is a good estimate of β when the funnel is symmetric, but it also works when it has a censoring asymmetry. When the funnel has several tops, it normally converges to the highest one. This might be the best estimate of β, but it might also be the most biased estimate as discussed below. The typical result in simulations, where β is known, is that β_M is (much) closer to β than is $\underline{b}$. In most meta-studies, where asymmetries are found, they look like censoring at predictable points. Hence, β_M is a fine estimate of β.

10. The FAT is from Egger *et al.* (1997), while the FAT-PET MRA is developed by Stanley (2008).
11. Version (2b) is preferable to (2a) for estimation purposes as it has less heteroskedasticity.

2.4 The big gap is due to the hidden mining dimension of the J-set [12]

To find the regression giving b_i, the typical author makes a search of J_i regressions. A rational researcher will go on regressing as long as the marginal benefits of pushing on are larger than the very small costs. The J_i-set and the selection of b_i are private information of the researcher. The number of regressions made to produce the N-set is:

(3) $\sum_{i=1}^{N} J_i = N\underline{J}$, where all $J_i \geq 1$. My guess, based on informal polls, is that $\underline{J} \approx 25$.

This is the iceberg property of macroeconomic research: The visible part of the iceberg is the N-set, and the $N(\underline{J}-1)$ other regressions are the part of the berg that remains invisible below the water. It is much larger and known as the dangerous part, also in the present context. $\underline{J}$ is termed the mining ratio. If all $N\underline{J}$ regressions were available, they would give an ideal funnel.

The J_i-set is made to allow a search for a fine model by the priors of the researcher. With the *prior* $\beta > 0$, the researcher wants to choose an estimate with a positive b. And by the clarity prior he wants to choose an estimate with high t-ratio. They rarely occur for b's close to zero. This implies that the mean of the J_i-set, $\underline{b}_{Ji}$, is likely to be closer to β than the best value, b_i, selected.

(4) $SB_i = b_i - \beta \approx b_i - \underline{b}_{Ji} > 0$, SB_i is the *selection bias* for b_i for the prior $\beta > 0$

Thus, the positive sign prior gives a positive bias on the b's. In the same way a negative sign prior gives a negative bias. Thus, a prior exaggerate estimates in the direction of the prior, see Doucouliagos and Stanley (2012). The discussion of the exaggeration result is continued in section 3.2. The micro selection bias is easy to aggregate to the whole N-set:

(5) $PB = \sum_{i=1}^{N} SB_i / N = \sum_{i=1}^{N} (b_i - \beta) / N = \underline{b} - \beta \approx \underline{b} - \beta_M$, PB is the *publication bias*

With many different priors the selection bias may even out so that the $PB \approx 0$, but with a main prior, $\beta > 0$, most SBs will be positive. Thus, (5) becomes significant and positive. Also, if PB becomes significant, we know that the literature has a main prior.

The width of empirical funnels is an artifact because one of the selection rules used by researches is to select regressions with high t-ratios that are often at the rim of the $N\underline{J}$-funnel, so the width is an indication of the width of the $N\underline{J}$-funnel, where most of the estimates have

12. This section is a brief summary of Paldam (2013a and 2013c).

much lower *t*-ratios than the selected ones in the *N*-set.

2.5 *Selection of the best regression in practice*

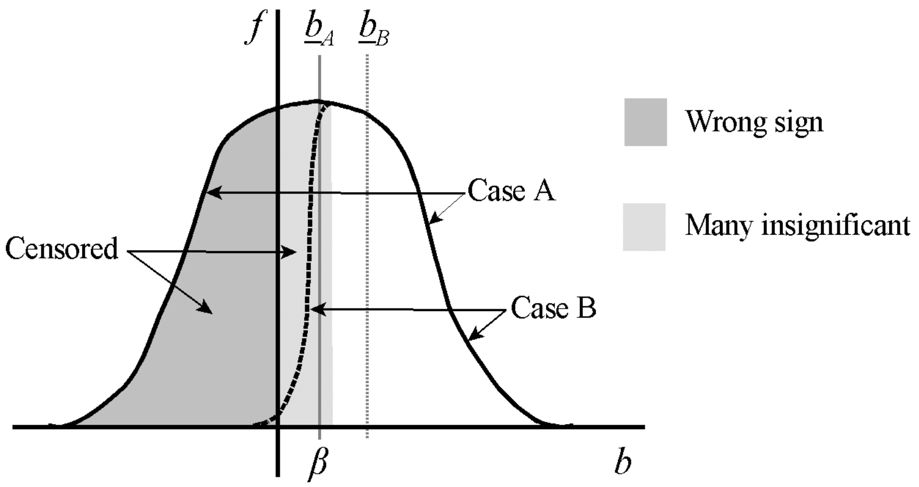
The main tool to obtain a large *J*-set of estimates to choose from is to vary the POCs. When β has been researched for some time many controls have been tried, so the number *L* is high (such as 50). Most regressions contain only a handful. If the researcher feels free to choose *K* controls from the trove of all *L* POCs, the number of possible combinations is:

$$(6) \qquad n(L, K) = \binom{L}{K} = \frac{L!}{(L-K)!K!}$$

This formula generates large numbers. With (*L*, *K*) = (50, 5) equation (6) gives 2,100,000 possible estimating equations.[13] Each of these provides an estimate of β, so the *trove model* enables the researcher to surround β with a range of possible *b*s to choose from. This is the first large number at level two of the meta-analysis. The wealth of defensible model variants makes it easy to do the *J* (hidden) experiments from section 2.4. From Sala-i-Martin *et al.* (2004) we know how the distribution of all the possible *b*'s look for one literature. It is fairly normal and that the vertical axis is often well within the distribution as shown on Figure 2.

Case A: The literature has a large number of POCs and all estimates are published. Here β_M is very close to $\underline{b}_A$, and the analysis concludes that $\underline{b}_A \approx \beta_M \approx \beta$.

Figure 2. The frequency (*f*) of possible choices with the trove approach



Note: Case A is without censoring as Figures 1A and 3. Here the full area below the black bell-curve is considered. Case B is with censoring as Figures 1C and 4. Here everything on the figure is relevant. The frequency distribution is another presentation of the same data as is used for the funnel.

13. Xavier Sala-i-Martin has calculated all these regressions in a concrete case and analyzed their distribution; see Sala-i-Martin (1997), and Sala-i-Martin *et al.* (2004).

Case B: The literature has a main prior ($\beta > 0$) so most estimates with the wrong sign are strongly underreported, and due to the clarity prior insignificant estimates are underreported as well. This gives an asymmetric distribution and too large t-ratios. The analysis concludes that $\underline{b}_B > \beta_M \approx \beta$. On the figure the exaggeration is about 2, i.e., $\underline{b}_B \approx 2\beta_M$.[14]

The trove model has been simulated simply by treating the choice of POCs as a stochastic noise in the *N*-set. When an estimate has been generated, it has then been submitted a selection rule, so that it is either accepted or rejected. This has been done until an N-set has been reached. The level one analysis has been made on these simulated *N*-set, see Stanley (2008) and Callot and Paldam (2011) and sections 5 and 6 below.

It should be noted that the simulation experiments in sections 4 and 5 use a much more moderate censoring than the one suggested by the theory of the hidden J-dimension above. Hence, it produces smaller publication biases than the ones often found in practice.

---

14. This is fairly typical. The meta-community has a rule-of-thumb: Expect an exaggeration of 2. Doucouliagos and Stanley (2012) report a rather large standard deviation around this rule.

3. Level two: The POCs and the third average, β_A

Section 3.1 introduces level two, and section 3.2 discusses false POCs. Section 3.3 presents a new tool for analyzing POCs. Section 3.4 looks at means of biased and unbiased estimates. Section 3.5 discusses the augmentation technique giving the A-set of averages, and section 3.6 derives a few general results for the A-set, while section 3.7 shows that the average of the A-set is $\underline{b}$, not β_M.

The data at level two includes the binary ($N\ x\ L$)-matrix for inclusion/non-inclusion of the POCs. It has the inclusion vector ω_k for control z_k as column k. The symmetrical exclusion vector is φ_k. If estimate i includes control z_k, $\omega_{ki} = 1$ and $\varphi_{ki} = 0$, and if z_k is excluded $\omega_{ik} = 0$ and $\varphi_{ki} = 1$. An augmented meta-average is reached if either ω_{ki} or φ_{ki} is added as a regressor in the FAT-PET MRA of equation (2). Each of the ω_ks or φ_ks can also be sorted over time to give the ω_{kt}-set and the φ_{kt}-set.

3.1 Introducing level two of the meta-analysis

The second level deals with the big gap between expected and empirical funnels: The key to understand the gap is the four types of POCs given (Ca1) to (Ca4) from the introduction. Each individual POC gives a separate top on the funnel when it is included: It might be a true top or a false one.

With a few (or some very important) POCs the tops may show. Each top is increased or deleted if the POC is included or excluded in all estimates, and thus the funnel becomes (more) symmetrical. However, this is wrong to do in (Ca1) and (Ca2).

With many POCs all tops normally add up to a smooth funnel. When the POCs are selected systematically to give the 'right' estimate of β the funnel will be smooth too, but it will often show censoring. When the POCs are selected to make the estimate of β more significant (statistically) is leads to polishing, making the funnel much wider than implied by the t-ratios.

The cp-controls should differ when the data set does. A model variable is either true or false. Hence, it should either be in all estimates or in no estimates. In both cases it gives a POC-bias that it is partly included only. A model variable gives an omitted variable bias on β_M if the POC is true and an inclusion bias if it is false. The two biases will be jointly discussed as POC-bias.

The augmentation builds on the *same-effect-assumption*: The effect of the POC is the same in the models where the POC is included as it would have been, if included, in the models where it is not, this can be assumed in case (Ca3). It can also be assumed in case (Ca4), but here the augmentation has to be different as will be explained.

If (Ca3) is false the augmentation is biased. If the analysis at level one finds that the *N*-set has a publication bias, this means that the same-effect-assumption is rejected. Hence, the augmentation is biased. With no censoring and a one topped symmetrical funnel $\underline{b} = \beta_M$, and no augmentation should be made. If the funnel is clearly two-topped and there is no sign of censoring, the meta-analyst can determine the reasons for the two tops.

3.2 Are most POCs true or false?

The above analysis has macroeconomics in mind as it is a field with rather limited data and large interests at play. Imagine that the data are 50 annual data from 40 countries, this is 2,000 observations. If it is an important relation, there may easily be 200 papers with on average 8 estimates in each. If the mining ration *J* is 25, a total of 40,000 regressions have been made to explore this relation. Some of these are the same as new researchers often run through some of the old models before they add some new twist, but the field is still heavily mined. Mining has an important consequence for the two types of errors: Type I errors are the rejection of the true model – mining reduces this type of errors. Type II errors are the acceptance of false models (variables) – mining increases this type of errors.

Hence, it is likely that heavily mined fields have many false variables. In the field of growth regressions about 400 variables have been tried to explain the growth rate (see the appendix to Durlauf *et al.* 2005). All of these variables have been found significant in one study or another, but several studies have found that only about 20 of these variables provide robust explanations (see Sala-i-Martin *et al.* 2004, Sturm and de Haan, 2005). Thus, it is arguable that the growth regression literature contains 20 true and 380 false variables. Some of the 380 variables are significant due to a random fluke, but most belong to groups of variants of the same factor or are intermediate variables in a complex of confluent variables.

Thus, the prime example of false variables occurs with multicollinearity. Think of five control variables which are all proxies for the same factor. The factor should be counted once, but not five times. Thus, the five variables should be treated as one true and four false variables in the analysis at level two.

3.3 A new tool: The time profile of inclusion

Like the b_i-set should also be sorted into the b_t-set, the ω_z-vector can be sorted into a ω_{zt}-set to study *the time-profile* for the inclusion of *z*. It allows the analyst to see if the use of the said control happens in a certain period, and if this period has similar movements in the b_t-set. In a few simple cases this allows important conclusion about the true model to be drawn:

Think of a model-POC, *z*, which is discovered at a certain time and becomes increasingly common from that time onwards. This should show up as a corresponding structural break in both the ω_{zt}-set and the b_t-set. In this case it is likely that *z* would have had the same effect before it was discovered, as it has had after. Thus it is a break-through control, where it is likely that that the same-effect-assumption holds. It should be treated as a true control.

Most researchers dislike variables that give small or insignificant coefficient. Thus, if a control was prominent in the literature, up to a point and then drop out, we may assume that the researchers have noted and tried this control, but that it ceases to produce a significant coefficient. Thus it is a drop-out control, where it is unlikely that the same-effect-assumption holds. It should be treated as a false control.

An interesting case is a control that comes and goes. My assessment is that such variables are the ones most likely used in the systematic part of the selection process. Here it is unlikely that the same-effect-assumption holds, so it should be treated as a false control.

To go through all POCs in this way is a major effort and it likely to yield unclear results for many variables. Also, it does not sort out the cp-POCs.

3.4 The mean for a model with one POC: Sorting the N estimates in the Q- and the S-part

Assume that the POC is randomly included (Ca3) or (Ca4). This gives a funnel with two tops. The *N* estimates are sorted so that the *Q*-part, with the share $q = Q/N$, is first part. Here all estimates include the POC. The remaining $S = N - Q$ estimates are without the POC. It is the *S*-part, with the share $s = S/N = 1 - q$. The *Q*-part gives one top and the last *S*-part gives another. Many examples of such funnels are found in Paldam (2013b). The mean becomes:

(7) $\underline{b} = \sum_{i=1}^{N} b_i / N = \sum_{i=1}^{Q} b_i \frac{Q}{QN} + \sum_{i=Q+1}^{N} b_i \frac{S}{SN} = q\underline{b}^Q + s\underline{b}^R$. In the weighted case it is

(8) $$\underline{b}^w = \sum_{i=1}^{N} w_i b_i / \sum_{i=1}^{N} w_i = \sum_{i=1}^{Q} w_i b_i \frac{\sum_{i=1}^{N} w_i}{\sum_{i=1}^{Q} w_i \sum_{i=1}^{N} w_i} + \sum_{i=Q+1}^{N} w_i b_i \frac{\sum_{i=Q+1}^{N} w_i}{\sum_{i=Q+1}^{N} w_i \sum_{i=1}^{N} w_i} =$$

$$q^w \underline{b}^Q + s^w \underline{b}^R, \text{ where } q^w = \sum_{i=1}^{Q} w_i / \sum_{i=1}^{N} w_i, \text{ and } s^w = \sum_{i=Q+1}^{N} w_i / \sum_{i=1}^{N} w_i,$$

Note that $q^W + s^W = 1$, for all weight sets, and if the *w*'s are positive, $0 < q^w < 1$. The reader can easily check that if the weights are 1 equation (8) becomes (7). So, equation (7) is a good approximation to (8) if the weights are not extreme and independent of the in/exclusion of z.

Three conclusions follow: (i) The mean $\underline{b}$ is always between $\underline{b}^Q$ and $\underline{b}^S$. (ii) If z is true, $\underline{b}^Q$ is the unbiased estimates, and it is thus best. (iii) If z is false, $\underline{b}^S$ is the best average. This pattern is given in Table 1. The key observation from the table is that if a false variable is treated as true or vice versa, the resulting 'best' estimate of β is in fact the 'worst' average, which is more biased than the average of all observations.

Table 1. The relation between the means
if the POC is a model variable and it is randomly included

	(1)	(2)	(3)	(4)	(5)	(6)
	The POC is true			The POC is false		
	S-set	All	Q-set	S-set	All	Q-set
POC z	Not in	Mixture	In	Not in	Mixture	In
Biased	Fully	Partly	None	None	Partly	Fully
Average	$\underline{b}^S$	$\underline{b}$	$\underline{b}^Q$	$\underline{b}^S$	$\underline{b}$	$\underline{b}^Q$

Some authors think that the POC is a true model variable and others think that it is not. The average (7) is weighted with number of researchers in both groups. It is thus the best average, except when the funnel is censored, see sections 5.3 and 6.4.

Obviously, if we are dealing with case (Ca1) or (Ca2) the mean does not cover biased and unbiased averages, so the mean is better the means of both the S-set and the Q-set.

3.5 The augmentation method to account for randomly included POCs

Formula (7) is easy to apply for one POC, but it is not suitable in situations with many. However, an elegant solution is to augment the FAT-PET MRA (2) with the a-vector that is either the ω- or the φ-vector for each z the (2) to get (9):

(9) $$b_i = \beta_A + \beta_F s_i + \lambda_1 a_{i1} + \ldots + \lambda_L a_{iL} + u_i$$

With two possibilities (ω_{ij} and φ_{ij}) for each a_{ij}, the number of possible augmentations is:

(10a) $n(L) = 2^L$ or

(10b) $\binom{L}{K} 2^K$ if K variables are chosen for the augmentation

These expressions get large rather quickly, so it is the second large number in the meta-analysis. When it is compared with equation (6) it produces even larger numbers. Think of the example above (in section 2.5) where $L = 50$ and $K = 5$. Here (6) gave $2.1 \; x \; 10^6$ estimates of β. Equation (10a) gives $1.1 \; x \; 10^{15}$ estimates of β_A.

Some of the POCs may be cp-controls that should not be augmented for, and for other POCs the time profile may tell us that they are either true or false, so instead of L we should look at $L^M < L$ in relation (10). But even for $L^M = 30$ equation (10) still gives $1.1 \; x \; 10^9$ possible augmentations.

The $n(L^M)$-set of augmented meta-averages is the A-set. It does have a considerable range, and if the meta-analyst searches this set, it is likely that he will find a 'nice' result. If there is no censoring, the $\beta_F s_i$-term is irrelevant so (9) becomes:

(11) $b_i = \beta_A + \lambda_1 a_{i1} + \ldots + \lambda_L a_{iL} + u_i$

With only one POC (11) gives the same result as (7). So that

(11a) $b_i = \beta_{A1} + \lambda \varphi_i$, gives $\beta_{A1} \approx \underline{b}_Q$ and

(11b) $b_i = \beta_{A2} + \lambda \omega_\text{i}$, gives $\beta_{A2} \approx \underline{b}_R$.

Recall that the sum of ω is Q and the sum φ is R, so (11a) and (11b) replicate (7) $\underline{b} = q\beta_{A1} + s\beta_{A2}$. Hence, if $\beta_{A1} > \underline{b}$, it follows that $\beta_{A2} < \underline{b}$, and vice versa. If $q = s = 0.5$, β_{A1} and β_{A2} are symmetrical around $\underline{b}$.

If z is true, we use φ_i to take out the results that are not controlled for z, and if z is false, we use ω_i to take out the results that are controlled for z.[15] In the simulations we know the DGP, data generating process. Thus, if z is in the DGP, (11a) is the *right* augmentation and (11b) is the *wrong* augmentation. However, if z is not in the DGP, the roles of (11a) and (11b) are reversed. In the simulations we write the augmented meta-averages β_{AR} and β_{AW} if they are rightly and wrongly augmented.

If we know if z is true or false, it is easy to make the right augmentation, but the very fact that z is a POC means that the researchers differ in their assessments.

3.6 Two general results about the A-set, given that the FAT is zero

It is possible to derive a few general results about the A-set. We first look at the case without censoring where we disregard the FAT as in equation 11. For each choice of an augmentation

15. It is a little confusing that the exclusion set φ is used if z is true, and the inclusion set ω is used if z is false. The reader should keep this point in mind.

such as (12a), the symmetrical augmentation (12b) also exists:

(12a) $b_i = \beta_{A1} + \lambda_1\, \omega_{i1} + \lambda_2\, \varphi_{i2} + \ldots + \lambda_{L\text{-}1}\, \omega_{iL\text{-}1} + \lambda_L\, \varphi_{iL} + u_i$

(12b) $b_i = \beta_{A2} + \lambda_1\, \varphi_{i1} + \lambda_2\, \omega_{i2} + \ldots + \lambda_{L\text{-}1}\, \varphi_{iL\text{-}1} + \lambda_L\, \varphi_{iL} + u_i$

For each pair of symmetrical terms such as $\lambda_2\, \varphi_{i2}$ and $\lambda_2\, \omega_{i2}$, we know that they move the estimate of β to the reverse side of $\underline{b}$. Since it holds for each z, it must hold for all L zs in (12). Hence, $\underline{b}$ is always in the interval between β_{A1} and β_{A2}. This proves that $\underline{b}$ is the median in the full *A*-set. If the *A*-set is normally distributed, as seems likely in most cases, the median is also the mean. This explains why augmentations do not make sense with many (uncertain) POCs. If the A-set is normal around the mean $\underline{b}$, the best choice is surely $\underline{b}$ as 50 % of the results are above β_M and 50 % are below.

Finally, it should be noted that each estimate (9) and (11) is an average of the *N*-set. Any average reduces the range. Hence, the CV of the *A*-set is smaller than the CV of the original *N*-set.

3.7 The average of the A-set is $\underline{b}$, not β_M

By a complete augmentation I mean that every POC is augmented – rightly or wrongly. All complete augmentations make the FAT = 0, and thus the estimates of (12a) and (12b) are the same whether or not they contain the FAT-term. This should be obvious from the formulas, and the reader will see that it is confirmed by the simulations made.

Thus, the FAT does not work in augmented FAT-PET MRAs. It might still work if some of the POCs are augmented for, but it certainly fails for the average POC.

Hence, the average of the A-set is $\underline{b}$, whether or not the estimates (9) contain the FAT-term. It means that if $\underline{b}$ biased, the average augmented meta-average is biased as well, which is the case if the FAT in the FAT-PET MRA differs from zero. This is another way to see that there is publication bias when the same-effect-assumption from 3.1 is broken.

If we start from a FAT-PET MRA, where the FAT shows asymmetry, i.e. $\beta_F \neq 0$, it is a risky business to start augmenting the FAT-PET MRA. It is a toss-up if it makes the estimate of β better or worse.

To summarize: The *N*-set of primary estimates generates one mean, one β_M-meta-average, and a large *A*-set of augmented meta-averages, β_A. The *A*-set has the mean as the central estimate, and the range of the *A*-set is much smaller than the one of the *N*-set. If the FAT $\neq 0$ in the FAT-PET MRA, the average augmented meta-average is biased.

4. The simulations 1: Level one disregarding the POCs

As level one disregards the POCs it is the case with zero or many POCs. Section 4.1 explains how the simulations are run. Then section 4.2 looks at the ideal funnel and section 4.3 simulates censoring.

4.1 The simulation technique used

The simulations use a known DGP, Data Generating Process and an EM, Estimating Model, which is estimated using OLS. The DGPs and EMs used are listed in Table 2. All DGPs and EMs contain the term of interest, βx, where $\beta = 0.25$, which is the true value of β throughout. The present section uses $DGP = EM = (1)$, disregarding the POCs.

Table 2. The four models with 0, 1 or 2 POCs: z_1 and z_2

DGP	Data Generating Process	***EM***	Estimating Model	Used in	POCs
(1)	$y_i = \beta x_i + \varepsilon_i$	(1)	$y_i = b x_i + u_i$	Section 4	0
(2a)	$y_i = \beta x_i + \gamma_1 z_{1i} + \varepsilon_i$	(2a)	$y_i = b x_i + g_1 \omega_1 z_{1i} + u_i$	Section 5	1
(2b)	$y_i = \beta x_i + \gamma_2 z_{2i} + \varepsilon_i$	(2b)	$y_i = b x_i + g_2 \omega_2 z_{2i} + u_i$	Section 5	1
(3)	$y_i = \beta x_i + \gamma_1 z_{1i} + \gamma_2 z_{2i} + \varepsilon_i$	(3)	$y_i = b x_i + g_1 \omega_1 z_{1i} + g_2 \omega_2 z_{2i} + u_i$	Section 5	2

Notes: The explanatory variables, x_i, z_{1i} and z_{2i}, are correlated, as proposed in section 3.2, and the ε's are generated noise, while the u's are noise in the EM. ω is a random generator of the numbers 0, 1. φ is symmetric to ω, so that $\omega_i + \varphi_i = 1$ for all i.

When we disregard the POCs the model becomes very simple: The models have no constant, so x has zero average, and so has the noise term ε. The variation in the experiments is generated by the two standard deviations: sd(x) and sd($\boldsymbol{\varepsilon}$) which in this section are both set at 1. This is so large relative to β that the funnels look realistic (compare with Stanley and Doucouliagos 2010).

In the typical β-literature most of the noise in x comes from different samples, and most of the noise in ε comes from model variation. With a large number of possible model variants ε will give an approximate stochastic noise. In the empirical funnels where no asymmetry is evident the funnels look rather like Figure 3.

All funnels have $N = 500$ observations generated by 500 regressions. I here follow the convention of starting regression 1 with 20 simulated data, and then regression 2 has 21 data, etc. till regression 500 that has 519 simulated data. This gives a realistic range of precisions as shown on the vertical axes of the funnels. In the censoring experiments some of these regressions are censored, so N is increased till 500 points are reached. In Tables 4 to 6 below each experiment (line) is calculated from R = 1,000 simulated funnels (thus it covers 500,000 regressions). However, in Paldam (2013b) each experiment covers R = 100 funnels only. This is also the case for Table 3.

4.2 An ideal funnel

The ideal funnel is shown as Figure 3 (that looks like Figure 1A). Here the $DGP = EM$ = (1), with $sd(x) = sd(\varepsilon) = 1$. When the noise terms are increased the funnel becomes shorter and broader, but keeps the same general shape.

Figure 3. The ideal funnel

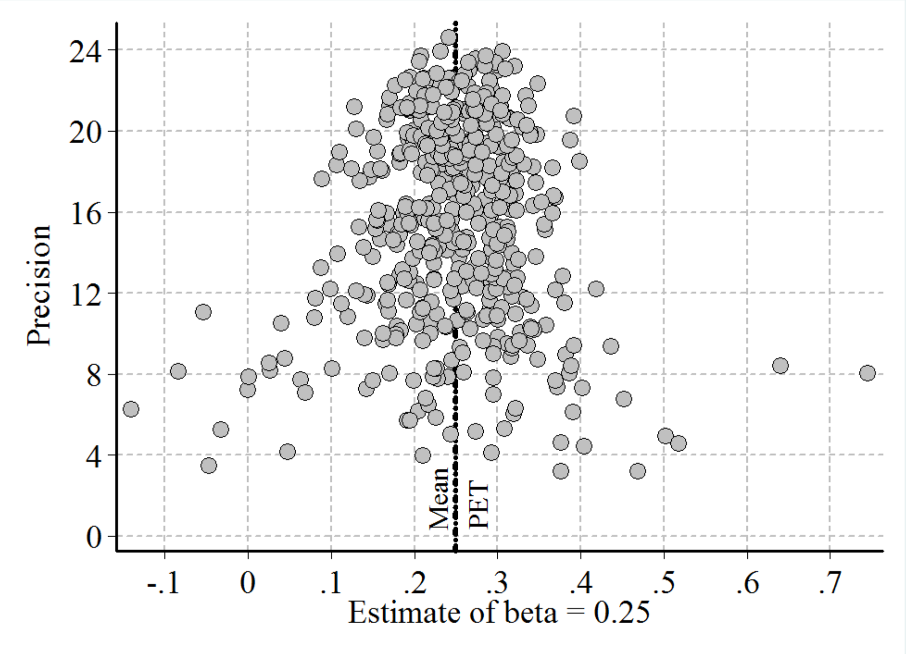


Table 3. Results for different *C*s, censoring point

C	Mean, $\underline{b}$	PET, β_M	*C*	Mean, $\underline{b}$	PET, β_M
-0.25	**0.250**	**0.249**	0.1	**0.257**	0.235
-0.2	**0.250**	**0.248**	0.125	**0.259**	0.232
-0.15	**0.251**	**0.248**	0.15	**0.264**	0.226
-0.1	**0.250**	**0.249**	0.17	0.265	0.227
-0.05	**0.250**	**0.249**	0.18	0.266	0.227
-0.025	**0.252**	**0.247**	0.19	0.266	0.228
0	**0.253**	**0.245**	0.2	0.275	0.230
0.025	**0.253**	**0.244**	0.25	0.294	**0.250**
0.05	**0.254**	**0.243**	0.3	0.338	0.291
0.075	**0.255**	**0.238**	0.35	0.406	0.336

Note: PET is the PET meta-average. Averages between 0.237 and 0.264 are bolded. R = 100 funnels are estimated for each row, so more than 1 million regressions was made to generate the table.

4.3 Censoring the funnel

Censoring of the funnel of Figure 3 is done by rejecting all points below a certain C-point and supplementing the number of simulation so that N remains at 500. The censoring experiments are run at the 20 C-points listed in Table 3. Table 3 shows that the mean is a little better than the PET if the censoring is small, but after $C \approx 0.18$ the PET becomes better.

Figure 4. The funnel from Figure 3 censored at the censoring point C = 0.2

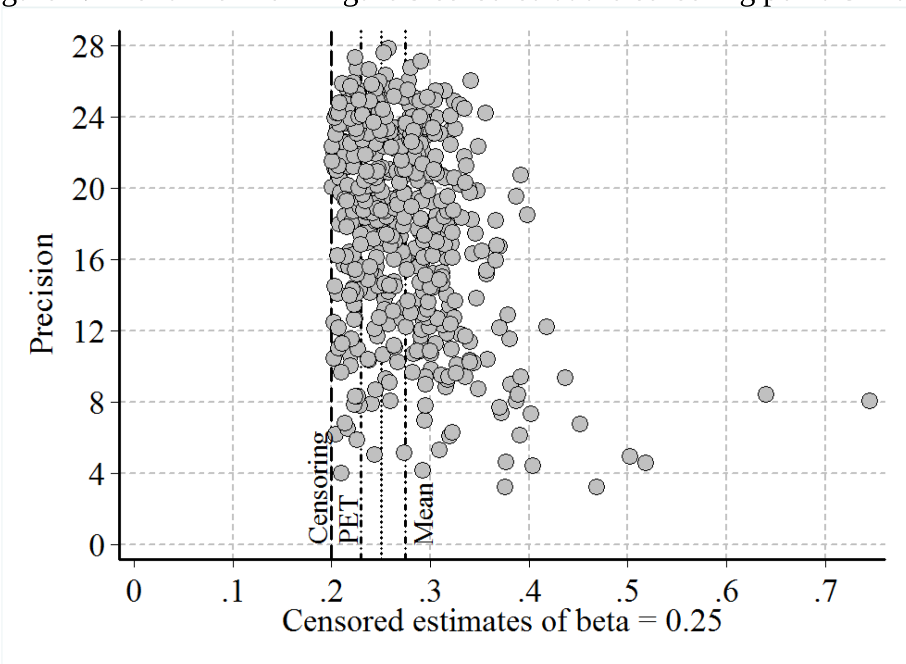


Figure 5. The bias of the mean, $\underline{b}$, and the PET, β_M from Table 3

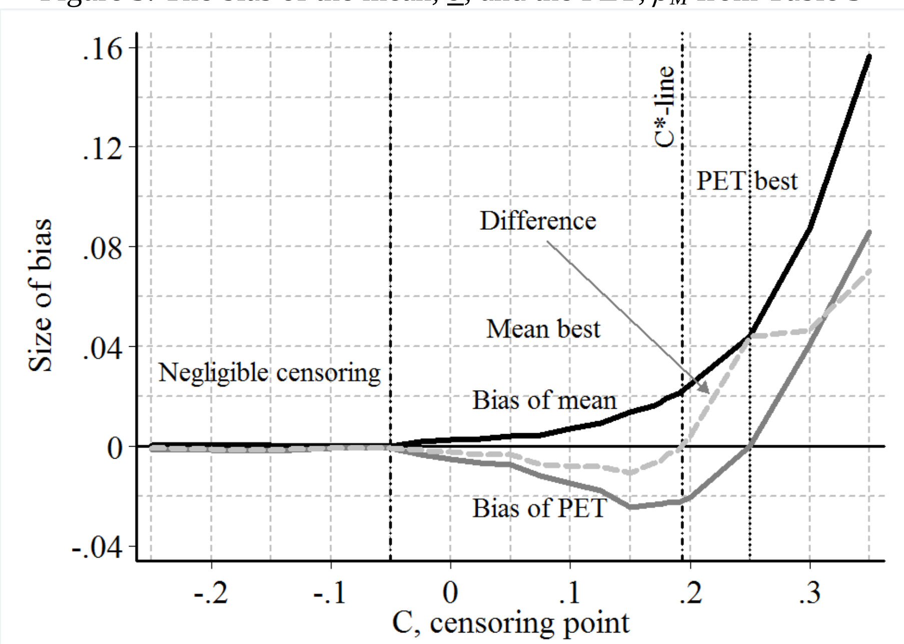


Figure 4 shows how the picture is when C = 0.2 where 22 % of the calculated estimates are censored. Here 639 regressions have been run. On the figure the PET is marginally better than the mean only.

This full pattern from Table 3 is showed on Figure 5. The mean is better than the PET for little to moderate censoring, but here the difference is small. However, with substantial censoring the PET is substantially better. Thus, it is better to use the PET when the funnel is censored. Note that if the censoring is at the true value β = 0.25 the PET gives a perfect estimate. When the censoring is larger also the PET is biased but not as much as the mean.

5. The simulations 2: Level two with one or two POCs

Section 5.1 surveys the cases covered, while section 5.2 describes the DGP and EMs used. Section 5.3 looks at cases with uncensored two-topped funnels, while 5.4 show what happens when these funnels are censored. Finally, section 5.5 identify a lucky case and conclude.

5.1 The cases covered

The three models with one or two POCs were shown in Table 2. They can be combined in the 9 ways listed in Table 4. To run the simulations with each combination four parameters (γ_1, γ_2, q, ρ) have to be chosen: γ_1 and γ_2 are the weights of the POCs in the DGP, q is the inclusion probability of the POC in the EM, and ρ is the correlation between x and the two POCs. This gives a large number of cases to cover even when both q and ρ are chosen to be the same for both POCs. Paldam (2013b) documents 182 systematic experiments (with R = 100). Fortunately, a rather simple pattern emerges.

Table 4. The nine combinations of the 3 equations with POCs from Table 2

	Combination		POC1, z_1			POC2, z_2			Right augmentation			Comment	
Case	DGP	EM	DGP	EM	Bias	DGP	EM	Bias	z_1	z_2	Rest	Same	Sym
(1)	(2a)	(2a)	Yes	Yes	OVB	No	No	Ok	φ_1	-	No	Yes	(5)
(2)	(2a)	(2b)	Yes	No	All	No	Yes	IVB	-	ω_2	Yes	No	(4)
(3)	(2a)	(3)	Yes	Yes	OVB	No	Yes	IVB	φ_1	ω_2	No	No	(6)
(4)	(2b)	(2a)	No	Yes	IVB	Yes	No	All	ω_1	-	Yes	No	(2)
(5)	(2b)	(2b)	No	No	OK	Yes	Yes	OVB	-	φ_2	No	Yes	(1)
(6)	(2b)	(3)	No	Yes	IVB	Yes	Yes	OVB	ω_1	φ_2	No	No	(3)
(7)	(3)	(2a)	Yes	Yes	OVB	Yes	No	All	φ_1	-	Yes	No	(8)
(8)	(3)	(2b)	Yes	No	All	Yes	Yes	OVB	-	φ_2	No	No	(7)
(9)	(3)	(3)	Yes	Yes	OVB	Yes	Yes	OVB	φ_1	φ_2	No	Yes	

Note: OVB is omitted variable bias, and IVB is included variable bias. 'Rest' is the remaining bias after the right augmentation. 'Same' means that the DGP and EM is the same. 'As' means that it is symmetrical to the other case mentioned.

5.2 The (DGP, EM)-combinations and the right augmentation if there is no censoring

Taking cases (1), (2) and (3) from Table 4 as examples will show what is going on:

In case (1) both the DGP and EM is (2a). As the POC z_1 is only included in Q of the N estimates, the FAT-PET has an OVB, omitted variable bias. It goes away when the relation is augmented with φ_1, while the augmentation with ω_1 is wrong.[16]

In case (2) the DGP is (2a) and the EM is (2b). Thus, the literature has not discovered

16. All 9 cases are estimated for various combinations of the 4 parameters in sections 4 and 6 in Paldam (2013b).

the right POC, z_1. Consequently, all estimates are biased. However, the EM uses z_2 instead. Both z_1 and z_2 are correlated with x so z_2 may turn significant.[17] Thus, z_2 is a false variable that gives an IVB, included variable bias, which is adjusted for by the ω_2-augmentation, but a bias remains due to the undiscovered z_1-POC. Here the φ_2-augmentation is wrong.

In case (3) the DGP is (2a) and the EM is (3). The EM thus contains one true and one false variable. The (φ_1, ω_2)-augmentation is right, and the (ω_1, φ_2)-augmentation is wrong, and so is the two remaining augmentations (ω_1, ω_2) and (φ_1, φ_2). They are both close to the mean.

When the EM contains one POC the number of possible augmentations is $2^1 = 2$, so the chance of doing it right is 50 %. When the EM contains two POCs the number of possible augmentations is $2^2 = 4$, so the chance of doing it right is 25 %. With 10 POCs the possible number of augmentations is $2^{10} = 1{,}024$, so the chance of getting it right is 0.1 %, etc.

The 9 combinations in Table 4 have four symmetrical cases, given in the 'Sym' column. So there are only 5 cases to study. When the number of POCs goes up the number of different cases rises very quickly.

5.3 *The two-topped funnels: without censoring*

All 9 cases are covered with examples and enough simulations in Paldam (2013b) to show the pattern. The simulations confirm the symmetries already discussed. Only 5 of the cases are different. Table 5 thus has five sections. The examples shown in this section use the parameters: Sd($\boldsymbol{x}$) = sd(z_1) =1, $\gamma_1 = 0.75$, $\rho = 0.7$, $q = 0.5$ and $N = 500$. Table 5 uses R =1,000 simulation of the funnel. Each line has thus needed ½ million simulated regressions. The results are stable $\pm$ 0.001.

The four *Avs* columns count the number of the 1,000 averages (in %) where it cannot be rejected, at the 5 % level, that the average is 0.25. It is reassuring to see that the bolded averages have *Avs*'es in the range from 92 to 95. In the same way the four *Fs*-columns are counts of the cases where the FAT = 0 is not rejected, so that symmetry is confirmed.

When a two-topped funnel occurs the two augmentations, β_{AR} and β_{AW}, find each top by taking out the other one. If the DGP is fully within the EM, one of these tops is the right one and one is false. It does not matter for the true top if the EM contains extra variables. However, if the EM misses a variable, all results are biased.

17. Both z_1 and z_2 are correlated with x, and hence they are also correlated with each other. With the correlation ρ^2 if ρ is large z_2 may work as a proxy for z_1, and thus actually improve the estimate. In most of our simulations $\rho = 0.5$ so $\rho^2 = 0.25$, and hence z_2 is a poor proxy for z_1.

Table 5. The 5 different cases, with different values of ρ

Case Tab 4	DGP	EM	ρ	Mean $\underline{b}$	Right augmentation β_{AR}	*Avs*	*Fs*	Wrong augmentation β_{AW}	*Avs*	*Fs*	PET meta-average β_M	*Avs*	*Fs*
(1.1)	(2a)	(2a)	0.7	0.513	**0.251**	94	94	0.776	0	94	0.806	0	0
(1.2)	(2a)	(2a)	0.2	0.363	**0.250**	93	95	0.474	0	95	**0.263**	92	0
(1.3)	(2a)	(2a)	-0.5	0.063	**0.250**	95	95	-0.125	0	95	0.096	0	90
(2.1)	(2a)	(2b)	0.7	0.775	0.776	0	95	0.776	0	95	0.775	0	95
(2.2)	(2a)	(2b)	0.2	0.475	0.474	0	95	0.474	0	95	0.474	0	95
(2.3)	(2a)	(2b)	-0.5	-0.125	-0.125	0	95	-0.124	0	95	-0.124	0	96
(3.1)	(2a)	(3)	0.7	0.513	**0.251**	94	94	0.776	0	94	0.950	0	0
(3.2)	(2a)	(3)	0.2	0.363	**0.250**	94	94	0.474	0	94	0.289	47	4
(3.3)	(2a)	(3)	-0.5	0.063	**0.249**	95	96	-0.125	0	96	-0.028	0	18
(7.1)	(3)	(2a)	0.7	0.031	-0.100	0	95	0.425	0	95	0.306	63	0
(7.2)	(3)	(2a)	0.2	0.157	0.099	0	94	0.324	0	94	0.099	0	13
(7.3)	(3)	(2a)	-0.5	0.406	0.500	0	94	0.125	0	94	0.412	0	98
(9.1)	(3)	(3)	0.7	0.294	**0.250**	94	94	0.425	0	94	0.431	0	0
(9.2)	(3)	(3)	0.2	0.269	**0.250**	94	94	0.324	0	94	0.243	91	47
(9.3)	(3)	(3)	-0.5	0.219	**0.250**	95	95	0.125	0	95	0.213	24	94

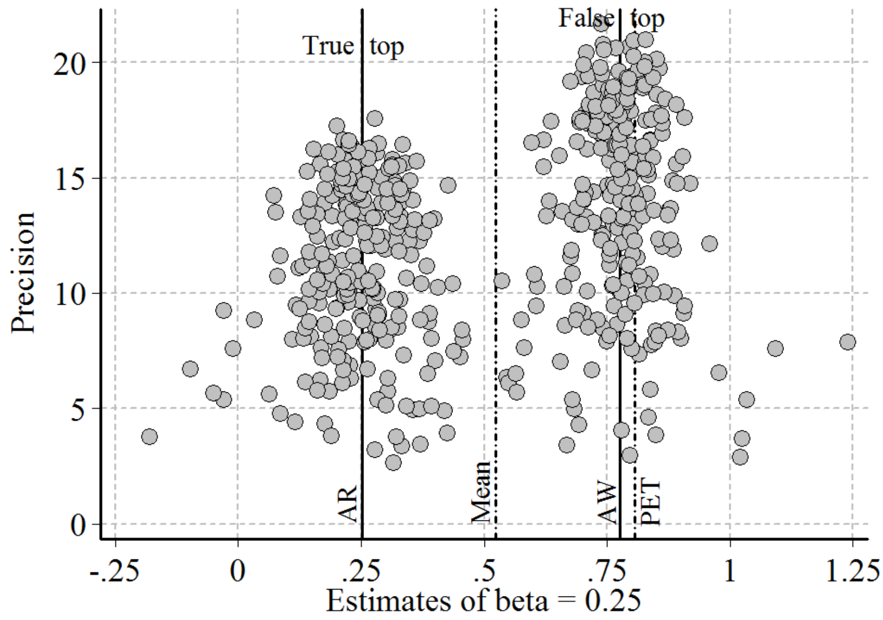
Notes: The left hand column refers to Table 4. Thus, (1.2) is the first experiment with case (1) from the table. Due to the symmetries only five of the 9 cases are covered. Averages from 0.237 to 0.264 are bolded. Each row reports averages for R = 1,000 funnels of N = 500 regressions. This is 7.5 mill regressions.

The table confirms that the mean $\underline{b}$ is always the average of β_{AR} and β_{AW}, while the PET is different. With two-topped funnels, β_{AR} and β_{AW}, find the two tops and β_M goes to the highest top – normally it overshoots that top. The table also confirms that the FAT-test is the same for the two augmentations. The *Fs*'s are in the range from 94 to 96 %. Fully augmented funnels are symmetric, so when the POCs are fully augmented the FAT turns powerless.

5.4 Two illustrations of cases from Table 5

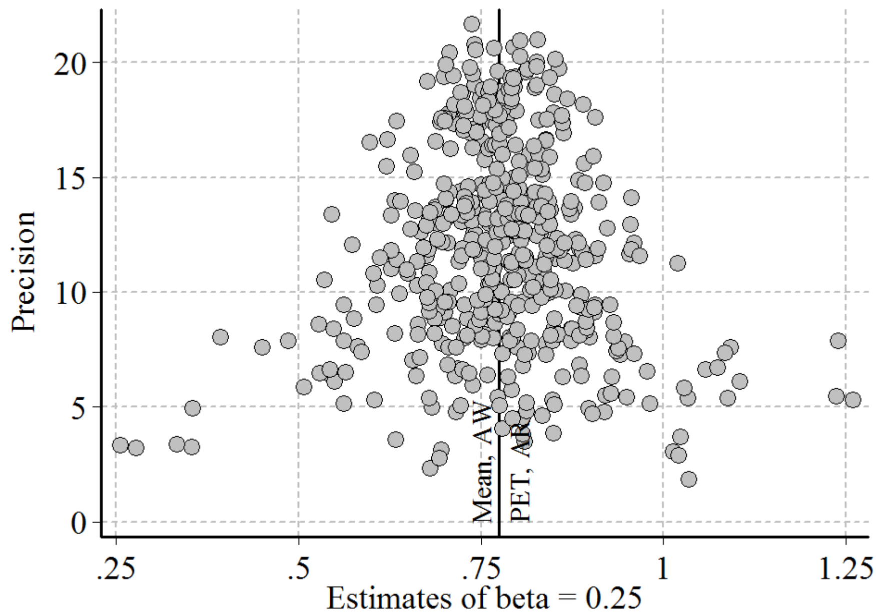
Case (1.1) with one true POC where the DGP = EM is illustrated by Figure 6, while the reverse case (2.1) with one false variable and a missing variable is shown as Figure 7. The averages included on the two figures are the ones from Table 5, and they look as they apply to the figures as well, so the two examples shown are typical of the 1,000 funnels estimated. Figure 7 shows that case (2) is puzzling. Here a variable is missing, and even if another variable is included it has a marginal effect only. However the funnel looks fine. All 9 FAT tests have *Fs*'es at 95. Much the same happened in case (7) but here the four averages differ.

Figure 6. Case (1.1) the DGP = EM = (2a)



Note: The case illustrated is the one of row (1.1) of Table 5. The false peak is at 0.25 + 0.7 *x* 0.75 = 0.775.

Figure 7. Case (2.1) the DGP = (2a), EM = (2b)



Note: The case of row (2.1) of Table 5. The peak is the same as the false peak on Figure 6.

5.5 Three cases of funnels with censoring

Finally, the two-topped funnels are censored. This assumes that here many more POCs with a small effect each, and one or two POCs that are really powerful. For easy intuition the experiments start with the two cases shown on Figures 6 and 7. Row (1.1) is the same as in

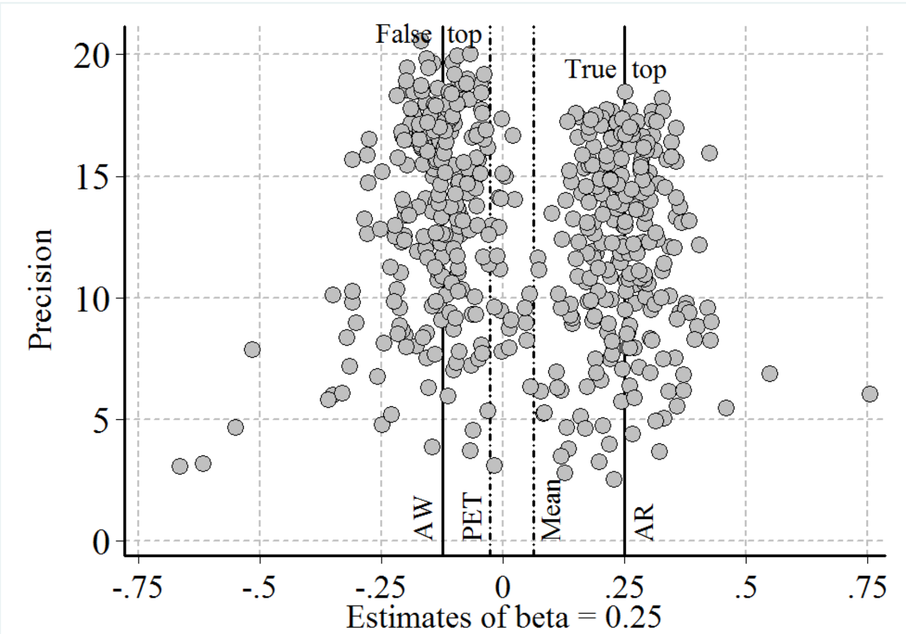
Table 5 and is also shown on Figure 6. Obviously, if the smallest estimates are censored, the false top becomes more prominent. This is precisely as happens. And as the two funnels have virtually no overlap case (1.5) shows how unlucky one can get.

Table 6. Three cases from Table 5 censored in two ways

Case Tab 5	DGP	EM	*C* censor	Mean $\underline{b}$	Right augmentation β_{AR}	*Avs*	*Fs*	Wrong augmentation β_{AW}	*Avs*	*Fs*	PET meta-average β_M	*Avs*	*Fs*
(1.1)	(2a)	(2a)	No	0.513	**0.251**	94	94	0.776	0	94	0.806	0	0
(1.4)	(2a)	(2a)	0.25	0.625	0.287	5	42	0.759	0	42	0.911	0	0
(1.5)	(2a)	(2a)	0.50	0.775	0.570	5	95	0.772	0	95	0.776	0	96
(2.1)	(2a)	(2b)	No	0.775	0.776	0	95	0.776	0	95	0.775	0	95
(2.4)	(2a)	(2b)	0.6	0.788	0.750	0	15	0.755	0	15	0.755	0	15
(2.5)	(2a)	(2b)	0.75	0.818	0.762	0	0	0.762	0	0	0.762	0	0
(3.3)	(2a)	(3)	No	0.063	**0.249**	95	96	-0.125	0	96	-0.028	0	18
(3.4)	(2a)	(3)	0	**0.251**	**0.239**	84	78	0.029	0	78	**0.244**	93	94
(3.5)	(2a)	(3)	0.25	0.398	**0.250**	95	0	0.207	98	0	**0.250**	95	0

Note: Each of the 3 sections starts with a row from Table 5. The next two lines are (x.4) and (x.5) that are new. They are the censored cases. *R* = 1,000 funnels are estimated for each row. This gives 3 mill new regressions.

Figure 8. Case (3.3) without censoring



The middle section of Table 6 is rather parallel to the story told in section 4.3. It only differs as it is all wrong because an important variable is missing in the EM. The censoring changes the weights of the number of rightly and wrongly augmented estimates, and both the mean and the PET may move outside the interval of the two augmented meta-averages. If we could have made a complete augmentation in this case then we would have got the mean as the average of any pair of symmetric complete augmentations.

5.6 A lucky illustration and concluding remarks

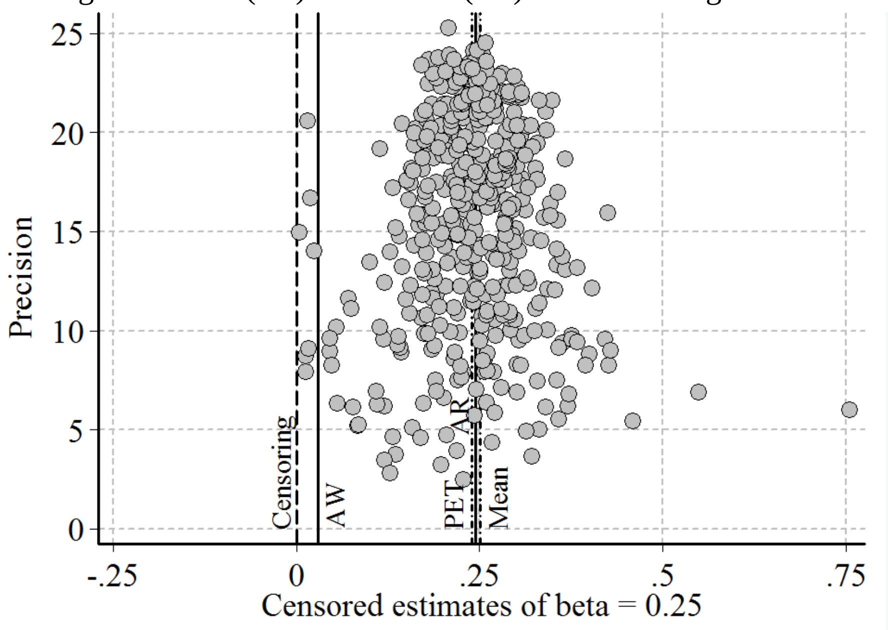
Figure 9 is case (3.3) from Tables 5 and Figure 8. The false top is at a lower β-level than the true top. As the main bias in the β-literature is that $\beta > 0$ Figure 9 shows a censoring at $C = 0$. This makes the funnel look ideal and all four averages are very close to $\beta = 0.25$.

There is no way to tell from the meta-analysis if we are in the lucky case of (3.4) where censoring rids the analysis of a false top or in the unlucky case of (1.1) where the censoring rids the analysis of the true top. This is surely a question for economic analysis.

The examples in this section show that with few POC only it is important to study the funnel and try to figure out what is going on. With few POCs this is normally easy, as there funnel has clear peaks and an augmentations exercise will tells exactly how the different POCs generate the peaks.

It has thus been demonstrated that while the PET meta-average works well adjusting for censoring, it is not the right tool to handle situations with a couple of POCs that give clearly multi-topped funnels. However, the FAT still rejects symmetry in these cases.

Figure 9. Case (3.4) that is case (3.3) with censoring at zero



6. Conclusion

A key concept in meta-analysis is the POC, partly omitted control variable, which is used in some primary studies and not in others. The nature of POCs is rarely stated in articles, so the inclusion of POCs is a gray area in economics. The paper has discussed three averages: the first two disregard the POCs. It is the mean and the PET meta-average.

The augmented meta-average corrects the mean for POC-biases, as should be done if the POC is a true model variable that is randomly excluded. The paper introduces a symmetrical augmentation if the POC is a false variable that is randomly included.

The reason why POCs are randomly included is often that authors disagree about models and hence they also disagree about the right augmentation. If the literature disagree about many variables very many augmented meta averages are possible – giving a large set of augmented meat averages. The average of this set is the mean (not the meta-average). The set has a substantial variation. That opens the meta-study for meta-mining that chooses augmentations that move the average in the direction wanted. This is a game that undermines the whole purpose of meta-analysis.

The set of possible augmentations can be somewhat reduced: Some POCs are *ceteris paribus* control for differences between data samples. They should differ between studies. It is wrong to augment the meta average with such variables.

Other POCs are parts of the model. They should be included in all or none of the models. In published results they are often omitted when they are insignificant. The average should not augment in such cases. They should be included as zero in the average precisely as done by the mean.

However, the choice of POCs may also be systematically influenced by their effect on the estimate of β giving a *selection loop* from the estimate to the choice of POCs. This loop is at the core of publication bias. It shows up as a censoring asymmetry for the funnel that is detected by the FAT.

If the FAT detects an asymmetry the mean and the PET meta average differs. If the asymmetry looks like a censoring at a predictable place, we can be confident that the literature has a publication bias. Then the PET meta-average is the best average. And we know that the effects of the POCs are different in the cases where they are included and excluded. This means that the augmentations give biased results.

References:

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

Doucouliagos, H., Paldam, M., 2013a. The robust result in meta-analysis of aid effectiveness: A response to Mekasha and Tarp. *Journal of Development Studies* 49, 584–87

Doucouliagos, H., Stanley, T.D., 2012. Are All Economic Facts Greatly Exaggerated? Theory Competition and Selectivity. *Journal of Economic Surveys* 27, 316-39

Durlauf, S.N., Johnson, P.A., Temple, J.R.W., 2005. Growth Econometrics. Chapter 8, pp 555-677 in Aghion. P., Durlauf. S., eds. *Handbook of Economic Growth.* North-Holland. Amsterdam

Egger, M., Smith, G. D., Schneider, M., Minder, C., 1997. Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal* 316, 629-34

MAER-Net (Meta-Analysis in Economics and Business), URL: http://www.hendrix.edu/maer-network/

Paldam, M., 2013. Simulating publication bias. Department of Economics WP 2013-27. Also available from http://www.martin.paldam.dk/Meta-method.php.

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

Paldam. M., 2013b. Net Appendix to: How do partly omitted control variables influence the averages used in meta-analysis in economics? Available from http://www.martin.paldam.dk/Meta-method.php.

Sala-i-Martin, X., 1997. I just ran two million regressions. *American Economic Review* 87, 178-83

Sala-i-Martin, X., Doppelhofer, G., Miller, R.I., 2004. Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach. *American Economic Review* 94, 813-35

Stanley, T.D., 2008. Meta-regression methods for detecting and estimating empirical effect in the presence of publication bias. *Oxford Bulletin of Economics and Statistics* 70, 103-27

Stanley, T.D., Doucouliagos, H., 2010. Picture this: A simple graph that reveals much ado about research. *Journal of Economic Surveys* 24, 170-91

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

Stanley, T.D., Doucouliagos, H., Giles, M., Heckmeyer, J.H., Johnson, R.J., Laroche, P., Nelson, J.P., Paldam, M., Poot, J., Pugh, G., Rosenberger, R.S., Rost, K., 2013. Reporting Guidelines for Meta-Regression Analysis in Economics. *Journal of Economic Surveys* 27, 390-4

Sturm, J.-E., de Haan, J., 2005. Determinants of long-term growth: New results applying robust estimation and extreme bounds analysis. *Empirical Economics* 30, 597–617

Appendix: The terminology of meta-analysis and the simulations

Terminology of the β-literature, the numbers M and N	
The parameter β	$\beta = \partial y/\partial x$ is the effect of interest. It is the effect of x on y
The β-literature	All papers containing estimates of β
Main prior in profession	Dominating prior in the literature. We assume $\beta > 0$
M-set of papers	All papers in the β-literature, within the search window of the meta-study
N-set of estimates: $(b_1, \dots, b_N)$	Each b_i has standard error, s_i, t-ratio, t_i,= b_i/s_i and precision, $p_i = 1/s_i$
The primary estimation equation and the POCs, the numbers L and K	
Estimating equation (linear)	$y_j = b_i x_j + [g_1 z_{1j} + \dots + g_K z_{Kj}] + u_j$. The []-bracket holds K POCs only
L and $K << L$	The β-literature uses L controls
POC, z, partly omitted control	The POC is included in $Q < N$ estimates, and excluded in $S = N - Q$
The funnel giving the distribution of the N-set	
Funnel: distribution of N-set	The $(b_i.\ p_i)$-scatter. It should be lean and symmetric, but often it is neither
Funnel width: CV of N-set	CV is coefficient of variation, the standard deviation divided by the mean
The meta-analytical tools	
FAT, Funnel Asymmetry Test	β_F, which is estimated jointly with the PET meta-average
FAT-PET MRA	Meta regression: $b_i = \beta_M + \beta_F\, s_i + u_i = \beta_M + \beta_F/p_i + u_i \rightarrow \beta_M$ if $p_i \rightarrow \infty$
Inclusion of z: $(\omega_1, \dots, \omega_N)$	$\omega_i = 1$ if z is included in estimate else 0, $\Sigma_N\, \omega_i = Q$
Exclusion of z: $(\varphi_1, \dots, \varphi_N)$	$\varphi_i = 1$ if z is excluded in estimate else 0, $\Sigma_N\, \varphi_i = S$. Thus, $\omega_i + \varphi_i = 1$
Augmented FAT-PET MRA	$b_i = \beta_A + \beta_F\, s_i + \lambda_1\, a_{i1} + \dots + \lambda_L\, a_{iL} + u_i$. For POC z_j a_{ij} is the φ-vector if z_j is a true model variable, or the ω-vector if z_j is a false model variable
The three averages and the publication bias	
Mean, $\underline{b}$	Arithmetic average. Disregards funnel asymmetries. May also be weighted
PET meta-average, β_M	Corrects the mean for funnel censoring
Augmented meta-average, β_A	From (i) or (ii). Corrects PET meta-average for random POC biases
Publication bias	$PB = \underline{b} - \beta \approx \underline{b} - \beta_M$. Main prior $\beta > 0 \Rightarrow \beta_F < 0 \Rightarrow \underline{b} > \beta_M \Rightarrow PB > 0$

Values in Simulations of the integers		
R	The number of funnels estimated	R = 1,000 in Tables 4 to 6 and R = 100 in Table 3
N	The number of estimates, b_i, in the funnel	N = 500. When censoring more regressions are run [b)]
L	The number of POCs in the experiment	L = 1, 2
D_i	D_i is the sample size used to estimate b_i	D_i = 20, 21, 22, …519 [b)]

Data generating process and estimating model

DGP $y_t = \beta\, x_t + [\gamma_1\, z^t_{1t} + \dots + \gamma_K z^t_{Kt}] + \varepsilon_t$ EM $y_t = b\, x_t + [g_1\, \omega_1\, z^t_{1t} + \dots + g_T\, \omega_T\, z^t_{Tt}] + u_t$,

Most of the POCs in the EM are the same as in the DGP, but there might be more or less

As the DGP has no constant all averages have zero mean, they have sd's of 1 in most cases

Parameters of the DGP and EM		Sizes chosen
β	Parameter of interest in DGP, $\beta = \partial$y/∂x	$\beta = 0.25$
ρ_j	Correlation of POC j and x in DGP	ρ_j = 0.25, 0.5 and 0.75
ε_t	Noise in DGP	Average 0, sd = 1
ω_j	Inclusion of POC j in EM	δ_j = 0.25, 0.5 and 0.75
z	If z is included in DPG, it is true, else false	K is 1 or 2
Q	$Q = \sum_N \omega$. The share $q = Q/N$ is inclusion probability	q = 0.5 in most experiments

Notes: When some estimates are censored N is increased, and thus the Ds increase as well. Note that s_i is uses both as the standard error of the estimates and as the share of the estimates, where the POC z is not included.