

Economics Working Paper

2010-1

Natural funnel asymmetries
A simulation analysis of the three basic
tools of meta analysis

Laurent Callot and Martin Paldam

School of Economics and Management
Aarhus University
Bartholins Allé 10, Building 1322
DK-8000 Aarhus C - Denmark
Phone +45 8942 1610
Mail: oekonomi@econ.au.dk
Web: www.econ.au.dk

Natural funnel asymmetries

A simulation analysis of the three basic tools of meta analysis

Laurent Callot¹ and Martin Paldam²

School of Economics and Management, Bartholins Allé 10 (building 1322)
Aarhus University, DK 8000 Aarhus C, Denmark

Abstract:

Meta-analysis studies a set of estimates of one parameter with three basic tools: The funnel diagram is the distribution of the estimates as a function of their precision; the funnel asymmetry test, FAT; and the meta average, where PET is an estimate. The FAT-PET MRA is a meta regression analysis, on the data of the funnel, which jointly estimates the FAT and the PET. Ideal funnels are lean and symmetric. Empirical funnels are wide, and most have asymmetries biasing the plain average. Many asymmetries are due to censoring made during the research-publication process. The PET is tooled to correct the average for censoring. We show that estimation faults and misspecification may cause natural asymmetries, which the PET does not correct. If the MRA includes controls for omitted variables, the PET does correct for omitted variables bias. Thus, it is important to know the reason for an asymmetry.

Keywords: Meta-analysis, funnel asymmetry, meta average

JEL: B4, C9

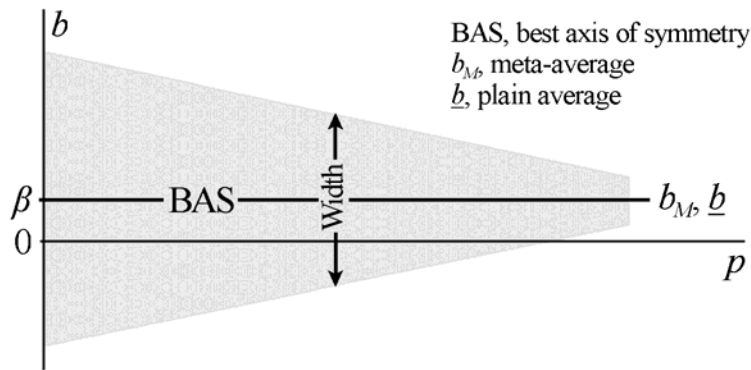
1. E-mail: lcallot@creates.au.dk

2. E-mail: mpaldam@econ.au.dk, URL: <http://www.martin.paldam.dk>

1. Introduction: The problem and the set-up of the analysis

Meta analysis summarizes a literature on one parameter of interest. The best average may be reached by giving more precise estimates a higher weight. Accordingly, the *funnel diagram* shows the distribution of the estimates over their precision. If the estimates are representative and properly estimated, the funnel has an ideal symmetric form. The *meta average*, b_M , is on the axis of symmetry of the ideal funnel. Many funnels are asymmetric, so the meta average has to be estimated by a method correcting for the asymmetry. The PET estimate of the meta average corrects the average for asymmetries due to censoring. The purpose of this paper is to study other ways in which funnels can be asymmetric, and how this affects the PET estimate.

Figure 1. Shape of the ideal funnel



1.1 Basic definitions and concepts

The N estimates of β are b_i with standard error s_i , and precision $p_i = 1/s_i$. The funnel diagram is the scatter of the (b_i, p_i) -points. The funnel narrows with increasing precision. We distinguish between *empirical* and *natural* funnels. Empirical funnels take the estimates from the literature on β . Natural funnels contain only estimates published with no loops of result-based corrections. We can only be sure that a funnel is natural in simulations.

Conditions are ideal if the N estimates are made with the right estimator on the true model. Ideal funnels have the form sketched on Figure 1. It is lean and symmetric around the BAS, best axis of symmetry, which intersects the b -axis in the meta average, b_M . It is close to the plain average, \underline{b} , and both are good estimates of the true value β .

The authors do not know of any field of economics where conditions are ideal. Researchers always disagree about models and estimators. This is evident in the *excess-variation* result: Empirical funnels are amazingly wide compared to the ideal. In addition, most

empirical funnels are asymmetric. Meta-analysts typically consider asymmetries to be due to priors which cause censoring that should be corrected. Increasingly precise methods have been developed to estimate the meta-average, when the funnel has been censored.

This paper is due to a nagging suspicion: Maybe there are natural funnel asymmetries due to estimation faults and model misspecification? If so, what happens to the meta-average?

1.2 The simulation set-up of Table 1: The DGP and the model

Consequently, we simulate the ideal case and a set of cases with known problems. Each experiment generates a funnel, a funnel asymmetry test and an estimate of the meta average. The simulations run each experiment 10,000 times to make statistics showing how often the test picks up the problem, and how often the meta average deviates from the true value.

Consequently, one experiment consists of N estimates where each is done on M observations from the same DGP, i.e. *data generating process*, simulating model and data uncertainty. Two conventions are used: (i) The parameter of interest $\beta = 1$. (ii) The estimates are OLS which assumes a linear *model* with residuals $\varepsilon_i = N(0, \sigma^2)$, i.e. they are normal, uncorrelated and has a constant variance.

If the DGP fulfils conditions (ii), the ideal funnel appears. We study what happens when the DGP deviates from the model. The field is new, so we chose simple, tractable assumptions about the DGP: It has non-normal residuals, it is non-linear, and it contains an explanatory variable that is correlated with x , which is omitted in the estimate.

Table 1. The simulation set-up for funnels of N estimates b of β

The two versions of the data generating process (DGP)		
DGP1	$y_j = \alpha + \beta x_j + \gamma_1 z_{1j} + \dots + \gamma_k z_{kj} + \varepsilon_j$	Parameter of interest: β
DGP2	$y_j = \alpha + \beta x_j^n + \varepsilon_j$, where $n \geq 1$	If $n > 1$, the DGP is non-linear
The controls	The z -variables in (1) are k controls	The potential number of controls is K : $k \leq K$
Two conventions	$\beta = 1$	All estimates are OLS with $n = 1$
One experiment is the simulation of one funnel with $N = R \cdot N_M$ estimates		
Estimates	One estimate: $i = 1, \dots, M$ observations	One set: $N_M = 25$ estimates. $M = 10, 20, \dots, 250$
One funnel	Each set is run $R = 10, 20$ times	One experiment: $N = R \cdot N_M = 250$ or 500 estimates
Data uncertainty, model uncertainty and estimation faults		
Data uncertainty	<i>Default</i> : Residuals $\varepsilon = N(0, \sigma^2)$	<i>Fault</i> : Residuals in DGP are non-normal
Model certainty	<i>Default</i> : $K = 0$. DGP2, with $n = 1$	<i>Fault</i> : DGP is non-linear, $n > 1$
Model uncertainty	$k < K$. DGP1, random choices of z 's	Omitted variable misspecification

Note: All experiments have been replicated 10,000 times, giving $N \cdot 10,000$ regressions for a simulation.

1.3 *The literature*

The medical literature uses meta-analysis routinely. Since Light and Pillemer (1984) it is common to meet brief discussions on the proper way to draw and interpret funnels. A number of recent papers discuss interpretations of actual funnels that look puzzling, see e.g., Tang and Liu (2000), Terrin, Schmid and Lau (2005) and Lau et al (2006); but they do not really discuss if asymmetric funnels could occur naturally.

Meta studies are increasingly common in economics, and many funnels have been published. Stanley and Doucouliagos (2009) is a fine introduction to the use of funnels in economics, with illustrative examples, showing that many funnels are asymmetric, see also Robers and Stanley (2005) for a collection of studies. The two closest predecessors to the present paper are: Koetse et al (2005) which discusses how the meta-average is affected by omitted variable biases. Stanley (2008) briefly discusses the effect of misspecification on the meta average, but uses a restrictive specification for the bias in his simulation. Where results overlap, we confirm their results.

1.4 *Content: Data dependency, estimation faults and misspecification*

Section 2 is a brief survey of the standard meta-theory of censoring. From there we push into (almost) virgin territory where the analysis is done by means of simulation as described.

Section 3 deals with *data dependency*. It is the common case in macroeconomics, where researchers have to use the same basic data set. It does expand over time, but it might also have important breaks. In this case funnel asymmetries are disturbingly common.

Section 4 deals with *estimation faults*: The residuals are non-normal and the true functional form is non-linearity. This is analyzed for model certainty, where $k = K = 0$, and the funnel is due to data uncertainty only. The funnel is robust when these faults are moderate.

Section 5 deals with *misspecification* in the form of omitted variables. This is the case of model uncertainty, where $k < K$. The funnel is due to both data and model uncertainty. This produces a range of funnel asymmetries, which in unlucky cases may look like censoring.

Both sections 4 and 5 demonstrate that the PET estimate of the meta-average often fails on funnels with natural asymmetries. Thus, it is important to distinguish between censoring and natural asymmetries.

Many important issues in meta analysis are disregarded. This applies to other weighting schemes, such as by the quality of journal, and by the date of the research. We also disregard the clustering of estimates by paper or author.

2. The standard theory of funnel asymmetry: It is due to censoring

The standard theory assumes that asymmetries are made by the research/publication process, which uses result-based loops where the researcher calculates one set of results, studies these result, makes correction in model, estimator and data and re-estimate.

2.1 *Research as a search process: Three stopping rules*

The very term *research* implies a search process, and empirical research means that the search takes place in a data set. It typically allows the researcher to report a *range of results*. In most cases it is possible to argue that various parts of the range are the proper one.

This is not only a theoretical observation. Everybody coming to meta-research appears to be amazed by the width of most empirical funnels. It is common that results reported in perfectly decent journals, for the same β , are statistically different at levels of significance such as 0.01%, or e.g. by a factor 5, so *b*-ranges are often wide. We refer to this observation as the *excess-variation* result.

When the research process leads to a paper reporting a result, it is consequently generated by the *stopping rule* ending the search. The rule surely differs from paper to paper, and it is often difficult even for the author to know the decisive stopping rule for a paper; but it appears that stopping rules are of three types which all have to be fulfilled:

- S1: The result fulfils the *priors* of the researcher, i.e., she likes it.
- S2: The result is *statistically satisfactory*, i.e., it satisfies current econometric standards.
- S3: The result is deemed to be *marketable* on the market for economic papers.

The researcher may think that he has found the truth, but the criteria that lead to the conclusion may also be that the result fulfils his priors. Perhaps S2 and S3 are more of the nature of constraints, but then S3 may also be the dominating concern.

2.2 *Priors lead to censoring: Systematic censoring give biases*

Four types of priors are commonly recognized:

- P1 *Theoretical priors*: Some part of the *b*-range cannot be true by theory. Hence, in large data-samples they ought to be impossible. This leads to censoring in small samples. Example: Censoring of positive price elasticities for (non-Giffen) goods.

- P2 *Political priors*: A part of the b-range is politically/morally unpalatable. Hence, it is censored. Example: Censoring of negative values for development aid effectiveness.
- P3 *Economic priors*: Researchers may work in areas where they have interests, and censor accordingly. Example: Discrimination by researchers of the Labor Movement against results showing that minimum wages generate unemployment.
- P4 *Polishing prior*: To reach marketable results, unclear results are censored. This follows from S3. This prior works in all directions and will not be discussed at present.

A censoring bias is caused by researchers, who, due to prior, prefer results in the right range. Think of the example from (P1) where $\beta = -1$ is the true price elasticity. Here +0.5 and -2.5 are the same distance from truth, and equally easy to reach. Maybe one identifying assumption gives +0.5 while another gives -2.5. The researcher knows that +0.5 cannot be true. Consequently, the second identifying assumption is better, and -2.5 is the result. Nothing in the process of choice is dishonest, or even unreasonable.

With randomly distributed priors (as in P4), they give variation in the results (i.e. wide funnels), but not asymmetries. However, if one prior dominates – as it may in the price elasticity example – the β -funnel is censored, and the plain average is biased. The typical censoring bias is caused by, e.g., 75% of the writers in the field having the same prior, while 25% have several other priors, so the funnel is not perfectly censored, but thin in certain parts. If the profession wants to know the best estimate from the β -literature, it has to turn to the meta-average.

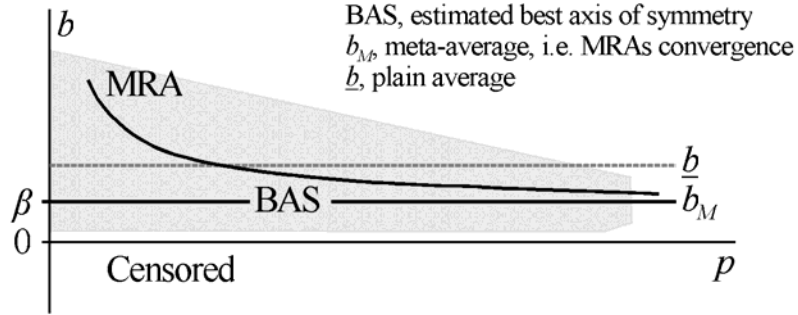
2.3 *The meta-average, b_M : The FAT-PET MRA of T.D. Stanley*³

If the literature search is exhaustive, the funnel represents all the work the profession has put into estimating β . It may be 100 studies, and consequently it represents more than 50 man-years of work. Thus, it seems almost unbearable not to take advantage and use the best average of the funnel as the best estimate of the value of β available.

Figure 2 is the same funnel as Figure 1, but it is censored for negative values and it looks obviously asymmetric. Censoring makes the funnel leaner and it causes the plain average, \underline{b} , of the funnel to be a bad estimate of β . If a likely prior against negative values exists, we are confident that we understand what is going on.

3. Meta-analysis is more than 30 years old in medicine and 20 years in economics, so it makes sense to speak of “standard” theory, but the *meta average* has been developed in the last few years notably by Stanley (2008). Equation (1) is termed the MRA in the paper, and we refer to its two terms as the FAT-term and the PET-term.

Figure 2. The funnel of Figure 1 censored for negative values



The meta-average, b_M , is still the intersection of the same BAS as on Figure 1 with the b-axis. But now we have to make an estimate that fills in the censored estimates. Stanley's idea is an expression that starts in the middle of the fat end of the funnel and converges to the BAS. The resulting MRA (meta regression analysis) is written in two equivalent ways:

$$\text{MRA:} \quad b_i = b_M + \gamma s_i + u_i \quad (3a) \quad \Leftrightarrow \quad t_i = b_M p_i + \gamma + v_i \quad (3b)$$

Here (3b) is reached from (3a) by dividing with s_i . The PET estimate of the meta-average is b_M , while γ is the FAT, i.e., the funnel asymmetry test, u_i and v_i are residuals. As a preview of the findings the PET is a good estimate of the meta average, if the asymmetry is due to censoring, but not normally in other cases, while the FAT works in all cases examined.

Formulation (3a) gives an easier intuition, while (3b) is preferable to estimate, as it has less heteroskedasticity. The estimate produces a hyperbola converging to the BAS for s falling (i.e. p rising); see Figures 3, 6b, 7 and 8b. That is, b_M is the point to which b converges for $p \rightarrow \infty$, as drawn on Figures 1 and 2. Even if the top part of the funnel or bottom part (as on Figure 2) is censored, the MRA still reaches the same result. However, if the highest precision part of the funnel is affected by the bias, the MRA fails to find the BAS, and hence b_M becomes a poor estimate of the true value, see section 5.

2.4 The MRA(k) expanded with n binary specification dummies k_1, \dots, k_n

Section 5 analyzes what happens if the estimating equation suffers from an omitted variable, giving the estimate an OV bias. We show that if some of the researchers detect the said variable(s) and others do not, the funnel will – in most instances – become asymmetric.

The best practice meta analysis to deal with the problem is to *detect* the OV, z , and including a binary control variable $k(z)$, with a value for all N estimates analyzed in the MRA.

If the estimate is controlled for z , $k(z) = 1$. Otherwise, it is zero. The coefficient to $k(z)$ is an estimate on the bias. This demands that some of the N estimates are controlled for x and others are not. The recommended approach is to form a set of binary control variables $k_1 = k(z_1)$ to $k_n = k(z_n)$ for all such possible OVs and run an expanded MRA:

$$\text{MRA(k):} \quad \begin{aligned} b_i &= b_M + \gamma s_i + \lambda_1 k_1 + \dots + \lambda_n k_n + u_i & (4a) \text{ or after division with } s_i \\ t_i &= b_M p_i + \gamma + \lambda_1 (k_1 / s_i) + \dots + \lambda_n (k_n / s_i) + v_i & (4b) \end{aligned}$$

The relation between (4a) and (4b) is the same as the one between (3a) and (3b).

Assume the true process is generated with a set of K controls, while each model is estimated with $k_i \leq K$ controls. If $K \subset \bigcup_i k_i$, i.e. the true model is included in the union of all the model estimated, it is possible to obtain a correction for the MRA, which makes it converge to the true value even when the funnel is plagued by OVB. The complement of k_i is termed k_i^c in the set K . If the MRA is modified with dummies for each element of k_i^c the control is zero if the estimate is controlled for x and 1 otherwise.

The MRA(k) allows us to see if each of the n z -variables matters. The paper introducing a new explanation, z , claims that it is important, and z is surely significant in that paper. Once it has been included in a set of papers, it is possible to see if z gives a robust improvement of the estimation model for β or not, by running the MRA(k) for $k(z)$. So in any case it is useful to run expanded MRAs.

The problem arises, if the OV remains undetected in the meta analysis. Hence, the MRA is uncontrolled for $k(z)$. Section 4 simulates cases with undetected OVs, to analyze what the MRA does. It is shown that it often causes the MRA to pick a wrong b_M . It is then demonstrated that the relevant MRA(k) does remove the problems.

As an aside, we may mention that (4) has also been used to study the effects of estimators. Does it matter to use the X-estimator instead of the simple OLS estimator? Such studies have shown that it rarely does, see e.g. Doucouliagos and Paldam (2009c).

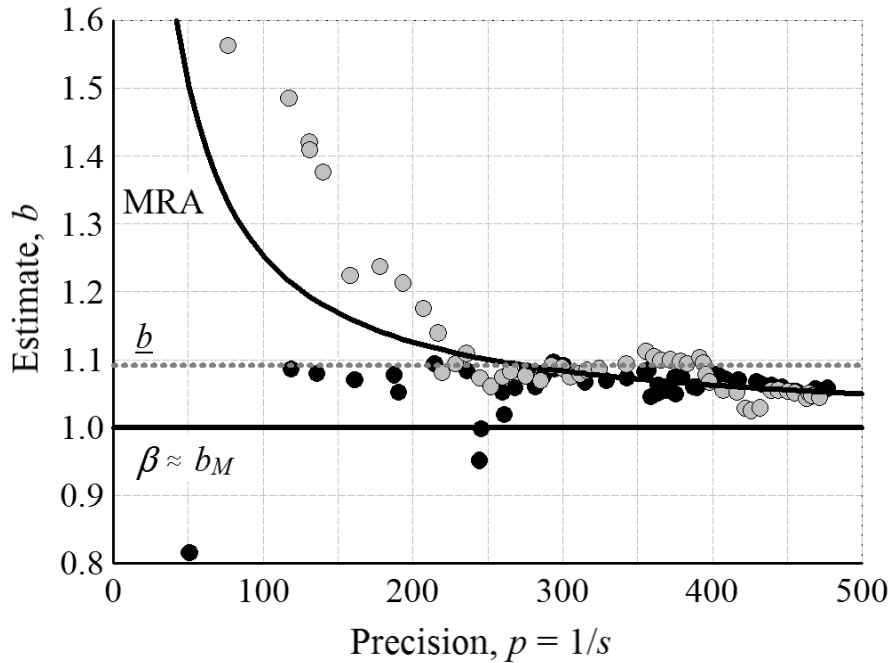
The method used in the paper is to simulate funnels from data generated with controlled DGPs, and to compare the form of the funnels with the ideal one. Also, we compare the true value, $\beta = 1$, with the plain averages, \underline{b} . Finally, we estimate the MRA or the MRA(k) to see if the FAT-statistic, γ , detects asymmetry, and if meta-averages, b_M , find the true value, β , or at least is closer to β than is the plain average, \underline{b} .

3. Data dependency: A problem in empirical macroeconomics

Sections 4 deals with estimation faults and section 5 considers misspecified models. Thus, they look at problems caused by mistakes of researchers. The present section considers a more innocent problem. In some fields of economics researchers are forced to use the same data, as time passes new data are added, but there is a great deal of data dependency. Our set-up allows us to study how long Type I errors may persist in such cases. Data dependencies may also create funnel asymmetries. Furthermore, we show that the likelihood of asymmetries increases, when dependent data have structural breaks.

Figure 3 is the first example of a simulated funnel used to illustrate the argument. The figure shows the format of the following funnels. It includes the MRA, as the black curve, estimated with equation (3); the plain average \underline{b} , as the gray dotted line; and the meta average, b_M , to which the MRA converges. In most cases it is shown as a black dotted line, but on Figure 3, it is hidden by the solid black line at $\beta = 1$.

Figure 3. Two dependent-data paths simulated as described in text, $N = 100$.



3.1 *Sequentially revealed data generates path dependency in the funnel*

Consider the sequence of data d_1, \dots, d_N , where $d_j = (x_1, \dots, x_{mj})$ and $d_{j+1} = (x_1, \dots, x_{mj}, x_{mj+1})$. M is the maximum sample size generated. The coefficient of interest is then sequentially

estimated on the subsample (x_1, \dots, x_t) then on (x_1, \dots, x_{t+1}) until (x_1, \dots, x_M) . It is well-known how such estimates look – with 95% certainty they start at a point within the 95% confidence interval around the true value of β , and then they converge to β .

Figure 3 shows the path of two such sequential estimates, chosen to be reasonably representative. We may think of them as the results based on the data from two countries. In the case with gray markers, the convergence of the curves to β is quite slow. In the case with black markers, it starts low, but then jumps up and down a few times. Here convergence is much faster. Thus, we should keep in mind that quirks may create amazing path dependency. Table 2 estimates how often it actually does.

Let us imagine that the points on Figure 3 are the estimates analyzed in the meta study – the funnel for these points suggests strong asymmetry. The MRA confirms this impression, but fortunately the MRA converges to $b_M \approx 1$. However, the plain average \underline{b} will for long differ significantly from 1.

In macroeconomics data-mining leads to Type I errors – acceptance of false models reached by polishing a quirk in the data – as witnessed by the excess significance result mentioned. The risk of Type I errors makes the usual scientific requirement of *independent replication*⁴ particularly important in macroeconomics. The path dependency of quirks means that just to add a few years of observations is not enough for a serious independent replication. Table 2 shows the consequence of this. It is quite easy to stay with a wrong model for a time.⁵

3.2 Structural breaks

Another frequent problem in time series is changes in the value of the parameter of interest, in the form of structural breaks, where β increases to β_{new} . It is obvious that the estimates of the parameter of interest fall in precision immediately after the structural break, compared to estimates obtained before the break, once the sample size after the structural break has increased sufficiently, the fall in precision turns. This has interesting consequences for the funnel and the MRA: If the break is located relatively early in the sample, the MRA will converge to β_{new} while if it is located in the end of the sample, the MRA will converge to β because the most precise estimates will be those made on the samples before the break.

4. *Independent replication* means replications of the same model by other researchers on new data. *Dependent replication* of a model is by other authors on the same data. In economics there are even cases of dependent replications that fail. The classical horror story is Dewald et al (1986). The ensuing discussion is surveyed in e.g. McCullough et al (2008).

5. Doucouliagos and Paldam (2009a) deal with meta-studies in a field of macroeconomics, which has seen a number of Type I errors, where a false model is accepted and dominates the literature for a period of 4-6 years.

Table 2. Overlapping sample cases without and with structural breaks

Section 1 Dimensions			Section 2 no breaks			Section 3 Break after 75%			Section 4 Break after 25%		
N	M	Nr series	Meta	Frequency		Meta	Frequency		Meta	Frequency	
			average	$\gamma \neq 0$	$b_M \neq 1$	average	$\gamma \neq 0$	$b_M \neq 1$	average	$\gamma \neq 0$	$b_M \neq 1$
20	100	2	0.999	0.204	0.210	1.112	0.351	-	2.119	0.999	-
100	100	10	0.999	0.206	0.206	1.113	0.350	-	2.120	0.999	-
50	250	2	1.000	0.371	0.364	1.135	0.849	-	2.147	1.000	-
250	250	10	1.000	0.373	0.363	1.136	0.853	-	2.148	1.000	-
200	1000	2	1.000	0.625	0.622	1.147	1.000	-	2.124	1.000	-
1000	1000	10	1.000	0.627	0.627	1.147	1.000	-	2.124	1.000	-
Averages			1.000	0.402	0.399	1.132	0.733	-	2.130	1.000	-

Note: Each line is repeated 10,000 times so each of the 3 sections contains 16.2 mil regressions. “Nr series” are the number of sequences. That is, Figure 3 shows two series. The break is made by increasing $\beta = 1$ to $\beta_{new} = 2$. Remember from section 1.1 that N is the number of points in the funnel, while M is the number of observations in each estimate.

3.3 The case of data dependency (section 2 of Table 2)

Table 2 contains a set of simulation experiments, each repeated 10,000 times. The table has four sections marked with a change of background color. The first section describes the dimensions of each experiment. N is the number of points in each funnel. M is the number of observations on which each regression (each point in the funnel) is run. Nr series are the number of dependent data sets included, the reader may think of each series as a country.

Section 2 looks at the case with no structural breaks, only data dependence. In average the meta average always converges to 1 as it should. Consequently, there is no systematic bias due to overlapping samples. However, funnels generated using the method described above are asymmetric much more often than in the independent sample case. Consequently, a biased meta average, b_M , occurs in an important share of the funnels. Interestingly, this increase with sample size, but not to the number of independent series used to generate the funnel.

Thus in about 40% of the cases data dependencies will generate a funnel where not only the plain average, but also the meta average is wrong. Thus the meta analysis will show a wrong result due to a quirk in the data which has not disappeared for the number of observations existing.

3.4 Data dependency and structural breaks (sections 3 and 4 of Table 2)

Sections 3 and 4 of Table 2 deal with structural breaks. In section 3 they take place late, i.e. after 75% of the sequence. In section 4 they occur early, i.e. after only 25% of the sequence.

The structural breaks are done by a shift of β from 1 to 2, at the breakpoint. Hence, $b_M \neq 1$ per definition, and it makes no sense to report the frequency of rejection of $b_M \neq 1$.

When the break is late, the funnel is asymmetric in most cases, and the MRA converges to a value close to the value of β in the early part of the sample. In fact, it converges to something like 1.13. When the break is early, the funnels are asymmetric in virtually all cases, and the MRA converges to the latter value of the parameter. However, it is not precisely 2, but about 2.13.

Before we leave the case of dependent data, we should state that it appears to give problems which are rather more serious than normally assumed. In some cases it helps to calculate the meta average – using the standard MRA – and not stick to the plain average.

4. Estimation faults: Analyzed for model certainty

This section deals with model certainty which is defined as $k = K$ in our set-up. It is modeled, with no loss of generality, by setting $K = 0$. Section 4.1 surveys the estimation faults analyzed in sections 4.2 to 4.4. In each case we bring illustrations showing typical specimens, so that the reader can see how everything looks. The MRA estimates for the funnels used in the illustrations are given in Table 3 in section in 4.5, while Table 4 in section 4.6 shows the percentage point for 10,000 replications.

4.1 The model faults considered

With $K = k = 0$ DGP1 and DGP2 from Table 1 are:

$$y_j = \alpha + \beta x_j + \varepsilon_j \quad (1) \quad \text{and} \quad y_j = \alpha + \beta x_j^n + \varepsilon_j \quad (2)$$

where $\partial y / \partial x = \beta$, in (1) and $\partial y / \partial x = n\beta x^{n-1}$ in (2)

Thus the funnel is due to the variation generated by the data and:

- (4.2) DGP1 and $\varepsilon_i = N(0, \sigma^2)$. The estimating model is right giving an ideal funnel.
- (4.3) DGP1, but ε_i is non-normal. We use log-normal for the main results. Within reason funnels keep the ideal form.
- (4.4) DGP2, with $n > 1$, so that the estimating model is misspecified. Within reason, funnels keep the ideal form, but the estimate of β goes bad, as it depends on the range of x-values covered by the data.

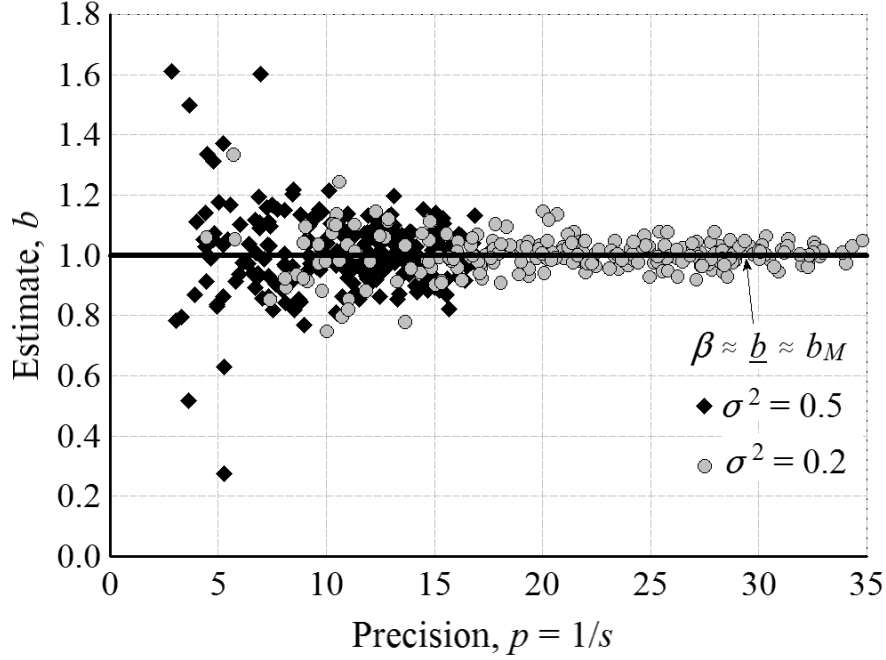
The words *within reason* in (4.3) and (4.4) refers to the fact that researchers know they should watch out for residual non-normality and model non-linearity, and econometric packages have diagnostic tests to help in the watch. If the problem is large, it will surely be detected. Thus, we only need to be concerned about *moderate* non-normality and *gentle* non-linearity.

4.2 Linear model with normal residuals: Ideal funnels are lean and symmetric

Here DGP1 is estimated with the right model, $N = 250$ times for $\sigma^2 = 0.2$ and $\sigma^2 = 0.5$. The two funnels look as they ideally should, so the plain average and the meta-average both give the same (good) estimate of the true value. In spite of the large residual variation both funnels are *lean* compared to the empirical funnels published in the typical meta-study (see Stanley

and Doucouliagos, 2009), and to the simulated funnels with estimation faults and model uncertainty. This is the precise meaning of the *excess-variation* result. It is also clearly visible if Figure 4 is compared to the later funnels of the paper.

Figure 4. Two funnels for model certainty and normally distributed residuals



Note: Each funnel has $N = 250$ points. The MRA is indistinguishable from the $\beta = 1$ -line.

The ideal case is easy to solve analytically for the shape of the reference funnel. Recall that the variance of the estimator of a linear model decreases linearly with the sample size.

$$\hat{\sigma} = \frac{1}{n} \sum_{i=1}^n u_i^2 \quad (5)$$

$$Var[b|x] = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (6)$$

Where n is the sample size. If the assumptions of the linear regression models are satisfied, it is an unbiased estimator. This explains the triangular form of the funnel converging to the true value of the parameter of interest.

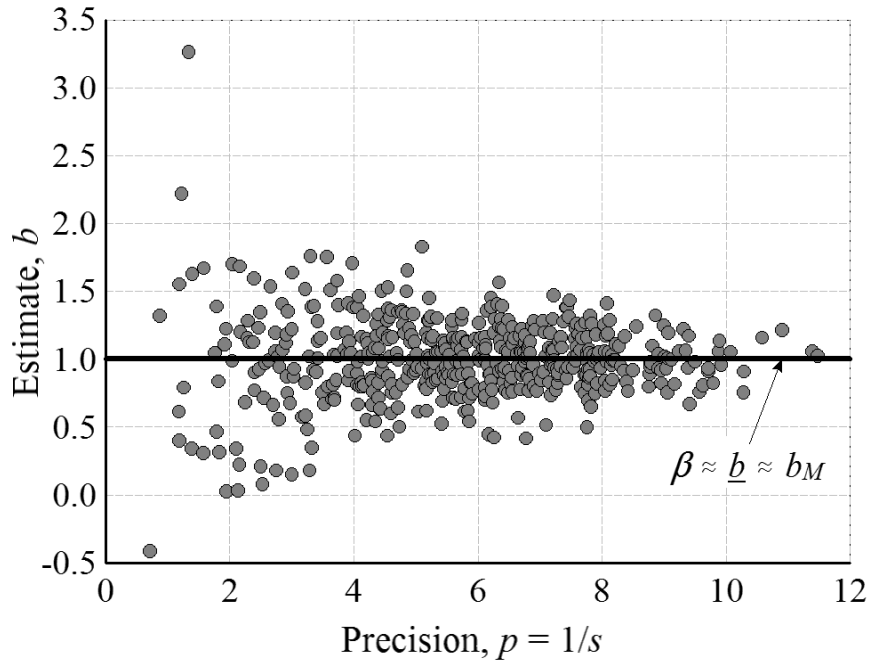
4.3 Estimation Fault 1: The DGP has non-normal residuals. The funnel form is robust

Within our set-up it is easy to make the ε -term moderately non-normal. Three groups of experiments were made: (i) log-normal ε 's, (ii) Weibull ε 's and (iii) t-distributed ε 's. The funnels were estimated by OLS, disregarding the problem of the residuals.

The t-distribution is symmetrical and so is the funnel. All it does is to make the funnel narrower in the middle. However, the log-normal and the Weibull distribution are both asymmetric. Nevertheless, all three experiments produced ideal-looking funnels with a horizontal BAS (best axis of symmetry) intersecting the b -axis close to 1. Thus, both averages (\underline{b} and b_M) are good estimates of the true value, see Table 4 below.

Figure 5 shows a log-normal ε -distribution that may escape detection. Here the model is, $y = x + \ln \varepsilon$, with $\sigma^2 = 0.5$. The funnel is perfectly symmetrical, but the fault in the residuals reduces the precision of the estimates, so the funnel is broader and shorter. However, with enough estimates both the plain and the meta-average are close to 1. Obviously one can generate asymmetric funnels by making the ε -term really skew, but then it will be detected by the β -researchers.

Figure 5. A funnel with a (disregarded) log-normal disturbance term



Note: Generated for $\sigma^2 = 0.5$ for $N = 500$. The MRA is indistinguishable from the $\beta = 1$ -line.

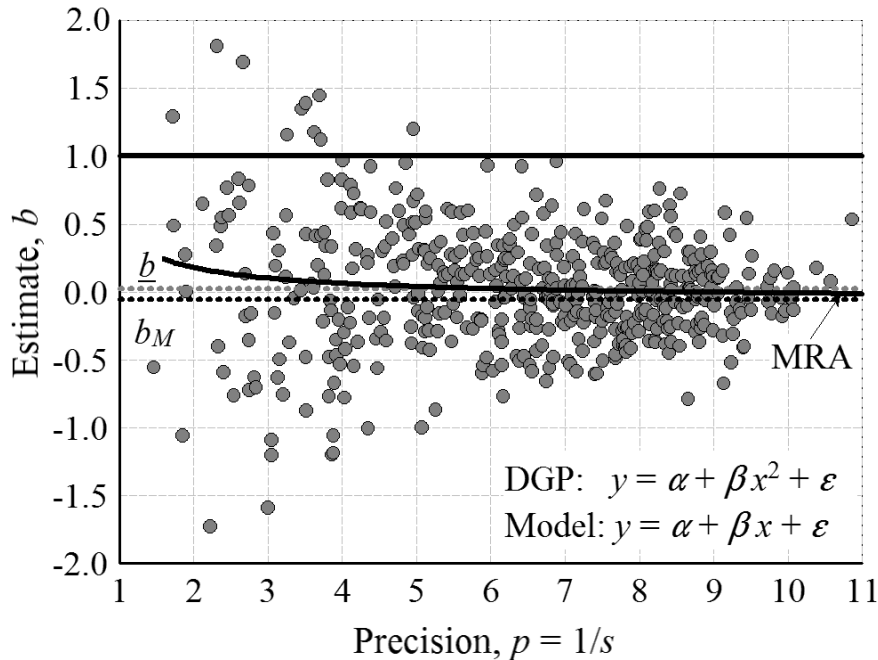
4.4 Estimation fault 2: The DGP is non-linear in x . The funnel form is still robust

We now turn to the case where the DGP is non-linear – using DGP2 for $n > 1$ – which is estimated by OLS as if it was $y_j = \alpha + \beta x_j + \varepsilon_j$. Here we use $\sigma^2 = 0.5$.

If the true form is so smooth that it can be approximated by a Taylor expansion, we only need to study what happens to the funnel if the function is in the power $n = 2, 3, 4, \dots$, where even $n = 2$ is high, but to get asymmetry, we continued to $n = 3$.

The quadratic DGP is $y_j = \alpha + \beta x_j^2 + \varepsilon_j$. Here the funnel looks as shown of Figure 6a. The estimation fault makes the funnel much wider. However, it still looks symmetrical. The MRA is the (vaguely) hyperbolic line shown. It detects no asymmetry (see Table 3) and converges to something close to the plain average, but even if the plot looks symmetrical, the density is much bigger on the higher part of the funnel so the symmetry is misleading. It causes the MRA to find almost the same meta-average as the plain average; but the two averages are $\underline{b} = 0.027$, and $b_M = -0.055$. Even if we multiply by 2, it is still rather far from the true value of 1.

Figure 6a. The funnel for a quadratic form misspecification (with $\sigma^2 = 0.5$ and $N=500$)

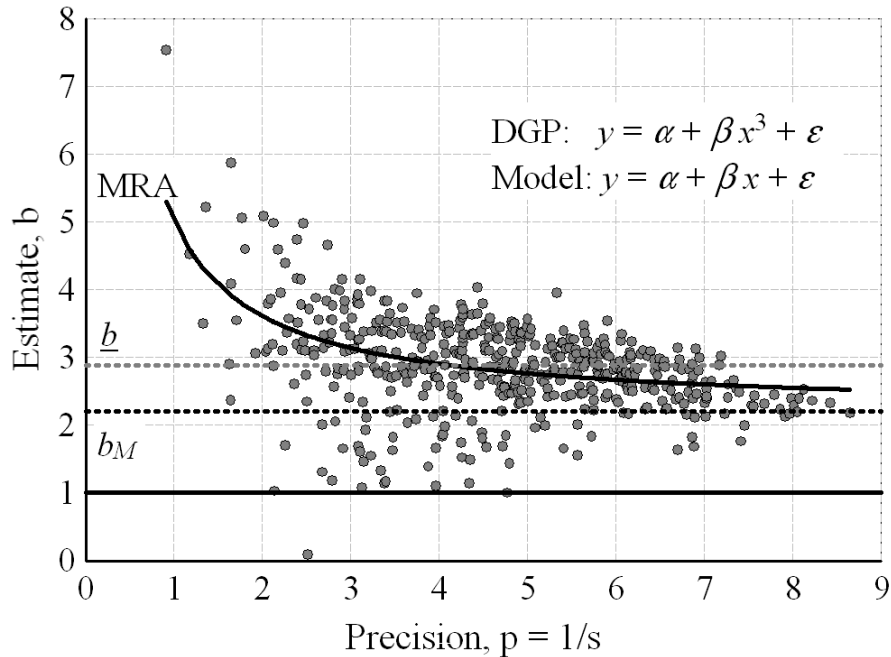


The cubic DGP is $y_j = \alpha + \beta x_j^3 + \varepsilon_j$. Figure 6b shows what happens in a typical experiment: the funnel becomes still wider, and finally it looks asymmetric. This is confirmed by the FAT-

part of the MRA which rejects symmetry – the reader may note that the MRA now looks much more hyperbolic, and the two averages are different from the true value $\beta = 1$, and, in addition, they differ from each other as $\underline{b} = 2.884$, and $b_M = 2.201$. These results get worse if we multiply by 3.

The MRA “looks” for censoring. Consequently, it treats the asymmetry as due to censoring below ap. 2, which causes the convergence to the meta-average mentioned. If the reader looks at the asymmetry of the funnel, it is not surprising that the MRA treats it this way, but it is not what is needed to find the true value.⁶

Figure 6b. The funnel for a cubic form misspecification (with $\sigma^2 = 0.5$ and $N = 500$)



For now we note that the MRA handles one type of misspecification quite badly. Below we shall see that it handles other misspecifications equally badly. This is not surprising as it was designed for a different purpose, but it is important to note.

The two functional form misspecifications are both rather large, and it is assuring that the misspecification has to be as bad as on Figure 6b before the funnel becomes asymmetric.

6. It is not obvious how censoring and natural asymmetries can be sorted out in the case of unnoticed non-linearity. The asymmetry on Figure 6b looks as bit, but not fully as a typical censoring. Also, it appears that economic theory has few priors against β 's below 2. Also the excessive width of the funnel points to something different from censoring which should make the funnel leaner.

4.5 Comparing the two averages in the cases of Figures 4 to 6

Table 3 shows what the MRA does in the 5 experiments. The graphs from 3 to 4 show funnels looking like the ideal form. Here both averages catch the true value of $\beta = 1$ rather well – the plain average being marginally better, as it should.⁷ In the two cases of model-form estimation fault, the MRA works rather badly when it tries to find the true value of β as already discussed.

Table 3. Comparing the true value, β , the plain average, b , and the meta-average, b_M

Figure	All estimates linear OLS		MRA-estimate		Plain avr. \underline{b}	True value β
	Model and funnel	Residual σ^2	Symmetry γ	Meta avr. b_M		
	4	Normal residuals, stubby	0.5	0.18 (0.9)	0.990	1.009
4	Normal residuals, lean	0.2	-0.10 (-0.5)	1.007	1.003	1
5	True residuals log-normal	0.5	-0.01 (-0.1)	1.010	1.005	1
6a	True equation quadratic	0.5	0.47 (1.4)	-0.055	0.027	1
6b	True equation cubic	0.5	2.82 (7.9)	2.201	2.884	1

Note: All estimates use Equation 2 for $k = K = 0$.

Finally, another estimation fault should be mentioned: It is the unit root bias. It may escape detection, but is easy to correct for once it is detected. Doucouliagos and Paldam (2009b) is a meta study of 212 estimates of a coefficient of inertia ϕ in a relation $h_t = \alpha + \phi h_{t-1} + u_t$, where α is a constant and u_t are the residuals. Here a univariate study of the data shows that ϕ is close to 1, so the estimate of ϕ does have a substantial unit root bias. It is demonstrated that this bias causes a natural funnel asymmetry. When the MRA is used to correct this bias, the result is worse than no correction.⁸

4.6 A systematic study the four main cases of the linear and non-linear DGPs

The results in 4.1 to 4.4 are based on one experiment with one funnel showing 500 estimates. Table 4 shows the results of adding of experiments and repeating each one 10,000 times. The Table has 5 sections where sections 2 to 5 corresponds to Figures 2 to 6b as indicated.

The ideal case of Figure 3 is considered in the left gray section 2. When the standard 5% level of significance is used, we want to get a 5% rejections of asymmetry, $\gamma \neq 0$, and the

7. The meta-average uses some degrees of freedom to correct for a problem that is not present, and hence it is marginally worse, but it really does not matter.

8. The case is more complicated as the literature also has a censoring bias, so the funnel is very asymmetric.

meta-average should reject that $b_M = 1$ in 5% of the cases. The rejection rates are slightly higher in both cases, so the MRA errs a marginally towards Type I error.

Table 4. Experiment with ideal funnel, residual non-normality and non-linearity

Section 1	Section 2			Section 3			Section 4			Section 5			
	The DGP is linear						The DGP is non-linear						
Dimensions	Ideal			Log-normal residuals			Squared x-term			Cubic x-term			
Each line is	Meta	Frequency of		Meta	Frequency of		Meta	Frequency of		Meta	Frequency of		
4 experiments	avr.	rejection		avr.	rejection		avr.	rejection		avr.	rejection		
N	b_M	$\gamma \neq 0$	$b_M \neq 1$	b_M	$\gamma \neq 0$	$b_M \neq 1$	b_M	$\gamma \neq 0$	$b_M \neq 1$	b_M	$\gamma \neq 0$	$b_M \neq 1$	
M													
40	100	0.999	0.062	0.065	1.001	0.062	0.070	-0.000	0.055	0.829	1.656	0.823	0.504
100	100	1.000	0.056	0.059	1.000	0.049	0.059	-0.001	0.047	0.998	1.618	0.999	0.909
1000	100	1.000	0.053	0.054	1.000	0.049	0.055	0.000	0.041	1.000	1.600	1.000	1.000
100	250	1.001	0.061	0.063	1.001	0.060	0.070	0.001	0.053	0.826	1.653	0.833	0.495
250	250	1.000	0.055	0.055	1.000	0.055	0.063	0.004	0.043	0.998	1.619	0.999	0.916
2500	250	1.000	0.050	0.051	1.000	0.052	0.060	0.000	0.046	1.000	1.600	1.000	1.000
400	1000	0.999	0.066	0.067	1.001	0.057	0.067	0.003	0.055	0.826	1.658	0.825	0.496
1000	1000	0.999	0.056	0.054	1.000	0.057	0.066	-0.001	0.044	0.998	1.621	0.999	0.920
10000	1000	1.000	0.055	0.055	1.000	0.053	0.061	-0.001	0.042	1.000	1.601	1.000	1.000
Averages		1.000	0.057	0.058	1.000	0.055	0.063	0.001	0.047	0.942	1.628	0.935	0.780

Note: N is the total number of points in each funnel, $\sigma^2 = 0.5$ in all simulations. The significance level for the rejection rates is the standard 0.05. Each experiment is repeated 10,000 times so the total number of regressions made to generate the table is about 600 million.

Next we look at section 3 reporting the simulations of the estimate with log-normal residuals, as illustrated on Figure 4. The rejection rates are virtually the same as for the ideal funnels. As the funnel is broader slightly fewer reject that $b_M = 1$, but slightly more find the funnel asymmetric. However, the deviations from the results for ideal funnel are amazingly small. In both cases the rejection rate falls with N , the number of points in the funnel.

In the two cases reported in sections 3 and 4 of the table of non-linearity, $b_M = 1$, is rejected. This is not surprising, but it is surprising that the FAT rejects symmetry less often for the squared form than in the ideal case – though only marginally so. In the cubic case the FAT does reject symmetry quite strongly. It is only accepted in 6.5% of the cases.

5. Misspecification, i.e. omitted variables: Model uncertainty

All meta-studies we know of show a lot of model-uncertainty. This appears to be the main explanation of the excess variability observation. Our set-up generate model uncertainty by using a DGP with K variable, and an estimate with $n < K$ control variables, z_1, \dots, z_n . As all K variables have some explanatory power, this generates omitted variable bias, OVB. We examine what an undetected OV does to the funnel and the MRA. Section 2.3 reported that the standard cure for this problem is to detect the OV, and generate the appropriate binary dummy and use the MRA(k), to account for the bias. This procedure is applied in 5.5.

5.1 One undetected OV (omitted variable)

With one OV model (1) becomes:

(3a) True model $y_j = \alpha + \beta x_j + \delta z_j + \varepsilon_j$, the data generating process, β is unbiased

(3b) Biased model $y_j = \alpha + \beta_z x_j + \varepsilon_j$, where β_z has an OV bias

The controls, z , are generated with a correlation coefficient ρ with respect to x using the following process:

$$z_j = \lambda x_j + \varepsilon_j, \text{ where } \lambda = \sqrt{\frac{\rho^2}{1-\rho^2}} \text{ and } \varepsilon_j = N(0, \sigma^2)$$

Model (3a) and (3b) are equivalent in two cases: The expected value $E(z) = 0$ iff $\lambda = 0 \rightarrow \rho = 0$. Here z and x are independent, and $\beta_z = \beta = 1$. In all other cases there is a positive OV bias. When β_z is estimated (by OLS) in model (3b), the bias is a simple linear relation:

$$\begin{aligned} \hat{b}_z &= (X'X)^{-1} X'Y = (X'X)^{-1} X'X\beta + (X'X)^{-1} X'Z\delta + (X'X)^{-1} X'\varepsilon \\ E[\hat{b}_z] &= \beta + (X'X)^{-1} X'Z\delta = \beta + \lambda\delta \end{aligned} \quad (7)$$

5.2 The leanest funnels are the ones with the biggest OV bias

The funnel for $\delta = 0$ is the *true* funnel, with average 1 as it should, but for other values of δ OVBs occur, and in many cases the biased $\beta_z > \beta$, and hence the MRA goes to β_z . This is, at a first consideration, counterintuitive, for it means that the most precisely estimated β 's are the most biased ones. Recall the estimator of the variance of b , the estimated value of β :

$$\text{Var}[b|x] = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (6)$$

This implies that the more variation in x , the more precisely estimated the coefficient is. Thus when no control is included in a model whose DGP contains a control, all the variation in the data is contained in x , implying a more precise estimated coefficient than when the model is estimated also with the control, where the variation in x becomes smaller. It becomes harder to predict the precision of b when several controls are used in the DGP and that they are randomly included or omitted as it depends mainly on the level of correlation between the control and x (and thus among the controls themselves).

5.3 Two examples of funnels with one OV in half the estimates

To understand what is going on, we start with the simplest setting. The true DGP is a model with the variable of interest and one control. The control is generated with an effect size termed δ and a correlation to the variable of interest denoted ρ . These two parameters determine the omitted variable bias (OVB) (see equation 7), and the variance bias. They enter as a product in the formula for the OVB, but their effect on the variance is less straightforward.

Figure 7a. Merged funnel for $\delta = 1, \rho = 0.5$ and $N = 500$ (line * in Table 5)

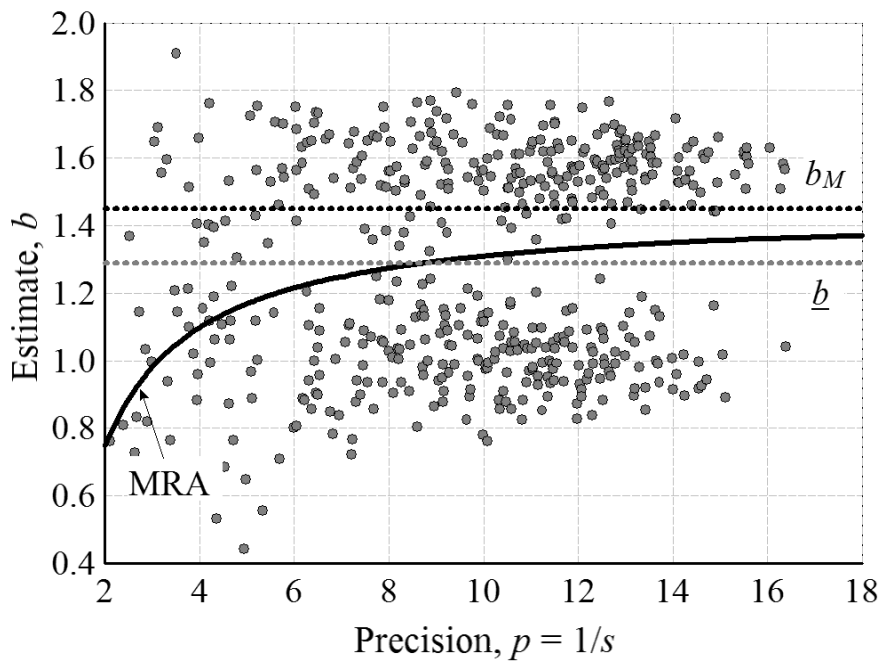
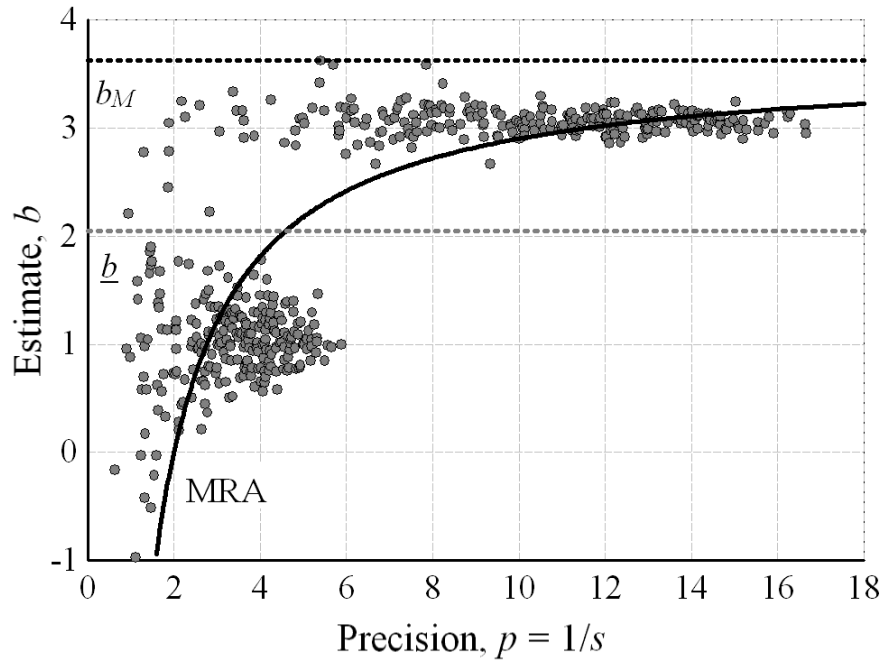


Figure 7b. Merged funnel for $\delta = 1, \rho = 0.9$ and $N = 500$ (line ** in Table 5).



Note: The two funnel both generated for 250 right estimates and 250 estimates with OVB.

The two illustrative cases, are shown as Figure 7a, where the z variable has $\delta = 1$ and $\rho = 0.5$, and Figure 7b, where the variable has $\delta = 1$ and $\rho = 0.9$. In both cases the reader can see that the funnel is bimodal,⁹ with two sub-funnels appearing as *peaks* in the horizontal direction. The sub-funnels at $b = 1$ are the right one, while the other is the OVB-peak at b_z . It is biased as the estimate is not controlled for the z -variable. The probability of inclusion of the control is 0.5, so the two sub-funnels have the same number of observations.

In both cases drawn the MRA converges to the OVB-peak and not to the right one, and consequently: $\beta < \underline{b} < b_M$. Consequently, if the meta analysis does not include a control for z in the MRA, the meta average is worse than the plain average. However, Table 5 shows that this is not a general result.

5.4 A systematic study of funnels with one undetected OVs occurring in 50% of the cases

The cases shown are systematically analyzed in Tabel 5. It studies what happens when the OV, $\delta = 0, 0.5, \dots, 3$ and the correlation is $\rho = 0.1, 0.5$ and 0.9 . The two cases of Figure 7a

9. If the two parameters, δ and ρ , are smaller, it may be difficult to see that the funnel is bimodal, and hence it is less clear that we are looking at a case where a fraction of the estimates has a missing variable.

and b are the lines with * and ** in the table. The reader will see that the two graphs are typical of the findings in the two cases.

The pattern is rather as expected from the graphs and equation 7: when ρ and δ are small the bias is small too and it causes small problems as well. However, b_M soon starts to move away from 1, and the FAT-test detect asymmetries ($\gamma \neq 0$) and then the fraction of bad estimates of b_M rises too.

It is interesting to note that in about half the cases \underline{b} is closer to 1 than is b_M so it is hard to predict if the meta average is better. Also, a rather strange pattern comes about in the rejection rate for funnel symmetry. Figure 7a shows a case where funnel asymmetry may and may not be rejected. While $b_M = 1$ is almost always rejected.

Table 5. Experiments with OV biases

Section 1 Parameters		Section 2 $N = 40$					Section 3 $N = 100$				
δ	ρ	Averages		Best	Fraction rejected		Averages		Best	Fraction rejected	
		\underline{b}	b_M		$\gamma \neq 0$	$b_M \neq 1$	\underline{b}	b_M		$\gamma \neq 0$	$b_M \neq 1$
0	0.1	1.000	1.000	Same	0.018	0.065	1.000	1.000	Same	0.009	0.057
0.5	0.1	1.025	1.020	b_M	0.020	0.162	1.025	1.020	b_M	0.018	0.519
1	0.1	1.050	1.014	b_M	0.099	0.090	1.050	1.013	b_M	0.517	0.164
1.5	0.1	1.075	0.994	b_M	0.444	0.052	1.075	0.995	b_M	0.996	0.036
2	0.1	1.100	0.977	b_M	0.767	0.116	1.100	0.979	b_M	1.000	0.274
2.5	0.1	1.126	0.965	b_M	0.923	0.225	1.126	0.968	b_M	1.000	0.669
3	0.1	1.151	0.958	b_M	0.966	0.324	1.151	0.961	b_M	1.000	0.897
0	0.5	1.000	1.000	Same	0.022	0.077	1.000	1.000	Same	0.016	0.068
0.5	0.5	1.145	1.249	\underline{b}	0.268	0.992	1.145	1.248	\underline{b}	0.811	1.000
1*	0.5	1.288	1.349	\underline{b}	0.024	0.908	1.289	1.346	\underline{b}	0.031	1.000
1.5	0.5	1.433	1.266	b_M	0.072	0.429	1.434	1.261	b_M	0.201	0.863
2	0.5	1.577	1.099	b_M	0.543	0.053	1.577	1.099	b_M	0.978	0.099
2.5	0.5	1.723	0.947	b_M	0.939	0.017	1.721	0.952	b_M	1.000	0.010
3	0.5	1.868	0.829	b_M	0.998	0.082	1.866	0.842	b_M	1.000	0.239
0	0.9	1.000	1.000	Same	0.431	0.098	1.000	1.000	Same	0.037	0.092
0.5	0.9	1.516	2.166	\underline{b}	1.000	1.000	1.516	2.158	\underline{b}	1.000	1.000
1**	0.9	2.034	3.352	\underline{b}	1.000	1.000	2.032	3.334	\underline{b}	1.000	1.000
1.5	0.9	2.547	4.490	\underline{b}	1.000	1.000	2.547	4.461	\underline{b}	1.000	1.000
2	0.9	3.066	5.445	\underline{b}	0.991	1.000	3.065	5.404	\underline{b}	1.000	1.000
2.5	0.9	3.581	6.059	\underline{b}	0.798	1.000	3.582	6.022	\underline{b}	0.999	1.000
3	0.9	4.096	6.256	\underline{b}	0.346	0.997	4.094	6.212	\underline{b}	0.828	1.000

Note: The selected significance level is 0.05. The results are based on 10,000 replications of each row. The line corresponds to Figure 7a has *. The line corresponds to Figure 7b has **.

5.5 Comparing undetected and detected cases with one OV: Using the MRA(k)

Given the pattern in Table 5, we examine a set of the cases to see if it helps if the OV is detected and the appropriate control is inserted in the MRA, so that we go from the MRA to the MRA(k) estimate. This is done in Table 6. The results are spectacular:

For the MRA with an undetected OV, the results are as before, but for detected OV, where the MRA(k) is used, the results are amazingly good: the average \underline{b} and b_M 's found are very close to 1 in all cases, and the number of $b_M \neq 1$ remains constantly around 6%, as in simulations of ideal funnels. If the OV is detected, it is easy to control for it, and the meta average works as well as it did before. The MRA does not pick up and correct one OV, but the MRA(k) does.

In both cases, the FAT is working rather well though it depends on the number of points in the funnel (N). For large N 's the number of detected asymmetries at the 5% quickly falls below 5%. Thus, the FAT works too well in funnels with many observations.

Table 6. Experiments with an undetected and a detected OV

Section 1 Dimensions				Section 2 Undetected OV MRA				Section 3 Detected OV MRA(k)		
OV characteristics				Averages		Frequency of Rejection		Average same	Frequency of rejection	
N	M	z	ρ	\underline{b}	b_M	$\gamma \neq 0$	$b_M \neq 1$	$\underline{b} = b_M$	$\gamma \neq 0$	$b_M \neq 1$
40	100	0.25	0.5	1.072	1.135	0.242	0.981	1.000	0.017	0.086
40	100	1	0.5	1.288	1.348	0.023	0.905	1.000	0.019	0.078
40	100	2	0.5	1.578	1.103	0.538	0.054	1.000	0.021	0.085
100	100	0.25	0.5	1.072	1.137	0.677	1.000	1.000	0.015	0.069
100	100	1	0.5	1.289	1.351	0.042	0.999	1.000	0.013	0.070
100	100	2	0.5	1.577	1.084	0.975	0.076	1.000	0.015	0.071
100	250	0.25	0.5	1.072	1.134	0.878	1.000	1.000	0.009	0.071
100	250	1	0.5	1.288	1.346	0.032	1.000	1.000	0.012	0.069
100	250	2	0.5	1.578	1.099	0.977	0.100	1.000	0.011	0.067
250	250	0.25	0.5	1.072	1.134	1.000	1.000	1.000	0.008	0.065
250	250	1	0.5	1.289	1.347	0.098	1.000	1.000	0.008	0.064
250	250	2	0.5	1.577	1.094	1.000	0.255	1.000	0.011	0.065
400	1000	0.25	0.5	1.072	1.133	1.000	1.000	1.000	0.007	0.065
400	1000	1	0.5	1.289	1.344	0.159	1.000	1.000	0.005	0.061
400	1000	2	0.5	1.577	1.101	1.000	0.532	1.000	0.007	0.063
1000	1000	0.25	0.5	1.072	1.133	1.000	1.000	1.000	0.006	0.066
1000	1000	1	0.5	1.289	1.345	0.605	1.000	1.000	0.007	0.061
1000	1000	2	0.5	1.577	1.100	1.000	0.952	1.000	0.007	0.063

5.6 More complex cases: many OVs

We now depart from the case of only one OV to study a more general setting. The DGP is composed of K controls where each is included or omitted with a certain probability which is set at 0.5 throughout. Each individual control is generated with a given correlation ρ_i and effect size, δ_i . The analytical form of the expected level and variance is much harder to work out in this setting as it depends on the cross correlation between included and omitted controls as well as with x .

Formula (1) $y_j = \alpha + \beta x_j + \gamma_1 z_{1j} + \dots + \gamma_k z_{kj} + \varepsilon_j$, can be made to generate as many peaks as necessary to make the funnel smooth-looking, by using higher K 's and k 's. The number of different biases – and hereby peaks – produced by (1) is:

$$\lambda(K, k) = \binom{K}{k} \quad (8)$$

which easily produces large numbers, if the two variables are large. For large λ 's and different z 's, the dents disappear to make a nice smooth funnel. We may get symmetry or asymmetry:

Symmetry demands that the set of controls is symmetrical in coefficients and correlation and they are selected with the same probability. The resulting funnel appears like an ideal funnel. When some controls are included, and others omitted, it becomes harder to predict the precision of the estimated β .

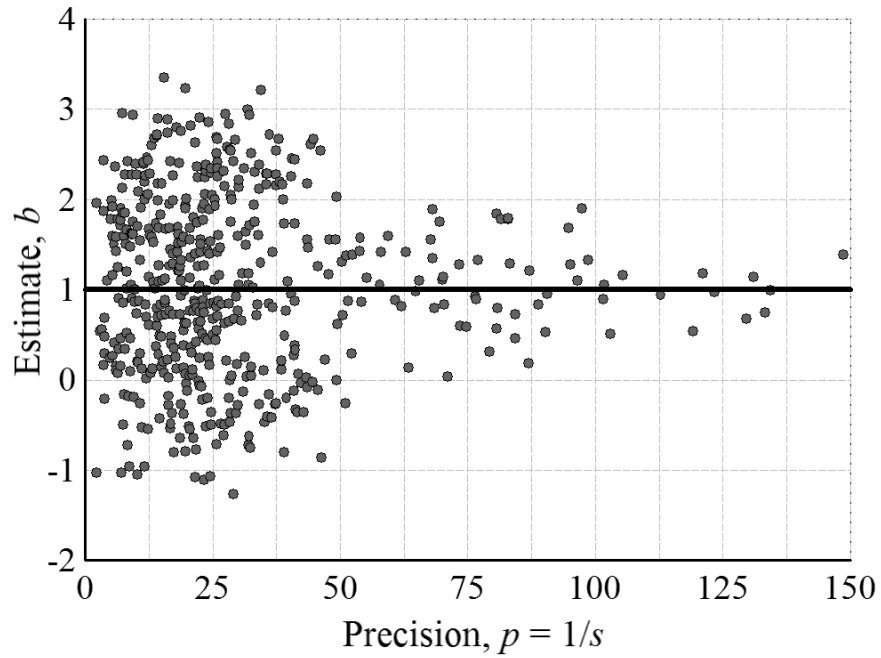
Asymmetry occurs in all other cases. By playing with the control set, the correlation coefficients and the frequency of selection, it is possible to generate a wide range of funnels. There does not seem to be any regularity in the shape of funnels subject to omitted variable bias to could indicate that we are dealing with a funnel generated with such a bias.

If there are less negative controls than positive ones, the most precisely estimated points will be biased above the true value of the parameters of interest, while the second most precise estimates will be those estimated with the full model thus closer to the true value. The funnel will thus appear to be censored above.

5.7 Two examples of funnel with 6 OVs: A symmetric and an asymmetric

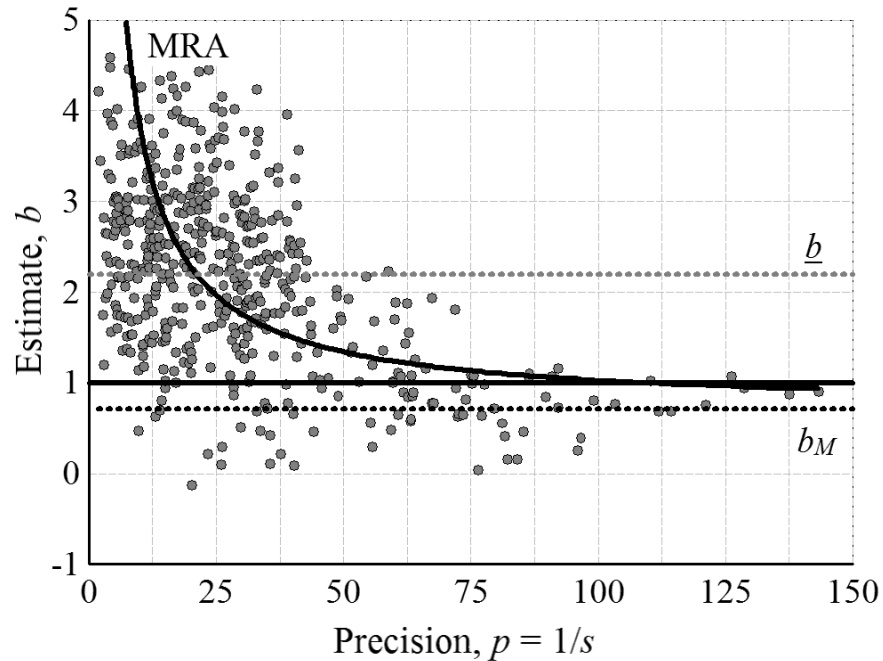
To show what can be done, consider Figures 8a and b. Controls are randomly included with probability 0.5. They have $\rho = \sigma^2 = 0.5$ with x . In the symmetric case $\delta = 2, 1, 0.5, -0.5, -1$ and -2 , while in the asymmetric case they are $\delta = 2, 1, 0.5, -0.5, -1$ and -2 .

Figure 8a. A smooth symmetric funnel made as described in text ($N = 500$).



Note: The MRA is indistinguishable from the $\beta = 1$ -line.

Figure 8b. An smooth asymmetric funnel made as described in text ($N = 500$).



The reader may think of the wide funnel on Figure 8a as a realistic case of model uncertainty. It certainly produces the excess variation result, with lots of estimated b s that differ significantly. Finally, Figure 8b shows an asymmetric funnel. It is treated by the MRA as a

case of downward censoring. We want our story to have a happy end, and Figure 8b is the last funnel in the paper. Consequently, the peak is made at 1, so the MRA converges to almost the right value, and in addition the b_M is a better average than \underline{b} .

A set of simulations were also made with the two cases of Figure 8. Also in these cases the MRA(k) worked very well. These results are available on demand.

6 Summary

A funnel shows the distribution of a set of estimates of the same parameter. All empirical funnels we have seen are amazingly wide, and most are asymmetric as well. If the funnel is due to noise in the data only, it is symmetric and lean. Consequently, funnel wideness is due to model variation and funnel asymmetries are due to biases, which are common in the set of estimates. The *meta average* is the average corrected for the asymmetry. The paper studies a range of asymmetries that may occur in funnels. We distinguish between *censoring* biases caused by the research and publication process, and *natural* biases that occur when researchers overlook a modeling problem.

The standard tool in meta analysis is the FAT-PET MRA, which works on the data of the funnel. It contains two terms: The FAT-term is the funnel asymmetry test. We showed that the FAT-term is a general tool of considerable power irrespective of the cause of the asymmetry. The PET-term is one estimate of the meta average. The PET estimate is tooled to detect and adjust funnel asymmetries for censoring biases. It does so rather well. However, the PET-term can be trusted to correct censoring biases only. With natural biases it may even increase the distance to the true value compared with the plain average. Three types of problems have been analyzed.

First, the case of dependent – but expanding data – was considered. This is a typical case in macroeconomics. Here the results did converge to the true value, but often quite slowly, so a surprisingly large number of cases rejected the true value. Also the FAT often rejected symmetry. The problems increased with a structural break in the data.

Second, estimation faults were studied. One fault occurs when it is disregarded that the residuals in the regression are non-normal. It appears a fairly harmless problem. Another fault occurs if the true model is non-linear. This typically gives a wrong estimate of the parameter, but the funnel stays symmetrical till the non-linearity is rather strong.

Third, omitted variables were analyzed. If everybody fails to include the said variable, the funnel becomes symmetric, and the meta average and the plain average both come to include the OV-bias. However, if some of the researchers in the field fail to include a certain control variable while other researchers do not, it typically causes a funnel asymmetry. Meta analysis recommends that the problem is cured by detecting the OV and inserting a binary control in the MRA that then becomes the MRA(k). This method is shown to work well.

Hence, it is important to know if an observed asymmetry is natural or due to censoring. We consequently recommend the meta analyst to study the funnel, and try to determine the reason for the asymmetry is, and then take the appropriate step to adjust for that asymmetry. If the adjustment made is the wrong one, the meta average may be worse than the plain average.

Acknowledgements:

One of the authors (MP) has been tutored in meta-analysis by Chris Doucouliagos and wants to thank him. The paper was presented at the 2009 MAER workshop, University of Oregon, Corvallis, Oregon. We are grateful to the discussants – especially to Tom Stanley for very constructive advice. Also, we wish to thank Bent Jesper Christensen, who helped us starting up the paper, and Erich Gundlach for a useful discussion.

References:

- Dewald, W.G., Thursby, J.G., Anderson, R.G., 1986. Replication in Empirical Economics: The Journal of Money, Credit and Banking Project. *American Economic Review* 76, 587-603
- Doucouliagos, H., Paldam, M., 2009a. The aid effectiveness literature. The sad results of 40 years of research. *Journal of Economic Surveys* 23, 433-61
- Doucouliagos, H., Paldam, M., 2009b. Development aid inertia. Stylized facts and a meta study. Working paper, available from <http://www.martin.paldam.dk>
- Doucouliagos, H., Paldam, M., 2009c. Development Aid and Growth: An association converging to zero. Working paper, available from <http://www.martin.paldam.dk>
- Koetse, M.J., Florax, R.J.G.M., Groot, H.L.F., de, 2005. Correcting for Primary Study Misspecifications in Meta-Analysis. Tinbergen Institute Discussion Paper TI 2005-029/3
- Lau, J., Ioannidis, J. Terrin, N., Schmidt, C., Olkin, I., 2006. The case of the misleading funnel plot. *British Medical Journal* 333 (7568), 597-600
- Light, R.J., Pillemer, D.B., 1984. *Summing Up: The Science of Reviewing Research*. Harvard University Press, Cambridge, Mass
- MAER. Meta-Analysis of Economic Research Network: <http://www.hendrix.edu/MAER-Network/>
- McCullough, B.D., McGeary, K.A., Harrison, T.D., 2008. Do economics journal archives promote replicable research? *Canadian Journal of Economics* 41, 1406-20
- Roberts, C.D., Stanley, T.D., 2005. *Meta-Regression Analysis: Issues of Publication Bias in Economics*. Series: Surveys of Recent Research in Economics. Blackwell, Oxford, UK
- 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., 2009. Picture This: A Simple Graph that Reveals Much Ado about Research. *Journal of Economic Surveys*, forthcoming
- Tang, J., Liu, J., 2000. Misleading funnel plots for detection of bias in meta-analysis. *Journal of Clinical Epidemiology* 53, 477-84
- Terrin, N., Schmid, C., Lau, J., 2005. In an empirical evaluation of the funnel plot, researchers could not visually identify publication bias. *Journal of Clinical Epidemiology* 58, 894-901

Economics Working Paper

- 2009-05: Martin Paldam: The welfare state and Baumol's law
- 2009-06: Erich Gundlach and Martin Paldam: The agricultural and the democratic transitions - Causality and the Roundup model
- 2009-07: Torben M. Andersen: Fiscal policy and the global financial crisis
- 2009-08: Helena Skyt Nielsen: Causes and Consequences of a Father's Child Leave: Evidence from a Reform of Leave Schemes
- 2009-09: Marianne Simonsen and Lars Skipper: The Family Gap Reconsidered: What Wombmates Reveal
- 2009-10: Olaf Posch and Klaus Wälde: On the non-causal link between volatility and growth
- 2009-11: Torben Sørensen and Rune Vejlin: The Importance of Worker, Firm and Match Fixed Effects in the Formation of Wages
- 2009-12: Nisar Ahmad: State Dependence in Unemployment among Danish Immigrants
- 2009-13: Nisar Ahmad and Rayhaneh Esmailzadeh: Immigrant-Native Differences in Earnings Mobility Processes: Evidence from Canadian and Danish Data
- 2009-14: Nisar Ahmad and Michael Svarer: The Effect of Sanctions and Active Labour Market Programmes on the Exit Rate From Unemployment
- 2009-15: Martin Paldam and Erich Gundlach: The religious transition - A long-run perspective
- 2009-16: Torben M. Andersen and Joydeep Bhattacharya: Unfunded pensions and endogenous labor supply
- 2009-17: Hristos Doucouliagos and Martin Paldam: Development Aid and Growth: An association converging to zero
- 2009-18: Christian Bjørnskov and Martin Paldam: The spirits of capitalism and socialism. A cross-country study of ideology
- 2010-1: Laurent Callot and Martin Paldam: Natural funnel asymmetries. A simulation analysis of the three basic tools of meta analysis