2 August 2011

The problem of natural funnel asymmetries

A simulation analysis of meta-analysis in macroeconomics

Laurent Callot[1] and Martin Paldam[2]

School of Economics and Management, Bartholins Allé 10

Aarhus University, 8000 Aarhus C, Denmark

Abstract:

Effect sizes in macroeconomic are estimated by regressions on data published by statistical agencies. Funnel plots are a representation of the distribution of the resulting regression coefficients. They are normally much wider than predicted by the t-ratio of the coefficients and often asymmetric. The standard method of meta-analysts in economics assumes that the asymmetries are due to publication bias causing censoring, and adjusts the average accordingly. The paper shows that some funnel asymmetries may be 'natural' so that they occur without censoring. We investigate such asymmetries by simulating funnels by pairs of DGPs (data generating processes) and EMs (estimating models), in which the EM has the problem that it disregards a property of the DGP. The problems are data dependency, structural breaks, non-normal residuals, non-linearity, and omitted variables. We show that some of these problems generate funnel asymmetries. When they do, the standard method often fails.

Keywords: Meta-analysis, funnel asymmetry, meta average

JEL: B4, C9

Accepted by *Research Synthesis Methods*

1. E-mail: lcallot@creates.au.dk, at the Center for Research in Econometric Analysis of Time Series.

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

1. Introduction: Meta analysis in macroeconomics

This paper deals with meta analysis in macroeconomics.[3] The introduction has two parts. Section 1.1 explains why meta analysis needed a special development to be useful in macro-economics, and section 1.2 presents the problem analyzed by the paper.

1.1 Macroeconomics: Regression coefficients estimated on limited national data

Economic policies of governments rely on beliefs about parameters in macroeconomic relations. Below we consider one such parameter, β, which is assumed to have one true value.[4] The reader may think of β as the effect of a treatment, e.g., the effect of the interest rate on unemployment; the effect of development aid on development; or the effect of a common currency on the income of a group of countries. Political decisions based on beliefs about β may therefore affect millions of people and involve billions of €. This has two consequences:

(i) The size of β is politically important. This generates priors affecting research.

(ii) A whole body of literature studies β. It reports the B-set of N estimates.[5]

Once N is substantial (such as 200), the literature is likely to contain much more information about β than any new primary study possibly can. Consequently, the B-set of estimates begs for a meta study, trying to extract the best *meta average*, $\hat{b}_M \approx \beta$.

Researchers are not allowed to make controlled experiments with national economies. Empirics in macroeconomics are therefore based on econometric inference from official data compiled by statistical agencies. Economic theory contains hypotheses and logical inferences about the true relation, guiding researchers in picking the relations estimated. Econometrics provides a set of regression techniques developed to fit these models to such data. The B-set thus consists of regression coefficients standardized to a comparable scale.[6] The b's in the B-set all come with t-ratios, which are normally well above 2, suggesting that β is well known.

The primary data available for estimating β are often quite limited, such as 40 annual

3. Macroeconomics has been the subject of about 200 meta studies. Also, many microeconomic meta studies have appeared. The second author has been co-author of 10 macroeconomic meta-studies and participated in the meetings at the MAER network (references) where another 100 meta studies in economics have been discussed.

4 The author typically assumes that all differences between countries and time periods are controlled for – in panel data this is often done by sets of fixed effects.

5. For ease of exposition we assume that each study reports one estimate. The β-literature is thus the literature which tries to establish the value of β. It is arguable that any β is unstable over time and across countries. The literature assumes that the instability can be modeled by appropriate controls.

6. The most desirable standardization is to convert the coefficients into elasticities (relative effects). In practice this is rarely possible. The second-best alternative is to use partial correlations. Normally, this is possible.

observations from 12 countries – the data expands every year by one new observation per country, so they are *sequentially dependent*. All papers in the β-literature have to use these data. Instead of controlled experiments the researcher includes control variables in the estimating equations and uses various estimation techniques.[7] Often, the number of possible controls is large. The researcher has to choose a subset of these controls based on economic theory and statistical testing. However, many controls are possible, authors can only try so many, and various subsets of the data typically show that different ones matter. Therefore the β-literature contains a great deal of model variation, which is normally done around one central model.

In medicine the meta-analysis typically starts with a *B*-set generated by independent, controlled experiments. In macroeconomics the analyst starts with a *B*-set of regression coefficients generated by many related model variants on dependent data.

As usual the meta analysis starts by looking at the funnel of the *B*-set.[8] Meta studies in macroeconomics inevitably find funnels to be (much) *wider than expected* from the t-ratios of the estimates. Thus β is much less well known than suggested by the individual studies. In addition most funnels are *asymmetric*. Due to the importance of prior beliefs, it is the standard assumption that the asymmetry is due to publication bias, i.e., by priors causing a censoring of results. The meta-analysis tries to find the best *meta average* by including variables explaining the excess width, and correcting for the censoring.

In economics the standard method is the ***FAT-PET MRA*** technique, developed by T. D. Stanley to handle this case, see section 2.3. It is the ***M***eta ***R***egression ***A***nalysis, with two terms. The FAT is the (usual) ***F***unnel ***A***symmetry ***T***est. The PET is the ***P***recision ***E***stimate of the censoring-corrected estimate of the meta average, $\hat{b}_M$,[9] with a ***T***est to show if it is statistically different from zero.

This process of meta research is simulated by generating funnels using two tools: The ***DGP*** is a known ***d***ata ***g***enerating ***p***rocess simulating the unknown 'true' relation. The ***EM*** is the ***e***stimation ***m***odel used on these data, including the estimator. ***Ideal*** funnels emerge in the limiting case when the DGP and the EM are the same. They are caused by variation in the primary data only. It is well-known from statistical theory how regression coefficients are

7. Thus $\beta = \partial y/\partial x$ is the effect of the variable of interest *x*. A control is a variable included in the estimation equation to take care of something that influences *y*, such as country or time period differences.
8. The funnel is the distribution of the standardized regression coefficients over their precision, which is the inverse standard error of the regression coefficients.
9. We take the term *meta average* to be the best average of β from the *B*-set based on some assumptions about its distribution. At present we do not discuss other meta averages than the one estimated by the FAT-PET MRA.

distributed in the ideal case: They are *symmetrical* and as *narrow* as predicted from the t-ratios of the coefficients. If funnels are ideal, asymmetries can only be due to censoring.

1.2 Funnels without censoring: Do non-ideal funnels have natural asymmetries?

The difference between the narrow and symmetrical ideal funnels and the wide and asymmetrical empirical funnels is the observation that started this research. The purpose of the paper is to study if asymmetries can occur 'naturally' without censoring, when funnels are non-ideal. It appears that they actually can, and therefore, a question arises: Does the FAT-PET MRA still catch the true average?

To be sure that no publication bias occurs funnels are simulated by *DGP-EM*-pairs with a *problem,* i.e., a known difference. This simulates the situation where the true model has a property that is undetected by most researchers. The simulated funnels are used to study how the problem affects the width of the funnel, how it biases the estimates of β, and if it causes an asymmetry. Also, FAT-PET MRAs are estimated on the data of the funnels to see if the bias is due to asymmetries detected by the FAT, and if the PET-estimate finds the true value of β.

Section 2 defines the basic terms and concepts, describes the way the simulations are set up, the origin of the FAT-PET MRA etc. From there we push into (almost) virgin territory where the analysis is done by means of simulation as described. The field is new, so simple, tractable assumptions are chosen for the DGP.

Section 3 deals with the problem of *data dependency*. It is the common case in macroeconomics where researchers have to use the same data set. It expands over time, but it might also have important breaks. In these cases natural funnel asymmetries are common. Section 4 deals with *estimation faults:* The DGP has the property of a non-normal disturbance term or non-linearity. The EM misses the said property. The funnel is due to data uncertainty only, but it is affected by the estimation fault. They always make the funnel wider, but the funnel remains symmetric, when the fault is symmetric. This is not the case when the fault is an undetected no-linearity.

Section 5 deals with the problem of *misspecification* in the form of omitted variables in the EM, where the funnel is due to data and model uncertainty. This produces a wide range of asymmetries, which in unlucky cases may look like censoring. Sections 3 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.

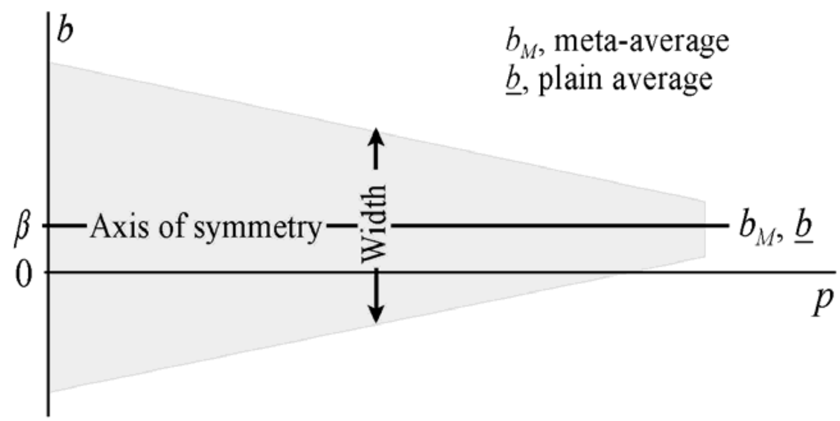
2. Concepts and methods

Section 2.1 gives definitions and concepts, 2.2 summarizes the argument why censoring asymmetries are likely, 2.3 considers the FAT-PET MRA and the censoring-corrected estimate of the meta average, 2.4 discusses the extended version and the concept of omitted variables, 2.5 gives a brief reference to the literature, and 2.6 presents the simulation framework used from section 3 onwards.

2.1 Basic definitions and concepts

One meta-analysis considers *N* estimates, b_i, where $i = 1, \ldots, N$, of the parameter β. Estimate b_i, has the standard error s_i, the precision $p_i = 1/s_i$, and the t-ratio $t_i = b_i/s_i$. The funnel diagram is the (b_i, p_i)-scatter. The arithmetic or 'plain' average is $\underline{\hat{b}} = \sum_{i=1,N} b_i / N$.[10]

Figure 1. Shape of the ideal funnel[11]



Ideal funnels, where the DGP and the EM are the same, have the form sketched in Figure 1 as shown in section 4.2. They are narrow and symmetric around a horizontal axis of symmetry. The discussion turns around three averages: The true value, β, the plain average, $\underline{\hat{b}}$, and the PET estimate of the meta average, $\hat{b}_M$, which is found as the intersection of the axis of symmetry and the *b*-axis, as demonstrated in section 2.3.

10. Many weighting schemes can be used to calculate the average. One uses the precision of the estimate, another uses the quality of the publication, etc. In the interest of simplicity, the unweighted or plain average is used for comparisons.
11. In the literature an almost equal fraction of funnels are presented with the axis reversed. We think of funnels as graphs showing the width of the distribution as a function of the precision of the estimate.

If the funnel is symmetric $\hat{b}_M$ is close to $\underline{\hat{b}}$ and both are good estimates of the true value *β*. If the funnel is asymmetric $\underline{\hat{b}}$ is a bad estimate of *β*, but the meta average should still be a good estimate of *β.* We study if this is the case for the PET estimate.

2.2 Priors beliefs lead to censoring: Systematic censoring gives biases

From 2.6 onwards we only discuss natural asymmetries, so it is only fair to briefly summarize why economists are concerned with censoring asymmetries. Politically important parameters are surrounded by emotions (P1), interests (P2), and much theorizing (P3), which give prior beliefs – that most researchers disregard.[12] Four types of such priors are commonly recognized in meta analysis:

P1 *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.

P2 *Economic priors*: Researchers may work in areas where they have interests, and censor accordingly. Example: Censoring of results showing that minimum wages generate unemployment by researchers of the Labor Movement.

P3 *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.[13]

P4 *Polishing prior*: To reach marketable results and satisfy the researchers own craving for clarity, unclear results are censored. This prior works in all directions to give *excess significance* of results, i.e., excess width of funnels. It will not be discussed at present.

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 links up to the excess width observation. It is common in economics that results reported in well-respected journals, for the same *β*, are statistically different at levels of significance such as

12. It is rare that economists state their priors and take them into consideration explicitly. A school of Baysian econometrics is well established, but the Baysian methods are cumbersome to use, and it is not always clear that the true priors are considered. So the school is a minority, even when the Baysian challenge has been forcefully presented to the profession already by Leamer (1983).

13. When the price of a good goes up people normally buy less, so the elasticity (the relative price effect) is negative. The Giffen case is the most well-known exception, in which the price of the staple food rises in a poor country, causing income to fall so much that people cannot afford any other food.

0.01 %, or, e.g., by a factor of 5, so the width of funnels is indeed excessive.

When the research process leads to a paper reporting a result, it is consequently generated by the *stopping rule* ending the search. 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., he 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 economics papers.

The rule applied differs from paper to paper, and it is often difficult, even for the author, to know the decisive stopping rule for a paper. 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.

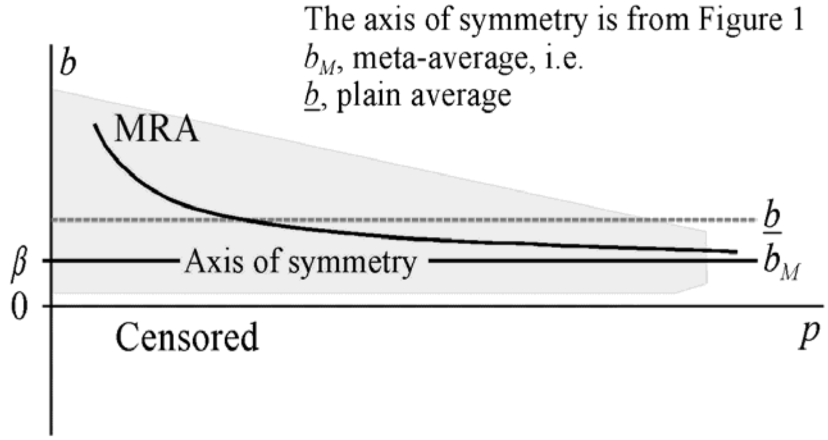
A censoring bias is caused by researchers who have priors for results in a certain range. Think of the example from (P3) where $\beta = -1$ is the true price elasticity. Here +0.5 and −2.5 are the same distance from truth, and presumably equally easy to reach. Maybe one identifying assumption gives +0.5, while another gives −2.5. From his theoretical prior 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, but it does bias the result substantially.

Randomly distributed priors (as in P4) give variation in the results, i.e., wide funnels, but not (necessarily) asymmetries. 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, and the best practice tool in economics is developed by Stanley (2008).

2.3 The FAT-PET MRA and the censoring corrected estimate of the meta average, $\hat{b}_M$

Figure 2 is the same funnel as Figure 1, but it is censored for negative values and it looks obviously asymmetric. Censoring narrows the funnel, so it does not help explain the excessive width observation, but if a likely prior against negative values exists, we are still confident that we can explain the asymmetry, see, e.g., Doucouliagos and Paldam (2008).

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



Censoring causes the plain average, $\hat{\underline{b}}$, to be a bad estimate of β. The meta average, $\hat{b}_M$, is still the intersection of the 'true' axis of symmetry from Figure 1 with the b-axis. But now $\hat{b}_M$ has to be estimated by an estimator that adjusts for the censored estimates. The resulting FAT-PET MRA is written in two equivalent ways:

$$b_i = \hat{b}_M + \hat{\gamma}(1/p_i) + u_i \quad \text{(1a)} \qquad \Leftrightarrow \qquad t_i = \hat{b}_M p_i + \hat{\gamma} + v_i \quad \text{(1b)}$$

Where u_i and v_i are residuals, and (1b) is reached from (1a) by dividing with s_i. The censoring adjusted meta average is $\hat{b}_M$, which comes with a t-test of significance. The FAT is the t-test of significance for $\hat{\gamma}$ indicating asymmetry – the sign on $\hat{\gamma}$ points to the direction of the asymmetry. We use OLS to estimate (1), but many meta-analysts use weighted regressions.

Formulation (1a) gives an easier intuition, while (1b) is preferable to estimate as it has less heteroskedasticity. The estimate produces a hyperbola converging to the axis of symmetry for p rising; see Figures 3, 6b, 7 and 8b below. If the upper or lower part (as Figure 2) of the funnel is censored, the MRA still reaches the same result. If the highest precision part of the funnel is affected by the bias, the convergence becomes less precise, but $\hat{b}_M$ is still closer to β than is $\hat{\underline{b}}$. Thus, the PET is a fine estimate of the meta average in the typical censoring case. This is shown by Stanley (2008) and confirmed below. That is, the difference between β and the estimated $\hat{b}_M$ is insignificant at the 5 % level in almost 95 % of the cases.

2.4 The extended FAT-PET MRA: Model uncertainty

Macroeconomics has a great deal of model uncertainty, which typically consist in variation of the control variables, which are included in the central model. Econometrics treat this case as a true model with K variables, which is estimated with k variables, where $k \leq K$, where the k variables are always true ones. That is, the researcher can determine if a variable is significant, but may not try all the right ones, and therefore, some are omitted.

We distinguish between model certainty where $k = K$, and model uncertainty, where $k < K$. Sections 3 and 4 deals with model certainty. In section 4 the data has a sequential dependency and the residuals or the functional form is wrong in the EM. Section 5 deals with model uncertainty, where the functional form and the residuals are rightly specified, but the EM suffers from $K - k$ omitted variables, giving biases. We show that if some of the researchers detect the right variable 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 omitted variables, $z_1, \ldots, z_n$ and code a set of binary control (or conditioning) variables $k_1, \ldots, k_n$, where $k_1 = 1$ is the estimate controlled for z_1, and zero otherwise, etc. with the remaining ks. In addition to variables that appear in the primary studies many meta-studies add variables to account for author characteristics and the quality of the publication.[14] The coded variables are then added to the MRA with the ks to obtain the extended MRA(k):

$$b_i = \hat{b}_M + \hat{\gamma}/p_i + \hat{\lambda}_1 k_1 + \ldots + \hat{\lambda}_n k_n + u_i \quad \text{(2a)} \quad \text{or} \quad t_i = \hat{b}_M p_i + \hat{\gamma} + \hat{\lambda}_1 k_1/s_i + \ldots + \hat{\lambda}_n k_n/s_i + v_i \quad \text{(2b)}$$

The relation between (2a) and (2b) is the same as the relation between (1a) and (1b). Section 5 demonstrates that the $\hat{\lambda}$'s are good estimates of the bias that occurs when the z's are omitted, and it also provides good estimates of $\hat{b}_M$ and $\hat{\gamma}$. While DGP has K variables, the different EMs have only $k_i \leq K$ variables.[15] If $K \subset \bigcup_i k_i$, the DGP is included in the union of all the EMs estimated, and 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 omitted variable bias. The complement of k_i is termed k_i^c, with the union $\bigcup_i k_i^c$. If the MRA is extended with dummies for each element of $\bigcup_i k_i^c$, i.e. all the controls that are included in any study, the problem is solved.

The MRA(k) allows us to see if each of the *z-variables* matters. The paper introducing a new explanatory variable, z, claims that it is important, and z is surely significant in that

14. Consider a β-literature with e.g. N = 500 estimates of β. It typically contains a great many k-variables, such as 50, which have to be coded. However, by using the testing-down procedure (see e.g., Kennedy, 2003) the results are normally considerably reduced, and therefore, easier to present.

15. If $k > K$ there may be problems in small samples, but the limit coefficients to excess variables will be zero. The meta analysis should show this. Below we only discuss the case of $k \leq K$.

paper. Once it has been included in a set of papers, it is possible to see if z gives a robust improvement of the EM for β or not, by running the MRA(k_z). Thus, it is useful to run expanded MRAs.[16]

The problem arises if the omitted variable remains undetected in the meta-analysis. Hence, the MRA is uncontrolled for k_z. Section 4 simulates cases with undetected omitted variables to analyze what the MRA does. It is shown that it often causes the MRA to pick a $\hat{b}_M$ that is further from β than is $\underline{\hat{b}}$.

2.5 *A brief look at the literature*

Meta-analysis originates from medicine. Here the problems are somewhat different, as many meta studies deal with independent samples of experiments designed to isolate the β-effect; but even then the tools used are closely related.

It appears that Light and Pillemer (1984) introduced the funnel diagram. Stanley and Doucouliagos (2010) survey the use of funnels in economics, with illustrative examples, showing that many funnels are asymmetric, see also Roberts and Stanley (2005) for a collection of studies stressing publication biases in economics.

The FAT originates from Egger et al. (1997). The PET is the latest version of a family of estimators of the meta average. It is developed by Stanley (2008), who also follows the track that led to his estimate.

We have found few studies of natural funnel asymmetries, but some recent papers in medicine 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). They suggest that 'strange' looking funnels are due to model heterogeneity, as discussed in section 5. Also, some new working papers such as Doucouliagos and Paldam (2009b) contain clear examples.

The two closest predecessors to the present paper are: Koetse et al. (2005), which discuss how the meta average is affected by omitted variable biases, and Stanley (2008), which 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.

2.6 *The simulation set-up: The ideal DGP = EM and the problem DGP ≠ EM*

The parameter of interest $\beta = 1$ in the DGP, except for the experiments in sections 3.2 and 3.4,

16. As an aside, is should be mentioned that (2) has also been used to study the effects of estimators. Does it matter to use a new estimator instead of OLS estimator? Such studies have shown that it rarely does, see, e.g., Doucouliagos and Paldam (2010).

where β changes. The estimator in the EM is OLS (Ordinary Least Square), which assumes that the DGP fulfils the following *ideal* conditions: It is linear, and the residuals are $\varepsilon_t = N(0, \sigma^2)$, i.e., they are normal, uncorrelated and have a constant variance.

A simulated funnel has N estimates of β, where each is estimated by the EM on M observations generated by the DGP. The experiments are done in series with different values of N and M. Each funnel gives one set of estimates – $\hat{\gamma}$, $\hat{b}_M$ and $\hat{\underline{b}}$ – for the FAT, the PET estimate of the meta average and the plain average respectively.

An experiment uses the same DGP-EM pair to generate L = 10,000 funnels. The L funnels are used for estimating rejection rates, R_{FAT}, $R_{\underline{b}}$ and R_{bM}. They show how often the FAT rejects symmetry, and how often the two averages, $\hat{b}_M$ and $\hat{\underline{b}}$, reject the true value β.

In this paper the level of significance is always 5 %. In experiments with ideal funnels, where the DGP and EM are the same, we expect R_{FAT}, $R_{\underline{b}}$ and $R_{bM} \approx 0.05$. This is close to the result found. In all other experiments, the DGP and EM differ. This should cause the three frequencies to increase, and they do in most cases.

Sections 3 to 5 deal with 3 sets of problems that might cause natural asymmetries: The sequential dependency of macroeconomic data, estimation faults, and omitted variables. The problems are analyzed following the same format. First, the problem is set up and illustrated by a few typical graphs of simulated funnels. Then the full set of experiments follows, with many replications used for calculating the frequencies of rejection: R_{FAT}, $R_{\underline{b}}$ and R_{bM}.

The funnel-figures in the next sections all use the same format (see Figure 3 overleaf). The scatter of the funnel is indicated with circles, and the FAT-PET MRA estimated by equation (1b) from section 2.3 is drawn as a black curve. The graphs include three horizontal lines: A solid black horizontal line at β = 1, a gray dotted line at the plain average, $\hat{\underline{b}}$, and a black dotted line at the meta average, $\hat{b}_M$, to which the MRA converges.

3. The sequential dependency of macroeconomic data

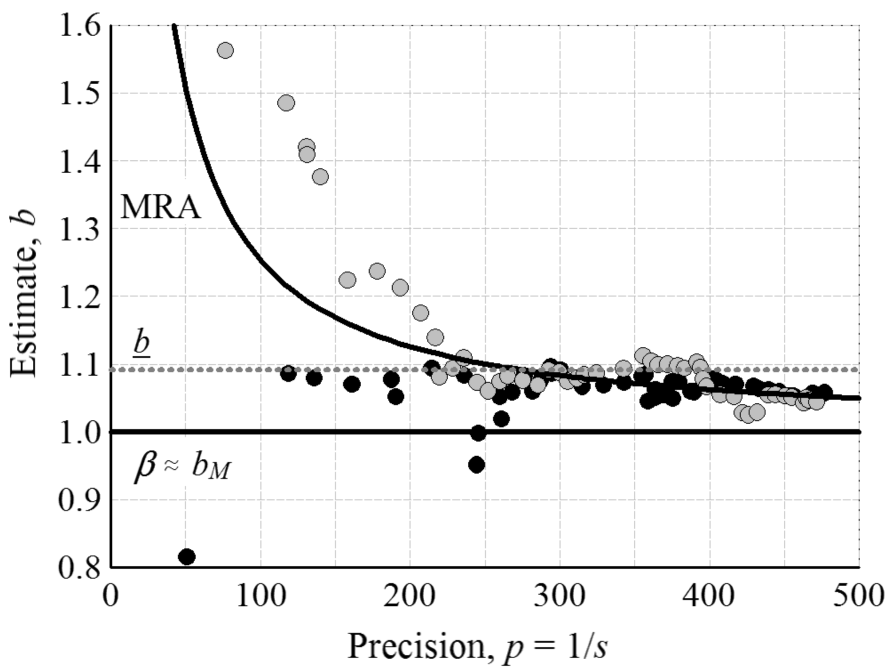
In macroeconomics researchers are forced to use the same data. As time passes, data expands by new observations, so they have a sequential dependency. We study if this dependency creates funnel asymmetries, and if the likelihood of asymmetries increases when the data have structural breaks.

3.1 Sequentially expanding data generate path dependency in the funnel

Consider the funnel of N estimates, published in N papers by different researchers, on a sequence of expanding data. As the papers are independently published they typically have some other difference apart from the added data. To isolate the effect of the data dependency, the *DGP* is equal to the *EM*, so that the same model is used. Hence all that changes is that the data expands by one more observation, and then one more estimate is produced.

The data are the sequence: $(x_1, \ldots, x_t)$, $(x_1, \ldots, x_{t+1})$,…, $(x_1, \ldots, x_T)$, where each x_t is a set of observations for the variables entering the model, and the number of observations, M, in the sequence grows from time t to T.

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



Note: The format of the figure is explained in the last paragraph of section 2

It is well-known how the estimates on such sequential data look. The 95 % confidence interval around β is $I_{95}(M) = \beta \pm (s_M / b_M) \cdot t_{5,M} \approx 1 \pm (\sigma / \sqrt{M}) \cdot t_{5,M} \to 1$, for $M \to \infty$, where $t_{5,M}$ are the 5 % (two tailed) percentage point for the t-distribution. Figure 3 shows the path of two sequential estimates depicted with gray and black circles.

The DGP is $y_i = \alpha + \beta x_i + \varepsilon_i$, where $\varepsilon = N(0, \sigma^2)$. The EM is $y_i = \hat{a} + \hat{b} x_i + e_i$
Here $\beta = 1$ and $\sigma = 0.5$. The sequence has $M = 50$ estimates for $t = 51$ to $T = 100$.

We may think of them as the results based on expanding annual data from two countries during 50 years of research, where one regression is done every year in each country using the same model. In Figure 3 the line for the meta average, $\hat{b}_M$, is hidden by the solid black line at $\beta = 1$. The figure illustrates the problem with data dependency: The sequence of b_i's does converge to β, but the estimation errors $|b_i - \beta|$ are not *i.i.d.* In this case the OLS estimate of β using the estimated b_i is biased.

In the case with gray markers, the convergence of the curves, when N increases, 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, one should keep in mind that quirks in the data may create amazing path dependency. Table 2 below 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 unfortunately the MRA does converge very slowly to $\hat{b}_M \approx 1$, and in many cases it is actually found to be statistically different from 1 even in large samples.

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 in section 2.2. The risk of Type I errors makes the usual scientific requirement of *independent replication*[17] particularly important in macroeconomics. The path dependency of quirks means that adding 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

17. *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 the Journal of Money, Credit and Banking project of Dewald et al. (1986). The JMCB journal request documented data for the articles accepted. Dewald et al studied 54 data sets submitted over 2 years of the said journal, and only found 8 cases where data was perfectly documented. Further they tried to replicate the results of 9 studies where only two replicated perfectly and major differences appeared in five cases. The ensuing discussion is surveyed in, e.g., McCullough et al. (2008).

for some time.[18]

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 size of the sample has increased sufficiently beyond the break, the fall in precision turns up. 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 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.

3.2 Systematic simulations: Without and with structural breaks

Table 2 contains a set of simulation experiments, each repeated 10,000 times. The table has four panels marked with a change of background color. Panel 1 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 independent data sets included. The reader may think of each series as a country.

Panel 2 looks at the case with no structural breaks, only data dependence. The meta average is always very close to 1 on average, so there is no *systematic* bias in this case.

Table 2. Overlapping sample cases without and with structural breaks

Panel 1 Dimensions			Panel 2 No breaks			Panel 3 Break after 75 %			Panel 4 Break after 25 %		
N	*M*	*Nr*	Average	Rejection rates		Average	Rejection rates		Average	Rejection rates	
			$\hat{b}_M$	R_{FAT}	R_{b_M}	$\hat{b}_M$	R_{FAT}	R_{b_M}	$\hat{b}_M$	R_{FAT}	R_{b_M}
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 million regressions. *Nr* is 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 2.6 that *N* is the number of estimated 'points' in the funnel, while *M* is the number of observations used for each estimate.

18. 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 dominates the literature for a period of 4-6 years.

However, there are random biases in 40 % of the cases with an equal probability to both sides. The funnels generated using the method described above are asymmetric much more often than in the independent sample case. In nearly all these cases a biased meta average, $\hat{b}_M$, occurs – so the two rejection rates are very similar. Interestingly, the likelihood of asymmetry increases with sample size (*M*), but not with the number of independent series (*Nr*) used to generate the funnel.

Thus the meta-analysis will show a wrong result due to a quirk in the data which has not disappeared over the sequence.

Panel 3 looks at the case where the structural break takes place late, i.e., after 75 % of the sequence. In panel 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, $\hat{b}_M \neq 1$ per definition, and it makes no sense to report the rejection rates for $\hat{b}_M \neq 1$.

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 about 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, it is important to note that it gives 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 only the plain average; but in most cases it is certainly not enough.

4. Estimation faults

This section deals with model certainty, where there are no omitted variables. 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 fraction of rejections for 10,000 replications.

4.1 The model faults considered

With $K = k = 0$ the two DGPs used are:

$$y_j = \alpha + \beta x_j + \varepsilon_j \quad (3) \qquad \text{and} \qquad y_j = \alpha + \beta x_j^n + \varepsilon_j \quad (4)$$

where $\partial y / \partial x = \beta = 1$, in (3) and $\partial y / \partial x = n\beta x^{n-1} = nx^{n-1}$ in (4). Note that this implies that the average value of β found for (4) depends upon x and n.

The EM is (3) estimated with OLS. Thus, the funnel is due to the variation generated by the data and the estimation fault:

Section 4.2 DGP is (3) and $\varepsilon_t = N(0, \sigma^2)$. As this is the EM as well, the funnel is ideal.

Section 4.3 DGP is (3), but ε_t is non-normal. We use log-normal for the main results. Within reason funnels keep the ideal form.

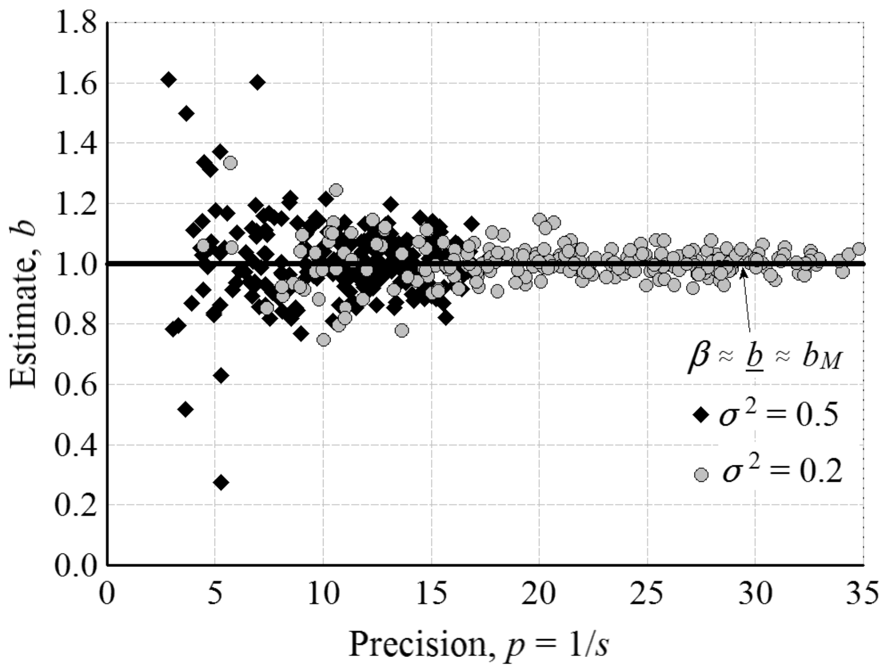
Section 4.4 DGP is (4), with $n > 1$, and $\varepsilon_t = N(0, \sigma^2)$, so the EM is misspecified. Within reason, funnels keep the ideal form, but the estimate of β goes bad, as it depends on the range of x-values covered by the data.

The words *within reason* in (4.3) and (4.4) refer to the fact that researchers know they should watch out for residual non-normality and non-linearity of the DGP, 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 DGP with normal residuals: Ideal funnels are narrow and symmetric

On Figure 4 DGP (3) is estimated with the right EM, $N = 250$ 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 *narrow* compared to the empirical funnels published in the typical meta study (see Stanley and Doucouliagos, 2010), 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 ideal funnels



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

The ideal case is easy to solve analytically for the shape of the reference funnel. For a linear model $y_i = \beta x_i + \varepsilon_i$ the variance of the estimator decreases linearly with the sample size

$$\hat{\sigma} = \frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i^2 \qquad (5)$$

$$Var\left[b|x\right] = \frac{\sigma^2}{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2} \rightarrow 0, \text{ for } n \rightarrow \infty \qquad (6)$$

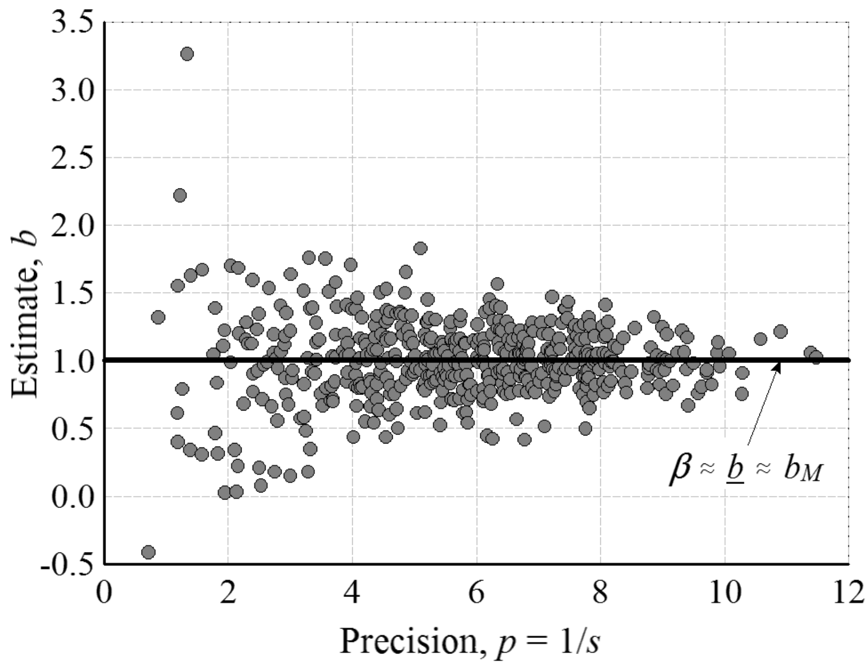
where n is the sample size. If the assumptions of the linear regression models are satisfied, b is an unbiased estimator. The two formulas explain the triangular and symmetrical form of the funnel, and how it converges to the true value of the parameter of interest.

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 in the DGP 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. The log-normal and the Weibull distribution are both asymmetric. Nevertheless, all three experiments produced ideal-looking funnels with a horizontal axis of symmetry intersecting the b-axis close to 1. Thus, both averages ($\hat{\underline{b}}$ and $\hat{b}_M$) are good estimates of the true value, see Table 4 below.

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



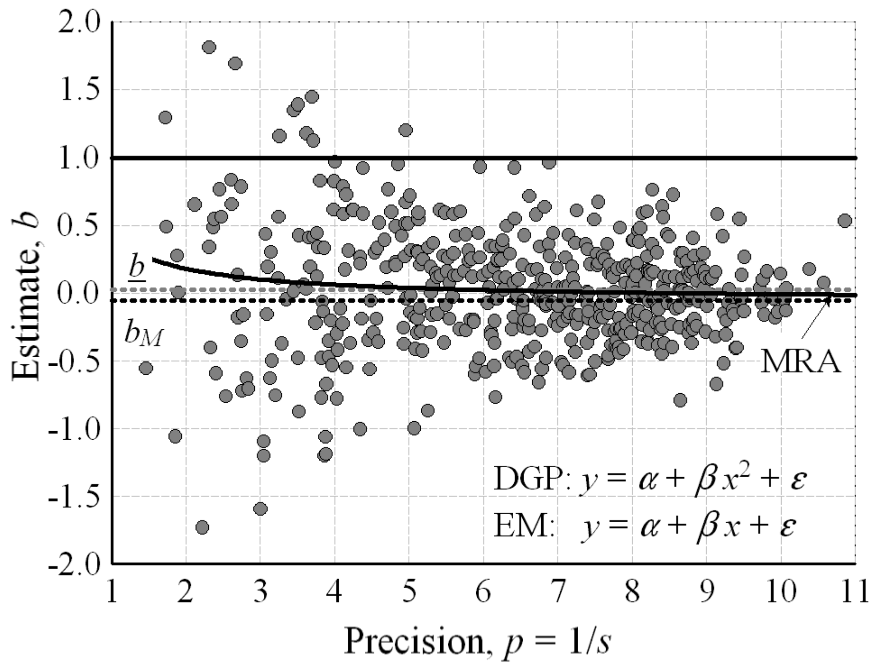
Note: Generated for $\sigma^2 = 0.5$ for $N = 500$. The MRA is indistinguishable from the $\beta = 1$-line.

Figure 5 shows a log-normal ε-distribution that may escape detection. Here the DGP 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 wider 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.

4.4 Estimation fault 2: The DGP is non-linear in x

We now turn to the case where the DGP is non-linear, while the EM is still (3) estimated with OLS. Here we use $\sigma^2 = 0.5$. Most functional forms seen in economics are so smooth that they can be approximated by a Taylor expansion. Hence, we cover most functional forms by analyzing what happens to the funnel if the function is in the power n = 2, 3, 4 … , where even n = 2 is high. To get asymmetry, we continue to n = 3.

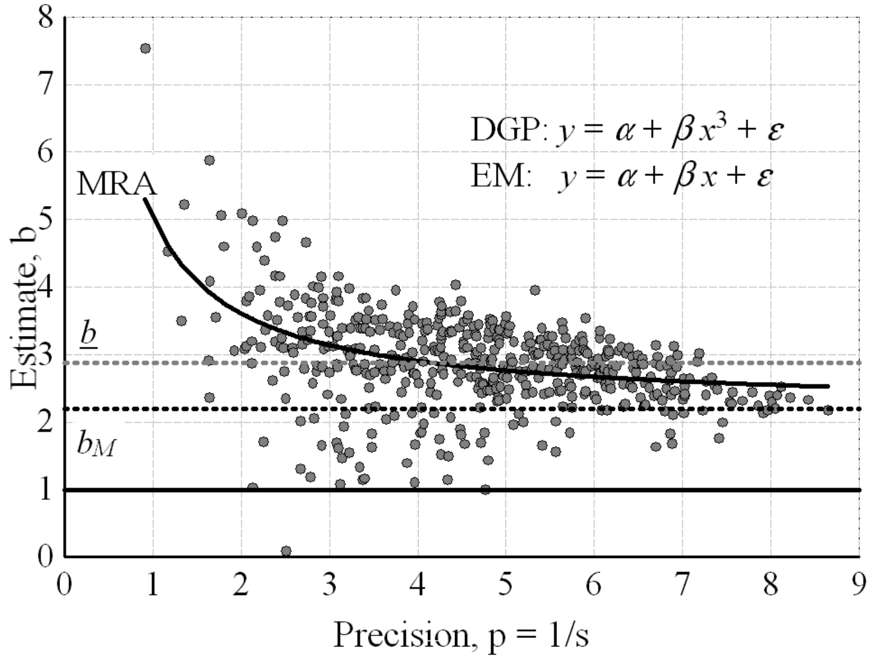
Figure 6a. The funnel for a quadratic form misspecification (with $\sigma^2 = 0.5$ and $N = 500$)



The quadratic DGP is $y_j = \alpha + \beta x_j^2 + \varepsilon_j$, where β = 1 and $\partial y / \partial x = 2x$. Here the funnel looks as shown on Figure 6a. The estimation fault makes the funnel much wider, but 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 $\underline{\hat{b}} = 0.027$, and $\hat{b}_M = -0.055$.

The cubic DGP is $y_j = \alpha + \beta x_j^3 + \varepsilon_j$, where β = 1 and $\partial y / \partial x = 3x^2$. Figure 6b shows what happens in a typical experiment: the funnel becomes still wider, and finally it looks asymmetric.

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



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 $\underline{\hat{b}} = 2.884$, and $\hat{b}_M = 2.201$. The MRA 'looks' for censoring, and consequently, it treats the asymmetry as due to censoring below 2.2. This 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.[19]

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 in Figure 6b before the funnel becomes asymmetric.

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 funnels graph in Figures 4 and 5

19. 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 like a censoring of most points below 2. Fortunately, 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.

are symmetric around $\beta = 1$. Therefore, the averages catch the true value of $\beta = 1$ rather well – the plain average being marginally better, as it should.[20] In the two cases of model-form estimation fault in Figures 6a and b, the MRA works rather badly when it tries to find the true value of β as already discussed. [21]

Table 3. Comparing the true value, β, the plain average, b, and the meta average, $\hat{b}_M$

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

Note: All estimates use Equation 1b. The true value of $\beta = 1$.

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

The results in sections 4.1 to 4.4 are based on one experiment with one funnel showing 500 estimates. Table 4 shows the results of a set of experiments each repeated 10,000 times. The Table has 5 panels where panels 2 to 5 correspond to Figures 2 to 6b as indicated in table heads. The ideal case of Figure 3 is considered in the left gray panel 2. As explained in section 2.6, we expect that R_{FAT} and R_{bM} are close to 0.05. The rejection rates are slightly higher in both cases, so the MRA errs marginally towards overestimating Type I error.

Next, we look at panel 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. The funnel is wider, so that slightly more funnels reject that $\hat{b}_M = 1$, but fewer funnels are found asymmetric. In both panel 2 and 3 the rejection rate falls with *N,* the number of points in the funnel. To disregard non-normality thus seems a rather innocent fault.

20. The meta average uses some degrees of freedom to correct for a problem that is not present.

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

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

Panel 1		Panel 2			Panel 3			Panel 4			Panel 5		
		The DGP is linear						The DGP is non-linear					
Dimensions		Ideal			Log-normal residuals			Squared x-term			Cubic x-term		
		Avr.	Rejection rates		Avr.	Rejection rates		Avr.	Rejection rates		Avr.	Rejection rates	
N	M	$\hat{b}_M$	R_{FAT}	R_{bM}	$\hat{b}_M$	R_{FAT}	R_{bM}	$\hat{b}_M$	R_{FAT}	R_{bM}	$\hat{b}_M$	R_{FAT}	R_{bM}
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, σ^2 = 0.5 in all simulations. The significance level for the rejections 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.

Panels 4 and 5 report the cases of non-linearity. Here $\hat{b}_M$ = 1 is rejected very often, as it should. It is surprising that the FAT rejects symmetry less often for the squared form than in the ideal case – though only marginally so. Our interpretation is that the width increases, and therefore, it becomes more difficult to reject asymmetry. In the cubic case, the FAT does reject symmetry in 93.5 % of all cases.

5. Omitted variables

All meta-analysis studies in economics, we know of, show a lot of model-uncertainty. This appears to be the main explanation of the excess variability observation. Our set-up generates model uncertainty by using a DGP with a set **C** of K variables, and an EM with a subset **c** of **C** containing k elements. If $k < K$ the estimator suffers from omitted variables bias.

Sections 5.1 to 5.5 examine the case with one omitted variable, where $k = 0$ and $K = 1$. The analytics are considered in 5.1, which also show that the PET estimate is likely to find the most biased estimate; 5.2 gives two examples, 5.3 explores a range of possibilities, 5.4 shows that the extended FAT-PET MRA does handle omitted variables rather well, and finally 5.5 and 5.6 look at cases with many omitted variables.

5.1 One undetected omitted variable

With one omitted variable the DGP-EM set becomes:

DGP: $y_j = \alpha + \beta x_j + \delta z_j + \varepsilon_j$, where $\varepsilon_t = N(0, \sigma^2)$ (7a)

EM: $y_j = \alpha + \beta_z x_j + \varepsilon_j$, where β_z has an omitted variable bias (7b)

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

$$z_j = \lambda x_j + \upsilon_j, \text{ where } \lambda = \sqrt{\frac{\rho^2}{1-\rho^2}} \text{ and } \upsilon_j = N(0, \sigma^2) \tag{8}$$

The DGP (7a) and EM (7b) are equivalent if the expected value $E(z|x) = 0$, i.e., iff $\lambda = 0$ ➔ $\rho = 0$. Here z and x are independent, and $\beta_z = \beta = 1$. In all other cases, there is a positive omitted variable bias. When β_z is estimated (by OLS) by model (3b), the bias is a simple linear relation, where to simplify α is set at zero:

$$\hat{b}_z = (x'x)^{-1} x'y = (x'x)^{-1} x'x\beta + (x'x)^{-1} x'z\delta + (x'x)^{-1} x'\varepsilon$$
$$E\left[\hat{b}_z\right] = \beta + (x'x)^{-1} x'z\delta = \beta + \lambda\delta \tag{9}$$

The funnel for $\delta = 0$ is the *true* funnel, with average 1 as it should be. For other values of δ an omitted variable bias occur, and in many cases the biased $\beta_z > \beta$. Here the MRA estimates β_z. This is, at a first consideration, it means that the most precisely estimated β's are the most biased ones, but recall the estimator of the variance of b, the estimated value of β:

$$Var\left[b|x\right] = \frac{\sigma^2}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} \tag{10}$$

This implies that the more variation in x, the more precisely estimated the coefficient is. Thus, when no control is included in an EM, while the DGP contains a control, all the variation in the data is contained in x. This gives a more precisely estimated coefficient than when the true EM is estimated, 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 when they are randomly included or omitted. In this case, the precision depends mainly on the level of correlation between the control and x and thus among the controls themselves.

5.2 Bimodal funnels: One omitted variable in half the estimates

To understand what is going on, we start with equation (7). The DGP contains 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 (see equation 9) and the variance bias. They enter as a product in the formula for the omitted variable bias, but their effect on the variance is less straightforward.

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

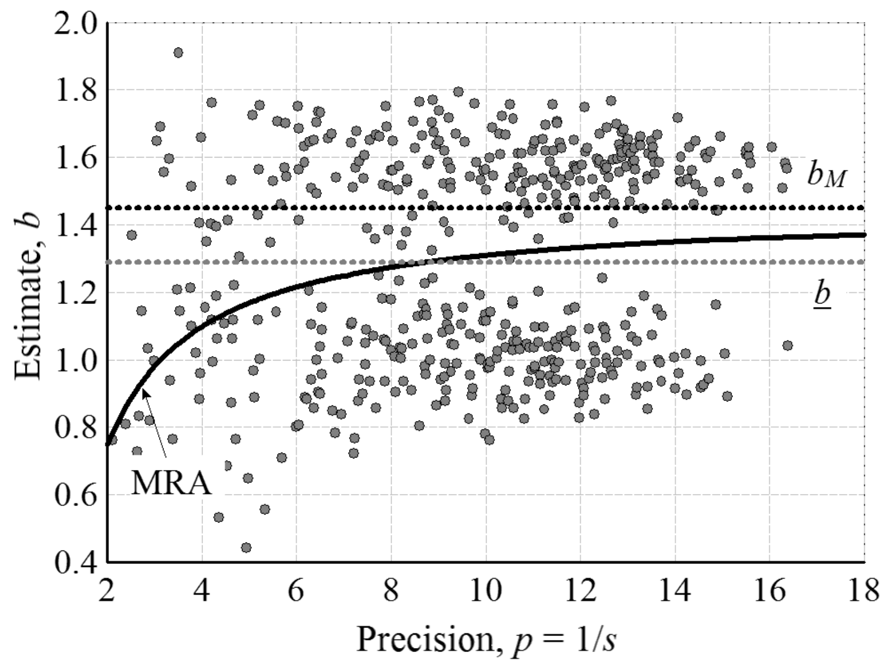
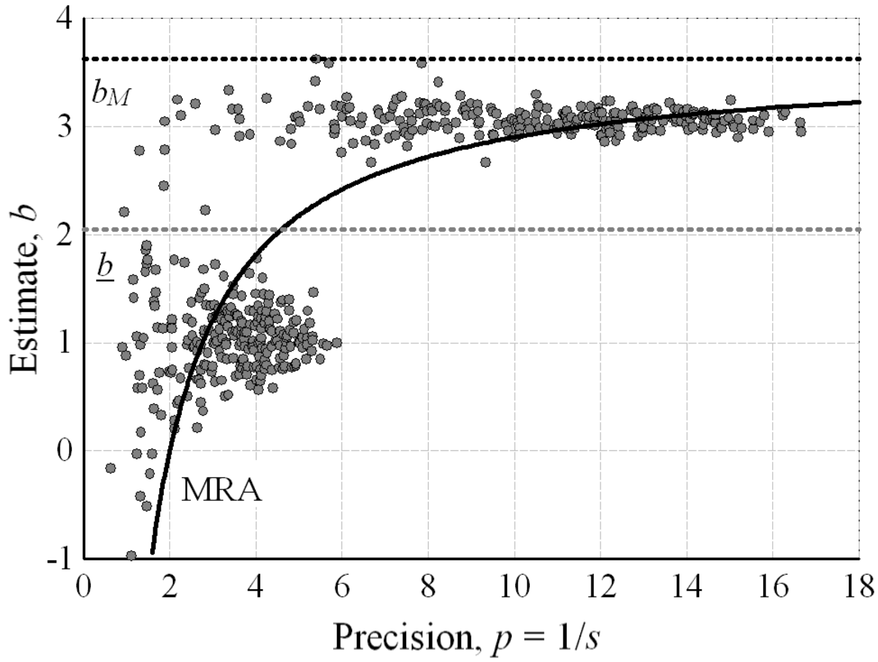


Figure 7b. Merged funnel for $\delta = 1$, $\rho = 0.9$ and $N = 500$ (line ** in Table 5)



Note: The two funnels both generated for 250 right estimates and 250 estimates with omitted variable.

The two illustrative cases are shown as Figure 7a, where $\delta = 1$ and $\rho = 0.5$, and Figure 7b, where $\delta = 1$ and $\rho = 0.9$. The 50% rate of omission is chosen for illustrative purposes as it is the one giving the clearest bimodal funnel. In both cases the reader can see that the funnel is bimodal,[22] with two sub-funnels appearing as *peaks* in the horizontal direction. The sub-funnels at $b = 1$ are correct, while the other is the omitted variable 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 chosen as 0.5, so the two sub-funnels have the same number of observations.

The two examples give bimodal funnels, with a highly notable "dent" between the two peaks. Once the funnel is displayed it is clear that something is amiss. However, if the two the two peaks are close and if there are fewer observations for the second peak it might look less obvious what is going on.

In both cases drawn, the MRA converges to the omitted variable peak and not to the correct one, so $\beta < \underline{\hat{b}} < \hat{b}_M$. Consequently, if the meta-analysis fails to include a control for z in the MRA, the meta average is worse than the plain average. Table 5 shows that this is not a general result.

22. 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.

5.3 A systematic study of funnels with one omitted variable in half the estimates

The cases shown are systematically analyzed in Table 5. It studies what happens when the omitted variable, $\delta = 0, 0.5, \ldots, 3$ and the correlation is $\rho = 0.1$, 0.5 and 0.9. The two cases of Figures 7a and 7b 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.

Table 5. Experiments with bias from omitted variables

Panel 1 Parameters		Panel 2 $N = 40, M = 500$					Panel 3 $N = 100, M = 500$				
		Averages			Rejection ratios		Averages			Rejection ratios	
δ	ρ	$\underline{\hat{b}}$	$\hat{b}_M$	Best	R_{FAT}	R_{bM}	$\underline{\hat{b}}$	$\hat{b}_M$	Best	R_{FAT}	R_{bM}
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	0431	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 that corresponds to Figure 7a has *****. The line that corresponds to Figure 7b has ******.

The pattern is as expected from the graphs and equation 7: When ρ and δ are small, the bias is also small, and it causes small problems as well. However, as $\hat{b}_M$ move away from 1, the FAT-test starts to detects asymmetries ($\hat{\gamma} \neq 0$), and the fraction of bad estimates of $\hat{b}_M$ rises.

It is interesting to note that in about half the cases $\underline{\hat{b}}$ is closer to 1 than is $\hat{b}_M$, 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 $\hat{b}_M$ = 1 is almost always rejected.

5.4 *Undetected and detected cases with one omitted variable: Using the MRA(k)*

Given the pattern in Table 5, we examine a set of cases to see if it helps, if the omitted variable 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:

Table 6. Experiments with an undetected and a detected omitted variable

Panel 1 Dimensions				Panel 2 Undetected MRA				Panel 3 Detected MRA(k)		
		Omitted variable characteristics		Averages		Frequency of Rejection		Average	Frequency of rejection	
				plain	meta			same		
N	M	δ	ρ	$\underline{\hat{b}}$	b_M	R_{FAT}	R_{bM}	$\underline{\hat{b}} = \hat{b}_M$	R_{FAT}	R_{bM}
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

For the MRA with an undetected omitted variable, the results are as before, but if it is detected and the MRA(k) is used, the results are amazingly good: the average $\underline{\hat{b}}$ and $\hat{b}_M$ *'s* found are very close to 1 in all cases, and the number of $\hat{b}_M \neq 1$ remains constantly around 6 %, as in simulations of ideal funnels. Hence, if the omitted variable 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 omitted variable, 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 % level quickly falls below 5 %.

5.5 Model variability: The case with too many potential controls

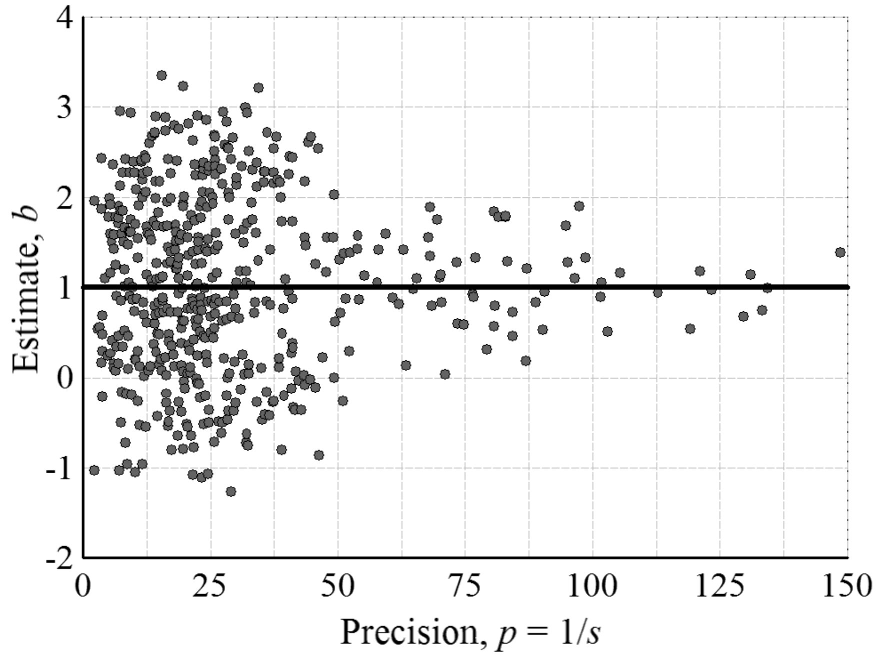
We now turn to more general setting to see how funnels may look in a world where the true relation has many variables and the operational models presented in the literature have fewer, but often different, variables.[23] The simple bimodal case discussed suggests that the multi-modal case generates wide funnels that are often asymmetric. We are worried that this may generate funnels that look deceptively like censored funnels, with a flat top or bottom and many peaks that overlap to give a distribution without obvious dents.

This is a large field of study, so we just make a few experiments, showing that the range of possible outcomes is wide indeed. The set-up for the experiments are: The DGP is composed of *K* controls where each is included or omitted with a certain probability which we have 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.

Assume the DGP has *K* controls, while each EM is estimated with $k_i \leq K$ controls. Define $\boldsymbol{C}$ as the set of controls in the true model. If $\boldsymbol{C} \subset \bigcup_i \boldsymbol{c_i}$,where $\boldsymbol{c_i}$ is the set of controls included in model i. Then the DGP is included in the union of all EMs 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 omitted variable biases. The complement of $\boldsymbol{c}_i$ is termed $\boldsymbol{c}_i^c$ in the set $\boldsymbol{C}$. If the MRA is modified with binary dummies equal to 1 for each element of $\boldsymbol{c}_i^c$ and otherwise 0. By using higher *K*'s and *k*'s, formula (7) expands into:

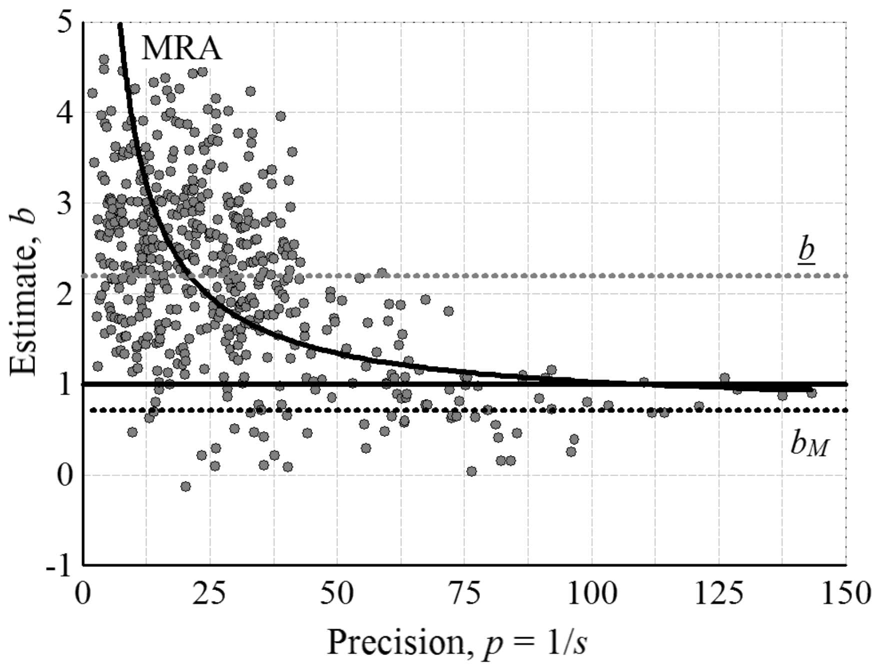
23. Cross-country studies of economic growth rarely include more than a dozen explanatory variables, but the literature has used more than 100 variables often in a handful of versions, which have all been found significant in at least one article, see Appendix B in Durlauf, Johnson and Temple (2005).

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. A smooth asymmetric funnel made as described in text ($N = 500$)



DGP: $y_j = \alpha + \beta x_j + \delta_1 z_{1j} + ... + \delta_K z_{Kj} + \varepsilon_j$ (11a)

EM: $y_j = \alpha + \beta_z x_j + \delta_1 z_{1j} + ... + \delta_k z_{kj} + \varepsilon_j$, where $K - k$ variable are omitted (11b)

The number of different biases – and hereby peaks – produced by (7) on β by different selection of the k controls are:

$$\lambda(K,k) = \binom{K}{k} \tag{12}$$

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 which 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 closer to the true value. The funnel will thus appear to be censored above, and the PET-estimate of the meta average will converge to the most biased estimate.

5.6 Two funnel with six omitted variables: 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 δ = 2, 1, 0.5, −0.5, −1 and −2, while in the asymmetric case they are δ = 2, 2, 1, 0.5, −0.5, and −1.

The reader may think of the wide funnel in Figure 8a as a realistic case of model uncertainty. It certainly produces the excess variation results with lots of estimated bs 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 $\hat{b}_M$ is a better average than $\underline{\hat{b}}$.

Furthermore, a set of simulations were made with the two cases of Figure 8. In these cases the MRA(k) also worked very well. These results are available from the authors.

6. Summary

A funnel shows the distribution of a set of estimates of the same parameter. If the funnel is due to noise in the data only, it is symmetric and narrow. All empirical funnels in economics have considerable excess width, and most are asymmetric as well. Most meta-analysts assume that funnel width is due to model variation, and funnel asymmetries are due to censoring caused by a common prior.

The best practice tool in meta-analysis in macroeconomics is the FAT-PET MRA, which works on the data of the funnel. It contains a FAT-term that tests for funnel asymmetry. We confirm that the FAT is a powerful test irrespective of the cause of the asymmetry. The PET-term is tooled to adjust the average for censoring biases. We confirm that it does an excellent job in this respect. Unfortunately, censoring asymmetries are not the only asymmetries that occur in funnels. The paper studies a range of problems that may generate 'natural' asymmetries without censoring: Data dependencies, estimation faults, and omitted variables. In all cases considered, the problem widens the funnel.

First, section 3 studied the case of dependent but expanding data. Here the results did converge to the true value, but often quite slowly, so in a surprisingly large number of cases the FAT rejects symmetry, and both the PET and the plain average differ from the true value. The problems increased with a structural break in the data.

Second, section 4 analyzed estimation faults. One fault occurs when it is disregarded that the residuals in the regression are non-normal. It appears an innocent 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, section 5 considered omitted variables. 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 omitted variable 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. If the omitted variable is controlled for by a binary control in the MRA, the asymmetry disappears.

Hence, it is important to know if an observed asymmetry is natural or due to censoring. We recommend the meta-analyst to study the funnel and try to determine the reason for the asymmetry, 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 that led to a major revision. We wish to thank Bent Jesper Christensen, who helped us start up the paper, and Erich Gundlach for a useful discussion. Also, we are grateful to the editor and the other referee for detailed and very constructive advice.

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., 2010. The Ineffectiveness of development aid on growth: An update. *European Journal of Political Economy* doi:10.1016/j.ejpoleco.2010.11.004

Doucouligos, H., Paldam, M., 2008. Aid effectiveness on growth. A meta study. *European Journal of Political Economy* 24, 1-24 – see also update in Doucouliagos and Paldam (2010)

Durlauf, S.N., Johnson, P.A., Temple, J.R.W., 2005. Growth econometrics. Cpt. 8, pp 555-677 in Aghion, P., Durlauf, S.N., eds. *Handbook of Economic Growth*, Vol 1A. 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

Kennedy, P., 2003. A guide to econometrics. (5th ed.). MIT Press, Cambridge, MA

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

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

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., 2010. Picture this: A simple graph that reveals much ado about research. *Journal of Economic Surveys* 24, 170–91

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