Faculty of Business and Law

SCHOOL OF ACCOUNTING, ECONOMICS AND FINANCE

School Working Paper - Economic Series 2006

SWP 2006/25

Two-Stage Precision-Effect Estimation and Heckman Meta-Regression for Publication Selection Bias†

T.D. Stanley

Two-Stage Precision-Effect Estimation and Heckman Meta-Regression for Publication Selection Bias †

by
*T.D. Stanley**

Abstract: This study offers a simple meta-regression method for estimating genuine empirical effects in research literatures tainted by publication selection. Two-stage precision-effect (PETS) corrects for the misspecification of conventional meta-regression models and provides a viable strategy for estimating empirical economic effects. C13, B40.
Keywords: Meta-regression analysis, publication bias, Heckman regression, sample selection.

**Department of Economics, Hendrix College, 1600 Washington St., Conway, AR USA (e-mail:Stanley@Hendrix.edu)* fax: 501-450-1400
† I gratefully acknowledge the support of a US Environmental Protection Agency STAR (Science To Achieve Results) grant #RD-832-421-01. Although the research has been funded in part by the US-EPA, it has not been subjected to the Agency's peer and policy review and therefore does not necessarily reflect the views of the Agency. Any remaining error or omission is solely my responsibility.

1. Introduction

Publication selection bias has long been recognized as a serious threat to an informed understanding of economics (DeLong and Lang, 1992; Card and Krueger, 1995). In its more benign form, it is the analysis and re-analysis of research data until a statistically significant result is found. Notorious cases of the suppression of damaging research results (for example, surrounding the use of Paxil and Vioxx) have caused leading medical journals to require prior registration of clinical trials as a condition for later publication (Krakovsky, 2004).

In economics, it has become common practice to use meta-regression analysis (MRA) to test and model publication selection (Card and Krueger, 1995; Ashenfelter et al. 1999; Roberts and Stanley, 2005; Mookerjee, 2006). Meta-regression analysis collects the reported estimates from each study and attempts to explain their wide study-to-study variation (Stanley and Jarrell, 1989). Although MRA models provide valid tests for the identification of publication selection and a genuine effect beyond publication selection (Stanley, 2007), they are mis-specified and produce biased estimates. The purpose of this paper is to identify the source of this misspecification, to explore the implications, and to offer an estimation strategy (two-stage precision-effect—PETS) that is not subject to this misspecification bias.

2. The meta-regression models of publication selection

Following Card and Krueger (1995), publication selection bias has often been modeled by a meta-regression of a study's reported effects on its standard error.

$$e_i = \beta_1 + \beta_0 S_i + \varepsilon_i \qquad (1)$$

(Ashenfelter et al. 1999; Mookerjee, 2006). Where e_i is an estimated effect (*e.g.*, elasticity), and S_i is e_i's standard error. With obvious heteroscedasticity, the WLS version of MRA model (1) is preferred.

$$t_i = e_i/S_i = \beta_0 + \beta_1(1/S_i) + u_i \tag{2}$$

Where $1/S_i$ is 'precision.' These MRA models are widely used in medical research to identify publication bias (Egger et al., 1997; Sutton et al., 2000), and economists use H_0: β_1=0 to test for the presence of a genuine effect beyond publication selection (Stanley, 2005; Roberts and Stanley, 2005). Simulations show that these MRAs provide a valid and powerful test of an empirical effect corrected for publication bias (Stanley, 2007).

3. Heckman regression of publication selection[1]

Publication selection bias is a special case of sample selection bias. The 'Heckman regression,' which is the second step of his two-stage method (Heckman, 1979), is very similar to the MRA model (1). The Heckman regression can be written as:

$$e_i = \beta_1 + \rho\sigma\, I(T_e/\sigma - c) + v_i \tag{3}$$

(Davidson and MacKinnon, 2004, pp. 486-89, Wooldridge, 2006).[2] Where T_e is the 'true' effect, σ is the standard error of estimating T_e, ρ is the correlation between the estimation errors and the random errors of publication selection, $I(T_e/\sigma - c)$ is the inverse Mills ratio, and c is the critical value from the t-distribution. The inverse Mills ratio is the ratio of the standard normal probability density function, evaluated at $T_e/\sigma - c$, to its cumulative probability.

The Heckman two-stage method requires a sample containing both published and unpublished effects to estimate the inverse Mills ratio. This is where approaches to publication selection must depart from sample selection because we do not observe unpublished and unreported estimates. Thus, Heckman's two-step method cannot be applied to publication selection.

---

[1] The idea that the Heckman selection regression serves as the foundation for the above MRA model of publication selection comes from (Doucouliagos and Stanley, 2006).

Fortunately, there is another approach to identifying the variation in the second term of Heckman's regression, $\rho\sigma I(T_e/\sigma - c)$. Heteroscedasticity of estimated effects in economic research will cause variation this second term of the Heckman regression, which allows for its estimation and leads to the MRA model (1). That is, for publication selection, we expect σ to vary from study to study. To make this connection between the Heckman regression and equation (1) explicit, assume that the inverse Mills ratio is constant, say k. In this case, we can set $\beta_0=\rho k$, and S_i replaces σ. With these substitutions, equation (3) becomes equation (1).

If the inverse Mills ratio were constant and independent of S_i, then MRA model (1) would be correctly specified and could provide consistent estimates of the true effect corrected for publication bias. Unfortunately, neither is true, which explains the low power of using H_0: $\beta_0=0$ as a test for publication bias (Egger et al., 1997). The inverse Mills ratio depends on S_i.[3] Thus, the relationship between the observed effect and its standard error will be nonlinear in the presence of publication selection.

This connection to Heckman regression explains why MRA models (1) and (2) give biased estimates. Furthermore, if this nonlinearity of reported effect to its standard error were the true source of these MRA models' misspecification, then it should be seen in the research record. Autocorrelation found in MRA model (1) when research results are sorted by S_i would confirm this nonlinearity. All four economic MRAs for which I have the necessary data and that exhibit publication selection contain positive autocorrelation.[4]

This nonlinearity with respect to S_i becomes the basis for a new approach to estimation, corrected for publication bias—two-stage precision-effect estimate (PETS).

---

[2] β_1 in equation (3) may also be replaced by an explanatory model, $Z\beta$.

[3] An exception to this rule is for the case when there is no genuine effect, T_e=0. In general, it can be shown that the derivative of inverse Mills ratio w.r.t. S_i is: $[T_e I(T_e/S_i - c)/S_i^2][(T_e/S_i - c) + I(T_e/S_i - c)]$. See Davidson and MacKinnon (2004) and Wooldridge (2006, p. 598). Because this derivative is positive when T_e >0, the inverse Mills ratio increases with S_i and is not constant. But conversely, when T_e =0, the relationship between the observed effect and its standard error will be linear.

[4] Details are available from the author.

Typically, nonlinear relations are estimated using a power series. Thus, PETS begins with the square of S_i as the second term of MRA model (1).

4. Simulation results for two-stage precision-effect[5]

Table 1 reports simulation results of 10,000 replications of the average observed effects as estimated by a number of obvious estimators. The simple average is $\Sigma e_i/L$, where L is the number of studies in the research literature. Fixed-effects and random-effects are well-known weighted averages that use different measures of e_i's variance as the weight (Sutton et al., 2000). $\hat{\beta}_1$ is the MRA estimate from equation (2), and PETS is the two-stage precision-effect estimate.

Very precise studies easily produce statistically significant t-values (hence publishable), while imprecise studies require much selection before a significant result is manufactured. Thus, publication bias will be nonlinear. To model this nonlinear relationship, the first stage of PETS estimates a quadratic version of MRA equation (2):

$$e_i = \beta_1 + \beta_0 S_i^2 + \xi_i \tag{4}$$

From this first-stage, $\hat{\beta}_0 S_i^2$ is used to estimate the magnitude of study i's publication bias. Next, these estimated publication biases are subtracted from the reported estimated effects, and a second MRA model is run on equation (2) using corrected t-values.[6] The coefficient on precision ($1/S_i$) in this second stage version of equation (2) defines the PETS estimate of effect corrected for publication bias.

---

[5] Details of the simulation design are given in Stanley (2007) or may be obtained from the author. σ_{bias} represents the magnitude of the random misspecification biases introduced into these simulations. Of course, PETS, as well as the other estimators, perform better when there are no misspecification biases. Values of σ_{bias} were chosen to be realistic, yet conservative.

[6] In the second stage, the MRA model is forced through the origin, because publication bias has already been filtered from the estimates. A one-stage MRA version, which divides equation (4) by S_i, also does a good job in estimating empirical effect corrected for publication bias. However, the confidence intervals produced by PETS more accurately reflect their nominal levels, due to the obvious multicollinearity of the one-stage MRA.

{Insert Table 1 about here}

When there are genuine effects and publication selection, Table 1 shows that PETS has the smallest bias. However, the simple MRA estimate of β_1 from equation (2) has the smallest bias when there is no underlying empirical effect (*i.e.*, β_1=0). The superiority of $\hat{\beta}_1$ when T_e=0 is an implication of our interpretation the MRA model of publication selection as a Heckman regression. When T_e=0, the relation between an estimated effect and its standard error will be linear, and MRA model (2) is correctly specified (recall note 2). Thus, two-stage precision-effect should be used only when then there is evidence of a genuine effect (*i.e.*, we can reject H_0: β_1=0).

5. Conclusions

The purpose of this paper is to identify a misspecification in the linear meta-regression models commonly used to identify and correct publication selection bias and to offer a solution. Due to the nonlinearity of the inverse Mills ratio in the Heckman regression, these MRA models are mis-specified and can produce biased estimates. The two-stage precision-effect (PETS), which first estimates and removes the nonlinear publication bias has smaller bias than conventional methods. The validity of this interpretation of the MRA models of publication selection is also confirmed by the autocorrelation found among previous MRAs in economics and by the superiority of the simple MRA estimate of β_1 from equation (2) when there is no genuine empirical effect.

References:

Ashenfelter, O., C. Harmon and H. Oosterbeek, 1999. A review of estimates of the schooling/earnings relationship, with tests for publication bias, Labour Economics 6, 453-470.

Card, D. and A.B. Krueger, (1995) Time-series minimum-wage studies: A meta-analysis, American Economic Review 85, 238-43.

Davidson, R. and J.G. MacKinnon, 2004. Econometric Theory and Methods, Oxford University Press, Oxford.

De Long, J.B. and K. Lang, 1992. Are all economic hypotheses false? Journal of Political Economy 100,1257-72.

Doucouliagos, H. and T.D. Stanley, 2006. Identifying and correcting publication selection bias in the democracy-growth literature: Heckman meta-regression, in review.

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

Heckman, J.J., 1979. Sample Selection Bias as a Specification Error, Econometrica 47, 153-62.

Krakovsky, M., 2004. Register or perish, Scientific American 291(Dec.), 18-20.

Mookerjee, R., 2006. A Meta-Analysis of the Export Growth Hypothesis. Economics Letters 91, 395-401.

Roberts, C. J. and T.D. Stanley (eds.), 2005. Meta-Regression Analysis: Issues of Publication Bias in Economics, Blackwell, Oxford.

Stanley, T.D., 2005. Beyond publication bias. Journal of Economic Surveys 19, 309-45.

Stanley, T.D., 2007. Meta-regression methods for detecting and estimating empirical effect in the presence of publication selection, Oxford Bulletin of Economics and Statistics, forthcoming.

Stanley, T.D. and S.B. Jarrell, 1989. Meta-regression analysis: A quantitative method of literature surveys, Journal of Economic Surveys, 3: 54-67.

Sutton, A.J. et al., 2000. Methods for Meta-analysis in Medical Research. John Wiley and Sons, Chichester.

Wooldridge, J.M., 2006. Introductory Econometrics: A Modern Approach. 3rd ed. Mason, South-Western.

Table 1: Mean Effects of PETS and Conventional Estimators

Hetero-geneity	True effect	Selection Incidence	Sample Size	Simple Average	Fixed-effect	Random-effect	$\hat{\beta}_1$	PETS
σ_{bias}=.25	0	0%	20	-.0005	.0002	.0003	.0033	-.0004
	0	0%	80	-.0012	.0007	.0007	.0029	-.0005
	0	25%	20	.2334	.1953	.2140	.0415	.0915
	0	25%	80	.2345	.1973	.2158	.0391	.0912
	0	50%	20	.4676	.3927	.4304	.0608	.1892
	0	50%	80	.4677	.3920	.4300	.0619	.1839
	0	75%	20	.7004	.5846	.6273	.0699	.2856
	0	75%	80	.7016	.5857	.6288	.0738	.2824
	1	0%	20	1.0012	.9980	.9983	.9969	1.0059
	1	0%	80	1.0001	.9997	.9996	1.0002	.9993
	1	25%	20	1.0660	1.0361	1.0429	.9289	.9763
	1	25%	80	1.0653	1.0385	1.0450	.9272	.9773
	1	50%	20	1.1320	1.0761	1.0876	.8495	.9545
	1	50%	80	1.1318	1.0762	1.0871	.8489	.9569
	1	75%	20	1.1988	1.1150	1.1297	.7693	.9372
	1	75%	80	1.1971	1.1133	1.1253	.7711	.9394
σ_{bias} =.50	0	0%	20	.0027	-.0025	-.0022	.0033	.0004
	0	0%	80	.0011	.0001	-.0003	-.0019	-.0024
	0	25%	20	.2701	.2267	.2546	.0449	.1114
	0	25%	80	.2706	.2258	.2545	.0391	.1089
	0	50%	20	.5393	.4486	.5043	.0874	.2384
	0	50%	80	.5399	.4510	.5062	.0847	.2301
	0	75%	20	.8115	.6802	.7461	.1373	.3734
	0	75%	80	.8095	.6803	.7478	.1358	.3644
	1	0%	20	.9996	.9996	.9999	.9943	.9981
	1	0%	80	.9991	.9993	.9994	.9993	1.0006
	1	25%	20	1.0962	1.0667	1.0815	.9388	.9921
	1	25%	80	1.0966	1.0652	1.0810	.9438	.9972
	1	50%	20	1.1958	1.1301	1.1590	.8827	.9980
	1	50%	80	1.1933	1.1298	1.1596	.8798	.9916
	1	75%	20	1.2901	1.1937	1.2317	.8171	1.0007
	1	75%	80	1.2909	1.1937	1.2331	.8096	.9910