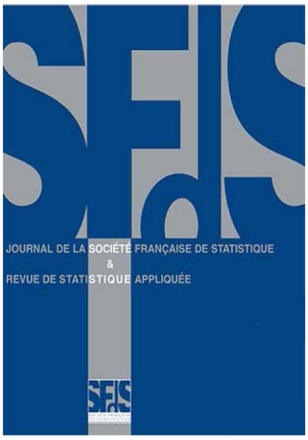
L'EXPLORATION STATISTIQUE DU BIAIS DE PUBLICATION



Patrice LAROCHE [1]

Journal de la Société Française de Statistique, tome 148, n° 4, 2007

RÉSUMÉ

Cet article présente les principales techniques statistiques permettant d'identifier et de corriger le biais de publication. Un biais de publication apparaît lorsque la publication d'une recherche dépend de la significativité et/ou de la direction des résultats obtenus par le chercheur. Plusieurs études ont ainsi montré que certaines revues académiques avaient tendance à un certain degré de publication sélective en ne retenant à la publication qu'un certain type d'études, notamment celles qui présentent des résultats statistiquement significatifs. Dès lors, une synthèse de la littérature qui s'appuierait uniquement sur les études publiées pourrait s'avérer biaisée en faveur d'études présentant des effets empiriques plus larges qu'ils ne sont réellement. Avec la méta-analyse, des techniques statistiques – comme le test d'asymétrie des entonnoirs et le test de méta-significativité – permettent d'identifier et de corriger ce type de biais.

Mots-clés : Biais de publication, test d'asymétrie des entonnoirs, test de méta-significativité, méta-analyse de régression.

1. CEREFIGE Nancy-Université, UFR AES – Université Nancy 2, 4 rue de la Ravinelle C.O. n°26 54 035 NANCY Cedex. Tel. 03.83.19.27.50., Fax. 03.83.19.27.63.,
e-mail : Patrice.Laroche@univ-nancy2.fr

L'auteur remercie les deux relecteurs anonymes et les co-éditeurs de la revue pour leurs remarques constructives ainsi que les professeurs Chris Doucouliagos et T.D. Stanley pour le partage de leur expertise. L'auteur tient également à exprimer ses remerciements à Antony Kuhn pour la lecture attentive qu'il a bien voulu faire de la première version de l'article. L'auteur porte évidemment seul la responsabilité des éventuelles erreurs et omissions contenues dans l'article.

Le développement des connaissances en sciences sociales, entre autres, rend de plus en plus indispensable l'utilisation de la méta-analyse pour synthétiser efficacement les résultats de la recherche scientifique. La méta-analyse est un terme générique qui désigne un ensemble de techniques statistiques pour combiner les résultats de plusieurs études s'intéressant à la même question (Rosenthal, 1991). Cette analyse simultanée d'un ensemble d'études produit un gain de puissance statistique dans la recherche du lien entre deux variables et permet, en cas de résultats discordants, d'obtenir une vue globale de la situation (Cucherat, 1997). En principe, les procédures méta-analytiques doivent s'appuyer sur une synthèse exhaustive de la littérature afin de combiner les résultats des études existantes. Or, le caractère souvent sélectif de la recherche bibliographique pose d'emblée le problème de la représentativité de l'échantillon d'études. En effet, le manque d'exhaustivité dans la collecte des travaux de recherche fait courir le risque d'introduire un biais dans la méta-analyse, appelé *biais de publication*. Un biais de publication peut apparaître lorsque la publication d'une recherche dépend de la significativité et/ou de la direction des résultats obtenus (Sutton *et al.*, 2000a). Plusieurs études ont ainsi montré que certaines revues académiques avaient tendance à un certain degré de publication sélective en ne retenant à la publication qu'un certain type d'études (Begg et Berlin, 1988 ; Begg, 1994).

La publication sélective résulte, d'une part, de l'autocensure des chercheurs et d'autre part, de la sélection des comités de lecture des revues (Hedges, 1992). En premier lieu, les auteurs peuvent s'autocensurer en considérant d'emblée que leurs résultats non significatifs sont sans intérêts et ont peu de chance d'être publiés (Gilbody *et al.*, 2000). Les auteurs peuvent également choisir de ne présenter à la publication que les résultats de leurs études qui confortent un positionnement théorique particulier. Dans les deux cas de figure, le comportement des chercheurs conduit à une publication sélective de leurs travaux de recherche. En second lieu, plusieurs études ont montré que certaines revues se laissaient influencer par les résultats d'un travail pour décider de le publier (Coursol et Wagner, 1986 ; Dickersin *et al.*, 1987 ; Begg et Berlin, 1988). Certaines revues auraient tendance à privilégier la publication d'études dont les résultats sont statistiquement significatifs. La publication sélective peut aussi refléter un positionnement théorique voire idéologique du comité de lecture des revues lui-même (Sterling, 1959 ; Smart, 1964 ; Mahoney, 1977). Dès lors, le problème de la non inclusion des études non publiées est un problème d'échantillonnage fréquent qui peut entacher la validité de la méta-analyse. Pour renforcer les conclusions de la méta-analyse, il convient alors de recourir à plusieurs techniques permettant de s'assurer que les études retenues sont représentatives de l'ensemble des études existantes. L'objet de cet article est de présenter les principaux outils d'investigation du biais de publication à la disposition du méta-analyste. Pour ce faire, les techniques d'identification et de correction de la publication sélective seront illustrées à partir de données issues de trois méta-analyses :

- celle de Doucouliagos et Laroche (2003) consacrée au lien syndicat/productivité [Union] ;

- celle de Dalhuisen, Florax, DeGroot et Nijkamp (2003) sur l'élasticité-prix de la demande d'eau [Elast] ;
- et celle de Card et Krueger (1995) portant sur la relation salaire minimum/emplois [MinWage].

Dans une première partie, sont exposés les outils d'investigation graphique du biais de publication. La seconde partie est consacrée à la présentation de la méta-régression et des outils statistiques permettant d'identifier et de corriger le biais de publication. Enfin, la troisième partie présente un procédé, proposé par Rosenthal (1979), qui permet d'évaluer la robustesse des résultats de la méta-analyse à l'introduction d'études non publiées.

1. Les outils d'investigations graphiques du biais de publication

Plusieurs méthodes graphiques ont été proposées afin d'explorer les données issues de méta-analyses. Ces techniques peuvent être employées afin de mettre en évidence l'existence d'un biais de publication mais également afin d'obtenir, de manière plus générale, une meilleure compréhension des phénomènes étudiés. On peut distinguer deux principaux types de représentations graphiques : les graphiques en entonnoir ou « *funnel graphs* » (1.1.) et les graphes radiaires de Galbraith (1.2.).

1.1. Les graphiques en entonnoir ou « funnel graphs »

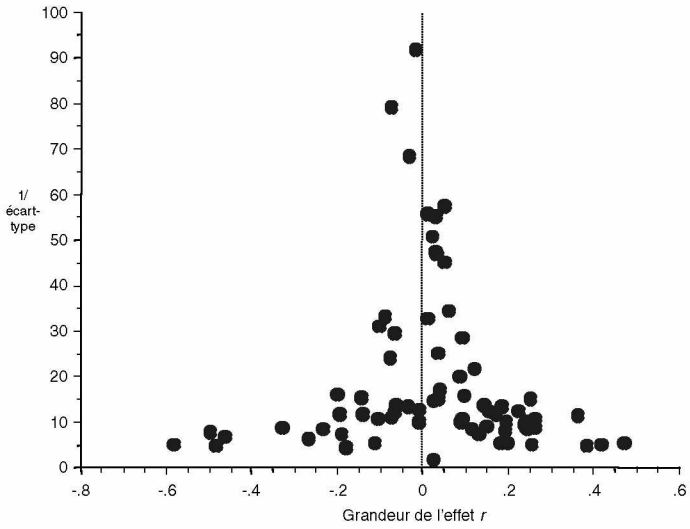
La méthode la plus simple et la plus utilisée par les méta-analystes s'appuie sur l'examen attentif d'une représentation dite « *funnel plot* » ou « *funnel graph* » (« graphiques en entonnoir ») (Sutton *et al.*, 2000a, page 1574) [2]. Ce type de représentation graphique consiste à représenter les *estimations des grandeurs d'effet* [3] tirées de chaque étude en fonction de la taille de l'échantillon de chaque étude. Le nom de « *funnel plot* » provient du fait que la précision de l'estimation de la « vraie » grandeur d'effet augmente avec la taille n de l'échantillon – quel que soit l'indicateur utilisé (n ou $\sqrt{n}$) – ou avec l'inverse de l'écart-type [4] de l'estimation issue de chaque étude i ($1/\sigma_i$). Dès lors, lorsque la valeur de $1/\sigma_i$ de chaque étude est représentée sur l'axe vertical (des Y), les résultats issus de petits échantillons d'observations vont varier avec une plus grande amplitude autour de la vraie grandeur d'effets. D'un

2. D'autres techniques purement descriptive permettent de rechercher l'existence d'un biais de publication à partir des données d'une méta-analyse. La simple vérification visuelle de l'allure gaussienne d'un histogramme permet un diagnostic rapide. En effet, la publication des résultats peut être vue comme un phénomène d'échantillonnage aléatoire au sein de la population de tous les résultats possibles obtenus.

3. La *grandeur de l'effet* peut se définir comme l'estimation du degré de relation entre deux variables d'intérêt. Les méta-analystes parlent également de *taille de l'effet* ou d'*importance de l'effet*.

4. L'écart-type d'une distribution d'échantillonnage est généralement appelé *erreur standard* ou *erreur type* de cette distribution. Afin de ne pas embrouiller les lecteurs, le terme écart-type a été privilégié tout au long de l'article et le symbole adopté est σ_i (écart-type de l'estimateur tiré de l'étude i).

point de vue graphique, ces résultats vont se répartir très largement à la base du graphe alors que les résultats issus de grands échantillons vont se répartir plus étroitement dans la partie haute du graphe. En l'absence de biais de publication, les différents résultats obtenus par les études empiriques vont être répartis de façon homogène autour de la vraie grandeur d'effet. La figure 1 représente une situation de ce type, obtenue à partir des données de la méta-analyse de Doucouliagos et Laroche (2003). On observe bien que les estimations issues des études sur petits échantillons varient avec une plus grande amplitude autour de la « vraie » grandeur d'effet que celles comportant un plus grand nombre d'observations. En l'absence de biais, le nuage de points est alors symétrique et conduit à une forme « d'entonnoir » sur le graphique.

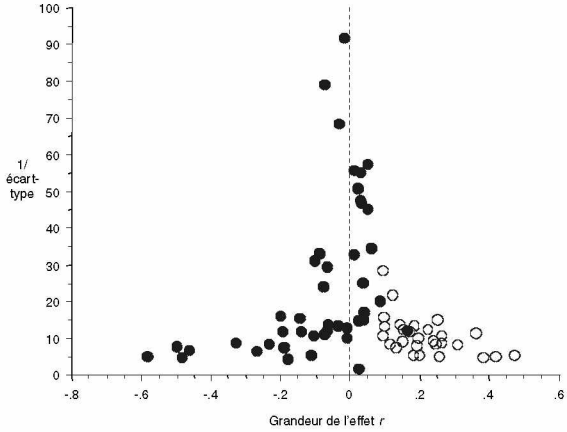


Source : Doucouliagos C., Laroche P. et Stanley T.D. (2005)

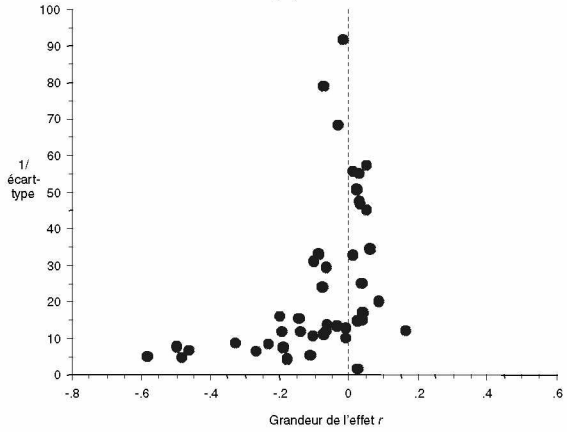
FIG 1. — Graphique en entonnoir en l'absence de biais de publication.

En présence d'un biais de publication, la répartition des observations n'est pas homogène. En effet, supposons un ensemble d'études empiriques dont une partie n'est pas publiée – en général, il s'agit plutôt d'études reposant sur de petits échantillons – (cercles ouverts sur la figure 2a), on observera alors un « *funnel plot* » asymétrique, avec un espace libre en bas à droite du graphique (figure 2b) dans le cas d'un biais en faveur des études présentant des résultats négatifs. En définitive, la sur-représentation d'études d'un côté ou de l'autre du graphique peut être considérée comme une preuve de l'existence d'un certain degré de publication sélective de type I, c'est-à-dire privilégiant un sens de la relation.

Dans le cas de la figure 2 (cas extrême), la grandeur de l'effet calculée à partir de la méta-analyse va surestimer l'importance des études présentant des effets négatifs (*cf.* Figure 2b) et faire état d'une grandeur d'effet inférieure à zéro (droite en pointillés) sans pour autant que cet effet existe réellement dans la



(a)



(b)

FIG 2. — Graphiques en entonnoir en présence de biais de publication.

population (l'effet moyen est en réalité plutôt proche de zéro, *cf.* droite pleine sur le graphe). Plus l'asymétrie sera importante et plus le biais de publication sera significatif. L'identification du biais de publication à partir de l'asymétrie des « *funnel graphs* » défend l'idée que la publication sélective favorise *a priori* une direction particulière (par exemple, un effet négatif plutôt qu'un effet positif).

Si les « *funnel graphs* » sont d'une grande utilité pour visualiser les éventuels biais de publication de type I, il n'en demeure pas moins que leur utilisation présente plusieurs limites. En premier lieu, l'existence d'une dissymétrie peut venir non pas d'un biais de publication mais simplement d'une forte dissymétrie du problème statistique sous-jacent. Par exemple, si plusieurs chercheurs tentent d'estimer la moyenne d'une loi exponentielle avec des échantillons indépendants de tailles différentes alors la distribution de ces moyennes et les graphiques associés seront de fait dissymétriques. Dès lors, il serait er-

roné d'imputer cette dissymétrie à un biais de publication [5]. En second lieu, l'hétérogénéité des résultats entre les études peut provenir de la différence de qualité pouvant exister entre elles (Schultz *et al.*, 1995) [6]. Egger *et al.* (2003) font remarquer, par exemple, que les recherches portant sur de petits échantillons sont plus souvent entachées de failles méthodologiques qui nuisent à la qualité de la méta-analyse. Par conséquent, le fait d'incorporer ce type d'études dans une méta-analyse peut conduire à une dissymétrie sur un graphe en entonnoir. Enfin, la variabilité d'estimateurs d'une même quantité est non seulement due à des problèmes de taille d'échantillon et des problèmes statistiques sous-jacents mais aussi et surtout à l'utilisation de données différentes (données de panel, données en coupe instantanée, séries temporelles) et/ou de techniques d'estimation variées (fonction Cobb-Douglas, translog pour les études sur la productivité, etc.). Dès lors, l'asymétrie observée sur un « *funnel plot* » peut résulter du manque d'homogénéité des méthodes et des instruments utilisés pour étudier un même phénomène [7]. A ce propos, les méta-analystes offrent aujourd'hui des outils statistiques permettant d'incorporer dans les tests d'asymétrie des indicatrices susceptibles de contrôler les caractéristiques méthodologiques des études analysées. C'est le principe de la méta-régression explicative sur lequel nous reviendrons par la suite.

En définitive, le biais de publication n'est pas la seule source d'asymétrie pouvant apparaître sur un graphe en entonnoir et il est donc difficile d'affirmer l'existence d'un biais de publication avec les techniques graphiques. Néanmoins, il faut connaître ces méthodes et savoir les utiliser lorsque les conditions s'y prêtent.

Au delà du biais de publication de type I, les chercheurs, les réviseurs ou les rédacteurs en chef des revues académiques peuvent parfois favoriser les résultats statistiquement significatifs, quel que soit le sens de l'effet observé (*i.e.* sélection de type II). La représentation radiaire de Galbraith (1988) permet d'identifier la publication sélective de type II.

1.2. La représentation radiaire de Galbraith (1988) et la publication sélective de type II

Comme nous l'avons déjà évoqué, une publication sélective de type II est une tendance à publier davantage les études qui affichent des résultats significatifs et ce quel que soit le sens (positif ou négatif) de l'effet observé. La sélection de type II va conduire à une sur-représentation des études présentant des valeurs

5. L'auteur remercie un des relecteurs anonymes pour cet exemple précis et pertinent.

6. La question de la qualité des études peut être partiellement éludée en tenant compte du poids scientifique des revues. Par exemple, Doucouliagos et Laroche (2003) ont pondéré les résultats de chaque étude incorporée dans leur méta-analyse par le facteur d'impact du Social Science Citation Index (SSCI) de la revue dans laquelle l'étude avait été publiée. Le facteur d'impact d'une revue (mais cela peut être aussi un classement réalisé par une institution comme le CNRS) peut ainsi être utilisé pour estimer la qualité scientifique d'une recherche.

7. Il convient aussi d'être vigilant sur le caractère indépendant des études collectées. En effet, il n'est pas rare que des chercheurs publient plusieurs études à partir des mêmes données. Le méta-analyste doit alors s'assurer de l'indépendance des études en retirant de son échantillon les études concernées.

élevées de t de Student (si l'on opte pour ce test statistique). Par construction, la statistique $|(\text{Effet}_i - \text{TE})/\sigma_i|$, ne doit pas excéder 1,96 dans plus de 5 % des cas (Stanley, 2005). TE représente la grandeur d'effet réelle dans la population (*True Effect*) et peut être estimée par différents tests, après avoir corrigé la publication sélective (voir ci-après).

La figure 3 présente un graphique radiaire de Galbraith réalisé à partir des données collectées par Doucouliagos et Laroche (2003). La représentation radiaire de Galbraith (1988) consiste à représenter chaque étude par un point dont l'abscisse x est égale à l'inverse de l'écart-type de l'estimateur de l'effet issu de l'étude i $(1/\sigma_i)$ et dont l'ordonnée y est l'estimateur de l'effet standardisé par son écart-type $(\text{Effet}_i/\sigma_i)$ [8]. En fait, il s'agit d'un « *funnel plot* » ayant subi une rotation à 90°, ajusté pour éliminer l'hétéroscédasticité. Ce type de représentation est utile pour analyser l'homogénéité des résultats.

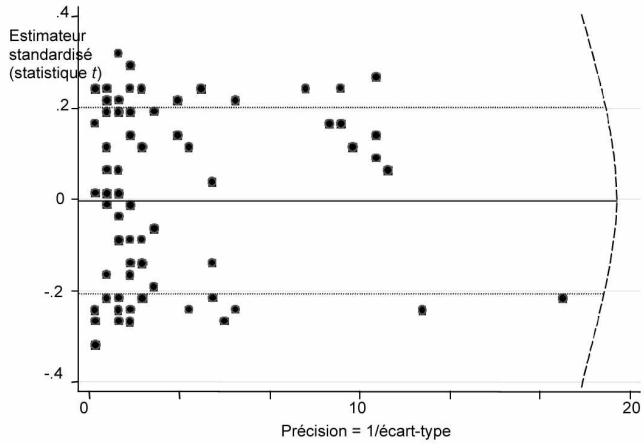


FIG 3. — Représentation radiaire de Galbraith (1988).

L'observation du graphe radiaire indique une répartition très inégale des t de Student autour de zéro (*cf.* Figure 3). En considérant qu'il n'existe pas de lien systématique entre la présence syndicale et la productivité du travail (TE = 0), seules 5 % des études collectées devraient reporter des valeurs de t supérieures à 1,96 en valeur absolue. Or, on constate que 33 études sur 73 – soit 45 % des études – font état de t de Student supérieurs à la valeur critique de t au seuil de risque de 5 % (Stanley, 2005). Un test du khi-deux permet de rejeter l'hypothèse nulle selon laquelle il n'existe pas de publication sélective de type II ($\chi_{(1)} = 248,4$; $p < 0,0001$). En effet, on constate une surreprésentation d'études empiriques qui reportent des valeurs absolues de t élevées ($|(\text{Effect}_i - \text{TE})/\sigma_i| > 1,96$). Au sein de ce domaine de recherche consacré au lien syndicat/productivité du travail, il semble bien qu'il existe un biais de sélection en faveur des études affichant des résultats significatifs (Doucouliagos *et al.*, 2005).

8. Il s'agit bien souvent du t de Student en économétrie puisque $t = (effet_i - TE)/\sigma_i = effet_i/\sigma_i$ en considérant l'hypothèse nulle $H_0 : TE = 0$.

Si l'utilisation de graphes peut permettre, avec beaucoup de précautions, d'identifier les différentes formes de publication sélective, il est souvent nécessaire d'y associer des tests statistiques pour quantifier la probabilité d'existence d'un biais. Parmi les outils statistiques les plus récents, la méta-régression (*MRA – Meta-Regression Analysis*) fait l'objet de nombreuses applications, notamment en économétrie (Stanley et Jarrell, 1989 ; Stanley, 2001).

2. La méta-régression et l'analyse du biais de publication

L'objectif général de la méta-régression est de modéliser, d'estimer et d'analyser les écarts de résultats entre les études empiriques. La méta-régression permet de savoir si le processus de recherche lui-même, à travers les différents choix du chercheur en termes de plan d'échantillonnage, de données, de techniques statistiques, etc, n'est pas à l'origine des écarts observés entre les résultats des études existantes. La méta-régression permet non seulement d'identifier les facteurs susceptibles de faire varier les résultats des études mais également d'identifier et de corriger le biais de publication (Stanley et Jarrell, 1989). L'objet de cette seconde partie est donc de présenter les derniers développements méthodologiques destinés à identifier et à corriger le biais de publication à partir de modèles de (méta-)régression. Tout d'abord, le test d'asymétrie des entonnoirs (ou FAT – *Funnel Asymmetry Test*), qui repose sur un modèle de régression simple, est présenté (2.1.). Ensuite, le test de méta-significativité (ou MST – *Meta-Significance Testing*), dont l'objectif est de déceler l'existence d'une « vraie » relation entre deux variables d'intérêt, est exposé (2.2.). Enfin, après avoir montré la complémentarité du FAT et du MST, la méta-régression explicative, susceptible d'incorporer ces deux tests, est abordée (2.3.).

2.1. Test d'asymétrie et correction du biais de publication

L'utilisation de tests statistiques permet au chercheur d'identifier de manière plus rigoureuse le biais de publication. Les méta-analystes, notamment en économie, ont développé une modélisation de la publication sélective qui s'appuie sur un modèle de (méta-)régression simple entre, d'une part, les estimations des grandeurs d'effet tirées de chaque étude (souvent l'estimation d'une élasticité ou d'un coefficient de régression en économétrie) et, d'autre part, leur écart-type (Card & Krueger, 1995 ; Ashenfelter *et al.*, 1999 ; Görg et Strobl, 2001) :

$$\text{Effet}_i = \beta_i + \beta_0 \sigma_i + \varepsilon_i \qquad (1)$$

En l'absence de biais de publication, les effets observés varient aléatoirement autour de la « vraie » valeur, β_1, indépendamment de l'écart-type. Dans le cas d'une sélection rigoureuse des études, le biais de publication évoluera proportionnellement à l'écart-type, $\beta_0 \sigma_i$ [9]. Toutefois, parce que les études s'appuient

9. Lorsque seuls les résultats statistiquement significatifs sont publiés (échantillon tronqué), Begg & Berlin (1988, pages 431-432) montrent que le biais de publication est proportionnel à l'inverse de la racine carré de la taille de l'échantillon, $n^{-1/2}$, qui est lui même proportionnel à l'écart-type de l'échantillon : « *Theoretical models predict that publication bias is strongly and inversely related to sample size* » (Berlin *et al.*, 1989, page 383).

sur des échantillons de tailles différentes et des modélisations variées, les erreurs ε_i vont présenter une hétéroscédasticité, remettant en question une des hypothèses principales du modèle de régression linéaire. Cependant, par une singularité inhabituelle en économétrie, la variable indépendante, σ_i, est une estimation de l'écart-type de l'échantillon de l'erreur aléatoire ε_i. Avec une mesure de l'hétéroscédasticité aisément disponible, l'estimation par les moindres carrés pondérés est la méthode qui permet d'obtenir des estimations non biaisées (Stanley, 2005). Rappelons que les moindres carrés pondérés divisent l'équation de régression par les écart-types, σ_i [10]. Ce faisant, on obtient :

$$t_i = \text{Effet}_i/\sigma_i = \beta_0 + \beta_1(1/\sigma_i) + e_i \tag{2}$$

où t_i représente la valeur de l'effet standardisé de l'étude i. Soulignons que la constante et le coefficient de régression sont inversés et que la variable indépendante devient l'inverse de sa valeur. L'équation (2) peut maintenant être estimée par la méthode des moindres carrés ordinaires et sert de base au test d'asymétrie des entonnoirs (FAT – *Funnel Asymmetry Test*).

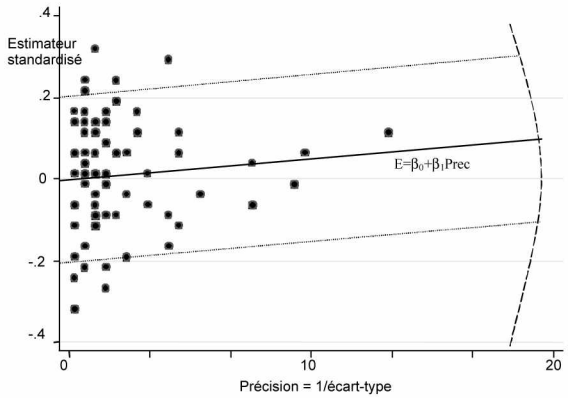
Egger *et al.* (1997) montrent que le test statistique associé à la constante de l'équation (2), β_0, est un test du biais de publication et que son estimation, b_0, indique la direction de ce biais. Ainsi, tester β_0 revient à tester l'asymétrie du *funnel plot* (Sutton *et al.*, 2000a).

En l'absence d'asymétrie du « graphe en entonnoir », les points sur une représentation radiaire de Galbraith (1988) vont se répartir autour d'une ligne qui part de l'origine (qui correspond à β_0) avec une pente β_1 indiquant l'intensité et le sens de l'effet. Cette situation est celle d'un « *funnel plot* » en l'absence de biais (*cf.* figure 4).

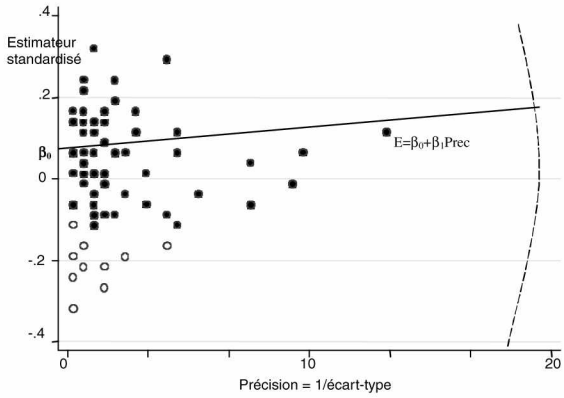
S'il existe une asymétrie dans le « *funnel plot* », l'ordonnée à l'origine de la droite de régression ne passera pas par zéro et la constante de l'équation de la droite de régression va finalement mesurer cette asymétrie. Plus l'ordonnée à l'origine s'écarte de zéro et plus l'asymétrie est importante. Le FAT correspond au test de l'hypothèse nulle ($\beta_0 = 0$, c'est-à-dire absence d'asymétrie) qui est déduit de l'estimation de la régression simple.

Le tableau 1 présente, en guise d'illustration, les résultats des tests d'asymétrie des entonnoirs (FAT) de plusieurs méta-analyses réalisées en économie. La colonne 1 présente, tout d'abord, le test d'asymétrie mené à partir des études empiriques consacrées au lien syndicat/productivité du travail (Union). On observe une faible publication sélective au sein de ce corpus empirique (H_0 : $\beta_0 = 0\,;\, t = 1,72\,;\, p > 0,05$). En revanche, il existerait un biais de publication significatif au sein de la littérature consacrée à l'élasticité-prix de la demande d'eau (Elast) (H_0 : $\beta_0 = 0\,;\, t = -7,27\,;\, p < 0,0001$).

10. Dans les MCG (Moindres Carrés Généralisés), on minimise une somme pondérée de carrés résiduels, $w_i = 1/\sigma_i^2$ jouant le rôle de poids. Dans les MCG, le poids assigné à chaque observation est inversement proportionnel à sa variance, ce qui signifie que les observations provenant d'une population avec un plus fort écart-type auront un poids relativement plus faible et celles issues d'une population avec un plus petit écart-type seront affectés d'un poids plus important lors de la minimisation de la somme des carrés résiduels (SCR). (Gujarati, 2003, page 401).



…sans biais de publication



…avec biais de publication

FIG 4. — Représentations radiaires du FAT.

Le test d'asymétrie des « *funnel plots* » est très utilisé par les méta-analystes bien qu'on lui connaisse plusieurs limites. Sterne *et al.* (2000) ont par exemple montré, à l'aide d'une simulation, qu'il fallait une taille d'échantillon plus grande au FAT pour qu'il ait une puissance supérieure au test de corrélation des rangs de Begg et Mazumdar (1994). Dans la pratique, le test de Begg et Mazumdar est souvent plus utilisé sur de petits échantillons puisque les tests non paramétriques sont habituellement moins puissants que les tests paramétriques à taille d'échantillon égale. Le FAT a manifestement un rapport « puissance/efficience » inférieur au test de corrélation des rangs.

Le test d'asymétrie est donc réputé pour sa faible puissance statistique (Egger *et al.*, 1997) mais il est aussi connu pour être biaisé (Sterne *et al.*, 2000 ; Macaskill *et al.*, 2001). En effet, la variable indépendante, $1/\sigma_i$, intègre une erreur d'échantillonnage parce qu'elle est estimée à partir d'un échantillon représentatif de la population étudiée. Par ailleurs, l'estimation

TABLEAU 1. — Tests d'asymétrie des entonnoirs (FAT – *Funnel Asymmetry Tests).*

	Variable dépendante $= t$			
Variables	Colonne 1 : Union	Colonne 2 : Elast	Colonne 3 : MinWage	Colonne 4 : MinWage
Constante	0,65 (1,72)*	−2,86 (−7,27)***	−2,87 (−2,56)**	−2,01 (−3,49)***
$1/\sigma$	−0,0179 (−1,06)	−0,0817 (−5,34)***	–	0,002 (0,06)
$\sqrt{(ddl)}$	–	–	0,104 (0,74)	–
n	73	110	15	14
R^2	0,024	0,356	0,035	0,0002
Écart-type de la régression	2,13	4,18	0,918	0,969

Source : Stanley (2005, page 28).

par les moindres carrés ordinaires utilise des approximations qui peuvent contenir des erreurs de mesure. Or, les erreurs de mesure posent de sérieux problèmes lorsqu'elles sont présentes dans les variables explicatives parce qu'elles rendent impossible l'estimation de paramètres cohérents. Là aussi, les remèdes économétriques sont bien connus.

Tout d'abord, on peut remplacer $1/\sigma_i$ par la racine carré de la taille de l'échantillon ou par le nombre de degrés de liberté. Rappelons que la théorie statistique lie le biais de publication à la taille de l'échantillon (Begg et Berlin, 1988). Par conséquent, l'utilisation de la racine carré de la taille de l'échantillon dans l'équation (2) peut fournir des estimations plus robustes du biais de publication. En effet, la taille de l'échantillon comme le nombre de degré de liberté ne font pas l'objet d'erreur de mesure. La colonne 3 du tableau 1 présente les résultats du FAT sur les données de Card et Krueger (1995) en utilisant la racine carré du nombre de degré de liberté comme mesure de la précision. Le FAT met en évidence un biais de publication négatif parmi les études consacrées aux effets du salaire minimum ($t = -2,56$; $p < 0,05$). En utilisant $1/\sigma_i$, on confirme la présence d'un biais de publication ($t = -3,49$, $p < 0,001$). Ainsi, le constat de Card et Krueger (1995) selon lequel il existerait un biais de publication dans cette littérature est confirmé par ces différents tests.

Un autre remède proposé par les économètres est d'utiliser des variables instrumentales qui ne sont pas corrélées avec les termes d'erreur d'équation et de mesure (Davidson et MacKinnon, 2004, page 314), proposant une approche alternative au FAT. La racine carrée de la taille de l'échantillon peut être considérée comme un instrument de $1/\sigma_i$. Pour des raisons statistiques, $\sqrt{n}$ est fortement corrélé à $1/\sigma_i$ ($r = 0,93$ pour les données syndicat/productivité, par exemple), tout en étant non corrélé avec l'erreur d'estimation de σ_i et, par conséquent, exogène aux erreurs de la méta-régression, e_i. Partant de ce constat, Stanley (2005) propose un test complémentaire appelé FAIVE-HR (*Funnel Asymmetry Instrumental Variables Estimator – heteroscedasticity-robust*) qui utilise les variances et écart-types corrigés de l'hétéroscédasticité

de White (appelés aussi écart-types robustes) afin d'obtenir des estimations cohérentes (au sens statistique) des variances et covariances des estimateurs MCO même s'il existe de l'hétéroscédasticité [11].

En tout état de cause, Stanley (2005, p. 30) rappelle qu'il faut être précautionneux dans l'utilisation du FAT dans la mesure où il s'agit d'un procédé attaché aux grands échantillons et que les estimateurs obtenus par cette procédure peuvent ne pas être très performants. Ioannidis et Trikalinos (2007) ont d'ailleurs montré très récemment que l'utilisation des tests d'asymétrie des « *funnel graphs* » était souvent inadaptée pour identifier la publication sélective et que leur interprétation était souvent erronée dans les méta-analyses en épidémiologie.

Pour autant, lorsque le biais de publication est convenablement modélisé sous la forme d'une relation systématique entre l'intensité de l'effet observé et son écart-type, il peut facilement être corrigé. Pour neutraliser le biais de publication de type II, un simple aménagement du modèle FAT est suffisant. Rappelons que la publication sélective peut être indifférente au sens de la relation entre deux variables d'intérêt, en privilégiant uniquement la significativité des résultats. Dans le cas d'une publication sélective de type II, l'intensité de l'effet observé va dépendre de σ_i. Ainsi, les équations (1) et (2) vont pouvoir s'écrire :

$$|\text{Effet}_i| = \beta_1 + \beta_0 \sigma_i + v_i \tag{3}$$

$$|t_i| = \beta_0 = +\beta_1 (1/\sigma_i) + u_i \tag{4}$$

L'équation (4) peut maintenant être utilisée pour estimer l'intensité du biais de publication, quel que soit le sens de l'effet observé. Si les erreurs, ε_i, sont normalement distribuées, les erreurs, v_i, associées aux valeurs absolues sont susceptibles de présenter une asymétrie, notamment s'il y a de nombreuses petites valeurs de t. Après avoir estimé l'équation (4), la contraction vers zéro de chaque effet observé en simplifiant par le coefficient de régression (0 permet de corriger le biais de publication ($|t_{\text{corrigé}}| = |t_i| - \beta_0 = \beta_1(1/\sigma_i) + u_i$ ou $t_{\text{corrigé}} = \delta_1(1/\sigma_i) + u_i$). La figure 5 présente les estimations des grandeurs d'effet observés dans chaque étude puis les grandeurs d'effets corrigées du biais de publication.

On remarque que les grandeurs d'effets corrigées sont beaucoup plus symétriques. Par ailleurs, elles présentent de manière plus équilibrée des résultats à la fois positifs et négatifs. Sans ses ajustements, le « *funnel plot* » de l'élasticité-prix de la demande d'eau (figure du bas) serait asymétrique à gauche (privilégiant les résultats négatifs). Toutefois, après élimination du biais de publication, une partie des estimations devient positive et le graphe présente cette fois-ci une forme en entonnoir. Une configuration du même type est perceptible avec les données de la méta-analyse consacrée au lien syndicat/productivité (graphe du haut) mais beaucoup moins visible, compte tenu

11. Pour plus de détails, consulter Davidson et McKinnon (2004, p. 315-335). Il convient d'indiquer que le test FAIVEHR peut s'avérer difficile à utiliser si la taille de l'échantillon n'est pas suffisamment corrélée avec $1/\sigma_i$.

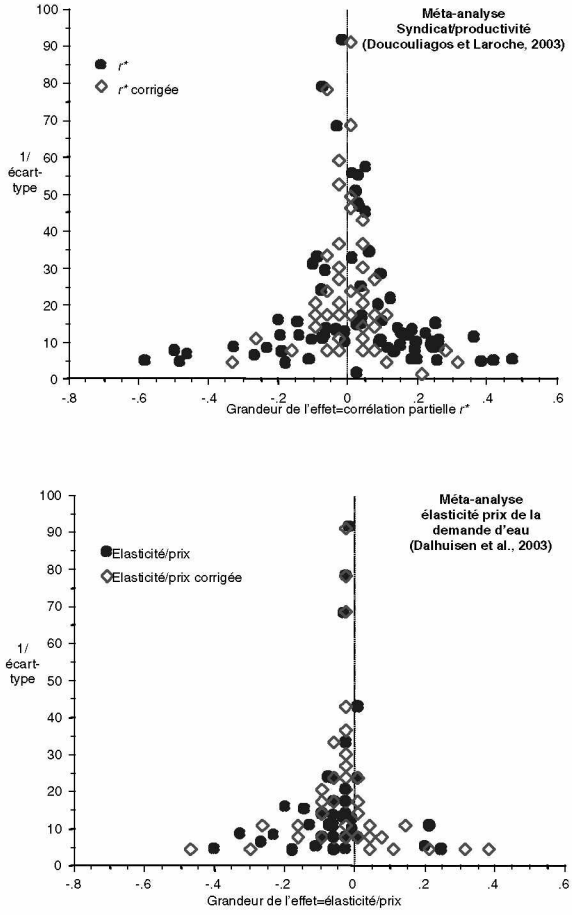


FIG 5. — Funnel Plot avec les corrélations r observés et corrigés du biais de publication.

de la faible asymétrie initiale, que nous avions déjà souligné. Néanmoins, cette procédure permet de corriger le biais de publication de type II qui était très présent au sein de ce corpus d'études (*cf.* figure 3).

Quel que soit le type de biais de publication – type I ou type II –, cette procédure se révèle appropriée pour éliminer la plupart des biais systématiques. Dans la section suivante, une extension de ce test est proposée dont l'objectif est d'identifier l'existence d'un effet «réel» dans la population à partir des estimations des grandeurs d'effet de chaque étude.

2.2. Le test de méta-significativité ou MST (*Meta-Significance Testing*)

Jusqu'ici, nous avons présenté les différentes façons d'identifier et d'éliminer la publication sélective. Toutefois, la question essentielle demeure celle de l'existence d'une « vraie » grandeur d'effet dans la population étudiée, indépendamment du biais de publication. Autrement dit, l'objectif ici est d'examiner si l'effet observé entre deux variables d'intérêt dans plusieurs échantillons permet de faire ressortir un effet « réel » dans la population, tout en contrôlant le biais de publication. Stanley (2001) propose d'utiliser la relation entre l'effet standardisé de chaque étude et le nombre de degré de liberté comme un moyen d'identifier la relation « véritable » entre deux variables d'intérêt. L'idée est simple et fondée sur une propriété bien connue de la puissance statistique (Cohen, 1969) : l'ampleur de l'effet standardisé varie positivement avec la grandeur de l'échantillon [12] dans le seul cas où il existe un effet réel dans la population étudiée [13]. Cette caractéristique de la puissance statistique sert de fondement à la technique du « test de méta-significativité » ou MST (*Meta-Significance Testing*) dont l'objectif est de déceler la présence d'un effet empirique réel entre deux variables d'intérêt (Stanley, 2005). Selon Card et Krueger (1995), la statistique t est liée au degré de liberté (ou à la taille de l'échantillon) de la façon suivante :

$$E(\ln|t_i|) = \alpha_0 + \alpha_1 \ln df_i \tag{5}$$

avec $\alpha_1 = 0$ sous l'hypothèse nulle H_o de non effet et $\alpha_1 = 1/2$ lorsqu'il existe un effet. Ce test de méta-significativité, noté MST, permet d'identifier l'existence d'une vraie grandeur d'effet si l'on rejette l'hypothèse $H_0 : \alpha_1 \leqslant 0$. Lorsqu'il existe un effet réel dans la population, on peut s'attendre à une relation logarithmique entre la valeur absolue de la statistique t de l'étude et son degré de liberté dont le coefficient de régression devrait être précisément 1/2 (*cf.* Greene, 1990 ; Davidson et MacKinnon, 2004).

L'observation d'une relation positive entre le degré de liberté et la grandeur des effets estimés au sein d'un ensemble d'études permet de confirmer l'existence d'une relation réelle entre deux variables d'intérêt. En revanche, si l'on observe une relation négative entre le degré de liberté et la statistique t ($\alpha_1 < 0$), cela signifie que des études portant sur de petits échantillons affichent des statistiques t plus élevées que les études disposant d'échantillons plus larges, laissant penser qu'il existe un biais de sélection.

Le test de méta-significativité peut-il être affecté par la présence d'un biais de publication ? Begg et Berlin (1988, p. 431-432) ont montré que le biais de publication est proportionnel à l'inverse de la racine carrée de la taille de l'échantillon, $n^{-1/2}$. Dès lors, en présence d'un biais de publication, l'effet observé dans la littérature sera inversement lié à la taille de l'échantillon ou

12. En fait, avec sa racine carrée, d'où la valeur 1/2 du coefficient α_1 introduit ci-dessous.

13. L'idée est simple : lorsque la taille de l'échantillon augmente, la précision des estimations s'accroît également (*i.e.* les écarts-types diminuent) et, par conséquent, la statistique t augmente...dans le seul cas où il existe réellement une relation entre deux variables.

au nombre de degrés de liberté. En revanche, en l'absence de publication sélective, l'effet standardisé (la statistique t ici) sera positivement associé au degré de liberté et proportionnel à la racine carrée du degré de liberté (*cf.* équation 5). Parce que les effets observés sont incorporés au numérateur de l'effet standardisé ($t_i = \text{Effet}_i/\sigma_i$), la relation négative entre l'estimateur standardisé et le nombre de degré de liberté, induite par la présence de publication sélective, va atténuer l'association positive entre t et le degré de liberté.

À l'extrême, la publication sélective pourrait éliminer toute trace de puissance statistique (et faire accepter $H_0 : \alpha_1 = 0$). Toutefois, la présence de publication sélective aurait plutôt tendance à réduire α_1 en dessous de 1/2 tout en restant supérieur à zéro lorsqu'il existe un effet « réel » (Stanley, 2005). Ainsi, une valeur de α_1 comprise entre 0 et 0,5 indique qu'il existe une relation réelle entre les deux variables d'intérêt ainsi qu'un biais de publication.

Le tableau 2 présente les résultats d'une série de MST afin d'illustrer l'utilisation de ce type d'outil statistique. En premier lieu, on observe l'absence d'effet empirique « réel » au sein de la littérature consacrée au lien syndicat/productivité (Union / acceptation de $H_0 : \alpha_1 \leqslant 0$; $t = -0,006$; $p >> 0,05$). De même, les résultats présentés dans la colonne 3, n'indiquent pas de lien « réel » entre le salaire minimum légal et l'emploi (MinWage / acceptation de $H_0 : \alpha_1 \leqslant 0$; $t = -1,01$; $p > 0,05$). En revanche, les résultats obtenus sur l'échantillon d'études consacrées au lien élasticité-prix de la demande d'eau révèlent un effet empirique réel (Elast / rejet de $H_0 : \alpha_1 \leqslant 0$; $t = 4,36$; $p < 0,0001$). Une analyse plus fine des résultats de la colonne 2 permet de révéler le sens de la publication sélective. Si l'élasticité-prix est négative alors $\alpha_1 = 1/2$. Dans le cas de l'élasticité-prix, on peut rejeter l'hypothèse nulle $H_0 : \alpha_1 = 1/2$ ($t = -7,93$; $p < 0,0001$) : les résultats obtenus font donc état d'une élasticité-prix négative de la demande d'eau qui est atténuée par un biais de publication ($0 < a_1 = 0,177 < 0,5$; a_1 étant l'estimation de α_1).

TABLEAU 2. — Tests de Méta-Significativité (MST – Meta-Significance Testing).

	Variable dépendante = $\ln\|t\|$		
Variables	Colonne 1 : Union	Colonne 2 : Elast	Colonne 3 : MinWage
Constante	0,39 (1,27)	0,35 (1,61)	2,03 (1,39)
$\ln(n)$	−0,0003 (−0,006)	0,177 (4,36)***	–
$\ln(df)$	–	–	−0,40 (−1,01)
n	73	110	15
R^2	$3,6 \times 10^{-7}$	0,144	0,093
Écart-type de la régression	0,83	0,791	0,510

Source : Stanley (2005, page 36).

En définitive, le test de méta-significativité met en exergue un effet « réel » au sein d'un seul corpus empirique : l'élasticité-prix de la demande d'eau. Toutefois, peut-on affirmer qu'il n'existe pas de lien entre la présence syndicale et la productivité, d'une part, et entre le salaire minimum et l'emploi, d'autre part ?

En fait, il convient d'utiliser d'autres tests afin de renforcer les résultats des tests de méta-significativité (MST). Rappelons ici que, dans les équations (1) et (2), β_1 peut être considéré comme un effet « corrigé » (Sutton *et al.*, 2000 ; Macaskill *et al.*, 2001). Selon ces modèles de publication sélective, lorsque la taille de l'échantillon tend vers l'infini (ou lorsque σ tend vers zéro), l'effet observé tend vers β_1. Ainsi, le modèle de méta-régression simple utilisé pour tester l'asymétrie de l'entonnoir (FAT) contient également une estimation de l'effet empirique « réel » (*cf.* tableau 1). Dans le tableau 1, le coefficient de régression de la précision ($1/\sigma_i$), c'est-à-dire β_1, est une estimation de l'effet réel après correction du biais de publication. Partant de cette constatation, Stanley (2005) propose un test de l'hypothèse nulle $H_0 : \beta_1 = 0$, qu'il appelle un test de la précision de l'effet (PET – *Precision Effect Test*) et qui a pour objet de confirmer ou d'infirmer la présence d'un effet empirique réel.

Les résultats du FAT présentés dans le tableau 1 sont conformes à ceux obtenus avec les tests MST. Pour la littérature consacrée au lien syndicat/productivité, l'estimation de β_1 est égale à $-0,018$ et nous acceptons l'hypothèse nulle selon laquelle il n'y a pas d'effet « réel » entre la présence syndicale et la productivité du travail (acceptation de $H_0 : \beta_1 = 0 ; t = -1,06 ; p > 0,05$) (*cf.* colonne 1 du tableau 1). De la même manière, il n'est pas possible de prouver l'existence d'un lien entre le niveau d'emploi et le salaire minimum légal (acceptation de $H_0 : \beta_1 = 0 ; t = 0,06 ; p > 0,05$). En revanche, seules les études consacrées à l'élasticité-prix de la demande d'eau permettent d'identifier un effet « réel » dans la population (rejet de $H_0 : \beta_1 = 0 ; t = -5,34 ; p < 0,0001$).

Par ailleurs, une troisième approche corrobore les résultats des deux tests précédents (MST et PET). Nous avons vu précédemment comment éliminer le biais de publication (équations 3 et 4). L'obtention des effets corrigés peut servir de base à un autre test de l'effet empirique « réel ». Le tableau 3 présente les résultats de trois méta-régressions de l'effet corrigé ($T_i$corrigé $= \delta_i(1/\sigma_i) + u_i$). On remarque que les effets estimés, corrigés du biais de publication, sont très proches de ceux observés auparavant (*cf.* tableau 1), à la différence près que les écarts-types des estimations sont beaucoup moins importants. Pour les études consacrées au lien syndicat /productivité (colonne 1, tableau 3) l'effet empirique est maintenant existant ($\beta_1 = -0,0093^{**}$; $p < 0,05$) et laisse apparaître un lien négatif (très faible) entre la présence syndicale et la productivité du travail [14].

Cette méthode d'identification et de correction du biais de publication est si puissante que le coefficient de régression β_1 peut s'avérer statistiquement

14. L'estimateur de l'effet commun est ici un coefficient de corrélation partielle calculée à partir des informations tirées de chaque étude. Cette corrélation s'obtient à partir du ratio t utilisé pour tester l'hypothèse selon laquelle le coefficient associé à la variable explicative x (ici la présence syndicale) dans la régression multiple est nul : $r^*_{yx} = \sqrt{t_x^2/t_x^2 + ddl}$.

TABLEAU 3. — Tests avec effets filtrés *(Filtered Effects Tests)*.

	Variable dépendante = t corrigé		
Variables	*Colonne 1 : Union*	*Colonne 2 : Elast*	*Colonne 3 : MinWage*
$1/\sigma$	*−0,0093 (−1,99)***	*−0,083 (−10,73)****	*0,0018 (0,13)*
n	*73*	*110*	*14*
R^2	*0,052*	*0,516*	*0,001*
Écart-type de la régression	*1,07*	*3,61*	*0,931*

Source : Stanley (2005, page 37).

significatif même si sa valeur est très faible. Cependant, dans le cas présent, l'interprétation des résultats obtenus n'est pas affectée par l'amélioration de cette précision. Bien que le coefficient β_1 soit négatif et statistiquement significatif, la valeur observée n'a aucune signification économique et il est difficile d'affirmer l'existence d'une relation entre la présence syndicale et la productivité du travail. Néanmoins, il est important de constater combien une analyse de méta-régression simple peut éclairer les chercheurs sur les liens statistiques pouvant exister entre deux variables d'intérêt.

TABLEAU 4. — Tests du biais de publication et de l'effet empirique réel.

Test	**Modèle MRA**	**H_1 et ses implications**
Funnel asymetry Precision-effect (FAT)	$t_i = \beta_0 + \beta_1(1/\sigma_i) + \varepsilon_i$	$\beta_0 \neq 0$ Biais de publication $\beta_1 \neq 0$ Effet empirique réel
Publication bias filtered-effect	$t_i = \beta_0 + \beta_1(1/\sigma_i) + v_i$ $t_{i_{\text{corrigé}}} = \delta_1(1/\sigma_i) + u_i$	$\beta_0 > 0$ Biais de publication $\delta_1 \neq 0$ Effet empirique réel
Meta-significance (MST)	$\ln\lvert t_i\rvert = \alpha_0 + \alpha_1 \ln df_i + v_i$	$\alpha_i > 0$ Effet empirique réel $0 < \alpha_i < 0,5$ Biais de publication + effet empirique réel
Joint precision-effect (FAT)/Meta-significance (MST)	$t_i = \beta_0 + \beta_1(1/\sigma_i) + \varepsilon_i$ $\ln\lvert t_i\rvert = \alpha_0 + \alpha_1 \ln df_i + v_i$	$\beta_1 \neq 0$ et $\alpha_1 > 0$ Effet empirique réel $0 < \alpha_1 < 0,5$ Biais de publication + effet empirique réel

Source : Stanley (2005, page 38).

Le tableau 4 présente une synthèse des différents tests statistiques du biais de publication et de la grandeur d'effet. En définitive, la complexité du processus de publication sélective nécessite d'utiliser plusieurs tests pour s'assurer de la robustesse des résultats. Stanley (2005) propose d'utiliser conjointement le

test d'asymétrie des entonnoirs et son test de précision (FAT-PET) ainsi que le test de méta-significativité (MST).

2.3. La méta-analyse de régression explicative (*MRA – Meta-Regression Analysis*)

Les tests d'asymétrie des « funnel graphs » (FAT) et de méta-significativité (MST) peuvent être révélateurs de publication sélective malgré leurs limites respectives. Cependant, les travaux de recherche en sciences sociales font souvent état de la complexité des relations pouvant exister entre des variables d'intérêt. Il est donc difficile d'expliquer l'évolution d'une variable à l'aide de modèles de régression simple, c'est-à-dire en n'examinant qu'une seule variable explicative à la fois. Dès lors, l'omission de certaines variables dans les modèles de méta-régression peut conduire à des résultats difficilement exploitables car totalement biaisés par l'absence de ces variables. Les revues de la littérature, qu'elles soient narratives ou méta-analytiques, mettent souvent en avant le caractère contingent et contradictoire des résultats observés par les études empiriques. Les statistiques issues des études empiriques sont souvent affectées par des biais de sélection et de mauvaise spécification avant d'être éventuellement déformées par un biais de publication. Dès lors, il s'avère souvent plus judicieux d'incorporer les tests du biais de publication dans des modèles de (méta-)régression multiple qui tiennent compte également des autres facteurs explicatifs de la variation des résultats des études collectées.

L'objectif de cette modélisation est d'examiner l'effet simultané de plusieurs variables modératrices sur la grandeur de l'effet. Il s'agit ici d'estimer des modèles de régression multiple qui ont les formes suivantes :

$$t = \beta_0 + \beta_1(1/\sigma) + \delta_1 K_1 + \gamma_1 X_1 + \ldots + \gamma_k X_k + u \tag{6}$$

$$\ln|t| = \alpha_0 + \alpha_1 \ln df_i + \delta_i K_1 + \gamma_1 X_1 + \ldots + \gamma_k X_k + v \tag{7}$$

avec β_0 et α_0 qui sont les constantes, $X_1 ... X_k$ sont les variables indicatrices de modalités représentant certaines caractéristiques associées à l'étude i ; K est une variable quantitative et u et v sont les perturbations aléatoires.

À titre illustratif, le tableau 5 présente deux modèles de méta-régression explicative pour illustrer cette approche qui consiste à incorporer des tests FAT ou MST dans des modèles de régression multiple. Il s'agit ici de deux méta-régressions, réalisées à partir des données de Doucouliagos et Laroche (2003), dont l'objectif est d'examiner l'influence de la nature des données utilisées [15] sur les résultats des études consacrées au lien syndicat/productivité.

Les résultats du FAT-MRA présentés dans le tableau 5 indiquent que les études menées sur des données américaines font ressortir plus souvent un lien positif entre la présence syndicale et la productivité du travail ($t = 2,08$, $p < 0,05$). Cependant, la valeur du coefficient de $1/\sigma$, supposée offrir une estimation de l'effet empirique au sein de la population des études, est non

15. Il s'agit ici d'une indicatrice codée 1 s'il s'agit de données américaines et 0 s'il s'agit de données d'autres pays industrialisés (variable US dans le tableau 5).

TABLEAU 5. — FAT-MRA et MST-MRA

	Variables dépendantes	
Variables explicatives	t FAT	$\ln\|t\|$ MST
Constante	−.21 (−.40)	−.74 (−1.26)
ln(df)	—	.20 (2.38)**
$1/\sigma$	−.020 (−1.01)	—
US	1.22 (2.08)**	1.46 (2.18)**
N	73	73
R^2	.061	.035
Ecart-type de la régression	2.08	.82

significative ($t = -1,01$, $p >> 0,10$), laissant entendre qu'il n'existe pas de relation entre la présence syndicale et la productivité du travail. S'agissant du biais de publication, les résultats obtenus montrent qu'il n'existe pas de biais de publication ($t = -0,40$, $p > 0,10$) au sein de cette littérature. Le modèle MST-MRA, quant à lui, confirme le lien positif entre la présence syndicale et la productivité aux USA. Toutefois, les résultats obtenus semblent indiquer à la fois l'existence d'une relation réelle entre les deux variables d'intérêt et la présence d'un biais de publication ($0 < a_1 = 0,20 < 0,50$; a_1 étant l'estimation de α_1). Etant donné que la variable dépendante correspond à la valeur absolue de la statistique t, il n'est pas possible d'identifier le sens de la relation entre les deux variables. Toutefois, dans d'autres circonstances, il est possible d'établir le sens de la relation en examinant le signe du coefficient de $1/\sigma$ du FAT. Ici, il se trouve que le coefficient est non significatif. Dans l'ensemble, il convient d'interpréter ces résultats avec précaution puisqu'ils s'avèrent contradictoires. Une série d'autres tests serait nécessaire pour confirmer ou infirmer l'existence d'un biais de publication et pour établir la nature exacte de la relation entre les deux variables.

Cet exemple, volontairement simplifié, montre combien le choix des variables explicatives revêt une importance considérable si l'on souhaite obtenir des résultats suffisamment robustes. Les informations existantes sur l'influence (connue ou supposée) de telle ou telle caractéristiques constituent un matériau précieux pour mener à bien le choix des variables explicatives à retenir. En fait, l'approche méta-analytique soulève un certain nombre de difficultés, liées notamment au codage des caractéristiques des études. Sur cet exemple précis, les travaux de Doucouliagos et Laroche (2003) et ceux de Doucouliagos *et al.* (2005) tiennent compte de nombreuses différences qui caractérisent le contexte spécifique de chaque recherche empirique. Remarquons que la multiplication des variables explicatives peut aussi constituer une impasse, ne serait-ce que dans la mesure où elle renvoie à des problèmes de respect des hypothèses fondamentales de la régression. En outre, la volonté de décrire de façon toujours plus précise les contextes des études risque de conduire

assez rapidement à des problèmes d'informations manquantes et *in fine* à la diminution du nombre d'études méta-analysées.

En définitive, la méta-régression apporte des réponses, au moins partielles, aux principales interrogations critiques suscitées par l'existence de différentes formes de publication sélective. La réalisation d'une méta-analyse de régression soulève cependant aussi des difficultés dont la méconnaissance est de nature à annihiler tous ses apports potentiels. En considérant les effets de plusieurs caractéristiques des études, l'utilisation des modèles de méta-régression se révèle néanmoins bien adaptée pour traiter le problème de la publication sélective.

Il existe d'autres approches qui apportent une aide à l'identification et au traitement du biais de publication (Dear et Begg, 1992). Par exemple, Begg et Mazumdar (1994) proposent d'utiliser la corrélation des rangs entre les effets estimés et leurs variances pour tester l'existence d'un biais de publication. Il s'agit en fait d'une version non paramétrique du test d'asymétrie du « *funnel plot* ». On peut évoquer aussi les travaux bayésiens utilisant des techniques d'augmentation de données (Givens *et al.*, 1997) et, en particulier, la méthode non paramétrique dite « *Trim and Fill* » de Duval et Tweedie (2000). Le principe de cette méthode est relativement simple : quand un « funnel plot » présente une asymétrie, les études asymétriques sont gommées (*Trim*). Le « *funnel plot* » est ensuite à nouveau complété (*Fill*) avec les études gommées et avec leur image en miroir. Le « *funnel plot* » devenu symétrique peut fournir une nouvelle moyenne. Un intervalle de confiance corrigé peut ainsi être calculé sur la base du « *filled funnel plot* ».

Si cette technique permet de corriger artificiellement le biais de publication, elle ne fait pas l'unanimité parmi les méta-analystes. Certains chercheurs sont opposés à l'idée d'introduire des études « fictives » dans la méta-analyse (Begg, 1994). Pour les méta-analystes employant cette technique, il s'agit essentiellement d'évaluer la sensibilité des résultats de la méta-analyse en incorporant fictivement les études manquantes. Dans la pratique, ce sont souvent les chercheurs en médecine qui utilisent cette technique. En sciences sociales, les méta-analystes préfèrent calculer le nombre d'études dans le tiroir (*Fail-Safe N* ou *File-Drawer Numbers*) pour s'assurer de la robustesse des résultats obtenus par la méta-analyse.

3. L'investigation des études manquantes dans la méta-analyse : le calcul du nombre d'études dans le tiroir

Le calcul du nombre d'études dans le tiroir (*file-drawer number*) est un procédé astucieux suggéré par Rosenthal (1979). Il s'agit d'un des dispositifs les plus anciens permettant d'évaluer la robustesse des résultats d'une méta-analyse et, en sciences sociales, certainement celui qui est le plus répandu. Becker et Morton (2002) ont d'ailleurs montré que plus d'un tiers des 28 méta-analyses publiées en psychologie et en sciences de l'éducation entre 1999 et 2002 présente le calcul du nombre d'études dans le tiroir (« N additionnel »).

Comme nous l'avons déjà évoqué, plusieurs études ont montré que les chercheurs pouvaient être amenés à ne pas publier leurs recherches si celles-ci présentaient des résultats non significatifs et à les laisser « au fond d'un tiroir » (d'où l'expression de Rosenthal). Si tel est le cas, alors les études publiées ne sont pas représentatives de l'ensemble des études qui ont été menées et les conclusions tirées de leur synthèse peuvent être fallacieuses. Rosenthal pense que s'il existe un grand nombre de résultats non significatifs (non publiés) et qu'on les introduit dans la méta-analyse, les conclusions peuvent changer. Par conséquent, Rosenthal se demande combien d'études supplémentaires faudrait-il introduire dans la méta-analyse pour que le seuil de significativité statistique ne soit pas atteint.

Avant d'aller plus loin, il convient d'être plus explicite sur la méthode de Rosenthal et les hypothèses sous-jacentes permettant de calculer le nombre d'études additionnelles (fail-safe N). Tout d'abord, supposons une série de k études indépendantes et une étude i qui teste l'hypothèse nulle $\mathrm{H}_0 : \theta_i = 0$. Le paramètre θ_i représente une corrélation entre deux variables d'intérêt ou une différence de moyenne standardisée entre deux groupes, ou un rapport de côte (odds ratio). Chaque étude fournit également une probabilité observée (unilatérale) p_i issue d'un test de l'hypothèse nulle unidirectionnelle selon lequel $\theta_i \leqslant 0$ (*i.e.* l'hypothèse alternative est $\theta_i > 0$).

Rosenthal commence par calculer l'effet standardisé qui correspond à un *z-score* (appelé aussi variable standardisée) dont Stouffer *et al.* (1949) ont proposé le mode de calcul (« sum of Z_s ») :

$$Z_s = \frac{\sum_{i=1}^{k} z_i}{\sqrt{k}}$$

où z_i est la valeur de la distribution normale associée à la probabilité p_i de l'étude i. L'hypothèse nulle pour cette série d'études correspond à l'ensemble des k valeurs de θ_i égales à zéro. Si l'hypothèse nulle est vraie, Z_s est la valeur de la distribution normale standardisée et le test de l'hypothèse $\mathrm{H}_0 : \theta_i = ... = \theta_k = 0$ est réalisé en comparant Z_s à une table des valeurs de la distribution normale.

Par exemple, Z_s est significatif au seuil α de 5 % s'il est supérieur à $Z_\alpha = 1,645$. Rosenthal se demande alors, pour une valeur observée de Z_s déjà supérieure à Z_α, combien d'études avec des valeurs Z_i, en moyenne, égale à zéro (*i.e.* avec $p_i = 0,5$) doivent être introduites pour ramener la valeur Z_s à Z_α. Il se demande finalement quelle valeur de N pourrait satisfaire la formule suivante :

$$\frac{\sum_{i=1}^{k} Z_i}{\sqrt{k+N}} < Z_\alpha$$

avec Z_α, valeur critique de la distribution normale au seuil α. La formule peut se traduire sous une forme différente afin de sortir N. Une des formules

proposée se fonde sur le test Z_s de Stouffer :

$$N > k \left[\frac{Z_s}{Z_\alpha} \right]^2 - k$$

En substituant $z_s = \dfrac{\sum z_i}{\sqrt{k}}$ dans la formule précédente, on obtient une forme alternative impliquant la somme des Z_i, c'est-à-dire :

$$N > \left[\frac{\sum Z_i}{Z_\alpha} \right]^2 - k$$

D'un point de vue arithmétique, l'addition de N études avec une moyenne de $Z_i = 0$ est équivalent à l'addition de N études ayant chacune un $Z_i = 0$. Toutefois, ces deux ensembles d'études peuvent être très différents dans la réalité. En effet, un ensemble d'effets dont la moyenne est égale à zéro peut afficher en réalité une variance très élevée. Il est donc peu probable que, dans la pratique, l'on puisse observer un ensemble d'études avec des résultats nuls identiques.

Le calcul du N *additionnel* repose donc sur plusieurs hypothèses discutables. Tout d'abord, les études omises ou non publiées présentent des résultats nuls, en moyenne. Plusieurs chercheurs se sont interrogés sur le bien fondé de cette hypothèse alors qu'il est possible de trouver des études dans le tiroir qui affichent des résultats très différents de zéro et statistiquement significatifs. Begg et Berlin (1988), par exemple, montrent que le fait de fixer à zéro l'effet moyen des études non publiées est susceptible de biaiser l'analyse. Il est intéressant de se demander ce qui se passerait si l'on ajoutait des études qui obtiennent, en moyenne, des résultats contraires à ceux des études de la méta-analyse. Il faudrait alors beaucoup moins d'études pour rendre non significatif le Z_s. En définitive, ajouter exactement le même nombre d'études (k) avec des z-score de signe opposé à ceux observés conduirait à réduire considérablement la valeur du $Z_s = 0$ (avec $p = 0,5$). De même, il suffirait de quelques études présentant des résultats contradictoires avec ceux observés pour ramener en dessous de Z_α la valeur de Z_s. Ensuite, Sutton *et al.* (2000a) précisent que le N additionnel ne tient pas compte directement de la taille de l'échantillon de chaque étude. Le fait d'ajouter N études (d'effet moyen égal à zéro) de 5 observations est équivalent à l'introduction de N études de 5000 observations. Enfin, la méthode proposée par Rosenthal surestime de beaucoup le nombre d'études non publiées et ne tient pas compte de la qualité scientifique des études. Comme on peut le constater, le calcul du nombre d'études dans le tiroir repose sur des hypothèses extrêmement fortes qui rendent difficile son utilisation par les méta-analystes. Certains vont jusqu'à dire que cette technique est inutilisable (Scargle, 2000). D'autres ont proposé un certain nombre d'aménagements à la procédure de Rosenthal afin de combler les lacunes existantes (Rosenberg, 2005). Parmi les méthodes récentes, celle proposée par Orwin (1983) s'appuie sur la différence

des moyennes standardisées des grandeurs d'effet. Plutôt que de fixer un seuil de significativité précis, cette méthode permet de calculer le nombre d'études additionnelles, N_0, qu'il faudrait introduire pour réduire la moyenne observée des grandeurs d'effet à une valeur minimum fixée par le méta-analyste.

Conclusion

Cet article avait pour objectif de présenter les différents outils statistiques permettant d'identifier et de corriger le biais de publication. Nous avons vu que les études en sciences sociales comme en sciences médicales sont souvent entachées par ce type de biais, lié à la publication sélective des travaux de recherche. Des investigations récentes ont montré l'omniprésence de la publication sélective en économie (Stanley et Roberts, 2005) et en médecine (Rothstein *et al.*, 2005). L'existence de biais dans le processus de sélection des travaux scientifiques n'est guère surprenant. Les discussions sur les politiques éditoriales des revues académiques et sur la qualité des relectures sont largement répandues au sein de la communauté scientifique (Abramowitz *et al.*, 1992; Shepherd, 1995). En fait, la «science» n'est en rien un objet bien circonscrit et stable dans le temps mais plutôt un ensemble de réalités institutionnelles de production, de valeurs et de normes qui s'inscrit dans un contexte particulier. Bruno Latour (2001) a bien montré comment se déroule le processus qui conduit à la publication d'un travail de recherche. Selon Latour (2001, page 50), les chercheurs font de multiples concessions qui construisent un compromis : «*si j'écris ceci, Dupont va dire que...mais si je mets cela, l'objection tombe; ici citons Durand, ça calmera les critiques, etc. Le résultat final est quelque chose d'extraordinairement stratégique, sain et intéressant (...) qui se rapproche du compromis*». Par ailleurs, la dynamique des connaissances est le résultat de tension toujours renouvelée entre des discours contradictoires et des écoles de pensée en opposition à l'intérieur du monde scientifique. En sciences sociales, la difficulté réside dans le fait qu'à la controverse scientifique se mêlent souvent des considérations au mieux épistémologiques au pire idéologiques. En définitive, le biais de publication résulte aussi du fait que certains chercheurs, grâce à la force de leurs alliances, ont pu se constituer des réseaux d'influence leur permettant de diffuser plus facilement leur contribution que d'autres. Comme le rappelle à juste titre Latour (2001, page 44), si la validation de votre travail – par les cher(e)s collègues – n'a pas lieu «*le statut de votre énoncé reste comme suspendu dans l'antichambre de la science, entre la fiction et la reconnaissance*». L'auteur montre finalement comment les scientifiques sont engagés dans des réseaux socio-techniques qui se jouent des frontières établies entre science et politique.

La méta-analyse ne serait-elle donc qu'un mythe supplémentaire destiné à protéger l'activité scientifique de toute pollution par les idéologies, les intérêts, les passions et les modes? Loin de là notre idée en présentant les différents outils statistiques permettant d'examiner le biais de publication. Les procédures méta-analytiques offrent aujourd'hui une approche plus rigoureuse de synthèse de la littérature. Incontestablement, la méta-analyse offre de multiples avantages par rapport aux revues narratives traditionnelles (Laroche

et Schmidt, 2004). Il s'agit d'un formidable outil de présentation globale des résultats de plusieurs études et, à ce titre, elle mérite d'être davantage utilisée par les chercheurs. Cependant, bien qu'il s'agisse d'un instrument puissant, la méta-analyse n'est pas pour autant plus propre, plus objective que les autres méthodes de recherche si elle n'est pas manipulée avec précaution et avec rigueur (Thompson et Pocock, 1991). Les perspectives d'exploitation de la méta-analyse ne doivent pas conduire à abuser de ces techniques qui souffrent encore de nombreuses limites (Muller, 1988). De la même façon, il ne faut pas non plus surestimer l'influence du biais de publication et le problème de la non-inclusion des études non publiées. D'autres éléments peuvent rendre difficile l'intégration des données et expliquer les différences de résultats obtenus entre les études empiriques. Le problème de l'intégration d'études très différentes sur le plan méthodologique n'est d'ailleurs pas entièrement résolu et il convient de contrôler davantage les différentes sources de variation afin d'approcher la valeur vraie de la relation étudiée. En définitive, bien que ces outils soient particulièrement intéressants, ces techniques statistiques sont encore en voie d'affinement et doivent être utilisées avec précaution. D'aucun diront que « *l'analyse statistique des données est une aide à la réflexion et non une alternative* » et « *doit permettre d'améliorer plutôt que de remplacer une discussion intelligente des résultats critiques* » (Green et Hall, 1984, page 41).

Bibliographie

ABRAMOVITZ M. *et al.* (1992). « A plea for pluralistic and rigorous economics », *American Economic Review*, 82, 25.

ASHENFELTER O., HARMON C. et H. OOSTERBEEK. (1999). « A review of estimates of the schooling/earnings relationship, with tests for publication bias », *Labour Economics*, 6, 453-470.

BAUCHAU V. (1997), « Is there a file drawer problem" in biological research ? », *Oikos*, 79, 407-409.

BEGG C.B. (1994). « Publication bias », in *The Handbook of Research Synthesis*, Russel Sage Foundation.

BEGG C. B. et J. A. BERLIN (1988), « Publication Bias : a problem in interpreting medical data », *Journal of Royal Statistical Society*, Series A, 151(3), 419-463.

BEGG C.B. et MAZUMDAR M. (1994). « Operating characteristics of a rank correlation test for publication bias », *Biometrics*, 50, 1088-1101.

BECKER B.J. et MORTON S.C. (2002). « *Publication bias : Methods and practices in evidence-based medicine and social science* », Mathematical Sciences Research Institute, Berkeley, California.

BERLIN J.A., BEGG C.B. et LOUIS T.A. (1989). « An assessment of publication bias using a sample of published clinical trials », *Journal of the American Statistical Association*, 84, 381-392.

CARD D. et KRUEGER A.B. (1995). « Time-Series Minimum-Wage Studies : A Meta-Analysis », *American Economic Review*, 85(2) :238-43.

COHEN J. (1969). « *Statistical power analysis for the behavioral sciences* », (1st edition), Hillsdale, NJ : Erlbaum.

COOPER H.M. (1998). « *Synthesizing research : a guide for literature reviews* », Thousand Oaks, CA : Sage.

COURSOL A. et WAGNER E.E. (1986). « Effect of positive findings on submission and acceptance rates : a note on meta-analysis bias », *Professional Psychological Research & Practice*, 17, 447-452.

CUCHERAT M. (1997). « *Méta-analyse des essais thérapeutiques* », Paris, Masson.

DALHUISEN J., FLORAX R.J.G.M., DEGROOT H.L.F. et NIJKAMP P. (2003). « Price and income elasticities of residential water demand : a meta-analysis », *Land Economics*, 79, 292-308.

DAVIDSON R. et MACKINNON J.G. (2004). « *Econometric theory and methods* », Oxford : Oxford University Press.

DEAR K.B.G. et BEGG C.B. (1992). « An approach for assessing for publication bias prior to performing a meta-analysis », *Statistical Science*, vol.7, n°2, 237-245.

DICKERSIN K., CHAN S., CHALMERS T.C. SACKS H.S. et SMITH H. Jr. (1987). « Publication bias and clinical trials », *Controlled Clinical Trials*, 8, 343-353.

DICKERSIN K., MIN Y.I., et MEINERT C.L. (1992). « Factors influencing publication of research results. Follow-up of applications submitted to two institutional review boards », *Journal of the American Medical Association*, 267, 374-378.

DOUCOULIAGOS C. et LAROCHE P. (2003). « What do Unions do to Productivity : A Meta-Analysis », *Industrial Relations*, 42(4), 650-692.

DOUCOULIAGOS C., LAROCHE P. et STANLEY T.D. (2005). « Publication Bias in Union-Productivity Research », *Relations Industrielles/Industrial Relations*, Vol. 60, n°2, 320-347.

DUVAL S.J. et TWEEDIE R.L. (2000). « A non-parametric 'trim and fill' method of accounting for publication bias in meta-analysis », *Journal of the American Statistical Association*, 95, 89-98.

EGGER M., SMITH G.D., SCHEIDER M., et MINDER C. (1997). « Bias in meta-analysis detected by a simple, graphical test », *British Medical Journal*, 316, 629-34.

EGGER M., JUNI P., BARTLETT, C. HOLENSTEIN F. et STERNE J. (2003). « How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study », *Health Technology Assessment*, 7, 1-68.

GALBRAITH J.K. (1988). « A note on graphical presentation of estimated odds ratios from several clinical trials », *Statistics in Medicine*, 7, 177-207.

GERBER A.S., GREEN D.P. et NICKERSON D. (2001). « Testing for Publication Bias in Political Science », *Political Analysis*, 9, 385-392.

GILBODY S.M., SONG F., EASTWOOD A.J. et SUTTON A. (2000). « The causes, consequences and detection of publication bias in psychiatry », *Acta Psychiatrica Scandinavia*, 102, 241-249.

GIVENS G.H., SMITH D.D. et TWEEDIE R.L. (1997). « Publication bias in Meta-analysis : A bayesian data-augmentation approach to account for issues exemplified in the passive smoking debate », *Statistical Science*, 12, 221-250.

GLASS G.V., MCGAW B. et SMITH ML. (1981). « *Meta-analysis in social research* », Beverly Hills, CA : Sage Publication.

GOLDFARB R.S. (1995). « The Economist-as-audience needs a methodology of plausible inference », *Journal of Economic Methodology*, 2(2), 201-22.

GORG H. et E. STROBL (2001). « Multinational Companies and Productivity : A Meta-analysis », *The Economic Journal*, 111(475) :723-39.

GREEN B. et HALL J. (1984). « Quantitative methods for literature review », *Annual Review of Psychology*, 35, 37-53.

GREENE W.H. (1990). « *Econometric Analysis* ». New York : MacMillan.

GUJARATI D.N. (2003). « *Basic Econometrics* », 4ème édition, MacGrawHill/Irwin.

HEDGES L.V. (1992). « Modeling publication selection effects in meta-analysis », *Statistical Science*, 7, 227-236.

HEDGES L.V. et OLKIN I. (1985). « *Statistical methods for meta-analysis* », Boston : Academic Press.

HEDGES L.V. et VEVEA J.L. (1996). « Estimating effect size under publication bias : small sample properties and robustness of a random effects selection models », *Journal of Educational and Behavioral Statistics*, 21, 299-332.

HUNTER J.E. et SCHMIDT F.L. (1990). « *Methods of Meta-Analysis : correcting error and bias in research findings* », Sage publications.

IOANNIDIS J.P.A. et TRIKALINOS T.A., (2007). « The appropriateness of asymmetry tests for publication bias in meta-analyses : a large survey », *Canadian Medical Association Journal*, April, 1091-1096.

JARRELL S. et T.D. STANLEY (1990). « A Meta-Analysis of the Union-Nonunion Wage Gap », *Industrial and Labor Relations Review*, 44, 54-67.

LAROCHE P. et SCHMIDT G. (2004). « *La méta-analyse en sciences de gestion : utilités, utilisations et débats* », Academy of Management, Division « Méthodes de Recherche », Crossing Frontiers in Quantitative and Qualitative Research Methods, ISEOR, Lyon, 18-20 Mars.

LATOUR B. 2001. « *Le métier de chercheur, regard d'un anthropologue* », Paris, INRA Editions.

LIGHT R.J. et PILLEMER D.B. (1984). « *Summing up : The science of reviewing research* », Cambridge, MA : Harvard University Press.

MACASKILL P., WALTER S.D., et IRWIG L. (2001). « A comparison of methods to detect publication bias in meta-analysis », *Statistics in Medicine*, 20, 641-654.

MAHONEY M.J. (1977). « Publication prejudices : An experimental study of confirmatory bias in the peer review system », *Cognitive Therapy & Research*, 1, 161-175.

MULLER J.L. (1988). « Pour une revue quantitative de la littérature : les méta-analyses », *Psychologie Française*, 33-4, décembre, 295-303.

NEUMARK D. et W. WASCHER (2002). « Meta-Analysis », *Journal of Economic Perspectives*, 16(3), 226-7.

ORWIN R.G. (1983). « A fail-Safe N for effect size in meta-analysis », *Journal of Educational Statistics*, 8, 157-159.

ROBERTS C.J. et STANLEY T.D. (Eds.), « *Meta-regression analysis : Issues of publications bias in economics* », Blackwell Publishing.

ROSENBERG M.S. (2005). « The File-drawer problem revisited : a general weighted method for calculating fail-safe numbers in meta-analysis », *Evolution*, 59 (2), 464-468.

ROSENBERG M.S., D.C. ADAMS et J. GUREVITCH. (2000). « *MetaWin : Statistical Software for Meta-Analysis* », Version 2.0. Sinauer Associates, Sunderland, Massachusetts.

ROSENTHAL R. (1979), « The "file drawer problem" and tolerance for null results », *Psychological Bulletin*, 86 :638-41.

ROSENTHAL R. (1991). « *Meta-analytic procedures for social research* », (ré-édition de l'ouvrage publié initialement en 1984), Newbury Park, CA : Sage.

ROTHSTEIN H.R., SUTTON A.J. et BORENSTEIN M. (2005). « *Publication bias in meta-analysis : prevention, assessment and adjustment* », John Wiley & Sons.

RUST R.T., D.R. LEHMANN et J.U. FARLEY (1990). « Estimating Publication Bias in Meta-Analysis », *Journal of Marketing Research*, 27, 220-226.

SCARGLE J.D. (2000). « Publication Bias : the "file-drawer" problem in scientific inference », *Journal of Scientific Exploration*, 14, 91-106.

SCHULZ K.F., CHALMERS I., HAYES R.J., et ALTMAN D.G. (1995). « Empirical evidence of bias : Dimensions of methodological quality associated with estimates of treatment effects in controlled trials », *Journal of the American Medical Association*, 273, 408-412.

SHEPHERD G.B. (Ed.) (1995). « *Rejected : Leading economists ponder the publication process* », Sun Lakes : Thomas Horton and Daughters.

SIGELMAN L. (1999). « Publication Bias Reconsidered », *Political Analysis*, 8, 201-210.

SMART R.G. (1964). « The importance of negative results in psychological research », *Canadian Psychologist*, 5, 225-232.

SOMMER B. (1987). « The file drawer effect and publication rates in menstrual cycle research », *Psychology of Women Quaterly*, 11, 233-242.

STANLEY T.D. (2001). « Wheat from chaff : Meta-Analysis as quantitative literature review », *The Journal of Economic Perspectives* 15(3), 131-150.

STANLEY T.D. (2005). « Beyond Publication bias », *Journal of Economic Survey*, 19, 309-337.

STANLEY T.D. et JARRELL S.B. (1989). « A meta-regression analysis : A quantitative method of literature surveys », *Journal of Economic Surveys*, 3, 54-67.

STANLEY T.D. et ROBERTS C.J. (2005). « *Meta-regression analysis : Issues of publication bias in economics* », Oxford : Blackwell Publishing Ltd.

STANLEY T.D, FLORAX R.J.G.M., et DEGROOT H.L.F. (2003). « It's all about power : Differentiating genuine empirical significance from the artifact of publication bias », Discussion Paper 3, Center for Entrepreneurial Studies, Hendrix College.

STERLING T.D. (1959). « Publication decisions and their possible effects on inferences drawn from tests of significance – or vice versa », *Journal of the American Statistical Association*, 54, 30-34.

STERLING T.D. ROSENBAUM W.L., et WEINKAM J.J. (1995). « Publication decisions revisited : The effect of the outcome of statistical tests on the decision to publish », *American Statistician*, 49, 108-112.

STERNE J.A.C., GAVAGHAN D. et EGGER M. (2000). « Publication and related bias in meta-analysis : Power of statistical tests and prevalence in the literature », *Journal of Clinical Epidemiology*, 53, 1119-1129.

STOUFFER S.A., SUCHMAN E.A., DEVINNEY L.C., STAR A. et WILLIAMS R.M. (1949). « *The American soldier : adjustment during army life* », vol.1, Princeton, NJ : Princeton University Press.

SUTTON A.J., DUVAL S.J., TWEEDIE R.L., ABRAMS K.R. et JONES D.R. (2000a). « Empirical assessment of effect of publication bias on meta-analyses », *British Medical Journal*, 320, 1574-1577.

SUTTON A.J., F. SONG, S. M. GILBODY et K. R. ABRAMS (2000b). « Modelling Publication Bias in Meta-Analysis : a review », *Statistical Methods in Medical Research*, 9, 421-45.

SUTTON A.J., ABRAMS K.R., JONES D.R., SHELDON T.A., et SONG F. (2000c). « *Methods for Meta-analysis in Medical Research* », Chichester : John Wiley and Sons.

THOMPSON S.G. et POCOCK S.J. (1991). « Can meta-analyses be trusted ? », *Lancet*, 338, 1127-1130.