School Working Paper

DEAKIN
UNIVERSITY AUSTRALIA

www.deakin.edu.au

ECONOMICS SERIES

SWP 2008/06

Theory Competition and Selectivity:
Are All Economic Facts Greatly Exaggerated?

Hristos Doucouliagos and
T.D. Stanley

The difference is **Deakin University.**

Theory Competition and Selectivity:
Are All Economic Facts Greatly Exaggerated?

Hristos Doucouliagos and T.D. Stanley

Abstract

There is growing concern and mounting evidence of selectivity in empirical economics. Most empirical economic literatures have a skewed (or truncated) distribution of results. The aim of this paper is to explore the links between publication selectivity and theory competition. In research areas where theory supports a wide range of outcomes, empirical evidence is less likely to be affected by selectivity. However, in those areas where theory is consistent with only one qualitative effect (e.g, a negative effect of price on quantity demanded), selectivity is more likely and its bias, more severe. This hypothesis is supported through the analysis of 65 distinct empirical economics literatures, involving approximately two thousand separate empirical studies, which in turn collectively contain many more thousands of estimates. Our meta-meta-analysis shows that publication selection is widespread, but not universal. It distorts scientific inference with potentially adverse effects on policy making, but competition and debate between rival theories reduces this selectivity and thereby improves economic inference. All literature reviews, whether traditional or quantitative (meta-analysis), need to adopt explicit selection correction methods.

Keywords: publication selectivity, theory competition, meta-meta-analysis, empirical economics

Doucouliagos: School of Accounting, Economics and Finance, Deakin University, 221 Burwood Highway, Burwood, 3125, Victoria, Australia. Email: douc@deakin.edu.au. Phone: 61 03 9244 6531.

Stanley: Department of Economics, Hendrix College, 1600 Washington St., Conway, AR, 72032, USA. Email: Stanley@hendrix.edu. Phone: 501 450 1276.

Theory Competition and Selectivity: Are All Economic Facts Greatly Exaggerated?

1.   Introduction

The aim of empirical analysis is to inform researchers and policy makers about the magnitude of key economic relationships and thereby serve the public interest. Science takes stock from the findings of numerous independent studies (Hunter and Schmidt, 2004). As research accumulates, the underlying associations are revealed with greater clarity and precision. However, valid inferences require representative observations. Over the past decades, economists have been concerned with the representativeness of publicly available empirical estimates. Do authors submit, and journals accept, only statistically significant findings or those findings that are in accord with economic theory? Is empirical economics distorted by publication selection? If so, the validity of inferences and the policy implications that can be drawn from an empirical literature would be cast in doubt.[1]   DeLong and Lang (1992) pose the question: 'Are all economic hypotheses false?' and find evidence of selection bias among the top economic journals. The distorting effects of publication selectivity are widely recognized by medical researchers and have caused the best medical journals to adopt explicit editorial policies to mitigate its effects (Krakovsky, 2004).

Indeed, most empirical investigations into the selectivity of economics research detect its presence. Examples include: Card and Krueger's (1995) analysis of the employment effects of minimum wages; Ashenfelter *et al.* (1999) for the returns to education; Görg and Strobl (2001) for the spillover effects of multinational companies; Ashenfelter and Greenstone (2004) for the value of life; Abreu *et al.* (2005) for growth-convergence; Doucouliagos (2005) for the impact of economic freedom on economic growth; Nijkamp and Poot (2005) for wage curves; Rose and

Stanley (2005) for the common currency effect on trade; Stanley (2005b) for water price elasticities; Waldorf and Byun (2005) for the impact of age structure on fertility; and Mookerjee (2006) for openness and economic growth.

Although the majority of investigations have found distinctive statistical evidence of selection, there are exceptions: the union-productivity literature (Doucouliagos and Laroche, 2003; Doucouliagos, Laroche and Stanley, 2005); the allocation of aid on the basis of economic growth (Doucouliagos and Paldam, 2007); the impact of corporate social responsibility on financial performance (Allouche and Laroche, 2005), and the effect of immigration on wages (Longhi, Nijkamp and Poot, 2005). It is the magnitude and variation of selectivity across empirical economic literatures that are the focus of our paper. We wish to extend and clarify the observations made by DeLong and Lang (1992).

The purpose of this study is to investigate the extent and impact of publication selection in economics and then to identify those factors that influence the severity of publication selectivity. The issue of publication selectivity is important, because the degree of selectivity affects the accuracy of any empirical inference and, hence, the usefulness of research for policy making. We explore the tentative hypothesis that areas of research where mainstream economic theory supports a specific directional effect are more likely to contain publication selection and bias. Theory drives empirical results, and the degree of competition among rival theories determines the degree of selectivity in empirical research. Specifically, where economic theory is consistent with a wider distribution of results, we predict and find less selectivity. We investigate the impact of theory competition among 65 distinct areas of research, involving approximately two thousand empirical studies and distributed widely across micro-, macro-, labor, and international economics.

The paper is organized as follows. Section 2 discusses the links between theory competition and publication selectivity. Section 3 presents our empirical framework and how to measure the severity of publication selectivity. Section 4 presents our meta-meta-analysis of empirical economics. Section 5 concludes.

2.   From Competition to Selectivity

Theoretical debates often generate an associated empirical literature designed to resolve related controversies and to test competing theoretical explanations. Indeed, empirical research makes its greatest contribution when it helps to settle such theoretical disputes. The intensity of theoretical contention varies across research literatures. In some literatures, there may be near universal agreement about the direction and magnitude of an effect, while in others even the direction of an effect may be hotly contested. Our conjecture is that the extent of theoretical competition shapes the distribution of reported empirical findings, *literally*. Theoretical debate provides avenues for the presentation of conflicting empirical findings. Where the monopoly power of the dominant theory is the strongest, we expect to find a large degree of selectivity. Studies that report results that are at odds with dominant theory will find it harder to get published. Conversely, where there is greater theory competition, or pluralism, we expect to find little or no selectivity, as all possible results are theoretically acceptable. With theory competition, there will be greater diversity and symmetry among reported research findings.[2]

Denote the degree of theory competition as κ, the degree of publication selectivity as π, and the reported average effect as η.[3] Selectivity is influenced by theory competition $\partial\pi/\partial\kappa < 0$, the *selection effect,* while publication selectivity, in turn, affects the reported economic effects, $\partial\eta/\partial\pi > 0$, the *inference effect*. Theory competition reduces selectivity, while selectivity

3

erroneously magnifies estimated empirical effects. Together, these two effects imply that the absence of theory competition exaggerates the reported empirical effect (*i.e.*, $\partial \eta / \partial \kappa < 0$). That is, dominant economic theories distort their empirical research literatures and the associated policy inferences.

2.1 The selection effect

The selection effect represents the link between publication selectivity and theory competition. Researchers often analyze data until they are satisfied with the results. It is as if empirical results are generated by a stopping rule, whereby researchers cease analyzing data when they have reached what they believe to be the 'truth,' or a sufficiently close approximation to it. However, what a researcher believes to be the 'truth' is likely to be influenced by what is consistent with prevailing theory. That is, theory defines the parameters of what is 'acceptable' and hence what might be publishable. Theory competition widens the set of acceptable results and thereby relaxes this stopping rule.

A stopping rule that results in selection bias will reveal itself in the form of a truncated or skewed distribution of reported findings. Theoretically 'unacceptable' results go either unreported or under-reported. Even if the unreported results are 'incorrect' (in the sense that they conflict with the 'true' theory), selectivity will exaggerate a literature's average of a key parameter, for example a price elasticity. Such distortions can make it appear that there is an empirical effect with full statistical rigor, where there is none, or make the effect appear much larger. In any case, the selection of what gets reported can lead to faulty inference. Such faulty inferences are illustrated in the next section on the inference effect.

Competition in economic theory emerges through the actions of researchers and journals. Often there are rival research groups that are engaged in active research competition with each other, defending their theory and testing their rivals' (*i.e.*, independent replications). For example, in the union-productivity effects literature, the "Harvard School" associated with Richard Freeman and James Medoff proposed the so-called "two-faces view" of unions (Freeman and Medoff, 1984). They and their supporters report evidence in favor of positive union-productivity effects. In contrast, neoclassical economists hold a more conventional model and find negative union-productivity effects (Hirsch, 1991 and 2004).

In the democracy-growth literature, some argue that democracy is detrimental to growth (Huntington, 1968; Haggard, 1990), while others argue that it is conducive to it (North, 1990; Olson, 1993). In the aid-effectiveness literature, two rival schools have emerged that claim to have identified the condition through which aid contributes to growth (Jensen and Paldam, 2006). These schools are the so-called 'good policy' (Burnside and Dollar, 2000) and the 'medicine' models (Hadjimichael et al., 1995). With respect to the effect of government deficits on the economy, Robert Barro and his followers argue that there is none ('Ricardian equivalence theorem'), while most of the rest provide theoretical reasons and empirical evidence for an effect (Barro, 1974; Bernheim, 1987; Barro, 1989).

In other cases, competition may be quite circumscribed. For example, there is no real debate about the direction of the effect that the adoption of a common currency will have on subsequent bilateral trade. However, many researchers question the very large estimates of currency unions' trade effects that are typically found, including those researchers that produce them (Rose and Stanley, 2005). At one extreme, where there is theoretical agreement on both the direction of the effect and its magnitude (*e.g.*, economists agree that the price elasticity of water

demand should be negative and inelastic), we can expect a high degree of selectivity. At the other extreme, where there is theoretical disagreement on both the direction of an effect and its magnitude, we expect to find little or no selectivity.[4] Where theory allows all possible results, we should find the empirical literature to be relatively free of selectivity.[5]

Theory dominance makes it harder to justify empirical results that contrast with the prevailing theory. Researchers will be less willing to submit results that are in sharp contrast to accepted theory. For example, it is difficult to defend a finding of a statistically significant negative effect of economic freedom on growth, as theory predicts a robust positive association. In contrast, a negative effect of political freedom on growth is not as problematic, because it is explicitly 'allowed' by theory. The absence of theory competition also means that referees will be more critical of conflicting findings and editors will be less willing to publish findings not sanctioned by prevailing theory. Publishing contrary results is not impossible, but authors will need to work harder to justify them, increasing the cost and reducing the relative frequency of contrary empirical results in the literature.

2.2   *The inference effect*

Publication selection is not, in itself, our concern. Nor is it necessarily a problem for economic science. Selectivity is a problem only when it distorts the inferences that can be made from a given research literature. Stanley (2005b) shows that even if publication selection arises for the best of reasons, it can have an adverse effect on inference. For example, in the case of water demand elasticities, publication selection leads to a fourfold increase in the estimated elasticity. Potentially worse still, is the case of currency union effect on international trade.
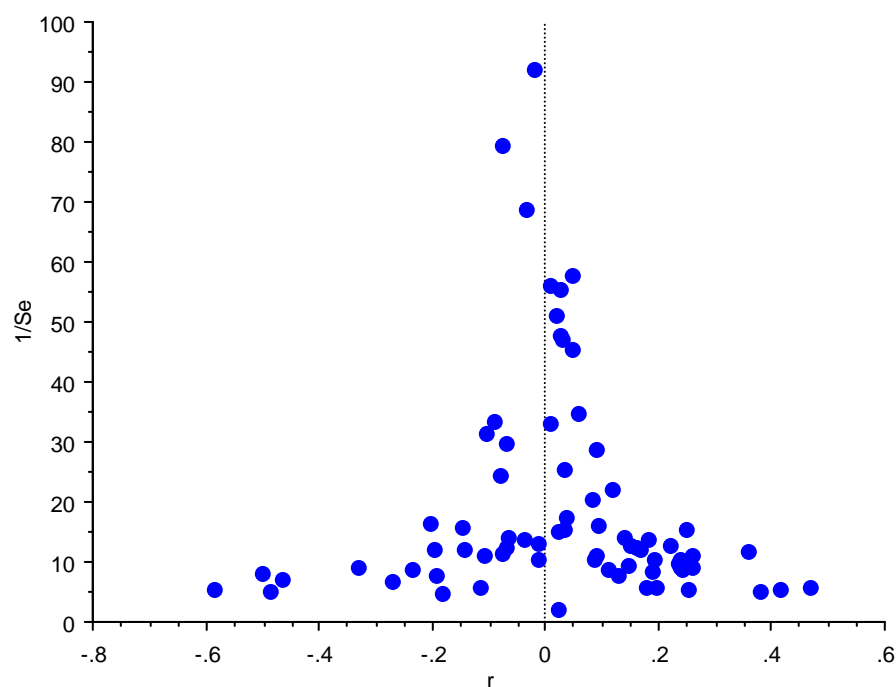
Rose and Stanley (2005) conduct a meta-analysis of 754 reported estimates of the effect of currency union contained in 34 studies. The simple average of these 754 estimates suggests that joining a common currency increases trade by 136%; that is, it more than doubles the trade among the countries involved. As mentioned previously, weighted averages which use the inverse of the estimate's variance are often used to mitigate the effects of selectivity. However, in this case, it makes little difference, because the random-effects estimate suggests that trade will increase by 123%.[6] Obviously, such an estimate could have profound policy implications for countries which are considering whether to join a currency union. If this research assessment were regarded as a hard fact, the United Kingdom would find it difficult to resist the Euro in spite of nationalistic reservations and unwillingness to relinquish control over UK's economy.

However, there are clear signs of publication selectivity in this area of research (Rose and Stanley, 2005). In contrast to this seeming large effect, a meta-regression correction for publication selectivity estimates this trade effect to be only 1.2%.[7] This difference could have a profound effect on desirable policy, and this difference is both 'economically' and 'practically' significant' (McCloskey, 1985; McCloskey, 1995; Ziliak and McCloskey, 2004). Clearly, selectivity can have important practical policy implications for currency unions. Nor is this the only example where the consequences of selectivity might be large.[8] Needless to say, when the estimate of price elasticity is off by a factor of four, a specific water conservation policy might be surprisingly ineffectual.

The best way to *see* how selection can distort inference is through a funnel plot. A funnel plot is a scatter diagram of an estimate's precision (*i.e.*, the inverse of its standard error) vs. the size of the estimate. Funnel graphs have been widely used by medical researchers to identify publication selection (Sutton et al., 2000; Rothstein et al., 2005), and they are rapidly being

adopted by economists (Roberts and Stanley, 2005). Figures 1, 2, and 3 display the distribution of estimates in three of the 65 areas of economic research used in our meta-meta-analysis. Figure 1 reflects a nearly symmetric distribution of union-productivity estimates (as measured by a partial correlation coefficient, r).[9] Areas of research where there is little selection (recall the 'two-faces' view) should look, more or less, like an inverted funnel.

Figure 1: Funnel Plot: Union-Productivity Partial Correlations

In contrast, the common-currency effect is highly skewed to the right (Figure 2). It is as if the left portion of a symmetric funnel were cut off.[10] Selection for large positive common-currency effects effectively removes the left portion of the full distribution of estimates, thereby causing any summary to be greatly exaggerated.

Figure 2: 5% Trimmed Funnel Plot: Common-Currency Effect on Trade
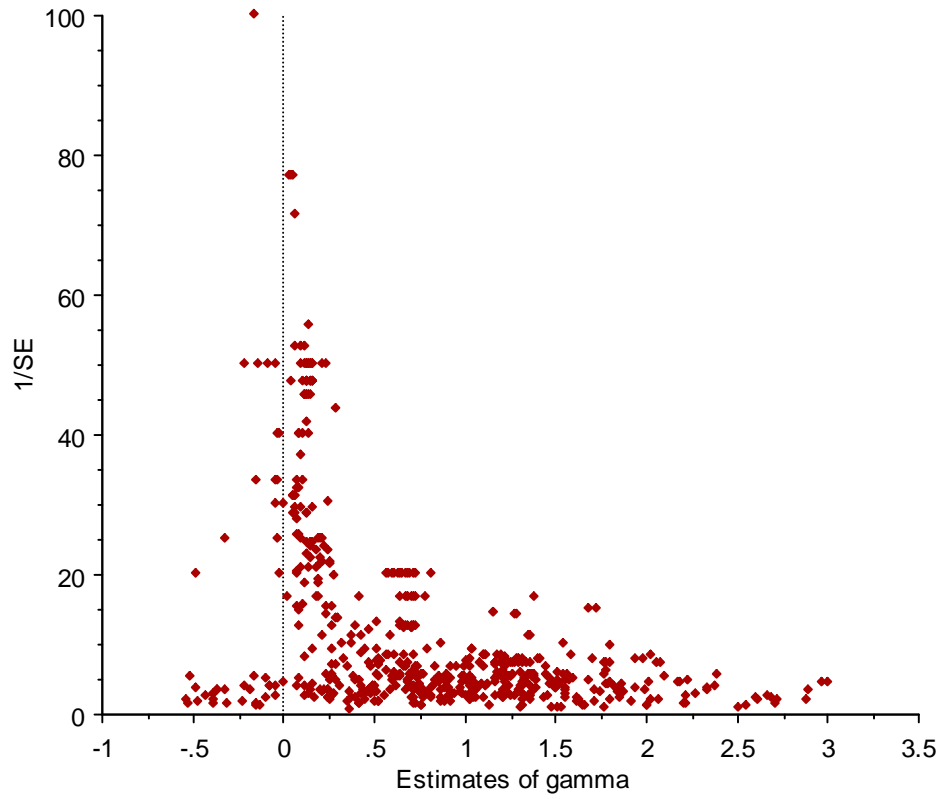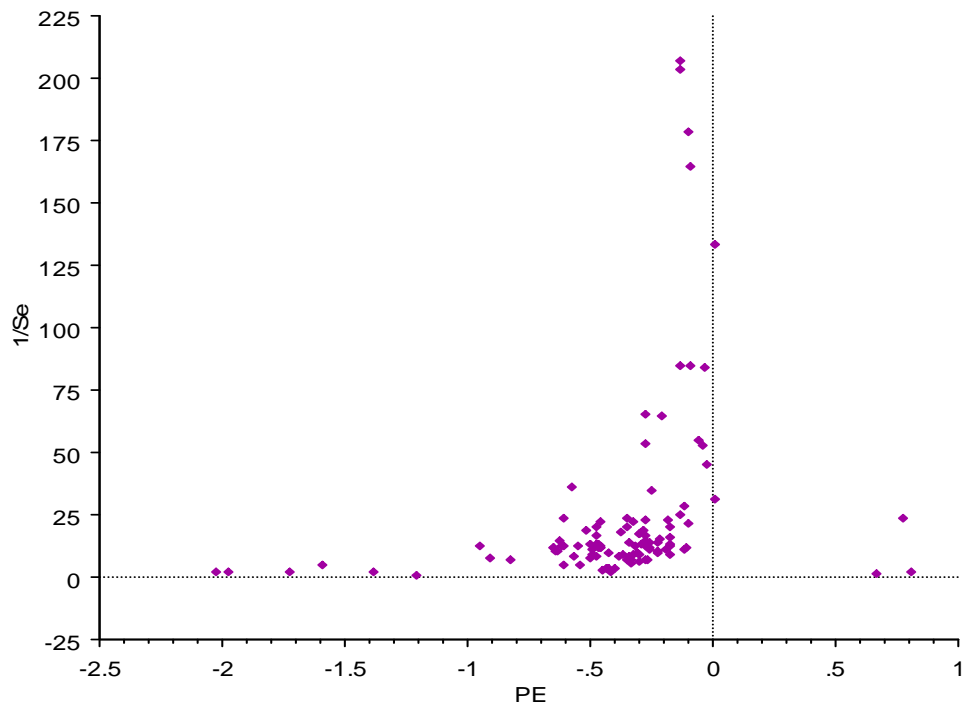

Figure 3: Funnel Plot: Price Elasticities for Water Demand

A similar distortion is seen among the estimates of the price elasticity (PE) of water demand (Figure 3). Here too, the funnel graph is highly skewed, but now to the left, and the average elasticity (-0.39) is several times larger than a corrected estimate (-0.082). Notice that the top of the graph points to rather small values of elasticity— about -0.10. The 'spout' of a funnel plot is an indicator of the 'true' effect, untrammelled by selection.

Theory competition increases the accuracy of inference by reducing the degree of selectivity. In this way, competition aids 'truth' revelation and policy relevance. In this sense, competition among rival theories advances science and increases welfare.[11]

Journalists and politicians have often characterized economists as strongly disagreeing with one another. As the old adage goes: "Ask 2 economists, and you get 6 different opinions." Although economists can be highly critical analysts, the magnitude of discord is not constant across all economic issues or areas of economic research. Indeed, survey evidence indicates that economists do not have as strong a propensity to disagree as the general public may believe. For example, Fuller and Geide-Stevenson (2003) found that among members of the American Economics Association there was strong consensus for 8 propositions (18%), substantial consensus for 18 (41%), modest consensus for 13 (30%), and no evidence of consensus in only 4 propositions (9%). Using data from the Survey of Americans and Economists on the Economy (1996), Caplan (2002, p. 439) shows that: "As a rule, the public is more pessimistic than economists, ranking problems' severity higher, the benefit of change lower, and outlook for progress worse". Fuchs (1996) found that consensus among health economists was significantly higher than it was amongst health professionals. Yet on that proverbial other hand, Klein and Stern (2006) found that there was less consensus among economists than there was in anthropology, history, philosophy, political science, and sociology.

3. Size Matters: Identifying, Measuring, and Explaining Publication Selectivity

3.1   FAT-MRA

The channel through which theory competition affects reported empirical results is selectivity. But how can selection be measured?   Fortunately, publication selection leaves a statistical trace that can be identified by simple meta-regression analysis (MRA) of the estimates reported in a given area of research.[12]   And, this MRA can also be used to estimate the magnitude of the publication selectivity.

$$\hat{\eta}_i = \beta_1 + \beta_0 SE_i + e_i \tag{1}$$

Where SE_i is the standard error associated with the estimated effect, $\hat{\eta}_i$, (see Card and Krueger, 1995; Egger $et\ al.$, 1997; Stanley, 2005b; Stanley, 2008).[13]

Clearly, equation (1) will have heteroskedasticity, and SE_i is its empirical estimate. Dividing equation (1) by SE_i gives the weighted least squares version of this meta-regression model of publication selectivity.[14]

$$t_i = \beta_0 + \beta_1 (1/SE_i) + \varepsilon_i \tag{2}$$

Economists and medical researchers have widely used these equations to test for publication selection (H$_0$: β_0=0) (Sutton $et\ al.$, 2000; Rothstein $et\ al.,$ 2005). This method for identifying publication selection has been called a funnel-asymmetry test (FAT) due to its relation to funnel graphs (Sutton $et\ al.$, 2000; Stanley, 2005b), and its validity has been confirmed in simulations and in several economic applications (Doucouliagos, 2005; Knell and

Stix, 2005; Longhi, Nijkamp and Poot, 2005; Rose and Stanley, 2005; Stanley, 2005a; Stanley, 2008; Cipollina and Salvatici, 2006; Coric and Pugh, 2006; Gammill *et al.*, 2007).

In the absence of publication selection, the estimated effect will be randomly distributed around its true value, β_1.[15] In equation (1), $\beta_0 SE_i$ represents systematic selection for statistical significance. Studies with smaller samples and, hence, larger standard errors (SE_i) will need to run and re-run their models more intensely to achieve a statistically significant result. More precise studies (*i.e.*, those with smaller SE_i) will require less searching and less selection to obtain the desired significant result.[16] Thus, the telltale signal of publication selection is a systematic relation of reported effects with their standard errors (Stanley, 2005b). Because the FAT coefficient, β_0, has no units of measurement, it may be further used to measure the magnitude of publication selectivity.

3.2   On the Importance of Being Selective

In a series of papers spanning two decades, McCloskey (1985) has reminded economists that there is an important distinction between statistical significance and genuine economic relevance (McCloskey, 1995; Ziliak and McCloskey, 2008). Economic importance is an entirely different matter that hinges on the magnitude of an empirical effect, not merely on its sign or statistical significance. This distinction and the concomitant abuse of statistical significance have caused debates throughout the social sciences, notably psychology (Thompson, 1996; Harlow, et al., 1997; Thompson, 2004).

Meta-regression analysis provides an alternative strategy for empirical inference. The above MRA models, equations (1) and (2), also contain a test for genuine empirical effect, after filtering selection for statistical significance (H$_0$: $\beta_1 = 0$). This precision-effect test (PET) offers

a viable methodology for empirical inference, one which is often robust to publication selectivity and the widespread abuse of statistical significance (Stanley, 2005a; Stanley, 2008).

In other social sciences, this abuse of statistical significance has caused researchers to develop and adopt a unit-less measure of the size of an empirical effect— the effect size. Although there are several statistical measures of effect size, the simplest is the ratio of the difference of the means between the control group and the treatment group to some pooled estimate of the standard deviation (σ). In psychology, this 'significance controversy' has caused a change in scientific practice, including the routine reporting of effect sizes (Thompson, 2004). But what do these effects sizes mean? How should we interpret them? For this purpose, Cohen (1988) offered the well-known and plausible guidelines: 0.2σ for a small effect, medium (0.50σ), and anything larger than 0.8σ is a large effect.

We have an analogous problem in understanding publication selectivity. It is not sufficient to determine whether publication selection exits, or not. Is it large enough to be a concern? Will the distortion that it causes make a practical difference? To address these issues, we offer the following guidelines regarding the practical significance of publication selectivity.

1. If FAT is statistically insignificant or if $|\hat{\beta}_0| < 1$, then selectivity is 'little to modest'.

2. If FAT is statistically significant and if $1 \leq |\hat{\beta}_0| \leq 2$, then there is 'substantial' selectivity.

3. If FAT is statistically significant and if $|\hat{\beta}_0| > 2$, then there is 'severe' selectivity.
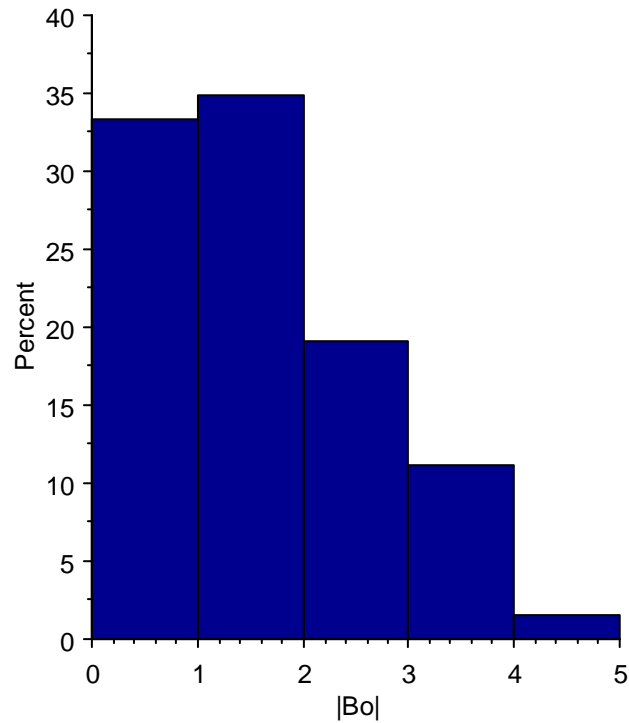
These guidelines are the results of Monte Carlo simulations over a wide range of design conditions that control for: whether or not there is an effect; the incidence of publication selectivity (0%, 25%, 50%, 75%, or 100%); the number of empirical studies available to the

meta-analysis (20 or 80); and the magnitude of variation caused by alternative model specifications. See Appendix 1 for the design details and the simulation results.

Consider the case where 100% of the research literature is selected for statistical significance but there is no genuine empirical effect. Then, reported t-values will be approximately 2, plus or minus random sampling error (Card and Krueger, 1995). With the reported t-values randomly varying around 2, the estimate of β_0, $\hat{\beta}_0$, from equation (2) will also be approximately 2. Thus, it seems appropriate to label values of $\hat{\beta}_0$ larger than 2 as 'severe.' As discussed previously, both literatures on water elasticity and common currency effects greatly distort their averages, and both have 'severe' publication selectivity ($\hat{\beta}_0$ = {2.86, 3.85}). Based on the 65 areas of economics research analyzed in this paper, we find that on average in economics, publication selectivity is 'substantial' (average $|\hat{\beta}_0|$ = 1.65). Approximately one-third of these areas of economics research display each of these three categories of selectivity (see Figure 4). Clearly, selectivity's potential to distort inference poses a serious a problem for empirical economics.

Simulations also show that there is a close association between the magnitude of $\hat{\beta}_0$ and the extent of selection's distorting effects (see Figure 5 in Appendix 1). Ninety-five per cent or more of the variation in the estimated bias from publication selection can be explained by $\hat{\beta}_0$ alone. Thus, this MRA estimate of $\hat{\beta}_0$ is an indicator of publication bias and thereby of the severity of the 'inference effect.'

Figure 4: Distribution of Selectivity in Empirical Economics

In contrast, if the absolute value of the FAT coefficient in MRA model (2) is less than one, selectivity is not practically significant, whether or not it is statistically significant. For example, recall Figure 1 and the union-productivity literature. Here, there is 'little to modest' selectivity ($\hat{\beta}_0$ = 0.56), and, as a result, there is little if any distortion to the average effect. In this area of research, the average partial correlation is 0.021, the weighted average is 0.002, and the corrected effect is -0.016; none of which are statistically or practically significant.[17]

Even moderate publication selectivity can have a practical effect. For example, in the efficiency wage literature the FAT coefficient is 1.25 (*i.e.,* 'substantial' selectivity), and the corrected estimate of the wage elasticity of output is approximately half this literature's average (0.30 vs. 0.61). If uncorrected, this difference is sufficient to give the appearance of consistency

with profit maximization (profit maximization requires that this elasticity be equal to labor's share), when the actual effect is much less—practically and statistically (Stanley and Doucouliagos, 2007).

Clearly there is much nuance and variation to selectivity in economics; the standard deviation of $\hat{\beta}_0$ among our 65 studies is 1.03. Thus, it is important to explain the substantial variation of selectivity found in economic research. This is precisely the task that Section 4 addresses, while the next section offers a framework in which to analyze selectivity.

3.3   Explaining Selectivity

Thus far, we have offered an objective way to identify publication selectivity and to measure its severity. Once such an empirical phenomenon is discovered, it becomes the researcher's duty to attempt to explain it. The general econometric model we use is given by:

$$\hat{\beta}_0 = \alpha_0 + \alpha_1 \kappa_j + \delta X_j + u_j \tag{3}$$

where κ_j is a measure of theory competition, j denotes the jth empirical literature, X_j is a vector of other variables, and u_j is a Gaussian error term. Equation 3 relates observed selectivity, $\hat{\beta}_0$, directly to competition, as well as to other factors that may influence selectivity independently of theory competition.

Measuring competition
The key explanatory variable that we wish to explore is theory competition. Ideally, this should be measured by a continuous variable, such as a concentration index. Because no such index exists, we are forced to use rough proxies and to construct our own measure. However, we

believe that these proxies capture some of the important dimensions of theory competition, and our empirical results confirm this optimism.

The first proxy we consider is a binary variable, **Competition**, where 1 is assigned to a literature where theory permits both positive and negative effects. In order to code this variable, we draw upon two sets of information. First, we refer to the key literature reviews for each area. Virtually all the extant reviews of empirical studies and meta-analysis discuss the underlying theory or theories. Second, we refer to surveys of economists' opinions. Examples of such surveys include Alston, Kearl and Vaughan (1992), Fuchs, Krueger and Poterba (1998), Fuller and Geide-Stevenson (2003), and Klein and Stern (2006). When there remains some doubt about whether theory permits both positive and negative effects, we code **Competition** one way and **Competition2** the opposite. We hypothesize that areas of research that have **Competition** will be less selective. Hence, the coefficient on κ, α_1, is expected to be negative.

Our second measure of theory competition is an indicator of the intensity of controversy, **Debate**, surrounding the hypothesis being tested. **Debate**=1, if there is considerable debate in the profession about this area of research; zero, otherwise. **Competition** and **Debate** are not used in the same meta-regression model because they measure essentially the same dimension of theory competition. **Debate**=1 is largely a subset of **Competition**=1.

A third measure, **Demand**, codes research areas that are based on demand. Standard microeconomic and macroeconomic theory states that demand functions are unlikely to have a positive slope.[18] That is, there is strong agreement on the direction of the price effect on the quantity demanded. Demand theory is well-developed and widely accepted by economists. Demand is a dominant theory, and its monopoly power is unsurpassed. Thus, we would expect that areas of research based on demand theory to possess greater selectivity.

A fourth indicator is a simple composite of **Competition**, **Debate,** and **Demand**. **Demand** = 1 is a subset of **Competition** = 0 and **Debate** = 1 is nearly a subset of **Competition** = 1. Thus, we define a composite index of theory competition as: **TC index** = **Competition** + **Debate** – **Demand** +1. **TC index** takes on values from 0 to 3. **TC index** = 3 represents the most competitive area of research {**Competition** = 1, **Debate** = 1 and **Demand = 0**}, while economic theory is expected to have the least competition when **TC index** = 0 {**Competition** = 0, **Debate** = 0 and **Demand** = 1}.

Endogenous competition

The above four measures of theory competition assume that competition is strictly exogenous. Perhaps, theory competition is itself endogenous to the research record? That is, researchers' views of a given theory may be influenced by the strength of the actual empirical evidence for or against it. We use $\hat{\beta}_1/SE$ to assess the strength of empirical evidence for the effect of a particular area of research, corrected for publication bias. Like effect size, $\hat{\beta}_1/SE$ can be considered a standardized measure of effect, one that is corrected for publication selection. $\hat{\beta}_1$ is the estimated coefficient on $1/SE_i$ in equation (2), and it serves as a correction for publication bias (Sutton et al., 2000; Stanley, 2008; Stanley and Doucouliagos, 2007). It is divided by the average standard error of the reported effects to serve as a relative measure of the strength of the underlying empirical phenomenon that can be compared across different areas of research. Individual meta-analyses use different measures of empirical effect: estimated regression coefficients, elasticities, partial correlation coefficients, and other functions of regression coefficients. Thus, effect must somehow be standardized if we wish to compare across the diverse meta-analyses that have been conducted in economics.

Next, theory competition may be defined endogenously as a function of our other variables and measures. **EndoComp** is the predicted probability that **Competition**=1 from a Logit model that uses $\hat{\beta}_1/SE$, **Debate**, and **Macroeconomics** as independent variables.[19] Thus, we estimate a recursive model.   The Logit model is first estimated and used to define **EndoComp**, which, in turn, replaces κ_j in equation 3 and employed to help explain a research area's level of publication selection.[20] None of these alternative measures of theory competition have a significant correlation with $\hat{\beta}_1/SE$ (Table 1); thus, it appears that theory competition is not influenced by the underlying empirical magnitude of the phenomenon in question. Nonetheless, theory competition may depend on other factors such as **Debate** and **Macroeconomics,** and $\hat{\beta}_1/SE$ may have a small marginal effect on theory competition after these other factors are taken into account. See Table 1 for the correlations among these measures of theory competition.

Table 1: Correlation Matrix for Alternative Measures of Theory Competition[21]

	Comp	Comp2	Debate	Demand	TC index	$\hat{\beta}_1/SE$
Competition	1.0000					
Competition2	0.8351	1.0000				
Debate	0.8137	0.7237	1.0000			
Demand	-0.3994	-0.4783	-0.3314	1.0000		
TC index	0.9062	0.8294	0.8818	-0.6784	1.0000	
$\hat{\beta}_1/SE$	-0.1571	-0.1542	-0.1643	-0.1676	-0.0717	1.0000

Other factors

In addition to competition, there could be other factors that shape selectivity. One such factor could be the field of study. Hence, we construct a binary variable, **Macroeconomics**. Several surveys of economists' opinions have found that there is less agreement on macroeconomic

issues and a greater degree of consensus with microeconomics (Fuller and Geide-Stevenson, 2003). Interestingly, however, this does not seem to apply to European economists (Frey and Eichenberger, 1992). **Macroeconomics** is not significantly correlated with our measures of theory competition. Thus, if it is a measure of theory competition, it is measuring an entirely different dimension than our other proxies. Nonetheless, **Macroeconomics** may have its own separate effect on the severity of publication selection, regardless of theory competition. Perhaps macroeconomics is more (or less) genuinely empirical and its tradition more (or less) reliant on the selection of reported empirical results?

Lastly, we distinguish whether the economic theory under examination is consistent with the null hypothesis, **AcceptHo** = 1 or the alternative hypothesis. When support for the tested economic theory is found through rejecting the null hypothesis, there are two reasons for selecting a significant result—statistical significance and pressure to present evidence consistent with theory. When economic theory is consistent with accepting Ho, these two motivations conflict, and we would expect less selection. Thus, those areas of research where theory is consistent with the acceptance of the tested empirical hypothesis should contain less publication selection.

4.   A Meta-Meta-Regression Analysis of Economics Research

4.1   Economics Research Data

Our data comprises the fruit of sixty-five meta-analyses of different areas of economics research. This data is the raw material for a meta-meta-regression analysis (M2RA) that seeks to detect patterns of economics research across different research literatures. It is difficult to estimate the number of distinct economics literatures, but the list must surely be very long. Many

of the included research literatures also have a traditional narrative literature review. Narrative reviews cannot be used in this meta-meta-regression analysis because they do not contain the quantitative measures that our empirical investigation requires. Hence, by necessity, our data are drawn from the population of meta-analyses.

Florax and Poot (2007) have recently identified 125 meta-analyses in economics, and we include as many of these as possible. However, the majority of these meta-analyses do not use MRA model (2) and therefore do not contain estimates of our dependent variable, $\hat{\beta}_0$. In order to calculate $\hat{\beta}_0$ from equation (2), we need estimates of the empirical effect from each of the approximately two thousand studies, as well as all of their associated standard errors. Then, $\hat{\beta}_0$ is estimated separately for each of these 65 literatures. In the past, meta-analysts rarely collected standard errors or used a FAT-PET-MRA model, equation (2), to control for selection bias, although practice has been changing in recent years. Thus, in the final analysis, we include 65 meta-analyses that either report the degree of selection bias or have made their data publically available. When the necessary information was not initially available, we contacted authors directly, most of whom either provided the data or ran meta-regression (2) on our behalf.[22] Hence, we are confident that our analysis is representative of the extant pool of publicly available meta-analyses. If these meta-analyses are representative of the population of economics literatures, then our results generalize beyond the existing quantitative reviews of empirical economics. These 65 economics literatures are listed in Appendix 2, together with the number of empirical studies, the values of $\hat{\beta}_0$, and **Competition**.

4.2 Meta-Meta-Results

Estimates of simple meta-meta-regression models (3) are presented in Table 3, where only one measure of theory competition is used in each M2RA to explain publication selectivity. All of our indicators of theory competition are significantly and negatively related to the observed degree of selection among empirical results.[23] Though simplistic, these results confirm our hypothesis that greater theory competition leads to less publication selection.

Table 3: Theory Competition and Publication Selectivity
(Dependent variable = $\hat{\beta}_0$)

Variable	(1)	(2)	(3)	(4)	(5)	(6)
Constant	2.04	2.08	2.08	1.44	2.13	2.45
	$(13.21)^*$	$(12.61)^*$	$(13.60)^*$	$(10.71)^*$	$(12.2)^*$	$(12.57)^*$
Competition	-1.10	-	-	-	-	-
	$(-5.39)^*$					
Competition2	-	-0.99	-	-	-	-
		$(-4.47)^*$				
Debate	-	-	-1.16	-	-	-
			$(-5.80)^*$			
Demand	-	-	-	0.97	-	-
				$(3.41)^*$		
EndoComp	-	-	-	-	-1.35	-
					$(-5.25)^*$	
TC index	-	-	-	-	-	-0.53
						$(-6.44)^*$
R^2	0.27	0.23	0.30	0.15	0.27	0.35
Standard error	0.89	0.91	0.87	0.96	0.89	0.84
k	65	65	65	65	61	65

* and ** one-tail statistically significant at the 1% and 5%, respectively. k is the number of studies included. Reported t-values are based on heteroskedasticity-robust standard errors. Some observations are lost in column 5 due to lack of information.

Between one-fourth and one-third of the observed variation in publication selection can be explained by a single index or measure of theory competition. Though less than ideal, this explanatory power is rather remarkable when one remembers that the funnel-asymmetry test has low power (Sutton et al., 2000; Macaskill et al., 2001; Stanley, 2008). Hence, our dependent

variable in all of these M2RAs, $\hat{\beta}_0$, will have correspondingly large *random* sampling errors, which inherently cannot be explained. On average, the observed standard error of this random sampling error is 0.735. Thus, the **TC index**, for example, is responsible for explaining 65% of the potential systematic variation among observed $\hat{\beta}_0$ s.[24]

Needless to say, publication selection is likely to be a more complex phenomenon than can be adequately captured by any single indicator. Table 4 reports the multivariate meta-meta-regression models of publication selection, equation (3). All models contain a measure of theory competition along with **Demand**, **Macroeconomics,** and **AcceptHo**.[25] All of these variables have the expected signs and are statistically significant.[26] Together, these indicators of theory competition explain between thirty and forty percent of the observed variation in publication selectivity.

Note that there is remarkable consistency in the coefficients across these M2RA models (Table 4). Each measure of theory competition is again statistically and practically significant in explaining publication selection, and their coefficients have roughly the same magnitude.[27] No matter how we code theory competition, as **TC index**, **Debate**, **Demand**, **Competition**, **Competition2** or endogenously defined, greater competition leads to significantly less selectivity. The magnitudes of these M2RA coefficients are also large enough to have a practical effect on empirical economic inference. Theory competition can transform 'severe' selection to 'substantial', and 'substantial' selection may be rendered essentially harmless through a competition of ideas.

All of these estimated models reflect essentially the same explanation of publication selectivity in economics. Studies grounded on demand theory are the most selective, because there is a strong professional consensus about its validity. Empirical results in apparent conflict

with the law of demand will tend to go unreported or dismissed as obvious error. As we discussed previously, even if demand theory were *entirely* valid, such selection can cause a large distortion to the reported elasticities. By itself, **Demand**=1 causes an area of research to have 'severe' selection.[28] Thus, the magnitudes of **Demand**'s M2RA coefficients are not only statistically significant but are also large enough, in conjunction with the intercept, to be practically important.

Table 4: A Meta-Meta-Regression of Economics Research,
(Dependent variable = $\hat{\beta}_0$)

Variable	(1)	(2)	(3)	(4)	(5)
Constant	1.60	1.60	1.63	1.57	2.17
	(7.24)*	(6.40)*	(7.43)*	(6.94)*	(6.85)*
Competition	-0.91	-	-	-	-
	(-3.97)*				
Competition2	-	-0.70	-	-	-
		(-2.69)*			
Debate	-	-	-0.96	-	-
			(-4.33)*		
EndoComp	-	-	-	-1.11	-
				(-3.86)*	
TC index	-	-	-	-	-0.51
					(-4.55)*
Demand	0.74	0.75	0.79	0.87	-
	(2.37)**	(2.23)**	(2.50)**	(2.72)*	
Macroeconomics	0.50	0.45	0.46	0.63	0.48
	(2.16)**	(1.78)	(2.03)**	(2.65)*	(2.11)**
AcceptHo	-0.96	-0.65	-0.44	-0.52	-0.68
	(-1.83)	(-2.51)**	(-1.85)	(-1.38)	(-1.77)
R^2	0.37	0.31	0.40	0.41	0.40
Standard error	0.84	0.88	0.83	0.82	0.82
k	65	65	65	61	65

* and ** one-tail statistically significant at the 1% and 5%, respectively. k is the number of meta-analyses included. Reported t-values are based on heteroskedasticity-robust standard errors. Some observations are lost in column 4 due to missing information. Appendix 2 lists the literatures included in the analysis. See text for definitions of the various measures of competition and other explanatory variables.

Similar results are uncovered for **Macroeconomic** research. We find that macroeconomic research studies contain greater selection than microeconomic applications. Together, **Macroeconomic Demand** studies are virtually guaranteed to be subjected to 'severe'

selection. In contrast, areas of research where mainstream economics is consistent with the null hypothesis under examination (*e.g.*, rational expectations) are likely to exhibit less selection. When selectivity is expected to favor the null hypothesis (**AcceptHo** = 1), there is systematically less selection for statistical significance (Table 4). In all cases and M2RA models, we find the same effects on publication selection.[29]

Between 30 and 40 percent of the variation in the observed severity of publication selection can be explained regardless of which dimensions are used to identify theory competition. Because the remaining unexplained standard error of our meta-meta-regression is only between 0.82 and 0.88, there is little potential systematic variation left to explain.[30]

These research results can provide interesting predictions for other areas of research that may yet to be meta-analyzed. For example, our M2RA estimates predict 'substantial' selectivity ($\hat{\beta}_0 \approx 1.60$) for microeconomic research not involving demand when there are no rival theories (**Macroeconomics** = **Demand** = **AcceptHo** = **Competition** = **Debate** = 0). Microeconomics research not based on demand theory but with some theory competition (**Competition** =1) is expected to have little selection. On the other hand, areas of research involving either the theory of demand or macroeconomics are predicted to have 'severe' publication selection.

No doubt there will be exceptions to our simple model of economics research. For example, one would expect that the 'true' magnitude of the empirical effect might be expected to have an inverse effect on publication selectivity. That is, if the empirical economic effect were truly large, little selection will be needed to produce the desired significant effects. However, our empirical measure of the 'true' effect size, $\hat{\beta}_1/SE$, has no discernible effect on $\hat{\beta}_0$ when added to any of these M2RAs.

Robustness

Table 5 reports the results of several additional regressions designed to test the robustness of the M2RA results presented in Table 4. Each cell reports the results of a given regression method for a particular measure of theory competition. For ease of comparison, row 1 repeats the key results from Table 4. In row 2, we use robust regression but find essentially the same results. Some of the meta-analyses are conducted by the same authors. This raises the question of whether there might exist some statistical dependence in the measure of publication bias. This is a problem that is more likely to be of concern at the level of an individual meta-analysis, rather than in a meta-meta-analysis. Nevertheless, we estimate a multi-level (mixed-effects or REML) model, using author identifiers to identify clusters of observations. The results are reported in row 3 and are basically the same as those from OLS.

Table 5: A Meta-Meta-Regression of Economics Research, Robustness Checks
(Dependent variable = $\hat{\beta}_0$)

Variable	Competition	Competition2	Debate	EndoComp	TC index
Base specification (1)	-0.91 (-3.97)*	-0.70 (-2.69)*	-0.96 (-4.33)*	-1.06 (-3.69)*	-0.51 (-4.55)*
Robust regressions (2)	-0.87 (-3.42)*	-0.70 (-2.58)*	-0.92 (-3.73)*	-1.02 (-3.14)*	-0.47 (-3.76)*
REML (3)	-0.90 (-3.79)*	-0.65 (-2.44)*	-0.95 (-4.12)*	-1.04 (-3.39)*	-0.49 (-4.03)*
With research interest (4)	-0.94 (-3.71)*	-0.74 (-2.50)**	-0.99 (-4.09)*	-1.11 (-3.45)*	-0.53 (-4.25)*
With field dummies (5)	-0.89 (-3.84)*	-0.72 (-2.82)*	-0.98 (-4.35)*	-1.07 (-3.70)*	-0.51 (-4.52)*

Each cell reports the coefficient on the competition variable. Base specification coefficients reproduced from Table 4. * and ** one-tail statistically significant at the 1% and 5%, respectively. REML denotes the random-effects multi-level model

In row 4, we add the variable **Research Interest.** This was constructed by dividing the number of studies included in each meta-analysis by the time span covered. *Ceteris paribus*, the greater the number empirical studies and empirical findings that are made available, the more interest the profession reveals in the research area and potentially the greater the competition. A

full symmetry scatter of empirical results takes time to develop. A literature may ultimately be free of selection effects but it takes time for studies to become publicly available. This variable always had the expected negative sign but was not statistically significant with one exception. When competition is measured by **Debate**, **Research Interest** has a coefficient of -0.01 and a t-value of -1.75.

In row 5 we add three field dummy variables. That is, in addition to macroeconomics, we add dummy variables for economic growth studies, labor economics, and industrial studies. These dummies are not statistically significant and do not affect the results for the other variables.

5.   Conclusion

The aim of this paper is to investigate the links between competition in economic theory and publication selection. Publication selectivity is widespread in economics and has potentially large and practical effects on empirical economic findings. Among 65 separate areas of economic research, contained in approximately two thousand empirical economic research studies, selectivity is, on average, 'substantial.' Much of the variation in the severity of publication selection can be explained by theory competition or its absence. Macroeconomic research and research based on demand theory are found to contain severe selectivity and are therefore predicted to impart significant distortions to the empirical findings reported in these areas of research. Our findings can be interpreted as testable hypotheses about economics research, which can be further corroborated in areas of research outside our current research database. That is, new meta-analyses can be conducted in areas of research that depend on

demand or involve macroeconomics to see if they contain severe selectivity, as predicted by our M2RA results.

Regardless, there are important implications of the current study for economic research and policy. As discussed above, a moderate amount of selectivity (*e.g.*, $\hat{\beta}_0$=1.25) can cause a doubling of the reported elasticity (recall efficiency wages). For demand studies, we expect publication selectivity to be 'severe' and the reported elasticities to be exaggerated many fold. Should we, as a result, suspect all empirical estimates of price elasticity? Even if the law of demand were true and universally applicable, the selection of what researchers (or reviewers) regard as acceptable elasticities can be expected to bias reported estimates greatly.[31] Thus, policy makers and researchers might wish to re-evaluate their beliefs about price sensitivity, as well as other economic 'facts' that are taken for granted. Ironically, the empirical 'facts' that enjoy the strongest consensus are those that can be expected to contain the greatest exaggeration and hence will be the least accurate.

Economists are firm supporters of competition in the economy and its markets. Thus, it should come as no surprise that the absence of theory competition distorts research and potentially invalidates empirical economic inference, no doubt creating a corresponding deadweight loss. The implications of our findings are obvious and potentially profound. In medical research, a similar concern about widespread publication selection caused the top medical journals to adopt an explicit publication policy to deal with this problem (Krakovsky, 2004).

This research also has implications for the conduct of meta-analysis in economics. Meta-analysts should always employ some method to detect and correct publication selection, because publication selectivity and its concomitant bias are widespread. Publication bias detection and

correction methods have already been widely adopted but they need to become standard practice. These methods include the MRA model (2), which we use to estimate $\hat{\beta}_0$, variations on this MRA model (Stanley 2008; Stanley and Doucouliagos, 2007), and Hedges' maximum likelihood, publication selection estimator (Hedges, 1992; Ashenfelter et al., 1999). Without some explicit selection correction strategy, all summaries of empirical economics, whether produced by conventional narrative reviews or by meta-analysis, must be regarded as suspect. 'Are all economic hypotheses false?' Of course not, but reports of economic facts are greatly exaggerated.

References

Abreu, M., de Groot, H.L.FR. and Florax R.G.M. 2005. A meta-analysis of beta-convergence: The legendary two-percent. *Journal of Economic Surveys* 19: pp. 389-420.

Allouche, J. and Laroche, P. 2005. A meta-analytical examination of the link between corporate social and financial performance. *Revue de Gestion des Ressources Humaines* 57: 18-41.

Alston, R. M., Kearl, J. R. and M.B. Vaughan. 1992. Is there a consensus among economists in the 1990's? *American Economic Review*, 82(2):203-09.

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

Ashenfelter, O. and Greenstone, M. 2004. Estimating the value of a statistical life: The importance of omitted variables and publication bias. *American Economic Review* 94:454-60.

Barro, R. 1974. Are government bonds net wealth? *Journal of Political Economy* 82: 1095-117.

Barro, R. 1989. The Ricardian approach to budget deficits. *Journal of Economic Perspectives*. 3: 37–54.

Bernheim, B. D. 1987. Ricardian equivalence: An evaluation of theory and evidence. In S. Fischer, ed. *NBER Macroeconomic Annual: 1987*. Cambridge, Mass.: MIT Press.

Burnside, C. and Dollar, D. 2000. Aid, policies and growth. *American Economic Review* 90: 847-68.

Caplan, B. 2002. Systematically biased beliefs about economics: Robust evidence of judgmental anomalies from the survey of Americans and economists on the economy. *Economic Journal* 112: 433-58.

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

Cipollina, M. and Salvatici, L. 2006. Reciprocal trade agreements in gravity models: A meta-analysis. TradeAG Working Paper No. 06/12, Social Science Research Network.

Cohen, J. 1988. *Statistical Power Analysis in the Behavioral Sciences*. 2nd ed. Hillsdale: Erlbaum.

Coric, B. and Pugh, G. 2006. The effects of exchange rate variability on international trade: A meta regression analysis. IESR Working Paper 01-2006.

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

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

Dobson, S., Ramlogan, C. and E. Strobl. 2006. Multinational companies and productivity: A meta-analysis, *Scottish Journal of Political Economy* 53:153-73.

Doucouliagos, C. 2005. Publication bias in the economic freedom and economic growth literature. *Journal of Economic Surveys* 19:367-88.

Doucouliagos, C. and Laroche, P. 2003. What do unions do to productivity: A meta-analysis. *Industrial Relations* 42: 650-691.

Doucouliagos, C. Laroche, P. and Stanley T.D. 2005. Publication bias in union-productivity research. *Relations Industrielles/Industrial Relations* 60:320-346.

Doucouliagos, C. and Paldam, M. 2007. Explaining development aid allocation by growth. Working Paper, Department of Economics, University of Aarhus.

Egger M, Davey Smith G, and Schneider M. 1997. Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal* 315:629–34.

Feige, E.L. 1975. The Consequence of Journal Editorial Policies and a Suggestion for Revision. *Journal of Political Economy* 83: 1291-5.

Florax, Raymond and Jacques Poot. 2007. Learning from the Flood of Numbers: Meta-analysis in Economics. Paper presented at the Aarhus Colloquium of Meta-analysis in Economics, September 27-30, Sandbjerg Manor, Sønderborg, Denmark.

Freeman R. and Medoff J. 1984. *What do unions do?* New York: Basic Books.

Frey, B. and Eichenberger, R. 1992. Economics and economists: A European perspective. *American Economic Review* 82:216-20.

Fuchs, V. 1996. Economics, values, and health care reform. *American Economic Review* 86:1-24.

Fuchs, V., A. Krueger and J. Poterba. 1998. Economists' views about parameters, values, and policies: Survey results in labor and public economics. *Journal of Economic Literature* 36:1387-425.

Fuller, D. and D. Geide-Stevenson. 2003. Consensus among economists: Revisited. *Journal of Economic Education,* Fall, 369-87.

Gammill, M.C., Costa-Font, J., and McGuire,A. 2007. In search of a corrected prescription drug elasticity estimate: A meta-regression approach. *Health Economics,* www.interscience.wiley.com.

Görg, H. and E. Strobl. 2001. Multinational companies and productivity: A meta-analysis, *The Economic Journal* 111:723-39.

Haggard, S. 1990. *Pathways from the Periphery: The politics of growth in the Newly Industrializing Countries.* Ithaca, New York: Cornell University Press.

Harlow, L.L., Mulaik, S.A., Steiger, J.H. (eds.) 1997. *What If There Were No Significance Tests?* Erlbaum, Mahwah, NJ.

Hadjimichael, M.T., Ghura, D., Mühleisen, M., Nord, R., Ucer, E.M., 1995. Sub-Saharan Africa: Growth, savings, and investment, 1986-93. *IMF Occasional Paper*, No. 118.

Heckman, J.J. 1977. Sample selection bias as a specification error. *Econometrica* 47: 153-62.

Hedges, L. V. 1992. Modelling publication selection effects in meta-analysis. *Statistical Science 7:* 246-255.

Hirsch B. 1991. *Labor Unions and the Economic Performance of Firms*. W.E. Upjohn Institute for Employment Research, Kalamazoo, Michigan.

Hirsch, B. 2004. What do unions do for economic performance? *Journal of Labor Research* 15: 415-55.

Hunter, J., and F. Schmidt. 2004. *Methods of Meta-Analysis: Correcting error and bias in research findings*. Sage, London.

Huntington, S. 1968. *The Political Order in Changing Societies*. New Haven: Yale University Press.

Huntington, S. and J. Dominguez. 1975. Political Development. In *Macropolitical Theory: Handbook of Political Science.* Volume 3. Edited by Fred I. Greenstein and Nelson W. Polsby. Reading, MA: Addision-Wesley.

Jensen, P. and Paldam, M. 2006. Can the Two New Aid-Growth Models be Replicated? *Public Choice*, 127:147-75.

Klein, D.B. and Stern, C. 2006. Economists' policy views and voting. *Public Choice*, 126:331-42.

Knell, M. and Stix, H. 2005. The income elasticity of money demand: A meta-analysis of empirical results. *Journal of Economic Surveys* 19:513-33.

Krakovsky, M. 2004. Register or perish. *Scientific American* 291:18-20.

Krassoi-Peach, E. and Stanley T.D. 2007. Are workers paid too much? A meta-regression analysis of the efficiency wage hypothesis. manuscript.

Laird, N. and Mosteller, F. 1988. Discussion of the paper by Begg and Berlin. *Journal of the Royal Statistical Society* (Series A) 151: 456.

Longhi, S., Nijkamp, P. and J. Poot. 2005. A meta-analytic assessment of the effect of immigration on wages. *Journal of Economic Surveys* 19: 451-77.

Lovell, M.C. 1983. Data Mining. *The Review of Economics and Statistics* 65: 1-12.

Macaskill, P., Walter, S.D., and Irwig, L. 2001. A comparison of methods to detect publication bias in meta-analysis. *Statistics in Medicine* 20:641-654.

Merton, R.K. and A. S. Kitt, 1950. Contributions to the theory of Reference Group Behaviour, in: Merton, R.K. and P.F. Lazarsfeld, eds., *Continuities in Social Research: Studies in the Scope and Method of "The American Solider",* (The Free Press: Glencoe), 40-105.

McCloskey, D.N. 1985. The loss function has been mislaid: The rhetoric of significance tests. *American Economic Review* 75:201-05.

McCloskey, D.N. 1995. Insignificance of statistical significance. *Scientific American* 272:32-33.

Mookerjee, R. 2006. A meta-analysis of the export growth hypothesis. *Economics Letters* 91:395-401.

Nelson, J.P. 2006. Cigarette advertising regulation: A meta-analysis. *International Review of Law and Economics* 26:195-226.

Nijkamp, P., and Poot, J. 2005. The last word on the wage curve? A meta-analytic assessment. *Journal of Economic Surveys* 19: 421-450.

North, D. 1990. *Institutions, Institutional Change and Economic Performance*. Cambridge, Cambridge University Press.

Olson, M. 1993. Dictatorship, democracy, and development. *American Political Science Review* 87:567-76.

Peloza, J. and P. Steel. 2005. The price elasticities of charitable contributions: A meta-analysis. *Journal of Public Policy and Marketing* 24:260-72.

Phillips, J.M. 1994. Farmer education and farmer efficiency: A meta-analysis. *Economic Development and Cultural Change* 42:149-65.

Phillips, J. M. and Goss, E. P. 1995. The effects of state and local taxes on economic development: A meta-analysis. *Southern Economic Journal* 62: 2–29**.**

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

Rose, A.K. and Stanley, T.D. 2005. A meta-analysis of the effect of common currencies on international trade. *Journal of Economic Surveys* 19:347-65.

Rothstein, H.R., Sutton, A.J. and Bornenstein, M. 2005. *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments*. Chichester: John Wiley and Sons.

Secord, P.F. and C. W. Backman. 1974. *Social Psychology*. 2nd edn, (McGraw Hill Kogakusha: Tokyo).

Simon, H. 1982. *Models of Bounded Rationality: Behavioral Economics and Business Organization, Vol. 2*. (MIT Press: Cambridge, Mass).

Stanley, T.D. 1998. New wine from in old bottles: A meta-analysis of Ricardian equivalence. *Southern Economic Journal* 64: 713-27.

Stanley, T.D. 2001. Wheat from chaff: Meta-analysis as quantitative literature review. *Journal of Economic Perspectives* 15: 131-50.

Stanley, T.D. 2005a. Integrating the empirical tests of the natural rate hypothesis: A meta-regression analysis. *Kyklos* 58:587-610.

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

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

Stanley, T.D. and Jarrell, S.B. 1989. Meta-Regression Analysis: A quantitative method of literature surveys. *Journal of Economic Surveys* 3: 161-170.

Stanley T.D. and C. Doucouliagos (2007). Identifying and Correcting Publication Selection Bias in the Efficiency-Wage Literature: Heckman Meta-Regression**.** School Working Paper, Economics Series 2007-11, Deakin University.

Steichen, T.J, Egger, M, and J. Sterne. 1998. Tests for publication bias in meta-analysis. *Stata Tech Bull* 44:19.1.

Sutton, A.J., Abrams, K.R., Jones, D.R., Sheldon, T.A. and Song, F. 2000. *Methods for Meta-Analysis in Medical Research*. Chichester: John Wiley and Sons.

Thompson, B. 1996. AERA Editorial policies regarding statistical significance testing: Three suggested reforms. *Educational Researcher* 25:26-30.

Thompson, B. 2004. The "significance crisis" in psychology and education. *Journal of Socio-Economics* 33:607-13.

Tullock, G. 1959. Publication Decisions and Tests of Significance – A Comment. *Journal of the American Statistical Association* 54: 593.

Waldorf, B. and P. Byun. 2005. Meta-analysis of the impact of age structure on fertility. *Journal of Population Economics*, 18:15-40.

Weiser, R. 2005. Research and development productivity and spillovers: Empirical evidence at the firm level. *Journal of Economic Surveys* 19:587-621.

Whaples, R. 1996. Is there consensus among American labor economists? Survey results on forty propositions. *Journal of Labor Research* 17:725-34.

Ziliak, S.T. and D.N. McCloskey 2008. *The Cult of Statistical Significance.* University of Michigan Press, Ann Arbor.

Appendix 1: Simulating Publication Selectivity and its Measurement

These simulations are all based on research literatures that test a given regression coefficient (*i.e.,* H_0:β_1=0). Such simple regression tests are meant only as a paradigm for testing other effects, in general. Similar statistical properties for these MRA tests should be found when an empirical literature uses other specific statistical tests.

The basic structure of these meta-regression simulations may be sketched as:

1. Generate the regression variables randomly.
2. Use OLS to estimate and test H_0:β_1=0. Select significant test results. Each selected test of H_0:β_1=0 comprises one study's reported result in our hypothetical empirical literature.
3. Simulate the FAT-MRA by repeating the previous steps either 20 or 80 times. At this stage, meta-regression model (2) is estimated to provide *one* estimate of $\hat{\beta}_0$.
4. Repeat all of the above steps 1,000 times while tracking various outcomes for FAT.

The first step defines the data-generating process. The independent variable (X_1) for each study is simulated by a random uniform variable (100, 200). As long as the independent variable is stationary, its distribution will not matter. Y is then generated from:

$$Y_i = 100 + \beta_1 X_{1i} + \beta_2 X_{2i} + 100\ e_i \qquad\qquad i=1, 2, \ldots n \qquad\qquad (4)$$

$e_i \sim$ NID(0,1). The effect in question, β_1, is assumed to be either 0 or 1. When β_1=1, the average R^2 is approximately 9%, and the correlation coefficient is about 0.3. Changing the values of β_1 only affects the power of the FAT-PET tests. The larger one makes β_1, the higher the power. The $\beta_2 X_{2i}$ term induces misspecification bias, in general, and omitted-variable bias, in particular.

As is widely known, omitting a relevant variable from a regression model causes the estimate of β_1 to be biased and inconsistent. Because this bias remains in large samples, it can be mistaken for genuine effect, potentially causing problems for any summary or statistical test.

Random misspecification bias is induced by making β_2 in equation (4) a random normal variable, $N(0, \sigma_{bias})$ and $\sigma_{bias}=\{.25, 1.0, 4.0\}$. This random misspecification bias acts as 'heterogeneity,' which has been recognized as a key parameter by other meta-analysts (Hedges and Vevea, 1996; Sutton et al., 2000a). The most influential dimension for the performance of these meta-regression methods is the size of the typical misspecification bias (σ_{bias}) relative to the sampling error. The larger the ratio of the standard deviation of these misspecification biases (σ_{bias}) to the standard deviation of the sampling errors (σ_{b_1}), the more leverage selectivity has. The larger the typical omitted-variable bias, *ceteris paribus*, the larger $\hat{\beta}_0$ can become.

The standard error of the estimate in MRA model (2) serves as rough guide to the amount of random misspecification bias available. This standard error measures the unexplained variation and will be equal to one in the absence of contaminating misspecification bias. Larger σ_{bias} increases this standard error. Large values of σ_{bias} are chosen because they represent a 'worse case scenario,' and they encompass the observed values of MRA model (2)'s standard error observed in our 46 areas of research.

In order to induce omitted-variable bias, X_{2i} is made equal to X_{1i} plus a random normal error. Thus, β_2, itself, becomes the omitted-variable bias. Given these magnitudes and the randomness of both the omitted-variable bias and the sampling error, bias and error will often

overwhelm a study's results. This becomes all the more true when there is also publication selection.

The meta-regression models are assumed to be estimated using either 20 or 80 studies. Twenty is chosen because it is a rather small sample size for any regression estimate, while eighty is both practically feasible and gives these MRA tests power to spare. Sample sizes chosen for the original studies and used to test $H_0: \beta_1 = 0$ are {30, 50, 75, 100, 200}. Publication bias is simulated as selecting a statistically significant positive b_1. That is, if the random estimate does not provide a significantly positive t-value, a new sample is taken and the original regression is run again with different random errors and random biases until a significant t-value is obtained by chance. For example, the 50% publication selection condition assumes that exactly half of the studies estimate and re-estimate their regression models until a random, yet statistically significant, estimate is found and reported. For the other half, the first random estimate, significant or not, is reported.

In practice, not all reported results that are published will have been selected for statistical significance. Therefore, to be comprehensive, it is assumed that the incidence of publication selection is either: 0%, 25%, 50%, 75% or 100%.

Appendix Table 1 reports the average value of $\hat{\beta}_0$ from 1,000 replications of MRA model (2). Note that value of $|\hat{\beta}_0| > 2$ occur only when 75% or more of the study results are selected for statistical significance. Values notably greater than one require 50% publication selection, while values of $\hat{\beta}_0$ less than or equal to approximately one occur only when publication selection is 25% or lower. Of course, these simulations only provide a rough guide

to the likely values of $\hat{\beta}_0$, and these estimates will themselves exhibit considerable variation, especially for large σ_{bias}. Nonetheless, we believe that our guidelines provide useful, if very approximate, benchmarks for the importance of being selective.
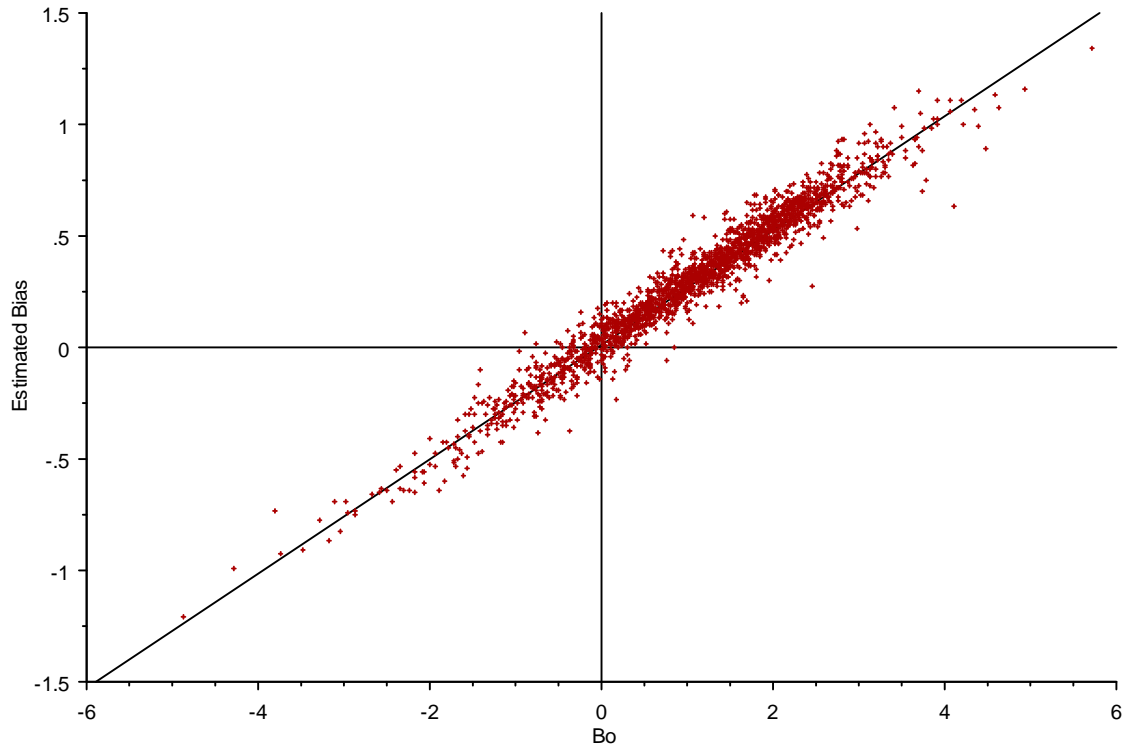
Appendix 1: Table 1

Average $\hat{\beta}_0$

Mis-specification	Publication Selectivity	No Effect *i.e.,* $(\beta_1=0)$		Effect *i.e.,* $(\beta_1=1)$	
		n=20	n=80	n=20	n=80
$\sigma_{bias}=.25$	0%	.045	.011	-.022	-.009
	25%	.422	.457	.370	.314
	50%	.967	.949	.620	.650
	75%	1.483	1.473	.957	1.011
	100%	2.081	2.062	1.339	1.347
$\sigma_{bias}=1.0$	0%	-.018	.069	-.004	.008
	25%	.695	.685	.393	.473
	50%	1.318	1.273	.930	.935
	75%	1.774	1.751	1.419	1.461
	100%	2.106	2.062	1.857	1.963
$\sigma_{bias}=4.0$	0%	-.112	-.035	-.058	.029
	25%	1.012	1.032	.926	.913
	50%	1.964	1.944	1.788	1.876
	75%	2.819	2.737	2.834	2.890
	100%	3.334	3.401	3.871	3.994

However, there is a second dimension to the importance of being selective, the magnitude of its distorting effect. That is, researchers and policy makers are likely to be more interested in how far off the average reported empirical effect might be. How is the magnitude of this publication selection bias and $\hat{\beta}_0$ likely to be related? The magnitude of the FAT-MRA intercept, $\hat{\beta}_0$, is highly correlated with the magnitude of the estimated bias of the average reported estimate. Figure 5 plots the estimated values of the publication bias and the FAT intercept, $\hat{\beta}_0$, for 2,000 replications. Clearly, the estimated FAT coefficient, $\hat{\beta}_0$, is a good indicator of the likely distortion of publication selection. When average values of the publication

bias and the estimated FAT coefficient are used the association only becomes stronger. Thus, the MRA-FAT intercept also captures the importance of being selective when measured by its effects.

Figure 5: Individual values of $\hat{\beta}_0$ and the Estimated Publication Bias ($\sigma_{bias}=1.0$)

Note: R^2 is 95.5% and n=2,000.

Appendix 2: Meta-Meta-Data

Field	k	$\|\hat{\beta}_0\|$	Comp	Field	k	$\|\hat{\beta}_0\|$	Comp
Beta-convergence	48	4.31	0	Immigration & employment	9	1.50	0
Reciprocal trade agreements	75	3.92	0	Participation & productivity	25	1.28	0
Common currency	34	3.85	0	Farmer education & efficiency	27	1.26	0
Tobacco price elasticity	46	3.59	0	Minimum wages	15	1.25	0
Alcohol price elasticity	78	3.13	0	Efficiency wages	14	1.25	0
Profit sharing & productivity	19	3.12	0	Democracy & growth	79	1.17	1
Multinationals & productivity spillovers	16	3.10	0	FDI & economic growth	49	1.17	0
Medicine model (Aid)	16	3.03	0	Immigration & wages	18	1.16	0
beta-convergence – Strobl study	25	2.92	0	Per capita income & allocation of aid	124	1.10	1
Participation & satisfaction	41	2.92	0	CEO pay & firm size	32	1.06	0
Economic freedom & growth	45	2.88	0	Hospital ownership and performance	30	0.94	1
Water price elasticity	64	2.86	0	Ricardian equivalence	28	0.92	1
Immigration & unemployment	10	2.59	0	Good policy model (Aid)	28	0.91	0
Wage curve	17	2.53	0	Population and growth	30	0.88	1
Tobacco income elasticity	35	2.49	0	Growth & allocation of aid	30	0.87	1
Alcohol advertising bans	8	2.48	0	Development aid & growth	68	0.86	1
Business cycle correlations	35	2.40	1	Market orientation and business performance	49	0.84	0
Drug price elasticity	60	2.24	0	Cigarette advertising elasticity	49	0.82	0
Economic freedom & investment	10	2.14	0	Bureaucratic aid Allocation	35	0.79	0
Unions & productivity growth	26	2.03	1	Ownership & productivity	17	0.73	1
Economic reform and growth	43	1.99	0	Population & Aid	94	0.73	1
Development aid & savings	24	1.92	1	Exchange rate variability	49	0.66	1
Corporate social responsibility	82	1.90	1	Downsizing	41	0.63	1
CEO pay-performance	40	1.89	0	Unions & capital formation	11	0.60	1
Price elasticity of beer	95	1.88	0	Unions & productivity levels	77	0.56	1
Inflation and Central Bank independence	59	1.88	0	Natural rate hypothesis	9	0.47	0
R&D spillovers	9	1.83	0	Unions & profits	45	0.44	1
Price elasticity of wine	90	1.81	0	Inflation and voting intentions	44	0.29	1
Price elasticity of spirits	91	1.8	0	Inequality and growth	48	0.22	1
Development aid & investment	29	1.75	1	Board composition and firm performance	66	0.19	1
Airport noise & property values	20	1.66	0	R&D output elasticity	28	0.12	0
Exports & economic growth	76	1.54	0	Board duality and performance	19	0.08	1
Participation & productivity (experimental)	25	1.51	0				

A full reference list of these 65 meta-analyses is available from the authors. k denotes the number of studies included in the original meta-analysis. $\hat{\beta}_0$ is the estimated absolute degree of selectivity. **Comp** is short for **Competition.** If **Com** =1, the literature enables both positive and negative effects. See text for alternative measures of competition and alternative coding of some of the literatures.

[1] Publication selectivity can distort any literature review, be it a traditional narrative review or a quantitative one, meta-analysis (Laird and Mosteller, 1988; Phillips and Goss, 1995; Stanley, 2001). The issue has been of longstanding concern to economists (e.g. Tullock, 1959; Feige, 1975; Lovell, 1983). These concerns are not unique to economics. Researchers and policy analysts in other disciplines, particularly medicine, have been equally, if not more, concerned about publication selection's distorting effects.

[2] As discussed below, symmetry is the key characteristic of an absence of publication selection. This does not mean that all research should be equally publishable. Even when there is great controversy, it is critical for a study to employ appropriate research methodology and valid data. Rigor and the symmetry of reported results are entirely unrelated.

[3] Typically, η is estimated by the weighted average of the results; where the chosen weights are some measure of precision, such as the sample size or the inverse of the variance. However, the distorting effect of selectivity affects all types of averages or summaries, whether weighted or not.

[4] For example due to the 'two-faces view,' the union-productivity literature shows little publication selectivity. Indifference is also possible. Researchers may take no particularly strong position on an issue. For example, a variable can be included in an econometric study purely as a 'control' variable. We group theory agnosticism together with theory competition, as their effect on selectivity should be roughly the same.

[5] In between are various degrees of competition and selectivity. For example, theories can agree on the direction of an effect but disagree on the mechanisms that produce the effect.

[6] In this application, the 'random-effects' estimate is to be preferred over the 'fixed-effects' estimate because there is clear evidence of heterogeneity across studies. See Sutton *et al.* (2000) for a discussion of these weighted averages. The random-effects estimate of studies' medians also gives approximately the same estimate even though it is robust to outliers and extreme values (Rose and Stanley, 2005). Rose and Stanley (2005) focus on each study's average trade effect, in part, because they are less distorted by publication selection.

[7] Medical researchers point out that the estimate of β_1 in $t_i = \beta_0 + \beta_1(1/SE_i) + \varepsilon_i$ provides a correction for publication selection (Sutton et al., 2000; Macaskill et al., 2001; Stanley, 2008). Section 3.1 below discusses this MRA model in greater detail. Of course, a comprehensive analysis of this area of research is more complex than any brief discussion can describe. For example, additional explanatory variables are needed to understand the variation among the reported estimates of the common currency effect. However, doing so also produces corrected estimates of the effect of currency unions much smaller than the reported average effect.

[8] The meta-analyses reported in Section 4 contain literally dozens of examples.

[9] Actually, the left side of this funnel plot is somewhat more sparse than its right side. In some meta-regression models, this funnel asymmetry is statistically significant, which indicates the existence of some selectivity (Doucouliagos, Laroche and Stanley, 2005). Nonetheless, the magnitude of this selectivity, if it exists, is too small to do much harm.

[10] Gamma is the common-currency effect on the logarithm of trade. To convert to the percent increase in trade, calculate: $(e^{\gamma} -1)*100$. A few positive and negative outliers (5%) are trimmed to reveal the shape of the great bulk of this funnel graph. This trimming is done only to allow the reader to see how the vast majority of these estimates are distributed. In all analyses presented here, the full set of available estimates is used. Publication selectivity is so strong in currency-union research that its identification is very robust (Rose and Stanley, 2005).

[11] Of course, it is possible that contestability or the threat of theory competition leads to similar results as actual theory competition. However, the econometric evidence presented below suggests that it is the actual competition (or its absence) that influences selectivity.

[12] By now, there have been many dozens of meta-regression analyses conducted in economics. See Stanley and Jarrell (1989), Stanley (2001), and Stanley (2005b) for primers.

[13] Publication selection is analogous to sample selection biases and produces the conventional 'Heckman regression': $\eta_i = \beta_1 + \rho\sigma I(K_i\hat{\alpha}) + e$, where $I(K_i\hat{\alpha})$ is the inverse Mills ratio, and σ is the standard error of $\hat{\eta}_i$ (Davidson and MacKinnon, 2004, pp.486-88). The inverse Mills ratio, $I(K\hat{\alpha})$, may be interpreted as an omitted variable in the estimation of β_1. 'Fortunately', the estimated standard error of $\hat{\eta}_i$, SE_i, is likely to vary greatly from one study to the next. Like the nonlinearity of the inverse Mills ratio that permits the identification of the Heckman regression in the typical economic application (Davidson and MacKinnon, 2004, p.489), it is this heteroscedasticity of empirical economic estimates that makes possible the identification and estimation of this MRA model of publication selection. Replacing the inverse Mills ratio term with $\beta_0 SE_i$ gives equation (1) (Stanley and Doucouliagos, 2007).

[14] This MRA can also be regarded as the feasible generalized least squares estimator (Davidson and MacKinnon, 2004).

[15] Of course, the true effect, β_1, may itself depend on other factors, Z, which may be modeled by adding these terms to the above meta-regression models. See Stanley and Doucouliagos (2007) for a more detailed discussion and illustration of multivariate FAT-MRA modeling.

[16] In this sense, the absence of theory competition increases research effort. Authors have to try harder to: (1) find statistically significant results; (2) uncover results that are consistent with dominant theory; or (3) get unconventional results published. Furthermore, it is likely that some of this extra effort represents a deadweight loss.

[17] For small effects, the effect size is approximately twice the correlation coefficient (Hunter and Schmidt, 2004), and 0.042 is a very small effect size.

[18] Of course, there are famous counter-examples such as 'Giffen goods,' but nonetheless the law of demand remains a strongly held theoretical expectation.

[19] **Demand** could not be used in this Logit because it perfectly predicts **Competition** =0. But this poses no insurmountable problem. We interpret **Demand** =1 as the most severe form of theory monopoly and include it along with alternative measures of theory competition in our meta-meta-regression models.

[20] Ideally, we would like a panel dataset that traces changes in theory competition and selectivity within and between economic literatures. Such a dataset is, however, unavailable.

[21] The data used to estimate these correlations is described below.

[22] We especially wish to thank Craig Gallett, Patrice Laroche, John List, Simonetta Longhi, Raymond Florax, Jarko Fidrmuc, Iikka Korhonen, Jacques Poot, Marin Gammill, T.D. Stanley, Chris Doucouliagos, Martin Paldam, Jan Babetskii, Nauro Campos, Stian Ludvigsen, Jakob de Haan, Geoff Pugh, Jeroen Klomp, and Andrew Rose for providing their datasets or estimates of equation 2.

[23] **Demand** has a positive effect on publication selectivity because **Demand** is inversely related to **Competition**.

[24] **Competition, EndoComp, Debate,** and **TC index** are all able to explain roughly half or more of this systematic variation in observed publication selectivity. Potential systematic variation is roughly approximated by subtracting the average sampling error variance of $\hat{\beta}_0$ from the observed variance of $\hat{\beta}_0$.

[25] **Demand** is not included along with **TC index** because it is already included in this index.

[26] The lone exception is **AcceptHo**, which is not statistically significant when used together with **TC index.**

[27] The difference among these estimates of the effect of theory competition is within the margin of sampling error, and none are statistically different than -1.0. However, the magnitude of the coefficient on **TC index** is an exception because it measures the magnitude of theory competition as one of four rather than two levels.

[28] When **Demand** =1, **Competition** and **Competition2** are always zero. Adding **Demand**'s coefficient to the estimated intercept in any of the first 4 columns of Table 4 predicts $\hat{\beta}_0$ to be greater than 2. For column 5, the intercept already exceeds 2, and it represents **Demand** =1 and **TC index**=0.

[29] As stated previously, the exception is **AcceptHo,** which is not statistically significant when we use **TC index** as our measure of theory competition.

[30] For example, the M2RA models reported in either column 4 or column 5 of Table (4) explain 73% of the potential systematic variation in observed publication selection severity.

[31] The magnitude of the selection bias is expected to depend on the magnitude of 'true' elasticity relative to the typical variation in the estimates of elasticity. For highly elastic goods that can be estimated accurately, this bias is likely to be smaller.