Meta-Regression Analysis:
Motivation and Introduction

Tom Stanley
Julia Mobley Professor of Economics,
Hendrix College, USA: Stanley@Hendrix.edu
Meta-Analysis Professor, Deakin University

“Meta-analysis refers to the **statistical** analysis of a **large collection** of results from individual studies for the purpose of **integrating** the findings. It connotes a **rigorous alternative** to the casual, narrative discussions of research studies that typify our attempt to make sense of the rapidly expanding research literature.” –Glass (1976)

Part I: Motivation and Overview: Meta-Regression Analysis

Problems of the Social Science Research

"The house of social science research is sadly dilapidated. It is strewn among the scree of a hundred journals and lies about in the unsightly rubble of a million dissertations"

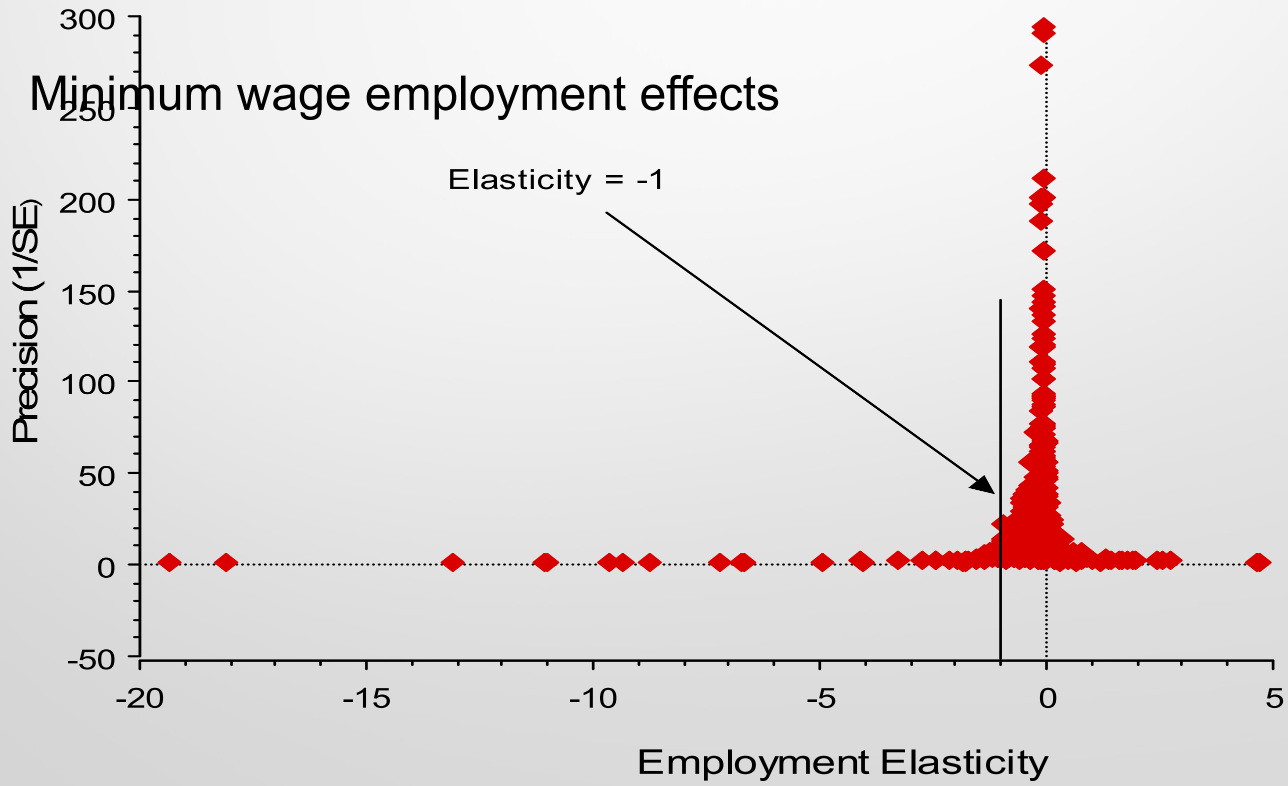
– Glass, McGaw and Smith (1981:11)

Empirical results reported in economics journals are selected from a large set of estimated models. Journals, through their editorial policies, engage in some selection, which in turn stimulates extensive model searching and prescreening by prospective authors. . . . As a consequence, statistical analyses are either greatly discounted or completely ignored.

– Leamer and Leonard (1983: 306)

Conflicting Empirical Findings

- Rarely do single studies provide definitive answers upon which to base policy or to settle theoretical disputes.
- Very large research variation is the norm.



Answer: **Meta-Regression Analysis is Regression of Regression Estimates**

It identifies and filters biases and heterogeneity

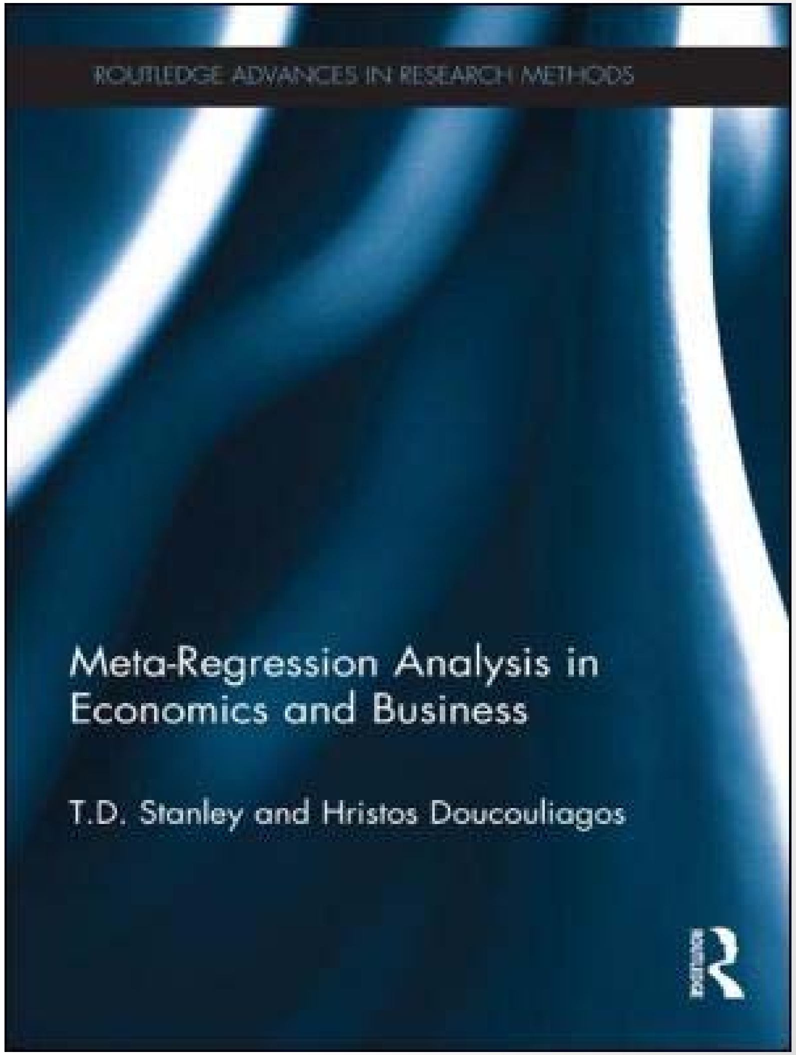
- Stanley, T.D., Jarrell, S.B. (1989) "Meta-regression analysis: A quantitative method of literature surveys." ***Journal of Economic Surveys*** 3: 54-67.
- Stanley, T.D. (2001) "Wheat from Chaff: Meta-Analysis as Quantitative Literature Review," ***Journal of Economic Perspectives***, *15*: 131-50.
- Ioannidis, J.P.A., Stanley, T.D. and Doucouliagos, C. "The power of bias in economics research," ***The Economic Journal***, forthcoming, 2017.

Our Book

"In our view, the central task of meta-regression analysis is to filter out systematic biases, largely due to misspecification and selection, already contained in economics research."

–Stanley & Doucouliagos (2012, p.16) ***Meta-Regression Analysis in Economics and Business.***

Now in Paperback!



Is Meta-Regression Analysis the Answer?

MRA:

- Uses **replicable** methods to identify and code **all** relevant research.
- Employs rigorous statistical techniques to summarize and explain all research results.
- Accommodates and 'corrects' selection, misspecification biases and heterogeneity routinely found among reported findings.

Meta-Analysis of Economics Research- Network (MAER-Net)

Colloquia

- Zeppelin University, Germany (Oct 12-14)
- Hendrix College, 2016
- Prague, CZ—2015
- Athens, Greece (2014)
- Greenwich, UK (2013)
- Perth, Australia (2012)
- Cambridge University, UK (2011)
- Corvallis, Oregon US (09)
- Nancy, France (08)
- Sonderborg, Denmark (07)

"Believing is seeing"

–Demsetz (1974, p. 164).

- In Psychology, this is called: confirmation (or experimenter) bias
- For example, when asked "Are you **happy** with your social life?"—Y(73%); N(22%); Undec(5%)
- When asked: "Are you **unhappy** with your social life?" — Y(65%); N(27%); Undec(8%)
- In economics, ideology and the commitment to ruling theory frames and constrains our research expectations and perceptions.

Believing is Seeing II

Publication Selection Bias

- Reviewers and editors are predisposed to accept papers consistent with the conventional view.
- Researchers may treat the conventionally expected result as a model selection test.
- Everyone may have a preference for ‘statistically significant’ findings.

—Card and Krueger (1995)

Publication Selection Paradox

- Many economists have difficulty seeing that the suppression of positive minimum wage elasticities (or positive price elasticities, or negative values of a statistical life is somehow 'biased.'
- It may be individually rational for economists to 'validate' their estimates by suppressing empirical results that do not 'make any sense.'
- **However**, when many researchers suppress these 'wrong' signs, our aggregate knowledge worsens.→
- **Fallacy of Composition {ecological fallacy}**

Warning! Research findings may appear **larger** than they actually are.

- 10^{+}-fold distortion of the adverse employment effect of minimum wage
- A tripling of the marriage wage premium.
- A 6-fold exaggeration of the value of a statistical life (VSL)

Register or Perish! All top Medical Journals require the prior registration of clinical trials, **and we should all file pre-analysis plans**

Figure 1: Marriage Wage Premium

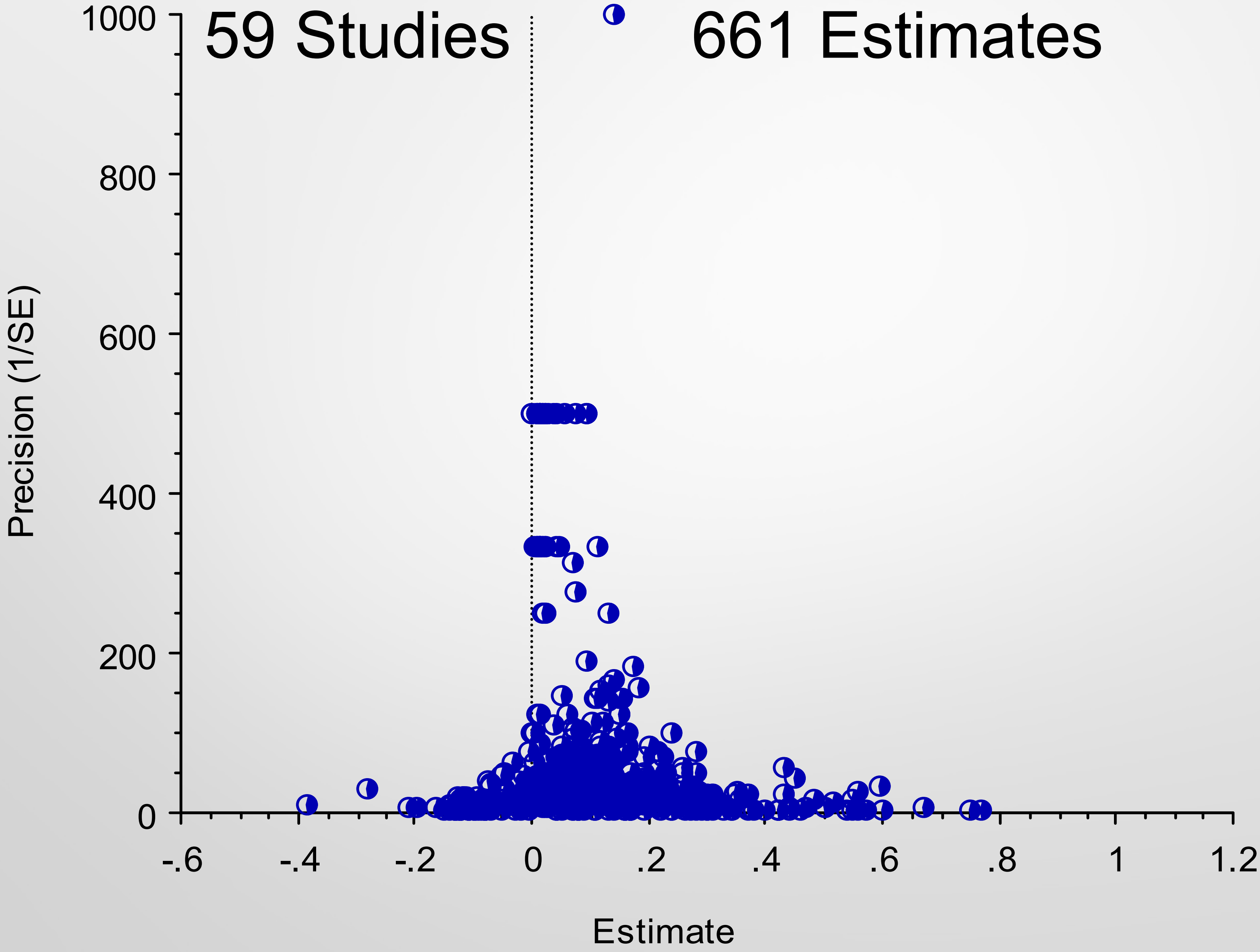


Figure 2: Baby Penalty
48 Studies & 1895 Estimates

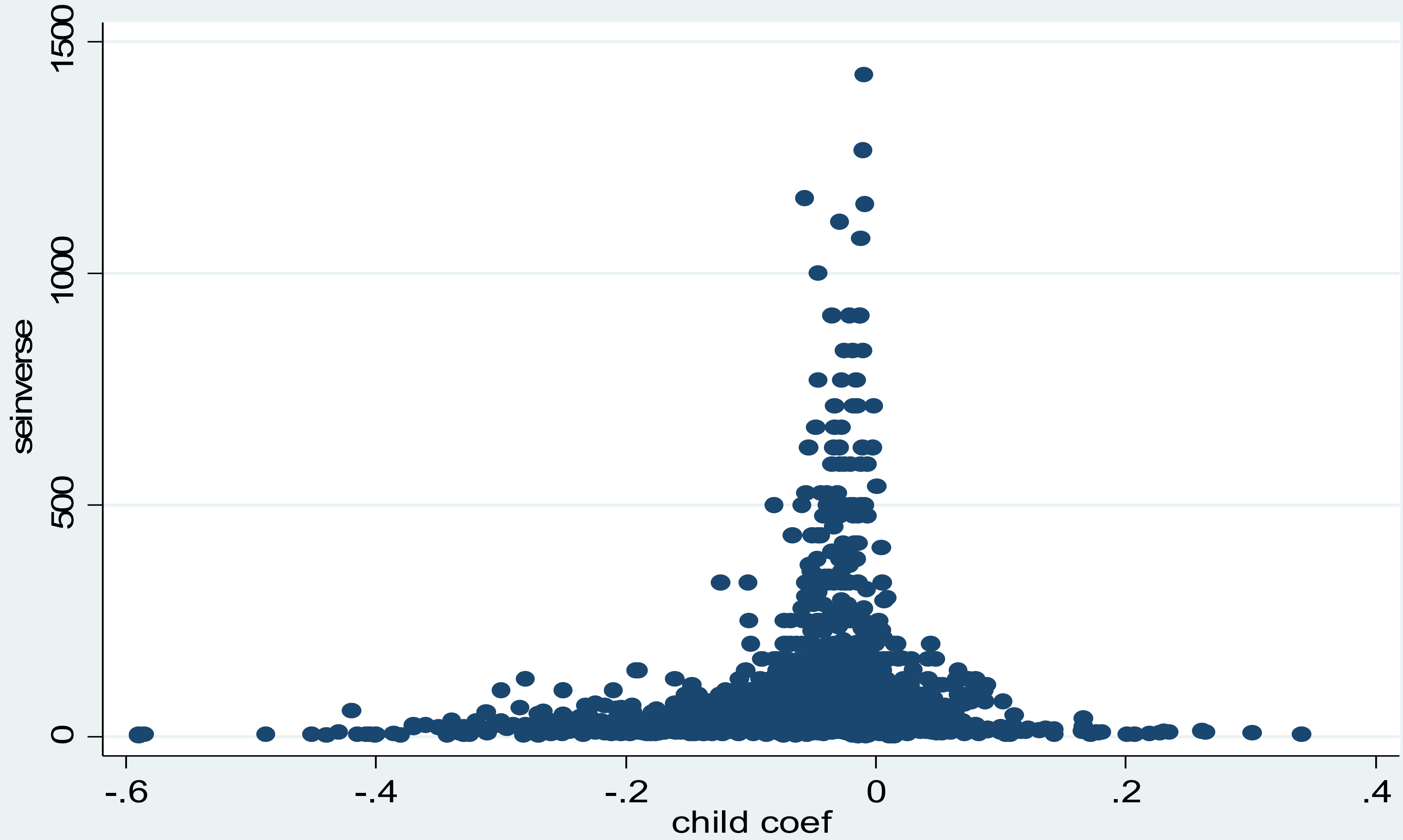
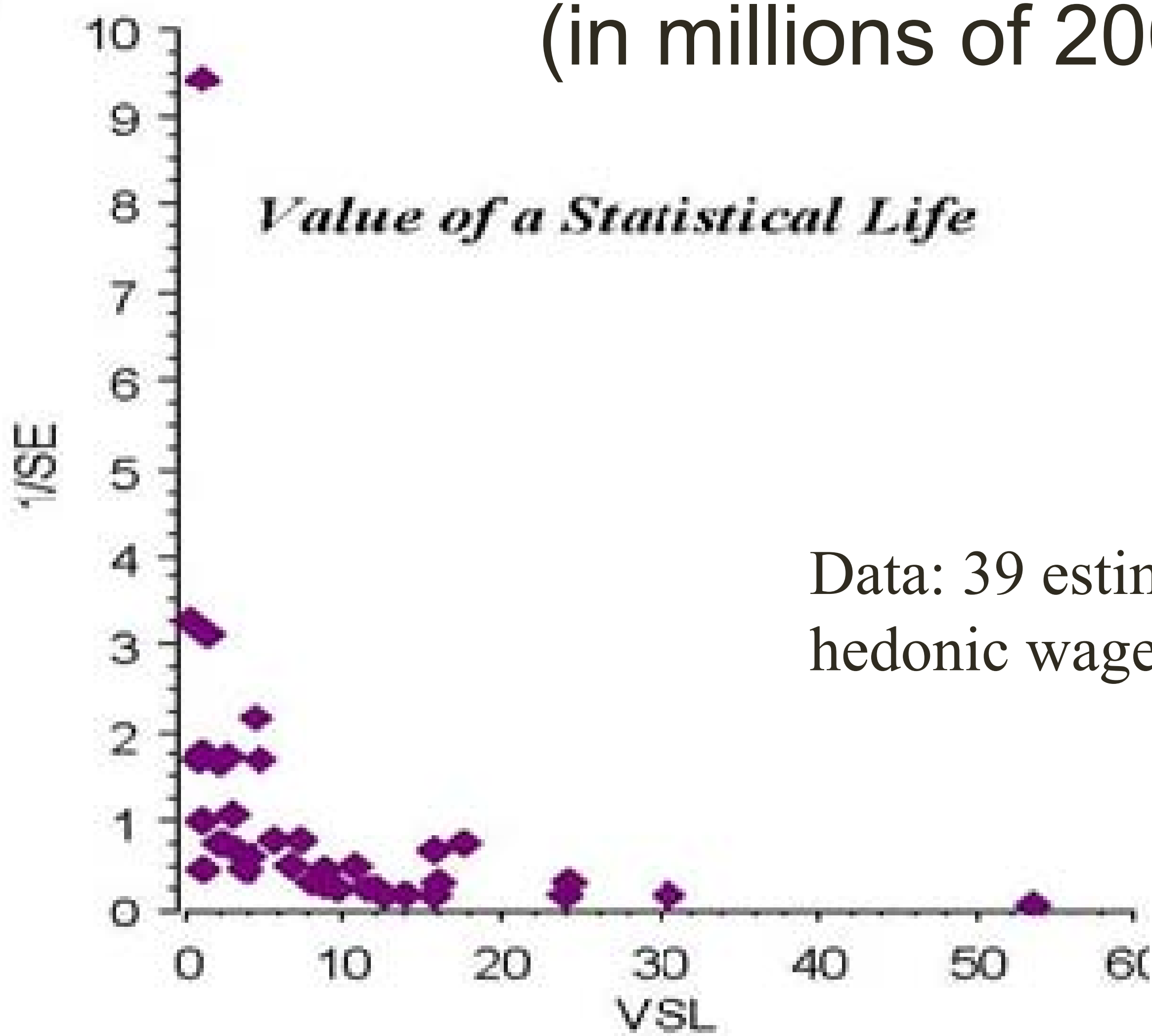


Figure 3: Value of Statistical Life (in millions of 2000 US $s)



Data: 39 estimates of VSL from hedonic wage equations.

Source: Doucouliagos, Stanley, and Giles (2012), *J. Health Econ.*

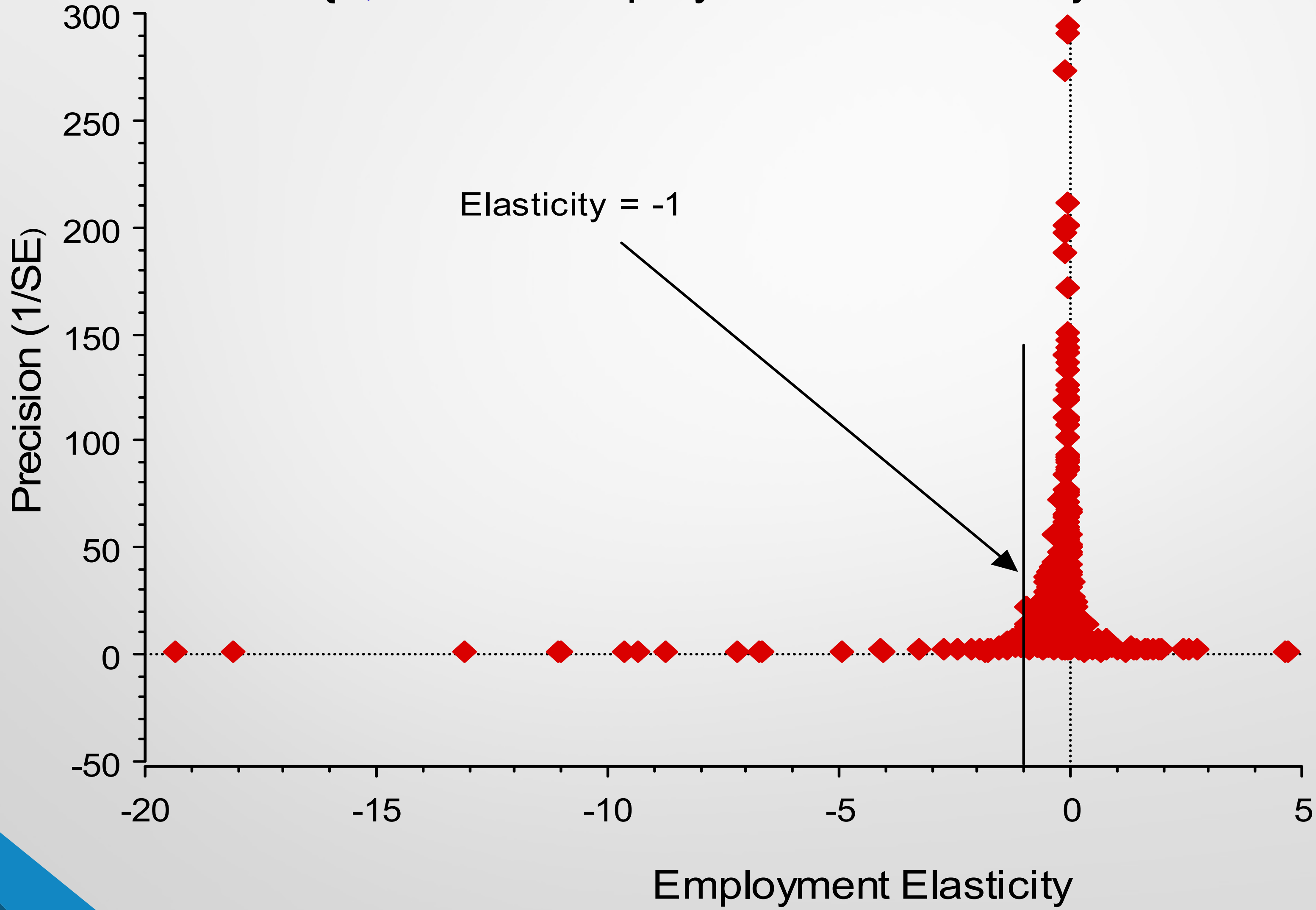
Conventional Economic Theory:

- predicts a strict **LAW of DEMAND** for labor {*i,e*., an inverse relation between wages and employment}.
  - Therefore, **raising the minimum wage** must **reduce employment**.
  - All real economists know that **government intervention** into the free market always has **negative consequences**. . . .
  - Recall the controversy in the 1990s around Card and Krueger's 'Myth and Measurement."

Fig 4: Minimum-Wage Employment Effects:

Economic Report of the President 2013

{**1,474** US employment elasticities}



Source: Doucouliagos and Stanley (2009) Brit. J. of Ind. Relations

Testing Selective Reporting Bias and Identifying Genuine Empirical Effect

FAT-PET-MRA

$$\textit{effect}_j = \beta + \beta_0 Se_j + \varepsilon_j \quad (j=1, 2, \ldots L) \quad \textbf{(1)}$$

{Comment: Always use weighted least squares (**WLS)** with $1/Se^2$ as the weight}

FAT (funnel-asymmetry test): t-test of β_0 is a test for publication bias (Egger et al., 1997)

PET (precision-effect test)**:** t-test of β is a test of authentic effect, beyond publication bias (Stanley, 2005; 2008; S&D, 2014)

FAT-PET-MRA-WLS

Table 1: MRA Tests of PB & Genuine Effect

Estimate	Marriage-Wage Premium	Baby Penalty	Minimum Wage	VSL
Se (β_0)	2.02 (6.99)	-1.05 (-8.27)	-1.6(-4.49)	3.20 (6.67)
Intercept β	0.037 (17.2)	-0.024(-30.85)	-.01 (-1.09)	0.81 (3.56)
Ave. Effect	0.121	-0.053	-0.19	$9.5 mil
WAAP	0.042	-0.027	-0.013	$1.47 mil
PEESE	0.044	-0.028	-.036	$1.67mil
n	661	1,895	1,474	39

What about Heterogeneity?

$$effect_i = (\beta + \Sigma\alpha_k Z_{ik}) + (\beta_0 Se_i + \Sigma\gamma_j K_{ij} Se_i) + \varepsilon_i \quad (3)$$

{Heterogeneity} {Publication Bias}

- Reported research results are likely to be dependent upon any number of other factors
  - Misspecification biases, poor designs, flawed methods, confounding stimuli, etc.
  - Genuine differences caused by variations in culture, country, stage of development, etc.

Enter Meta-Regression Analysis

Table 2: Minimum-Wage Moderator Variables

22 Z-var's +
22 K-var's =
44 Var's

K & Z Variable	Definition	Mean (standard deviation)
Elasticity	is the estimated effect, the dependent variable, and an elasticity	-1.69 (2.83)
Se	is the elasticity's standard error	22.76 (28.32)
Panel	=1, if estimate relates to panel data with time-series as the base	0.45 (0.50)
Cross	=1, if estimate relates to cross-sectional data with time-series as the base	0.13 (0.34)
Adults	=1, if estimate relates to young adults (20-24) rather than teenagers (16-19)	0.14 (0.35)
Male	=1, if estimate relates to male employees	0.07 (0.26)
Non-white	=1, if estimate relates to non-white employees	0.05 (0.22)
Region	=1, if estimate relates to region specific data	0.10 (0.30)
Lag	=1, if estimate relates to a lagged minimum wage effect	0.13 (0.34)
Hours	=1, if the dependent variable is hours worked	0.07 (0.25)
Double	=1, if estimate comes from a double log specification	0.42 (0.49)
AveYear	is the average year of the data used, with 2000 as the base year	-19.17 (11.90)
Agriculture	=1, if estimates are for the agriculture industry	0.01 (0.11)
Retail	=1, if estimates are for the retail industry	0.08 (0.27)
Food	=1, if estimates are for the food industry	0.13 (0.34)
Time	=1, if time trend is included	0.37 (0.48)
Yeareffect	=1, if year specific fixed effects are used	0.30 (0.46)
Regioneffect	=1, if region/State fixed effects are used	0.34 (0.47)
Un	=1, if a model includes unemployment	0.56 (0.50)
School	=1, if model includes a schooling variable	0.15 (0.35)
Kaitz	=1, if the Kaitz measure of the minimum wage is used	0.40 (0.49)
Dummy	=1, if a dummy variable measure of the minimum wage is used	0.17 (0.38)
Published	=1, if the estimate comes from a published study	0.85 (0.35)

Table 3: Multiple Meta-Regression of Minimum-Wage Employment Elasticities—General-to-Specific

Heterogeneity

Doucouliagos and Stanley (2009), *British Journal of Industrial Relations.*

Publication Selection

Variables:	*Cluster-Robust*	*FE-Panel*
Genuine empirical effects (Z-variables)		
Intercept (β)	0.120 (4.39)	0.102 (6.04)
Panel	-0.182 (-4.72)	-0.149 (-10.5)
Double	0.064 (3.20)	0.041 (5.42)
Region	0.040 (0.92)	0.090 (6.34)
Adult	0.024 (2.68)	0.021 (3.72)
Lag	0.026 (1.60)	0.010 (1.59)
AveYear	0.004 (4.34)	0.003 (6.38)
Un	-0.042 (-3.04)	-0.042 (-5.79)
Kaitz	0.052 (3.06)	0.032 (3.88)
Yeareffect	0.069 (1.98)	0.067 (7.44)
Published	-0.041 (-2.69)	-0.037 (-4.89)
Time	-0.022 (-2.08)	-0.017 (-2.46)
Publication bias (K-variables)		
Se (β_0)	-0.359 (-0.11)	-1.374 (-5.94)
Double x Se	-1.482 (-3.23)	-1.073 (-3.90)
Un x Se	-0.840 (-1.87)	1.164 (3.08)

Robustness, Robustness, Robustness

- Focus only on those findings that are robust to any reasonable variation in your methods. Include only these robust findings as part of your ‘Results.’

Robustness, 'Prediction' and Implications

- Many sets of robustness checks were investigated; all confirming our findings.
- Once publication selection is accommodated, **No** evidence of an **adverse employment effect remains**
- Substituting any defensible notion of 'Best Practice Research' into the estimated multivariate MRA coefficients finds: **No support** for a **practically significant adverse employment effect!**

Replicated by 2 Independent Teams &

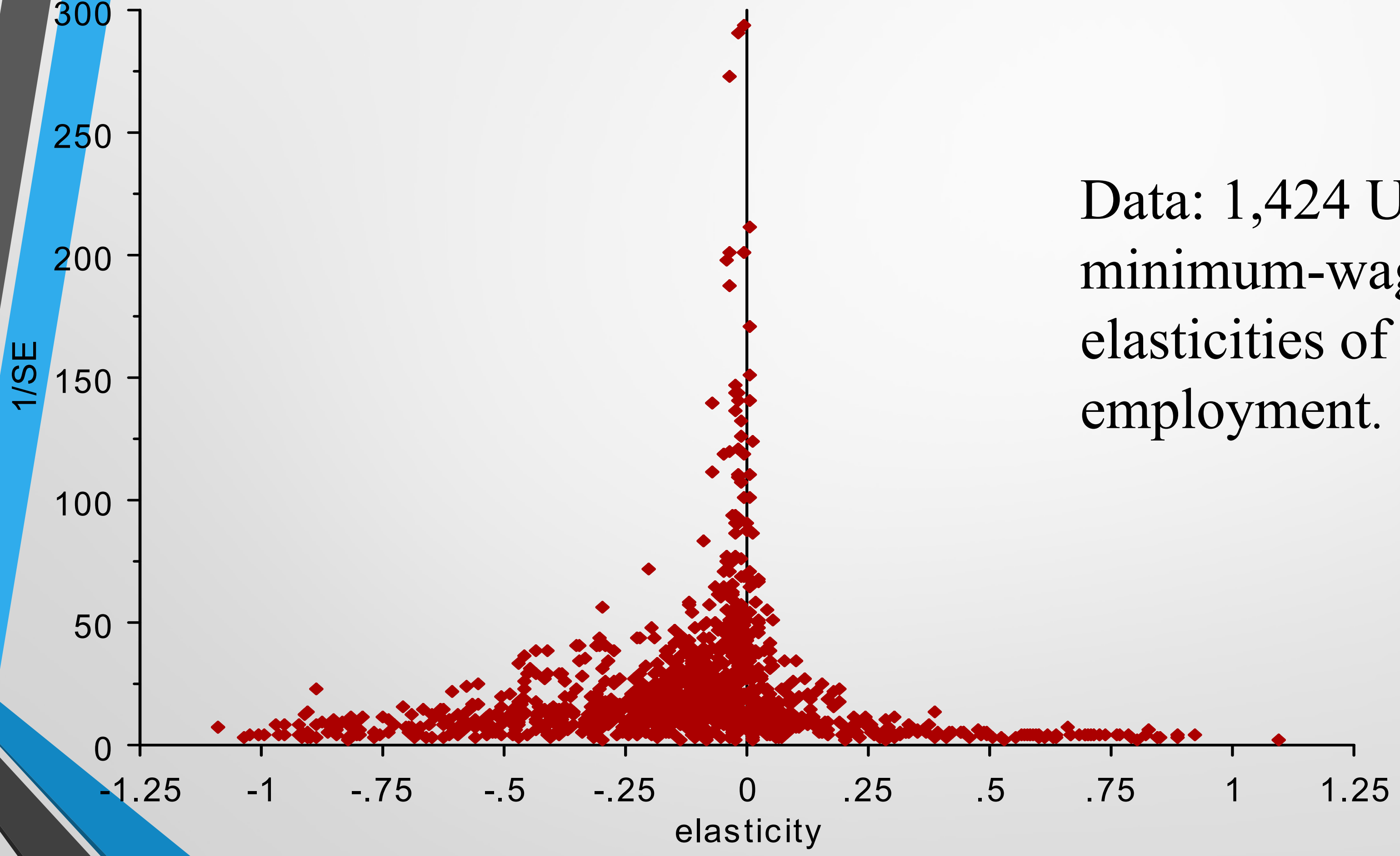
A Meta-Analysis of **UK's Minimum Wage**
Leonard et al. (2014), "Does the UK minimum wage reduce employment? *Brit J. of Industrial Relations*

Followed MAER-NET (2013) guidelines

- Confirmed our US findings
  - No evidence of adverse employment
  - But no evidence of publication bias in UK

Minimum Wage Employment Elasticities

{trimmed by |Elasticity|<1.1}



Data: 1,424 US minimum-wage elasticities of employment.

Source: Doucouliagos and Stanley (2009) Brit. J. of Ind. Relations

Part II: Conceptualization, Literature Search and Coding

Robert J. Johnston
George Perkins Marsh Institute

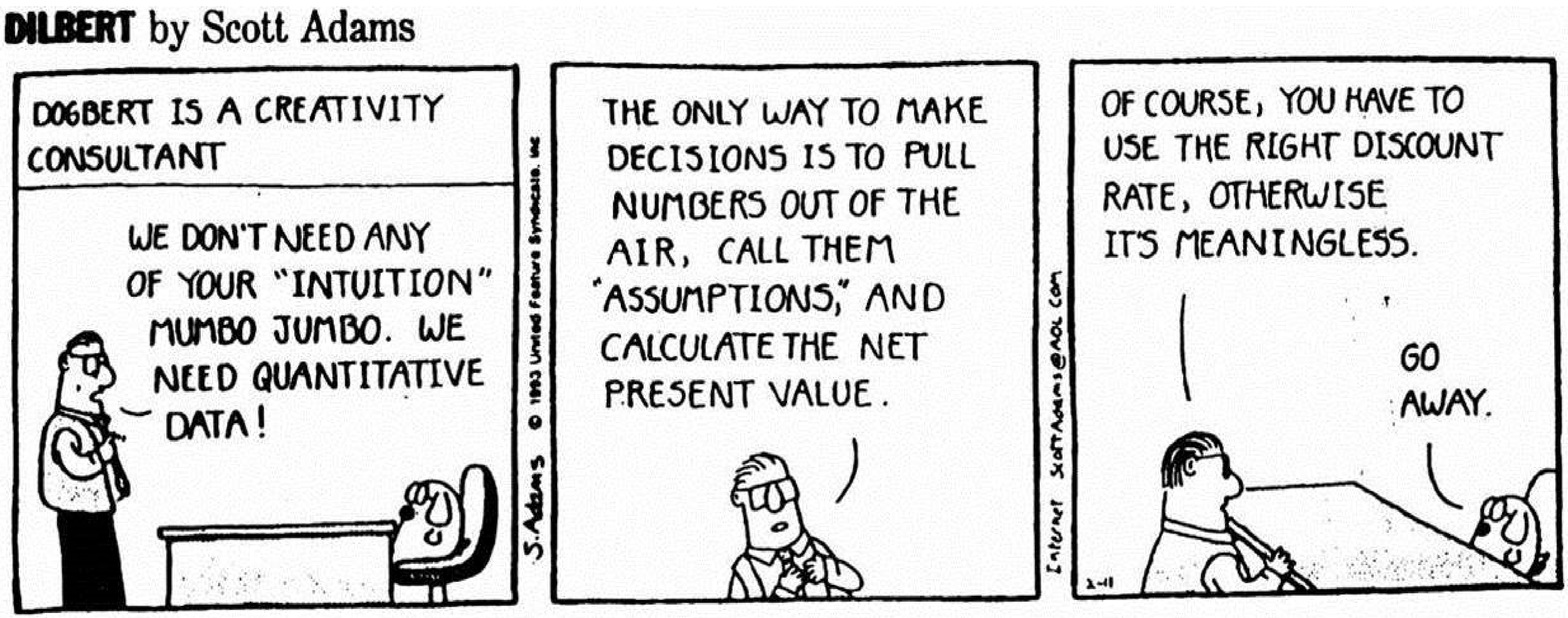
February 21, 2017

CLARK UNIVERSITY

Meta-Analysis is About the *Data*!

- The validity of any meta-analysis depends on what happens *before* econometric models are specified.
- Econometrics have limited capacity to correct for a poorly conceptualized research question or bad data.
- Economists are trained in econometrics, but often given minimal training in metadata development.
- The quality and relevance of the metadata depend on:
  - A well-defined research question
  - Protocols for the literature search
  - Protocols for identifying and the coding the data
  - Methods for data inspection and summary

Models Are Only As Good as the Data



Defining the Question

- Meta-analysis begins with a definition of the question to be answered.
  - Grounded in economic theory and the literature.
  - Includes *subject*, *intervention*, and *outcome*.
    - Subject: What is the unit or group under study?
    - Intervention: What is the action being studied and the default?
    - Outcome: What is the change being evaluated (the *effect size*)?
  - Must be answerable in scientific and empirical terms.
- Search for *prior meta-analyses* on related topics, if any exist.
- Will your analysis extend/improve these prior analyses?

Defining Scope

- How broadly or narrowly defined are your subject, intervention and outcome (effect size)?
- Broad scopes may increase perceived relevance and possible sample sizes, but increase heterogeneity in your data.
  - What is the price elasticity of residential water demand in the Eastern US? (narrower)
  - What is the price elasticity of residential water demand across developed countries? (wider)
- There are always tradeoffs when defining scope.
  - Larger N (desirable) versus more heterogeneity (generally undesirable).

An Example: Willingness to Pay per Fish (Johnston et al. 2006)

- Question: What is the willingness to pay (WTP) per marginal fish caught by US recreational anglers?
  - Subject: Recreational anglers in the US
  - Intervention: Marginal change in catch (one fish)
  - Effect Size: Marginal WTP/fish
- Narrower scope: What is the willingness to pay (WTP) per marginal fish caught by US recreational *fresh water* anglers?
- Which is a better scope for meta-analysis?
- There is an emerging literature on "optimal scope" for economic meta-analysis (Moeltner and Rosenberger 2014; Johnston and Moeltner 2014).

Optimal Scope in Recreational Fishing Data: Johnston and Moeltner (2014; *ERE*)

Table 4: Fish data esti

	prob. of pooling			
	round 1		round 2	
meta-category	overall	within	overall	within
Saltwater, big game				
hicks	0.07	-	0.84	-
Saltwater, small game				
marsh.	0.91	0.04	-	
hicks	0.05	-	0.15	-
Freshwater, big game				
hicks	0.15	-	0.90	-
Freshwater, small game				
hicks	0.91	-	-	-

Characteristics of the Literature

- Question definition (including scope) depends on the characteristics of the literature.
- How extensive and heterogeneous is the literature on your research question?
  - Are there a large number of studies testing the same hypothesis across comparable subject groups?
  - Are effect sizes comparable and measured in the same way?
  - Do studies provide more than one observation on the research question that can be used for MRA?
- Smaller literatures generally require broader scopes, given sample size requirements for MRA.

Data Reporting and Sample Size

- Data reporting is an important determinant of question definition and metadata construction (e.g., Loomis and Rosenberger 2006).
- How extensive is data reporting in the literature being studied?
- Any variable or effect size included in an MRA must be present for *all included studies and observations*.
- Incomplete data reporting often implies that increases in sample size for MRA come at a cost of reductions in the set of moderator variables that can be included.
- Moeltner et al. (2007) refer to this as the "N vs. K" problem.
- Accounting for (more) heterogeneity requires larger K!

Data Reporting and Sample Size

- Consider the types of data that will be necessary or desirable for your particular MRA. Are they available?
  - Effect size and measures of precision?
  - Variables necessary to account for heterogeneity?
- The available data determine the type of meta-analysis that can be conducted.
- Detailed understanding of the literature can prevent wasted effort in subsequent stages of the meta-analysis.

Identifying and Coding the Data

- Data identification and coding is among the most important steps of a meta-analysis.
- Stanley et al. (2013) provide guidelines for research reporting in MRA, including reporting on search protocols.
- Also see Chapter 2 in Stanley and Doucouliagos (2012).
- Prior, high quality meta-analyses can also be used as a starting point.


JOURNAL OF ECONOMIC SURVEYS

doi: 10.1111/joes.12008

META-ANALYSIS OF ECONOMICS RESEARCH REPORTING GUIDELINES

T.D. Stanley
Hendrix College

Hristos Doucouliagos
School of Accounting, Economics and Finance, Deakin University

Margaret Giles
Centre for Innovative Practice, Edith Cowan University

Jost H. Heckemeyer
Business School, University of Mannheim

Robert J. Johnston
George Perkins Marsh Institute and Department of Economics, Clark University

Patrice Laroche
CEREFIGE, Université de Lorraine

Jon P. Nelson
Pennsylvania State University

Martin Paldam
Århus University

Jacques Poot
NIDEA, University of Waikato

Geoff Pugh
Staffordshire University

Journal of Economic Surveys (2013) Vol. 27, No. 2, pp. 390–394

Identifying Relevant Studies

- Meta-analyses require systematic and transparent literature review.
  - Conducted on the *full population* of studies that satisfy a set of search criteria.
- Identifying the population of studies is not trivial.
  - How to identify gray (unpublished) literature?
  - Narrowing/restricting scope (e.g., by time period, study method, subject population).
  - Random samples of the literature can be used if the literature is very large (this is uncommon).
  - Can start with existing reviews or meta-analyses.

Search Protocols and Selection Criteria

- The literature search should always be guided by an *explicit set of search protocols and selection criteria*.
- There are many published sets of guidelines for literature searches and systematic reviews (e.g., Stroup et al. 2000; Higgins and Green 2008; Campbell and Cochrane Collaborations; MAER-Net).
- Search protocols should be documented clearly
  - Search engines
  - Search keywords
  - Treatment of peer reviewed versus gray literature.
  - Screening criteria and quality controls

Search Protocols and Selection Criteria

- Search protocols should be documented clearly
  - Required effect sizes
  - Other minimum data reporting required (e.g., key variables)
  - Other methods (beyond automated searches) used to identify literature.
    - Past meta-analyses, review of bibliographic listings from identified articles, contacts with key researchers, etc.
- Reporting on search protocols should also state:
  - Date(s) when search was conducted
  - Who conducted the search and coding (2+ individuals).

What Type of Studies to Include?

- *In general*, it is safest to begin with a comprehensive set of all studies that satisfy the search criteria.
  - Published studies (peer reviewed, e.g., journal articles)
  - Published studies (non-reviewed, e.g., most book chapters)
  - Government and consulting reports
  - Dissertations / theses
  - White, conference and working papers
- Differences can then be coded and tested in MRA.
  - Variables can be coded for study type & quality indicators.
- Ex ante exclusions risk the loss of relevant data.

Published versus Unpublished Studies

- "In our experience, there is … no detectable difference in quality between published and unpublished papers as measured by the objective statistical criteria of precision" (Stanley and Doucouliagos 2012, p. 18).
- Many studies find no difference between patterns identified by published and unpublished papers.
- Some studies find different patterns associated with published versus unpublished papers. A few examples include:
  - Minimum wage effects (Doucouliagos and Stanley 2009).
  - Value of Statistical Life (Kluve and Schaffner 2008).
  - Value of water quality (Johnston et al. 2016).

The Quality Conundrum

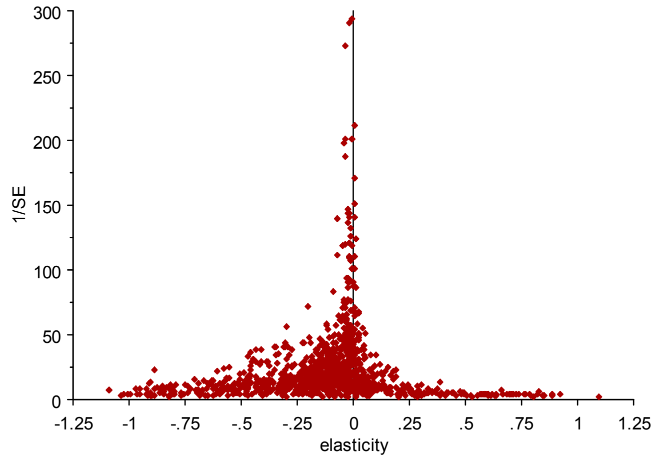
- There is sometimes disagreement over whether studies should be screened for quality.
- Should "low quality" studies be excluded from the metadata?
- It is appropriate to establish screens for *minimally acceptable methods*, where these exist. (But where to draw the line?)
  - There are some methods that may be deemed unacceptable by nearly universal consensus of the literature in question.
- It is *not appropriate* to screen for quality based on *whether the study obtains an expected result*.
  - Doing so introduces additional selection bias.
  - <u>*An "expected result" is not an indicator of quality*</u>.

The Quality Conundrum

- Risks of discarding low quality studies often outweigh benefits.
- "Studies should be omitted only by objective criteria that are applied evenly across the literature" (Stanley 2001, p. 147).
- Statistical precision is a key and objective quality measure, and is used to weight the observations (not screen studies).
- Other common quality measures have shortcomings.
  - Published versus unpublished papers (see above).
  - Journal reputation or journal/paper impact factor.
- *"From a pure statistical point of view, we should include all relevant estimates and code for all observable differences in quality and methods"* (Stanley and Doucouliagos 2012, p. 35).

The Selection Paradox

- But why is it not okay to eliminate studies that produce seemingly erroneous findings from metadata (e.g., negative value of statistical life)?
- The answer is the *Selection Paradox*, or fallacy of composition.
- Due to expected statistical variation (or imprecision), some effect sizes will naturally have unexpected sizes or signs.
- This is shown by a standard funnel plot.
- Eliminating such findings leads to bias when viewing the literature as a whole (e.g., calculating the mean effect size).
- This is the core of publication bias.



Trimmed Minimum Wage Employment Elasticities (Doucouliagos and Stanley 2009)

Choosing an Effect Size

- An *ideal* effect size should measure an theoretically meaningful effect thought to be conditionally invariant.
  - Clear economic interpretation.
  - Conditionally invariant—it makes theoretical sense to compare the effect across studies and contexts.
  - Akin to a partial derivative—empirical effect of A on B holding all other variables constant.
  - Comparable within and between different studies (Stanley and Jarrell 1989; Becker and Wu 2007).
- Common examples of effect size: elasticities, welfare measures, partial correlations, t-statistics.

Notes on Partial Correlations

- Partial Correlations (-1 to +1): $r = {}^{t}/_{\sqrt{t^2+df}}$
  - Unitless and readily calculated; enables large sample sizes. Holds other variables constant. Standard error readily calculated ($\sqrt{(1-r^2)/df}$).
  - Distribution is not normal close to -1 and +1 (Fisher's *z*-transform is an alternative).
  - Statistical rather than economic measure. It can be useful to supplement with an analysis of the economic effect.
    - Example: Doucouliagos and Laroche (2009) meta-analyze the effect of unions on profits (N=45; partial correlation).
    - Percentage reduction in profits at average unionization calculated for N=12 studies (economic effect of interest).

Notes on T-Statistics

- Directly comparable across studies and readily calculated.
- Like partial correlations, t-statistics are a statistical rather than economic measure.
- Economic interpretations can be unclear.
- Direct use of a t-statistic as the dependent variable can complicate interpretation—take care!
  - Weighting by the standard error (as traditionally done in MRA) transforms the dependent variable into a t-statistic.
  - Simultaneously transforms RHS variables.
  - Using a t-statistic as the LHS variable *without* the RHS transform leads to a *different interpretation* of the MRA.

Notes on Elasticities

- Widely used measure of economic effect with clear interpretation: The % effect on Y caused by a % increase in X.
- Common effect size in meta-analysis.
- Readily calculated from statistical regression results. Ease of calculation and data needs depends on functional form.
  - For example, means of LHS and RHS variables are required for specifications other than double-log.
- For functional forms other than double-log, standard errors must be calculated (e.g., Delta or Fieller methods).
  - Alternative measures of precision can sometimes be used.
- Care should be taken to ensure comparable elasticity measures (e.g., short-run versus long-run elasticities) or code differences.

Reconciling Effects Across Studies

- In many cases, not all studies will report directly comparable effect sizes.
- Transformations and researcher judgment is required.
- How "close" is "close enough" to be an identical effect size? This relates to the issue of scope.
- How can (or should) effect sizes be reconciled across different studies (Nelson and Kennedy 2009; Rolfe et al. 2015; Smith and Pattanayak 2002)?
- Particularly challenging for some types of applications, e.g., environmental economics.

Chapter 16
Meta-analysis: Rationale, Issues and Applications

John Rolfe, Roy Brouwer and Robert J. Johnston

Abstract This chapter reviews the key reasons for using meta-analysis for benefit transfer and provides an illustrative case study application. The case study involves a meta-analysis of values for improved river health in Australia from 2000 to 2009. To minimize potential problems of commensurability and methodology, we restrict the analysis to consider only values drawn from choice experiments. Different measures and scales of river health across studies were reconciled by transforming implicit prices into a comparable standard of willingness to pay (WTP) per kilometer of river in good health. Ordinary least squares and random effects meta-regression models were used to identify systematic relationships between the dependent variable (WTP/km) and explanatory variables characterizing sites, populations, affected resources, and primary study methodology. The case study illustrates both advantages and challenges involved in the application of meta-analysis to benefit transfer.

Keywords Benefit transfer · Choice experiments · River health · Willingness to pay · Meta-regression analysis

J. Rolfe (✉)
School of Business and Law, CQUniversity, Rockhampton, QLD, Australia
e-mail: j.rolfe@cqu.edu.au

R. Brouwer
Department of Environmental Economics, Institute for Environmental Studies, Vrije Universiteit, Amsterdam, The Netherlands
e-mail: roy.brouwer@vu.nl

R.J. Johnston
Department of Economics and George Perkins Marsh Institute, Clark University, Worcester, USA
e-mail: rjohnston@clarku.edu

 357

R.J. Johnston et al. (eds.), *Benefit Transfer of Environmental and Resource Values*, The Economics of Non-Market Goods and Resources 14,
DOI 10.1007/978-94-017-9930-0_16

Reconciling Effects Across Studies

- Example: Commodity consistency in MRA of willingness to pay (WTP) for non-market goods and services.
- Consider WTP for water quality changes (Johnston et al. 2003, 2005, 2016; Van Houtven et al. 2007).
- Primary studies measure WTP for various measures of water quality change. Examples include:
  - Water quality index (0-100) or ladder (0-10) units.
  - Use suitability (boatable, fishable, swimmable).
  - Pollutants or chemical properties (dissolved oxygen).
  - Clarity (Secchi depth)
  - Effects on sensitive species (e.g., able to support trout).

Reconciling Effects Across Studies

- To what extent can these distinct quality measures be transformed to a comparable metric of water quality change, for which WTP can be calculated?
- For example, Johnston et al. (2003, 2005, 2016) map water quality measures to a comparable water quality ladder or index units, then calculate WTP/unit.
  - Some transformations are straightforward (e.g., use suitability or pollutant concentrations to water quality index).
  - Others require stronger assumptions and should likely be avoided (e.g., relationships between clarity and water quality index depend on many generally unknown variables).

An Example of Reconciled Measures: The Water Quality Index

Pollutant	Unit	Freshwater WQI Weight	Estuarine WQI Weight
Dissolved Oxygen (DO)	Mg/L	0.24	0.26
Fecal Coliform (FC)	Lbs/100mL	0.22	0.25
Total Nitrogen (TN)	Mg/L	0.14	0.15
Total Phosphorous (TP)	Mg/L	0.14	0.15
Total Suspended Solids (TSS)	Mg/L	0.11	0.11
Biochemical Oxygen Demand (BOD)	Mg/L	0.15	
Chorophyll-a (ChA)	µg/L	---	

$$WQI = \prod_{i=1}^{6} Q_i^{W_i}$$

WQI Value	Water Quality Classification
95	Drinking without treatment
70	Swimming
50	Game fishing (food)
45	Rough fishing (non-food)
25	Boating

Parameter	Value	Subindex
DO	DO ≤ 3.3	10
	3.3 < DO < 10.5	-80.29 + 31.88*DO – 1.401*DO^2
	10.5 ≤ DO	100
FC	FC ≤ 50	98
	50 < FC ≤ 1600	98 * exp(-0.00099178*(FC–50))
	1600 < FC	10
TN	TN ≤ 3	100 * exp(-0.4605*TN)
	3 < TN	10
TP	TP ≤ 0.25	100 – 299.5*TP – 0.1384*TP^2
	0.25 < TP	10
TSS	TSS ≤ 28	100
	28 < TSS ≤ 168	158.48 * exp(-0.0164*TSS)
	168 < TSS	10
BOD	BOD ≤ 8	100 * exp(-0.1993*BOD)
	8 < BOD	10
ChA	ChA ≤ 40	100 * exp(-0.05605*ChA)
	40 < ChA	10

Coding

- Coding requires expertise and judgement. There is almost always ambiguity.
- As discussed above, LHS and RHS variables may require transformations to be comparable across studies.
- Some variables will be missing (or must be approximated or proxied) for some observations.
- Data coding must balance the need for data and sample size with risks of measurement error.
- Transparency is critical—all coding protocols and assumptions should be documented clearly and should be replicable.
- Stanley et al. (2013) provide concise guidelines. Stanley and Doucouliagos (2012) provide a broader introduction to coding.

Coding

- In general, research assistants should not be responsible for coding, and should not establish protocols.
- At a minimum, protocols should be established and coding should be verified by the principal researcher and subject matter expert (likely you).
- As a general rule, at least two coders should be involved to validate the coding.
- All assumptions should be documented.

The Data

- Data can be categorized depending on its importance to the central question.
- Familiarity with the literature and theory is necessary.
  - Essential data: The effect size, measures of precision as appropriate, study/author identification, bibliographic data.
  - Desirable data: Explain heterogeneity in the effect size. Most important are core variables suggested by theory.
    - Examples: Types of data used, study methodology, characteristics of the study sample or population (income), contextual variables, regional identifiers, geospatial variables, transformations made during coding, omissions of original study.

Missing Data

- Data that are missing from a document (e.g., published article) can often be obtained via alternative approaches.
  - Online appendices or accompanying technical manuscripts.
  - Direct contacts with the authors (e.g., to obtain raw data or supplementary results).
  - Imputing / approximating results from graphs.
  - External data sources (e.g., Census data for information on studied populations; geographic information system data layers for data on land use/land cover).
- The N versus K problem—sometimes it is better to drop a study that is missing too many key variables.

Multiple Estimates from Each Study

- Among the coding decisions that must be decided is the treatment of multiple estimates from each study.
  - Best-set: One "best" estimate from each study. But, this can magnify publication bias.
  - Average-set: Weighted average of all estimates from each study. But, discards information on within-study variation.
  - All-set: All estimates from each study. Often a superior approach, but should include statistical methods that account for within-study correlation.
    - This gives studies with more estimates greater weight in the analysis, *ceteris paribus*. Should weighting be used to account for this (Nelson and Kennedy 2009)?

Multiple Estimates from Each Study

- Independent-set: All *conceptually independent* estimates from each study (e.g., same authors but different data; same data but different authors).
  - But requires researcher to determine “sufficient” independence.
- Practice has evolved to favor all-set estimates (or perhaps independent-set).
- Average-set estimates are a natural extension of all-set data.
- Best option: Code for all-set, with independent variables to identify (a) the originating study, (b) best- and independent-set features.

Summarizing the Data: First Steps

- Once the data are coded (but before MRA), it is important to inspect and summarize the data.
- This is useful to visualize initial patterns and diagnose possible coding errors.
- In addition to standard approaches (e.g., statistical moments, outlier analyses, box-whisker plots), funnel graphs are a central part of initial data inspection.
  - Provides initial perspective on possible publication bias.
  - Can help detect possible coding errors.
  - Can assist in the visual identification of outliers and leverage points for subsequent evaluation.
- Chronological ordering can provide additional insight.

Descriptive Statistics: FEE and REE

- Initial data analysis often includes a simple precision-weighted average effect size.

$$\hat{\mu} = \frac{\sum_i w_i (effect)_i}{\sum_i w_i}$$

- This is statistically preferable to an unweighted average.
- Optimal weights (w_i) are the inverse of estimates' variances, but other weights can be used if necessary (e.g., sample size).
- Fixed-Effects Estimator (FEE) assumes estimates drawn from the same population with common mean: $w_i = {}^1/_{\sigma^2}$.
- Random-Effects Estimator (REE) assumes estimates drawn from several populations with different means: $w_i = {}^1/_{\sigma^2+\tau^2}$.
- τ^2 measures between-study variance.

Descriptive Statistics: FEE and REE

- FEE and REE *only* provide unbiased effect sizes in the absence of publication bias.
- Even in the presence of heterogeneity, FEE is less biased than REE when publication bias is present.
- Hence, prior to drawing conclusions from either FEE or REE, publication bias should be tested formally.
- Stanley and Doucouliagos (2015) show that Weighted Least Squares (WLS) is at least as good as both FEE and REE under a wide range of conditions: $t_i = \beta({}^{1}/_{SE_i})$.
- In general, MRA is a superior option. But data summary is an important preliminary step.

Testing for Heterogeneity

- Heterogeneity implies that any average effect size will not fully reflect the economic phenomenon under evaluation.

A Practical Approach

- Estimate WLS: $t_i = \beta({}^{1}/_{SE_i})$. The sum of squared errors is the χ^2 distributed Q-test statistic for heterogeneity, with N-1 degrees of freedom.
  - Homogeneity is almost *always* rejected (Stanley and Doucouliagos 2012).
- Proceed with MRA using coded moderator variables on the RHS to explain observed heterogeneity.
- Random-, fixed-effects or multi-level MRA can capture additional heterogeneity not explained by coded variables.

Conclusion (Part II)

- Development of the metadata is the most difficult component of meta-analysis, and can be subject to unseen errors.
- No statistical method can fully overcome bias caused by a poorly conceptualized research question, ambiguous definition of effect sizes, or incomplete/erroneous coding.
- Transparency in literature search and coding is critical.
- Data inspection and summary, including formal testing for heterogeneity, is a critical initial step.
- Beware of naïve interpretations of weighted averages (FEE and REE)—WLS is almost always more informative.
- Heterogeneity is always found—leading to multiple MRA.

Part III. The Theory and Practice of Meta-Regression Analysis

A. The Theory of Meta-Regression Analysis
B. How Basic Econometrics is changing evidence-based practice & research synthesis in the medical and social sciences
C. Further Applications

A. The Theory of MRA. . .

. . is Econometric Theory

- **Conventional econometrics:**

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\alpha} + \mathbf{u} \quad \& \quad (4)$$

$$\hat{\boldsymbol{\alpha}} = (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{Y} \qquad \text{OLS}$$

- **MRA:** $\mathbf{e} = \mathbf{M}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ (5) Where:

e is a Lx1 vector of all the reported empirical effects,

M is a LxK matrix of moderator variables

β is a Kx1 vector of MRA coefficients $\hat{\alpha}_{1i}$

ε is a Lx1 vector of sampling errors in $\hat{\alpha}_{1i}$

From econometrics

{and the assumptions that researchers make},

we know that:

- $\hat{\alpha}_{1i}$ will be asymptotically normal
- $(\hat{\alpha}_1 - \alpha_1)/S_{\hat{\alpha}_1}$ will have a t-distribution
- there will often be misspecification biases (e.g., omitted-variable bias, simultaneity bias, incorrect functional forms, *etc*.). These biases largely define **M** in our multiple MRA.

B. How Basic Econometrics is changing research synthesis in the medical and social sciences

- **Weighted Least Squares** as the central approach for both meta-analysis and meta-regression analysis
- **Meta-Regression Analysis (MRA)** accommodates and corrects selective reporting bias (aka, publication bias; small-sample bias) and misspecification biases, in general.

Weighted Least Squares (WLS-MRA)

- MRA model (5) should never be estimated by OLS, because there is much variation among the reported SEs of $\mathbf{y}_i$ or $\hat{\alpha}_{1j}$

- Enter **WLS**: $\hat{\boldsymbol{\beta}} = (\mathbf{M}^t\boldsymbol{\Omega}^{-1}\mathbf{M})^{-1}\mathbf{M}^t\boldsymbol{\Omega}^{-1}\mathbf{y}$ **(6)**

Where:

$$\Omega = \sigma^2 \begin{vmatrix} \sigma_1^2 & 0 & \bullet & \bullet & 0 \\ 0 & \sigma_2^2 & & & 0 \\ \bullet & & \bullet & & \bullet \\ \bullet & & & \bullet & \bullet \\ 0 & 0 & \bullet & \bullet & \sigma_L^2 \end{vmatrix} \quad (7)$$

The Gauss-Markov Theorem
{Proportional Heterogeneity Invariance}

- As long as Ω in (6) is known up to some unknown proportion, σ^2, WLS $\{\hat{\boldsymbol{\beta}}\}$ is **Best {Minimum Variance} Linear Unbiased Est.**
- Invariance to proportional excess heterogeneity is a robustness property of the Gauss-Markov Theorem and WLS. It is **not an assumption**.

Basic Econometric Unrestricted WLS

replaces Ω with:

$$\hat{\Omega} = S^2 \begin{vmatrix} SE_1^2 & 0 & \bullet & \bullet & 0 \\ 0 & SE_2^2 & & & 0 \\ \bullet & & \bullet & & \bullet \\ \bullet & & & \bullet & \bullet \\ 0 & 0 & \bullet & \bullet & SE_L^2 \end{vmatrix} \qquad (5)$$

and S^2 is estimated by the WLS residuals, automatically

Meta-Analysis is Simple Weighted Averages

$$\hat{\mu} = \Sigma w_i effect_i / \Sigma w_i$$

- Conventional Meta-Analysis: Fixed- & Random-Effects
- FE: $w_i = 1/\sigma_i^2$ restricts σ^2 to be equal to 1.
- RE: $w_i = 1/(\sigma_i^2 + \tau^2)$
- **Weighted Least Squares (WLS)** weighted average
  - Simple regression: **t vs. 1/SE ;** no intercept
  - Always as good as FE and RE
  - Better than RE if there is selective reporting
  - Better than FE if there is heterogeneity.
  - Stanley and Doucouliagos (2015): "Neither Fixed nor Random: Weighted Least Squares Meta-Analysis," ***Statistics in Medicine, 34*****:** 2116-27.

Simulations and statistical theory prove that:

- Our Unrestricted WLS MRA (**WLS-MRA**):
  - is always as good as and usually better than fixed and random-effects meta-regression used in medicine and social sciences— Stanley, T.D. and Doucouliagos, C. (2016) "Neither fixed nor random: Weighted least squares meta-regression analysis," ***Research Synthesis Methods.***
- Basic econometrics improves systematic reviews and evidence-based practice in all disciplines!
- **WAAP**—Weighted Average of the Adequately Powered— improves systematic reviews of research using stat. power—Ioannidis, J.P.A., Stanley, T.D. and Doucouliagos, C. "The power of bias in economics research," ***The Economic Journal***, forthcoming, 2017. Stanley, T.D., Doucouliagos, C. and Ioannidis, J.P.A. "Finding the Power to Reduce Publication Bias," ***Statistics in Medicine***, 2017.

Disciplinary Silos

- GLS and WLS are basic tools in economics.
- Rarely do we have any outside estimate of σ_j^2 just assume it's proportional to something or use the squared regression residuals.
- Here, we have all this additional, independent information— L estimates of variance, $SE_j^2 \rightarrow \sigma_j^2$
- **Why divide by root MSE as fixed effect does**? We always have excess systematic heterogeneity.
- **Why add another random term as random effect does**? If a moderator variable is correlated with the errors, **E(ε/M)≠0**, MRA coeff's are biased.

Accommodating and Correcting Publication Bias

- 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*.
- Stanley, T. D and Doucouliagos, C. (2014) "Meta-regression approximations to reduce publication selection bias," *Research Synthesis Methods*.
- Stanley, T.D. (2017) "Limitations of PET-PEESE and other meta-analysis methods." *Social Psychology and Personality Science*, forthcoming.

Reducing Selective Reporting Bias

PEESE **(precision-effect estimate with standard error)** **(S&D, 2014).** When PET finds evidence of some effect beyond pub'bias, run

$$effect_j = \beta_2 + \beta_0 Se^2_j + \varepsilon_j \quad (j=1, 2, \ldots L) \quad \textbf{(2)}$$

{Comment: Always use **WLS**; 1/*Se*^{2} as the weight}

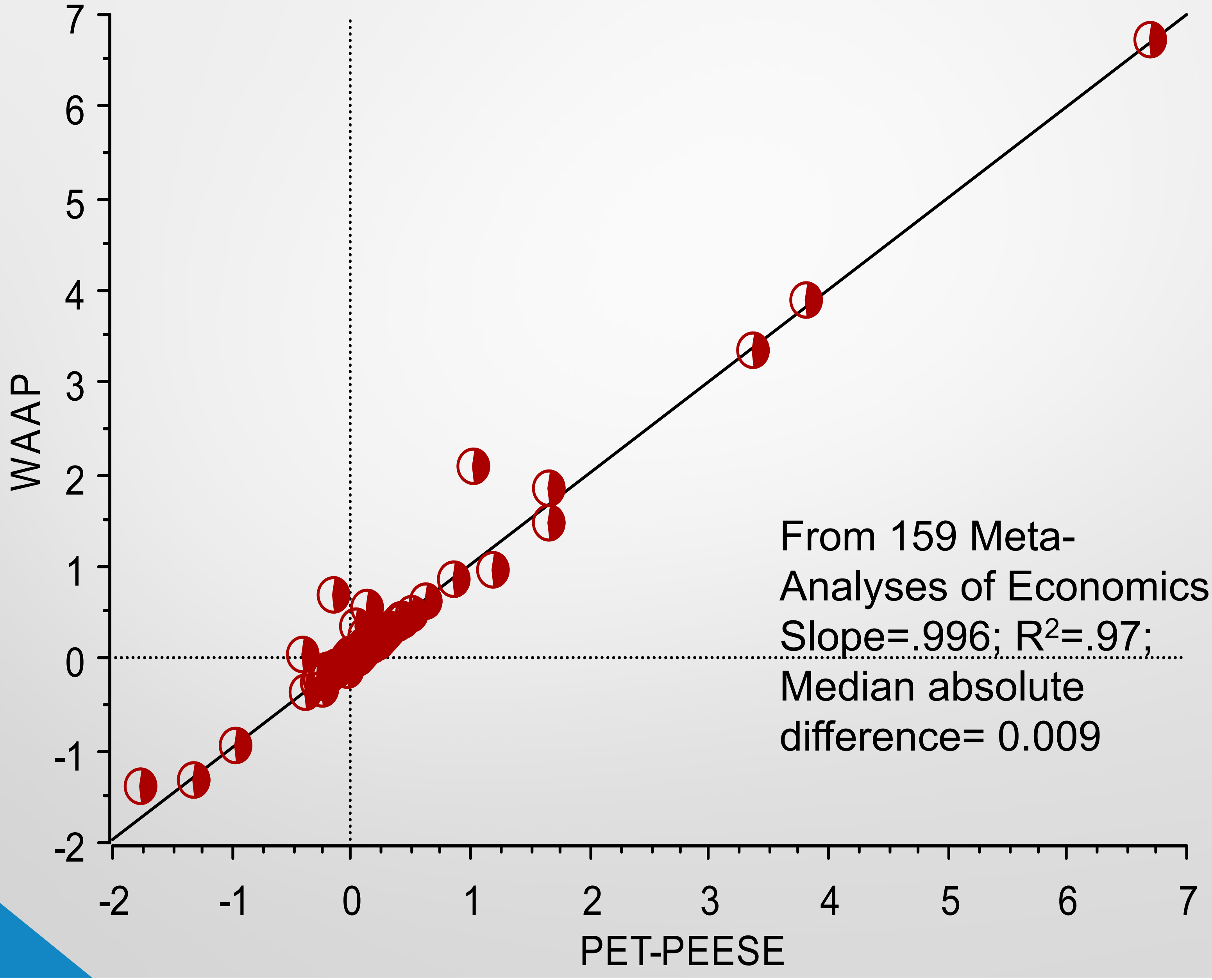
WAAP (Weighted Average of the Adequately Powered): Simple WLS weighted average of only those studies that are adequately powered >80%

FAT-PET-MRA-WLS

Table 1: MRA Tests of PB & Genuine Effect

Estimate	Marriage-Wage Premium	Baby Penalty	Minimum Wage	VSL
Se (β_0)	2.02 (6.99)	-1.05 (-8.27)	-1.6(-4.49)	3.20 (6.67)
Intercept β	0.037 (17.2)	-0.024(-30.85)	-.01 (-1.09)	0.81 (3.56)
Ave. Effect	0.121	-0.053	-0.19	$9.5 mil
WAAP	0.042	-0.027	-0.013	$1.47 mil
PEESE	0.044	-0.028	-.036	$1.67mil
n	661	1,895	1,474	39

Predicting WAAP from PET-PEESE



From 159 Meta-Analyses of Economics
Slope=.996; R^2=.97;
Median absolute difference= 0.009

When WAAP and PET-PEESE agree:

- We have a very good summary estimate!
- Robust to selective reporting bias.
- Vulnerable only to systematic heterogeneity.

C. Further Applications of (MRA)

Reported Econometric Estimates:

- Always have excess variation i.e., variation in excess of their standard errors.
- have ubiquitous misspecification biases and excess heterogeneity.

A Second Example of Multiple MRA:

Marriage-Wage Premium

Recall:

$$effect_i = (\beta + \Sigma \alpha_k Z_{ik}) + (\beta_0 Se_i + \Sigma \gamma_j K_{ij} Se_i) + \varepsilon_i \quad (3)$$

{Heterogeneity} **{Publication Bias}**

{Comment: Always use **WLS**; $1/Se^2$ as the weight}

Table 2: Marriage-Wage Premium (MWP)

Variables	WLS (1)	Cluster-Robust (2)	RE Panel (3)	FE Pa (4)
Constant	0.045*	0.045*	0.037*	0.064*
	(0.012)	(0.017)	(0.013)	(0.023
SE	0.937*	0.937*	0.936*	-0.411
	(0.176)	(0.363)	(0.283)	(0.280
two thousands	0.034*	0.034*	0.033*	0.040*
	(0.005)	(0.007)	(0.006)	(0.009
fixed effects	-0.021*	-0.021*	-0.019*	-0.016
	(0.003)	(0.004)	(0.003)	(0.003
restricted age range	-0.044*	-0.044*	-0.036*	-0.030
	(0.005)	(0.013)	(0.005)	(0.007
panel	0.013*	0.013*	0.013*	0.002
	(0.004)	(0.006)	(0.004)	(0.005
include years married	-0.015*	-0.015	-0.014	-0.012
	(0.002)	(0.010)	(0.002)	0.002
US data	0.072*	0.072*	0.079*	0.086*
	(0.007)	(0.014)	(0.008)	(0.012
omit occupation	-0.016*	-0.016*	-0.009	0.035*
	0.005	0.007	0.006	(0.009
omit union	0.055*	0.055*	0.050*	0.011
	(0.008)	(0.015)	(0.010)	(0.014
omit region	0.016*	0.016	0.019	0.035*
	(0.004)	(0.011)	(0.006)	(0.010
omit education	-0.062*	-0.062*	-0.054*	0.009
	(0.012)	(0.018)	(0.015)	(0.025
omit veteran	-0.032*	-0.032*	-0.039*	-0.080
	(0.008)	(0.013)	(0.010)	(0.014
omit urban	-0.033*	-0.033*	-0.021*	0.016
	(0.008)	(0.015)	(0.010)	(0.014
not about marriage	0.032*	0.032*	0.032*	0.053*
	(0.008)	(0.014)	(0.010)	(0.017
Adjusted R^2	0.772	0.772	0.831	0.649

MRA Modeling Issues—Cont.

- Always accommodate:
  - heteroscedasticity (WLS) and
  - within-study dependence
    - Cluster-robust standard errors
    - Fixed and Random-effects panel models {roughly the same as multilevel models}

Table 2: MRA of MWP, again

Variables	WLS (1)	Cluster-Robust (2)	RE Panel (3)	FE Pa (4)
Constant	0.045*	0.045*	0.037*	0.064*
	(0.012)	(0.017)	(0.013)	(0.023)
SE	0.937*	0.937*	0.936*	-0.411
	(0.176)	(0.363)	(0.283)	(0.280)
two thousands	0.034*	0.034*	0.033*	0.040*
	(0.005)	(0.007)	(0.006)	(0.009)
fixed effects	-0.021*	-0.021*	-0.019*	-0.016
	(0.003)	(0.004)	(0.003)	(0.003)
restricted age range	-0.044*	-0.044*	-0.036*	-0.030
	(0.005)	(0.013)	(0.005)	(0.007)
panel	0.013*	0.013*	0.013*	0.002
	(0.004)	(0.006)	(0.004)	(0.005)
include years married	-0.015*	-0.015	-0.014	-0.012
	(0.002)	(0.010)	(0.002)	0.002
US data	0.072*	0.072*	0.079*	0.086*
	(0.007)	(0.014)	(0.008)	(0.012)
omit occupation	-0.016*	-0.016*	-0.009	0.035*
	0.005	0.007	0.006	(0.009)
omit union	0.055*	0.055*	0.050*	0.011
	(0.008)	(0.015)	(0.010)	(0.014)
omit region	0.016*	0.016	0.019	0.035*
	(0.004)	(0.011)	(0.006)	(0.010)
omit education	-0.062*	-0.062*	-0.054*	0.009
	(0.012)	(0.018)	(0.015)	(0.025)
omit veteran	-0.032*	-0.032*	-0.039*	-0.080*
	(0.008)	(0.013)	(0.010)	(0.014)
omit urban	-0.033*	-0.033*	-0.021*	0.016
	(0.008)	(0.015)	(0.010)	(0.014)
not about marriage	0.032*	0.032*	0.032*	0.053*
	(0.008)	(0.014)	(0.010)	(0.017)
Adjusted R^2	0.772	0.772	0.831	0.649

ROBUSTNESS, ROBUSTNESS, ROBUSTNESS

- **Interpret your 'results' to be only those findings that are seen no matter how your look at them.**

Robust Effects on Marriage-Wage Premium

Variables	WLS	Cluster-Robust	RE Panel	FE Panel
	(1)	(2)	(3)	(4)
Constant	**0.045***	**0.045***	**0.037***	**0.064***
	(0.012)	(0.017)	(0.013)	(0.023)
SE	**0.937***	**0.937***	**0.936***	**-0.411**
	(0.176)	(0.363)	(0.283)	(0.280)
two thousands	**0.034***	**0.034***	**0.033***	**0.040***
	(0.005)	(0.007)	(0.006)	(0.009)
fixed effects	**-0.021***	**-0.021***	**-0.019***	**-0.016***
	(0.003)	(0.004)	(0.003)	(0.003)
restricted age range	**-0.044***	**-0.044***	**-0.036***	**-0.030***
	(0.005)	(0.013)	(0.005)	(0.007)
panel	**0.013***	**0.013***	**0.013***	**0.002**
	(0.004)	(0.006)	(0.004)	(0.005)
US data	**0.072***	**0.072***	**0.079***	**0.086***
	(0.007)	(0.014)	(0.008)	(0.012)
not about marriage	**0.032***	**0.032***	**0.032***	**0.053***
	(0.008)	(0.014)	(0.010)	(0.017)
Adjusted R^2	0.772	0.772	0.831	0.649

What does this say about the Marriage Wage Premium?

- Including fixed effects lowers MWP by about 2%; ***selection*** is not the entire reason for MWP
- Publication Bias explains most of MWP
  - FE panel shows clear ***differential PB***
- The US effect is 7-8%; thus, MWP is largely a US phenomenon.
- Our MRA findings favor the 'Married with children' hypothesis: that married men are more stable and reliable or perceived to be so.

What would a benchmark study find?

{using MRA to 'predict' study findings}

- The benchmark (or best-practice) study:
  - employs fixed-effects panel model
  - does not omit any potentially relevant variable,
  - Nor rely upon a restricted age range.
  - has an infinite sample size: SE→ 0
- MWP = 9.4% {CI = (7.3%; 11.4%)} for the US
- MWP = 3.6% {CI = (-0.1%; 7.4%)} for non-US

Part IV-The Crisis of Confidence and Replication in the Social Sciences

- "The power of bias in economics research," *Economic Journal*, forthcoming, Ioannidis, Stanley and Doucouliagos.
- We calculate statistical power and research inflation for more than **64,000** estimates from over **6,700** studies aggregated into **159** economics meta-analysis.
- All meta-analyses need to calculate power!

What did we find?

- The typical area of economics has about **90%** of its estimates **underpowered** (power<80%)
  - That is, the median proportion that is adequately powered is, at best, just over 10%
- The median of the median of median powers= **10.9%**
- The typical economic reported results is **inflated by 100%** or more; **1/3rd** are **exaggerated by a factor of four or more.**

References

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

De Linde Leonard, M., Stanley, T. D and Doucouliagos, C. "Does the UK minimum wage reduce employment? A meta-regression analysis." *British Journal of Industrial Relations* 52 (2014), 499-520.

De Linde Leonard, M. and Stanley, T. D "Married with children: What remains when observable biases are removed from the reported male marriage wage premium." *Labour Economics* 33 (2015), 72-80.

Doucouliagos, C.(H) and Stanley, T.D. 2009. 'Publication selection bias in minimum-wage research? A meta-regression analysis," *British Journal of Industrial Relations* 47: 406-29.

Doucouliagos, H., T.D. Stanley and M. Giles. "Are Estimates of the Value of a Statistical Life Exaggerated?" Journal of Health Economics, 31(2012): 197-206.

Egger, M. et al. 1997. Bias in meta-analysis detected by a simple, graphical test, *British Medical Journal* 316: 629-34.

Glass, G.V. 1976. Primary, secondary, and meta-analysis of research. *Educational Researcher* 5: 3-8.

Ioannidis, J.P.A., Stanley, T.D. and Doucouliagos, C. "The power of bias in economics research," *The Economic Journal*, forthcoming, 2017.

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

Stanley, T.D., 2005. 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 bias. *Oxford Bulletin of Economics and Statistics* 70:103-127.

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

Stanley, T.D. and Doucouliagos, H(C) 2012 *Meta-Regression Analysis in in Economics and Business*, Oxford: Routledge.

Stanley, T. D and Doucouliagos, C. "Meta-regression approximations to reduce publication selection bias," *Research Synthesis Methods* 5 (2014), 60-78.

Stanley, T.D. and Doucouliagos, C. "Neither fixed nor random: Weighted least squares meta-analysis," Statis*tics in Medicine* 34 (2015), 2116-27.

Stanley, T.D. and Doucouliagos, C. "Neither fixed nor random: Weighted least squares meta-regression analysis," *Research Synthesis Methods*, http://onlinelibrary.wiley.com/doi/10.1002/jrsm.1211/abstract , 2016.

Stanley, T.D., Doucouliagos, C. and Ioannidis, J.P.A. "Finding the Power to Reduce Publication Bias," *Statistics in Medicine,* 2017. DOI: 10.1002/sim.72

Using Meta-Regression Analysis for Environmental Benefit Transfer

Robert J. Johnston
George Perkins Marsh Institute

February 21, 2017

CLARK UNIVERSITY

Practical Applications of Meta-Analysis

- Meta-analysis is often discussed in terms of its relevance for understanding the scholarly literature.
- Results can (at least in concept) inform decisions in the real world, but sometimes impacts are indirect or unclear.
  - Example—Do minimum wages affect employment?
  - MRA results challenge common wisdom.
- MRA can also provide direct inputs for policy analysis. Here, the effect is more clear.
- Multiple examples are found in environmental economics.
  - VSL is a good example. This is frequently used as a direct input in benefit-cost analysis (BCA).

Non-Market Valuation

- MRA is commonly used to provide estimates of non-market values for use within BCA and other types of policy analysis.
- Non-market valuation provides estimates of economic value for environmental goods and services that are not exchanged in markets.
  - Ecosystem service values are often non-market values.
  - Common examples include the value of improved air quality, water quality, fish stocks, wildlife stocks and many others.
  - These values are often measured using estimates of willingness to pay (WTP), reflecting Hicksian compensating surplus or variation.

Example—Non-Market Value of Recreational Fishing

- What is the true value of recreational fishing to an angler (a recreational fisherman)?
- How much more would an angler be willing to pay (in time and travel costs) to go fishing at a site where he expects to catch one more fish compared to current sites?
- The angler cannot directly "buy" improved fishing quality.
- There is no market, so this is a non-market value.
- But, the observed tradeoff between time/travel and additional catch reveals an economic value.
- This value can be estimated by analyzing fishing behavior.

A Meta-Analysis of Recreational Fishing Value

- Johnston et al. (2006): Mean willingness to pay per fish caught.

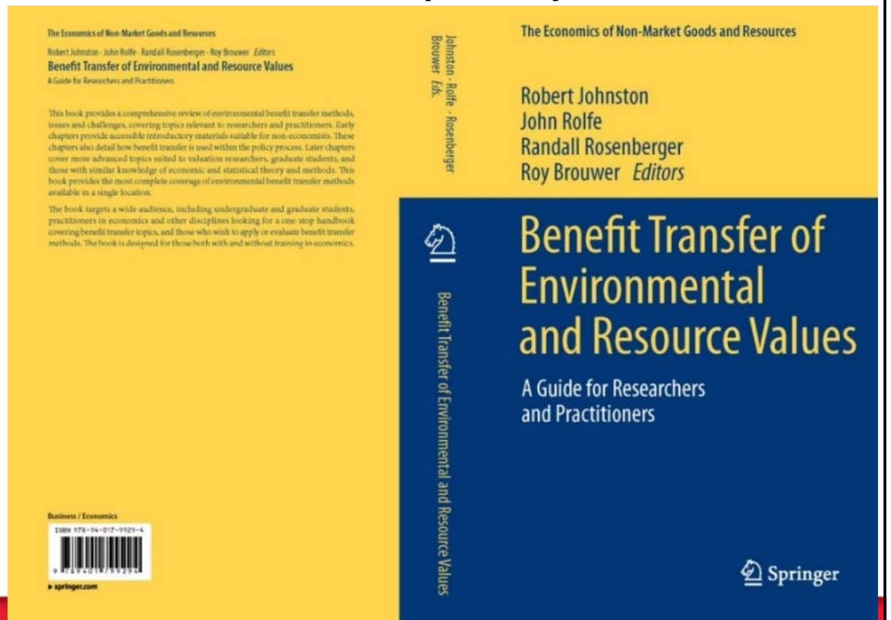
Marginal Value per Fish, by Region and Species							
Species	California	North Atlantic	Mid-Atlantic	South Atlantic	Gulf of Mexico	Great Lakes	Inland
big game	$12.32	$6.19	$5.95	$13.57	$13.26		
small game	$6.38	$5.22	$5.19	$5.03	$4.95		$4.71
flatfish	$8.57	$5.24	$4.94	$4.93	$4.82		
other saltwater	$2.60	$2.62	$2.56	$2.50	$2.44		$2.54
salmon	$13.67					$11.66	$13.88
steelhead	$11.25					$12.57	$11.42
musky						$61.37	$64.71
walleye/pike						$3.61	$3.60
bass						$7.52	$7.92
panfish			$0.93	$0.93		$1.17	$0.93
rainbow trout						$7.38	$2.84
other trout						$8.29	$2.48
generic freshwater						$5.46	$1.96
generic saltwater	$2.73	$2.64	$2.85	$2.51	$3.22		$2.79

Environmental Benefit Transfer

- The time and money required for high quality primary valuation research has led to the common use of *benefit transfer* to estimate values for policy analysis.
- Benefit transfer uses results from prior research at one or more study sites to predict value estimates at other policy sites for which value estimates are unavailable.
- Benefit transfer involves transfer errors, but is often the only option to estimate non-market benefits or costs for environmental policy analysis.
- Benefit transfer is a nearly universal component of large-scale BCA in the US, EU and other countries (Johnston et al. 2015).

Environmental Benefit Transfer

- With a few uncommon exceptions, benefit function transfers are more accurate than comparable unit value transfers (Boyle et al. 2010; Johnston et al. 2015).
- Benefit function transfer uses a parameterized function estimated at the study site to predict value at the policy site.
- This allows transferred benefits to be updated or adjusted for observed conditions at the policy site (e.g., income, environmental conditions, spatial factors).



Environmental Benefit Transfer

Transferred Value Estimate or Benefit Function

$$\text{Value}_A = f(X_A, \beta_A)$$

Study Site A
(Economic Value Measured Here by Prior Primary Research)

Policy Site B
(Value Estimate Required for BCA)

Coastal Management, 30:47–65, 2002
Copyright © 2002 Taylor & Francis
0892-0753/02 $12.00 + .00

Valuing Estuarine Resource Services Using Economic and Ecological Models: The Peconic Estuary System Study

ROBERT J. JOHNSTON
THOMAS A. GRIGALUNAS
JAMES J. OPALUCH
MARISA MAZZOTTA
JERRY DIAMANTEDES

Department of Environmental and Natural Resource Economics
University of Rhode Island
Kingston, Rhode Island, USA

This article summarizes four integrated economic studies undertaken to contribute to resource preservation and restoration decisions for the Peconic Estuary System of Suffolk County, NY. Completed as part of the National Estuary Program, the studies apply distinct resource valuation methods to a wide range of resource issues. The principal goals of this article are to highlight different methodologies that may be used to assess nonmarket economic values in a coastal management context, and characterize differences in the results that one may expect from each approach. We also emphasize potential relationships among values estimated by different nonmarket methodologies, and comment on the implications of these relationships for the interpretation and use of economic value estimates.

Keywords estuarine resources, nonmarket valuation, Peconic, resource values

Introduction

Coastal managers increasingly request quantitative methods for prioritizing management actions that explicitly account for economic and social factors (Chua 1993; Christie & White 1997; Ruddle, 1988). One common method of linking economic and ecological outcomes involves assessments of the economic consequences of ecological management outcomes—often using either market or nonmarket valuation techniques (Freeman, 1993). Several methods may be used to assess the value of coastal resource goods and services, depending upon the resources in question and the specific issues of concern. However, one limitation of valuation methods used in isolation is that they may quantify values for only a subset of the many relevant resource goods and services, providing only a partial view

Received November 15, 2000; accepted July 9, 2001.
This research was funded by the Peconic Estuary Program and the Rhode Island Agricultural Experiment Station (AES #3881). Opinions belong solely to the authors and do not imply endorsement by the funding agencies or the U.S. Government.
Address all correspondence to Robert J. Johnston, Department of Environmental and Natural Resource Economics, University of Rhode Island, Kingston, RI 02881, USA. e-mail: rjohnston@uri.edu.

47

$$\textbf{Transfer Value} = f(X_B, \beta_A)$$

Observed Conditions at Policy Site B

Environmental Benefit Transfer

- Benefit functions (used for benefit transfer) can be
  - transferred directly from one prior study, or
  - estimated using information from many prior studies in the literature.
- Meta-regression models (MRMs) are often used to estimate these benefit functions.
- Use of MRMs enables benefit functions that are more flexible and generally applicable than benefit functions taken from a single published study.

MRMs for BT

- The dependent variable in a benefit transfer MRM is a comparable measure of economic value drawn from similar studies addressing the same good at many different sites.
  - Most often mean willingness to pay (WTP) from revealed or stated preference valuation studies.
- Independent variables characterize site, resource, population and methodological attributes hypothesized to explain variation in value.
- The goal is a statistical benefit function able to predict economic values at sites where no primary valuation studies have been conducted.

Technical Structure of a Non-Market Valuation MRM

Primary Study *a*

$$Y_a = \alpha_a + \sum_j \beta_{aj} x_{aj} + \varepsilon_a$$

Behavioral Measure (e.g., trips)

Primary Study Calculates

$$WTP_{ak} = \beta_{ak} / \beta_{a\_cost}$$

WTP from Multiple Studies (1…A)

Policy, Site, Population & Resource Variables

Methodological Variables

MRM

$$\begin{bmatrix} WTP_{1k} \\ WTP_{2k} \\ \cdots \\ WTP_{Ak} \end{bmatrix} = \delta_k + \sum_m \gamma_{mk} \begin{bmatrix} z_{1mk} \\ z_{2mk} \\ \cdots \\ z_{Amk} \end{bmatrix} + \sum_s \mu_{sk} \begin{bmatrix} w_{1sk} \\ w_{2sk} \\ \cdots \\ w_{Ask} \end{bmatrix} + \varepsilon_k$$

Technical Structure of a Non-Market Valuation MRM

Benefit Function

$$\hat{WTP}_k = \hat{\delta}_k + \sum_m \hat{\gamma}_{mk} z_{mk} + \sum_s \hat{\mu}_{sk} \overline{w}_{sk}$$

Policy Site Data (Site *b*)

$\uparrow$ z_{bk}

$\uparrow$ **Means or Selected Values**

Predicted Welfare (Value) Estimate

$$\hat{WTP}_{bk}$$

Some Differences Between Traditional MRA and MRMs for Benefit Transfer

- Goal of MRMs used for BT is to *predict (or forecast)* a value of the dependent variable *out-of-sample*.
  - Not to estimate a mean treatment effect.
- Standard errors (or other comparable measures of precision) are generally not available for all WTP estimates in the sample.
- Selection biases are often addressed using different methods (Rosenberger and Johnston 2009).
- Valuation MRMs face a range of challenges
  - Example—commodity and welfare consistency
  - Effect of spatial and other variables that are unreported by primary studies

MRMs of Environmental Value

- There roughly 200 published MRMs in the environmental economics literature (Nelson and Kennedy 2009; Johnston et al. 2015). Examples include MRMs on the value of:
  - Water quality (Johnston et al. 2005, 2016; Johnston and Thomassin 2010; Poe et al. 2001; Van Houtven et al. 2007).
  - Wetlands (Brouwer et al. 1999; Woodward and Wui 2001; Ghermandi and Nunes 2013; Brander et al. 2012).
  - Coral reefs (Brander et al. 2007; Londoño and Johnston 2012).
  - Outdoor recreation (Bateman and Jones 2003; Johnston et al. 2006; Rosenberger and Loomis 2000a,b; Moeltner et al. 2007; Moeltner and Rosenberger 2008, 2014; Stapler and Johnston 2009).

An Illustration—MRM for Water Quality Benefit Transfer

- The estimation and use of MRMs is illustrated using an MRM for US water quality benefit transfer.
  - Provides estimates of per household WTP for water quality changes in US rivers, lakes and bays.
  - Developed as part of US Environmental Protection Agency efforts to generate MRMs for use in regulatory policy analysis.
  - The model predicts WTP as a function of:
    - Water quality change and baseline (non-improved) starting point.
    - Spatial characteristics of the water quality change and affected areas
    - Spatial and economic characteristics of the affected populations
    - Affected uses and ecosystem services

The Metadata

- Metadata are drawn from primary studies that estimate willingness to pay (WTP) for water quality changes in US water bodies that support non-consumptive ecosystem services.
- Include studies that estimate total (use & nonuse) values and use generally accepted stated preference methods.
- Protocols of Stanley et al. (2013) for coding, study identification and research reporting.
- Metadata include 148 observations from 53 stated preference studies published between 1985 and 2016.
- Published and unpublished literature included.
- Monetary values adjusted to 2016 US dollars.

The Metadata

- The metadata combine primary study information with geospatial data from geographic information system (GIS) data layers and other external sources.
- Model requires a comparable measure of water quality for each study included in the metadata.
- Water quality changes measured for each study using a 100-point Water Quality Index (WQI).
- WQIs combine information on multiple physical and chemical water quality parameters into a single index linked to the presence of aquatic species and suitability for human uses.

MRM Specification

- Dependent variable: natural log of (WTP/household/unit change), where units are measured on a standard 100-point water quality index (WQI) (Abassi 2012).
- 25 independent variables characterizing: (1) study methodology, (2) populations, (3) market areas and study sites, (4) water bodies and (5) water quality change.
- Unweighted OLS regression with cluster robust standard errors. Other estimation methods provide similar results.
- Results may also be weighted according to various factors (e.g., number of observations per study; standard error of estimates—although the latter are rarely available).

Independent Variables

Variable	Definition
ce	Binary variable with a value of one for studies that are choice experiments
thesis	Binary variable with a value of one for studies developed as thesis projects or dissertations
lnyear	Natural log of the year in which the study was conducted (converted to an index by subtracting 1980, before making the log transformation). Note that the year of the study is not the same as the year of publication.
volunt	Binary variable indicating that WTP was estimated using a payment vehicle described as voluntary
nonparam	Binary variable indicating that WTP was estimated using non-parametric methods.
non_reviewed	Binary variable indicating that the study was not published in a peer-reviewed journal
lump_sum	Binary variable indicating that payments were to occur on something other than an annual basis over an extended or indefinite period of time
northeast	Binary variable indicating that the survey included respondents from the USDA Northeast region
central	Binary variable indicating that the survey included respondents from the USDA Midwest or Mountain Plains regions
south	Binary variable indicating that the survey included respondents from the USDA Southeast or Southwest regions
nonusers	Binary variable indicating that the survey was implemented over a population of nonusers
lnincome	Natural log of median income for the sample area of each study based on U.S. Census data.

Independent Variables

Variable	Definition
mult_bod	Binary variable that takes on a value of 1 if the studied system includes multiple water body types (e.g., lakes and rivers), and zero otherwise
river	A binary variable that takes on a value of 1 if the studied system includes rivers, and zero otherwise
swim_use	Binary variable identifying studies in which changes in swimming uses are specifically noted in the survey
gamefish	Binary variable identifying studies in which changes in game fishing uses are specifically noted in the survey
boat_use	Binary variable identifying studies in which changes in boating uses are specifically noted in the survey
ln_ar_agr	Natural log of the proportion of the improved resource area which is agricultural based on the NLCD. Improved resource area includes all counties that intersect the improved resource
prop_chg	The proportion of water bodies of the same hydrological type improved by the water quality change, within affected state(s), measured by area (lakes and bays) or shoreline length (rivers).
q_avg	The linear mid-point between baseline and improved water quality valued by the study, specified on the 100-point water quality index (Abbasi 2012; McClelland 1974; Mitchell and Carson 1989).
lnquality_ch	Natural log of the change in mean water quality valued by the study, specified on the 100-point water quality index.
ln_size_ratio	The natural log of *size_ratio*, where *size_ratio* is defined as the ratio or quotient of (1) the total area of all counties that intersect the improved water resource(s) identified by the study, in square kilometers, and (2) the population-weighted average distance between US Census Block centroids and the nearest point on an affected water resource, in kilometers, for all Census Blocks sampled by the study.

Independent Variables

Variable	Definition
river_size_ratio	Multiplicative interaction between *river* and *size_ratio*.
ln_pop	The natural log of population within the affected resource area (counties intersecting improved water bodies), based on the 2010 US Census.

MRM Results

- Parameter estimates are jointly significant at $p<0.0001$, with $R^2 = 0.7480$. This is relatively good fit for a valuation MRM.
- Of 25 non-intercept parameter estimates in the MRM, 23 are statistically significant at $p<0.10$, with most significant at $p<0.01$
- All statistically significant variables have theoretically expected signs, where prior expectations exist.
- For example, WTP per unit of water quality improvement is negatively related to both the scope of water quality change (*lnquality_ch*), and to the midpoint of the change (*q_avg*).
    - This is an expected pattern reflecting diminishing marginal utility to water quality improvements.
- WTP also positively related to the size of improved water body.

Some Key Results

Spatial and Water Quality Change Variables

(standard errors in parentheses)

q_avg (WQI midpoint)	-0.0105** (0.0051)	Water Quality
lnquality_ch (Δ WQI)	-0.5776*** (0.1140)	
ln_sz_ratio (area / distance)	0.1373*** (0.0397)	Size / Distance
prop_chg (proportional affect)	1.1095*** (0.3237)	Proportional Effect
ln_ar_agr (fraction agriculture)	-0.3508*** (0.07459)	Land Use
ln_pop (population)	-0.1729*** (0.0517)	Population

Some Key Results

Regions and Populations

(standard errors in parentheses)

northeast	0.8065***	
	(0.2355)	
central	0.4947***	
	(0.1128)	Region
south	1.5661***	
	(0.1356)	
nonusers	-0.3750***	User vs Nonuser
	(0.1206)	
lnincome	1.1255***	Income
	(0.3783)	

Some Key Results

Methodological and Selection

(standard errors in parentheses)

thesis	0.7177***
	(0.2000)
lnyear	-0.5640***
	(0.09168)
volunt	-1.4817***
	(0.1757)
outlier_bids	-0.3694***
	(0.1285)
nonparam	-0.4100***
	(0.09350)
non_reviewed	-0.7027***
	(0.1801)

Original Study Methods

Illustrative Benefit Transfer

- Applications use these results to forecast WTP for site where no valuation study has been conducted.
- Assume we require an estimate of marginal WTP/household per unit of water quality change, for an unstudied policy site.
- Assume water quality midpoint and change at the mean for the metadata (*lnquality_ch*=2.8825; *q_avg*=52.8882).
- Change occurs to rivers in the northeast US and affects 5% of rivers in a state (say, Massachusetts).
- WTP desired for a general population of users and nonusers.
- Rivers are used for swimming but not fishing or boating.
- All other variables assumed at mean values for the metadata.

Illustrative Benefit Transfer

- Application of benefit function leads to predicted **WTP = $4.50/household**.
- These values change under different scenarios
  - If 10% of state's rivers are affected, WTP increases to $4.76.
  - If WTP of nonusers only is desired, WTP decreases to $3.09.
- Thus, MRM benefit transfers can easily adjust for different site, population and policy characteristics.
- Many of these adjustments are not possible using benefit functions drawn from a single study.
- This is the first MRM to enable adjustments for spatial variables such as average distance to the improved water body.

How Accurate Is Benefit Transfer?

- Like many economic phenomena, true WTP can never be observed, only estimated.
- Benefit transfer is only conducted when a primary study has not been conducted.
- Accuracy in actual situations is not known.
- But, if a primary valuation study *has been* conducted for a site, we can compare the value estimated using benefit transfer to the value estimated by the primary study.
  - This is called convergent validity testing.
  - Used to evaluate "how accurate" benefit transfer might be in actual policy uses.

Testing MRM Benefit Transfer

- To evaluate the out-of-sample accuracy of BT forecasts from the MRM (inversely related to transfer error), we apply an iterative leave-one-out convergent validity test.
  - Begin with metadata of n=1…N observations.
  - Omit n^{th} observation from the metadata.
  - Estimate MRM using the remaining N-1 observations.
  - Steps 2 and 3 iterated for each n=1…N observation, resulting in a vector of N unique sets of MRM parameter estimates, each corresponding to the omission of the nth observation.
  - For each iteration, results are used to forecast WTP for the nth omitted observation, resulting in N out-of-sample forecasts.
  - Evaluate transfer error for each iteration.

Convergent Validity Test Results

	Mean Absolute Value Error ($)	Std. Dev.	Mean Absolute Value Error (%)	Std. Dev. (%)
Model Accuracy Measures	$3.03	$4.09	**68.23%**	133.45%

- On average, one expects a mean (absolute value) error of approximately 68%, when the model is used for benefit transfer in actual situations (forecasting out of sample).
- This is a common magnitude of error for MRM benefit transfers.
- If greater accuracy is needed, primary valuation studies should be conducted.

Conclusions

- MRMs are commonly used within environmental economics to support BCA directly. This is a particularly influential use.
- The resulting MRMs
  - Can identify systematic patterns in WTP across studies that may not be reported in primary studies alone.
  - Enable benefit transfers that are more flexible than those that rely on functions from a single study.
- MRMs for benefit transfer are similar but not identical to meta-regression methods for other economic applications.
- Benefit transfers from MRMs are often more accurate than transfers using single-study functions or values, but benefit transfer in general is not particularly accurate compared to primary study valuation.

THANK YOU

Robert J. Johnston
Director, George Perkins Marsh Institute
Professor, Department of Economics
Clark University
950 Main St.
Worcester, MA 01610
Email: rjohnston@clarku.edu

CLARK UNIVERSITY

CHALLENGE CONVENTION. CHANGE OUR WORLD.