Recent Developments: Should We WAAP Economics' Low Power?

Tom Stanley
School of Business and Law,
Deakin University, Melbourne, Australia;
Stanley@Hendrix.edu &
Julia Mobley Professor of Economics, Emeritus
Hendrix College, USA

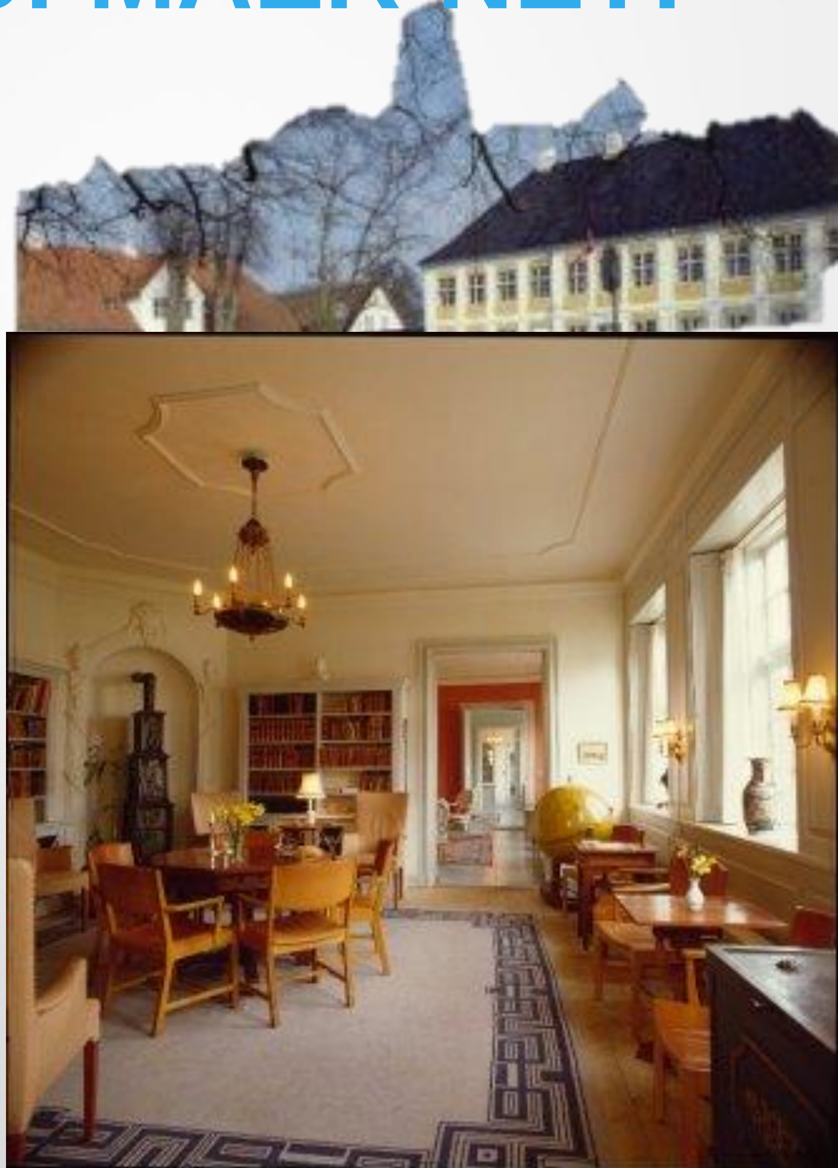
"In our view, the central task of meta-regression analysis is to filter out systematic biases, largely due to misspecification and selection, already contained in economics research."

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

It's the 10th Anniversary of the founding of MAER-NET!

Martin Pladam graciously hosted the 2007 meeting at Sønderborg, Denmark; where:

- Annual colloquia were planned &
- A core group formed to promote and to advance meta-analysis in economics that became—MAER-Net



Organization for this talk

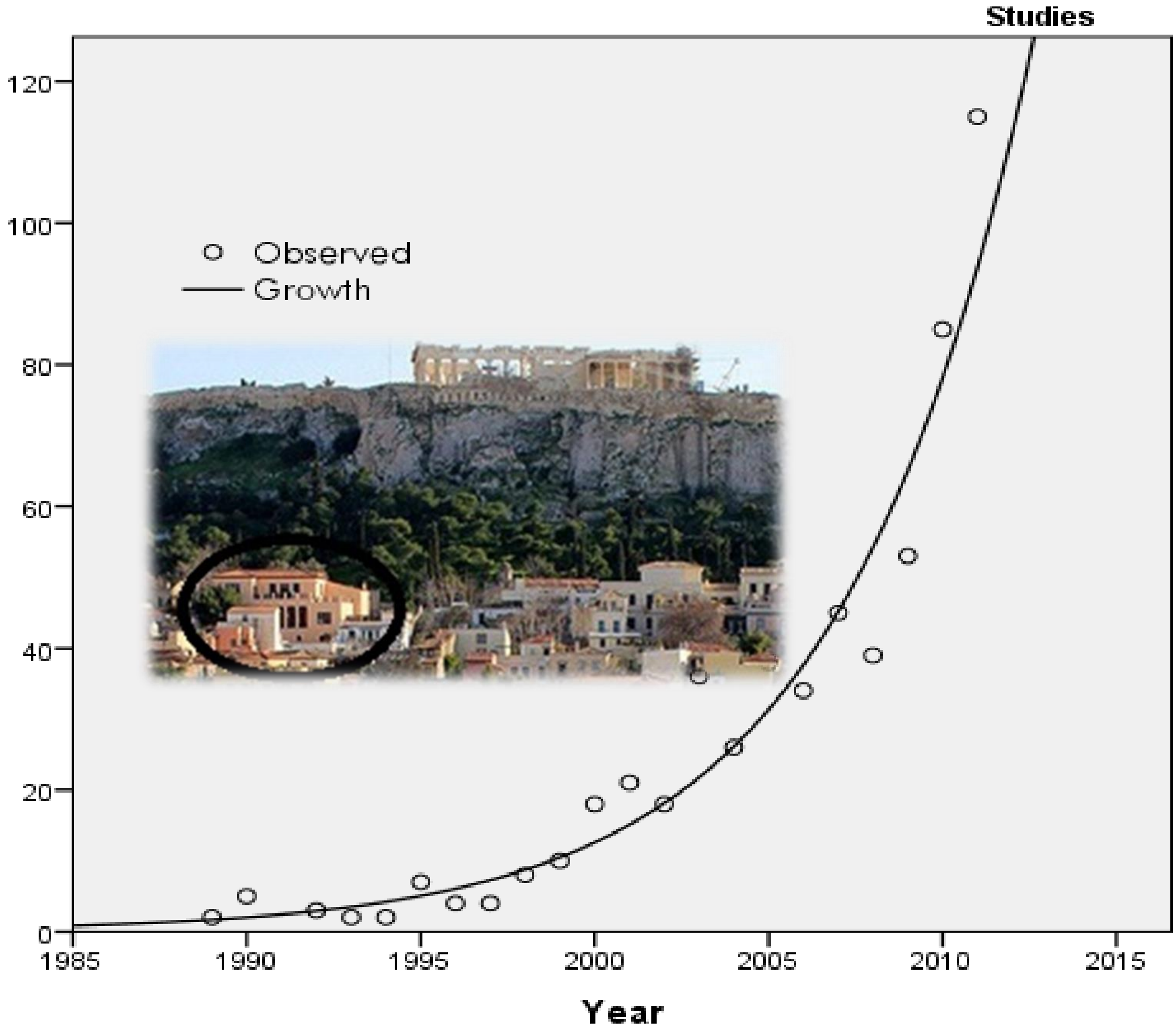
- **Part I: What's happened?**
  over the last decade
- Part II: Recent Developments
- Part III: Some Recommendations
  for practice going forward

A. We met in wonderful places and had a great time.

- **Zeppelin University 2017**
- **Hendrix College, 2016**
- **Prague, CZ—2015**
- **Athens, Greece (2014)**
- **Greenwich, UK (2013)**
- **Perth, Australia (2012)**
- **Cambridge University, UK (2011)**
- **Hendrix College, US (2010)**
- **Corvallis, Oregon US (09)**
- **Nancy, France (08)**
- **Sonderborg, Denmark (07)**


Melbourne Australia, DeLMAR Oct 4 ’18
London—2019
Tokyo—2020

B. Economic MRAs have flourished



Exponential
@
18%/year

C. We learned a lot, too!

Routinely, we find that economics research has:

- Publication Bias {or selective reporting, small-sample bias, p-hacking or **whatever you wish to call it**}
- Low Statistical Power
- High Heterogeneity

Next, we look more carefully at each

Publication Bias

- Funnel graphs are often quite skewed (FAT-2/3^{rds})
- Most likely due to selective reporting of statistically significant findings in the ‘right’ direction.
- Technically, we can never infer ‘publication bias.’ Rather, confirm or reject the socio-economic theory of researcher behavior about incentives for publishing and the preferences for statistical sig.
- Our big survey 159 meta-analyses with 64,076 estimates from 6,700 papers finds that economics is typically **inflated by 100%**- Ioannidis et al. (2017)

Low Power

Ioannidis, Stanley and Doucouliagos (2017). "The power of bias in economics research," current issue of the ***Economic Journal***

- Most areas of economics have about 90% or more of their reported results underpowered (using Cohen's convention of 80% as desired power).
- Median power is only about 11.5%

More about our survey of 64,076 economic estimates from 6,700 papers

Statistical Power

- Power is 1-β; where β is the probability of a type II error. The type II error is the mistake of accepting that there is no effect when, in fact, there is a genuine effect.
- Power is the probability that we can detect what we seek.
- It is analogous to the power of a telescope.

Why is Power Important?

Unless we increase the “power of (our) studies, the **published literature is likely to contain a mixture of apparent results buzzing with confusion.** . . .Not only do **underpowered studies lead to a confusing literature** but they also create a literature **that contains biased estimates** of effect sizes” (Maxwell, 2004, p.161).

- Without power, a single empirical finding is as likely to be bias as informative.
- It is **power, not p-value**, that is the real metric of the importance of an empirical result.

How do we calculate power?

- **Retrospectively** from these 159 meta-analyses.
- To be conservative, we use our **WLS** (which is the same as ‘fixed-effect’) as the proxy for ‘true’ effect.
- Not random-effects or the simple average: both are more biased if there is selective reporting bias (**PB**)
  - **WLS is also biased with PB, but much less so.**
- 3 other proxies for ‘true’ effect are used
  - **Top 10%:** WLS of the most precise 10%
  - **Top 1:** the single most precise estimate
  - **PET-PEESE:** PB corrected estimate, Stan & Douc (2014)

Impotence begets bias

- Low powered studies systematically report larger effects. Why?
  - How else will they be statistical significant?
  - And, most of them are statistically significant!
- **Idea!** Just use the high-powered estimates
  - Like our Top 10. . . but now with a justification

Let’s ‘WAAP’ Publication Bias

WAAP:

- is the **WLS W**eighted **A**verage of only the **A**dequately **P**owered estimates.
- WAAP is onomatopoeia
- It works kind of like……..

It was Yuge!

The Greatest Ever!


WAAP!
www.shutterstock.com · 630414290

The Greatest Ever!

WAAP!

OK, maybe it did not happen

{We wish!}

And, what did happen was an embarrassment

Let’s ‘WAAP’ Publication Bias

{At least, we can do that!}

WAAP:

- is the **W**LS **A**verage of only the **A**dequately **P**owered estimates.
- dominates RE and is somewhat better than WLS— Stanley, Doucouliagos & Ioannidis (2017) “Finding the power to reduce publication bias” *Statistics in Medicine*.
- offers a conservative assessment of bias; that is, an empirical lower bound for bias.
- consistently reduces bias; it does not eliminate it!
- Those meta-analysts who are skeptical of **PET-PEESE** have no reason to object to **WAAP**.

Back to Bias
Assessing ‘**Research Inflation**’

{the exaggeration due to selective reporting}

- **Research Inflation** is the relative difference of the **average reported effect** and some proxy for **‘true’ effect**. It calculates bias, empirically, as a ratio.
- We use **WAAP** and **PET-PEESE** as the proxies for the ‘true’ empirical effect.

How Biased is Economics?

- **The Paldam Principle** (just divide by 2) **is confirmed**
- The median research inflation is just over 100%—that is, typically economics is exaggerated by a factor of 2 or more.
- At least 1/3 of economics is exaggerated by a factor of 4 or more.

Implications

- It’s bad.
- It’s **REALLY bad**!
- Good news: Other disciplines are also **bad**!
  - Psychology routinely has low power.
    - When 100 psyc experiments are replicated, effects shrink by half (Open Science Collaboration; 2015).
    - Our survey of 12,065 effects from 200 psyc meta-analyses find that only 8% of psychological studies are adequately powered –R&R @ *Psyc Bull*
  - In 14,886 medical meta-analyses, the power to detect a medium-size effect is 13%

Follow the power!

All meta-analyses should report median power

C. We learned a lot!

Stylized economic meta-facts

Routinely, economics research has:

- Publication Bias {or selective reporting, small-sample bias, p-hacking or whatever}
- Low Statistical Power
- **High Heterogeneity**

Heterogeneity is our friend

- Otherwise, there is no need to conduct MRA.
- Economics research always has a lot of excess heterogeneity.
  - Among 35 meta-analyses of elasticity, the median heterogeneity is I^2=94% {93% from those 159 metas}
- MRAs of economics research often explain much of this excess heterogeneity.
  - Omitted-variable bias, publication selection and other misspecification biases are typically a large portion of this **systematic** heterogeneity.

High **Unexplained Heterogeneity** is not always our friend

It can:

- overwhelm any signal in the research record
  - Especially, when I^2>90%}
- cause PET to have inflated Type I errors and all methods to have notable remaining publication bias
  - Stanley (2008), D&S (2009), S&D (2012).
- make replication virtually impossible
  - Our survey of psychology finds that the average I^2=74% fully explains recent highly-publicized failures to replicate experiments.

New simulations confirm that high Het can cause the appearance of an effect where there is none.

{More on this later}

Part II: New Developments:

What's old is often new again

"Everything of importance has been said before by somebody who did not discover it." --Whitehead

A. **Unrestricted WLS**: MA & MRA

(S&D, 2015; 2017)

B. **WAAP** (I, S & D, 2017; S, D, & I, 2017)

C. **Meta-Omitted Variable Bias**

(Bruns, 2017)

D. **Limitations of PET-PEESE**

(Stanley, 2017 + new simulations)

A. Unrestricted Weighted Least Squares (WLS)

- Conventional meta-analysis relies on ‘fixed-’ (**FE**) and random-effects (**RE**) weighted averages.
- Because heterogeneity is endemic in research, **RE** is almost **always** used in medicine and psyc.
- **RE** can be very biased if there is publication bias.
- Answer: Unrestricted **WLS**. It
  - uses the same weights as does **FE**, and therefore WLS’s point estimate is always exactly the same as **FE**
  - does not force MSE to be 1, like **FE**, but allows excess heterogeneity to be estimated by the research record.

Unrestricted Weighted Least Squares (WLS)

- **MA**: Simple regression: **t vs. 1/SE ;** no intercept
- **MRA**: Use WLS package with weights, **$1/SE^2$**
- Simulations demonstrate that **WLS** is:
  - Always practically as good as **FE** and **RE**
  - Better than **RE** if there is pub'bias
  - Better than **FE** if there is heterogeneity.
- Implications: Little reason to use random-effects either in MA or MRA {exception: new sim's show that RE is better to test for an effect **if there is no pub'bias**}

Neither Fixed Nor Random:
Weighted Least Squares Meta-Regression and Meta-Analysis

- Stanley T.D. and Doucouliagos, C. 2015. Neither fixed nor random: Weighted least squares meta-analysis, *Statistics in Medicine*, *34:* 2116-27.
- Stanley, T.D. and Doucouliagos, C. 2017. Neither fixed nor random: Weighted least squares meta-regression analysis, *Research Synthesis Methods*, *8*:19-42.
- Stanley, T.D. and Jarrell, S. 1989. Meta-regression analysis: A quantitative method of literature surveys, *Journal of Economic Surveys*, *3*:54–67.

{*What's old is new again*}

B. WAAP (Weighted Average of the Adequately Powered):

- Simple **WLS** weighted average of only those studies that are adequately powered >80%
- **WAAP** consistently reduces publication bias if it is there and inflicts no harm if it is not.
- Systematic reviews of research across the disciplines can be improved by exploiting power.
- Stanley, T.D., Doucouliagos, C. and Ioannidis, J. 2017. "Finding the power to **reduce publication bias**," *Statistics in Medicine*

C. Meta-Omitted Variable Bias

{**Bruns, S. B. 2017**. Meta-regression models and observational research, *Oxford Bulletin of Economics and Statistics,79:*637-53.}

- Bruns' (2017) simulations show that the omission of omitted-variable dummies can cause MRA bias.
- S & D's (2017) sim's demonstrate how including omitted-variable dummies corrects this bias.
- Lessons:
  - Results from any simple MRA or MA can be biased and should not be trusted, in isolation.
  - Always conduct multiple MRA with omitted-variable dummies among many other moderators.

See the **MAER-Net's guidelines**

D. Limitations of PET-PEESE

{**P**recision-**E**ffect **T**est &
Precision **E**ffect **E**stimate with **S**tandard **E**rror}

1. Simulating Psychological Research. {Stanley, T. D. (2017). Limitations of PET-PEESE and other meta-analysis methods. *Social Psychology and Personality Science.*}
   - Simulates Cohen's d from randomized experiments
   - Follows Fraley and Vazire's (2014) large survey of psyc by setting median sample size at 50, and using their **sample size distribution** across studies.
   - Finds that PET can have Type I error inflation with high heterogeneity— Nothing new but important. (S, 2008; D&S, 2009; S&D, 2012),

Design			Average		Bias			Power/Type I Error		
d	**m**	σ_h	**Bias**	I^2	***RE***	***WLS***	*PET-PEESE*	***RE***	***WLS***	***PET***
0	10	0	.2489	.5113	.1957	.1674	.0014	.7977	.4828	0.0000
0	10	6.25	.2506	.5317	.2004	.1707	.0082	.7784	.4955	.0001
0	10	12.5	.2609	.5847	.2156	.1828	.0239	.7481	.4836	.0029
0	10	25	.2917	.7082	.2580	.2171	.0622	.6921	.4717	.0245
0	10	50	.3701	.8602	.3503	.2989	.1367	.6037	.4192	.0574
0	20	0	.2482	.5140	.1958	.1668	.0008	.9942	.9503	.0002
0	20	6.25	.2517	.5409	.2015	.1714	.0086	.9931	.9290	.0005
0	20	12.5	.2603	.6020	.2158	.1824	.0254	.9825	.8833	.0052
0	20	25	.2902	.7367	.2581	.2177	.0714	.9469	.7913	.0342
0	20	50	.3683	.8818	.3502	.2977	.1455	.8654	.6761	.0914
0	40	0	.2484	.5154	.1958	.1667	.0006	1.0000	1.0000	.0002
0	40	6.25	.2515	.5427	.2016	.1712	.0089	1.0000	.9999	.0009
0	40	12.5	.2614	.6102	.2170	.1832	.0275	1.0000	.9988	.0068
0	40	25	.2899	.7486	.2581	.2167	.0746	.9995	.9840	.0544
0	40	50	.3697	.8917	.3521	.2981	.1555	.9935	.9286	.1144
0	80	0	.2487	.5162	.1964	.1673	.0019	1.0000	1.0000	.0005
0	80	6.25	.2516	.5450	.2019	.1714	.0097	1.0000	1.0000	.0009
0	80	12.5	.2604	.6131	.2166	.1828	.0307	1.0000	1.0000	.0109
0	80	25	.2902	.7561	.2587	.2168	.0829	1.0000	1.0000	.0853
0	80	50	.3703	.8958	.3530	.2994	.1739	1.0000	.9990	.1941
Average type I error rate (size)								**.9198**	**.8247**	**.0342**
0.2	10	0	.1665	.2365	.0993	.0863	-.0616	1.0000	.9999	.1262
0.2	10	6.25	.1696	.2780	.1070	.0923	-.0499	.9998	.9993	.1459
0.2	10	12.5	.1808	.3780	.1252	.1045	-.0378	.9991	.9926	.1859
0.2	10	25	.2134	.6038	.1747	.1433	.0037	.9884	.9290	.2150
0.2	10	50	.2970	.8353	.2731	.2191	.0451	.9207	.7783	.2003
0.2	20	0	.1659	.2254	.0986	.0868	-.0345	1.0000	1.0000	.3056
0.2	20	6.25	.1704	.2759	.1065	.0919	-.0297	1.0000	1.0000	.3268
0.2	20	12.5	.1809	.4034	.1266	.1060	-.0126	1.0000	1.0000	.3423
0.2	20	25	.2136	.6509	.1756	.1418	.0167	1.0000	.9983	.3378
0.2	20	50	.2973	.8632	.2762	.2218	.0669	.9973	.9592	.2857
0.2	40	0	.1667	.2206	.0984	.0866	-.0043	1.0000	1.0000	.6265
0.2	40	6.25	.1701	.2797	.1066	.0920	.0019	1.0000	1.0000	.6271
0.2	40	12.5	.1810	.4212	.1269	.1055	.0147	1.0000	1.0000	.6121
0.2	40	25	.2130	.6754	.1763	.1409	.0397	1.0000	1.0000	.5390
0.2	40	50	.2974	.8761	.2768	.2202	.0863	1.0000	.9998	.4152
0.2	80	0	.1663	.2198	.0982	.0864	.0178	1.0000	1.0000	.9244
0.2	80	6.25	.1706	.2839	.1070	.0921	.0239	1.0000	1.0000	.9136
0.2	80	12.5	.1814	.4301	.1275	.1055	.0370	1.0000	1.0000	.8927
0.2	80	25	.2136	.6843	.1772	.1414	.0676	1.0000	1.0000	.8078
0.2	80	50	.2970	.8819	.2768	.2201	.1207	1.0000	1.0000	.6334
0.5	10	0	.0806	.1071	.0277	.0236	-.0243	1.0000	1.0000	.9528
0.5	10	6.25	.0824	.1429	.0301	.0238	-.0282	1.0000	1.0000	.9251
0.5	10	12.5	.0912	.2573	.0421	.0305	-.0339	1.0000	1.0000	.8259
0.5	10	25	.1188	.5344	.0797	.0545	-.0431	1.0000	.9995	.6137
0.5	10	50	.2033	.8102	.1769	.1247	-.0368	.9970	.9599	.3785
0.5	20	0	.0793	.0786	.0252	.0222	-.0239	1.0000	1.0000	.9997
0.5	20	6.25	.0834	.1282	.0300	.0249	-.0224	1.0000	1.0000	.9978
0.5	20	12.5	.0894	.2786	.0401	.0289	-.0209	1.0000	1.0000	.9854
0.5	20	25	.1193	.5905	.0819	.0552	-.0107	1.0000	1.0000	.8578
0.5	20	50	.2020	.8492	.1771	.1197	-.0151	1.0000	.9992	.5567
0.5	40	0	.0799	.0567	.0247	.0226	-.0240	1.0000	1.0000	1.0000
0.5	40	6.25	.0833	.1102	.0290	.0247	-.0229	1.0000	1.0000	1.0000
0.5	40	12.5	.0909	.2892	.0415	.0302	-.0190	1.0000	1.0000	.9998
0.5	40	25	.1192	.6227	.0824	.0542	.0007	1.0000	1.0000	.9838
0.5	40	50	.2021	.8624	.1789	.1196	.0234	1.0000	1.0000	.7905
0.5	80	0	.0804	.0395	.0241	.0227	-.0243	1.0000	1.0000	1.0000
0.5	80	6.25	.0826	.1012	.0280	.0244	-.0231	1.0000	1.0000	1.0000
0.5	80	12.5	.0906	.3028	.0416	.0300	-.0193	1.0000	1.0000	1.0000
0.5	80	25	.1208	.6370	.0846	.0556	.0029	1.0000	1.0000	.9998
0.5	80	50	.2027	.8694	.1800	.1208	.0514	1.0000	1.0000	.9568
Average			**.2016**	**.5083**	**.1575**	**.1291**	**.0175**	**.9976**	**.9904**	**.6822**

Results highlights
50% selection; **typical sample sizes**

Design			Average		Bias			Power/Type I Error		
d	**m**	σ_h	**Bias**	I^2	***RE***	***WLS***	***PET-PEESE***	***RE***	***WLS***	***PET***
0	80	0	.2487	.5162	.1964	.1673	.0019	1.0000	1.0000	.0005
0	80	6.25	.2516	.5450	.2019	.1714	.0097	1.0000	1.0000	.0009
0	80	12.5	.2604	.6131	.2166	.1828	.0307	1.0000	1.0000	.0109
0	80	25	.2902	.7561	.2587	.2168	.0829	1.0000	1.0000	.0853
0	80	50	.3703	.8958	.3530	.2994	.1739	1.0000	.9990	.1941
Average type I error rate (size)								**1.0000**	**.9998**	**.0583**
0.2	80	0	.1663	.2198	.0982	.0864	.0178	1.0000	1.0000	.9244
0.2	80	6.25	.1706	.2839	.1070	.0921	.0239	1.0000	1.0000	.9136
0.2	80	12.5	.1814	.4301	.1275	.1055	.0370	1.0000	1.0000	.8927
0.2	80	25	.2136	.6843	.1772	.1414	.0676	1.0000	1.0000	.8078
0.2	80	50	.2970	.8819	.2768	.2201	.1207	1.0000	1.0000	.6334
Average			**.2450**	**.5826**	**.2013**	**.1683**	**.0566**	**1.0000**	**1.0000**	**.8344**

Notes: RE, WLS denotes the random-effects and unrestricted weighted least squares meta-analysis averages, respectively, and PET-PEESE is the meta-regression publication bias corrected estimate.

Limitations/Lessons

- If there are **only a few research studies**, regression-based PET-PEESE is not reliable.
- **Extreme high heterogeneity** is a problem
  - PET can have high Type I errors– 19.4% vs 5%
  - Nothing new: (Stanley, 2008)
  - PET-PEESE is upwardly biased too, but much less so than conventional meta-analysis
  - Random-effects are very invalid with or without extreme heterogeneity– Type I error rates are 100%.
- PET's power is very low **when no study has adequate power**. {This is the only truly new result} Still, all other methods are much worse.

D. Limitations of PET-PEESE cont.

2. Simulating Economics Research with more realistic research parameters {work in progress}

- Calibrated from 35 economic elasticity meta-analyses.

Simulations duplicates the median values of:

- Research Inflation (Paldam Principle)
- **Distribution of SEs** {widely dispersed; this is new}
- I^2= 94% & other values {higher than past sims}
- Sample size {larger than past sims}

Table 1: Bias, MSE, power and level of alternative meta-methods with 50% selective reporting

Design			Bias					MSE				Power/Type I Error					Average
ξ	m	I^2	Mean	*RE*	*WLS*	*PET-PEESE*	*WAAP*	*RE*	*WLS*	*PET-PEESE*	*WAAP*	*RE*	*WLS*	*PET*	*WAAP*	*FAT*	*|WAAP-PP|*
0	100	.6753	.1602	.0587	.0266	.0075	.0207	.00348	.00073	.00018	.00048	1.0000	.9966	.1558	.7686	1.0000	.0146
0	100	.7385	.1598	.0663	.0335	.0199	.0257	.00442	.00120	.00072	.00081	1.0000	.9850	.4739	.6246	.9994	.0103
0	100	.8541	.1635	.0860	.0471	.0355	.0384	.00748	.00252	.00201	.00199	1.0000	.9497	.6200	.4813	.8967	.0094
0	100	.9443	.1813	.1248	.0750	.0628	.0658	.01579	.00682	.00618	.00619	1.0000	.8998	.6748	.4528	.4940	.0096
0	100	.9828	.2304	.1958	.1311	.1164	.1206	.03905	.02155	.02060	.02089	1.0000	.8721	.7004	.5989	.2361	.0102
0	400	.6779	.1598	.0583	.0266	.0127	.0167	.00340	.00071	.00028	.00032	1.0000	1.0000	.4303	.7637	1.0000	.0091
0	400	.7477	.1599	.0662	.0334	.0270	.0234	.00439	.00113	.00084	.00061	1.0000	1.0000	.8332	.7880	1.0000	.0067
0	400	.8659	.1635	.0856	.0465	.0417	.0365	.00736	.00224	.00191	.00151	1.0000	1.0000	.9171	.7982	1.0000	.0063
0	400	.9521	.1817	.1251	.0752	.0709	.0663	.01572	.00598	.00550	.00495	1.0000	1.0000	.9466	.8222	.9060	.0058
0	400	.9856	.2303	.1964	.1308	.1265	.1237	.03873	.01823	.01746	.01692	1.0000	.9989	.9597	.8886	.4833	.0044
0	1000	.9537	.1813	.1251	.0753	.0725	.0671	.01568	.00579	.00538	.00469	1.0000	1.0000	.9991	.9878	.9982	.0054
Average type I error rate (size) and Power for FAT												1.0000	.9729	.7010	.7250	.8194	
.15	100	.4157	.0943	.0116	.0034	.0008	.0005	.00016	.00005	.00004	.00004	1.0000	1.0000	1.0000	1.0000	.9926	.0007
.15	100	.6145	.0965	.0178	.0039	.0012	.0007	.00037	.00013	.00012	.00014	1.0000	1.0000	1.0000	1.0000	.9255	.0009
.15	100	.8213	.1030	.0346	.0090	.0063	.0054	.00130	.00049	.00047	.00051	1.0000	1.0000	.9999	.9999	.7111	.0014
.15	100	.9378	.1238	.0716	.0307	.0278	.0266	.00538	.00234	.00229	.00237	1.0000	1.0000	.9961	.9919	.4407	.0020
.15	100	.9819	.1762	.1430	.0845	.0789	.0798	.02121	.01180	.01213	.01190	1.0000	.9947	.9500	.9468	.2408	.0034
.15	400	.4294	.0946	.0114	.0033	.0007	.0004	.00014	.00002	.00001	.00001	1.0000	1.0000	1.0000	1.0000	1.0000	.0004
.15	400	.6372	.0963	.0179	.0040	.0014	.0010	.00033	.00005	.00003	.00004	1.0000	1.0000	1.0000	1.0000	1.0000	.0006
.15	400	.8421	.1032	.0349	.0090	.0064	.0057	.00124	.00019	.00015	.00016	1.0000	1.0000	1.0000	1.0000	.9884	.0008
.15	400	.9470	.1243	.0724	.0315	.0288	.0283	.00531	.00134	.00120	.00121	1.0000	1.0000	1.0000	1.0000	.8025	.0008
.15	400	.9846	.1761	.1435	.0849	.0823	.0823	.02080	.00840	.00803	.00808	1.0000	1.0000	.9996	.9996	.4516	.0008
.15	1000	.9487	.1246	.0727	.0314	.0288	.0284	.00531	.00113	.00098	.00096	1.0000	1.0000	1.0000	1.0000	.9840	.0006
0.3	100	.3156	.0635	.0034	.0011	-.0007	.0003	.00004	.00004	.00004	.00004	1.0000	1.0000	1.0000	1.0000	.7196	.0010
0.3	100	.5720	.0639	.0059	.0012	-.0007	.0003	.00009	.00012	.00013	.00012	1.0000	1.0000	1.0000	1.0000	.5445	.0010
0.3	100	.8244	.0674	.0132	.0017	-.0003	.0007	.00029	.00045	.00047	.00046	1.0000	1.0000	1.0000	1.0000	.3775	.0010
0.3	100	.9431	.0819	.0373	.0081	.0059	.0068	.00168	.00168	.00174	.00173	1.0000	1.0000	1.0000	1.0000	.3025	.0010
0.3	100	.9823	.1301	.0996	.0487	.0462	.0471	.01074	.00755	.00768	.00763	1.0000	1.0000	.9973	.9989	.2260	.0015
0.3	400	.3343	.0632	.0033	.0011	-.0008	.0002	.00002	.00001	.00001	.00001	1.0000	1.0000	1.0000	1.0000	.9951	.0010
0.3	400	.6033	.0639	.0060	.0012	-.0007	.0003	.00005	.00003	.00003	.00003	1.0000	1.0000	1.0000	1.0000	.9440	.0010
0.3	400	.8454	.0673	.0135	.0016	-.0003	.0007	.00021	.00011	.00012	.00012	1.0000	1.0000	1.0000	1.0000	.7271	.0010
0.3	400	.9515	.0823	.0379	.0082	.0061	.0071	.00151	.00049	.00048	.00048	1.0000	1.0000	1.0000	1.0000	.5519	.0010
0.3	400	.9848	.1295	.0994	.0485	.0462	.0474	.01010	.00363	.00348	.00357	1.0000	1.0000	1.0000	1.0000	.3778	.0012
0.3	1000	.9528	.0821	.0378	.0079	.0058	.0068	.00146	.00023	.00020	.00022	1.0000	1.0000	1.0000	1.0000	.8318	.0011
Average		.7954	.1267	.0660	.0341	.0292	.0297	.00737	.00325	.00306	.00301	1.0000	.9998	.9974	.9971	.6880	.0035

Bias, power and level of alternative meta-methods with 50% selective reporting

Design			Bias					Power/Type I Error					Ave.
ξ	m	I^2	**Mean**	***RE***	***WLS***	***PET-PEESE***	***WAAP***	***RE***	***WLS***	***PET***	***WAAP***	***FAT***	**|*WAAP-PP*|**
0	100	.9443	.1813	.1248	.0750	.0628	.0658	1.00	.8998	.6748	.4528	.4940	.0096
0	100	.9828	.2304	.1958	.1311	.1164	.1206	1.00	.8721	.7004	.5989	.2361	.0102
0	400	.9521	.1817	.1251	.0752	.0709	.0663	1.00	1.000	.9466	.8222	.9060	.0058
0	400	.9856	.2303	.1964	.1308	.1265	.1237	1.00	.9989	.9597	.8886	.4833	.0044
Average type I error rate (size) and Power for FAT								**1.00**	**.9729**	**.7010**	**.7250**	**.8194**	
.15	100	.9378	.1238	.0716	.0307	.0278	.0266	1.00	1.000	.9961	.9919	.4407	.0020
.15	100	.9819	.1762	.1430	.0845	.0789	.0798	1.00	.9947	.9500	.9468	.2408	.0034
Average		**.7954**	**.1267**	**.0660**	**.0341**	**.0292**	**.0297**	**1.00**	**.9998**	**.9974**	**.9971**	**.6880**	**.0035**

Notes: ξ is the true elasticity, **m** is the number of estimates, I^2 is the proportion of the observed variation among reported elasticities that cannot be explained by their reported standard errors, ***RE, WLS*** denotes the random-effects and unrestricted weighted least squares meta-analysis averages, respectively, ***PET-PEESE*** is the meta-regression publication bias corrected estimate, ***WAAP*** is the weighted average of the adequately powered, ***PET*** is the precision-effect test, ***FAT*** is the funnel-asymmetry test, |***WAAP-PP*** | is the average absolute difference between ***WAAP*** and ***PET-PEESE***.

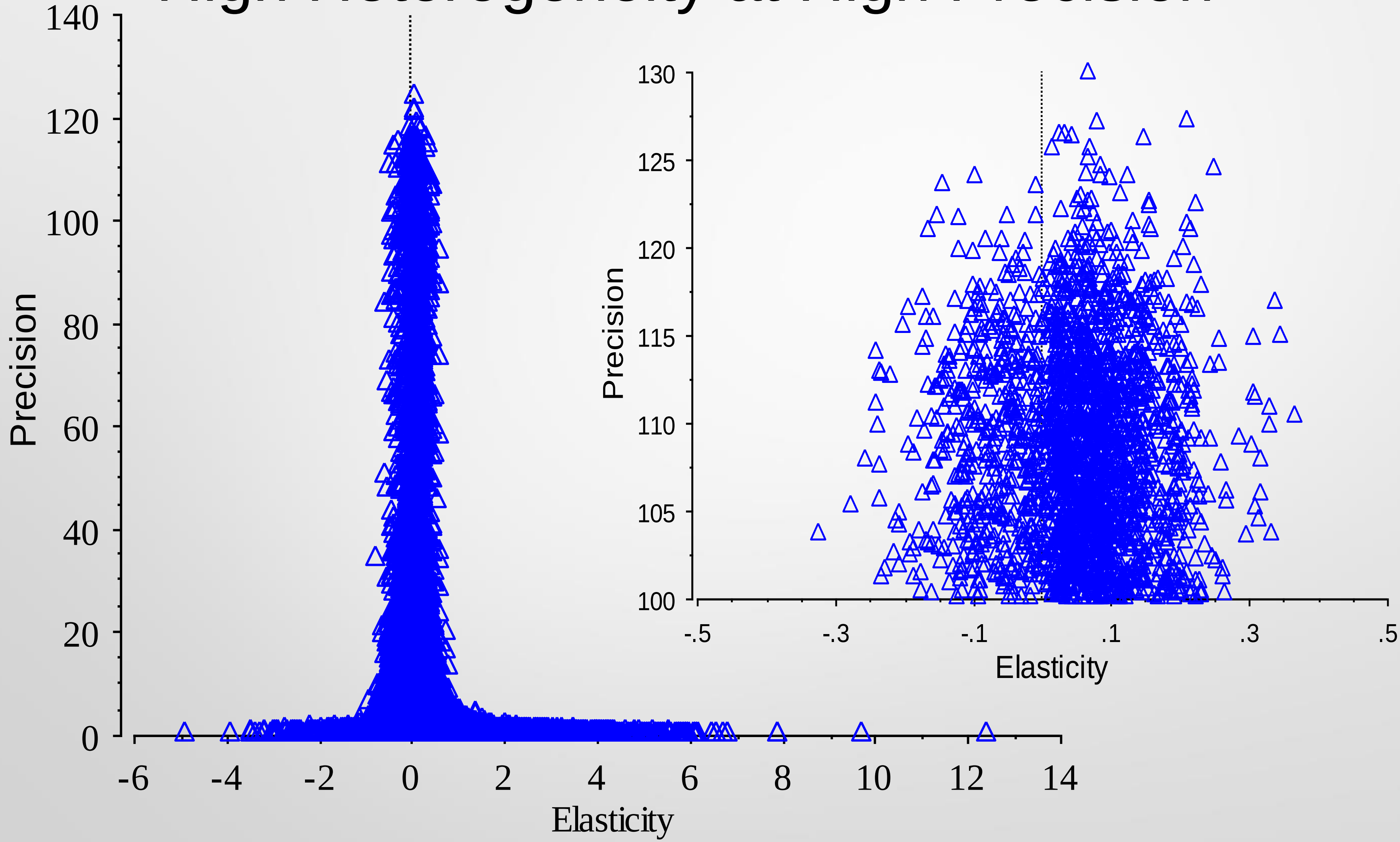
Results

PET-PEESE and WAAP:

- are virtually identical (difference is .01 or less)
- have half the bias as RE {even at the highest levels of Heterogeneity}
- reduce pub'bias by two-thirds {at highest levels of Heterogeneity} by three-fourths {overall}
- **are only as good as the best research!**
- **have unacceptable Type I errors** at high het, but **RE and other methods are worse**.
- **Implications: Cannot trust simple MRAs**

  {FAT is good; except at highest heterogeneity}

Cause of PET's Type I error inflation: High Heterogeneity at High Precision



In practice, there is less heterogeneity (and less selection?) at high precision.

Typically, it’s probably not this bad, but **Technical Solutions** are **not sufficient**, either.

- In most actual metas, heterogeneity is not constant, but correlated with SE.
  - simulations that account for this proportionality have lower Type I errors; but still a problem at highest het.
- Robust SEs help.
- Bias of PET-PEESE & WAAP is practically small.
- In practice, the question is always whether the effect is practically significant; not different than 0.
  - When testing against practical significance (.05 or .1), Type I error inflation largely goes away.

Implications for meta-analysis practice

- None: If you find no evidence of an overall effect, new sims only strengthen that finding.
- Can never know whether there is a genuine small effect vs. no effect, exaggerated by pub'bias
- Worse Case: Meta-analysis is merely a survey/ summary of the best economics research.
- Still: Well-conducted MRAs will often succeed in identifying a few of the genuine drivers of the research record and in reducing some of the largest biases.

III: Recommendations for Practice

{little has changed}

- Be modest about your meta-results. {Not New}
  - Be **especially cautious** if there is pub'bias, high heterogeneity and evidence of a small effect.
- Emphasize practical significance. {Not New}
- Interpret all simple MAs and MRAs as descriptive or indicative, not definitive. {Not really new}
- Routinely conduct multiple MRA with many moderator variables. {Nothing new there}
- Always report I^2 and median power

{**finally, something New!**}

Need for further research

We know that MRAs with a dummy variable and SE can remove omitted variable bias and greatly reduce most of the publication bias; however,

We do not know whether:

- using many dummy variables will also work, or whether some complex structure of their interactions will be required. {work in progress}
- panel methods are as good as theory suggests
- there are better, more resilient, pub'bias methods.

Thank You!

Stanley@Hendrix.edu

References

Bruns, S. B. 2017. Meta-regression models and observational research, *Oxford Bulletin of Economics and Statistics*,79:637-53.

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.

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

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

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. (2017). Limitations of PET-PEESE and other meta-analysis methods. *Social Psychology and Personality Science*. [Epub ahead of print] doi: 10.1177/1948550617693062

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.

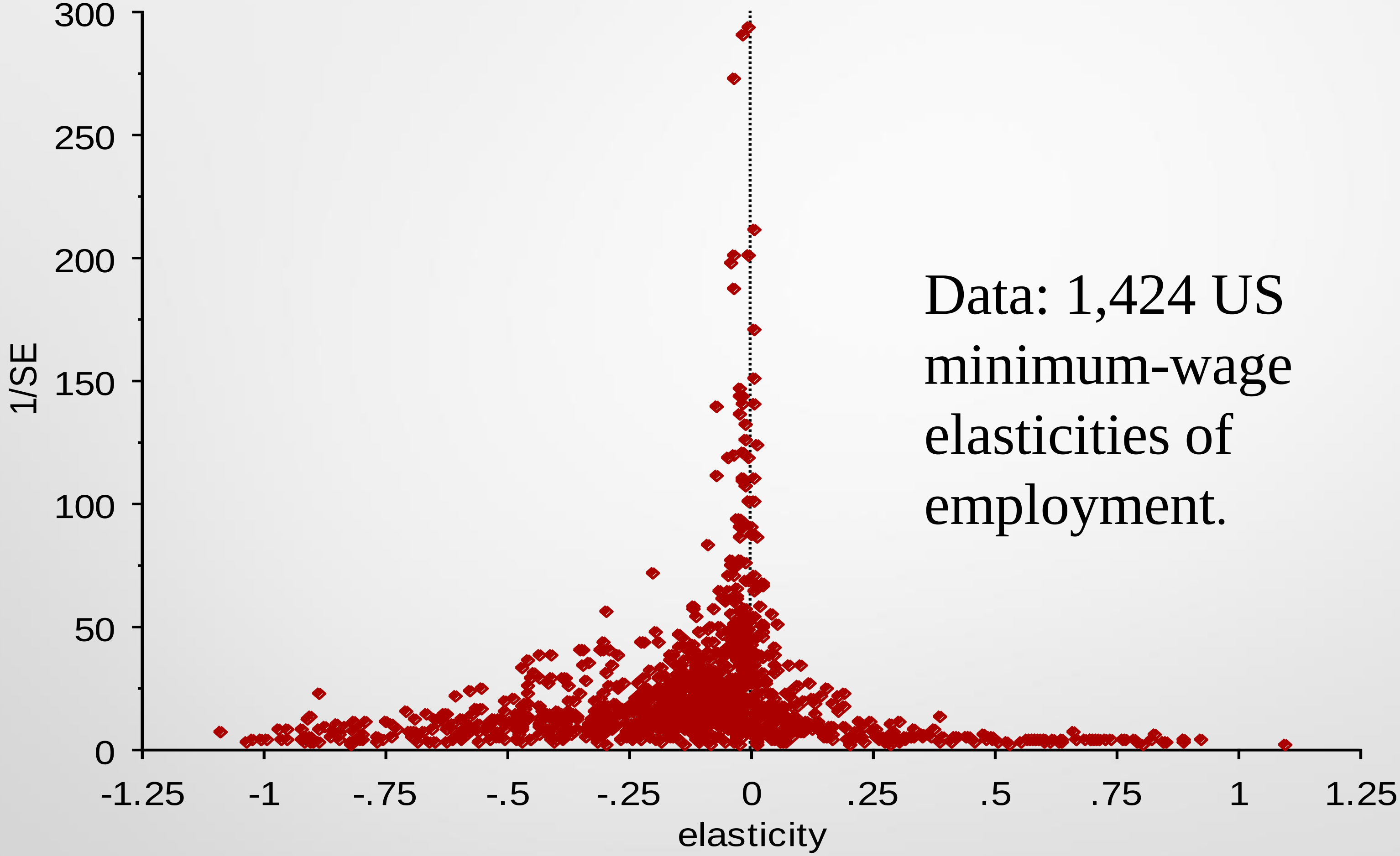
Stanley, T.D. and Doucouliagos, C. "Neither fixed nor random: Weighted least squares meta-analysis," Statis*tics in Medicine* 34 (2015), 2116-27.

Stanley, T.D. and Doucouliagos, C. 2017. "Neither fixed nor random: Weighted least squares meta-regression analysis," Res*earch Synthesis Methods*, 8, 19-42.

Stanley, T.D., Doucouliagos, C.& Ioannidis, J.P.A. 2017. "Finding the power to reduce publication bias." *Statistics in Medicine*, [Epub ahead of print]. doi: 10.1002/sim.7228

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