Star Wars: The Empirics Strike Back
Online Appendix

Abel Brodeur[a] Mathias Lé[b] Marc Sangnier[c]
Yanos Zylberberg[d]

June 2015

Reporting guidelines

Rules we used during the data collecting process are described below.[1] The reporting guidelines for each information are presented in table A1.

The main idea here is to keep things very simple and transparent even if it implies measurement errors.

- *Reporting concerns articles published in the American Economic Review, the Journal of Political Economy, and the Quarterly Journal of Economics between 2005 and 2011, except the AER Papers and Proceedings. All articles that include at least one empirical tests should be included in the sample.*
- *Report data as they appear in the published version.*
- *Report exactly what is printed in the paper. Do not allow the spreadsheet to round anything up (this may be done by setting the data format to text before entering data).*
- *Take all the variables of interest, even if the authors state in the comments that it is a good thing that a variable is not significant (an interaction term for example). Define variables of interest by looking at the purpose of each table and at comments of the table in the text. A rule of thumb may be that a variable of interest is commented in the*

[a]University of Ottawa
[b]Banque de France, ACPR, and Paris School of Economics)
[c]Aix-Marseille Univ. (Aix-Marseille School of Economics), CNRS & EHESS
[d]School of Economics Finance and Management, University of Bristol

[1]As highlighted in the introduction, collecting tests of central hypotheses necessitates a good understanding of the argument developed in the articles and a strict process avoiding any subjective selection of tests. The first observation restricts the set of potential research assistants to economists and the only economists with a sufficiently low opportunity cost were ourselves. We tackle the second issue by being as conservative as possible, and by avoiding interpretations of the intentions of the authors.

text alongside the paper. Explicit control variables are not variables of interest. However, variables use to investigate "the determinants of" something are variables of interest. When the status of a variable of a test is unclear, be conservative and report it.

- *Report solely tests from the main tables of the paper. That is, report results, but do not report descriptive statistics, group comparisons, and explicit placebo tests. Only report tests of the tables that are commented under the headings "main results", "extensions", or "robustness check". Let us impose that robustness checks and extensions should be entitled as such by the authors to be considered as non-main results. Regarding first-stages in a two-stage regression, keep only the second-stage except if the first-stage is presented as a contribution of the paper by itself.*
- *See the input mask for the reporting rules.*
- *For tests, fill up one spreadsheet per issue. For authors' data, look for their curriculum vitae using a search engine and fill up information by taking care that some are author-invariant (such as the PhD date), but other vary with published articles (such as status).*

Example of the reporting process

Below an excerpt of an imaginary example. We only report coefficients and standard deviations for the first variable. First-stage is not presented as a major finding by the imaginary researchers and other variables are explicitly presented as control variables. If the authors of this fictive table had put the last three columns in a separate table under the heading "robustness check", we would have considered this new table as a "non-main" table. Table A2 presents the displayed results of this imaginary example.

> *Some of the* 1024 *economists of our dataset engage in some project that requires heavy data capture. It would be optimal that only individuals with a low opportunity cost engage in such an activity. We want to investigate the relationship between the choice of this time-consuming activity and the intrinsic productivity of these agents in their main job. We first estimate the relationship between data capture activity and the number of research papers released by each economist (the number of papers is a proxy for productivity). As shown by column* 1 *in table A2, there is a significant and negative correlation between the probability to engage in data capture and the number of written papers. However, since data capture is a time-consuming activity, we suspect that it may also influence the number of papers written. Thus, the average number of sunny days per year at each department is used as an instrument for productivity. The first-stage regression is reported in column* 2 *and the second stage in column* 3*. The last columns reproduce the same*

exercise using additional co-variates. We conclude that the decision to engage in data capture is definitely negatively related to productivity. Note that controls have the expected sign, in particular the number of co-authors.

Construction of empirical inputs

The method to construct the distributions of test statistics from existing datasets is as follows: (i) We randomly draw a variable from the dataset. This variable will be the *dependent variable*. (ii) Conditional on the dependent variable being a numeric variable, we randomly draw three other variables that will serve as *independent variables*, excluding numeric variables whose correlation coefficient with the dependent variable exceeds 0.95. If any of the three dependent variables contains less than 50 distinct string or integer values, we treat it as a categorical variable and create as many dummies as distinct values. (iii) We regress the dependent variable on the independent variables using ordinary least squares and store the test statistic (the ratio of coefficient to the standard error) associated to the first independent variable (or to the second dummy variable derived from the first independent variable if the latter has been treated as a categorical variable).

We repeat this procedure until we obtain $2,000,000$ test statistics. This procedure has been applied to four different datasets: the World Development Indicators, the Quality of Government standard time series dataset, the Panel Study of Income Dynamics, and the Vietnam Household Living Standards Survey.

Discontinuities

In order to test if there are any discontinuities around the thresholds of significance, we create bins of width 0.00125 and count the number n_z of z-statistics ending up in each bin. This partition produces $1,600$ bins for approximately $20,000$ z-statistics between 0 and 2 in the full sample.

Before turning to the results, let us detail two concerns about our capacity to detect discontinuities. First, tests in an article are presented as a set of arguments which smooth the potential discrepancy around the threshold. Second, numbers are often roughly rounded in articles such that it is difficult to know whether a statistic is slightly above or below a certain threshold.

Figure A1 plots n_z for the full sample around 1.65 (10%) and 1.96 (5%). There does not seem to be strong discontinuities around the thresholds. The use of a regression around the thresholds with a trend and a quadratic term $P_2(z)$ confirms the visual impressions. To achieve this, we estimate the following expression:

$$n_z = \gamma \mathbb{1}_{z>\tilde{z}} + P_2(z) + \varepsilon_z, \quad z \in [\tilde{z} - \nu, \tilde{z} + \nu].$$

Table A3 details the different values taken by γ along the different thresholds $\tilde{z} \in \{1.65, 1.96, 2.57\}$ and the different windows $\nu \in \{0.05, 0.10, 0.20, 0.30\}$

around the significance thresholds. The four panels of this table document the estimated discontinuities for the *full* sample, the *eye-catcher* sample, the *no eye-catchers* sample and the *eye-catcher without theoretical contribution* sample.

The small number of observations in windows of width 0.10 and 0.05 does not allow us to seize any effect in these specifications, with very high standard errors. Nonetheless, even focusing on wider windows, the only sizable effect is at the 10% significance threshold: γ is around .02 for samples of tests reported with eye-catchers. Whichever the sample, the other discontinuities are never statistically different from 0. For articles using eye-catchers, the 10%-discontinuity is around 0.02. This implies a jump of density of $\frac{0.02}{0.24} \approx 8\%$ at the threshold.

Overall, the discontinuities are small and concentrated on articles reporting stars and around the 10% threshold. It is to be noted that, if the experimental literature tends to favor 1% or 5% as thresholds for rejecting the null, the vast majority of empirical papers now considers the 10% threshold as the first level for significance. Our results tend to illustrate the adoption of 10% as a norm.

These findings are of small amplitude, maybe because of the smoothing or because there are multiple tests in a same article. More importantly, even the 10% discontinuity is hard to interpret. Authors might select tests who pass the significance thresholds among the set of acceptable specifications and only show part of the whole range of inferences. But it might also be that journals prefer significant results or authors censor themselves, expecting journals to be harsher with unsignificant results.

An extension to the Benford's law

Following the seminal paper of Benford (1938) and its extension by Hill (1995, 1998), the leading digit of numbers taken from scale-invariant data or selected from a lot of different sources should follow a logarithmic distribution. Tests of this law are applied to situations such as tax monitoring in order to see whether reported figures are manipulated or not.

The intuition is that we are in the precise situation in which the Benford's law should hold: we group coefficients from different sources, with different units and thus different scales. According to Hill (1995, 1998), the probability for one of our collected coefficients to have a leading digit equal to $d \in \{0, \ldots, 9\}$ is $\log_{10}(1 + 1/d)$.

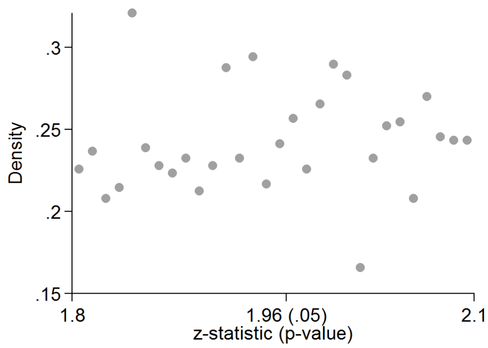
We group coefficients and standard deviations taken from our different articles (with different units). z-statistics are not used in this analysis as it is not scale-invariant and normalized across the different articles. For both coefficients and standard errors, we compute the empirical probability r_d to have a leading digit equal to $d \in \{0, \ldots, 9\}$.

The first columns of table A4 (panel A for the coefficients, panel B for the standard errors) display the theoretical probabilities and the empirical counterparts for three samples: the full sample, the sample of coefficients and standard errors for which the associated z-statistic was in the $[1.65, 6]$ interval, and the others. All samples seem incredibly close to the theoretical predictions. The distance to the theoretical prediction can be summarized by the statistic $U = \sqrt{\sum_{d=1}^{\infty}(r_d - log_{10}(1 + 1/d))^2}$.

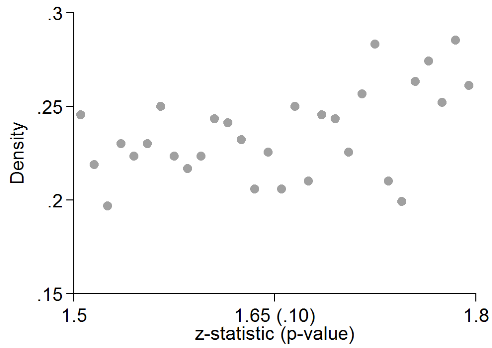
A concern is that we have no benchmark upon which we can rely: the distance to the theoretical prediction seems small but it may be because of the very large number of observations. We re-use our random regressions on WDI, and randomly extract approximately as many coefficients and standard deviations as in the real sample. We already know that the distribution of z-statistics are quite close; the decomposition of coefficients and standard errors with z-statistics between 1.65 and 6 should partition the placebo sample in similar proportions as the real sample. The three last columns of panels A and B present the results on this placebo sample.

The comparison of the two sets of results indicates that the distance of the full sample to the theoretical predictions is higher than with the placebo both for coefficients ($U_s = 0.0089$ against $U_p = 0.0058$) and for standard errors ($U_s = 0.0117$ against $U_p = 0.0046$). The analysis across subsample is less straightforward. As regards the coefficients, the relative distance compared to the placebo is particularly large between 1.65 and 6 ($U_s = 0.0161$ against $U_p = 0.0036$ for the $[1.65, 6]$ sample, and $U_s = 0.0081$ against $U_p = 0.0093$ for the rest). It seems however that this observation does not hold for standard errors. Naturally, part of this discrepancy can be explained by the differences in the number of observations: there are less numbers reported between 1.65 and 6 in the placebo test. To conclude, this Benford analysis provides some evidence indicating non-randomness in our universe of tests.

Figure A1: Discontinuitites around the thresholds (values grouped by bins of bandwith 0.1).



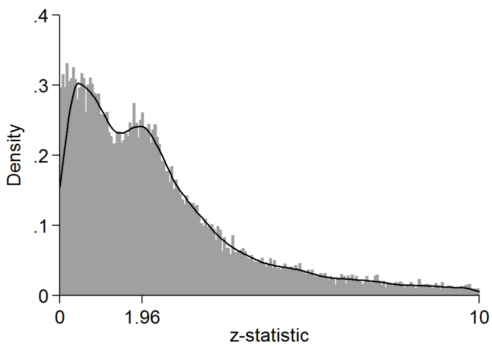
(a) Around 1.96 (.05).



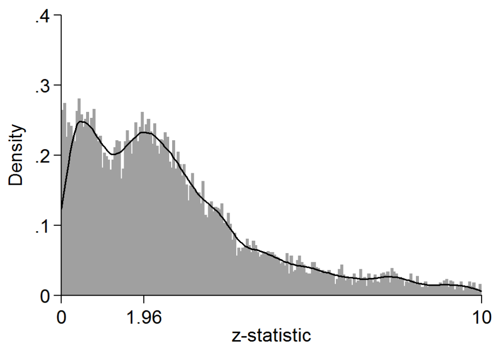
(b) Around 1.65 (.10).

Sources: AER, JPE, and QJE (2005-2011).

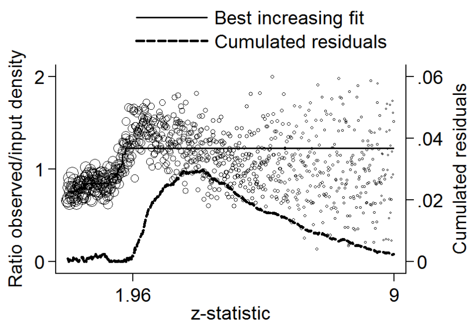
Figure A2: Distributions of z-statistics and estimations of inflation along aggregated research fields.



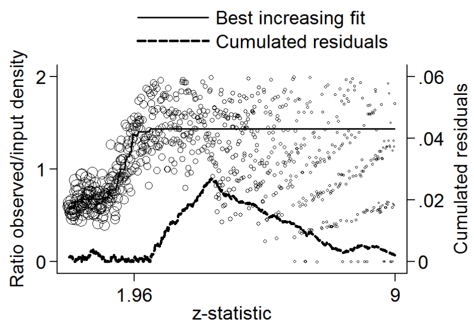
(a) Microeconomics.



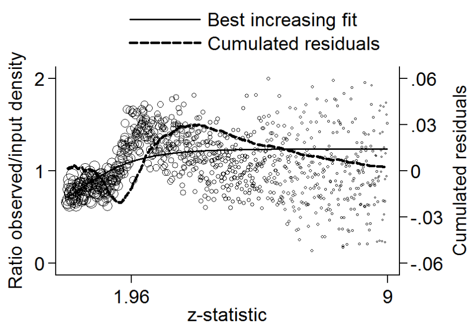
(b) Macroeconomics.



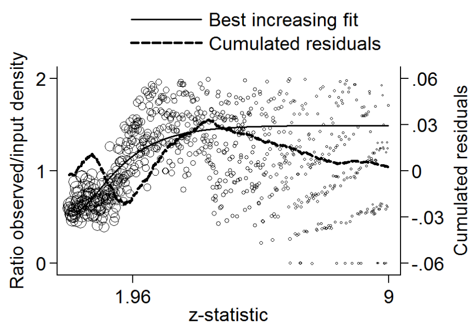
(c) Microeconomics, non-parametric estimation.



(d) Macroeconomics, non-parametric estimation.



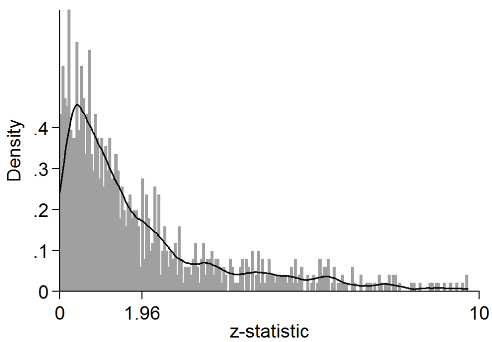
(e) Microeconomics, parametric estimation.



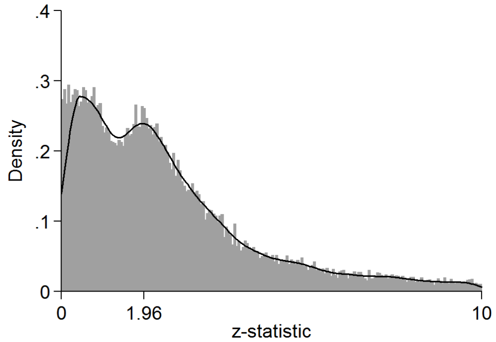
(f) Macroeconomics, parametric estimation.

Sources: AER, JPE, and QJE (2005-2011). The input used is the empirical WDI input.

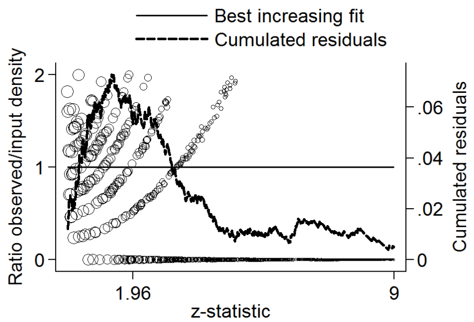
Figure A3: Distributions of z-statistics and estimations of inflation along the nature of empirical evidence.



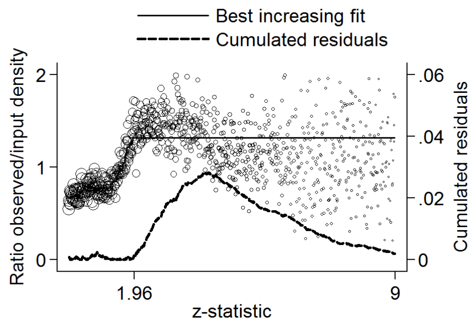
(a) Null results.



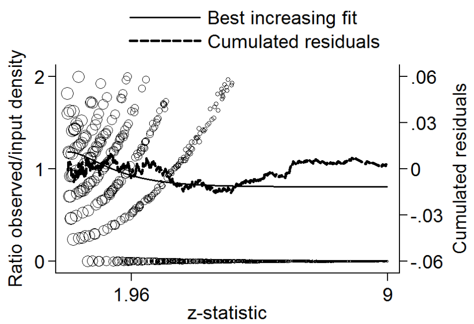
(b) Positive results.



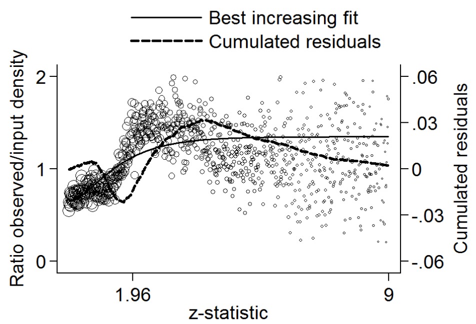
(c) Null results, non-parametric estimation.



(d) Positive results, non-parametric estimation.



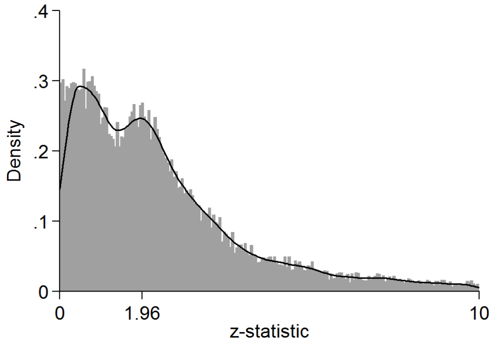
(e) Null results, parametric estimation.



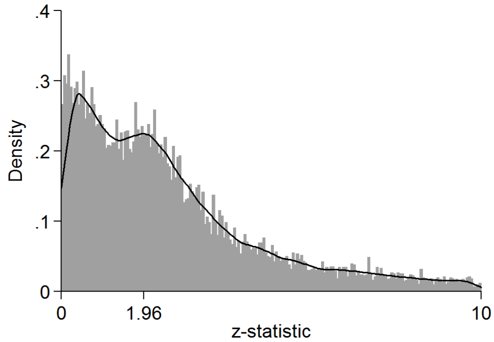
(f) Positive results, parametric estimation.

Sources: AER, JPE, and QJE (2005-2011). The input used is the empirical WDI input.

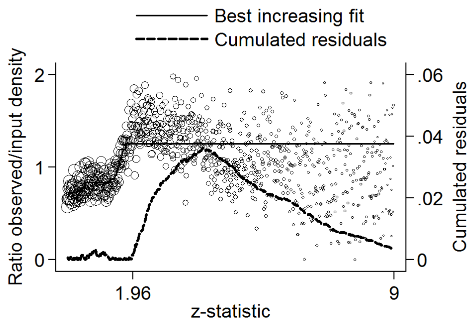
Figure A4: Distributions of z-statistics and estimations of inflation along the use of eye-catchers.



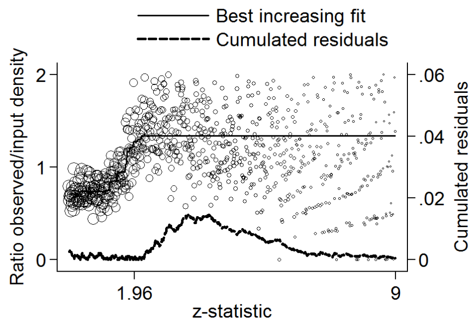
(a) Eye-catchers.



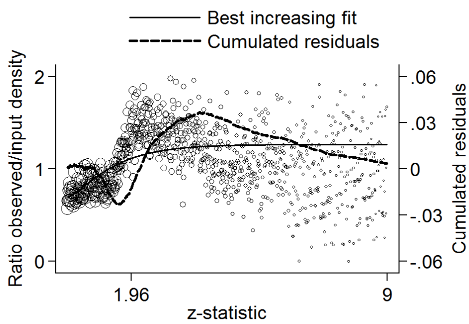
(b) No eye-catchers.



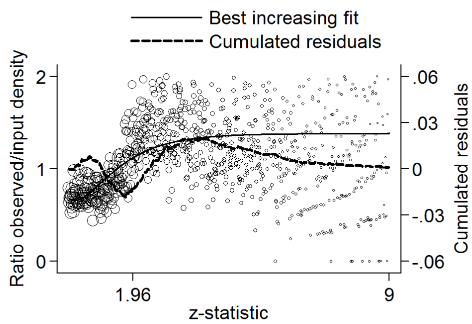
(c) Eye-catchers, non-parametric estimation.



(d) No eye-catchers, non-parametric estimation.



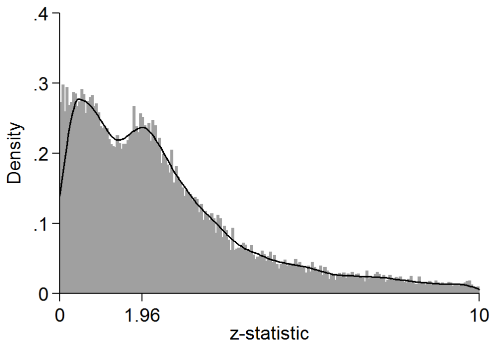
(e) Eye-catchers, parametric estimation.



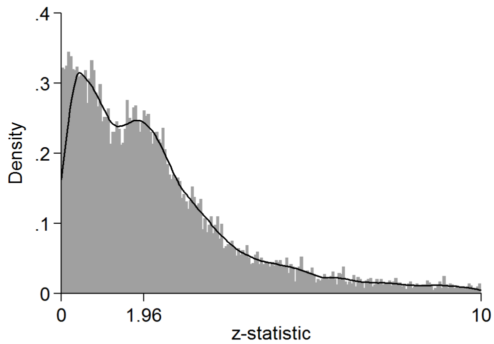
(f) No eye-catchers, parametric estimation.

Sources: AER, JPE, and QJE (2005-2011). The input used is the empirical WDI input.

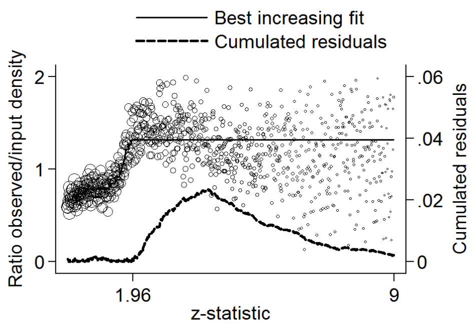
Figure A5: Distributions of z-statistics and estimations of inflation along status of results.



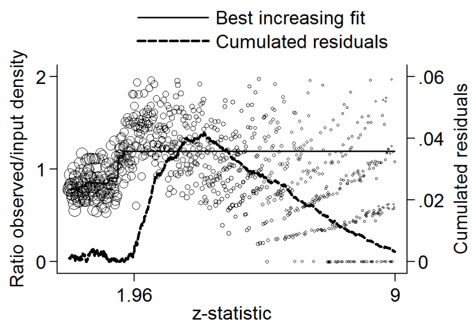
(a) Main results.



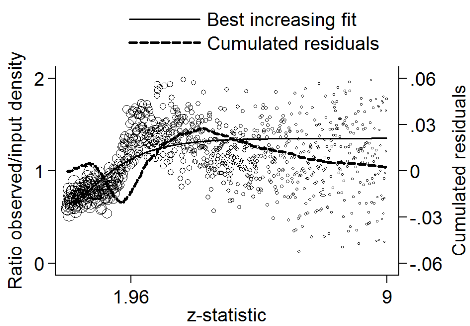
(b) Non-main results.



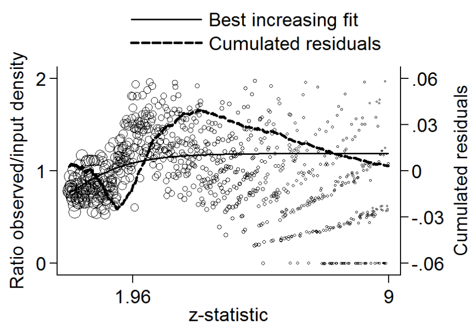
(c) Main results, non-parametric estimation.



(d) Non-main results, non-parametric estimation.



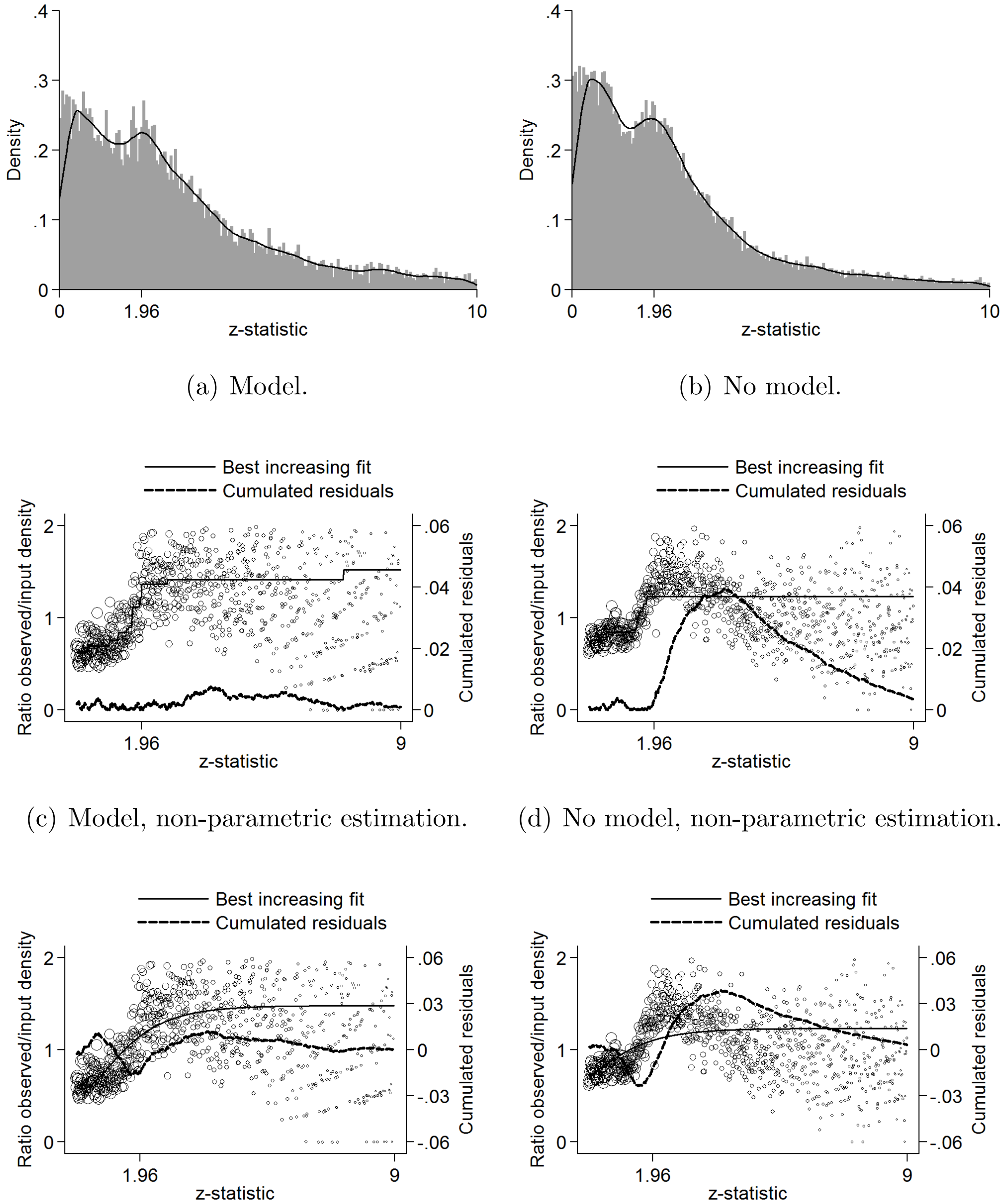
(e) Main results, parametric estimation.



(f) Non-main results, parametric estimation.

Sources: AER, JPE, and QJE (2005-2011). The input used is the empirical WDI input.

Figure A6: Distributions of z-statistics and estimations of inflation along the presence of a theoretical contribution.



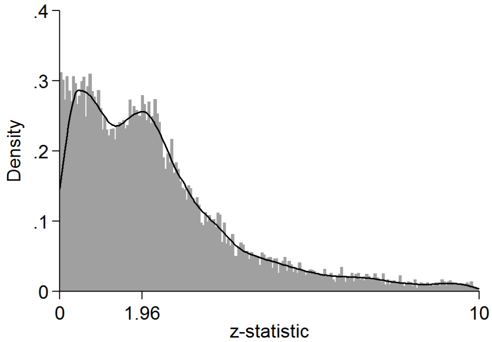
(a) Model. (b) No model.

(c) Model, non-parametric estimation. (d) No model, non-parametric estimation.

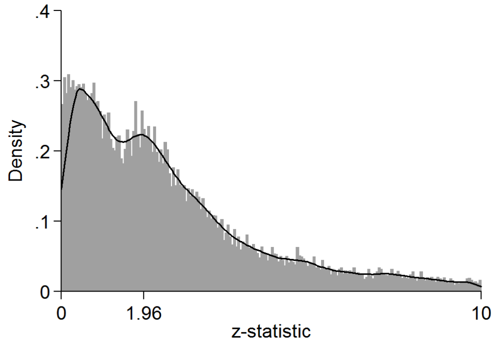
(e) Model, parametric estimation. (f) No model, parametric estimation.

Sources: AER, JPE, and QJE (2005-2011). The input used is the empirical WDI input.

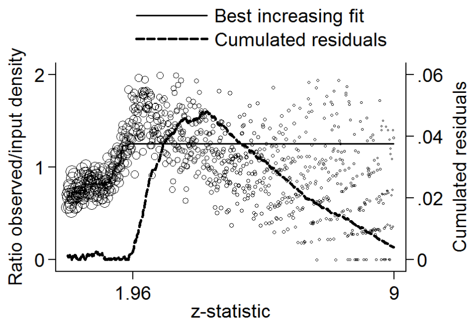
Figure A7: Distributions of z-statistics and estimations of inflation along PhD-age.



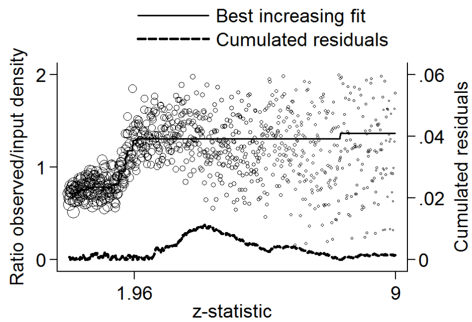
(a) Low PhD-age.



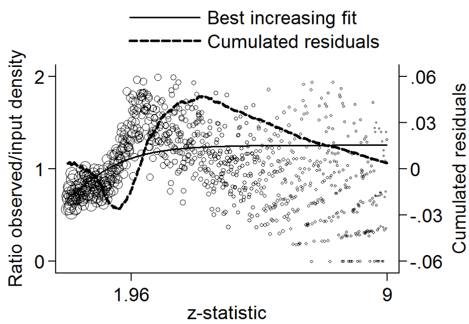
(b) High PhD-age.



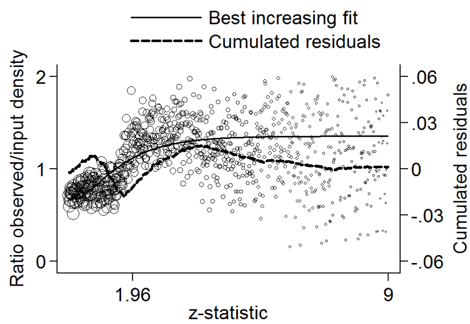
(c) Low PhD-age, non-parametric estimation.



(d) High PhD-age, non-parametric estimation.



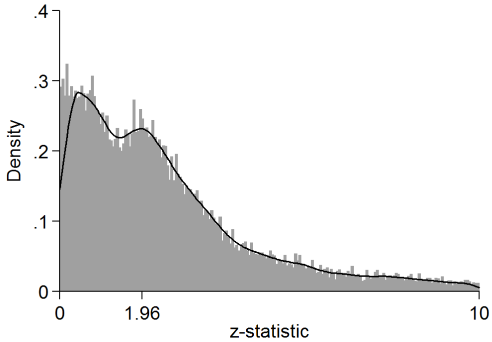
(e) Low PhD-age, parametric estimation.



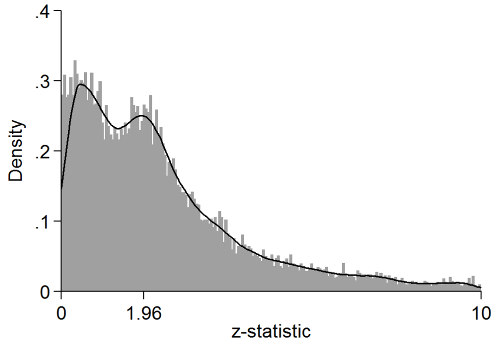
(f) High PhD-age, parametric estimation.

Sources: AER, JPE, and QJE (2005-2011). The input used is the empirical WDI input.

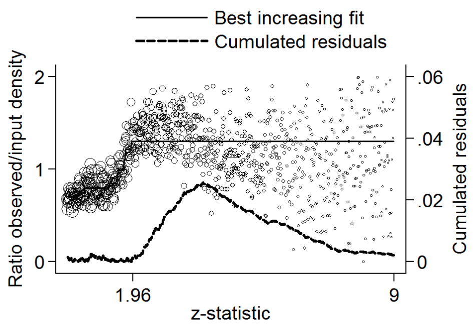
Figure A8: Distributions of z-statistics and estimations of inflation along the presence of editors among authors.



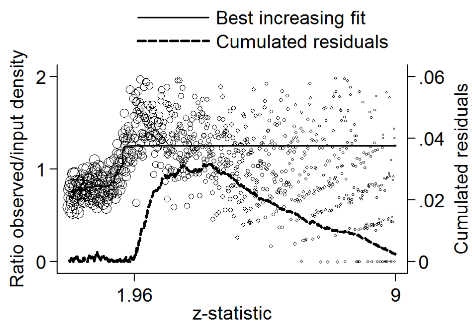
(a) At least one editor.



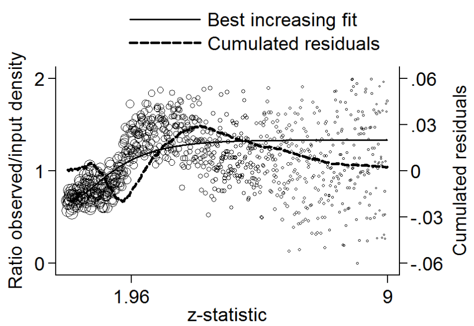
(b) No editor.



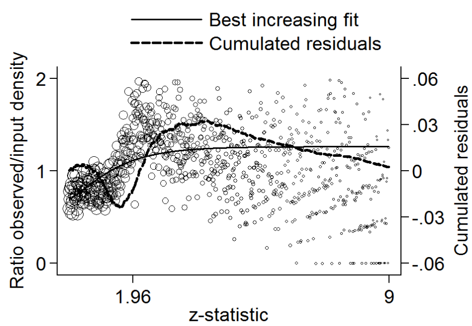
(c) At least one editor, non-parametric estimation.



(d) No editor, non-parametric estimation.



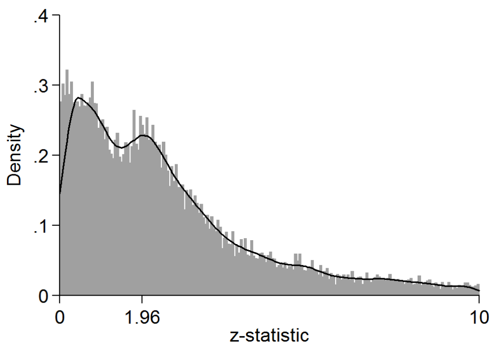
(e) At least one editor, parametric estimation.



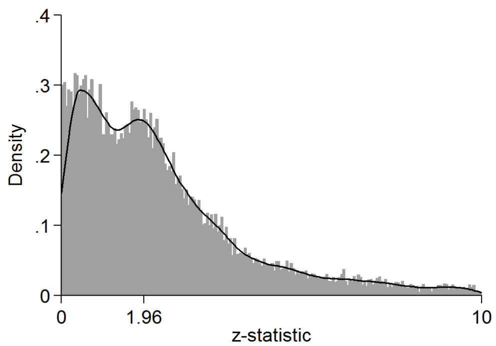
(f) No editor, parametric estimation.

Sources: AER, JPE, and QJE (2005-2011). The input used is the empirical WDI input.

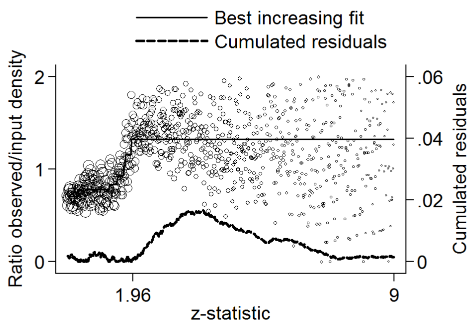
Figure A9: Distributions of z-statistics and estimations of inflation along the presence of tenured researchers among authors.



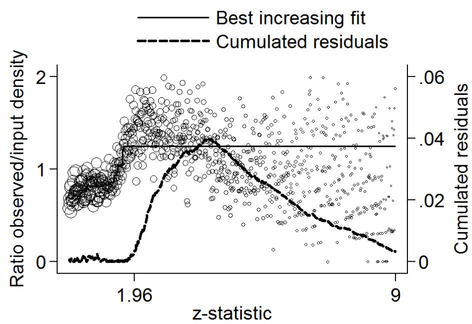
(a) At least one tenured author.



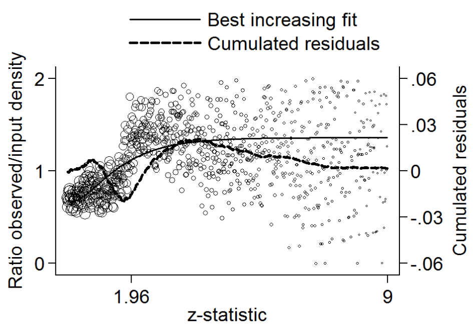
(b) No tenured author.



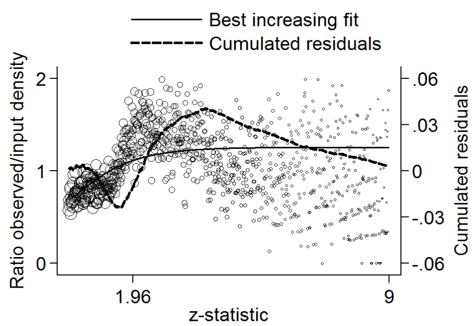
(c) At least one tenured author, non-parametric estimation.



(d) No tenured author, non-parametric estimation.



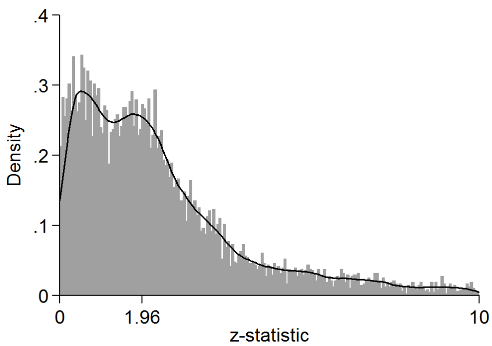
(e) At least one tenured author, parametric estimation.



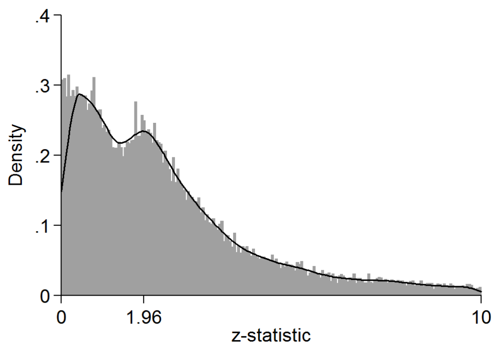
(f) No tenured author, parametric estimation.

Sources: AER, JPE, and QJE (2005-2011). The input used is the empirical WDI input.

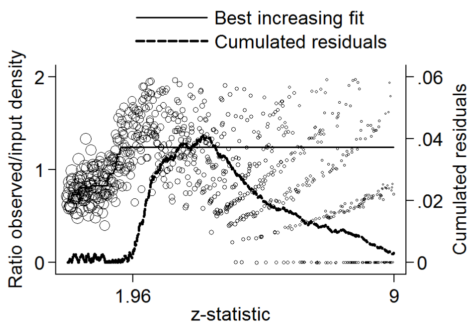
Figure A10: Distributions of z-statistics and estimations of inflation along co-authorship.



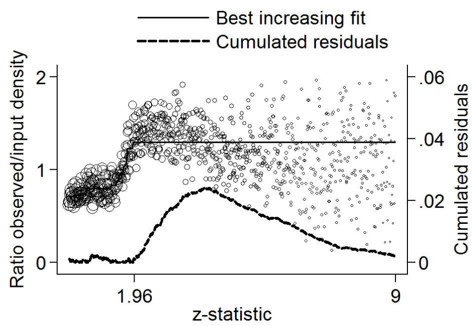
(a) Single authored.



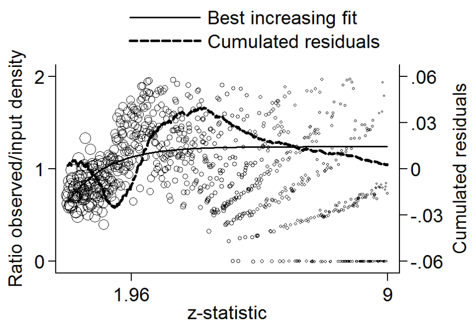
(b) Co-authored.



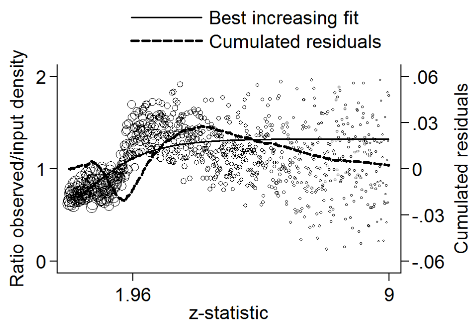
(c) Single authored, non-parametric estimation.



(d) No tenured author, non-parametric estimation.



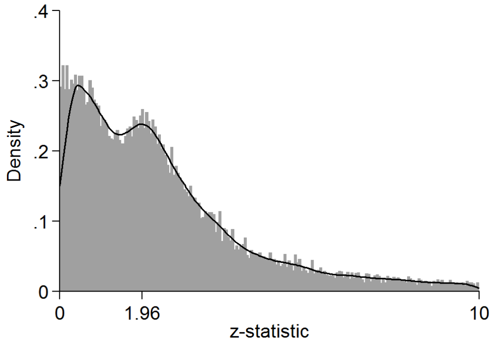
(e) Single authored, parametric estimation.



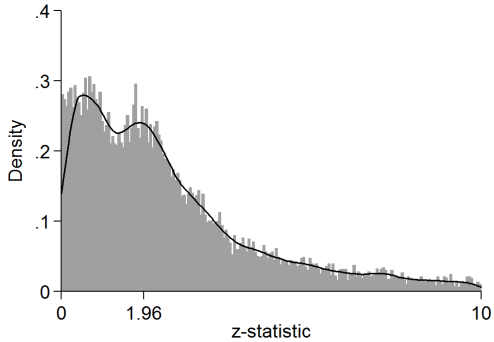
(f) Co-authored, parametric estimation.

Sources: AER, JPE, and QJE (2005-2011). The input used is the empirical WDI input.

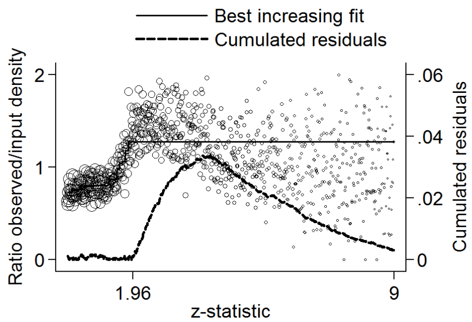
Figure A11: Distributions of z-statistics and estimations of inflation along the use of research assistants.



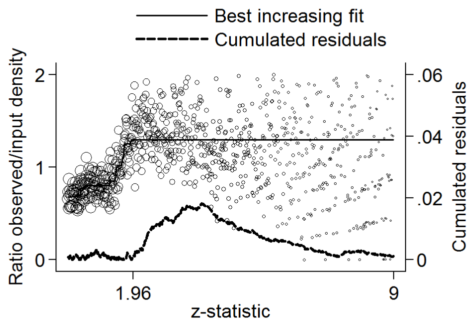
(a) With research assistants.



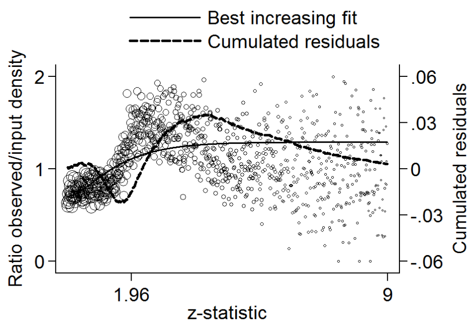
(b) Without research assistants.



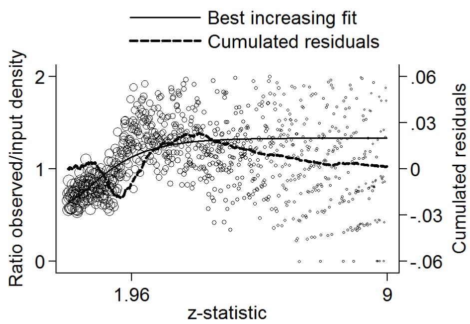
(c) With research assistants, non-parametric estimation.



(d) Without research assistants, non-parametric estimation.



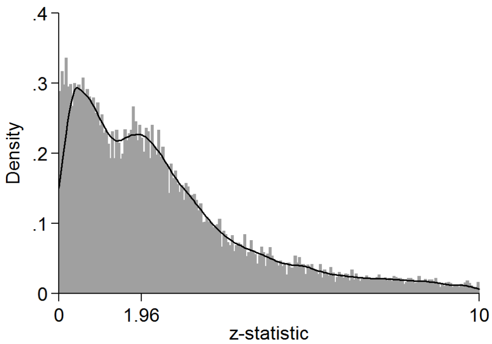
(e) With research assistants, parametric estimation.



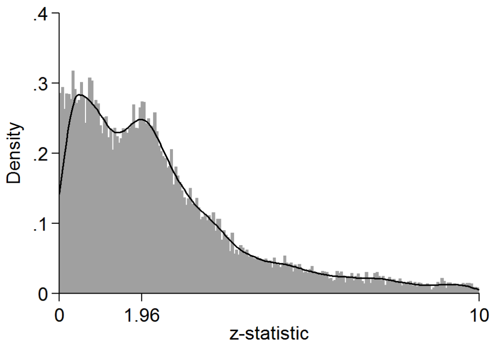
(f) Without research assistants, parametric estimation.

Sources: AER, JPE, and QJE (2005-2011). The input used is the empirical WDI input.

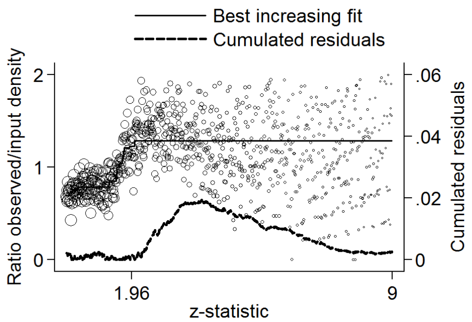
Figure A12: Distributions of z-statistics and estimations of inflation along the number of individuals thanked.



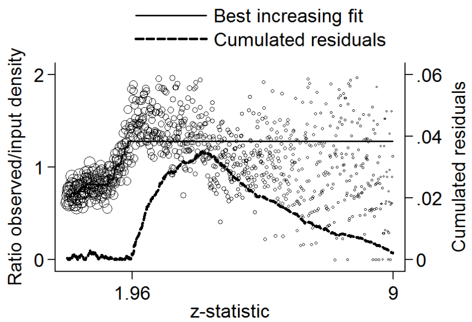
(a) Low number of thanks.



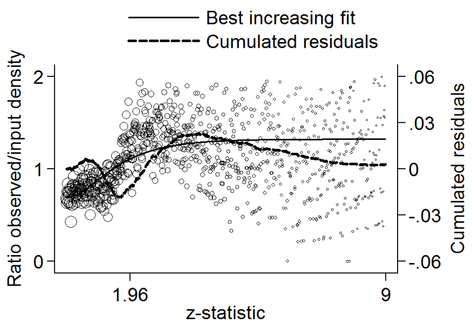
(b) High number of thanks.



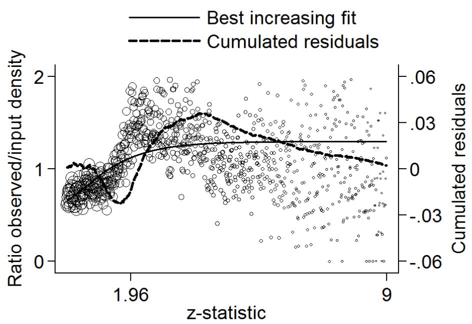
(c) Low number of thanks, non-parametric estimation.



(d) High number of thanks, non-parametric estimation.



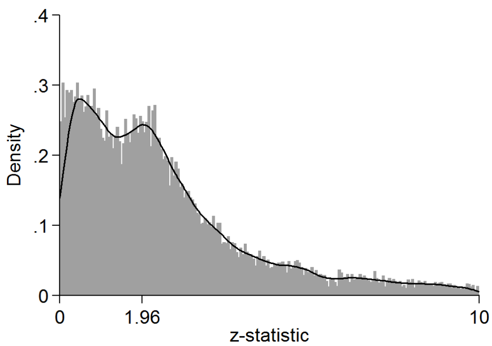
(e) Low number of thanks, parametric estimation.



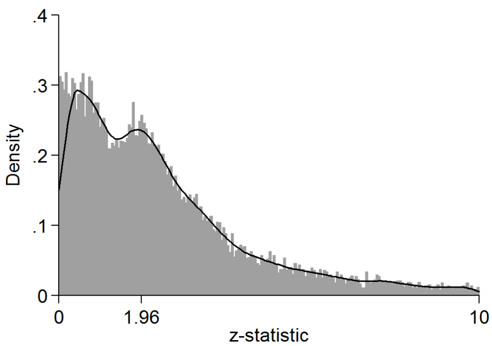
(f) High number of thanks, parametric estimation.

Sources: AER, JPE, and QJE (2005-2011). The input used is the empirical WDI input.

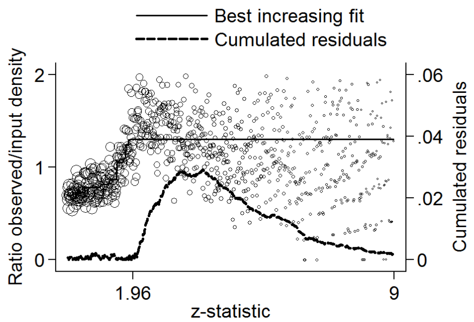
Figure A13: Distributions of z-statistics and estimations of inflation along the availability of data and codes on the journal's website.



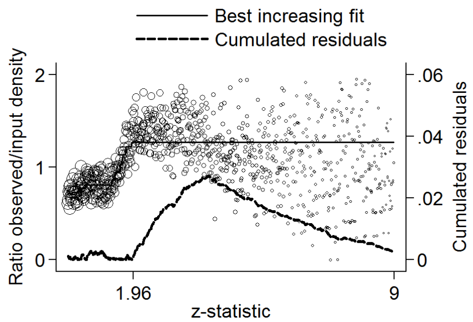
(a) Data and codes available.



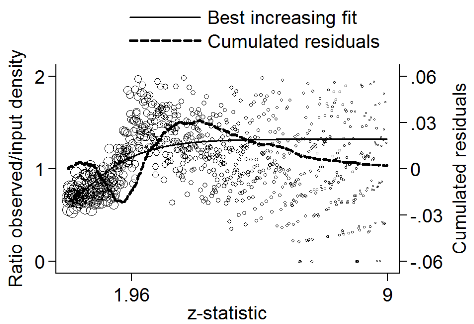
(b) Data or codes not available.



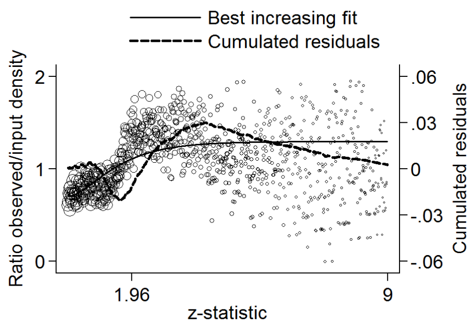
(c) Data and codes available, non-parametric estimation.



(d) Data or codes not available, non-parametric estimation.



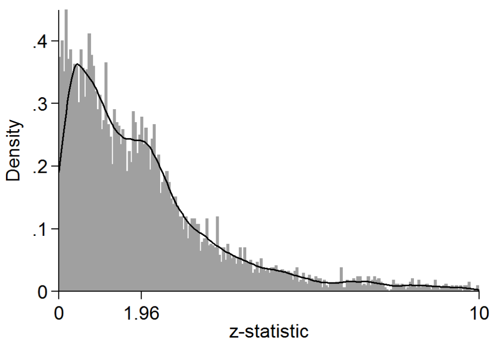
(e) Data and codes available, parametric estimation.



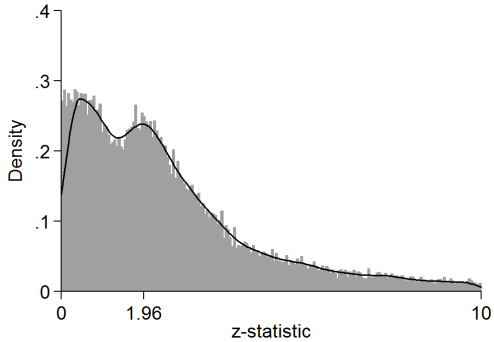
(f) Data or codes not available, parametric estimation.

Sources: AER, JPE, and QJE (2005-2011). The input used is the empirical WDI input.

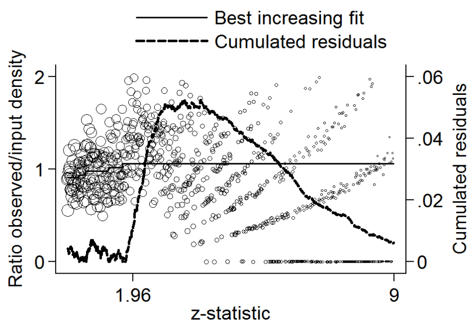
Figure A14: Distributions of z-statistics and estimations of inflation along the type of data used.



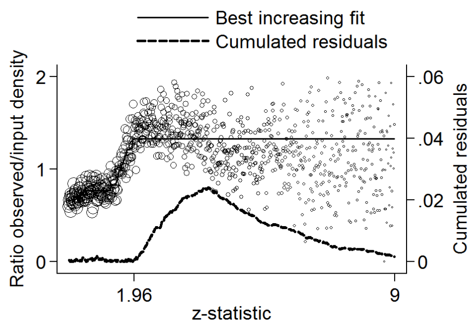
(a) Lab. experiments or RCT data.



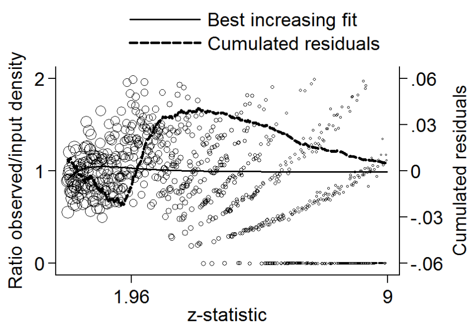
(b) Other data.



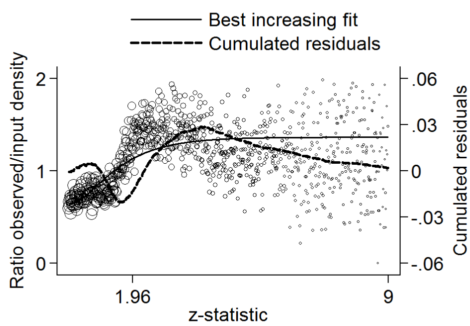
(c) Lab. experiments or RCT data, non-parametric estimation.



(d) Other data, non-parametric estimation.



(e) Lab. experiments or RCT data, parametric estimation.



(f) Other data, parametric estimation.

Sources: AER, JPE, and QJE (2005-2011). The input used is the empirical WDI input.

Table A1: Reported information.

Article-specific variables	
journal_id	Journal identifier. Use full text, e.g. “American Economic Review”, “Journal of Political Economy”, and “Quarterly Journal of Economics”.
issue	Enter as $XXX.Y$, where XXX is *volume* and Y is *issue*.
year	Enter as ♯♯♯♯.
article_page	Enter the number of the article’s first page in the issue.
first_author	Last name of the first author as it appears.
num_authors	Number of authors.
jel	Enter JEL codes as a dot-separated list if present.
model	Does the paper include a theoretical contribution? Answer by “yes” or “no”.
ras	Enter the number of research assistants thanked.
thanks	Enter the number of individuals thanked (excluding research assistant and referees).
negative_result	Does the paper explicitly put forward null results? Answer by “no”, “yes”, or “nes” if both positive and null results are presented as contributions.
field	Enter the rough field, i.e. “macroeconomics” or “microeconomics”.
field_2	Enter a finer field definition, e.g. agriculture, environment, health, development, finance, industrial organization, macroeconomics, international economics, microeconomics, labor, public economics, urban economics, or other (history, institutions, law and economics, miscellaneous).
Test-specific variables	
type	Type of data used. Three choices: “expe” for experimental data, “rct” for data obtained from randomized control trials, and “other” for any other type of data.
table_panel	Enter the table identifier. If a table includes several parts (e.g. panels), also enter the panel identifier. In case where results are presented only in the text, create a separate identifier for each “group” of results in the paper.
type_emp	Statistical method or type of test used.
row	Row identifier when referring to a table. Enter the page number when referring to results presented in the text.
column	Column identifier when referring to a table. Enter the order of appearance when referring to results presented in the text.
stars	Does the paper use eye-catchers (e.g. stars or bold printing) to highlight statistical significance? Answer by “yes” or “no”.

Continued on next page

Table A1: Reported information (continued).

main	Enter "yes" or "no". For simplicity, assign the same status to all tests presented in the same table or results'group.
coefficient	Enter the coefficient.
standard_deviation	Enter the standard deviation. Create a variable called "standard_deviation_2" if authors reported multiple standard deviations (e.g. first is clustered, the second is not).
t_stat	Enter the test statistic. Create a variable called "t_stat_2" if authors reported multiple statistics.
p_value	Enter the p-value. Create a variable called "p_value_2" if authors reported multiple p-values.
Author-specific variables	
phd	PhD institution. Enter "Unknown" if you cannot find the information.
phd_date	Enter as ♯♯♯♯ the year at which the PhD was awarded. Enter the expected date if necessary. Enter "Unknown" if you cannot find the information.
Author×article-specific variables	
author	Author's name
status	Enter the status of the author at time of publication. Enter "Unknown" if you cannot find the information.
status_1y	Enter the status of the author one year before publication.
status_2y	Enter the status of the author two years before publication.
status_3y	Enter the status of the author three years before publication.
editor	Is the author an editor or a member of an editorial board at time of publication? Answer by "yes" or "no". Enter "Unknown" if you cannot find the information.
editor_before	Was the author an editor or a member of an editorial board before publication? Answer by "yes" or "no".
affiliation_1	Enter the author's first affiliation as it appears on the published version of the article.
affiliation_2	Enter the author's other affiliations as they appear on the published version of the article.

Table A2: Imaginary example: Activity choice and productivity among economists.

Dependent variable:	(1) OLS Data capture	(2) First stage Papers	(3) IV Data capture	(4) OLS Data capture	(5) First stage Papers	(6) IV Data capture
Papers	-0.183***		-0.190	-0.189***		-0.243*
	(0.024)		(0.162)	(0.018)		(0.134)
Sunny days		-0.360***			-0.372***	
		(0.068)			(0.068)	
Gender				-0.028	0.090**	-0.067**
				(0.023)	(0.039)	(0.031)
Size of department				-0.000	0.001	-0.000
				(0.001)	(0.002)	(0.002)
Number of co-authors				0.223***	0.004	0.222***
				(0.008)	(0.014)	(0.010)
Country fixed effects				Yes	Yes	Yes
Observations	1,024	1,024	1,024	1,024	1,024	1,024
R-squared	0.054	0.027		0.475	0.042	

*** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses. Data are imaginary. The model is a linear probability model estimated with ordinary least squares. *Data capture* is a dummy equal to 1 if the economist captures data. *Papers* is the number of papers (including working papers). *Sunny days* is the number of sunny days per year at each department. *Size of the department* is the size of the department. *Number of co-authors* is the number of co-authors. In columns 3 and 6, *papers* is instrumented by *sunny days*. All regressions include a constant term.

Table A3: Amplitude of discontinuities around the 0.10, 0.05, and 0.01 significance thresholds.

Panel A	Full sample				Panel B	Eye-catchers			
Window width	0.30	0.20	0.10	0.05	Window width	0.30	0.20	0.10	0.05
Observations	480	320	160	80	Observations	480	320	160	80
0.10 threshold					0.10 threshold				
γ	0.0157	0.0033	0.0168	0.0262	γ	0.0268	0.0204	0.0206	0.0365
	(0.0115)	(0.0144)	(0.0207)	(0.0330)		(0.0150)	(0.0186)	(0.0266)	(0.0393)
	[0.232]	[0.232]	[0.232]	[0.228]		[0.238]	[0.238]	[0.242]	[0.240]
0.05 threshold					0.05 threshold				
γ	0.0049	0.0152	0.0064	0.0015	γ	0.0160	0.0187	0.0106	0.0018
	(0.0121)	(0.0147)	(0.0189)	(0.0247)		(0.0147)	(0.0180)	(0.0235)	(0.0310)
	[0.241]	[0.244]	[0.242]	[0.260]		[0.248]	[0.253]	[0.251]	[0.273]
0.01 threshold					0.01 threshold				
γ	0.0009	-0.0051	-0.0061	0.0103	γ	0.0106	0.0075	0.0047	0.0341
	(0.0104)	(0.0129)	(0.0187)	(0.0263)		(0.0125)	(0.0155)	(0.0232)	(0.0348)
	[0.187]	[0.187]	[0.182]	[0.176]		[0.188]	[0.187]	[0.182]	[0.184]
Panel C	No eye-catchers				Panel D	Eye-catchers and no model			
Window width	0.30	0.20	0.10	0.05	Window width	0.30	0.20	0.10	0.05
Observations	480	320	160	80	Observations	480	320	160	80
0.10 threshold					0.10 threshold				
γ	-0.0086	-0.0306	0.0130	0.0113	γ	0.0225	0.0280	0.0295	0.0293
	(0.0214)	(0.0266)	(0.0385)	(0.0613)		(0.0170)	(0.0211)	(0.0296)	(0.0448)
	[0.225]	[0.223]	[0.213]	[0.204]		[0.246]	[0.248]	[0.255]	[0.255]
0.05 threshold					0.05 threshold				
γ	-0.0279	-0.0003	-0.0148	-0.0116	γ	-0.0048	-0.0152	-0.0223	-0.0637
	(0.0222)	(0.0281)	(0.0405)	(0.0586)		(0.0173)	(0.0212)	(0.0284)	(0.0386)
	[0.230]	[0.232]	[0.232]	[0.244]		[0.251]	[0.254]	[0.251]	[0.272]
0.01 threshold					0.01 threshold				
γ	-0.0192	-0.0253	-0.0284	-0.0379	γ	0.0160	0.0236	0.0287	0.0420
	(0.0186)	(0.0229)	(0.0331)	(0.0504)		(0.0145)	(0.0179)	(0.0260)	(0.0385)
	[0.192]	[0.191]	[0.184]	[0.165]		[0.190]	[0.190]	[0.186]	[0.190]

Each element of the table is the result of a separate linear regression. $n_z = \gamma(z > \tilde{z}) + P_2(z) + \varepsilon_z, \quad z \in [\tilde{z} - \nu, \tilde{z} + \nu]$ where ν and $\tilde{z}$ are the window width and the threshold considered and n_z the density of z-statistics in each bin. Robust standard errors are reported in parentheses. The mean of the variable of interest–the density of the distribution in the window–is reported between brackets. The results are shown omitting the coefficients for the polynomial controls.

Table A4: Tests of Benford's law on the full sample, and partition $[1.65, 6]$ vs. $[0, 1.65) \cup (6, \infty)$.

Panel A		Coefficients					
			Collected data			Simulated data	
Leading digit	Theoretical	Full sample	$[1.65, 6]$	$[0, 1.65) \cup (6, \infty)$	Full sample	$[1.65, 6]$	$[0, 1.65) \cup (6, \infty)$
1	0.3010	0.3052	0.3095	0.3017	0.3016	0.3026	0.3011
2	0.1761	0.1692	0.1675	0.1706	0.1735	0.1774	0.1717
3	0.1249	0.1240	0.1166	0.1301	0.1283	0.1244	0.1302
4	0.0969	0.0981	0.0972	0.0989	0.0937	0.0973	0.0920
5	0.0792	0.0772	0.0767	0.0776	0.0809	0.0796	0.0816
6	0.0669	0.0666	0.0676	0.0657	0.0658	0.0669	0.0653
7	0.0580	0.0592	0.0604	0.0583	0.0586	0.0559	0.0599
8	0.0512	0.0524	0.0542	0.0509	0.0510	0.0520	0.0506
9	0.0458	0.0480	0.0503	0.0461	0.0464	0.0439	0.0477
U		0.0089	0.0161	0.0081	0.0058	0.0036	0.0093
Observations		43,954	19,892	24,062	43,297	14,015	29,282

Panel B		Standard errors					
			Collected data			Simulated data	
Leading digit	Theoretical	Full sample	$[1.65, 6]$	$[0, 1.65) \cup (6, \infty)$	Full sample	$[1.65, 6]$	$[0, 1.65) \cup (6, \infty)$
1	0.3010	0.2928	0.2955	0.2906	0.3002	0.3025	0.2992
2	0.1761	0.1751	0.1798	0.1711	0.1732	0.1730	0.1733
3	0.1249	0.1267	0.1283	0.1253	0.1277	0.1279	0.1276
4	0.0969	0.1046	0.1029	0.1060	0.0974	0.0945	0.0988
5	0.0792	0.0803	0.0803	0.0803	0.0801	0.0808	0.0798
6	0.0669	0.0649	0.0634	0.0662	0.0680	0.0664	0.0687
7	0.0580	0.0584	0.0556	0.0608	0.0569	0.0581	0.0564
8	0.0512	0.0510	0.0478	0.0537	0.0517	0.0533	0.0509
9	0.0458	0.0462	0.0466	0.0459	0.0448	0.0436	0.0453
U		0.0117	0.0111	0.0153	0.0046	0.0062	0.0053
Observations		42,422	19,200	23,222	42,792	14,015	28,777

This table reports the ratios of coefficients and standard errors starting with each digit $1, ..., 9$. The formula for the theoretical ratios is $\log_{10}(1 + 1/d)$ for the leading digit d. U is the square root of the sum of squares of differences between the actual and the theoretical ratio. Simulated data correspond to regressions using random variables drawn from the WDI dataset.

Table A5: Summary of non-parametric and parametric estimations using the Student(1) input for various sub-samples.

Sample	Maximum cumulated residual from non-parametric estimation	Maximum cumulated residual from parametric estimation
Macroeconomics	0.021	0.026
Microeconomics	0.018	0.027
Positive results	0.019	0.026
Null results[1]	0.015	0.009
Eye-catchers	0.025	0.033
No eye-catchers	0.008	0.014
Main results	0.015	0.022
Non-main results	0.030	0.038
With model	0.005	0.005
Without model	0.028	0.035
Low average PhD-age	0.036	0.044
High average PhD-age	0.006	0.009
No editor	0.021	0.029
At least one editor	0.017	0.024
No tenured author	0.029	0.038
At least one tenured author	0.008	0.014
Single authored	0.029	0.038
Co-authored	0.015	0.023
With research assistants	0.024	0.031
Without research assistants	0.009	0.018
Low number of thanks	0.012	0.018
High number of thanks	0.024	0.032
Data and codes available	0.019	0.027
Data or codes not available	0.018	0.026
Lab. experiments or RCT data[1]	0.040	0.042
Other data	0.015	0.023

Sources: AER, JPE, QJE (2005-2011) and authors' calculation. *Low average PhD-age* corresponds to articles written by authors whose average age since PhD is below the median of the articles' population. *Low number of thanks* corresponds to articles where the number of individuals thanked in the title's footnote is below the median of the articles' population. See notes of table 1 for the definitions of other categories.
[1]: These estimates are not reliable. In the case of articles reporting null results as their main contribution, the number of observations is way too low to apply our accounting method. In the case of laboratory experiments or randomized control trials, large z-statistics are less likely to appear which violates our methodological hypothesis that selection is increasing.

Table A6: Summary of non-parametric and parametric estimations using the Cauchy(0.5) input for various sub-samples.

Sample	Maximum cumulated residual from non-parametric estimation	Maximum cumulated residual from parametric estimation
Macroeconomics	0.016	0.013
Microeconomics	0.012	0.019
Positive results	0.013	0.017
Null results[1]	0.013	0.009
Eye-catchers	0.018	0.025
No eye-catchers	0.005	0.002
Main results	0.010	0.012
Non-main results	0.022	0.030
With model	0.005	0.001
Without model	0.021	0.028
Low average PhD-age	0.028	0.036
High average PhD-age	0.005	0.001
No editor	0.013	0.021
At least one editor	0.012	0.014
No tenured author	0.022	0.030
At least one tenured author	0.005	0.004
Single authored	0.022	0.030
Co-authored	0.010	0.013
With research assistants	0.017	0.023
Without research assistants	0.006	0.008
Low number of thanks	0.007	0.008
High number of thanks	0.016	0.023
Data and codes available	0.012	0.017
Data or codes not available	0.012	0.017
Lab. experiments or RCT data[1]	0.031	0.039
Other data	0.010	0.013

Sources: AER, JPE, QJE (2005-2011) and authors' calculation. *Low average PhD-age* corresponds to articles written by authors whose average age since PhD is below the median of the articles' population. *Low number of thanks* corresponds to articles where the number of individuals thanked in the title's footnote is below the median of the articles' population. See notes of table 1 for the definitions of other categories.

[1]: These estimates are not reliable. In the case of articles reporting null results as their main contribution, the number of observations is way too low to apply our accounting method. In the case of laboratory experiments or randomized control trials, large z-statistics are less likely to appear which violates our methodological hypothesis that selection is increasing.

Table A7: Summary of non-parametric and parametric estimations using the empirical VHLSS input for various sub-samples.

Sample	Maximum cumulated residual from non-parametric estimation	Maximum cumulated residual from parametric estimation
Macroeconomics	0.020	0.023
Microeconomics	0.025	0.022
Positive results	0.022	0.023
Null results[1]	0.091	0.004
Eye-catchers	0.029	0.028
No eye-catchers	0.013	0.013
Main results	0.017	0.018
Non-main results	0.035	0.031
With model	0.005	0.003
Without model	0.032	0.030
Low average PhD-age	0.040	0.038
High average PhD-age	0.010	0.006
No editor	0.027	0.025
At least one editor	0.020	0.020
No tenured author	0.032	0.031
At least one tenured author	0.014	0.013
Single authored	0.034	0.032
Co-authored	0.018	0.019
With research assistants	0.026	0.026
Without research assistants	0.015	0.015
Low number of thanks	0.015	0.015
High number of thanks	0.028	0.027
Data and codes available	0.025	0.024
Data or codes not available	0.019	0.021
Lab. experiments or RCT data[1]	0.049	0.033
Other data	0.018	0.020

Sources: AER, JPE, QJE (2005-2011) and authors' calculation. *Low average PhD-age* corresponds to articles written by authors whose average age since PhD is below the median of the articles' population. *Low number of thanks* corresponds to articles where the number of individuals thanked in the title's footnote is below the median of the articles' population. See notes of table 1 for the definitions of other categories.

[1]: These estimates are not reliable. In the case of articles reporting null results as their main contribution, the number of observations is way too low to apply our accounting method. In the case of laboratory experiments or randomized control trials, large z-statistics are less likely to appear which violates our methodological hypothesis that selection is increasing.

Table A8: Summary of non-parametric and parametric estimations using the empirical QOG input for various sub-samples.

Sample	Maximum cumulated residual from non-parametric estimation	Maximum cumulated residual from parametric estimation
Macroeconomics	0.018	0.022
Microeconomics	0.017	0.021
Positive results	0.016	0.021
Null results[1]	0.051	0.006
Eye-catchers	0.023	0.027
No eye-catchers	0.008	0.009
Main results	0.012	0.017
Non-main results	0.029	0.031
With model	0.004	0.001
Without model	0.026	0.029
Low average PhD-age	0.035	0.038
High average PhD-age	0.006	0.005
No editor	0.019	0.023
At least one editor	0.015	0.019
No tenured author	0.026	0.031
At least one tenured author	0.007	0.010
Single authored	0.028	0.031
Co-authored	0.012	0.017
With research assistants	0.021	0.025
Without research assistants	0.009	0.013
Low number of thanks	0.009	0.013
High number of thanks	0.021	0.026
Data and codes available	0.017	0.021
Data or codes not available	0.015	0.020
Lab. experiments or RCT data[1]	0.040	0.034
Other data	0.012	0.018

Sources: AER, JPE, QJE (2005-2011) and authors' calculation. *Low average PhD-age* corresponds to articles written by authors whose average age since PhD is below the median of the articles' population. *Low number of thanks* corresponds to articles where the number of individuals thanked in the title's footnote is below the median of the articles' population. See notes of table 1 for the definitions of other categories.

[1]: These estimates are not reliable. In the case of articles reporting null results as their main contribution, the number of observations is way too low to apply our accounting method. In the case of laboratory experiments or randomized control trials, large z-statistics are less likely to appear which violates our methodological hypothesis that selection is increasing.

Table A9: Summary of non-parametric and parametric estimations using the empirical PSID input for various sub-samples.

Sample	Maximum cumulated residual from non-parametric estimation	Maximum cumulated residual from parametric estimation
Macroeconomics	0.018	0.021
Microeconomics	0.016	0.026
Positive results	0.016	0.024
Null results[1]	0.016	0.017
Eye-catchers	0.019	0.030
No eye-catchers	0.006	0.008
Main results	0.010	0.017
Non-main results	0.025	0.035
With model	0.004	0.000
Without model	0.022	0.032
Low average PhD-age	0.031	0.041
High average PhD-age	0.005	0.004
No editor	0.016	0.026
At least one editor	0.013	0.019
No tenured author	0.023	0.034
At least one tenured author	0.006	0.009
Single authored	0.024	0.035
Co-authored	0.010	0.018
With research assistants	0.018	0.026
Without research assistants	0.007	0.014
Low number of thanks	0.008	0.013
High number of thanks	0.018	0.028
Data and codes available	0.015	0.023
Data or codes not available	0.012	0.021
Lab. experiments or RCT data[1]	0.036	0.043
Other data	0.011	0.018

Sources: AER, JPE, QJE (2005-2011) and authors' calculation. *Low average PhD-age* corresponds to articles written by authors whose average age since PhD is below the median of the articles' population. *Low number of thanks* corresponds to articles where the number of individuals thanked in the title's footnote is below the median of the articles' population. See notes of table 1 for the definitions of other categories.

[1]: These estimates are not reliable. In the case of articles reporting null results as their main contribution, the number of observations is way too low to apply our accounting method. In the case of laboratory experiments or randomized control trials, large z-statistics are less likely to appear which violates our methodological hypothesis that selection is increasing.

References

Benford, F.: 1938, The law of anomalous numbers, *Proceedings of the American Philosophical Society* **78**(4), 551–572.

Hill, T. P.: 1995, A statistical derivation of the significant-digit law, *Statistical Science* **10**(4), 354–363.

Hill, T. P.: 1998, The first digit phenomenon, *American Scientist* **86**(4), 358–363.