

Referees

Ivo Welch, UCLA

Mar 2018

Referee Objectives?

What is the journal guidance to you as referee?

- ▶ Should you use editorial taste?
(Whose? Your's? Journal's?)
- ▶ What should be your rejection threshold?
- ▶ How good/bad were you as a referee?
- ▶ Rewards? Repercussions?

Should you/journal care about

- ▶ Fairness to the Author?
- ▶ Maximum Cite Impact of the Journal?

reject simultaneous submissions by authors from less prominent institutions?

reject corrections/critiques, because they harm journal's earlier paper?

see JPE referee report below

Uncontroversial?

- ▶ Maybe author identity could matter,
- ▶ ... but few (not none!) would think referee identity **should** matter.
- ▶ ... yet many authors believe referee identity **does** matter.

Evidence?

Uncontroversial?

- ▶ Maybe author identity could matter,
- ▶ ... but few (not none!) would think referee identity **should** matter.
- ▶ ... yet many authors believe referee identity **does** matter.

Evidence?

Uncontroversial?

- ▶ Maybe author identity could matter,
- ▶ ... but few (not none!) would think referee identity **should** matter.
- ▶ ... yet many authors believe referee identity **does** matter.

Evidence?

Can and Cannot Do

Cannot measure scientific merit

Can measure “Reliability” (consensus)

Reliability

- ▶ Referees Choose Best Papers \Rightarrow Reliability
- ▶ Reliability $\not\Rightarrow$ Referees Choose Best Papers

Minimal Requirement

Can conclude bad process, but not good process (sci-merit)

All referees could like papers based on fonts, language, etc. not scientific merit. (If some referees like fonts, other language, then still no reliability.)

My paper is only about consensus (= reliability), i.e., paper outcome if referees were different.

Selection Problem

Journals do not randomly assign papers to referees,

...but in the SFS Cavalcade, I could and did!

Selection Problem

Journals do not randomly assign papers to referees,
...but in the SFS Cavalcade, I could and did!

Evidence Here

Same Paper, Multiple Referees at

- ▶ Econ&Fin Journals: ECMTA (75%), IER (85%), JEEA (74%), JET (50%), QJE (88%), RAND(81%), JF (20%), RFS (31%)
- ▶ SFS Cavalcade (100%; 8-10 refs each)

Study Inference: Nearly Identical.

Descriptive:

- ▶ \approx 56% REJ,
- ▶ \approx 18% Weak
- ▶ \approx 17% R&R
- ▶ \approx 10% ACC

4 Key Findings

1. Low (but Some) Agreement Among Referees
2. Editors follow referees
 - 2.1 Occasional editorial defensive claims that (studying) ratings is irrelevant because editors “read reports” are anecdotal, unscientific, and wishful thinking.
 - 2.2 \Rightarrow papers with more referees are less likely to survive.
3. Thresholds are different across referees,
4. but order preference switches are, too.

1. Low Agreement

Conditional Probabilities

	MR	SR	NR	SA	MA
Must Reject MR	0.177	0.41	0.26	0.14	0.018
Should Reject SR	0.092	0.37	0.32	0.19	0.021
Neutral NR	0.059	0.32	0.33	0.26	0.038
Should Accept SA	0.041	0.25	0.34	0.31	0.061
Must Accept MA	0.034	0.17	0.32	0.39	0.077
Unconditional	0.073	0.320	0.323	0.246	0.038

e.g., prob of one-ref MA: 3.8%. prob of two MA: < 0.3%.

e.g., cprob that a one-ref MA gets the same MA by the second ref: 7.7%

e.g., cprob that a one-ref MA gets a MR or SR by the second ref: 20%

→ not completely unreliable, but also not very reliable.

2. Editors Follow

- ▶ Editors: read letters, not ratings?!
- ▶ Possibly, but ratings and letters are highly congruent.
- ▶ ... and so are editors

	Editor R&R	Editor REJ	
Referee REJ	4%	96%	Wow?!
Referee WEAK	50%	50%	
Referee R&R	87%	17%	Wow?!
Referee ACC	90%	10%	Wow?!
Referee REJ,WEAK	25%	75%	≪WEAK
Referee REJ,R&R	39%	61%	≪R&R
Referee REJ,ACC	58%	42%	≪Weak

- ▶ Letters to editors may help explain the rest, but the study of referee ratings can explain most of the editorial decisions.

- ▶ ⇒ Unconditionally, heterogeneous referee nums: more refs → more rejections.

3. Referee Thresholds

Roughly: Run regression predicting each recommendation with

- ▶ other referees on same paper (ideally 1): ORSP
- ▶ same referee on other papers (ideally 0): SROP

(and normalize coefficients properly!)

	ECMTA	JEEA	JET	QJE	IER	RAND
ORSP	0.20	0.19	0.27	0.15	0.24	0.25
SROP	0.19	0.15	0.12	0.32	0.18	0.14

	RFS	JF	SFS
ORSP	0.14	0.10	0.40
OPSR	0.29	0.25	0.18

⇒ Draw intrinsically generous referees!!

3. Referee Thresholds

Roughly: Run regression predicting each recommendation with

- ▶ other referees on same paper (ideally 1): ORSP
- ▶ same referee on other papers (ideally 0): SROP

(and normalize coefficients properly!)

	ECMTA	JEEA	JET	QJE	IER	RAND
ORSP	0.20	0.19	0.27	0.15	0.24	0.25
SROP	0.19	0.15	0.12	0.32	0.18	0.14

	RFS	JF	SFS
ORSP	0.14	0.10	0.40
OPSR	0.29	0.25	0.18

⇒ Draw intrinsically generous referees!!

3. Referee Thresholds

Roughly: Run regression predicting each recommendation with

- ▶ other referees on same paper (ideally 1): ORSP
- ▶ same referee on other papers (ideally 0): SROP

(and normalize coefficients properly!)

	ECMTA	JEEA	JET	QJE	IER	RAND
ORSP	0.20	0.19	0.27	0.15	0.24	0.25
SROP	0.19	0.15	0.12	0.32	0.18	0.14

	RFS	JF	SFS
ORSP	0.14	0.10	0.40
OPSR	0.29	0.25	0.18

⇒ Draw intrinsically generous referees!!

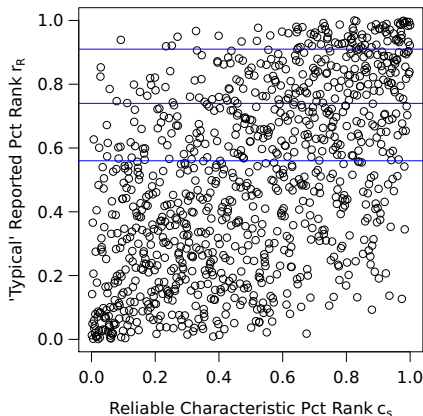
4. Ordering/Preference Reversals

Total pairs (duplicated)	4,416
Same distance between papers	1,121
Identical recommendations	456
Same difference, but offset	756
incl 0 diffs, same orderings	2,093
excl 0 diffs, same orderings	972
Pairwise Opposite orderings	702

(Some agreement, even randomly, of course. 972 same ordering, 702 opposite ordering. (omit no difference).)

Reliability To Decision?

- ▶ simple structural model in the paper.
- ▶ Prob acc/rej objectively good ($n \rightarrow \infty$) papers



ps: finance is much worse

Worse Possibilities

Could not be random noise, but likely worse!

Important **Conflicts of Interest**:

- ▶ Turf
- ▶ Vanity
- ▶ Suppression of contrary evidence
- ▶ Professional networks
- ▶ Ideology
- ▶ Personal likes and dislikes
- ▶ Professional likes and dislikes

+ Noise + Mood

Replicability — or Replication???

Critical Finance Review

- ▶ In most cases, if you write a critique of paper A, one of the **anonymous** referees consulted will often be an author of A!
- ▶ Honestly, what do you think this referee will write?
- ▶ Anonymous referees' conflicts of interest are **not** disclosed.

What Is At Stake?

Do we care about reliability, i.e., referee irrelevance ?

Do we care about author id irrelevance ?

Do we care about progress in objective and scientific economics?

Do we care about junior faculty?

Is the referee process (still) the best venue to discuss and decide merit?

Is it good? Or, like the B-Week rankings, great when successful and lousy when not?

So What Are We Doing About It?

Where are our system checks?

How do conflicts of interest manifest themselves?

Why do we not continuously study and improve our own critical processes?

I don't have all the answers, but why don't we try to find them??
There are other models, from ArXiv to Science (with its cadre of dedicated referees).

Delicious

JEP: First, finance conf refs \neq econ refs. Then, not our cup of tea and perhaps too critical of other processes.

AER: not interested in navel-gazing.

JPE: see next page

Referees also have biases. One of mine is that the **scientific investigations** I find compelling enough to merit publication in top journals **are theory-driven, not general searches for empirical findings** that might be of interest. I'm guessing that you must have had some sort of implicit theory that motivated your data collection, but it's your responsibility to present it, not my job to guess about it. Ultimately, this means **I found your paper unsatisfying as a scientific contribution.**

Specific:

p. 1, line 8: What you find ideal, at least in your prose, reflects your bias. I suspect, admittedly with no data from editors to confirm/deny, that editors as well as authors might also find it 'ideal' for the subject matter of a manuscript to be consistent with what the journal normally publishes. That is, there is a matching aspect that is non-trivial. **But no, to you objective quality of the paper would be the ONLY aspect that determines the editorial decision.**

p. 1, line 16: "more sinisterly... ." Give me a break. I have been told bluntly by more than one journal editor that, without apology, **they use their professional networks to identify good papers to publish. This strikes me as efficiency-enhancing, not sinister.** You really do have some pre-conceived positions that shadow your work, or at least your prose.

p. 4, last paragraph. There may not be much existing literature, but there is a lot more than you give credit to. I strongly encourage you to go find it; it isn't hard.

p. 25. Your conclusion and recommendations are infused with a strong assumption that the editorial/review process needs improvement. So under recommendation 3, we see that "Accurate and fair decisions become almost impossible" while at the bottom of p. 24 we read that "...it is not clear whether any other editorial processes are better." **Woe is us. How have we managed to survive for so long? The refereeing world stinks and there is nothing we can do about it... ..except embrace your fixes.**

Before I anoint anyone as Mr. Fixit with respect to how we do what we do, I'd like to see a little evidence in support of a proposition that the current editorial/review system is inefficient, in consideration of the fact that the production of refereeing services, as well as all other aspects of the production of science, is costly. Absent that, I am skeptical that the sky is falling.