Information Aggregation and Product Reviews in the Entertainment Industry: Executive Summary

Scott N. Swisher IV[*]

July 12, 2013

Purpose

Metacritic (`metacritic.com`), an important source of market information on entertainment product quality, does not disclose their weighted average methodology to the public. Specifically, the critic weights they use to calculate aggregate score (Metascore) from individual reviews are unknown. I estimate critic weights using a dataset of scores. I propose a simple economic theory to explain Metacritic's choice of weights. According to the theory, critic weights should be assigned proportional to review quality.

Method

Assuming that Metacritic uses a weighted average system with weights assigned to each publication, I estimate critic weights using a dataset of aggregate and individual scores. A data transformation method is used for estimation. I provide estimates of critic weight for a series of 2-year periods from January 2000 to April 2013. The most recent period is 2012-2013. The estimation method is applied to three product categories: video games, music, and film.

---

[*]Ph.D. Candidate, Department of Economics, University of Wisconsin-Madison.

Main Findings

Video games

Table 1 provides estimates for game critics over the 2012-2013 period.[1]

Table 1: Results (video games), 2012-2013, Top 20 Critics

Rank	Critic	Weight
1	Modojo	100
2	Edge Magazine	95
3	PC PowerPlay	92
4	Giant Bomb	92
5	Eurogamer	91
6	Metro GameCentral	90
7	Game Informer	89
8	Polygon	88
9	Telegraph	83
10	Nintendo Power	79
11	Guardian	78
12	games(TM)	78
13	TotalPlayStation	75
14	MacLife	73
15	AppGamer	73
16	148Apps	73
17	IGN	73
18	Apple'N'Apps	71
19	Post Arcade (National Post)	71
20	Gamezebo	70

I now provide a list of main findings for video game reviewers.

1. Metacritic does weight game reviews more if they are of high quality, but they do not weight higher-quality reviews enough according to the theory

2. Metacritic's weighting system is not responsive enough to variation in review quality across game critics: they underweight the best critics

3. Reviewers with more web traffic (according to Alexa rank) receive a higher weight from Metacritic, but this relationship is weak

4. Based on age of the publication or founding date, younger critics receive weakly higher weight

---

[1]Modojo's weight is set equal to 100 for comparison purposes. This table omits game critics when the estimation method cannot tell their weight apart from zero. Additional results are available in the summary or full versions of the paper on my website (`mywebspace.wisc.edu/sswisher/web`).

5. I find a very weak negative relationship between review length and critic weight

6. No compelling relationship exists between the number of reviews by a particular critic and their weight

7. Metacritic gives unbiased reviewers more weight

8. Critics with lower variance of review error (their score minus Metascore) receive higher weight

Focus attention on popular game review sites IGN (`ign.com`) and Gamespot (`gamespot.com`). The following table provides the relevant estimates.[2]

Critic	2000-2001	2002-2003	2004-2005	2006-2007	2008-2009	2010-2011	2012-2013
IGN	94	94	49	52	34	42	38
Gamespot	93	95	50	48	20	27	31

The two sites received very similar weight from Metacritic in the 2000-2007 period. However, starting in 2008 this relationship changed and IGN was assigned substantially more weight. Gamespot slowly recovered from this shock that manifested itself in the 2008-2009 period, gradually catching up to IGN in terms of weight, but still has not yet fully recovered. The shock that caused Gamespot's weight to suddenly decline is likely the 2007 *Kane & Lynch: Dead Men* incident (referenced extensively in the press) where Gamespot Editorial Director Jeff Gerstmann was fired for publishing an unfavorable review of the game due to external pressure from publisher Eidos Interactive. Metacritic viewed this event as signaling Gamespot's lack of independence and reduced their weight in the Metascore formula accordingly.

This paper's model always improves upon the unweighted average in terms of matching observed variation in Metascore. While the unweighted average usually misses Metascore by a point on average, the weighted average model reduces that to 0.4 points. Again, to interpret, the estimation procedure applied by this paper misses the actual Metascore by about 0.4 points (the average Metascore for video games is 68.83) on average. This method explains 99.8% of the observed variation in Metascore.

---

[2]Note that the critic normalized to 100 is changing each year.

Music

Table 2 provides estimates for music critics over the 2012-2013 period.[3]

Table 2: Results (music), 2012-2013, Top 20 Critics

Rank	Critic	Weight
1	Punknews.org (Staff)	100
2	The Source	58
3	Revolver	52
4	Pretty Much Amazing	27
5	Country Weekly	24
6	Okayplayer	22
7	Rock Sound	22
8	Expert Witness (MSN Music)	22
9	The A.V. Club	22
10	Billboard.com	21
11	Los Angeles Times	19
12	Kerrang!	18
13	Metal Hammer (UK)	16
14	Alternative Press	16
15	Chicago Tribune	15
16	Mojo	15
17	Uncut	15
18	Austin Chronicle	14
19	Magnet	14
20	The New York Times	14

I now provide a list of main findings for music reviewers.

1. Metacritic does weight music reviews more if they are of high quality, but they do not weight higher-quality reviews enough according to the theory

2. Metacritic's weighting system is not responsive enough to variation in review quality across music critics: they underweight the best critics

3. A negative relationship exists between the number of reviews by a particular critic and their weight

4. Metacritic gives unbiased music reviewers more weight

5. Music critics with lower variance of review error (their score minus Metascore) receive higher weight

---

[3]The weight of Punknews.org is set equal to 100 for comparison purposes. This table omits music critics when the estimation method cannot tell their weight apart from zero. Additional results are available in the summary or full versions of the paper on my website (`mywebspace.wisc.edu/sswisher/web`).

The weighted average model estimated for music reviews always outperforms the baseline unweighted average model in terms of predicting Metascore, but the weighted average model still misses by about 2 points on average (the unweighted average model misses by 2.5 points). This is likely due to the "normalization" procedure that Metacritic states that they use for music and film reviews on their site, which is still unclear. Whatever it is, this normalization goes beyond the standard weighted average procedure. It is difficult to provide an economic rationale for why such a normalization (arbitrarily increasing the dispersion of critic reviews before applying the weighted average formula) is appropriate. Still, the weighted average model employed here explains about 90% of the observed variation in Metascore, up from 83% for the unweighted average model (for the 2012-2013 period).

Film

Table 3 provides estimates for film critics over the 2012-2013 period.[4]

Table 3: Results (film), 2012-2013, Top 20 Critics

Rank	Critic	Weight
1	Rolling Stone	100
2	Wall Street Journal	76
3	Salon.com	67
4	Chicago Reader	67
5	Charlotte Observer	64
6	The New Yorker	63
7	The Guardian	59
8	Los Angeles Times	58
9	The A.V. Club	56
10	New York Daily News	56
11	The New York Times	49
12	Chicago Tribune	47
13	New York Magazine (Vulture)	45
14	Washington Post	45
15	Austin Chronicle	42
16	New York Post	42
17	Christian Science Monitor	40
18	Miami Herald	38
19	indieWIRE	37
20	Village Voice	36

---

[4]The weight of Rolling Stone is set equal to 100 for comparison purposes. This table omits film critics when the estimation method cannot tell their weight apart from zero. Additional results are available in the summary or full versions of the paper on my website (`mywebspace.wisc.edu/sswisher/web`).

I now provide a list of main findings for film reviewers.

1. Metacritic does weight film reviews more if they are of high quality, but they do not weight higher-quality reviews enough according to the theory

2. Metacritic's weighting system does not respond at all to variation in review quality across film critics: they underweight the best critics

3. A negative relationship exists between the number of reviews by a particular critic and their weight

4. Metacritic gives reviewers with negative bias (score below Metascore) more weight

5. No relationship is found between variance of review error (score minus Metascore) and weight

Similar to music reviews, the weighted average model estimated for film reviews always outperforms the baseline unweighted average model in terms of predicting Metascore, but the weighted average model still misses by about 2.5 points on average (the unweighted average model misses by 4 points). Nevertheless the weighted average model employed here explains about 96% of the observed variation in Metascore, up from 87% for the unweighted average model (for the 2012-2013 period).

Conclusion

I find that the most prolific publication in a genre is not necessarily assigned the most weight. Metacritic's assigned weight depends very weakly on critic age, number of reviews, and review length. The weighted average scheme in use at Metacritic weakly gives lower weight to less accurate reviewers in terms of the bias and noise embedded in their reviews. However, according to the economic theory they do not sufficiently weight critics with reviews of high quality. Metacritic can do better by making their weighting system more responsive to review quality.

From an economic perspective, Metacritic's weighting system is too egalitarian. Their weighted average scheme produces an aggregate score or Metascore too close to the standard unweighted average. Metacritic should give the best critics more weight.

Author: Scott N. Swisher IV

Affiliation: Ph.D. Candidate, Department of Economics, University of Wisconsin-Madison

Website: `mywebspace.wisc.edu/sswisher/web`