



International Journal of Environment, Workplace and Employment

ISSN online: 1741-8445 - ISSN print: 1741-8437

<https://www.inderscience.com/ijewe>

Investigating polarisation in critic and audience review scores via analysis of extremes, medians, averages, and correlations

Kyle Day, Jong-Min Kim

DOI: [10.1504/IJWE.2022.10051729](https://doi.org/10.1504/IJWE.2022.10051729)

Article History:

Received:	25 July 2022
Accepted:	23 September 2022
Published online:	19 July 2023

Investigating polarisation in critic and audience review scores via analysis of extremes, medians, averages, and correlations

Kyle Day and Jong-Min Kim*

Division of Science and Mathematics,
University of Minnesota-Morris,
Morris, MN 56267, USA

Email: day00090@umn.edu

Email: jongmink@morris.umn.edu

*Corresponding author

Abstract: This research's goal examines the relationship between four critic and audience review score categories (Rotten Tomatoes critics, Rotten Tomatoes fans, IMDb, metacritic) across a sample of 225 films released from 2002 to 2016. Minima and maxima analysis initially suggested intermixing between critic and audience scores. However, similar averages and medians suggested that critic scores were closer to each other than to audience scores. Correlational analysis confirmed that while each of the score categories were correlated to each other, the correlations were significantly stronger between critical score categories than with between critical and user score categories, suggesting polarisation. These correlations were found to persist over the course of five trienniums. In addition, correlations between site scores and box office grosses all supported the notion of polarisation, with audience scores having similar correlation coefficients to each other than to the critic scores.

Keywords: movie critic; audience review score; correlation analysis.

Reference to this paper should be made as follows: Day, K. and Kim, J-M. (2023) 'Investigating polarisation in critic and audience review scores via analysis of extremes, medians, averages, and correlations', *Int. J. Environment, Workplace and Employment*, Vol. 7, No. 1, pp.3–12.

Biographical notes: Kyle Day is an undergraduate student at the Computer Science Discipline, University of Minnesota at Morris, Morris, Minnesota, USA.

Jong-Min Kim is a Full Professor at the Statistics Discipline, University of Minnesota at Morris, Morris, Minnesota, USA. He received his BS from the Department of Mathematics Education, Cheongju University in 1994; MS from the Department of Mathematics, Chung-Ang University in 1996; and PhD from the Department of Statistics, Oklahoma State University, Oklahoma, USA, in 2002. His current research interests include copula, time series analysis, and survey sampling.

1 Introduction

The filmmakers' goal is to appease their audience. Through visual effects and storytelling to choreographed sequences and acting, their films are intended to entertain and resonate with the audience, making the time the audience spent watching seem worthwhile. But, ultimately, intended outcomes are different from actual outcomes. As hard as the filmmakers may try to appease their audience, the success or failure of the appeasement is determined by not filmmakers, but by audience. The audience may consist of both critics and general moviegoers, the latter of which is often referred to as simply the audience. Therefore, filmmakers are heavily relied on movie critics and reviews because movie critics and reviews are main factors in movie box office revenues. As some references to support how important movie critics and reviews are to movie box office revenues, Basuroy et al. (2003) studied how critics affected the box office performance of films and considered how the effects may be moderated by stars and budgets. Reinstein and Snyder (2005) found that positive reviews have a particularly large influence on the demand for dramas and narrowly released movies. Boatwright et al. (2007) investigated the impact of individual film critics on the market performance of movies, where specific key critics and reviewers may serve as market gatekeepers, and where various critics may have different types of impacts on product performance. Dhar and Weinberg (2016) found that critics' ratings had larger impact during economic downturns than during periods of economic expansion. Pang et al. (2022) found that results from a dataset of 408 US movies support the influence of consumer ratings on critic ratings; this influence increases when a movie receives more consumer reviews and decreases when critics publish reviews in media outlets for the entertainment industry. Even though it is important for the consumers to use review comments of ecommerce products when they purchase some best buy priced match products, there is a side effect of review comments such that there has been fake review comment issue in ecommerce businesses. In a film business, metacritic as a recommender system has produced metacritic score which aggregates reviews of entertainment products. But metacritic is controversial in its process for gathering, translating, and aggregating reviews (Perano et al., 2021). Therefore, we want to look at the relationship between four critic and audience review score categories (Rotten Tomatoes critics, Rotten Tomatoes fans, IMDb, metacritic). Through review sites such as Rotten Tomatoes, the International Movie Database (IMDb), and metacritic, they can convey their thoughts on the film in public. The 2016 film *Red* (Warner Bros.) garnered Rotten Tomatoes critics and fans (RT critics and RT fans) scores of 72% each. The film's IMDb score, assembled with reviews from critics and audience members, was on average 7.0/10 stars, not far off from its metacritic score (metacritic) of 60%, assembled from critical reviews. Meanwhile, the 2017 film *Star Wars: Episode VII – The Last Jedi* (Walt Disney) had a RT critics score of 91% and an RT fans score of 42%. All the while, the film had an IMDb score of 6.0/10 stars and a metacritic score of 84%. *Red* and *The Last Jedi* suggest two intuitive generalisations about the relationship between critics and the general audience: critics and audiences are either closely aligned or significantly unaligned in their scores on films. Which generalisation is more correct? With intuition alone, the answer is not easily determinable. To get closer to finding an answer, one must review a sampling of films and analyse their scores to determine where the similarities and differences lie. In 2016, a user named Anthony Allen uploaded a data test to Kaggle with a sample of films with various review scores, box office grosses, and genre information. Analyses of averages, medians, and

correlations were conducted to determine if Rotten Tomatoes critics and metacritic scores were closer aligned than with Rotten Tomatoes fan and IMDb scores, which would suggest polarisation. The remainder of this paper is organised as follows: Section 2 provides a description of the data, Section 3 summarises the findings of the data analysis, and Section 4 presents our conclusions.

2 Data description

The sample used for the research was collected from Anthony Allen's 'comparing numerical movie review scores' data set from the following site: <https://www.kaggle.com/datasets/antallen/comparing-numerical-movie-review-scores>.

The set consists of a sample of 225 films released from 2002 to 2016. They are equally distributed across each year at 15 films per year. Each film contains four review score categories: Rotten Tomatoes critics, Rotten Tomatoes fans, IMDb, and metacritic (metascores). Rotten Tomatoes critics' scores and metascores are averages of critic scores, Rotten Tomatoes fans scores are averages of audience scores, and IMDb scores are averages of both critic and audience scores. All review score categories with exception to IMDb are out of 100; IMDb is out of 10. Box office grosses are also included in full integer form. Furthermore, the films are all classified into 11 genres by binary integers of 0 and 1. Films may belong to multiple genres. Table 1 is an example of *Treasure Planet's* entry, including the first two genres. Film ratings are also included.

Table 1 Film *Treasure Planet*

<i>Titles</i>	<i>Rating</i>	<i>RT critic</i>	<i>RT fan</i>	<i>IMDb</i>	<i>Metacritic</i>	<i>Box office</i>	<i>Year</i>	<i>Action</i>	<i>Animation</i>
Treasure planet	PG	69	70	7.1	60	38,120,554	2002	1	1

The main statistics examined were minima, maxima, averages, medians, and correlations, all of which were calculated in Excel spreadsheets. Minima and maxima were gathered to determine how similar the highest and lowest scores were between different review scores. Averages were calculated from there, providing overviews of how the general film was rated in each category. Medians were also collected to serve as an indicator of how well films were rated. After that, correlations were taken between each review score category to determine how close each of the categories were to each other without implying causality. Correlations were then taken between each review score category and box office gross to provide an idea of how close each category was in terms of influence.

These statistics were first retrieved from all 225 films. Then, they were compared against each other based on the review score category. To better illustrate the change occurring over time, they were all examined over five trienniums, consisting of 45 films a piece, with line graphs. Scatter plots were created to illustrate the linear differences between the Rotten Tomatoes critics and Fans correlation and the Rotten Tomatoes critics and metacritic correlation.

3 Data analysis

3.1 Extremes

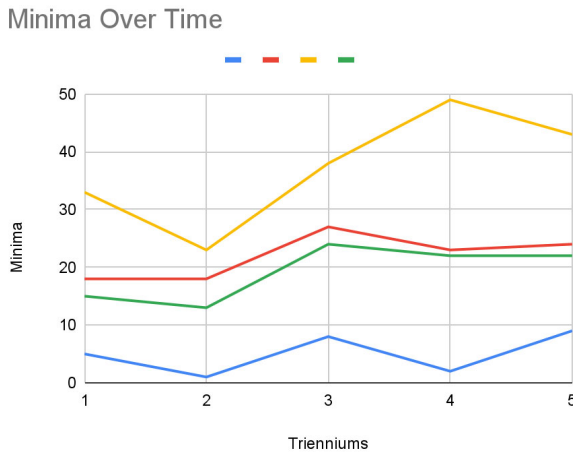
The overview minima and maxima for each review score category are shown in Table 2.

Table 2 Minima and maxima for each review score category

Categories	RT critics	RT fans	IMDb	Metacritic
Minima	1	18	2.3	13
Maxima	98	92	8.3	90

For overview minima, RT critics were a minimum outlier, being 12/100 away from the lowest metascore. The other three review score categories clustered in intervals of 5/100 away from each other when each was divided over the maximum possible scores for their respective categories. IMDb had the highest minimum rating of 2.3/10 stars, or 23/100. However, IMDb also had the lowest maximum rating: 8.3/10 stars, or 83/100. RT critics, on the other hand, had the highest score of 98/100. It was also no longer a significant outlier. Only RT fans and metacritic were noticeably clustered together, at 2/100 apart. The consistent clustering of these two categories seemed to suggest them to be the most related thus far.

Figure 1 Minima found for review score categories over five trienniums (see online version for colours)

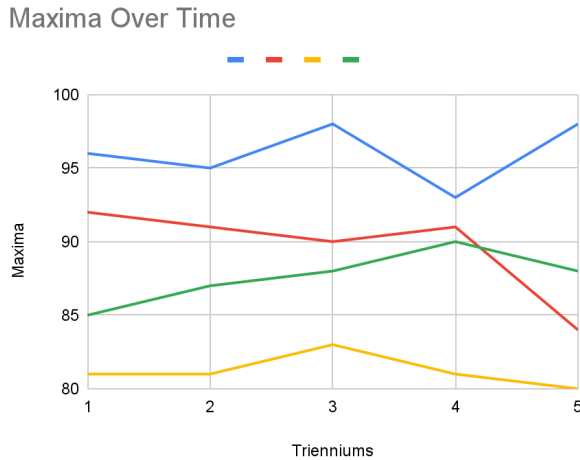


In Figure 1, IMDb’s star ratings (yellow) were upsized by a scale of 10 in order to show relevant similarities and differences. This review category in particular increased in minima over time, signifying an increase in positive reviews up until after triennium 4. It grew more distant from RT fans (red) and metacritic (green), which almost converged at triennium 4, before dispersing in opposite directions, though slightly. RT critics (blue) contained the lowest minima out of any category, increasing into triennium.

In Figure 2, IMDb had a relatively stable change in maxima over time compared to the other categories. Albeit, after peaking at triennium 3, the category began a steady decline in score minima. RT fans (red) and metacritic (green) converged a little past

triennium 4, before then diverging into triennium 5 like IMDb. Rotten Tomatoes would be the only review score category to increase into triennium 5.

Figure 2 Maxima found for review score categories over five trienniums (see online version for colours)



Based on minima and maxima alone, one may be convinced that RT fans and metacritic scores are the most related. They grew closer to each other over time, and they appeared much more clustered in terms of maxima. Due to the fact they are of audience and critic review score categories, one may go on further to claim that polarisation does not exist between the critics and the audience.

3.2 Medians

Overview review score category medians that were determined may be seen in Table 3.

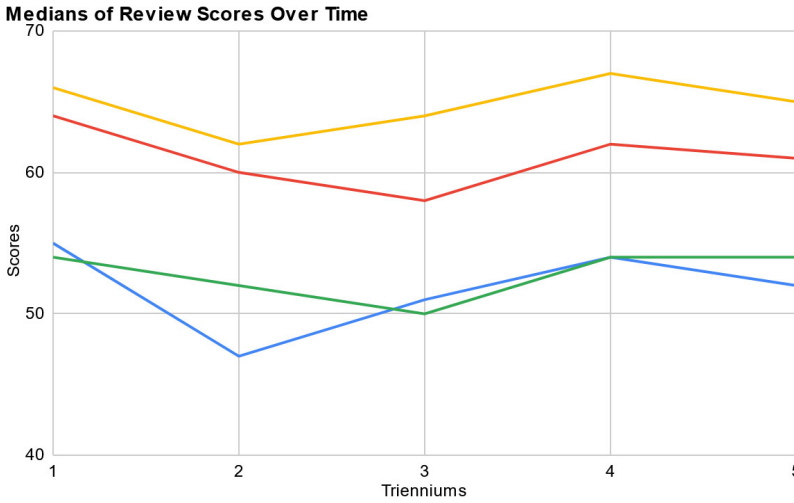
Table 3 Review score category medians

Categories	RT critics	RT fans	IMDb	Metacritic
Medians	52	56	6.5	53

The determined medians suggested a very different idea of how the review score categories compared to each other. While RT critics appeared to be rather isolated from metacritic based on minima and maxima, their median scores was only 1/100 off. Meanwhile, metacritic and RT fans, which were shown to have similar minima and maxima, had a slightly wider gap at 3/100. IMDb’s median was not clustered like the rest, but it was closest to RT fans.

When examined over the course of five trienniums in Figure 3, the relationship between RT critics and metacritic appeared to be much closer related. They grew closer together over time, until converging at triennium 4. The low differences between the scores suggest a more intimate intertwining than the RT fans and IMDb, which, while consistently close, never converged.

Figure 3 Medians found for review score categories over five triennium (see online version for colours)



In contrast to minima and maxima, the median scores suggest that RT critic and metacritic were much closer related, suggesting that the critic review score categories were similar. Meanwhile, RT fans and IMDb, though not exactly alike (average of audience vs. average of all), are more similar to each other than to the other categories. The median analysis suggests that critics polarised against the mixes of audiences. Albeit, like minima and maxima, medians were only single scores within the sample. The question of polarisation needed to be picked apart more closely, which values that represented all films, as an overview or for certain periods of time.

3.3 Averages

To get a better idea of how the films in the sample behaved in terms of whether their review score categories were polarised or not, average were taken into account, as seen in Table 4.

Table 4 Review score category averages

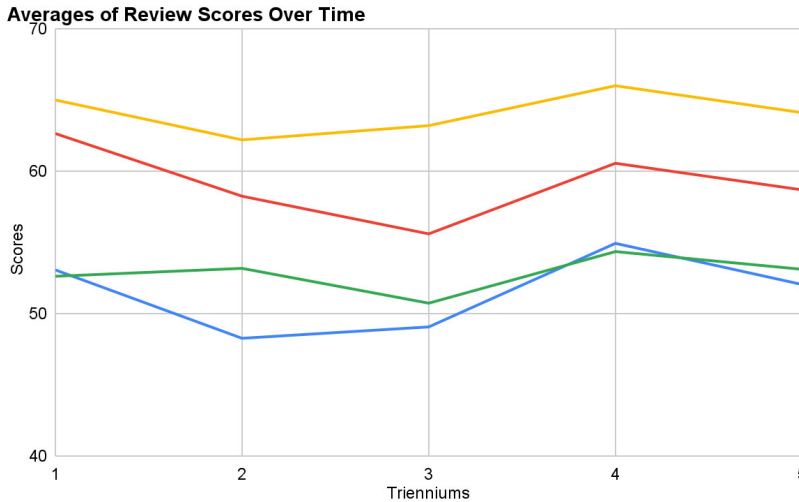
Categories	RT critics	RT fans	IMDb	Metacritic
Averages	51.5	59.9	6.41	52.8

Between the points of extremes and medians, the averages held a similar answer to the polarisation inquiry from an overview. RT critics and metacritic had a 1/100 difference in average scores. Although, RT fans and metacritic had a much greater difference in averages than with medians: around 7/100 vs. 3/100. On the other hand, RT Fans and IMDb were much closer in terms of average differences and median differences: 4/100 vs. 9/100. The relationship divides between critic and audience scores appeared to be growing. Over five trienniums, Figure 4 shows that the relationship of averages between the review score categories held, with some notable observations.

IMDb and RT fans started out close between the range of 62 and 65 in triennium 1, before growing farther apart into the range of 55 and 63 in triennium 3. Although, after

the transition between trienniums 2 and 3, they began to increase and decrease in the same direction, resting at the range of 58 and 64. Meanwhile, RT critics and metacritic started out with similar averages, at 53 and 52, before then growing further apart in triennium 2, at 48 and 53. From there on, they grew closer together up until triennium 4, at 54 each, where they diverged to 52 and 53. The relationship between the critic review score categories was much stronger than the one between the audience and mixed-audience score categories.

Figure 4 Averages found for review score categories over five triennium (see online version for colours)



4 Correlations

With how similar the averages are between critics than between audiences, the evidence for critics and audiences being polarised seemed more likely than not. However, correlation coefficients are an even stronger means of determining the degree to which the review scores are related. The correlation coefficients between each review score category are shown in Table 5.

Table 5 Correlation coefficients between each review score category

Categories	RT critic	RT fans	IMDb	Metacritic
Highest	Metacritic ~ 0.94	IMDb ~ 0.82	RT fans ~ 0.82	RT critics ~ 0.94
Lowest	RT fans ~ 0.72	Metacritic ~ 0.66	Metacritic ~ 0.77	RT critics ~ 0.66

All correlations were strong and positive. However, there were clear differences in classification between the highest and lowest correlations per review score category. RT critics and metacritic had a near-perfect correlation of 0.94. Meanwhile, these categories had diminished strength with RT fans, at coefficients of 0.72 and 0.66 respectively.

In Figure 5, the scatter plot for RT critics and RT fans scores exhibited a greater deal of noise than the one for RT critics’ scores and metascores. Less clusters were present.

Due to the noise, the direction of the correlation is much less distinct on the former scatterplot than the latter without a trend line. Furthermore, there are several more values which do not align as well in the former than in the latter.

Figure 5 Scatterplots for the scores of RT critics with RT fans and metacritic (see online version for colours)

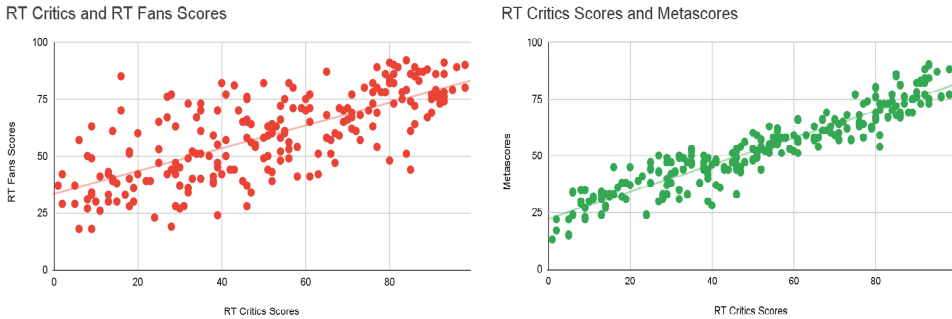


Figure 6 Correlations of review scores over time (see online version for colours)

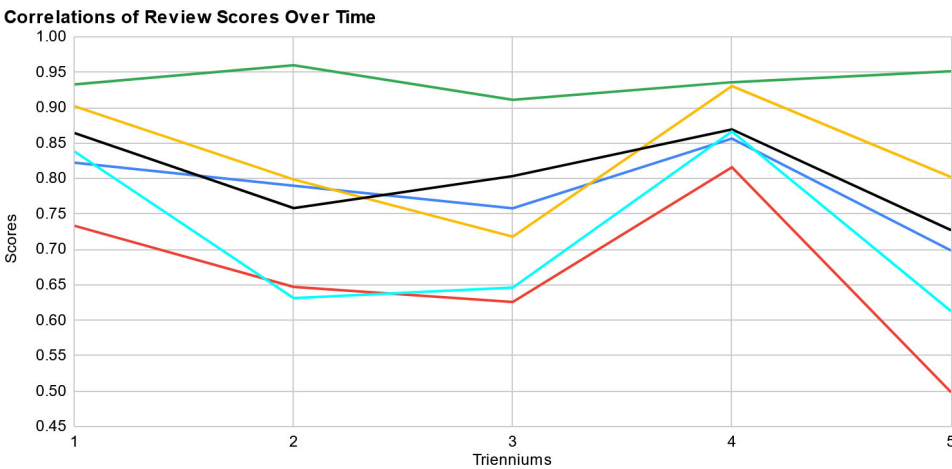


Figure 6 shows that all correlations other than RT critics and metacritic (green) ended in decline over time. While RT critics and metacritic peaked at triennium 2, all other correlations peaked at triennium 4. RT critics' and fans' (cyan) correlation and RT fans' and metacritic's correlation fell significantly from triennium 4 into triennium 5, from 0.87 to 0.61 and 0.82 to 0.50 respectively. The decline suggested a significant misalignment between critic and audience scores. Critics and mixed audiences did not fare much better from trienniums 4 to 5. RT critics' correlation with IMDb fell from 0.87 to 0.73 (black), while metacritic's correlation with IMDb decreased from 0.86 to 0.70 (blue). Furthermore, RT fans and IMDb (yellow) had a falling out in correlation from 0.94 in triennium 4 to 0.80 in triennium 5. The critical opinion which influenced IMDb's scores may be to blame for the sudden drop, considering how drastic the audience and critic divides were between RT fans and RT critics or metacritic.

Correlations, although strong in general, were in decline between review score categories where critics and audiences are involved in some capacity. Meanwhile, critic review score categories exhibited stable, strong, and positive change over time, never having a correlation below 0.90. With the critic review score categories remaining significantly more correlated with themselves than with the audience review score categories over time, polarisation had become apparently obvious. To further explore how similar and different review score categories behaved, correlation coefficients were calculated using them and the box office scores of each film and shown in Table 6.

Table 6 Box office correlation coefficients

<i>Category</i>	<i>RT critics</i>	<i>RT fans</i>	<i>IMDb</i>	<i>Metacritic</i>
Box office correlation	0.216	0.252	0.255	0.197

Although all correlations appear to be low with this sample, RT Fans and IMDb were found to be closer in range with higher correlations than RT critics and metacritic. For this sample, both mixtures of audience scores had a more positive correlation with the box office than the critic scores. With how similar each of the two category types is in terms of correlation, polarisation had set them apart.

5 Discussion

The filmmaker ultimately leaves the viewers to decide the receptive fate of his or her film. These viewers may consist of critics and a casual audience. While they share an experience in viewing films, they tend to form different opinions on it. The four different review score categories analysed (RT critics, RT fans, IMDb, and metacritic) appeared to be related to each other. In fact, analyses of minima and maxima seem to suggest that the review score categories had similarities between audience and critic review score categories. However, through analyses of medians, averages, and correlations, critic review score categories were found to be much more related to each other than to any mixture of audience review scores (only audience or audience mixed with critics). Furthermore, breaking down these analyses over time, audience scores grew further apart from critic scores over time. RT fans itself had fallen further away from any mixture of critic scores between 2014 and 2016 (the fifth triennium), although it retained the strongest relationship with IMDb, to the point where their box office correlations were only a few thousandths off. IMDb, the middle ground between critics and audiences, showed to be slightly better correlated with box office success. Meanwhile, RT critics and metacritic remained strongly intertwined, although not as close in box office correlations. These differences in the behaviour of the review score categories suggest that critic and audience scores are not generally closely aligned, but clustered and polarised. Hence, while creating their films, filmmakers should consider the differences between critics and the general audience much more closely when trying to create films which resonate with them. From there, the level of resonation between both may very well lead to box office success.

References

- Basuroy, S., Chatterjee, S. and Ravid, S.A. (2003) 'How critical are reviews? The box office effects of film critics, star power, and budgets', *Journal of Marketing*, Vol. 67, No. 4, pp.103–127.
- Boatwright, P., Basuroy, S. and Kamakura, W. (2007) 'Reviewing the reviewers: the impact of individual film critics on box office performance', *Quantitative Marketing and Economics*, Vol. 5, No. 4, pp.401–425.
- Dhar, T. and Weinberg, C.B. (2016) 'Measurement of interactions in non-linear marketing models: the effect of critics' ratings and consumer sentiment on movie demand', *International Journal of Research in Marketing*, Vol. 33, No. 2, pp.392–408.
- Pang, J., Liu, A.X. and Golder, P.N. (2022) 'Critics' conformity to consumers in movie evaluation', *Journal of the Academy of Marketing Science*, Vol. 50, pp.864–887 [online] <https://link.springer.com/article/10.1007/s11747-021-00816-9>.
- Perano, M., Casali, G.L. and Liu, Y. (2021) 'Professional reviews as service: a mix method approach to assess the value of recommender systems in the entertainment industry', *Technological Forecasting and Social Change*, August, Vol. 169, p.120800.
- Reinstein, D. and Snyder, C. (2005) 'The influence of expert reviews on consumer demand for experience goods: a case study of movie reviews', *Journal of Industrial Economics*, Vol. 53, No. 1, pp.27–51.