Do Professional Critics Diverge from Public Opinion? Evidence from Twitter

Yu-Hsi Liu*

Suffolk University

May 31, 2014

Very preliminary.

Please do not quote, cite, post, or circulate without author's permission.

Abstract

This paper tests whether expert judgment diverges from collective opinion in social networks. The conventional thinking suggests that in cultural offerings, the public usually have "bad taste," which differs from professional evaluation. Some literature has focused on the role of professional critics in mass markets, but little research examines how public opinion converges or diverges from professional criticism dynamically. Using Twitter data collected from May 2011 to Aug 2011, this paper measures how movie tweets relate to professional film critics' review on a weekly basis. After controlling for movie-specific effects, this paper finds that the ratio of positive tweets for a movie is significantly correlated with professional criticism. My results confirm that there are positive correlations between professional judgment, ordinary evaluation and public appeal. Moreover, if we control the unobserved movie quality by a fixed-effects model, the gap between ordinary evaluation and professional judgment decreases over time. It is could be a sign that ordinary consumers revise their expectation and converge to experts' opinion after the opening weekend, suggesting little influence effect of critics. This paper confirms that the online public appears to have "good taste" in movie consumption.

JEL Classifications: L10, D83, L82

*Suffolk Univ., 8 Ashburton Pl., Boston, MA 02109. Correspondence to: laches2@gmail.com

1. Introduction

In the consumption of cultural goods, whether expert judgment diverges from public opinion evaluation is a key question for economics, marketing, and even cultural policy. Using data from social network websites, this paper empirically addresses the following questions: Do ordinary consumers have the same taste with experts? Do mass audiences learn from professional critics? Is professional criticism a leading indicator of public opinion rather than an influencer? Will a positive public evaluation lead to a strong demand? As the internet becomes a fast channel of collective opinion, these issues need to be revisited using the data from social network websites.

The movie exhibition market provides a good example of the divergence/convergence between public evaluation and expert judgment. A movie is a typical experience good, whose product quality is ascertained by consumers only after a purchase. On such experience goods, consumers' decisions highly depend on consumers' information sets. Professional movie criticism is usually published before or on the opening weekend, and it comes to consumers' information sets before word-of-mouth communication starts. Producers and media view criticism as an important indicator, but there has been a controversy over the real influence of professional critics on consumers' decisions.

Two competing hypotheses, "good taste" and "no taste", explain the relationship between professional judgment and consumer decision. "No taste" means public opinion is not significantly correlated with expert judgment, while "good taste" means ordinary consumers display a taste consistent with professional opinion. Some empirical literature shows that good professional reviews correlate with high box-office revenue, while the conventional thinking suggests that film with high expert opinion does not guarantee its appeal to popularity.

A rich literature has investigated the relationship between expert opinion and popular appeal using data from movie industry, but few measures ordinary evaluation directly. Eliashberg and Shugan (1997) find that for a movie, movie criticism significantly correlates with late box-office revenue but not with early box-office revenue. Reinstein and Snyder (2005) find that a positive expert review does have an influence on opening weekend box office revenue in some cases, but the significance of the “influence effect” varies by movie type. Although they do not measure ordinary evaluation directly, their research suggests professional criticism could be positively correlated with public appeal.

Eliashberg and Shugan (1997), Reinstein and Snyder (2005) all suggest that in some cases, professional criticism could be only a leading factor of public opinion rather than an influencer. As Reinstein and Snyder (2005) pointed out, high demand could be a response to high movie quality; it may not be a result of good review. In this case, professional criticism has only a “prediction effect” but no “influence effect”. A positive correlation between ordinary evaluation and professional criticism consistent with “good taste” hypothesis does not necessarily imply causal relationship. If we can control for movie quality and measure ordinary evaluation directly, the spurious correlation between high market performance and good expert review will be cleared.

Since high ordinary evaluation does not always lead to good market performance, box-office revenue may not be a good proxy of ordinary evaluation. We need to find direct proxy of ordinary evaluation. In some cases in industry, consumers may stand in a long line to see a “bad” movie that receives unfavorable judgments from both ordinary people and critics. It shows that there are not always positive correlations between the three indicators: expert judgment, ordinary evaluation, and consumption decisions. As Holbrook (2005) concludes, the intervention of ordinary evaluation between professional judgment and popular appeal complicate the issue of how professional critics influence ordinary consumers.

One of the challenges in empirical study is that movie quality is often unobservable and difficult to measure. Previous literature often uses expert review as a proxy of quality, but this proxy does not apply since we are interested in the relationship between expert review and ordinary evaluation. Moreover, unlike the other goods, there are no objective criteria for the aesthetic quality of a culture good. Either ordinary evaluation or expert judgment is a subjective opinion rather than an objective criterion. A fixed effects model may control the unobservable quality for each movie, but it will omit the time-constant explanatory factors such as expert review at the same time.

To address the issue, other than a fixed-effect model, I use the dynamics of my Twitter data that allows me to measure the gap between expert review and ordinary evaluation every week. Previous study has employed IMDB ratings as a proxy of ordinary evaluation (Holbrook, 2005) or followed an experimental approach (West and Broniarczyk ,1998; Holbrook, 2003) to measure ordinary evaluation. However, the IMDB rating is time-invariant and hence it provides cross-sectional data across films rather than panel data showing both time and movie variance. My data from the social network website, Twitter, enables me to measure ordinary evaluation directly by counting the weekly ratio of positive tweets, which is called the “valence” of online information. The gap between expert review and Twitter valence directly shows how consumers’ taste deviate from experts’ opinion every week, while the time-constant quality can be eliminated by the differencing. Also, the weekly changes of valence show how consumers learn as the other information such as word-of-mouth updates their information set. Movie reviews usually come before the priors, and their influence is supposed to be weakened after consumers learn from word-oh-mouth communication. Compared to the previous studies, my data characteristics will enable me to measure the probably spurious relationship between ordinary evaluation and professional criticism.

In this paper, I estimate the panel data models of weekly gap between expert review and Twitter valence and investigate the impact of movie-specific factors on the gap, with a particular focus on under which conditions ordinary evaluation will diverge from expert opinion. In the following Section 2, I describe the dataset. In Section 3, I discuss the econometric methodology and present the results. I provide short conclusion and discussions in Section 4.

2. Data Description

This paper uses the same Twitter dataset with Liu (2012). The social networking website "Twitter" is a platform for users to post comments within 150 words, which are called “tweets”. Twitter does not follow a real-name system, so the networks among twitter users are supposed to be more open than facebook. Due to tremendous world-wide growth, as of 2011, Twitter has 100 million active users logging in at least once per month and generates over 200 million tweets per day.

The data in this paper comes from the daily counts of English-written tweets [1] on individual movies from May 16, 2011 to August 19, 2011. The variables of tweets for 57 movies are collected from TwitCritics[2]. These variables include the daily number of total tweets (by movie), the daily number of positive tweets (by movie), and the daily number of negative tweets (by movie). Data was collected at 6:00am and 12:00pm EST every day.

Previous studies have focused on online comments for various product review websites such as Yahoo!.com and Amazon.com. However, Twitter is different from these product review websites. Consumers search for movie comments on movie review websites with a clear intention of learning. In contrast, consumers encounter information about movies randomly and unintentionally with Twitter. Communication in Twitter is closer to the real-world word-of-mouth communication that is presented in the learning models.

Table 1 summarizes the tweet data. Fifty-seven movies in theater were included in the data set. These 57 movies are all widely-released commercial movies. On average, each movie had daily tweet data for 77.22 days. This data duration is longer than the average time a movie is in theaters (about 50 days), enough to catch the fade-out pattern of movie tweets. Table 2

[1] Although 60% of Twitter users reside outside the United States, it is still appropriate to project English tweets to the North American movie market since overseas users are mainly tweeting in Japanese, Portuguese and Spanish, not in English.

[2] TwitCritics (*www.twittercritics.com)* was a website that compiled positive and negative tweets about movies and created a list of the movies currently in theaters. It showed the number of tweets and the ratio of positive tweets to all tweets on a movie. Unfortunately, the website service stopped on October 2012. You can still find thousands of articles about TwitCritics on the internet, or find their “remains” at http://www.facebook.com/twitcritics.

presents the comparison of the movies in theater and movies in my dataset. The set of 57 movies is a subset of the 286 movies playing in theaters during the time period studied. Selection bias is suspected since the 57 movies have higher average box-office revenue, larger production budgets and less missing data. However, this is not necessarily a limitation of my data set. Out of 286 movies, about 50 movies are only at local theaters with limited release. Since the aggregate Twitter data reflect the movie market as a whole, mass-marketed movies reflect aggregate Twitter opinions better than limited release movies. In this sense, my data selection is appropriate because it rules out the local movies with lower box-office revenue.

The two key factors, ordinary evaluation and professional judgment, are measured from separated data source. Professional judgment on a movie is measured by Metascore. Metascore is a movie-variant, time-invariant variable collected from *Metacritic.com*, a website that summarizes influential reviews of movies from publications and public media such as *New York Times* and *Chicago Sun*. It ranges from 0 to 100, and it is a weighted average in that the website assigns different weights to different critics. In contrast to tweets, Metascore is much more oriented towards professional, authoritative critics, whose reviews are delivered in public before the priors. There is no time dynamic of Metascore on each movie.

Ordinary evaluation on is measured by tweet valence, which is defined as the ratio of the positive tweets to total tweet on a movie. Unlike Metascore, tweet valence is a movie-variant, time-variant variable which ranges from 0 to 1. By counting the ratio of positive tweet every week, I am able to catch the dynamic change of ordinary evaluation. Whether a tweet is positive or negative is judged by computer programs from TwitCritic. Due to the technological limitations, measurement error may exist in computerized content analysis. However, Morris

(1994) shows that human-coded and computerized content analysis may be equally effective. The precision of sentiment analysis is beyond the scope of this paper.

Table 3 shows the descriptive statistics for the valence of tweet, Metascore and the gap between Metascore and valence. The gap is defined as

Gap = | (valence*100-Metascore) |

I use the absolute value of the difference between Metascore and valence because I focus on whether ordinary evaluation diverges from expert review, not the negative or positive sign of the difference between Metascore and valence. The statistics of the difference between Metascore and valence without the absolute value are also reported in Table 3.

Table 4 presents the covariance matrix of the key variables. In the first glance of descriptive statistics, Metascore is positively correlated with tweet valence, which suggests that ordinary evaluation is positively correlated with professional judgment. If we divide Metascore by 100 to make it within the same range with valence, I observe an interesting fact in the dataset: in most cases, Metascore is lower than tweet valence on a movie. It implies that in general, ordinary people on Twitter tend to give higher review than experts do. Considering the fact that there are all commercial movies in my data set, it is consistent with Pierre Bourdieu’s (1979) perspective on taste. On Bourdieu’s perspective, one’s taste is related to his possession of cultural capital. Acknowledged experts display “good taste” because they have acquired more social capital through long-time training or higher intellectual status. On the other hand, experts have the incentive to “legitimate” their good taste by appreciating the cultural goods that require audience to have more social capital to enjoy them. It may lead experts to systematically underrate commercial films which do not require audience to have much social capital to get into.

The North American weekend box-office data was collected from Boxofficemojo.com from May 13, 2011 through August 21, 2011. The revenue data shows consistency with commercial movie revenue characteristics and trends. Weekend box-office revenue per film declines exponentially over time, as does total box-office revenue across film.

3. Estimation Methods and Results

3.1 Hypothesis

Many competing theories try to answer the following questions: does ordinary evaluation converge to expert opinion? If yes, does ordinary evaluation have an influence effect on ordinary consumers? To answer those questions on theatrically released movies, I propose and test the following hypotheses:

1. If ordinary consumers' opinion does not differ from the professional judgment, the professional judgment, Metascore, should have explanatory power on tweet valence. In this case, we will confirm that professional judgment has a prediction effect on ordinary evaluation.
2. Following Eliashberg and Shugan (1997), I assume Metascore to have more explanatory power on tweet valence during the early weeks than in the later weeks if professional judgment has an influence effect. It is because expert reviews come before or on the opening weekend and word-of-mouth starts only after the priors.

Consumers become better informed after the prior, hence the influence effect of professional judgment on ordinary information will be mitigated.

3. Under the condition that professional judgment is consistent with ordinary evaluation, the gap between Metascore and tweet valence will decrease over time if professional judgment has NO influence effect. In other words, when consumers have "good taste", ordinary evaluation differs from professional opinion in the beginning but converges to professional opinion because professional judgment has only prediction effect.

3.2 Public Appeal as a Dependant Variable

The baseline models using panel data estimate regression relating weekend revenue at week *t* to daily tweets at week *t-1* first, then estimate a regression equation relating tweet valence at week *t* to Metascore. The first regression is relating the popular appeal to professional judgment, while the second one is to relate professional judgment to ordinary evaluation. The second one would be the key result of this paper, but the first one can give us a glance about how consumers' and experts' evaluation relate to box-office revenue.

Following Liu (2012), The first model takes the North-American weekend box-office revenue as the dependent variable. To keep it simple, the weekend revenues of movie *j* at week *t* can be identified as

$$\ln(Rev_{jt}) = X_j\beta_1 + \beta_2 \ln VOLUME_{j,t-1} + \beta_3 VALENCE_{j,t-1} + \beta_4 (t^*Z_j) + \beta_5 \tau_t + \eta_{jt}, \quad (1)$$

for which *X* denotes factors that are fixed to the total revenue of movie *j*, including rating, genre and so on. Z denotes the movie-specific factors that influence the weekly declining rate of revenue. The impact of Z on weekend revenue is dynamic and reinforced by time, hence the coefficient of Z is multiplied by t. *VOLUME* is the weekly incremental number of total tweets for movie *j* from Sunday through Thursday prior to weekend *t*, while *VALENCE* is defined as the average ratio of positive tweets to total tweets from Sunday through Thursday before weekend *t*. Saturday tweets are not included because they are simultaneous with the weekend box-office revenue, not prior to revenue. Also, τ denotes the time-specific disturbance while η denotes the random disturbance for movie *j* at week *t*.

In regression (1), the issue of movie-specific effect needs to be addressed, but first-differencing the both-sides variables can wash out the movie-specific effect. Fixed-effects model would not be an ideal setting since it will omit all the time-invariant factors such as Metascore. Therefore, for t > 1, the first difference of the dependent variable can be defined as the natural log of movie *j*'s revenue change rate d_{jt} where $\ln(d_{jt}) = \ln \frac{\text{Rev jt}}{\text{Rev j,t}-1} = \ln(Rev_{jt}) - \ln(Rev_{j,t-1})$. Again, following Liu (2012), the model can be written as follows:

$$\ln d_{jt} = \ln(Rev_{jt}) - \ln(Rev_{j,t-1})$$

$$= \delta_0 + \delta_1 (\ln VOLUME_{j,t-1} - \ln VOLUME_{j,t-2}) + \delta_2 (VALENCE_{j,t-1} - VALENCE_{j,t-2}) + \delta_3 Z_j + \delta_4 T_{jt} + \varepsilon_{jt} \qquad (2)$$

where T_{jt} denotes all time-specific factors to d_{jt}; and ε_{jt} denotes disturbance for movie *j* at week *t*. T_{jt} includes "number of week" and "long weekend or not". Z_j denotes all the movie-specific factors that influence the rate d_{jt}.. Here we assume that although X_j is eliminated in equation (2).

However, some movie-specific variables, Z_j, still can influence the change rate of weekend box-office revenue. For example, we may assume that family comedy movies have lower revenue change rates because their online discussions fade out more slowly than other movies' (Liu, 2012.) Z_j may include Metascore, number of theaters, production budget, age factors, and genre. The equation (2) states that the dependent variable, the change rate of North-American weekend box-office revenue, is related to the weekly variation of Twitter word-of-mouth measures in the previous time period ($\Delta VALENCE_{jt-1}$ and $\Delta VOLUME_{jt-1}$).

Both in regression (1) and (2), the residual term ε_{jt} could be subject to a movie-specific source of correlation and a week-specific source of autocorrelation. I use Rogers clustered standard errors to control for the movie-specific effect on the residual term.

It is a question that whether the professional judgment, Metascore, should be categorized as X_j or Z_j. X_j are the movie-specific factors to total box-office revenue and it can be eliminated by first-differencing or a fixed-effects model; Z_j are the movie-specific factors which affect the declining pattern of box-office revenue and cannot be eliminated by the econometric settings. I try both models to see in which group Metascore has the significance in the regressions. The result on Table 5 shows that Metascore is significant in regression (1) but not in regression (2), supporting the assumption that Metascrore is more likely to belong to X_j rather Z_j. Considering the timing that professional criticism usually published before the prior, Metascore can predict the box-office revenue but may not affect the dynamic learning process after the opening weekends.

Table 5 present the full result of regression (1) and regression (2). In the regression (1) without tweet valence, Metascore is significantly and positively correlated with box-office revenue. It is consistent with the previous literature about the positive correlation between professional judgment and public appeal. Tweet valence is significant both with and without Metascore listed as independent variable, and it is also consistent with the assumption that ordinary evaluation is positively related to public appeal. However, Metascore loses its significance while adding word-of-mouth indicators, tweet valence and volume, into the regression. This finding implies that tweet opinions could be valued above professional critics, and professional critics cannot substitute or overwhelm the effect of learning from online word-of-mouth communication.

3.3 Ordinary Evaluation as a Dependant Variable

To explore the relationship between ordinary evaluation and professional judgment, a regression relating tweet valence to Metascore is estimated. It takes tweet valence as the dependent variable. The tweet valence of movie *j* at week *t* can be identified as

$$VALENCE_{jt} = \lambda_0 + X_j\lambda_1 + \lambda_2\, Metascore_j + \lambda_3 \ln VOLUME_{jt} + \lambda_4\, \theta_{tj} + \upsilon_{jt}, \quad (3)$$

for which X denotes movie-specific factors that are fixed to movie j. $VALENCE_{jt}$ is defined as the average ratio of positive tweets to total tweets from Sunday through Thursday before weekend t, while $VOLUME_{jt}$ is the tweet volume simultaneous with $VALENCE_{jt}$. $VOLUME_{jt}$ is usually positively correlated with the absolute level of box-office revenue, so adding $VOLUME_{jt}$ in the regression is a way to control for the simultaneous effect of the box-office revenue level at the same week. θ denotes the both time-specific and movie-specific factors such as the number of weeks and holiday weekend. υ denotes the random disturbance. Rogers clustered standard errors is applied to regression (3) to control for the movie-specific effect on the residual term.

Table 6 shows the result of regression (3). As the previous studies suggest, Metascore is significantly correlated with tweet valence. Considering the fact that Metascore ranges from 1 to 100 while valence ranges from 0 to 1, the positive coefficient of Metascore, 0.003, is not too small. It is consistent with the assumption of good taste that ordinary evaluation is positively correlated with professional critics.

It is also interesting that tweet valence is not significantly correlated with production budget. Production budget of a movie is usually positively correlated with its advertisement expenditure. The insignificant relationship between production budget and ordinary evaluation implies that ordinary evaluation is hard to be influenced by advertisement and promotions.

The temporal sequence of Metascore and tweet valence can be identified easily since professional criticism is published before the online word-of-mouth starts. However, it is still difficult to identify the causal relationship between Metascore and tweet valence. A way to indirectly observe the causal relationship is to see if professional critics have more influence on ordinary evaluation in the opening weekend than in the following weekends. If so, professional critics do have some impact on ordinary consumers when they do have much information about the movie. But my data does not allow me to simply regress $VALENCE_{j1}$ when t=1 on Metascore$_j$ because among 52 movies, only 18 movies have the opening weekend tweet data. Hence, I propose the following equation to identify the dynamic relationship between professional critics and ordinary evaluation:

$$Gap_{jt} = \gamma_{0} + X_j \gamma_1 + \gamma_2 \, Metascore_{j} + \gamma_3 \ln VOLUME_{jt} + \gamma_4 \, \theta_{jt} + \rho_{jt}, \quad (4)$$

for which *Gap* is defined as the absolute value of (valence*100-Metascore). *X*, $VALENCE_{jt}$,, $VOLUME_{jt}$ and θ have the same definition with regression (3). ρ denotes the random residual term which could be subjected to movie-specific correlations. Other a fixed-effects model, Rogers clustered standard errors is also applied to equation (4) to control for the movie-specific effect on the residual term. The results of the fixed-effects model and the Rogers clustered standard errors model will both be reported in Table 2.

Gap measures the degree that tweet valence deviate from Metascore. One may doubt that the absolute value could not be the best measure of the difference between professional critics and ordinary evaluation. But I am focusing on the question that whether ordinary evaluation diverge from professional criticism, not the magnitude of positive or negative difference between valence and Metascore. The absolute value of the difference between the valence and Metascore provides me a better proxy of the degree of divergence. On the other hand, in my dataset, the absolute value of (valence*100-Metascore) equals to the value of (valence*100-Metascore) in most of the cases. Twitter users tend to give a movie higher opinion than experts does. As discussed in section 2, it could be resulted in the fact that only commercial movies are selected to my dataset, and professionals tend to give lower scores to commercial movies than ordinary consumers do.

Included in θ_{jt}, "number of weeks" is a variable showing that how many weeks a movie has been in theater at time *t*. If "number of weeks" =1, it is the first week for the movie to be in theater. The sign of its coefficient shows whether the gap between tweet valence and Metascore increases or decreases over time. Under the presumption that valence is positively correlated with Metascore, if ordinary consumers learn from professional critics in the opening weekend, the gap between valence and Metascore is not expected to increase over time. It is due to the assumption that professional critics will have stronger influence on ordinary consumers during the opening weekend than in the following weekends. In other words, professional critics have influence effect on ordinary evaluation during the opening weekend but not after. If we observe a decreasing gap between tweet valence and Metascore, the consistency between Metascore and valence could be a result of a similar reaction to movie quality, not a result of learning from critics.

Table 6 presents the result of regression (4). The results of the fixed effects model and the OLS model with Roger clustered standard error are presented in column 2 and column 3. The fixed-effects model omit all the time-invariant variables hence the coefficients of some variables are not estimated. The coefficient of “number of weeks” is insignificantly negative in the OLS model, while it is significant and negative in the fixed-effect model. It implies that the gap between Metascore and tweet valence is not increasing over time. If we take the fixed-effect model as a way to control for the unobservable quality of a movie, we can go one step forward to say that the gap between Metascore and tweet valence decreases over time, supporting the hypothesis that ordinary consumers do not learn from professional critics. Ordinary evaluation may converge to professional criticism gradually, but it could be a result of word-of-mouth among ordinary consumers. Professional judgment and ordinary evaluation are positively correlated because they react to movie quality in the same direction, not because ordinary consumers are influenced by experts.

It is noteworthy that the coefficient of production budget is insignificant in regression (3). I also change the dependent variable to Metascore and re-esitimate regression (3) and present the result in the last column in Table 6, despite its very low R square. As presented in section 3.2, production budget is also insignificant in the regression of public appeal. Production budget is a proxy of advertisement expenditure of a movie. My panel data with the dynamic of box-office revenue and ordinary evaluation covers a long time period which allows consumers to revise their expectations by word-of-mouth, hence advertisement and promotions has little influence on both public appeal and ordinary evaluation in the whole data period. It is consistent with the result that Metascore is not correlated with production budget, supporting the conventional thinking that professional judgment is hard to be influenced by a movie's advertisement and promotions. Ordinary evaluation still shows consistency with professional judgment on producers' promotion.

Compared to the insignificance of production budget in regression (3), it is interesting that production budget is significant at a 0.1 significance level. If we assume that professional judgment is neutral of promotions and advertisement, the negative coefficient implies that movies with larger production budget or advertisement do not earn much favorable ordinary evaluation; in contrast, for those movies, ordinary evaluation is more likely to converge to professional critics than for low budget movies. One of the possible explanations is that high budget may help to raise ordinary evaluation in the beginning, but later consumers start to revise their evaluation after receiving information. The revising is greater for high-budget movies than low-budget movies, so the gap between ordinary evaluation and professional judgment becomes smaller.

An interesting question is that whether genre has something to do with the gap between ordinary evaluation and professional judgment. To make the degree of freedom not to shrink too much, I list only two dummy variables to divided movies into three genre groups: one is “whether it is action movie or not”, the other is “whether it is drama or not”. Comedy, romance and thriller are all categorized as “drama”. On table 6, “Action movie or not” is insignificant in regression (4), while “drama or not” is significantly positive. Ordinary consumers and professionals have less discrepancy of opinions on action movies than on drama. Of course, the result could be due to the loose definition of drama genre in my research. But according to Pierre Bourdieu’s (1979) perspective on taste, compared to action movies, drama movies may require audience to have more social capital to enjoy them. Experts are expected to have more social capital than ordinary consumers, hence it is possible that the gap of the social capital leads a gap between ordinary evaluation and professional judgment.

The coefficient of Metascore is significantly negative in regression (3) and (4). In other words, when Metascore is low, the gap between tweet valence and Metascore is larger. In other words, experts and ordinary consumers are more likely to have a discrepancy of opinion on the movies which receive unfavorable expert reviews. To see whether the difference between valence and Metascore is positive or negative, I use the gap without absolute value as the dependent variable and re-estimate the regression (4). The coefficient of Metascore is still significantly negative. It shows that ordinary consumers tend to give the movies which receive unfavorable expert reviews higher opinion than professionals do.

4. Conclusion

Many economists connect their research on professional critics to Pierre's (1984) perspective on taste. One of the main research questions is that whether the public have "good taste" which is consistent with professional judgment. The debate over taste has implications for economics, marketing and even policy: if ordinary consumers have "bad taste", subsidy for high art could be necessary when art of "good taste" has positive externalities.

Previous empirical studies focus on the relationship between professional judgment and public appeal. However, little research measures ordinary evaluation directly. The intervention of ordinary evaluation between professional judgment and public appeal can make the relationship hard to be observed. For example, ordinary consumers may follow experts to give a movie low opinion but still want to see this "bad" movie. The pairwise relationships between ordinary evaluation, professional judgment and public appeal need to be examined separately.

Using panel data from the social network websites, Twitter, I investigate the two questions: whether ordinary evaluation is consistent with professional criticism, and whether ordinary evaluation positively correlated with public appeal. My data allows me to measure the dynamics of ordinary evaluation directly by the ratio of positive tweets on a movie. The weekly change of the gap between twitter opinion and professional criticism helps us to investigate the influence effect of expert reviews.

My results confirm that there are positive correlations between professional judgment, ordinary evaluation and public appeal. Ordinary evaluation is highly consistent with professional judgment, supporting the "good taste" hypothesis. High production budget is insignificant to professional criticism, as it is insignificant to twitter opinion, too. Moreover, if we control the unobserved movie quality by a fix-effects model, the gap between ordinary evaluation and professional judgment decreases over time. It could be a sign that ordinary evaluation is influenced by promotions or advertisement in the beginning, but later ordinary consumers revise their expectation and converge to experts' opinion. In this case, professional critics may have little influence effect since they do not have more impact in earlier weeks than later.

My evidence suggests that ordinary consumers have good taste, not supporting Pierre's (1984) perspective on taste. There seems to be little distinction between public opinion and expert judgment. However, some details in my empirical research may still echo Pierre's argument on social capital. Twitter opinion tends to give those commercial higher opinions than expert do. Also, ordinary consumers and professionals have more discrepancy of opinions on drama than on action genre. It is possible that the gap of the social capital leads a gap between ordinary evaluation and professional judgment. The issue of social capital still needs further investigation.

References:

1. Bourdieu, Pierre. (1984). *"Distinction: A Social Critique of the Judgement of Taste" Cambridge: Harvard U. Press*
2. Eliashberg, J. ; Shugan, S. M., 1997, 'Film Critics: Influencers or Predictors?,*' Journal of Marketing*, 61, pp. 68–78.
3. Hennig-Thurau, Thorsten; Houston, Mark B.; Sridhar, Shrihari; 2006, "Can good marketing carry a bad product? Evidence from the motion picture industry", *Market Letters*, 17:205–219
4. Holbrook, Morris B; 2005, "The Role of Ordinary Evaluations in the Market for Popular Culture: Do Consumers Have "Good Taste"?", *Marketing Letters*, Vol. 16, No. 2 (Apr., 2005), pp. 75-86
5. Holbrook, Morris B. (2003). "Perish Or Publish: On the Cultivation of Bad Taste and Surviving Your Fifteen Minutes of Fame," Working Paper, Graduate School of Business, Columbia U.
6. Liu, Yong (2006), "Word-of-Mouth for Movies: Its Dynamics and Impact on Box Office Revenue," Journal of Marketing, 70 (3): 74-89.
7. Liu, Yu-Hsi (2012), "How Does Online Word-of-Mouth Influence Movie Revenue? Evidence from Twitter", Working Paper.
8. Reinstein, David A.; Snyder, Christopher M.; 2005, "The Influence of Expert Review on Consumer Demand for Experience Goods: A Case Study of Movie Critics", *The Journal of Industrial Economics*, Volume LIII No. 1: 27-51

9. West, Patricia M.; Susan M. Broniarczyk. 1998. "Integrating Multiple Opinions: The Role of Aspiration Level on Consumer Response to Critic Consensus." *Journal of Consumer Research,* 25(1):38–51.

Table 1: The Basic Information of the Tweet Dataset

Variable	Number of Observations
The number of movies in theater during the data period	286
The number of movies in the data set	57
Maximum data duration for a movie (days)	94
Minimum data duration for a movie (days)	3
Average of data duration for a movie (days)	77.22
Average total (accumulated) number of tweets per movie	7438.87
Average daily increment number of tweets per movie	110.47

Table 2: The Comparison of the Movies from the Two Data Sets

		Average	Max	Min
57 Movies in the Twitter Dataset	Total Box Office Revenue	$90,813,260	$366,007,900	$1,183,354
	Production Cost	$73,890,385	$260,000,000	$1,500,000
	* 6 movies out of 57 have their production cost data missing.			
		Average	Max	Min
286 Movies on theater in the same period	Total Box Office Revenue	$22,043,286	$366,007,900	$11
	Production Cost	$49,917,031	$260,000,000	$135,000
	* 189 movies out of 286 have their production cost data missing.			

Table 3: Descriptive Statistics of Variables of Interest

	Number of Observations	Mean	s.d.	Min	Max
Metascore	51	55.451	14.768	29	87
Tweet Valence	283	0.689	0.106	0.37	0.95
Gap between Metascore and Valence (without absolute value)	283	11.168	13.006	-15	40.333
Gap between Metascore and Valence (with the absolute value)	283	13.203	10.927	0	40.333
Production Budget (million)	45	74.564	64.834	1.5	260

Table 4: The Covariance Matrix of Variables of Interest

	Metascore	Tweet Valence	Gap (without the absolute value)	Production Budget
Metascore	1			
Tweet Valence	0.395	1		
Gap (without the absolute value)	-0.6921	0.3897	1	
Production Budget (million)	-0.165	-0.3384	-0.1005	1

Table 5: Public Appeal as the Dependent Variable (t or z-value beneath)

Variables	Regression (1) ln(Rev_{jt}) as the dependent variable			Regression (2) ln d_{jt} as the dependent variable	
Tweet Valence$_{jt}$	---	4.289824***	3.870926***	---	6.687643***
		(4.78)	(4.16)		(2.14)
Metascore$_j$	.0193786***	---	.0076652	.0047034	.0003148
	(2.70)		(1.10)	(1.61)	(0.20)
Tweet Volume$_{jt}$	---	.2260198***	. 2272448***	---	. 2051919***
		(7.13)	(7.07)		(3.42)
The Log Number of Theaters $_{jt}$	1.047739***	.8822426***	.8766562***	.0254635	.113932***
	(10.04)	(13.10)	(12.67)	(0.51)	(4.92)
Budget$_j$	.0050071***	.0063829***	.0063212***	-.0001576	.0000903
	(3.04)	(4.72)	(4.69)	(-0.27)	(0.22)
Age factor$_j$: less than 13	-.1206548	.1008298	.0946017	-.0892129	-.023859
	(-0.49)	(0.53)	(0.49)	(-0.89)	(-0.42)
Age factor$_j$: greater than 25	.0199043	.464287***	.3887583*	-.0187926	-.1428213**
	(0.09)	(2.22)	(1.86)	(-0.33)	(-2.05)
Action Dummy$_j$	-.2266014	-.3146439*	-.3402305*	-.0617022	-.1870169**
	(-0.95)	(-1.67)	(-1.93)	(-0.69)	(-2.08)
Drama Dummy$_j$	-.101933	-.3431935*	-.3421007*	.1031843	-.1075971
	(-0.53)	(-1.74)	(-1.84)	(1.26)	(-1.22)
Constant	-.1185821***	4.267048***	4.210253***	-1.109292***	-1.423735***
	-3.11	(5.22)	(5.22)	(-2.87)	(-7.19)

***: significant at 0.01 level. **: significant at 0.05 level. *: significant at 0.1 level.
The results of some independent variables are not reported.

Table 6: Ordinary Evaluation as the Dependent Variable (t or z-value beneath)

Variables	Regression (3) $VALENCE_{jt}$ as the dependent variable	Regression (4) GAP_{jt} as the dependent variable: Roger Clustered s.d.	Regression (4) GAP_{jt} as the dependent variable: FE model	Metascore as the dependent variable (not panel data; very low R^2)
$Metascore_j$	. 0036356*** (3.55)	-.5873585*** (-6.72)	---	---
Number of $Week_{jt}$	-.0025159* (-1.82)	-.1818972 (-1.24)	-.2073809*** (-4.06)	---
Tweet Volume $_{jt}$	-.00242 (-0.54)	-.279117 (-0.58)	-.2877691*** (-2.35)	---
Budget $_j$	. -.0004236 (-1.44)	-.0476874* (-1.90)	---	-.0026381 (-0.06)
Age factor $_j$: less than 12	.0049644 (0.11)	2.558464 (0.70)	---	-1.026785 (-0.15)
Age factor $_j$: greater than 25	-.1033068 (-1.56)	-2.778314 (-1.00)	---	-3.446447 (-0.45)
Action Dummy	.0047381 (0.07)	4.974019 (1.39)	---	-1.91241 (-0.24)
Drama Dummy	.0401407 (0.84)	6.368361** (2.09)	---	-.6728859 (-0.08)
Constant	.5470272*** 6.35	48.16437*** 6.67	15.22063*** (11.76)	55.7002*** (-2.89)

***: significant at 0.01 level. **: significant at 0.05 level. *: significant at 0.1 level.
The results of some independent variables are not reported.