

I Heard It Through the Grapevine:
The Effects of Online Word of Mouth on Music Sales

By
Elaine Chang

An honors thesis submitted in partial fulfillment
of the requirements for the degree of
Bachelor of Science
Undergraduate College
Leonard N. Stern School of Business
New York University
May 2007

Professor Marti G. Subrahmanyam

Faculty Advisor

Professor Vasant Dhar

Thesis Advisor

Abstract:

Word of mouth (WOM) has always been one of the most credible and influential sources of information for a consumer. With the advent of the Internet, new mediums of communication, such as blogs and social networking sites, have opened up, allowing more consumers to gather WOM information more easily than ever before. In this study, I examine the usefulness of online WOM, particularly from blogs and social networking sites, in predicting sales in the music industry. I track the changes in online WOM for a sample of 108 albums for four weeks before and after their release dates, and use a linear regression and regression tree to examine the relative significance of online WOM variables on their observation date in predicting album unit sales two weeks after that date. I find that the volume of blog posts about an album is the most consistently significant measure, with each additional blog post corresponding to higher album sales. In addition, I found that greater increases in an artist's Myspace friends week over week and higher average consumer ratings also corresponded to higher sales. Although evidence supports my hypothesis that online word of mouth can be an important predictor of sales, my study also found that traditional factors are still relevant – albums released by major labels and albums with a number of reviews from mainstream sources like *Rolling Stone* also tended to have higher sales.

1. Introduction

In a world increasingly saturated with advertising, the old-fashioned concept of word of mouth (WOM) has remained one of the most credible and influential sources of information for a consumer. Classic examples of WOM are finding out about a new restaurant because a friend raved about a recent dinner she had there or seeing a movie after hearing how good it was from five different coworkers. WOM is especially important for experience goods – products whose quality is difficult to observe before purchase and consumption. Consumers must rely on indirect information on such experience goods as movies and music to make a purchase decision. About 48% of consumers report that friends are a major source of music information and recommendations, so it is no surprise that the music industry places great value on word of mouth marketing.¹

Traditionally, the music industry has relied on four different methods of increasing word of mouth and sales: concerts, retail relationships with stores like Best Buy, radio airplay, and publicity through television advertising and other media. Through the Internet, consumers now have the ability not only to purchase music online in both physical CD and digital form, but also to find recommendations for it by word of mouth (or mouse, as some would say) about the album through new mediums of communication such as blogs and social networking sites.

Blogs are short for “web logs” – websites that tend to be written in the style of a journal, with posts arranged in reverse chronological order. Some blogs are little more than online diaries, while others are almost collections of essays on weighty topics like politics, religion, or health. The growing importance of blogs has been well-noted by the media, with surveys estimating that

¹ “JupiterResearch Finds That MySpace Music Community Activity Far Outpaces Online Music Sites.” May 31, 2006. http://www.jupiterresearch.com/bin/item.pl/press:press_release/2006/id=06.05.31-myspace_music_community.html.

almost 30% of the American population considers blogs an important source of information.¹

Social networking sites, meanwhile, are sites that enable users to create profiles and make connections to other users who live in the same area, share similar interests, or simply seem interesting. These networks generally allow users to create a public list of other users that are mutual friends – that is, both users have listed each other as a friend. One of the most important social networks to the music industry is Myspace (www.myspace.com), which provides a special music category that allows artists to create profile pages including anything from band biographies to upcoming tour dates and streaming music tracks. Through these band profiles, individual Myspace users can add artists that they enjoy to their friends, simultaneously bookmarking the artist's work for later perusal and promoting the artist to their other friends. In addition, the number of friends a band has is displayed on its Myspace page – almost like a public badge of popularity – so a band with a couple hundred thousand friends on Myspace is certainly more popular with Myspace users than a band with two. While the media has reported extensively on the growing influence of these new online mediums, their potential scope of influence on the success of a band is better illustrated through an example.

In early 2005, Clap Your Hands Say Yeah (CYHSY) was just another rock band trying to make it big. Since no record label had showed any interest in signing the group, CYHSY pressed their self-titled debut album on their own and sold it at CYHSY's concerts and on the Internet.² To get the album to Internet buyers, the band would go to the post office and mail each individual CD by hand. Things began to change when, on June 9, 2005, a popular music blog, Said the Gramophone (www.saidthegramophone.org), posted two MP3s of songs off CYHSY's debut

¹ "Online Survey Finds General Public, Media Conference Attendees Agree that Traditional News Outlets Could Do a Better Job." *PR Newswire*. February 17, 2007.

² "To Generate Buzz, Clap Your Hands on the 'Net!" NPR. Nov. 29, 2005. <http://www.npr.org/templates/story/story.php?storyId=5023133>.

album, commenting, “there are three *amazing* songs on this record, and bunch of other great ones... Just get the album, it's very worth it.”¹ With that, Said the Gramophone opened the door to a rush of buzz about the band, including a review of one of the songs on the album by Pitchfork (www.pitchforkmedia.com), an influential online music magazine, and reports of David Bowie attending a CYHSY concert from Gothamist (gothamist.com), a New York City area blog.² By the end of August, CYHSY had debuted on Billboard’s Top Independent Albums Chart – and the band still had not signed with a label.

So far, no other band has yet managed to successfully duplicate CYHSY’s “do it yourself” mode of operation. However, the music industry has already begun to adapt – artists are now regularly signed by record labels after being discovered on social networking sites like Myspace and elsewhere on the Internet. Despite this, there has not been any conclusive data on whether great online WOM is indicative of a potential hit-making band. Therefore, my aim in this study is to examine if online word of mouth is truly a valid predictor of music sales, or whether CYHSY was just an anomaly.

To do so, I trace the changes in online WOM for an album by tracking blog chatter, the number of friends an artist has on Myspace, and album reviews for four weeks before and after the release date. I control for the influence of external differences in promotion budgets and so on by introducing a dummy variable for whether an album is released by a major or independent label. I then use a linear regression and regression tree to examine the relative significance of online WOM variables on their observation date in predicting album unit sales two weeks after that date.

My findings suggest that, of all the online WOM measures, the volume of blog posts on an

¹ “You Will Pay for Your Excessive Charm.” Said the Gramophone. June 9, 2005. http://www.saidthegramophone.com/archives/you_will_pay_for_you.php.

² “Clap Your Hands Say Yeah: ‘This Home on Ice’ [track review].” Pitchfork Media. June 14, 2005. http://www.pitchforkmedia.com/article/track_review/30565-in-this-home-on-ice.

album is the most consistently significant, with each additional blog post corresponding to higher album sales. In addition, I found that greater increases in an artist's Myspace friends week over week and higher average consumer ratings also corresponded to higher sales. Although evidence supports my hypothesis that online word of mouth can be an important predictor of sales, my study also found that traditional factors are still highly relevant – albums released by major labels and albums with a number of reviews from mainstream sources like *Rolling Stone* also tended to have higher sales.

The remainder of the paper is organized as follows: after reviewing the relevant literature on the topic in Section 2, I describe the sample, the various variables included in the data analysis, and the sources I use to gather the data. Section 4 discusses my methods of analysis and the empirical results I obtain. I conclude the paper in Section 5 and propose topics worth further research in Section 6.

2. Relevant Literature

While few papers have been written about online word of mouth in the music industry, there is a large body of existing work examining word of mouth both offline and online in the movie and book industries.

Eliashberg and Shugan (1997) found statistical evidence that the percentage of positive and negative professional movie critics' reviews were predictive of total cumulative box office sales, though not of early box office sales (sales in the first four weeks).¹ In addition, they found that the number of reviews is a significant predictor of the first week's box office sales, but not of the

¹ Eliashberg, Jehoshua and Steven Shugan (1997). "Film Critics: Influencers or Predictors?" *Journal of Marketing*, Vol. 61, No. 2, 68-78.

movie's performance over its entire lifecycle. Moul (2006) examined a sample of theatrical movie admissions from 1990-1996, and found that approximately 10% of the variation in implied consumer expectations among movies is due to the effect of word of mouth.¹

While the previous studies used data from before the Internet became ubiquitous in the late 1990s, other researchers have since examined online word of mouth about movies. Dellarocas et al. (2004) explored more specifically the usefulness of online product ratings in predicting movie revenues.² The researchers first conducted a survey asking a random sample of people who did not rate movies online to rate a number of movies, and then compared the results to the online movie review data they gathered. The correlation between the ratings given by the two populations was significant, suggesting that online consumer reviews are a valid proxy for total word of mouth for a movie. Dellarocas et al. also found that weighted averages of user ratings were more informative in predicting future movie sales than professional critic reviews.

Duan et al. (2005) also explores the relationship of online user ratings to movie revenues.³ Duan et al. found that the average online user review rating does not have a significant impact on box office revenues. The number or volume of online postings, however, is significant. The researchers suggest that online word of mouth ratings do not influence consumers because consumers will make an independent judgment about the movie's quality by drawing upon many different sources. Increases in the number of postings, however, do affect consumers because the postings provide more information about the movie.

Chevalier and Mayzlin (2006) turn their attention towards the book industry, examining the

¹ Moul, Charles (2006). "Measuring Word of Mouth's Impact on Theatrical Movie Admissions." *Journal of Economics and Management Strategy*, forthcoming.

² Dellarocas, Chrysanthos, Neveen Farag Awad, and Xiaoquan (Michael) Zhang (2004). "Exploring the Value of Online Product Ratings in Revenue Forecasting: The Case of Motion Pictures." Working paper, University of Maryland. May 2004.

³ Duan, Wenjing, Bin Gu and Andrew B. Whinston (2005). "Do Online Reviews Matter? – An Empirical Investigation of Panel Data." Working Paper, the University of Texas at Austin. January 2005.

effects of online consumer ratings on book sales at Amazon.com and Barnes and Noble's website, BN.com.¹ They found that additional positive reviews at Amazon.com would increase sales at Amazon.com relative to BN.com and vice versa. Additional evidence suggested that very negative reviews had a greater impact on sales than very positive reviews.

My research differs from these studies in that I specifically focus on the music industry, and that I examine word of mouth variables aside from consumer reviews posted on review websites – I also look at new online communication mediums such as blogs and social networking sites.

3. Data

3.1 The sample

My data consist of album statistics and word of mouth data collected from publicly available information on websites such as Amazon (www.amazon.com). The data were either manually recorded or downloaded through automated Javascript programs. I compiled the sample of music albums by collecting the names of albums released in the United States between January 16 and March 6, 2007 from Pause & Play (www.pauseandplay.com), a website devoted solely to listing upcoming album releases. Old material, such as reissues and compilations, were excluded from the sample. I further cross-checked the release date given by Pause & Play with Amazon's page for the album in order to verify that the record label had not moved the release date, and if the album did not have a corresponding page on Amazon, it was eliminated from the sample. The final sample consisted of a total of 108 albums.

¹ Chevalier, Judith and Dina Mayzlin (2006). "The Effect of Word of Mouth on Sales: Online Book Reviews." *Journal of Marketing Research*, Vol. 43, No. 3, 345-354.

3.2 Dependent variable: album sales

Since information on digital music sales is difficult to obtain, I focused on physical CD sales. Although digital music sales have grown tremendously over the past few years, physical album sales still account for 85% of the music market.¹ Nielsen SoundScan would have been the ideal source for album sales data, as it is the industry standard tracking system for sales of music products in the US, but its data is proprietary and very expensive to obtain.

I chose Amazon.com sales ranks for my sales data instead, as Amazon is one of the largest online CD retailers and its sales ranks are easily observed.² Each product page on Amazon lists the product's sales rank relative to all products in its category. Sales ranks order albums according to quantity: the highest-selling album is ranked #1, the next highest #2, and so on. In addition, Amazon allows consumers to preorder, or purchase products that are announced far ahead of the actual release date. Products available for preorder are also given sales ranks; in notable instances, such as the Harry Potter series, products have made it to #1 on Amazon without even being released.³ The preorder feature on Amazon provides an important advantage for my study because it allows the effect of word of mouth to be tracked before *and* after the release date; I would expect that many consumers who are inspired to purchase a to-be-released album after reading positive reviews online would want to put in a preorder almost immediately instead of making a quickly-forgotten mental note to come back and make the purchase when the album is released.

One note that needs to be made is that if an album's sales are lower than a certain cutoff point, Amazon does not even calculate the sales rank, simply giving the album a ranking of "none".

¹ "Can Music Survive Inside the Big Box?" *Wall Street Journal*. April 27, 2007. B1.

² "Apples iTunes Slams Into Top Ten Music Retail Chart." *The Register*. November 22, 2005. http://www.theregister.co.uk/2005/11/22/itunes_us_retail_top_ten/.

³ "Where Harry Most Enchants." *The Washington Post*. April 17, 2007. B01.

The cutoff rank for each category is not published, but the ranks in my sample ranged from 1 to 603,638 and followed an exponential decay pattern with fewer albums receiving very high ranks. Therefore, for the purposes of data analysis, I set such “none” rankings to 700,000. In addition, Amazon calculations of sales ranks take into account sales in its Amazon Marketplace, which allows buyers to travel from a link on the product page to listings by individual merchants selling used and new copies of the product. Thus, the sales quantities I use may include sales that the record company will not capture revenue from.

I converted weekly sales ranks into weekly sales quantities using a technique previously developed by Brynjolfsson et al. (2003) and Chevalier and Goolsbee (2003).¹ Both studies found ample evidence of the following Pareto relationship between the ordinal sales rank of a book and its weekly sales quantities:

$$\ln Quantity = \alpha + \beta * \ln Rank$$

For the α parameter, I used the 10.526 calculated for book sales on Amazon by Brynjolfsson et al. (2003). Professor Anindya Ghose provided a β parameter of -1.61 that he calculated for DVDs.² Since these parameters are not calibrated specifically to the music category on Amazon, the actual estimates for the parameters could be significantly different. The possible effect of different parameter estimates is explored in detail in the results section of this paper.

In the data analysis, I use weekly sales quantities for two weeks after the observation date. The reason for this is to calculate the predictive abilities of the independent variables.

¹ Brynjolfsson, Erik, Yu (Jeffrey) Hu, and Michael D. Smith (2003). “Consumer Surplus in the Digital Economy: Estimating the Value of Increased Product Variety at Online Booksellers.” *Management Science*, Vol 49, No. 11, 1580-96; Chevalier, Judith and Austan Goolsbee (2003). “Measuring prices and price competition online: Amazon and Barnes and Noble.” *Quantitative Marketing and Economics*, Vol. 1, No. 2, June 2003, 203-222.

² Ghose, Anindya, Michael Smith, and Rahul Telang. “Internet Exchanges for Used Books: An Empirical Analysis of Product Cannibalization and Welfare Impact.” *Information Systems Research*, Mar. 2006, pp. 3-19.

3.3 Independent variables

3.3.1 Days since release

As noted, music albums, like movies, are an experience good – a product whose characteristics are difficult to observe before purchase and consumption. Therefore, I expect that sales would follow the pattern of movie revenues, with sales highest right after the release date. As the pool of potential consumers decreased, I would expect sales to fall.

To account for the effect of time on sales, I calculated “days since release” by subtracting the release date from the observation date. Interestingly enough, virtually all of the CDs in the sample were released on a Tuesday, which may be related to when Billboard chart information is released or when music stores receive their shipments.

3.3.2 Major label versus independent label releases

Four major record groups dominate the music industry: Universal Music Group, Sony BMG Music Entertainment, EMI Group, and Warner Music Group. Combined, these four groups account for an estimated 82% of the music market.¹ Independent labels not associated with the four major record groups make up the remaining 18%. Since albums released by major labels are more likely to have larger promotion budgets and are often by more popular and mainstream artists, I accounted for the “major label effect” by introducing a dummy variable where 1 corresponded to a major label release, and 0 to an independent label release.

¹ “The Net is a Boon for Indie Labels.” *The New York Times*. December 27, 2005.
<http://www.nytimes.com/2005/12/27/arts/music/27musi.html?ex=1293339600&en=7d783101017430f0&ei=5088&partner=rssnyt&emc=rss>.

3.3.3 Average number of reviews and average rating

A major source of information for consumers about an album is CD reviews. The level of buzz around an album can be measured by the volume of reviews an album received, and its quality can be measured by how positive the average rating is. I gathered review data by compiling a number of review websites that posted either numerical (e.g. 5 star scale) or letter grade ratings, and manually checking those websites for new album reviews every Sunday. I then converted the rating systems of the different websites to a 100 point scale. If albums were not rated, then they received a 0 rating. While doing so does decrease the mean rating significantly, a 0 rating takes into account the fact that the album is receiving no word of mouth. Most of the albums did not receive consumer reviews until after the album was released, which helps explain the relatively low mean rating in Figure 1.

To examine the relative influence of different types of reviews, I sorted review sources into three different categories: consumer reviews, online media reviews, and mainstream media reviews.

Figure 1: Descriptive Statistics for Review Data

| Variable | N | Mean | SE Mean | StDev | Minimum | Q1 | Median | Q3 | Maximum |
|---------------------------------|-----|--------|---------|--------|---------|----|--------|-------|---------|
| No of consumer reviews | 864 | 4.85 | 0.575 | 16.911 | 0 | 0 | 0 | 3 | 208 |
| No of online media reviews | 864 | 0.4063 | 0.0281 | 0.8257 | 0 | 0 | 0 | 0 | 3 |
| No of mainstream media reviews | 864 | 0.8009 | 0.0292 | 0.857 | 0 | 0 | 1 | 1 | 3 |
| No of total reviews | 864 | 6.057 | 0.598 | 17.566 | 0 | 0 | 1 | 5 | 211 |
| Average consumer rating | 864 | 35.44 | 1.49 | 43.91 | 0 | 0 | 0 | 90 | 100 |
| Average online media rating | 864 | 15.45 | 0.957 | 28.136 | 0 | 0 | 0 | 0 | 84 |
| Average mainstream media rating | 864 | 40.91 | 1.25 | 36.72 | 0 | 0 | 60 | 73.33 | 91.67 |
| Average total rating | 864 | 50.38 | 1.32 | 38.93 | 0 | 0 | 70 | 83.19 | 100 |

Consumers. Consumer reviews represent the truest form of word of mouth. The reviewers are usually individuals who have already purchased or listened to the album in question, and are not being compensated for writing the review. In my study, consumer reviews were represented by Amazon.com's "customer reviews" feature, which allows individual users to submit reviews on a product and rate it on a 5 star scale. I recorded the

total number of customer reviews and the average number of stars given for each album. I then converted the star rating to a 100 point scale by multiplying the Amazon rating by 20.

Online media. I categorized “online media reviews” as those posted by media websites that only have a presence on the Internet and do not publish “dead tree” magazines or other materials. I chose websites that are, at least anecdotally, considered influential on the music scene: Pitchfork Media (www.pitchforkmedia.com), PopMatters (www.popmatters.com), and Stylus Magazine (www.stylusmagazine.com). Pitchfork and PopMatters both rate albums on a 1-10 scale, so I multiplied their ratings by 10. Stylus Magazine rates albums on an A+ to F scale that includes + and – gradations for all letter grades besides F. To obtain a numerical rating, I set A+ to equal 100 and F equal to 0, leaving equal divisions between each possible rating. I then averaged the three ratings with equal weights to obtain the “average online media rating.”

Mainstream media. “Mainstream media reviews” differ from the other two in that they are reviews that are posted online by major print media publications. I chose Rolling Stone (www.rollingstone.com), Entertainment Weekly (www.ew.com), and Allmusic (www.allmusic.com). Allmusic is known primarily to consumers as an online music information and review database, but it also publishes All Music Guide reference books and licenses its data and reviews to Billboard so that an artist search on Billboard’s website will return an Allmusic review. For our intents and purposes, it will be considered a mainstream media site. Rolling Stone and Allmusic grade music albums on a 1-5 scale, and Entertainment Weekly uses an A+ to F scale like Stylus Magazine’s. The ratings for the

three were averaged, with each source weighted equally.

3.3.4 Blog chatter

A vast array of blogs exists on the Internet, spanning a variety of subjects; even within a subject, there are many different types of blogs. A music blog might focus on only one specific genre of music, or post only concert reviews, or offer MP3s for readers to download. Many music blogs might post CD reviews, but not all of them rate CDs on any sort of scale. This makes capturing blogs' average rating of an album extremely difficult, but it seems plausible that any mention of an album on a blog still constitutes word of mouth because it generates consumer awareness. The more blog posts about an album that are online and accessible to a consumer, the more likely a consumer will have enough information about the album to make a purchase decision.

Therefore, to measure "blog chatter," I looked at blog post volume. I did searches on Technorati, a search engine for blogs, for each album in the sample and noted the total number of blog posts that came up. I limited the search to only blogs in English with some authority – in Technorati terms, the more other blogs link to a blog post, the more authority that post has. This constraint was necessary to narrow the results down to legitimate blogs written by humans instead of spam sites that list all sorts of keywords in hopes of getting hits.

I also calculated the "change in chatter" from week to week, since a rapid increase in word-of-mouth may signal a critical amount of buzz surrounding an album, which would imply that sales may increase faster than normal.

3.3.5 Myspace

One measure of the popularity of an artist is the number of friends he or she has on Myspace. Myspace, a social networking site recently bought by News Corporation, has a strong reputation for promoting music artists. Approximately 80% of the artists in the sample – which ranged from John Mellencamp to American Idol runner-up Katherine McPhee – had an official Myspace page. I gathered Myspace friends data by visiting each artist's Myspace page and noting the number of Myspace users who had added them as friends, which is publicly displayed on each profile.

I also calculated the percentage change in Myspace friends from week to week, since I hypothesized that if word of mouth about the album is good, the artist's Myspace friends will increase more rapidly.

4. Analysis and results

To analyze my data, I regressed the independent variables on weekly unit sales of an album two weeks after the observation date. I first examined the correlations between the variables – similar to the results in Dellarocas (2004), I found that average mainstream media ratings showed low correlation with consumer ratings (see Figure 2).¹ Online media ratings showed an even lower correlation with mainstream media ratings, showing that online word of mouth variables provide new and different information from mainstream ratings. In addition, of all the variables, blog chatter showed the greatest correlation (0.495) with sales two weeks after the observation date.

¹ Dellarocas, Chrysanthos, Neveen Farag Awad, and Xiaoquan (Michael) Zhang (2004). "Exploring the Value of Online Product Ratings in Revenue Forecasting: The Case of Motion Pictures." Working paper, University of Maryland. May 2004.

Figure 2: Pearson correlation matrix for variables data

| | Days since release | Number of consumer reviews | Number of online media reviews | Number of mainstream media reviews | Total number of reviews | Average consumer rating | Average online media rating | Average mainstream media rating | Average total rating | ln (sales 2 weeks after obs date) | ln (blog chatter) | ln (number of Myspace friends) | ln (% change in Myspace friends) |
|------------------------------------|--------------------|----------------------------|--------------------------------|------------------------------------|-------------------------|-------------------------|-----------------------------|---------------------------------|----------------------|-----------------------------------|-------------------|--------------------------------|----------------------------------|
| Number of consumer reviews | 0.280 (0.000) | | | | | | | | | | | | |
| Number of online media reviews | 0.370 (0.000) | 0.255 (0.000) | | | | | | | | | | | |
| Number of mainstream media reviews | 0.513 (0.000) | 0.459 (0.000) | 0.548 (0.000) | | | | | | | | | | |
| Total number of reviews | 0.312 (0.000) | 0.997 (0.000) | 0.319 (0.000) | 0.516 (0.000) | | | | | | | | | |
| Average consumer rating | 0.682 (0.000) | 0.318 (0.000) | 0.349 (0.000) | 0.488 (0.000) | 0.347 (0.000) | | | | | | | | |
| Average online media rating | 0.389 (0.000) | 0.188 (0.000) | 0.839 (0.000) | 0.449 (0.000) | 0.242 (0.000) | 0.314 (0.000) | | | | | | | |
| Average mainstream media rating | 0.496 (0.000) | 0.187 (0.000) | 0.371 (0.000) | 0.795 (0.000) | 0.237 (0.000) | 0.446 (0.000) | 0.339 (0.000) | | | | | | |
| Average total rating | 0.621 (0.000) | 0.234 (0.000) | 0.321 (0.000) | 0.664 (0.000) | 0.273 (0.000) | 0.700 (0.000) | 0.370 (0.000) | 0.814 (0.000) | | | | | |
| ln (sales 2 weeks after obs date) | 0.067 (0.050) | 0.328 (0.000) | 0.196 (0.000) | 0.343 (0.000) | 0.341 (0.000) | 0.234 (0.000) | 0.110 (0.001) | 0.184 (0.000) | 0.177 (0.000) | | | | |
| ln (blog chatter) | 0.252 (0.000) | 0.332 (0.000) | 0.520 (0.000) | 0.511 (0.000) | 0.369 (0.000) | 0.319 (0.000) | 0.406 (0.000) | 0.371 (0.000) | 0.321 (0.000) | 0.495 (0.000) | | | |
| ln (number of Myspace friends) | 0.062 (0.112) | 0.189 (0.000) | 0.145 (0.000) | 0.210 (0.000) | 0.201 (0.000) | 0.123 (0.000) | -0.008 (0.830) | 0.169 (0.000) | 0.094 (0.017) | 0.359 (0.000) | 0.544 (0.000) | | |
| ln (% change in Myspace friends) | 0.000 (0.995) | 0.143 (0.001) | -0.009 (0.834) | 0.145 (0.001) | 0.145 (0.001) | 0.077 (0.074) | -0.012 (0.784) | 0.094 (0.029) | 0.097 (0.024) | 0.250 (0.000) | 0.127 (0.003) | -0.166 (0.000) | |
| Major/ Independent label | 0.001 (0.980) | 0.220 (0.000) | -0.013 (0.704) | 0.122 (0.000) | 0.217 (0.000) | 0.063 (0.063) | -0.100 (0.003) | 0.058 (0.086) | 0.012 (0.732) | 0.458 (0.000) | 0.210 (0.000) | 0.291 (0.000) | 0.265 (0.000) |

Pearson correlation
(P value)

4.1 Linear regression

In order to normalize the distributions of *BlogChatter* and *%ChgMyspace*, I took the logs of the variables. I then ran an ordinary least squares regression and estimated a model of the resulting form:

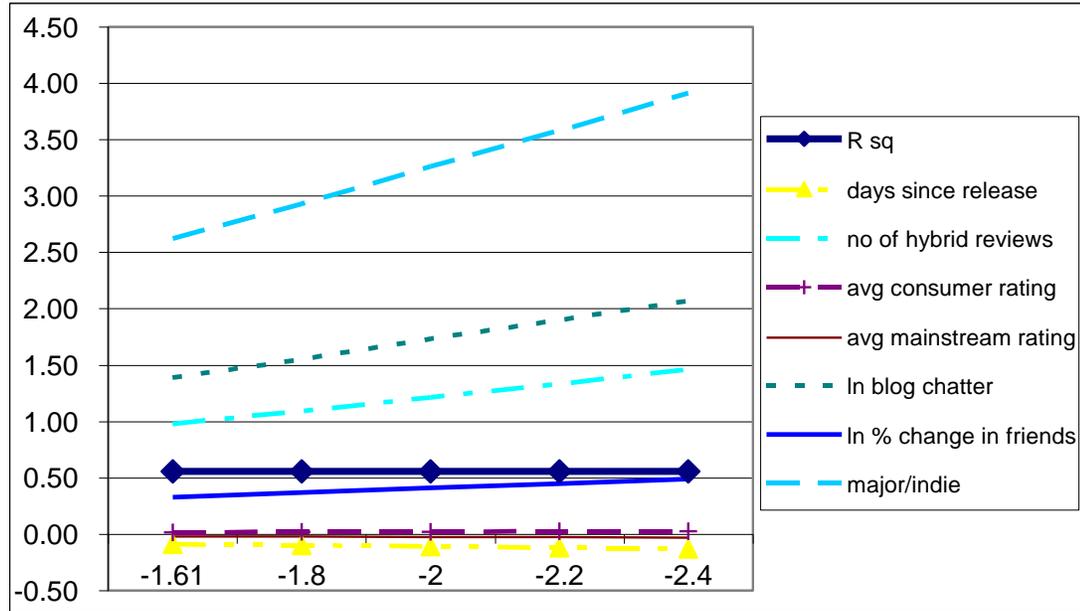
$$\begin{aligned} \ln(\text{Sales2WeeksAfter}) = & \alpha + \beta_1(\text{DaysSinceRelease}) + \beta_2(\text{MajorIndieLabel}) + \\ & \beta_3(\text{NoMainstreamReviews}) + \beta_4(\text{AvgMainstreamRating}) + \\ & \beta_5(\text{AvgConsumerRating}) + \beta_6 \ln(\text{BlogChatter}) + \\ & \beta_7 \ln(\text{\%ChgMyspace}) \end{aligned}$$

As stated in Section 3.2, the parameters in the Pareto relationship I used to convert sales rank into sales quantity are not calibrated to the CD category at Amazon. In order to ensure that the regression results were not distorted by this estimation, I ran a sensitivity analysis. I found that the overall R^2 of the linear regression equation remains unchanged no matter how the β parameter is adjusted (see Figure 3). Changes in the β parameter simply change the regression coefficients in a linear fashion (see Figure 4). The α parameter, since it is a constant parameter, does not affect the R^2 of the equation and only changes the constant in the regression equation.

Figure 3: A sensitivity analysis of the β parameter

| | | | | | |
|--------------------------|---------|---------|---------|---------|---------|
| B parameter | -1.61 | -1.8 | -2 | -2.2 | -2.4 |
| R sq | 0.56 | 0.56 | 0.56 | 0.56 | 0.56 |
| constant | -5.56 | -7.45 | -9.45 | -11.5 | -13.4 |
| days since release | -0.0889 | -0.0994 | -0.11 | -0.121 | -0.133 |
| no of mainstream reviews | 0.976 | 1.09 | 1.21 | 1.33 | 1.46 |
| avg consumer rating | 0.0168 | 0.0188 | 0.0209 | 0.023 | 0.0251 |
| avg mainstream rating | -0.0207 | -0.0231 | -0.0257 | -0.0283 | -0.0308 |
| ln blog chatter | 1.39 | 1.55 | 1.73 | 1.9 | 2.07 |
| ln % change in friends | 0.329 | 0.368 | 0.409 | 0.449 | 0.49 |
| major/indie label | 2.62 | 2.93 | 3.26 | 3.58 | 3.91 |

Figure 4: R² and independent variable coefficients versus β parameter



Since the R² of the regression equation is not sensitive to the Pareto relationship parameters, I used the previously stated α parameter of 10.526 and β parameter of -1.61 to obtain the results shown in Figure 5. The findings strongly suggest that higher blog post volumes and higher percentage changes in Myspace friends correspond to increased weekly sales in the future. This makes intuitive sense as higher numbers for either of these variables would indicate a greater consumer awareness of the album. In addition, both of these variables tend to have usable information *before* an album is released, while the majority of reviews in all review source categories only begin to appear within the first week surrounding the album release date.

Figure 5: Results for the linear regression

| Variable | Coefficient | T Statistics |
|----------------------------------|-------------|--------------|
| Days since release | -0.0904 | -7.62 |
| Number of mainstream reviews | 1.13 | 4.29 |
| Average mainstream review rating | -0.0241 | -4.44 |
| Average consumer review rating | 0.0179 | 4.69 |
| ln (Blog chatter) | 1.27 | 13.92 |
| ln (% change in Myspace friends) | 0.363 | 2.78 |
| Major or independent label | 2.54 | 8.99 |

* all significant at the 1% level

Interestingly, the coefficient for *BlogChatter* is larger than that for *%ChgMyspace*.

Therefore, an additional blog post corresponds to a greater increase in sales than an additional 1% increase in Myspace friends. Figure 6 and Figure 7 are based on a median album with median values for each independent variable. They clearly show that higher percentage changes in Myspace friends produce diminishing returns, which is not true for blog chatter.

There are a few possible explanations for this difference. One is that Myspace is considered a less credible source of word of mouth. Adding a friend on Myspace is a fairly simple process – a user need only click “add to friends” on the artist’s profile page and confirm it for the friend to be added. Therefore, it is possible that an artist’s Myspace friends count is bloated with individuals who only passed by the artist’s profile once, and are not very interested in the artist’s new album or in spreading word about the album. In addition, being a Myspace friend is a very passive process; no information, aside from the increased friend count, is produced for other potential consumers to view. On the other hand, blog writers have to spend some amount of time and effort on writing a blog post, and the post produced likely provides information about the album that the reader can use.

Figure 6: Blog chatter vs weekly sales quantities two weeks after the observation date

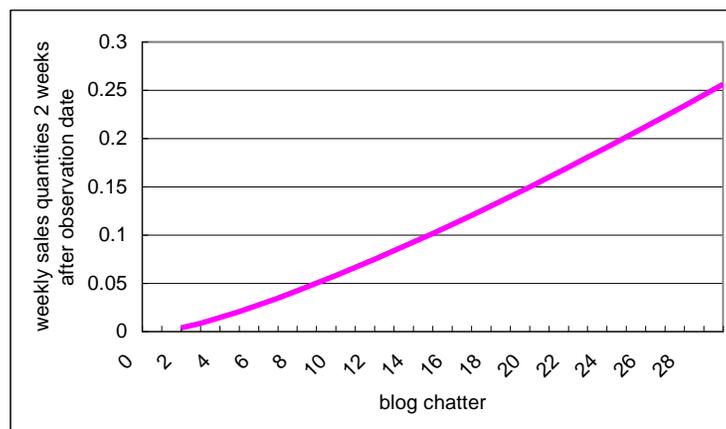
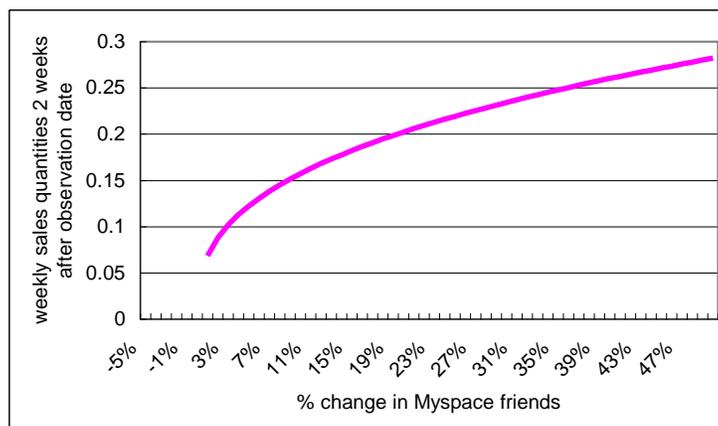


Figure 7: Percentage change in Myspace friends vs weekly sales quantities two weeks after observation date



The coefficients on the other independent variables are mostly as expected. The negative coefficient for *DaysSinceRelease* confirms my hypothesis that sales of an album fall over time after the release date. The positive coefficients for both *AvgConsumerRating* and *NoMainstreamReviews* suggest that album sales rise with higher consumer ratings and more mainstream reviews. Based on the coefficient for *MajorIndieLabel*, I estimate that a release by a major label has approximately twelve times the sale of a release by an independent label. Other sources have estimated that major label releases sell about six times more than independent label releases.¹ The discrepancy could be due to the small size of the sample I used – 108 albums in comparison to the approximately 35,000 albums released in a single year.² Nevertheless, the general trend is still correct.

The negative coefficient for *AvgMainstreamRating*, on the other hand, is surprising; generally, a higher critical rating would be assumed to correspond to higher sales. A closer examination of the data reveals a number of possible theories for such a coefficient. One is simply that critical acclaim may not correspond with popular appeal, and this seems to be supported by the

¹ Hull, Geoffrey. *The Recording Industry*. Routledge: New York, 2004, 136.

² Hull, 136.

relatively low correlation of 0.446 between the average mainstream media rating and average consumer rating shown in Figure 2. In addition, when I examined the five albums with the lowest mainstream media rating, all of them had high sales ranks (see Figure 8), indicating high popularity. It is difficult to say if this is representative of a trend, since the sample is very small; however, I speculate that there may be some sort of selection process occurring. Albums that will not be well-received by critics are only released by labels if the label believes the album will sell anyway; albums that are not surefire hits based solely on the artist's notoriety or other factors will only be released if they meet a certain minimum quality standard. Whatever the reason, the results seems to agree with the finding in Dellarocas et al. (2004) that average consumer ratings are better predictors than professional critics' ratings.

Figure 8: The five albums with the lowest mainstream media rating

| Artist | Album | Average Mainstream Rating | Sales Rank 2 Weeks After |
|------------------|------------------------|---------------------------|--------------------------|
| Celtic Woman | A New Journey | 41.67 | 10 |
| John Mellencamp | Freedom's Road | 50.00 | 16 |
| Kidz Bop Kids | Kidz Bop 11 | 50.00 | 136 |
| Katharine McPhee | Katharine McPhee | 53.33 | 48 |
| Art Garfunkel | Some Enchanted Evening | 54.17 | 149 |

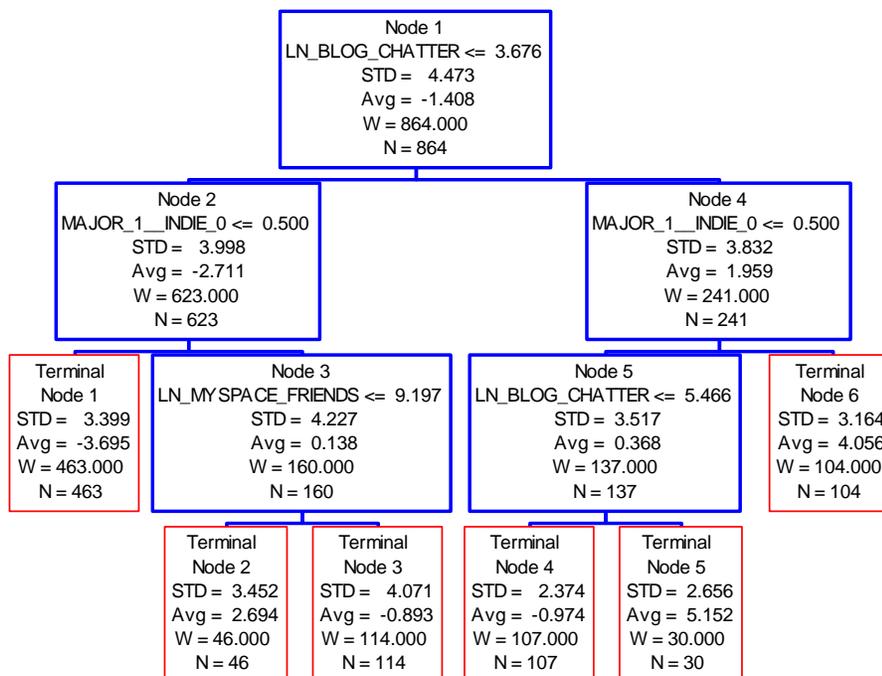
4.2 Regression tree

My regression tree analysis produces slightly different results (However, even if chatter is relatively high for an independent label (above 40 posts), sales will be higher than the average for the sample, but still relatively low if the 240 post level is not breached. An independent label with low blog chatter has very low sales, which is as expected.

Figure 9), as it only uses blog chatter, absolute number of Myspace friends, and type of label. It nevertheless confirms that blog chatter is a significant variable.

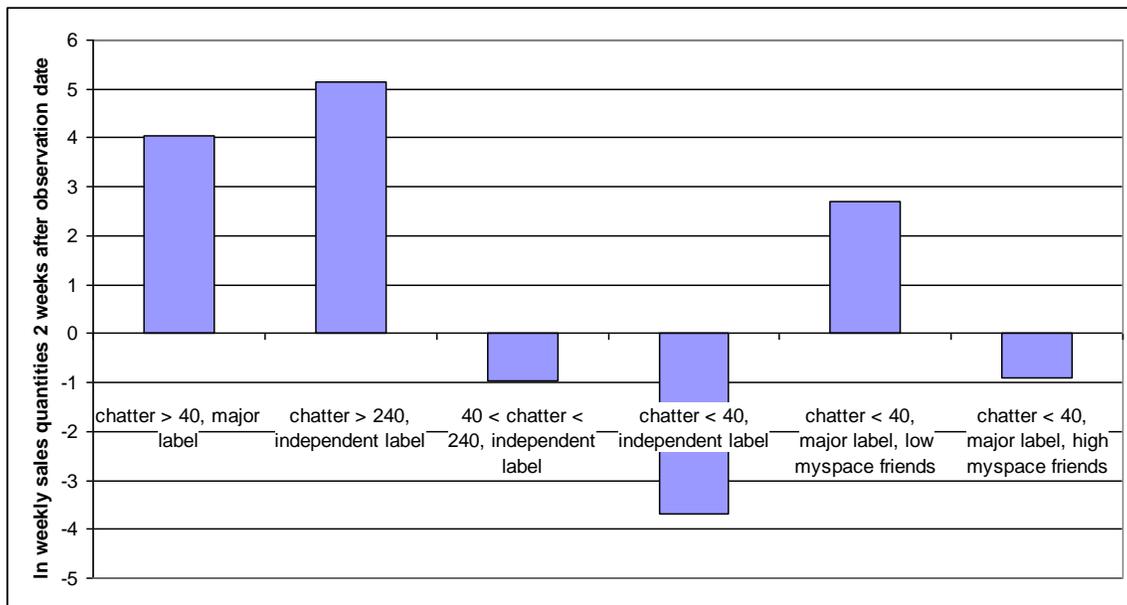
If an album has more than 40 blog posts, it will have an above average level of sales. If an album has more than 40 blog posts and is released by a major label, then it is likely to have very high sales. This is no surprise, as a large number of blog posts indicates a high level of online WOM, and being released by a major label means it is more likely that there will be significant promotion of the album through channels other than the Internet. Interestingly, though, if blog chatter is extremely high – above 240 posts – it is possible for an album to overcome the disadvantage of being released by an independent label. In fact, albums with such extreme highs in chatter correspond to sales even higher than major label, high chatter albums (see Figure 10). However, even if chatter is relatively high for an independent label (above 40 posts), sales will be higher than the average for the sample, but still relatively low if the 240 post level is not breached. An independent label with low blog chatter has very low sales, which is as expected.

Figure 9: Regression tree analysis



My findings also show that major label releases with low blog chatter and low numbers of Myspace friends have higher sales than major label releases with low blog chatter and high numbers of Myspace friends. This seems counter-intuitive at first, but examination of the sample found that major label releases without a Myspace page were considered to have 0 Myspace friends, which skewed the results. In addition, major label releases that had a Myspace page but few Myspace friends were from artists such as John Mellencamp and Art Garfunkel; I would presume that the majority of their audiences do not generally use Myspace, which may account for the discrepancy in the results.

Figure 10: Regression tree nodes



5. Conclusion

Via the Internet, consumers can find more word of mouth information on upcoming music releases than ever before. Previous research has found evidence that online consumer reviews can predict book and movie sales (Dellarocas et al. 2004, Duan et al. 2005, and Chevalier 2006), but no

study, as far as I know, has explored the effects of blogs and social networking sites as new mediums of online WOM communication.

In this study, I analyze the usefulness of WOM in these new mediums, as well as reviews in consumer, online media, and mainstream media, in predicting album sales in the four weeks before and after the album's release date. I found that the most significant online WOM variable is blog chatter or the volume of blog posts on an album, with higher numbers of posts corresponding to higher sales. Higher percentage changes in Myspace friends are also significant. Contrary to the conclusions in Duan (2005), I find that the average consumer rating is significant, while the number of consumer reviews is not. My results did confirm the findings in Dellarocas (2004), with average consumer ratings better predicting sales than average mainstream media ratings.

Although I found that online WOM can be good predictors of music album sales, my analysis showed that traditional factors cannot be ignored. While independent label releases with extremely high blog chatter can sell even more units than major label releases, my findings estimated that the average major label release sold approximately twelve times more than the average independent label release. I also found that the higher the number of mainstream media reviews, the greater the sales.

The results of this study suggest that online word of mouth should be an important factor to be considered by record labels. Most notably, since blog chatter and Myspace friend information is available *before* an album releases and ships, record labels can examine these two variables to predict future sales well in advance of when the album is available in stores.

6. Further research

Word of mouth has generated a number of interesting topics worth further research. One

would be a closer examination of the relative effects of online WOM by genre of music. Presumably, genres of music more popular with the younger generations – those most likely to use the Internet heavily – would be more influenced by online WOM than those popular with older generations, such as classical music. Alternatively, one might find that genres such as Scandinavian death metal that are more difficult to find information on through mainstream sources may be more strongly correlated with online word of mouth variables.

Another direction would be to examine not only blog post volume on an album, but also blog sentiment – how positive or negative are the opinions expressed in the blog post. Various studies, including Das and Chen (2006) and Ghose et al. (2006), have experimented with different ways to quantify sentiment expressed in text.¹² Extracting sentiment and converting it to a numerical score would help describe the kind of word of mouth an album is getting. In addition, in my experience, most blog posts tend to be positive overall – rarely will a blogger feel the need to spend time and effort to write a scathing review. If this hypothesis holds true, then if blog sentiment is generally negative on an album and blog post volume is high, it would seem likely that sales would decrease significantly.

A final avenue of study could be on the relative effect of online word of mouth on digital sales versus physical album sales if data for digital download sales could be obtained from iTunes or other similar digital download stores. Since digital downloads enable instant gratification and – by virtue of their lower price – are more likely to be purchased impulsively, online word of mouth may have a more immediate effect on digital download sales. It seems plausible that a consumer could read a highly favorable blog post or review online, and then decide to pay the 99 cents or so

¹ Das, Sanjiv Ranjan, and Mike Chen (2006). “Yahoo! For Amazon: Sentiment Extraction from Small Talk on the Web.” Working paper, Santa Clara University. January 2006.

² Ghose, Anindya, Panagiotis Ipeirotis, and Arun Sundararajan (2006). “The Dimensions of Reputation in Electronic Markets.” Working paper, New York University. February 2006.

to hear the song for him or herself.

Bibliography

- “Apples iTunes Slams Into Top Ten Music Retail Chart.” *The Register*. November 22, 2005.
http://www.theregister.co.uk/2005/11/22/itunes_us_retail_top_ten/.
- Brynjolfsson, Erik, Yu (Jeffrey) Hu, and Michael D. Smith (2003). “Consumer Surplus in the Digital Economy: Estimating the Value of Increased Product Variety at Online Booksellers.” *Management Science*, Vol 49, No. 11, 1580-96.
- “Can Music Survive Inside the Big Box?” *Wall Street Journal*. April 27, 2007. B1.
- Chevalier, Judith and Austan Goolsbee (2003). “Measuring prices and price competition online: Amazon and Barnes and Noble.” *Quantitative Marketing and Economics*, Vol. 1, No. 2, June 2003, 203-222.
- Chevalier, Judith and Dina Mayzlin (2006). “The Effect of Word of Mouth on Sales: Online Book Reviews.” *Journal of Marketing Research*, Vol. 43, No. 3, 345-354.
- “Clap Your Hands Say Yeah: ‘This Home on Ice’ [track review].” *Pitchfork Media*. June 14, 2005.
http://www.pitchforkmedia.com/article/track_review/30565-in-this-home-on-ice.
- “Clap Your Hands Say Yeah Sidesteps Label Deals.” *Billboard*. September 20, 2005.
http://www.billboard.com/bbcom/news/article_display.jsp?vnu_content_id=1001140150.
- Das, Sanjiv Ranjan, and Mike Chen (2006). “Yahoo! For Amazon: Sentiment Extraction from Small Talk on the Web.” Working paper, Santa Clara University. January 2006.
- Dellarocas, Chrysanthos, Neveen Farag Awad, and Xiaoquan (Michael) Zhang (2004). “Exploring the Value of Online Product Ratings in Revenue Forecasting: The Case of Motion Pictures.” Working paper, University of Maryland. May 2004.
- Duan, Wenjing, Bin Gu and Andrew B. Whinston (2005). “Do Online Reviews Matter? – An Empirical Investigation of Panel Data.” Working Paper, the University of Texas at Austin.

January 2005.

Eliashberg, Jehoshua and Steven Shugan (1997). "Film Critics: Influencers or Predictors?"

Journal of Marketing, Vol. 61, No. 2, 68-78.

"To Generate Buzz, Clap Your Hands on the 'Net!" *NPR*. Nov. 29, 2005.

<http://www.npr.org/templates/story/story.php?storyId=5023133>.

Ghose, Anindya, Michael D. Smith, Rahul Telang (2006). "Internet Exchanges for Used Books:

An Empirical Analysis of Product Cannibalization and Welfare Impact." *Information*

Systems Research, Vol 17, No. 1, 3-19.

Ghose, Anindya, Panagiotis Ipeirotis, and Arun Sundararajan (2006). "The Dimensions of

Reputation in Electronic Markets." Working paper, New York University. February 2006.

Godes, David and Dina Mayzlin (2004). "Using Online Conversations to Study

Word-of-Mouth Communication." *Marketing Science*, Vol 23, No. 4, 545-560.

Hull, Geoffrey. *The Recording Industry*. Routledge: New York, 2004.

"JupiterResearch Finds That MySpace Music Community Activity Far Outpaces Online Music Sites." JupiterResearch. May 31, 2006.

http://www.jupiterresearch.com/bin/item.pl/press:press_release/2006/id=06.05.31-myspace_music_community.html.

Moul, Charles (2006). "Measuring Word of Mouth's Impact on Theatrical Movie Admissions."

Journal of Economics and Management Strategy, forthcoming.

"The Net is a Boon for Indie Labels." *The New York Times*. December 27, 2005.

<http://www.nytimes.com/2005/12/27/arts/music/27musi.html?ex=1293339600&en=7d783101017430f0&ei=5088&partner=rssnyt&emc=rss>.

"Online Survey Finds General Public, Media Conference Attendees Agree that Traditional News

Outlets Could Do a Better Job.” *PR Newswire*. February 17, 2007.

“Where Harry Most Enchants.” *The Washington Post*. April 17, 2007. B01.

“You Will Pay for Your Excessive Charm.” *Said the Gramophone*. June 9, 2005.

http://www.saidthegramophone.com/archives/you_will_pay_for_you.php.