Newspaper Financial Performance: What Content Makes a Difference?

Hari Sridhar, You Li, Esther Thorson

Mays Business School, Texas A&M University

Abstract: This study explores the relationship between newspaper content and financial performance using econometric analysis of rare monthly financial and online news content data spanning ten years. Content is measured monthly in terms of the total number of words produced and divided into three substantive dimensions: topic category (news, sports, feature, or community announcement), geographic focus (international, national, regional, r local), and story origin (wire or staff-produced). The findings underscore the differential influence of content dimensions on various sources of revenue (e.g., online vs. print advertising revenues). The results can guide newspaper management to profitably allocate scarce newsroom resources. Content and revenues exhibit a non-linear relationship and the perils of looking at a simple linear relationship are also discussed.

Keywords: Newspaper revenue, Newspaper management, Newspaper economics, and Newspaper finance

Introduction

The welfare of the newspaper industry is a concern for both media scholars and industrial practitioners (H. Cho, Martin, & Lacy, 2006; McGinley, 2009; Stoll, 2009; Wade, 2009). Newspaper financial performance has greatly worsened with the economic downturn of the last three years. Advertising revenue for the newspaper industry was down 25% from $46.6 billion in 2006 to $34.7 billion in 2008 (Newspaper Association of America, 2008). Newspaper stocks, which had already lost 42% of their value between 2005 and 2007, dropped another 83% in 2008 (Mutter, 1 January 2009). In 2010, the newspaper industry lost 6.4% in revenue (Pew Center Research, 2011). There is an urgent need to understand what impacts newspaper financial performance to avoid ad hoc approaches during this business crisis. This study responds to such a call by seeking to connect core newspaper content to newspaper financial performance.

A financial commitment approach to the study of newspaper revenues suggests that increasing newsroom investment increases financial performance (Cho, Thorson, & Lacy, 2004; Chen, Thorson and Lacy, 2005; Mantrala, Naik, Sridhar & Thorson, 2007). Investments in advertising departments also generate strong positive effects, with investments in circulation having the least positive effect

(Mantrala, Naik, Sridhar & Thorson, 2007; Tang, Sridhar, Thorson, & Mantrala, 2011). The financial commitment approach, however, is limited to a dollars- to-dollars comparison, i.e., the relationships between dollars into the newsroom, advertising, and circulation with revenue dollars. What is unknown is how the dollars invested in newsrooms should be spent, i.e., what content pays off the most? Although newspaper professionals seem to have informal theories about what kinds of news content earn them circulation, web traffic, advertiser loyalty, and, as a result, revenues, there has not yet been an attempt to quantify the relationship between amount of various kinds of newspaper content, circulation, and newspaper revenue streams. It is that gap that this paper begins to fill.

This study adopts a content attribute approach to predicting newspaper financial variance. It measures the number of words produced in terms of story topics, geographic locations of stories, and story origins (wire vs. staff-produced). Using these simple metrics of word output by categories, we seek to predict a newspaper's financial outcomes in terms of circulation, circulation revenue, and advertising revenue from print and online. The content-revenue model is empirically tested with ten years' worth monthly longitudinal data of content and financial performance for a medium-sized newspaper in the Southwest region of the United States. This work contributes to the literature in several ways. First, the measures of newspaper content and financial performance reflect the multidimensional complexity of the industry. Second, using longitudinal data across t en years, we are able to generate better insights than is the case with cross-sectional snapshots predominantly used in the extant literature. Last but not least, the presence of linear and nonlinear effects of content attributes on financial performance affirm that "uphill and downhill" effect, i.e. the notion that diminishing returns exist with respect to the production of newsroom content and managers need to know precisely whether they are over or under - producing newspaper content (Mantrala, Naik, Sridhar, & Thorson, 2007). The amount and categories of contents on different platforms, print or online, also affect financial performance in different ways.

Literature Review

The financial commitment approach

The financial commitment approach argues that increased financial investment in newsroom yields gains in revenue Scholars have operationalized financial commitment by measuring dollars invested in the newsroom to predict revenue streams (S. Cho, Thorson, & Lacy, 2004; Cyr, Lacy, & Guzman-Ortega, 2005; Mantrala, et al., 2007; Rosenstiel & Mitchell, 2004). Breaking down its process, Lacy argued that financial expenditure in newsrooms improves newspaper quality, which in turn improves audience utility, and hence circulatio n (Lacy, 1989, 1992;

Lacy & Fico, 1991), with cuts in the newsroom causing the opposite effect (Lacy & Fico, 1991).

Rosenstiel and Mitchell (2004) found that news-editorial cost per copy (that is, how many dollars were invested in newsrooms) was strongly and positively correlated with total revenue, advertising revenue, and circulation revenue. Using financial data for around 300 newspapers from 1998 to 2002, Chen, Thorson and Lacy (2005) also found a positive relationship between newsroom investment and circulation revenue per copy, advertising revenue per copy, total revenue per copy, and gross profit per copy for newspapers with less than 86,000 circulation. The strongest relationships were between newsroom investment and advertising revenue per copy with total revenue per copy.

Researchers at the University of Missouri have done extensive modeling of the financial connections of investment in newsrooms, circulation and advertising departments to advertising and circulation revenues (Mantrala, Naik, Sridhar &Thorson, 2007; Tang, Sridhar, Thorson, & Mantrala, 2011). Their findings, also based on longitudinal newspaper financial data, show that newsroom investment in dollars is the most powerful predictor of newspaper profits, i.e., the marginal impact of investment in advertising and circualiton are less than the newsroom.

However, the use of financial commitment to predict newspaper financial performance has its limitations. Bogart (2004) was critical of the financial commitment approach, asserting that investments provide the funding for newsroom operations, but the management of those funds to create high quality journalism is independent of the investment itself. Bogart also suggested that investments in the newsroom are a relatively gross index of news quality, lacking the richness needed to understand what exactly goes on in the newsroom. For example, does revenue grow due to volume increases in contents about a certain topic category (e.g., sports)? Or is it because the investments allow a newsroom to hire better writers? The financial commitment approach cannot identify the relative effectiveness of various content types on circulation or revenue, and this limits its power to guide allocating financial resource in various content departments. For that reason, this study proposes to use newspaper content as an alternative approach to predict newspaper financial performance.

Advantages of using newspaper content to predict financial revenue

Newspaper content represents the output that a newspaper organization produces contingent on its resources. A multi-dimensional measure of newspaper content and its association with newspaper financial performance will shed light strategies for more effective allocation of resources in the newsroom. The newspaper content approach stresses the importance of measuring what readers read (e.g., content category, geographic location for the stories, and source of the stories) which connects a newspaper's product to its performance. The content approach produces a more fine-grained measure of newsroom initiatives than the financial commitment approach by focusing on multiple attributes of news creation. This approach affirms that news is a "credence good" that relies on consumers' trust and beliefs to make the sale (McManus, 1992). Contents with perceived social influence will improve the paper's credibility and strengthen reader's loyalty, which will then increase revenues (Picard, 2004; Sullivan, 2006).

Scholars have measured newspaper content in terms of topic, geographic focus (e.g., local, state, national, international), type of writing (e.g., inverted pyramid vs narrative styles), and type of "voice" (e.g., straight news, opinion, analysis, columnists, citizen journalism (Stone & Boudreau, 1995). Specifically, empirical studies have often measured the amount of content (i.e., size of news hole) and the ratios of different types of content, such as number of wire service stories (Bogart, 2004; Lacy & Fico, 1991), number of staff-written stories (Lacy & Fico, 1990, 1991), amount of local news (Gladney, 1996; Maguire, 2005; Pardue, 2004; Plopper, 1991), and the amount of non-advertising contents (Lacy & Fico, 1990, 1991). The selection and presentation of the news may follow the rules of proximity, relevance, interaction, and engagement Generally, a perceived good newspaper should maintain a high ratio of original content to purchased content, high ratio of localism to nationalism or internationalism, high ratio of news interpretations and backgrounders to spot news reports, and high non- advertising content to advertising (Bogart, 1981, 2004; Gladney, 1996; Lacy & Fico, 1990, 1991). It seeks comprehensiveness of news coverage (Bogart, 2004; Gladney, 1996), and diversities in the news agenda (Culbertson, 2007) and the frames embedded in stories (Porto, 2007).

To compensate for the limitations of the financial commitment approach, scholars have turned to measures of content to predict circulation and revenue. Indeed, Lacy and Sohn (1990) showed that it was not the total amount of content space that affected circulation, but the amount of certain types of content. Circulation had the highest correlation with the amount of space devoted to local sports, local editorials, and local societal news, though the order of importance varied across

markets (Lacy & Sohn, 1990). Lacy extended this line of research and found a positive effect on circulation using a quality index that included high ratio of staff-written copy to wire service and feature service copy, total amount of non-advertising copy, and a high ratio of interpretive to spot news (Lacy & Fico, 1991). Later studies also found more positive correlations between more local focus in stories and newspaper circulation (e.g.Schoenbach, 2004).

Larger circulation is believed to lead to a higher advertising rate. Not only the circulation but also the characteristics of the readers are critical to advertising revenue (Thompson, March 1989). The correlation between circulation and advertising revenue is grounded in the nature of newspaper being a two-sided platform (Evans, 2003). The values obtained by readers as a result of consuming newspapers create positive externalities for the advertisers. The profitability of the platform, i.e., revenue, depends on the elasticity of demand on the readers' side and on both sides as a result of indirect effects from the externalities (Evans, 2003). The logic so far has been clear: certain content correlates positively with newspaper circulation; larger circulation improves advertising revenue; circulation and advertising revenue altogether contribute to the total revenue. This logic emphasizes readers' utility and gratification obtained from consuming quality content, and again signifies that good journalism is good business. Following this logic, managing newspaper content creation is essential to newspaper financial performance. But no scholarly research has empirically tested the relationship between content attributes and advertising revenue and total revenue.

Given the limitations of the financial commitment approach and the strengths of the content approach, this study uses content attributes to predict revenue streams across circulation and advertising on both print and online platforms. It asks:

RQ1: What content attributes influence newspaper circulation?

RQ2: What content attributes influence online and print advertising revenue respectively?

RQ3: What content attributes influence print revenue?

RQ4: What content attributes influence total revenue?

Non-linear effects of contents on financial performance

Though financial commitment to newsrooms improves news quality it is not necessarily true that the relationship is linear. Some scholars propose that quality exhibits diminishing returns on firm performance, i.e., every incremental effort in news quality has a lower marginal impact on financial return because the costs of every marginal increase in quality also increase simultaneously (Kim & Meyer, 2005; Meyer, 2004; Picard, 2004). A recent study found that once the newsroom reaches an optimal level of spending for quality, any additional investment will

only yield a declining rate of return on investment (Mantrala, et al., 2007). This study therefore posits that content attributes are likely to demonstrate a nonlinear effect on newspaper revenue, that is, diminishing returns to increases in content volume.

Method

Data Setting and Measures

A privately-owned media company provided both financial data and the online content archive for one of its newspapers. The company wishes to remain anonymous. The newspaper under study is medium-sized, located in a Southwestern U.S. city, and produces both online and print versions of its newspaper. Its content and financial data were aggregated monthly, spanning a ten-year time frame from January, 1999, through December 2008 (120 months).

Financial performance measures

We used the company-supplied indicators of financial performance, i.e., online media revenue, circulation, total advertising revenue, print advertising revenue and total print revenue as the focal dependent variables.[i]

Newspaper content measures

Content was coded in terms of story topic (sports, news, feature, and community announcement), story geographic location (international, national, regional, or local), and story origin (wire or staff written). The measure of story topic relies on the newspaper's online archival system that assigned a unique theme tag to each story topic occurring for the ten-year period. The tag appeared in the story's web URL, which indicated the main topic of the story. A total of 64 story theme tags were detected. Tags of similar themes were then aggregated into four major story topic categories: news (17 tags), sports (24 tags), features (15 tags), and community announcements (8 tags)ii. To measure the geographic focus of the story, the city name that appeared before the main text of the story indicated where the story occurred. If the story was labeled as from cities outside U.S., it had an international focus. If the story was from major cities in states other than the newspaper's home state, it was coded as national focus. If the story was from cities in any of the three neighboring states of the newspaper, then it was coded as regional focus. If the story had no clear geographic identification, the newsroom manager suggested it most likely had a local focus. To measure the origin of the story, stories with a staff byline or newsroom email address were labeled as staff-produced, and stories with a wire service identity were labeled as wire-produced.

We next sampled time periods every sixth day starting on January 1, 1999. On average, four to five sample days' content composed a month of content. Eight constructed weeks of content formed each one-year content sample. Altogether, over 600 days were sampled for 120 months ranging from January 1999 to December 2008. This sampling strategy is consistent with previous literature, which suggests that constructed week sampling is more efficient than simple random sampling or consecutive day sampling (Lacy, Riffe, Stoddard, Martin, & Chang, 2001; Riffe, Aust, & Lacy, 1993).

Next, a computer program was developed, based on the above content measurement and sampling techniques, to count the number of words for each sampled story and then the number of stories in four topic categories (news, sports, features, or community announcements), in four geographic categories (international, national, regional, or local), and with two origins (wire or staff). These served as the independent variables of the study.iii The computer-aided content analysis approach is gaining more popularity because of its ability to produce efficient and reliable results (Dowling & Kabanoff, 1996).

Descriptive Statistics

Figures 1-1 to 1-6 show the overtime plots of all the content variables. The total number of words, regardless of story topic, decreased over time, particularly since 2005. Also, the relative amounts of the four kinds of content have changed. Sports and community content stayed in about the same proportion over the ten years, but the proportion of news shrank and the proportion of feature words increased. Local content, which already accounted for over 70% of the total content amount, increased to around 80% after 2005 (see Figures 1-3 and 1-4). Meanwhile, the percentage of staff-produced content decreased slightly from 2005. These changes will prove important to our understanding of the relationships between content and financial performance.

Table 1-1 displays the descriptive statistics of the content variables (IVs) and the financial variables (DVs). During the ten-year period, the average monthly number of news words was highest (150,974 words / month), followed by sports content (106,348 words / month), features content (106,181 words / month), and community announcement content (73,719 words / month). The average monthly words of local content (337,723 words / month) well surpassed those of regional, national, and international contents. The average of monthly staff-produced words (357,257 words / month) is greater than wire-produced words. Finally, the average monthly print advertising revenue was over $1,720,000, contributing about 80% to the average monthly total print revenue. The average monthly online revenue was around $150,000, adding less than 10% to the average monthly total advertising revenue.

Simple Linear Model

We began with simple linear models of the relationship between the content variable and financial performance. Tables 3-1 (story topic), 3-2 (geographic focus), and 3-3 (origin) show the results of simple linear regressions between the content variables on the financial variables. When we entered the content variables one at a time to predict each financial indicator without controlling the time trend effect, the results showed clearly that news and sports words had significantly positive effects on daily circulation and print advertising revenue, but a significantly negative effect on online revenue (See Table 3-1). Contents of all geographic foci positively predicted circulation and print advertising revenue, but negatively predicted online revenue (See Table 3-2). Feature and community announcement contents failed to predict any revenue stream. This pattern is consistent with previous findings in the correlation between certain types of newspaper content and circulation and advertising (e.g., Lacy & Sohn, 1990). Using circulation revenue as the dependent variable did not yield any significant prediction. Compared to circulation size, circulation revenue is a less stable metric, because the newspaper's copy price fluctuated across time and even across customers. Hence, we drop circulation revenue from further analysis, and instead add advertising revenue and total print revenue as additional financial indicators.

Perils of Using a Simple Linear Model

Figures 2-1 to 2-4 provide scatterplots between an arbitrarily chosen set of independent variables and financial performance. It would seem that there is no need to be concerned about curvilinear effects, since the scatter plots show dominant linear relations. Yet when we controlled for time effects through a trend variable in the simple linear regression model, the significant effects described above vanished. Specifically, we ran the simple regression again, but with time trend added as a control variable to each of the twelve regressions. These results showed that none of the content variables entered linearly explained any more significant proportion of the financial variance over and above time.

Thus, we suspected that the "linear" relationship shown between the content variables and financial performance in the scatter plots (Figures 2-1 to 2-4) was heavily confounded by the time trend. To affirm this, we parsed out time effects from each of the dependent variables and plotted the dependent variable residuals against the independent variables. These plots, shown in Figures 3-1 to 3-4, clearly show that the independent variables, after controlling for overtime effects, shared a curvilinear relationship with the financial performance variables.

In sum, it was dangerous to use a simple linear model in our situation, not only because content would be wrongly diagnosed as not influencing financial performance (after controlling for time) when modeled as a linear effect, but also because the explained variance due to content, when modeled to linearly affect revenue, is very low (see R-squared values in Tables 3-1 to 3-3). A common approach to address the curvilinear issue in regression is to use a log-log transformation, which we present next.

Log-Log Model

In econometric modeling, the common solutions to nonlinearity are to add polynomials to the independent variable or to use logarithmic transformation of variables (Murray, 2006). However, the coefficients are difficult to interpret in a polynomial model, whereas the coefficients in logarithmic transforms are easier to interpret in percentage terms. We applied the natural logarithm of both dependent and independent variables to the regression equation:

$$\text{Ln } (Yi) = \beta_0 + \beta_1 \text{ ln } (X_i) + \varepsilon_i$$

After the transformation, the linear regression function can then explain the nonlinear relationship in variables. The coefficient, β_1 has the interpretation of an elasticity, meaning a 1% change in X is associated with a β_1 % change in Y[iv]. When $\beta_1 < 0$, X has a negative effect on Y, at the current level of X and Y. When $\beta_1 > 0$, X has a positive effect on Y, at the current level of X and Y.

Multiple regression analysis is appropriate for this research, because it meets two objectives: 1) it assesses which set of content attributes are most predictive of newspaper financial variance by computing the total amount of variance in the dependent variable that can be explained by a combination of the independent variables; and 2) it calculates the magnitude of effect of each content attribute on the financial outcome (Tabachnick & Fidell, 2007). In our analysis, we seek to estimate twelve models, i.e., we have three substantive classifications of news content (content type, geographic focus and story origin) that serve as independent variables explaining five types of revenue (online media revenue, circulation, total advertising revenue, print advertising revenue and total print revenue).

We performed the log-log regression in two steps. First, only time trend was entered in the equation as the predictor to determine the association between trend and financial performance. This step tests whether the time trend, by itself, had a large effect on the five financial performance indicators. Second, after controlling the effect of time trend, the natural logarithm of each set of the content variables (i.e., story topic, geographic focus, and story origin) was regressed on the natural

logarithm of the five financial variables respectively. This step produced five regression functions for each of the three sets of content variables, i.e., with one regression each with the dependent variable being online media revenue, circulation, total advertising revenue, print advertising revenue and total print revenue respectively. Appendix A presents the full model specification for each regression.

Log-Log Regression results: Time-trend only

When time trend was entered alone, it significantly predicted all financial variables, explaining from 26% to 96% of the variance (See Model 1s in Tables 4-1, 4-2, and 4-3). In particular, the time trend predicted circulation, advertising revenue and total revenue of the print platform negatively, but it predicted online revenue positively overtime. These statistical results are consistent with the patterns in the industry.

Full Log-Log Regression results: Story Topic (Table 4-1)

Five log-log regressions were performed with the natural log of monthly words of sports, news, features, and community announcements and time as the independent variables and the natural log of each of the five financial indicators as the dependent variables (see Model 2's in Table 4-1). Overall, the full models significantly improved the prediction of all the financial variables over the models with time trend alone.

Any further increases in news words had a significantly negative effect on circulation (β_1=-.375, p<. 01), on advertising revenue (β_1=-.407, p<. 05), on print advertising revenue (β_1=-.383, p<. 01), and on total print revenue (β_1=-.372, p<. 01). Sports words exhibited a significantly positive yet diminishing rate of marginal effect on circulation (β_2=. 133, p<.1), on advertising revenue (β_2=.202, p<.1), on print advertising revenue (β_2=.196, p<.05), and on total print revenue (β_2=.207, p<. 05). Community words only significantly predicted online revenue (β_3=. 079, p< .01). Feature words had a significantly positive but diminishing effect on online media revenue (β_4=.052, p<. 05), on circulation (β_4=.111, p<.1), on advertising revenue (β_4=.177, p<.1), on print advertising revenue (β_4=.164, p<. 05) and on total print revenue (β_4=.152, p<. 05). Overall, a 1% increase in sports and feature contents positively contributed up to a .2% increase to revenue streams of both online and print. By contrast, a 1% increase in news content negatively influenced all revenue streams except for online revenues by about .4%.

Full Log-Log Regression results: Geographic focus (Table 4-2)

The entire group of variables significantly predicted all financial variables, and explained up to 96% of the variance. Local words positively predicted all revenue streams, whereas regional words negatively predicted all revenue streams. In particular, local words significantly contributed to online revenue (β_1=. 124, p< .01). Regional words exhibited a negative yet diminishing rate of effect on circulation (β_2=-.448, p< .01), advertising revenue (β_2=-.442, p< .05), print advertising revenue (β_2=-.373, p< .05), and total print revenue (β_2=-.395, p<.05). In other words, for every 1% increase in regional words, the newsroom would drop about .4% in revenue. National and international words, however, did not predict the revenue streams, except on one occasion where the international words positively predicted circulation (β_4=. 171, p< .05).

Full Log-Log Regression results: Story origins (Table 4-3)

Once again, the entire group of variables significantly improved the prediction compared with using the trend time variable alone. The full model explained up to 96% of the variance in revenue streams. Consistent with previous studies, staff-written content significantly predicted all revenue streams at a significance level of .1. Though the marginal effect was diminishing, the amount of staff-written content significantly contributed to online media revenue (β_2=. 096, p< .01), circulation (β_2=. 179, p<. 01), advertising revenue (β_2=. 176, p< .1), print advertising revenue (β_2=. 159, p<. 05), and total print revenue (β_3=. 167, p<. 05). Wire-produced words, by contrast, did not significantly predict revenue streams, except for its negative impact on circulation (β_1=-. 309, p<. 01).

Discussion & Implications

This study attempted to better understand the relationship between newspaper content and financial performance in order to offer insights about the allocation of financial resources in newsroom. Not surprisingly, time trend had a large impact on financial outcomes, and the relations between the combinations of independent variables and financial variables were not linear when the effect of time trend is present.

Given the perils of wrong inference due to the simple linear model, we used a log-log linear regression and controlled for the effect of time trend. This approach yielded better prediction, as it generally explained a large variance in financial outcomes. The first objective of the research question explored what content attributes predicted circulation. Overall, content of various attributes predicted circulation. After controlling out the effects of time trends, the combination of

the contents of the story topics, geographic attributes and origins explained an additional 3% variance in daily circulation. Thus, using content attributes predicted circulation variance slightly better than using the time trend variable alone. Specifically, the amount of news content, wire-produced content and content that with a regional geographic focus had significantly negative yet diminishing marginal effects on daily circulation.v The amount of sports, feature, staff-written, and contents with a local or international geographic focus significantly predicted circulation in a positive way.

Given a newspaper's two-sided nature, circulation's influence should carry through and affect advertising revenue. When controlling for time trends, four content topics predicted up to an additional 3% of variance in print advertising revenue, and contents of various geographic attributes predicted an additional of 2% variance in total advertising revenue. In particular, sports, features, staff-written contents and content with a local geographic focus positively predicted total advertising revenue and total print advertising revenue. News content, however, negatively predicted total advertising revenue and total print advertising revenue.

Consistent with the industrial dilemma, the empirical evidence showed that online media revenue grew steadily as time went by, even though most content attributes did not contribute to the online revenue in a significant way. Only feature content, community service content, staff-written content, and content with a local geographic focus significantly contributed to online revenue. It is not clear why categories like sports words, news words, and features did not predict online advertising revenues, it seems likely to be because of the way online is often sold—somewhat as an "add-on" to the print product. Thus advertisers are not specifically looking at the news product when they decide whether to buy online. The lack of connection between traditional measures of news quality, like how much news, sports and features there is, and online advertising revenues remains for future investigation.

Sports, features, staff-written and local contents were significant positive predictors of total print revenue. But any additional news content or content with a regional geographic focus led to a negative impact on total print revenue. One interpretation of this finding is more hard news content is not adding to the financial performance of this newspaper. It may be, however, that the conglomeration of all categories of hard news into one category masks more subtle but important impacts of news in various domains like crime, education, politics and so on. The negative impact of the news category may also be an anomaly of the newspaper for which the analyses were executed. The content that with a regional geographic focus may lack a clear target market and be less relevant to local readers.

This study provides some initial insights for management. First, it affirms that the content attribute approach provides superior predictions of financial performance. It empirically verified what literature defined as quality content attributes, such as sports, feature, local, and staff-written contents, were positive predictors of financial performance. But, the existence of null effects and negative effects in the relationship between content variables and performance should be interpreted with caution. This does not suggest that these content types do not increase performance, but merely suggests that their effects relative to other kinds of news content produced by the newspaper are low. Newspapers may be better off financially by becoming more niche content providers, focusing on certain areas of content and strengthening performance.

Second, the print and online platforms should differentiate in their allocations of types and amounts of contents. Some types of contents generate revenue in print form, but not in the digital form. For example, the amount of sports content has a positive effect on print revenues, but not much effect on online media revenue. The amount of community service content was a significantly positive predictor of online revenue, but was not a significant predictor of print revenue streams. The amount of news content did not have much impact on online revenue, but negatively predicted all other print revenue streams. Though it is beyond the scope of the study to explain why the content-revenue models are different for online and print platforms, the natures of the two media and readers' uses and gratification may contribute to the differences (Chyi & Lasorsa, 1999). This suggests that newsrooms should supply different information or packages of information to different forms.

Third, the effects of content on the financial outcomes are not always linear. Some content variables demonstrate significant quadratic effects (See Figure 7-14 for specific variables and effects). This means that the content amount of a certain attribute exhibits a bell curve influence on the financial outputs. The creation of certain contents generates revenue up to a turning point and then works in the opposite direction. This observation empirically verifies the theoretical assumptions in previous studies that have proposed a nonlinear relationship between newspaper quality and financial performance (Kim & Meyer, 2005; Meyer, 2004; Picard, 2004). Therefore, the claim that managers should maintain newspaper qualit y at a level that meets but does not exceed expectations is still reasonable (Picard, 2004).

References

Bogart, L. (1981). Press and public : *Who reads what, when, where, and why in American newspapers.* Hillsdale, NJ: L. Erlbaum Associates.

Bogart, L. (2004). Reflections on Content Quality in Newspapers. *Newspaper Research Journal,* 25(1), 40-53.

Chen, R., Thorson, E., & Lacy, S. (2005). The impact of newsroom investment on newspaper revenues and profits: Small and medium newspapers, 1998-2002. *Journalism & Mass Communication Quarterly,* 82(3), 516-532.

Cho, H., Martin, H. J., & Lacy, S. (2006). An industry in transition: Entry and exit in daily newspaper markets, 1987-2003. *Journalism & Mass Communication Quarterly,* 83(2), 381-396.

Cho, S., Thorson, E., & Lacy, S. (2004). Increased circulation follows investments in newsrooms. *Newspaper Research Journal,* 25(4), 26-39.

Chyi, H. I., & Lasorsa, D. (1999). Access, use and preferences for online newspapers. *Newspaper Research Journal,* 20(4), 2-14

Culbertson, H. (2007). Agenda diversity indicates newspaper quality. *Newspaper Research Journal,* 28(1), 40-54.

Cyr, C. S., Lacy, S., & Guzman-Ortega, S. (2005). Circulation increases follow investments in newsrooms. *Newspaper Research Journal,* 26(4), 50-60.

Dowling, G. R., & Kabanoff, B. (1996). Computer-aided content analysis: What do 240 advertising slogans have in common? *Marketing Letters,* 7(1), 63-75.

Evans, D. S. (2003). Some empirical aspects of multi-sided platform industries. *Review of Network Economics,* 2(3), 191-209.

Gladney, G. A. (1996). How editors and readers rank and rate the importance of eighteen traditional standards of newspaper excellence. *Journalism & Mass Communication Quarterly,* 73(2), 319-331.

Hester, J. B., & Dougail, E. (2007). The efficiency of constructed week sampling for content analysis of online news. *Journalism & Mass Communication Quarterly,* 84(4), 811-824.

Kim, K.-H., & Meyer, P. (2005). Survey yields five factors of newspaper quality. *Newspaper Research Journal,* 26(1), 6-15.

Lacy, S. (1989). A model of demand for news: Impact of competition on newspaper content. *Journalism Quarterly,* 66(1), 40-128

Lacy, S. (1992). The financial commitment approach to news media competition. *Journal of Media Economics,* 5(2), 5-21.

Lacy, S., & Fico, F. (1990). Newspaper quality & ownership: Rating the groups. *Newspaper Research Journal,* 11(2), 42-56.

Lacy, S., & Fico, F. (1991). The link between newspaper content quality & circulation. *Newspaper Research Journal,* 12(2), 46-57.

Lacy, S., Riffe, D., Stoddard, S., Martin, H., & Chang, K.-K. (2001). Sample size for newspaper content analysis in multi-year studies. *Journalism & Mass Communication Quarterly,* 78(836-845).

Maguire, M. (2005). Changes in ownership affect quality of oshkosh paper. *Newspaper Research Journal,* 26(4), 74-86.

Mantrala, M. K., Naik, P. A., Sridhar, S., & Thorson, E. (2007). Uphill or downhill? Locating the firm on a profit function. *Journal of Marketing,* April, 26-44.

McGinley, M. (2009). All together now (II). *Columbia Journalism Review,* 48(3), 11-11.

McManus, J. H. (1992). What kind of commodity is news? *Communication Research,* 19(6), 787-805.

Meyer, P. (2004). The Influence model and newspaper business. *Newspaper Research Journal,* 25(1), 66-83.

Murray, M. P. (2006). *Econometrics : A modern introduction.* Boston: Pearson Addison Wesley.

Mutter, A. (1 January 2009). Newspaper share value fell $ 64B in 2008. Reflections of a Newsosaur, from http://newsosaur.blogspot.com/2008/12/newspaper-share-value-fell-64b-in-08.html

Pardue, M. J. (2004). Quality key to highest city aone penetration in U.S. *Newspaper Research Journal,* 25(4), 13-25.

Picard, R. G. (2004). Commercialism and newspaper quality. *Newspaper Research Journal,* 25(1), 54-65.

Plopper, B. L. (1991). Gannett and the Gazette. *Newspaper Research Journal,* 12(2), 58-71.

Porto, M. P. (2007). Frame diversity and citizen competence: Towards a critical approach to news quality. *Critical Studies in Media Communication,* 24(4), 303-321.

Pew Research Center Publications (March 14, 2011). State of the news media 2011. Retrieved July 29, 2011, from http://pewresearch.org/pubs/1924/state-of-thenews- media-2011.

Riffe, D., Aust, C., & Lacy, S. (1993). Random, consecutive day and constructed week samples in newspaper content analysis. *Journalism Quarterly,* 70, 133-139.

Rosenstiel, T., & Mitchell, A. (2004). The impact of investing In newsroom resources. *Newspaper Research Journal,* 25(1), 84-97.

Schoenbach, K. (2004). A balance between imitation and contrast: What makes newspaper successful? A summary of internationally comparative research. *Journal of Media Economics,* 17(3), 219-227.

Stoll, M. (2009). No profit, no problem. *Columbia Journalism Review,* 47(6), 37-38.

Sullivan, D. (2006). Business service model promotes better newspapers. *Newspaper Research Journal,* 27(4), 66-78.

Stone, G. C., & Boudreau, T. (1995). Comparison of reader content preferences. *Newspaper Research Journal,* 16(4), 13-28.

Sridhar, Shrihari, Mantrala, Murali, Naik, Prasad A., & Thorson, Esther (In press). Dynamic marketing budgeting for platform firsm: Theory, evidence and Application. *Journal of Marketing Research.*

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston: Pearson/Allyn & Bacon.

Thompson, R. S. (March 1989). Circulation versus advertiser appeal in the newspaper industry: An empirical investigation. *The Journal of Industrial Economics,* XXXVII(3), 259-271.

Tang, Yihui (Elina), Sridhar, Shrihari, Thorson, Esther, & Mantrala, Murali (In press). The bricks that build clicks: Newsroom investment and online news performance. *International Journal of Media Management.*

Wade, J. (2009). The end of an era. *Risk Management,*56 (3), 16-17.

Zaller, J. (2003). A new standard of news quality: Burglar alarms for the monitorial citizen. *Political Communication,* 20(2), 109.

Figure 1-1: Annual distribution of total number of words by topic category (in raw number): January 1999-December 2008

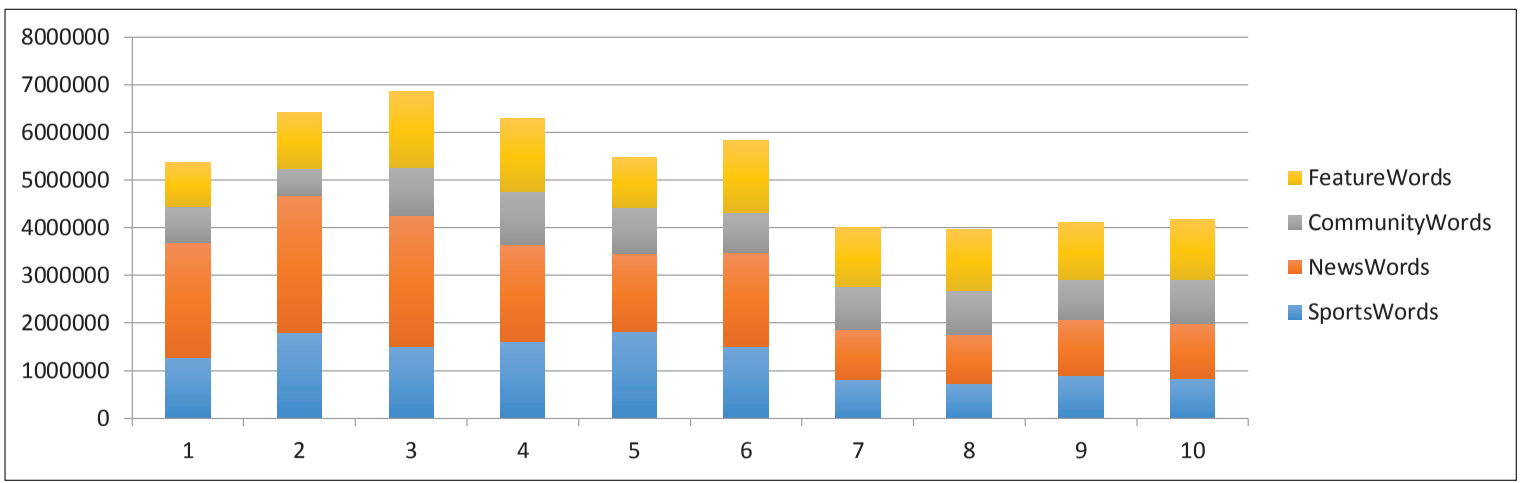

Figure 1-2: Annual distribution of total number of words by topic category (in percentage): January 1999-December 2008

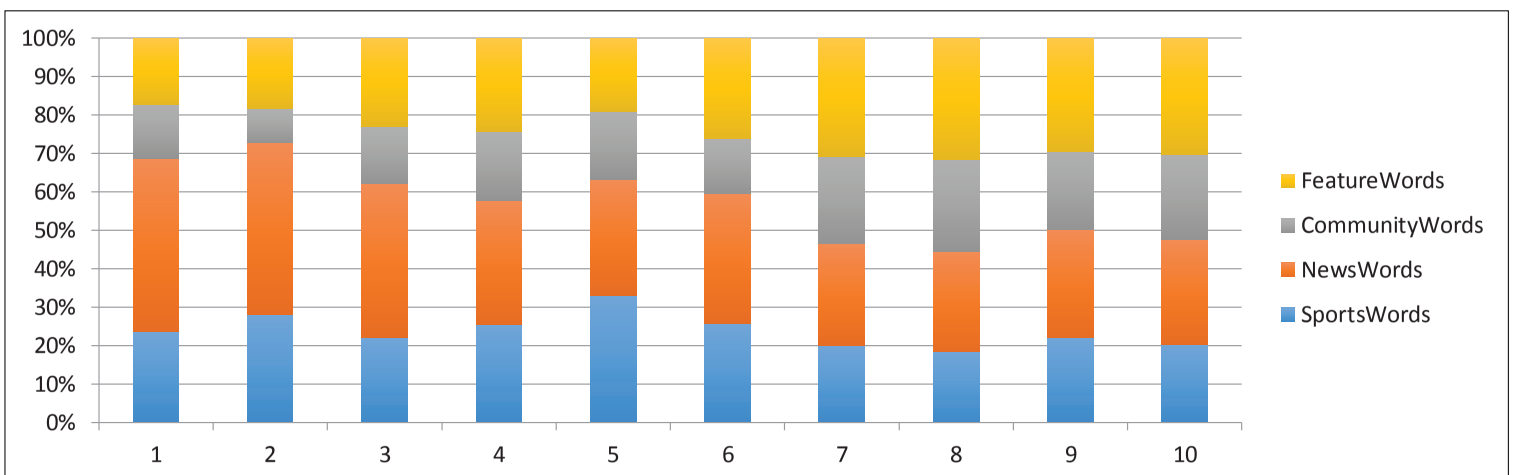

Figure 1-3: Annual distribution of total number of words by geographic focus (in raw number): January 1999-December 2008

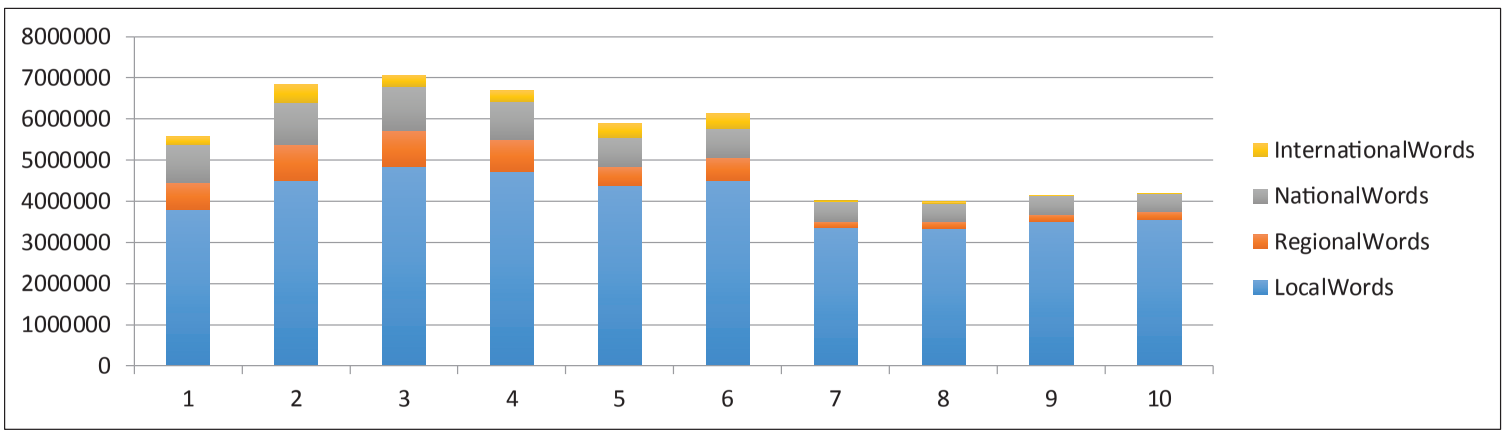

Figure 1-4: Annual distribution of total number of words by geographic focus (in percentage): January 1999-December 2008

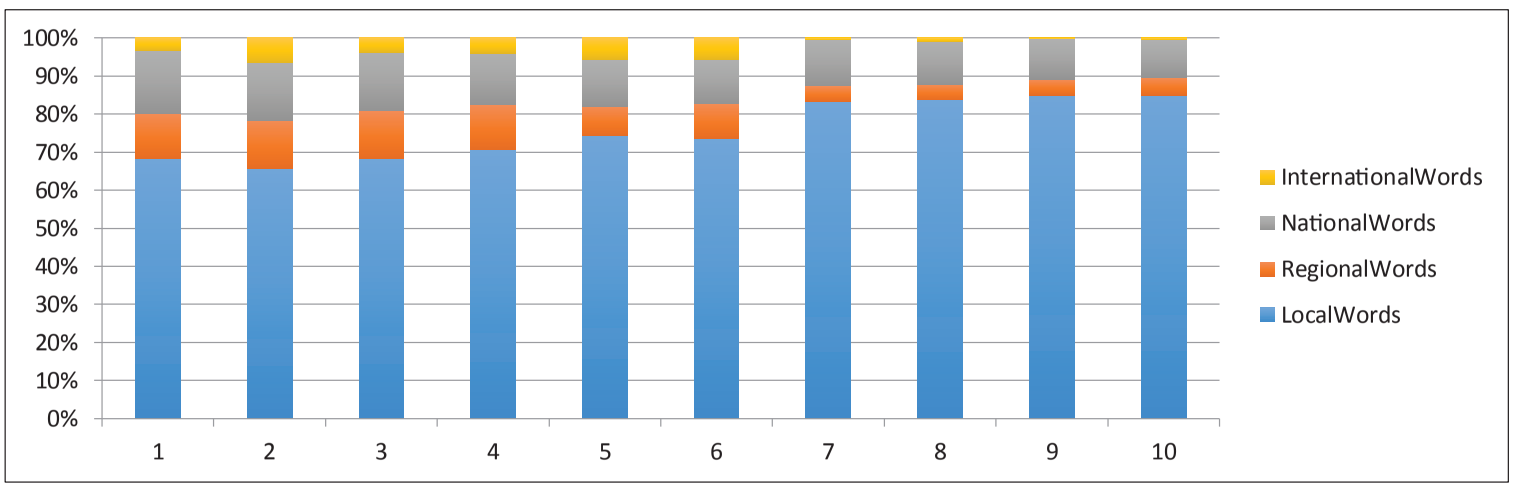

Figure 1-5: Annual distribution of total number of words by story origin (in raw number): January 1999-December 2008

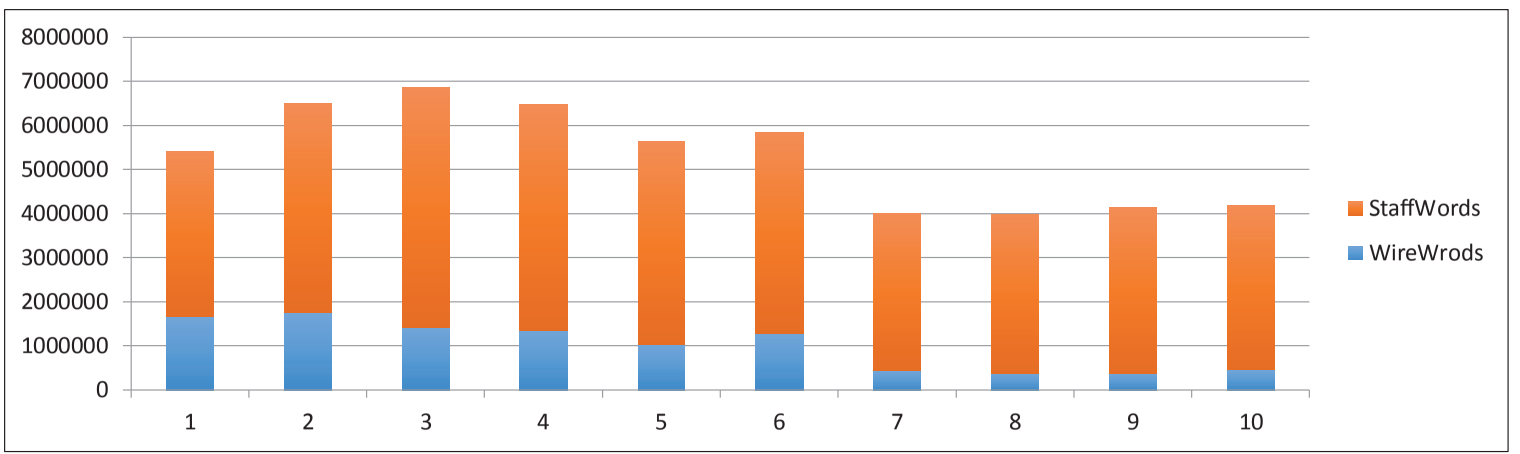

Figure1-6: Annual distribution of total number of words by story origin (in percentage): January 1999-December 2008

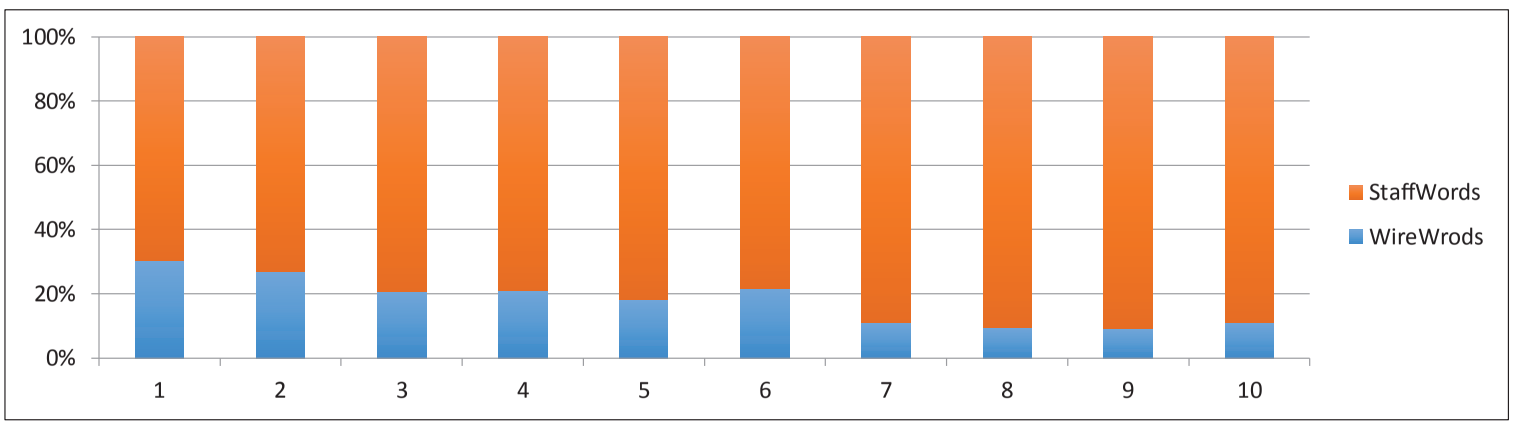

Figure 2-1 Scatterplot of sports word amount against circulation

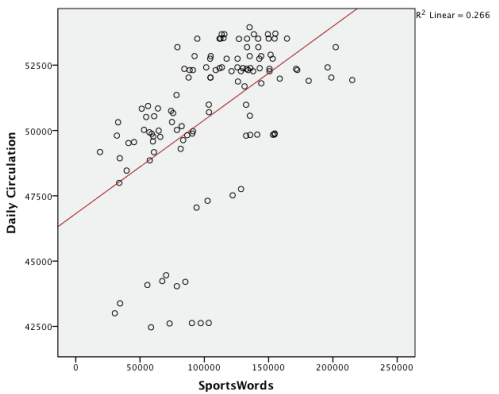

Figure 2-2. Scatterplot of news words amount against circulation

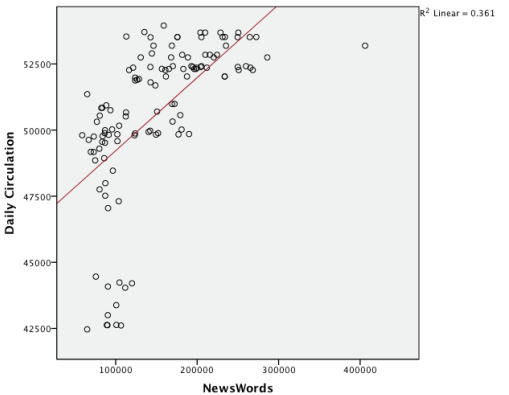

Figure 2-3. Scatterplot of sports word amount against print ad revenue

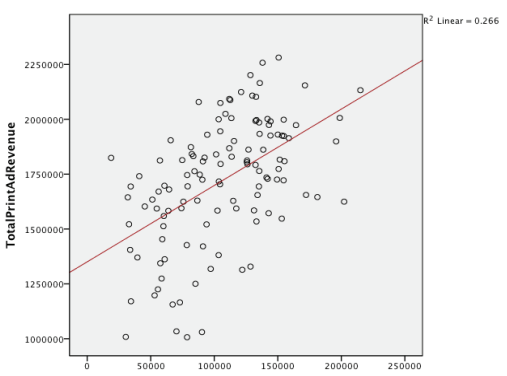

Figure 2-4. Scatterplot of news word amount against print ad revenue

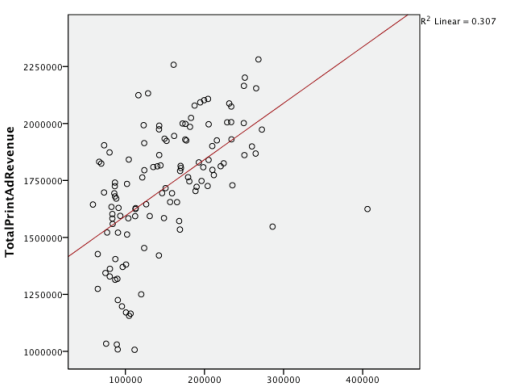

Figure 3-1: Residual scatterplot of themes predicting circulation after controlling for trend variable.

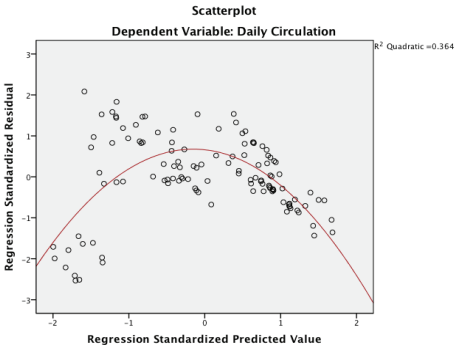

Figure 3-2: Residual scatterplot of geographic focuses predicting circulation after controlling for trend variable

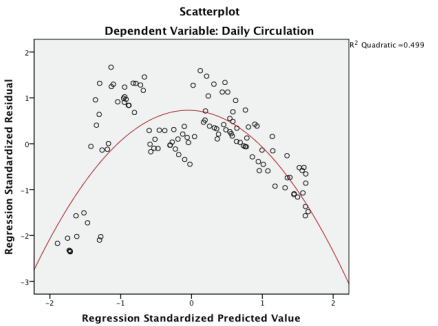

Figure 3-3: Residual scatterplot of themes predicting print advertising revenue after controlling for trend variable

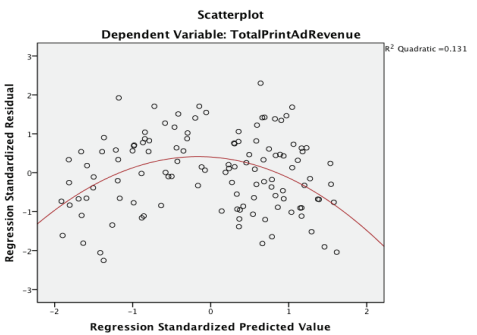

Figure 3-4: Residual scatterplot of geographic focus predicting print advertising revenue after controlling for trend variable

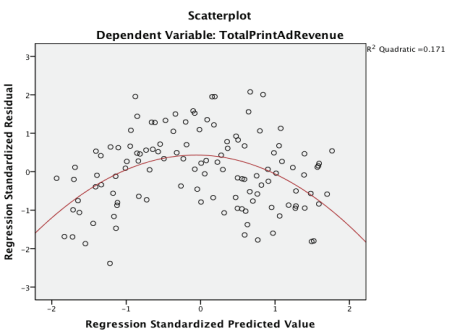

Table 1-1: Descriptive of Content and Financial variables (original)

	N	Minimum	Maximum	Mean	Std. Deviation	Variance
Sports	120	18846	215008	106348.44	42020.466	1.766E9
News	120	58952	406079	150973.96	63556.408	4.039E9
Community	120	10150	117041	73718.83	18132.264	3.288E8
Feature	120	56212	210656	106180.86	25988.672	6.754E8
Local	120	225602.00	593863.00	337722.5750	66579.39598	4.433E9
Regional	120	4747.00	131110.00	40725.8167	26145.38179	6.836E8
National	120	19364.00	164518.00	59635.2333	24158.75799	5.836E8
International	120	.00	196790.00	16019.4917	22807.43410	5.202E8
Wire	120	12979.00	314028.00	84532.7667	51939.83901	2.698E9
Staff	120	223743.00	586944.00	357257.0333	72879.77600	5.311E9
Online media revenue	120	18028	357389	150457.78	107482.431	1.155E10
Advertising revenue	120	1294957	2327045	1871043.30	211247.877	4.463E10
Daily circulation	120	42465	53951	50628.38	2920.366	8528539.512
Circulation revenue	120	399512	500398	458546.56	18026.033	3.249E8
Print advertising revenue	120	1006786	2281262	1720585.52	283534.240	8.039E10
Total print revenue	120	1458444	2734740	2179132.07	284819.241	8.112E10
Valid N (listwise)	120					

Note: The independent variables were entered using the absolute amount of words. The dependent variables were entered using the absolute dollars amount.

Table 1-2: Descriptive of Content and Financial variables (Natural Log Transformation)

	N	Minimum	Maximum	Mean	Std. Deviation	Variance
LNSports	120	9.84	12.28	11.4794	.46974	.221
LNNews	120	10.98	12.91	11.8380	.42132	.178
LNCommunity	120	9.23	11.67	11.1682	.31582	.100
LNFeature	120	10.94	12.26	11.5439	.24241	.059
LNLocal	120	12.33	13.29	12.7111	.19465	.038
LNRegional	120	8.47	11.78	10.3600	.77402	.599
LNNational	120	9.87	12.01	10.9123	.41984	.176
LNInternational	120	.00	12.19	7.7512	3.43603	11.806
LNStaff	120	12.32	13.28	12.7658	.20276	.041
LNWire	120	9.47	12.66	11.1297	.69787	.487
LNDaily Circulation	120	10.66	10.90	10.8305	.06040	.004
LNAdvertising Revenue	120	14.07	14.66	14.4354	.11750	.014
LNOnlineRevenue	120	9.80	12.79	11.5866	.90155	.813
LNPrintAdvertising Revenue	120	13.82	14.64	14.3432	.17874	.032
LNCirculation Revenue	120	12.90	13.12	13.035	.03947	.002
LNPrintRevenue	120	14.19	14.82	14.5853	.13821	.019
Valid N (listwise)	120					

Note: The independent variables were entered using the natural log of the original amount of words. The dependent variables were entered using the natural log of the absolute dollars amounts.

Table 2: Correlation Matrix of all variables (N=120)

Variables	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1 Sports	1														
2 News	.66**	1													
3 Community Service	0	-.12	1												
4 Feature	.11	.2*	.34**	1											
5 Local	.79**	.73**	.29**	.56**	1										
6 Regional	.68**	.92**	-.02	.26**	.68**	1									
7 National	.72**	.9**	.04	.25**	.72**	.87**	1								
8 International	.51**	.48**	-.05	.14	.41**	.51**	.45**	1							
9 Staff-written	.75**	.68**	.35**	.59**	.96**	.66**	.72**	.41**	1						
10 Wire-written	.68**	.91**	-.11	.15	.65**	.91**	.85**	.5**	.53**	1					
11 Circulation Revenue	.04	-.01	.04	-.16	.01	-.05	0	-.120	.03	-.06	1				
12Advertising Revenue	.38**	.35**	-.11	.06	.3**	.37**	.4**	.19*	.3**	.35**	.06	1			
13 Daily Circulation	.52**	.6**	-.03	-.01	.41**	.64**	.66**	.37**	.42**	.59**	.02	.58**	1		
14 Online Revenue	-.61**	-.77**	.11	.01	-.49**	-.8**	-.79**	-.45**	-.47**	-.78**	.02	-.53**	-.82**	1	
15 Print Advertising Revenue	.52**	.55**	-.12	.04	.4**	.58**	.6**	.32**	.4**	.56**	.04	.95**	.74**	-.78**	1
16 Total Print Revenue	.52**	.55**	-.12	.03	.4**	.57**	.6**	.31**	.4**	.55**	.10	.95**	.74**	-.77**	1**

****p<.01, *p<.05**

Table 3-1 Simple Linear Regression Story topic of contents

	Circulation Revenue				Daily Circulation				Online Revenue				Print Advertising Revenue			
Model / IVs	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Sports	.041				.516**				-.611**				.516**			
News		-.01				.601**				-.767**				.554**		
Community			.036				-.033				.107				-.124	
Feature				-.159				-.009				.007				.041
Adjusted R^2	-.007	-.008	-.007	.017	.26	.356	-.007	-.008	.367	.585	.003	-.008	.260	.301	.007	-.007
F value	.199	.012	.157	3.08	42.8**	66.8**	.131	.009	70.12**	168.54**	1.378	-.006	42.772**	52.33**	1.847	.199

**p<.01, *p<. 05

Note: Each content variable (IV) was entered using its absolute value one at a time to predict revenue

Table 3-2 Simple Linear Regression: Geographic location of contents

	Circulation Revenue				Daily Circulation				Online Revenue				Print Advertising Revenue			
Model / IVs	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Local	.009				.412**				-.486**				.404**			
Regional		-.046				.636**				-.799**				.575**		
National			-.001				.657**				-.792**				.600**	
International				-.120				.367**				-.451**				.315**
Adjusted R^2	-.008	-.006	-.008	.006	.163	.399	.426	.128	.230	.635	.625	.197	.156	.325	.355	.092
F value	.010	.245	.000	1.717	24.097**	80.153**	89.49**	18.399**	36.489**	207.626**	198.972**	30.180**	23.046**	58.292**	66.399**	13.032**

**p<.01, *p<. 05

Note: Each content variable (IV) was entered using its absolute value one at a time to predict revenue

Table 3-3 Simple Linear Regression: Origin of contents

	Circulation Revenue		Daily Circulation		Online Revenue		Print Advertising Revenue	
Model / IVs	1	2	1	2	1	2	1	2
Staff	.025		.424**		-.468**		.399**	
Wire		-.057		.593**		-.778**		.556**
Adjusted R^2	-.008	-.005	.173	.346	.213	.602	.152	.303
F value	.075	.379	25.832**	64.091**	33.159**	180.812**	22.295**	52.678**

**p<. 01, *p< .05.

Note: Each content variable (IV) was entered using its absolute value one at a time to predict revenue

Table 4-1: Summary of Log-Log Linear Regression for Content Variables Predicting Newspaper Revenue: Story Topic

	Online Media Revenue		Circulation		Advertising Revenue		Print Advertising Revenue		Total Print Revenue	
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
Variables	β	β	β	β	β	β	β	β	β	β
Month	.976***	.952***	-.819***	-.906***	-.511***	-.632***	-.738***	-.801***	-.738***	-.799***
Local		.124***		.151*		.230*		.193*		.200**
Regional		-.081*		-.464***		-.442**		-.373**		-.395**
National		-.048		.142		.095		.105		.119
International		.041		.171**		.092		.108		.116
Adjusted R^2	.95	.96	.67	.70	.26	.27	.54	.56	.54	.56
F value	2404.97***	549.11***	241.05***	55.34***	41.79**	9.81**	141.47***	30.87***	141.03**	31.23***
***F* for change in R^2**	4.94***		3.60***		1.61		2.01*		2.27*	

*p<.1, **p < .05. ***p < .01.

The independent variables were entered at their natural logs of the monthly amount of words. The dependent variables were entered at their natural logs of the monthly dollars amount.

Table 4-2: Summary of Log-Log Linear Regression for Content Variables Predicting Newspaper Revenue: Geographic Focus

	Online Media Revenue		Circulation		Advertising Revenue		Print Advertising Revenue		Total Print Revenue	
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
Variables	β	β	β	β	β	β	β	β	β	β
Month	.976***	.923***	-.819***	-1.085***	-.511***	-.752***	-.738***	-.966***	-.738***	-.949***
News		-.049		-.375***		-.407**		-.383***		-.372***
Sports		.024		.133*		.202*		.196**		.207**
Community		.079***		.072		-.014		.009		.011
Feature		.052**		.111*		.177*		.164**		.152**
Adjusted R^2	.95	.96	.67	.70	.26	.28	.54	.57	.54	.57
F value	2404.97***	611.94***	241.05***	56.56***	41.79***	10.22***	141.47***	32.9***	141.03***	32.66***
***F* for change in R^2**	8.609**		4.1***		1.98		3.17**		3.08**	

*p<. 1, **p < .05. ***p < .01.

The independent variables were entered at their natural logs of the monthly amount of words. The dependent variables were ent ered at their natural logs of the monthly dollars amount.

Table 4-3: Summary of Log-Log Linear Regression for Content Variables Predicting Newspaper Revenue: Origin

	Online Media Revenue		Circulation		Advertising Revenue		Print Advertising Revenue		Print Revenue	
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
Variables	β	β	β	β	β	β	β	β	β	β
Month	.976***	.970***	-.819***	-.995***	-.511***	-.606***	-.738***	-.793***	-.738***	-.796***
Wire		-.057		-.309***		-.206		-.149		-.156
Staff		.096***		.179***		.176*		.159**		.167**
Adjusted R^2	.95	.96	.67	.70	.26	.27	.54	.55	.54	.55
F value	2404.97***	919.82***	241.05***	93.36***	41.79***	15.5***	141.47**	49.95**	141.03***	50.19***
***F* for change in R^2**	9.24***		7.09***		2.00		2.45*		2.72*	

**p<.1, **p < .05. ***p < .01.

The independent variables were entered at their natural logs of the monthly amount of words. The dependent variables were ent ered at their natural logs of the monthly dollars amount.

Appendix A: Specification of Full Log-Log Regression Model

We first specify the relationship between newspaper revenue (both online and in print) over time and the words amount of content on story topic dimension as follows:

$$OnlineRevenue_t=\alpha+\beta_0 Time_t+\beta_1 \ln(Sports_t) + \beta_2 \ln(News_t) +\beta_3 \ln(Feature_t)+\beta_4 \ln(Community_t)+\varepsilon_t \quad (1)$$

$$Circulation_t=\alpha+\beta_0 Time_t+\beta_1 \ln(Sports_t) + \beta_2 \ln(News_t) +\beta_3 \ln(Feature_t)+\beta_4 \ln(Community_t)+\varepsilon_t \quad (2)$$

$$AdvertisingRevenue_t=\alpha+\beta_0 Time_t+\beta_1 \ln(Sports_t) + \beta_2 \ln(News_t) +\beta_3 \ln(Feature_t)+\beta_4 \ln(Community_t)+\varepsilon_t \quad (3)$$

$$PrintAdRevenue_t=\alpha+\beta_0 Time_t+\beta_1 \ln(Sports_t) + \beta_2 \ln(News_t) +\beta_3 \ln(Feature_t)+\beta_4 \ln(Community_t)+\varepsilon_t \quad (4)$$

$$PrintRevenue_t=\alpha+\beta_0 Time_t+\beta_1 \ln(Sports_t) + \beta_2 \ln(News_t) +\beta_3 \ln(Feature_t)+\beta_4 \ln(Community_t)+\varepsilon_t \quad (5)$$

In above equations 1-5, $OnlineRevenue_t$ represents online revenue in dollars amount obtained by a newspaper in period *t*; Circulationt represents the average daily circulation in a given time period *t*; $AdvertisingRevenue_t$ represents advertising revenue in dollars amount in a given month. $PrintAdRevenue_t$ represents the total advertising revenue in dollars amount obtained by the print edition. We calculate this number by subtracting online revenue amount from the total revenue in a given month. Print $Revenue_t$ represents the total revenue in dollars amount obtained from the print edition in a given time frame. We calculate the number by adding print advertising revenue amount and circulation revenue in a given month. α represents the intercept. We consider the effect of control variable Time on online revenue. We use the natural logarithms of word amount in sports, news, feature, and community announcement categories to reflect the diminishing returns to content amount overtime. β_1, β_2, β_3, β_4 capture the impact of content amounts in four story topics on the revenue indicators

Next, we specify the relationship between newspaper revenue indicators (both online and in print) over time and the amount of story contents on geographic locations as follows:

$$\text{Online Revenue}_t=\alpha+\beta_0 Time_t+\beta_1 \ln(Local_t) + \beta_2 \ln(Regional_t) +\beta_3 \ln(National_t)+\beta_4 \ln(International_t)+\varepsilon_t \quad (6)$$

$$Circulation_t=\alpha+\beta_0 Time_t+\beta_1 \ln(Local_t) + \beta_2 \ln(Regional_t) +\beta_3 \ln(National_t)+\beta_4 \ln(International_t)+\varepsilon_t \quad (7)$$

$$\text{Advertising Revenue}_t=\alpha+\beta_0 Time_t+\beta_1 \ln(Local_t) + \beta_2 \ln(Regional_t) +\beta_3 \ln(National_t)+\beta_4 \ln(International_t)+\varepsilon_t \quad (8)$$

$$PrintAdRevenue_t=\alpha+\beta_0 Time_t+\beta_1 \ln(Local_t) + \beta_2 \ln(Regional_t) +\beta_3 \ln(National_t)+\beta_4 \ln(International_t)+\varepsilon \quad (9)$$

$$\text{Print Revenue}_t=\alpha+\beta_0 Time_t+\beta_1 \ln(Local_t) + \beta_2 \ln(Regional_t) +\beta_3 \ln(National_t)+\beta_4 \ln(International_t)+\varepsilon \quad (10)$$

In equations 6-10, the measurements of the dependent variables stay the same as equations 1-5. $Local_t$ represents the word amount of stories that have a focus on the local community, city, and county in a given time frame t. $Regional_t$ represents the word amount of stories that primarily focus on three neighbouring states in a given time frame t. $National_t$ represents the word amount of stories that primarily focus on cities or regions in states other than the newspaper's home state or neighbouring states. $International_t$ represents the word amount of stories that primarily focus on countries or regions outside of the United States. For these independent variables, we use the natural logarithms of word amount in local, regional, national, and international categories to reflect the diminishing returns to content amount overtime. β_1, β_2, β_3, β_4 capture the impact of content amounts in four geographic focuses on the revenue indicators.

Last, we specify the relationship between newspaper revenue indicators (both online and in print) over time and the amount of story contents on source of story origin as follows:

$$OnlineRevenue_t=\alpha+\beta_0 Time_t+\beta_1 \ln(Wire_t) + \beta_2 \ln(Staff_t) +\varepsilon_t \quad (11)$$

$$Circulation_t=\alpha+\beta_0 Time_t+\beta_1 \ln(Wire_t) + \beta_2 \ln(Staff_t) +\varepsilon_t \quad (12)$$

$$Advertising\ Revenue_t=\alpha+\beta_0 Time_t+\beta_1 \ln(Wire_t) + \beta_2 \ln(Staff_t) +\varepsilon_t \quad (13)$$

$$PrintAdRevenue_t=\alpha+\beta_0 Time_t+\beta_1 \ln(Wire_t) + \beta_2 \ln(Staff_t) +\varepsilon_t \quad (14)$$

$$PrintRevenue_t=\alpha+\beta_0 Time_t+\beta_1 \ln(Wire_t) + \beta_2 \ln(Staff_t) +\varepsilon_t \quad (15)$$

Wire$_t$ represents the word amount of stories that were written or produced in a given time frame t by wire services such as Associated Press, Reuters, etc. Staff$_t$ represents the word amount of stories that were written or produced in a given time frame t by the newsroom staff. With the same measurement of dependent variables, we use the natural logarithms of word amount in wire-written and staff-written to reflect the diminishing returns to content amount overtime. β_1, β_2 capture the impact of content amounts in two sources origins of content on the revenue indicators.

---

[i] We finally chose average daily circulation as the indicator of circulation, and dropped circulation revenue. As the correlation matrix shows, the circulation revenue has low to none correlation with most of the content variables. This may due to the fact that the newspaper changed its price per copy across years and across markets. The average daily circulation is a more stable and reliable indicator of circulation performance than the circulation revenue. Hence, we dropped circulation revenue for the log-log linear regression.

[ii] Sports category includes codes of hockey, sports, basketball, high school sports, local hockey, tennis, high school sports, volleyball, national football league, football, master's tournament, NASCAR, Major League Baseball, softball, college sports, track, Olympics, women's basketball, wresting, Dallas baseball team, professional basketball, soccer, NBA. News category includes codes of business, community news, Texas news, tri-state, auto, technology, terrorism, news index, district news, legal and crime, news, homicide, growth, national politics, year 2000, business, sex education. Feature category includes codes of entertainment, feature, outdoor, program on air, faith, belief, active life, art, religion, book review, letters, opinion, how to, food, health. Community announcement category includes codes of birthday, obituary, engagement, friendship and neighborhood, bride, anniversary, wedding.

[iii] We decide to use the number of words per each coding category as the measurement of content outputs, because it directly represents the newsroom's content volume and yields most significant regression results. We also tried several alternative measurements, such as using the factor scores generated by the exploratory factor analysis of all content categories, newsroom's two-dimensional categorization (i.e. local / non news, local/non local sports, local/non local feature), and average amount of words amount per each category. We dropped these later approaches, because they yield less consistent or significant regression results that were difficult to explain.

[iv] Ln (Y)=β_0+β_1Ln(X). By the chain rule, the derivative of the Ln(Y) with respect to X is the same as the derivative of the Ln(Y) with respect to Y times the derivative of Y with respect to X. Also, since the derivative of the Ln(X) with respect to X is 1/X, we can write: $\frac{dLn(Y)}{dX} = \frac{dLn(Y)}{dY}\frac{d(Y)}{dX} = \frac{1}{Y}\frac{dY}{dX}$, taking the derivative of [β_0+β_1Ln(X)] with respect to X, we can then write $\frac{1}{y}\frac{dY}{dX} = \frac{\beta_1}{x}$, which implies that the content elasticity of revenue with respect to content amount is equal to β_1. Therefore, $\frac{dY}{Y}\frac{X}{dX} = \frac{dY/Y}{dX/X} = \frac{\%changeinY}{\%changeinX}$

[v] We caution practitioners against a knee jerk reaction to a negative effect in this case. The negative effect due to a content variable indicate that, at the current level of content allocation and financial performance, the marginal impact of extra quantity towards that content variable (e.g. news) is less impactful than another content area (e.g., sports). Also, with respect to the negative effect of news content, we note that we are superimposing the impact of our content variables on two big changes over time: number of words (which itself is decreasing) and the relative proportion of words in the different categories. Further research should also model why news content amount is also decreasing and control for this feature while estimating the impact of news content on financial performance. However, even given this caveat, we believe management can get a fair idea of the effectiveness of each content type in affect financial performance given our approach.