The Internet and the Market for Local Newspapers

Lisa M. George
Hunter College and The Graduate Center, CUNY

August 31, 2006

PRELIMINARY – PLEASE DO NOT CITE OR QUOTE

Abstract

The internet has dramatically reduced the cost of distributing information goods quickly over large areas, and a growing economics literature documents that electronic media have drawn consumers from traditional media markets. Less work has been done to examine *how* the internet has altered either the audience for traditional media or the products themselves. Using zipcode-level data on newspaper sales from the Audit Bureau of Circulations and internet penetration rates from the Current Population Survey (CPS), this paper documents that the internet has differentially attracted younger, more educated and more urban individuals away from daily newspapers. Higher internet penetration is not associated with lower newspaper sales among blacks and Hispanics, who thus far have been less likely to connect. There is evidence that the internet has driven changes in coverage, with greater emphasis on local topics such as education and crime and on less time-sensitive areas such as investigative reporting.

I. Introduction

The early days of the internet produced headlines predicting the demise of traditional media. Bold-faced headings such as "Are Newspapers Doomed" and "The Future of Books" attest to concern that access to timely, specialized and often free information from electronic sources would drive traditional media from market. After a decade of use, it is clear that the internet has not replaced newspapers, television, books or other traditional information sources. Yet the pull of faster, better, cheaper information clearly remains on the minds of producers across markets.

A growing literature in economics seeks to document the extent to which electronic sources substitute for traditional products. Early papers considered the effect of the internet on retail markets (Goolsbee 2001; Sinai and Waldfogel 2004). An active debate continues on the relationship between free music exchanged through file sharing and sales of CD's (Zetner 2003; Oberholzer and Strumpf 2004; Blackburn 2004; Rob and Waldfogel 2004) and concert tickets (Mortimer and Sorensen 2005). More recently, the relationship between print and on-line newspaper and magazine sales has been studied (Filistrucci 2005; Kaiser and Ulrich 2005; Simon 2005; Gentzkow 2006). Yet despite growing evidence that the internet has drawn consumers from traditional media, little work has been done to determine *how* the audience for traditional media has changed in response to the internet. Changes in market composition are particularly important for media products because research has shown that, unlike traditional product markets, the number and types of products available depends on the distribution of consumer tastes. In traditional product markets, firms that find themselves competing against new products can reduce output, producing just enough to satisfy profitable niche markets. In media markets, large fixed costs limit the number of products that can survive in a market. Thus technologies that reduce the size or alter the composition of the audience for media can mean smaller, fewer or different products for remaining consumers. The tendency for consumers to affect each other through product markets has been documented in radio, newspaper and television.[1]

In the context of the internet, if groups with particular tastes are more likely than others to connect, these groups might also be more likely to jettison traditional media in favor of electronic sources. For example, if younger individuals with a college degree are more likely to have ready access to the internet, they might also be more likely to shift consumption to internet news. Individuals who switch are better off, but the effect on consumers who do not switch is uncertain. All else equal, the presence of fixed costs would suggest that a smaller audience for traditional media would mean fewer products with less variety, making remaining consumers worse off. But traditional media sources would be expected to reposition themselves in response to internet penetration, altering content to better suit the new distribution of types. An example might be reducing coverage of topics that appeal to a younger urban audience such as live entertainment and adding content appealing to older suburbanites, perhaps education. Changes might also take the form of shifting coverage toward topics less adaptable to the internet, for example newspapers shifting away from breaking news stories to in-depth analysis.

[1] The mechanism through which consumers with different tastes affect each other in product markets is developed in Waldfogel (2003). Empirical evidence of preference externalities is offered for radio in Waldfogel (200x), for newspapers (George and Waldfogel 2003) and television Waldfogel (2004).

This paper aims to document the extent to which the internet has altered the audience for daily newspapers and examine how coverage has changed in response. Newspapers are studied for several reasons. First, despite circulation declines over many decades, newspapers remain the news source of record in the US at both the national and local level. Second, the number of products available in any market tends to be small, aiding analysis of product positioning. Third, the availability of comprehensive circulation data at the zipcode level allows inference about consumption within particular demographic groups.

Using zipcode-level data on circulation at 1,025 newspapers in 1996, 2000 and 2004 combined with market-level data on internet penetration, analysis shows that the spread of the internet has not led to widespread circulation declines in newspaper markets. However greater internet penetration is associated with circulation declines among younger, urban, college-educated consumers relative to less youthful, less educated and less urban consumers. Evidence indicates that the internet has led to larger declines in readership among whites than among blacks and Hispanics. There is evidence that newspaper coverage has changed in response, increasing resources devoted to investigative reporting and local topics such as education and crime.

The paper proceeds as follows. The next section summarizes the data used in the analysis. Section III describes the empirical strategy and presents results. Section IV concludes the paper.

II. Data

Two sets of data are used in the analysis. The structure and composition of each is described in turn.

A. Circulation Data Set

The basic data used to estimate the effect of the internet on the composition of the newspaper audience is a zipcode-level panel of per-capita newspaper sales in 182 newspaper markets in 1996, 2000 and 2004, along with population characteristics by zipcode from the 2000 Census and internet penetration at the market level.

Internet penetration for metropolitan statistical areas (MSA's) or consolidated metropolitan statistical areas (CMSA's) is constructed using individual data from the Current Population Survey (CPS) Computer and Internet Use Supplement for years 2000 and 2003. The CPS supplements record internet access along with demographic characteristics for approximately 50,000 households each year. Geographic identifiers are not available in smaller markets (MSA population less than 20 million), so the sample is restricted to the approximately 35,000 individuals living in 191 MSA's or CMSA's that can be linked to circulation data. Internet penetration in the market is defined as the fraction of households with an internet connection in the home. In some specifications, all types of internet connections are considered. In others, only high-speed (broadband or DSL) connections are studied. Demographic characteristics are based on heads of household in the survey.

Aggregate zip code-level circulation is constructed from underlying sales for approximately 1,000 daily general-interest newspapers. Newspaper circulation data are published by the Audit Bureau of Circulations (ABC), a membership organization providing independent audits of newspaper and magazine circulation data for use by advertisers. Circulation across newspapers is summed to create zip code-level totals. Because some papers do not report circulation data every year, to allow inclusion of more papers in the sample, the zipcode totals are created from daily circulation averaged over the years 1996/1997, 2000/2001 and 2004/2005. Circulation for these years is merged with internet data from 2000 and 2003, with internet penetration in 1996 set to zero. Several major papers do not report circulation data to ABC, so the home markets for these papers (9 markets) are excluded from the analysis.[2] National newspapers such as the *Wall Street Journal*, *Christian Science Monitor*, and *USA Today* also do not report zipcode sales, so these papers are excluded as well.

B. Newspaper Positioning Data

The basic data for estimating the effect of the internet on newspaper coverage is a market-level panel of the reporting topics covered by daily newspapers along with internet penetration in 1993, 1997, 2000 & 2004.

Market-level coverage data is constructed from newspaper-level data on the job titles of individual reporters and editors listed in *Burrelle's Media Directory*. The *Directory* is a listing of newspaper staff produced by Burrelle's Information Services, a media monitoring organization. Each edition reports the job title of more than 30,000 reporters and editors at the 1,400-1,500 daily newspapers published across the US. To match with circulation and internet data, the analysis considers 933 papers published within 182 MSA's. For this study, job titles are linked by keyword to about 100 detailed reporting topics, which are then consolidated into 28 mutually-exclusive reporting categories. The distribution of reporters across categories is used to construct market-level measures of newspaper composition over time. For some specifications, the full set of papers in a market is studied. For others, only the largest paper in each market is considered.

The positioning data is matched by MSA to the internet penetration data from the CPS described above. Because the positioning data is available over a longer time period than the circulation data, the analyses are conducted over four time periods rather than three. Reporter data from 1997 are matched with internet penetration for 1998, reporter data from 2000 are matched with internet penetration for 2000 and reporter data for 2004 are matched with internet data from 2003. Internet penetration in 1993 is defined as zero.

C. Sample Characteristics

Tables 1-2 summarize data used in the study. Table 1 documents circulation and internet trends across markets over time.[3] Average *per capita sales* in the MSA drop from 19% to 17%

---

[2] The following markets are excluded because circulation for the dominant paper is not available at the zipcode level: Beaumont-Port Arthur (TX), Lake Charles (LA), Little Rock-North Little Rock (AR), Myrtle Beach (SC), Nashville (TN), New York – New Jersey (NY-NJ) Oklahoma City (OK), Scranton-Wilkes-Barre-Hazleton (PA) and Tulsa (OK).

[3] Internet penetration is measured in 2000 & 2003. Internet penetration is assumed to be zero in 1996, although home internet access was available in some markets during this time. The potential for error introduced by this assumption is discussed in section x. Internet penetration in 2003 is matched with newspaper circulation in 2004.

between 1996 and 2004, with a range from about 11% at the 5th percentile to 25% at the 95th percentile. Internet penetration spread rapidly over this period, with all types of access reaching an average of 42% in 2000 and 55% in 2003 and high speed growing from 4% in 2000 to 20% in 2003. There is substantial variation across markets in all years. In 2000, the penetration ranges from 25% at the 5th percentile to 58% at the 95th percentile for all types of access and 0-10% for high speed access. The spread is similar in 2003, with all types of access ranging from 40% to 70% at the 5th and 95th percentiles for regular access and 8% to 33% for high speed access.

Table 2 shows home internet use by demographic group in 2000 & 2003. Overall internet penetration had reached 44% by 2000 and 57% by 2003. High speed access spread from just 5% in 2000 to 23% in 2003. Differences in internet penetration across groups is apparent from the table. Households headed by individuals with a college degree comprise about 30% of the population in both years, but 45% of those with internet access and 51% of those with high speed access in 2000. Even with rapid expansion of internet access, college educated individuals make up a larger fraction of the internet audience than the general population in 2003, 42% and 49%. High income individuals are also more likely than others to have an internet or high speed connection, making up 19-20% of the population and 30-40% of the internet and high-speed audience. The internet audience is also less racially and ethnically diverse than the population as a whole, with blacks and Hispanics constituting a smaller fraction of internet households than in the general population. Younger households (ages 18-34) are also more likely and older households (60+) less likely to have home internet access. The CPS does not include a specific indicator for urban or rural residency, but a comparison of internet use within and outside of large MSA's shows higher penetration rates in larger markets.[4]

Table 2 documents significant differences across groups in the tendency to connect to the internet. If the internet is a substitute for daily newspapers, circulation declines would be predicted among groups with higher tendency to connect relative to groups with lower tendency. The next section seeks evidence in support of this prediction.

III. Empirical Strategy & Results

This section outlines the empirical strategy for estimating the relationship between internet penetration and audience composition and presents results. The empirical approach outlined here is similar to that detailed in George and Waldfogel (2006). Readers are referred to that paper for more extensive exposition.

A. Local Newspaper Circulation

[4] The relationship between market size and the tendency to connect to the internet is considered in more detail in Sinai & Waldfogel 2004.

Three distinct strategies are used to study the effect of the internet on newspaper circulation. The first approach examines the overall relationship between internet penetration and local newspaper circulation in an MSA. If the internet were drawing readers from newspapers overall, greater internet penetration would be associated with lower per capita newspaper circulation. However, because cross-sectional relationships of this sort are vulnerable to the effects of unobserved heterogeneity, the analyses focus on longitudinal relationships. Specifically, if the internet draws readers away from daily newspapers, a larger *increase* in internet penetration should be associated with a greater *reduction* in circulation.

The next step is to examine whether groups with a greater tendency to connect to the internet are more likely to see declines in newspaper circulation as the internet becomes more widely available. Table 2 suggests that the internet is more widely used among youthful, affluent, college-educated individuals living in large metro areas, so the largest circulation declines would be predicted among these groups. Table 2 also suggests that black and Hispanic households constitute a smaller fraction of internet users than the general population, indicating that smaller circulation declines would be expected among blacks and Hispanics. To empirically estimate the effect of the internet on newspaper sales among different groups, the paper makes use of zipcode-level variation in the fraction of the population in these groups. The simplest way to do this is to divide MSA's into "zones" reflecting areas of the market with high and lower percentages of each group. For example, the high-education zone is defined as all zipcodes where more than 68% of the population has a college degree, which is the top 10% of zips. The relationship between changes in internet penetration and changes in newspaper circulation in the high education zone then can be compared with the relationship in the low education zone. If the internet exerts a stronger pull on college educated individuals, the relationship should be more negative in the high education zone than in the low education zone. The exercise of defining zones and estimating the effect of internet growth inside and outsize of the zone is repeated using the zipcode fraction urban, fraction young, fraction black, etc.

The zone approach is intuitively appealing, but is limited by the need to consider groups one by one. Also, since some groups of interest constitute a small fraction of the population, the zones are a very imprecise measure of behavior among particular groups. A number of markets also may not have zipcodes which fit the criteria for the zone, reducing the sample size and power in the estimation. A third empirical strategy, the continuous analogue of the "zone" regressions, avoids these limitations. With this approach, the zipcode fraction of the population in each demographic category of interest is interacted with internet penetration. Coefficient estimates on these interaction terms show the extent to which internet affects local newspaper circulation among the groups with a higher tendency to connect relative to groups with a lower tendency to connect.

1. Overall Effect of the Internet

As a first step, local newspaper sales are aggregated to the MSA level. The overall relationship between internet penetration and per capita local newspaper sales in a market is estimating according to the following:

$$(1) \qquad \frac{S_M}{Pop_M} = \alpha_0 + \alpha_1 \frac{Internet_M}{Pop_M} + \varepsilon_M$$

where S_M / Pop_M is per capita newspaper sales in an MSA and $Internet_M / Pop_M$ is per capita internet penetration in that MSA. In this simple model, α_1 shows the relationship between internet penetration and newspaper circulation $\left(\frac{\partial S_M}{\partial Internet_M}\right)$, and ε is an error term. Table 1 presents results of the estimation in 2000 and 2004. The first two columns shows the effect of all types of internet use, while the third and fourth columns show results for high-speed internet connections. The results of the cross-sectional regressions are generally positive though not highly significant, suggesting no clear relationship between internet penetration and newspaper circulation.

However, simple cross-sectional models are vulnerable to biases from unobserved heterogeneity. For example, if consumers in markets with greater internet penetration also have a greater appetite for news, then estimates of equation (1) may produce positive coefficients even if the internet functions as a substitute for daily newspapers. A better approach would be estimation of a longitudinal extension of equation 1, namely:

$$(2) \qquad \frac{S_{Mt}}{Pop_M} = \alpha_0 + \alpha_1 \frac{Internet_{Mt}}{Pop_M} + \phi_t + \mu_M + \varepsilon_{Mt}$$

Inclusion of year and MSA fixed effects (ϕ_t and μ_M) allows estimation of the effect of *changes* in internet penetration on *changes* in per capita newspaper sales. This specification produces unbiased estimates of α_1 as long as unobserved market attributes do not change over time. The second panel of table 3 shows results of the estimation. Both time dummies are negative, reflecting declining per capita newspaper sales overall. The coefficient estimates for internet penetration (α_1) is about 0.03, suggesting that markets with larger increases in internet penetration are not losing circulation as fast as those with slower increases in internet use.

This result is counter-intuitive in that it suggests that internet expansion *increases* newspaper circulation. One interpretation of these results is that internet expansion is correlated with other factors affecting newspaper sales. If that is the case, it is not possible to identify the overall effect of the internet on newspaper sales from the aggregate regressions. However it is possible to identify how the audience for newspapers has been affected by internet expansion. This is done by studying how the gap in sales across different groups of readers changes as the internet expands. This identification strategy is valid as long as the endogenous change in newspaper circulation correlated with internet expansion is the same for different types of consumers. The next section details the approach.

2. The Effect of the Internet on the Newspaper Audience (Zone Estimates)

Aggregate estimates such as those above obscure any possible differential impact internet penetration on newspaper sales. If local newspaper sales were available for different demographic groups, equations (1) and (2) could be estimated separately for each group. Direct comparison of α_1 across groups would reveal whether the internet differentially attracted particular types away from local newspaper markets.

While data on local newspaper sales by demographic category is not available, zipcode data can be used to infer how the internet has altered the newspaper audience. As a first step, each MSA is divided into zones based on the characteristics in table 2. Aggregating local newspaper circulation over zips with a high fraction in the group of interest and over zips with a low fraction in the group allows estimation of the relationship between changes in internet penetration and changes in per capita local newspaper sales across zones.

More formally, equation 2 is modified to include a dummy variable for each zone, alone and interacted with internet penetration and the year dummies. This produces:

(3)

$$\frac{S_{zone}}{Pop_{zone}} = \alpha_0 + \alpha_1 \frac{Internet_M}{Pop_M} + \alpha_2 Zone + \alpha_3 Zone * \frac{Internet_M}{Pop_M} + \phi_t YD + \varphi_t YD * Zone + \mu_M + \varepsilon_M$$

Equation (3) is estimated for each demographic category of interest. The effect of internet penetration in areas with a low group fraction (outside the zone) is given by α_1 and the year effects by ϕ_1 and ϕ_2. The effect of internet penetration in areas with a high group fraction (inside the zone) is given by the sum of the internet variable and internet-zone interaction, $\alpha_1 + \alpha_3$. The year effects inside the zone are similarly given by the sum of the coefficients on the year dummies and year-zone interaction terms. The average difference in per capita sales across zones is given by the zone coefficient α_2. As before, longitudinal estimation with market fixed effects shows the relationship between *changes* in internet penetration and *changes* in per capita sales. The common fixed effect assumes that unobserved market characteristics are the same across zones.

In defining the zones, a zipcode is generally classified as high education, high urban, high youth, high black, etc. if the zipcode fraction of residents in the category is in the top 10% of all zipcodes. For example, a zipcode is included in the high education zones if the fraction of individuals with a college degree is at least 68 percent, which is the top 10% of zips. The youth zone includes all zip codes where the fraction of the population ages 18-34 exceeds 30%. Markets which contain no zips in the target zone are excluded from the zone regressions. In some categories, the number of markets with zipcodes in the target zone is quite small, in which case a looser cutoff of 20% is used. In general, tighter cutoffs provides a stronger link to the group under study. For example, comparing zipcodes with more than 50% Hispanic to those with less would be expected to be more informative than a comparison of zips with more or less than 15% Hispanic. However, only 43 MSA's have any zipcodes with more than 50% Hispanic, with fewer than half of these with more than 10 zips in the zone. To include a larger number of markets, a lower cutoff of 15% Hispanic is used, which is the top 20% of zips.[5]

Table 4 presents estimates of equation 3 for the seven demographic categories of readers delineated in table 2. The top row of the table identifies the criteria for inclusion in each zone. The first column in each category shows results for all types of internet access and the second column considers only high speed internet access. A quick scan of estimates on the interaction term α_3 shows that virtually all of the coefficient signs are consistent with the hypothesis that per capita newspaper sales are falling among groups with higher internet penetration relative to groups with lower internet penetration, although a number of the estimates are not statistically significant. Specifically, the coefficient estimates for groups more likely to have internet access (highly educated, urban, youthful and high income readers) are virtually all negative, suggesting that increases in internet penetration in a market reduce newspaper sales among these groups relative to the general population. Similarly, the coefficient estimates for groups less likely to have internet access (older, black and Hispanic readers) tend to be positive, suggesting that increases in internet penetration in a market increase newspaper sales among these groups. The one group for which results are opposite of those predicted are for blacks. Increased internet penetration in zips with a high fraction black see decreases in readership relative to markets with a lower fraction black. The results for high speed internet penetration tend to be larger and more significant than estimates considering all types of internet penetration, but this does not hold in all cases.

To estimate the magnitude of the effect, it is useful to consider an increase in internet penetration of 25 percentage points, which is approximately equal to moving from the 25th to the 75th percentile from 2000-2003 for both regular and high speed internet access. The results in column (2) show that per capita newspaper sales are 6.7 percentage points higher in the high-education zone (α_2). The effect on an increase in internet penetration in the low ed zone is about 0.1 (α_1), while the effect in the high ed zone is -0.15 ($\alpha_1+\alpha_3$). An increase in internet penetration of 0.25 from 2000-2004 would be expected to *increase* circulation in the low ed zone by 0.025 and *decrease* circulation in the high ed zone by 0.0375, reducing the newspaper sales gap to less than 1 percentage point across zones. Results in several other categories are also striking. The estimates in columns (4) show per capita sales in the urban and non-urban zip codes. Per capita sales in urban areas are higher by 3.7 percentage points. An increase in internet penetration of 0.25 is associated with an increase in per capita sales in the non-urban zone 0.25*.084 = 0.021. Inside the urban zone per capita sales would be predicted to decline by 0.25 * -0.012 = -0.003. The gap in per capita sales across zones is estimated to shrink by 2.4 percentage points. For Hispanics, results in column 11 indicate than in increase in internet penetration of 25 percentage points would increase per capita sales outside the Hispanic zone by 0.005 and inside the Hispanic zone by 0.032. The sales gap shrinks by 0.027, from 5.7 percentage points to 3 percentage points. Note that for the case of senior citizens, internet penetration exacerbates an existing gap in newspaper sales. Column (8) predicts that an increase in internet penetration of 25 percentage points would increase the gap in per capita sales of 8.6 percentage points to more than 13 percentage points.

---

[5] In general, results in table 4 are not highly sensitive to the cutoffs. However, restrictive cutoffs reduces the sample of MSAs considerably, which does affect estimates.

In sum, the results in table 4 indicate that an increase in internet penetration is associated with a reduction in newspaper sales among highly educated, urban, youthful and non-Hispanic readers relative to less highly educated, less urban, older and Hispanic readers. No effect is apparent for other groups, and some coefficient estimates are not statistically significant. One reason for the lack of clear results in some areas may be that may be that the fraction of individuals in the target zones is in some cases not very high. For example, a zipcode is assigned to the black or Hispanic zone if the fraction of the population in the zip is greater than or equal to 14%, which is not particularly high even though these represent the top 10 or 20% of zips. Other market factors correlated with the fraction black or Hispanic, such as urbanization or income, might overshadow the demographic effects. A similar effect might be at work for zipcodes with a high fraction elderly. The number of markets with zipcodes in the target zone is also low in some categories, further reducing the power of the estimation. Using the full range of variation across zipcodes eliminates some of these concerns, and the next section details these estimations.

3. The Effect of the Internet on the Newspaper Audience (Zipcode Estimates)

Variation in zipcode composition can be exploited to obtain more precise estimates of the effect of internet penetration on newspaper sales across groups. To see this, first consider the very simple regression of per capita circulation in a zipcode on the zipcode fraction in group *g*:

$$(4) \qquad \frac{S_z}{Pop_z} = \gamma_0 + \gamma_1 g_z + \varepsilon_z ,$$

where S_z is newspaper sales in the zipcode and g_z is the fraction of persons in each zipcode in the group of interest. While the coefficient γ_1 must be interpreted with some care, under some assumptions it provides a measure of the group tendency to purchase a daily newspaper.[6] Of interest here is how the internet alters the relationship between group membership and newspaper sales, in other words whether γ_1 is different in zips with higher levels of internet penetration. The relationship between γ_1 and internet penetration can be studied by adding internet penetration to equation (4), both on its own and interacted with the zip-group fraction. More specifically, the estimation equation becomes:

$$(5) \quad \frac{S_z}{Pop_z} = \gamma_0 + \gamma_1 g_z + \gamma_2 \frac{Internet_M}{Pop_M} + \gamma_3 g_z * \frac{Internet_M}{Pop_M} + \varepsilon_z .$$

Here, γ_2 shows the effect of internet penetration on newspaper sales overall, and γ_3 indicates how the group-newspaper relationship varies with internet penetration. A negative coefficient on the interaction term would indicate that the relationship between group membership and newspaper sales is smaller in markets with higher internet penetration.

[6] Interpreting $\gamma 1$as the differential between the tendency for individuals across groups to purchase a daily newspaper requires an assumption that the demographic characteristics observed do not reflect other factors associated with tastes for newspapers. For example, it might be that limited English skills rather than Hispanic status *per se* drives the difference in per capita newspaper sales in zipcodes with a high and low fraction Hispanic. However, for the analysis here, as long as the underlying factors that determine readership for each demographic group are the same across markets, γ_1 provides a valid measure of the group-readership differential. More importantly, γ_3 provides a measure of how that differential is altered by the spread of the internet.

Cross-sectional regressions are vulnerable to unobserved heterogeneity, so equation (5) must be modified to make use of longitudinal data. Adding market or zipcode fixed effects show how increased internet penetration affects changes in the relationship between the zipcode group faction and zipcode newspaper sales. The equation can also be expanded to consider the full set of demographic groups affected by the spread of the internet. The estimation equation becomes:

$$(6) \qquad \frac{S_{zt}}{Pop_z} = \gamma_0 + \gamma_2 \frac{Internet_{Mt}}{Pop_M} + \Sigma_i \left[\gamma_{1i} g_{iz} + \gamma_{3i} g_z * \frac{Internet_{Mt}}{Pop_M} \right] + \mu_M + \phi_z + \phi_t + \varepsilon_{zt}$$

where γ_{1i} shows the relationship between per capita newspaper sales and the zipcode fraction in group *i*, and γ_{3i} shows how changes in internet penetration alters the group-sales relationship. With zipcode fixed effects (ϕ_z), none of the γ_1 coefficients can be estimated because they are observed at only one point in time at the zipcode level.

Table 5 presents estimates of equation 6 for both general internet access and high speed access. Columns (1) and (3) include MSA fixed effects, while columns (2) and (4) include zipcode fixed effects. The positive and significant coefficients on internet penetration (γ_2) in all columns show that increases in internet penetration lead to increases in daily newspaper circulation in zips where the g_z are all zero. The coefficients on zipcode group fractions (γ_{1i}) in columns (1) and (3) show that greater internet penetration is associated with higher sales in zipcodes with a higher fraction college educated, higher fraction age 18-24 or age 60+, and higher fraction urban. Zipcodes with a higher fraction black or Hispanic are associated with lower sales. The effect for income is positive but not significant. In most cases, these coefficients are consistent with evidence on the literature on newspaper sales and readership.[7]

More interestingly, the interaction coefficients (γ_{3i}) reinforce some results in the zone regressions of table 4. The negative interaction coefficients for fraction college, fraction urban and fraction youth indicate that markets with greater increases in internet penetrations see smaller gaps in newspaper consumption among educated, urban, youthful residents. The positive interaction for Hispanics also indicates a smaller gap in markets with greater growth in internet penetration. In contrast to the zone regressions, the estimate for blacks is also positive, indicating that higher internet penetration shrinks the gap in per capita sales more in markets with a greater fraction black.

Results for two groups are not consistent with the hypothesis of differential distraction from daily newspapers (or with results in Table 4). Greater internet penetration appears to be associated with a larger gap in per capita sales between high income and low income zips, and with a smaller gap in sales between zips with a high and low fraction senior citizens. One explanation might be a tendency for papers facing greater internet competition to more heavily target readers most valuable to advertisers, who tend to be wealthy and young.[8] It is worth noting that the estimates with zip fixed effects are larger and more significant than those estimated with MSA fixed effects only. This suggests that unobserved zip attributes might have affected results in table 4.

---

7 Citations

8 Advertising cites.

B. Newspaper Targeting

With evidence in tables 4 and 5 that the internet has changed the composition of the newspaper audience, it is worthwhile to consider the extent to which newspapers have changed as a result. Data on the fraction of reporters and editors devoted to different topics can be used to characterize reporting over time and study whether markets with greater internet penetration have seen different trends than markets with a less rapid spread. Table 6 presents results of two sets of 26 regressions showing the relationship between changes in internet penetration and changes in the fraction of reporters devoted to different general reporting areas. For each of 26 reporting areas, the fraction of reporters assigned to the topic in the MSA was regressed on MSA internet penetration, a set of year dummies, and MSA fixed effects. Only the constant and internet effects are shown, along with t-statistics for internet penetration. The left panel considers the fraction of reporters in each category across all papers in a market, while the right panel considers only the dominant urban paper.[9] The categories were aggregated to help ensure topics would be mutually exclusive. These regressions include reporter data for 1993, 1997, 2000 & 2004 matched with internet penetration for 1998, 2000 & 2003. (Internet penetration in 1993 is set at zero). High speed internet access was not available (and not studied) prior to 2000, so is not included here.

Most of the coefficients in the table are not highly significant, however a number of the results are interesting. While the fractions of reporters assigned to local, regional and national news show no significant relationship with internet penetration, the fraction of coverage devoted to crime and education coverage does increase with internet penetration at statistically significant (or borderline significant) levels. Increases in internet penetration are also associated with increases in the fraction of reporters and editors devoted to investigative reporting, an area more suited to the slower but more thorough content generation at print newspapers than other sources. Interestingly, the fraction of reporters devoted to social policy & minority issues, which includes immigration, race, social services and welfare reporters, also shows increases in markets with higher internet penetration. Several other topics such as consumer issues and classified & real estate listings (across all papers) show a relationship to internet penetration, though the mechanism behind the effect is more opaque.

IV. Conclusions

The spread of the internet has not been the first shock to newspaper markets. Radio, then television, triggered steep declines in both the number and circulation of daily papers. By comparison, the overall effects of internet penetration documented here are small. But, at least so far, the internet appeals to more narrow demographics than radio and especially television, which diffused rapidly to a wide cross-section of the US population.[10] Differences across groups in the tendency to connect to the internet thus has the potential to alter the market for newspapers in ways different than earlier media innovations.

This paper provides evidence that the "digital divide" between groups more and less likely to connect to the internet has consequences in traditional product markets. Individuals

---

[9] In the case of mergers between two competing dailies, both papers are used to calculate the reporting shares.
[10] For example, while television owners in the late 1940's were more affluent than the general population due to the high cost of sets, the 1960 census shows penetration rates of about 90% across all demographic groups.

more likely to connect to the internet are growing less likely to purchase daily newspapers relative to groups with a lower tendency to connect. There is some evidence that traditional newspapers have shifted as a result, devoting a greater fraction of reporting resources to some local topics, such as education and crime, and toward social policy topics such as immigration and diversity of particular interest to groups such as Hispanics and blacks. These shifts perhaps ameliorate some ill-effects that have been associated with differential internet access.

In the long run, as the internet spreads to a more comprehensive set of households, differential targeting by traditional media producers of users less likely to connect to the internet would likely decline. Shifts toward topics more intrinsically suited to print than electronic journalism are likely to increase. The increase in the fraction of resources devoted to investigative reporting documented here indicates some shift has already occurred. More is likely to follow, at least until the next innovation comes along.

Bibliography

Audit Bureau of Circulations. *Circulation Data Bank*. New York: Audit Bureau of Circulations, 1999, 2006 editions.

Burrelle's Information Services. *Burrelle's Media Directory.* Livingston: Burrelle's Information Services. 1994, 1996, 2000, 2005 editions.

Filistrucci, Lapo. 2005. "The Impact of Internet on the Market for Daily Newspapers in Italy." European University Institute mimeo.

Gentzkow, Matthew. 2006. "Valuing New Goods in a Model with Complementarity: Online Newspapers." University of Chicago mimeo.

George, Lisa and Joel Waldfogel. 2003. "Who Affects Whom in Daily Newspaper Markets?" *Journal of Political Economy,* 11(4), pp. 765-785.

George, Lisa. 2006. "What's Fit to Print: The Effect of Ownership Concentration on Product Variety in Daily Newspaper Markets." Hunter College Mimeo.

Kaiser, Ulrich and Hans Christian Kongsted. 2005. "Do Magazines' 'Companion Websites' Cannibalize the Demand for the Print Version? " University of Southern Denmark mimeo.

Mortimer, Julie Holland and Alan Sorensen. 2005. "Supply responses to digital distribution: Recorded music and live performances." Harvard University mimeo.

Oberholzer-Gee, Felix and Koleman Strumpf. 2004. "The effect of file sharing on record sales: An empirical analysis." Harvard Business School mimeo.

Oberholzer-Gee, Felix and Joel Waldfogel, 2006. "Media Markets and Localism: Does Local News en Español Boost Hispanic Voter Turnout?" NBER Working Papers 12317.

Rafael Rob & Joel Waldfogel, 2004. "Piracy on the High C's: Music Downloading, Sales Displacement, and Social Welfare in a Sample of College Students," NBER Working Papers 10874.

Simon, Daniel. 2005. "The Effect of a Magazine's Digital Content on its Print Circulation: Cannibalization or Complementarity?" Cornell University mimeo.

Sinai, Todd and Joel Waldfogel. 2004. "Geography and the Internet: is the Internet a substitute or a complement for cities?" *Journal of Urban Economics* 56(1), pp. 1-24.

Zentner, Alejandro. 2003. "Measuring the effect of music downloads on music purchases." University of Chicago mimeo.

Waldfogel, Joel. 2003. “Preference Externalities: An Empirical Study of Who Benefits Whom in Differentiated-Product Markets” *RAND Journal of Economics*, 34(3), pp. 557-568.

Waldfogel, Joel. 2004. “Who Benefits Whom in Local Television Markets?” *Brookings-Wharton Papers on Urban Affairs, 2004*, pp. 257-305.

Table 1: Sample Statistics

MSA Statistics	**Year**	**MSA's**	**Mean**	**5%**	**25%**	**50%**	**75%**	**95%**
Per Capita Daily Newspaper Sales (ABC)								
	1996	182	0.189	0.121	0.161	0.186	0.219	0.259
	2000	182	0.183	0.121	0.154	0.182	0.207	0.257
	2004	182	0.171	0.109	0.147	0.168	0.195	0.235
Internet Penetration - All Types (CPS)								
	1998	182	0.271	0.129	0.213	0.273	0.326	0.425
	2000	182	0.423	0.245	0.352	0.428	0.488	0.577
	2003	182	0.555	0.395	0.496	0.561	0.622	0.699
Internet Penetration - High Speed (CPS)								
	1998	182	0.000	0.000	0.000	0.000	0.000	0.000
	2000	182	0.044	0.000	0.020	0.044	0.065	0.099
	2003	182	0.200	0.077	0.150	0.196	0.250	0.328
Zipcode Statistics	**Year**	**Zips**	**Mean**	**5%**	**25%**	**50%**	**75%**	**95%**
Per Capita Daily Newspaper Sales (ABC)								
	1996	9,849	0.200	0.052	0.117	0.174	0.239	0.395
	2000	9,849	0.194	0.056	0.118	0.170	0.229	0.370
	2004	9,849	0.184	0.054	0.113	0.162	0.216	0.342
Demographics (2000 Census)								
Fraction College		9,849	0.336	0.075	0.163	0.270	0.452	0.816
Fraction Urban		9,849	0.717	0.000	0.496	0.954	1	1
Fraction Youth (Age 18-34)		9,849	0.226	0.124	0.182	0.215	0.254	0.361
Fraction Old (Age 60+)		9,849	0.164	0.070	0.119	0.155	0.195	0.282
Fraction Black		9,849	0.103	0	0.005	0.025	0.100	0.544
Fraction Hispanic		9,849	0.100	0	0.011	0.029	0.101	0.493
Fraction Rich		9,839	0.133	0.021	0.054	0.096	0.177	0.379

Table 2: Internet Access by Group, 2000 & 2003

	All Households		Internet Households		High Speed Households	
	2000	2003	2000	2003	2000	2003
All Households	100%	100%	44%	57%	5%	23%
College Degree	29%	31%	45%	42%	51%	49%
High Income (75k+)	19%	20%	33%	31%	40%	40%
White	83%	82%	87%	86%	87%	86%
Black	13%	13%	7%	8%	6%	7%
Hispanic	10%	10%	5%	6%	5%	6%
Age 18-34	23%	23%	25%	23%	27%	26%
Age 60+	26%	26%	14%	17%	11%	12%
MSA's	72%	71%	76%	74%	85%	82%

Note. Sample includes only households in large MSA's (population of at least 20 million). Number of households in 2000 is 45,246; number in 2003 is 39455.

Table 3: MSA Internet Penetration and Per Capita Newspaper Sales

	Cross-Sectional Analysis				Longitudinal Analysis	
	All Internet		*High Speed Internet*		*All Internet*	*High Speed Internet*
	2000	2004	2000	2004	1996, 2000, 2004	1996, 2000, 2004
	(1)	(2)	(3)	(4)	(5)	(6)
Internet Fraction	***0.066***	***0.093***	***-0.008***	***0.059***	***0.030***	***0.028***
	(1.87)+	***(2.98)******	***(0.08)***	***(1.62)***	***(2.32)****	***(1.54)***
Year 2000	-	-	-	-	-0.019	-0.007
	-	-	-	-	(3.18)**	(4.30)**
Year 2004	-	-	-	-	-0.034	-0.023
	-	-	-	-	(4.46)**	(5.35)**
Constant	0.155	0.119	0.183	0.159	0.189	0.189
	(10.19)**	(6.81)**	(36.51)**	(19.68)**	(184.94)**	(182.78)**
MSA's	182	182	182	182	182	182

Notes: Dependent variable is aggregate per capita local newspaper circulation in each MSA. Overall internet fraction is measured as homes with any type of internet connection, high-speed internet fraction is measured as homes with cable, DSL or other broadband access. T-statistics in parentheses: +significant at 10 percent level; * significant at 5 percent level; ** significant at 1 percent level. Constants in fixed effects regressions represent the average value of the fixed effects. Standard errors clustered by MSA in cross-sectional specifications, MSA-year in longitudinal specifications.

Table 4: The Effect of Internet Penetration on the Local Newspaper Audience by Zone

	High Ed Zone (Fr. College >68%)		Urban Zone (Fr. Urban >95%)		Youth Zone (Fr. Age 18-34 >27%)		Old Zone (Fr. Age 60+ >21%)		Black Zone (Fr. Black >14%)		Hispanic Zone (Fr. Hispanic >14%)		High Income Zone (Fr. Inc.75k>20%)	
	All Internet	*High Speed*	*All Internet*	*High Speed*	*All Internet*	*High Speed*	*All Internet*	*High Speed*	*All Internet*	*High Speed*	*All Internet*	*High Speed*	*All Internet*	*High Speed*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	***(12)***	***(13)***	***(14)***
Internet	0.101	0.110	0.057	0.084	0.015	0.209	0.021	-0.037	-0.020	0.165	-0.056	0.021	0.165	0.189
	(1.68)+	(1.16)	(2.98)**	(3.09)**	(0.26)	(2.25)*	(0.44)	(0.52)	(0.22)	(2.19)*	(1.08)	(0.29)	(1.59)	(1.19)
Zone	0.067	0.067	0.037	0.037	0.015	0.015	0.086	0.086	-0.030	-0.030	-0.057	-0.057	0.034	0.034
	(3.20)**	(3.20)**	(9.91)**	(9.91)**	(1.27)	(1.27)	(8.15)**	(8.15)**	(2.99)**	(2.99)**	(5.65)**	(5.65)**	(1.81)+	(1.81)+
Internet*Zone	***-0.062***	***-0.255***	***-0.058***	***-0.096***	***0.042***	***-0.290***	***0.047***	***0.176***	***0.040***	***-0.334***	***0.142***	***0.110***	***-0.263***	***-0.322***
	(0.54)	***(1.52)***	***(2.20)****	***(2.29)****	***(0.47)***	***(1.64)***	***(0.76)***	***(2.36)****	***(0.32)***	***(1.77)+***	***(1.68)+***	***(0.75)***	***(1.29)***	***(0.98)***
Year 2000	-0.050	-0.010	-0.022	-0.001	-0.012	-0.014	-0.014	-0.004	0.002	-0.013	0.026	0.002	-0.080	-0.016
	(1.50)	(0.69)	(2.46)*	(0.42)	(0.46)	(1.34)	(0.67)	(0.57)	(0.07)	(1.34)	(1.03)	(0.20)	(1.46)	(0.92)
Year 2004	-0.076	-0.042	-0.032	-0.017	-0.024	-0.058	-0.027	-0.007	-0.008	-0.052	0.014	-0.021	-0.116	-0.062
	(2.25)*	(1.48)	(2.83)**	(2.59)**	(0.68)	(2.27)*	(1.02)	(0.51)	(0.17)	(2.44)*	(0.47)	(1.23)	(1.78)+	(1.40)
Yr 2000*Zone	0.032	0.017	0.011	-0.010	-0.023	0.009	-0.026	-0.014	-0.021	0.011	-0.068	-0.014	0.135	0.035
	(0.48)	(0.53)	(0.82)	(1.74)+	(0.59)	(0.43)	(0.83)	(0.95)	(0.44)	(0.65)	(1.65)	(0.87)	(1.26)	(1.02)
Yr 2004*Zone	0.038	0.058	0.002	-0.011	-0.036	0.047	-0.052	-0.061	-0.020	0.070	-0.078	-0.022	0.189	0.108
	(0.56)	(1.13)	(0.11)	(1.04)	(0.66)	(0.93)	(1.45)	(3.27)**	(0.29)	(1.29)	(1.59)	(0.66)	(1.49)	(1.19)
Constant	0.183	0.183	0.171	0.171	0.188	0.188	0.175	0.175	0.193	0.193	0.194	0.194	0.185	0.185
	(19.76)**	(19.79)**	(83.87)**	(83.82)**	(29.52)**	(29.35)**	(41.11)**	(41.23)**	(32.43)**	(32.44)**	(34.15)**	(34.37)**	(18.56)**	(18.57)**
MSA's	100	100	172	172	170	170	162	162	104	104	64	64	117	117

Notes: Dependent variable is aggregate per capita local newspaper circulation in each zone. Overall internet fraction is measured as homes with any type of internet connection, high-speed internet fraction is measured as homes with cable, DSL or other broadband access. T-statistics in parentheses: +significant at 10 percent level; * significant at 5 percent level; ** significant at 1 percent level. Constants in fixed effects regressions represent the average value of the fixed effects. Standard errors clustered by MSA-year.

Table 5: The Effect of Internet Penetration on the Newspaper Audience, Zipcode Regressions

	All Internet		*High Speed Internet*	
	(1)	(2)	(3)	(4)
Internet Fraction	*0.173*	*0.162*	*0.346*	*0.321*
	*(4.70)***	*(6.01)***	*(4.38)***	*(5.93)***
Year 2000	-0.029	-0.028	-0.008	-0.008
	(3.66)**	(3.11)**	(3.31)**	(2.72)**
Year 2004	-0.050	-0.048	-0.030	-0.029
	(4.72)**	(3.98)**	(4.27)**	(3.45)**
Fraction Hi Ed	0.076	-	0.079	-
	(4.26)**	-	(5.66)**	-
Fraction Black	-0.074	-	-0.074	-
	(10.69)**	-	(13.56)**	-
Fraction Hispanic	-0.113	-	-0.103	-
	(10.31)**	-	(10.83)**	-
Fraction Urban	0.074	-	0.065	-
	(10.96)**	-	(12.69)**	-
Fraction Age 18-34	0.088	-	0.066	-
	(2.14)*	-	(2.02)*	-
Fraction Age 60+	0.530	-	0.512	-
	(14.37)**	-	(18.08)**	-
Fraction Income $70k+	0.044	-	0.043	-
	(0.98)	-	(1.20)	-
*Internet * Hi Ed*	*-0.015*	*-0.035*	*-0.086*	*-0.139*
	(0.44)	*(1.71)+*	*(1.20)*	*(2.95)***
*Internet * Black*	*0.014*	*0.013*	*0.052*	*0.057*
	(0.93)	*(1.29)*	*(1.46)*	*(2.37)**
*Internet * Hispanic*	*0.065*	*0.060*	*0.133*	*0.113*
	*(3.54)***	*(3.91)***	*(3.47)***	*(2.84)***
*Internet * Urban*	*-0.076*	*-0.077*	*-0.195*	*-0.206*
	*(5.04)***	*(8.42)***	*(5.75)***	*(9.28)***
*Internet * Age 18-34*	*-0.150*	*-0.112*	*-0.304*	*-0.183*
	(1.70)+	*(4.20)***	*(1.55)*	*(2.55)**
*Internet * Age 60+*	*-0.169*	*-0.161*	*-0.405*	*-0.386*
	*(2.30)**	*(4.05)***	*(2.70)***	*(3.68)***
*Internet * High Income*	*0.027*	*0.074*	*0.113*	*0.252*
	(0.32)	*(1.56)*	*(0.62)*	*(2.20)**
Fixed Effect	MSA	Zip	MSA	Zip
Constant	0.005	0.181	0.019	0.181
	(0.39)	(114.37)**	(1.67)+	(110.25)**
MSA's	182	182	182	182
Zips	9,866	9,866	9,866	9,866

Notes: Dependent variable is aggregate per capita local newspaper circulation in each zip code. Overall internet fraction is measured as homes with any type of internet connection, high-speed internet fraction is measured as homes with cable, DSL or other broadband access. T-statistics in parentheses: +significant at 10 percent level; * significant at 5 percent level; ** significant at 1 percent level. Constants in fixed effects regressions represent the average value of the fixed effects. All regressions population-weighted with standard errors clustered by MSA-year.

		All Papers			*Major Papers*		
		Constant	Internet	*t-statistic*	Constant	Internet	*t-statistic*
1	Agriculture	0.002	0.003	[0.28]	-0.002	0.012	[0.86]
2	Arts & Entertainment	0.045	0.034	[1.26]	0.041	0.047	[1.28]
3	Auto	0.006	-0.005	[0.62]	0.005	-0.003	[0.31]
4	Business & Industry	0.078	-0.008	[0.30]	0.079	-0.004	[0.15]
5	*Consumer Interests*	*-0.003*	*0.007*	*[1.89]*	*-0.004*	*0.009*	*[1.97]*
6	*Crime & Courts*	*0.007*	*0.013*	*[1.65]*	*0.008*	*0.014*	*[1.48]*
7	Economics & Finance	0.008	0.002	[0.32]	0.008	0.003	[0.43]
8	Editorial	0.071	-0.032	[1.59]	0.073	-0.029	[1.24]
9	*Education*	*0.015*	*0.028*	*[1.77]*	*0.007*	*0.042*	*[2.53]*
10	Family	0.021	-0.008	[0.53]	0.017	-0.002	[0.09]
11	Features	0.055	0.011	[0.45]	0.062	-0.001	[0.04]
12	Food & Dining	0.02	-0.003	[0.21]	0.032	-0.026	[1.42]
13	Health	0.029	-0.017	[1.22]	0.033	-0.022	[1.40]
14	Home & Fashion	0.064	0.004	[0.15]	0.068	-0.005	[0.15]
15	*Investigative Reporting*	*0.01*	*0.017*	*[1.60]*	*0.011*	*0.019*	*[1.66]*
16	Labor & Workplace	0.002	-0.003	[0.98]	0.003	-0.003	[0.93]
17	*Listings*	*0.033*	*-0.031*	*[1.84]*	*0.026*	*-0.021*	*[1.03]*
18	News - Local	0.24	-0.015	[0.40]	0.218	-0.006	[0.13]
19	News - National	0.05	-0.004	[0.25]	0.058	-0.012	[0.57]
20	News - Regional	0.038	0.012	[0.68]	0.055	-0.003	[0.12]
21	Photography	0.023	0.004	[0.28]	0.018	0.008	[0.47]
22	Science & Technology	0.016	-0.005	[0.39]	0.01	0.003	[0.22]
23	*Social Policy & Minorities*	*0.002*	*0.019*	*[1.50]*	*-0.0003*	*0.025*	*[1.46]*
24	Sports	0.154	-0.026	[0.91]	0.152	-0.034	[1.16]
25	Travel	0.011	0.003	[0.31]	0.02	-0.009	[0.56]
26	Weekend	0.002	-0.001	[0.13]	0.004	-0.003	[0.52]