HARDCOPY AND INTERNET NEWSPAPERS: COMPLEMENTS OR SUBSTITUTES?

Matthew A. Gentzkow, Department of Economics, Harvard University
Dr. James. H. Collins, Scarborough Research

Introduction

The idea of technological change is captured, for many of us, by a series of well-known images. Travel on foot gives way to travel by horse, the horses are hitched to buggies, and the buggies become cars. Smoke signals are replaced by telegraphs then telephones, and recorded music moves from records, to cassettes, to CDs. New technologies replace inferior technologies, and progress marches on.

In keeping with this conception, the introduction of online news has been accompanied by dire predictions about the fate of more primitive media, with print newspapers singled out as the most likely victims. In his 1999 speech entitled "The Death of Print," Time, Inc. Editor Daniel Okrent commented, "Twenty, thirty, at the outside forty years from now, we will look back on the print media the way we look back on travel by horse and carriage" [1]. Warren Buffet echoed the point, saying “I love newspapers… But that is not the way the world is going… Newspapers are very threatened by the internet” [2].

While these statements fit neatly into a familiar story of progress, the reality of technological change—and change in media industries in particular—is often messier than the analogy to cars and buggies would suggest. In the early days of many now established medium predictions were offered regarding the demise of one or more already established ones; radio would kill newspapers, television would supplant radio, cable would displace broadcast television. However, the outcome in each case has been more complex. Similarly, print and online newspapers clearly compete on some dimensions, and coexist peacefully or are even complementary on others. How these complex interactions stack up, and what the ultimate effect of the new technology will be, is a fundamentally empirical question.

In this paper, we present new evidence on the impact of online newspapers on their traditional print counterparts, using data collected by Scarborough Research on the Washington DC market. The results, which primarily summarize the analysis of Gentzkow [3], show how tools of empirical economics can be combined with rich audience measurement data to disentangle the relationship between two different media products in consumers' choices. We also quantify the benefits consumers have derived from online news, moving beyond profit and loss to look at the broader social impacts of technological change. While limited to a single market and time period we offer the strategy developed herein as a model for the analysis of the increasingly complex nature of media and audience relationships.

Existing Evidence on Print-Online Competition

The interaction of print and online has understandably been a topic of great interest to newspaper readership researchers, and there is no shortage of studies addressing it. A handful, including a widely cited study commissioned by the Newspaper Association of America and conducted by Belden Associates, ask consumers directly how their print readership has changed since they began reading online [4]. The results are generally varied: significant fractions of consumers say their reading has either increased or decreased, while a majority report no change. This evidence is suggestive, and certainly helps dispel the notion that online news is crowding out print. Drawing more detailed inferences from such results is problematic, however, since the data are limited by consumers' ability to introspect on hypothetical choices (what they would have done in the absence of the internet). The questions also offer no way of assessing the magnitude of the changes, either positive or negative, in print readership.

The more common approach has been to look at data on consumers' actual print readership, as measured by audience data, and compare consumers who use online news to those who don't. Here, too, the results provide evidence against strong crowding out. Hague and Asquith [5] report that "based on the last 12 months of UK NRS data, 56% of all adults read a national newspapers yesterday and a strikingly similar 57% of people who use the internet daily are newspaper readers." Other studies, looking at a wide range of newspapers, find that consumers who read online are actually more likely to read print newspapers [6, 7, 8].

Looking at data on actual readership avoids many of the pitfalls of asking consumers about hypothetical choices. But simply comparing the online and non-online groups has its own problems which are if anything more severe. The key point is that online and non-online consumers differ along many dimensions, beyond just their internet usage. We cannot be sure how much of the difference in print choices is driven by online readership per se, and how much by these other characteristics.

To make the issue concrete, consider looking at similar data on the consumption of Mercedes cars and Chateau Rothschild wine among a random sample of US adults. We might well find that Mercedes drivers are on average more likely to drink expensive wine, but we would certainly not want to interpret this as saying that the relationship is causal and that the two products are therefore complementary. Rather, we would think that other characteristics of consumers—income, for example—drive consumption of both.

To get a clear picture of the causal relationship between print and online news consumption, one must go beyond simply comparing the average behavior in the two groups. Controlling for heterogeneity in this type of analysis is a classic topic in statistics, and is one of the key foci of applied work in economics. In the following sections, we show how these tools can be applied in the readership setting.

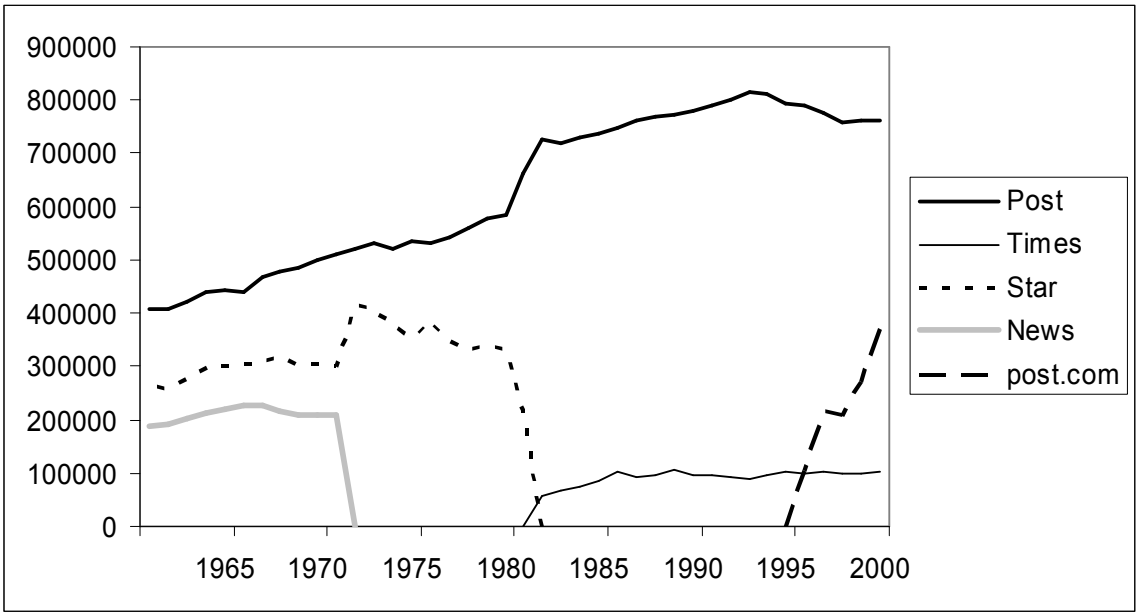
Data

Scarborough Research conducts research studies in 75 of the top local markets in the United States. In addition to extensive demographic information, Scarborough collects data on newspaper readership, radio listening, television viewing and a variety of shopping and lifestyle related activities. Beginning with the release of data in 2000 for the Washington, DC market Scarborough has collected audience information for both print and online versions of the major market newspapers. It is upon this audience data supplemented by other sources that much of this analysis will rest.

Washington DC has two major daily newspapers each with its own online version, the Washington Post and the Washington Times. The former is substantially larger than the latter with respective daily print circulations (2001) of over 750,000 and approximately 100,000 respectively. With respect to the online versions for the comparable time period the washingtonpost.com site received approximately 370,000 Washington area visitors per day with the washingtontimes.com receiving fewer than 40,000. (Because of the limited number of observations available for the washingtontimes.com it will be excluded from further analysis).

Figure 1 displays the daily circulations of Washington, DC's print and online newspapers since 1961 (with several minor exceptions). With respect to the question at issue the most important feature of the graph is that the introduction (in 1996) and growth of the washingtonpost.com has been associated with only a very modest decline in the circulation of the print version.

Figure 1: Newspaper Circulation and Newspaper Site Exposure in Washington DC (1961-present)



Source: Audit Bureau of Circulations' data on print newspapers and Media Metrics data on washingtonpost.com

Of course it is entirely possible that the washingtonpost.com did depress circulation of the print edition, but that the effect is obscured by other factors that were changing over time. It might well be that in the absence of the internet, the Post's print circulation would actually have increased substantially.

Controlling for Consumer Heterogeneity

In order to get a clearer picture of the print-online relationship, we turn to the individual-level consumption data collected by Scarborough in the DC market. Table 1 shows that the widely observed pattern of online readers being more likely to read the print newspaper than those who do not read online holds in this market as well: 59% of consumers who say they read the washingtonpost.com in the last 24 hours also read the Post print edition in the last 24 hours, as compared to only 40% of consumers who did not read the washingtonpost.com.

Table 1: Washington Post and washingtonpost.com Readership

Number of Consumers		**Read post.com:**	
		No	*Yes*
Read Post:	*No*	4481	255
	Yes	2932	371

Source: Scarborough Research, Washington 2001 survey.
Notes: Data is for reported readership in last 24-hours.

As discussed above, however, interpreting such relationships as causal is problematic, since those who read online are likely to be different in many ways from those who don't (they display a taste for news, for example). We can see this clearly in the current setting, by comparing the 19% difference in the total sample between those who read online and those who don't to smaller sub-samples. If we look only at consumers in households making more than $50,000 per year, for example, the difference falls to 14%; if we also restrict the sample to college graduates, it is only 8%.

A standard statistical methodology for separating consumer heterogeneity from causal effects of interest is multiple regression. Since the Scarborough data provides an extremely rich set of consumer characteristics, ranging from age and income to occupation, computer use, and party affiliation, a natural approach would be to regress print readership on these characteristics along with readership of the online edition. If the measured characteristics completely capture the relevant consumer heterogeneity, the coefficient on online consumption could be interpreted as a causal effect.

Coefficients from such a regression are presented in Table 2. Each number reflects the increase in the probability of reading the Post yesterday for a one-unit increase in the given variable. The results accord with both intuition and previous results on the determinants of readership: older, wealthier, more educated adults are more likely to read the Post. Whites are more likely to read, as are subway riders and Democrats (the last effect is consistent with the Post's reputation as a relatively liberal paper).

Table 2: Multiple Regression Coefficients
Dependent Variable: Read Post Yesterday

Age	.0053
Female	-.074
HH Income	.0087
High-School	.217
College	.304
Grad-School	.351
White	.031
Subway	.07
Democrat	.054
Read post.com	.115
N	8627

Notes: Coefficients are marginal effects from a probit regression, evaluated at the mean of the data. All coefficients are significant at the 1% level, with the exception of "White" which is significant at the 5% level. Regression also includes controls for years lived in the DC area and the portion of the DMA in which the respondent lives.

As expected, controlling for these characteristics causes the estimated effect of online readership on print readership to fall. The average effect is 11%, compared with 19% when no controls are included.

Results from a Full Economic Demand Model

Even controlling for observable characteristics of consumers in a regression framework does not eliminate the possibility of bias in the results. There may be other dimensions of heterogeneity that make a consumer more likely to consume both types of news. Indeed, a widely observed fact in readership studies is that some consumers just seem to be "news junkies"—they consume news with high frequency, and the difference between them and non-junkies cannot be explained by observable characteristics alone.

A great deal of recent research in economics has been directed at the problem of controlling for this kind of unobserved consumer heterogeneity. There are two kinds of information in the data that will make this possible. To understand the first, consider that the ideal way to answer our question would be to perform an experiment: randomly select some consumers and make it easier for them to access the online newspaper. If these consumers also begin reading more print newspapers, we could be confident that the products are truly complementary; if the consumers read less, we would know the products crowd each other out.

Literally carrying out such an experiment is difficult, but we can take advantage of "natural experiments" in the data that may perform the same function. Consider, for example, whether or not a consumer has access to the internet at work. If we thought such access was truly uncorrelated with other characteristics of consumers, we could treat groups with and without access at work just like the groups in the experiment. We may doubt that this variable is in fact uncorrelated, but we could at least hope that (i) the most important covariates (occupation, income) are accounted for in the data and (ii) whether or not consumers have access is not directly related to their taste for news. In this case, we could use access at work as an "instrument" for the real variable we care about—readership of the washingtonpost.com—and thus recover the true causal effect. Other variables in the data that may function the same way include high-speed internet connections at home, and subway ridership (the latter is equivalent to the experiment in reverse, serving as an instrument for print readership).

The second important feature of the Scarborough data is that it measures readership on both a last-24-hours and last-7-days basis. It would be natural to assume that the unobserved characteristics of consumers we care about are constant over the course of the week. If so, observing repeated choices can allow us to infer something about these unobserved traits.

To see this intuitively, consider watching the day-to-day choices of two readers. Reader A is a news junkie who likes both the print and the online edition, but sees the two goods as independent—having read one has no direct effect on her likelihood of reading the other. Reader B, on the other hand, sees the two as complements—he likes reading them both on the same day, so making the online more available would have a positive causal effect on his print readership.

Both consumers would be more likely on average to read either print or online on a given day. But the distribution over several days should be different. Reader A's distribution should be essentially random: some days she reads print, some days online, and some days by chance she reads both. Reader B, on the other hand, should only read the two together, and only rarely read one or the other separately. Although the data does not literally record choices on each day of the week, some information on repeated choice is contained in the 1- and 7-day questions, and we can potentially use this to separate consumers who look like readers A and B.

Gentzkow [3] uses an economic model to combine these different kinds of information and extract an overall estimate of the relationship between print and online newspapers. The details of that analysis are beyond the scope of this paper, and the interested reader is referred to the original for more information. The main results, however, are summarized in Table 3.

Table 3: Results from the Full Demand Model

Experiment: Take out Washington Post		
Washington Post Readers	1,712,270	
Change in Circ of post.com	-21,823	(-6.6%)
Experiment: Take out washingtonpost.com		
washingtonpost.com Readers	329,045	
Change in Circ of Post	15,558	(.9%)
Consumer Benefits		
Per Consumer as % of Print	66%	
Per Consumer in $ per day	$0.27	
Total as % of Print	13%	
Total in $ per day	$88,174	
Total in $ per year	$32,183,510	

Notes: Results from Gentzkow (2003).

The key conclusion is that when unobserved heterogeneity is properly accounted for, the apparent positive relationship between Post and washingtonpost.com readership disappears: the model predicts that if the washingtonpost.com went offline tomorrow, the number of print readers would actually increase. The scale of the effect is tiny, however, since only 16,000 of the more than 300,000 washingtonpost.com readers would increase their readership of print edition, resulting in a less than 1% change in print readership. On the whole, the products appear to be roughly independent.

Table 3 also shows results for the reverse scenario, removing the print edition and considering the effects on online readership. Interestingly, the effect in this direction is reversed—without the print edition, the washingtonpost.com would have fewer, rather than more readers. This reflects the fact that a large fraction of consumers have a strong preference for the print edition and get a small additional gain from adding the online. Without the print option, however, they prefer reading neither (i.e. switching to another source of news entirely).

Finally, Table 3 presents rough estimates of the benefit accruing to consumers from the free provision of online news. These numbers can be interpreted as the amount an average consumer would be willing to pay to read one day's worth of online news. The estimated value, 27 cents, is just over half the estimated average value of a single copy of the print edition. Adding this up over all Washington-area consumers over an entire year, we find that the online edition provides a total social benefit of just over 32 million dollars.

Conclusions

Two conclusions emerge strongly from this analysis. On the one hand, the positive correlation between print and online readership appears to be an artifact of consumer heterogeneity. It is unlikely that adding the online edition substantially increases print readership. On the other hand, we also find no evidence that the online edition substantially crowds out print readership. Barring major future changes in the nature of online media, we expect that print newspapers have little to fear in the long-run from their online counterparts.

Overall, the impact of online newspapers on the overall news markets resembles that of previous innovations like radio and television. In certain limited dimensions, the old and the new compete directly, but for the most part they are qualitatively different products. Radio clearly had an edge over newspapers in covering breaking news, but newspapers were still superior for in-depth coverage and classified advertising. Television clearly caused the end of serial comedies on the radio, but remains a decidedly inferior entertainment when one is driving a car. Similarly, consumers appear to use print and online news in different places and for different purposes, with relatively little crowding out.

We estimate that the free provision of online news has had substantial benefits to consumers. Though these do not exceed the operating costs incurred by online editions at the peak of the internet boom, they are likely higher than the level of costs today. In the long run, then, online news probably represents an important net gain for society.

Endnotes

1. Dan Okrent, "The death of print?" Hearst New Media Lecture, Columbia University.

2. Warren Buffet, USA Today, 2000.

3. Matthew Gentzkow, "Valuing New Goods in the Presence of Complementarities: Online News," working paper available by request.

4. Newspaper Association of America, 2001.

5. “Hearts and Minds – Managing a Newspaper Brand in the Digital Age”. 1999, Anita Hague and Richard Asquith, Proceedings of the Worldwide Readership Research Symposium #9, pages 501-507.

6. “Not Drowning but Waving: Researching the Internet”. 1997, Nigel Jacklin and Peter Highland, Proceedings of the Worldwide Readership Research Symposium #8, pages 227-230.

7. “Online Advertising Media in a Market and Multimedia Study: Measurement Methods and Findings on the Interaction Between Print and Online Consumption”. 1999, Johannes Schneller, Proceedings of the Worldwide Readership Research Symposium #9, pages 557-577.

8. “Print Media Internet Portals: A Boon or Bust for Print Media Consumption”. 2001, Jochen Hansen, Proceedings of the Worldwide Readership Research Symposium #10, pages 377-391.