Valuing New Goods in a Model with Complementarity: Online Newspapers

Matthew Gentzkow*

University of Chicago

May 5, 2005

Abstract

The rapid growth of information technology has focused attention on the extent to which new platforms for delivering information or services either crowd out or complement existing platforms. Recent structural demand models rule out complementarity by assumption, so their applicability to this problem has been limited. I develop a new model that relaxes this restriction, and use it to study the specific case of competition between print and online newspapers. Using micro data from the Washington DC market, I find that the major online paper reduces print readership by 30,000 per day and print profit by $5.7 million per year. This conclusion depends on properly controlling for consumer heterogeneity, as models without such controls imply that the print and online papers are either strong complements or independent. The overall welfare effect is a gain of $42 million for consumers and a loss of roughly $20 million for firms. I estimate that the profit-maximizing price of the online paper would be $.20 per day, in contrast to the current price of zero. However, I show that reasonable expectations of online advertising growth and a small transaction cost of online payments could rationalize both the introduction of the online paper and its low price.

JEL classification: C25, L82

Keywords: discrete choice, complementarity, online newspapers

*I would like to offer special thanks to Bob Cohen and Jim Collins of Scarborough Research for giving me access to the data for this study. I am also grateful to John Asker, Richard Caves, Gary Chamberlain, Judith Chevalier, Karen Clay, Liran Einav, Gautam Gowrisankaran, Ulrich Kaiser, Larry Katz, Julie Mortimer, Jen Paniza, Jesse Shapiro, Andrei Shleifer, Minjae Song, and especially to Ariel Pakes for advice and encouragement. I thank the Social Science Research Council for financial support.

"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."

Dan Okrent, Editor at Large of Time, Inc., speaking in 1999.

1 Introduction

The effect of online news on traditional newspaper markets has been the subject of much speculation and debate. Some see Internet news as a superior substitute whose continuing growth may curtail or even eliminate the market for traditional newspapers.[1] Others have disagreed, arguing that online news need not crowd out print consumption and could even complement it.[2] On the consumer side, few doubt that online news provides substantial benefits, but the magnitude of these benefits and the extent to which consumers will ultimately be willing to pay for them remain uncertain.[3]

The newspaper debate echoes a broader question that has attracted a great deal of interest from both economists and policy makers: when does the development of a new platform for the delivery of information or services either complement or crowd out existing platforms? Examples analyzed in recent work include the links between file-sharing services and sales of recorded music (Oberholzer and Strumpf 2004; Blackburn 2004; Rob and Waldfogel 2004), online and offline retailing (Goolsbee 2000, 2001; Sinai and Waldfogel 2004, Chiou 2004), print and online magazines (Kaiser 2002), television and newspapers (Gentzkow 2004), television and the Internet (Owen 1999), and movies released on VHS and DVD (Mortimer 2004).

In all these contexts, analyzing cross-platform competition boils down to measuring the impact of a new good. Recent literature in industrial organization suggests one approach to this problem: first estimate the parameters of a discrete choice demand system and a model of supply, taking account explicitly of observed and unobserved heterogeneity among consumers. Then ask how outcomes change when the new good is removed from the choice set. This framework allows rich

[1] In a 2000 speech, Warren Buffet said, "I love newspapers... But that is not the way the world is going... Newspapers are very threatened by the internet" (Henry 2000). Dan Okrent said in the same 1999 speech quoted above, "I believe... all forms of print are dead. Finished. Over" (Okrent 1999). Dick Brass, Microsoft's Vice President of Technology Development, predicted that the last print issue of the New York Times will appear in 2018 (Gates 2000). A study of the U.K. market by the firm Screen Digest concludes that the "dramatic decline in consumption of newspapers and books will continue as more consumers switch to electronic media" (Screen Digest 2002).

[2] A widely cited 2000 survey of U.S. online readers by Belden Associates found that 21% reported buying more print copies since they began using the internet, 7% reported buying less, and 72% reported no change. The same survey found that 7% had started a new print subscription since beginning to read online, while only 3% had stopped a print subscription (Runett 2002).

[3] For the discussion of the decision whether or not to charge for online content, see Seelye (2005), Kirsner (1997), and Gates (2002).

specifications of both the demand and supply sides of the market, and permits the researcher to look at the effect of the new good on demand for other products, consumer welfare, and firm profits.

These techniques have not, however, been widely applied to issues of cross-platform competition.[4] A key reason is that standard discrete choice demand models assume that each consumer chooses exactly one good from the available set, and as a consequence, all goods in these models are restricted *a priori* to be substitutes. In the settings mentioned above, consumers frequently choose multiple options simultaneously—they both download music and purchase CDs, buy books from amazon.com and Borders, and read newspapers in print and online. Whether goods are substitutes or complements is not something to be assumed *a priori*, but the key question of interest.

The goal of this paper is to analyze the impact of online newspapers using a discrete demand model that allows for both multiple choices and complementarity. I estimate the model using new individual-level data on the print and online newspaper readership of consumers in Washington DC, and look at the interaction among the Washington Post, the Post's online edition (the post.com), and the city's competing daily (the Washington Times). I then use the fitted model to ask whether the print and online newspapers are substitutes or complements, whether firms could make more money by charging positive prices for online content that is currently free, and how the introduction of online news has affected the welfare of consumers and newspaper firms.

The central empirical challenge in answering these questions is separating true substitutability or complementarity of the goods from correlation in consumer preferences. This remains true whatever the modeling framework. Observing that frequent online readers are also frequent print readers—or that file sharers buy more CDs—might be evidence that the two platforms are complementary. It might also reflect the fact that unobservable tastes for the goods are correlated—for example, that some consumers just have a greater taste for news or music overall which leads them to consume more heavily on all platforms.[5]

In the first section below, I discuss the identification problem in the context of a structural demand model. I show that the substitutability of goods is unidentified with data on consumers' choices and observable characteristics alone. I then point out two natural sources of additional information that can aid identification. The first is variables that can be excluded *a priori* from the utility of one or more goods. The most obvious example of such a variable would be price. In the newspaper market, there is no variation in price which can be plausibly separated from changing demand conditions over time, but I observe other variables such as whether consumers have Internet access at work which shift the utility of the online edition without affecting the utility

[4] An exception is the analysis of competition between cable and direct broadcast satellites by Goolsbee and Petrin (2004). In this case, it is reasonable to assume that each household chooses either cable or satellite but not both, and that the products are therefore perfect substitutes at the individual level. The challenge of handling simultaneous choices does not arise.

[5] This identification problem is closely related to Manski's (1993) reflection problem.

of the print edition. The second potential source of identification is panel data. If the correlated unobservables we are concerned about (i.e. the taste for news) are constant for a given consumer over time, observing repeated choices by the same consumer can allow us to separate correlation and complementarity. In the application, I have data on which newspapers consumers read in the last twenty-four hours, and also in the last five weekdays. This provides a limited form of panel data which I exploit in the estimation.

The results show that properly accounting for consumer heterogeneity changes the conclusions of the model substantially. Both reduced-form OLS regressions and a structural model without heterogeneity suggest that the print and online editions of the Post are strong complements, with the addition of the post.com to the market *increasing* profits from the Post print edition by almost $8 million per year. A structural model with observed heterogeneity only would lead us to infer that the print and online editions are independent, with the effect of the post.com on the Post not significantly different than zero. In contrast, when I estimate the full model with both observed and unobserved heterogeneity, I find that the print and online editions are significant substitutes.

In the end, I find that the growth of the online edition has crowded out print readership by a moderate amount. I estimate that raising the price of the Post by $.10 would increase post.com readership by about 2%, and that removing the post.com from the market entirely would increase readership of the Post by 30,000 readers per day, or 1.7%. Overall, the estimated $33.2 million of revenue generated by the post.com comes at a cost of about $6 million in lost Post readership.

With the estimates of the full model in hand, I also ask how profits would change if the price of the online edition were positive rather than zero. Whether positive prices for online content will become common in the future is a subject of frequent speculation in the industry. I find that the optimal price is indeed positive, at $.20 per day, and the loss from charging the sub-optimal price of zero is about $9 million per year. One possible explanation is that there are either real or perceived transaction costs of online payments. I show that a zero price would be optimal for any transaction cost greater than or equal to $.14 per day.

Turning to overall welfare effects, rough estimates of online operating costs, combined with the $33.2 million of post.com revenue and the $6 million of lost print sales already mentioned, suggest that the post.com reduced producer surplus in 2000-2003 by roughly $20 million per year. For consumers, the online edition generated a per-reader surplus of $.28 per day, implying a total gain of $42 million per year.

Finally, I ask what might explain the introduction of the post.com given that it appears to have reduced firm profits. I show that online advertising revenues increased substantially in 2004. A back-of-the-envelope calculation indicates that introducing the online edition would represent a net gain for the firm at this level of advertising. Furthermore, the optimal online price at the higher advertising level is only 9 cents per day and the loss only $2 million per year. This is some evidence for the intuitive proposition that introducing the online edition and keeping its price low were dynamic decisions motivated by an expectation of future market growth.

A number of previous papers have estimated discrete choice demand models that allow agents to select multiple goods.[6] The main difference between the current framework and those used in the past is that I allow a more flexible specification of the way goods interact in utility and the correlation of unobservable tastes. I show below that the functional forms for observable and unobservable utility that have been used in the past impose strong restrictions on substitution patterns—for a given set of observations, one could choose models from the literature that would imply that the goods are strong substitutes, independent, or strong complements. I also point out that the existence of both direct and indirect substitution effects makes the prior information necessary to support such restrictions stronger than is sometimes assumed. In certain settings, such assumptions will be justified, and making them has the obvious benefit of allowing the researcher to analyze much larger choice sets than the one considered here. In other settings, the necessary prior information is not available, and it will be critical to allow a more flexible structure and address directly how substitution patterns are identified by the data.

The next section analyzes the general problem of identifying substitution patterns in a discrete demand model with multiple choices, and provides some additional discussion of the existing literature. Section 3 introduces the data and presents reduced-form results on the relationship between print and online demand. Section 4 specifies the empirical model and estimation strategy, section 5 presents the results, and section 6 concludes.

2 Substitution patterns and identification

2.1 An illustrative model

In this section, I use a simple example to examine identification of substitution patterns in a discrete choice setting where consumers can choose multiple goods. I then discuss the extension to a more general setting.

Suppose there are two goods, labeled A and B, and consumers can choose at most one unit of each. We observe the choices of a large number of heterogeneous consumers. For each vector of observable consumer characteristics x, we can potentially measure three quantities: $P_A(x)$ (the probability of choosing A but not B); $P_B(x)$ (the probability of choosing B but not A); and $P_{AB}(x)$

[6] See the nested logit models of Manski and Sherman (1980) and Train, McFadden, and Ben-Akiva (1987); the multiple discrete choice framework of Hendel (1999) and Dube (2004); and the discrete-continuous model of Kim, Allenby, and Rossi (2002). Other papers that include multiple choices are Chan (2004), Augereau, Greenstein, and Rysman (2004), Smith (2004), and Gandal, Markovich and Riordan (2003).

(the probability of choosing both). The final probability—choosing neither—is linearly dependent so does not provide any additional information.

The goal is to estimate the various own and cross-price elasticities. These may in turn be inputs into the analysis of the welfare from new goods, the effect of a merger, or the change in profits from offering a different mix of products.

Denote the prices of discrete goods A and B by p_A and p_B. Income not spent on A or B is used to purchase a continuous composite commodity. Utility from q units of this commodity is αq which enters overall utility linearly. The expected utility for a consumer with characteristics x from consuming neither A nor B, A alone, B alone, and both A and B is given by:

$$\begin{aligned} u_0 &= 0 \\ u_A &= \bar{u}_A(x) - \alpha p_A + \nu_A \\ u_B &= \bar{u}_B(x) - \alpha p_B + \nu_B \\ u_{AB} &= u_A + u_B + \Gamma(x) \end{aligned} \tag{1}$$

Here utility of consuming neither A nor B is normalized to zero. This means that $\bar{u}_A(x)$, for example, is the difference between the mean utility of consuming A and the mean utility of consuming neither good. The difference between the utility from the composite commodity when A is purchased ($\alpha(y - p_A)$) and when neither good is purchased (αy) is just $-\alpha p_A$. This normalization is standard and does not restrict the model, since utility is defined only up to a monotonic transformation.

The terms ν_A and ν_B represent unobservable variation in the utility from goods A and B, which may be correlated. To make the discussion concrete, I assume their distribution is:

$$\begin{bmatrix} \nu_A \\ \nu_B \end{bmatrix} \sim N\left(0, \begin{bmatrix} 1 & \sigma \\ \sigma & 1 \end{bmatrix}\right).$$

The normalization of one of the variance terms to one is without loss of generality since we can divide all utilities by a constant and not change any of the choice probabilities. The normalization of the other is purely to simplify exposition—I discuss identification of an arbitrary covariance matrix below.

The term $\Gamma(x)$ is an interaction that determines how the utility from the bundle AB differs from the sum of the underlying utilities.

2.2 Substitution patterns

Let $F_x(\mathbf{u})$ be the distribution of $\mathbf{u} = [u_A \; u_B \; u_{AB}]$ implied by the assumptions above, conditional on characteristics x. Assuming consumers choose the bundle of goods with the highest utility, choice probabilities will be given by:

$$P_A(x) = \int_{\mathbf{u}} I(u_A \geq 0) I(u_A \geq u_B) I(u_A \geq u_{AB}) \, dF_x(\mathbf{u})$$

$$P_B(x) = \int_{\mathbf{u}} I(u_B \geq 0) I(u_B \geq u_A) I(u_B \geq u_{AB}) \, dF_x(\mathbf{u})$$

$$P_{AB}(x) = \int_{\mathbf{u}} I(u_{AB} \geq 0) I(u_{AB} \geq u_A) I(u_{AB} \geq u_B) \, dF_x(\mathbf{u}). \tag{2}$$

Denote expected demand for goods A and B conditional on observable characteristics x by $Q_A(x) = P_A(x) + P_{AB}(x)$ and $Q_B(x) = P_B(x) + P_{AB}(x)$. Because the quasi-linear specification of utility caused income to drop out, there are no wealth effects. The elements of the Slutsky matrix are then just the cross-derivatives of demand, and so by the standard definition:

Definition 1 *Goods A and B are substitutes if $\partial Q_A(x) / \partial p_B > 0$, independent if $\partial Q_A(x) / \partial p_B = 0$, and complements if $\partial Q_A(x) / \partial p_B < 0$.*

Figure 1 shows demand for the goods as regions of (u_A, u_B) space. The first panel shows the case of $\Gamma = 0$, the second panel shows the case of $\Gamma > 0$, and the third panel shows the case of $\Gamma < 0$. To see how the model determines the cross-price derivatives, observe first that increasing p_B is equivalent to shifting probability mass downward. That is, for any point (a, b) in this space, it increases the probability that $u_B \leq b$ given that $u_A = a$.

Consider the first panel. Increasing p_B causes marginal consumers such as m to switch from buying the bundle AB to buying A alone. It also causes marginal consumers such as n to switch from buying B alone to buying neither good. Neither of these changes has any effect on the demand for good A, however—the increase in P_A is exactly offset by a decrease in P_{AB}. This implies that when $\Gamma = 0$, the cross-derivatives of demand for the products will be zero, and that they are therefore independent.

Next, consider the second panel. Increasing p_B causes consumers m and n to switch as before. There will now be consumers such as o, however, who will switch from buying the bundle AB to buying nothing. This means that the drop in P_{AB} will be larger than the increase in P_A, and so demand for good A falls. In the case of $\Gamma > 0$, therefore, the goods are complements.

In the third panel, there are no consumers indifferent between buying AB and buying neither good, but consumers such as o are indifferent between buying A alone and buying B alone. Increasing p_B causes them to switch from buying B to buying A, so that the increase in P_A is larger than the drop in P_{AB}. We therefore find that $\Gamma < 0$ implies the goods must be substitutes.

This discussion suggests the quite intuitive result that the interaction term Γ is the key parameter for determining the substitutability of goods in a multivariate discrete choice model. Formally, we can substitute into the definition of $Q_A(x)$ and take the derivative with respect to p_B to show that:

$$\frac{Q_A}{dp_B} = \int_{\mathbf{u}} [I(u_A = u_B) I(-\Gamma \geq u_A, u_B \geq 0) \;\; - I(u_A + u_B = -\Gamma) I(u_A \leq 0) I(u_B \leq 0)] dF_x(\mathbf{u}) \,. \tag{3}$$

The first term inside the integral represents points on the dark diagonal line segment in the third panel of figure 1, along which consumers are indifferent between buying A alone and B alone. The second term represents points on the dark diagonal segment in the second panel, along which consumers are indifferent between the bundle AB and buying neither good.

Inspection of equation 3 immediately implies the following result.

Proposition 1 *For a consumer with characteristics x, goods A and B are substitutes if $\Gamma(x) < 0$, independent if $\Gamma(x) = 0$, and complements if $\Gamma(x) > 0$.*

2.3 Identification

Under the assumptions made so far, the model is not identified. At each value of the vector x, there are three observable data points and five independent parameters: $\bar{u}_A(x)$, $\bar{u}_B(x)$, $\Gamma(x)$, α, and σ. Having variation in x clearly does not help: observing choices at both x and x' adds three new moments (P'_A, P'_B, and P'_{AB}) but also three new parameters ($\bar{u}_A(x')$, $\bar{u}_B(x')$, and $\Gamma(x')$).

2.3.1 Price coefficient

The price coefficient α is nonparametrically identified from choice data if and only if there is variation in prices for each vector of observable characteristics x. To see this, note that all predicted probabilities would be the same if we replace the parameters $(\bar{u}_A(x), \bar{u}_B(x), a)$ by $(\bar{u}_A(x) + \alpha p_A, \bar{u}_B(x) + \alpha p_B, 0)$ for all x. With two observed price vectors for a given x, on the other hand, we gain three additional moments—any one of these would be sufficient to identify α given the other parameters of the model.

Even if we are willing to make parametric assumptions on the form of the mean utilities, α will be unidentified without price variation in any specification that allows a constant term for each good. However parametric assumptions will allow us to relax the requirement that observable x be identical at both price vectors.

A final possibility is that we have independent information about α from some other source. In the application below, this will come from firms' first-order conditions. In this case only the sums $\bar{u}_A(x)+\alpha p_A$ and $\bar{u}_B(x)+\alpha p_B$ are identified from demand data, but there will be a unique α such that the first-order condition is satisfied at the observed price.

I will assume in the discussion that follows that α is identified from one of these sources.

2.3.2 Covariance and interaction parameters

The remaining issue is how to separately identify the interaction term, $\Gamma(x)$, and the covariance of the unobservables, σ. As discussed in the introduction, this represents an identification problem akin to the reflection problem of Manski (1993). Intuitively, the mean utilities $\bar{u}_A(x)$ and $\bar{u}_B(x)$ will be identified by the marginal probabilities $Q_A(x)$ and $Q_B(x)$. The remaining moment in the data will be how often the goods are consumed together (whether P_{AB} is high relative to P_A and P_B). A high value of P_{AB} can be explained by either a high value of $\Gamma(x)$ or a high value of σ, and there is nothing left in the data to separate these.

Furthermore, proposition 1 shows that this leaves the substitution patterns in the model severely unidentified. Without some additional information, the same data could be fit by assuming that the goods are nearly perfect substitutes ($\Gamma \approx -\infty$ and σ high) or nearly perfect complements ($\Gamma \approx \infty$ and σ low). A model which "solves" the problem by imposing an *ad hoc* restriction on one of these two parameters will be unlikely to provide a basis for reliable inference about any quantity in which substitutability of the goods plays an important role.

There are, of course, many ways that more moments could be added to the data in order to identify the model. I will briefly discuss two that seem likely to arise frequently in practice and will play a key role in the application below. I will assume that the necessary technical conditions are satisfied such that the model is identified if and only if the number of moments is greater than or equal to the number of parameters.[7]

The first possible source of identification is exclusion restrictions. Suppose in particular that there is some variable x_j which is allowed to enter the utility of one good, say $\bar{u}_A(x)$, but does not enter either $\bar{u}_B(x)$ or $\Gamma(x)$. The most obvious candidate would of course be the price of good A. In the newspaper application considered in this paper, there is no price variation, but there are consumer specific observables such as having Internet access at work that affect the utility of online but not print newspapers. Having observations at a second value of such an x_j (call this

[7] To be more precise, write the vector of probabilities as a function of the parameters: $P = H(\theta)$. It should be clear that H is continuously differentiable in θ so long as the distribution of the errors is continuous and has full support—this can be shown by pointing out that the probabilities can be rewritten as integrals over different regions of the parameter space where the parameters enter either the boundaries of integration or the distribution function. The necessary assumption is that the Jacobian of H has full column rank—i.e. that there are no columns that are linearly dependent. This has to be checked on a case-by-case basis, but is likely to be satisfied so long as the distribution of the errors is not pathological.

new vector x') would add three new moments ($P_A(x')$, $P_B(x')$, and $P_{AB}(x')$) but only one new parameter ($\bar{u}_A(x')$). The model would therefore be formally identified.

Furthermore, the intuitive basis of the identification is quite strong. Suppose, for example, that the goods are frequently consumed together (P_{AB} is high relative to P_A and P_B). If this is the result of a high Γ, the goods are complements, and shifting up the utility of good A should also increase the probability of consuming good B. If Γ is zero and the observed pattern is the result of correlation, the probability of consuming good B should remain unchanged.[8]

The second possible source of identification is panel data. Suppose we extend the model slightly and assume that the observables (ν_A, ν_B) are made up of two components—a possibly correlated random effect term $(\tilde{\nu}_A, \tilde{\nu}_B)$ which is constant within consumers over time, and an additional time-varying component (ϵ_A, ϵ_B) which is assumed to be i.i.d. (one could also allow for time-varying shocks that affect bundles separately from the individual goods). In the newspaper application, this would amount to assuming that unobserved correlation in the utilities of different papers is driven by consumer characteristics such as a general taste for news that are constant over the course of a week, and that the additional shocks that lead consumers to read on Monday but not Tuesday are uncorrelated.[9]

Now, if we observe each consumer's choice at two different points in time, we have increased the number of moments from three to fifteen.[10] Under the assumption that $(\tilde{\nu}_A, \tilde{\nu}_B)$ is constant over time, this is sufficient for formal identification. Intuitively, the argument is just a variant of the usual one for the identification of random effects from panel data. Suppose again that goods A and B are frequently consumed together. If this is the result of correlated random effects, we should see some consumers likely to consume both and some consumers likely to consume neither, but conditional on a consumer's average propensity to consume each good the day-to-day variation should be uncorrelated across goods. If it is the result of a high Γ, on the other hand, the day-to-day variation should be strongly correlated—a given consumer might consume both on one day and neither on another day but would be unlikely to consume either one alone.

2.4 Generalizing the model

The simple example above was designed to illustrate as clearly as possible the basic identification problem that arises when consumers can choose more than one good simultaneously. Of course in

[8] Keane (1992) presents monte carlo evidence on the role of this kind of exclusion restriction in identifying the covariance parameters in a multinomial probit model. Since a multinomial probit model defined over bundles effectively nests equation 1, this evidence is relevant. He shows that including exclusion restrictions greatly improves the accuracy of the model.

[9] In the actual application, I also allow for correlation in these daily shocks.

[10] With observations at two points in time, the moments would be the probability of each possible combination of choices over the two periods. When there are four choices, this gives sixteen possible combinations. The number of moments is one less than this because the probabilities must sum to one.

almost any real application of such a model, including the one in this paper, we will want to use a richer model. In this section, I discuss the issues that arise when the choice set is larger than two goods, and when the model includes a more complex error structure.

Suppose we consider consumers' choices among J goods, indexed by $j \in \{1, ..., J\}$. Index all possible bundles of these goods by $r \in \{0, 1, ..., 2^J\}$ and denote the corresponding set of goods J^r. Assume the bundles are ordered so that $J^0 = \varnothing$ and for any $r \in [1, J]$, J^r is a singleton set consisting only of good $j = r$. To allow explicitly for panel data, index time periods by t.

Extending equation 1, I assume the utility of bundle r is:

$$\begin{aligned} u_{rt} &= \epsilon_{rt} && \text{if } r = 0 \\ &= \bar{u}_r(x) - \alpha p_r + \nu_r + \tau_{rt} + \epsilon_{rt} && \text{if } r \in [1, J] \\ &= \sum_{j \in r} \left[\bar{u}_j(x) - \alpha p_j + \nu_j + \tau_{jt}\right] + \Gamma_r(x) + \epsilon_{rt} && \text{if } r > J \end{aligned} \tag{4}$$

The errors are now divided into three components. The first, ν_r, is an unobservable taste for each good which is constant over time. The second, τ_{jt}, is a time-varying unobservable taste for each good. The third, ϵ_{rt}, is a time-varying taste for each *bundle*. I assume that these errors have a parametric joint distribution and that the ν_r have full support.

The first question is whether the identification arguments above remain valid. From an intuitive point of view, it seems clear that combining some number of exclusion restrictions and panel data should still allow us to separately identify the Γ_r interaction terms and the covariance matrix of the ν_r. We can be more precise, however. Suppose, for example, that the error distribution has the same number of parameters as the covariance matrix of a joint normal. For a given x, the model now has J $\bar{u}_r$ terms, $2^J - J - 1$ Γ_r terms, $J(J+1)/2 - 1$ covariance parameters (with one element of the covariance matrix normalized to 1), and a single price coefficient, for a total of $2^J + J(J+1)/2 - 1$ parameters. The number of moments is $2^J - 1$.

Assume again that the technical conditions are satisfied so that we can compare the number of moments and parameters. Assume also that there is either price variation or information from the firm's first-order condition that allows us to identify α. The model then has $J(J+1)/2 - 1$ unidentified parameters.

Consider, first, an excluded variable x_j that only enters the utility of a single good (i.e. only a single $\bar{u}_r(x)$ and none of the $\Gamma_r(x)$). This adds a single parameter to the model, but with observations at two values of x_j adds $2^J - 1$ new moments. It is easy to check that $2^J - 1 > J(J+1)/2$ for any $J \geq 2$ and so this is sufficient to identify the model. If the error structure were more flexible, identification would require either a larger number of goods or additional exclusion restrictions.

Next, suppose we have panel data with observations at two points in time for each consumer. This adds $2^{2J} - 2^J$ new moments (for every possible combination of choices on the two points in time), which is clearly greater than the $2^J - 1$ new moments added by an exclusion restriction, so this too is sufficient to identify the model.

The second question is how substitution patterns change in the model with multiple goods. To build some intuition, it is useful to recall how this works in standard (i.e. continuous) demand theory. When utility is quasi-linear and there are two goods, the goods will be complements if and only if the cross-partial of utility is positive. This is the analogue of proposition 1 above. As discussed by Samuelson (1974), the simple relationship between utility and demand no longer holds when there are three or more goods. To use Samuelson's example, suppose that the cross-partial of utility is negative for coffee and tea, positive for tea and cream, and positive but smaller for coffee and cream. Then coffee and cream will be substitutes by the standard Hicksian definition. Ogaki (1990) has shown that the matter can be clarified by decomposing the substitution effect into direct and indirect components—in the case of coffee and cream, the direct complementarity between the products will be outweighed by indirect substitutability working through consumption of tea.

It is possible to show that the same intuition holds in a general discrete choice model given by equation 4. Suppose, for example, that there are three goods, A, B, and C. Denote the interaction terms for each pair by $\Gamma_{AB}(x)$, $\Gamma_{AC}(x)$, and $\Gamma_{BC}(x)$ (i.e. $\Gamma_{AB}(x)$ is the $\Gamma_r(x)$ term in equation 4 for the bundle consisting only of goods A and B). Assume that the interaction term for the three-good bundle is $\Gamma_{ABC}(x) = \Gamma_{AB}(x) + \Gamma_{AC}(x) + \Gamma_{BC}(x)$.[11] Assume also that the ϵ_{rt} terms are equal to zero (or that we examine substitution patterns conditional on ϵ_{rt}). The following proposition, which is proved in the appendix, shows that the substitutability of goods A and B will depend on the sum of two terms—a direct component that is just $\Gamma_{AB}(x)$, and an indirect component that will have the same sign as the product $\Gamma_{AC}(x)\Gamma_{BC}(x)$.

Proposition 2 *For a consumer with characteristics x, goods A and B are substitutes if $\tilde{\Gamma}(x) < 0$, independent if $\tilde{\Gamma}(x) = 0$, and complements if $\tilde{\Gamma}(x) > 0$, where,*

$$\tilde{\Gamma}(x) = \Gamma_{AB}(x) + I(x),$$

and the indirect substitution term $I(x)$ has the same sign as $\Gamma_{AC}(x)\Gamma_{BC}(x)$.

Proof. *See appendix.* ■

So in the coffee-tea-cream example, we would have $\Gamma_{cof,cream} > 0$, $\Gamma_{tea,cream} > 0$, and $\Gamma_{cof,tea} < 0$, implying a positive direct effect and a negative indirect effect between coffee and cream..

[11] To understand why this restriction is intuitive, note that the two-good interaction terms such as Γ_{AB} are analogous to cross-partial derivatives of utility in a continuous model. Setting $\Gamma_{ABC} = \Gamma_{AB} + \Gamma_{AC} + \Gamma_{BC}$ is then analogous to setting the third-order derivative equal to zero, or in other words, to taking a second-order approximation of utility.

A key implication of this discussion is that imposing *a priori* restrictions on the interaction terms based on "intuitive" criteria will be dangerous. Suppose, for example, that we had a strong prior that the marginal utility of good A does not depend on consumption of good B. While it would be tempting to use this information to impose the restriction $\Gamma_{AB}(x) = 0$, this would only be correct if we believed we had included in the choice set all other goods that could potentially interact with either A or B (and all goods that could interact with these goods, and so on). If not, there could be indirect substitutability or complementarity that would cause the true cross-price elasticity to be non-zero—the model would either be unable to generate this, or would be forced to fit it by biasing other interaction terms.[12]

2.5 Relationship to past literature

It is worth pausing to consider how the model of equation 4 relates to other approaches in the literature for estimating discrete choices when multiple goods are chosen simultaneously.[13] The first point to make is that the earlier models are by no means all special cases of equation 4. They allow variously for multiple units of discrete goods, discrete-continuous choice, and relaxation of the quasi-linearity assumption, none of which are treated either in this model or in the application below. On the other hand, these models often place strong restrictions on either the interaction among goods (the Γ_r in the terminology of the model above) or the covariance of the unobservables (the ν_r). Looking at these restrictions may shed some light on the tradeoffs among the models, and the way the substitution patterns they generate depend on their underlying assumptions.

To make the discussion concrete, suppose we have micro data on demand for two goods, A and B. Suppose that the frequency with which the goods are consumed together is high relative to the frequency with which either is consumed alone.

Consider, first, the multiple-discrete choice approach pioneered by Hendel (1999) and applied by Dube (2004). These models assume that the data is generated by an aggregation over a number of individual choice problems, or "tasks." Each agent chooses a single good for each task, which

[12] Another way to understand this point is that when relevant goods are omitted from the choice set, the utility of each choice contains an implicit maximization over the omitted goods. If good C is omitted from the model, for example, the term u_A is implicitly $\max\{u_A, u_{AC}\}$ and u_{AB} is implicitly $\max\{u_{AB}, u_{ABC}\}$. The "correct" interaction term is then

$$\tilde{\Gamma}_{AB} = \max\{u_{AB}, u_{ABC}\} - \max\{u_A, u_{AC}\} - \max\{u_B, u_{BC}\} + \max\{u_0, u_C\}.$$

This may be non-zero even if the marginal utility of A does not depend on consumption of B (i.e. $u_{AB} = u_A + u_B$).

[13] I will focus on models that use micro data on individual consumer choices. There is of course a large literature on representative-agent models of aggregate demand, and in many settings this provides a plausible alternative to a discrete choice approach. The drawback is that these models do not provide a natural way to incorporate the identifying power of micro data—especially valuable in settings such as the newspaper market where we do not observe substantial exogenous price variation over time. Representative agent models also do not allow us to consider the distribution of choice behavior over the population, nor do they let us impose intuitive restrictions on the substitution parameters, such those in characteristics-based demand models like Berry, Levinsohn, and Pakes (1995).

makes the task-level problem analogous to equation 4 with $\Gamma_{AB} = -\infty$.[14] Because the utility from using a given good in one task does not depend on what good was chosen for the any other task, aggregating over a large number of these tasks is similar to aggregating over a population of heterogeneous consumers in a standard multinomial discrete choice model. The model therefore restricts the goods to be substitutes.[15]

A second approach is the multivariate probit (applied, for example, by Augereau, Greenstein, and Rysman 2004). Here, consumption of each good is assumed to be driven by a separate probit equation, with errors possibly correlated across equations. This is exactly equivalent to equation 4 with $\Gamma_{AB} = 0$, and so restricts all goods to be independent in demand (all cross-elasticities are zero).[16]

A third approach is to estimate a logit or nested logit model defined over the set all possible bundles. Papers that take this approach include Train, McFadden, and Ben-Akiva (1987) and Manski and Sherman (1980). Because each bundle's utility is parameterized separately, the Γ_{AB} term could be estimated freely (although both of these papers restrict the interactions as a parametric function of the goods' characteristics). The unobservables, on the other hand, are either assumed to be uncorrelated (in the case of the logit) or have a correlation structure dictated by the nests, which is too restrictive to allow the kind of correlation implied by equation 4 with $\sigma \neq 0$. Given the hypothetical data, we would expect such a model to find $\Gamma_{AB} > 0$, implying that the goods would be complements.

Thus, given a single dataset on choices of two goods, it would be possible to select models from the existing literature that would imply that the goods are strong complements, independent, or strong substitutes. Of course the parametric restrictions these models make pay off in terms of computational simplicity, and increase the size of the choice set that can be analyzed. But it is crucial to verify that these restrictions are motivated by valid prior information rather than the computational benefits *per se*. Furthermore, the discussion of indirect substitution effects

[14] Both papers allow consumers to choose multiple units of each good, so the task-level choice is more complicated than a standard multinomial discrete choice problem. But the utility specification implies that consumers will choose at most one type of good for each task.

[15] A different parametric restriction on the Γ interaction terms underlies the model of Chan (2004). He defines goods to be a bundle of characteristics, and assumes that the utility of a bundle is a function of the sum of each characteristic across the different goods. The bundle consisting of a bottle of Diet Coke and a bottle of Diet Pepsi, for example, consists of two units of the characteristic "cola," two units of the characteristic "diet," and one unit each of the characteristics "Coke" and "Pepsi." Because utility is assumed in the main specification to be concave in the total of each characteristic, it is sub-additive across goods, meaning $\Gamma_{AB} < 0$. This would again imply that the products must be substitutes. (Chan does find complementarity among some products in a specification with many goods which appears to result from indirect substitution effects as described above.)

[16] The discrete-continuous framework of Kim, Allenby, and Rossi (2002) also assumes the equivalent of $\Gamma_{AB} = 0$ (that the utility of a bundle is simply the sum of the utilities of the underlying goods). The conclusion that the goods must be independent does not hold here, however, because the utility of the outside composite commodity is allowed to be concave rather than linear. This implies that all goods will be substitutes, though with a single curvature parameter governing all the cross-elasticities as well as the elasticity of total expenditure on the inside goods.

suggests that even a strong prior intuition about the way consumption of one good affects the marginal utility of another may be insufficient justification for imposing such restrictions *a priori*. In situations where sufficiently strong prior information is lacking, it will be preferable to begin from a more flexible utility structure.

3 A first look at the data

3.1 The Scarborough survey

The empirical analysis is based on a survey of 16,179 adults in the Washington DC Designated Market Area (DMA), conducted between March 2000 and February 2003 by Scarborough Research. The Washington DC DMA includes the District of Columbia itself, as well as neighboring counties in Virginia, West Virginia, Pennsylvania, and Maryland. The data include a range of individual and household characteristics of the respondents, as well as information on various consumption decisions. Most importantly for the current application, these include an enumeration of all print newspapers read over the last twenty-four hours and five weekdays, as well as readership of the major online newspapers over the same periods.

Washington DC has two major daily newspapers, the Washington Post and the Washington Times. The former is dominant: average daily readership of the Post was 1.8 million in 2000-2003, compared to 256,000 for the Times. The key difference between the two papers is their political stance: the Times is generally thought to be much more conservative than the Post. The main online newspaper is the post.com, which had an average of 406,000 area readers per day.[17]

I will therefore define the goods in the model to be daily editions of the Post, the Times, and the post.com. The outside alternative will include other print and online newspapers, other news sources such as television and radio, and the choice not to consume news at all. I interpret *all* choices in the model to represent an implicit maximization over these outside goods—the observed choice to read the Post only, for example, includes consumers who both read the Post and the New York Times, or both read the Post and watched TV news.

Table 1 gives summary statistics for the Scarborough data along with corresponding census figures for the Washington DC DMA. The survey is approximately a .4% sample, and is broadly representative, with some over-representation of older, more educated, and more wealthy individuals. The survey includes sampling weights to correct for this over-representation. I will use the

[17] Readership figures are based on the Scarborough survey used in this study. Note that these readership numbers are larger than circulation figures for the same papers, reflecting the fact that multiple consumers read each copy. The Times also has an online edition, the washingtontimes.com, but its readership is very small and there are only 373 readers in my sample. In practice, this turns out to be too few to accurately estimate utility parameters for the washingtontimes.com, and so I omit it from the analysis.

un-weighted data for estimation, and use weights when I simulate aggregate effects.[18]

3.2 The newspaper product

Before turning to the data, it will be worth considering in principle why different newspaper products might be either substitutes or complements. In the case of the print and online editions of a single paper, the fact that the online edition generally contains close to 100% of the content available in the affiliated print edition underlies a common intuition that the products should be substitutes. Viewing newspapers as a bundle of information, and recognizing that consuming the same information twice should have close to zero value, we might predict that the marginal value of reading the print edition to someone who has already read the online edition should be close to zero.

Offsetting this, however, are a number of factors that make the products differentiated. Most obviously, one or the other medium may be convenient at different times and places—print newspapers at home over breakfast, say, and online newspapers at one's desk at work. The print edition offers particular advantages such as portability, lower eye strain, and the tactile experience of a printed page. Similarly, the online edition offers access to breaking news, continually updated financial information, archived stories, searchable classifieds, and special interactive features. All these factors would tend to weaken the degree of substitutability among the products. If reading a story about a news item in the morning increases the value of hearing breaking updates on the item later in the day, if consumers like to start a story at work and finish it at home, or if they buy the paper to read in depth analysis of the market movements or sports scores they had been tracking online, the two could even be complementary.

Another set of factors come into play when we consider the interaction of two print editions. On one hand, their content is different, weakening the argument for substitutability. Consumers might well gain from reading two newspapers that are particularly strong in different subject areas, or offer different perspectives on the same story. On the other hand, they are not differentiated in terms of convenience, as print and online editions are. We might well expect that a pair of "up-market" and "down-market" newspapers such as the New York Times and Daily News or the Boston Globe and Herald would be substitutes, though their interaction would be weak since they appeal to different segments of the population. A pair of newspapers that appeal to more similar consumers but have different strengths and ideologies might be more likely to be complements, as

[18] In addition to including weights, the raw Scarborough data also corrects for respondents who filled out an initial questionnaire but not the longer survey by filling in a small number of these consumers' survey responses using the responses of other consumers matched by demographics. These "ascribed" observations are easy to identify because the probability of two respondents with the same sampling weight matching perfectly on all survey responses by random chance is very low. I omit these observations (about 6 percent of the initial sample) in all estimation, but include them in the policy simulations in order to get the correct match to aggregate demographics.

reading both could provide a variety of viewpoints on the same issue.

3.3 Reduced-form results

Figure 2 displays the daily readership of Washington DC's print and online newspapers since 1961. The first thing to note is that the rapid increase in post.com readership since its introduction in 1996 has not been accompanied by a large drop in Post readership. A simple OLS regression of Post readership on post.com readership gives a significantly negative coefficient, but suggests that it takes four post.com readers to reduce Post readership by one.

This graph also provides evidence about the extent of substitutability among different print papers. The exits of the Washington Star in 1981 and the Washington News in 1973 both led to increases in the readership of the remaining papers, suggesting some substitutability. In both cases, however, the exits led to declines in total readership, and fewer than half of the readers of the exiting paper appear to have switched to one of the remaining papers. In terms of the Post and the Times, the time-series provides no evidence of a negative relationship. A linear regression of Post readership on Times readership actually gives a positive coefficient (though insignificant), even when a time trend is included. Of course, these regressions do not distinguish substitutability from changes in demand or characteristics of the products over time.

Turning to the micro data, the first thing to note is that readership of multiple papers is common. 48% of consumers reported reading at least one of the Post, Times, or post.com in the last twenty-four hours. Of these consumers, 18% reported reading two of the papers, and 1% reported reading all three. Over a five-day window, 65% of consumers read at least one of the papers; of these 27% read two papers and 3% read all three.

Table 2 reports raw and partial correlation coefficients for each pair of papers. The partial correlations control for age, sex, education, industry of employment, employment status, income, political party, date of survey, location of residence within the DMA, and number of missing values in the survey.[19] Readership of the Post and post.com are significantly positively correlated over both twenty-four-hour and five-day windows. Controlling for observable characteristics reduces this correlation by about two thirds, but it remains significant at the .1% level. The correlation between readership of the Post and the Times is also significantly positive in the raw data, but this disappears when controls are added. The partial correlation is zero over a twenty-four-hour window, and significantly negative over a five-day window. The correlation between the Times and the post.com is never significantly different from zero.

What can we conclude from these results? The basic fact in the raw data is that a consumer

[19] These correlations drop consumers for whom either print or online papers were excluded from the choice set as discussed in the demand specification below.

who reads any one paper is on average more likely to have also read a second paper. If all heterogeneity in utilities was uncorrelated across papers, this would be strong evidence that all three are complements.

Of course an alternative explanation is that the kind of consumers who get a lot of value from reading the Post also get a lot of value from reading the post.com and the Times. The fact that the positive correlation decreases dramatically when we partial out the effect of observables provides direct evidence for this. The question is whether the remaining correlation—in particular the positive correlation between the Post and the post.com—represents true complementarity or additional correlation in tastes which is unobserved.

As discussed above, one way to identify such unobserved tastes is to use variables that can plausibly be excluded from the utilities of one or more goods. I will use several variables that should intuitively have a strong effect on the utility of reading the online newspaper but should have no direct effect on the utility from reading in print. The first of these is a dummy variable measuring whether the consumer has Internet access provided by her employer. Being able to access the Internet at work clearly reduces the time cost of reading online, but should not directly affect the utility from reading in print. A second variable is a dummy for having high-speed Internet access at home. Finally, I use dummy variables for other tasks the consumer performs on the Internet: e-mail, chatting, job search, research, work-related tasks. Performing these tasks should lead consumers to be more familiar with the Internet and spend more time at their computers, both of which should decrease the effective cost of reading news online.

Each of these variables is unlikely to directly effect the utility of reading print newspapers. One might still worry, however, that they could be correlated with unobservables that in turn affect utility from print. For example, having the Internet at work will clearly be correlated with occupation, and high speed Internet access might be related to income. Many of these characteristics are things that I can control for explicitly–for example, I observe occupation and income. As a specification check, I can also verify that the results do not change substantially when I add increasingly detailed occupational controls. However, I will have to rely on the panel component of the data to identify any remaining unmeasured correlation.

One way to see the effect of these excluded variables is to use them to instrument for online reading in a linear probability model of print reading. There are several problems with such a specification—it will be biased because it does not restrict probabilities to be between zero and one, it restricts the cross-derivative of utility between print and online to be the same for all consumers, and it does not use all of the information from the other choice equations. However, it shows in an intuitive way how the exclusion restrictions contribute to identification.

Table 3 shows estimates from a linear probability model of readership over the five-day window, using the same controls as in the partial correlations above. Reflecting the positive correlation noted earlier, the first column shows that reading the post.com is positive and significant in an OLS regression. When I instrument with the excluded variables, however, the coefficient becomes significantly negative. The magnitude suggests that if we could do an experiment and randomly assign individuals to read the online paper (at zero time cost), they would be on average 40% less likely to read the print paper, though the limitations of the linear probability model mean this number must be interpreted with caution.

One possible concern is that the instruments drive not only post.com readership but readership of other online news sources as well. If these are substitutes for print readership, the IV coefficients could overstate the effect of the post.com. The Scarborough survey asks directly whether consumers use the Internet for news apart from reading online newspapers. As shown in column 3, adding this dummy variable as a control decreases the magnitude of the post.com coefficient by less than a fourth, suggesting that most of the effect does indeed work through the post.com. Of course it may be that the instruments affect the intensity of other online news readership in a way not picked up by the dummy variable, in which case we would have to treat the magnitude of the coefficient as an upper bound.

The final column adds dummy variables for forty-seven possible occupation categories (as opposed to the two dummies for computer and white-collar employment included before). Although these variables are jointly very significant, they change the magnitude of the post.com coefficient only slightly, providing evidence that it is not being driven by omitted variables whose correlation with Internet access works through occupation.

Table 4 shows first-stage regressions for the two-stage least squares specifications in table 3. The instrument with the most explanatory power is having Internet access at work, followed by chatting, having a fast Internet connection, and work-related tasks. One surprising result is that with controls for other Internet news, e-mail use has a negative coefficient. The results are robust to changing the set of instruments—the magnitude and significance of the post.com coefficient remains similar when each instrument is dropped in turn, with the exception that dropping the Internet at work variable which causes the magnitude to increase. The coefficient also remains similar when the Internet at work variable is used as the sole instrument.

3.4 Moving to a structural model

There are several advantages to moving beyond this reduced-form analysis to estimate a structural model of demand. First, observed choice patterns are determined by a web of interactions among the different goods. Simple regression frameworks require estimating each piece of this system separately.[20] A structural model will allow all the interactions to be estimated jointly. Second, the structural model will allow me to control explicitly for the possibility of unobserved correlation of tastes—the key identification problem highlighted above—and bring to bear multiple features of the data to identify them, while still allowing for non-linearity. Finally, a structural model is necessary to do welfare analysis, and to solve for the optimal price of the online edition.

4 Empirical specification

4.1 Demand

To specify the structural demand model, I impose several restrictions on the general form of equation 4. First, I assume that the interaction terms Γ_r do not vary with x. I have experimented with specifications that relax this assumption, and have not found that they change the results substantially. Second, because the data lacks sufficient price variation to freely estimate heterogeneous price elasticities, I will assume that the price coefficient α_i is constant across consumers.[21] And finally, motivated by the analysis of substitution patterns above, I assume that the interaction term for the three-good bundle (Post-post.com-Times) is simply the sum of the component two-good interactions (Post-post.com, post.com-Times, and Post-Times). In the robustness section I show that relaxing this restriction does not substantially change the results.[22]

Index consumers by $i = \{1, ..., N\}$, and define the mean utility of an individual good—$\bar{u}_r(x)$ for $r \in [1, J]$—to be a linear function of x with good-specific coefficients β, and a constant term denoted by δ. Utility functions are then given by:

$$\begin{aligned} u_{irt} &= \epsilon_{irt} && \text{if } r = 0 \\ &= -\alpha p_r + \delta_r + \sum_k x_{ik}\beta_{rk} + \nu_{ir} + \tau_{it} + \epsilon_{irt} && \text{if } r \in [1, J] \\ &= \sum_{j \in r}\left[-\alpha p_j + \delta_j + \sum_k x_{ik}\beta_{jk} + \nu_{ij} + \tau_{it}\right] + \Gamma_r + \epsilon_{irt} && \text{if } r > J \end{aligned} \tag{5}$$

[20] If we were to extend a simple framework by jointly estimating a system of equations and allowing for discrete choices and correlated errors, we would arrive at the structural model estimated below.

[21] Allowing heterogeneous price coefficients would not change the estimates of producer surplus, or of consumer surplus of the online edition relative to the print. The only case in which it would clearly change the key estimates of the dollar value of consumer surplus is if the average income of Post readers (the group whose behavior determines the estimate of α) was very different from that of post.com readers. From inspection of the data, this does not appear to be the case.

[22] While this restriction only eliminates a single parameter in the current three-good specification, it could be a natural way to restrict the growth of the parameter space as the choice set gets large. Letting S^r be the set of the indices of all two-good subsets of the bundle r, the general restriction would be $\Gamma_r = \sum_{s \in S^r} \Gamma_s$. This is equivalent to taking a second-order approximation of an arbitrary utility function, in the spirit of Christensen, Jorgenson, and Lau (1975), and it reduces the number of interaction terms from $2^J - J - 1$ to $\binom{J}{2}$, a substantial savings if J is large.

As discussed below, I will estimate the term $(-\alpha p_j + \delta_j)$ as a single constant and then use the supply model to separate its component parts.

I also make several parametric assumptions on the form of the errors. The first error component, ν_{ir}, is meant to capture time-constant consumer characteristics including overall taste for news (which would show up as positive correlation across the ν_{ir}), brand loyalty to a particular paper, and current subscription status. I assume ν_{ir} has a J-dimensional multivariate normal distribution with a free covariance matrix.[23] The second component, τ_{it}, captures variation over time in the utility of all papers. This could be caused by shocks to the overall quality of news—i.e. days when there is a major event such as September 11—as well as variation in individuals' cost of time. To represent these factors as simply as possible, I assume that τ_{it} has a two-point discrete distribution—it is equal to zero with some probability $(1-\gamma)$ and equal to some value $\tau > 0$ with probability γ. The final component, ϵ_{irt}, is time-varying utility for each bundle, which will rationalize all remaining day-to-day variation in choices. I assume that the ϵ_{irt} are distributed 2^J-dimensional type-I extreme value.[24]

The choice set includes three goods: the Washington Post print edition, the post.com, and the Washington Times. The vector x includes the following characteristics: age, sex, education, industry of employment, employment status, log income, political party, date of survey, location within the DMA, and number of variables on the questionnaire coded as missing.[25] Dummy variables for web access at work and use of the web for e-mail, chatting, research, and work-related tasks are allowed to enter utility of the post.com but are excluded from the utility of the two print editions.

Two additional variables define whether or not either print or online newspapers were part of a consumer's choice set. For print newspapers, I use data from the question "What newspaper sections do you generally look at?" For the approximately 10% of consumers who answered that they did not read any newspaper sections, I assume that print newspapers do not enter their choice set.[26] Similarly, I assume that if a consumer does not report using the Internet at all during the last thirty days, online newspapers are not part of his choice set. I include these variables because I want to condition on any characteristics that affect demand for newspapers and are unlikely to change

[23] Note that the variance of the errors will be pinned down by the distribution of the ϵ_{irt} so it is not necessary to impose a normalization on the covariance matrix of the ν_{ir}.

[24] I show below that the ϵ_{irt} account for a small portion of the estimated variance in utility, and so the strong assumptions of the extreme value distribution are unlikely to bias the results substantially .

[25] The education dummies are completed high-school, completed college, and completed a post-graduate degree. Occupation dummies are for white-collar workers and computer workers. Employment status is a dummy for being employed full-time. Political party includes dummies for being a registered Republican or Democrat. Dummies for survey dates are each six month interval between March, 2000 and February, 2003. Dummies for location are based on six county-groups comprising the District of Columbia proper, the metro area, and non-metro surrounding areas. The number of missing observations ranges from 0 to 7 out of approximately 65 questions.

[26] Of consumers who reported reading no newspaper sections, 95 percent also reported reading no print newspapers in the last week. The small number who reported reading no sections but did report reading some paper are dropped from the sample.

over the one-week horizon of the data. To allow for the possibility that unobserved characteristics might be correlated with the choice set measures, I include dummies for each as a control in the utility of the goods it does not affect directly (i.e. the "no print" variable enters utility of the online paper and the "no online" variable enters utility of both print papers).[27]

Recall that the survey includes summary data on consumption over a period of five weekdays. In light of Hendel (1999), a natural question is how many independent choices the five-day period represents. In other words, how frequently does the consumer receive a new draw on the time-varying unobservables τ_{it} and ϵ_{irt}. I assume that the number of independent choices over 5 days ranges from one to five and has a binomial distribution with parameter π. This could be interpreted to mean that on each day with probability π the consumer draws new unobservables while with probability $(1-\pi)$ she simply repeats the previous day's choice.[28]

The model will generate probabilities of each possible choice on each day. I sum over these probabilities to match the form of the data: a dummy variable for whether each product was consumed in the last twenty-four hours and a separate dummy for whether or not it was consumed in the last five weekdays. To write this formally, denote the vector of all model parameters by θ, and denote the one-day choice probabilities and distribution of ν conditional on parameters by $Q_{ir}(p,\nu;\theta)$ and $F(\nu;\theta)$ respectively. Let $q^s \in \left\{0,1,...,2^J\right\}^s$ be a vector indicating the bundle chosen on each of s days. Let $\mathbf{r} = \left(r^{24}, r^5\right)$ be a consumer's reported consumption over the twenty-four-hour and five-day windows, and let $\Omega\left(\mathbf{r},s\right)$ be the (possibly empty) set of q^s that are consistent with reported consumption $\mathbf{r}$.[29] The probability of observing consumer i choose $\mathbf{r}$ is then:

$$P_{i\mathbf{r}}(p,\nu;\theta) = \sum_{s=1}^{5} b\left(s-1;\pi\right)\left[\sum_{q\in\Omega(\mathbf{r},s)}\prod_{t=1}^{s} Q_{iq_t}(p,\nu;\theta)\right],$$

where $b\left(s;\pi\right)$ is the binomial coefficient $\binom{4}{s}\pi^s\left(1-\pi\right)^{4-s}$. The number of trials is four rather than five because I assume the data always comes from at least one independent choice.

[27] Otherwise, substitution patterns would also be identified by variation in the choice set. While this form of identification is valid in many settings, it would be hard to argue here that the choice set variation is exogenous and uncorrelated with other unobserved tastes. It therefore seems safer to include the choice set dummies as controls.

[28] The π parameter will ultimately be identified by the degree of consistency of each consumer's choices over time. Note that a high variance of the components of ν_i would also imply greater degree of day-to-day consistency. While the two are probably not separately identified nonparametrically, the fit of the model is improved by combining them to produce a more flexible parametric structure.

[29] For example, suppose the consumer reported reading no newspapers in the last 24 hours and only the Post print edition at some point in the last 5 weekdays. Suppose 1 is the index of the bundle containing only the post. Then $\Omega\left(\mathbf{r},3\right)$ would be $\{[1\ 0\ 0],[0\ 1\ 0],[1\ 1\ 0]\}$ and $\Omega\left(\mathbf{r},1\right)$ would be the empty set.

4.2 Supply

One of the challenges of the current application is that limited variation in the prices of the Post and the Times, and the zero price of the post.com, make it infeasible to estimate the price coefficient α directly. I will instead use industry information to approximate advertising revenue and cost parameters, and then use the firm-side pricing equation for the Washington Post to back out α.

One complication is how to treat the fact that the Post is sold both as single copies and by subscription. I will abstract from the dynamic choice problem this implies for consumers, and assume each pays a single price per copy (as noted above, the effect of subscriptions on the choice to read or not will be captured by the ν_r unobservables). The cover price of the Post increased from \$.25 at the beginning of the sample period (late 2000) to \$.35 at the end of the period (early 2003). The subscription price increased from \$11.16 to \$12.60 for 4 weeks of daily home delivery. Pricing out Sunday editions at their cover price of \$1.50, the implied price per daily copy ranged from \$.22 to \$.25. I will use the average cover price of \$.30 in the estimation. In the final section, I show that the results are robust to varying the assumed price between \$.25 and \$.35.[30]

Another distinction that has not arisen yet, but will be important here, is the difference between the readership of the paper and the number of copies sold. Comparing the micro data used in this study (which measures readership) to circulation figures from the Audit Bureau of Circulations suggests that the average issue of the Post is read by 2.4 adults. This may be accurate, considering the large fraction of papers that are delivered either to multi-occupant households or to businesses. Alternatively, it may reflect some over-reporting in the survey data. It turns out that under plausible assumptions, both scenarios would have the same implications for the model.[31] For simplicity, I will therefore describe the setup of the model assuming there is no over-reporting.

I suppose that firm costs are made up of a fixed first-copy cost for both the print and web editions. I assume that the marginal cost of an additional web reader is zero in a reasonable neighborhood of the observed number of readers, and the marginal cost of either print edition is constant. Finally, I will use a highly simplified model for advertising demand, namely a constant revenue-per-reader for both print and online. Putting these together, I can specify profits for the

[30] Because there is nothing in the data that would allow me to separate these price changes from demand shocks that occurred over the same period, I do not use this price variation to identify substitution patterns. In the results section below, however, I show that the reactions to these changes that the model would predict match the actual changes in the time series quite closely.

[31] In either case, we would assume that the revenue derived from a single reported reader is price minus marginal cost divided by λ (the number of readers per paper). In the former case, a natural assumption would be that the cost of the issue is shared equally among all readers, meaning the the consumer pays p_j/λ rather than p_j. This will mean the estimated α is scaled up by λ, and thus the final utility estimate is scaled down by the same factor. In the latter case, we would assume that each reader pays p_j, but we would want to divide the final welfare estimate by λ to take account of over-reporting. The final answer would thus be the same in either case.

Washington Post Company:

$$\Pi(p) = a_p N_p(p,\alpha) + \frac{(p-c_p)}{\lambda} N_p(p,\alpha) + a_w N_w(p,\alpha) - \Psi \tag{6}$$

where p is the print edition's price; a_p and a_w are advertising revenue per reported reader for print and online respectively; $N_p(p,\alpha)$ and $N_w(p,\alpha)$ are the number of print and online readers (as a function of price and the price coefficient);[32] c_p is the marginal printing and distribution cost per copy; λ is the number of readers per copy; and Ψ is the fixed first copy cost. Based on the assumption that λ represents multiple readership, I assume that the price which enters utility in Equation 5 is p_j/λ. In the results section, I will plug in demand estimates from the model and approximations of the marginal advertising revenue and marginal cost terms taken from industry data. The first order condition for maximization of $\Pi(p)$ then uniquely defines the price coefficient α.

4.3 Estimation

Given characteristics (x_i) and the observed choices for each consumer ($\mathbf{r}_i$), a natural way to estimate θ would be to find the value that maximizes the log-likelihood:

$$L(x,\mathbf{r},p,\theta) = \sum_i \ln \int_\nu P_{i\mathbf{r}_i}(p,\nu;\theta)dF(\nu;\theta) \tag{7}$$

As usual with a random-coefficients model, however, the integral in equation 7 does not have a closed-form solution. I will therefore use simulation draws on the distribution of the ν_i to form consistent estimates of these probabilities. The simplest way to do this would be to average the conditional probabilities, $P_{i\mathbf{r}_i}(p,\nu;\theta)$, over S draws ν_{is} from the multivariate normal distribution of ν. In the actual estimation, I use an importance-sampling variant of this simulator to generate approximations of the true integral. The simulator re-weights the normal distribution to increase the likelihood of drawing ν_{is} for which the observed choices are relatively likely.[33]

[32] Note that $N_w(p,\alpha)$ may include readers both inside and outside of Washington DC. I will use the Post's own estimate of the fraction of readership outside of DC (about 25 percent) to scale up the survey-based estimate of DC readership to total readership.

[33] To take importance-sampling draws for a particular consumer i, I use a preliminary estimate of the parameters, θ_0, to generate a simulated approximation of the probability of i's observed choice $\mathbf{r}$: $\hat{P}_{i\mathbf{r}}(p;\theta_0) = \int_\nu P_{i\mathbf{r}}(p,\nu;\theta_0)dF(\nu;\theta_0)$. Since this is calculated only once for each i, it can be approximated with a large number of simulation draws. I then draw ν_{is} from the distribution:

$$h(\nu|i,\mathbf{r},\theta_0) = \frac{P_{i\mathbf{r}}(p,\nu;\theta_0)f(\nu;\theta_0)}{\hat{P}_{i\mathbf{r}}(p;\theta_0)}. \tag{8}$$

The simulated probabilities are formed as a weighted average over S draws from $h(\nu|i,\mathbf{r},\theta_0)$:

$$\bar{P}_{i\mathbf{r}}(p;\theta) = \frac{1}{S}\sum_s P_{i\mathbf{r}}(p,\nu_s;\theta)W(i,\mathbf{r},\nu_s;\theta_0) \tag{9}$$

An extensive literature considers techniques for constructing simulation estimators in discrete choice models. The two leading approaches are the simulated maximum likelihood (SML) estimator first discussed by Lerman and Manski (1981) and the method of simulated moments (MSM) proposed by Pakes and Pollard (1989) and McFadden (1989). The main advantage of the latter is that it is consistent for a fixed number of simulation draws as N goes to infinity, while the former also requires the number of simulation draws, S, to approach infinity with $\sqrt{S}/N = O(1)$. On the other hand, efficiency of MSM requires the optimal instruments to be calculated exactly, and its performance can be quite poor when the preliminary parameter estimate used to generate the instruments is inaccurate (see Gourieroux and Monfort 1996, 43-44, for a discussion). Also, because the MSM estimator requires calculation of the probabilities and derivatives of all possible observed choices, it is computationally costly in cases where the choice set is large—a situation especially likely to arise with panel data. In preliminary tests, I estimated both models and found the parameters did not differ dramatically. Since computation of MSM took ten to twenty times longer than computation of SML, I will use SML for the final estimation. The SML estimator is:

$$\hat{\theta}^{SML} = \arg\max_{\theta} \left\{ \sum_i \ln \left[\frac{1}{S} \sum_{\nu_s} P_{i\mathbf{r}_i}(p, \nu_s; \theta) \right] \right\} \tag{10}$$

The final estimates are based on 300 simulation draws and adjust equation 10 slightly to incorporate the first-order bias correction suggested by Gourieroux and Monfort (1996, 45).[34] I calculate standard errors using the robust asymptotic approximation proposed by McFadden and Train (2000). Standard errors for welfare estimates and other statistics computed from the model in later stages are estimated by taking draws on the estimated distribution of $\hat{\theta}^{SML}$, computing the statistic in question at each draw, and calculating the sample variance.

where $W(i, \mathbf{r}, \nu; \theta_0) = \hat{P}_{i\mathbf{r}}(p; \theta_0)/P_{i\mathbf{r}}(p, \nu; \theta_0)$. Drawing from $h()$ is simple since it is just the distribution of ν conditional on choice $\mathbf{r}$, and draws can thus be taken using an acceptance-rejection method.

[34] The bias correction can be derived as follows. Let $\tilde{f}$ be the simulated value of a choice probability and let f be the true value. Bias in SML arises because even though $E(\tilde{f}) = f$, it is not true that $E(\ln \tilde{f}) = \ln f$. A second order expansion of $\ln \tilde{f}$ around f shows

$$E \ln \tilde{f} = \ln f - \frac{1}{2} E \frac{(\tilde{f} - f)^2}{f^2}.$$

This suggests replacing the log term in Equation 10 (i.e. $\ln \tilde{f}$) with a consistent estimator of $\ln \tilde{f} + (1/2)E\ (\tilde{f} - f)^2/f^2$. The bias-corrected estimator is thus:

$$\hat{\theta}^{SML} = \arg\max_{\theta} \sum_i \left\{ \ln \left[\frac{1}{S} \sum_{\nu_s} P_{i\mathbf{r}_i}(p, \nu_s; \theta) \right] + \frac{1}{2} \frac{\sum_{\nu_s} \left[P_{i\mathbf{r}_i}(p, \nu_s; \theta) - \frac{1}{S} \sum_{\nu_s} P_{i\mathbf{r}_i}(p, \nu_s; \theta) \right]^2}{\left[\frac{1}{S} \sum_{\nu_s} P_{i\mathbf{r}_i}(p, \nu_s; \theta) \right]^2} \right\}$$

5 Results

5.1 Demand parameters

Table 5 displays SML estimates of the coefficients on observable characteristics. Most of the coefficients in the utility of the Post and post.com are significant, as are about half the coefficients in the utility of the Times. On the whole, the results correspond closely to expectations.

The coefficients in the utility of the Post are consistent with its reputation as a relatively high-brow, liberal newspaper. Education and income both have a positive and significant effect. Considering a consumer with characteristics at the mean of the data, college attendance, graduate school attendance, and doubling household income increase the probability of choosing the Post by 32, 35, and 9 percentage points respectively. Being a registered Democrat increases the probability by 3 points on the margin, while being a registered Republican reduces the probability by 2 points. Age has a positive impact, with an additional ten years of age adding 8 points to the probability. Being employed full-time decreases the probability of choosing the Post by 14 points, having a white-collar job not related to computers decreases it by 6 points, and having a computer-related job decreases it by 8 points.

The coefficients of post.com utility are generally of the same sign as the Post coefficients, though their magnitudes in terms of marginal effects are smaller (reflecting the lower probability of choosing the post.com overall—16% versus 53% for the Post). Two notable exceptions are age and employment: adding ten years of age *decreases* the probability of a mean consumer choosing the post.com by 1 percentage point; being employed full-time increases the probability by 2 points, having a white-collar job increases it by 1 point, and having a computer-related job increases it by 3 points. The age coefficient is consistent with a widely cited belief in the industry that the online edition has the potential to reach out to younger consumers. The employment coefficients reflect the fact that the online edition is frequently read at work.

The coefficients in the utility of the Times are consistent with its reputation as more conservative than the Post. Registered Republicans are significantly more likely to choose the Times, with a marginal effect of 3 percentage points, which is large viewed relative to the average probability of choosing the Times, which is 8 percent.

Tables 6-8 show estimates of the other parameters in the model. Table 6 shows coefficients on the variables that were excluded a priori from utility of the print editions, all of which have the predicted signs, and half of which are significant. Table 7 shows the values of the interaction terms (Γ) and the non-linear parameters (τ, γ, and π). Both of the print-online interaction terms are significantly negative. The Post-Times term is negative but not significant. This is implies that all products are likely to be substitutes, consistent with the IV reduced-form regressions; the substitution patterns are discussed in more detail below. The τ and γ parameters are equal to 6.7 and .06 respectively. This means that on 6% of days the utility of all three products increases by about 7 points—a natural interpretation is that major news stories which occur relatively rarely increase demand for news overall. The π parameter is equal to .95, meaning that with 95% probability the consumer draws a new set of time-varying unobservables. In other words, the weekly data is best fit as the product of four or five independent choices. Finally, table 8 shows that the mean quality of the post.com is estimated to have increased steadily over the sample period.

Table 9 shows the sample covariance of the estimated utility from observables, the estimated covariance matrix of the unobservables ν_i, and the variance of the τ_i and ϵ_i unobservables. The correlation of utilities for the Post and post.com, and for the post.com and the Times, are more positive than the observables alone would predict. The variance of the unobservable component of Times utility is quite large, reflecting the fact that Times consumption is more consistent over days than the small probabilities from the model without unobservables would allow. Finally, comparing the relative magnitudes of the different utility components shows: (i) most of the variance is explained by observables and the ν_i; (ii) the variance of the ν_i is substantial relative to the observables; (iii) the τ_i unobservables contribute relatively little to the variance (although the difference in utility on the occasional "major news days" when τ_i is positive is substantial); and (iv) the role of the ϵ_i is quite small, meaning the usual concerns about incorporating logit errors in estimating the value of new goods are not likely to be a major issue.

The overall fit of the model can be measured in a number of ways. One is to look at the fit of the aggregate predictions. A simulation of choices from the model matches the aggregate shares closely: the overall MSE is .0014. We can also look on a more micro level at how well the model is able to fit consumer heterogeneity. One way to see this is to compare the predicted probabilities of a particular choice for consumers who actually consumed that choice to the predicted probabilities for those who did not. The predicted probability of those who made a particular choice is on average 2.6 times as great as the probability for those who did not make that choice.

5.2 Supply parameters and price coefficient

To calculate the marginal advertising revenue a_p and a_w, I assume that the quantity of advertising space both in print and online is fixed (at least over small variations in readership), and the print and online advertising markets are competitive with a fixed price per-reader per-day. Note that this abstracts from many important features of the advertising market, including differential values for different types of consumers, the extent to which the same reader on two consecutive days is valued differently from two different readers, and possible market power of the Post.

I will estimate the advertising parameters by averaging over observed revenue and readership of the Post for 2001 and 2002. For 2001 and 2002, the Post had total print advertising revenue of \$574.3 million and \$555.7 million respectively (Washington Post Company 2002). Apportioning this by the percentage of circulation accounted for by the daily edition gives \$1.5 and \$1.4 million per day. For daily circulation of 771,614 in 2001 and 768,600 in 2002, we have an average value of $a_p = 1.91$.

Online advertising revenue is not made public, but I employ two sources of information to get a ballpark figure. First, total revenue for the Post's online division was \$30.4 million and \$35.9 million for 2001 and 2002 respectively (Washington Post Company 2002), of which the majority was revenue for the post.com. The average daily online readership of the post.com within the Washington DC DMA (estimated from the Scarborough data) is 450,457; Post financials state that roughly 75% of the post.com's total readership is within the DMA. Dividing 75% of the average online division revenue by the DC readership gives a per-reader advertising revenue of $a_w = .151$. Second, Competitive Media Research tracks online advertising spending for major websites. While they do not track the post.com, they provided me with an estimate of June 2001 advertising revenue for the New York Times online edition. If we assume that this month was representative, nytimes.com revenue for 2001 was \$32.4 million, which, combined with nytimes.com readership statistics from Media Metrix, yields $\alpha_w = .160$.[35] I will use the .151 figure in estimation, and then check the robustness of the results to apportioning a smaller fraction of post.com revenue to online advertising.

The next parameter is the marginal cost of printing and distribution for an additional print copy. According to industry sources, the largest component of marginal cost is newsprint. The Post's average annual newsprint consumption in 2001-2002 was 226,796 metric tons (Editor and Publisher 2001), and the average price was \$541 per metric ton (Bureau of Labor Statistics 2005). Using these figures along with the Post's circulation and average number of pages per issue, I estimate that the newsprint cost was 37 cents per daily copy. Considering that there are additional marginal costs of ink and distribution, I will estimate $c_p = \$.40$ and check robustness to values ranging from \$.30 to \$.50.

The final parameter is the Post's price. As discussed above, I will abstract from the details of subscription and single-copy pricing and assume that each reader faces the same per-copy price, and use the average cover price of \$.30 in the estimation. Below, I show that the results are robust to varying the assumed price between \$.25 and \$.35.

Combining the price, marginal cost, and advertising numbers, I estimate that the Washington

[35] Competitive Media Research reports that June 2001 advertising revenue for the nytimes.com was \$2,701,085. Media Metrix reports the number of unique readers of the nytimes.com in July 2001 at 5,034,000. Assuming that the ratio of monthly to daily readership at the nytimes.com was the same as at the post.com, this implies a daily nytimes.com readership of 562,883 (this is probably a lower bound, since the daily figure used for the post.com only counts readers in the DC metro area, and so overstates the monthly-daily ratio).

Post Company's marginal profit per day from an additional print reader is \$.75. I then calculate the price coefficient directly from the optimality condition obtained by setting the derivative of equation 6 with respect to p equal to zero. Plugging the above values into the first order condition yields a point estimate of α equal to -10.57.

5.3 Substitution patterns

One of the main questions that motivated this paper was whether the introduction of the post.com has had a significant crowding out effect on demand for the Post. The reduced-form results provided some evidence: in a simple OLS linear probability model, all pairs of goods appeared to be strong complements; in the IV specification, the sign of the Post-post.com interaction switched, suggesting substitutability. We can now see how these conclusions hold up in the full model.

The key findings are summarized in table 10. The table shows three indicators of the degree of complementarity or substitutability between the print and online editions of the Post: (i) the cross-price derivative of demand; (ii) the change in Post readership when the post.com is removed from the choice set; and (iii) and the cost of this lost readership in terms of profits from the print edition. The first panel shows the results when the model is estimated with no consumer heterogeneity, the second panel shows the results for a model with observable consumer characteristics only, and the final panel shows the results for the full model.

The table suggests two key conclusions. The first is that the print and online papers are significant substitutes. I estimate that a \$.10 (33%) increase in the price of the Post would lead to an increase in post.com readership of 9,990 (2%). Comparing actual Post demand with the counterfactual simulation in which the post.com is removed from the choice set, I find that introducing the post.com reduces Post readership by about 30,000 readers per day and reduces Post profits by approximately \$5.7 million per year. These effects are all precisely estimated and significantly different from zero. Their magnitude is moderate, however, and certainly does not suggest anything close to a one-for-one crowding out of print readership. The second conclusion is that properly accounting for observed and unobserved consumer heterogeneity is critical to obtaining accurate results. A model that includes consumer observables but no correlated unobservables (i.e. the ν_r and τ_{tr} are fixed at zero) would lead us to conclude that the products are essentially independent: the cross-price derivative is not significantly different from zero, the change in Post readership is only 6% of what we estimated was its true value, and the effect on Post profits is only \$340,000. A model with neither observed nor unobserved heterogeneity would lead us to conclude that the products are strong complements, with introducing the post.com *increasing* Post profits and readership by a significant amount.

In table 11, I report all of the estimated price derivatives of demand. The own-price derivatives show that a \$.10 increase in own price would lead to a loss of 251,343 Post readers, 127,365 post.com readers, and 58,734 Times readers. This is equivalent to 14% of Post readership, 27% of post.com readership, and 23% of Times readership, and implies own-price elasticities of .41 and .57 for the Post and Times respectively (we cannot calculate an elasticity for the post.com since its price is zero).[36] Note that the fact that advertising revenue per copy exceeds marginal cost implies an equilibrium elasticity of less than one when firms set prices optimally.

One way to cross-check the validity of the estimated elasticities is to compare them to an aggregate time series of the Post's price and readership over a longer period. I calculate a price series for 1994-2003 using a weighted average of the single-copy and subscription prices.[37] I then calculate counterfactual readership, beginning in 1994 and assuming no change in demand other than reaction to price changes at the estimated elasticity of .41. The results of this simulation are shown in figure 3. The fit of the counterfactual is good, implying that the estimated elasticity is in the right ballpark. The fact that the counterfactual slightly over or under-predicts the demand change in some years probably reflects the fact that other demand shocks unrelated to price were occurring over the period.

The relationship of both the Post and the post.com to the Times is weaker than their relationship to each other. Both pairs of products are substitutes, but the Post-Times elasticity is not significantly different from zero and the post.com-Times elasticity implies that a \$.10 increase in the price of the Times would increase post.com readership by 3,424.

5.4 Welfare analysis: consumer surplus

From the estimated demand parameters, it is a straightforward exercise to calculate the welfare effects of an individual product. For clarity in what follows, I denote the portion of utility excluding the ϵ_{irt} by $\hat{u}_{ir}(\nu, \tau)$. Let the space of possible bundles be $\mathcal{R} = \{0,1\}^J$ and let $\mathcal{R}^{-j} \subset \mathcal{R}$ be the set of bundles that do not include good j. We can then calculate consumer i's expected gain from adding good j to the choice set as:

$$\Delta_i^j = E_{\nu,\tau,\epsilon}\left[\max_{r\in\mathcal{R}} \hat{u}_{ir}(\nu,\tau) + \epsilon_{irt}\right] - E_{\nu,\epsilon}\left[\max_{r\in\mathcal{R}^{-j}} u_{ir}(\nu,\tau) + \epsilon_{irt}\right], \tag{11}$$

Note that this is the expected value from the perspective of the econometrician, given all observed data. Since ν is constant for a given consumer over time, the choice observed in the data will contain information about the expected value of ν, and the expectation should therefore integrate over ν conditional on i's choice. The shocks ϵ and τ, on the other hand, are i.i.d. across consumers

[36] Since the Post cover price was \$.25 in the first half of the sample period and \$.35 thereafter, the Post elasticity is calculated using the average of the two prices.

[37] In 2001, single copies accounted for 24% of Post sales (Ahrens 2001).

and periods, and observed choices contain no information about its expected value in the future. Thus, the integral over ϵ and τ should be unconditional. A standard result on the expectation of the maximum over extreme value errors allows us to re-write the expectation in equation 11 as:

$$E_{\nu,\tau,\epsilon}\left[\max_{d\in\mathcal{C}} \hat{u}_{ir}(\nu,\tau)+\epsilon_{irt}\right] = E_{\nu,\tau}\left[\ln\sum_{d\in\mathcal{C}} e^{\hat{u}_{ir}(\nu,\tau)}\right]. \tag{12}$$

We can estimate the expectation over ν and τ by simulation:

$$E_{\nu,\tau}\left[\ln\sum_{d\in\mathcal{C}} e^{\hat{u}_{ir}(\nu,\tau)}\right] \approx \sum_{s=1}^{S}\left[\gamma\ln\sum_{d\in\mathcal{C}} e^{\hat{u}_{ir}(\nu,\hat{\tau})} + (1-\gamma)\ln\sum_{d\in\mathcal{C}} e^{\hat{u}_{ir}(\nu,0)}\right] R(\nu_s|r) \tag{13}$$

where d_i is i's observed choice, $\hat{\tau}$ is the fitted value of τ, and each ν_s is an independent draw from the estimated (unconditional) distribution of unobservables. The conditional probability of ν_s given the observed choice r is:

$$R(\nu_s|r) = \frac{Q_{ir}(p,\nu_s)f(\nu_s)}{\bar{Q}_{ir}(p)}. \tag{14}$$

Here, $\bar{Q}_{ir}(p)$ is a consistent estimator of the unconditional expectation of $Q_{ir}(p,\nu)$ over ν which can easily be generated by simulation. The figures in utils can be converted to dollars by multiplying by the price elasticity and then aggregating to find the total change in consumer surplus:

$$\Delta CS(j) = \sum_i \alpha\Delta_i^j. \tag{15}$$

A number of caveats should be emphasized about the validity of these figures. First, we don't actually see prices varying over a large range so the estimates depend heavily on the assumption of quasi-linearity that allows utils to be converted into dollars at a constant rate. It is variation in consumer characteristics that allows us to map out the demand curve against the scale given by the normalized ϵ_{it} distribution. Alternative error distributions, or specifications that allowed the variance of ϵ_{it} to differ across bundles, could yield different results. Second, the analysis here ignores the welfare of consumers who may read the post.com outside of the Washington DC area. Finally, the model does not take account of changes in the prices or characteristics of other products in response the post.com's introduction. If the added competition from the post.com caused the quality of competing papers to be improved, for example, we would understate the consumer welfare gains.

The results of the consumer surplus calculation are presented in Table 12. We can get some idea of the welfare effects of the post.com independent of the estimated price coefficient by looking at the surplus to post.com readers relative to the surplus of print readers. I find that the average post.com reader's surplus is 70% of the average print reader's surplus. Aggregating over readers, this means that the total consumer welfare gain from the post.com is 16% of the total print surplus. In dollars, I find that the average post.com reader would be willing to pay $.28 per day. The total consumer surplus gain from the post.com is $114,959 per day, or $42 million per year.

5.5 Welfare analysis: producer surplus

The calculation of producer surplus is also straightforward. Given estimates for the advertising and marginal cost parameters, the gain to a given company from the addition of a particular product can be calculated by simulating the change in demand when that product is removed from the choice set, and calculating the resulting change in profit from equation 6.

Similar caveats to the above will apply here. Once again, the results will be sensitive to the stylized supply-side model, and especially the values used for advertising revenue. Furthermore, the producer surplus calculations only apply to the welfare of the Post and Times companies. This neglects a number of other firms that are affected by the post.com. The most notable example is advertisers. Other examples include other newspaper companies, television and radio companies, and other online news providers.

In table 13, I present a more detailed breakdown of the counterfactual changes in readership when the post.com is removed from the choices set. As already mentioned, this would increase Post readership by about 29,000 per day. The bottom rows of this table show the change in the choices of two specific groups of consumers: those who read both the post.com and the Post and those who read the post.com alone. Of consumers in the first group, almost 80% would switch to reading the Post alone if the post.com were removed, while a smaller number would stop reading entirely.[38] Of consumers in the second group, 27% would start reading the Post if the post.com is removed, while almost 70% would switch to the outside good.

These demand effects are translated into dollars in table 14. I estimate that the introduction of the post.com cost about $5.7 million a year in lost print profits, compared to the $33.2 million dollars of revenue generated directly. The effect on the Washington Times is smaller, with the customers who substitute to the post.com decreasing profits by about $2.8 million. The total gain in producer surplus from the online edition (before taking account of post.com operating costs) is thus $24.7 million.

A final question that the structural model allows us to ask is how profits of the Post company would change if it charged a positive price for the online edition. The direct revenue gains would be

[38] To understand why some readers switch to the outside good, recall that the ϵ_{irt} unobservables make the interaction of goods in utility heterogeneous. This means that for some consumers the products are actually complements, even though they are substitutes in aggregate demand.

offset by reduced advertising revenue. Based on the results above, they would also be bolstered by an increase in print readership. The net effects of these different factors are summarized in figure 4, which shows that the estimated profit-maximizing price of the online edition is $.20 per day, or about $6 per month. The loss from charging a price of zero is approximately $9 million, or more than a quarter of post.com advertising revenue. The $6 per month estimate is close to the prices of the few online newspapers that do charge for access: in 2003, the online subscription prices for two such papers, the Wall Street Journal and the Tulsa World, were $6.95 and $3.75 respectively.

A common explanation for the frequency of zero prices online is that consumers and firms face real or perceived transactions costs. Real costs could arise on the consumer side from the hassle of entering a credit card number or fear of online fraud. For firms, they would be primarily the cost of processing credit card transactions. Consumers might have additional perceived costs if an online newspaper charging positive prices deviates from the prevailing norm and is thus seen as unfair. How big would transaction costs have to be to rationalize the observed price? Simulations show that the figure is approximately $.14 per day. Whether this is large or small is a matter of speculation, but it does not seem out of the realm of possibility.[39]

Based on these results, can we understand the decisions of the Post Company to introduce the post.com? The answer, of course, hinges on the operating costs ignored in the previous calculations. While the Post does not break out costs separately for the online division, the New York Times does. Reported costs of New York Times Digital were $63.5 million and $67.7 million for 2001 and 2002 respectively (New York Times Company 2003). Given the larger scale of the New York Times online operations, we could take the cost figure of $60 million as an upper bound on the costs of the post.com. On the other hand, the Post reported in 2001 that its online division is losing money, so a reasonable lower bound would be $30 million. Taking the midpoint of these bounds as a ballpark figure, and combining this with the figures in table 12, the net annual effect of the post.com on producer surplus was a loss of roughly $20 million. This makes the decision seem puzzling.

On the other hand, online revenues almost doubled between 2002 and 2004, driven in large part by an improved market for online advertising. If we project the 2004 revenues back to the 2000-2003 period (assuming that the only change is a higher advertising price per reader), the picture changes substantially. The total gain to the Post Company from introducing the post.com is now estimated to be about $56 million per year, the effect net of operating costs is a gain of $11 million per year, and the total change in producer surplus is an increase of $8 million per year (again assuming $45 million per year operating costs). Furthermore, the more favorable advertising market reduces the incentives to raise the online edition's price: the optimal price is now estimated to be $.09, and the lost profits from charging a price of zero are just $2 million. Although these calculations may

[39] Transaction costs would depend on whether consumers had to pay every day, or could pay at infrequent intervals, as with a subscription.

overstate the improvement in advertising revenues somewhat, they suggest that a large part of the initial losses earned by the online paper can be seen as an investment in anticipation of future market growth, and that a zero price is probably not far from the long-run optimum, especially in the presence of small transactions costs.

5.6 Robustness checks

In table 15, I present results from the model under a number of alternative values for the primitive supply parameters. As mentioned earlier, the values used in the estimation were based on rough estimates from industry data, and it is important to know how sensitive the results are to alternative assumptions.

The first two experiments use values of α derived from the assumption that the price of the Post was \$.25 and \$.35 respectively. Recall that since the former was the price through the end of 2001, and the latter was the price for 2002-2003, the estimates presented earlier were based on the average of these two α values. The table shows that varying α in this range does not substantially change any of the estimates. The next two experiments vary the marginal cost per print copy of the Post between \$.30 and \$.50 (\$.40 was the value used in estimation). The table shows that the changes in this case are slightly larger—the optimal price varies by \$.03, and consumer surplus varies by \$5 million—but none of the qualitative results are affected.

The following row uses a lower estimate of post.com revenue than the \$33 million assumed earlier. The figure used, \$25 million, is a low estimate taking account of the fact that the Post's online division includes several ventures other than the post.com, including the online edition of Newsweek magazine. Predictably, this change does have a significant effect on the estimated surplus from introducing the post.com, reducing it by about a third. Other effects are smaller, however: the estimated optimal price increases by \$.03 and estimated consumer surplus is essentially unchanged.

The final row relaxes the restriction that the Γ_r interaction term in the utility of the three-good bundle is equal to the sum of the component two-good interactions. The resulting changes in the quantities of interest are again small, suggesting that the restriction does not substantially constrain the model's ability to fit the data.

Overall, the results seem robust to varying the supply parameters and form of the utility function, with none of the qualitative results from the model changing under the alternative assumptions considered.

6 Conclusions

This paper began with three questions: Are print and online newspapers substitutes or complements? How might demand respond if papers were to charge positive prices for online content that is currently provided free of charge? And how has the introduction of online news affected the welfare of consumers and newspaper firms?

I find, first, that print and online papers are clearly substitutes. The apparent positive relationship in the data is an artifact of unmeasured consumer heterogeneity, and disappears in the full demand model. The magnitude of the crowding out of print readership is non-negligible. It is also small, however, relative to some earlier predictions. Barring dramatic changes in tastes or technology, online news does not appear to threaten the survival of print media.

Second, the increased revenue the firm could gain by raising online prices would outweigh the loss in readership. The potential gain was relatively large in the period analyzed. However, the gain is much smaller with current advertising rates and would disappear with a small transaction cost of online payments.

Finally, the welfare benefits of the online newspaper appear to outweigh its costs. Consumers gain $42 million a year from free provision of the online paper, and although the firm appeared to suffer a net loss during the 2000-2003 period, an improved advertising market means that the current annual effect on firm profits is probably positive. The net gain to society is small relative to the more enthusiastic predictions of the late 1990's. On the other hand, continuing Internet adoption and technology improvement may make the long-run benefit substantially larger.

The methodological contributions of the paper are to develop a demand model that allows both multiple choices and complementarity, and to clarify the identification of substitution patterns. Although many details of the implementation are specific to the online news setting, I hope that the framework may prove useful in other contexts where standard discrete choice assumptions do not apply.

Appendix: proof of proposition 2

For simplicity I will suppress the dependence of all terms on characteristics x. Assume that the ϵ_{rt} terms are zero, or that we condition on a particular value of ϵ_{rt} and redefine the u and Γ terms accordingly. The utility of a bundle of two goods—say A and B—is then:

$$u_{AB} = u_A + u_B + \Gamma_{AB},$$

and the utility of the bundle of all three goods is:

$$u_{ABC} = u_A + u_B + u_C + \Gamma_{AB} + \Gamma_{AC} + \Gamma_{BC}.$$

Suppose we define

$$\begin{aligned} \tilde{u}_0 &= \max\{u_0, u_C\} \\ \tilde{u}_A &= \max\{u_A, u_{AC}\} \\ \tilde{u}_B &= \max\{u_B, u_{BC}\} \\ \tilde{u}_{AB} &= \max\{u_{AB}, u_{ABC}\}. \end{aligned}$$

Then the same steps used to derive equation 3 imply:

$$\begin{aligned} \frac{Q_A}{dp_B} &= \int_{\mathbf{u}} [I(\tilde{u}_A = \tilde{u}_B) I\left(-\tilde{\Gamma} \geq \tilde{u}_A, \tilde{u}_B \geq 0\right) \\ &\qquad - I\left(\tilde{u}_A + \tilde{u}_B = -\tilde{\Gamma}\right) I(\tilde{u}_A \leq 0) I(\tilde{u}_B \leq 0)] dF_x(\mathbf{u}). \end{aligned} \tag{16}$$

where $\mathbf{u} = [\tilde{u}_0\ \tilde{u}_A\ \tilde{u}_B\ \tilde{u}_{AB}]$ and $\tilde{\Gamma} = \tilde{u}_{AB} - \tilde{u}_A - \tilde{u}_B + \tilde{u}_0$. It is clear, then, that the cross-derivative will be zero if $\tilde{\Gamma}$ is everywhere zero, positive if $\tilde{\Gamma} \geq 0$ everywhere with strict inequality at some region of $\mathbf{u}$ with positive measure, and negative if $\tilde{\Gamma} \leq 0$ everywhere with strict inequality at some region of $\mathbf{u}$ with positive measure.

Substituting in the definitions of the different utilities, we can show that $\tilde{\Gamma}$ is:

$$\tilde{\Gamma} = \Gamma_{AB} + \max\{-u_C, \Gamma_{AC} + \Gamma_{BC}\} - \max\{-u_C, \Gamma_{AC}\} - \max\{-u_C, \Gamma_{BC}\} + \max\{-u_C, 0\}.$$

Call Γ_{AB} the "direct substitution effect" and call the remaining part the "indirect substitution effect." Then given that $F_x(\mathbf{u})$ has full support, the following lemma establishes that the indirect substitution effect has the same sign as the product $\Gamma_{AC}\Gamma_{BC}$.

Lemma 1 *Consider the expression:*

$$\max\{x, a+b\} - \max\{x, a\} - \max\{x, b\} + \max\{x, 0\}.$$

This is

- *zero if* $ab = 0$*;*
- *weakly positive and strictly positive for an interval of* x *if* $ab > 0$;
- *weakly negative and strictly negative for an interval of* x *if* $ab < 0$.

Proof. Assume without loss of generality that $a \geq b$.

If either a or b equals 0 the expression inside the integral is trivially equal to 0, so assume $a, b \neq 0$.

There are three cases (i) $a, b > 0$; (ii) $a > 0 > b$; (iii) $0 > a, b$.

Case (i): If either $x \geq a + b$ or $x \leq 0$ the expression is 0. If $a + b > x \geq a$ the expression is $a + b - x > 0$. If $a > x \geq b$, the expression is equal to $b > 0$. If $b > x > 0$, the expression is equal to $x > 0$. Therefore the integral must be strictly positive.

Case (ii): If either $x \geq a$ or $x \leq b$ the expression is 0. If $a > x \geq \max\{a + b, 0\}$, the expression is $x - a < 0$. If $\max\{a + b, 0\} > x \geq \min\{a + b, 0\}$, the expression is either b, $-a$, both of which are < 0. If $\min\{a + b, 0\} > x > b$, the expression is $b - x < 0$. Therefore the integral must be strictly negative.

Case (iii): If either $x \geq 0$ or $x \leq a + b$, the expression is 0. If $0 > x \geq a$, the expression is $-x > 0$. If $a > x \geq b$, the expression is $-a > 0$. If $b > x > a + b$, the expression is $x - (a + b) > 0$ Therefore the integral must be strictly positive. ■

References

[1] Ahrens, Frank. 2001. The Post to raise paper's price to 35 cents a copy. *Washington Post.* December 14.

[2] Augereau, Angelique, Shane Greenstein, and Marc Rysman. 2004. Coordination vs. differentiation in a standards war: 56K modems. Boston University mimeo.

[3] Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. Automobile prices in market equilibrium. *Econometrica* 63 (4): 841-90.

[4] Blackburn, David. 2004. Online piracy and recorded music sales. Harvard University mimeo.

[5] Bureau of Labor Statistics. 2005. Commodity price index. Series ID WPU09130291. Retrieved 5 March, 2005 from www.bls.gov.

[6] Chan, Tat Y. 2004. Estimating a continuous hedonic choice model with an application to demand to soft drinks. Washington University mimeo.

[7] Chiou, Lesley. 2004. Empirical analysis of retail competition: Spatial differentiation at Wal-Mart, Amazon.com, and their competitors. Massachusetts Institute of Technology mimeo.

[8] Christensen, Laurits R., Dale W. Jorgenson, and Lawrence J. Lau. 1975. Transcendental logarithmic utility functions. *American Economic Review* 65(3): 367-83.

[9] Dube, Jean-Pierre. 2004. Multiple discreteness and product differentiation: demand for carbonated soft drinks. *Marketing Science* 23(1): 66.

[10] Editor and Publisher. 2001. *Editor & Publisher International Yearbook.* New York: Editor and Publisher Co.

[11] Gandal, Neil, Sarit Markovich, and Michael Riordan. 2003. Ain't in 'suite'? Strategic bundling in the PC office software market. Columbia University mimeo.

[12] Gates, Dominic. 2000. E-book evangelist: Microsoft's brash Dick Brass leads the revolution that could bury dead trees. *Industry Standard, 25* September. Retrieved 8 November, 2002 from http://www.thestandard.net/article/0,1902,18591,00.html.

[13] ——. 2002. News sites hustle for profitability. *USC Annenberg Online Journalism Review,* July 11, 2002. Retrieved 8 November, 2002 from http://www.ojr.org/ojr/future/1026348638.php.

[14] Gentzkow, Matthew. 2004. Television and voter turnout. University of Chicago mimeo.

[15] Goolsbee, Austan. 2000. In a world without borders: the impact of taxes on Internet commerce. *Quarterly Journal of Economics* 115 (2): 561-576.

[16] ——. 2001. Competition in the computer industry: online versus retail. *Journal of Industrial Economics* 49 (4): 487-499.

[17] Goolsbee, Austan and Amil Petrin. 2004. The consumer gains from direct broadcast satellites and the competition with cable TV. *Econometrica* 72 (2): 351-81.

[18] Gourieroux, Christian and Alain Monfort. 1996. *Simulation-Based Econometric Methods.* New York: Oxford U. Press.

[19] Hendel, Igal. 1999. Estimating multiple-discrete choice models: An application to computerization returns. *Review of Economic Studies* 66: 423-46.

[20] Henry, David. 2000. Is Buffet too quick to write off newspapers? *USA Today*, 4 May.

[21] Kaiser, Ulrich. 2002. The effects of website provision on the demand for German women's magazines. NBER Working Paper #8806.

[22] Keane, Michael P. 1992. A note on identification in the multinomial probit model. *Journal of Business & Economic Statistics* 10 (2): 193-200.

[23] Kim, Jaehwan, Greg M. Allenby, and Peter E. Rossi. 2002. Modeling consumer demand for variety. *Marketing Science* 21 (3): 229-50.

[24] Kirsner, Scott. 1997. Profits in site. *American Journalism Review,* December.

[25] Lerman, Steven R. and Charles F. Manski. 1981. On the use of simulated frequencies to approximate choice probabilities. In *Structural Analysis of Discrete Data with Econometric Applications*, edited by Charles F. Manski and Daniel McFadden. Cambridge, MA: MIT Press.

[26] Manski, Charles F. 1993. Identification of endogenous social effects: the reflection problem. *Review of Economic Studies* 60 (3): 531-42.

[27] Manski, Charles F. and Leonard Sherman. 1980. An empirical analysis of household choice among motor vehicles. *Transportation Research* 14A: 349-66.

[28] McFadden, Daniel. 1989. A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica* 57 (5): 995-1026.

[29] McFadden, Daniel and Kenneth Train. 2000. Mixed MNL models for discrete response. *Journal of Applied Economics* 15: 447-70.

[30] Mortimer, Julie Holland. 2004. Price discrimination and copyright law: Evidence from the introduction of DVDs. Harvard University mimeo.

[31] New York Times Company. 2003. The New York Times Company reports fourth-quarter and full-year results. Press release, 28 January. Retrieved 31 October, 2003 from http://www.nytco.com/press.html.

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

[33] Ogaki, Masao. 1990. The indirect and direct substitution effects. *American Economic Review* 80 (5): 1271-75.

[34] Okrent, Dan. 1999. The death of print? Columbia University lecture, 14 December. Retrieved 8 November, 2002 from http://www.digitaljournalist.org/issue0002/okrent.htm.

[35] Owen, Bruce M. 1999. The Internet challenge to television. Cambridge and London: Harvard University Press.

[36] Pakes, Ariel and David Pollard. 1989. Simulation and the asymptotics of optimization estimators. *Econometrica* 57 (5): 1027-57.

[37] Rob, Rafael and Joel Waldfogel. 2004. Piracy on the high C's: Music downloading, sales displacement, and social welfare in a sample of college students. University of Pennsylvania mimeo.

[38] Runett, Rob. 2002. Yes circulation directors, they're still buying the paper: Single-copy purchases up during Belden Q3 survey. *The Digital Edge*, January.

[39] Samuelson, Paul A.. 1974. Complementarity: An essay on the 40th anniversary of the Hicks-Allen revolution in demand theory. *Journal of Economic Literature* 12 (4): 1255-89.

[40] Screen Digest. 2002. Mediaphile 2010: UK media market could be worth 84.26 billion pounds by 2010. Retrieved 18 November, 2002 from http://www.screendigest.com/press_mediaphile.htm

[41] Seelye, Katharine Q. 2005. Can papers end the free ride online? *New York Times* 14 March.

[42] 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: 1-24.

[43] Smith, Howard. 2004. Supermarket choice and supermarket competition in market equilibrium. *Review of Economic Studies* 71: 235-63.

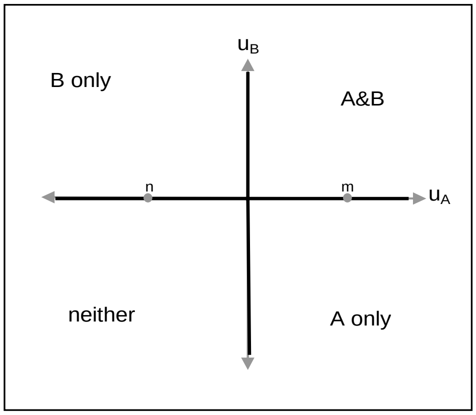
[44] Train, Kenneth E., Daniel L. McFadden, and Moshe Ben-Akiva. 1987. The demand for local telephone service: a fully discrete model of residential calling patterns and service choices. *RAND Journal of Economics* 18 (1): 109-23.

[45] Washington Post Company. 2002. *Annual Report 2002.*

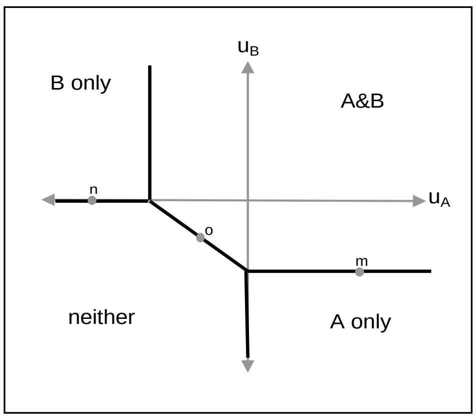
Figure 1

Illustration of substitution patterns in a model with two goods

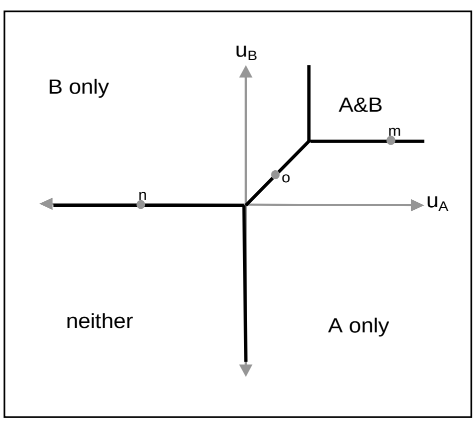
Case 1: Γ=0



Case 2: Γ>0

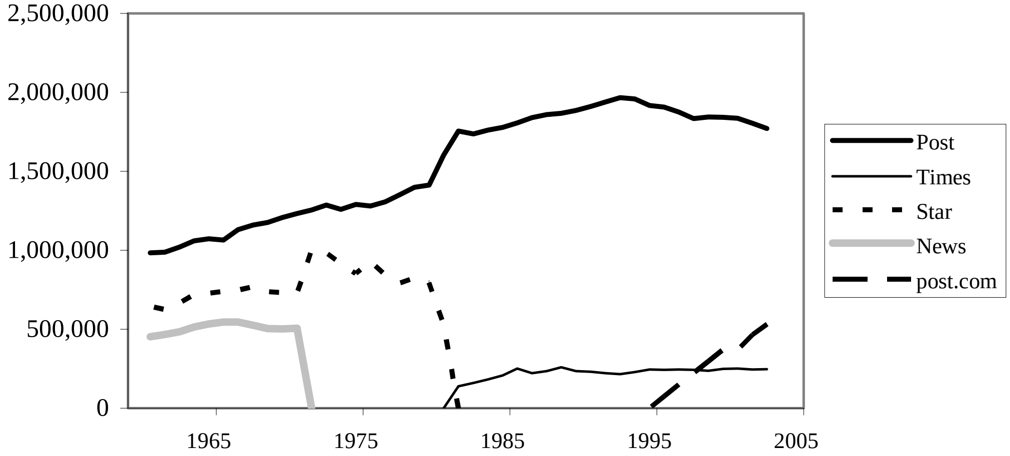


Case 3: Γ<0



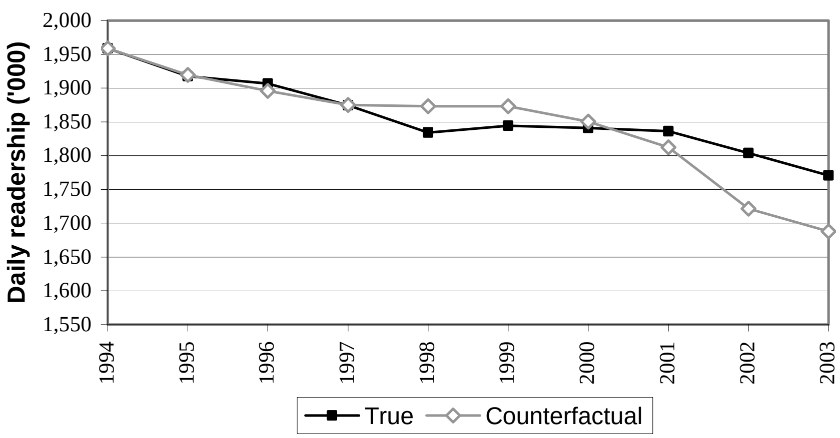
Notes: Figures show the regions of U_A-U_B space in which the consumer would choose the bundles $A\&B$, B alone, A alone, or neither good. The first panel shows the case where the interaction between the two goods in utility is zero, the second panel the case where it is positive, and the third panel the case where it is negative.

Figure 2
Readership of newspapers in Washington DC (1961-present)



Source: Audit Bureau of Circulations; Scarborough Research.
Notes: Readership figures are derived by using historical circulation data and the ratio of readership to circulation in the 2000-2003 Scarborough data.

Figure 3
Simulation of Post readership using estimated own-price elasticity



Source: Audit Bureau of Circulations. Counterfactual readership simulated from estimated model as described in the text.
Notes: True readership is calculated from circulation figures and the ratio of readership to circulation in the 2000-2003 Scarborough data. Counterfactual circulation is calculated by taking 1994 as a base year and finding the percentage change in each year based on actual price changes and the estimated own-price elasticity of .41. This assumes that no other demand shocks occurred over the period. The price series used for the counterfactual is calculated as a weighted average of single-copy and subscription prices.

Figure 4
Yearly Post Company variable profit for different prices of the post.com

Source: Estimated from Scarborough Research data as described in the text.
Notes: The figure shows counterfactual estimates based on the full demand model. The vertical axis shows total variable profits from the Post and post.com, defined to be the sum of circulation and advertising revenue minus marginal printing and distribution costs. This figure does not include any other components of costs such as website or administrative overhead.

Table 1:
Summary statistics

	Scarborough Survey	Washington DC DMA (Census)
N	16,179	4,203,621
Median income	$62,500	$60,774
Black	20.6%	23.5%
Hispanic	6.4%	7.9%
Female	57.9%	52.1%
Age distribution:		
18-29	17.5%	21.3%
30-39	22.6%	23.4%
40-49	22.2%	21.7%
50-59	17.9%	15.8%
60+	19.8%	17.7%
Highest schooling:		
< High school	7.7%	14.4%
High School	47.0%	42.1%
College	27.2%	26.3%
Graduate	18.0%	17.2%

Source: Scarborough Research data; 2000 Decenial Census.
Notes: The Scarborough survey is a randomized sample of residents of the Washington DC DMA eighteen years of age and older. All census figures refer to to the population of individuals eighteen years of age and older, except percent Black and Hispanic, which are proportions of all residents. Median income is the population-weighted mean of the median incomes of counties in the Washington DC DMA.

Table 2
Correlation coefficients

	24-hour		5-day	
	Raw	Partial	Raw	Partial
Post-post.com	.0989**	.0364**	.1579**	.0673**
Post-Times	.0632**	.0035	.0450**	-.0623**
Times-post.com	.0146	.0090	.0184	.0066

Source: Scarborough Research data.
Notes: ** indicates a coefficient significantly different from zero at the .1% level. The table displays correlation coefficients between dummy variables for reading the Washington Post, post.com, and Washington times. In the first two columns, the variable is equal to one if a respondent read in the last twenty-four hours. In the second two columns, the variable is equal to one if a respondent read in the last five weekdays. Partial correlations are correlations in the residuals from regressions of each consumption variable on controls for age, sex, education (four categories), industry of employment (three categories), employment status, income, political party, date of survey, location of residence within the DMA (six categories), and dummy variables for the number of missing values. Observations where either print or online newspapers were not in the choice set (consumer reports that she generally reads no newspaper sections or did not use the Internet in the last thirty days) were dropped.

Table 3
Linear probability model of Post consumption

	OLS	IV (1)	IV (2)	IV (3)
	Dependent variable: read Post print edition last 5 days			
post.com	.0464	-.426	-.348	-.377
	(.0090)	(.106)	(.129)	(.201)
Other Internet news			.0133	.0034
			(.0181)	(.0250)
Detailed occupation controls	No	No	No	Yes
N	14313	14313	14313	10544
R-squared	.333	.208	.246	.204

Source: Scarborough Research data.
Notes: Robust standard errors are in parentheses. The first row gives coefficients on a dummy for reading the Post print edition in the last five weekdays. IV regressions instrument for post.com consumption with dummy variables for Internet access at work, fast Internet connection, and the following reported Internet uses: e-mail, chatting, research/education and work-related tasks. Other Internet news is a dummy for online news use other than online newspapers. Detailed occupation controls are dummies for forty-seven occupation categories. All regressions include controls for Washington Times readership, age, sex, education (four categories), industry of employment (three categories), employment status, income, political party, date of survey, location of residence within the DMA (six categories), and dummy variables for the number of missing values. The regressions omit observations where print newspapers are not in the choice set (consumer reports that she generally reads no newspaper sections) and control for presence of online newspapers in the choice set (whether the consumer used the Internet in the last thirty days).

Table 4

First-stage regressions

	(1)	(2)	(3)
	Dependent variable: read post.com		
Internet at work	.0909	.0722	.0490
	(.0083)	(.0084)	(.0098)
Fast connection	.0218	.01708	.01143
	(.0093)	(.0092)	(.0112)
Use for e-mail	-.00807	-.0315	-.0319
	(.0121)	(.0121)	(.0149)
Use for chatting	.0252	.01667	.00947
	(.0106)	(.0105)	(.0125)
Use for research	.00154	-.00495	-.00517
	(.00994)	(.0099)	(.0118)
Use for work-related	.0206	.0103	.0106
	(.0110)	(.0109)	(.0125)
Other Internet news		.1140	.1084
		(.0074)	(.0089)
N	14313	14313	10544
R-squared	.172	.185	.192

Source: Scarborough Research data.
Notes: Robust standard errors are in parentheses. The table presents first-stage coefficients for the IV regressions in table 3. The regression includes all the controls described in the notes to table 3. Fast connection indicates consumers with DSL, cable modem, or T1 connections at home. Use variables were responses to the question "In what ways do you use online services?" Other Internet news is as in table 3.

Table 5
Parameter estimates from full model: observable characteristics

	Post	post.com	Times
Age	.667	-.528	.700
	(.0410)	(.0696)	(.112)
Female	-.505	-.453	-3.23
	(.0942)	(.152)	(.326)
High school	2.01	2.98	1.24
	(.214)	(.635)	(.668)
College	2.60	4.13	1.45
	(.236)	(.655)	(.719)
Grad school	2.84	4.23	1.31
	(.253)	(.666)	(.762)
Computer job	-.625	1.21	.449
	(.226)	(.338)	(.676)
White collar job	-.477	.453	-.492
	(.120)	(.198)	(.396)
Full-time	-1.13	.949	-.289
	(.144)	(.235)	(.465)
Log income	.722	.231	.905
	(.0744)	(.125)	(.245)
Democrat	.243	.304	-.087
	(.100)	(.165)	(.337)
Republican	-.192	-.0472	2.787
	(.120)	(.197)	(.391)
Constant	6.78	-2.40	-8.27
	(.908)	(1.24)	(2.06)
N	16179	16179	16179

Source: Scarborough Research data.
Notes: Standard errors in parentheses. Details of the model are given in the text. Age is measured in units of ten years. High school, college, and graduate school are mutually exclusive categories. Computer and white-collar job are dummies for reported occupations in these categories and are also mutually exclusive. Full-time is a dummy for full-time employment. Democrat and Republican indicate registered members of the parties. Additional model parameters are shown in tables 6-8. Not shown in any table are dummies for the number of missing observations, location of the respondent's residence within DC, and having print and online papers in the choice set.

Table 6
Parameter estimates: excluded variables

	post.com
Internet at work	1.339
	(.183)
Fast connection	.188
	(.199)
Use for e-mail	.637
	(.336)
Use for chatting	.446
	(.205)
Use for research	.272
	(.213)
Use for work-related	.558
	(.226)

Source: Scarborough Research data.
Notes: Standard errors in parentheses. Each of these variables is excluded from utility of Post and Times print editions. The table gives coefficients in the utility of the post.com. Fast connection indicates consumers with DSL, cable modem, or T1 connections at home. Use variables were responses to the question "In what ways do you use online services?"

Table 7
Parameter estimates: interaction terms and nonlinear parameters

Interaction terms (average value over all consumers)	
Post-post.com	-1.4046
	(.2557)
Post-Times	-.2255
	(.3025)
post.com-Times	-1.6453
	(.5211)
Nonlinear parameters	
τ	6.6577
	(.4928)
γ	.0565
	(.0269)
π	.9467
	(.1715)

Source: Scarborough Research data.
Notes: Standard errors in parentheses.

Table 8
Parameter estimates: changes in demand for the post.com

Sep. 2000-Feb. 2001	-.0525
	(.264)
Mar. 2001-Aug. 2001	1.012
	(.298)
Sep. 2001-Feb. 2002	1.696
	(.298)
Mar. 2002-Aug. 2002	1.608
	(.300)
Sep. 2002-Feb. 2003	2.332
	(.320)

Source: Scarborough Research data.
Notes: Standard errors in parentheses. The table shows coefficients on dummy variables for the given dates, measured relative to the March 2000-August 2000 period. Similar coefficients for the Post and the Times are included in the model but not shown.

Table 9
Variance and covariance of consumer characteristics

Covariance of observable utility

	Post	*post.com*	*Times*
Post	14.4		
post.com	8.2	8.6	
Times	8.1	5.14	10.4

Covariance of ν unobservables

	Post	*post.com*	*Times*
Post	10.9		
post.com	4.3	23.4	
Times	.195	6.23	84.5

Variance of τ unobservables	2.36
Variance of ε unobservables	1.64

Source: Scarborough Research data.
Notes: The table shows covariance of different componenets of utilities at the estimated parameter values.

Table 10
Impact of the online edition on demand for print

Case 1: full model	
Cross-price derivative	9,990
	(1,682)
Change in print readership	-29,184
	(4,406)
Change in print profits	-$5,688,467
	(858,754)
Case 2: model with observable characteristics only	
Cross-price derivative	544
	(1,220)
Change in print readership	1,757
	(3,912)
Change in print profits	-$342,515
	(762,451)
Case 3 model with no heterogeneity	
Cross-price derivative	-9,204
	(1,257)
Change in print readership	40,395
	(5,468)
Change in print profits	$7,873,560
	(1,065,881)

Source: Scarborough Research data.
Notes: The table shows three measures of the online edition's impact. The cross-price derivative is the change in post.com readership when the Post's price is increased by $.10. Change in print readership and print profits are the total changes for the Post when the online edition is added to the choice set. The table shows the estimated values in three models. Case 1 is the estimates from the full model. Case 2 is a model with observable consumer characteristics but no unobservables other than the i.i.d. logit errors. Case 3 is a model with no observable or unobservable consumer heterogeneity except the i.i.d. logit errors.

Table 11
Own and cross-price derivatives of demand
(Change in daily readership per $.10 change in price)

	Post	*post.com*	*Times*
Post	-251,343		
	(8,600)		
post.com	9,990	-127,365	
	(1,682)	(5,591)	
Times	547	3,424	-58,734
	(911)	(1,065)	(4,750)

Source: Scarborough Research data.
Notes: Standard errors in parentheses.

Table 12
Consumer surplus from the post.com

Per Consumer as % of Print	70.1%
Per Consumer in $ per Day	$0.28
Total as % of Print	16.0%
Total in $ per Day	$114,959
Total in $ per Year	$41,959,898
Standard Error	($6,792,630)

Source: Scarborough Research data.
Notes: The table shows the loss in consumer surplus that would result if the post.com was removed from the choice set. "Per consumer in $ per Day" is the total daily loss divided by the number of post.com consumers. "Per consumer as % of print" is this value as a percentage of the total loss from removing the Post print edition from the choice set divided by the number of Post readers.

Table 13
Changes in readership when *post.com* is removed

post.com readers	405,643
Change in Post readership	29,184
	(4,406)
Consumers who read post.com and Post	
Switch to outside good	*17.3%*
Switch to Post only	79.1%
Switch to other bundles	3.6%
Consumers who read post.com only	
Switch to outside good	*68.8%*
Switch to Post only	27.3%
Switch to other bundles	3.9%

Source: Scarborough Research data.
Notes: Standard errors in parentheses. The table shows results from a counterfactual simulation where the post.com is removed from the choice set. The first row shows the total estimated number of DC-area post.com readers. The final rows show the distribution of new choices in the counterfactual by consumers whose observed choices were either the Post-post.com bundle or the post.com alone.

Table 14
Producer surplus effects of the post.com
(Before accounting for post.com operating costs)

Δ Post Profit	-$5,688,467
Δ Times Profit	-$2,786,317
Total post.com Revenue	$33,150,000
Δ Producer Surplus	$24,675,217
Standard Error	($1,727,413)

Source: Scarborough Research data.
Notes: The first two rows show changes in firm profits caused by adding the post.com to the choice set. The next row shows the estimated advertising revenue generated by the post.com (including consumers both inside and outside of DC). The producer surplus figure is the effect of the post.com on producer profits, before accounting for the operating costs of the post.com. As discussed in the text, a rough estimate of these costs is approximately $40 million.

Table 15
Robustness Checks

	Post Company Surplus ($ mil)	*Cross-Price Derivative*	*Optimal Price of post.com*	*Consumer Surplus per Year ($ mil)*
Actual estimates	$27.46	9,990	$0.20	$41.96
Price = $.25	$28.51	10,100	$0.19	$41.49
Price = $.35	$28.39	9,883	$0.20	$42.38
Marginal cost = $.30	$28.14	9,462	$0.21	$44.32
Marginal cost = $.50	$28.77	10,581	$0.18	$39.09
Online revenue = $25m	$19.15	9,968	$0.23	$42.38
Third order interaction	$29.23	10,023	$0.19	$42.20

Notes: The table shows estimates from the model for alternative values of the supply parameters. The first row repeats the estimates from the actual model that were presented in tables 5-14. The next two rows vary the Post cover price used to derive the price coefficient. The following rows vary the value of the marginal cost used. The next row assumes a lower value for the post.com's advertising revenue. The final row allows the utility of the bundle of all three goods to vary independently of the utilities of the component two-good bundles.