



MIT Sloan School of Management

**Working Paper 4305-03
April 2003**

Consumer Surplus in the Digital Economy: Estimating the Value of Increased Product Variety at Online Booksellers

Erik Brynjolfsson, Michael D. Smith, Yu (Jeffrey) Hu

© 2003 by Erik Brynjolfsson, Michael D. Smith, Jeffrey Hu. All rights reserved.
Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission, provided that full credit including © notice is given to the source.

This paper also can be downloaded without charge from the
Social Science Research Network Electronic Paper Collection:

<http://ssrn.com/abstract=400940>

Consumer Surplus in the Digital Economy: Estimating the Value of Increased Product Variety at Online Booksellers

Erik Brynjolfsson^{*}, Michael D. Smith^{**}, and Yu (Jeffrey) Hu^{***}

This Version: April 2003

ABSTRACT

We present a framework and empirical estimates that quantify the economic impact of increased product variety made available through electronic markets. While efficiency gains from increased competition significantly enhance consumer surplus, for instance by leading to lower average selling prices, our present research shows that increased product variety made available through electronic markets can be a significantly larger source of consumer surplus gains.

One reason for increased product variety on the Internet is the ability of online retailers to provide a large number of products for sale. For example, the number of book titles available at Amazon.com is over 23 times larger than the number of books on the shelves of a typical Barnes & Noble superstore and 57 times greater than the number of books stocked in a typical large independent bookstore.

Our analysis indicates that the increased product variety of online bookstores enhanced consumer welfare by \$731 million to \$1.03 billion in the year 2000, which is at least five times as large as the consumer welfare gain from increased competition and lower prices in this market. There may also be large welfare gains in other SKU-intensive consumer goods such as music, movies, consumer electronics, and computer software and hardware.

Acknowledgements: We thank Matthew Gentzkow, Austan Goolsbee, Jerry Hausman, Lorin Hitt, Guillermo Israilevich, participants at the University of Maryland, University of Pittsburgh, the 2001 Workshop on Information Systems and Economics (WISE), the 2003 American Economic Association meetings, the 2003 International Industrial Organization Conference and the referees and editors of this journal for valuable comments on this paper. We thank Jeff Bezos, Al Greco, Vicki Jennings, Mary Ellen Keating, Jason Kilar, Steve Riggio, Madeline Schnapp, Jeff Wilke and representatives of the American Booksellers Association, the American Wholesale Booksellers Association, the Book Industry Study Group, Barnes and Noble, Amazon.com, MIT Press and various other publishers and booksellers for providing valuable information on the book industry. Generous financial support was provided by the Center for eBusiness at MIT under a grant from Fleet Bank and the National Science Foundation (grant number IIS-0085725).

* Sloan School of Management, Massachusetts Institute of Technology; email: erikb@mit.edu; web: <http://ebusiness.mit.edu/erik>

** H. John Heinz III School of Public Policy and Management, Carnegie Mellon University; email: mds@cmu.edu; web: <http://www.heinz.cmu.edu/~mds>

*** Sloan School of Management, Massachusetts Institute of Technology; email: yuhu@mit.edu

1. Introduction

“Clearly, new goods are at the heart of economic progress. But that realization is only the beginning of an understanding of the economics of new goods. The value created by new goods must somehow be converted into an exact quantitative measure...”

Timothy F. Bresnahan and Robert J. Gordon (1997, p. 1)

“The Internet is providing access for people who just can’t find the book they are looking for in a store.”

Nora Rawlinson, editor of Publishers Weekly, quoted in Investors Business Daily, June 2, 1999.

Information technology facilitates the delivery of many new products and services over electronic networks. As these electronic networks develop and mature, it will be important to quantify their value for customers, merchants, shareholders, and society. While much of the attention in academic research and in the press has been on the relative operational efficiency of the online channel versus traditional channels, we believe that important benefits lie in new products and services made available through these channels. While it has been relatively easy to quantify the operational costs of each channel, the value of new products and services made available through electronic networks has remained unquantified. By default, this value has been ignored, effectively treating convenience and selection as if its value were zero.

Our research focuses on increased product variety, which is one category of new products and services made available through electronic networks. Internet retailers have nearly unlimited “virtual inventory” through centralized warehouses and drop-shipping agreements with distributors (e.g., Bianco 1997 and Mendelson and Meza 2002). Because of this, they can offer convenient access to a larger selection of products than brick-and-mortar retailers can. Table 1 shows the difference between the number of items available at Amazon.com and a typical large

brick-and-mortar retailer for several consumer product categories.¹ For example Amazon.com and Barnesandnoble.com provide easy access to each of the 2.3 million books in print (and millions more used and out of print titles) while conventional brick-and-mortar stores carry between 40,000 and 100,000 unique titles on their shelves. Thus, online consumers are easily able to locate, evaluate, order, and receive millions of books that are not available on the shelves of local bookstores. Large differences in product variety are also seen in music, movies, and consumer electronics products. Even Wal-Mart Supercenters, which range in size from 109,000 to 230,000 square feet, only carry one-sixth of the number of SKUs that are carried by Walmart.com (Owen 2002).

Table 1: Product Variety Comparison for Internet and Brick-and-Mortar Channels

<i>Product Category</i>	<i>Amazon.com</i>	<i>Typical Large Brick-and-Mortar Store</i>
Books	2,300,000	40,000 – 100,000
CDs	250,000	5,000 – 15,000
DVDs	18,000	500 – 1,500
Digital Cameras	213	36
Portable MP3 players	128	16
Flatbed Scanners	171	13

While some of these products may be available from specialty stores or special ordered through brick-and-mortar stores, the search and transaction costs to locate specialty stores or place special orders are prohibitive for most consumers.² In effect, the emergence of online retailers

¹ Inventory values for Amazon.com were obtained from discussions with industry executives, industry estimates and Bowker's Books in Print database (books), wholesale suppliers to Amazon.com (CDs), and direct counting of normally stocked items. Inventory values for brick-and-mortar stores were obtained from interviews with managers and direct observation of inventory for Barnes and Noble Superstores (Books, CDs, DVDs), Best Buy (CDs, DVDs, Digital Cameras, Portable MP3 Players, Flatbed Scanners) and CompUSA (Digital Cameras, Portable MP3 Players, Flatbed Scanners).

² To illustrate this difference, on November 26, 2001 one of the authors ordered the same book through Barnesandnoble.com and through a special order at a local Barnes & Noble Superstore. The Barnesandnoble.com order process took 3 minutes to place, arrived in 3 days, and cost \$31.99. The Barnes and Noble order took nearly 1 hour to place, took 8 days to arrive, and cost \$37.45. The store was located 4.6 miles away from the author's house (note that Brynjolfsson and Smith (2000) found that the average person in the United States lives 5.4 miles away

places a specialty store at every shopper's desktop. This improves the welfare of these consumers by allowing them to buy specialty products they otherwise cannot buy or will not buy due to high transaction costs and low product visibility. This effect will be especially beneficial to those consumers who live in remote areas that do not have specialty retailers.

As one might expect, the lower transactions costs offered by the Internet have led to increased orders for many titles not previously stocked in brick-and-mortar stores. Frank Urbanowski, Director of MIT Press, attributed the 12% increase in sales of backlist titles directly to increased accessibility to these titles through the Internet (Professional Publishing Report 1999). Similarly, Nora Rawlinson, the editor of Publishers Weekly, observes:

“Publishers are finding that books on their backlists are suddenly selling well. Bookstores are great for browsing but they are difficult places to find a specific title...The Internet is providing access for people who just can't find the book they are looking for in a store.” (Lyster 1999)

The differences in variety reflect underlying differences in the technology and economics of conventional and Internet retailers. As noted by Saul Hansell in the New York Times:

“The average book may sit on the shelf of a store for six months or a year before it is bought. The cost of this inventory in a chain of hundreds of stores is huge. Amazon can keep just one or two copies in its warehouse — and still make the title available to the whole country — and restock as quickly as customers buy books.” (Hansell 2002)

Further, anecdotal evidence suggests that consumers place a high value on the convenience offered by Internet retailers when locating and purchasing obscure products. For example, Yahoo/ACNielsen's 2001 Internet Confidence Index lists “wide selection of products” as one of the top three drivers of consumer ecommerce based on a survey of Internet purchasers. However,

from the closest bookstore). The time to place the order included 21 minutes of driving time (round-trip) to place the order; 8 minutes to park, search for the book in the store, search for a sales person, and place the order; 20 minutes of driving time to pick up the order; and 9 minutes to park and pay for the special order.

no systematic estimates exist to quantify the dollar value consumers place on the increased product variety available through Internet markets.

This paper represents a first effort to apply a methodology for estimating this value to one prominent category of products offered by Internet retailers — obscure book titles. Our methodology uses a small set of generally available statistics that track how consumers “vote with their dollars,” and thus may find application in a variety of product categories. The resulting estimates of consumer surplus will have important economic and public policy implications, especially as investors and managers try to understand and evaluate the value proposition of Internet-based commerce.

The remainder of the paper proceeds as follows. Section 2 presents the economic literature pertaining to consumer welfare gains from new goods and increased product variety. Section 3 develops a methodology to measure consumer welfare from increased product variety offered in Internet markets. Section 4 applies this methodology to obscure book sales over the Internet and section 5 concludes with some broader implications. Appendix A summarizes both the model development (Section 3) and the data necessary to calibrate the model in a general market environment (Section 4).

2. Literature Review

The development of an empirical methodology to estimate the welfare change resulting from price changes can be traced to Hicks’ (1942) compensating variation measure. Historically compensating variation has been difficult to measure because it involves integration of the unobservable Hicksian compensated demand curve. However, Hausman (1981) develops a closed-form solution for measuring compensating variation under standard linear or log-linear

demand functions. More recently, Hausman (1997a) shows that the welfare effect of the introduction of a new product is equivalent to the welfare effect of a price drop from the product's "virtual price", the price that sets its demand to zero, to its current price. Applying this technique, he estimates that the FCC's decision to delay the introduction of two telecommunication services has reduced U.S. consumer welfare by billions of dollars a year. Subsequently researchers have examined the welfare effects of other new products in traditional markets, using similar or more refined models. Examples include Hausman (1997b), Nevo (2001), Goolsbee and Petrin (2001), Petrin (2001), Hausman and Leonard (2001). In addition, Bresnahan (1986) and Brynjolfsson (1995) have looked at welfare gains from information technology investments.

Researchers in the field of macroeconomics have also started to pay attention to new products or new varieties of products. Bils and Klenow (2001) find that consumer spending has shifted away from products that have shown little variety gain. The Stigler commission (NBER 1961) and the Boskin Commission (Boskin Commission Report 1996) conclude that the greatest flaw in the Consumer Price Index is its failure to account adequately for new goods and quality improvements in existing goods.

It is also worth noting that there is a large body of marketing literature examining the relationship between perceived variety and actual assortment. Most researchers agree that consumers generally prefer more variety when given a choice (e.g., Baumol and Ide 1956 and Kahn and Lehmann 1991). More recently, researchers have shown that consumers' perception of variety is influenced not only by the number of distinct products, but also by the repetition frequency, organization of the display, and attribute differences (e.g., Dreze, Hoch and Purk 1994; Broniarczyk, Hoyer and McAlister 1998; Hoch, Bradlow and Wansink 1999, 2002; Van

Herpen and Pieters 2002). In this paper, we focus on the impact that increased availability of products in the online channel has on consumers' actual purchase behavior. Thus, questions of shelf space and consumer perceptions are muted relative to the actual assortment of products and observed consumer behavior.

3. Methodology

This paper applies and extends existing welfare estimation techniques to measure the consumer welfare gain from the increased product variety made available through electronic markets. To do this, we define the total effect of the introduction of new products in online markets on consumer welfare as the difference in the consumer's expenditure function before and after the introduction, measured at the level of post-introduction utility:

$$CV = e(p_{e0}, p_{n0}, u_1) - e(p_{e1}, p_{n1}, u_1), \quad (1)$$

where p_{e0} and p_{e1} are the vectors of pre- and post-introduction prices of existing products respectively, p_{n0} is the virtual price of the new product (the price that sets demand to zero), p_{n1} is the post-introduction price of the new product, and u_1 is the post-introduction utility level. In effect, equation (1) measures how much a pre-Internet consumer would need to be compensated in order to be just as well off as they would be after the emergence of online markets.

We then follow Hausman and Leonard's (2001) derivation to break the total effect into the variety effect resulting from the availability of the new product and the price effect resulting from changes of prices of existing products:

$$CV = [e(p_{e1}, p_{n0}, u_1) - e(p_{e1}, p_{n1}, u_1)] + [e(p_{e0}, p_{n0}, u_1) - e(p_{e1}, p_{n0}, u_1)]. \quad (2)$$

When the vector of prices of existing products does not change before and after the introduction of the new product, i.e., $p_{e0}=p_{e1}=p_e$, one only needs to measure the variety effect and we can redefine the expenditure function such that $e(p_e, \dots) \equiv e'(\dots)$:

$$CV = e(p_e, p_{n0}, u_1) - e(p_e, p_{n1}, u_1) = e'(p_{n0}, u_1) - e'(p_{n1}, u_1). \quad (3)$$

The assumption that $p_{e0}=p_{e1}=p_e$ appears to be valid in our empirical context because the overwhelming majority of book prices charged by brick-and-mortar stores have not changed as a result of the emergence of online markets. Nearly all brick-and-mortar stores sold most titles at the manufacturer's suggested list price before the emergence of online markets and continue to do so today. Moreover, most studies have shown that, if anything, Internet retailers tend to increase competition and place downward pricing pressure on brick-and-mortar retailers (e.g., Brynjolfsson and Smith 2000; Scott Morton, Zettelmeyer, and Silva-Risso 2001; Brown and Goolsbee 2002). Thus, if brick-and-mortar prices were to change at all, we would expect them to decline. Our calculations under the zero price change assumption would therefore underestimate true gains in consumer surplus.

To apply equation (3) in practice, we specify a standard log-linear demand function for the new product made available by the Internet,

$$x(p, y) = Ap^{\mathbf{a}} y^{\mathbf{d}}, \quad (4)$$

where p is the price of the new product, y is the income, \mathbf{a} is the price elasticity, and \mathbf{d} is the income elasticity. This specification is the most widely used specification in the literature of

demand estimation and it fits a wide variety of data well (e.g., Brynjolfsson 1995, Hausman 1997a, 1997b, and Hausman and Leonard 2001).³

Following Hausman (1981), we can use Roy's identity to write equation (4) as

$$x(p, y) = -\frac{\partial v(p, y)/\partial p}{\partial v(p, y)/\partial y}, \quad (5)$$

where $v(p, y)$ is the indirect utility function.

Solving this partial differential equation gives

$$v(p, y) = -A \frac{p^{1+a}}{1+a} + \frac{y^{1-d}}{1-d} \quad (6)$$

and the expenditure function

$$e(p, u) = \left[(1-d) \left(u + \frac{Ap^{1+a}}{1+a} \right) \right]^{1/(1-d)}. \quad (7)$$

Using equations (3) and (7), it can be shown (Hausman 1981) that the welfare impact of the introduction of a new product is given by

$$CV = \left[\frac{1-d}{1+a} y^{-d} (p_{n0}x_0 - p_{n1}x_1) + y^{(1-d)} \right]^{1/(1-d)} - y, \quad (8)$$

where CV is the compensating variation, d is the income elasticity estimate, a is the price elasticity, y is income, (p_{n1}, x_1) are the post-introduction price and quantity of the new product, and (p_{n0}, x_0) are the pre-introduction virtual price and quantity of the new product.

³ The final result of welfare estimates will depend on the adopted specification of demand function. However, earlier research (e.g. Brynjolfsson 1995 and Hausman and Newey 1995) finds that using a nonparametric specification with complete freedom to fit the data may not significantly improve the accuracy of welfare estimates over estimates using a standard log-linear specification.

Prior research has shown that income elasticity effects can be ignored for typical consumer products where purchases are a small fraction of the consumer's annual income (e.g. Hausman 1997a, Brynjolfsson 1995). Applying this assumption, i.e. $d=0$, equation (8) simplifies to

$$CV = -\frac{p_n x_1}{1 + a}, \quad (9)$$

since the pre-introduction quantity is zero and $p_{n0} x_0=0$. If income elasticity were positive, as is likely for books, including income elasticity in our calculations would increase our consumer surplus estimates by a small amount (Varian 1992).

4. Data and Results

We use this methodology to measure the consumer surplus gain in Internet markets from access to books not readily available through brick-and-mortar retailers. As noted above, for many consumers, these obscure books can properly be categorized as “new” products because, while they are readily available in Internet markets, the transactions costs necessary to acquire these goods in physical markets are prohibitively high. The availability of these books to Internet consumers reflects, in part, the increased inventory carrying capacity of Internet retailers.

Furthermore, recommendation lists, customer and industry reviews, images of the book jacket and selected book pages, and convenient search facilities allow Internet consumers to discover and evaluate obscure books that likely would have remained undiscovered in conventional retail environments where these books would be unavailable for browsing.

This product category also provides a useful starting point for surplus measurement because it represents a relatively mature Internet market, and because prior research has already measured

the reduction in prices from increased competition on the Internet (e.g., Brynjolfsson and Smith 2000), providing a point of reference for our surplus measurements.

In the following sections, we discuss how we estimate the parameters necessary to calculate the consumer surplus resulting from increased accessibility to obscure books on the Internet: the price elasticity of demand and the price and quantity of sales of obscure books on the Internet.

4.1. Elasticity of Demand

The most straightforward approach to calculate elasticity of aggregate demand would be through direct empirical estimation. In the conclusion section we discuss how elasticity of demand might be obtained by partnering with a book publisher or with a retailer with dominant market share to conduct a direct pricing experiment. Unfortunately, we were unable to obtain cooperation from either publishers or retailers to conduct such an experiment. In the absence of this data, we estimate the elasticity of aggregate demand by taking advantage of the characteristics of the book industry structure and available industry statistics on gross margins.

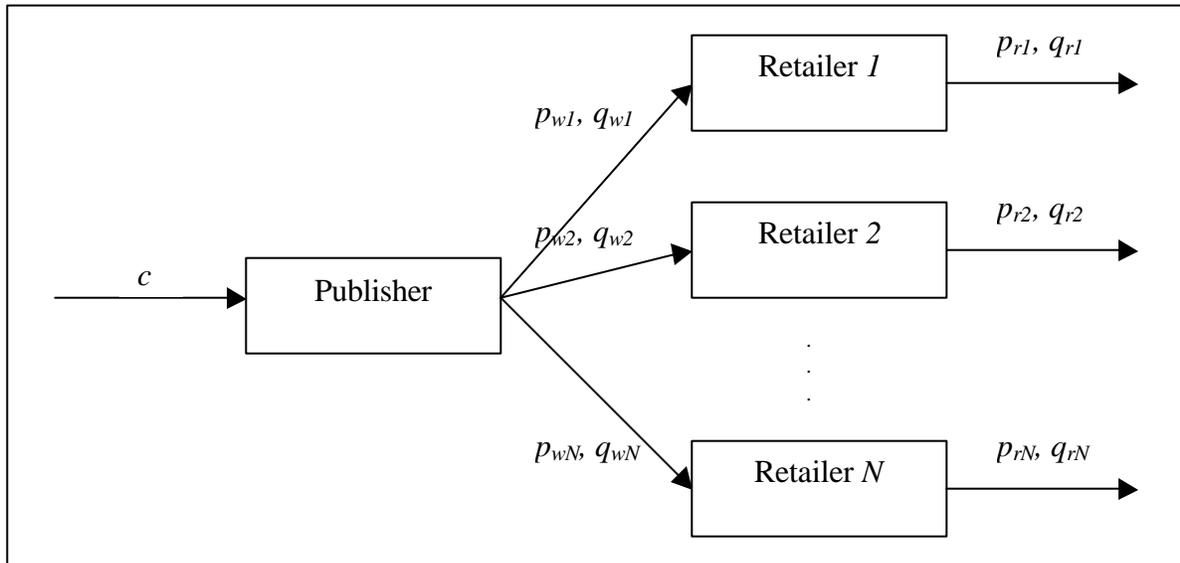
To do this, we first note that the book industry is vertically structured as shown in Figure 1, where c is the marginal cost of a book and p_{wi} , q_{wi} , p_{ri} , and q_{ri} are wholesale price and quantity and retail price and quantity for retailer i ($i=1,2,\dots,N$) respectively.⁴

Publishers set both list prices and wholesale prices of the books they publish. They sell books to retailers, either directly or through distributors, at prices that are a set percentage of the books' list prices, typically between 43-51% of list prices. Thus, a change in the list price of a book

⁴ This structure is accurate for the vast majority of consumer purchases, which are made through bookstores. However, in a few cases, customers choose to special-order books directly from the publisher. For our purposes, as long as these books are available from the publisher, at the same prices both before and after Internet book retailers introduce the increased selection of books, publisher special-orders will not affect our underlying model. This follows because, as noted above, the unchanged prices of existing products are irrelevant to consumer surplus calculations from the introduction of new products.

would also result in a proportional change in the wholesale price of the book.⁵ Further, wholesale prices charged on an individual book are almost the same across retailers, regardless of the channel that the retailer operates in or the size of the retailer (e.g., Clay, Krishnan, and Wolff 2001). Thus we have $p_{wi}=p_w$ for $i=1,2,\dots,N$.

Figure 1: Vertical Industry Structure in Book Sales



Retailer i ($i=1,2,\dots,N$) receives books from either publishers or distributors, and then sells these books to consumers at some discount off list price. Books in different categories are sold at different pre-set discounts off their list prices. For a particular book, the discount off its list price will not change as a result of a change in its list price.⁶ Most obscure books are sold at their full list prices.⁷ Therefore, for a given obscure title, there exists a stable relationship between the

⁵ Source: conversation with Vicki Jennings, MIT Press, July 23, 2002.

⁶ Moreover, this discount off list price is usually set by retailers in multiples of 10%. For example, in a representative sample of 23,744 titles sold at Amazon.com in late 1999, 88.5% of them follow such a pricing pattern — 29.5% have 0% discount, 1.4% have 10% discount, 34.3% have 20% discount, 18.4% have 30% discount, 1.6% have 40% discount, 3% have 50% discount, and 0.1% have 60, 70, 80, or 90% discounts. (See Smith 2001 for more information on this sample of titles.)

⁷ We selected 100 books at random from a sample of all customer searches at Dealtime.com on July 2, 2001. Among the 37 books with Amazon.com sales ranks greater than 100,000, 86% are sold at their respective list prices at Amazon.com (versus 41% for the remaining titles). Lee and Png (2002) also collect data showing that bookstores typically offer zero discounts on non-bestseller titles.

book's wholesale price and retailer i 's retail price, $p_w = k_i p_{ri}$ where k_i is a constant between 0 and

1. In addition, according to several publishers we interviewed, as well as the American

Wholesale Booksellers Association, the vast majority of books are sold on consignment.

Typically, retailers can return unsold books to publishers or distributors without penalty (except

for return shipping). Because of this, $q_{wi} = q_{ri}$. If we define $q_r \equiv \sum_{i=1}^N q_{ri}$ as the total quantity sold

by retailers to consumers and $q_w \equiv \sum_{i=1}^N q_{wi}$ as the total net quantity sold by the book's publisher to

retailers, we easily get $q_r = q_w$.

If we define retailer i 's market share on this book as $s_i \equiv \frac{q_{ri}}{q_r}$, then the weighted retail market

price can be written as $\bar{p} \equiv \sum_{i=1}^N s_i p_{ri} = \sum_{i=1}^N \frac{s_i}{k_i} p_w$. One can show that the elasticity of aggregate

demand in the retailing market equals the elasticity of demand faced by the publisher:⁸

$$\frac{\overline{pdq_r}}{q_r d\bar{p}} = \frac{\overline{pdq_w}}{q_w d\bar{p}} = \frac{\left(\sum_{i=1}^N \frac{s_i}{k_i} p_w \right) dq_w}{q_w \left(\sum_{i=1}^N \frac{s_i}{k_i} dp_w \right)} = \frac{\left(\sum_{i=1}^N \frac{s_i}{k_i} \right) p_w dq_w}{q_w \left(\sum_{i=1}^N \frac{s_i}{k_i} \right) dp_w} = \frac{p_w dq_w}{q_w dp_w}. \quad (10)$$

Since the publisher of a particular title has total control over establishing the title's list and

wholesale price, it is reasonable to apply the well-known Lerner index formula to estimate the

price elasticity of demand faced by the publisher:⁹

⁸ The elasticity of aggregate demand can be thought of as the percentage change in total market sales if all retailers changed their price by 1 percentage. In general, this will differ from the elasticity of demand faced by a particular retailer acting independently.

⁹ This form of the Lerner index is applicable to single product monopolists, multiproduct monopolists who maximize profits on a per product basis, or in instances where cross elasticity is zero. In the more general

$$\frac{p_i - C_i'}{p_i} = -\frac{1}{\mathbf{a}_{ii}}. \quad (11)$$

Publishers sell books to both online retailers and brick-and-mortar retailers, either directly or through distributors, at wholesale prices that are a set percentage of books' list price, typically between 43-51% of list prices. Publishers incur the same production costs whether books are sold to an online retailer or to a brick-and-mortar retailer. Therefore, publishers sell obscure books to online retailers and brick-and-mortar retailers at the same gross margins. Discussions with various publishers indicate that the gross margin of a typical obscure title is between 56-64%.¹⁰ Thus, using (11), the elasticity of demand faced by the publisher is between -1.56 and -1.79, and by (10) this is also the aggregate demand in the retailing market.

This estimate can also be compared with what other researchers have obtained, albeit using retailer data. For example, Brynjolfsson, Dick, and Smith (2002) use shopbot data to calculate an own-price elasticity of -1.47 for retailers listing their prices at a popular shopbot, which is somewhat lower in absolute value than our estimates. Similarly, Chevalier and Goolsbee (2003) estimate a demand system for two online book retailers: Amazon and BarnesandNoble.com. The imputed demand elasticity using their calculations is also lower than our elasticity estimate. As noted in (9), a smaller elasticity will translate to a larger consumer surplus estimate.

multiproduct monopolist case, the Lerner index for product i is given by $\frac{p_i - C_i'}{p_i} = -\frac{1}{\mathbf{a}_{ii}} - \sum_{j \neq i} \frac{(p_j - C_j')q_j \mathbf{a}_{ji}}{p_i q_i \mathbf{a}_{ii}}$, where i

and j are indexes for products. However, in the Internet book market all available evidence suggests that prices are set on an individual book basis and thus we use estimates based on equation (11) for our elasticity calculations.

¹⁰ For example, data from the American Association of Publishers suggest that the gross margin of a typical book is between 56-58% depending on whether shipping is included.¹⁰ A typical MIT Press book has a gross margin of approximately 63% (source: conversation with Vicki Jennings, MIT Press). A large publisher of technical books revealed that each of their books had gross margins of between 58-64% over the past several years. A large publisher of trade books revealed that each of their books had gross margins of approximately 60%.

4.2. Sales of Obscure Titles on the Internet

Internet retailers are extremely hesitant about releasing specific sales data, and we were unable to obtain sales data from a major Internet retailer that would allow us to estimate the sales of obscure titles on the Internet. However, we were able to obtain data from one publisher that allow us to estimate the proportion of sales of obscure titles in total sales at Amazon.com. This proportion can apply to the overall Internet book market, given that Amazon.com has approximately a 70% share of the Internet book market (Ehrens and Markus 2000).

This publisher provided data matching the publisher's weekly sales for 321 titles to the sales rank observed at Amazon.com's web site during the same week. According to Amazon.com's frequently asked questions page:

“The [rank] calculation is based on Amazon.com sales and is updated regularly. The top 10,000 best sellers are updated each hour to reflect sales in the preceding 24 hours. The next 100,000 are updated daily.”¹¹

These data, gathered for three weeks in the summer of 2001, provide a fairly robust basis for correlating sales and sales rank at Amazon.com. The observed weekly sales range from 1 to 481 units sold and the observed weekly sales rank ranges from the 238 to 961,367. Summary statistics for this data are shown in Table 2.¹²

Table 2: Summary Statistics for Amazon Sales-Rank Data

<i>Variable</i>	<i>Obs.</i>	<i>Mean</i>	<i>S.D.</i>	<i>Min</i>	<i>Max</i>
Weekly Sales	861	19.17	30.63	1	481
Weekly Sales Rank	861	31,532.85	58,350.92	238	961,367

¹¹ Available at <http://www.amazon.com/exec/obidos/tg/browse/-/525376/104-2977251-9307125>. Further experimentation demonstrated that the sales rank does not include used book sales or sales through Amazon's marketplace sellers.

¹² The panel of titles changed somewhat during the sample period and as a result not all titles were tracked in all weeks.

We fit our data on sales and sales rank to a log-linear (Pareto) distribution:

$$\log(\text{Quantity}) = \mathbf{b}_1 + \mathbf{b}_2 \cdot \log(\text{Rank}) + \mathbf{e} , \tag{12}$$

where e is orthogonal to $\log(\text{Rank})$ and it is spherical, following the standard OLS assumptions.

This approach was suggested by Madeline Schnapp of O'Reilly Books who reported excellent success estimating competitors' unit books sales by comparing their books' sales ranks to O'Reilly's. Chevalier and Goolsbee (2003) also fit sales and sales rank data to a (slightly different) log-linear distribution with good success. Earlier applications include Pareto (1896), who found that income can be approximated such a log-linear distribution, and Zipf (1949) who suggested that city size also follows a log-linear distribution with a slope of -1.

Regressing $\log(\text{Quantity})$ onto a constant and $\log(\text{Rank})$, we obtain an estimate of 10.526 for \mathbf{b}_1 and -0.871 for \mathbf{b}_2 as shown in Table 3 below.

Table 3: Regression Results Amazon Sales-Rank Data¹³

<i>Variable</i>	<i>Coefficient</i>
Constant	10.526 (0.156)
Log(Rank)	-0.871 (0.017)
R ²	0.8008

The coefficients in this regression are highly significant and the R² value suggests that our model is precisely estimated. Furthermore, the estimates lead to plausible sales-rank results. Given our estimates, a book with a rank of 10 is estimated to get 5,000 sales per week and a book with a

¹³ A graphical analysis suggests that the size of the residuals increases in rank, and a Breusch-Pagan test confirms the presence of heteroskedasticity in the residuals. Thus we use White heteroskedasticity consistent estimator (see Greene 2000, p.463) to estimate both coefficients. We also performed a test for structural change by interacting $\log(\text{Rank})$ with a dummy variable that took on the value of one for ranks larger than 40,000. The coefficient on this variable was positive (but statistically insignificant) suggesting that our results would, if anything, be strengthened if we based our \mathbf{b}_2 on only high rank books.

rank of 100,000 gets, on average, 1.6 sales per week. Likewise, integrating under the curve for titles with rank from 1 to 2.3 million suggests that Amazon.com was selling books at a rate of 99.4 million per year in the summer of 2001. This estimate compares well with industry statistics.¹⁴

These estimates also compare favorably with Pareto slope parameter estimates obtained by Chevalier and Goolsbee (2003) using a clever and easily executed experiment. To conduct this experiment, they first obtained information from a publisher on a book with relatively constant weekly sales, then purchased 6 copies of the book in a 10-minute period, and tracked the Amazon rank before and shortly after the purchases. Using the sales and sales rank before and after the experiment, they estimated b_2 as -0.855 (note that the q reported by Chevalier and Goolsbee corresponds to $-1/b_2$). They also estimated b_2 from similar sales-rank data reported by Weingarten (2001) and Poynter (2000) as -0.952 and -0.834 respectively. We performed a similar purchase experiment in September 2002 and calculated b_2 as -0.916.¹⁵ It is significant that while these parameter estimates rely on only 2 points, they are remarkably similar to the results calculated in Table 3 above, which are based on over 800 points.

We can use the Pareto slope parameter estimate from our data to calculate the proportion of unit sales at Amazon that fall above a particular rank as

¹⁴ The 2001 Book Industry Trends lists 2000 total unit sales of books at 2.5 billion and their study also shows that the Internet makes up 6% of total book purchases. Amazon.com has a 70% share of the Internet book market (Ehrens and Markus 2000).

¹⁵ We selected a book whose Amazon.com rank on September 13, 2002 was 606,439. We checked the rank of this book each day between September 14 and September 30 and noted 3 changes: on Monday September 16 the book jumped from 606,439 to 596,625; on Monday September 23 the book dropped from 596,625 to 599,352; and on Monday September 30 the book dropped from 599,352 to 601,457. We infer from this that Amazon updates its sales rankings on low selling books each Monday and that a sale occurred sometime during the week ending September 15 and no sales occurred during the remaining weeks. On September 30 one of the authors ordered 5 copies of this book using 5 different Amazon user accounts. The next morning the book had a sales rank of 4,647.

$$r(x, N) = \frac{\int_1^N \mathbf{b}_1 t^{b_2} dt}{\int_1^N \mathbf{b}_1 t^{b_2} dt} = \frac{N^{(b_2+1)} - x^{(b_2+1)}}{N^{(b_2+1)} - 1}, \quad (13)$$

where x is the rank, and N is the total number of books available.

What rank cutoff is appropriate for our purposes? As noted above, we wish to estimate the gain in consumer surplus from access to books on the Internet that are not normally stocked by brick-and-mortar stores. Thus, our rank figure should approximate the average number of books a consumer could readily locate in local physical stores.

At one end of the spectrum one would want to consider consumers who do not have easy access to bookstores with a broad selection of titles. In Appendix C of Brynjolfsson and Smith (2000) the authors calculated that the average consumer in the United States lives 5.4 miles away from the closest general selection bookstore. Using the same dataset, we find that 14% of U.S. consumers live more than 10 miles away from the nearest general selection bookstore and 8% live more than 20 miles from their nearest bookstore. For such customers the relevant rank might be near 0. That is, without the Internet such customers are not able to easily purchase even general selection books.

More typically, consumers will have at least one and possibly multiple bookstores close-by. However, these brick-and-mortar bookstores vary significantly in size. Small bookshops and mall-based stores stock approximately 20,000 unique titles, large independent booksellers stock approximately 40,000 unique titles, Barnes and Noble and Borders superstores stock

approximately 100,000 unique titles, and the Barnes and Noble superstore in New York City, reported to be the “World’s Largest Bookstore,” carries 250,000 unique titles on its shelves.¹⁶

Table 4: Proportion of Sales from Obscure Titles at Amazon

<i>Sales Rank</i>	<i>Proportion in Total Sales</i>	<i>Standard Error</i> ¹⁷
>40,000	47.9%	2.7%
>100,000	39.2%	2.5%
>250,000	29.3%	2.0%

In Table 4 we estimate the proportion of total sales at Amazon.com that lies above a particular rank (i.e., titles that are not available at a typical brick-and-mortar bookstore) for each of the reference points discussed above. These calculations are based on equation (13) along with the estimate from Table 3 for b_2 and 2,300,000 (the number of books in print) for N . This table shows that 47.9% of Amazon’s unit sales fall in titles with ranks above 40,000 and 39.2% of sales fall in titles with ranks above 100,000, as Figure 2 illustrates. It is unlikely that every consumer will live within reasonable driving distance to the largest Barnes and Noble superstore in New York City and have access to the 250,000 titles stocked there, but using that number as the cutoff point only reduces the proportion down to 29.3%.

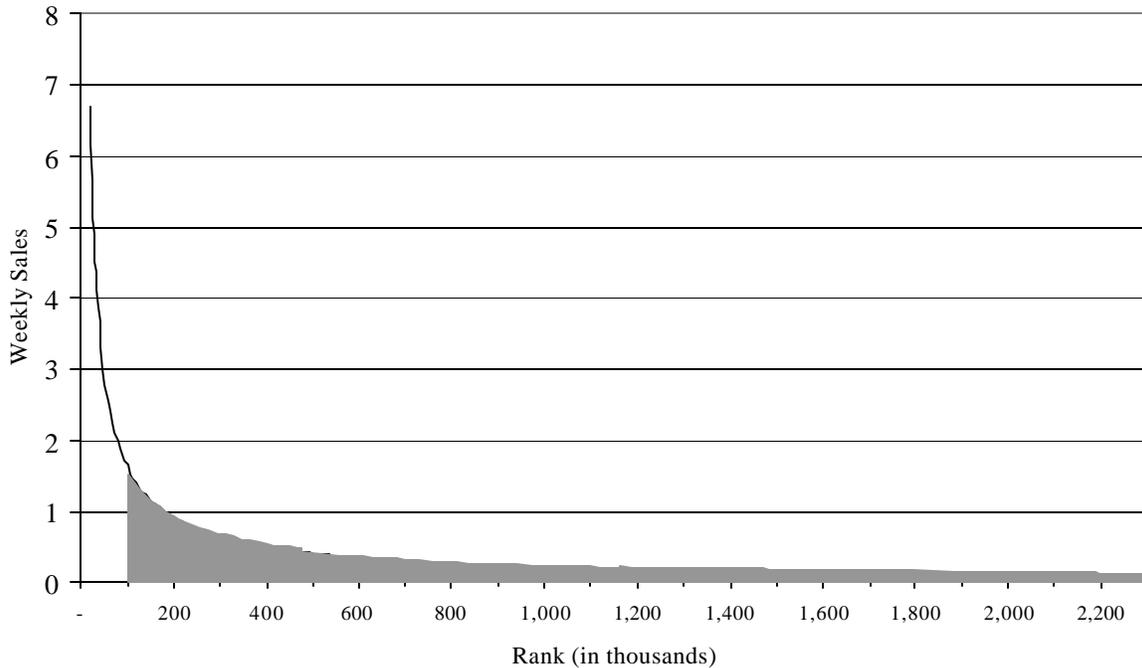
In subsequent calculations, we use a rank of 100,000 as our point-of-reference for consumer surplus estimates. This cutoff can be interpreted either in terms of the average stock levels at a Barnes and Noble or Borders superstore or as taking into account the possibility that consumers shop at multiple smaller independent stores. For example, if there were only a 50% overlap in

¹⁶ Stock figures for Barnes and Noble were obtained from correspondence with Mary Ellen Keating, Senior Vice President of Corporate Communications and Public Affairs, Barnes and Noble, December 3, 2001. Stock figures for independent stores were obtained from multiple industry sources and discussions, including Ritchie (1999).

¹⁷ Since the proportion of Amazon unit sales that fall in titles with ranks above x is a function of b_2 and we obtain the standard error of b_2 from the regression, we calculate the standard error of our estimate using the Delta Method (see Greene 2000, p.118).

stocked titles at large independent bookstores, a consumer would have to shop at a minimum of five such stores to have access to 100,000 titles.

Figure 2: Share of Amazon Sales Above Rank 100,000



This large cutoff point seems fairly conservative on two dimensions. First, it is unlikely that most consumers, particularly rural consumers as mentioned above, have access to this number of unique titles through local bookstores. Second, even if all consumers had access to these larger stores, the 100,000 cutoff will underestimate true consumer surplus if, as seems likely, superstores do not stock exactly the same 100,000 most popular books that Amazon.com stocks.

4.3. Consumer Welfare

According to 2001 Book Industry Trends, book revenue in year 2000 was \$24.59 billion (Book Industry Study Group 2001). Given that the Internet makes up 6% of total book sales (Book Industry Study Group 2001), we estimate that the total Internet book sales in 2000 were \$1.475

billion. If we assume that obscure titles account for about the same proportion of total sales at other Internet book retailers as at Amazon, the sales of titles that are not available at a typical brick-and-mortar bookstore are \$578 million based on the estimates in Section 4.2.

Since these estimates are based on aggregate figures, it is further necessary to ensure that the average prices of obscure books sold on the Internet are not lower than the average prices of more popular books sold on the Internet. If this were not true, we would overestimate the true consumer surplus by using aggregate figures. To analyze the relative prices of obscure and more popular books we selected 100 books at random from a sample of all customer searches at Dealtime.com on July 2, 2001. We then categorized the books into obscure and regular titles based on whether their Amazon.com sales rank was greater than (obscure) or less than (regular) 100,000.¹⁸ Table 5 illustrates that the prices of obscure books are greater than the prices of regular titles. Thus, if anything, our estimates using aggregate sales figures will underestimate true consumer surplus from sales of obscure titles on the Internet.

Table 5: Price Comparison for Obscure Titles and Regular Titles on the Internet

<i>Amazon Sales Rank</i>	<i><100,000</i>	<i>>100,000</i>
Average List Price	\$34.53	\$42.18
Average Amazon Price	\$29.26	\$41.60
Average Price at Dealtime	\$29.52	\$39.06
Average Minimum Price at Dealtime	\$20.03	\$29.52
Observations	63	37

With these estimates of elasticity and revenue from obscure book sales, we use equation (9) to calculate that the introduction of obscure books in online markets has increased consumer

¹⁸ Analogous results are obtained using a cutoff of 40,000, the number of books stocked at a typical large independent bookseller.

welfare by between \$731 million and \$1.03 billion in the year 2000 alone, with standard errors of \$46.7 million and \$65.8 million respectively.¹⁹

It is worth noting that, our log-linear demand curve does not restrict consumers' valuation to be below a certain dollar amount. One concern, therefore, is that our consumer surplus estimates could be driven by a small number of consumers with very high valuations. It might be reasonable to exclude some of these very high value consumers on the assumption that they might have been motivated to find a way to gain access to the book without using the Internet, even if entailed significant personal effort. On the other hand, they might never have learned of the book in the first place without the recommendation engines, search tools and other aids provided by successful online booksellers. Discussions with a publisher suggest that this latter effect is more important than any substitution away from conventional channels by high value consumers.

Nonetheless, as a check on the robustness of our results, we can also conduct an analysis in which we restrict our consumer estimates by excluding high value consumers. Excluding all consumers with valuations above five times a book's current price would reduce our current consumer surplus estimates by 28.0%-40.6% while excluding consumer valuations above ten times a book's current price would reduce our current consumer surplus estimates by 16.2%-27.5%.

We also calculate the consumer surplus gain from increased competition and operational efficiency in Internet markets as a point of reference to the consumer surplus gains estimated

¹⁹ Using a cutoff of 250,000 would reduce our consumer surplus estimates to between \$547 and \$772 million in the year 2000, with standard errors of \$37.3 million and \$52.7 million respectively. Using a cutoff of 40,000 would increase our consumer surplus estimates to between \$894 and \$1.26 billion in the year 2000, with standard errors of \$50.4 million and \$71.1 million respectively.

above. Brynjolfsson and Smith (2000) calculated that prices on the Internet were 6-16% lower than prices in brick-and-mortar retailers due to increased competition and increased operational efficiency. A fractional price change of f will lead to a f^*a change in quantity, according to the definition of elasticity of demand. Thus we have

$$CV = -\frac{p_1x_1 - p_0x_0}{1+a} = -\frac{(1-f)p_0(1-f^*a)x_0 - p_0x_0}{1+a}, \quad (14)$$

where CV is the change in consumer surplus, a is the price elasticity, (p_1, x_1) are the price and quantity after the price change, and (p_0, x_0) are the price and quantity before the price change.

Using the same estimates as were used above (i.e., p_0x_0 =\$1.475 billion and a between -1.56 and -1.79), equation (14) shows that the consumer welfare gain from a 6-16% drop in price for all titles on the Internet is between \$73.7 million and \$150 million. Thus, the consumer welfare gain from the introduction of obscure books in online markets is between five to fourteen times as large as the consumer welfare gain from increased competition and lower prices on the Internet.

4.4. Discussion

While the magnitude of the consumer welfare gain from increased variety is large both in absolute terms and relative to the savings from lower prices, our approach is imperfect and is likely to underestimate the total welfare benefits for a number of reasons.

First, it is important to note that the book market is just one of many markets affected. Online sales of other consumer product categories, like music CDs, movies, and electronic products, are likely to also show significant gains in consumer surplus. Furthermore, gains in all product categories will increase as more customers gain access to the Internet channel and as new technologies such as print-on-demand, digital content delivery, mobility services, and broadband

access further reduce consumer search and transactions costs. Finally, it is possible that the ability to sell obscure books through Internet channels that would not have been stocked in physical stores will allow some books to be published that otherwise would not have been viable.²⁰

Second, there is some evidence that the Internet may have reduced the effective cost of special orders even in offline stores, including the consumer time and effort required to identify the relevant books. Some obscure titles were available in brick-and-mortar stores through customer initiated special orders, even before the rise of the Internet as a channel for books. However, according to several bookstore owners we spoke to, special orders for items not normally stocked account for less than 1% of sales through the physical channel. This low level of special orders should not be surprising given that the special order process in a conventional store is inconvenient and time-consuming, as discussed above.

However, it is interesting to note that the availability of obscure titles on the Internet has apparently led to somewhat *increased* sales through special orders at brick-and-mortar stores.

Several brick-and-mortar retailers we spoke to said that the Internet has allowed *brick-and-mortar* customers to locate and evaluate books they would not have been able to find otherwise.

Mary Ellen Keating, Barnes and Noble Senior Vice President of Corporate Communication and Public Affairs put it as follows with regard to sales in Barnes and Noble's brick-and-mortar stores:

²⁰ While making more and more titles available online will result in higher sales, it is important to note that our calculations demonstrate that there are diminishing returns to adding titles. For example, according to our Pareto curve estimates, titles ranked from 100,000 to 200,000 account for 7.3% of sales at Amazon.com while titles ranked between 200,000 to 300,000 account for only 4.6% of sales.

“Sales from special orders are up, and customers are ordering a broader range of titles in a number of different categories. What some customers tend to do is their own research on the Web and then special order the book from our stores.”²¹

If the cost of special orders is unaffected by the Internet, then our consumer surplus calculations can ignore changes in the quantity of special orders, while our estimates will be too low if the effective cost of special orders were reduced as suggested by the preceding quotation.

Lacking precise data on the costs or quantities of special ordering sales of obscure titles at brick-and-mortar stores, this potential consumer welfare gain is left out in our calculation. However, given that the Internet has apparently led to a net increase in special order sales through the physical channel, our calculations will underestimate the true consumer surplus from the availability of obscure titles on the Internet.

The Internet may also lower the cost of placing special orders in other ways. For example, on October 26, 2000 Barnes and Noble announced a plan to install Internet service counters in all its superstores. These service counters would allow in-store customers to place orders from Barnes and Noble’s Internet site for home delivery. While not included in our calculations, the availability of this service will increase consumer surplus by providing Internet access in new and convenient locations and thus lowering the cost of placing special orders to in-store consumers and consumers who otherwise would not have access to the Internet.

Last but not least, our calculations only focus on consumer welfare gains. There may also be significant gains in producer welfare from the additional sales. Indeed, retailers like Amazon, book publishers, printers, and authors, all stand to benefit and earn a slice of the growing pie created by lower search and transactions costs. In contrast, consumer welfare gains from lower

²¹ Source: E-mail communication with Mary Ellen Keating, December 3, 2001.

prices come largely at the expense of producers. This suggests that increased product variety creates a total welfare gain, including both consumer and producer welfare, which exceeds the total welfare gain from lower prices by even more than the ratio we estimated for consumer welfare gains alone. It would be interesting to adapt the methods of this paper to also explore the implications for producer welfare.

5. Conclusions

While lower prices due to increased market efficiency in Internet book markets provide significant benefits to consumers, we find that the increased online availability of previously hard-to-find products represents a positive impact on consumer welfare that is five to fourteen times larger. Limited shelf-space in conventional retail outlets constrains the types of products that can be discovered, evaluated, and easily purchased by consumers. Limits on the number of titles Internet retailers can present and sell to consumers are substantially lower. As a result, Internet customers have easy access to millions of products that they could not easily locate or purchase through brick-and-mortar retailers.

To date, the economic effect of increased product variety on the Internet has been ignored, effectively setting to zero the value consumers place on increased selection at Internet retailers. Recent econometric advances have allowed for the measurement of the economic impact of such new products. Our research applies and extends these methodologies to quantify an important welfare impact of online markets. Preliminary calculations for one product category sold in U.S. markets show that the welfare gains are between \$731 million and \$1.03 billion for the year 2000 alone. These welfare gains dwarf the consumer welfare gain from increased competition and lower prices uncovered in previous research (Brynjolfsson and Smith 2000).

There are a variety of ways our results can be extended by future research. First, while our results use the well-known Lerner index to obtain elasticity calculations, it may be possible for future research to directly estimate these elasticities using an experiment in cooperation with a publisher of obscure books or possibly a retailer with a dominant market position. Such an experiment would change wholesale prices on a randomly selected set of titles and track the resulting levels of demand. These price changes would be exogenous if both the titles and price change levels (positive and negative) were selected at random for the purposes of this experiment.

It also should be possible to extend this methodology to measure welfare contributions of other product categories sold on the Internet or other new products made available through Internet markets. For example, one could easily extend our results to the online sale of movies, music CDs, or consumer electronics products. It also might be possible to estimate consumer surplus gains from the distribution of digital products such as downloadable e-books, music, movies, and software. Moreover, consumers should also benefit from easy access to formerly localized markets such as RealAudio broadcasts of local radio stations or eBay auctions for products that would otherwise have been sold in yard sales. The results of this paper suggest that ultimately the most important benefits of Internet retailing are not fully reflected in lower prices, but rather are due to the new goods and services made readily available to consumers.

References

- Baumol, William, and Edward A. Ide. 1956. Variety in Retailing. *Management Science* **3**(1) 93-101.
- Bianco, Anthony. 1997. Virtual Bookstores Start to Get Real. *Business Week*, October 27.
- Bils, Mark, and Peter J. Klenow. 2001. The Acceleration in Variety Growth. *American Economic Review Papers and Proceedings*, **91**(2) 274-280.
- Boskin Commission Report. 1996. Toward a More Accurate Measure of the Cost of Living, Final Report to the Senate Finance Committee from the Advisory Commission to Study the Consumer Price Index. (Available at <http://www.ssa.gov/history/reports/boskinrpt.html>.)
- Book Industry Study Group. 2001. *Book Industry Trends 2001*. R. R. Donnelley and Sons, Allentown, PA.
- Bresnahan, Timothy F. 1986. Measuring Spillovers from Technical Advance: Mainframe Computers in Financial Services. *American Economic Review* **76**(4) 742-755.
- Bresnahan, Timothy F., and Robert J. Gordon. 1997. *The Economics of New Goods*. The University of Chicago Press, Chicago, IL.
- Broniarczyk, Susan M., Wayne D. Hoyer, and Leigh McAlister. 1998. Consumers' Perceptions of the Assortment Offered in a Grocery Category: The Impact of Item Reduction. *Journal of Marketing Research* **35**(2) 166-176.
- Brown, Jeffrey R., Austan Goolsbee. 2002. Does the Internet Make Markets More Competitive? Evidence From the Life Insurance Industry. *Journal of Political Economy* **110**(3) 481-507.
- Brynjolfsson, Erik. 1995. The Contribution of Information Technology to Consumer Welfare. *Information Systems Research* **7**(3) 281-300.
- Brynjolfsson, Erik, Astrid A. Dick, Michael D. Smith. 2002. Consumer Heterogeneity and Search Costs on the Internet, Working Paper, MIT, Cambridge, MA. (Available at <http://www.heinz.cmu.edu/~mds/chsci.pdf>.)
- Brynjolfsson, Erik, and Michael Smith. 2000. Frictionless Commerce? A Comparison of Internet and Conventional Retailers. *Management Science* **46**(4) 563-585.
- Chevalier, Judith, and Austan Goolsbee, 2003 (Forthcoming). Measuring Prices and Price Competition Online: Amazon and Barnes and Noble. *Quantitative Marketing and Economics* **1**(2).
- Clay, Karen, Ramayya Krishnan, and Eric Wolff. 2001. Price strategies on the web: Evidence from the online book industry. *Journal of Industrial Economics* **49**(4) 521-540.

- Dreze, Xavier, Stephen J. Hoch, and Mary E. Purk. 1994. Shelf Management and Space Elasticity. *Journal of Retailing* **70**(4) 301-326.
- Ehrens, Scott, Adria Markus. 2000. Amazon.com: There's An "R" in E-Tailing. Epoch Partners Consumer Internet Company Report, November 13, Page 4.
- Goolsbee, Austan, and Amil Petrin. 2001. The Consumer Gains from Direct Broadcast Satellites and the Competition with Cable Television. NBER Working Paper No. W8317, Cambridge, MA.
- Greene, William H. 2000. *Econometric Analysis*. Prentice-Hall, Upper Saddle River, NJ.
- Hansell, Saul. 2002. News Analysis: Amazon Looks for an Encore. *New York Times*, Business/Financial Desk, Section C, Page 2, Column 4, January 26.
- Hausman, Jerry A. 1981. Exact Consumer's Surplus and Deadweight Loss. *American Economic Review* **71**(4) 662-676.
- Hausman, Jerry A. 1997a. Valuing the Effect of Regulation on New Services in Telecommunications. *Brookings Papers: Microeconomics*, 1-38.
- Hausman, Jerry A. 1997b. Valuation of New Goods under Perfect and Imperfect Competition. In Bresnahan and Gordon, eds. *The Economics of New Goods*. The University of Chicago Press, Chicago, IL, 209-237.
- Hausman, Jerry A., and Greg Leonard. 2001. The Competitive Effects of a New Product Introduction: A Case Study. *Journal of Industrial Economics* Forthcoming.
- Hausman, Jerry A., and Whitney Newey. 1995. Nonparametric Estimation of Exact Consumers Surplus and Deadweight Loss. *Econometrica* **63**(6) 1445-1476.
- Hicks, John R. 1942. Consumers' Surplus and Index Numbers. *Review of Economic Studies* **9**(2) 126-137.
- Hoch, Stephen J., Eric T. Bradlow, and Brian Wansink. 1999. The Variety of an Assortment. *Marketing Science* **18**(4) 527-546.
- Kahn, Barbara, and Donald R. Lehmann. 1991. Modeling Choice Among Assortments. *Journal of Retailing* **67**(3) 274-299.
- Lee, Tom S.Y., and Ivan Png. 2002. Buyer Switching Costs: Online vis-à-vis Conventional Retailing. Working Paper, National University of Singapore, Singapore. (Available at <http://www.comp.nus.edu.sg/~ipng/research/books.pdf>.)
- Lyster, Michael. 1999. Printed Words Get Redefined In Digital Age. Investor's Business Daily, Page A6, Computers and Technology Section, June 2.
- Mendelson, Haim, Philip Meza. 2002. Amazon.com: Marching Toward Profitability. Stanford Graduate School of Business, Case EC-25, July.

- National Bureau of Economic Research. 1961. *The Price Statistics of the Federal Government: Review, Appraisal, and Recommendation*. Price Statistics Review Committee, General Series, No. 73, New York, NBER.
- Nevo, Aviv. 2001. *New Products, Quality Changes and Welfare Measures Computed From Estimated Demand Systems*. NBER Working Paper No. W8425, Cambridge, MA.
- Owen, Thomas. 2002. *Lord of the Things*, *Business 2.0*, March.
- Pareto, Vilfredo. 1896. *Cours D'Economie Politique*. In Bousquet, G.H. and G. Busino eds. *Oevres Completes de Vilfredo Pareto*, 1. Librairie Droz, Geneva, 1964. Originally published 1896.
- Petrin, Amil. 2001. *Quantifying the Benefits of New Products: the Case of the Minivan*. NBER Working Paper No. W8227, Cambridge, MA.
- Poynter, Daniel, "Publishing Poynters," April-June 2000, <<http://parapub.com/getpage.cfm?file=newsletter/News0400.html>>, accessed February 26, 2003.
- Professional Publishing Report. 1999. "University Presses Credit Internet For Increased Sales," Volume 3, Number 2, January 29.
- Ritchie, Harry. 1999. *Fully Booked*. *Mail on Sunday*, Page 22, February 28.
- Scott Morton, Fiona M., Florian Zettelmeyer, Jorge Silva-Risso. 2001. *Internet Car Retailing*. *The Journal of Industrial Economics*, **49**(4) 501-520.
- Smith, Michael D. 2001. *The Law of One Price? Price Dispersion and Parallel Pricing in Internet Markets*. Working Paper, Carnegie Mellon University, Pittsburgh, PA.
- Varian, Hal. 1992. *Microeconomic Analysis*, Page 167-168. Norton & Company. New York, NY.
- Van Herpen, Erica, and Rik Pieters. 2002. *Research Note The Variety of an Assortment: An Extension to the Attribute-Based Approach*. *Marketing Science* **21**(3) 331-341.
- Weingarten, Gene. 2001. *Below the Beltway*, *Washington Post*, June 17, 2001.
- Zipf, George. 1949. *Human Behavior and the Principle of Least Effort*. Addison-Wesley, Cambridge, MA.

Appendix A: Summary of Derivation and Necessary Data

1. Derivation of the formula to calculate consumer surplus

(Follows Hausman 1981 and Hausman and Leonard 2000)

$$\begin{array}{l}
 CV = e(p_{e0}, p_{n0}, u_1) - e(p_{e1}, p_{n1}, u_1) \\
 \downarrow \\
 CV = [e(p_{e1}, p_{n0}, u_1) - e(p_{e1}, p_{n1}, u_1)] \\
 \quad + [e(p_{e0}, p_{n0}, u_1) - e(p_{e1}, p_{n0}, u_1)] \\
 \downarrow \\
 \text{Assume } p_{e0} = p_{e1} = p_e \\
 \downarrow \\
 CV = e(p_e, p_{n0}, u_1) - e(p_e, p_{n1}, u_1) \\
 \quad = e'(p_{n0}, u_1) - e'(p_{n1}, u_1)
 \end{array}
 \qquad
 \begin{array}{l}
 \text{Assume } x(p, y) = Ap^a y^d \\
 \downarrow \\
 x(p, y) = -\frac{\partial v(p, y) / \partial p}{\partial v(p, y) / \partial y} \\
 v(p, y) = -A \frac{p^{1+a}}{1+a} + \frac{y^{1-d}}{1-d} \\
 \downarrow \\
 e(p, u) = \left[(1-d) \left(u + \frac{Ap^{1+a}}{1+a} \right) \right]^{1/(1-d)}
 \end{array}$$

$$\begin{array}{c}
 \swarrow \quad \searrow \\
 CV = \left[\frac{1-d}{1+a} y^{-d} (p_{n0} x_0 - p_{n1} x_1) + y^{(1-d)} \right]^{1/(1-d)} - y \\
 \downarrow \quad \text{Assume } d=0 \quad p_{n0} x_0 = 0 \\
 CV = -\frac{p_{n1} x_1}{1+a}
 \end{array}$$

$$\begin{array}{l}
CV = e(p_{e0}, p_{n0}, u_1) - e(p_{e1}, p_{n1}, u_1) \\
\downarrow \\
CV = [e(p_{e1}, p_{n0}, u_1) - e(p_{e1}, p_{n1}, u_1)] \\
\quad + [e(p_{e0}, p_{n0}, u_1) - e(p_{e1}, p_{n0}, u_1)] \\
\downarrow \\
\text{Assume } p_{e0} = p_{e1} = p_e \\
\downarrow \\
CV = e(p_e, p_{n0}, u_1) - e(p_e, p_{n1}, u_1) \\
= e'(p_{n0}, u_1) - e'(p_{n1}, u_1)
\end{array}
\qquad
\begin{array}{l}
\text{Assume } x(p, y) = Ap^a y^d \\
\downarrow \quad x(p, y) = -\frac{\partial v(p, y) / \partial p}{\partial v(p, y) / \partial y} \\
v(p, y) = -A \frac{p^{1+a}}{1+a} + \frac{y^{1-d}}{1-d} \\
\downarrow \\
e(p, u) = \left[(1-d) \left(u + \frac{Ap^{1+a}}{1+a} \right) \right]^{1/(1-d)}
\end{array}$$

$$\begin{array}{c}
\swarrow \quad \searrow \\
CV = \left[\frac{1-d}{1+a} y^{-d} (p_{n0}x_0 - p_{n1}x_1) + y^{(1-d)} \right]^{1/(1-d)} - y \\
\downarrow \quad \text{Assume } d=0 \quad p_{n0}x_0=0 \\
CV = -\frac{p_{n1}x_1}{1+a}
\end{array}$$

Notes:

- If post-introduction prices of existing products are lower than pre-introduction prices (i.e., $p_{e0} > p_{e1}$) results under equality assumptions will underestimate true consumer surplus.
- If $d > 0$ (i.e., if the good is a luxury good as opposed to a necessity good) results under $d=0$ will underestimate true consumer surplus.

2. Derivation of the price elasticity of aggregate demand

Assume $p_{wi}=p_w$ → $p_w=k_i p_{ri}$

Assume $p_{wi}=k_i p_{ri}$ → $\bar{p} \equiv \sum_{i=1}^N s_i p_{ri} = \sum_{i=1}^N \frac{s_i}{k_i} p_w$

$s_i \equiv \frac{q_{ri}}{q_r}$ → $\bar{p} \equiv \sum_{i=1}^N s_i p_{ri}$

Assume $q_{wi}=q_{ri}$ → $q_r=q_w$

$q_r \equiv \sum_{i=1}^N q_{ri}$ → $q_r=q_w$

$q_w \equiv \sum_{i=1}^N q_{wi}$ → $q_r=q_w$

$$\frac{\bar{p} dq_r}{q_r d\bar{p}} = \frac{\bar{p} dq_w}{q_w d\bar{p}} = \frac{\left(\sum_{i=1}^N \frac{s_i}{k_i} p_w \right) dq_w}{q_w \left(\sum_{i=1}^N \frac{s_i}{k_i} dp_w \right)} = \frac{\left(\sum_{i=1}^N \frac{s_i}{k_i} \right) p_w dq_w}{q_w \left(\sum_{i=1}^N \frac{s_i}{k_i} \right) dp_w} = \frac{p_w dq_w}{q_w dp_w}$$

3. Derivation of the formula to calculate total sales of obscure products on the Internet

$$\log(\text{Quantity}) = \mathbf{b}_1 + \mathbf{b}_2 \cdot \log(\text{Rank}) + \mathbf{e}$$

↓
Estimate of β_2

$$r(x, N) = \frac{\int_1^N \mathbf{b}_1 t^{\mathbf{b}_2} dt}{\int_1^x \mathbf{b}_1 t^{\mathbf{b}_2} dt} = \frac{N^{(\mathbf{b}_2+1)} - x^{(\mathbf{b}_2+1)}}{N^{(\mathbf{b}_2+1)} - 1}$$

↓

$$p_{n1} x_1 = (\overline{PQ}) * r(x, N)$$

4. Data Requirements and Potential Sources:

Data Requirement	Potential Data Sources
Elasticity of Aggregate Demand (a)	<ul style="list-style-type: none"> • Lerner index and information on publisher/manufacturer margins. • Experimental estimation through partnership with publisher/manufacturer (see pp. 10, 25-26). • Experimental estimation through partnership with retailer with a dominant market position (see pp. 10, 25-26). • Industry estimates of elasticities.
Total sales of products on the Internet (\overline{PQ})	<ul style="list-style-type: none"> • Market research firms. • Department of commerce reports. • Industry consortia.
Proportion of sales of obscure products at Internet retailers ($r(x,N)$)	<ul style="list-style-type: none"> • Estimation using log-linear relationship between sales rank and sales and data obtained from publisher/manufacturer (pp. 14-16). • Experimental estimation using power-law relationship between sales rank and sales and observation of changes in rank after purchase of goods (Chevalier and Goolsbee 2003, pp. 16-17). • Direct observation from representative publisher/manufacturer. • Direct observation from retailer (or retailers) with dominant market share.
Total sales of obscure products on the Internet ($p_n x_I$)	<ul style="list-style-type: none"> • Total sales of product category on the Internet * proportion of sales of obscure products at Internet retailers. • Direct observation from representative publisher/manufacturer. • Direct observation from retailer (or retailers) with dominant market share.

Potential data sources in bold indicate those used in this paper.