Available online at www.sciencedirect.com
SCIENCE DIRECT®

Journal of Urban Economics 56 (2004) 1–24

JOURNAL OF Urban Economics

www.elsevier.com/locate/jue

Geography and the Internet: is the Internet a substitute or a complement for cities?

Todd Sinai and Joel Waldfogel *

The Wharton School, University of Pennsylvania, Philadelphia, PA, USA
NBER, Cambridge, MA, USA

Received 6 October 2003; revised 29 March 2004

Available online 28 May 2004

Abstract

We study the tendency to connect to the Internet and the online and offline shopping behavior of connected persons. We document that larger markets have more locally-targeted online content and that individuals are more likely to connect in markets with more local online content, suggesting the Internet is a complement to cities. Yet, holding local online content constant, people are less likely to connect in larger markets, indicating that the Internet is also a substitute for cities. Finally, we find that individuals connect to overcome local isolation, in the form of racial isolation or distance to retail stores.

Traditionally, markets for news and information as well as some retail goods have been predominantly local. As a result, consumers' welfare has been limited by the size of their local market, and agglomeration of persons sharing similar preferences has improved their welfare by facilitating the provision of products they want.[1] By agglomerating persons around the country—indeed, around the world—into a single market, the Internet offers the potential to radically alter consumption possibilities. In particular, the Internet may

* Corresponding author.
E-mail address: waldfogj@wharton.upenn.edu (J. Waldfogel).

[1] We are aware of other arguments for agglomeration besides consumption. Henderson [11] presents a model in which city size balances production benefits against congestion costs. See, for example, Ciccone and Hall [2] or Rosenthal and Strange [13] for recent empirical evidence on the production benefits of agglomeration as well as additional references. There is little research on consumption benefits of agglomeration. Glaeser et al. [9] is an exception.

0094-1190/$ – see front matter
doi:10.1016/j.jue.2004.04.001

serve as a substitute for urban agglomeration by leveling the consumption playing field between large, variety-laden and small, variety-starved markets. However, leveling the field requires that content on the Internet be similarly attractive to persons in large and small markets. If the Internet offers local, as well as general, information, then its role as a substitute for agglomeration will be undermined. Indeed, if local online content is sufficiently attractive—and if it is more prevalent in larger markets—then the Internet may be a *complement* for urban agglomeration.[2]

In this paper we use information on how Internet connection tendencies, and connected persons' online and offline retail spending, are related to the attractiveness of their online and offline options to draw inferences about whether the Internet serves as a substitute or a complement for urban agglomeration. First, we examine the relationship between market size and available locally-targeted online content. To this end we characterize the nature of available content using Media Metrix data on 16.5 million web page visits by about 32,000 households in 326 Metropolitan Statistical Areas (MSAs) in August 2000. We document that substantially more online local content is available in larger markets, suggesting that Internet use may complement urban agglomeration.

We then examine how the tendency to connect to the Internet varies with metro area population—a proxy for available product variety—and locally targeted online content. Combining the Current Population Survey's August 2000 Computer and Internet Use module with our Media Metrix-derived data on local online content and Census data on local market size (as a proxy for offline product variety), we document that local online content attracts people to connect. When we separately account for both local online content and our measure of local offline variety (population), we find statistically significant direct evidence of both complementarity and substitutability. Individuals are more likely to connect in markets with more local online content, and holding local online content constant, are less likely to connect in larger markets. On balance we find that these effects offset each other so that the Internet neither promotes nor discourages agglomeration in larger markets.

Second, we shed light on the Internet's substitution for offline variety by examining the relationship between the tendency to connect and individuals' local isolation from persons sharing their preferences. The problem of dissatisfaction with offline options can be particularly acute for persons either isolated from products generally or with preferences unlike their neighbors'. It is clear from existing research that blacks and whites have sharply different preferences in media products. As a result, the local availability of products that blacks value may have been limited not simply by the absolute size of their local markets, but also by the size of the local community sharing their preferences.[3] If isolated blacks are more likely than blacks living among large concentrations of blacks to connect to the

[2] We are not the only authors to pose this question. See Kolko [12], as well as Gaspar and Glaeser [7], who find that telephones complement agglomeration because phone conversation complements face-to-face (two-way) communication. Our focus in this paper is on one-way communication over the web rather than two-way communication, but our question is similar. Finally, Forman et al. [5] examine a related question, the effect of urban location on diffusion of Internet technologies.

[3] For example, larger markets have more and better local newspapers (George and Waldfogel [8]), radio broadcasts (Waldfogel [18]), and television stations (Waldfogel [19]). These studies document that black and

Internet, then we can infer not only that the Internet is a substitute for cities but also that it is a substitute for groups of people within a market sharing in common their locally atypical product preferences. Again using the CPS data, we find that blacks are more likely to connect, relative to whites, as blacks have smaller local population, suggesting that the Internet is a substitute for the agglomeration of persons sharing similar preferences.

Finally, we examine effects of isolation measured by distance to retail establishments among a sample of Internet-connected respondents to the Wharton Virtual Test Market (WVTM) survey. Using 1997 Economic Census data on the location and size of book, music, and clothing stores, matched by zip code to data on WVTM respondents' online, catalog, and offline spending in each of those categories, we document that people buy more books and clothing online or via catalogs, relative to their offline expenditure in local stores, as they live farther from their nearest book and clothing store, respectively. This evidence shows that, at least in some categories, consumers use the Internet as a substitute for proximity to retail stores.

Friedman [6] has argued that each person gets what she wants through market allocation, so that markets avoid the tyranny of the majority endemic to collective choice. Friedman's argument holds literally only when production can take place at arbitrarily small scale, so that available product variety does not depend on the size, or the preference composition of potential customers in the market. When fixed costs are sizable, the number of available products, and the resultant welfare of consumers in local markets can depend on the size and composition of the local market. By agglomerating consumers into larger markets, the Internet allows locally isolated persons to benefit from the product variety made available for consumers elsewhere. By increasing the size of markets relative to fixed costs, the Internet may therefore bring market allocation nearer to the ideal in which an individual's welfare does not depend on either the number of her neighbors or their product preferences.

The paper proceeds in six sections. Section 1 reviews available evidence on product variety and market size and characterizes the decision to use the Internet, as a function of one's preference type and the quality of local options. Section 2 describes the Current Population Survey (CPS), Media Metrix (MM), WVTM, and Economic Census data used in this study. Section 3 employs the MM data to quantify local content on the Internet and, in particular, to characterize how the availability of local content varies with market size. Section 4 employs the CPS data to characterize the demand for Internet connection. Section 5 examines how the balance of online and offline retail spending varies with connected consumers' distance to retail establishments. A brief conclusion follows.

1. How does the Internet function as a substitute or complement for cities?

1.1. The Internet as a substitute for cities

When production entails fixed costs and preferences differ across individuals, the number of differentiated product options available locally will increase in the size of

white consumers' welfare, in their capacity as media consumers, increase in the size of their own respective group populations.

the market.[4] Larger markets have more local product variety than small markets, and this greater variety draws a higher fraction of persons to consumption of local offline products. In this way persons benefit each other through what has elsewhere been termed a "preference externality" (Waldfogel [18]).

By aggregating people in disparate locations into a single market, the Internet has the capacity to increase market size relative to fixed costs.[5] This can, in turn, raise the number of available products and reduce the dependence of consumption options on the number and mix of consumers in one's local market. That is, consumers in small offline markets can instead turn to the Internet for products unavailable offline locally. What sorts of sites make the Internet a substitute for cities? Retail sites provide one clear possibility: a person who has no store nearby can instead buy online. We also have in mind sites that offer content that is not geographically specific but which may have greater appeal in smaller markets with less offline product variety. For example, Spinner.com offers 140 channels of streaming music programming, over twice the number of radio stations available in any of the largest US markets. Spinner.com may appeal to listeners in both large and small markets but presumably provides more of a benefit to listeners in small markets with few traditional radio stations. News sites, such as CNN.com or MSNBC.com, present domestic and international news of interest to individuals in cities of all sizes. But because small markets tend to have slender local newspapers, people who live in them may place a higher value on the availability of news on the Internet.

1.2. The Internet as a substitute for agglomeration of like-minded persons

The paucity of offline product variety is not determined solely by the total population in an area. To the extent that preferences differ across types of individuals, the number of like-minded persons in a local area will determine the size of the offline market and the amount of locally available offline products that would appeal to those people. Since the distribution of types differs across geographic markets, we expect persons to be more likely to connect to the Internet to satisfy their locally unfulfilled tastes when they are "preference minorities," that is, part of a group with distinct preferences that makes up a small number of the local population. For example, it is well documented that blacks and whites have sharply different preferences in some categories of products. In particular, the radio programming formats attracting two thirds of black listening collectively attract less than 2 percent of nonblack listening. In major cities with both tabloid and non-tabloid newspaper options, the tabloid attracts about three quarters of black readers, compared with about a third of nonblack readers.

[4] This is an example of what Spence [14,15] terms the "product selection problem." See also Dixit and Stiglitz [3].

[5] Computer technology may also reduce the absolute size of at least the exogenous component of the fixed and sunk costs of operating a business. Given the large advertising expenditures of web retailers such as Amazon.com, it is not clear whether web businesses have lower fixed costs than bricks and mortar businesses, when endogenous fixed costs are taken into account. See Sutton [16] for extensive discussion of endogenous fixed costs.

1.3. The Internet as a complement for cities

In addition to providing universal content that appeals to individuals in any size market, the Internet may also be a local medium. The Internet can provide information that helps people to navigate cities, and may deliver other goods and services that improve city life. For example, city portals, such as boston.citysearch.com, provide information about events, restaurants, and movie listings. Match-making sites, such as boston.matchmaker.com, help users in large cities meet each other. And sites associated with local newspapers or television stations provide another distribution channel for local news. If there are fixed costs associated with producing such content, then the number and variety of local sites may increase in the size of the local market, making the Internet more useful to people in larger markets and mitigating the Internet's role as a substitute for local offline product variety. Moreover, much local online content, for example at local newspaper sites, largely recycles the content produced for their traditional offline products. To the extent that the underlying traditional products grow richer in market size, the online products will as well.

Just as Internet access can complement cities generally, web retailers can complement proximity to offline retailers. Many web retailers, such as gap.com, allow items purchased online to be returned or exchanged at their offline stores.[6]

These considerations motivate the four questions that this study addresses. How much web content is local? Is there more local online content in larger markets? How does the tendency to connect—and how do uses of the Internet—vary with one's local offline and online options?

2. Data

Data for this study are drawn from four sources, the August 2000 Current Population Survey Computer and Internet Use Supplement, an August 2000 Media Metrix data extract, the Wharton Virtual Test Market (WVTM), and the 1997 Economic Census. The CPS supplement has information on Internet connection, as well as demographic and geographic information, for roughly 50,000 households in August 2000. We reduce our sample to 29,027 by restricting our attention to those households that live in metropolitan statistical areas (MSAs) that can be matched to MSAs in our Media Metrix data set.

The first column of Table 1 reports sample characteristics. Because we are interested in factors that may affect the likelihood of a household using the Internet, we first define a household as Internet connected if the household reports Internet access from home. Almost 44 percent of the households in our sample have such connections. Just over half the sample has one or more computers at home. Of the household heads in the sample, more than 13 percent are black, and 30 percent are at least college-educated.

[6] According to the Gap: "All online merchandise can be returned to any US Gap store or by mail. Make sure to choose the "gift receipt" option at checkout if you are sending the gift directly to the recipient" (according to http://www.gap.com/asp/home_gap.html?wdid=0, accessed on September 17, 2003).

Table 1
CPS and Media Metrix sample characteristics

	August 2000 CPS		MM	WVTM
	All heads of households	Connected household heads		
Internet at home	43.7	100.0	100.0	100.0
Computer(s) at home	54.3	98.8	100.0	
White	82.2	87.3	88.7	85.1
Black	13.4	7.4	4.5	3.4
Native American	0.6	0.4	N/A	0.5
Asian	3.8	4.9	3.0	5.6
Less than high school	14.7	4.1	2.0	1.8
High school	28.5	20.1	9.7	19.3
Some college	26.7	30.3	27.0	43.4
College	19.6	28.7	36.4	23.4
Post graduate	10.5	16.7	24.8	12.1
Observations	29,027	12,685	17,104	21,309

Note. CPS sample includes only households in MSAs that match with Media Metrix sample (and contain at least 20 MM households). Media Metrix sample includes only households in MSAs. In both samples, included observations must have valid entries for age, education, and race of the household head. Media Metrix education is the maximum educational attainment of either spouse if the household is married.

The household heads in the portion of the sample that is Internet-connected is disproportionately white and is more highly educated than the population as a whole.[7] Only about 7 percent of the connected sample is black-headed, and approximately 45 percent of the households are headed by a college-educated person. Naturally, nearly all of the connected sample of households has a home computer.

Our second data source follows the Internet usage behavior of a panel of households. Media Metrix collects data on all web page visits by a representative sample of households in 326 MSAs by placing recording software on panelists' computers. In our extract, which covers August 2000, each visit to a web page, or page "hit," by a household is a separate record—with 16.5 million page visits in total during the month.[8] Media Metrix appends basic information about their Internet-connected panelists, such as MSA, income category, educational attainment, and race, to their data on web surfing. The third column of Table 1 reports the education distribution for the Media Metrix sample, and it is similar to the CPS sample of households with Internet connections at home (in column 2). The Media Metrix sample is a little more highly educated than the CPS sample, but that result is generated in part by our applying the highest level of educational attainment of any member in the household to the whole household.

[7] These results are consistent with evidence elsewhere on the digital divide. See US Department of Commerce [17, p. 8] (accessed at http://www.ntia.doc.gov/ntiahome/fttn99/FTTN.pdf).

[8] A "site," in our extract, is typically a three-level name, such as www.aol.com. The data contain other sites at America On-line (AOL), such as members.aol.com, as separate "sites," even though they are in the same "domain."

In addition, for each site visit, we observe the URL, or "address" of the web page, which Media Metrix classifies into one of 27 categories.[9] The first column of Table 2 reports the distribution across these categories of page hits and numbers of visited sites.[10] The category with the most hits is portals, with 21.9 percent of the total hits, followed by services (15.9), entertainment (13.7), adult content (8.0), retail (5.4), and auctions (4.3). Business-to-business sites have the fewest hits by our sample of residential households with only 0.1 percent of the total. Turning to the number of unique sites visited in each category, reported in column 3, "adult content" has the largest share of sites with 23.0 percent of the 22,432 total sites the Media Metrix panelists visited. However, the portals category, which received a large portion of the total number of page hits, comprises a very small fraction of the sites in the sample, just 1.0 percent. This pattern indicates that there is more concentration in the portals, with a few sites each receiving a large amount of traffic. We return to the latter part of Table 2, on local sites, later in the paper.

The third data source is the WVTM, which is an Internet-based survey of more than 20,000 Internet users. For this project, we used data from the third and fourth waves, which were collected during 1999 and 2000, respectively. The WVTM is designed to be representative, but is not random. Survey participants were recruited by banner advertisements. The ad placements were targeted to recruit a demographically representative population. This survey is well-suited for our purposes since it breaks down how much money the respondent spent over the last two months via Internet sites, catalogs, and physical stores on books, music, and clothing. The WVTM also records the respondent's zip code, which enables us to match the spending data to geographical characteristics. In addition, the survey records a host of demographic information, including gender, race, income, and education. The fourth column of Table 1 shows demographic characteristics of the individuals in the WVTM sample. The education and race distributions are at least broadly similar across connected samples.

The fourth and final data source in the study is the 1997 Economic Census, one wave of a census of business establishments that is collected every five years by the Bureau of the Census. For 1997, the Census department surveyed more than 5 million establishments with employees and collected administrative data on 15.5 million additional small establishments. For each establishment, which is a physical location providing a service or making or distributing goods, Census recorded its location, industry, dollar volume of sales, number of employees, and payroll.

For each line of business (NAICS code) and zip code, the Census reports the number of stores in each of the following size classes, by thousands of dollars of annual retail sales: 0–100, 100–200, 200–500, 500–1000, and over 1 million. In this study we are interested in books (NAICS code 451211), music (NAICS code 45122), and clothing stores

[9] The Media Metrix categories are: Adult Content, Auctions, Automotive, Business to Business, Business/Finance, Careers, Community, Corporate Presence, Directories/Resources, Education, Entertainment, Government, Health, Hobbies, Lifestyle, News/Information, Portals, Real Estate, Regional/Local, Retail, Search/Navigation, Sports, Technology, Travel, Services, All other.

[10] In order to be included in the total hits or sites calculations, a site must be estimated to have received at least 5000 hits during the month.

Table 2
Distribution of hits and sites by category

Category	All sites				Local sites					
	Hits	Pct. of total (%)	Sites	Pct. of total (%)	Hits	Pct. of total (%)	Pct. of cat. hits (%)	Sites	Pct. of total (%)	Pct. of cat. sites (%)
Adult content	1,317,683	8.0	5154	23.0	106,362	12.2	8.1	1642	22.6	31.9
Auctions	715,491	4.3	71	0.3	3106	0.4	0.4	24	0.3	33.8
Automotive	115,769	0.7	215	1.0	1984	0.2	1.7	42	0.6	19.5
Business to business	24,218	0.1	83	0.4	943	0.1	3.9	25	0.3	30.1
Business/Finance	508,931	3.1	491	2.2	43,067	4.9	8.5	122	1.7	24.8
Careers	144,138	0.9	130	0.6	1086	0.1	0.8	49	0.7	37.7
Community	338,323	2.1	324	1.4	29,845	3.4	8.8	137	1.9	42.3
Corporate presence	505,089	3.1	1061	4.7	11,663	1.3	2.3	266	3.7	25.1
Directories/Resources	417,353	2.5	1258	5.6	43,754	5.0	10.5	462	6.4	36.7
Education	234,527	1.4	628	2.8	87,553	10.0	37.3	206	2.8	32.8
Entertainment	2,259,803	13.7	2460	11.0	77,602	8.9	3.4	1061	14.6	43.1
Government	146,018	0.9	190	0.8	19,060	2.2	13.1	38	0.5	20.0
Health	78,117	0.5	302	1.3	4386	0.5	5.6	82	1.1	27.2
Hobbies	171,222	1.0	341	1.5	6831	0.8	4.0	103	1.4	30.2
Lifestyle	479,021	2.9	970	4.3	46,083	5.3	9.6	338	4.6	34.8
News/Information	264,938	1.6	395	1.8	40,043	4.6	15.1	195	2.7	49.4
Portals	3,607,436	21.9	227	1.0	16,884	1.9	0.5	59	0.8	26.0
Real Estate	71,982	0.4	67	0.3	2728	0.3	3.8	15	0.2	22.4
Regional/Local	72,081	0.4	230	1.0	26,426	3.0	36.7	122	1.7	53.0
Retail	896,480	5.4	1313	5.9	20,122	2.3	2.2	323	4.4	24.6
Search/Navigation	409,879	2.5	197	0.9	17,107	2.0	4.2	61	0.8	31.0
Sports	279,152	1.7	468	2.1	17,922	2.0	6.4	229	3.1	48.9
Technology	152,928	0.9	280	1.2	2942	0.3	1.9	105	1.4	37.5
Travel	262,862	1.6	324	1.4	2389	0.3	0.9	62	0.9	19.1
Services	2,616,051	15.9	2489	11.1	145,265	16.6	5.6	733	10.1	29.4
All other	406,822	2.5	2764	12.3	100,180	11.4	24.6	772	10.6	27.9
Total	16,496,314	100.0	22,432	100.0	875,333	100.0	5.3	7273	100.0	32.4

Note. To be included in this table, a local site must have received at least 5000 hits when scaled up to match the Internet-connected population.

Table 3
Number of large and total stores nearby

Within:	Books		Music		Clothing	
	Stores	Large stores (> $1M)	Stores	Large stores (> $1M)	Stores	Large stores (> $1M)
1 mile	1.37	0.47	0.98	0.33	9.65	2.90
2 miles	3.59	1.29	2.55	0.87	26.15	8.18
5 miles	10.29	3.87	7.51	2.62	80.36	25.59
10 miles	27.26	10.41	20.34	7.11	224.57	71.62
20 miles	67.73	25.96	52.00	18.07	583.95	186.85

Note. $N = 18{,}069$.

(NAICS code 4481).[11] Matching the zip codes of the WVTM respondents to the zip code distribution of stores, we can calculate the numbers of book, music, and clothing stores within any distance for each individual in the sample.[12] Table 3 describes some of these data. For example, the first entry in the first column indicates that persons in the sample have an average of 1.37 bookstores within a one-mile radius. If we restrict attention to large bookstores (with over $1 million in annual sales), the number is 0.47 bookstores. Music stores are slightly less proximate than bookstores—an average person has 1 within a mile, while clothing stores are much more dense: the average is nearly 10 within a mile.

3. How much content is local?

To determine whether the Internet is a complement to cities, we need to measure the amount of local content targeted at each metropolitan area. As our metric, we count the number of sites that produce content that appeals primarily to one particular market. Unfortunately, there exists no comprehensive list of sites by locale from which one could characterize local content.[13] Indeed, one cannot determine the localness of a site's targeting from the registration location of a site, or where the parent company's headquarters are located, since the site's visitors could be from anywhere.

Fortunately, we can use the Media Metrix data to measure the geographic focus of a site. By recognizing that a locally targeted site must have a primarily local audience, we can use the geographic origin of a site's visitors as reported in the Media Metrix data to estimate the extent of its local focus and which market it primarily serves. In essence, after calculating every site's share of each market's total page hits, we presume that a site that has a sufficiently high proportion of its total market shares across all markets coming from just one market must be providing content of local interest to that market.

[11] We experimented with alternate definitions: including department stores, superstores, and electronics stores in music, and including department and superstores with clothing. In all cases, results were very similar to those reported.

[12] Distance is measured between the centroids of the two zip codes.

[13] Kolko [12] uses the list of registered domain names and shows higher "domain density" in larger markets. Domain registration indicates the geographic location of the registrant, not the site users, however, so that unlike our study, Kolko provides no direct evidence on how consumer Internet use relates to market size.

To make this more concrete, we compute each site's index of "site localness," *LI*, as an inverse HHI of site localness. To see how this works, first consider a simplified version:

$$LI_i = 1 \Big/ \sum_j \left(\frac{p_{ij}}{\sum_j p_{ij}} \right)^2,$$

where p_{ij} is page visits to site i from MSA j. The numerator is something like visits to cnn.com from Boston, while the denominator is total visits across all MSAs to cnn.com. Since cnn.com gets visitors from all over, the denominator will be a large number, and the localness index will be low.[14] On the other hand, if all visits to a site come from one market, the denominator is 1, as is the localness index.

The actual index we use is slightly more complicated because the Media Metrix sample size varies across MSAs in rough proportion with population. Differing market sizes would skew the raw index toward making sites appear local. To correct this problem, we normalize the pages visits to each site from each MSA by the number of MSA households in the sample, giving:

$$LI_i = 1 \Big/ \sum_j \left(\frac{p_{ij}/HH_j}{\sum_j (p_{ij}/HH_j)} \right)^2,$$

where HH_j is the number of Media Metrix households in MSA j. We define a site as local if it has a "localness index" of two or less, and we attribute the site to the locale, j, that contributes the site's largest market share.[15]

There is one more complication. Because our measure of localness depends on the composition of a site's audience, our accuracy in classifying sites as local diminishes for sites with very few visits from the Media Metrix panel during the month. For example, many sites have one (possible stray) page visit in the entire sample. Since the single page visit comes, by definition, from someone in a particular market, our approach will treat a rarely visited site as local, even if it is not. To counteract that problem, we require that sites exhibit a minimum absolute level of interest, as measured by page hits, in the MSA that they target in order to be considered local to that market. Since Media Metrix does not sample exactly proportionally to market population, we estimate the total number of hits to a site by multiplying the number of hits per household in Media Metrix by the number of connected households in the MSA, as calculated from the CPS. We then include only sites with more than 5000 overall estimated hits. Finally, we also exclude MSAs with fewer than 20 Media Metrix households from these calculations in order to reduce the error in measuring hits per household. That leaves us with 113 MSAs that we are able to match to the CPS data. These MSAs vary in size from about 200,000 people to 18 million. The median MSA has 641,000 people, and the 25th and 75th percentile MSAs in our sample have 393,000 and 1.3 million, respectively.

[14] If the tendency to visit is equal across markets, the localness index will be the number of markets.

[15] The most obvious type of local site that we misclassify as not local are sites which contain local information for a number of locales. Since we require that a site be targeted to one locale to be defined as local, these sites do not qualify. However, we suspect the narrowly targeted sites are a reasonable proxy for the localness patterns we would find if we used a more broad definition.

When we look at the sites with the largest and smallest adjusted localness indices across Media Metrix categories, the index produces reasonable results. The least local sites have localness indices over 20 and include such familiar sites as CBS.com (entertainment), microsoft.com (corporate presence), ebay.com (auctions), autobytel.com (automotive), theglobe.com (communities), SSA.gov (government), MSNBC.com (news/information), netscape.com (portals), buy.com (retail), jobsonline.com (careers), google.com (search/navigation), and SmarterLiving.com (travel).[16]

The most local sites have localness indices close to one and tend to refer to locales in their URL. For example, siliconalleyjobs.com (careers), state.de.us (government), ncweb.com (communities), baltimoresun.com (news/information), sanantonio360.com (portal), corealty.com (real estate), searchchicago.com (regional/local), and uscfootball (sports) all have localness indices equal to one meaning that all their hits came from one MSA. Many other local sites are associated with local newspapers or television stations.

Columns 5–10 of Table 2 show the distributions of hits to local sites, and local sites themselves, across Media Metrix categories. On average, 5.3 percent of all hits are to the 7273 sites that we deem local. In column 7, more than a third of the hits to sites in the education and regional/local categories are deemed local. Nearly half of the news/information sites are local, as reported in the last column, though they account for just 15 percent of the category's hits. Auctions has the lowest share of its hits considered to be local, followed by portals. Travel and automotive has the lowest rate of its sites being local.

The average number of local sites per MSA, reported in Table 4, follows the same general pattern. There is an average of 64 local sites in each MSA, with a maximum of 841 local sites in the New York–Northern New Jersey–Long Island, NY–NJ–CT–PA CMSA. The largest individual categories are Adult Content and Entertainment.

We finish our description of the local site data by asking how the amount of local content varies with the size of the local market. If there are fixed costs of providing a local online site, then the quantity of local online content will increase in the size of the local connected market. And, in turn, the heightened content availability in larger markets will reinforce the Internet as a complement for cities. In traditional media, as we have mentioned above, larger markets have more local content (more radio stations, more and better local newspapers, more television stations). Are there similarly more local online sites in larger markets?[17]

Table 5 reports regressions of MSA local online content, overall and by the larger categories, on total MSA population. An additional million residents in an MSA adds 48 total local sites. To get a sense of the magnitudes, note that the 75th percentile population MSA (with 1.3 million people) is predicted to have about 45 more local sites than the 25th percentile MSA (with 0.4 million), and the mean number of local sites is 64. The relationship between MSA population and the number of local sites is positive and significant for all of the larger Media Metrix categories. That larger markets have more Internet content indicates that the Internet is not simply a leveler of utility across geography

[16] We have suppressed the "www" site name prefix for clarity.

[17] The empirical question addressed in this section mirrors the question of the entry literature: how does the number of firms (or products) vary with market size. See Bresnahan and Reiss [1], for example.

Table 4
Average number of local sites per market, August 2000

Category	Average	Minimum	Maximum
Total local sites	64.34	0	841
Auctions	0.21	0	5
Automotive	0.37	0	5
Business/Finance	1.08	0	21
Careers	0.43	0	8
Community	1.21	0	12
Corporate presence	2.35	0	32
Directories/Resources	4.09	0	40
Education	1.82	0	21
Entertainment	9.36	0	121
Government	0.34	0	6
Health	0.73	0	8
Hobbies	0.91	0	11
Lifestyle	2.99	0	39
News/Information	1.73	0	29
Portals	0.52	0	8
Real Estate	0.13	0	3
Regional/Local	1.08	0	15
Retail	2.86	0	46
Search/Navigation	0.54	0	10
Sports	2.03	0	22
Technology	0.93	0	9
Travel	0.55	0	11
Services	6.49	0	79
Adult content	14.53	0	191
Business to business	0.22	0	7
All other	6.83	0	95

Note. To be included in this table, a site must have received at least 5000 hits when scaled up to match the Internet-connected population.

and, indeed, may be a city complement. As in local media, the extent of online content increases in the size of the market.[18]

4. The demand for Internet connection

This section examines how the tendency to connect to the Internet varies with measures of the extent of local online and offline options.[19] Our basic measure of the extent of

[18] To the extent that local online content is recycled from local offline uses, this relationship arises directly if larger markets have more or better offerings.

[19] One might in principle study demand for Internet connection as a function of price or availability of ISPs. Greenstein [10] indicates that by 1998 Internet access is widely available in all MSAs. The price of Internet access also varies little across MSAs. A regression of the 1998 CPS measure of monthly ISP costs (hesiu9) on 1990 MSA population gives a constant of \$17.46 (s.e. = \$0.21) and a population coefficient indicating that the price paid for access increases by \$0.043 (\$0.018) per million of additional population.

Table 5
Local sites and market size

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Total local sites, Aug. 2000	Directories/ Resources	Education	Entertain-ment	Lifestyle	News/ Information	Regional/ Local	Retail	Services	Adult content	Miscellaneous
Pop. '90 (mil.)	48.3976	2.6879	1.1084	6.5455	2.0243	1.2532	0.7306	2.1823	5.0263	11.6647	5.4720
	(0.6655)**	(0.0867)**	(0.0559)**	(0.1414)**	(0.0878)**	(0.0603)**	(0.0459)**	(0.0864)**	(0.1045)**	(0.2319)**	(0.1069)**
Constant	−2.6580	0.3678	0.2888	0.3022	0.1890	−0.0090	0.0683	−0.1625	−0.4710	−1.6159	−0.7427
	(1.8502)	(0.2411)	(0.1554)	(0.3931)	(0.2440)	(0.1676)	(0.1276)	(0.2401)	(0.2905)	(0.6446)*	(0.2971)*
Observations	113	113	113	113	113	113	113	113	113	113	113
R-squared	0.98	0.90	0.78	0.95	0.83	0.80	0.70	0.85	0.95	0.96	0.96

Note. Standard errors in parentheses.
* Significant at 5%.
** Significant at 1%.

offline options is total local population, which is presumed to increase the variety of goods and services available. In our second approach, a resident's relevant product variety is determined by the size of the market of people who share her preferences. In our estimation, we will proxy for the size of that market with the population—and population share—of one's group, where the groups are blacks and nonblacks.

Our measure of local online product variety is the number of local sites. Although Table 5 shows that the Internet provides more local content in bigger places, it does not say whether the Internet actually enhances city life. For that to be true, people must want local content. There is ample evidence in traditional local media that the greater variety brought forth in larger markets attracts a higher fraction of the population to consumption. The radio listening, television viewing, and newspaper reading shares are higher in larger markets. The greater quality and variety of options in traditional media provide a reason why persons' welfare, in their capacity as media consumers, may be higher in larger markets. What about Internet use? Does the greater variety of online options targeted at big-city consumers attract a higher fraction of them to the Internet? If so, then the Internet functions as a city complement.

4.1. Internet connection and the extent of local offline and online options

We examine how the tendency to connect is affected by local offline variety by using the CPS data to estimate probits of an individual having a home Internet connection as a function of MSA population or its log:

$$C_i = \alpha + \beta \cdot POP_m + \delta \cdot X_i + \varepsilon_{im}. \tag{1}$$

The left-hand side variable, C, takes the value of one if household i has an Internet connection at home. Our basic measure of local offline product availability, MSA population, is denoted by *POP* and varies only at the MSA level, m. In some specifications, we add a large set of household level demographic controls, X_i, including race of head, gender, household head education and age dummies, and the number of children in the household. In those specifications, the estimated coefficient on *POP* will measure the effect of offline variety in the local market on households' decisions to connect to the Internet even after accounting for their own characteristics.

The first and fifth columns of Table 6, which include only population or its log as a measure of market size, show that overall the probability of connecting to the Internet does not vary with market size. The point estimate in the level specification (5) is slightly negative but indistinguishable from zero, and the log estimate is positive and insignificant.[20] By itself the absence of a relationship between market size and connection indicates that substitution and complementarity effects of the Internet are either nonexistent or offsetting. Supplementing these specifications with individual-level controls for education, age, and presence of children in columns (3) and (7) has little substantive effect. The market size coefficients remain insignificant.

We distinguish the substitute and complement effects by adding a measure of the number of local online options. Columns (2) and (6) add the total number of local sites

[20] The standard errors in all of these and subsequent regressions are adjusted for clustering by MSA.

Table 6
Does the household connection tendency depend on market size and local sites?

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log pop.	0.0089	−0.1076	0.0061	−0.0516				
	(0.0120)	(0.0353)**	(0.0094)	(0.0350)				
Log local sites		0.1025		0.0508				
		(0.0297)**		(0.0305)				
Pop. (mil.)					−0.0008	−0.0686	0.0007	−0.0147
					(0.0016)	(0.0350)	(0.0013)	(0.0350)
Local sites						0.0015		0.0003
						(0.0008)		(0.0008)
Constant	−0.2890	0.9463	−1.4707	−1.2091	−0.1534	−0.1719	−1.3836	−1.7423
	(0.1744)	(0.3924)*	(0.3305)**	(0.4514)**	(0.0212)**	(0.0209)**	(0.2944)**	(0.2653)**
Controls	No	No	Yes	Yes	No	No	Yes	Yes
Observations	29,027	29,027	29,027	29,027	29,027	29,027	29,027	29,027

Notes. Probit estimates with household connection to the Internet (hesiu3) as the dependent variable. Robust standard errors in parentheses (clustered on MSA). Controls include education of household head, presence of children, and household head age dummies. Households are the unit of observation. Data on local sites are calculated by the authors from Media Metrix data, and the remaining data are drawn from the August 2000 CPS Computer and Internet Use supplement.

* Significant at 5%.

** Significant at 1%.

($LOCAL_m$) (or its log) in the MSA to the specifications in columns (1) and (5), resulting in the following equation:

$$C_i = \alpha + \beta \cdot POP_m + \gamma \cdot LOCAL_m + \delta \cdot X_i + \varepsilon_{im}. \tag{2}$$

In this regression, β measures the sensitivity of connection to offline options (as measured by market size), and γ reflects the sensitivity of connection to local online content. If the greater local online content documented for larger markets is valuable to people, it should attract them to connect. In the log specification without controls (column (2)) the coefficient on local population is negative and significant, and the coefficient on local sites is positive and significant. Results in levels are similar, although less significant. Finally, columns (4) and (8) include controls and show the same pattern (negative coefficient on market size, positive coefficient on local sites) but are not statistically significant in conventional two-sided tests. While statistical significance levels are not high, the results suggest that the greater local online content in larger markets attracts households to connect to the Internet but that households in larger markets are otherwise less likely to connect.

How large are the estimated relationships? All else equal, a move from the 25th percentile market in local sites (with log local sites = 2.1) to the 75th percentile (log local sites = 4.1) increases the probability of connectedness by 0.2 according to the estimates without controls (column (1)) and by roughly half that according to the estimates with controls (column (4)). In contrast, conditional on the number of local sites, moving from the 25th percentile market in population (log population = 12.9) to the 75th percentile (log population = 14.1) decreases the rate of connectedness by 0.12. Of course, all else is not equal: log local sites and log population have a correlation of 0.88, so the overall relationship between market size and connectedness (in columns (1) and (3)) is flat.

The patterns of results in Table 6 are noteworthy. First, the absence of a relationship between connection tendency and market size stands in clear contrast to the city-complementing relationships documented for traditional local media (radio, television, and daily newspapers). Second, when both market size and local sites are included in the estimation, there is some evidence that the lack of overall effect is due to distinct and offsetting substitute and complement effects of the Internet on agglomeration. The results at least suggest that local content, which is more plentiful in larger markets, attracts people to the Internet. Holding constant the amount of local online content, people are less likely to connect as their local offline options, proxied by population, are more appealing. The Internet functions as both as a substitute and a complement for cities.

4.2. Internet connection and racial isolation

While the Internet does not function, on balance, as a substitute for cities generally, it may still allow locally isolated households to surmount the limitations of their local offline markets. To put this another way, the Internet may be a substitute for the agglomeration of like-minded persons. To investigate this we ask whether racially isolated individuals are more likely to connect to the Internet. We implement this in Table 7 by asking whether blacks (nonblacks) are less (more) likely to connect as blacks face less appealing local offline product options.

Depending on the nature of local products, their appeal to blacks might reflect either the proportion or absolute number of blacks in the local area. For products with large fixed costs relative to market size, a local market supplies few options, and positioning relative to black preferences depends on the fraction black in the local market.[21] For products with smaller fixed costs relative to market size, a local market can supply multiple options, and the appeal of the local offline product options depends on the absolute number of blacks.[22] Because the local alternatives to content offered over the Internet may take either form, we perform tests of whether blacks use the Internet to overcome racial isolation using both absolute levels of black population and percentages.

First, we perform the test in levels, running probits of the connection dummy on the MSA populations of blacks and nonblacks. We also run the test in percentages, substituting the black share for the market populations. Columns (1) and (2), and (5) and (6), of Table 7 report separate black and nonblack regressions of the tendency to connect on either the absolute numbers of blacks and nonblacks or the fraction black in the MSA. All of the specifications in Table 7 include the full set of individual-level controls. As columns (1) and (2) show, while the nonblack connection tendency does not vary significantly with black or nonblack population, the black connection tendency decreases in black population and increases in nonblack population. The relationships of interest in column (2) are statistically significant or nearly so. Columns (5) and (6) perform the tests in percentages.

[21] George and Waldfogel [8] document this as the mechanism for local daily newspapers. The appeal of the few daily papers in a market depends on the fraction black in the market. Similarly, Waldfogel [19] documents that the absolute amount of black-targeted local television programming varies with the fraction black in the market.

[22] Waldfogel [18] documents that the number of black-targeted radio stations, as well as the black tendency to listen to the radio, vary with the size of the local black population.

Table 7
Does racial isolation explain connection?

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Nonblacks	Black			Nonblacks	Black		
Black dummy			−0.6589	−0.1987			−0.5369	−0.1649
			(0.0316)**	(0.0084)**			(0.0655)**	(0.0167)**
Black pop. (mil.)	0.0027	−0.0602						
	(0.0317)	(0.0362)						
Nonblack pop. (mil.)	0.0013	0.0221						
	(0.0069)	(0.0070)**						
Black dummy × Black pop. (mil.)			−0.0972	−0.0294				
			(0.0232)**	(0.0069)**				
Black dummy × Nonblack pop. (mil.)			0.0276	0.0080				
			(0.0042)**	(0.0012)**				
MSA % black					−0.0474	−0.5390		
					(0.1674)	(0.2893)		
Black dummy × MSA % black							−0.4697	−0.1363
							(0.3068)	(0.0748)
Observations	25,135	3892	29,027	29,027	25,135	3892	29,027	29,027
	Probit	Probit	Probit	LPM	Probit	Probit	Probit	LPM
			MSA FE	MSA FE			MSA FE	MSA FE

Notes. Probit and LPM estimates with household connection to the Internet (hesiu3) as the dependent variable. Robust standard errors in parentheses (clustered on MSA). All specifications include the following controls: education of household head, presence of children, and household head age dummies. Households are the unit of observation. Data on local population are drawn from the 1990 Census, and the remaining data are drawn from the August 2000 CPS Computer and Internet Use supplement.

** Significant at 1%.

The nonblack tendency to connect is invariant with the percent black in the MSA, while the black connection tendency is (almost significantly) smaller in MSAs with larger black shares.

It is possible, however, that local market-level unobserved factors, such as local offline traffic congestion or the quality of local offline media products, affect the tendency for persons in that market to connect to the Internet. We can accommodate this by including a local market fixed effect. When we do this, we cannot identify effects of market-level factors, such as group population, or the percent black in the market. We can, however, identify the *difference* between the effect of the MSA's populations of blacks and nonblacks (or the MSA black share) on black and white tendencies to connect via a coefficient on a black dummy interacted with, say, the MSA percent black.

Columns (3) and (4) of Table 6 report estimates of the MSA fixed effects specifications, both probits and linear probability models, in absolute population levels; and the results are significant in the anticipated direction. Markets with more blacks have a lower tendency for blacks to connect to the Internet, relative to the nonblack tendency to connect; and markets with more nonblacks have a higher tendency for blacks to connect. Columns (7) and (8) report analogous MSA FE specifications in percentages. They show a negative, although not significant, coefficient on the interaction of the black dummy with the black population share (the coefficients are over 1.5 times their standard errors). These results provide evidence that persons are more likely to connect, the more they are isolated locally.

How large are the estimated relationships? For ease of interpretation, consider the linear probability model in column (8). In a market with the 90th percentile black share (0.273), blacks would be 20.2 percentage points less likely than whites to be connected ($0.1649 + 0.2733 * 0.1363$), while in a market with the 10th percentile black share (0.0195), blacks would be 16.8 percentage points less likely to be connected.

4.3. The ironic digital divide?

Evidence above indicates that blacks are more likely to connect as they are more isolated. Blacks make up about 15 percent of the population in sample MSAs and so are in some sense isolated in most places (the median black share across sample MSAs is 8.9 percent, and the 25th and 75th percentile shares are 4.8 and 17.4, respectively). If people connect to avoid isolation and blacks are isolated, blacks should be more likely than whites to connect. Yet, the large and significant negative coefficients on the black dummy in columns (3), (4), (7), and (8) confirm the existence of the digital divide.[23] We can offer two explanations for this puzzle. First, ample evidence indicates that blacks have sharply different content preferences in other media. Even if there were no digital divide, the black Internet audience would be much smaller than the white one. Consequently there would be less absolute content of interest to blacks, giving rise to a smaller black connection tendency.

Second, we have documented that local content attracts people to connect. Because our MM sample contains too few blacks per locale, our data do not allow us to quantify local

[23] See Fairlie [4] for comprehensive documentation that blacks are less likely than whites to be connected to the Internet, after accounting for a host of observable variables.

group-targeted content. Still, it is safe to assume that there is little black-targeted local content on the Internet that would attract connection.

5. Retail spending and isolation from retail stores

We now turn to our final test for substitution, based on retail spending. The WVTM data allow us to ask whether individuals overcome isolation, in this case from nearby retail establishments, by making purchases on the Internet. For each individual in the WVTM sample, we have data on purchases of three types of merchandise (books, music, and clothing) via two channels, offline stores and online (which we define as Internet and catalog). Thus, for each individual i we observe 6 spending variables. In addition, we know each individual's zip code location, and from the Economic Census we know the size distribution of retail stores in each merchandise category by zip code. As a result we can calculate the number of stores in given radii from each individual in the sample.

To this point we have examined overall Internet connection, which is driven by all possible online activities, including both information acquisition as well as retail spending and other tasks. Here we examine online and offline retail spending directly. Geography plays very different roles in offline retail spending and information acquisition. Regardless of one's location within a metro area, media products are delivered to one's door, either via the airwaves, cable, or newspaper carrier. Offline retail spending, by contrast, requires a trip to the store. As a result, consumers' distance to stores may affect their relative preferences for online and offline retail activity.

Table 8 examines online[24] and offline spending on books, music, and clothing, by proximity to offline stores in the category. For example, persons living within a mile of a bookstore spend an average of \$99 per year at bookstores, about 46 percent of it online. As distance from bookstores increases, total spending on books declines, presumably reflecting lower tastes for books by people who live in isolated places. Individuals over 20 miles from a bookstore spend an average of \$68, but the fraction spent online is nearly 53 percent. Finally, total spending on music also declines with distance to stores. Except for those most distant from music stores, the fraction spent online increases with distance from music stores.

Clothing accounts for more spending, and a higher fraction—about two thirds—is offline. The online fraction increases from 35 to 38 percent as distance from clothing stores increases, although the increase is not monotonic. Individuals whose nearest store is 1–2 miles away spend the highest fraction (nearly 41 percent) online.

While Table 8 provides suggestive evidence that people use the Internet to overcome isolation from retail establishments, there are stronger tests available with these data, to which we now turn. Table 9 documents this via regressions of category spending on a measure of distance to the nearest category store, with and without controls (dummy

[24] "Online" here means online and catalog.

Table 8
Online and offline spending by merchandise category and distance to nearest stores

Distance	Books			Music			Clothing		
	Total ($)	Internet/ catalog ($)	Store ($)	Total ($)	Internet/ catalog ($)	Store ($)	Total ($)	Internet/ catalog ($)	Store ($)
1 mile	98.79	45.16	53.63	93.19	46.63	46.55	266.06	93.40	172.66
2 miles	92.33	45.48	46.85	84.92	43.49	41.42	293.70	119.19	174.51
5 miles	91.80	46.04	45.76	85.90	45.39	40.51	260.35	92.27	168.08
10 miles	77.29	37.56	39.73	78.11	40.94	37.17	238.80	85.33	153.48
20 miles	70.20	37.26	32.94	74.19	41.67	32.53	238.07	89.81	148.26
Over 20	68.21	35.83	32.38	75.49	38.85	36.64	239.13	90.92	148.21

Distance	Books		Music		Clothing	
	Internet/ catalog (%)	Store (%)	Internet/ catalog (%)	Store (%)	Internet/ catalog (%)	Store (%)
1 mile	45.71	54.29	50.04	49.96	35.11	64.89
2 miles	49.25	50.75	51.22	48.78	40.58	59.42
5 miles	50.15	49.85	52.84	47.16	35.44	64.56
10 miles	48.59	51.41	52.42	47.58	35.73	64.27
20 miles	53.07	46.93	56.16	43.84	37.73	62.27
Over 20	52.52	47.48	51.46	48.54	38.02	61.98

Table 9
Regressions of spending on distance to nearest store

Coefficient on miles to nearest category store	(1)	(2)	(3)	(4)	(5)	(6)
	Books		Music		Clothes	
	Internet or catalog	Offline	Internet or catalog	Offline	Internet or catalog	Offline
No controls	−0.5225	−0.8557	−0.4276	−0.4499	−0.5740	−1.2305
	(0.1024)**	(0.0868)**	(0.0978)**	(0.0719)**	(0.2793)	(0.2741)**
With controls	−0.1966	−0.5307	−0.3212	−0.3559	−0.0562	−0.3976
	(0.1107)	(0.0928)**	(0.1064)**	(0.0773)**	(0.3005)	(0.2845)

Notes. Both specifications include an indicator variable for the year of the survey. Controls in the second row include 17 household income dummies and 8 education dummies. The sample includes all respondents who provided valid answers to all the questions about spending via the Internet, catalog, or stores within a merchandise category. In the first row, $N = 21{,}968$ in columns (1) and (2); $N = 21{,}967$ in columns (3) and (4); and $N = 21{,}966$ in columns (5) and (6). The sample in row two is further restricted to those households who answered the questions about income and education and thus $N = 19{,}033$ in all columns.

** Significant at 1%.

variables for household income and education level).[25] Each coefficient in the table is from a separate regression. The first two columns, for example, show how online and offline

[25] For these regressions we code distance as the top of the cell. For example, if one has a store in the 0–1 mile radius, we code distance as 1 mile. We code distance somewhat arbitrarily as 30 miles for persons without stores within 20 miles. We experimented with other values (40, 50, 100 miles) and found similar results.

book spending vary across persons by their distance from an offline store. Both online and offline spending decline with distance, although offline spending declines more sharply, particularly for books and clothing.

In the specifications without controls, both online and offline spending are smaller for individuals living farther from stores. For example, the 25th percentile household lives no more than one mile from the nearest store in a merchandise category while the 75th percentile household lives five miles away. That difference corresponds to about a \$2.10 reduction in online spending on books and about a \$3.40 reduction in spending on books in local stores, from a base average spending on books in our data of approximately \$43. The pattern in music and clothing looks similar, with declines in online and offline spending in music being about \$1.70 and \$1.80, respectively, and \$2.30 and \$5.00 in clothing. The average spending on clothing in our sample is higher than for the other merchandise categories, at \$93, whereas music averages just \$44.

When controls are included, online spending does not decline significantly with distance for books or clothing, although it does for music. Importantly, though, the offline decline with distance is greater (for books and clothing) than the online spending reduction. For example, the same interquartile jump in distance as before, from one to five miles from the nearest store, would lead to a \$0.80 reduction in spending on books online but a \$2.10 fall in spending on books in stores.

Since spending through every channel declines with distance, we can infer that individuals' who live further from stores have endogenously lower 'tastes for consumption.' Thus the fact that offline spending declines in distance (in even-numbered columns of Table 9) does not provide clean evidence that sparse local options induce people to purchase online. Rather, it is equally plausible that people who like consumption relatively less, choose to live further from stores.

These varying tastes for consumption across individuals can be addressed by examining the difference between the online and offline spending gradients, within categories. Even if people who live further from stores are those who prefer to spend less, the relative paucity of local stores may make them more likely to shop online when they do shop. To this end we estimate a model with individual fixed effects:

$$S_{io} = \alpha_0 + \alpha_1 D_i + \alpha_2 \delta_I + \alpha_3 \delta_I D_i + \mu_i + \varepsilon_{io}, \tag{3}$$

where S_{io} is individual's spending in a merchandise category (say, books) though outlet o (online or off), D_i is the individual's distance to the nearest store in that merchandise category (e.g. bookstore), δ_I is a dummy for online, and μ is an individual fixed effect (i.e., the taste for books, independent of channel).

When this model is estimated with individual fixed effects, only α_2 and α_3 are estimable since the distance to the nearest store in a merchandise category is constant for a given person. The parameter α_2 shows how much more people spend on books online than off. The parameter of interest α_3 shows how spending-distance gradient varies online as opposed to offline. If α_3 is positive, it indicates that, relative to offline spending, online spending rises with distance to the nearest store.[26]

[26] The individual-level controls are subsumed by the individual fixed effect.

Table 10
Spending and distance to nearest store, FE specifications

	(1)	(2)	(3)
	Books	Music	Clothes
Internet or catalog dummy	−4.2854	4.5451	−74.0254
	(1.0080)**	(0.9770)**	(1.9661)**
Distance × Internet dummy	0.3398	0.0241	0.6757
	(0.1027)**	(0.0865)	(0.3172)*
Fixed effects?	Person	Person	Person
Observations	43,936	43,934	43,932

Notes. Standard errors in parentheses. The sample includes all respondents who provided valid answers to all the questions about spending via the Internet, catalog, or stores within a merchandise category.
* Significant at 5%.
** Significant at 1%.

The three columns of Table 10 run specification (1) on each of the merchandise categories, and the estimates of α_3 are positive and significant for books and clothing (zero for music).[27] These estimates indicate that after accounting for individuals' tendency to spend in a category, online spending increases in distance relative to their offline spending (in books and clothing). In particular, for books, moving from one to five miles away from the nearest store would lead to the gap in online spending and offline spending on books growing by about $1.40. For clothing, the difference in online and offline spending would rise by nearly $2.60, nearly 3 percent of the average spending amount of $93. More succinctly, these results indicate that for clothing and books at least, isolation from retail establishments induces Internet-connected persons to spend relatively more money over the Internet.[28]

6. Conclusion

It is apparent from our results that, in spite of more and better local online options in larger markets, the tendency to connect to the Internet is not affected by market size. This result stands in sharp contrast to relationships in traditional media, which reinforce the welfare advantages of larger markets for consumption. In the case of the Internet, local

[27] The point estimates for the "Internet or catalog dummy" show that people who live right next to a store on average spend $4.30 less on books online than offline, $4.50 more on music online than off, and $74 less on clothing online than off.

[28] Another way of controlling for individual heterogeneity in the propensity to consume would be to regress the share of spending in a merchandise category that was online on the distance to the nearest store in that category. We have tried this approach, with similar results to the levels regressions we report in Table 10. As the distance to the nearest store increases, the share of spending that is online goes up as well. The effect is statistically significant in all merchandise categories, and is virtually unaffected by the addition of income and education dummies (although the standard errors rise when the controls are added so the clothing coefficient is no longer significantly different from zero). The magnitudes of the economic effects are smaller in the shares regression, typically about one-fourth (at the mean) of the ones reported in Table 10.

content does encourage increased connection in larger markets, as with traditional media. However, unlike traditional local media, the Internet also provides access to a national level of variety for small places, mitigating the advantage of larger markets over smaller ones. This kind of effect is especially clear in blacks' tendency to use the Internet relative to whites to overcome preference isolation. However, despite black isolation and the tendency toward Internet use to overcome isolation, blacks remain relatively unlikely to use the Internet, possibly for lack of appealing Internet content.

The Internet holds out the promise of erasing differences in consumption opportunities arising from isolation, either from living in small markets or near small groups of people with similar preferences. There is some evidence that the Internet accomplishes both of these functions. All else equal, the Internet is more attractive in small markets than in large ones. And online retail spending on books and clothing is higher, relative to offline spending in those categories, as people live farther from offline stores. Thus, there is evidence from a variety of results in the paper that the Internet functions for consumers as a substitute for agglomeration. Yet, all else is not equal. There is more local Internet content in larger markets, so the Internet is more attractive in larger places than in smaller ones. The complementarity of local sites with local agglomeration offsets the Internet's substitute function.

The Internet is a young technology whose diffusion is not complete. It is perhaps surprising that we find any relationship between connection tendencies and various kinds of isolation. It is probably too early to sensibly determine whether the Internet has fulfilled its promise, but it seems fair to observe, at this point, that the Internet has not yet rendered geography irrelevant. To paraphrase Mark Twain, reports of the death of cities at the hands of the Internet may be greatly exaggerated.

Acknowledgments

We are grateful to the Wharton e-Business Initiative (WeBI) for giving us access to, and technical assistance with, the Media Metrix data, and to Stephen Hoch for the Wharton Virtual Test Market data. We received useful comments from the editor and an anonymous referee, as well as seminar participants at Wharton, the University of Illinois, Duke, the Federal Reserve Banks of New York and Philadelphia, and the 2002 AEA Meetings in Atlanta. Sinai received financial support from the Ballard Scholars Program at the Zell-Lurie Real Estate Center, and Waldfogel thanks WeBI. Sam Chandan and Dan Simundza provided excellent research assistance.

References

[1] T.F. Bresnahan, P.C. Reiss, Entry and competition in concentrated markets, Journal of Political Economy 99 (1991) 977–1009.
[2] A. Ciccone, R.E. Hall, Productivity and the density of economic activity, American Economic Review 86 (1996) 54–70.
[3] A.K. Dixit, J.E. Stiglitz, Monopolistic competition and optimum product diversity, American Economic Review 67 (1977) 297–308.

[4] R. Fairlie, Race and the digital divide, Mimeo, University of California at Santa Cruz, 2002, available at http://econ.ucsc.edu/~fairlie/papers/digitaldivide7.pdf.
[5] C. Forman, A. Goldfarb, S. Greenstein, How did location affect adoption of the commercial Internet? Global village, urban density, and industrial composition, NBER working paper 9979, September 2003.
[6] M. Friedman, Capitalism and Freedom, Univ. of Chicago Press, Chicago, 1962.
[7] J. Gaspar, E.L. Glaeser, Information technology and the future of cities, Journal of Urban Economics 43 (1998) 136–156.
[8] L. George, J. Waldfogel, Who affects whom in daily newspaper markets? Journal of Political Economy 111 (2003) 765–784.
[9] E.L. Glaeser, J. Kolko, A. Saiz, Consumer city, Journal of Economic Geography 1 (2001) 27–50.
[10] S. Greenstein, Building and delivering the virtual world: Commercializing services for Internet access, Mimeo, Kellogg School, 2000.
[11] J.V. Henderson, Optimum city size: The external diseconomy question, Journal of Political Economy 82 (1974) 373–388.
[12] J. Kolko, The death of cities? The death of distance? Evidence from the geography of commercial Internet usage, Mimeo, Harvard University, 1999.
[13] S.S. Rosenthal, W. Strange, Evidence on the nature and sources of agglomeration economies, in: Handbook of Urban and Regional Economics, vol. 4, Elsevier, Amsterdam, in press.
[14] A.M. Spence, Product selection, fixed costs, and monopolistic competition, The Review of Economic Studies 43 (1976) 217–235.
[15] A.M. Spence, Michael, Product differentiation and welfare, American Economic Review 66 (1976) 407–414.
[16] J. Sutton, Sunk Costs and Market Structure: Price Competition, Advertising, and the Evolution of Concentration, MIT Press, Cambridge, MA, 1991.
[17] US Department of Commerce, National Telecommunications and Information Administration, Falling through the net: Defining the digital divide, A report on the Telecommunications and Information Technology Gap in America, July 1999.
[18] J. Waldfogel, Preference externalities: An empirical study of who benefits whom in differentiated product markets, RAND Journal of Economics 34 (2003) 557–568.
[19] J. Waldfogel, Who benefits whom in local Television markets? Brookings–Wharton Papers on Urban Affairs, in press.