

Geography and the Internet:

Is the Internet a Substitute or a Complement for Cities?

Todd Sinai
Joel Waldfogel
The Wharton School
University of Pennsylvania
and NBER

December 31, 2001

ABSTRACT

By combining persons around the world into a single market, the Internet may serve as a *substitute* for urban agglomeration. That is, the Internet may level the consumption playing field between large, variety-laden and small, variety-starved markets. However, if local content on the Internet is more prevalent in larger markets, then the Internet may be a *complement* for urban agglomeration. Characterizing the nature of available content using Media Metrix web page visits by about 32,000 households during August 2000, we document that substantially more online local content is available in larger markets. Combining this with CPS August 2000 Internet use data, we find 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. We also find that individuals connect to overcome local isolation: notwithstanding a large digital divide, blacks are more likely to connect, relative to whites, when they comprise a smaller fraction of local population, making the Internet is a substitute for agglomeration of preference minorities within cities, if not cities themselves. On balance we find that the substitution and complementarity effects offset each other so that the Internet does not promote or discourage agglomeration in larger markets.

We are grateful to the Electronic Commerce Forum at Wharton for giving us access to, and technical assistance with, the Media Metrix data. Sinai received financial support from the Ballard Scholars Program at the Zell-Lurie Real Estate Center, and Waldfogel thanks the Electronic Commerce Forum for financial assistance. Sam Chandan provided excellent research assistance.

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 serve as a substitute for urban agglomeration by leveling the consumption playing field between large, variety-laden and small, variety-starved markets. But this is not necessary. 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 ask whether the Internet serves as a substitute or a complement for urban agglomeration. Existing research indicates that larger markets provide greater, and smaller markets less, offline product variety, particularly in information products. As a result, a consumers' welfare has traditionally been limited by the size of her local market. The Internet allows persons unsatisfied by their local options access to products and information geared to a global market. If we observe that individuals are more likely to connect as they live in smaller markets, then we can infer that the Internet functions as a substitute for cities.

¹ We are aware of other arguments for agglomeration besides consumption. Henderson (1974) presents a model in which city size balances production benefits against congestion costs. See Ciccone and Hall (1996) for recent empirical evidence on the production benefits of agglomeration as well as additional references.

² We are not the first to pose this question. See Kolko (1999), as well as Gaspar and Glaeser (1998), 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.

The problem of dissatisfaction with offline options can be particularly acute for persons with preferences unlike their neighbors'. It is clear from existing research that blacks and whites have sharply different preference 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 Internet, then we can infer not only that the Internet is a substitute for cities but also that it is a substitute for “product ghettos,” or groups of people within a market sharing in common their locally atypical product preferences.

By connecting people across geographic markets, the Internet holds the promise of liberating isolated individuals from their neighbors' tastes. Yet, the existence of a digital divide, a lower tendency for blacks to connect to the Internet, is well known; and our results reflect this. Notwithstanding an overall digital divide, we are looking for behavioral evidence that the divide narrows in local markets where blacks are more isolated.

In this paper we ask a number of questions that shed light on these concerns. First, how much Internet content is local? Second, is there more locally targeted content in larger markets? If so, then Internet content may complement urban agglomeration. This leads to our third question: does local content attract people to connect? The push of local offline product paucity – and the pull of local online variety – have opposing effects on whether the Internet is a substitute or a complement for cities. We separately document these competing effects and determine, on

³ For example, larger markets have more and better local newspapers (George and Waldfogel, 2000), radio broadcasts (Waldfogel, 1999), and television stations (Waldfogel, 2001). These studies document that black and white consumers' welfare, in their capacity as media consumers, increase in the size of their own respective group populations.

balance, which is stronger. Finally, we ask whether persons experiencing local preference isolation (isolated blacks) are more likely than less isolated blacks to connect.

We first examine evidence about possible complementarity of the Internet with urban agglomeration. Characterizing 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. Combining the Current Population Survey's August 2000 Computer and Internet Use module with the Media Metrix-derived data on local online content, 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 does not promote or discourage agglomeration in larger markets. The relationship between Internet connection and market size stands in stark contrast with evidence on local media markets, where the tendency to consume increases in the size of the market. We find no evidence that persons in larger markets are more likely to connect.

We further explore whether the Internet allows individuals to overcome local isolation by asking how the tendency to connect varies across geography with the extent of individuals' local offline options, measured by the fraction of local persons of their race. We find that blacks are more likely to connect, relative to whites, as blacks have a smaller local population, suggesting that the Internet is a substitute for agglomeration of preference minorities within cities.

Friedman (1962) 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 CPS and Media Metrix 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 Current Population Survey (CPS) data to characterize the demand for Internet connection. In particular, we examine how individuals' Internet connection tendencies vary with the extent of one's local and online options. A brief conclusion follows.

I. How Does the Internet Function as a Substitute or Complement for Cities?

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 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, 1999).

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 consumer welfare on the number and mix of consumers in her 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? We 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 is 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 (George and Waldfogel, 2000), people who live in them may place a higher value on the availability of news on the Internet.

⁴ This is what Spence (1976a,b) terms the “product selection problem.” See also Dixit and Stiglitz (1977). Waldfogel (1999) presents an empirical characterization of how product availability varies with the size and demographic composition of local markets.

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 (1991) for extensive discussion of endogenous fixed costs.

2. The Internet as a Substitute for Product Ghettos

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.⁶ In these contexts, the satisfaction of blacks as consumers of local products has typically depended on the number or share of blacks in the local population, not the overall size of the local market.⁷

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

⁶ Based on views about public spending blacks and whites have substantially different political preferences as well (Oberholzer Gee and Waldfogel, 2000).

⁷ With some products the absolute number of local persons determines the number of available options. This is true for radio broadcasting. See Berry and Waldfogel (1999ab) and Waldfogel (1999). For other products, such as daily newspapers, fixed costs rise endogenously with market size, keeping the number of products small. George and Waldfogel (2000) present evidence that fraction black in the market determines the how black-targeted the paper(s) is (are). Other statistics in this paragraph are also drawn from these sources.

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 people. 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.

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 vary with one's local offline and online options? Which effect dominates?

II. Data

Data for this study are drawn from two sources, the August 2000 Current Population Survey Computer and Internet Use Supplement and an August 2000 Media Metrix data extract. 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.

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 31 percent are at least college-educated.

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.⁸ Only about 7 percent of the connected sample is black-headed, and approximately 45 percent 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 by a household is a separate record – with over 16.5 million page visits in total during the month.⁹ 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. 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. 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.

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.¹⁰ The first column of table 2 reports the

⁸ These results are consistent with evidence elsewhere on the digital divide. See <http://www.ntia.doc.gov/ntiahome/ftn99/contents.html> .

⁹ 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.”

¹⁰ The Media Metrix categories are: Auctions, Automotive, 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, Adult Content, Business to Business, All Other, and Religion.

distribution across these categories of page hits and numbers of visited sites.¹¹ The category with the most hits is portals, with 21.9 percent of the total hits, followed by services (15.9), entertainment (12.5), 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 sites visited in each category, reported in column 2, “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 suggests that there may be more concentration in the portals, with a few sites each receiving a large amount of traffic.

III. 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.¹² 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

¹¹ In order to be included in the total hits or sites calculations, a site must be estimated to have received at least 5,000 hits during the month.

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.

To make this more concrete, we compute each site’s index of “site localness” as

$$1 / \sum_j \left[\frac{p_{ij} / HH_j}{\sum_j (p_{ij} / HH_j)} \right],$$

where p_{ij} is page visits to site i from MSA j and HH_j is the number of

households in MSA j . This formula is essentially an inverse HHI scaled to reflect the fact that population varies across MSAs. If we did not normalize by each MSA’s total households, sites would appear to be targeted towards larger MSAs even if they truly were equally appealing to any individual in any part of the country. When all visitors are from a single MSA, the index is 1. If a site has equal average hits per household in two MSAs, the index is 2. The larger the index, the less local a site. A perfectly geographically neutral site’s index equals the number of MSAs in the sample. We then attribute a local site, which we define as having a “localness index” of two or less, to the locale that contributes the site’s largest market share.¹³

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. To counteract that problem, we require that sites exhibit a minimum 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 proportionally to market

¹² Kolko (1999) uses the list of registered domain names and shows higher “domain density” in larger markets, which is at least suggestive that web content is complementary with cities. Domain registration indicates the geographic location of the registrant, not the site users, however.

¹³ 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

population, we estimate the actual number of hits 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. A cutoff of 5,000 hits is applied to the resulting estimate. This way, even if one city has fewer households than another in the Media Metrix panel, our threshold maintains the same economic importance across locations. 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.

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).¹⁴

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.

The second panel 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 7,273

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.

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 have the lowest rate of its sites being local.

The average number of local sites per MSA, reported in table 3, follows the same general pattern. There are 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. 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?¹⁵

Table 4 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. 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 and, indeed, may be a city complement. As in local media, the extent of online content increases in the size of the market.

¹⁴ We have suppressed the “www” site name prefix for clarity.

IV. The Demand for Internet Connection

If the Internet is a substitute for cities, the probability that a person connects to the Internet should increase as the variety or quality of local offline options decline. If it is a complement, it should be more prevalent in larger cities and people should be more likely to use it where there are more city-specific options. This section asks how the tendency to connect to the Internet varies with measures of the extent of local online and offline options.¹⁶ Our basic measures of the extent of offline options are total local population, which is presumed to increase the variety of goods and services available. In addition, 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 measure 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 4 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 part of the reason why persons' welfare, in their capacity as media consumers, is higher in larger markets. What about Internet use? Does the greater variety of

¹⁵ 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 (1991), for example.

¹⁶ One might in principle study demand for Internet connection as a function of price or availability of ISPs. Greenstein (1999) 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.

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.

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:

$$C_i = \mathbf{a} + \mathbf{b} \cdot POP_m + \mathbf{d} \cdot X_i + \mathbf{e}_{im} \quad (1)$$

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 5, 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.¹⁷ 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

¹⁷ The standard errors in all of these and subsequent regressions are adjusted for clustering on MSAs.

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

$$C_i = \mathbf{a} + \mathbf{b} \cdot POP_m + \mathbf{g} \cdot LOCAL_m + \mathbf{d} \cdot X_i + \mathbf{e}_{im} \quad (2)$$

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. 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 at least suggest that local content attracts households to connect to the Internet but that households in larger markets are otherwise less likely to connect.

The patterns of results in table 5 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.

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 “product ghettos.” To investigate this we ask whether racially isolated individuals are more likely to connect to the Internet. We implement this in table 6 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 the local product, its 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.¹⁸ 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.¹⁹ 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 percents.

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 percents, substituting the black share

¹⁸ George and Waldfogel (2000) 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 (2001) documents that the absolute amount of black-targeted local television programming varies with the fraction black in the market.

for the market populations. Columns (1) and (2), and (5) and (6), of table 6 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 6 include the full set of 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. 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 as the 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

¹⁹ Waldfogel (1999) 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.

MSA FE specifications in percentages. Although the coefficients are over 1.5 times their standard errors, we find a negative although not significant coefficient on the interaction of the black dummy with the black population share. These results provide evidence that persons are more likely to connect, the more they are isolated locally, or that the Internet is a substitute for product ghettos. The large and significant negative coefficients on the black dummy in columns (3), (4), (7), and (8) confirm the existence of the digital divide.

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. If people connect to avoid isolation and blacks are isolated, why is there a digital divide? We can offer two reasons. 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 group-targeted content. Still, it is safe to assume that there is little black-targeted local content on the Internet that would attract connection.

4. Computer Ownership, Market Size, and Racial Isolation

Connection to the Internet requires two separate but presumably related decisions. First, one must purchase a computer or Web TV.²⁰ Second, given PC ownership, one must purchase Internet connection. If the decision to own a home computer were orthogonal to the Internet connection decision, it would make sense to run the regressions in the previous sections only on the sample of persons with a home computer. Instead, we suspect that these are related decisions (one reason to purchase a computer is to connect to the Internet). In table 7 we explore how the decision to own a home computer varies with our measures of the attractiveness of local offline and online options. The table is entirely parallel to table 5. Columns (1), (3), (5), and (7) show that computer ownership is essentially invariant with market size (using population and its log, with and without controls). Columns (2), (4), (6), and (8) ask how computer ownership varies with local content and market size. The sign pattern generally echoes the Internet connection results. Households are more likely to have computers in smaller markets and markets with more local online content. Significance levels are higher, except in the levels specification with controls. These results suggest that part of the Internet's relationship with city size operates through computer ownership.

Table 8 examines how computer ownership varies with racial isolation. The specifications in the table are parallel to table 6. There are almost no statistically significant results in the table. Columns (3) and (4) show the only partial exception: black computer ownership is significantly more likely, relative to nonblack ownership, in markets with more nonblacks. These results suggest that the Internet's role in allowing persons to overcome preference isolation operates only slightly through the decision to own a computer.

V. Conclusion

²⁰ According to the 1998 CPS, 0.70 percent of the non-computer-owning population has Web TV.

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 content does encourage increased connection in larger markets, as with traditional media. However, unlike traditional 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. In spite of black isolation and the tendency toward Internet use to overcome isolation, blacks remain relatively unlikely to use the Internet, possibly for a lack of appealing Internet content.

The Internet holds out the promise of erasing welfare differences arising from isolation, either from living in small markets or near small groups of preference compatriots. 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. That is, there is evidence 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.

All else equal, isolated blacks are more likely than other blacks to connect to the Internet. That is, Internet functions as a substitute for living among larger groups of persons sharing one's preferences. Yet, even blacks facing extreme isolation are far less likely than whites to connect to Internet.

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.

References

- Berry, Steven T. and Joel Waldfofel. "Free Entry and Social Inefficiency in Radio Broadcasting." *RAND Journal of Economics*. 1999.
- Berry, Steven T. and Joel Waldfofel. "Public Radio in the United States: Does it Correct Market Failure or Cannibalize Commercial Stations?" *Journal of Public Economics*, 1999.
- Bresnahan, Timothy F; Reiss, Peter C. "Entry and Competition in Concentrated Markets." *Journal of Political Economy*. Vol. 99 (5). p 977-1009. October 1991.
- Ciccone, Antonio and Robert E. Hall. "Productivity and the Density of Economic Activity." *American Economic Review*. Vol. 86 (1). p 54-70. March 1996.
- Dixit, Avinash K; Stiglitz, Joseph E. "Monopolistic Competition and Optimum Product Diversity." *American Economic Review*. Vol. 67 (3). p 297-308. June 1977.
- Friedman, Milton. *Capitalism and Freedom*. Chicago: University of Chicago Press, 1962.
- Gaspar, Jess and Edward L. Glaeser. "Information Technology and the Future of Cities." *Journal of Urban Economics*. Vol. 43 (1). p 136-56. January 1998.
- George, Lisa and Joel Waldfofel. "Who Benefits Whom in Daily Newspaper Markets?" Mimeo. University of Pennsylvania, August 2000.
- Greenstein, Shane. "Building and Delivering the Virtual World: Commercializing Services for Internet Access." Mimeo. Kellogg School. 2000.
- Henderson, J. V. "Optimum City Size: The External Diseconomy Question." *Journal of Political Economy*. Vol. 82 (2). p 373-88. Part I, March-April 1974
- Kolko, Jed. "The Death of Cities? The Death of Distance? Evidence from the Geography of Commercial Internet Usage." Mimeo. Harvard University. 1999
- Oberholzer Gee, Felix and Joel Waldfofel. "Tiebout Acceleration." Mimeo. University of Pennsylvania. August 2000.
- Spence, Michael. "Product Selection, Fixed Costs, and Monopolistic Competition." *The Review of Economic Studies*. Vol. 43 (2). p 217-35. June 1976.
- Spence, Michael. "Product Differentiation and Welfare." *American Economic Review*. Vol. 66 (2). p 407-14. May 1976.
- Sterngold, James. "A Racial Divide Widens on Network TV." *New York Times*. December 29, 1998. P. A1.

Sterngold, James. "For All the TV Pilots, There's Just Not Enough Youth to Go Around." *The New York Times*, May 17, 1999. Page E1.

Sutton, John. *Sunk costs and market structure: Price competition, advertising, and the evolution of concentration..* Cambridge, Mass. and London: MIT Press. 1991.

Waldfogel, Joel. "Preference Externalities: An Empirical Study of Who Benefits Whom in Differentiated Product Markets." NBER Working paper 7391, October 1999.

Waldfogel, Joel. "Who Benefits Whom in Local Television Markets?" Mimeo. The Wharton School. November 2001.

Table 1: CPS and Media Metrix Sample Characteristics

	August 2000 CPS		MM
	all heads of households	connected household heads	
Internet at Home	43.7	100.0	100.0
Computer(s) at Home	54.3	98.8	100.0
White	82.2	87.3	88.7
Black	13.4	7.4	4.5
Native American	0.6	0.4	N/A
Asian	3.8	4.9	3.0
Less than High School	14.7	4.1	2.0
High School	28.5	20.1	9.7
Some College	26.7	30.3	27.0
College	19.6	28.7	36.4
Post Graduate	10.5	16.7	24.8
Observations	29027	29027	17104

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.

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
Auctions	715,491	4.3%	71	0.3%	3,106	0.4%	0.4%	24	0.3%	33.8%
Automotive	115,769	0.7%	215	1.0%	1,984	0.2%	1.7%	42	0.6%	19.5%
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%	1,086	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%	1,061	4.7%	11,663	1.3%	2.3%	266	3.7%	25.1%
Directories/Resources	417,353	2.5%	1,258	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%	2,460	11.0%	77,602	8.9%	3.4%	1,061	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%	4,386	0.5%	5.6%	82	1.1%	27.2%
Hobbies	171,222	1.0%	341	1.5%	6,831	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%	2,728	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%	1,313	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%	2,942	0.3%	1.9%	105	1.4%	37.5%
Travel	262,862	1.6%	324	1.4%	2,389	0.3%	0.9%	62	0.9%	19.1%
Services	2,616,051	15.9%	2,489	11.1%	145,265	16.6%	5.6%	733	10.1%	29.4%
Adult Content	1,317,683	8.0%	5,154	23.0%	106,362	12.2%	8.1%	1,642	22.6%	31.9%
Business to Business	24,218	0.1%	83	0.4%	943	0.1%	3.9%	25	0.3%	30.1%
All Other	406,822	2.5%	2,764	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%	7,273	100.0%	32.4%

Note: To be included in this table, a site must have received at least 5,000 hits when scaled up to match the Internet-connected population.

Table 3: 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 5,000 hits when scaled up to match the Internet-connected population.

Table 4: Local Sites and Market Size

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Total Local Sites, Aug. 2000	Directories/ Resources	Education	Entertainment	Lifestyle	News/ Information	Regional/ Local	Retail	Services	Adult Content	Misc
Pop. '90 (mil.)	48.3976 (0.6655)**	2.6879 (0.0867)**	1.1084 (0.0559)**	6.5455 (0.1414)**	2.0243 (0.0878)**	1.2532 (0.0603)**	0.7306 (0.0459)**	2.1823 (0.0864)**	5.0263 (0.1045)**	11.6647 (0.2319)**	5.4720 (0.1069)**
Constant	-2.6580 (1.8502)	0.3678 (0.2411)	0.2888 (0.1554)	0.3022 (0.3931)	0.1890 (0.2440)	-0.0090 (0.1676)	0.0683 (0.1276)	-0.1625 (0.2401)	-0.4710 (0.2905)	-1.6159 (0.6446)*	-0.7427 (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

Notes: Standard errors in parentheses. * significant at 5%; ** significant at 1%

Table 5: Does the Household Connection Tendency Depend on Market Size and Local Sites?

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log Pop	0.0089 (0.0120)	-0.1076 (0.0353)**	0.0061 (0.0094)	-0.0516 (0.0350)				
Log Local Sites		0.1025 (0.0297)**		0.0508 (0.0305)				
Pop (mil.)					-0.0008 (0.0016)	-0.0686 (0.0350)	0.0007 (0.0013)	-0.0147 (0.0350)
Local Sites						0.0015 (0.0008)		0.0003 (0.0008)
Constant	-0.2890 (0.1744)	0.9463 (0.3924)*	-1.4707 (0.3305)**	-1.2091 (0.4514)**	-0.1534 (0.0212)**	-0.1719 (0.0209)**	-1.3836 (0.2944)**	-1.7423 (0.2653)**
Controls	No	No	Yes	Yes	No	No	Yes	Yes
Observations	29027	29027	29027	29027	29027	29027	29027	29027

Notes: Probit estimates with household connection to the Internet (hesiu3) as the dependent variable. Robust standard errors in parentheses (clustered on MSA). * significant at 5%; ** significant at 1%. 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.

Table 6: Does Racial Isolation Explain Connection?

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Non-Blacks	Black			Non-Blacks	Black		
Black Dummy			-0.6589 (0.0316)**	-0.1987 (0.0084)**			-0.5369 (0.0655)**	-0.1649 (0.0167)**
Black Pop (mil)	0.0027 (0.0317)	-0.0602 (0.0362)						
Non-Black Pop (mil)	0.0013 (0.0069)	0.0221 (0.0070)**						
Black Dummy x Black Pop (mil)			-0.0972 (0.0232)**	-0.0294 (0.0069)**				
Black Dummy x Non-Black Pop (mil)			0.0276 (0.0042)**	0.0080 (0.0012)**				
MSA % Black					-0.0474 (0.1674)	-0.5390 (0.2893)		
Black Dummy x MSA % black							-0.4697 (0.3068)	-0.1363 (0.0748)
Observations	25135 Probit	3892 Probit	29027 Probit	29027 LPM	25135 Probit	3892 Probit	29027 Probit	29027 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). * significant at 5%; ** significant at 1%. 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.

Table 7: Does Local Content Induce Households to Have Computers?

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log Pop	-0.0009 (0.0141)	-0.1427 (0.0409)**	-0.0058 (0.0103)	-0.0926 (0.0405)*				
Log Local Sites		0.1245 (0.0329)**		0.0762 (0.0343)*				
Pop (mil.)					-0.0032 (0.0021)	-0.0813 (0.0334)*	-0.0022 (0.0017)	-0.0299 (0.0309)
Local Sites						0.0017 (0.0007)*		0.0012 (0.0007)
Constant	0.1223 (0.2066)	1.6262 (0.4616)**			0.1256 (0.0241)**	0.1045 (0.0238)**		
Controls	No	No	Yes	Yes	No	No	Yes	Yes
Observations	29027	29027	29027	29027	29027	29027	29027	29027

Notes: Probit estimates with household computer ownership as the dependent variable. Robust standard errors in parentheses (clustered on MSA). * significant at 5%; ** significant at 1%. 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.

Table 8: Does Racial Isolation Explain Computer Ownership?

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Non-Blacks	Black			Non-Blacks	Black		
Black Dummy x MSA % Black							-0.0675 (0.2970)	-0.0371 (0.0796)
Black Pop (mil)	-0.0268 (0.0241)	-0.0148 (0.0515)						
Non-Black Pop (mil)	0.0027 (0.0054)	0.0120 (0.0104)						
Black Dummy x Black Pop (mil)			-0.0177 (0.0313)	-0.0080 (0.0105)				
Black Dummy x Non-Black Pop (mil)			0.0162 (0.0058)**	0.0052 (0.0019)**				
Constant	-1.7087 (0.1442)**	0.3084 (0.3043)	-2.0936 (0.2763)**	0.0717 (0.0436)	-1.6808 (0.1473)**	0.3983 (0.2780)	-1.7510 (0.2901)**	0.0709 (0.0435)
Black Dummy			-0.6440 (0.0316)**	-0.2033 (0.0097)**			-0.5691 (0.0649)**	-0.1792 (0.0178)**
MSA % Black					-0.2913 (0.1741)	-0.2914 (0.3368)		
Observations	29027 Probit	29027 Probit	29027 Probit	29027 LPM	29027 Probit	29027 Probit	29027 Probit	29027 LPM

Notes: Probit and LPM estimates with household connection to the Internet (hesiu3) as the dependent variable. Robust standard errors in parentheses (clustered on MSA). * significant at 5%; ** significant at 1%. 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.