Measuring Activity-Based Internet Use Inequality in U.S. Counties and Metropolitan Areas

Innovative Broadband Use Evaluation and Research Data Seed Grant, 2014-2015

William W. Franko
Auburn University
October 30, 2015

This document provides an overview of the estimation procedures used to develop new measures of activity-based Internet use inequality for counties and metropolitan areas in the United States, as well as a codebook that describes the variables included in the associated data files.

Purpose

The development of broadband technology in the United States has substantial consequences for America's economic and social progress. The growth of the Internet provides direct and indirect benefits across multiple policy domains, such as employment, education, health care, public safety, mass transit, energy, and political participation. At the same time, these benefits are not equally accessible to everyone. While high-speed Internet is available to nearly 95% of Americans (FCC 2010), broadband adoption rates are considerably lower in various parts of the country. This well-established digital divide—that is, the inequality in technology access and skill largely based on income, education, race, and ethnicity—has structured opportunity in a way that exacerbates the separation between the haves and have-nots (e.g., see Mossberger, Tolbert, and Stansbury 2003; Norris 2001).

In addition to the disparities in broadband adoption stemming from socioeconomic and demographic factors, recent research shows that inequalities in Internet access are also shaped by geography. The effect of place on adoption is partially due to the lack of broadband infrastructure in rural locations (FCC 2010), but focusing only on infrastructure and availability ignores the geography-based obstacles to high-speed Internet adoption found in America's largest cities (Mossberger and Tolbert 2009). As a result of issues like racial and economic segregation in large metropolitan areas (Fong and Cao 2008; Jargowski 1997; Massey and Denton 1993; Wilson 1987; 1996), some inner-city neighborhoods have home broadband adoption rates below 40%, which is significantly lower than the national adoption rate of over 65% (Mossberger, Tolbert, and Franko 2012). It is important to note that these neighborhood effects are separate from the influence of the individual characteristics mentioned above. In other words, simply living in a poverty stricken neighborhood makes an individual less likely to have a high-speed Internet connection at home, on top of any other personal barriers one faces such as a modest income or low education levels (Mossberger et al. 2012).

Making matters worse, the effects of individual and place-based factors are not limited to the adoption of broadband. The importance of having an effective online skill set has recently been emphasized by Mossberger, Tolbert, and Franko's (2012) evaluation of individuals who rely on mobile devices to access the Internet. Despite popular rhetoric that mobile access can close the digital divide, the authors show that those who do not have broadband access in their homes but do access the Internet using a smart device tend to use the Internet for a relatively smaller set of tasks. When examining how people use the Internet when they are online, less-connected individuals with a mobile-only connection are substantially less likely to use the Internet for activities such as searching for jobs, taking an online class, online banking, and accessing the news (Mossberger, Tolbert, and Franko 2012).

These findings suggest recent advancements in technology are creating a new system of social stratification based on access to electronic information *and* the ability to effectively use this information. The most disadvantaged in this system are individuals with no access to the Internet at home and those with second-class access (e.g., relying on mobile devices to use the Internet). Individuals with the highest quality broadband connections who have the most developed online skills are those likely benefiting most from the system. But to more thoroughly understand how the structure of technology inequality affects society, scholars must be able to use a robust measure that adequately accounts for these disparities. The measure should not only represent differences in broadband access and online activity, but it should also allow researchers to examine how technology use varies geographically.

With the importance of place, Internet access, and how the Internet is used in mind, this project has developed measures of activity-based Internet use inequality at the local level. These novel measures were created using the 2011 Current Population Survey (CPS) Computer and Internet Use supplement and consist of several aggregate county- and metro-level estimates of online activities across the U.S. for various economic groups.

Estimating Activity-Based Internet Use Inequality at the Local Level

Questions from the 2011 CPS Computer and Internet Use supplement asking respondents whether they use the Internet to participate in economic, health-related, and other information-seeking activities are used to estimate the differences in online use across counties and metro areas. Several questions assess whether individuals use the Internet for these activities by asking if respondents "rely on the Internet" (1) to "work from home or for telecommuting," (2) for "financial services," (3) for "personal communications," (4) for "healthcare," and (5) for "general information."

Aggregate estimates of each question at the local level have been estimated by subsets of respondents according to their economic status. This allows for an assessment of the unequal structure of technology use for a potentially important set of online activities. These income-based measures of Internet use will allow researchers to more carefully analyze who benefits from the current composition of digital literacy.

As an example of how studying income differences in Internet-based activities can give us a better sense of digital inequality, consider two counties where 40% of the public in both areas uses the Internet for work. When observing the income breakdown of this particular online skill, the poor and rich residents in the first county use the Internet for work at rates of 30% and 50%, respectively. Yet the income breakdown of the second county reveals that only 10% of poor residents use the Internet for work and 70% of rich residents do the same. This simple illustration demonstrates how two areas can have similar rates of an Internet use on average, but examining how different income groups participate in the same activity can help us understand the structure of digital inequality. Of course, using income to examine inequalities in online activity will not completely account for all of the factors that contribute to biases in Internet use such as education, race, and age. Income will, however, be correlated with many of these factors and at a minimum can be used to assess the economic disparities in online skill.

The CPS is used to measure Internet activity-based inequality for two main reasons. First, the 2011 version of the survey asks the questions required to estimate the various Internet activities discussed above. Second, the CPS interviews approximately 130,000 U.S. residents for each supplement, which provides a large enough sample to measure aggregate behavior at the county and metro level using advanced estimation techniques (discussed in detail below). While the CPS does collect a variety of geographic information about the survey respondents, indicators at the lowest levels of geography are only included for some respondents in the data released for public use due to confidentiality concerns. In the case of the 2011 CPS Internet supplement, the survey includes 281 and 248 publicly available county and metro codes, respectively. The combined population of the counties with available indicators is approximately 143.5 million (46% of U.S. population) and 228.8 million (74% of U.S. population) for all available metro areas.

Now the question becomes how CPS samples of American adults can be used to estimate aggregate county- and metro-level online activities. The most common approach to obtaining aggregate measures using individual survey responses in small geographic areas in the absence of polls specifically designed to sample these small populations (which are quite rare) is to use some form of disaggregation. Disaggregation generally involves combining many national surveys to accumulate enough respondents in each area of interest and then disaggregating responses. Even though the CPS samples are relatively large, disaggregation is still problematic since not nearly enough individuals are surveyed from each local area to produce representative samples at the county and metro level. Recent advances in estimation, however, have given researchers an alternative to disaggregation when studying survey responses at the local level. Multilevel regression and poststratification (MRP) is a measurement strategy that allows for the estimation of local activities using typical national opinion polls. Research has shown that MRP provides accurate estimates of state and county opinion, for instance, even when using a single national survey (Lax and Phillips 2009a; 2009b; 2012; Park et al. 2006). This is the approach that is used here to create unique measures of online activity inequalities for counties and metro areas.

Estimating aggregate Internet use via MRP involves two steps. The first is to model individual responses to the survey question of interest—in this case, whether the individual used the Internet for one of the five activities being analyzed—using multilevel regression. These models include basic demographic and geographic characteristics of the survey respondents. Similar to pervious work, this study will use the following characteristics to model Internet activities: income (bottom quartile, second quartile, third quartile, or top quartile of the income distribution), race (black, white, or other), gender (female or male), age (18-29, 30-44, 45-64, or 65+), education (less than high school graduate, high school graduate, some college, or college graduate), county/metro of residence, and county/metro unemployment rates. The results of the model are then used to predict the probability of participating in the online activity for every possible individual type (e.g., a white female who is 30-44 years of age with some college education living in Cuyahoga County, Ohio).

To summarize, each dependent variable *y* is modeled as a function of individual (*i*) and county or metro (*j*) characteristics using the following multilevel structure:

$$\Pr\left(Activity_{ij} = 1\right) = logit^{-1}(\gamma_{00} + \gamma_{10} race_{ij} + \gamma_{20} gender_{ij} + \gamma_{30} age_{ij} + \gamma_{40} educ_{ij} + \gamma_{10} income_{ij} + \gamma_{01} unemp_{1j} + u_{0j} + u_{1j} income_{ij} + e_{ij})$$

where the variable *Activity* represents each of the online activities being evaluated, and is equal to 1 if the respondent uses the Internet for the activity and 0 if not. The model nests respondents within counties/metros and the county/metro estimates are also modeled as a function of unemployment rates. The γ estimates can be thought of as the fixed portions of the model while the *u* terms can be considered random effects. The u_{0j} term simply indicates that the model intercept can vary by geographic area. The second random term included in the model, u_{1j}, allows the effect of income on issue priorities to also vary by county or metro (this is estimated as the difference from the overall average effect of income, which is provided by the γ_{10} term). Modeling income as a random component is necessary to create the measures since this will allow for the estimation of not only whether use of the Internet is different for low- and high-income groups, but also if these differences are more pronounced in some geographic areas.

The probabilities estimated from the models described above are then used in the second step of the estimation, which is poststratification. Poststratification is the process of weighting each individual type probability estimate by the actual proportion of each type in the population using data from the U.S. Census. This part of the procedure adjusts for any differences between the individuals surveyed in each area and the actual population of each county and metro area according to Census data.

The final result is aggregate county and metro estimates of the five online activities by income subgroup (i.e., bottom, second, third, and top income quartile). Each variable ranges from a minimum of 0% to a maximum of 100% and represents the percent of residents in each county and metro who participate in each of the online activities. The details of datasets that include the measures are provided below.

Codebook

The activity-based Internet use inequality measures discussed above are available in two datasets:

```
InternetActivityIneq-CO.csv
InternetActivityIneq-MSA.csv
```

The first file contains the county estimates and the second file contains the metro area estimates. Table 1 provides the question wording used in the 2011 CPS Internet supplement for each of the survey questions used to create the aggregate measures of online activity. The variables included in each data file are described in Tables 2 and 3. Note that there are four separate estimates for each Internet activity, one for each income quartile. Income quartiles are indicated using the numerals 1-4. For instance, the variable `netrely_work_1` contains the estimates for the percentage of those in the bottom income quartile who rely on the Internet for work. For the same online activity, `netrely_work_2`, `netrely_work_3`, and `netrely_work_4` contain the estimates for those in the second, third, and top income quartiles, respectively.

Table 1: Questions from 2011 CPS Used to Create Measures of Activity-Based Internet Use Inequality

Variable Name	Question Wording: "Do you rely on the Internet for any of the following…"
PEPR3A1	"Working from home or telecommuting?"
PEPR3A3	"Financial services (such as banking, investing, or trading)?"
PEPR3A6	"Personal communications (such as email, instant messaging, social networking, blogging, or sharing photos)?"
PEPR3A7	"Healthcare?"
PEPR3A8	"General Information (such as news, weather, sports, maps, or government)?"

Table 2: County-Level Variables Measuring Activity-Based Internet Use Inequality (file name: `InternetActivityIneq-CO.csv`)

Variable Name	Variable Description	Mean	Std. Dev.	Min.	Max.
fips_stco	Combined FIPS state and county code				
state_name	State name				
county_name	County name				
netrely_work_1	% relying on Internet for work: bottom income quartile	7.74	2.01	3.83	16.87
netrely_work_2	% relying on Internet for work: second income quartile	13.84	2.81	7.52	24.09
netrely_work_3	% relying on Internet for work: third income quartile	25.19	4.08	15.56	38.85
netrely_work_4	% relying on Internet for work: top income quartile	40.01	5.30	26.29	58.49
netrely_finance_1	% relying on Internet for financial services: bottom income quartile	22.16	6.10	9.91	45.32
netrely_finance_2	% relying on Internet for financial services: second income quartile	33.32	7.10	17.21	55.44
netrely_finance_3	% relying on Internet for financial services: third income quartile	48.82	8.27	25.90	68.10
netrely_finance_4	% relying on Internet for financial services: top income quartile	63.37	8.40	37.29	81.36
netrely_comm_1	% relying on Internet for personal communication: bottom income quartile	42.05	8.63	22.36	68.71
netrely_comm_2	% relying on Internet for personal communication: second income quartile	54.81	8.13	32.84	74.72
netrely_comm_3	% relying on Internet for personal communication: third income quartile	69.21	7.40	46.25	85.55
netrely_comm_4	% relying on Internet for personal communication: top income quartile	80.04	6.08	58.28	92.53
netrely_health_1	% relying on Internet for healthcare: bottom income quartile	11.77	3.87	2.99	28.30
netrely_health_2	% relying on Internet for healthcare: second income quartile	16.39	4.88	4.53	35.97
netrely_health_3	% relying on Internet for healthcare: third income quartile	23.45	6.28	7.42	46.72
netrely_health_4	% relying on Internet for healthcare: top income quartile	31.91	7.74	10.88	59.00
netrely_geninfo_1	% relying on Internet for general information: bottom income quartile	32.57	7.94	11.72	59.49
netrely_geninfo_2	% relying on Internet for general information: second income quartile	44.54	9.15	16.84	69.83
netrely_geninfo_3	% relying on Internet for general information: third income quartile	59.12	10.28	24.20	79.81
netrely_geninfo_4	% relying on Internet for general information: top income quartile	71.27	10.20	31.67	87.49

Note: Each variable has 281 observations.

Table 3: Metro-Level Variables Measuring Activity-Based Internet Use Inequality (file name: `InternetActivityIneq-MSA.csv`)

Variable Name	Variable Description	Mean	Std. Dev.	Min.	Max.
fips_msa	MSA FIPS code				
msa_name	MSA name				
netrely_work_1	% relying on Internet for work: bottom income quartile	7.62	2.17	4.03	16.94
netrely_work_2	% relying on Internet for work: second income quartile	13.74	2.83	8.03	24.05
netrely_work_3	% relying on Internet for work: third income quartile	25.18	3.84	16.72	38.64
netrely_work_4	% relying on Internet for work: top income quartile	40.13	4.63	28.22	53.58
netrely_finance_1	% relying on Internet for financial services: bottom income quartile	22.26	6.14	11.85	44.54
netrely_finance_2	% relying on Internet for financial services: second income quartile	33.45	6.71	20.22	54.85
netrely_finance_3	% relying on Internet for financial services: third income quartile	49.10	7.23	32.79	67.97
netrely_finance_4	% relying on Internet for financial services: top income quartile	63.80	6.89	46.38	79.51
netrely_comm_1	% relying on Internet for personal communication: bottom income quartile	41.84	7.84	23.61	66.06
netrely_comm_2	% relying on Internet for personal communication: second income quartile	55.23	7.04	35.29	72.55
netrely_comm_3	% relying on Internet for personal communication: third income quartile	70.28	6.01	50.45	83.14
netrely_comm_4	% relying on Internet for personal communication: top income quartile	81.40	4.68	63.97	90.30
netrely_health_1	% relying on Internet for healthcare: bottom income quartile	11.01	3.16	4.02	26.38
netrely_health_2	% relying on Internet for healthcare: second income quartile	15.84	3.92	6.25	32.86
netrely_health_3	% relying on Internet for healthcare: third income quartile	23.33	5.07	10.38	44.59
netrely_health_4	% relying on Internet for healthcare: top income quartile	32.52	6.24	15.42	55.66
netrely_geninfo_1	% relying on Internet for general information: bottom income quartile	32.36	7.45	12.83	57.96
netrely_geninfo_2	% relying on Internet for general information: second income quartile	44.56	8.20	18.93	68.69
netrely_geninfo_3	% relying on Internet for general information: third income quartile	59.54	8.85	27.79	79.15
netrely_geninfo_4	% relying on Internet for general information: top income quartile	71.97	8.53	36.92	87.20

Note: Each variable has 248 observations.

References

Fong, Eric, and Xingshan Cao. 2008. "Bridges across the Racial Digital Divide: Residential Ecology of Internet Use." *Canadian Studies in Population* 35 (2): 243–68.

Lax, Jeffrey R., and Justin H. Phillips. 2009a. "Gay Rights in the States: Public Opinion and Policy Responsiveness." *American Political Science Review* 103 (03): 367–86.

———. 2009b. "How Should We Estimate Public Opinion in The States?" *American Journal of Political Science* 53 (1): 107–21.

———. 2012. "The Democratic Deficit in the States." *American Journal of Political Science* 56 (1): 148–66.

Massey, Douglas S., and Nancy A. Denton. 1993. *American Apartheid: Segregation and the Making of the Underclass*. Cambridge, MA: Harvard University Press.

Mossberger, Karen, and Caroline J. Tolbert. 2009. "Digital Excellence in Chicago: A Citywide View of Technology Use." http://www.cityofchicago.org/city/en/depts/doit/supp_info/digital_excellenceinchicagoacitywideviewoftechnologyuse.html.

Mossberger, Karen, Caroline J. Tolbert, and William W. Franko. 2012. *Digital Cities: The Internet and the Geography of Opportunity*. New York, NY: Oxford University Press.

Mossberger, Karen, Caroline J. Tolbert, and Mary Stansbury. 2003. *Virtual Inequality: Beyond the Digital Divide*. Washington, D.C.: Georgetown University Press.

Norris, Pippa. 2001. *Digital Divide: Civic Engagement, Information Poverty, and the Internet Worldwide*. Cambridge, MA: Cambridge University Press.

Park, David K., Andrew Gelman, and Joseph Bafumi. 2006. "State-Level Opinions from National Surveys: Poststratification Using Multilevel Logistic Regression." In *Public Opinion in State Politics*, edited by Jeffrey E. Cohen, 209–28. Stanford, CA: Standford University Press.

Wilson, William Julius. 1987. *The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy*. Chicago, IL: University of Chicago Press.

———. 1996. *When Work Disappears: The World of the New Urban Poor*. New York, NY: Alfred A. Knopf.