The Relationship between Broadband Access and Online Course enrolment at Postsecondary Institutions with Open Admissions Policies

Hanjing Huang

2020/12/9

GitHub Repo: `https://github.com/juliehuang216216/STA304-Project`

Keywords

Distance education, Online Courses, Broadband, Bayesian modeling

Introdcution

Since the 21st century, the proportion of online education in post-secondary education has gradually increased (Snyder et al. 2016; Radford and Weko 2011). Although the mass media often pay more attention for large-scale online open courses (MOOCs) (for example, Pappano 2012), relatively speaking, only a few number of college students only receive distance education credits. Usually, they take the courses on campus and few online courses as supplement through their home institution (Snyder et al. 2016; Snyder and Dillow 2015). With the outbreak of the epidemic (2020 Covid-19), in the case of mandatory e-learning, it is particularly important to study the factors affecting the enrollment of online courses.

This research is a reproduction of the research 'Making the connection: Broadband access and online course enrolment at public open admissions institutions' (Skinner, Ben, 2019). Through reproducing, to explore the relationship between high-speed broadband access and the number of students in public universities / community colleges who take certain courses online (Skinner, Ben, 2019). The enrollment data is collared from the Integrated Postsecondary Education Data System (National Center for Education Statistics Department of Education 2015) and census block-level measures of broadband access is collared from the National Broadband Map (National Telecommunications and Information Administration 2011) from 2012 to 2014 (Skinner, Ben, 2019). I assign each school broadband measures that are the population/inverse distance-weighted averages of those recorded in surrounding census block groups to estimate the download speeds and number of providers of the average student enrolled at each institution in the sample (Skinner, Ben, 2019).

Research conclusions and discussions will be appended at the end of the research.

Methodology (Data and Model)

Data

The following data will be used for the project:
American Community Survey / Census
Broadband(12, 13, 14)
Geographic Data

The Integrated Postsecondary Education Data System (IPEDS)
State appropriations by year come from the State Higher Education Executive Office.(SHEEO)

The specific description of the data is listed below, because the broadband data is huge and takes a lot of time to decompress, and the collation, induction and cleaning of other data are also very complicated. In practical application, the cleaned version is used.

Note: The original link to the SHEEO data has been invalidated.

Institution Data

Data on the number of students enrolled in online courses come from Integrated Post-secondary Education Data System (National Center for Education Statistics Department of Education 2015).

Although the IPEDS survey has been asking for years whether institutions are mainly distance learning schools (the answer is yes or no), since the fall of 2012, it has only required institutions to break down the total number of students who have tried distance learning courses.

Table 1: Descriptive statistics of the institution sample

	Mean/(SD)
Total enrollment	7725
	(7819)
Some online enrollment	1361
	(1536)
Two year institution	0.88
	(0.33)
Has on-campus housing	0.25
	(0.44)
Non-white enrollment	0.42
	(0.24)
Women enrollment	0.42
	(0.07)
Pell grant recipients	0.42
	(0.15)
Part-time enrollment	0.57
	(0.15)
Aged 25 years and older	0.37
	(0.11)
2013	0.4
	(0.49)
2014	0.3
	0.46
N (2012)	750
N (2013)	1003
N (2014)	741

Notes. Total enrollment and some online enrollment represent the average number of students rounded to nearest student. Other rows are proportions. Standard deviations are shown in parentheses. Schools included in the sample are public, open admissions postsecondary institutions that report at least one student who took some distance education courses.(Skinner, Ben, 2019)

Table 1 shows the average and standard deviation of the number of students enrolled in credit programmes, the number of students participating in some online courses, and the covariates of other institutions.

Schools in Alaska and Hawaii were removed from the final analysis data set because of their unique background-Alaska is a vast but sparsely populated state. Hawaii is an island group-which may be biased towards the weight used to build the broadband metrics I assign to each institution.

Because data from all branches of the Ivy Tech (Ivy Tech), a two-year public college in Indiana, are concentrated under a single identifier, these institutions are also removed from the data set because broadband measures cannot be accurately assigned to them. The final estimates represent 1017 unique public open admissions agencies observed over the past three years.

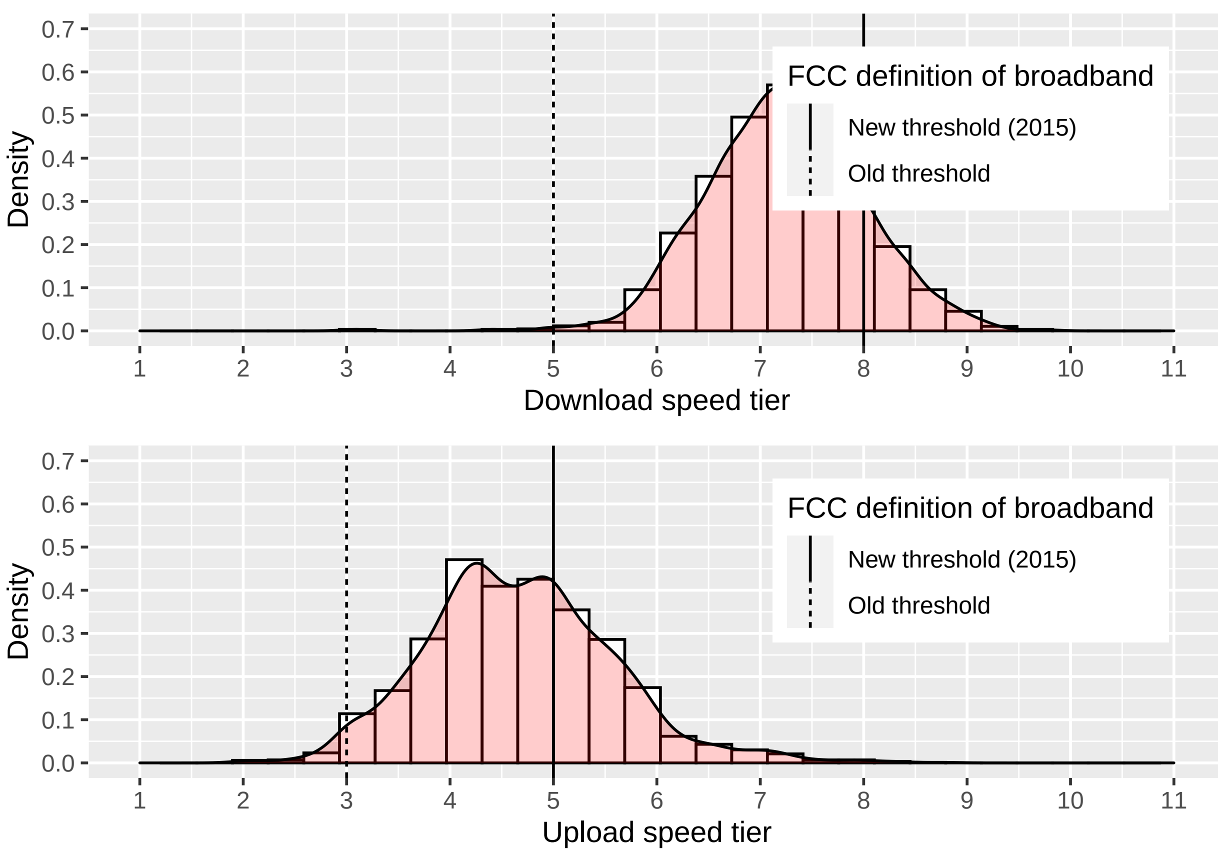
Broadband Data

Broadband data is collected from the National Broadband Map website. At the request of the National Telecommunications and Information Administration, in cooperation with the Federal Communications Commission (FCC), these data are collected from. The Internet service provider (ISP) in each state is provided by a designated authorized authority.

Each service provider provides information on upload and download rates at the census district level, as well as information on the number and type of community anchors (usually libraries, K-12 schools, college, etc.). In the region (National Telecommunications and Information Administration 2011).

These data are verified with other broadband information sources and released to the public. Use the data released in June 2012, 2013 and June 2014, as they are roughly the same time period as IPEDS data collection and represent the best estimates of broadband connectivity around the research facility during the study period.

Geographic and Demographic Data



When estimating multilevel models, I use additional state level covariates to aid in fitting second level parameters. These include measures of statewide unemployment rates, which were taken from data provided by the Bureau of Labor Statistics (Bureau of Labor Statistics 2012, 2013, 2014). To account for potential differences that funding structures could have on the availability of online courses, measures of state appropriations per full-time equivalent student in each year were gathered from a report produced by the State Higher Education Executive Officers Association (SHEEO) (Carlson et al. 2015). I also include a measure of the proportion of 2-year open admissions public institutions within each state that I computed using data from IPEDS. These variables were intended to account for potential differences across states in the number of students who

Model

I used Bayesian univariate regression model and multilevel regression model to estimate the correlation between broadband access download speed, upload speed, number of providers and online course enrollment. In this case, there are two main reasons why Bayesian paradigms, contrary to the frequently occurring paradigms, occur.

In the analysis, I first estimate a number of single-level Bayesian linear regression models that take the form

$$log(y_i) \sim N(\alpha + \beta broadband_i + X_{\gamma, {}^2_y})$$

where y_i is the number of students enrolled in some online courses; α is a constant term; β is a parameter of interest to broadband, which is the broadband measurement assigned by the institution; and X is the covariable data value matrix as its corresponding parameter vector γ. Because the number of students taking some online courses is correct, I am in line with the logarithmic conversion of these values. An additional benefit of using the natural logarithm of the results is that each additional unit in the broadband interest measurement can indicate a percentage change in the number of students attending some online courses (Greene 2012).

With the multilevel model, each state, represented by α_j, is allowed to have its own intercept. It takes the form

$$log(y_i) \sim N(\alpha + \beta broadband_i + X_{\gamma, {}^2_y})$$

$$\alpha_j \sim N(\delta_s Region_s + Z_\varphi, \sigma_s^2)$$

in which X represent a vector of school and county level covariates. Each state intercept, α_j,is modeled using state-level covariates, Z , and a region-specific intercept, δ_s . As with the single-level models, all unknown parameters (β, γ, φ, δ, σ_y^2 ,σ_s^2) are given diffuse priors.

Results

In order to generate my results, I use a computation-based Markov chain Monte Carlo algorithm, which is repeatedly applied to each Bayesian model. Although simple Bayesian models can be solved analytically, nontrivial equations are often too complex or have no closed form solution (Gelman et al. 2014)To solve these problems, computer programs use iterative processes to suggest, compare, and accept or reject parameter values. With enough iterations, the parameters generated by this process will come from the real posterior distribution (Brooks et al. 2011).Although there is no test to determine whether enough samples have been taken, thus effectively summarizing the real a posterior results.

Single-Level Models

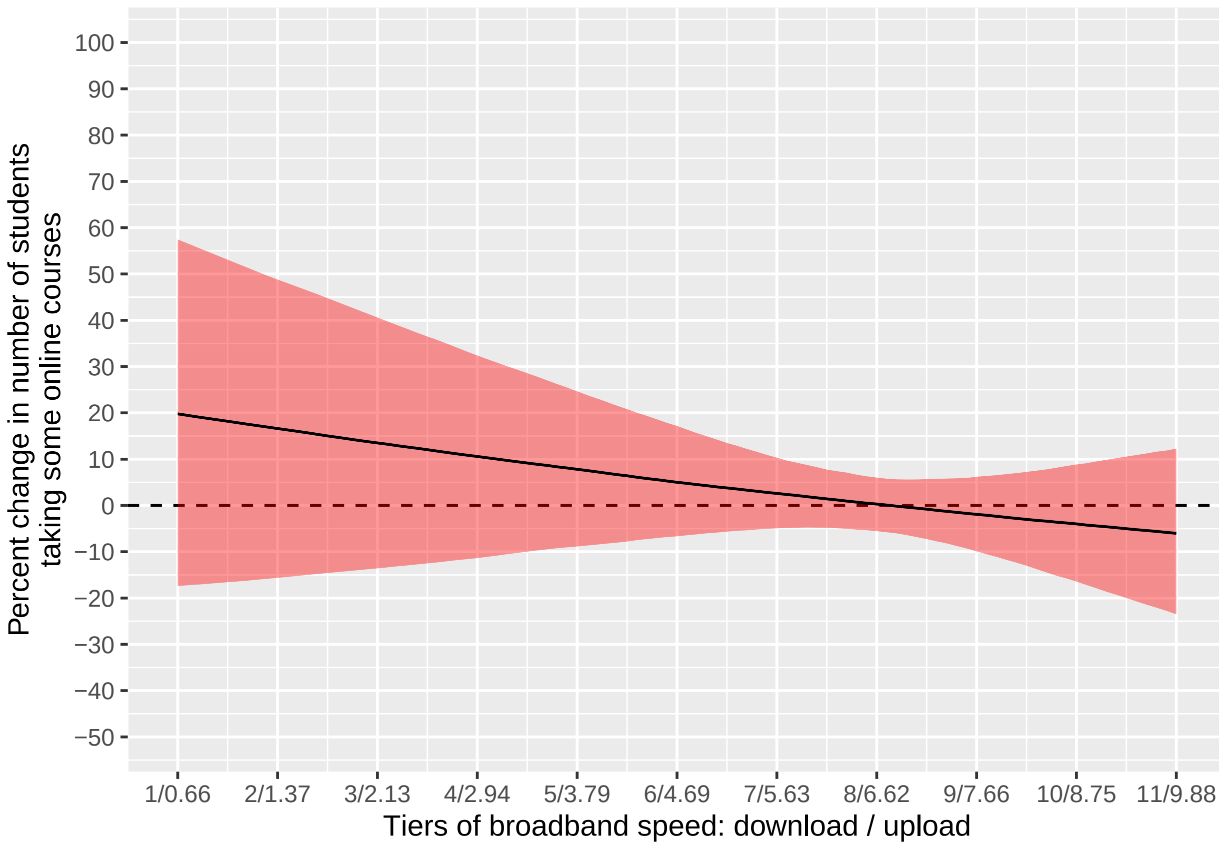


Table 2: Single level Bayesian beta regressions of percentage of students who enrolled in some distance education courses on broadband measures.

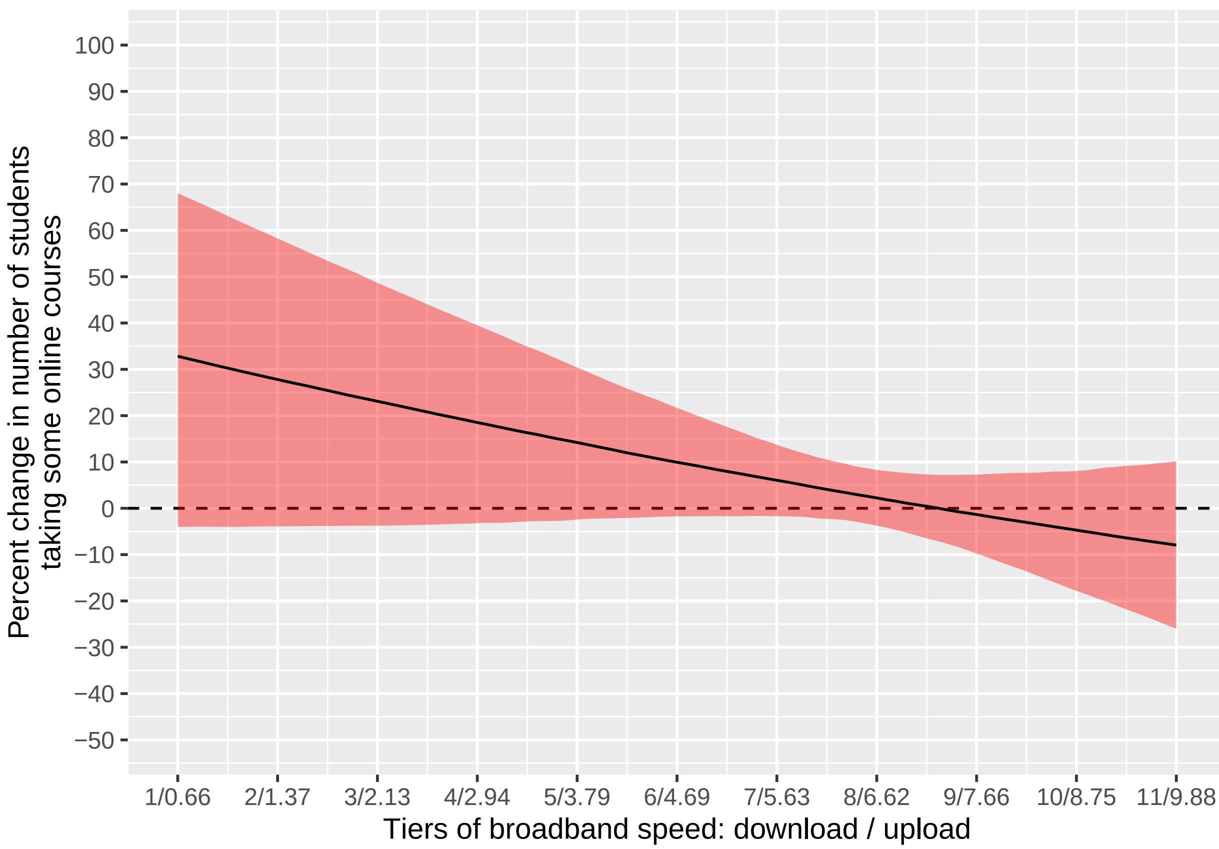
	(1)	(2)	(3)	(4)
Download speed	0.333			0.41
	$[-0.027, 0.701]$			$[-0.007, 0.82]$
Download speed^2	-0.025			-0.027
	$[-0.05, 0]$			$[-0.056, 0.002]$
Upload speed		-0.077		-0.214
		$[-0.272, 0.114]$		$[-0.44, 0.021]$
Upload speed^2		0.003		0.016
		$[-0.016, 0.022]$		$[-0.007, 0.039]$
# Providers			0.007	0.029
			$[-0.102, 0.112]$	$[-0.084, 0.144]$
# Providers^2			-0.003	-0.004
			$[-0.015, 0.01]$	$[-0.017, 0.009]$
Two year institution	0.058	0.067	0.065	0.058
	$[-0.016, 0.131]$	$[-0.007, 0.139]$	$[-0.009, 0.14]$	$[-0.015, 0.132]$
Has on-campus housing	-0.049	-0.067	-0.048	-0.063
	$[-0.12, 0.024]$	$[-0.136, 0.004]$	$[-0.12, 0.024]$	$[-0.133, 0.006]$
log(Total enrollment)	1.154	1.156	1.155	1.157
	$[1.117, 1.191]$	$[1.119, 1.193]$	$[1.119, 1.189]$	$[1.12, 1.195]$
Prop. non-white	-0.633	-0.63	-0.635	-0.611
	$[-0.755, -0.509]$	$[-0.746, -0.509]$	$[-0.75, -0.519]$	$[-0.73, -0.492]$
Prop. women	-2.219	-2.201	-2.237	-2.203
	$[-2.653, -1.799]$	$[-2.624, -1.784]$	$[-2.656, -1.82]$	$[-2.626, -1.765]$
Prop. Pell grant	0.614	0.568	0.608	0.578
	$[0.395, 0.832]$	$[0.353, 0.786]$	$[0.4, 0.818]$	$[0.353, 0.801]$
Prop. part-time	-0.476	-0.495	-0.469	-0.488
	$[-0.697, -0.247]$	$[-0.717, -0.273]$	$[-0.695, -0.241]$	[-0.708, -0.266]
Prop. 25 years and older	0.397	0.405	0.408	0.413
	$[0.155, 0.638]$	$[0.158, 0.646]$	$[0.156, 0.662]$	$[0.171, 0.665]$
log(Pop. density)	-0.091	-0.084	-0.087	-0.087
	$[-0.118, -0.064]$	$[-0.111, -0.058]$	$[-0.114, -0.06]$	$[-0.114, -0.06]$
2013	0.103	0.11	0.1	0.105
	$[0.045, 0.159]$	$[0.054, 0.165]$	$[0.045, 0.158]$	$[0.047, 0.164]$
2014	0.15	0.164	0.142	0.159
	$[0.088, 0.211]$	$[0.102, 0.225]$	$[0.08, 0.204]$	$[0.095, 0.222]$
(Intercept)	6.276	6.276	6.276	6.276
	$[6.703, 6.749]$	$[6.702, 6.75]$	$[6.701, 6.751]$	$[6.702, 6.749]$
Unique institutions	1017	1017	1017	1017
N	2494	2494	2494	2494

Notes. Bayesian point estimates represent posterior mean values. Values in the square brackets are 95% credible intervals. Covariates not reported include indicators for USDA urban/rural community codes. Parameter distributions in each model are the combination of four independent MCMC chains of 1000 draws each (with 1000 initial draws discarded as burn-in) for a total of 4000 draws. All models were estimated using the Stan NUTS sampler with a reparameterized beta likelihood sampling statement. The outcome measure in all models is the percentage of students at each institution who enrolled in some distance education courses.(Skinner,Ben, 2019)

Table 2 shows the results of the single-level model. Bayesian point estimates represent the average of a posterior distribution, and the accompanying numbers in square brackets represent a 95% confidence interval. Models 1-3 in turn use each broadband metric (download speed, upload speed, and number of providers) and its quadratic power. Model 4 uses all broadband measures and their squares in the same equation. All models include indicators of two-year colleges, housing availability, year, and USDA urban and rural continuum code (not reported); the total number of institutional enrollment and its natural logarithm of proportion among students of color, women, Pell grant recipients, part-time students, and students aged 25 and older; and population density measured at the county level (on record). In all models, the dependent variable is the natural logarithm of the number of students taking some online courses. Therefore, the posterior distribution of parameters represents the percentage change in the number of online students sometimes when one unit in the covariable changes.

Multilevel Models

To better explain the nesting nature of the data and the differences in the context of state higher education policies, I fitted a multi-level model that allows each state to have its own interception (Gelman et al. 2014). Like the single-layer linear regression model, the dependent variable is the logarithmic conversion of the number of students who try online courses. In addition to the covariates used in the single-stage equation, I include some secondary covariates to help predict the unique interception of each state. These indicators include the state average unemployment rate, the average funding per full-time equivalent student across the state, the proportion of the state's two-year public colleges and universities that open enrollment, and. The population-weighted average distance to the nearest public open admissions agency.



In terms of space, the second-level parameters and state-specific interception Bayesian point estimation are not reported.

Discussion

This paper provides some evidence that there is a positive correlation between the download speed of broadband access and the number of online course enrollment in public colleges and universities in the United

States.

Among the preferred model specifications, I found that the average increase in download speed in the area around the institution corresponds to an increase in the number of students taking some online courses, which can be as high as 33.2% at the lowest speed and 2.3% under the new broadband speed threshold.

At the highest speed level, the average change of margin may be negative, although the confidence interval in this range is fully distributed above zero, indicating that the changes in both cases are very small.

Although the range of potential marginal values is wide-which means that specific values should be carefully interpreted-according to FCC, the probability of a positive marginal correlation between download speed and sometimes online students is greater than 50% between the lowest speed layer and the currently defined broadband.

References

Skinner, Ben, 2019, ‘Making the connection: Broadband access and online course enrollment at public open admissions institutions’, Research in Higher Education

Snyder, T. D., de Brey, C., & Dillow, S. A. (2016). Digest of Education Statistics: 2015. NCES 2016-014. Washington, D.C.: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Snyder, T. D., & Dillow, S. A. (2015). Digest of Education Statistics: 2013. NCES 2015-011. Washington, D.C.: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Radford, A. W., & Weko, T. (2011). Learning at a Distance: Undergraduate Enrollment in Distance Education Courses and Degree Programs. NCES 2012-154. Washington, D.C.: U.S. Department of Education, National Center for Education Statistics, Institute of Education Sciences.

Carlson, A., Ott, C., Armstrong, J., Zaback, K., & Auer, G. (2015). State Higher Education Finance FY 2014. Tech. rep. Boulder, CO: State Higher Education Executive Officers Association.

Pappano, L. (2012). The Year of the MOOC. In The New York Times.

Greene, W. H. (2012). Econometric Analysis (7th ed.). Boston, MA: Prentice Hall.

Gelman, A. (2014). Bayesian data analysis (3rd ed.). Boca Raton: CRC Press.