PHOENIX
CENTER
FOR ADVANCED LEGAL & ECONOMIC PUBLIC POLICY STUDIES
www.phoenix-center.org

PHOENIX CENTER POLICY BULLETIN NO. 62

George S. Ford, PhD

November 2022

THE BROADBAND TRIBAL GAP: AN EMPIRICAL EVALUATION

Abstract: In this BULLETIN, I study broadband deployment over the years 2014-2020 in Tribal and non-Tribal census tracts using the Federal Communications Commission's Form 477 data to quantify progress. This "Tribal Gap" is measured as the difference in average broadband availability between Tribal and non-Tribal census tracts. Unmatched and matched sample are used, and a sample of census tracts within 30 miles of a Tribal area are also analyzed with and without matching. In all cases, the gap between Tribal and non-Tribal census tracks has been getting closer to zero over time and by 2020 (the last year data are available) the Tribal Gap was near zero in all cases, especially when the deployment differences are conditioned on a few covariates. Indeed, the Tribal Gap is nearly fully explained by differences in demographic characteristics. These results are encouraging and suggests efforts to close the Tribal Gap are meeting with some success, though many factors that determine deployment largely are beyond regulatory remedy (e.g., population density).

I. Background

Deploying broadband to all Americans has long been a goal of the U.S. Government.[1] Some progress has been made and the bi-partisan Infrastructure Investment and Jobs Act of 2021

---

1 *See, e.g.*, Section 706 of the Telecommunications Act of 1996, 47 U.S.C. § 1302 ("The Commission and each State commission with regulatory jurisdiction over telecommunications services shall encourage the deployment on a reasonable and timely basis of advanced telecommunications capability to all Americans….").

provides a sizable subsidy of nearly $45 billion to complete the task.[2] Broadband deployment to Tribal areas in particular is a longstanding problem, and thus has received considerable attention by the Federal Communications Commission ("FCC").[3] Has this focus been productive? In this BULLETIN, I study broadband deployment over the years 2014-2020 in Tribal and non-Tribal census tracts using the FCC's Form 477 data to quantify progress.

As has been discussed in some detail, the Form 477's assumption that a census block is "served" if a single customer has access (or could have access in a short time) could lead to a substantial overstatement of availability in the relatively large areas covered by blocks in Tribal areas.[4] Two approaches aim to address the relatively large square mileage of Tribal tracts. First, Coarsened Exact Matching ("CEM") is used to select a sample of non-Tribal tracts that are, in many respects, identical to Tribal areas, including square mileage. Second, the sample is limited to blocks (aggregated to tracts) within 30 miles of a Tribal area; these census blocks presumably are more like Tribal blocks in several dimensions, especially related to the costs of deployment. Across the three approaches, the evidence is clear that the Tribal Gap is shrinking. Moreover, when conditioned on a few covariates (population, the tract's land area, median income, and unemployment), the Tribal Gap is zero or nearly so. Thus, the Tribal Gap can be explained by relatively few factors.

Several recent studies look at Internet availability and use on Tribal lands. For instance, Howard and Morris (2020) conduct a small primary survey which suggests that residents on Tribal lands predominantly use mobile wireless service to connect to the Internet, with many others using public Wi-Fi or shared connections.[5] In an econometric analysis using multiple large

---

2 *Congress Approves Infrastructure Bill with $65 Billion for Broadband Programs*, JDSUPRA (Wiley Rein LLP) (November 10, 2021) (available at: https://www.jdsupra.com/legalnews/congress-approves-infrastructure-bill-6672990).

3 *See, e.g., Report on Broadband Deployment in Indian Country Pursuant to the Repack Airwaves Yielding Better Access for Users of Modern Services Act of 2018, Federal Communications Commission, Prepared by the Consumer & Governmental Affairs Bureau, Wireless Telecommunications Bureau and Wireline Competition Bureau, Submitted to the Senate Committee on Commerce, Science, and Transportation House of Representatives Committee on Energy and Commerce* (May 2019) (available at: https://docs.fcc.gov/public/attachments/DOC-357269A1.pdf) (hereinafter "*FCC Tribal Broadband Report*").

4 G.S. Ford, A *Quality Check on Form 477 Data: Errors, Subsidies, and Econometrics*, PHOENIX CENTER POLICY PERSPECTIVE No. 21-05 (October 27, 2021) (available at: https://www.phoenix-center.org/perspectives/Perspective21-05Final.pdf); G.S. Ford, *Quantifying the Overstatement in Broadband Availability from the Form 477 Data: An Econometric Approach*, PHOENIX CENTER POLICY PERSPECTIVE NO. 19-03 (July 11, 2019); *Tribal Broadband: FCC's Data Overstate Access, and Tribes Face Barriers Accessing Funding*, Statement of Mark Goldstein, Director, Physical Infrastructure Issues, Government Accountability Office, GAO-19-134T (October 3, 2018) (available at: https://www.gao.gov/products/gao-19-134t).

5 B. Howard and T. Morris, *Tribal Technology Assessment: The State of Internet Service on Tribal Lands*, American Indian Policy Institute, Arizona State University (Fall 2019) (available at: https://aipi.asu.edu/sites/default/files/Tribal_tech_assessment_compressed.pdf).

datasets, Bauer, Feir, and Gregg (2022) use data on adoption, speed tests, and prices to look for Tribal disparities in these dimensions. The study does not look at availability, which is the focus here, but reports Tribal Gaps in several dimensions, though often a good portion of these gaps can be explained by a few factors.[6] Most akin to this BULLETIN, Mack, et al. (2022) use the FCC's Form 477 data to quantify the pattern in the Tribal Gap over time, though the authors measure the Tribal Gap as a difference in provider counts between Tribal and non-Tribal areas rather than availability of broadband service.[7] With average number of providers exceeding four providers on Tribal lands (and in some cases above six providers), the results are unbelievable. Many Tribal areas have no broadband provider and I suspect few have as many as four providers. By my count, less than 20% of Tribal areas have more than two providers and only 1.2% have as many four providers. The variable measuring the tract-level count of providers in Mack, et al., (2022) seems improperly constructed, an error discussed in Flamm and Vargas (2017) a few years earlier.[8] Most importantly, the number of providers does not measure broadband availability; it could be that two providers cover only 20% of homes, or one provider covers 100% of homes. While availability and provider count (properly constructed) are correlated ($\rho = 0.50$), the two outcomes are not measuring the same thing.

An important question is why availability may differ between Tribal and non-Tribal areas. As the FCC has observed, Tribal areas present several challenges for broadband deployment including: (1) rugged terrain; (2) complex permitting processes governing access to Tribal lands; (3) jurisdictional issues involving states and sovereign Tribal governments; (4) a predominance of residential, rather than business, customers; (5) high poverty rates and low-income levels on Tribal lands, and (6) cultural and language barriers.[9] I attempt to control for some of these differences, but there may be other barriers to deployment, including securing an arrangement with Tribal governments. In any case, while the Tribal Gap is mostly explained by a few

---

[6] M.T. Gregg, A. Bauer, and D. Feir, *The Tribal Digital Divide: Extent and Explanations*, Center for Indian Country Development Working Paper Series No. 2021-03 (revised June 2022) (available at: https://www.minneapolisfed.org/~/media/assets/papers/cicdwp/2021/cicd-wp-2021-03.pdf).

[7] E.A. Mack, E. Helderopb, T. Keene, S. Loveridge, J. Mann, T. H. Grubesic, B. Kowalkowski, M. Gollnow, *A Longitudinal Analysis of Broadband Provision in Tribal Areas*, TELECOMMUNICATIONS POLICY (forthcoming) (available at: https://doi.org/10.1016/j.telpol.2022.102333).

[8] K. Flamm and P. Varas, *The Evolution of Broadband Competition in Local US Markets: A Distributional Analysis*, Paper presented at the 46th Research Conference on Communications, Information and Internet Policy (TPRC) (August 15, 2018) (available at: https://ssrn.com/abstract=3142329).

[9] *FCC Tribal Broadband Report, supra* n. 3 at p. 1 (available at: https://www.fcc.gov/document/report-broadband-deployment-indian-country); *see also* G.S. Ford, T.M. Koutsky and L.J. Spiwak, *Competition After Unbundling: Entry, Industry Structure and Convergence,* 59 FEDERAL COMMUNICATIONS LAW JOURNAL 331 (2007) (available at: https://www.phoenix-center.org/papers/FCLJCompetitionAfterUnbundling.pdf).

demographic features, I make no specific claims about why a Tribal Gap may exist; my interest is limited to the Tribal Gap's magnitude.

II. Empirical Framework

My task is to estimate the "Tribal Gap," which I define as the means difference in the average availability of broadband between Tribal and non-Tribal census tracts. If, for example, non-Tribal areas have an average availability rate of 90% and Tribal areas 70%, then the Tribal Gap is 20 basis points. Ideally, the data would permit the average availability for a person or a household, but the Form 477 data are reported at the census block level. As the best approximation of availability, I use the lowest level of aggregation (the census block) to construct the availability variable and then aggregate to the census tract using a population-weighted mean. These outcomes are interpreted as the expected experience of a person living in the census tract. For instance, an availability rate of 0.90 indicates the typical household has a 90% probability of having access to a broadband network.

To estimate the Tribal Gap, the means differences over time t are estimated using,

$$y_{it} = \sum_{t=1}^{7} \Delta_t A_i \cdot D_t + \beta X_i + \mu_t + \varepsilon_{it}, \tag{1}$$

where y_{it} is the broadband availability (per capita) in census tract i in year t, A_i is an indicator of Tribal tracts (= 1 for Tribal, zero otherwise), D_t is a dichotomous indicator for each year t, X_i is a set of covariates, μ_t is a year fixed effect, and ε_i is the econometric disturbance term. Since availability is bound on the unit interval and is often close to 1.0, the model is estimated using a Generalized Linear Model ("GLM") of the binomial family with a logit link so the predictions remain on the unit interval. Standard errors are clustered at the county level and the regression is weighted by population (for year 2019) so that the estimates represent the typical experience of a person. The GLM coefficients are not directly interpretable, so the gap is measured by the difference in the mean predicted y_t between Tribal and non-Tribal tracts. Hypothesis tests are conducted on these predictions.[10]

Excluding the X covariates produces a simple, unconditional means difference. That is, the means differences are a simple comparison of averages. This is akin to the statistic reported in the FCC's *Broadband Reports*, though the spread may differ slightly due to differences in population values and Tribal area indicators used to construct the means. Such differences may be informative as they measure the observed difference in availability.

---

[10] Stata 17 is used for all estimations. Predictions are generated using the -margins- command and hypothesis tests are based on the delta method.

When including the *X*, the means difference is conditioned on those variables. That is, how much of the unconditional Tribal Gap may be explained by relevant factors such as cost and demand? The means difference plausibly may be conditioned on many factors. Here, I take a somewhat limited approach and include factors indicated by the FCC as relevant to the deployment differences: (1) population; (2) square mileage (the two components of population density); (3) the median household income; and (4) the unemployment rate. The first three are measured in natural log form. Of course, many other covariates might be considered (e.g., education, and so forth), though several of the potentially relevant variables are correlated with income.

A. *Data*

Broadband deployment of consumer services is measured using the December releases of the FCC's 477 data over the years 2014 through 2020, a period over which the Commission had control of broadband data collection.[11] My analysis occurs at the census tract level, so tract-level values are population-weighted means of data measured at aggregation levels smaller than the tract..[12] The data are restricted to a balanced panel, but only a few tracts are excluded due to missing values. A Tribal tract is marked by joining census blocks to Tribal areas using a Tribal tract shape file. Demographic data are from Safegraph (2019 version), which are based on the American Community Survey ("ACS") for years 2015-2019 (thus time invariant in this sample).[13] These data are native to the census blocks group; the values are aggregated to the tract using a population-weighted mean or sum in case of population, area, and so forth.

B. *Alternate Samples*

Many factors affect broadband deployment, and the comparison of broadband deployment across Tribal and non-Tribal areas may be best quantified by comparing similar census tracts across the two groups. The number of Tribal census tracts is small but total census tracts counts are very large; Tribal tracts are about 1% of total tracts and such a small share can sometimes lead to imprecise estimates. Also, rural Tribal census blocks often cover relatively large areas, and the overstatement of the one-all assumption of the Form 477 data is presumably larger in

---

[11] Data available at: https://www.fcc.gov/general/broadband-deployment-data-fcc-form-477. Prior to December-2014, states collected broadband availability data. In my experience, using data before-and-after the switch to FCC control is problematic as there are clear disparities at the join of the data sets.

[12] Analysis was also conducted at the block group level, but differences were small.

[13] Data available at: https://docs.safegraph.com/docs/open-census-data.

geographically-larger census areas.[14] To ensure reasonable comparisons and reliable estimates, I obtain a matched sample using Coarsened Exact Matching ("CEM").[15] The matching variables are population, square mileage (two variables that also define population density), median household income, the share of persons with a tertiary education, and the share of population that is minority (Black or Hispanic). The matching algorithm aims at a 2:1 match, though the unique character of Tribal areas prevents a 2:1 match in some cases.[16] The final match is 1.9:1 non-Tribal to Tribal tracts.

In addition to the CEM sample, I also analyze a sample of tracts within 30 miles of a Tribal area. Given the proximity to Tribal areas these tracts are expected to be similar in several dimensions to Tribal tracts, some perhaps unmeasurable. CEM is also applied to this sample to ensure apples-to-apples comparisons, especially in relation to the size of the tract.

C. *Sample Descriptive Statistics*

Some descriptive statistics for the three samples are provided in Table 1. Means and standardized differences are provided. The standardized difference is a measure of the difference in the distributions across the Tribal and non-Tribal groups. A standardized differences more than 0.25 is viewed as "large," though some researchers prefer a difference of 0.10.

---

[14] This relationship is demonstrated in G.S. Ford, *A Quality Check on Form 477 Data: Errors, Subsidies, and Econometrics*, PHOENIX CENTER POLICY PERSPECTIVE No. 21-05, *supra* n. 4. If it is assumed the overstatement is equal in same-sized census blocks between non-Tribal and Tribal areas, then the errors cancel in the means difference. That is, say broadband is measured as $b_i + e_i(s_i)$ where e_i is an error related to the size of the area s_i. When s_i is the same and the e have a zero mean, then $b_i + e_i(s_i) - (b_k + e_j(s_j)) = b_i - b_j$ if $s_i = s_j$.

[15] S.M. Iacus, G King, and G. Porro, *Causal Inference Without Balance Checking: Coarsened Exact Matching*, POLITICAL ANALYSIS ADVANCE ACCESS (2011).

[16] In a few instances, it was not possible to find a pair of good matches.

Table 1. Descriptive Statistics

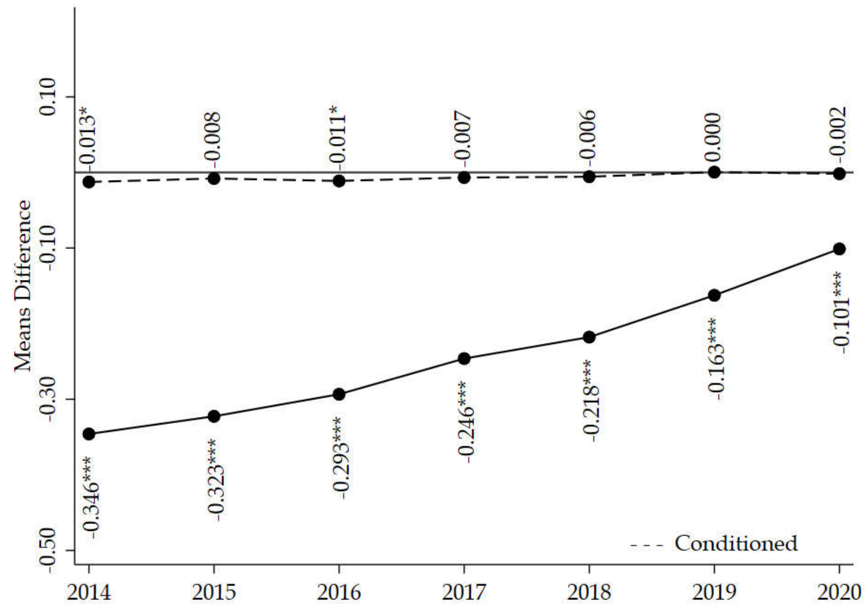
	Full Sample			CEM Sample		
	Tribal	**Non-Tribal**	**St. Diff.**	**Tribal**	**Non-Tribal**	**St. Diff.**
Population	4,104	4,502	0.173	4,084	4,091	0.004
Sq. Miles	219.76	25.63	0.686	161.3	154.4	0.213
Median Income	53,758	68936	0.550	55,662	56,558	0.049
College	0.129	0.189	0.633	0.135	0.136	0.002
Minority	0.144	0.299	0.645	0.128	0.124	0.019
Obs.	723	71,261		568	1,077	
	Within 30 Miles Sample			**CEM Sample**		
Population	4,104	4,597	0.221	4,045	4,087	0.023
Sq. Miles	219.76	26.56	0.705	149.50	126.66	0.117
Median Income	53,758	69,306	0.597	57,328	58,602	0.072
College	0.129	0.193	0.770	0.140	0.143	0.050
Minority	0.144	0.285	0.611	0.108	0.100	0.057
Obs.	723	16,843		446	775	

For the full sample, the standardized differences are quite large for all but population. CEM performs its task by making the differences much smaller and below the 0.25 threshold. Likewise, for tracts within 30 miles of a Tribal area the standardized differences are large, but CEM again makes for comparable groups. Note, however, that some tracts are lost in the matching, indicating that the difficulty in finding non-Tribal tracts that match the unique characteristics of Tribal areas. Still, the samples remain large.

III. Results

To start, I begin by analyzing the entire dataset. In Figure 1, the trends in the Tribal Gap are illustrated for the unmatched sample with and without covariates along with indicators for statistical significance (* 10%, ** 5%, and *** 1%). The unconditioned means of the outcome are listed at the bottom of the figure (A = 1 for Tribal areas). The conditioned means differences are indicated by the dashed line. Conditioning on the covariates shrinks the Tribal Gap since part of the difference is explained by the covariates.

Figure 1. Means Differences, Availability Rate
Full Unmatched Sample



	2014	2015	2016	2017	2018	2019	2020
A=0	0.915	0.918	0.937	0.951	0.959	0.968	0.985
A=1	0.569	0.596	0.644	0.704	0.741	0.806	0.884

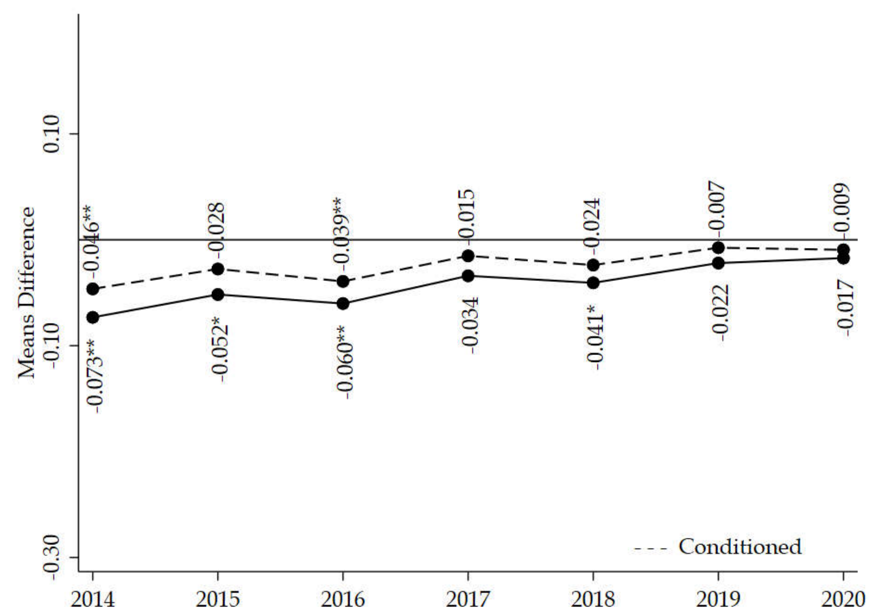
Plainly, the deployment differences between non-Tribal and Tribal areas are shrinking over time. For the unconditioned means difference, the availability rate difference between Tribal and non-Tribal areas was -0.346 in 2014, but the difference had fallen to -0.101 by 2020.[17] A 10-percentage point difference is not small, and warrants attention. All the unconditioned differences in the unmatched sample are statistically significant at the 1% level. When conditioned on the covariates (population, square miles, and median income), the gap is essentially eliminated, suggesting Tribal areas are not treated differently than non-Tribal areas with similar characteristics. In 2020, the Tribal Gap is -0.002, a difference that is not statistically different from zero. Even in 2014, the Tribal Gap is small (-0.013) though statistically different

---

[17] The Tribal Gaps reported in the FCC's *2020 Broadband Deployment Report* are: 2014 (-0.323); 2015 (-0.321); 2016 (-0.288); 2017 (-0.254); and 2017 (-0.221). These closely match those reported in Figure 1. *See In the Matter of Inquiry Concerning Deployment of Advanced Telecommunications Capability to All Americans in a Reasonable and Timely Fashion,* FCC 20-50, 2020 BROADBAND DEPLOYMENT REPORT, 35 FCC Rcd 8986 (rel. April 24, 2020) (available at: https://www.fcc.gov/reports-research/reports/broadband-progress-reports/2020-broadband-deployment-report); *Tribal Broadband: National Strategy and Coordination Framework Needed to Increase Access*, Government Accountability Office GAO-22-104421 (2022) (available at: https://www.gao.gov/products/gao-22-104421).

from zero at the 10% level. A test of whether the Tribal Gap is equal in 2014 and 2020 is rejected at the 1% level for the unconditioned sample and 10% for the conditioned sample.

Figure 2. Means Differences, Availability Rate
Full Matched Sample



A=0	0.681	0.684	0.745	0.784	0.824	0.865	0.933
A=1	0.608	0.632	0.685	0.750	0.784	0.843	0.916

In the matched sample, illustrated in Figure 2, the unconditional differences are much smaller, but the pattern is the same.[18] When conditioned on the covariates, the Tribal Gaps are slightly larger than those for the unmatched sample. Since the matching algorithm includes the regression covariates, the differences between unconditioned and conditioned Tribal Gaps are smaller. The unconditioned means difference falls from -0.073 to -0.017 over the interval, a statistically significant changes at the 5% level. The change between 2014 and 2020 is statistically different from zero at the 10% for the conditioned means. By 2020, the Tribal Gaps of -0.017 for the unconditioned mean and -0.009 for the conditioned mean, both of which are quite small. These results are encouraging, suggesting that broadband availability in Tribal areas is becoming closer or equal to non-Tribal areas over time, and that any broadband gap is largely the result of economic characteristics and not the disparate treatment of Tribal areas.

---

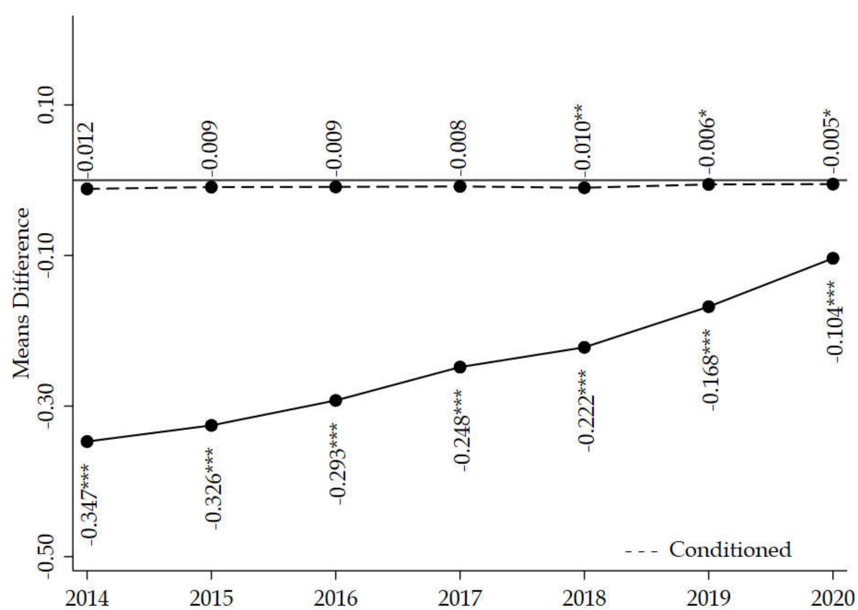
18 The equality of the predicted means in 2014 and 2015 is due to rounding. The means differ by a very small amount.

The influence of matching is observed by comparing Figures 1 and 2. In Figure 1, the conditioned means difference is near zero in all years, which is not the case for Figure 2. This discrepancy reflects the large means differences in the explanatory variables in the unmatched sample and the functional form of the model. When those differences are tamed with matching, the projection between groups has less severe consequences for the predictions.

A. *Within 30 Miles Sample*

Figure 3 illustrates the trend in the availability gap for the sample including only tracts within 30 miles of a Tribal area. The Tribal Gaps are comparable to those from the full sample. The Tribal Gap is closing. For the unconditioned means differences, the Tribal Gap closes from -0.347 in 2014 to -0.104 in 2020. When conditioned on the covariates, the Tribal Gap is essentially zero. A test of whether the Tribal Gap is equal in 2014 and 2020 is rejected at the 1% level for the unconditional means difference. The conditioned means differences are small across the board.

Figure 3. Means Differences, Availability Rate
Within 30 Miles of Tribal Area, Unmatched Sample

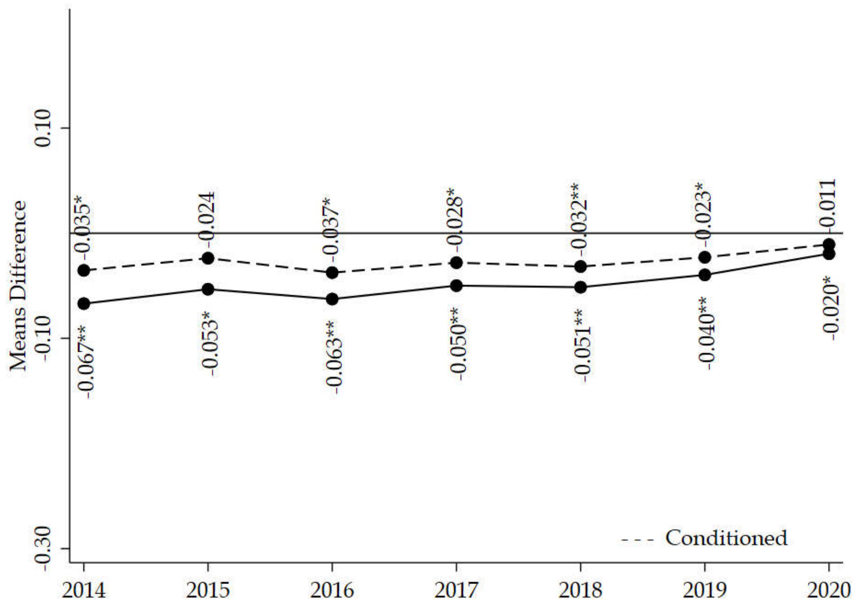


	2014	2015	2016	2017	2018	2019	2020
A=0	0.916	0.921	0.936	0.953	0.963	0.974	0.988
A=1	0.569	0.596	0.644	0.704	0.741	0.806	0.884

The matched sample (shown in Figure 4) also shows the Tribal Gap is shrinking. The unconditioned means difference falls from -0.067 to -0.020 over time. The conditioned means differences are relatively small, though sometimes statistically different from zero at the 10%

level. A test for the equality of the Tribal Gap in 2014 and 2020 is rejected at the 10% level for the unconditional means difference but not the conditioned means difference.

Figure 4. Means Differences, Availability Rate Within 30 Miles of Tribal Area, Matched Sample



	2014	2015	2016	2017	2018	2019	2020
A=0	0.701	0.713	0.766	0.818	0.855	0.888	0.946
A=1	0.634	0.659	0.703	0.768	0.804	0.848	0.972

When conditioned on the covariates, the Tribal Gap is -0.011 in 2020 and not statistically different from zero. Even with the unconditioned means difference, the Tribal Gap is relatively small at -0.020 in 2020, and even smaller when conditioned (-0.011). The relatively small Tribal Gaps in the matched sample suggest that similar tracts near Tribal areas are not much different in broadband availability than are Tribal census tracts. It appears that the observed gaps merely reflect the economic challenges of serving Tribal areas.

IV. Conclusion

In this BULLETIN, I address the issue of the Tribal Gap. This gap is measured as the difference in average broadband availability between Tribal and non-Tribal census tracts. Unmatched and matched sample are used, and a sample of census tracts within 30 miles of a Tribal area are also analyzed with and without matching. In all cases the Tribal Gap is getting closer to zero over time and is by 2020 is near zero when conditioned on covariates or in matched samples. The unconditioned differences, however, shows about a 10-percentage spread in availability in Tribal areas. These results are largely encouraging and suggests efforts to close the Tribal Gap are meeting with some success, though work remains to be done. These results do not imply that broadband is ubiquitous in either Tribal or non-Tribal areas; instead, these results simply demonstrate that the difference in availability between Tribal and non-Tribal areas is shrinking and that this difference is mostly explained by a few demographic characteristics.